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

IDENTIFICATION OF SPECIES-SPECIFIC ALLOSTERIC BINDING SITES IN PHOSPHOFRUCTOKINASE FOR DRUG DESIGN STUDIES

MERVE AYYILDIZ

MASTER'S THESIS

ISTANBUL, JUNE, 2020

IDENTIFICATION OF SPECIES-SPECIFIC ALLOSTERIC BINDING SITES IN PHOSPHOFRUCTOKINASE FOR DRUG DESIGN STUDIES

MERVE AYYILDIZ

MASTER'S THESIS

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

ISTANBUL, JUNE, 2020

DECLARATION OF RESEARCH ETHICS

METHODS OF DISSEMINATION

I, MERVE AYYILDIZ, 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 "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/project will be accessible from everywhere by all means.

MERVE AYYILDIZ

_____

KADIR HAS UNIVERSITY

SCHOOL OF GRADUATE STUDIES

ACCEPTANCE AND APPROVAL

This work entitled **IDENTIFICATION OF SPECIES-SPECIFIC ALLOSTERIC BINDING SITES IN PFK AND ITS USE IN DRUG DESIGN STUDIES** prepared by **MERVE AYYILDIZ** has been judged to be successful at the defense exam held on **----------** and accepted by our jury as **MASTER THESIS**.

APPROVED BY:

Prof. E. Demet Akdoğan (Advisor) _____

Kadir Has University

Asst. Prof. Şebnem Eşsiz _____

Kadir Has University

Asst. Prof. Özge Kürkçüoğlu Levitaş _____

Istanbul Technical University

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

_____

(Title, Name and Surname)

Dean of School of Graduate Studies

DATE OF APPROVAL: (Day/Month/Year)

TABLE OF CONTENTS

IDENTIFICATION OF SPECIES-SPECIFIC ALLOSTERIC BINDING SITES IN PHOSPHOFRUCTOKINASE FOR DRUG DESIGN STUDIES

ABSTRACT

Phosphofructokinase (PFK), one of the most critical enzymes in glycolytic pathway was targeted for use in species-specific drug design studies and its allosteric binding sites in three species, bacteria, parasite and human, were investigated. As allosteric regions are evolutionarily less conserved than active sites, they became powerful target sites for species-specific drug design studies. In accordance, we developed a novel approach which combines well-known computational methods including solvent-mapping, elastic network modeling and sequence/structural alignments.

First, all binding regions were explored by computational solvent mapping. The majority of the regions detected in mapping were located at the interface regions in between monomeric subunits of the receptor. Then, percent frequency shift calculations based on elastic network model were conducted to scan the enzymes in order to determine the areas which are more likely to perturb the global dynamics upon ligand binding. Some of these regions were found to be located near well-reported active and allosteric regions supporting the accuracy of our methodology. Furthermore, sequence and structural differences between human and bacteria/parasite were investigated by sequence and structural alignment methods. Despite high structural similarity, low sequence similarity which increase the specificity between bacteria/parasite and human was observed. As a result, novel allosteric regions in all both bacteria and parasite have been identified. Finally, virtual screening was performed via docking for potential species-specific drug molecules that can be used against *S. aureus*. FDA approved and World-not-FDA approved subsets extracted from ZINC database which contains 1416 and 2922 drug molecules respectively were docked to potential allosteric sites in *S. aureus* PFK. Six potential drug molecules were identified from both subsets which will be later used in *in-vitro* studies for further assessment.

Keywords: Drug Design, Phosphofructokinase, Allosteric Sites, Elastic Network Model, Docking

İLAÇ TASARIM ÇALIŞMALARI İÇİN FOSFOFRUKTOKİNAZ ENZİMİNDE TÜRE ÖZGÜ ALLOSTERİK BAĞLANMA BÖLGELERİNİN BELİRLENMESİ

ÖZET

Glikolitik yolaktaki en kritik enzimlerden biri olan fosfofruktokinaz (PFK), türe özgü ilaç tasarım çalışmalarında kullanılmak üzere hedeflenmiş ve bakteri, parazit ve insan olmak üzere üç türdeki allosterik bağlanma bölgeleri araştırılmıştır. Evrimsel süreçte allosterik bölgeler, proteinlerin aktif bölgesinden daha az korunmuş oldukları için, türe özgü ilaç ilaç tasarım çalışmaları için güçlü hedef bölgeler haline geldiler. Bu doğrultuda, çözücü haritalama, elastik ağ modeli, dizi ve yapı hizalama gibi iyi bilinen yöntemleri birleştiren yeni bir yaklaşım geliştirdik.

İlk olarak, tüm bağlanma bölgeleri hesaplamalı çözücü haritalama yöntemiyle araştırıldı. Bu haritalama sonucunda tespit edilen bölgelerin çoğu, reseptörün monomerik alt birimleri arasındaki arayüz bölgelerinde bulundu. Daha sonra, ligand bağlanmasına bağlı olarak proteinin global dinamiklerinde daha fazla değişikliğe sebep olabilecek bölgelerin belirlenmesi için elastik ağ modelini baz alan yüzde frekans kayması hesaplamaları ile enzimler tarandı. Bu bölgelerin bazılarının, metodolojimizin doğruluğunu destekleyerek iy rapor edilmiş aktif ve allosterik bölgelerin yakınında olduğu bulunmuştur. Ayrıca, sekans ve yapısal hizalama yöntemleri ile insan ve bakteri/parazit arasındaki sekans ve yapısal farklılıklar araştırıldı. Yüksek yapısal benzerliğe rağmen, bakteri/parazit ve insan arasındaki özgüllüğü artıran düşük sekans benzerliği gözlenmiştir. Sonuç olarak, hem bakteri hem de parazitte yeni allosterik bölgeler tanımlanmıştır. Son olarak, *S. aureus'a* karşı kullanılabilecek potansiyel türe özgü ilaç molekülleri için yerleştirme yoluyla sanal tarama gerçekleştirildi. Sırasıyla 1416 ve 2922 ilaç molekülleri içeren ZINC veri tabanından çıkarılan FDA onaylı ve World-not-FDA onaylı alt kümeler, *S. aureus* PFK'daki potansiyel allosterik bölgelere yerleştirildi. Her iki alt gruptan altı potansiyel ilaç molekülü tanımlandı ve bunlar daha sonra *in-vitro* çalışmalarda daha fazla değerlendirme için kullanılacaktır.

Anahtar Sözcükler: İlaç Tasarımı, Fosfofrüktokinaz, Allosterik Bölgeler, Elastik Ağ Modeli, Doking

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere appreciation to my thesis adviser, Prof. Demet AKTEN for her guidance, support and patience. She has thought me the methodology of good research. Without her encouragement, the goal of this thesis would not have been possible. Besides my adviser, I am grateful to Asst. Prof. Şebnem EŞSİZ for her valuable comments and everything she thought me.

I would also like to thank our group members for their help and encouragements. Especially, I am grateful to my friends and colleagues Serkan Çeliker, Fatih Özhelvacı, and Melis Gencel who were always there to help and support me.

Last but not least, I would like to express my extreme gratitude to my parents for supporting me in all circumstances throughout my life.

I would like to acknowledge Kadir Has University for my Master Scholarship.

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

To My Parents

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

Å	Angstrom
α	Alpha
Δ	Delta
ΔG	Gibbs Free Energy
λ	Lambda
λ_i	Eigenvalue of Mode i
$\%s_i$	Percentage of Frequency Shift for Mode i
ADP	Adenosine diphosphate
ATP	Adenosine three phosphate
ASP	Astex Statistical Potential
Cα	Carbon Alpha
CS	Consensus Site
ENM	Elastic Network Model
F6P	fructose 6-phosphate
FDA	US Food and Drug Administration
GA	Genetic Algorithm
GADPH	glycerate 3-phosphate dehydrogenase
GABA	gamma-Aminobutryic acid
GOLD	Genetic Optimization for Ligand Docking
GLY	Glycine
MD	Molecular Dynamics
NAD	Nicotinamide Adenine Dinucleotide

NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
PFK	Phosphofructokinase
PK	Pyruvate Kinase
RMSD	Root Mean Square Deviation
rSASA	Relative Solvent Accessible Surface Area
Vdw	Van der Waals

1. INTRODUCTION

1.1 Glycolysis

Glycolysis is one of the most essential metabolic pathway in all living cells, where glucose is broken down into three-carbon pyruvate via many enzymatic processes (Romano and Conway 1996). Oxygen is not required, thus it is a common pathway in both anaerobic and aerobic organisms and it is the first step of cellular respiration. Also, glycolytic pathway is an essential pathway for the survival of organisms. It may be separated into two main phases. In the first phase, two ATP molecules are used up in order to give phosphate molecules needed. In the second phase, four ATP molecules are produced with one NADH (Barnett 2003; Meyerhof and Junowicz-Kocholaty 1943). Discovered in the 1940s (Campa, McAnulty, and Laurent 1990), the glycolytic pathway consist of ten distinct biochemical reactions catalysed by ten different enzymes of which three are known to be allosteric (Li et al. 2015). As illustrated in Figure 1.1, these enzymes are hexokinase, phosphoglucose isomerase, phosphofructokinase (PFK), aldolase, triphosphate isomerase, glyceraldehyde 3-phosphate dehydrogenase (GAPDH), phosphoglycerate kinase, phosphoglycerate mutase, enolase and pyruvate kinase (PK) where PFK, GADPH and PK are known as allosteric. Phosphofructokinase is one of the most important enzymes in the pathway, since it catalyses a rate-limiting step (Boscá and Corredor 1984).

Glycolytic pathway is a common metabolic process in all living organism; hence, a specified strategy is needed. Species-specific drug design is one of the most important strategies for developing effective drugs for the treatment of life-threatening diseases by targeting the disease-related organism without harming humans. These types of drugs are developed by targeting one of the critical enzymes which is crucial for the survival of the infecting organism.

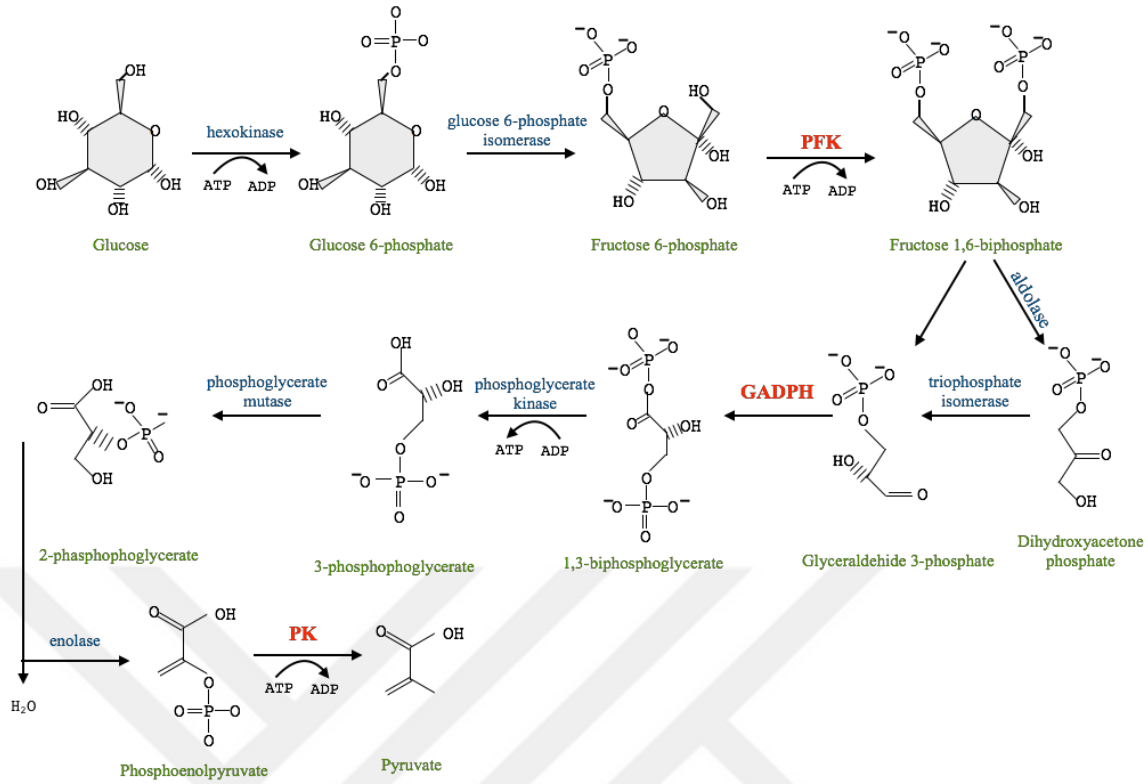


Figure 1.1 Enzymes in glycolytic pathway.

1.2 Species-Specific Allosteric Drugs

Allosteric drugs have crucial roles in affecting protein conformation via binding to a region in the receptor distant from the active site (Jacques Monod, Changeux, and Jacob 1963; Perutz 1989; Daniel E. Koshland and Hamadani 2002). These drug molecules have a remote effect on the active site by inducing a conformational change and consequently increasing (positive modulation) or decreasing (negative modulation) the binding properties of small molecules at the active site (Figure 1.2) (May and Christopoulos 2003). The term allosteric effect starts when Niels Bohr explained the conformational change in hemoglobin as a result of binding of two different molecules to the protein (Bohr, Hasselbalch, and Krogh 1904). Two allosteric models which are called MWC and sequential KNF model were proposed for the first time in 1960s (MONOD and JACOB 1961; Jacque Monod, Wyman, and Changeux 1965; D. E. Koshland, Nemethy, and Filmer 1966). Although proteins are often considered as allosteric or non-allosteric, conformation of a protein may be altered as a result of mutations or binding of a drug

molecule at a specific site. In this respect, all proteins can be considered as potentially allosteric (Gunasekaran, Ma, and Nussinov 2004).

Although most current drugs are developed to target orthosteric site, recent studies have focused on the design of allosteric inhibitors which have significant advantages over orthosteric inhibitors. Since active sites of the same biological targets among species are more conserved than any other sites, allosteric inhibitors have high specificity with less side effects (Conn, Christopoulos, and Lindsley 2009). The residues in the active site have evolutionary pressure to be conserved and for enzymes to work in the same way as before for the continuity of the organism's life. This pressure is weaker for allosteric sites, therefore, residues encounter more mutations among species (Christopoulos et al. 2004) Consequently, differences in the amino acid sequence and structure level are significantly high in the regions where allosteric drugs bind. Flurazepam (benzodiazepine derivative) is one of the allosteric drugs which binds to GABA-A receptor and cause conformational change in the orthosteric site. It is used to treat insomnia disorder (Möhler, Fritschy, & Rudolph, 2002; Spurny et al. 2012). Another allosteric drug is Maraviroc which was approved by the FDA in 2007 and used as an allosteric modulator for HIV treatment (Abel et al. 2008).

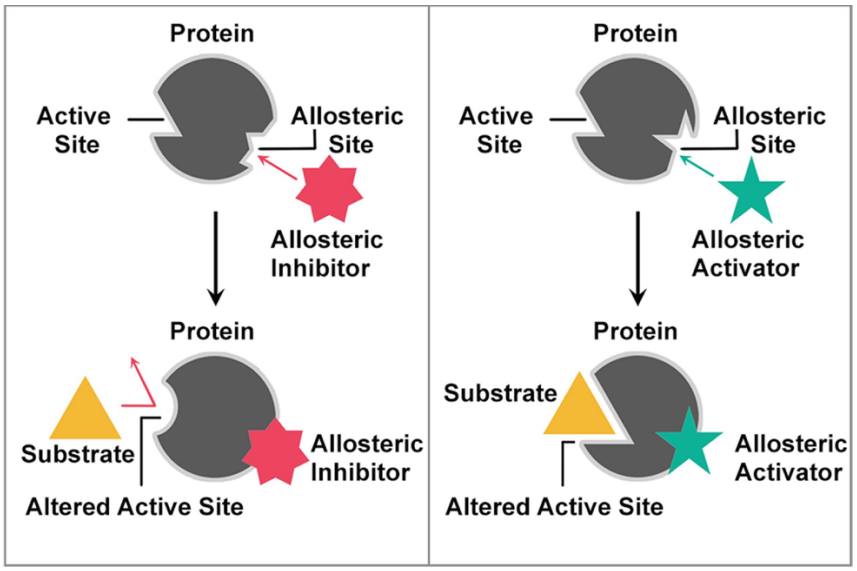


Figure 1.2 Allosteric drug mechanism (Cheng et al. 2019)

1.3 Structure and Function of Phosphofructokinase

Phosphofructokinase is one of the three allosteric enzymes which catalyses the phosphorylation of fructose-6-phosphate (F6P) to fructose-1,6- bisphosphate in glycolytic pathway (Figure 1.1). PFK could allosterically regulate glycolysis to answer the requirements of energy in the cell via increasing or decreasing glucose ratio in the cell. In this thesis, three species of PFK have been studied including, *Homo sapiens* PFK *(h*PFK), *Trypanasoma brucei* PFK *(Tb*PFK*) and Staphylococcus aureus* PFK *(Sa*PFK). In a biological cell, PFK shows activity in tetrameric form but human tetrameric form differs than bacterial/parasitic tetramer. The enzyme shares a common scaffold structure in all three species, that is, the homotetramer structures consist of two dimers. Each dimer has a catalytic region to which ATP and F6P (substrate) are bound and an effector region for allosteric binding (Shirakihara and Evans 1988). Apart from these, the gene that synthesize the human phosphofructokinase has undergone gene duplication during evolutionary process. For this reason, each *h*PFK is synthesized twice as much as other species. Accordingly, one monomer of *h*PFK corresponds to two monomers in bacterial and parasitic PFK. (Brüser et al. 2012). Figure 1.3 shows tetramer structures of bacterial and parasitic PFK, as well as both dimer and tetramer structures for human PFK. It should be noted that dimer structure of the human is actually composed of four subunits (Kloos et al. 2015). In this thesis, it is aimed to determine species-specific allosteric regions, so that human and bacteria/parasite should be compared reasonably. As a result, the studies were done with the tetrameric PFK of bacteria/parasite and corresponding dimer structure of human PFK.

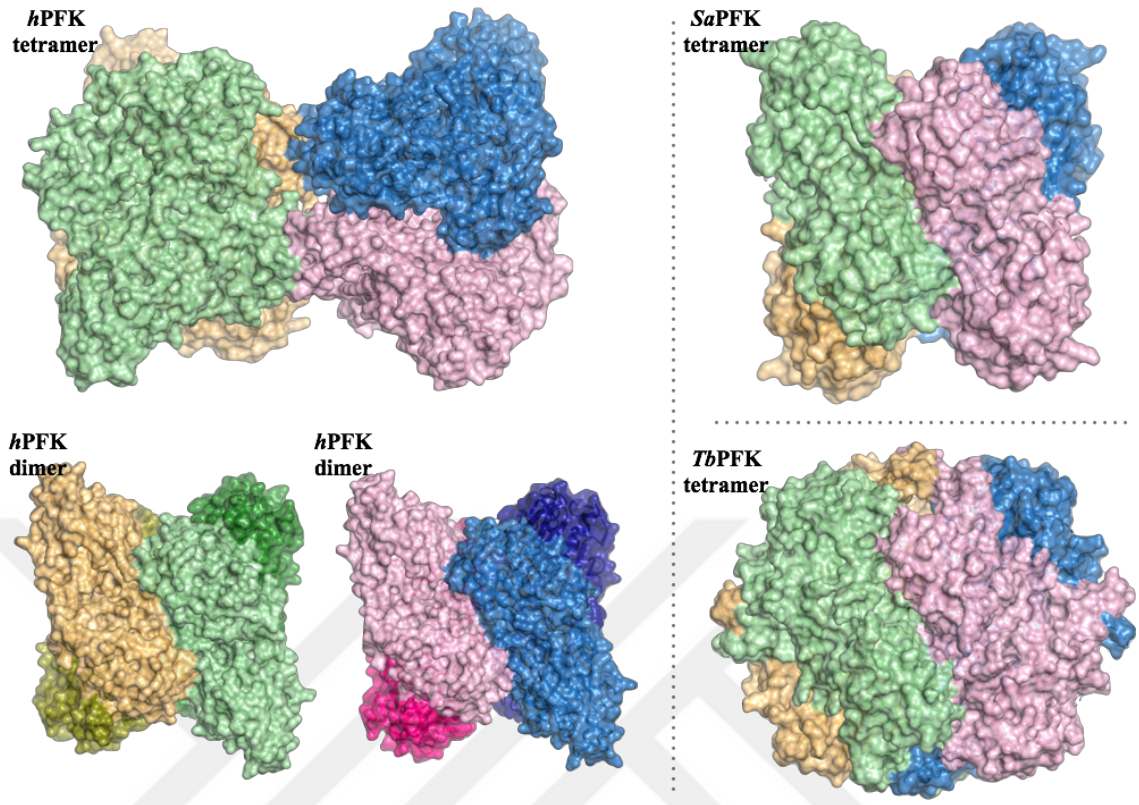


Figure 1.3 Tetrameric structures of *Sa*PFK (1288 residues) and *Tb*PFK (1948 residues) and *h*PFK (3136 residues).

1.4 Allosteric Regulation of Phosphofructokinase

PFK catalyses the rate limiting step which is the slowest step of glycolysis. Regulation of PFK will eventually regulate the glycolysis. PFK is an enzyme with two conformational states called R (active) and T (inactive) that are in equilibrium. ATP is the substrate but also the inhibitor of PFK, so PFK has two binding sites for ATP. If the ATP is bound to the inhibitory site, structure will be found in the inactive state, and if the ATP is bound to the ligand binding site then the structure will be in the active state. Additional to ATP, there is another substrate of PFK which is F6P and it also binds to the R-state. Energy (ATP) is a result of glycolysis, and when ATP production is too high, ATP will be an inhibitor of phosphofructokinase and this will stop the production of ATP. Citrate is also an allosteric inhibitor as well as ATP. ADP is an allosteric activator. General scheme about allosteric regulation is illustrated as in Figure 1.4.

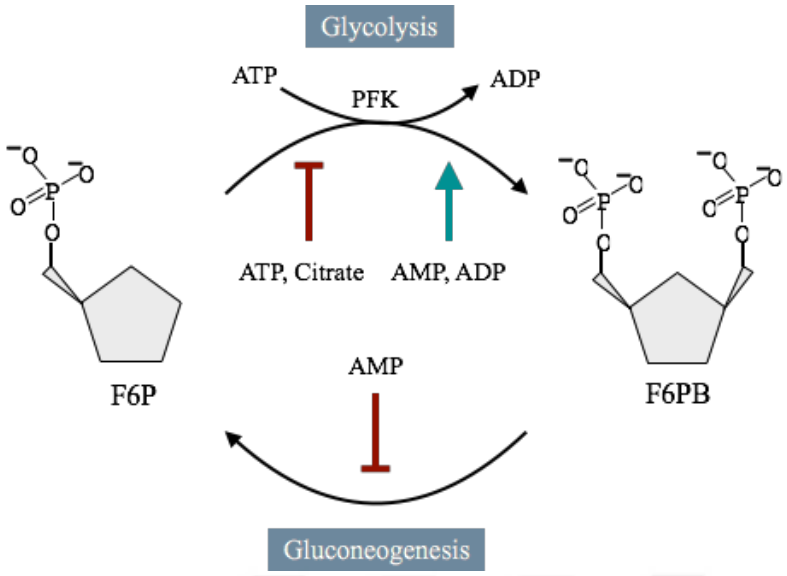


Figure 1.4 Known allosteric modulators of PFK.

2. MATERIAL AND METHODS

2.1 System Preparation

In this thesis, phosphofructokinase enzymes of *H. sapiens*, *T. brucei* and *S. aureus* have been studied. The crystal structures of these enzymes were obtained from Protein Data Bank (Martinez-Oyanedel et al. 2007; McNae et al. 2009; Webb et al. 2015; Kloos et al. 2015; Tian et al. 2018). The table below lists all the crystal structures that belong to *Sa*PFK, *Tb*PFK, *h*PFK in PDB. The structures reported as tetrameric form which displays the biological activity were selected for subsequent studies. As shown in Table 2.1, there exist more than one X-ray crystal structure for tetrameric form of *H. sapiens* and *T. brucei*, and there is no tetrameric structure of *S. aureus* reported so far. However, it is necessary to determine one structure for each species that can be representative among the existing tetramers.

Table 2.1 PDB IDs of PFK in three species.

Phosphofructokinase	***Homo sapiens***	***Trypanasoma brucei* (Parasite)**	***Staphylococcus aureus* (Bacteria)**
Monomer	N/A	N/A	2JG5
Dimer	4OMT	N/A	5XOE, 5XZ6, 5XZ7, 5XZ8, 5XZ9, 5XZA
Tetramer	4RH3, 4U1R, 4WL0, 4XZ2, 4XYJ, 4XYK	3F5M, 2HIG	N/A

Prior to a decision making between the existing structures for each species, it should be noted that human and bacterial PFK structures are more challenging than the parasite structure and they require modifications. Since human structure doubles compared to any other species as a result of gene duplication during evolution, it is quite problematic to make structural and/or sequence comparison between them. Human tetrameric structure for PFK consists of dimer of dimers and each monomer of dimers is identical. Besides, this homotetrameric form is also valid for bacterial/parasitic PFK. In addition, it should be noted that one dimer of human PFK corresponds to one tetramer in *Sa*PFK and *Tb*PFK in terms of size and shape. Apart from this challenge, six crystal structures have been released in PDB. Human PFK has six tetramer entries in protein data bank, one is in apo form (PDB id: 4WL0) and the remaining five are ligand-bound forms. All six entries are active forms of *h*PFK. Besides, structural similarities of these structures with each other in order to select a representative were determined by structural alignment (Table 1.2). Considering that RMSD values of the structures are below one Angstrom (Å) except the structure with PDB id: 4XYK, any of them could be representative. The structure with PDB id: 4RH3 with a resolution of 3.02 Å, which has missing residues at the beginning and at the end of the sequence mostly instead of somewhere in the middle of sequence was chosen for human PFK.

Table 2.2 Structural alignment results of *h*PFK tetramers with each other.

RMSD	4RH3	4U1R	4WL0	4XYJ	4XYK	4XZ2
4RH3	**0**					
4U1R	0.273	**0**				
4WL0	0.430	0.340	**0**			
4XYJ	2.668	2.539	2.422	**0**		
4XYK	0.904	0.735	0.652	2.226	**0**	
4XZ2	0.356	0.323	0.306	2.447	0.681	**0**

For *Sa*PFK, there is no released tetrameric structure in the PDB. It is known that PFK must be in tetrameric form in order to show biological activity in a cell. In the light of the paper related to the structure with PDB id: 5XZ7, the conversion was made from dimer to tetramer. (Tian et al. 2018). Two different strategies were followed for that conversion. First, asymmetric units of the released dimer structure were generated with a distance of 4 Å using PyMOL. This method was used as the proposed method of the paper. Secondly,

a tetramer structure of another bacterium released in 1981 with 2.4 Å resolution, of Bacillus stearothermophilus PFK (*Bs*PFK, PDB id: 4PFK) was chosen for structural alignment with the dimer structure of the *Sa*PFK. *Sa*PFK dimer structure was first aligned to one half of the *Bs*PFK and then aligned to the other half. Then, these aligned dimer structures were merged and *Sa*PFK tetramer structure was obtained. This second method followed for tetrameric *Sa*PFK is summarized in the Figure 2.1. Then, in order to compare the tetrameric *Sa*PFK structures obtained from both methods, structural alignment was performed and it was seen that RMSD value was lower than 1 Å. As a result, the tetrameric *Sa*PFK obtained from the first method was chosen as the representative for further studies.

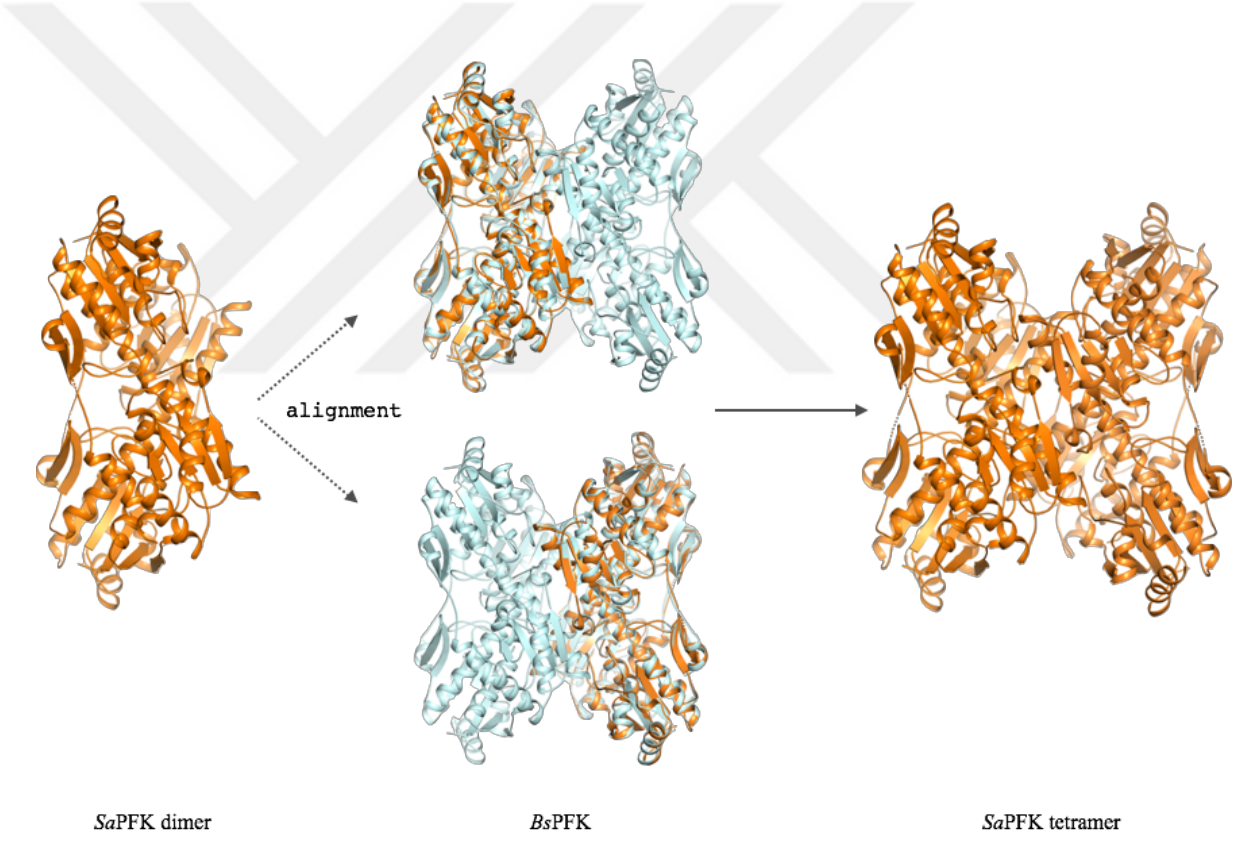


Figure 2.1 Dimer to tetramer conversion. Alignment with one of the close-family structure. PDB id: 4PFK was used for PDB id: 5XZ7.

Two tetrameric structures were released in PDB for parasitic PFK. One of these structures is in holo (Pdb id: 5XZ7), and the other is in apo (PDB id: 2HIG) form. These structures were aligned with a root mean standard deviation (RMSD) value of 0.5 Å. Accordingly, structures were quite similar to each other and any of them could be selected as representative structure. *Tb*PFK (PDB id: 3F5M) was selected for subsequent studies, as it is ligand-bound and the *Sa*PFK and *h*PFK are also ligand-bound structures. As a result, structures of *Sa*PFK, *Tb*PFK and *h*PFK that were selected as representatives of PFK for each species were shown as in Figure 2.2.

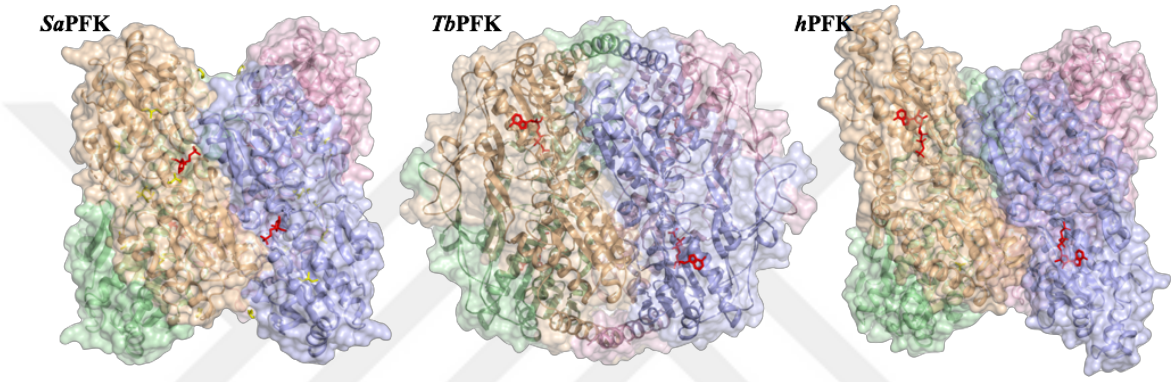


Figure 2.2 Crystal structures of *Sa*PFK, *Tb*PFK and *h*PFK respectively. Native ligands are shown with red sticks (F6P in *Sa*PFK and ATP in *Tb*PFK/*h*PFK). Wheat, green, pink and blue colors represent A, B, C and D chains in bacterial/parasitic PFKs and corresponding four units in human PFK.

2.2 Sequence and Structural Alignment

Sequence alignment is the primary and essential method to determine proteins' evolutionary relatedness with other proteins in phylogenetic analysis. If the similarity is high between two same proteins from different species, then it shows evidence for common ancestor. Besides, high sequence similarity/identity in a specific region of the protein between species will provide us the importance of that area for the organism's survival. It also means that crucial regions such as active site of the protein will be conserved between species during evolution, and an unknown active site for one species may be predicted depending on its close homolog. In this thesis, Needleman-Wunsch alignment algorithm (Needleman and Wunsch 1970) was used for the pairwise global alignment of human and bacterial/parasitic species. EMBOSS-Needle (Madeira et al. 2019) webserver tool which uses this algorithm for alignment was used with the default

parameters. Some of these parameters are; Blosum62 as scoring matrix, 10 as gap open penalty, 0.5 as gap extend penalty. Aim of using sequence similarity was to find dissimilar regions between human and bacterium/parasite which can be proposed as target sites for species-specific drug design studies since only targeting infecting organism without harming human should be considered.

Besides, structural similarity was also determined to identify differences between species. In this thesis, we used *super* module of PyMOL graphics visualization tool (Schrödinger 2015). *Super* module superposes two structures irrespective of its sequence similarity. It minimizes RMSD that measures distance between the backbone atoms between two aligned structures via several refinement cycles.

As shown in Figure 1.3, tetrameric *h*PFK is twice the size of tetrameric *Sa*PFK and *Tb*PFK. Thus, only half of the *h*PFK (dimeric structure) is structurally aligned with the tetrameric *Sa*PFK/*Tb*PFK. Finally, structures for bacterial and parasitic PFK were coloured based on sequence and structural similarity with respect to human PFK to be able to identify differences visually.

2.3 Computational Solvent Mapping (CS-Mapping)

Computational solvent-mapping is a useful tool to identify possible binding pockets on the protein surface upon docking drug-like organic molecules. It is known that binding pockets consist of regions which are known as hot spots, and have huge impact on associated binding energies (Hajduk, Huth, and Fesik 2005). There are various computational tools to identify binding sites that could also be used for allosteric binding sites identification. These tools can be summarized in three subtitles according to the method used; first is *geometry-based pocket detection* which tries to detect cavities according to the surface area where a probe molecule rolls along. The areas between probe and receptor are considered as binding cavity. POCASA (Yu et al. 2009) can be given as an example. Second type of tool is *knowledge-based pocket detection* which tries to find pockets according to the evolutionary and structural similarity with other receptor via scanning large libraries. 3DLigandSite can be given as an example (Wass, Kelley, and Sternberg 2010). Third type is *energy-based pocket detection*. In this method, a small

molecule will be docked to the receptor and analysed. FTMap is a widely-used energy-based method used in this thesis. It is a mapping algorithm with a web server that predicts hot spots based on NMR or X-ray structures. (Brenke et al. 2009; Kozakov et al. 2015). It was shown that FTMap (https://ftmap.bu.edu/) generally shows good agreement with experimental methods (Wakefield et al. 2019). FTMap places 16 small organic molecules (*isopropanol, acetaldehyde, phenol, benzaldehyde, urea, dimethyl ether, acetonitrile, ethane, acetamide, benzene, methylamine, cyclohexane, ethanol, N, N-dimethylformamide, isobutanol and acetone*) which are varying in chemical and structural properties and called as probes. At the beginning of the process, these probes are placed by rigid body docking, and the most favourable positions for each are searched. For each locations, probes are rotated 500 times, thus, creates billions of probe positions. To accelerate this calculation, FTMap uses Fast Fourier Transform (FFT) correlation. For each probe, 2000 poses with the lowest energy level were kept and energy minimization was performed using CHARMM (Brooks et al. 1983) energy function. Besides, terms as van der Waals, or electrostatics interaction were calculated for energy levels of probes. Each probe type was clustered once, and ranked according to their average energies. Then the clusters within 4 Å apart to each other were clustered to obtain consensus sites. In addition, a region was considered as druggable more than one CS were located close to each other (4 Å) (Figure 2.3). For each mapped protein structure, FTMap gives consensus regions by ranking those with the highest number of probe clusters.

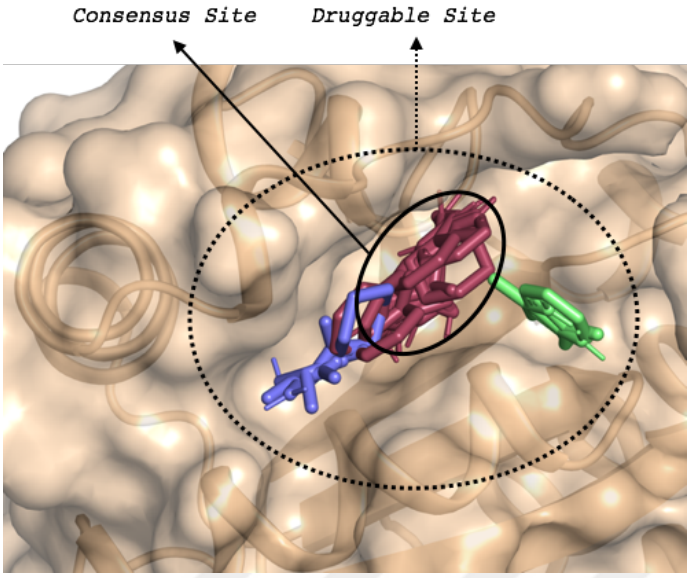


Figure 2.3 An example for FTMap result. Pink, green, blue colored probes represent three different consensus site. Since they are located close to each other, that area is called as druggable site.

FTMap web-server was used to map the X-ray structures of PFK from *H. sapiens* and target organisms which are *T. brucei* and *S. aureus*. In order to show all possible binding regions, crystal structures have been processed by FTMap server as homotetramer structure, and also as single chains. However, each monomeric (domain) *h*PFK, corresponds to the dimeric *Sa*PFK/*Tb*PFK. Thus, when one chain of *h*PFK was used for solvent mapping, two chains of *Sa*PFK/*Tb*PFK was necessarily mapped for compatibility (Figure 2.4). Domain separation was done according to the information provided by PDB. To map as single chain and single domain can turn inaccessible regions to accessible. Therefore, probe molecules can be clustered in these inaccessible regions. In tetrameric structures, these molecules cannot bind and thus they need to be eliminated. In addition, some probes from chain-by-chain mapping were eliminated as well if they overlap with the probes from overall-mapping.

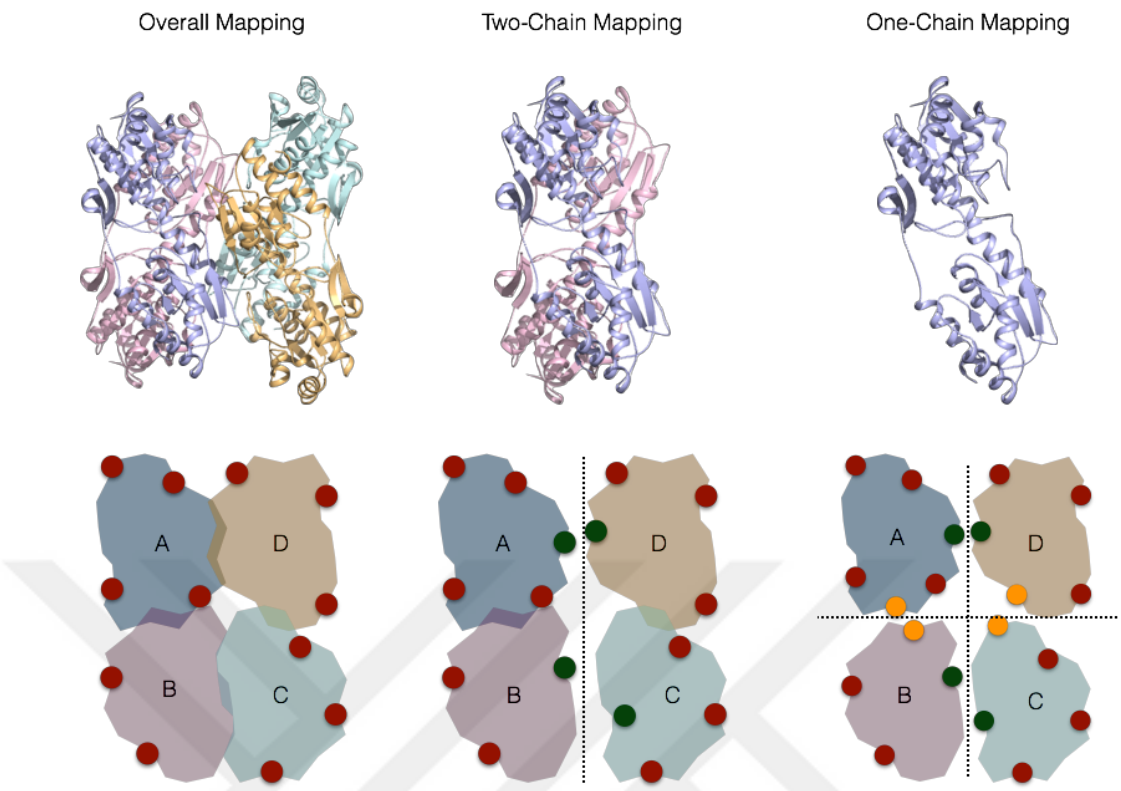


Figure 2.4 Mapping strategy using FTMap. Overall mapping, Two-chain mapping and one-chain mapping was done according to get as many as possible binding sites.

2.4 ENM-Based Residue Scanning

Elastic network model (ENM) is a powerful theoretical approach that is used to predict the global dynamics of biomolecular structures. It is used to establish the relationship between structure and the functional mechanism (Atilgan et al. 2001; Doruker, Atilgan, and Bahar 2000). In this model, the protein was represented as a collection of beads connected by Hookean springs corresponding to a collection of atoms connected by fluctuating bonds. For simplification, only backbone α-Carbon (Cα) atoms were taken into consideration. Furthermore, the springs connected the atoms only if they were closer than a predefined cut-off distance of 15 Å in the native structure. In this study, we used a residue-based scanning method that was developed based on standard ENM (Kurkcuoglu et al. 2015). In this new approach, each residue represented by its backbone Cα as a single node was redefined such that side-chain heavy atoms will be included as extra nodes. It

was proposed that these additions will mimic the presence of a bound ligand interacting with that residue. In this method, each residue was considered as a single node. The effect of the extra nodes was quantified by the change in the *i*th collective mode's eigenvalue upon adding the extra nodes to the selected residue.

$$\%shift\ for\ mode\ i = \frac{\lambda i(modified) - \lambda i(original)}{\lambda i(original)}(x100)$$

The percentage shift for each residue was determined as an average over the 20 slowest modes for PFK. The regions with positive values were investigated further for potential allostery, as they have the largest effect on the global dynamics of the receptor.

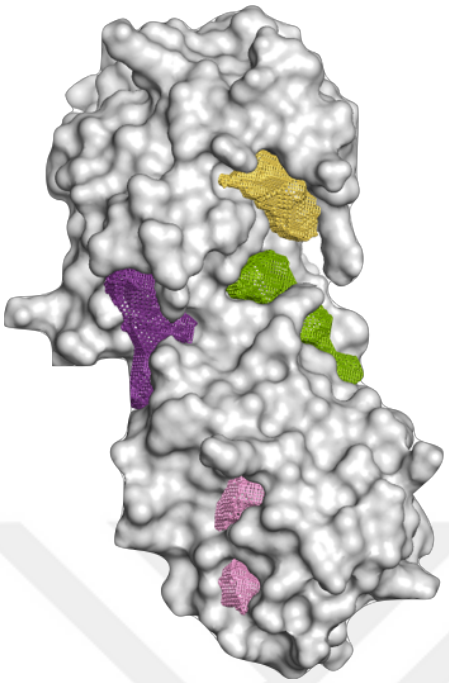
2.5 Determination of Interface Regions using Relative Solvent Accessible Surface Area (rSASA)

Interface regions are known as to be less conserved between species compared to the active site of proteins. Thus, it was used as target regions for species-specific drug design studies. In this study, interface regions were determined based on relative solvent accessibility surface area (rSASA). rSASA can be calculated with normalization which is the division of residue's accessible surface area (ASA) in the protein with maximum accessible surface area (ASA) of that residue (Rose et al. 1985). To get rSASA values for each residues, ASA values are needed and the way that is followed to get it is to place X residue (any residue to be checked) between Gly-X-Gly tripeptide and then evaluate surface area around X (Tien et al. 2013). Division of surface and interface regions were defined with a 25% rSASA threshold. The residues having rSASA less than this threshold were accepted to be part of the interface region. Otherwise, they are accepted to be at the surface.

In the literature, there exist several methods to identify interface regions, but in this study a *tcl* script developed in-house was used. Interface and surface regions were determined for bacteria (PDB id: 5XZ7), parasite (PDB id: 3F5M) and human (PDB id: 4RH3 and compared/merged with solvent mapping and ENM results. Our aim was to demonstrate whether consensus sites would be located close to interface. This would support the likelihood of druggable sites to be near interface regions (Clackson and Wells, 1995).

2.6 Supportive Methods

DogSiteScorer. To evaluate druggability of a protein is crucial for target-based studies, in view of the fact that the identification of small molecule as a lead compound approximately has 60% failure (Brown and Superti-Furga 2003). Therefore, it is important to support findings with other methods/tools. In this study, we aimed to propose possible binding areas that will be inhibitory for targeted species (bacteria/parasite) without harming human. For that purpose, a webserver (https://proteins.plus) where DoGSiteScorer is implemented (Volkamer, Kuhn, Rippmann, et al. 2012) was used. DoGSiteScorer is a grid-based pocket detection tool which predicts binding site and evaluates protein structure in terms of its druggability (Volkamer et al. 2010). First, it detects pockets on the protein surface and then it makes its calculation based on size, shape, and physico-chemical properties of pockets (Volkamer, Kuhn, Rippmann, et al. 2012). It makes use of a supervised machine learning method which is a support vector machine. DoGSite ranks the prediction according to the Drug Score which vary between 0 and 1. A site with a score value closer to 1 is more likely to become druggable site. Besides, it gives druggability score and assumes results higher than 0.5 as possible allosteric site (Volkamer, Kuhn, Grombacher, et al. 2012). Furthermore, it has user-friendly interface, and was incorporated as a plugin inside PyMOL. In this study, DoGSiteScorer were used via Proteins*Plus* web portal (Fährrolfes et al. 2017). It analyzes structure of macromolecules, with the aim of for users to access and process various structure-based tools.



	Name	Volume Å³	Sur-face Å²	Drug Score
	P_0	527.87	679.23	0.71
	P_1	407.68	585.85	0.62
	P_2	361.54	567.29	0.77
	P_3	349.44	423.13	0.69

Figure 2.5 An example of DoGSiteScorer results. Possible druggable sites are shown as spheres on the left side, while corresponding values of each spheres are shown on the right side.

AlloSigma. AlloSigma is a web server that predicts allosteric free energy on each residue of protein as a result of ligand binding, mutation or combination of these two cases (Guernera, Tan, Zheng et al. 2017). AlloSigma tool investigates energetics of allosteric communication and in this respect, it differs from other webservers. It follows three basic steps; first, normal modes of C-alpha atoms of a protein are used to obtain the dynamics of ligand-bound and ligand-unbound structures. Then, the elastic energy which is related with the local structural changes are assessed for each residue. Lastly, an average free energy taken over all nearby residues' free energy is determined for a specific region. Results are evaluated according to following information; blue color shows increase of dynamics which means destabilization (positive ΔG values), while red color indicates decrease of dynamics, means stabilization (negative ΔG values) (Figure 2.6).

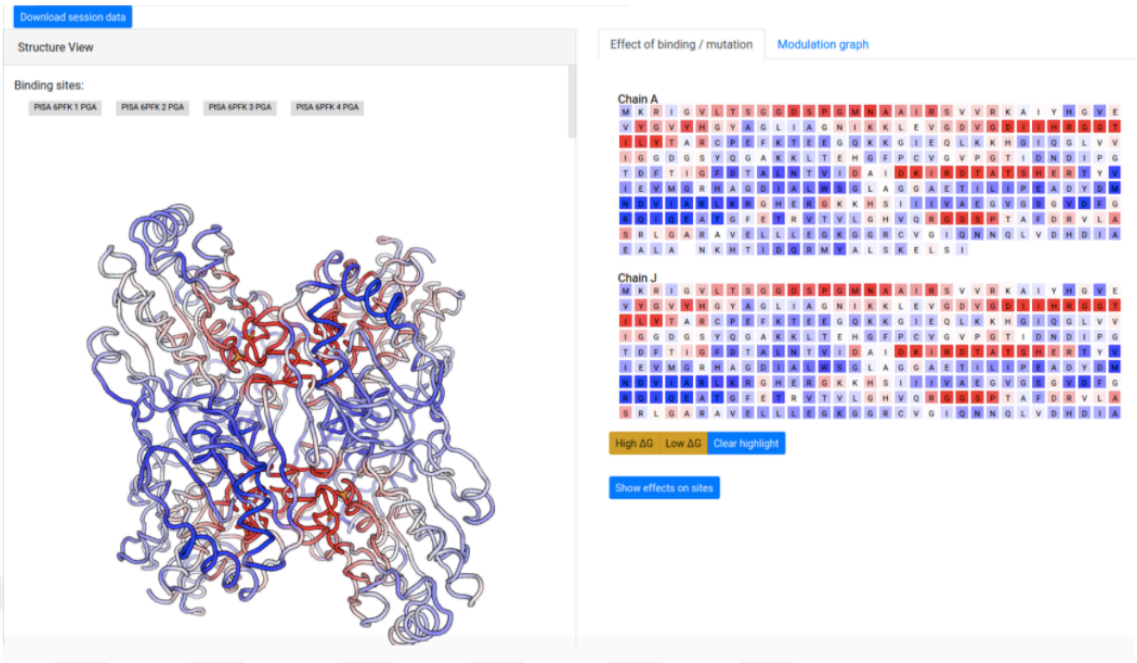


Figure 2.6 An example result page of AlloSigma. PGA binding in PFK (Guernera, Tan, Zheng et al. 2017. To illustrate the energy values of each residue, colors are used. Blue and red shows increase (destabilization) and decrease (stabilization) of conformational change, respectively. White color indicates unaffected residues upon effector binding.

2.7 Virtual Screening through Docking Studies

Docking programs seek to predict the binding affinity of a small molecule called *ligand* to a target receptor. The main purpose of docking is to accurate prediction of most favourable binding modes of ligand to receptor. Docking consists of the following steps; approximate orientation of a ligand to defined binding site, modifying the position and checking maximum favourable interactions which is called as pose evaluation by using a scoring function. AutoDock (Goodsell and Olson 1990), GOLD (Jones et al. 1997), and Glide (Friesner et al. 2004) can be given as the most popular docking software tools. One of the most common application of docking is virtual screening for *hit* identification. It automatically docks a library that contain an extensive amount of small molecules to a drug target and search for the best pose of the molecule (Rester 2008). The best pose with the highest binding affinity is important for giving insight to structure-activity relationship and for further analysis with molecular dynamics simulations. In this study, we used GOLD (Genetic Optimization for Ligand Docking) software for virtual screening. It is well-known and proven as outperformed than other programs (Nagarajan

et al. 2012) . It uses genetic algorithm (GA) based on Darwin's evolutionary theory and specifically theory of natural selection (Holland 1975). GAs have emerged as highly effective methods to uncover all possible ligand binding models (Mitchell 1996). As scoring function GOLD uses *GoldScore, ChemScore* and *ChemPLP* and *ASP (the Astex Statistical Potential)*. Goldscore is based on calculating protein-ligand and intra-ligand van der Waals (vdW) energies, protein-ligand hydrogen bond energies and ligand torsional strain energies. ChemScore is an estimation that utilizes the binding affinities calculated from 82 experimental protein-ligand interactions. It calculates the total energy change between ligand and protein.

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hbond} + \Delta G_{metal} + \Delta G_{lipo} + \Delta G_{rot}$$

Unlike other three scoring functions, ASP considers information about the frequency of interaction between ligand and protein atoms by analyzing them over existing protein-ligand structures. In this respect, it is a knowledge-based scoring function. In this thesis, we used ChemPLP which incorporates the ligand flexibility and hydrophobic interactions similar to other scoring functions, but gives faster results. Additionally, ChemPLP is preferred and suggested by GOLD official website "www.ccdc.cam.ac.uk" as the most accurate scoring function for virtual screening. During docking studies, generally default parameters were used in the program. First of all, a grid box that includes residues 10 Å away from the given coordinates determined for the ligand binding region. The selection that restricts atom selection to solvent-accessible surface was unselected. GOLD optimizes fitness score using genetic algorithm; in here default parameters used as well (GA runs – 10, GA search option- slow).

In this study, by virtual screening method, drug repositioning has been performed for possible druggable regions proposed by previous methods. Drug repositioning is readily performed to estimate the binding affinity via docking an FDA approved drug molecule used for a specific disease and to repurpose those molecules with high affinity to *in-vitro* studies. The biggest advantage of drug repositioning is to use a drug molecule which has already passed all kinds of tests that have high cost for approval. Molecules for drug repositioning in this project were obtained from ZINC15 (https://zinc15.docking.org) free database (Irwin et al. 2012). ZINC15 database has more than 35 million purchasable

molecules, however, here only 1416 FDA (U.S. Food and Drug Administration) approved and 2922 World-not-FDA (approved by different institutions around the world but not FDA) molecules were separately used for virtual screening. It is important that molecules should preferentially bind to bacterial PFK, but not to human PFK since we target bacteria without harming human.

First, the proposed regions have been docked with two databanks as FDA (1416 molecules) and World-not-FDA (2922 molecules). Results were ranked according to the bacteria based on their score value, from high to low and the molecules in the top 2% will be selected. This 2% number is tentative and may be changed to 5% or %10 depending on the number of molecules left. Finally, common molecules with high scores for all designated regions were searched and their corresponding scores in *h*PFK were checked. The remaining molecules will be determined for further steps, such as MD simulations and *in-vitro* studies. To sum all up, table 2.3 gives all methods, tools and webservers that were used in this thesis.

Table 2.3 Summary of methods used.

Name of Method	Summary	Website
Computational Solvent Mapping (FTMAP)	Energy-based method to identify possible druggable site over a receptor.	https://ftmap.bu.edu/login.php (Free)
ENM-based residue scanning	Side-chain of residues included as extra nodes to mimic ligand and then determine change of proteins dynamics.	In-house written source code
Interface Region Determination	Ligand is likely to change dynamic of a protein if it is bound to interface regions.	In-house written source code
Sequence Alignment (EMBOSS-Needle)	Pairwise alignment to reveal differences of bacteria/parasite from human.	https://www.ebi.ac.uk/Tools/psa/emboss_needle/ (Free)
Structural Alignment (Pymol-Super Module)	Superposition of two species to reveal structural differences.	https://pymol.org/2/ (Free)
Supportive Method (DoGSiteScorer)	Commonly-used grid-based method to determine possible druggable regions.	https://proteins.plus (Free)
Supportive Method (AlloSigma)	İnvestigates energetics of allosteric communication upon ligand binding.	https://allosigma.bii.a-star.edu.sg (Free)
Virtual Screening (GOLD)	Drug repositioning of approved drug molecules for proposed druggable sites.	https://www.ccdc.cam.ac.uk/solutions/csd-discovery/Components/Gold/ (Licensed)

3. RESULTS AND DISCUSSION

In this thesis, several methods were employed and combined to identify several potential allosteric sites in human, bacterial and parasitic PFK. As these allosteric sites in parasitic PFK and bacterial PFK were distinctive from those in human PFK, the specificity of these newly proposed binding areas was suggested to be potential druggable sites for species-specific drug design.

3.1 Solvent-Mapping and Interface Region Analysis

Consensus sites (CS sites) of three species were concluded via mapping the structures in three ways. Chain-by-chain and overall tetrameric structure mapping were employed in order to reach all possible binding sites for small molecules to bind. Additionally, domains (chain AB and CD in the Table 3.1) were also given to the FTMap server, because of the fact that human PFK exists as dimer and each chain corresponds to bacteria/parasite's two chains. As given in the second column of Table 3.1, CS that are located in the inaccessible regions in tetrameric form and overlapping molecules were eliminated (See Method 2.3) Then, second elimination was done according to the probable significant positive changes observed in the protein dynamics, i.e. frequency shift. Last columns of Table 3.1, state that approximately 50% of the total CSs were able to make significant changes (frequency shift >25%). In addition, Table 3.2 shows all druggable sites which consist of at least two CS. Druggable sites are shown in three ways; those who were eliminated by displaying 25% frequency shift were shown with two stars, those who were eliminated by displaying 50% frequency shift were shown with a single star and those who are determined for further studies by passing 50% frequency shift are shown without star. In addition, druggable sites that has the CS that were passing 50% frequency shift were also shown separately in Table 3.3. It is also important to note that bacteria and parasites have more populated and fewer druggable sites than humans. Top druggable sites of three species on the list were given in bold character. Among three

species, *S. aureus* exhibited the highest amount of consensus site in the top druggable sites with seven CS and six CS in the second top druggable site, while the lowest amount is two CS in a druggable site. By using these two methods together, it is ensured that the possible binding sites are considered only if they are above 50% frequency shift.

Table 3.1 Total number of clusters for PFK in three species.

Species	**Total Number of Clusters / Non-Overlapping Solvent-Accessible Clusters / After ENM filtering (frequency shift > 25% / 50%)**							
	Tetramer	Chain A	Chain B	Chain C	Chain D	Chain AB	Chain CD	TOTAL
Human	13 13 13 / 13	11 9 7 / 3	12 11 9 / 5	13 12 9 / 5	9 8 8 / 5	10 2 2 / 2	11 2 2 / 2	79 66 50 / 35
S. aureus	17 17 17 / 17	12 9 5 / 4	11 8 4 / 3	11 8 4 / 3	11 8 5 / 4	11 4 4 / 4	12 4 4 / 4	85 58 43 / 39
T. brucei	13 13 13 / 8	10 10 8 / 7	11 10 8 / 3	9 9 6 / 3	10 10 7 / 3	13 4 4 / 3	12 5 5 / 5	78 61 51 / 32

Table 3.2 Distribution of consensus sites (CS) among druggable sites in phosphofructokinase (PFK).

Druggable Site ID	***S. Aureus* (PDB id: 5XZ7)**	***T. Brucei* (PDB id: 3F5M)**	***H. Sapiens* (PDB id: 4RH3)**
1	1A* – 6A*	2A – 3A – 5A – 8A	1A – 3A** – 9A** –
2	**2A – 8A** – 10A – 11A**	6A – 7A	2A** – 4A**
3	5A* – 12A*	4B* – 7B*	2B** – 5B – 10B
4	1B* – 5B*	3B** – 5B** – 6B** – 8B**	3C** – 4C – 8C
5	**2B – 10B**	1C* – 3C** – 6C**	1D** – 5D – 6D – 7D
6	6B* – 11B*	4C* – 5C** – 7C*	6 – 11
7	1C* – 6C*	2D** – 3D** – 6D**	3 – 5 – 9
8	2C – 11C**	4D* – 5D**	10 – 12
9	5C* – 10C*	2* – 5 – 6CD – 9CD – 11CD	1B** – 4B – 9B
10	1D* – 7D*	3 – 12 – 13AB	7B* – 11B* – 12B**
11	**2D – 9D – 11D**	4** – 7 – 11 – 5CD – 12CD	2C** – 6C – 9C
12	**1 – 2 – 3 – 6 – 11 – 16**	1 – 9 – 13 – 6AB – 7AB	2D – 4D**
13	**4 – 5 – 7 – 8 – 12 – 15 -**	10B – 11B	6B – 7AB – 1
14		7D – 10D	4 – 7 – 8 – 11CD
15			2 – 5C

*Clusters with FS < 25% are eliminated.
** Clusters with 25% < FS < 50% are eliminated.

Table 3.3 Top druggable sites in three PFK species. Top druggable sites that has the most amount of CS of species are given in the first line.

Druggable Site ID	*S.aureus*	*T. brucei*	*H. sapiens*
1	**4-5-7-8-12-15-17**	**1-9-13-6AB-7AB**	**4-7-8-11CD**
2	1-2-3-6-11-16	7-11-5CD-12CD	5D-6D-7D
3	2A-10A-11A	5-6CD-9CD-11CD	6B-7AB-1
4	2D-9D-11D	2A-3A-5A-8A	3-5-9
5	2B-10B	3-12-13AB	5B-10B
6	-	6A-7A	4C-8C
7	-	10B-11B	6-11
8	-	7D-10D	10-12
9	-	-	4B-9B
10	-	-	6C-9C
11	-	-	2-5C

The location of all remaining consensus sites was displayed in Figure 3.1, 3.2 and 3.3 for *Sa*PFK, *Tb*PFK and *h*PFK, respectively. It should be noted that most of the consensus sites were located at/near interface regions. In the Figure 3.1, 3.2 and 3.3, orange sticks represent ATP molecules bound to active site, while allosteric regions were indicated by yellow color. Magenta and green sticks are the CS that were coming from chain-by-chain/domain-by-domain and overall mapping, respectively. Although there is no known allosteric region for the parasitic PFK in the literature, presence of CS in experimentally known regions in human and bacterial PFK shows success of new methodology used in this study (Figure 3.1, 3.3). It is possible to say that probe molecules were bound to corresponding areas in three species with small differences. Since structures are similar to each other, this result was expected thus, differentiation became more challenging.

In addition, top druggable sites of three PFK enzyme were shown in Figure 3.4 in more detail. Top druggable regions in bacteria and parasite were located directly at the interface regions, whereas any interface region is absent in the corresponding site of human PFK (Figure 3.4). This difference might offer some advantage for designing species-specific drug molecules.

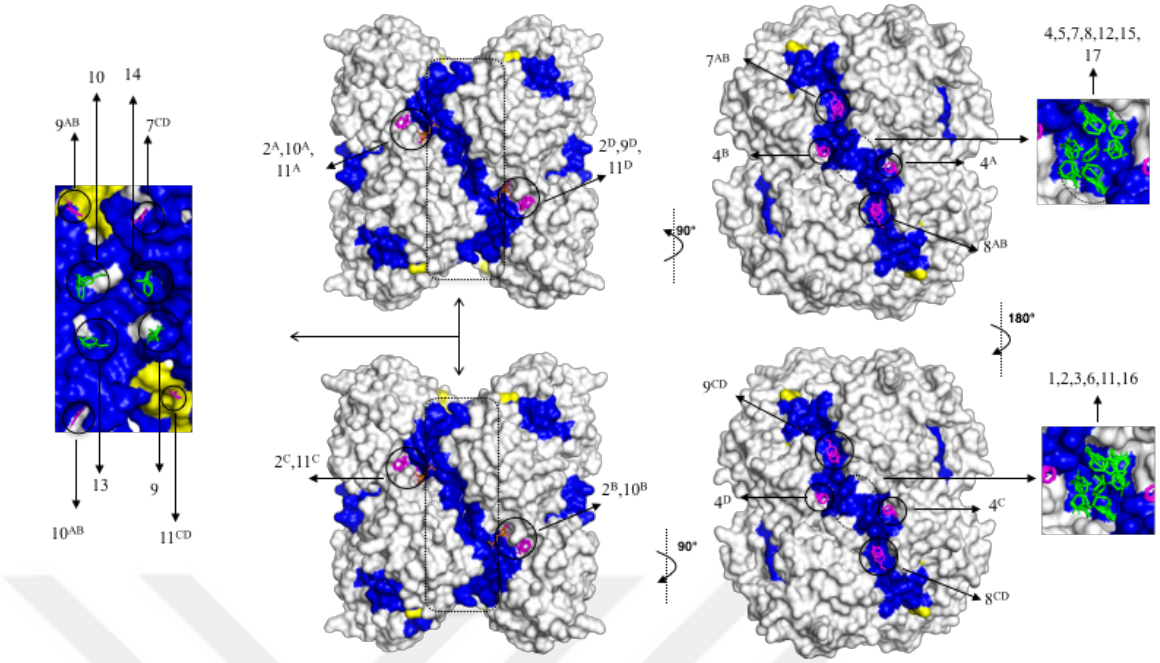


Figure 3.1 *S. aureus* PFK possible ligand binding regions. Yellow color indicates known allosteric sites, while orange sticks are native ligands. Magenta sticks are CS from one-chain and two-chain mapping, and green sticks are CS from overall mapping (Ayyildiz *et al.* 2020).

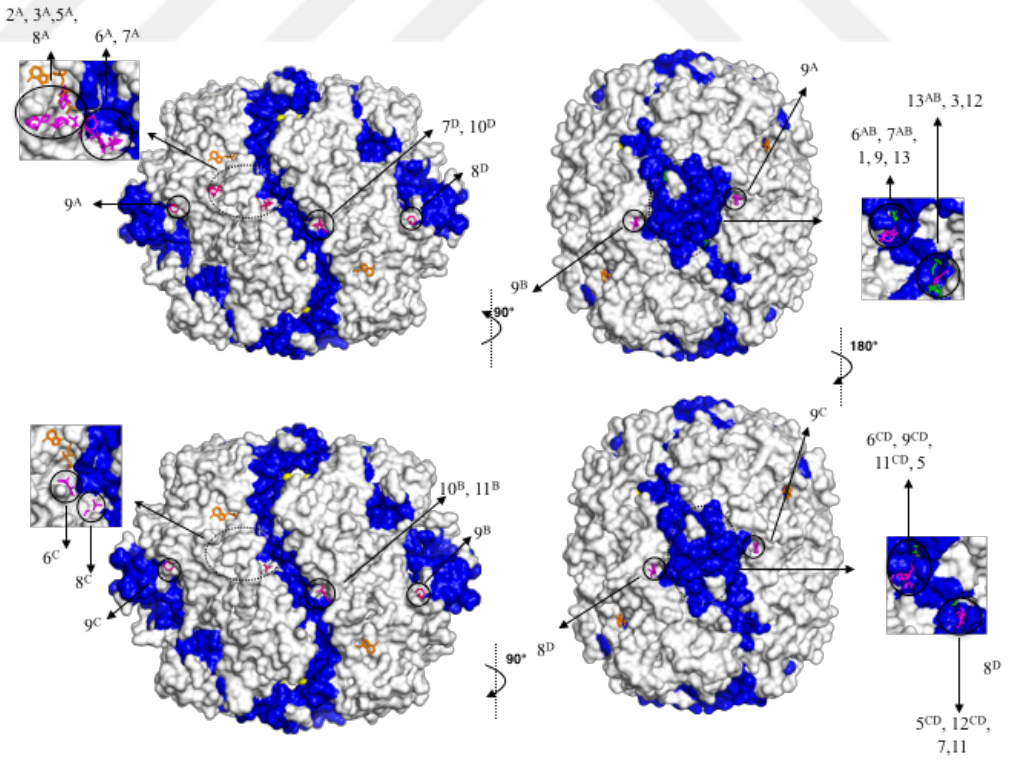


Figure 3.2 *T. brucei* PFK possible ligand binding regions. Yellow color indicates known allosteric sites, while orange sticks are native ligands. Magenta sticks are CS from one-chain and two-chain mapping, and green sticks are CS from overall mapping (Ayyildiz *et al.* 2020).

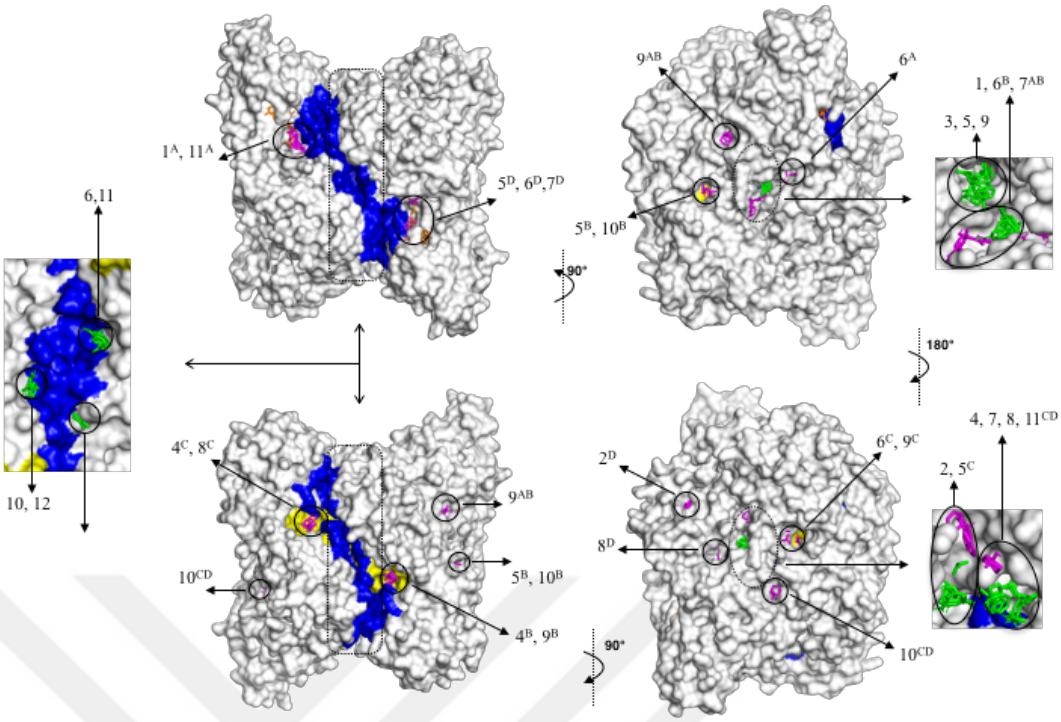


Figure 3.3 *H. sapiens* PFK possible ligand binding regions. Yellow color indicates known allosteric sites, while orange sticks are native ligands. Magenta sticks are CS from one-chain and two-chain mapping, and green sticks are CS from overall mapping.

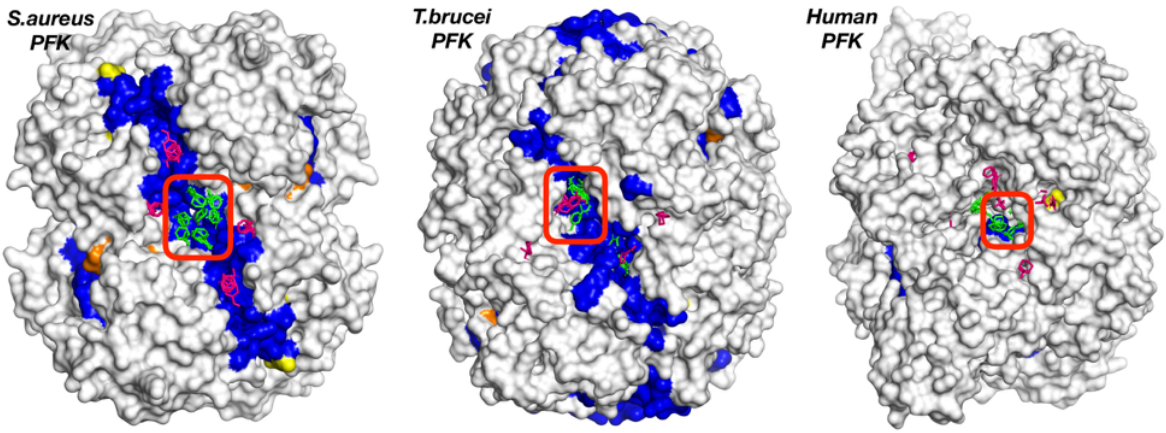


Figure 3.4 Top druggable sites in three enzymes. Blue regions indicate interface regions, while orange and yellow regions show active and allosteric regions, which are known experimentally, respectively (Ayyildiz *et al.* 2020).

3.2 Frequency Shift Analysis via Elastic Network Model

In this study, the elastic network model was employed for the tetrameric form of bacterial/parasitic structures as well as human structure. In Figure 3.5a and 3.5b bacterial and parasitic tetrameric PFK structures were illustrated, whereas in Figure 3.5c and 3.5d human dimer and tetramer structures were shown from three different views. Change in frequency shift is indicated by a color gradient. Accordingly, blue represents decrease and red represents increase in frequency shift, while white indicates that the change is negligible. In addition, increase or decrease in frequency shift means increase or decrease in global dynamics. As can be seen in Figure 3.5, basically interface regions were more likely to affect proteins' global dynamics, additionally the same regions in three species are more likely to have the same affect depending on the structural similarity between them. Besides, the highest and lowest scores belong to the bacterial structure. In addition to different upper and lower scores that species have, structures were recolored according to a fix range (Figure 3.6). The reason behind this was that while distinct limits allow us to evaluate structures within themselves, fixing helps us to understand the relative magnitude of frequency shift between species.

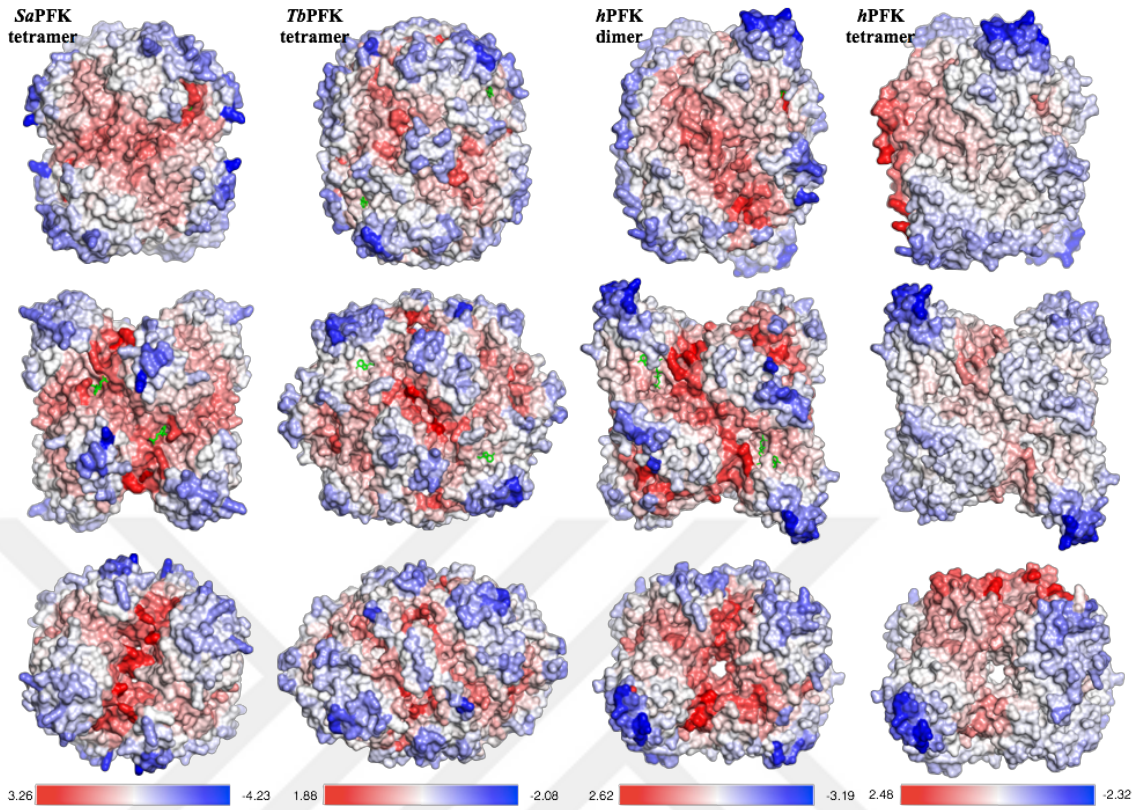


Figure 3.5 ENM-based residue scanning results. Upper and lower values for each species are given below. Green sticks are native ligands of each structure.

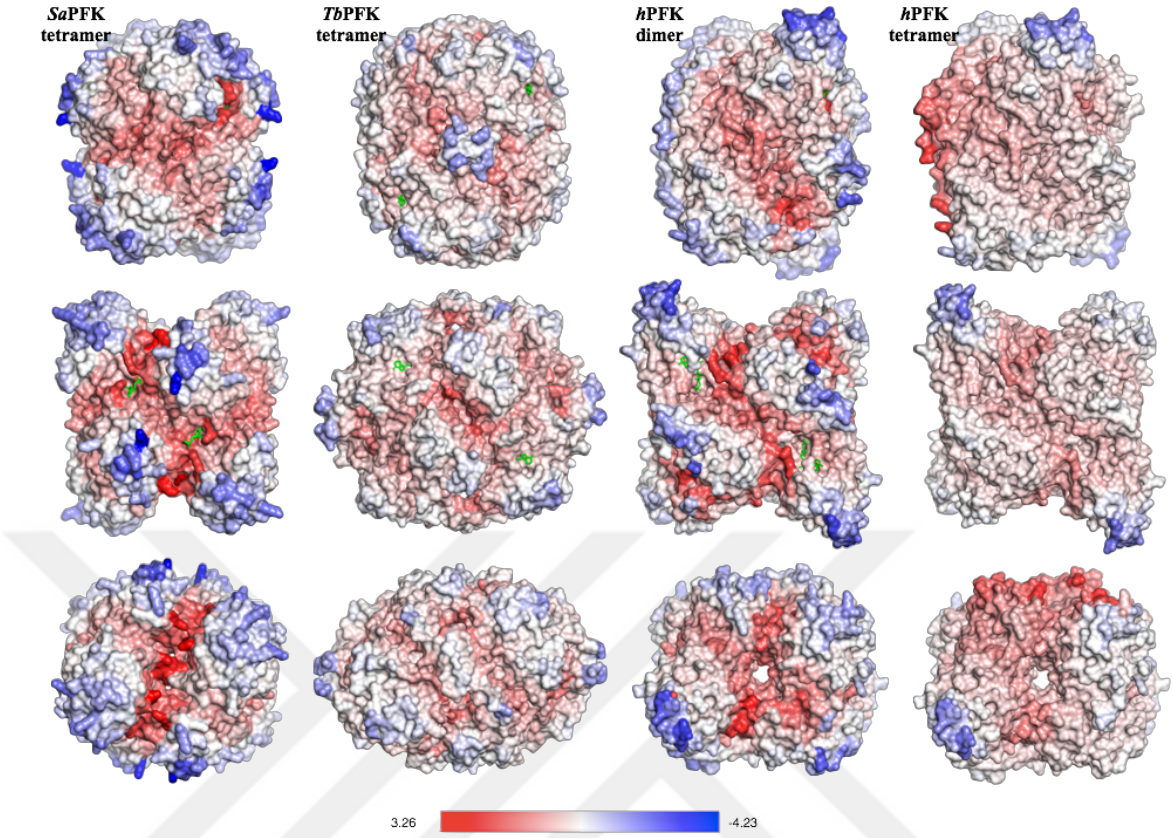


Figure 3.6 ENM-based residue scanning results. Upper and lower values all species have, are given below (values belong to *S. aureus* PFK). Green sticks are native ligands of each structure.

Moreover, when we look at both figures, the negative change can be clearly seen in parasite. Besides, there is significant changes in the both dimer (inactive) and tetramer (active) structures of human. Consequently, it can be said that when a small molecule is bound, bacteria is more likely to be affected or it has more allosteric regions than other two species.

Apart from that, figures clearly show the difference between active and inactive structures in human PFK. The tetramer structure of human PFK was formed by combining two dimers, with the intersection region encircled as shown in Figure 3.7b. Clearly, our method successfully suggested this site as allosteric thus, overlaps with the literature. Besides, change in frequency shifts in *h*PFK tetramer structure is different than *h*PFK dimer structure. Red colored areas are less intense in tetrameric form which indicates less change in global dynamics upon ligand binding. This can be considered as an advantage

for species specific drug design, since in the cell, we see human PFK mostly in active state (tetramer). We further analyzed corresponding areas in bacteria/parasite top druggable site to human active and inactive structures via computational solvent mapping method (Figure 3.8), and no consensus sites were found in the corresponding area in the active human PFK (Figure 3.8b).

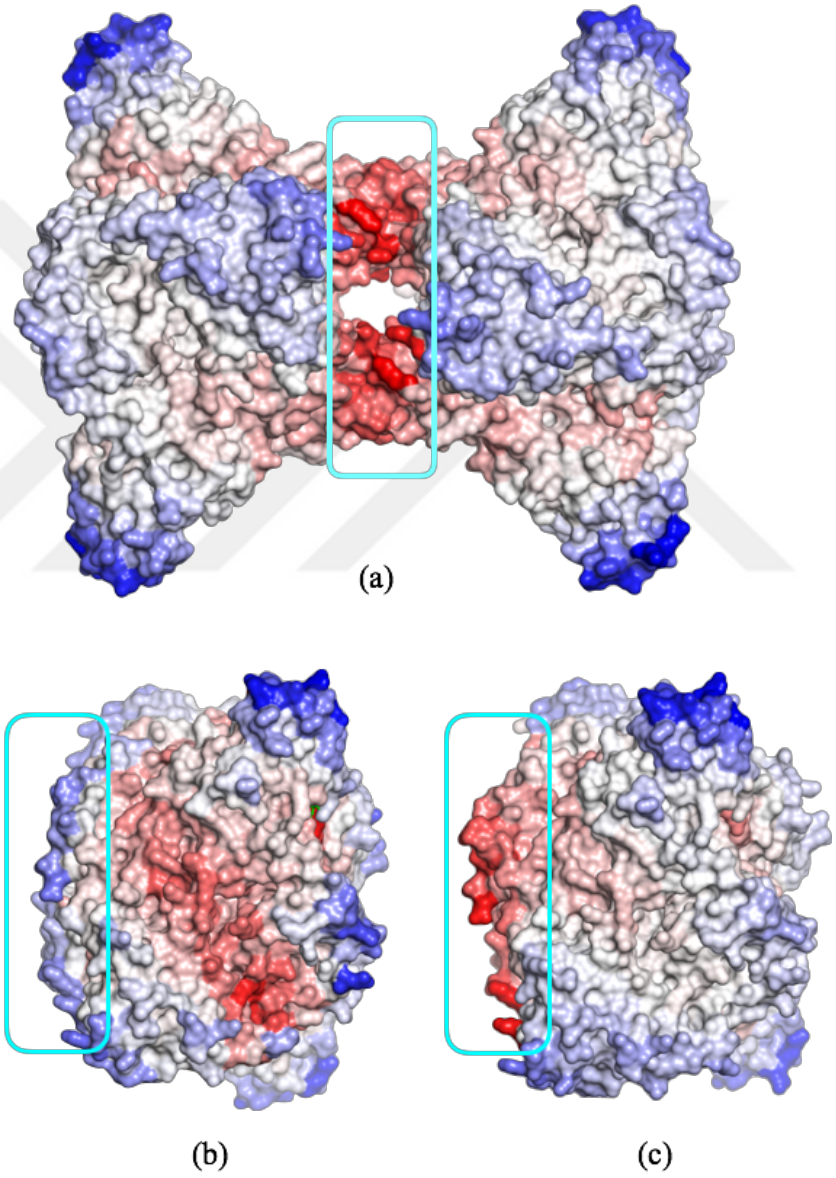


Figure 3.7 Cyan encircled areas are the intersection region of two dimers of *h*PFK. a) Result of frequency shift analysis for tetrameric *h*PFK, b) dimeric form of *h*PFK from separate calculations and c) one half of tetrameric for tetramer *h*PFK.

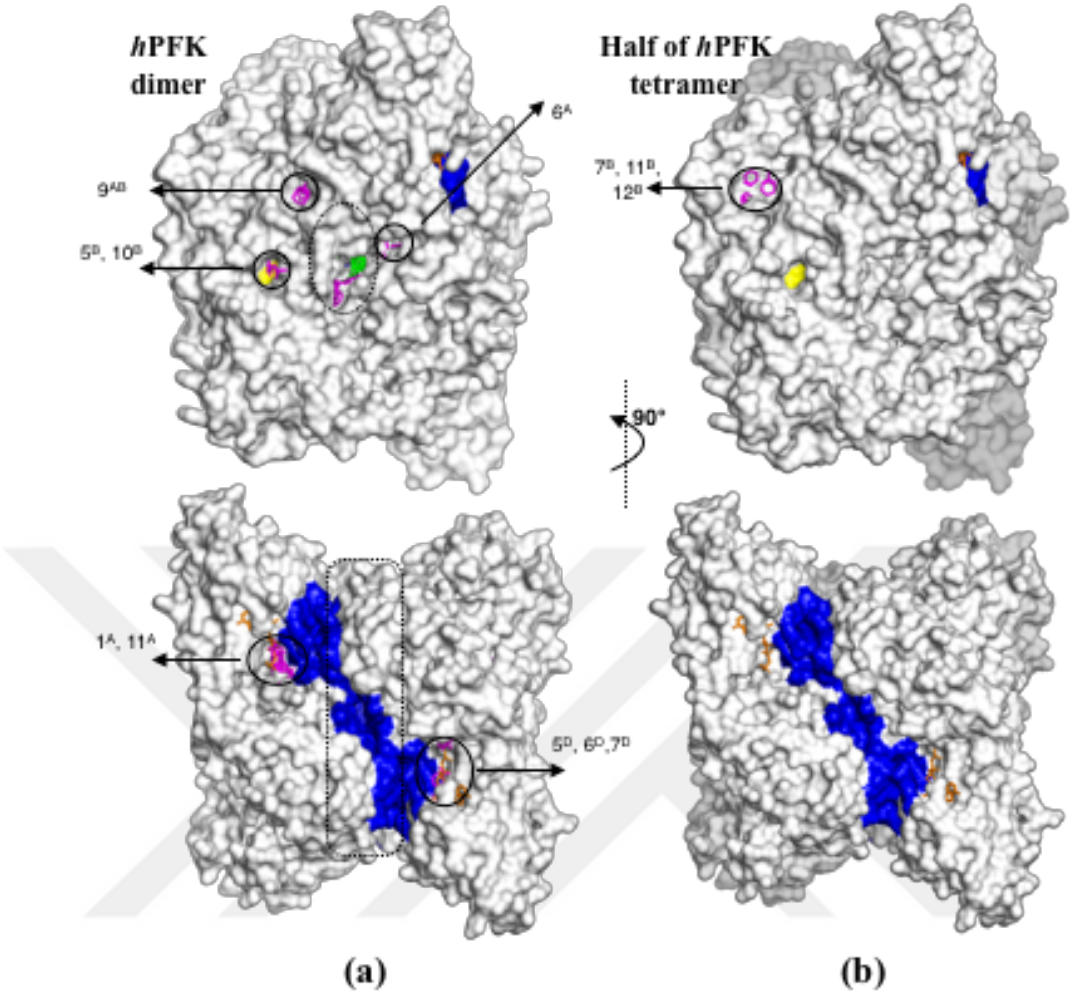


Figure 3.8 Consensus sites comparison of dimer *h*PFK (a), and tetramer *h*PFK (b). Magenta sticks represent consensus sites obtained from one-chain and two-chain mapping.

3.3 Proposed Allosteric Regions with Sequence-Structure Similarities

As mentioned in the Table 3.1 in Method 3.1, fewer and more populated druggable regions for bacteria was observed. As a result of the high structural similarity (RMSD: 1.56 Å, the druggable regions showed high correlation in human and bacteria. However, when looking at the structures from above in Figure 3.9, it can be seen that bacteria has four consensus sites when corresponding area in human has none. However, this region is close to known allosteric regions, thus, could not be proposed. But, this is still an indication that our approach gives accurate results.

Our alternative region for bacteria is shown in Figure 3.3 (side view of Figure 3.9). While this region that was being proposed, had one smaller area with seven consensus sites for

the bacteria, the same region contains fewer consensus sites in humans and remaining CSs (outside of the cyan circled area) were not very close to each other. CS that were considered as druggable site if they are at most 4-5 Å distance from each other. In addition, whereas the interface region in bacteria passed through this region, the same was not the case for human. Hence, this area was highlighted for its allosteric and druggability potential. In order to propose alternative region where was mentioned above as species-specific, sequence and/or structural dissimilarities had to be revealed. As shown in Figure 3.10, although there was sequence similarity in the desired region, the difference could be seen to be satisfactory. However, overall structural similarity was only 1.56 Å and even lower for the desired region. Despite this high structural similarity between *h*PFK and *Sa*PFK, it is likely for this region to be a possible allosteric region considering the low sequence identity between species (Figure 3.10) and being located at interface region of bacteria. For *Sa*PFK, thus, this novel region that we propose and the region that was already known were chosen for further docking studies.

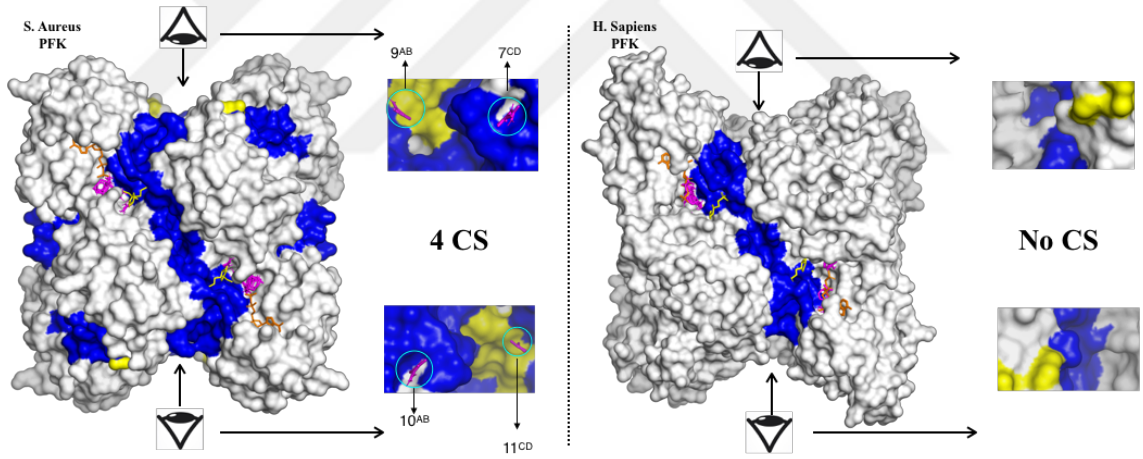


Figure 3.9 Comparison of consensus sites between *Sa*PFK and *h*PFK. a) *Sa*PFK consensus sites located at/near interface regions and known allosteric region. b) *h*PFK has no consensus site when compared with bacterial PFK (Ayyildiz *et al.* 2020).

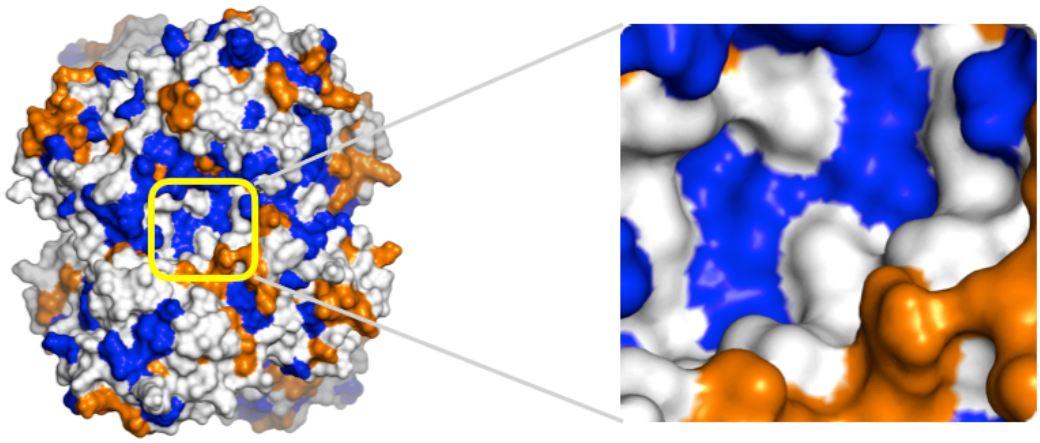


Figure 3.10 Sequence similarity between *h*PFK and *Sa*PFK illustrated on a snapshot Similar, identical and dissimilar residues colored in orange, blue and white, respectively. Top site residues encircled in yellow in sequence alignment (Ayyildiz *et al.* 2020).

In *Tb*PFK, two main possible allosteric regions were identified. As illustrated in Figure 3.11, cyan encircled areas contain several CS, however, these areas were located either in/near the active site. Therefore, they could not be proposed as novel region. However, these two regions were close to each other and the area between them was directly at the interface region and the same region was comparably small in human structure which further increases its specificity. Moreover, if sequence similarity is considered for the region, white and orange colored residues attract attention. Considering that the blue, orange and white colored residues demonstrate identical, similar and dissimilar residues respectively, it is possible to say that the similarity was significantly low. Taking all into account, it is convincing to propose this region as an ideal target region for species-specific drug design studies. The second region was identified as top druggable site for parasite which is located between A and B chains (Figure 3.11) and interface region passes through this site as in bacteria, and consists of five CS. Also, as an outcome of structure and sequence alignment, this site does not exist in human structure. As shown in Figure 3.11, residues were all white in this site. This significantly strengthens the site's probability of being species-specific.

As a summary, two sites in bacteria, one is known and the other is top druggable site, were determined for further studies. In this thesis, no further studies on parasite have been carried out, however, it can be said that two sites are suitable, as one alternative site and the other as the top druggable site.

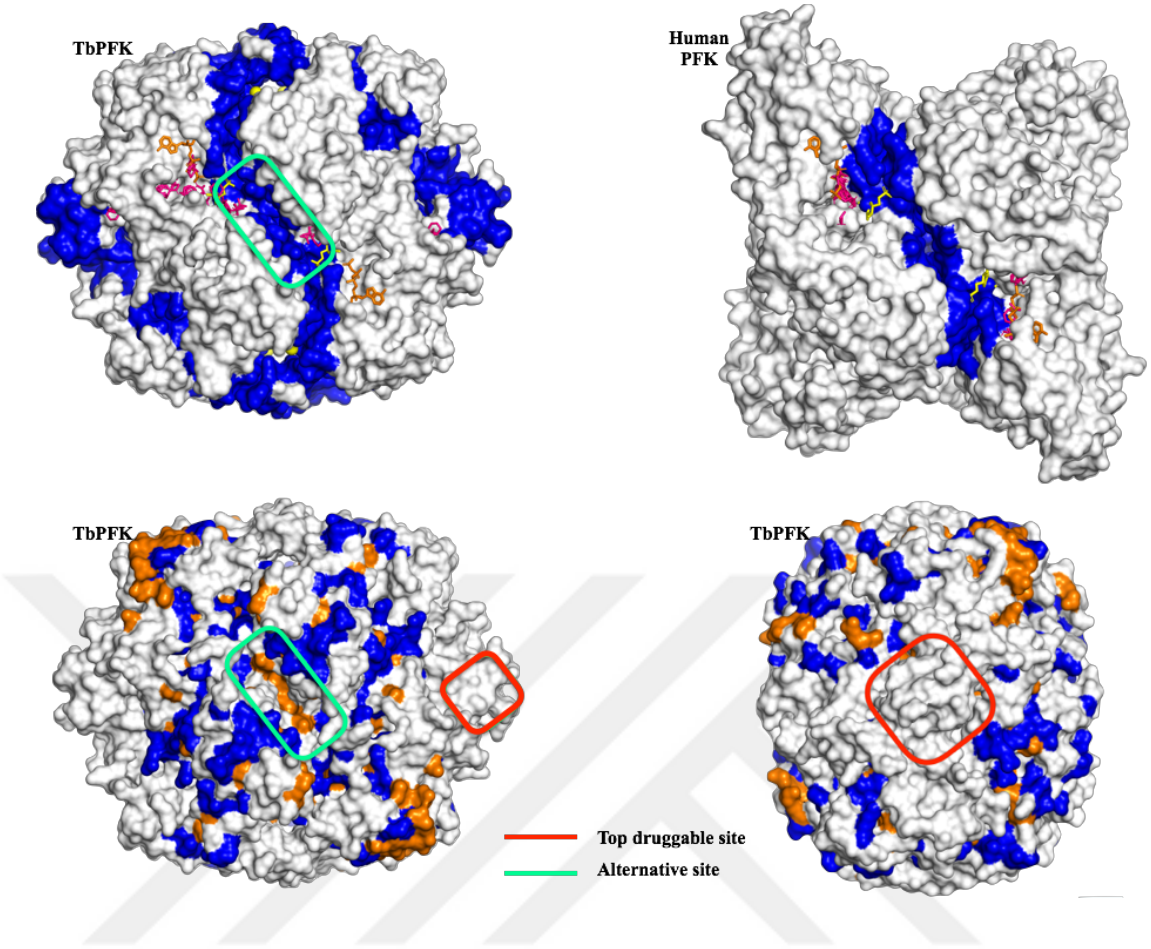


Figure 3.11 Snapshots of *Tb*PFK top druggable site and an alternative site (Ayyildiz *et al.* 2020).

3.4 Critical Assessments with DoGSiteScorer and AlloSigma Tools

In the following step, we compared our proposed druggable regions with potential druggable regions suggested by DoGSiteScorer tool which is a grid-based method. Our druggable sites for *Sa*PFK/*Tb*PFK was given in Table 3.2, and these two species re-evaluated according to the scores of DogSiteScorer tool. DrugScore's of pockets vary between 0 to 1 and closer value to 1 means higher probability for possible binding region. Accordingly, in Table 3.4 shows that high scored pockets of DogSiteScorer tool overlapped with our proposed regions (0.81 in *Sa*PFK allosteric region and 0.8 in *Tb*PFK allosteric region). In Figure 3.12 and Figure 3.13, for *Sa*PFK and *Tb*PFK top results of DogSiteScorer pockets was shown as red spheres with their DrugScore values. The first two pockets given by DoGSiteScorer for *Sa*PFK were in the active regions of the receptor, and DrugScore values of these pockets were slightly higher than the DrugScore values of

the region that we propose (Figure 3.12a, b) However, the same values of the pockets in the symmetrical active regions of the SaPFK receptor was lower than the proposed region, with 0.79 and 0.78. When DrugScore values of active regions and proposed region were compared, it can be seen that our result and DoGSiteScorer results overlap.

The region corresponding to the top druggable region for *Sa*PFK was also top druggable region for *Tb*PFK. This proposed region for *Tb*PFK, was suggested as the fourth highest probable region by DoGSiteScorer with the DrugScore value of 0.8 (Table 3.4). Also, in *Tb*PFK, DoGSiteScorer pockets in its active regions have slightly higher DrugScore values than our proposed region (Figure 3.13a, b). In addition, the second alternative region that we propose for *Tb*PFK was also given by DoGSiteScorer as a third highest probable region with 0.81 DrugScore value which is similar to its active regions' values (Figure 3.1b). Interestingly, DoGSiteScorer proposed inside the four monomers of *Tb*PFK as the highest probable binding region where has not been previously referred in the literature and also was not found by our method as well.

Table 3.4 Druggable sites with DoGSiteScorer results of *S. aureus* PFK and *T. brucei* PFK.

S.aureus	Region	Score/ Rank	Parasite	Score/ Rank	Region
4-5-7-8-12-15-17	Allosteric	0.81 / 3	1-9-13-6AB-7AB	0.80 / 4	Allosteric
1-2-3-6-11-16	Allosteric	0.81/ 3	7-11-5CD-12CD	0.80 / 4	Allosteric
2A-10A-11A	Catalytic	0.49 /13	5-6CD-9CD-11CD	0.80 / 4	Allosteric
2D-9D-11D	Catalytic	0.49 / 13	2A-3A-5A-8A	0.81 / 3	Catalytic
2B-10B	Catalytic	0.78 / 6	3-12-13AB	0.80 / 4	Allosteric
			6A-7A	N/A	Catalytic
			10B-11B	0.53 / 16	Allosteric

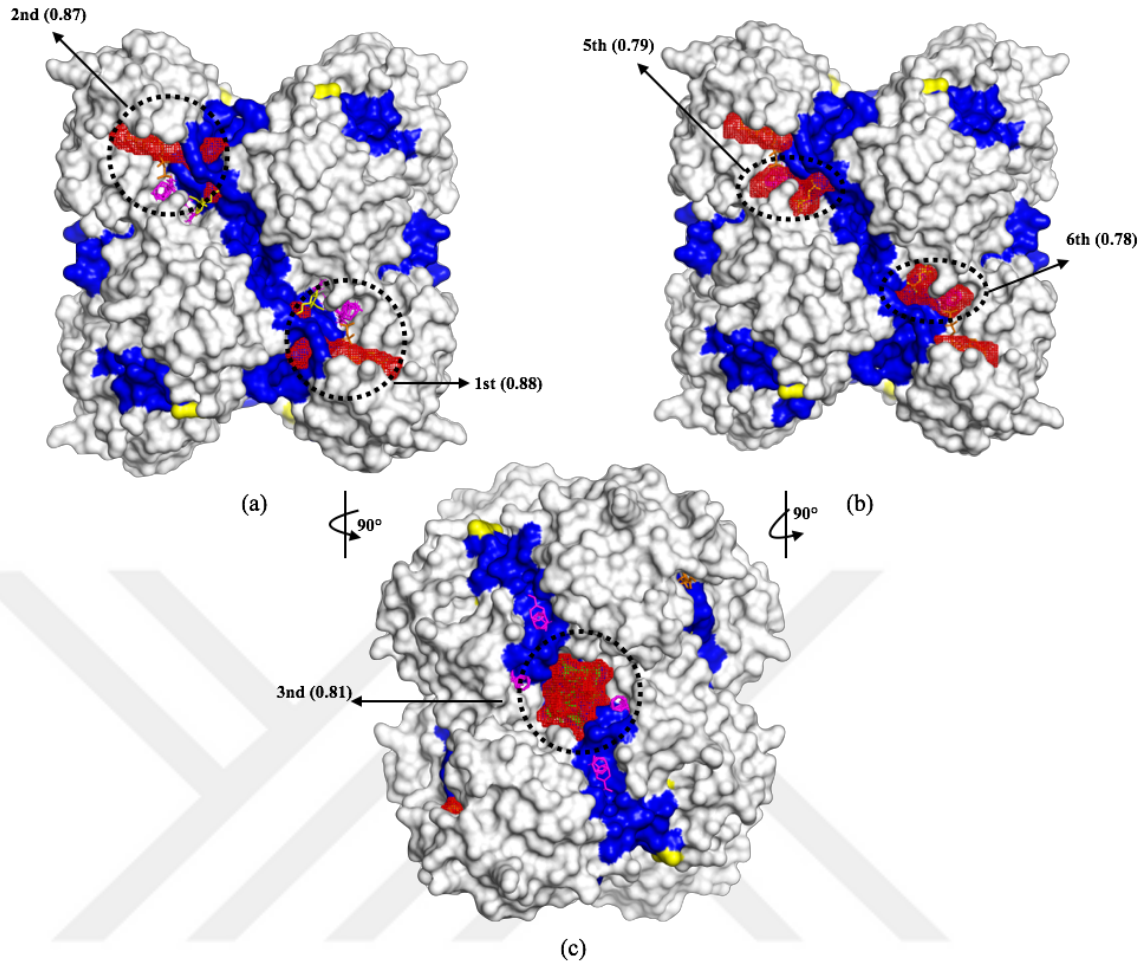


Figure 3.12 Snapshots showing the location of the binding pockets predicted by DoGSite in *S.aureus* phosphofructokinase (Ayyildiz *et al.* 2020). a), b) DoGSiteScorer pockets in active regions, c) DoGSiteScorer pockets in proposed region with their DrugScore values.

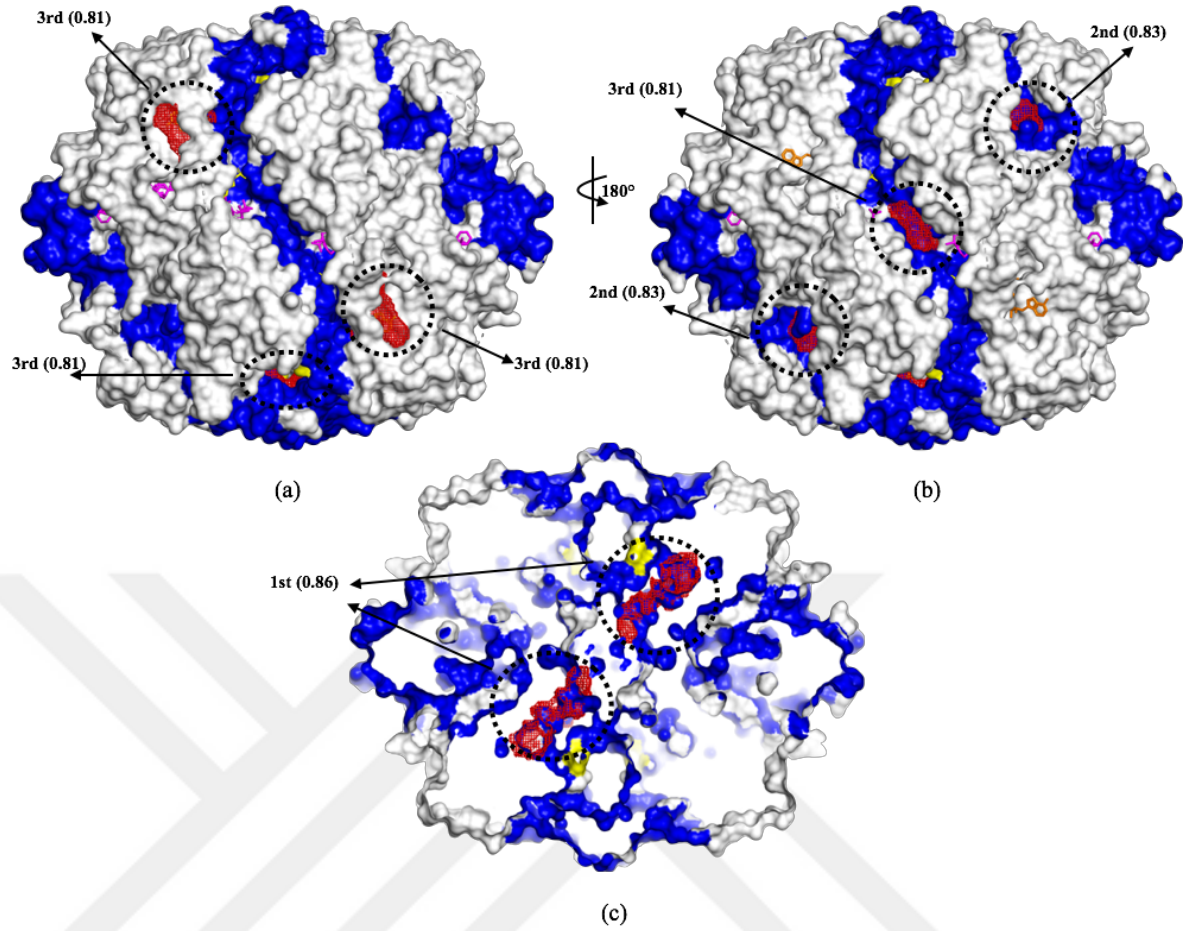


Figure 3.13 Snapshots showing the location of the binding pockets predicted by DoGSite in *T.brucei* phosphofructokinase (Ayyildiz *et al.* 2020).

For *Sa*PFK, AlloSigma tool has been used in four different regions, two symmetric top druggable and two symmetric known allosteric regions. In the results, blue color shows increase of dynamics which means destabilization, while red color indicates decrease of dynamics, means stabilization. First, residues of two symmetric druggable regions (top and second top druggable) were selected for AlloSigma tool analysis, since possible inhibitor may bind both regions at the same time (Figure 3.14). Here, catalytic regions of all four homotetramer chains were rather affected than other regions. When top druggable region and its symmetric region were examined in SaPFK, ΔG values given by AlloSigma are between -0.3 and -0.02 for the ATP binding regions and -0.1 for F6P binding regions. ΔG values for all four catalytic regions, both for ATP and F6P ligands were negative, which indicates possibly stabilized, restricted proteins.

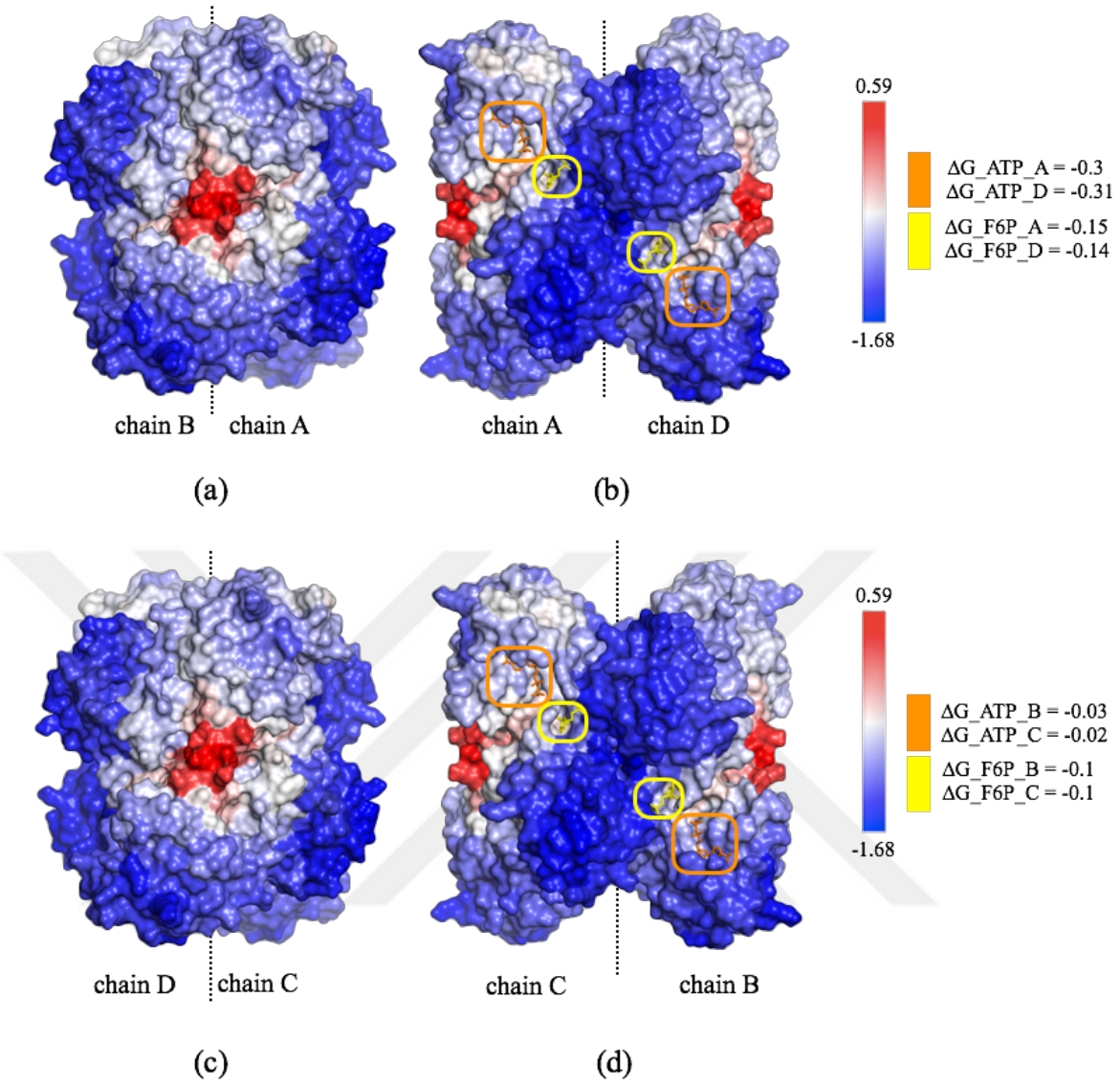


Figure 3.14: Change in the catalytic region when two top druggable regions residues are selected for AllaSigma tool analysis. a), c) Top druggable region and second top druggable region. b), d) Catalytic regions are shown in orange (ATP) and yellow (F6P) rectangles.

Lastly, the region where known as allosteric activator binding region were selected examined by AlloSigma tool for only comparison purpose and the results are shown in Figure 3.15. This region is known for effector binding which increase the activity of protein. Therefore, ΔG values for both binding region in monomers (ATP and F6P) were expected to be positive. Although ATP binding regions of A and D monomers had negative ΔG values, whereas ATP binding sites in B and C monomers and all F6P binding sites had positive ΔG values as expected.

In this thesis, allosteric effect of possible ligand-binding regions was evaluated with AlloSigma tool for only *Sa*PFK. The same analysis could be also used for potential druggable regions in *Tb*PFK in future studies.

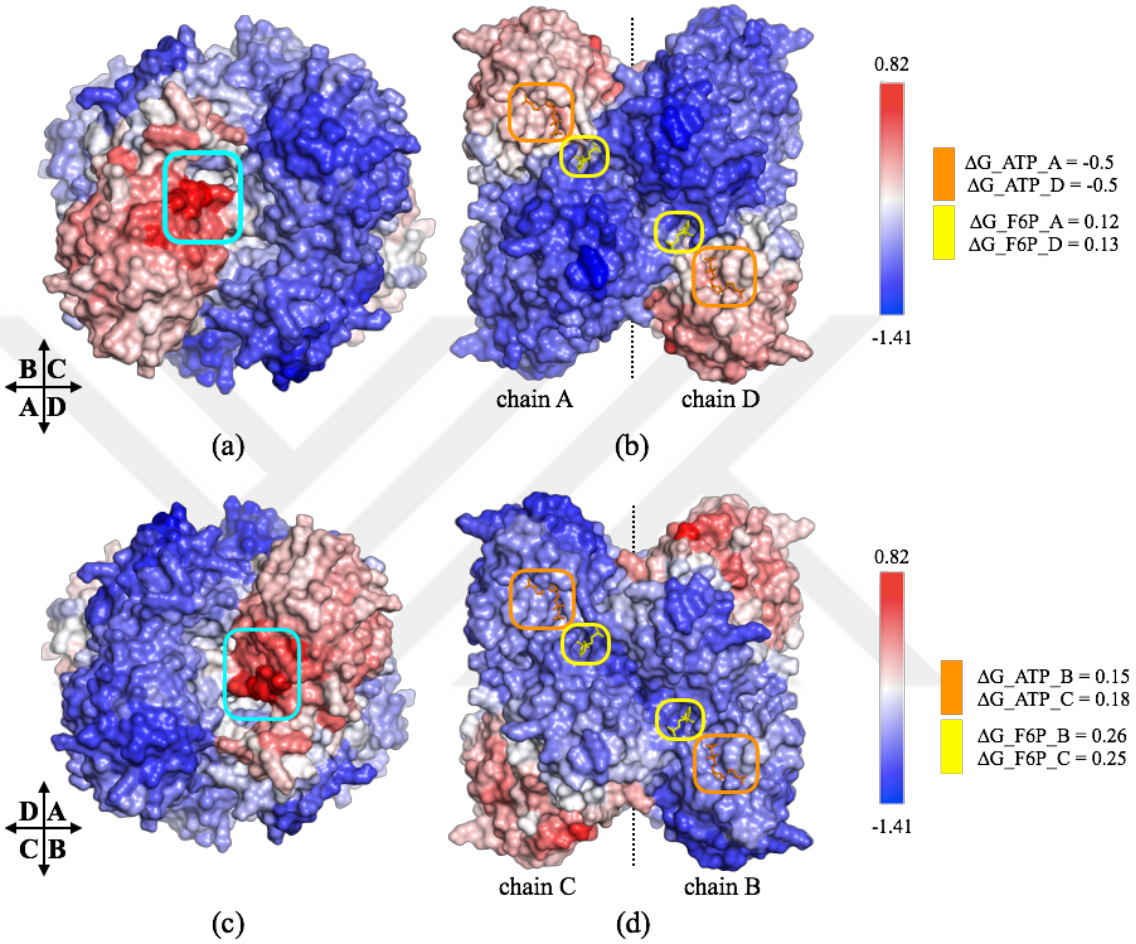


Figure 3.15: Change in the catalytic region when two known allosteric regions residues are selected via AlloSigma. a), c) Two symmetric allosteric activator binding regions. b), d) Catalytic regions are shown in orange (ATP) and yellow (F6P) rectangles.

3.5 High-Throughput Virtual Screening via Docking

In the following step, drug repositioning was performed for the selected sites in bacteria via GOLD docking software. Two different subsets from ZINC15 database were used for one proposed and one known allosteric region in bacteria and for corresponding regions in human. First, FDA approved and then World-not-FDA drug molecules were studied.

The FDA subset contains 1416 molecules, while World-not-FDA subset has 2922 molecules. Default GOLD/ChemPLP parameters were used. The coordinates of the grid box corresponding to the regions in bacteria had to be given for the human. Therefore, structural alignment was done once more and coordinates were given by predicting exactly where the CS in the bacteria would be placed in the human structure (Table 3.5). There is minor difference between the coordinates given in human and bacteria for docking. The reason for this difference is to ensure that the regions that are determined for the bacteria (top druggable and known allosteric) are used exactly in the human and to prevent small shifts caused by alignment of human to bacteria. ChemPLP scores for *S. aureus* were ranked from highest to lowest. For each region determined in PFK, the first 28 and 58 molecules in the top 2% were taken into consideration (number of FDA approved and World-not-FDA approved molecules, respectively). It is not known exactly which region these drug molecules can bind and inhibit the enzyme. For this reason, in order to reveal the molecules that are bound to both regions with high scores, common ones were taken from these 28 and 58 molecules determined for each region. After screening, there are 6 molecules left from FDA, and other 6 molecules from World-not-FDA subsets. A general summary of these stages is summarized as in Figure 3.16.

Table 3.5 Coordinates of grid boxes of for virtual screening that were used in bacteria and human.

Coordinates	**Bacteria**			**Human**		
	X	**Y**	**Z**	**X**	**Y**	**Z**
First site (Top druggable)	36.21	39.28	22.78	38.31	39.18	25.01
Second site (known allosteric)	30.41	57.01	7.4	30.41	57.1	3

As shown in Figure 3.17, for PFK enzyme, 1416 FDA molecules were ranked by score values of bacteria, also the difference between bacteria and human was given in the y-axis on the right side. Accordingly, interesting conclusions can be thought; in the proposed allosteric region the difference between bacteria and human was negative for the molecules (Figure 3.17a). In other words, molecules that bind to bacteria with lower scores bind better to human. However, for the known allosteric region, it is seen that the high-scores molecules on the left side of Figure 3.17b are bound to bacteria with higher scores than human, which highlights their potential to be species-specific. Interestingly,

the same trend with FDA approved set applies to World-not-FDA approved molecule set as seen in Figure 3.18. Then, for both subsets, the first 28 and 58 molecules with the highest value were selected for both bacteria and human (FDA approved, World-not-FDA approved, respectively) (28 and 58 are %2 of 1416 and 2922 molecules). Afterwards, common molecules in bacteria and human were identified (Figure 3.18c, d, 3.18c, d) In the graphs showing the results, purple line, blue line and green squares shows scores of bacteria, differences of bacteria-human scores and the molecules that are commonly has high scores. These common molecules and their properties (its score in bacteria and in human molecular weight etc. are given in Table 3.6, 3.7.

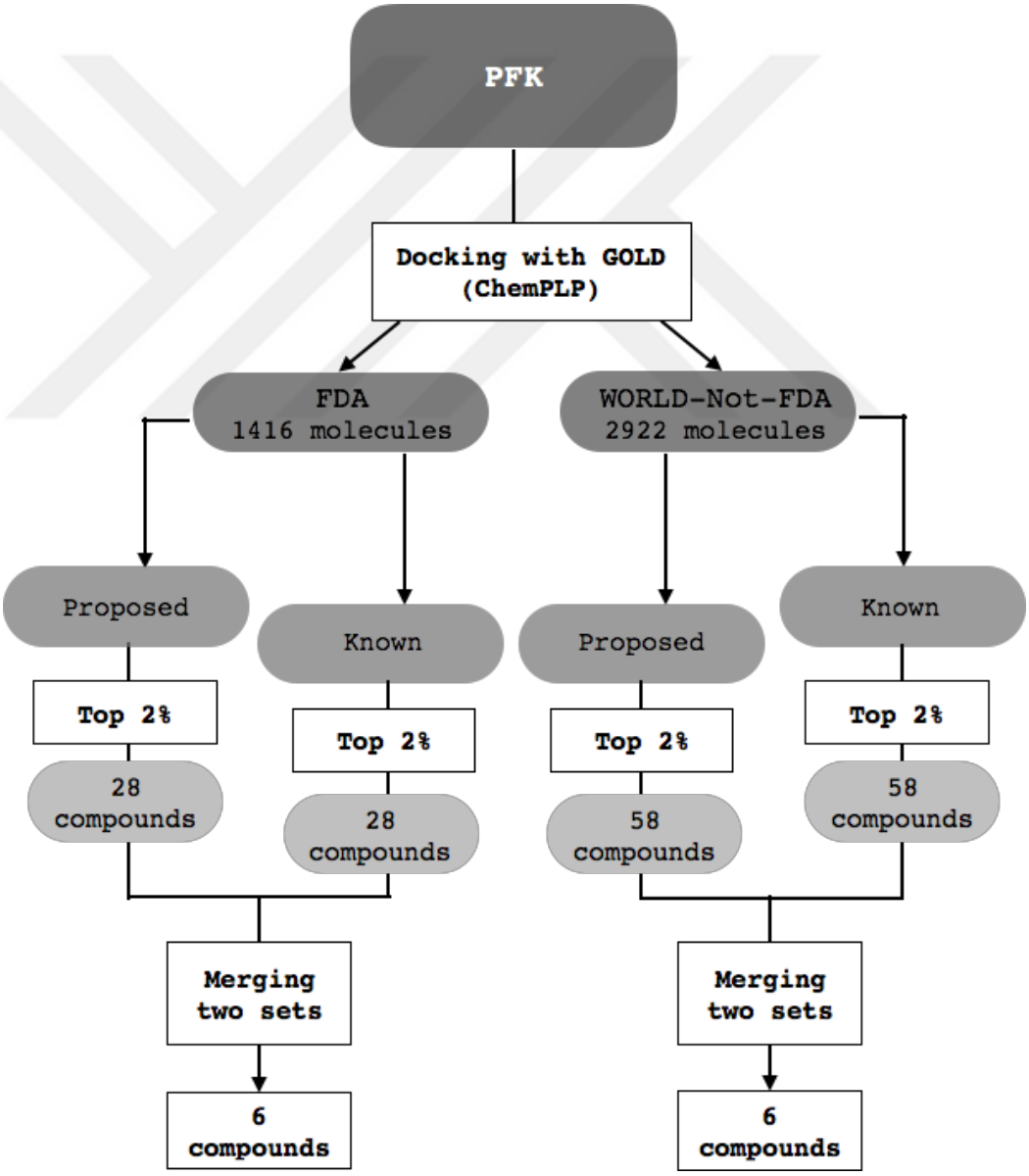


Figure 3.16 Flowchart followed for docking studies.

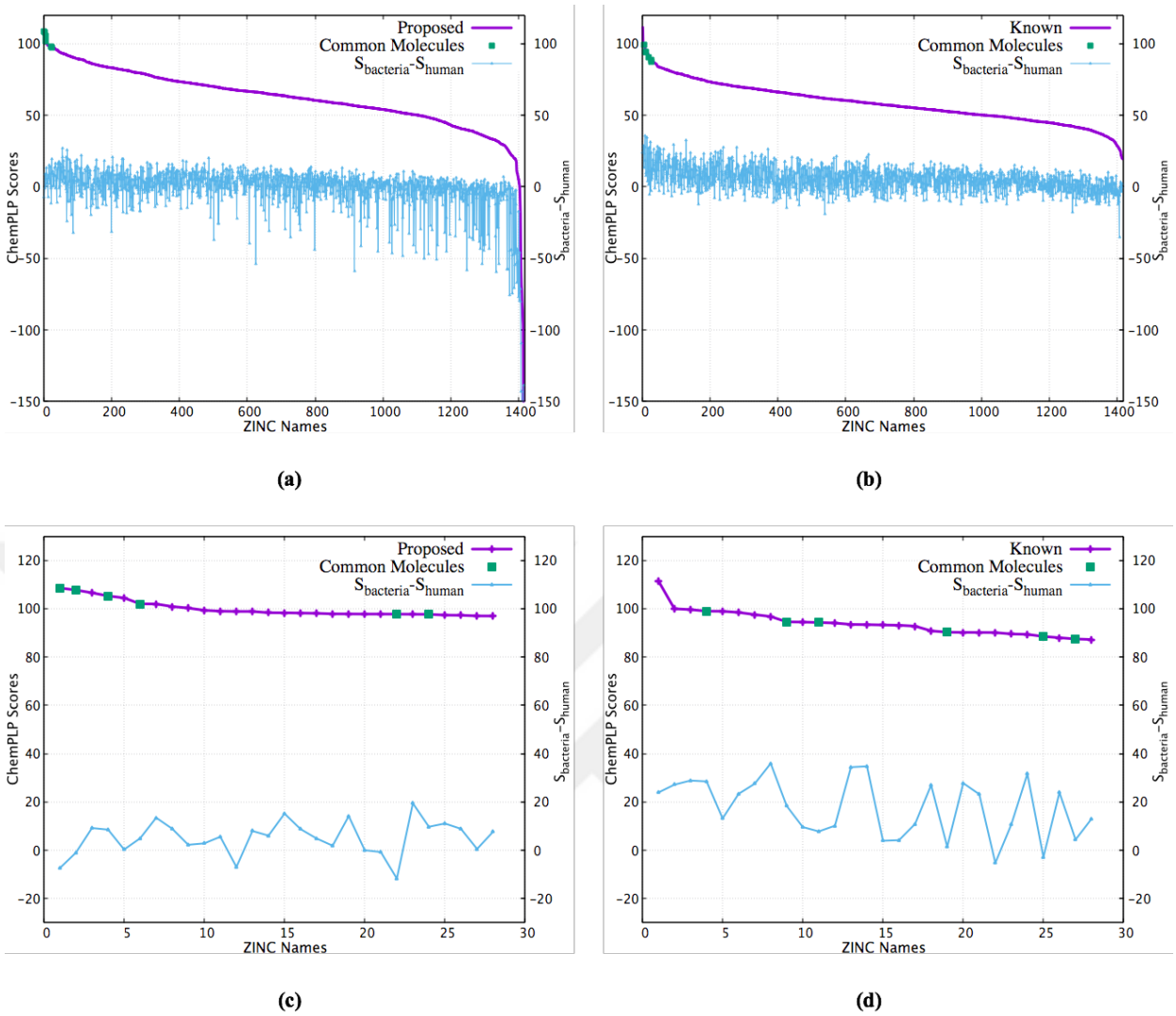


Figure 3.17 FDA-approved molecules ranked by scores of bacteria. a), b) score of 1416 molecules ranked by bacteria and their differences with human (purple and green) in proposed and known allosteric regions. c), d) 28 molecules in the top 2% of total ranked by bacteria and their differences with human (purple and green) in proposed and known regions.

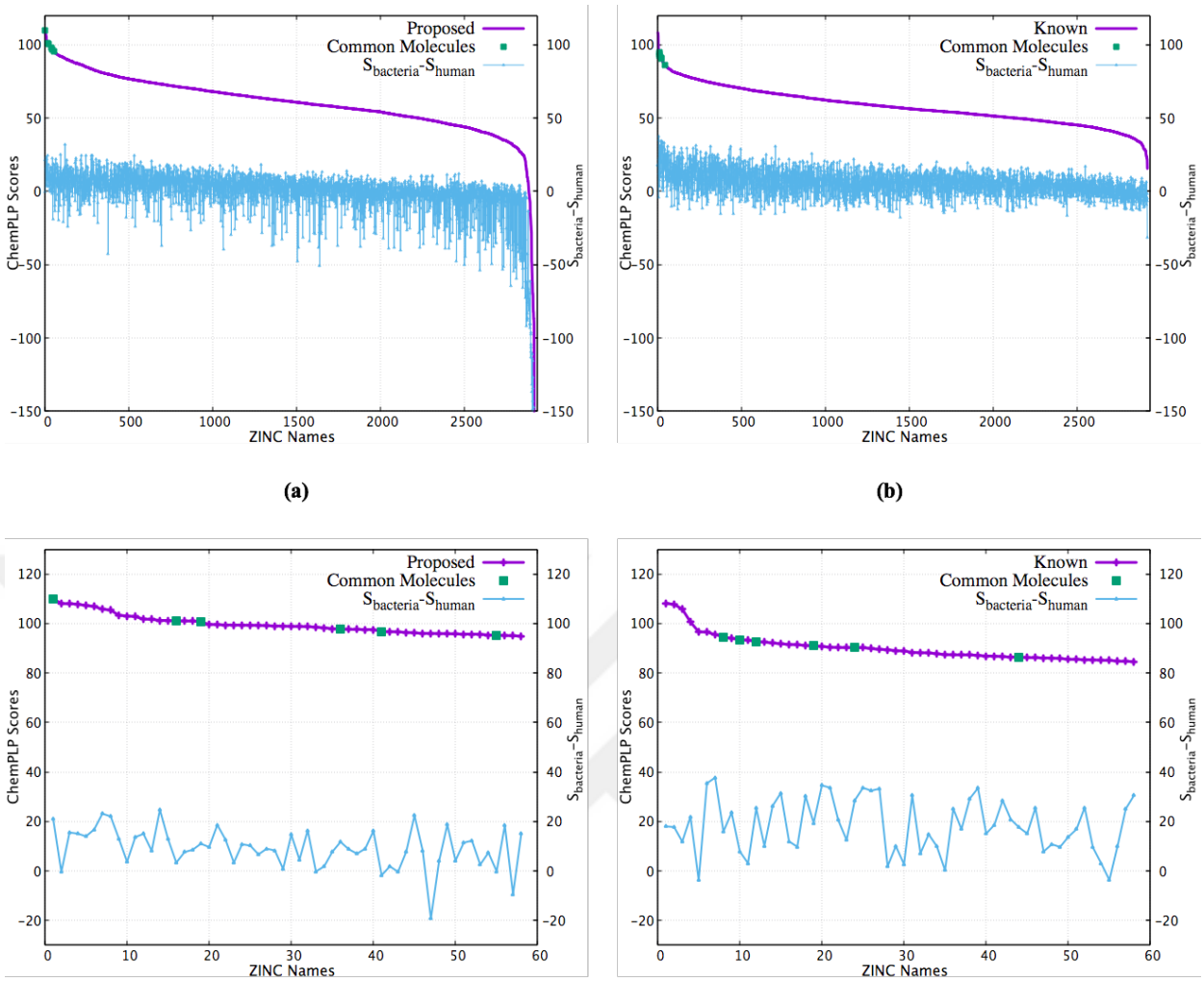


Figure 3.18 World-not-FDA molecules ranked by scores of bacteria. a), b) score of 2922 molecules ranked by bacteria and their differences with human (purple and green) in proposed and known allosteric regions. c), d) 58 molecules in the top 2% of total ranked by bacteria and their differences with human (purple and green) in proposed and known regions.

Table 3.6 Common 6 molecules (FDA approved) docked to both proposed and known allosteric sites with high scores.

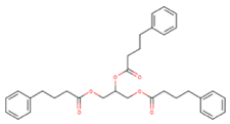
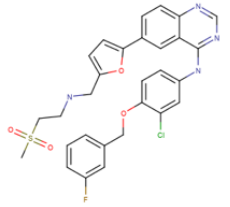
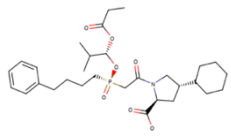
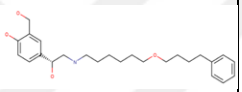
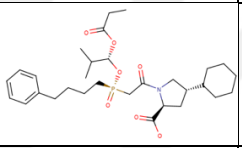
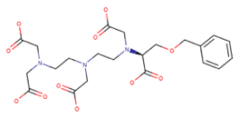
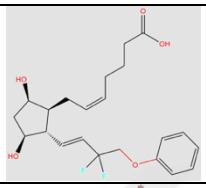
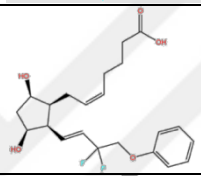
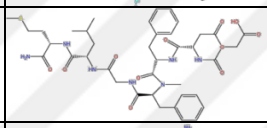
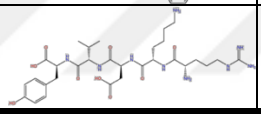
ZINC ID	Molecule Name	2D Structure	Molecular Weight	Bacteria		Human	
				Proposed	Known	Proposed	Known
ZINC000038945666	Ravicti		530,661	108.56	88.51	115.95	91.39
ZINC000001550477	Lapatinib		581,069	107.7	98.97	108.72	70.51
ZINC000004097309	ZINC000004097309		563.672	105.29	94.53	96.76	76.24
ZINC000003799072	Salmeterol		415.574	102	87.47	97.05	83.10
ZINC000004097310	ZINC000004097310		563.672	97.77	90.25	109.59	88.82
ZINC000021982937	Multihance		513.500	97.68	94.28	88	86.5

Table 3.7 Common 6 molecules (World-not-FDA approved) docked to both proposed and known allosteric sites with high scores.

ZINC ID	Molecule Name	2D Structure	Molecular Weight	Bacteria		Human	
				Proposed	Known	Proposed	Known
ZINC000001655706	Dequalinium		450.678	110.02	92.59	88.91	67.24
ZINC000100090021	ZINC000100090021		454.662	101.05	90.22	97.69	61.66
ZINC000027644813	Tafluprost Free Acid		410.457	100.67	86.41	89.70	68.83
ZINC000256090847	ZINC000256090847		410.457	97.72	90.96	85.98	71.60
ZINC000095615286	Senktide		841.985	96.71	93.42	98.44	85.58
ZINC000003938642	Thymopentin		679.776	95.30	94.58	95.49	78.70

4. CONCLUSION AND FUTURE WORK

The aim of this thesis was to identify possible allosteric binding sites in phosphofructokinase enzyme in bacteria, parasite and human for use in species-specific drug design studies. Therefore, a novel approach that uses and combines several well-established and widely used methods such as computational solvent mapping, elastic network modelling, sequence and structural alignments and also additional analysis methods such as the relationship of interface regions to ligand binding regions have been employed. This novel approach has successfully found nearly all reported catalytic and allosteric regions for PFK in all three species, thereby supporting the accuracy of our findings. In addition, alternative allosteric regions have been proposed for use in species-specific drug design studies. Subsequently, one proposed and one known allosteric region in bacteria and two proposed regions in parasite were revaluated with supportive methods. These regions were located in the interface regions and also overlapped with DoGSiteScorer pockets that had high DrugScore values. Moreover, when the proposed allosteric region and its symmetrical counterpart were examined by AlloSigma tool for only *Sa*PFK, ligand binding caused a significant decrease in dynamics of all catalytic regions which further supports our findings. The same analysis was also conducted to the known allosteric region and its symmetrical counterpart, but on the contrary, an increase in dynamics was observed which was not a desired allosteric outcome. Additionally, virtual screening studies were done for *Sa*PFK via using FDA approved and World-not-FDA approved subsets that includes 1416 and 2922 molecules, respectively. In *Sa*PFK, docking was performed for top druggable site that we proposed and the known allosteric site. Corresponding regions in human PFK were also docked. Several small molecules with slightly lower docking values in human and higher values in bacteria were identified. Finally, we have decided to six molecules from FDA approved subset and six molecules from World-not-FDA subset, which bound to both regions with high docking scores.

For further studies, drug molecules obtained from previous work and changes in the dynamics of the enzyme will be determined by the binding of the molecules via molecular dynamics simulation. For this, MD simulations will be performed for the apo form of the enzyme and ligand-bound form of the enzyme. Its purpose is to reveal the effect of ligand binding on the dynamics of the receptor and eventually the allosteric capacity of the

ligand. Therefore, the effect of the proposed region as allosteric region will be demonstrated and finally drug molecules proposed via computational methods will be investigated in further stages of *in-vitro* studies.

In this study, we focused only one of the allosteric enzymes in the glycolytic pathway. However, the remaining seven non-allosteric enzymes can be also investigated via using the same approach to unravel the hidden allosteric sites. Considering that supposedly all enzymes have allosteric regions, the same approach can be extended and applied to all enzymes to contribute to the development of effective allosteric drugs.

REFERENCES

Abel, Samantha, Timothy M. Jenkins, Lyndsey A. Whitlock, Caroline E. Ridgway, and Gary J. Muirhead. 2008. "Effects of CYP3A4 Inducers with and without CYP3A4 Inhibitors on the Pharmacokinetics of Maraviroc in Healthy Volunteers." *British Journal of Clinical Pharmacology* 65 (SUPPL. 1): 38–46. doi.org/10.1111/j.1365-2125.2008.03134.x.

Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. 2001. "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model." *Biophysical Journal*. doi.org/10.1016/S0006-3495(01)76033-X.

Ayyildiz, M., Celiker, S., Ozhelvaci, F., Akten, E. D. (in review) "Identification of Alternative Sites in Glycolytic Enzymes for Potential Use as Species-Specific Drug Target" *Frontiers in Molecular Biosciences* doi:10.3389/fmolb.2020.00088.

Barnett, James A. 2003. "A History of Research on Yeasts 5: The Fermentation Pathway." *Yeast*. doi.org/10.1002/yea.986.

Bohr, Chr, K. Hasselbalch, and August Krogh. 1904. "Ueber Einen in Biologischer Beziehung Wichtigen Einfluss, Den Die Kohlensäurespannung Des Blutes Auf Dessen Sauerstoffbindung Übt." *Skandinavisches Archiv Für Physiologie*. doi.org/10.1111/j.1748-1716.1904.tb01382.x.

Boscá, Lisardo, and Carlos Corredor. 1984. "Is Phosphofructokinase the Rate-Limiting Step of Glycolysis?" *Trends in Biochemical Sciences*. doi.org/10.1016/0968-0004(84)90214-7.

Brenke, Ryan, Dima Kozakov, Gwo Yu Chuang, Dmitri Beglov, David Hall, Melissa R. Landon, Carla Mattos, and Sandor Vajda. 2009. "Fragment-Based Identification of Druggable 'hot Spots' of Proteins Using Fourier Domain Correlation Techniques." *Bioinformatics* 25 (5): 621–27. doi.org/10.1093/bioinformatics/btp036.

Brooks, Bernard R., Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. 1983. "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations." *Journal of Computational Chemistry*. doi.org/10.1002/jcc.540040211.

Brown, David, and Giulio Superti-Furga. 2003. "Rediscovering the Sweet Spot in Drug Discovery." *Drug Discovery Today*. doi.org/10.1016/S1359-6446(03)02902-7.

Brüser, Antje, Jürgen Kirchberger, Marco Kloos, Norbert Sträter, and Torsten Schöneberg. 2012. "Functional Linkage of Adenine Nucleotide Binding Sites in Mammalian Muscle 6-Phosphofructokinase." *Journal of Biological Chemistry*. doi.org/10.1074/jbc.M112.347153.

Campa, Juan S., Robin J. McAnulty, and Geoffrey J. Laurent. 1990. "Application of High-Pressure Liquid Chromatography to Studies of Collagen Production by Isolated Cells in Culture." *Analytical Biochemistry* 186 (2): 257–63. doi.org/10.1016/0003-2697(90)90076-L.

Christopoulos, A, T May, A Avlani, and M Sexton. 2004. “GPCR Allosterism and Accessory Proteins Biochemical Society Focused Meeting GPCR Allosterism and Accessory Proteins : New Insights into Drug Discovery.” *Biochemical Society Transactions* 32 (July): 873–77.

Conn, P Jeffrey, Arthur Christopoulos, and Craig W Lindsley. 2009. “Allosteric Modulators of GPCRs: A Novel Approach for the Nat Rev Drug Discov . 2009 January ; 8(1): 41–54. Doi:10.1038/Nrd2760. Treatment of CNS Disorders.” *Nat Rev Drug Discov.* 8 (1): 41–54. doi.org/10.1038/nrd2760.Allosteric.

Doruker, Pemra, Ali Rana Atilgan, and Ivet Bahar. 2000. “Dynamics of Proteins Predicted by Molecular Simulations and Analytical Approaches: Application to α-Amylase Inhibitor.” *Proteins: Structure, Function and Genetics*. doi.org/10.1002/1097-0134(20000815)40:3<512::AID-PROT180>3.0.CO;2-M.

Fährrolfes, Rainer, Stefan Bietz, Florian Flachsenberg, Agnes Meyder, Eva Nittinger, Thomas Otto, Andrea Volkamer, and Matthias Rarey. 2017. “Proteins Plus: A Web Portal for Structure Analysis of Macromolecules.” *Nucleic Acids Research*. doi.org/10.1093/nar/gkx333.

Friesner, Richard A., Jay L. Banks, Robert B. Murphy, Thomas A. Halgren, Jasna J. Klicic, Daniel T. Mainz, Matthew P. Repasky, et al. 2004. “Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy.” *Journal of Medicinal Chemistry*. doi.org/10.1021/jm0306430.

Goodsell, David S., and Arthur J. Olson. 1990. “Automated Docking of Substrates to Proteins by Simulated Annealing.” *Proteins: Structure, Function, and Bioinformatics*. doi.org/10.1002/prot.340080302.

Gunasekaran, K., Buyong Ma, and Ruth Nussinov. 2004. “Is Allostery an Intrinsic Property of All Dynamic Proteins?” *Proteins: Structure, Function and Genetics*. doi.org/10.1002/prot.20232.

Hajduk, Philip J., Jeffrey R. Huth, and Stephen W. Fesik. 2005. “Druggability Indices for Protein Targets Derived from NMr-Based Screening Data.” *Journal of Medicinal Chemistry* 48 (7): 2518–25. doi.org/10.1021/jm049131r.

Holland, J H. 1975. *Adaptation in Natural and Artificial Systems : An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. *Ann Arbor University of Michigan Press 1975*.

Irwin, John J., Teague Sterling, Michael M. Mysinger, Erin S. Bolstad, and Ryan G. Coleman. 2012. “ZINC: A Free Tool to Discover Chemistry for Biology.” *Journal of Chemical Information and Modeling*. doi.org/10.1021/ci3001277.

Jones, Gareth, Peter Willett, Robert C. Glen, Andrew R. Leach, and Robin Taylor. 1997. “Development and Validation of a Genetic Algorithm for Flexible Docking.” *Journal of Molecular Biology*. doi.org/10.1006/jmbi.1996.0897.

Kloos, Marco, Antje Brüser, Jürgen Kirchberger, Torsten Schöneberg, and Norbert Sträter. 2015. “Crystal Structure of Human Platelet Phosphofructokinase-1 Locked in an Activated Conformation.” *Biochemical Journal* 469 (3): 421–32. doi.org/10.1042/BJ20150251.

Koshland, D. E., J. G. Nemethy, and D. Filmer. 1966. “Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits.” *Biochemistry*. doi.org/10.1021/bi00865a047.

Koshland, Daniel E., and Kambiz Hamadani. 2002. “Proteomics and Models for Enzyme Cooperativity.” *Journal of Biological Chemistry*.doi.org/10.1074/jsbc.R200014200.

Kozakov, Dima, Laurie E. Grove, David R. Hall, Tanggis Bohnuud, Scott E. Mottarella, Lingqi Luo, Bing Xia, Dmitri Beglov, and Sandor Vajda. 2015. “The FTMap Family of Web Servers for Determining and Characterizing Ligand-Binding Hot Spots of Proteins.” *Nature Protocols*. doi.org/10.1038/nprot.2015.043.

Kurkcuoglu, Zeynep, Doga Findik, Ebru Demet Akten, and Pemra Doruker. 2015. “How an Inhibitor Bound to Subunit Interface Alters Triosephosphate Isomerase Dynamics.” *Biophysical Journal*. doi.org/10.1016/j.bpj.2015.06.031.

Madeira, Fábio, Young mi Park, Joon Lee, Nicola Buso, Tamer Gur, Nandana Madhusoodanan, Prasad Basutkar, et al. 2019. “The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019.” *Nucleic Acids Research*. doi.org/10.1093/nar/gkz268.

Martinez-Oyanedel, José, Iain W. McNae, Matthew W. Nowicki, Jeffrey W. Keillor, P. A M Michels, Linda A. Fothergill-Gilmore, and Malcolm D. Walkinshaw. 2007. “The First Crystal Structure of Phosphofructokinase from a Eukaryote: Trypanosoma Brucei.” *Journal of Molecular Biology* 366 (4): 1185–98. doi.org/10.1016/j.jmb.2006.10.019.

May, Lauren T, and Arthur Christopoulos. 2003. “Allosteric Modulators of G-Protein-Coupled Receptors.” *Current Opinion in Pharmacology* 3 (5): 551–56. doi.org/10.1016/S1471-4892(03)00107-3.

McNae, Iain W., José Martinez-Oyanedel, Jeffrey W. Keillor, P. A M Michels, Linda A. Fothergill-Gilmore, and Malcolm D. Walkinshaw. 2009. “The Crystal Structure of ATP-Bound Phosphofructokinase from Trypanosoma Brucei Reveals Conformational Transitions Different from Those of Other Phosphofructokinases.” *Journal of Molecular Biology* 385 (5): 1519–33. doi.org/10.1016/j.jmb.2008.11.047.

Meyerhof, O., and R. Junowicz-Kocholaty. 1943. “The Equilibria of Isomerase and Aldolase, and the Problem of the Phosphorylation of Glyceraldehyde Phosphate.” *Journal of Biological Chemistry*.

Mitchell, Melanie. 1996. *An Introduction to Genetic Algorithms*. MIT Press Cambridge, MA, USA.

MONOD, J., and F. JACOB. 1961. “Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation.” *Cold Spring Harbor Symposia on Quantitative Biology*. doi.org/10.1101/sqb.1961.026.01.048.

Monod, Jacque, Jeffries Wyman, and Jean Pierre Changeux. 1965. “On the Nature of Allosteric Transitions: A Plausible Model.” *Journal of Molecular Biology*. doi.org/10.1016/S0022-2836(65)80285-6.

Monod, Jacques, Jean Pierre Changeux, and François Jacob. 1963. “Allosteric Proteins and Cellular Control Systems.” *Journal of Molecular Biology* 6 (4): 306–29. doi.org/10.1016/S0022-2836(63)80091-1.

Nagarajan, Shanthi, Dimitrios A. Skoufias, Frank Kozielski, and Ae Nim Pae. 2012. “Receptor-Ligand Interaction-Based Virtual Screening for Novel Eg5/Kinesin Spindle Protein Inhibitors.” *Journal of Medicinal Chemistry* 55 (6): 2561–73. doi.org/10.1021/jm201290v.

Needleman, Saul B., and Christian D. Wunsch. 1970. “A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins.” *Journal of Molecular Biology*. doi.org/10.1016/0022-2836(70)90057-4.

Perutz, M. F. 1989. “Mechanisms of Cooperativity and Allosteric Regulation in Proteins.” *Quarterly Reviews of Biophysics* 22 (2): 139–237. doi.org/10.1017/S0033583500003826.

Rester, U. 2008. “From Virtuality to Reality - Virtual Screening in Lead Discovery and Lead Optimization: A Medicinal Chemistry Perspective.” *Current Opinion in Drug Discovery and Development*.

Romano, A. H., and T. Conway. 1996. “Evolution of Carbohydrate Metabolic Pathways.” In *Research in Microbiology*. https://doi.org/10.1016/0923-2508(96)83998-2.

Rose, George D., Ari R. Geselowitz, Glenn J. Lesser, Richard H. Lee, and Micheal H. Zehfus. 1985. “Hydrophobicity of Amino Acid Residues in Globular Proteins.” *Science*. doi.org/10.1126/science.4023714.

Schrödinger, LLC. 2015. “The PyMol Molecular Graphics System, Versión 1.8.” *Thomas Holder*. doi.org/10.1007/s13398-014-0173-7.2.

Shirakihara, Yasuo, and Philip R. Evans. 1988. “Crystal Structure of the Complex of Phosphofructokinase from Escherichia Coli with Its Reaction Products.” *Journal of Molecular Biology*. doi.org/10.1016/0022-2836(88)90056-3.

Tian, Tian, Chengliang Wang, Minhao Wu, Xuan Zhang, and Jianye Zang. 2018. “Structural Insights into the Regulation of Staphylococcus Aureus Phosphofructokinase by Tetramer-Dimer Conversion.” Research-article. *Biochemistry* 57 (29): 4252–62. doi.org/10.1021/acs.biochem.8b00028.

Trion, M. M., ben-Avraham, D. 2015. "Atomistic torsional modal analysis for high-resolution proteins." *Physical Review E*. doi.org/10.1103/PhysRevE.91.032712.

Tien, Matthew Z., Austin G. Meyer, Dariya K. Sydykova, Stephanie J. Spielman, and Claus O. Wilke. 2013. “Maximum Allowed Solvent Accessibilites of Residues in Proteins.” *PLoS ONE*. doi.org/10.1371/journal.pone.0080635.

Volkamer, Andrea, Axel Griewel, Thomas Grombacher, and Matthias Rarey. 2010. “Analyzing the Topology of Active Sites: On the Prediction of Pockets and Subpockets.” *Journal of Chemical Information and Modeling*. doi.org/10.1021/ci100241y.

Volkamer, Andrea, Daniel Kuhn, Thomas Grombacher, Friedrich Rippmann, and Matthias Rarey. 2012. “Combining Global and Local Measures for Structure-Based Druggability Predictions.” In *Journal of Chemical Information and Modeling*. doi.org/10.1021/ci200454v.

Volkamer, Andrea, Daniel Kuhn, Friedrich Rippmann, and Matthias Rarey. 2012. “Dogsitescorer: A Web Server for Automatic Binding Site Prediction, Analysis and Druggability Assessment.” *Bioinformatics*. doi.org/10.1093/bioinformatics/bts310.

Wakefield, Amanda E., Jonathan S. Mason, Sandor Vajda, and György M. Keserű. 2019. “Analysis of Tractable Allosteric Sites in G Protein-Coupled Receptors.” *Scientific Reports* 9 (1): 1–14. doi.org/10.1038/s41598-019-42618-8.

Wass, Mark N., Lawrence A. Kelley, and Michael J E Sternberg. 2010. “3DLigandSite: Predicting Ligand-Binding Sites Using Similar Structures.” *Nucleic Acids Research*. doi.org/10.1093/nar/gkq406.

Webb, Bradley A., Farhad Forouhar, Fu En Szu, Jayaraman Seetharaman, Liang Tong, and Diane L. Barber. 2015. “Structures of Human Phosphofructokinase-1 and Atomic Basis of Cancer-Associated Mutations.” *Nature* 523 (7558): 111–14. doi.org/10.1038/nature14405.

Yu, Jian, Yong Zhou, Isao Tanaka, and Min Yao. 2009. “Roll: A New Algorithm for the Detection of Protein Pockets and Cavities with a Rolling Probe Sphere.” *Bioinformatics*. doi.org/10.1093/bioinformatics/btp599.

CURRICULUM VITAE

Personal Information

Name Surname : Merve AYYILDIZ

Place and Date of Birth : Diyarbakır-Turkey, 07.07.1993

Education

Undergraduate Education : Nevzat Ayaz Anatolian High School

Graduate Education : Inonu University – Molecular Biology and Genetics

Foreign Language Skills : English (Professional Working), German (Basic)

Contact:

Telephone: +90 (543) 355 8813

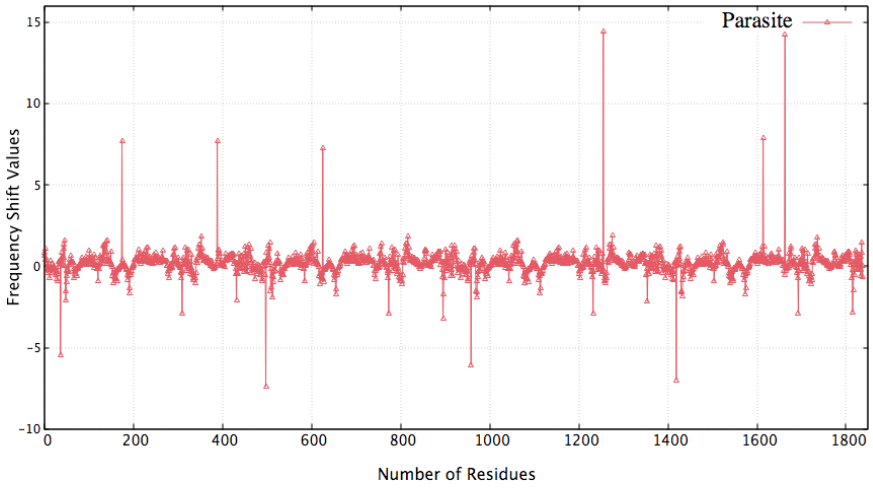
E-mail Address: ayyildizmeri@gmail.com / 20171109035@stu.khas.edu.tr

APPENDIX A

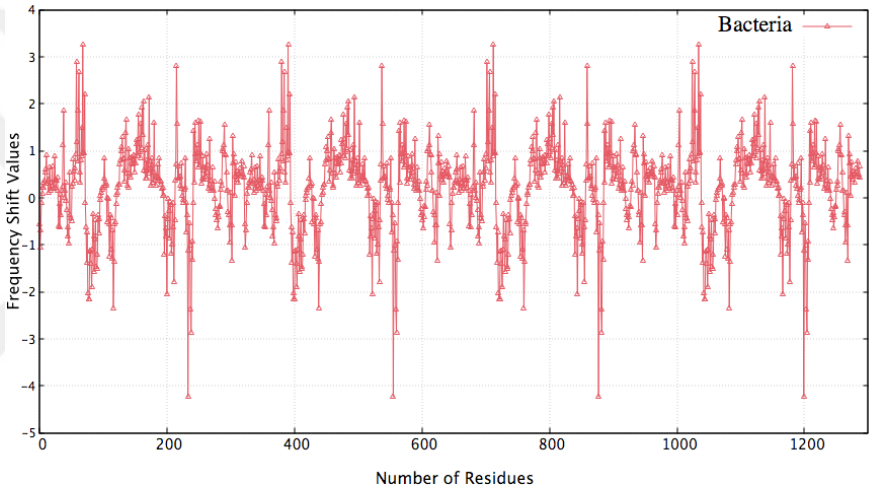
A.1 Elastic Network Model calculations can cause outlier residues.

As a result of elastic network model-based analysis, the frequency shift values for each residue vary within a certain range (See Result 3.2). This frequency shift values were calculated as the average of 20 slowest modes for each residue. This calculation was done for bacteria, parasite, human tetrameric structures as well as human dimer structure. However, some residues in parasites and human enzymes were found to have very different values than the average, which were called as outlier residues. There outlier residues make up only 1% of the entire enzyme in parasite (19 outliers out of 1839 total residues) (Figure A.1a). These outlier residues correspond only 0.3% in human dimer structure which corresponding tetramer structures of bacteria/parasite tetramer structures (5 outliers out of 1462 total residues (Figure A.1b). However, there was no outlier residue in bacterial PFK (Figure A.1c).

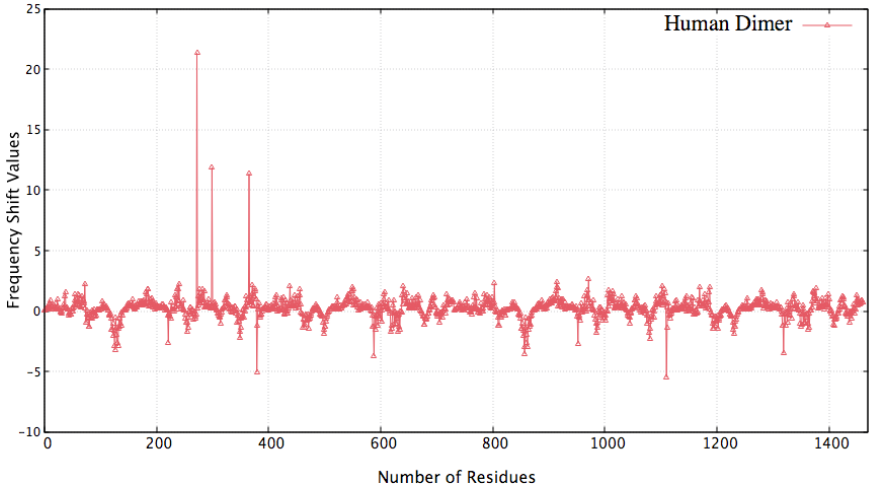
Additionally, when human tetramer structure was investigated by same method, it was seen that number of outlier residues significantly increased (50 outlier residues). The ratio of outliers to the entire structure also increased (1.71% of 2924 total residues) (Figure A.2a). In ENM-based calculations, each residue was considered as node and the interactions of nodes with each other were investigated. An interaction distance cutoff value (Rc) was considered as 15 Å as default, but can be increased according to the size of the receptor. (Trion and Avraham, 2015). For this reason, we increased the Rc value from 15 Å to 20 Å due to the large size of human tetramer PFK. In the end, number of outlier residues decreased from 50 to 12 as expected (Figure A.2b).



(a)

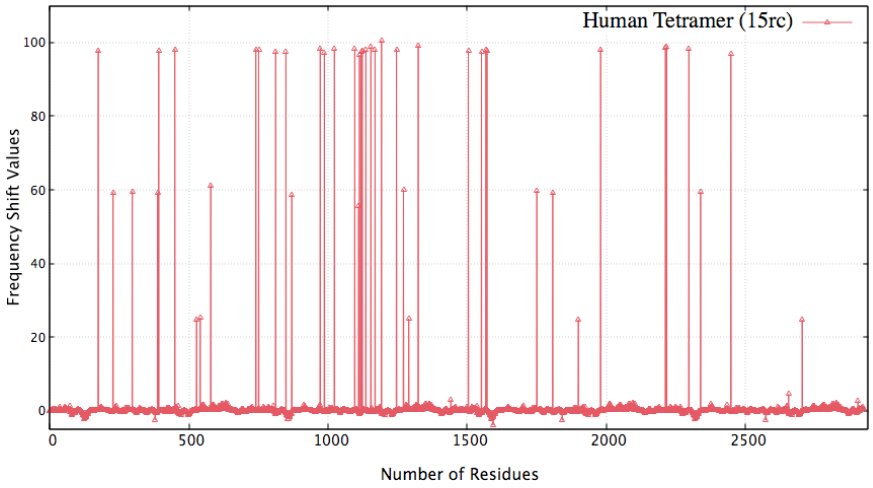


(b)

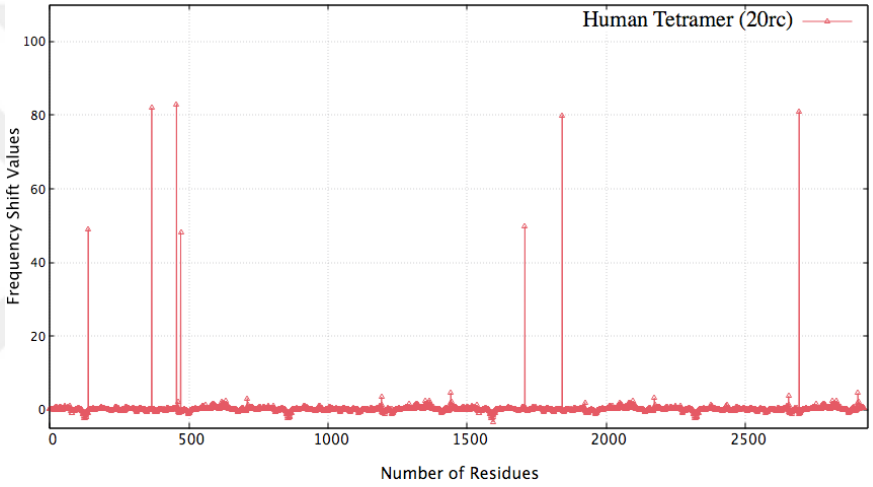


(c)

Figure A.1 Outlier residues after ENM-based residue scanning in a) *Tb*PFK, b) *Sa*PFK and c) *h*PFK dimer.



(a)



(b)

Figure A.2 Outlier residues after ENM-based residue scanning via using Rc = 15 Å and Rc = 20 Å values in a) *h*PFK tetramer, b) *h*PFK tetramer, respectively.