DETERMINATION OF SPECIES-SPECIFIC ALLOSTERIC BINDING SITES IN PYRUVATE KINASE AND ITS USE IN DRUG DESIGN STUDIES

by

Mehmet Fatih Özhelvacı

B.S., Molecular Biology and Genetics, İnönü University, 2016

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

Graduate Program in Computational Science and Engineering

Boğaziçi University

2020

ACKNOWLEDGEMENTS

First and foremost, I would like to express my greatest appreciation to my thesis co-supervisor Prof. E. Demet Akten Akdoğan for the sensitivity, the effort, and the courage that she has shown me since the beginning of my thesis studies. I would not have proceeded that much without her guidance and could not stand up again when I lost my hope.

Besides, I am very grateful to my former supervisor Prof. Pemra Doruker, who has helped me from the other side of the world and has been my thesis advisor before her retirement.

I am grateful to my thesis supervisor Prof. Levent Kurnaz for carrying out the process without any problems. I would also like to thank Assoc.Prof. Mustafa Mert Sözen for helping with all the issues related to the thesis.

I would like to thank my group members participating in the study. I am especially grateful to Serkan Çeliker and Merve Ayyıldız who stayed up all night with me and helped me in any condition throughout the study. I would also like to express my thanks to Ahmet Vatansever, Selin Pekel, Ezgi Çaylı, and Ayberk Türkyılmaz who cheered me up and supported me in this difficult process when I did not feel very well about myself.

Last but not least, I would like to thank my family whose support and efforts cannot be expressed with words and my lovely niece who makes me the happiest person in the world with his smile.

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

To My Parents ...

ABSTRACT

DETERMINATION OF SPECIES-SPECIFIC ALLOSTERIC BINDING SITES IN PYRUVATE KINASE AND ITS USE IN DRUG DESIGN STUDIES

Knowledge of structural and sequential differences in allosteric regions and conformational changes occurring in catalytic sites as a result of allosteric coupling had a great impact in the design of species-specific allosteric inhibitors. It is the most critical strategy to design effective drugs by targeting the disease-causing organism only without harming the infected organism. Glycolytic pathway is the most essential metabolic pathway in almost all living organisms which converts glucose into pyruvate to produce energy, thus it is often targeted in drug design studies. The objective of this study was to develop a new approach to identify all possible species-specific allosteric binding sites that will inhibit the activity of critical allosteric enzyme in the glycolytic pathway which was Pyruvate Kinase (PK) . The ultimate goal was to propose drug molecules that will bind to the proposed allosteric sites and inhibit the activity of bacterial/parasitic PK without effecting human PK.

In addition to sequence and structural comparisons, the computational approach consists of several steps; computational solvent-mapping to identify all possible ligand binding sites, identification of interface regions, elastic network modeling (ENM) to predict the regions which would have the highest effect on global or essential dynamics upon ligand binding; screening approved drug molecules via docking studies using GOLD tool to propose possible drug candidates. Accordingly, several FDA approved and world - wide used molecules were identified that bind to bacterial PK with high binding affinities without much effect on human PK.

ÖZET

PİRÜVAT KİNAZ ENZİMİNDE TÜRE ÖZGÜ ALLOSTERİK BAĞLANMA BÖLGELERİNİN BELİRLENMESİ VE İLAÇ TASARIMI ÇALIŞMALARINDA KULLANIMI

Allosterik bölgelerdeki yapısal ve sıralı farklılıklar ve allosterik eşleşmenin bir sonucu olarak katalitik bölgelerde meydana gelen konformasyonel değişikliklere ilişkin bilgiler, türe özgü allosterik inhibitörlerin tasarımında büyük bir etkiye sahip olmuştur. Enfekte organizmaya zarar vermeden yalnızca hastalığa neden olan organizmayı hedeflemek, etkili ilaçlar tasarlamanın en kritik stratejisidir. Glikolitik yolak, hemen hemen tüm canlı organizmalarda bulunan ve enerji üretmek için glikozu piruvata dönüştüren en temel metabolik yolaktır, bu nedenle genellikle ilaç tasarım çalışmalarında hedeflenmektedir. Bu çalışmanın amacı, glikolitik yolağın kritik allosterik enzimlerinden biri olan Piruvat Kinaz' ın (PK) aktivitesini inhibe edecek olası tüm türe özgü allosterik bağlanma yerlerini tanımlamak için yeni bir yaklaşım geliştirmektir. Nihai amaç, önerilen allosterik bölgelere bağlanacak ve insan PK'sini etkilemeden bakteriyel/parazitik PK aktivitesini inhibe edecek ilaç molekülleri önermektir.

Dizi ve yapısal karşılaştırmalara ek olarak, hesaplamalı yaklaşım birkaç adımdan oluşmaktadır. Bunlar; tüm olası ligand bağlanma bölgelerinin tanımlanması amacıyla hesaplamalı çözücü haritalaması, arayüz bölgelerinin tanımlanması, ligand bağlanmasına ilişkin küresel veya temel dinamikler üzerinde en fazla etkiye sahip olan bölgelerin tahmin edilmesi amacıyla elastik ağ yapı modeli (ENM), olası ilaç adaylarını önermek için GOLD programı kullanılarak kenetlenme çalışmaları yoluyla onaylanmış ilaç moleküllerinin taranmasıdır. Bu doğrultuda, insan PK'si üzerinde çok fazla etki olmaksızın bakteriyel PK'ye yüksek bağlanma afiniteleri ile bağlanan FDA onaylı ve dünya çapında kullanılan birkaç molekül tanımlanmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

α	Alpha
β	Beta
Δ	Delta
λ	Lambda
μ	Mu
σ	Sigma
Å	Angstrom
%s	Average Frequency Shift Value
μ ΔG	Average of ΔG values
Cα	Carbon Alpha
R_c	Cutoff value
ΔS	Differences of Score Values
λ_i	Eigenvalue of Mode i
${\lambda_i}^0$	Eigenvalue of Reference Mode i
ΔG	Gibbs Free Energy
$\%s_i$	Shift for Mode i
σ ΔG	Standard Deviation of ΔG values

LIST OF ACRONYMS/ABBREVIATIONS

ADP	Adenosine dihosphate
ASA	Acsessible Surface Area
ATP	Adenosine three phosphate
C.17	Consensus Site 17
CS	Consensus Site
DS	Druggable Site
E	Enzyme
ENM	Elastic Network Model
FBP	Fructose-1,6-bisphosphate
FDA	Food and Drug Administration
FS	Frequency Shift
GA	Genetic Algorithm
GLY	Glycine
GTP	Guanosine triphosphate
*h*PK	*H. sapiens* Pyruvate Kinase
IS-130	N'-[(1E)-1-(1H-benzimidazol-2-yl)ethylidene] -5-bromo-2-hydroxybenzohydrazide
K^+	Potassium ion
kDa	Kilodalton
*Lm*PK	*L. mexicana* Pyruvate Kinase
MD	Molecular Dynamics
Mg^{2+}	Magnesium ion
MOL.	Molecule
MRSA	Methicillin Resistant *Staphylococcus aureus*
NZT	N-(4-[4-(pyrazin-2-yl)piperazin-1 yl]carbonylphenyl) (quinoline-8-sulfonamide)
OXL	Oxalate
PEP	Phosphoenolpyruvate
PDB	Protein Data Bank

PK	Pyruvate Kinase
PKL	Liver isoform of Pyruvate Kinase
PKM1	M1 isoform of Pyruvate Kinase
PKM2	M2 isoform of Pyruvate Kinase
PKR	Red blood cell isoform of Pyruvate Kinase
PO_4^{3-}	Phosphate ion
PRO	Proposed Allosteric Region
RMSD	Root Mean Square Deviation
RSA	Relative Solvent Accessibility
rSASA	Relative Solvent Accessible Surface Area
S	ChemPLP Docking Score
*Sa*PK	*S. aureus* Pyruvate Kinase
SO_4^{2-}	Sulfate ion

1. INTRODUCTION

1.1. Glycolysis and Glycolytic Pathway

Glucose is converted into pyruvate to produce energy through glycolysis which is a metabolic process; however, it is also essential for the design of species-specific drugs. This process followed by one of two different conditions, which are described as cellular respiration and fermentation. Cellular respiration occurs in an environment with high oxygen concentration; however, fermentation takes place in the environments lack of oxygen to produce ATP. As shown in Figure 1.1, the glycolytic pathway has ten steps in which ten distinctive enzymes have separate roles in producing energy. Three of all these enzymes are allosteric enzymes, which are phosphofructokinase (used on the third stage), glyceraldehyde 3 phosphate dehydrogenase (used on the sixth stage), and pyruvate kinase (on the last step of glycolysis). Pyruvate kinase (PK) is a significant enzyme that plays a significant role in controlling metabolic flux from fructose-1,6-bisphosphate (FBP) to pyruvate (Mattevi *et al.*, 1996). PK is a fundamental enzyme for not only the glycolytic pathway but also cellular mechanism and it catalyzes the movement of phosphates from phosphoenolpyruvate (PEP) to ADP to produce pyruvate and ATP (Zhong *et al.*, 2017). Both substrates and activators of PK have significant roles in many biosynthetic pathways. Different types of X-ray crystal structures of PK according to various species (e.g., *Escherichia coli*, *Leishmania mexicana*, *Staphylococcus aureus*, cat, rabbit muscle, human erythrocyte, and yeast) demonstrate that they have conserved structures (Zoraghi *et al.*, 2011).

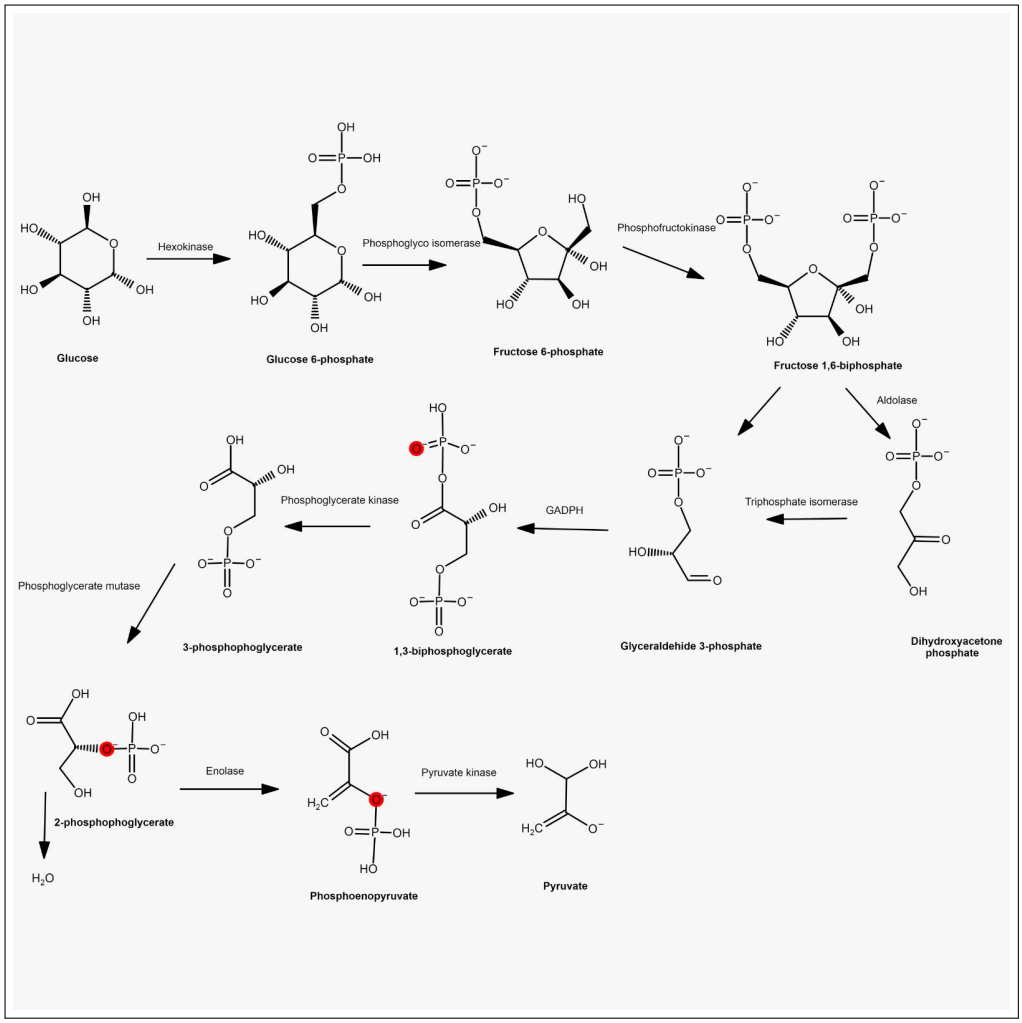


Figure 1.1. Ten reactions catalyzed by ten enzymes in glycolytic pathway.

Glycolysis is the only energy source for parasites and bacteria such as *S.aureus*, *T.brucei*, *L.mexicana*, *T.cruzi*, and it locates itself in the glycosome; therefore, the enzymes taking part in the glycolytic pathway are significantly important in terms of drug target. These parasites cause several diseases. For instance, *S.aureus* gives rise to toxic shock syndrome by contributing to the production of superantigens in the bloodstream and long term seated infections such as endocarditis and osteomyelitis (Foster, 1996). *L.mexicana* causes leishmaniasis, which is a result of the transmission of intracellular parasites into the human body by a simple bite of a sandfly (Torres-Guerrero *et al.*, 2017).

Different allosteric mechanisms are observed in the pyruvate kinase of humans, parasites, and bacteria. The protein, in the tense state, is a monomer and becomes a tetramer in a relaxed state in humans. During the transition from tense state to a relaxed state, three additional domains of bacteria rotate and create salt bridges. The formation of salt bridges is stabilized through the rotation of A- and C- domain cores in parasites as seen in Figure 1.2 (Morgan *et al.*, 2014).

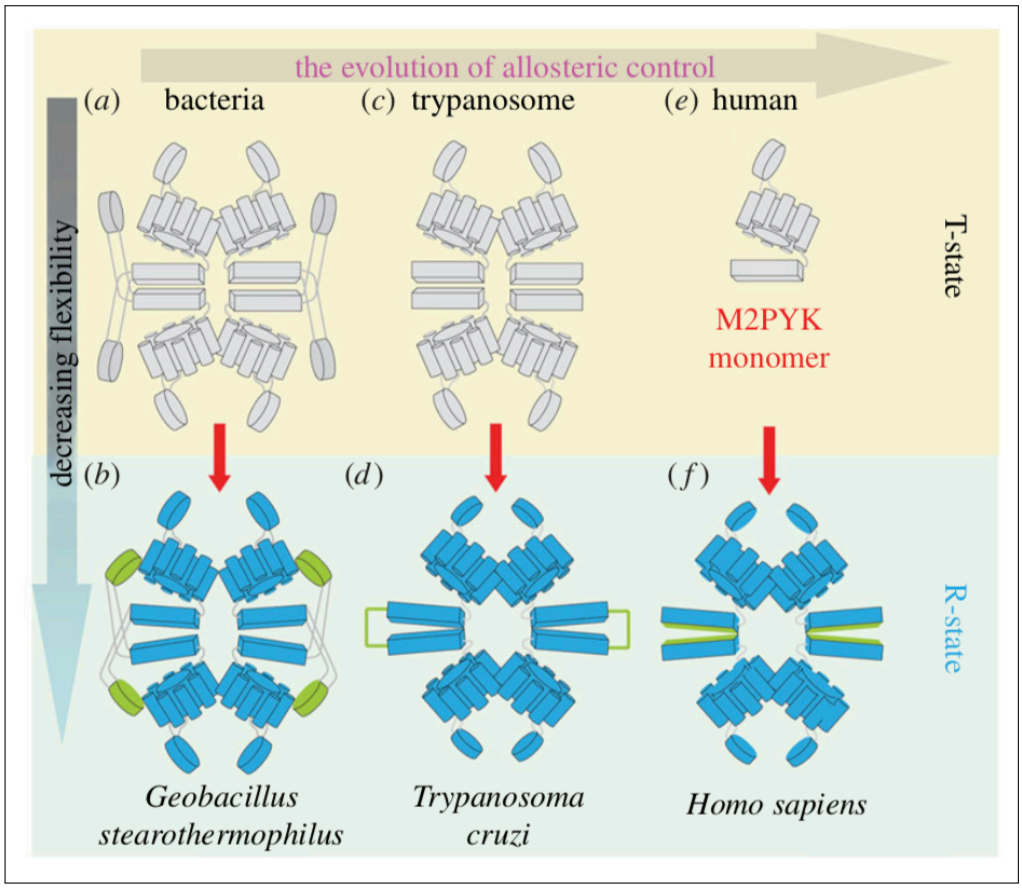


Figure 1.2. Evolutionary process of PK in terms of allosteric control. (a) T-state representation of bacteria, (b) The active state of *Geobacillus stearothermophilus*, (c) and (d) stabilization between T and R-state of parasitic PK, (e) The inactive form of human monomeric PK and (f) R state representation of *H.sapiens* (Morgan *et al.*, 2014).

Allosteric enzymes have the capability to recognize several peculiar conditions thanks to their effector binding sites. Each enzyme has active sites - which is a location of an enzyme where substrate molecules bind and pass on a chemical reaction - apart from this situation, allosteric enzymes have allosteric sites too. Allosteric sites are parts of allosteric enzymes that enhance the activity of enzymes through the binding of non-substrate molecules (Mattevi *et al.*, 1996).

Most currently contemplated drugs aim to induce transient changes in the function of biomolecular structures, such as enzymes or receptors, by targeting regions defined as the active site (also known as orthosteric) to which the substrate binds directly. This ensures that many drugs have high affinity and specificity for targeting a specific disease or disorder. However, this situation has some disadvantages. It is the similarity and evolutionary protection of active sites of biomolecular structures such as enzymes or receptors with similar biological functions. Another disadvantage is that biomolecular structures may cause drugs to act as full activators or as full inhibitors.

Allostery can be expressed as a large-scale conformational change of the protein dynamics by inducing a region far from the active regions of the protein, called the allosteric region, by a ligand or another protein defined as an effector. Nowadays, the approach that takes into account the effects of the second binding region (allosteric region) in drug design has been gaining popularity. This approach does not consider the primary active regions of biomolecules, such as enzymes and receptors. However, rather it takes into account the macromolecular structure motions created by the second binding regions (allosteric regions) of biomolecules (Guo and Zhou, 2016). In this context, conformational change in the biomolecule (allosteric modulation) resulting from the successful binding of drugs is identified and has different effects on biomolecules such as enzymes or receptors. It occurs when a drug that binds to the allosteric site by increasing (positive modulation) or decreasing (negative modulation) the orthosteric reaction rate (Mattevi *et al.*, 1996). It is known that there are many examples of molecules that allosterically induce modulation in the target protein.

The identification of allosteric modulator regions on the target protein may not indicate that each region will be a suitable candidate for the drug. In this sense, the determination of structural and sequential differences in allosteric regions and the design of appropriate molecules for species-specific drug design has great importance in the formation of macromolecular conformation affecting orthosteric regions.

1.2. Structure of Pyruvate Kinase

The last step of glycolysis converts phosphoenolpyruvate into pyruvate with the help of pyruvate kinase, which is an allosteric enzyme. Catalysis yields a pyruvate and an ATP molecule. Pyruvate kinase has dimeric and tetrameric forms. It has both active (R) and inactive (T) states. In its R state conformation, which has a high binding affinity to PEP, it has a molecular weight of 228 kDa. It has a molecular weight as a tetramer consisting of four equal subunits of 57 kDa. On the other hand, T state conformation with low affinity to PEP has 114 kDa, and it is available as a dimer consisting of two equal subunits with a molecular weight of 57 kDa.

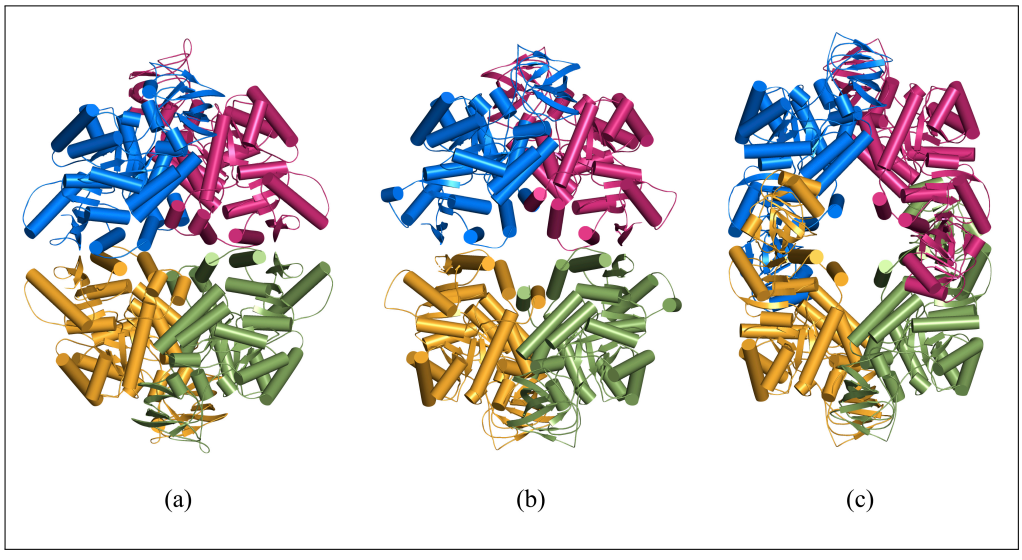


Figure 1.3. Structures of Pyruvate Kinase a)*h*PK (Kung *et al.*, 2012) b)*Lm*PK (Rigden *et al.*, 1999) c) *Sa*PK (Axerio-Cilies *et al.*, 2012).

PK has four mammalian isoforms (PKM1, PKM2, PKR, PKL), and each of these isoforms has multiple conformations (Israelsen and Vander Heiden, 2015). The regulation of PK is crucial for controlling levels of ATP, GTP, and glycolytic intermediates in the cell. Allosteric regulation of PK has a vital feature to control the cell cycle. PK shows activity in its tetrameric conformation, and the amino acid sequences are highly conserved among species. PK can be obtained by using lactate dehydrogenase, which is coupled pyruvate with NADH oxidation. It might also be noted that compounds having lactate dehydrogenase, chelate, Mg^{2+}, and K^{+} aggressively with Mg^{2+},K^{+}, ADP, and PEP prevent the activity. These mechanisms cannot be allosteric in nature (Fenton, 2012). On the other hand, fructose 1,6-bisphosphate (FBP) described as a significant allosteric activator of PK (Israelsen and Vander Heiden, 2015). PK structures for *H.sapiens*, *L.mexicana*, and *S.aureus* seen in Figure 1.3, respectively.

1.3. Pyruvate Kinase as A Drug Target

Computational approaches have critical beneficial effects that reduce economic expenses as well as the quantity of drug candidates used during drug discovery. Repurposing of a drug used for a specific disease means to apply its effects on different kind of diseases. Approved drugs may have undiscovered molecular effects along with discovered ones. Reducing the experimental and economic expense of the process are the pros of drug repurposing due to the experienced dosage use and found side effects. The time spent for clinical trials - phase I and Phase II - can be reckoned without performing when the correct amount of drug is used in drug repurposing (Oprea and Mestres, 2012).

The fact that a hub protein appeals to the drug target is an elementary feature of antibiotic-resistant strains. Thus a glycolytic enzyme called pyruvate kinase is used for the target protein. Pyruvate kinase enzyme maintains the simplification of the final step of glycolysis, which includes the transformation of ADP to ATP. Not only a hub protein appears in interconnecting with 243 proteins, but also it takes an integral part in the glycolysis process for the life of the cell. Consequently, Pyruvate kinase is essential for a drug target (Axerio-Cilies *et al.*, 2012).

B-domain including the active site, A-domain and C-domain consisting of the effector site are included in pyruvate kinase. The active site is attached to ATP-oxalate and the effector site is attached to Fructose 2,6 bisphosphate (F-2,6-BP). After the attachment of ATP and/or F-2,6-BP, a transition from the tense (inactive) state to relaxed (active) state occurs. It has been observed that a 7-fold increase the transition rate of *Leishmania mexicana* pyruvate kinase is caused by the effector binding. It is thought that the formation of the salt brigdes helps effector binding to stabilize the protein (Morgan *et al.*, 2010). With the help of these investigations, in this study, *H.sapiens*, *S.aureus*, and *L.mexicana* were determined in PK to examine the presence of potential allosteric sites for species-specific drug design.

2. MATERIALS AND METHODS

2.1. Data Preperation

In this thesis, three different tetrameric crystal structures of pyruvate kinase (PK) enzyme belonging to *H.sapiens*, *S.aureus*, and *L.mexicana* were examined. Protein Data Bank (Berman *et al.*, 2002) database was used to determine all enzymes belonging to all species. As illustrated in previous studies, binding affinity might vary among formations of protein structures. For instance, PK tetrameric structure has a high affinity for substrates more than dimeric and monomeric structures. By contrary, the dimeric formation of PK has a low affinity to PEP (Riefke *et al.*, 2007). In this term, homotetrameric structures of PK selected species in Table 2.1 have used in this study to prepare data. Pyruvate kinase enzyme has 36, 3, and 12 tetrameric crystal structures for *H.sapiens*, *S.aureus*, and *L.mexicana*, respectively. Firstly, similarities between structures were determined by using the PyMOL graphics visualization tool (Schrödinger, 2015) for the selection of one representative structure of each species. *Super* module of the PyMOL graphics visualization tool (Schrödinger, 2015) superposes the two structures based on the backbone positions of alpha carbon atoms regardless of their amino acid identity, and it calculates root-mean-square-deviation (RMSD) between them. Secondly, the RMSD values of the structures belong to the human, parasite, and bacteria were calculated and listed in a matrix. These calculations were done separately for 36 different structures of human, three different structures of bacteria, and 12 different structures of the parasite.

R is a language and environment used for statistical and graphical computing. R includes a broad spectrum of mathematical (e.g., linear and nonlinear modeling, traditional predictive experiments, time series analysis, grouping, clustering) and graphical methods, which is exceptionally extensible (Team, 2019). R programming language was used to generate "Cluster Dendrogram" and "Heatmap" for identifying differences among crystal structures (see Figure A.1, A.2, A.3, and A.4). Making use of these results, three different structures of Pyruvate kinase were selected from Protein Data

Bank (PDB) according to resolution, activity state, and ligand of enzymes listed in Table 2.2. The selection of the structures was carefully made to get the correct results.

Table 2.1. PDB IDs of Pyruvate Kinase from Protein Data Bank Database for three species.

Pyruvate Kinase	***H.sapiens***	***L.mexicana***	***S.aureus***
Monomer	5M4E 5M4H 5M4K 5M4M 5M4N	3HQP	N/A
Dimer	2BTZ 2BU2 2BU5 2BU6 2BU7 2BU8 2E0A 2Q8F 2Q8G 2Q8H 2ZDX 2ZDY 2ZKJ 3D2R 4MP2 4MP7 4MPC 4MPE 4MPN 4V25 4V26 5J6A 5J71 6B6U	3HQO	N/A
Tetramer	1T5A 1ZJH 2VGB 2VBF 2VBG 2VGI 3BJF 3BJT 3G2G 3GQY 3GR4 3H6O 3ME3 3SRD 3SRF 3SRH 3U2Z 4FXF 4FXJ 4G1N 4IMA 4IP7 4JPG 4QG6 4QG8 4QG9 4QGC 4RPP 4WJ8 4YJ5 5X0I 5X1V 5X1W 6GG4 6GG5 6GG6	1PKL 3E0V 3E0W 3HQN 3HQQ 3IS4 3KTX 3PP7 3QV6 3QV7 3QV8 3SRK	3T0T 3T05 3T07

In line with the information given above, crystal structures of pyruvate kinase enzyme of *Homo sapiens* (PDB id: 4g1n) (Kung *et al.*, 2012), crystal structures of pyruvate kinase enzyme of *Staphylococcus aureus* (PDB id: 3t0t) (Axerio-Cilies *et al.*, 2012), and crystal structures of pyruvate kinase enzyme of *Leishmania mexicana* (PDB id: 1pkl) (Rigden *et al.*, 1999) were investigated to reveal differences among species.

Table 2.2. Classifying of crystal structures of Pyruvate Kinase according to their ligands, solvents, metals/ions and activity states .

PDB NAME	**LIGANDS**	**SOLVENTS**	**METAL/IONS**	**STATES**
4g1n	NZT OXL	N/A	Mg^{2+}	R - STATE
1pkl	N/A	N/A	SO_4^{2-}	T - STATE
3t0t	IS-130	N/A	PO_4^{3-}	R - STATE

2.2. Sequence and Structural Alignment

Sequence and structural alignment procedures are essential parts of the study to compare species. In sequence alignment, the higher the sequence similarity is the greater the similarity between structures, and it shows conservation between species. Sequence alignment also shows the known and unknown regions of species. For instance, by looking at the similarities in a structures which have experimentally known active regions, we might learn the activity state of an unknown region. On the other hand, we can also predict allosteric regions of enzymes. In this study, for sequence alignment, EMBOSS-Needle webserver (Rice *et al.*, 2000) was used. EMBOSS-Needle webserver (Rice *et al.*, 2000) uses Needleman-Wunsch global alignment algorithm (Needleman and Wunsch, 1970), it tries to achieve the best global alignment, so it always considers entire input with different parameters. These parameters are; Blosum62 as matrix (Henikoff and Henikoff, 1992), 10 as gap open, 0.5 as gap extend, false as end gap penalty, 10 as end gap open, and 0.5 as end gap extend. This web server was used to identify differences between targeted regions in drug design.

It is crucial to know the structures of enzymes in order to understand the structural differences and dynamics of the enzyme. This approach might give us plenty of information about the differences between species. When the structures are aligned, the differences are visible. Structural alignment is more reliable than sequence alignment as different structures might reveal even if they have the sequence similarity. There are several types of web servers using different structural alignment algorithms. In this thesis, the PyMoL visualization program (Schrödinger, 2015) was used to indicate differences between species. PyMoL program (Schrödinger, 2015) has the "Super" module to align two selections. It does sequence-independent structural alignment; it also considers positions of backbone carbon alpha (CA) atoms regardless of their amino acid identity through refinement cycles (Ayyildiz *et al.*, 2020).

The result of sequence and structural alignments were colored via the Pymol program to show the dissimilarities clearly between species. Finally, the EsPript program (Robert and Gouet, 2014) was used for better illustration of sequence similarity.

2.3. Computational Solvent Mapping (CS-Map)

Computational Solvent Mapping is a method to predict potential binding sites of proteins (Brenke *et al.*, 2009). This method allows fast examination for potential binding sites and druggability of proteins. There are several various tools for identifying binding sites of proteins. Computational mapping strategies put molecular probes (small molecules or functional groups) on a protein surface to distinguish the leading favorable authoritative positions by calculating its potential (Goodford, 1985). Numerous methods have created for the prediction of ligand-binding sites. FTMap server frequently used in the computational solvent mapping. FTMap is one of the static mapping applications which used to detect hot spots of macromolecules. For instance, it can identify the areas of the surface with significant contributions to the ligand - binding free energy. FTMap employed a combination of a protein, DNA, or RNA structure which was used in PDB format. The probe, which consists of extreme amounts of tiny sampled organic molecules, is scored, with the help of energy expressions, by FTMap. FTMap uses 16 different organic molecules as probes, which are ethanol, isopropanol, isobutanol, acetone, acetaldehyde, dimethyl ether, cyclohexane, ethane, acetonitrile, urea, methylamine, phenol, benzaldehyde, benzene, acetamide, and N, N dimethylformamide. The region where several different probes are bound is called "consensus sites" (CSs), and the area with the most significant number of probes considered as the central hot spot; all other regions considered secondary hot spots (Kozakov *et al.*, 2015). When consensus sites gathered together in a specific area, seen in Figure 2.1, it is called "druggable site".

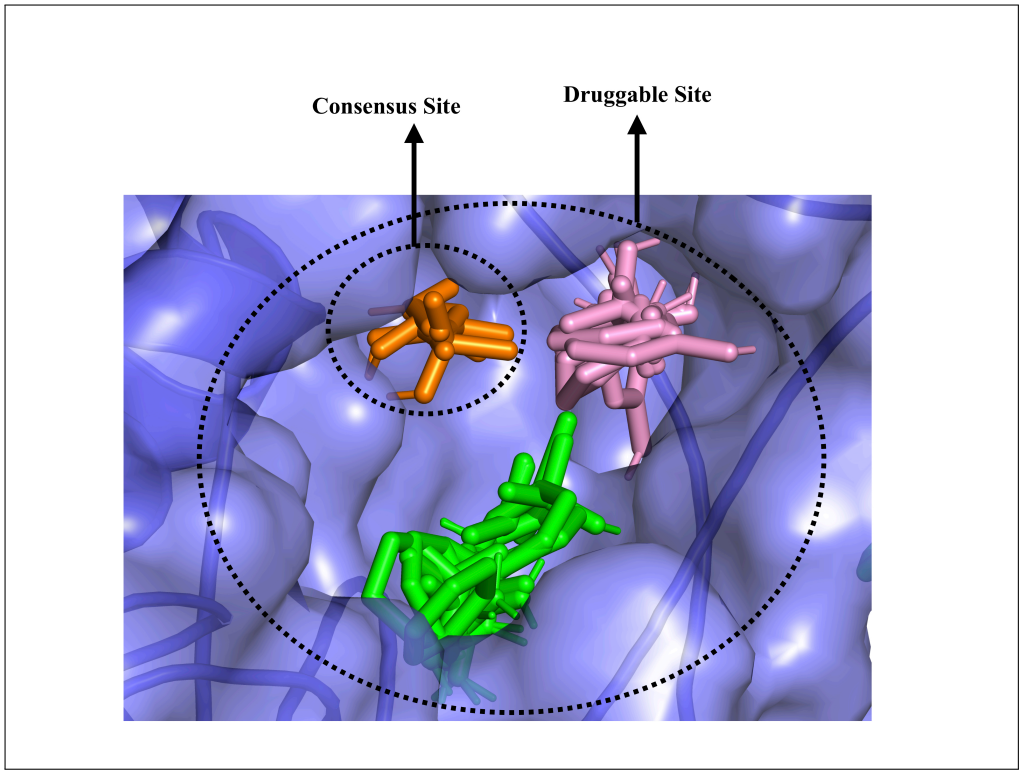


Figure 2.1. Representation of Consensus and Druggable Sites. The green, orange and pink colors are each called Consensus Site separately. When they come together, they form druggable site.

By using FTMap, the solvent – mapping process was applied to both the overall tetrameric structure and each chain separately, shown in Figure 2.2. When the entire structure mapped; some regions were inaccessible; however, when structures were given chain by chain, these regions were obtained that shows the purpose of this application. Alternative binding sites were obtained with the use of this approach. By using monomeric decomposition and mapping approach the drawback was alleviated. Taking into consideration that an X-ray structure in time only represents an instantaneous state of the receptor. If the clusters, which could emerge with this approach, were located at the interface of monomeric subunits, they would be inaccessible from outside, and for this reason, they should be eliminated (Ayyildiz *et al.*, 2020).

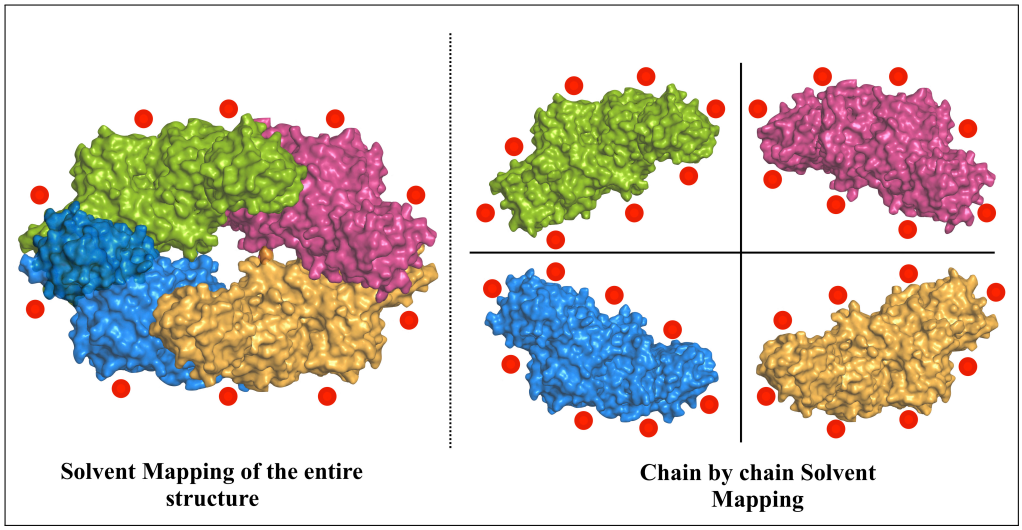


Figure 2.2. Representation of the mapping methods of FTMap. Differences between overall and chain by chain mapping.

2.4. Elastic Network Modeling (ENM) Based Residue Scanning

ENM is a method describing interactions between closely ($\leq$7.0 Å) positioned α-carbon atom pairs in folded proteins (Bahar *et al.*, 1997). The approach for this model is sufficient to model the Dynamics of folded proteins. The global dynamics of biomolecular structures estimate through this model, which is a powerful theoretical approach. Following that, the relationship between structure and the functional mechanism built through the model (Doruker *et al.*, 2000).

ENM provides an analytical solution for the normal mode spectrum, which can be accessed uniquely for each protein/complex. This solution was also used to extract soft modes (Zhang *et al.*, 2020). ENM was used to illuminate the structural movements of proteins. In this model, each amino acid assumed to be interconnected by a flexible spring. Generally, each amino acid represents α-carbon atoms; this can change under the desired detail (side chains or all atoms). As the distance of atoms increases, their interactions with each other decrease, and this model calculates the interactions of atoms in the whole structure according to this information. Protein acts as if it is beads

that connected through a spring. In this way, the dynamic structure of the protein tried to be precise (such as how it moves, channel opening and closing, allosteric regulation, protein-protein interactions).

In this study, heavy atoms that belong to each residue were taken into consideration. A residue-based cutoff distance (R_c) between nodes was determined as 20 Å to observe if α-carbon atoms contained each residue or specific residues of side-chain atoms, therefore the effect of the side chains were overlooked in this phase (Kurkcuoglu *et al.*, 2015). Then, for each scanning residue which belongs to the eigenvalue (λ_i) of i^{th} collective mode was calculated.

$$\%shift\ for\ mode\ i\ (\%s_i) = \frac{\lambda_i(modified) - \lambda_i(original)}{\lambda_i{}^0(original)}. \tag{2.1}$$

Because of the observation of 20 modes representing more than 90% of global dynamics of protein, the percentage shift of the calculation was adjusted among 20 modes for each one of the residues. With the examination of not only 15 Å, but also 20 Å cutoff distances were employed to predict differences between beta factors. As a result of the evaluations, it was observed that there was not a big difference between the β factors values obtained for two different R_cs. When two values were compared, 20 Å was used to eliminate the outliers obtained from 15 Å. A greater $R_c$is appropriate for the ANM compared to GNM (Atilgan *et al.*, 2001). The protein's X - ray structure was displayed average value (see Formula 2.2) that was maintained with the aid of color gradients.

$$(\%s) = \sum_{i=1}^{20}(\%s_i)/20 \tag{2.2}$$

H.sapiens, *S.aureus*, and *L.mexicana* have different PDB files, and each of them contains different coordinates, amino acids, and B-factors. The B-factors obtained as a result of the ENM calculation were replaced with the-B factors in the original file. Then, the color scale (blue-white-red from lower to higher, respectively) was adjusted using the PyMol (Schrödinger, 2015) program for a more detailed review. Percentage shift values obtained from each residue contribute to protein-ligand interaction

dynamics. When the calculation was implemented, all percentage shift values from large to small were taken into account, but minimal values were ignored and described as "outliers". Moreover, these outliers were not taken into consideration while coloring on protein. In order to compare the differences between the three species, the highest and the smallest percentage shift values of each were recorded, and the color scale was adjusted accordingly. In this way, the differences between the species were observed more clearly.

2.5. Merging FT-Map and ENM Based Residue Scanning Results

The cross-clusters obtained using FTMap were included in the calculation, and the residues closer to 5 Å to the protein were calculated. Then, as seen in Formula 2.3., the average of the frequency shift values was divided by the total number of residues. Since higher average shift frequency values were more critical for our method, those below a particular value was ignored. The threshold value was set at 25% for three species, and clusters below this value were eliminated. Several cross-clusters that do not affect the dynamics of the protein have been eliminated in this way and have led to a more detailed examination (Ayyildiz *et al.*, 2020).

$$(\%s_{\mathrm{i}}) = \sum_{i=1}^{n} (\%s_{\mathrm{i}})/n \tag{2.3}$$

2.6. Determination of Interface Regions with the Use of Relative Solvent Accessible Surface Area(rSASA)

Many studies show that the binding interface between protein is usually extensive so that it can carry out many intermolecular contacts (Clackson and Wells, 1995). Additionally, a degree of residue solvent exposure is demonstrated through the relative solvent accessible surface area (rSASA). The normalization of the total accessible surface area (ASA) of the residues in a protein structure provides calculation of rSASA utilizing the ASA of the residues at the highest level of exposure to a solvent molecule

(Rose *et al.*, 1985). Relatively accessible surface area or relative solvent accessibility (RSA) is a measure of a protein residue. Residue solvent exposure were calculated with the following formula:

$$RSA = \frac{ASA}{maxASA} \tag{2.4}$$

In this formula, ASA is a solvent accessible surface area, and maxASA is the highest possible solvent accessible surface area, so both ASA and maxASA were commonly measured. In order to calculate the relative solvent accessibility of the residue, ASA values were needed, the way to do that was to place the X residue (any residue to be controlled) between the Gly-X-Gly tripeptide and then the surface area around X was evaluated (Levy, 2010; Tien *et al.*, 2013).

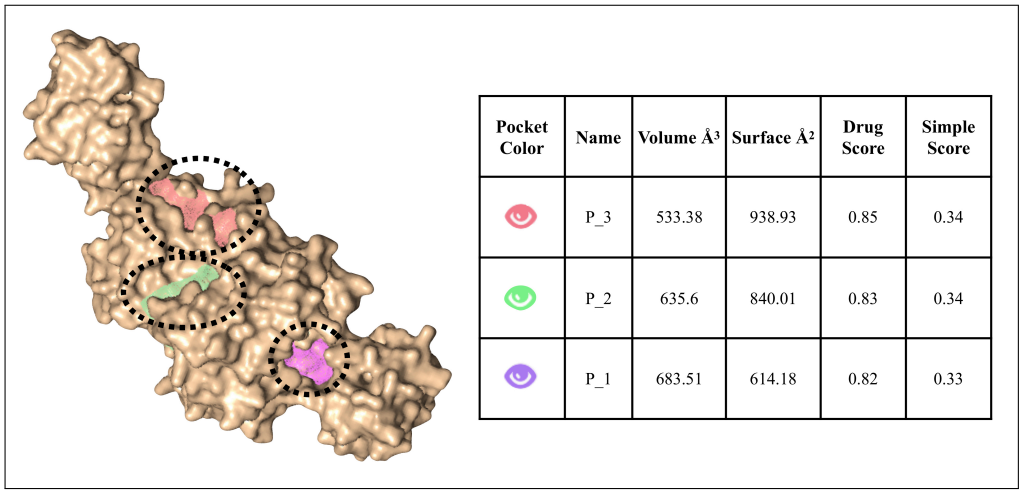
In this study, interface, surface, and internal regions on protein were determined using the calculations mentioned above. Surface and interface regions were found according to a determined threshold value, which is 25%. Accordingly, it was shown that regions with a threshold value less than 25% were in the interface, and regions with a threshold higher than 25% were on the surface. The interior region has been ignored to make the area on the surface more visible.

When the literature was observed, there were several programs and online tools that performed such calculations. In our study, a code was written to make these calculations. The β-factor values we obtained as a result of this code were replaced with the original β-factor values in PDB structures (4g1n, 1pkl, 3t0t), and each of them were colored with the help of the PyMol program according to β-factor values. In the coloring from blue to white, the blue color represented the area in the interface, and the white color represented the remaining regions.

2.7. Programs Used to Support the Method: DogSiteScorer and Allosigma

2.7.1. DogSiteScorer

Calculations for the druggability of proteins is a significant procedure in target-based drug design studies. Druggability plays an essential role in this study in the targeted design of the protein and how vital the binding pockets are. The DoGSiteScorer method estimates potential pockets on the protein surface by considering atomic coordinates of protein. Many geometric and physicochemical properties are automatically calculated for predictive pockets and other subpockets, and the volume and surfaces of these predictive pockets are calculated by counting the grid points that make up the pocket size, shape, or surface and multiplying this number by the grid box volume or surface, respectively. Machine learning algorithms also used to estimate the druggability of the protein. This tool gives values between 1 and 0. The pocket with the score closest to 1 is the pocket with the highest druggability effect on the protein (Volkamer *et al.*, 2012).



Pocket Color	Name	Volume Å³	Surface Å²	Drug Score	Simple Score
	P_3	533.38	938.93	0.85	0.34
	P_2	635.6	840.01	0.83	0.34
	P_1	683.51	614.18	0.82	0.33

Figure 2.3. Representation of possible druggable pockets and different scores ranked by "Drug Score" according to DogSiteScorer results.

This program sorts the pockets according to the scores, seen in Figure 2.3. The figure on the left shows potentially druggable areas in one chain of the homotetrameric protein. Each pocket seen in the figure represents a druggable area. Each pocket shown in the form of a mesh corresponds to one druggable region. On the other hand, the table on the right shows the different score values belong to these pockets. These pockets were compared with the results in the methods we used in ENM based residue scanning methods. Therefore, the significance of the results obtained from ENM based residue scanning were better understood with the comparison of DogSiteScorer results.

2.7.2. AlloSigMA

The allosteric free energies which act based on a single residue resulting from either ligand binding are estimated with a tool called the AlloSigMA server. The AlloSigMA server aims to provide a numerical tool for the analysis of allosteric communication energies by quickly estimating the effects of ligand binding, mutations, and their combinations on the allostery. Numerical investigation of allosteric effects facilitates experimental design for protein activity and provides a reasoned guideline in the selection of allosterically binding sites and mutations (Guarnera *et al.*, 2017). This server calculated deltaG (ΔG) free energy according to increased dynamics (positive ΔG, local destabilization) and decreased dynamics (negative ΔG, local stabilization), respectively.

Calculations were made using this program for the known allosteric and active site of the PK. Structural alignment procedure was applied for *Sa*PK and *Lm*PK enzymes by taking PEP (from PDB id: 4hyv) (Zhong *et al.*, 2013) and ADP (from PDB id: 6p0y) (Schormann *et al.*, 2019). Considering the PEP and ADP substrates in the allosteric and active region of the enzyme, various calculations were made in the light of the results obtained from this server. In the program, binding regions were determined, and the results from this server were replaced with the data in the original PDB file for each 3 species. Structures were colored according to the scale of the values as in the ENM based residue scanning methods according to β-factor values (blue-white-red from lower to higher, respectively).

2.8. Docking Procedure

Molecular docking algorithms are one of the essential methods used in structure - based drug design studies. The molecular docking approach helps us to understand the behavior of small molecules in the binding site of target proteins. It can also be used to model the atomic level interaction between a small molecule and protein, which allows us to illuminate basic biochemical processes (McConkey *et al.*, 2002). The docking procedure consists of two fundamental steps, which are the estimation of ligand conformation and evaluation of its position, direction, and binding affinity in the regions (Meng *et al.*, 2012).

There are many docking algorithms and programs were created using these algorithms. Some of these programs were AutoDock (Morris *et al.*, 1998), GOLD (Verdonk *et al.*, 2003), GLIDE (Friesner *et al.*, 2004), etc. This thesis was carried out by employing the GOLD program. The essential parts of the GOLD program are considered to be scoring function and algorithm. Along with these two main parts, locating the ligand in the binding site with a specific approach is another aspect of the program. In general, the object of these fitting points is to combine protein along with ligand structure in hydrogen bonding groups, which results in reorientation between acceptor and donor groups. This system forms a new degree of freedom in the binding sites that the importance of the binding energy values can be observed. This program includes three different scoring functions: ChemPLP, ChemScore, and Goldscore. The scoring function used is ChemPLP developed based on the force field, which gives extremely fast and precise results (Korb *et al.*, 2009). It is a widespread score function, especially since it has a high success rate for virtual screening. Because of the fast and precise prediction, the ChemPLP scoring function was used in this study. Moreover, in this program, some parameters were used during the docking application. A grid box containing 10 Å distance from the ligand-binding site was created, taking into account the coordinates. Detect cavity and allow early termination options were unselected. GOLD uses a Genetic Algorithm (GA) (Jones *et al.*, 1997), GA settings directly affect the timing of the placement process and the possibility of finding a global optimum. Therefore, GA runs were employed as 10 and slow GA search options.

Structure-based virtual screening is a crucial strategy for early drug discovery and development. Along with target and ligand - based approaches, virtual screening has emerged in the prioritizing of compounds for high throughput screening. It is also observed that lead compounds are used as an effective methodology in structure-guided optimization. Structure-based virtual screening enables fast and cost-effective evaluation of interactions between large compound libraries and biomolecular targets (Ma *et al.*, 2013). In this thesis, with the use of virtual screening, drug repositioning was employed to find new drugs and solutions to diseases that have not been treated. For this purpose, the FDA approved and approved around the world but not FDA (World-not-FDA) compounds in the ZINC15 (Sterling and Irwin, 2015) database were taken into consideration. The purpose of this method is to find the existing binding affinity of a drug currently used for a different disease, rather than the discovery of a new molecule with high binding affinity to PK bacteria. The main purpose in this phase is to obtain the compound binds the bacteria with high affinity, but with a lower affinity to human for species-specific drug design. GOLD docking process was applied for the region known in the literature and the 2 proposed regions, respectively. Then, by comparing the obtained scoring values, compounds that can affect the species-specific were determined.

To sum up, all the methods employed in this thesis are shown in Figure 2.4.

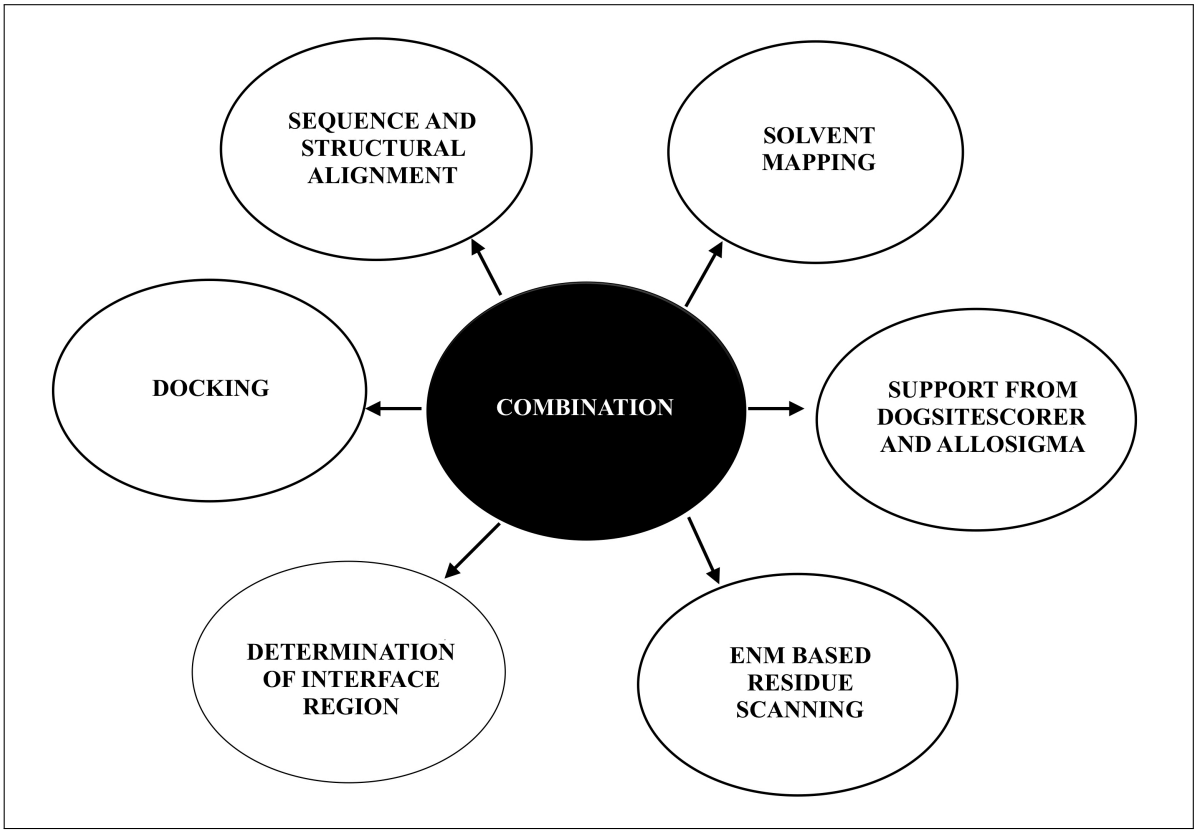


Figure 2.4. Different methods used in our current computational approach.

3. RESULTS AND DISCUSSIONS

3.1. Selections of PDB Structures

In this study, 3 different PDB structures belong to *H.sapiens*, *S.aureus* and *L.mexicana* species were used to shed light on potential allosteric binding sites. In line with this, 32 crystal structures of *H.sapiens*, 12 crystal structures of *L.mexicana*, and three crystal structures of *S.aureus* were extracted from Protein Data Bank (**?**). The RMSD values were calculated by the alignment of each structure belonging to the species within itself. For instance, 36 structures of H.sapiens were aligned one by one and obtained RMSD values were recorded. As the human and parasite species had many structures with the help of R programming language (Team, 2019), clustering dendrograms and heatmap were created according to calculated RMSD values to find the best X - ray crystallographic structure. However, this method was not implemented for *S.aureus* since it had only three structures. Low RMSD between *S.aureus* species for the first cycle was calculated as 1.48 Å; on the other hand, high RMSD value was calculated as 2.05 Å during the first cycle. For *H.sapiens* the low RMSD value was calculated as 0.3 Å and high value as 5.2 Å. Finally, for *L.mexicana*, the low value was computed 0.28 Å and high as 2.57 Å. A table, heatmap, and clustering dendrograms were represented in Figure A.1, A.2, A.3, and A.4 to examine in detail. As a result of these calculations, X-ray crystallographic structure for bacteria (PDB id: 3t0t), human (PDB id: 4g1n), and parasite (PDB id: 1pkl) were selected, and computations were continued through these structures. Not only heatmap and cluster dendrograms were taken into account, but also resolutions belonging to structures were taken into consideration. These RMSD values calculated for *H.sapiens*, *L.mexicana*, and *S.aureus* had 2.3 Å, 2.35 Å, and 3.10 Å, respectively.

3.2. Sequence and Structural Alignment

Sequence and structural alignment aided to reveal differences between structures of species; therefore, chosen three X - ray crystallographic structures were aligned with each other. Firstly, between *Sa*PK – *h*PK and *Lm*PK - *h*PK structures were sequentially aligned to each other to observe the conserved region between species. Sequence identity, similarity, and the gap between human and bacteria were calculated as 34.7%, 48.8%, and 24.1%, respectively. On the other hand, sequence identity, similarity, and the gap between human and parasite were calculated as 48.7%, 64.4%, 4.0%, respectively. As shown in Figure 3.1, the identity, the similarity, and the differences between sequences were represented as blue, orange, and white color, ordinarily.

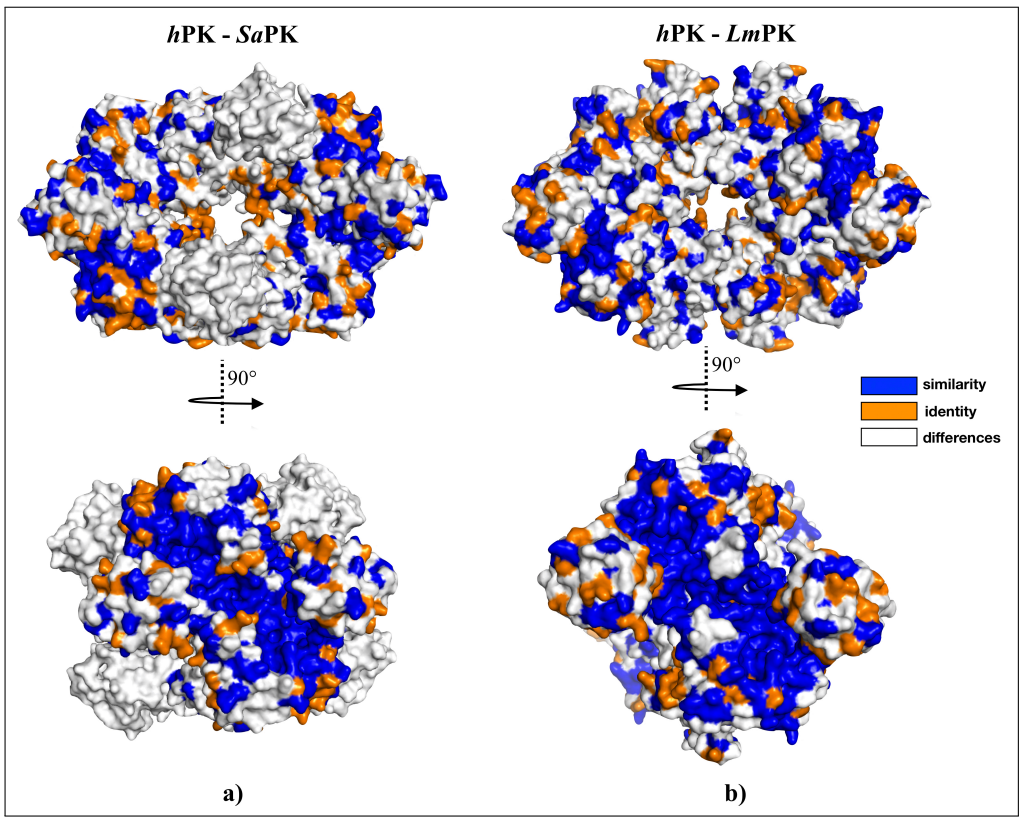


Figure 3.1. Sequence alignment results. a) Sequence alignment between *H. sapiens* and *S. aureus* b) Sequence alignment between *H. sapiens* and *L. mexicana*.

The results of the sequence alignment showed that the residues in the active sites were the same for human and bacteria species. However, looking at the whole structure, there were differences between the residues in the middle part of the bacteria structure. These results showed that active sites had conserved residues. Alternative binding regions could be selected by looking at the differences between the sequences. In this thesis, the analysis of the sequence alignment was performed as the first step that it guided the areas we needed to focus on and provided us insight knowledge about the method of the study. The same results were obtained in the sequence alignment which was made using human and parasite structures. In addition to the results of sequence analysis, structural analysis was conducted to observe structural differences between species. Therefore, with the use of the "*ColorbyRMSD*" script of PyMol graphical tool (Schrödinger, 2015), the indication of the C- alpha atom pairs according to their distances among each other was colored with three different colors. The "*ColorbyRMSD*" module uses the "*Super*" module for alignment and gives the pairwise RMSD values; pairwise RMSD values can be used for large ensembles; they use for clustering structures. The alignment algorithms were used to find the best alignment value, to compare sequences and structures with each other, and to reveal differences between structures. Blue color highlighted the minimum pairwise RMSD, and the red one indicated the highest level. Unaligned residues were colored grey, as shown in Figure 3.2. ColorbyRMSD module calculates deviations in color to show variable regions in a protein structure. Minimum, maximum, and average distance values between *h*PK and *Sa*PK were 1.90, 40.74, 5.95, respectively. Contrarily, minimum, maximum, and average distance values between *h*PK and *Lm*PK were 0.86, 18.33, 4.52.

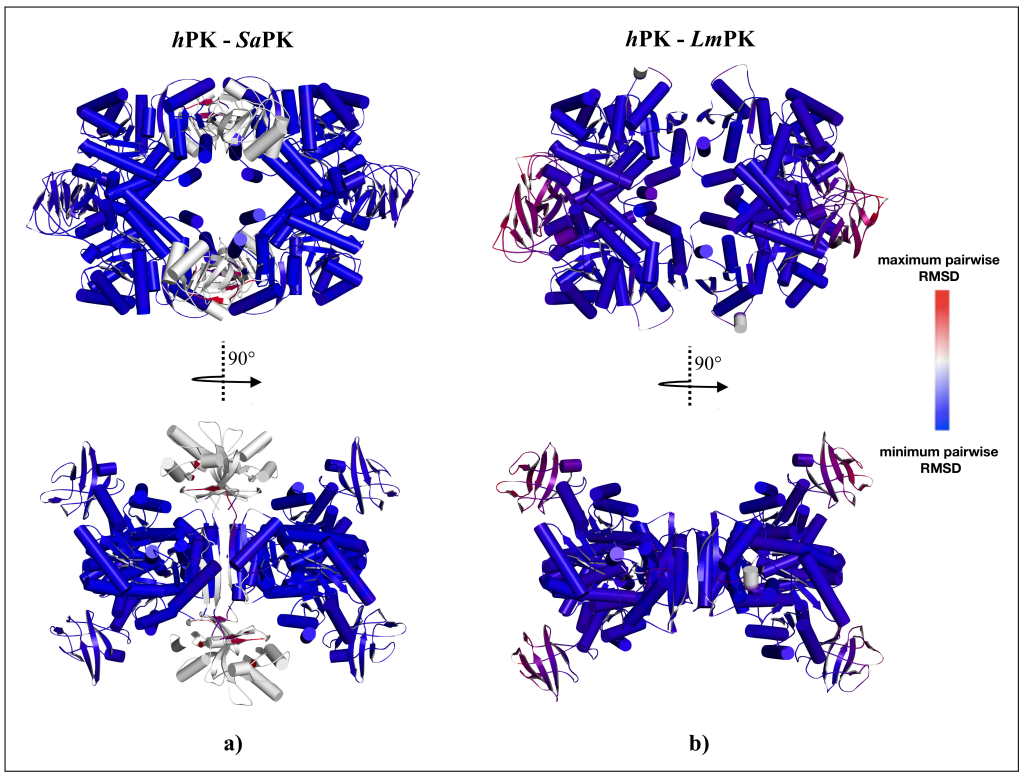


Figure 3.2. Structural alignment results. a) Structural alignment between *H. sapiens* and *S. aureus* b) Structural alignment between *H. sapiens* and *L. mexicana*.

The light of the information given above, sequence and structural alignment strategy was an essential task for highlighting potential binding regions. Alignment results provided high predictions for the continuation of the study.

3.3. Computational Solvent Mapping Analysis with the Use of FTMap

Protein X-ray crystallographic structures were given to the FTMap program for processing both tetramer and chain by chain structures. The clusters which were obtained as a result of calculations were employed to the existence of alternative potential binding sites on the structure. Both the results of the overall structure and the chain by chain structure were employed to reach every region on the protein structure. All the CSs on the protein structure were examined, and their properties were revealed. For instance, it could be overlap and inaccessible. As listed in the first row of Table 3.1, the total number of CSs were examined 58 for *S.aureus*, 53 for *H.sapiens* and 53

for *L.mexicana*. After the elimination of overlapped and inaccessible CSs, 49, 55 and 40 CSs were left for *S.aureus*, *H.sapiens* and *L.mexicana* as listed in the second row of Table 3.1, respectively. The next elimination procedure was implemented according to frequency shift values, both lower than 25%, and 50% CSs were eliminated. Each fields in the remaining list, as stated in the Methods section, was calculated the average percentage frequency shift value for all residues in the immediate vicinity and evaluated according to the calculated value (<%s>). In addition, CSs with an average of frequency values less than 50% were eliminated at the first stage. Then, 25% threshold was also used to increase the number of alternative binding sites. As listed in the third row of Table 3.1, 44, 45, and 32 CSs were left for *S.aureus*, *H.sapiens*, and *L.mexicana*, respectively, when the threshold was 25% . However, when threshold was increased to 50% the remaining CSs were 23, 26, 21 for *S.aureus*, *H.sapiens*, and *L.mexicana*, respectively.

Table 3.1. Determination of Consensus Sites (CSs) before and after eliminating procedure (Ayyildiz *et al.*, 2020).

Enzyme	**Species**	**1) Total Number of Clusters** **2) Non-Overlapping Solvent-Accessible Clusters** **3) After ENM filtering (frequency shift >25% / 50%)**					
		Tetramer	**Chain A**	**Chain B**	**Chain C**	**Chain D**	**TOTAL**
PK	***H.sapiens***	12	12	10	10	9	53
		12	11	9	9	8	49
		12 / 5	9 / 7	8 / 4	7 / 3	8 / 4	44 / 23
	S.aureus	18	8	12	10	10	58
		18	8	11	9	9	55
		18 / 10	6 / 4	7 / 4	7 / 4	7 / 4	45 / 26
	L.mexicana	15	9	11	9	9	53
		15	6	9	5	5	40
		15 / 6	4 / 3	7 / 6	5 / 5	2 / 1	32 / 21

Accordingly, all CSs belong to *Sa*PK, *Lm*PK, and *h*PK were represented in Table 3.2 to highlight all druggable sites on the protein structure. All druggable areas with two or more CSs that caused a frequency shift of more than 50% and 25% in global dynamics were listed in Table 3.2. According to Table 3.2, the CSs obtained after eliminating those below 25% were marked with a single star(*) sign, moreover, between

25% and 50% CSs were marked with a double star(**) and above 50% were marked without the star.

Table 3.2. Distribution of CS among druggable sites in Pyruvate Kinase (PK) (Ayyildiz *et al.*, 2020).

Consensus Sites			
Druggable Site ID	***S.aureus* (PDB id: 3t0t)**	***L.mexicana* (PDB id: 1pkl)**	***H. Sapiens* (PDB id: 4g1n)**
1	1A – 2A -6A	4A* – 5A* – 6A*	1A**–4A*-8A–12A
2	3A* – 5A*	8 – 8A	3A – 9A
3	1B-2B–8B	1B – 9B -11B	5A* – 7A*
4	3B*–5B* -7B*	4B* – 5B*	1B – 9B
5	6B**–9B**	4C - 5C – 6C	4B**–7B**–3B**
6	2C–3C–5C	9C – 5 - 14	1C -5C -10C**
7	1D -6D*	4D* – 5D* – 6D*	3C **– 9C**
8	2D -3D -4D	1**- 10**	4C*–8C*
9	8D* -9D**	2** - 15**	2D**– 4D**
10	1** – 2 – 3 - 16 -18	3** – 9** - 11**	6D** – 7D - 9D
11	4–5–11**-13–15	4** - 6**	1–3**–7**–8**–9**-10**
12	7** - 12	7 - 13	2 -4 -5–6–11**-12**
Clusters with FS <25% are eliminated.** * Clusters with 25% <FS <50% are eliminated.** **CS which is not part of any druggable site is not included in this table** **(8A,10C, 12B, 17 in S. aureus** **3A, 12, 7B, 7C, 9D, 9A,2B,10B in L.mexicana** **3D, 6A, 7C, 8B, 10B, 10A 11A, 8D in H. sapiens).**			

Besides, for broad research, Table 3.3 was created containing the druggable sites according to higher frequency shift values. In this table, *L.mexicana* and *S.aureus* had the same amount of highest CSs in druggable sites; however, parasite and bacteria had the smallest CSs. Detection of druggable sites of proteins was a method needed to find alternative allosteric regions. These sites had a crucial effect on protein dynamics. They were observed at the protein interface, as mentioned earlier, interface regions had a significant impact on species-specific drug design.

Table 3.3. Druggable sites of PK enzyme (Ayyildiz *et al.*, 2020).

Enzyme	**Druggable Site ID**	***S.aureus***	***L.mexicana***	***H.sapiens***
PK	**1**	2-3-16-18	4C-5C-6C	2-4-5-6
	2	4-5-13-15	1B-9B-11B	3A-8A-9A-12A
	3	1A-2A-6A	9C-5-14	1B-9B
	4	1B-2B-8B	8-8A	1C-5C
	5	2C-3C-5C	7-13	7D-9D
	6	2D-3D-4D		

For a better understanding of the CSs contained in the druggable areas, the names of the residues contained in the druggable sites were shown in Table 3.4. Looking at the CSs and druggable sites which were obtained as a result of FTMap, we could have easily observed how much of which area was conserved according to residue comparison. The known allosteric site were located in the same area as the alternative region proposed acoording to the results. It accommodated the inhibitor IS-130 (N'-[(1E)-1-(1H-benzimidazol-2-yl)ethylidene]-5-bromo-2-hydroxybenzohydrazide) which was a potential allosteric inhibitor targeting methicillin-resistant *S. aureus* (MRSA) (Axerio-Cilies *et al.*, 2012). Alignment results also focused attention on that, this region was differed for both *h*Pk and *Sa*PK species. Moreover, for *L.mexicana*, there was also a newly proposed allosteric target site.

Table 3.4. Residues in druggable sites (Ayyildiz *et al.*, 2020).

E	**ID**	***S. aureus***	***L. mexicana***	***H. sapiens***
PK	**1**	**B:**K260/N299/Y302/D303/A337-D339/Y340/K342/L343/D346 **C:**K260/R264/N267/N299/Y302/D303/D346/R347 **D:**Y302/Q338-Y340/K342/L343	**C:**R19/R22/I23/L40/I41-S46/V76-I78/C420/T427/C428-T434/V437-S439	**A:**F26/L27/H29/M30/L33/K311/C326/N350/V352-G355/A388-H391/Q393-F395/E397 **B:**F26/L27/M30/G52/K311/N350/L353/D354/A388-L394/E397
	2	**A:**K260/R264/N267/N299/Y302-G304/A337-Y340/K342/L343/D346/R347 **D:**K260/R264/N299/Y302/D303/A337-A337-Y340/K342/L343/D346	**B:**G86-R90/D145/G176-P180/C182/V184/L186/F212/R214/E240-Q243/D264/L265/V267/E268	**A:**N75/S77/H78/K115/P117-R120/D178/K207-P212/A214/V216/F244/R246/E272/N273/G295-L297/I299/E300
	3	**A:**R32/N34/S36/D66-R73/D125/K156-L160/V165/L167/A190/S192/F193/K219/E221/N222/D245/M246/V248/E249	**C:**S53/K85-R90/G176-C182/V184/L186/F212/R214/E240/N241/D264/V267/E268	**B:**R73/N75/S77/H78/D113-G116/E118-R120/Y175-G179/K207-N210/F241/S243/F244/K270/E272/A293/G295-L297/I299
	4	**B:**R32/N34/S36/D66-R73/D125/G157-L160/V165/L167/A190/S192/F193/E221/N222/D245/M246/V248/E249	**A:**P87-I89/V177-C182/V184/L186/F212/R214/E240/N241/D264/L265/V267/E268	**C:**R73/N75-S77/D113-G116/E118/S243/F244/K270/E272/M291-L297/A327/T328/M360/S362
	5	**C:**R32/N34/S36/H37/D66-R73/D125/K156-L160/V165/L167/A190/S192/F193/K219/E221/N222/D245/M246/V248/E249	**D:**P87-I89/G176-V184/L186/F212/R214/E240-Q243/V267/E268	**D:**A42-I47/N70/D357/C358/P449/I450/A463-P471/F502
	6	**D:**R32/N34/S36/D66-R73/D125/K156-L160/V165/L167/A190/S192/F193/K219/E221/N222/G244-M246/V248/E249		

As shown in Figure 3.3, for *S.aureus*, there were several CSs, when the whole structure was given to the FTMap program, the obtained CSs were shown in green color. The chain by chain structure results were shown in hotpink color. Known allosteric sites were colored with the yellow and active site as orange. Moreover, the IS-130 inhibitor was colored as yellow sticks.

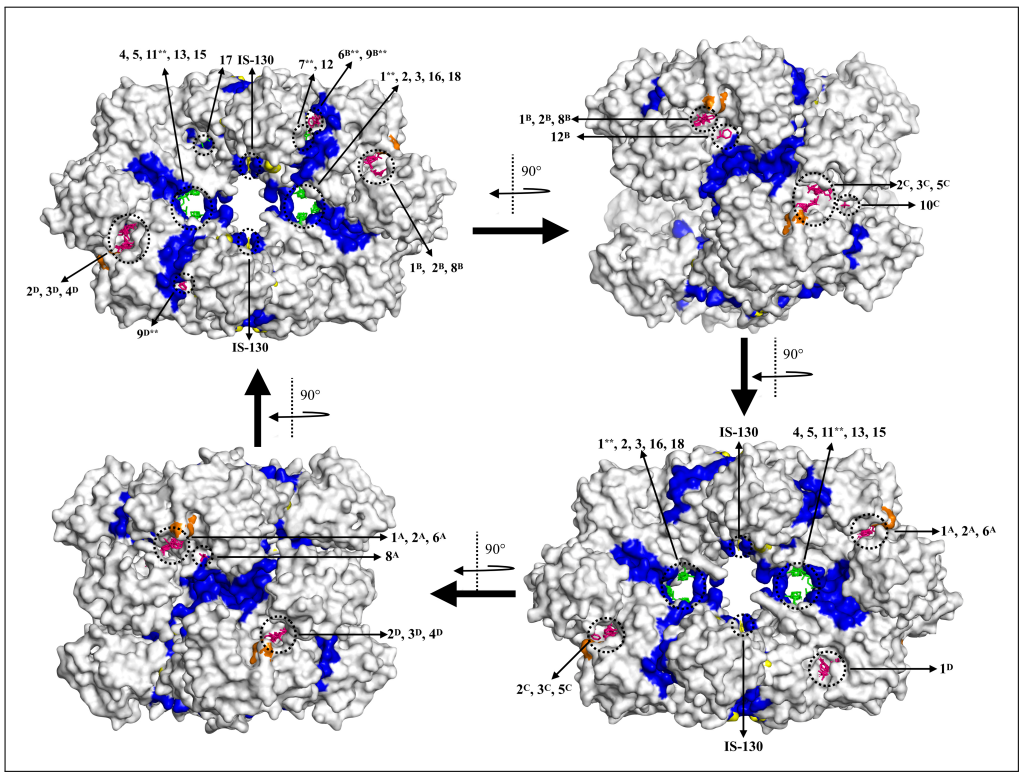


Figure 3.3. Representation of each CSs on the *Sa*PK protein structure according to FTMap results.

As it is shown in Figure 3.4 for *H.sapiens* there were several CSs, known allosteric sites were colored with the yellow and active site as orange. Moreover, NZT (N-(4-[4-(pyrazin-2-yl)piperazin-1-yl]carbonylphenyl)quinoline-8-sulfonamide) was colored with orange and Oxalate(OXL) as yellow.

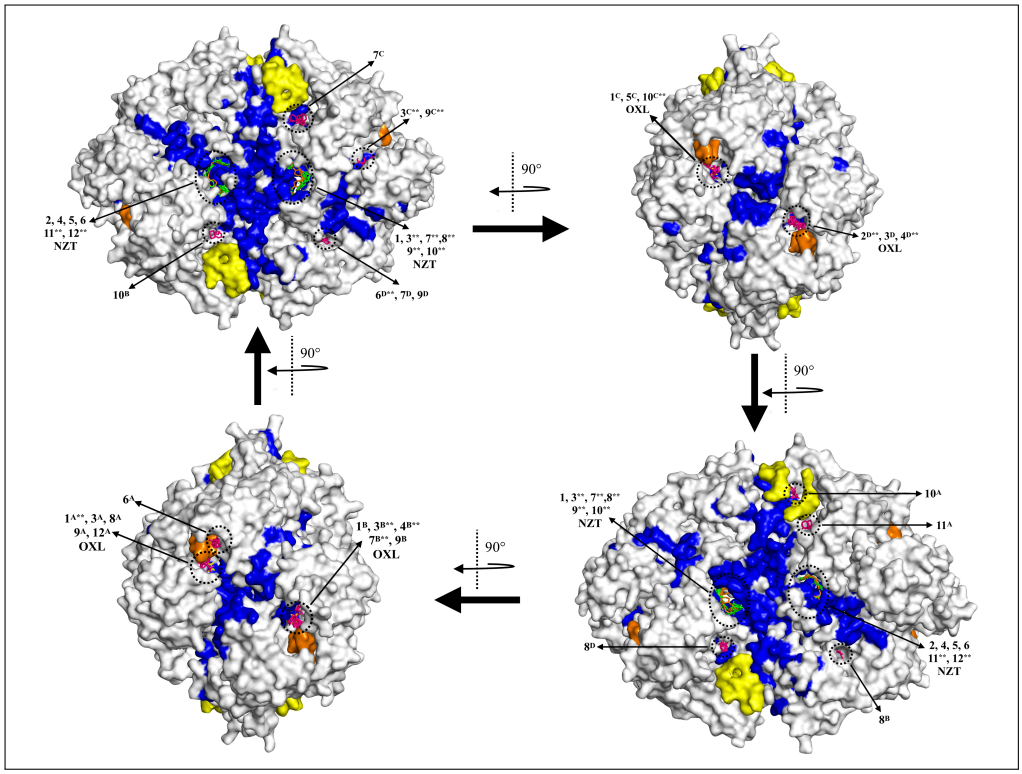


Figure 3.4. Representation of each CSs on the *h*PK protein structure according to FTMap results.

As it is shown in Figure 3.5, for *L.mexicana* there were several CSs, known allosteric sites were colored with yellow and active site as orange.

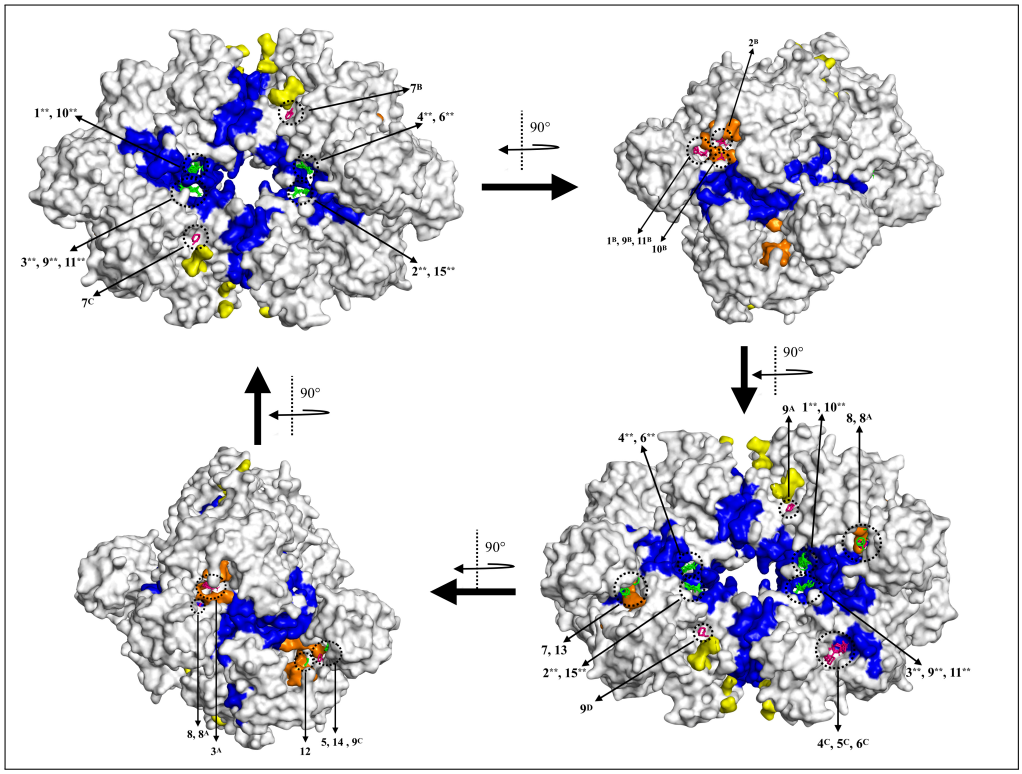


Figure 3.5. Representation of each CSs on the *Lm*PK protein structure according to FTMap results.

When the CSs belonging to 3 species were compared one by one, the following information might be obtained in summary: considering the frequency shift over 50% (in other words extracting double star CSs), the CSs in the middle region (orifice) of the bacteria were different from the human and this region might be an alternative allosteric region. The new proposed target site, shown in green bars on the right and left sides of the protein structures, it passed directly through the large interface area, so it eventually affected the allosteric potential required for activation. There was a potentially druggable site in the region corresponding to the human pyruvate kinase. However, the center of the receptor was in an inaccessible position and had a different shape than the structure of bacteria. Additionally, sequence alignment between *H.sapiens* and *S.aureus* structures showed that there were many different residues between the species in this region. Comparing the parasite and human species, looking at the *Lm*PK, as seen in Figure 3.6, a separate druggable region was observed around an interface area, it accomodated away from both the catalytic regions and the

central region. This area was described as druggable site as seen in Figure 3.6, the residues in this region were likely to provide a separate binding site for specific drug candidates, based on a low sequence of conservation. The remaining four druggable sites listed in Table 3.3 were identified in each of the four catalytic regions. Therefore, these were not considered as target areas. Also, it was observed that there were no druggable sites in the center that met the 50% frequency shift threshold, as in humans or bacteria. On the other hand, four isolated consensus sites were identified in the center of the same locations as humans or bacteria. However, it remained in the 25% -50% range to affect - change the frequency of normal modes. Therefore, considering CSs above 50%, the remaining sites were not very meaningful. Consequently, consensus sites above the 50% frequency shift value threshold were shown in Figure A.5 entirely. In conclusion, the residues of druggable sites were shown in the Table 3.4.

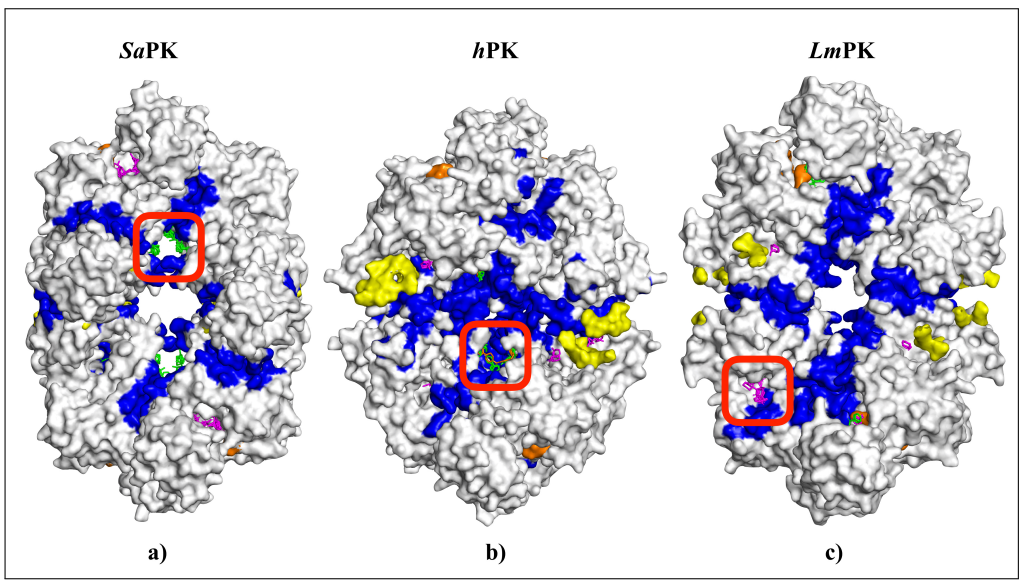


Figure 3.6. Top druggable sites in three species (a,b,c for *Sa*PK, *h*PK, and *Lm*PK). Blue regions indicate interface regions, while orange and yellow regions show active and allosteric regions, which are known experimentally, respectively (Ayyildiz *et al.*, 2020).

3.4. ENM-Based Residue Scanning

In this part of our study, the regions that affected protein dynamics were determined by using an elastic network model. One of the elimination methods used in solvent mapping, which was discussed in the previous section, was a process performed according to the 50% frequency shift method. Therefore, in the previous stage, inaccessible and overlapping CSs were eliminated. The remaining CSs were evaluated together with the average frequency shift results specified in the method, taking into account all residues in the crystal structures of each species. Before proceeding to the elimination process, the global dynamics affecting the structure were calculated and colored according to the frequency shift value. This coloring process was as follows: Frequency shifts smaller than 0 were blue, frequency shifts higher than 0 were red, and frequency shifts very close to 0 were colored as white. ENM was used in the prediction of alternative allosteric regions as a result of the combination of our methods and it gave us information about the global dynamics of the protein.

Some extreme frequency shift values were obtained as a result of ENM calculations. These values were called as "outliers". When the effect of the frequency shift value obtained on the global dynamic was examined, in summary: It was observed that there were both a small number of all residues, and the interesting values coincided with the blue region, which had little effect on the protein dynamics. As shown in Figure 3.7, for *Sa*PK, 16 of 2332 residue defined as outliers (0.68% outliers). Potential dynamics of protein was seen in the Figure 3.7. The same calculations were done for the *h*PK, and the effect of global dynamics was colored and shown in Figure 3.8. As shown in the Figure 3.8, 26 of 2072 residue defined as outliers. (1.25% outliers). For *Lm*PK the calculations were shown in Figure 3.9. Furthermore, 15 of 1965 residue defined as outliers (0.76% outliers).

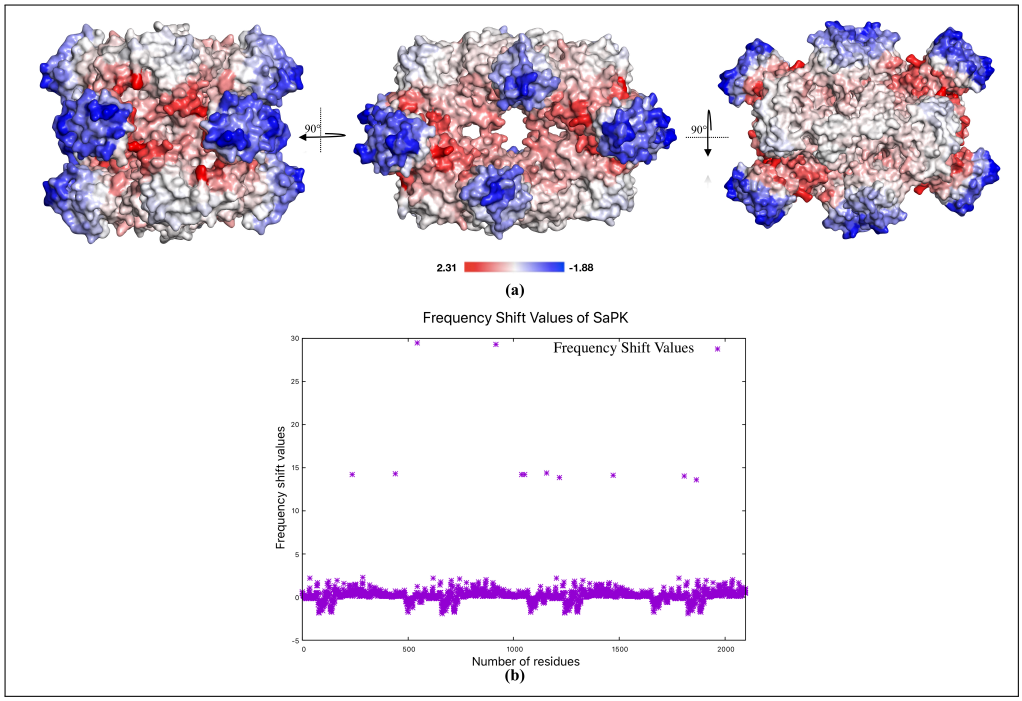


Figure 3.7. Combination of ENM results for *Sa*PK (a) ENM results (b) Plot of frequency shift value.

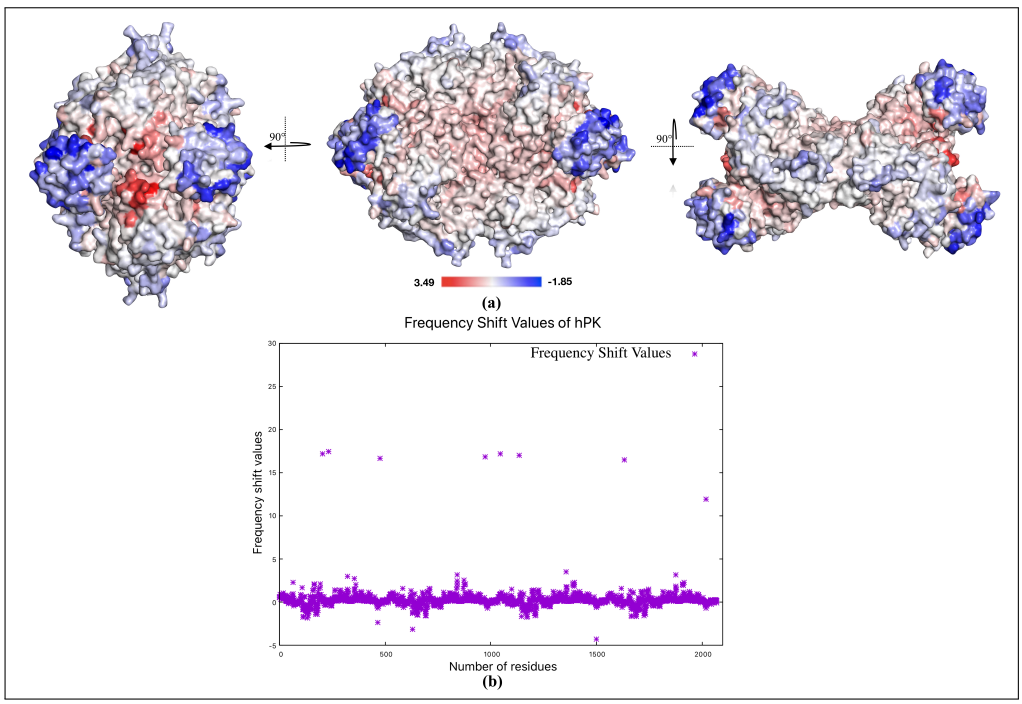


Figure 3.8. Combination of ENM results for *h*PK (a) ENM results (b) Plot of frequency shift value.

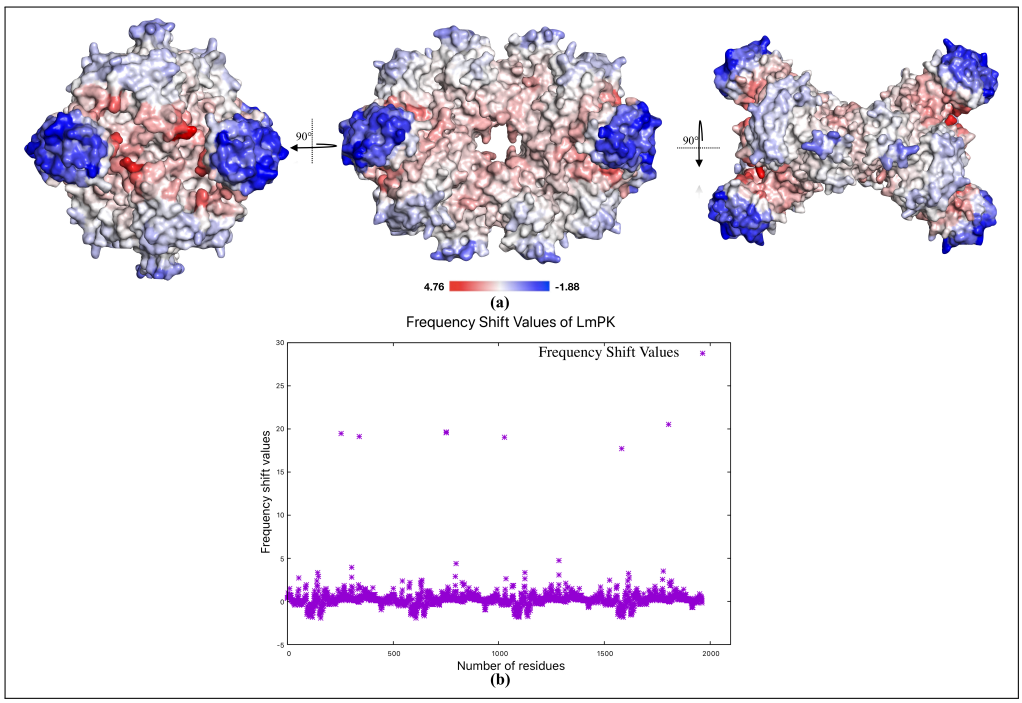


Figure 3.9. Combination of ENM results for *Lm*PK (a) ENM results (b) Plot of frequency shift value .

In Figure 3.10, tetramer structures for all species were given. The percentage of allostery increased from blue to red, and white colors represented being neutral. For the PK enzyme, the allosteric effect in the regions was observed by coloring according to the ENM results obtained in each species. Accordingly, the species with the lowest and highest frequency shift value was observed as *Lm*PK. The color scale was reduced to the same level for all three species for a more detailed examination and to see the differences between the species more clearly. As seen in Figure 3.11, this color scale was between 4.76 and -1.88. Fixing the color scale did not affect in terms of ENM and conservation; this process was done to examine the differences entirely visually.

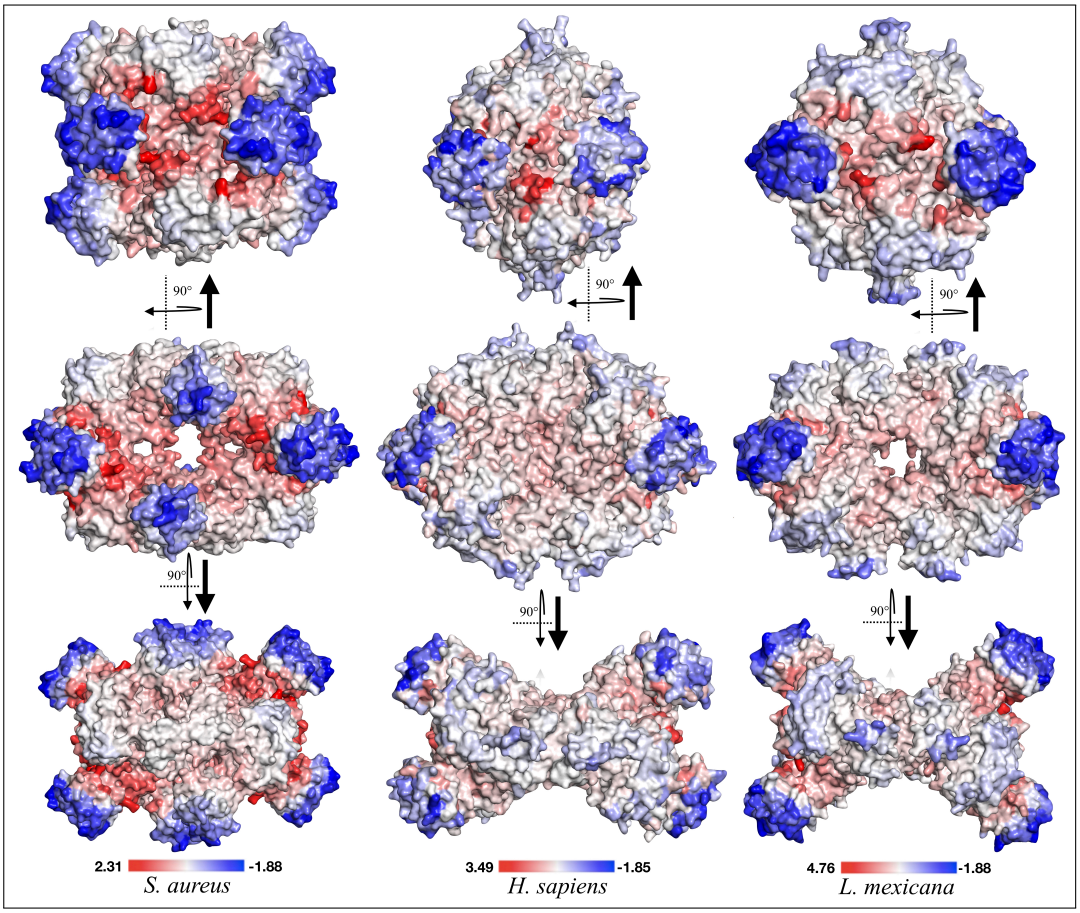


Figure 3.10. ENM based residue scanning results for all 3 species with scale specific to each species.

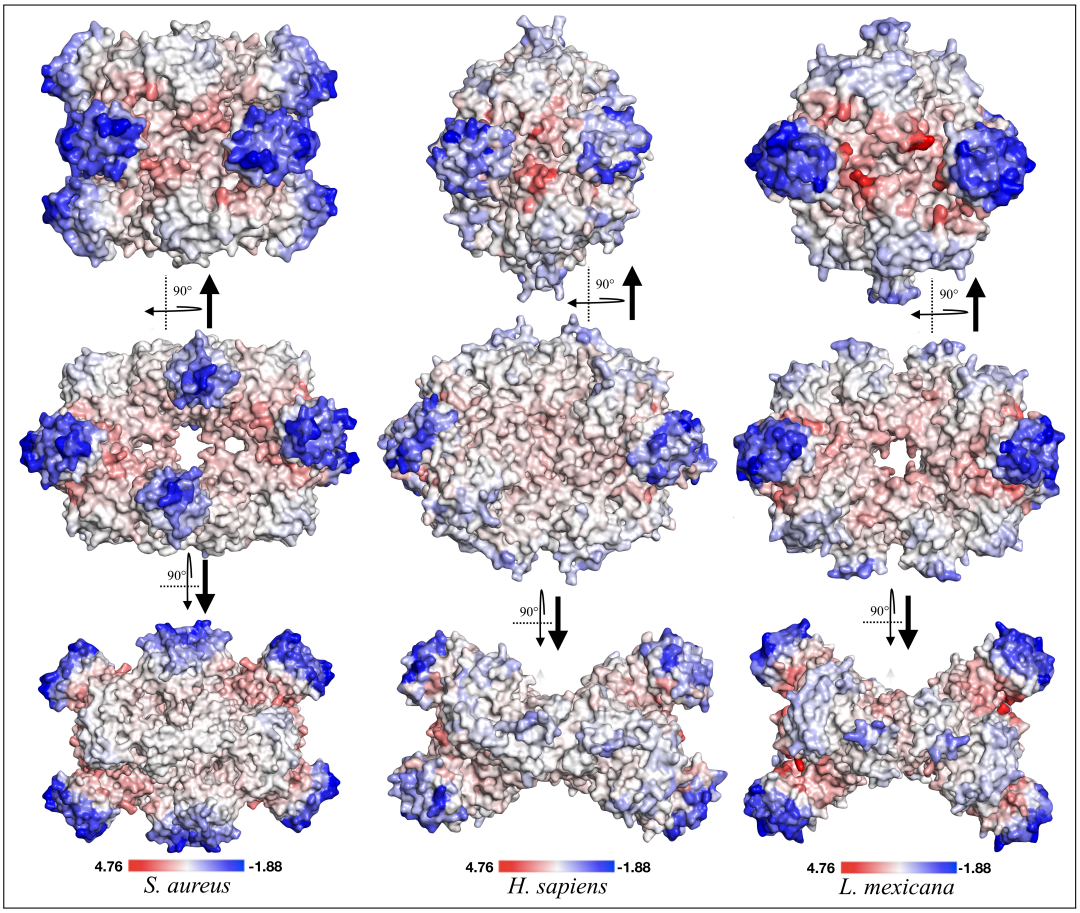


Figure 3.11. ENM based residue scanning results for all 3 species using the same scale.

In addition, looking at Figure 3.10 and Figure 3.11, changes between species were visible. When the scale was equalized, it was seen that the most significant change occured in the *Sa*PK. As a result, when a small molecule was bound to the bacteria, the dynamic was more likely to be affected, or the bacteria had more allosteric sites than the other two species.

Looking at the figures, the color change in the known active sites for each species and the color change in the alternative allosteric regions were observed in the same proportion. This result showed the accuracy of our method. For example, when the known active sites of the bacteria was observed the red colors were intense. The central (middle) region that recommended for the type of bacteria, this color scale was observed. Looking at the ENM results of the druggable sites mentioned in the previous

FTMap results (see Figure 3.6), these regions had effects on protein dynamics and each region might be a potential allosteric site.

3.5. Combination of ENM and FTMap Results

Considering the ENM results, the standard deviation and average frequency shift value were determined by taking the residues up to 5 Å distance of each CSs. As a result, three distinctions, above 50%, between 50% and 25% and below 25%, were used. By evaluating the results of ENM and FTMap, Table 3.5 was created for residues above 50%, and alternative allosteric regions were determined for *Sa*PK. The table was prepared considering the values created by giving the chain by chain and the whole structure to the FTMap separately.

Table 3.5. Values of *Sa*PK ENM results according to 5 Å distance.

Cluster ID	Sum of ENM Frequency Shift	Total Number of Residues	Mean Frequency Shift
3T0T CHAIN A			
1	9.05	15	0.60 ± 0.38
2	12.24	16	0.76 ± 0.49
6	6.13	9	0.68 ± 0.44
8	4.52	8	0.57 ± 0.28
3T0T CHAIN B			
1	9.22	15	0.61 ± 0.39
2	11.15	15	0.74 ± 0.51
8	6.31	9	0.70 ± 0.46
12	4.98	8	0.62 ± 0.24
3T0T CHAIN C			
2	11.30	16	0.71 ± 0.53
3	12.31	16	0.77 ± 0.49
5	6.19	9	0.69 ± 0.44
10	7.46	9	0.83 ± 0.41
3T0T CHAIN D			
1	16.48	18	0.92 ± 3.18
2	9.26	15	0.62 ± 0.39
3	12.44	16	0.78 ± 0.51
4	6.73	10	0.67 ± 0.44
3T0T ALL CHAIN			
2	5.98	8	0.74 ± 0.14
3	4.78	6	0.80 ± 0.12
4	5.96	8	0.74 ± 0.14
5	5.96	8	0.75 ± 0.14
12	31.86	10	3.19 ± 8.74
13	4.58	9	0.51 ± 0.21
15	2.97	5	0.59 ± 0.19
16	4.09	7	0.58 ± 0.16
17	16.98	9	1.89 ± 4.41
18	3.04	6	0.51 ± 0.16

For *h*PK and *Lm*PK species same table was created to observe possible allosteric druggable sites. Values of *h*PK for chain by chain structure and the entire structure were seen in Table 3.6 and *Lm*PK in Table 3.7, respectively.

Table 3.6. Values of *h*PK ENM results according to 5 Å distance.

Cluster ID	Sum of ENM Frequency Shift	Number of Residues	Mean Frequency Shift
4G1N CHAIN A			
2	42.13	14	3.01 ± 9.13
3	28.78	16	1.80 ± 4.07
6	9.01	15	0.60 ± 0.52
8	8.68	7	1.24 ± 0.60
9	23.69	14	1.69 ± 4.36
10	39.74	18	2.21 ± 8.06
11	6.58	10	0.66 ± 0.25
12	3.89	6	0.65 ± 0.47
4G1N CHAIN B			
1	8.43	17	0.50 ± 0.42
8	19.61	16	1.23 ± 4.02
9	8.16	16	0.51 ± 0.36
10	6.51	11	0.59 ± 0.20
4G1N CHAIN C			
1	22.80	16	1.43 ± 4.03
5	20.60	15	1.37 ± 4.18
7	7.08	13	0.54 ± 0.24
4G1N CHAIN D			
1	78.48	18	4.36 ± 11.05
3	10.31	18	0.57 ± 0.35
7	38.57	17	2.27 ± 8.28
8	40.50	9	4.50 ± 11.18
9	37.31	10	3.73 ± 10.56
4G1N ALL CHAIN			
1	78.14	15	5.20 ± 11.93
2	42.39	15	2.82 ± 8.85
4	42.76	16	2.67 ± 8.59
5	38.71	11	3.52 ± 10.25
6	39.32	9	4.37 ± 11.16

Table 3.7. Values of *Lm*PK ENM results according to 5 Å distance.

Cluster ID	Sum of ENM Frequency Shift	Number of Residue	Mean Frequency Shift
1PKL CHAIN A			
1	17.58	21	0.84 ± 0.41
2	11.40	14	0.81 ± 0.47
3	11.50	17	0.68 ± 0.65
7	63.22	18	3.51 ± 13.04
8	7.15	10	0.71 ± 0.39
9	62.16	7	8.88 ± 19.80
1PKL CHAIN B			
1	12.97	14	0.93 ± 0.69
2	9.65	17	0.57 ± 0.37
7	6.43	10	0.64 ± 0.32
9	7.94	10	0.79 ± 0.62
10	45.94	9	5.10 ± 7.76
11	8.37	12	0.70 ± 0.28
1PKL CHAIN C			
1	10.48	14	0.75 ± 0.46
2	12.29	16	0.77 ± 0.35
3	9.86	18	0.55 ± 0.42
4	44.03	21	2.10 ± 8.63
5	42.82	13	3.29 ± 10.80
6	41.82	13	3.22 ± 10.82
7	5.42	8	0.68 ± 0.35
9	7.06	7	1.01 ± 0.57
1PKL CHAIN D			
1	14.89	19	0.78 ± 0.52
2	12.83	18	0.71 ± 0.41
7	11.84	15	0.79 ± 0.44
9	4.19	8	0.52 ± 0.37
1PKL ALL CHAIN			
5	10.08	13	0.78 ± 0.46
7	7.48	11	0.68 ± 0.36
8	13.49	16	0.84 ± 0.39
12	7.41	11	0.67 ± 0.48
13	13.27	15	0.88 ± 0.37
14	9.03	12	0.75 ± 0.35

As a result, CSs with mean frequency shift values below 50% were eliminated. General information about average frequency shift values, standard deviations of remaining CSs and total residue numbers around CSs were shown in the table (see Table 3.3). After FTMap and ENM results were combined with each other, three regions different from human for *Sa*PK and one region for *Lm*PK were found. These regions were designated as hotspots for the studies, and the processes were carried out by continuing through these regions.

Proposing different allosteric regions other than the existing and known allosteric regions in the literature showed great progress in terms of achieving the purpose of the study. Focused on these proposed regions, differences between species were revealed, and target-based actions became more significant than other regions.

3.6. Identification of Hotspot Regions and Revealing Differences Between Species

To determine the differences between the species, sequence analyzes were performed by focusing on the proposed alternative allosteric regions, and it was revealed how much the sequences were conserved and how different they were from other species. This method is one of the methods employed to observe the differences between regions determined in species-specific drug design studies. As a result of the alignment, FTMap, and ENM studies, the regions proposed for each type were shown in Figure 3.12. Figure 3.12 a section contained the IS-130 (colored by yellow), which was already identified in the literature as an allosteric inhibitor, above and below the hole located in the midpoint of the *Sa*PK enzyme. There was also a new proposed alternative allosteric site, shown in red in the same central zone. This region was different from the human structure; as it appeared from the Figure 3.12, the region was clear and its located in a large interface. The drug molecules attached here had great importance for the activation of the enzyme.

Additionally, this proposed druggable site which were corresponded to the same pocket to which the quinolone sulfonamide activators were attached(Kung *et al.*, 2012). Besides, sequence alignment showed this area with a high amount of variation that further emphasized proposed region as an ideal target for species-specific drug design.

When the druggable sites were examined (see Table 3.3), the remaining four areas did not show an allosteric feature since they appeared around the catalytic area. A second alternative allosteric region in *Sa*PK emerged at the junction of the area A and C of a subunit, as illustrated with the orange color, cluster 17 in Figure 3.12 a section. Looking at the *h*PK of the same region, no cluster was observed in *h*PK. Furthermore, sequence similarity analysis showed a high degree of variation, indicating that this region was species-specific.

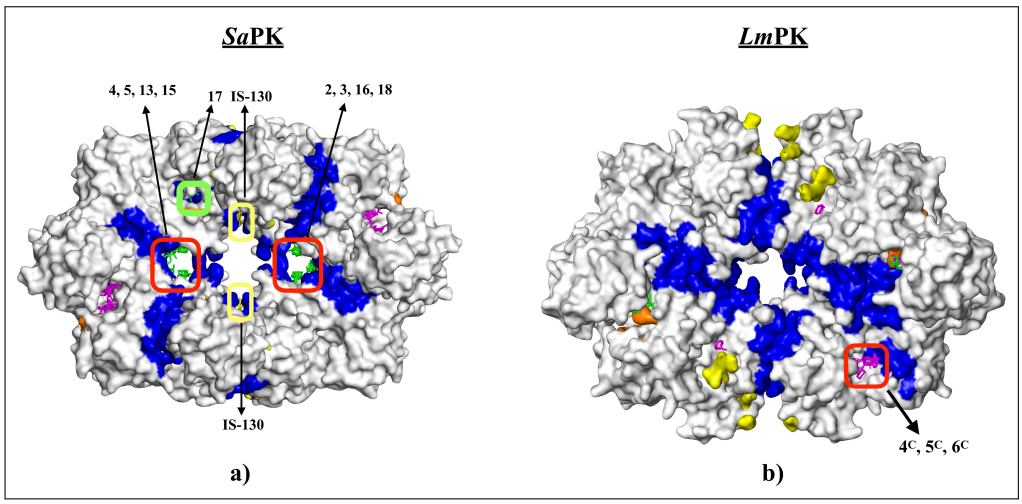


Figure 3.12. Proposed druggable sites for a) *Sa*PK and b) *Lm*PK species (Ayyildiz *et al.*, 2020).

As shown in Figure 3.12 b section, a separate druggable site on the LmPK, away from both the catalytic regions and the central region, was observed near an interface region. This region was likely to provide an alternative binding site for specific drug candidates, based on a low sequence of conservation observed compared to *h*PK. The remaining four druggable areas listed in Table 3.3 were observed in the catalytic site as in bacteria. Also, there was no druggable area in the center that met the 50% frequency

shift threshold, as in *h*PK or *Sa*PK. On the other hand, four isolated consensus sites were detected in the center in the same places as hPK or SaPK, but since the average frequency value found in these regions remained below 50%, it remained at a moderate level to affect the normal mode.

Sequence analysis of *h*PK and *Sa*PK species was carried out in order to examine the differences between the proposed allosteric regions. As a result of this sequence alignment analysis in 3 different regions determined for bacteria were shown in Figure 3.13.

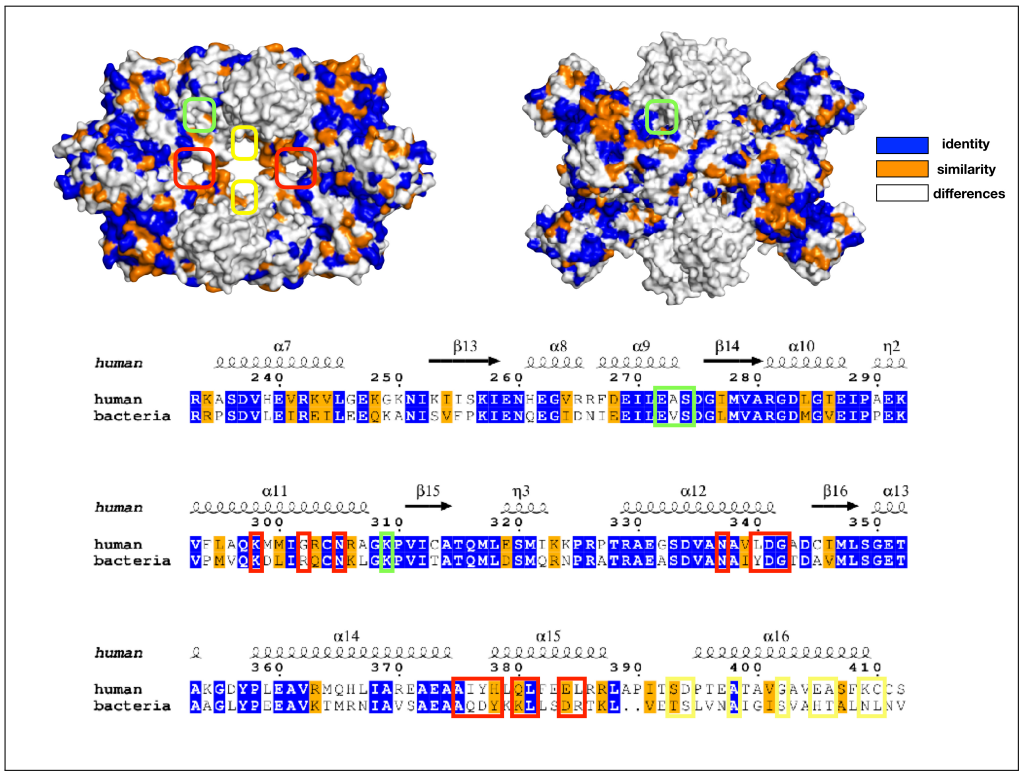


Figure 3.13. Sequence alignment results for proposed alternative allosteric regions for *Sa*PK, green circle represents CS-17, red circle represents CSs 2,3,16,18 and orange circle represents IS-130 inhibitor (Ayyildiz *et al.*, 2020).

In the Figure 3.13, the sequence similarity between the species was low in the proposed allosteric regions of the *Sa*PK species. The differences between species must be high in the recommended regions to prove the accuracy of our results.

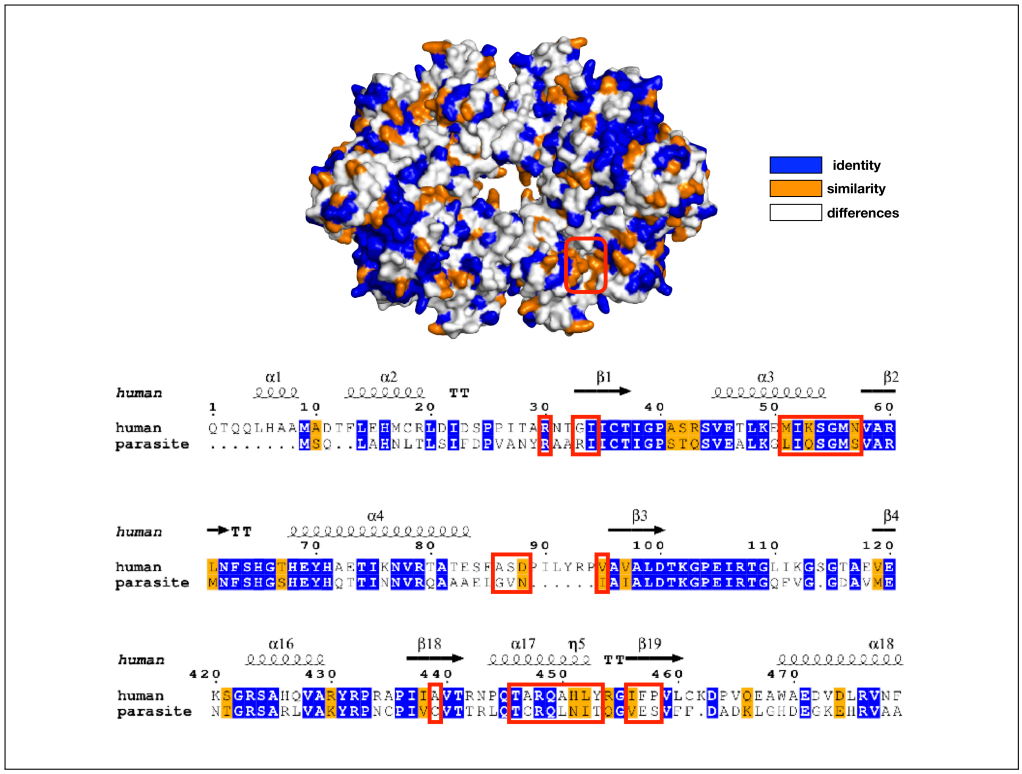


Figure 3.14. Sequence alignment results for proposed alternative allosteric regions for *Lm*PK, red circle represents CSs 4C, 5C, 6C (Ayyildiz *et al.*, 2020).

As it was shown in Figure 3.14, as in the *Sa*PK, it was observed that the sequence similarity between human and parasite was low. Since the conservation rate between sequences was also very small, this region was also highly likely to be an alternative allosteric region.

3.7. Supporting Found Results by DogSiteScorer and AlloSigMA

3.7.1. Determination of Binding Pockets via DogSiteScorer

The DogSiteScorer tool(Volkamer *et al.*, 2012) was used to support proposed alternative allosteric regions with other programs. Considering the results of this program, in summary: the regions we recommended for both species (*Sa*PK and *Lm*PK) showed a high drug score value. This tool gives the drug score in the range of 1 and 0. It showed that the pocket found as it approached an appropriate druggable region,

as it was shown in Figure 3.15. DogSiteScorer tool gave us different binding pockets according to drug score. The figure showed the top 3 binding pockets that the program gave as a result. When the values of binding pockets were evaluated with our proposed regions, the new proposed allosteric region for *S.aureus* PK, in the center of the structure adjacent to the large interface, showed a high binding rate with a value of 0.76. When the values given by the program for *Sa*PK are listed in Table 3.8, it was observed that the region in the first place had 0.84 value.

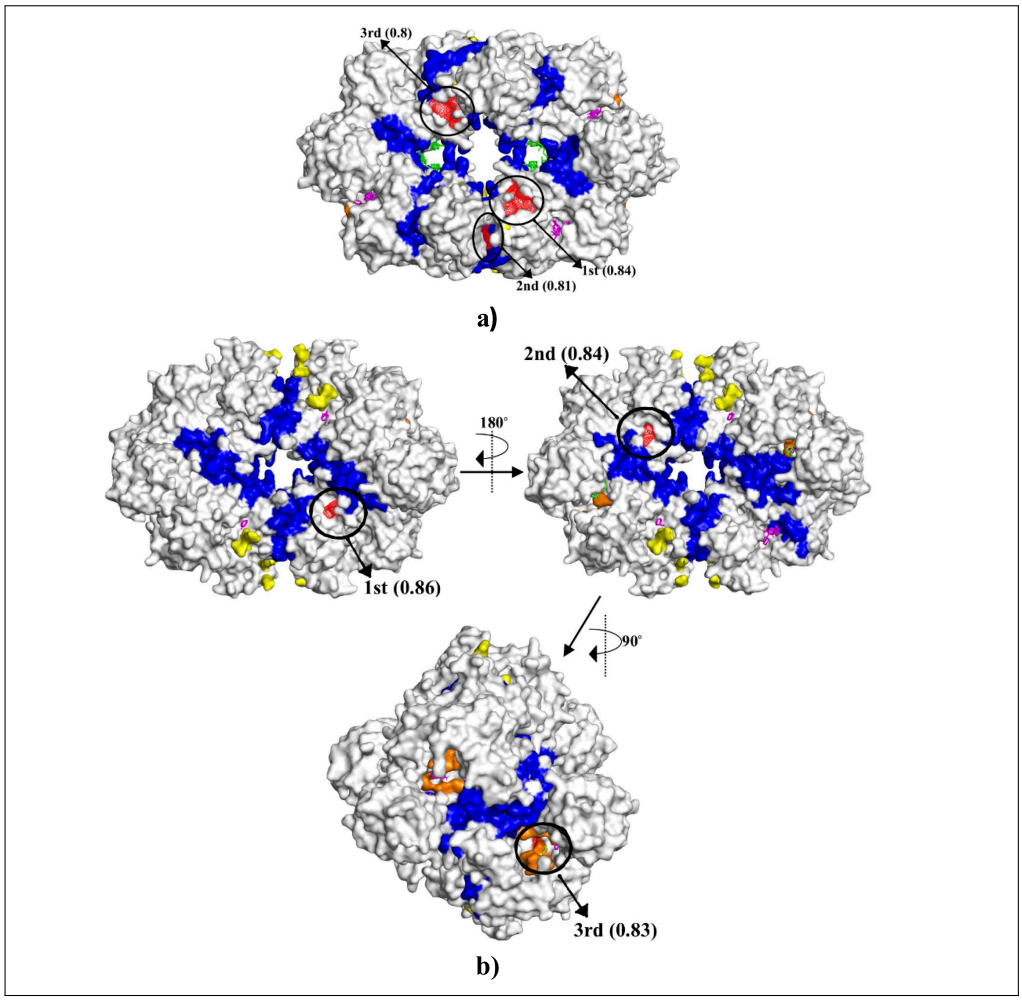


Figure 3.15. The location of the binding pockets predicted by DoGSite in a) *S.aureus* pyruvate kinase and b) *L.mexicana* pyruvate kinase (Ayyildiz *et al.*, 2020).

On the other hand, the catalytic regions that appear as druggable regions on our list (see Table 3.3) were not strongly recommended by DogSiteScorer. For *L.mexicana* PK, the proposed allosteric region, located away from the origin and near an interface, was not a suitable pocket with a value of 0.42 and in the ninth position for DogSiteScorer. On the other hand, the remaining four catalytic regions for *Lm*PK coincided with the results of DogSite with a high score.

Table 3.8. Top druggable for each species and their corresponding DogSite binding pockets with score and rank (Ayyildiz *et al.*, 2020).

Enzyme	*S.aureus*	Region	Score*/ Rank	Parasite	Score/ Rank	Region
PK	2-3-16-18	allosteric	0.76 / 6	4C-5C-6C	0.42 / 9	allosteric
	4-5-13-15	allosteric	0.71 / 8	1B-9B-11B	0.81 / 4	catalytic
	1A-2A-6A	catalytic	0.58 / 18	9C-5-14	0.83 / 3	catalytic
	1B-2B-8B	catalytic	0.57 / 19	8-8A	0.81 / 4	catalytic
	2C-3C-5C	catalytic	0.63 / 16	7-13	0.81 / 4	catalytic
	2D-3D-4D	catalytic	0.53 / 21			

In Table 3.8, *Sa*PK DogSiteScorer values were ranged between 0.15 and 0.84 and *Lm*PK values were ranged between 0.11 and 0.86.

3.7.2. Determination of Allosteric Effects with AlloSigMA

The AlloSigMA program was used to evaluate the stabilization in our proposed allosteric regions, observe the effect of ligand binding or mutation, occurring in that region, on all possible protein local configurations, and measure allosteric effects based on free energy per residue. In pyruvate kinase *S.aureus* and *L.mexicana* structure AllosigMA server was applied for top druggable binding sites. As a result of binding to these regions, changes in the active sites were taken into account. This server calculated ΔG free energy according to increased dynamics (positive sign, local destabilization) and decreased dynamics (negative sign, local stabilization), respectively. In line with results for *Sa*PK, when bindings were created on the proposed (top druggable region) region symmetrically, the mean of energy was measured between -0.00 and 0.17 for ADP

and the standard deviation between 0.34 and 0.44, it was seen in Figure 3.16. Mean energy values were between -0.63 and -0.09, and the standard deviation was between 0.40 and 0.51 for the PEP. These indicated that the dynamic in the PEP binding region decreased,whereas the dynamic in the ADP binding region increased. Moreover, for *Lm*PK, the mean of energy was measured between -0.92 and 0.65 for ADP and the standard deviation between 0.09 and 0.17. Mean energy values were between -0.82 and 0.69, and the standard deviation is between 0.09 and 0.27 for the PEP, shown in Figure 3.17. This indicated that dynamics had decreased in the active region close to the binding site, while increased dynamics in the other region. Looking at the results, it was observed that when a ligand was bound to the proposed region, the residue dynamic decreased in the catalytic regions. Since this protein was homotetramer in structure, it was estimated that the same effect will be observed when ligand is bound to the region located in the symmetry of the proposed binding site. When these two results were compared, it was observed that the change in the proposed regions was significant in terms of allostery.

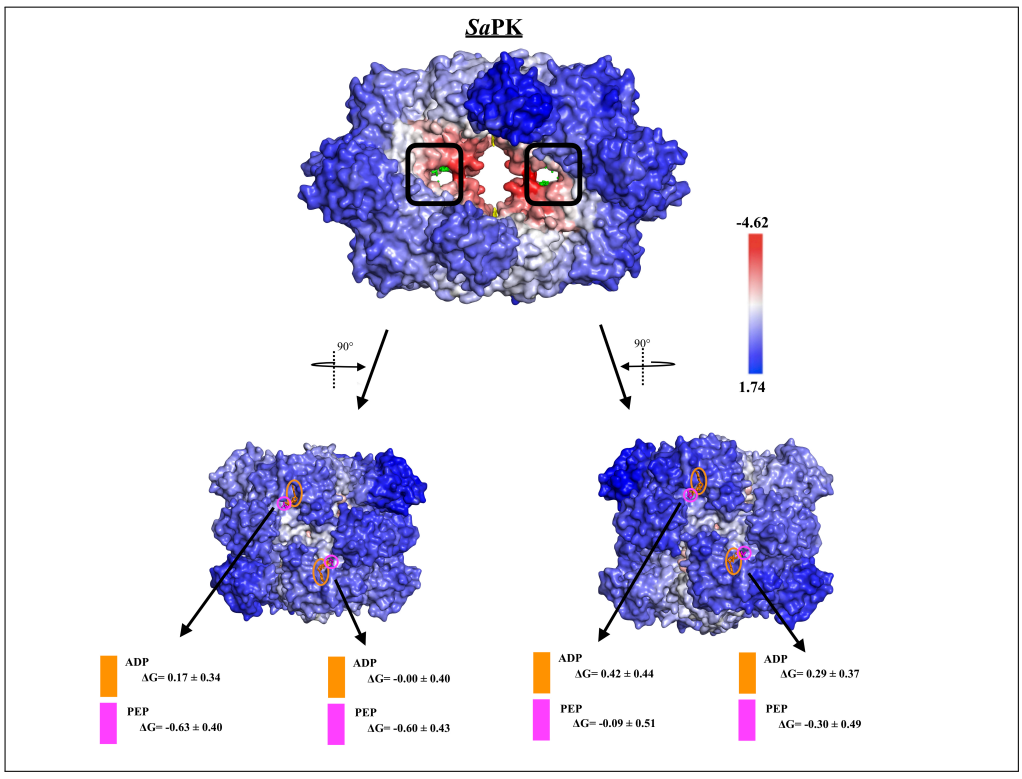


Figure 3.16. AlloSigMA results for *Sa*PK enzyme with PEP and ADP substrates (Ayyildiz *et al.*, 2020).

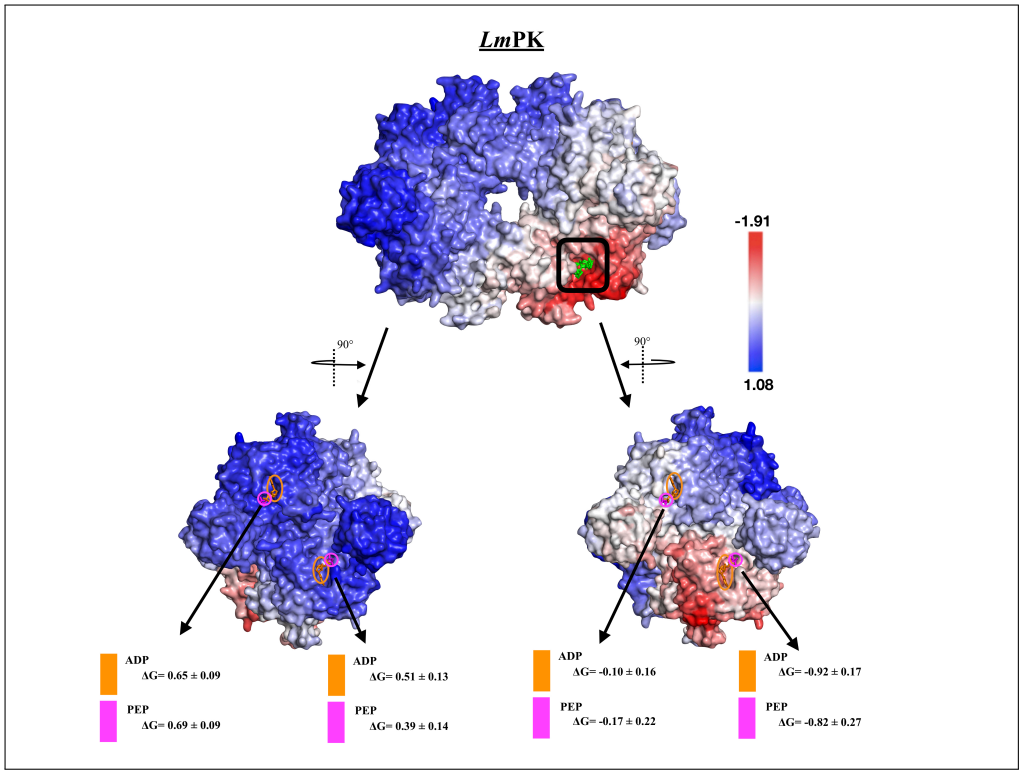


Figure 3.17. AlloSigMA results for *Lm*PK enzyme with PEP and ADP substrates.

Furthermore, the same procedures shown in Figure 3.16 and Figure 3.17 were applied to the known allosteric region. The binding site was determined as the region where the IS-130 inhibitor was bound, and a similar analysis was performed. Astonishingly, the allosteric effect showed the opposite effect for the proposed region, with an increased in dynamics in most of the PEP and ADP binding regions with mean ΔG values as high as 0.78$\pm$0.21. To sum up, considering the values obtained from the known allosteric region, the fact that the values obtained for the regions we proposed were more significant than known allosteric region values, and they increased the accuracy of the new potential allosteric regions we found.

3.8. Determination of Target Molecules by Docking

Based on the results obtained from all the methods performed so far, docking studies were started to find the most appropriate conformation of the ligand that would bind in the target allosteric regions. The score functions ChemPLP was used in these studies to list the docking conformations between the target allosteric region and ligand that interacts in terms of quality. It classified the different docking conformations of a single complex. Moreover, it calculated and sorted the binding affinities of the same protein with different ligands. To ensure the validation of docking parameters used in this thesis; it was noted that accurate RMSD values were obtained, the docking scores were ranked correctly, and the target regions determined in the studies were the most appropriate. The purpose of docking applications was to find positions where the interaction was maximum, and the energy was minimum.

In this regard, by using the ChemPLP scoring function, docking studies of *S.aureus* and *H.sapiens* species for 1416 FDA approved molecules, and 2922 World-not-FDA molecules obtained from the ZINC15 database were started. In order to find the coordinates of the human corresponding to the target allosteric regions determined on the bacteria, structural alignment was performed first. Grid boxes were placed according to the coordinates obtained as a result of structural alignment. The determined Grid box coordinates for *h*PK and *Sa*PK are shown in Table 3.9. These regions targeted for docking studies were divided into three as the proposed allosteric region, the region where the cluster 17 was located, and the regions where the IS-130 inhibitor was located, which was known allosteric region. A total of 12 different operations were performed as a docking study of FDA and World-not-FDA molecules for all three regions in *h*PK and *Sa*PK.

Table 3.9. Coordinates of grid boxes for *Sa*PK and *h*PK.

Regions	Bacteria			Human		
	X	Y	Z	X	Y	Z
First site (proposed allosteric)	15.41	-10.56	-4.34	15.41	-10.56	-4.34
Second site (cluster 17)	35.47	-31.63	-0.81	24.94	-21.97	-25.04
Third site (known allosteric)	30.10	-9.06	-9.11	27.18	0.64	-12.05

After the end of docking calculations, score functions obtained for each molecule were listed from higher to lower. Looking at the values acquiring from screening results, the values should be very well bound to bacteria, that was, they should have a high score value, but a low value in human. Therefore, firstly, molecules that observed below scoring function 60 in the bacteria, then molecules with the difference between bacteria and human are below 10 were eliminated. After the elimination procedure, the remaining molecules for each region were merged with each other, and three molecules were obtained for three regions in FDA and 13 molecules in World-not-FDA. The docking processes were shown in Figure 3.18 as a flow chart.

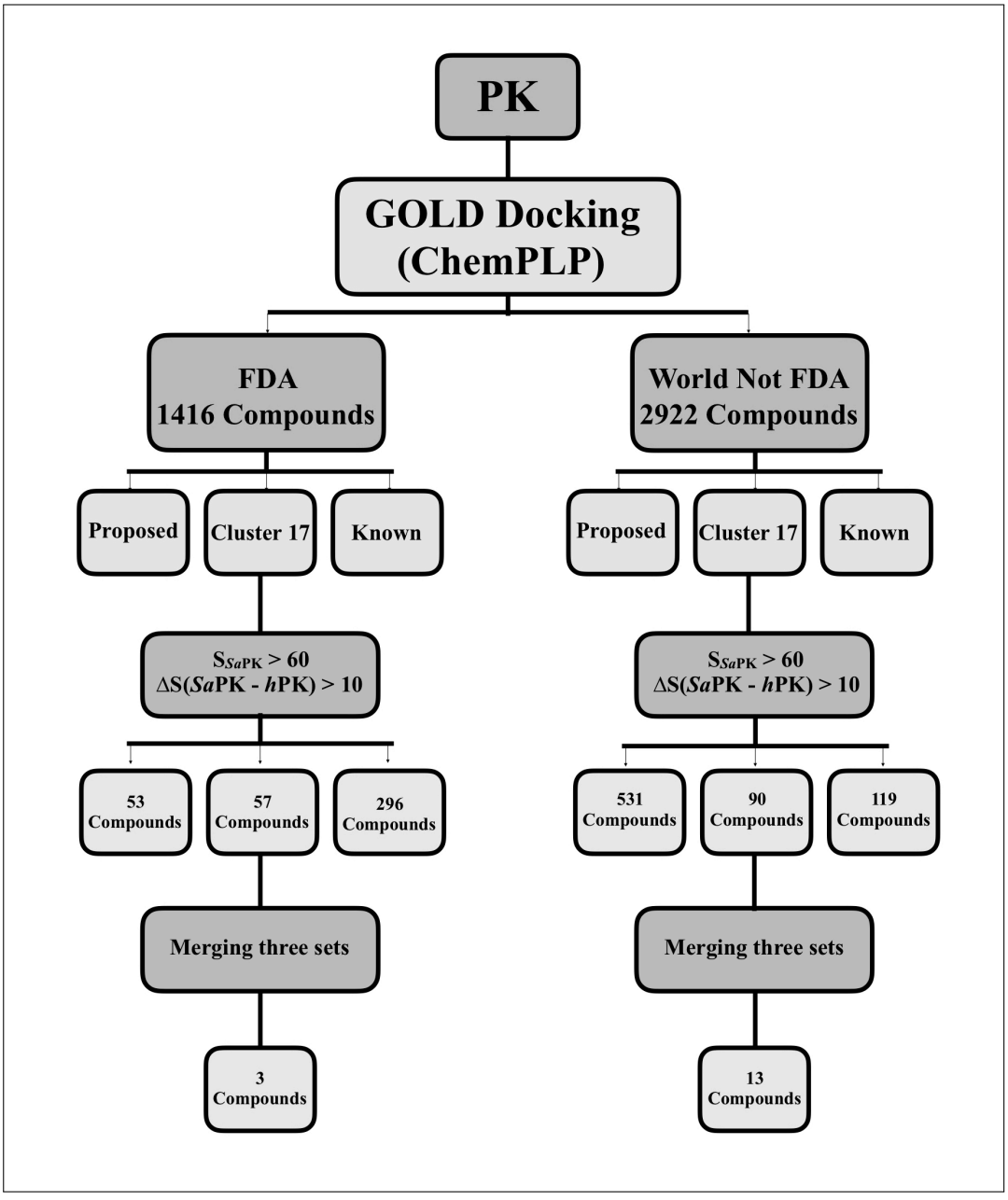


Figure 3.18. Flow chart used for docking experiments.

As seen in Figure 3.19, Figure 3.20, and Figure 3.21, 1416 FDA approved and 2922 World-not-FDA molecules were listed as proposed, cluster 17 and known regions according to results of docking studies between *Sa*PK and *h*PK species. The differences between *Sa*PK and *h*PK were given on the y-axis that appears on the right in the figures. Considering elimination method according to the results, it was seen in Figure 3.19 a section, while the FDA approved molecules bound bacteria to high score values,

they bound to human with low values, thus the differences in values between the two species was observed to have a positive distribution. In Figure 3.19 b section almost the same results were obtained for World-not-FDA molecules values. These values showed a slightly positive distribution. As a result of the study, the molecules that were common to both species are shown on the figure with a black square. It was observed that these common molecules also bound to bacteria with high score value and that the difference between common molecules of bacteria and human was high. The distribution of molecules obtained after the elimination step performed according to the flow chart was shown in Figure 3.19 c section and Figure 3.19 d section. A favorable distribution was also observed in these figures.

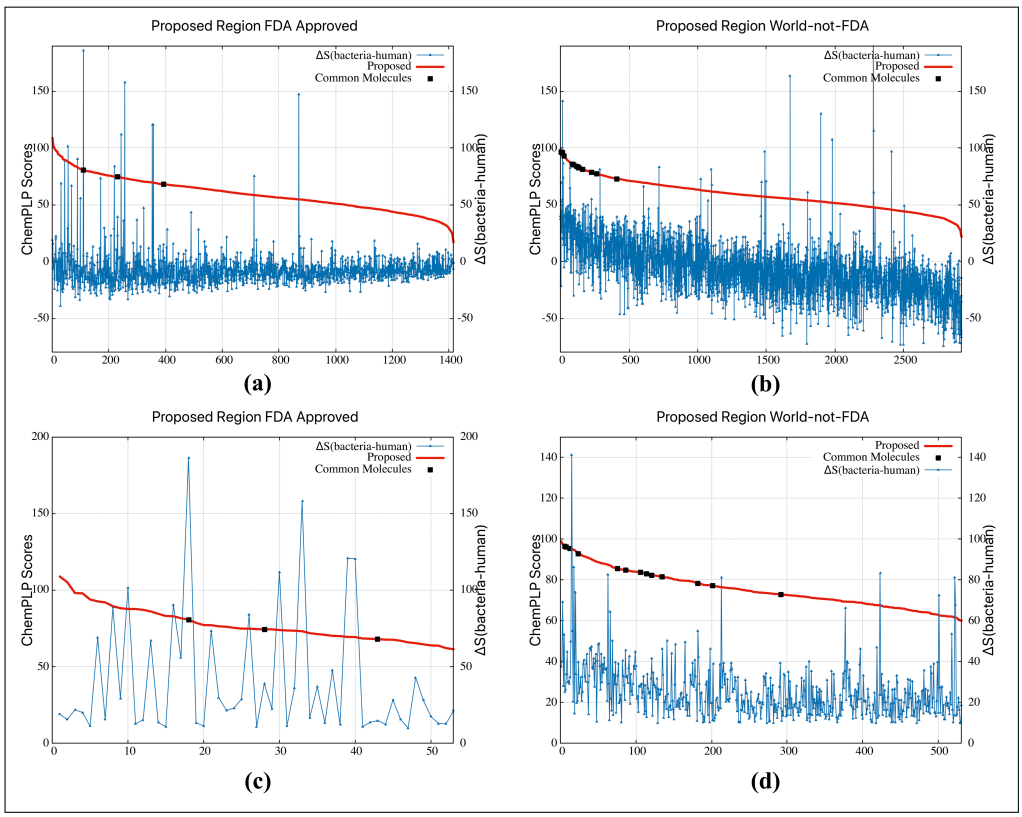


Figure 3.19. Docking result for Proposed region on *Sa*PK and hPK (a) FDA Approved Compounds, (b) World-not-FDA Compounds , (c) Remaining FDA Approved compounds after elimination and (d) Remaining World-not-FDA Compounds after elimination.

As a result of the examination of another proposed region, which was identified as Cluster 17, it was observed that the positive distribution in Figure 3.20 a section and Figure 3.20 b section for FDA and World-not-FDA molecules was not as high as another proposed region. However, looking at the difference between bacteria and human values, it was observed that values with high scores were still obtained. It was observed from Figure 3.20 c section and Figure 3.20 d section that some of the common molecules in both types were much better bound to the bacteria.

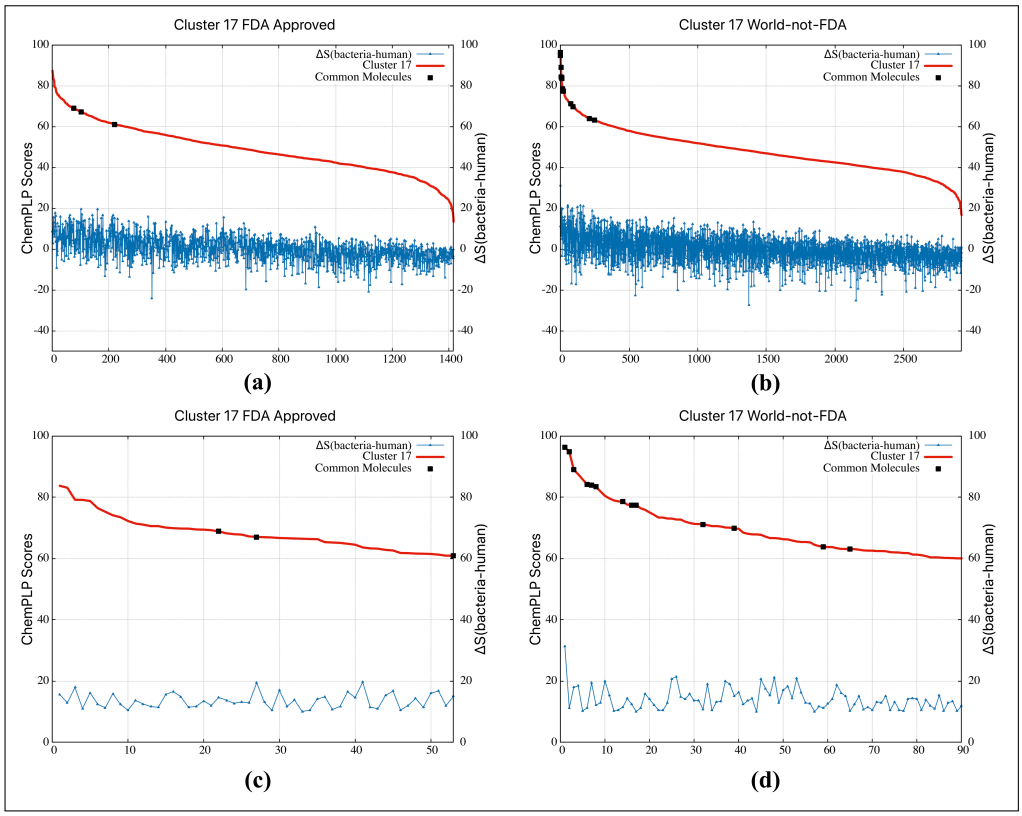


Figure 3.20. Docking result for Cluster 17 region on *Sa*PK and hPK (a) FDA Approved Compounds, (b) World-not-FDA Compounds , (c) Remaining FDA Approved compounds after elimination and (d) Remaining World-not-FDA Compounds after elimination.

These calculations were made for the other targeted region, which defined as the known region. It was mentioned in the previous stages that an already known inhibitor was bound to this region. Therefore, excellent positive distribution and high difference values were expected in this region. In Figure 3.21 a section and Figure 3.21 b section,

it was observed that the positive distribution was as high as expected. From Figure 3.21 c section and Figure 3.22 d section, after the elimination process, the results obtained for this known region were observed to be significant.

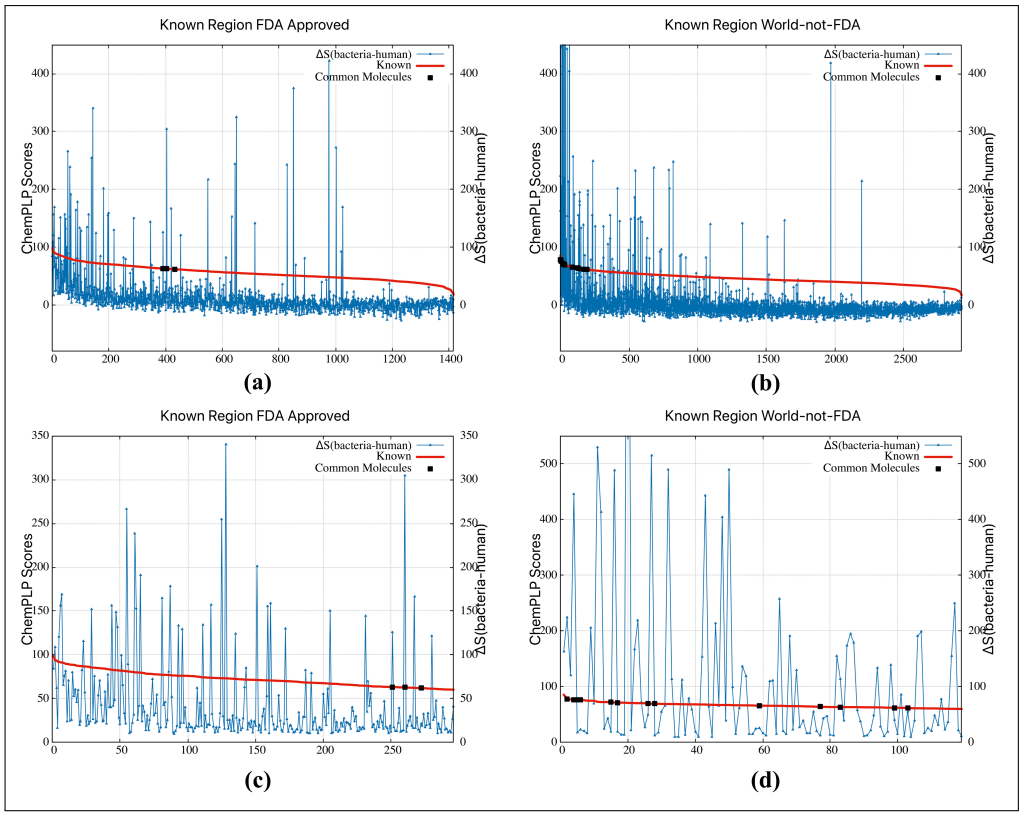


Figure 3.21. Docking result for Known region on *Sa*PK and hPK (a) FDA Approved Compounds, (b) World-not-FDA Compounds , (c) Remaining FDA Approved compounds after elimination and (d) Remaining World-not-FDA Compounds after elimination.

To sum up, positive distribution in the known allosteric region was higher. Direction of distribution in the proposed region (top druggable site) was almost similar to the distribution in the known allosteric region. Moreover, for cluster 17 region, distribution towards the positive and negative direction was slightly similar. For the graphs shown in the Figure 3.19, Figure 3.20, and Figure 3.21, blue, red and black colors symbolized differences between *Sa*PK and *h*PK scoring function values, *Sa*PK scoring functions, and common molecules after elimination processes, respectively. Our target molecule should have the following properties: It should be a molecule with a

high scoring function for bacteria, show a positive distribution and difference in score values between human and bacteria should be also high. Compounds obtained as a result of elimination processes were shown in the Table 3.10 and Table 3.11 not only for FDA approved but also World-not-FDA. Common molecules related to these tables were represented in Figure 3.22 and Figure 3.23, respectively.

Table 3.10. Compounds acquired as a result of docking with 1416 FDA molecules for 3 different regions. (PRO, 17, and 130 represent Proposed Region, Cluster 17 Region, and Known Region, respectively.)

PYRUVATE KINASE (PK) – FDA APPROVED										
ZINC ID	**MOL. W.**	**$S_{ChemPLP}$ (*S.aureus*)**			**$S_{ChemPLP}$ (*H.sapiens*)**			**ΔS (*S.aureus* - *H.sapiens*)**		
		PRO	**17**	**130**	**PRO**	**17**	**130**	**PRO**	**17**	**130**
ZINC000169289767	872.894	80.77	68.82	62.75	-105.33	54.13	-241.95	186.10	14.96	304.70
ZINC000068202099	485.506	74.47	67.02	63.17	35.48	47.58	-62.32	38.99	19.44	125.49
ZINC000001611274	421.453	68.03	60.90	61.96	53.07	45.88	40.08	14.96	15.02	21.88

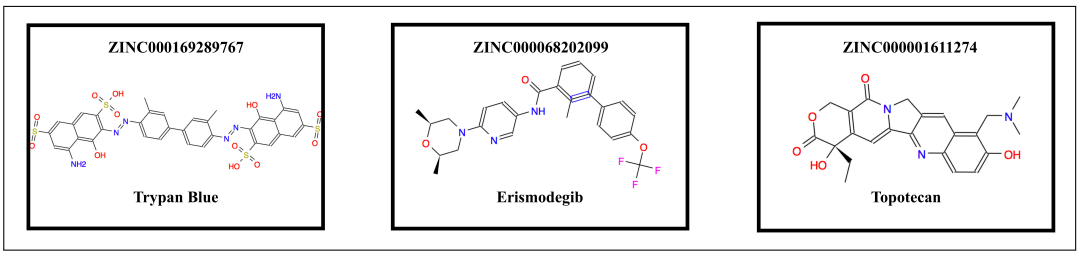


Figure 3.22. Common FDA molecules for 3 different regions.

Table 3.11. Compounds acquired as a result of docking with 2922 Worl-not-FDA molecules for 3 different regions. (PRO, 17, and 130 represent Proposed Region, Cluster 17 Region, and Known Region, respectively.)

PYRUVATE KINASE (PK) – World-not-FDA										
ZINC ID	**MOL. W.**	**$S_{ChemPLP}$ (*S.aureus*)**			**$S_{ChemPLP}$ (*H.sapiens*)**			**ΔS (*S.aureus - H.sapiens*)**		
		PRO	**17**	**130**	**PRO**	**17**	**130**	**PRO**	**17**	**130**
ZINC000003875376	556.790	96.24	96.27	76.11	62.19	64.94	52.89	34.05	31.33	23.22
ZINC000095604030	574.843	84.79	83.98	72.37	39.38	64.58	52.50	45.41	19.40	19.87
ZINC000011677890	627.826	96.48	88.99	77.98	71.13	71.05	-145.79	25.35	17.94	223.77
ZINC000001529444	416.690	78.42	69.74	69.42	23.27	54.58	56.44	55.15	15.16	12.98
ZINC000095617640	511.630	82.43	77.46	62.20	41.03	65.02	22.00	41.40	12.44	40.20
ZINC000004214779	530.790	92.72	83.36	70.16	52.98	71.17	19.76	39.74	12.19	50.40
ZINC000095617639	511.630	81.43	78.55	66.02	35.22	67.02	40.08	46.21	11.53	25.94
ZINC000001545566	444.324	72.87	63.84	71.40	55.46	52.57	51.91	17.41	11.27	19.49
ZINC000008221075	410.771	83.68	84.12	76.24	55.28	72.86	57.87	28.40	11.26	18.37
ZINC000008214943	536.888	95.47	94.81	77.05	63.09	83.62	-368.35	32.38	11.19	445.40
ZINC000043205655	574.991	85.69	71.15	61.52	52.57	60.29	6.04	33.12	10.86	55.48
ZINC000026671872	562.670	77.10	63.05	63.43	65.78	52.73	-50.80	11.32	10.32	114.23
ZINC000001566899	502.746	83.08	77.46	63.92	70.05	67.40	51.35	13.03	10.06	12.57

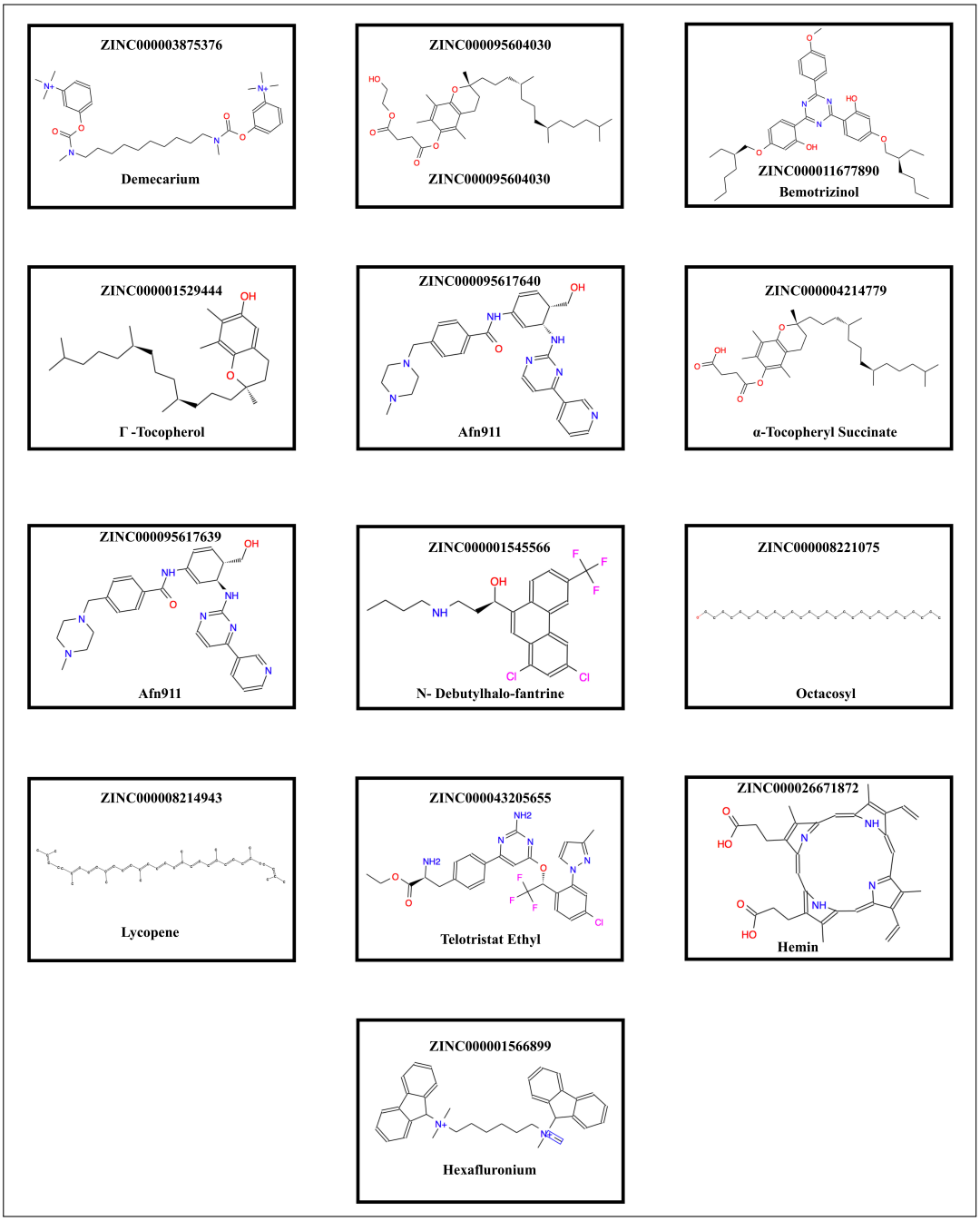


Figure 3.23. Common World-not-FDA molecules for 3 different regions.

4. CONCLUSION AND FUTURE WORKS

Glycolysis occurs in all living organisms. Allosteric enzymes have a crucial role in cells since these enzymes have two essential functions in the cell, which are catalysts of reactions in metabolic pathways and control the rates of these pathways. As mentioned in this whole study, PK is an allosteric enzyme and has significant roles in the cell. Therefore, it is considered as the target enzyme in species-specific drug design studies. In this thesis, which was created by using many computational methods together, newly allosteric regions were found to contribute to species-specific drug design. With the use of the elastic network modeling, hotspot regions have been determined, and the results have become more meaningful by combining these regions with the results obtained from other methods.

As the study progresses, it was observed that the effect of the proposed regions on protein dynamics was almost the same as interaction with the allosteric regions already present in the literature, and sometimes even better results were obtained. Our approach was successful in detecting almost all experimentally known allosteric binding sites. After a detailed analysis, three different regions were proposed for potential drug target sites for *S.aureus* and one region for *L.mexicana*. As a result, allosteric binding sites might be identified, and new drugs might be developed using this approach. In addition, this approach could be used for three allosteric enzymes in the glycolytic pathway, and drugs that are likely to bind to all three enzymes could be found via virtual screening.

Docking studies were carried out by using 1416 FDA molecules and 2922 World-not-FDA molecules. In *Sa*PK docking studies were performed for top druggable site, the second proposed site, and known allosteric site. Corresponding regions in *h*PK were also docked. As a result of these studies, 3 FDA approved and 13 World-not-FDA common molecules, which did not suitably bind to the human species but bound to the bacteria species, were obtained for the targeted three regions.

For further studies, it is recommended to perform Molecular Dynamics (MD) simulations in successful docking poses to observe the stabilization of the molecules found as a result of the docking. In this way, the effect of regions proposed from this study will be proved by MD studies and will demonstrate the accuracy of drug molecules that are considered to be used in the next stage.

REFERENCES

Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., and Bahar, I. (2001). Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophysical Journal*, 80(1):505–515.

Axerio-Cilies, P., See, R. H., Zoraghi, R., Worral, L., Lian, T., Stoynov, N., ... Cherkasov, A. (2012). Cheminformatics-driven discovery of selective, nanomolar inhibitors for staphylococcal pyruvate kinase. *ACS Chemical Biology*, 7(2):350–359.

Ayyildiz, M., Celiker, S., Ozhelvaci, F., and Akten, E. D. (2020). Identification of alternative allosteric sites in glycolytic enzymes for potential use as species-specific drug targets. *Frontiers in Molecular Biosciences*, 7(88).

Bahar, I., Atilgan, A. R., and Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding and Design*, 2(3):173–181.

Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., ... Zardecki, C. (2002). The protein data bank. *Acta Crystallographica Section D: Biological Crystallography*, 28(1):235–242.

Brenke, R., Kozakov, D., Chuang, G. Y., Beglov, D., Hall, D., Landon, M. R., ... Vajda, S. (2009). Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics*, 25(5):621–627.

Clackson, T. and Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science*, 267(5196):383–386.

Doruker, P., Atilgan, A. R., and Bahar, I. (2000). Dynamics of proteins predicted by molecular simulations and analytical approaches: Application to α-amylase inhibitor. *Proteins: Structure, Function and Genetics*, 40(3):512–524.

Fenton, A. W. (2012). Identification of allosteric-activating drug leads for human liver pyruvate kinase. *Methods in Molecular Biology*, 796:369–382.

Foster, T. (1996). *Staphylococcus*. Medical Microbiology. 4th edition., Galveston (TX): University of Texas Medical Branch at Galveston.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., ... Shenkin, P. S. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, 47(7):1739–1749.

Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of Medicinal Chemistry*, 28(7):849–857.

Guarnera, E., Tan, Z. W., Zheng, Z., and Berezovsky, I. N. (2017). AlloSigMA: Allosteric signaling and mutation analysis server. *Bioinformatics*, 33(24):3996–3998.

Guo, J. and Zhou, H. X. (2016). Protein allostery and conformational dynamics. *Chemical Reviews*, 116(11):6503–6515.

Henikoff, S. and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, 89(22):10915–10919.

Israelsen, W. J. and Vander Heiden, M. G. (2015). Seminars in cell developmental biology. *Methods in Molecular Biology*, 43:43–51.

Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, 267(3):727–748.

Korb, O., Stützle, T., and Exner, T. E. (2009). Empirical scoring functions for advanced protein-ligand docking with PLANTS. *Journal of Chemical Information and Modeling*, 49(1):84–96.

Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., ... Vajda, S. (2015). The ftmap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nature Protocols*, 10(5):733–755.

Kung, C., Hixon, J., Choe, S., Marks, K., Gross, S., Murphy, E., ... Dang, L. (2012). Small molecule activation of pkm2 in cancer cells induces serine auxotrophy. *Chemistry and Biology*, 19(9):1187–1198.

Kurkcuoglu, Z., Findik, D., Akten, E. D., and Doruker, P. (2015). How an inhibitor bound to subunit interface alters triosephosphate isomerase dynamics. *Biophysical Journal*, 109(6):1169–1178.

Levy, E. D. (2010). A simple definition of structural regions in proteins and its use in analyzing interface evolution. *Journal of Molecular Biology*, 403(4):660–670.

Ma, D. L., Chan, D. S. H., and Leung, C. H. (2013). Chemical society reviews. *Current Opinion in Structural Biology*, 42(5):2130–41.

Mattevi, A., Bolognesi, M., and Valentini, G. (1996). The allosteric regulation of pyruvate kinase. *FEBS Letters*, 389(1):15–19.

McConkey, B. J., Sobolev, V., and Edelman, M. (2002). The performance of current methods in ligand-protein docking. *Current Science*, 83(7):845–855.

Meng, X.-Y., Zhang, H.-X., Mezei, M., and Cui, M. (2012). Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer Aided-Drug Design*, 7(2):146–157.

Morgan, H. P., McNae, I. W., Nowicki, M. W., Hannaert, V., Michels, P. A., Fothergill-Gilmore, L. A. and Walkinshaw, M. D. (2010). Allosteric mechanism of pyruvate kinase from Leishmania mexicana uses a rock and lock model. *Journal of Biological Chemistry*, 285(17):12892–12898.

Morgan, H. P., Zhong, W., McNae, I. W., Michels, P. A., Fothergill-Gilmore, L. A., and Walkinshaw, M. D. (2014). Structures of pyruvate kinases display evolutionarily divergent allosteric strategies. *Royal Society Open Science*, 1(1):140120.

Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, 19(14):1639–1662.

Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453.

Oprea, T. I. and Mestres, J. (2012). Drug repurposing: far beyond new targets for old drugs. *The AAPS Journal*, 14(4):759–763.

Rice, P., Longden, L., and Bleasby, A. (2000). EMBOSS: The european molecular biology open software suite.

Riefke, B., Mumberg, D., Kroemer, G., Keun, H., Petersen, K., and Steger-Hartmann, T. (2007). Oncogenes meet metabolism. From deregulated genes to a broader understanding of tumour physiology. Preface. *Ernst Schering Found Symp Proc.*, 4:V–VII.

Rigden, D. J., Phillips, S. E., Michels, P. A., and Fothergill-Gilmore, L. A. (1999). The structure of pyruvate kinase from Leishmania mexicana reveals details of the allosteric transition and unusual effector specificity. *Journal of Molecular Biology*, 291(3):615–635.

Robert, X. and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1):320–324.

Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., and Zehfus, M. H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science*, 229(4716):834–838.

Schormann, N., Hayden, K. L., Lee, P., Banerjee, S., and Chattopadhyay, D. (2019). Protein science. *Current Computer Aided-Drug Design*, 28(10):1771–1784.

Schrödinger, L. (2015). The PyMol molecular graphics system, Versión 1.8.

Sterling, T. and Irwin, J. J. (2015). ZINC 15 - Ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55(11):2324–2337.

Team, R. C. (2019). R: A Language and Environment for Statistical Computing.

Tien, M. Z., Meyer, A. G., Sydykova, D. K., Spielman, S. J., and Wilke, C. O. (2013). Maximum allowed solvent accessibilites of residues in proteins. *PLoS ONE*, 8(11):e80635.

Torres-Guerrero, E., Quintanilla-Cedillo, M. R., Ruiz-Esmenjaud, J., and Arenas, R. (2017). Leishmaniasis: a review. *F1000Research*, 750(6).

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. *Proteins: Structure, Function and Genetics*, 52(4):609–623.

Volkamer, A., Kuhn, D., Rippmann, F., and Rarey, M. (2012). Dogsitescorer: A web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*, 28(15):2074–2075.

Zhang, Y., Doruker, P., Kaynak, B., Zhang, S., Krieger, J., Li, H. and Bahar, I. (2020). Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior. *Current Opinion in Structural Biology*, 62:14–21.

Zhong, W., Cui, L., Goh, B. C., Cai, Q., Ho, P., Chionh, Y. H., ... Dedon, P. C. (2017). Allosteric pyruvate kinase-based "logic gate" synergistically senses energy and sugar levels in *Mycobacterium tuberculosis*. *Nature Communications*, 8(1):1986.

Zhong, W., Morgan, H. P., McNae, I. W., Michels, P. A., Fothergill-Gilmore, L. A., and Walkinshaw, M. D. (2013). 'In crystallo' substrate binding triggers major domain movements and reveals magnesium as a co-activator of *Trypanosoma brucei* pyruvate kinase. *Acta Crystallographica Section D: Biological Crystallography*, 69(Pt 9):1768–1779.

Zoraghi, R., Worrall, L., See, R. H., Strangman, W., Popplewell, W. L., Gong, H., ... Reiner, N. E. (2011). Methicillin-resistant Staphylococcus aureus (MRSA) pyruvate kinase as a target for bis-indole alkaloids with antibacterial activities. *Journal of Biological Chemistry*, 286(52):44716–44725.

APPENDIX A: PDB STRUCTURES AND CONSENSUS SITES

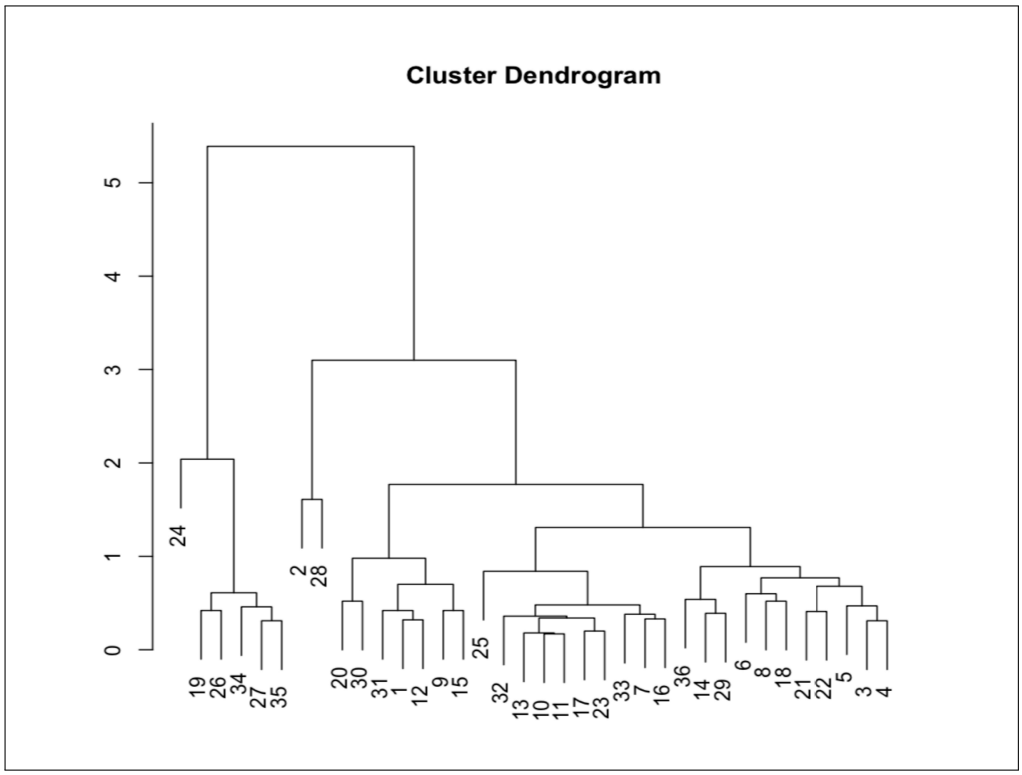


Figure A.1. Clustering Tree for *H.sapiens* Tetramer Structures according to RMSD Values.

”

Table A.1. Name of PDB structures in Figure A.1.

1	1T5A	7	3BJF	13	3ME3	19	4FXJ	25	4QG8	31	5X0I
2	1ZJH	8	3BJT	14	3SRD	20	4G1N	26	4QG9	32	5X1V
3	2VGB	9	3G2G	15	3SRF	21	4IMA	27	4QGC	33	5X1W
4	2VGF	10	3GQY	16	3SRH	22	4IP7	28	4RPP	34	6GG4
5	2VGG	11	3GR4	17	3U2Z	23	4JPG	29	4WJ8	35	6GG5
6	2VGI	12	3H6O	18	4FXF	24	4QG6	30	4YJ5	36	6GG6

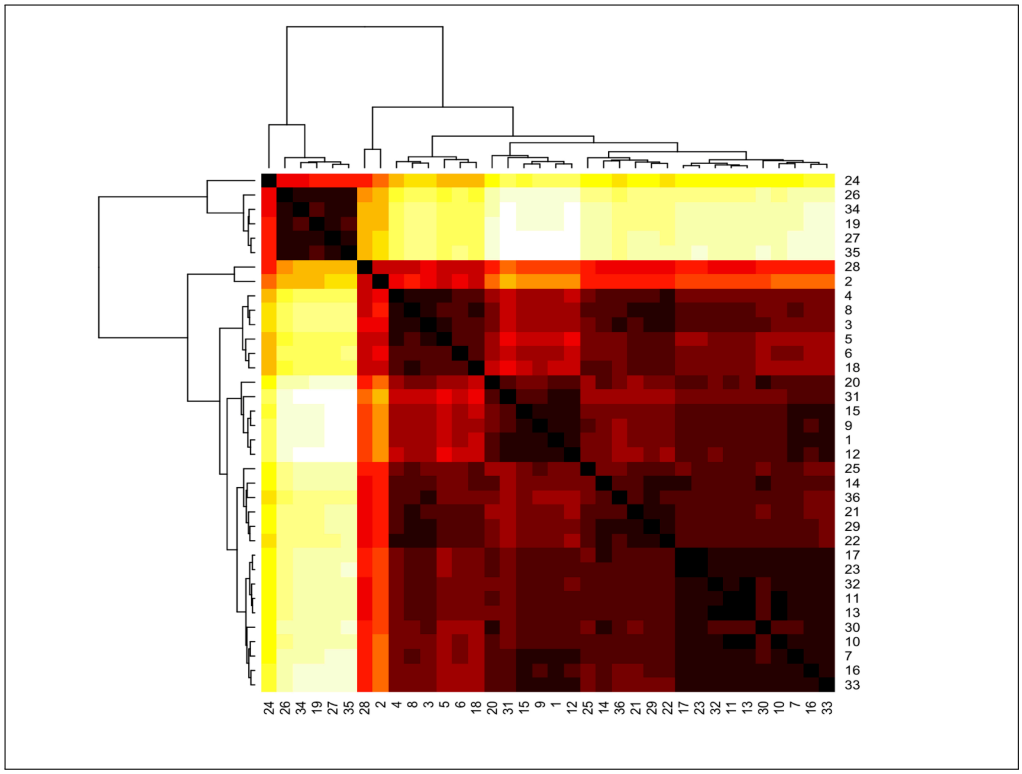


Figure A.2. Heatmap of *H.sapiens* Tetramer Structures.

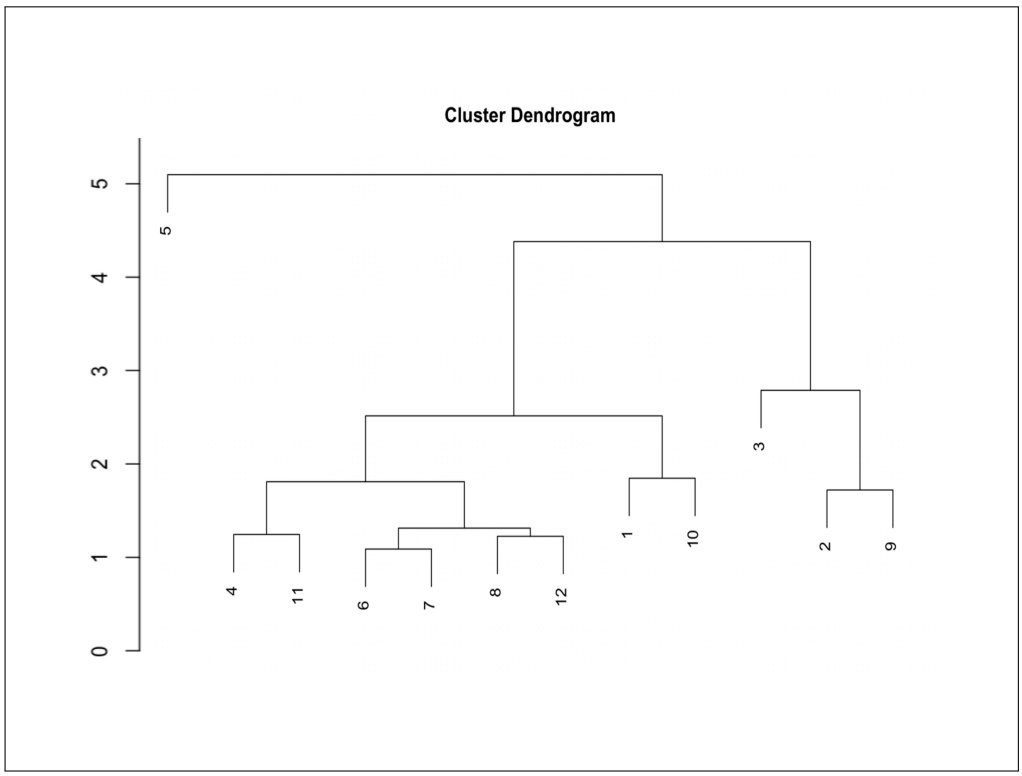


Figure A.3. Clustering Tree for *L.mexicana* Tetramer Structures according to RMSD Values.

Table A.2. Name of PDB structures in Figure A.3.

1	1PKL	4	3HQN	7	3KTX	10	3QV7
2	3E0V	5	3HQQ	8	3PP7	11	3QV8
3	3E0W	6	3IS4	9	3QV6	12	3SRK

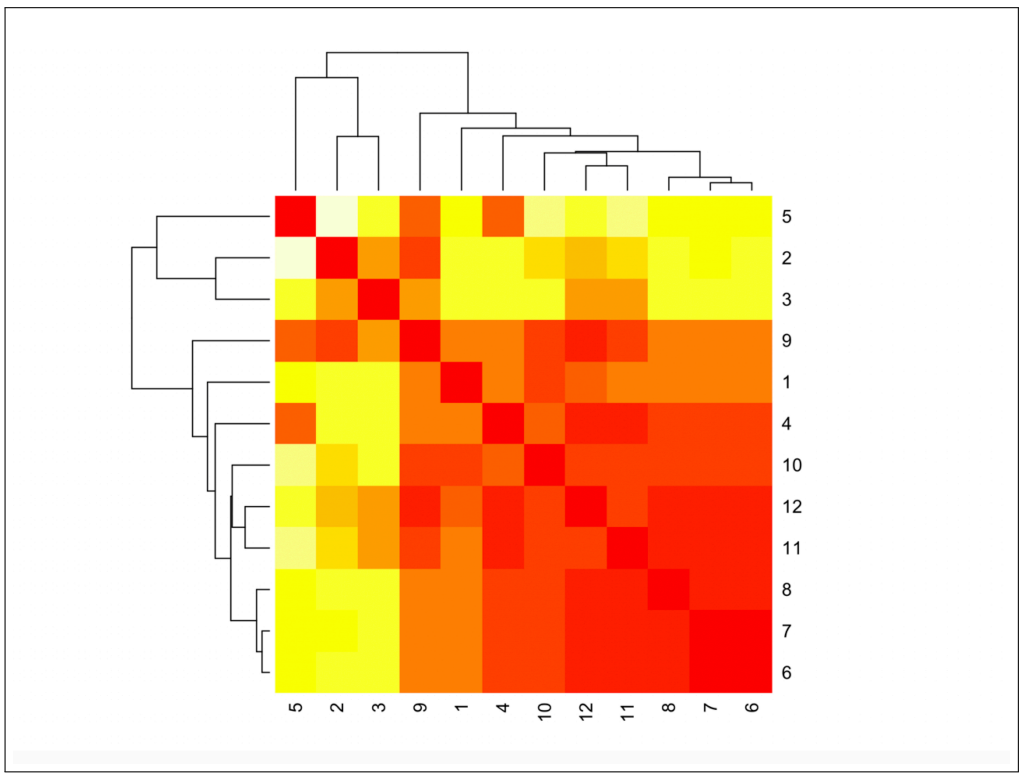


Figure A.4. Heatmap of *L.mexicana* Tetramer Structures.

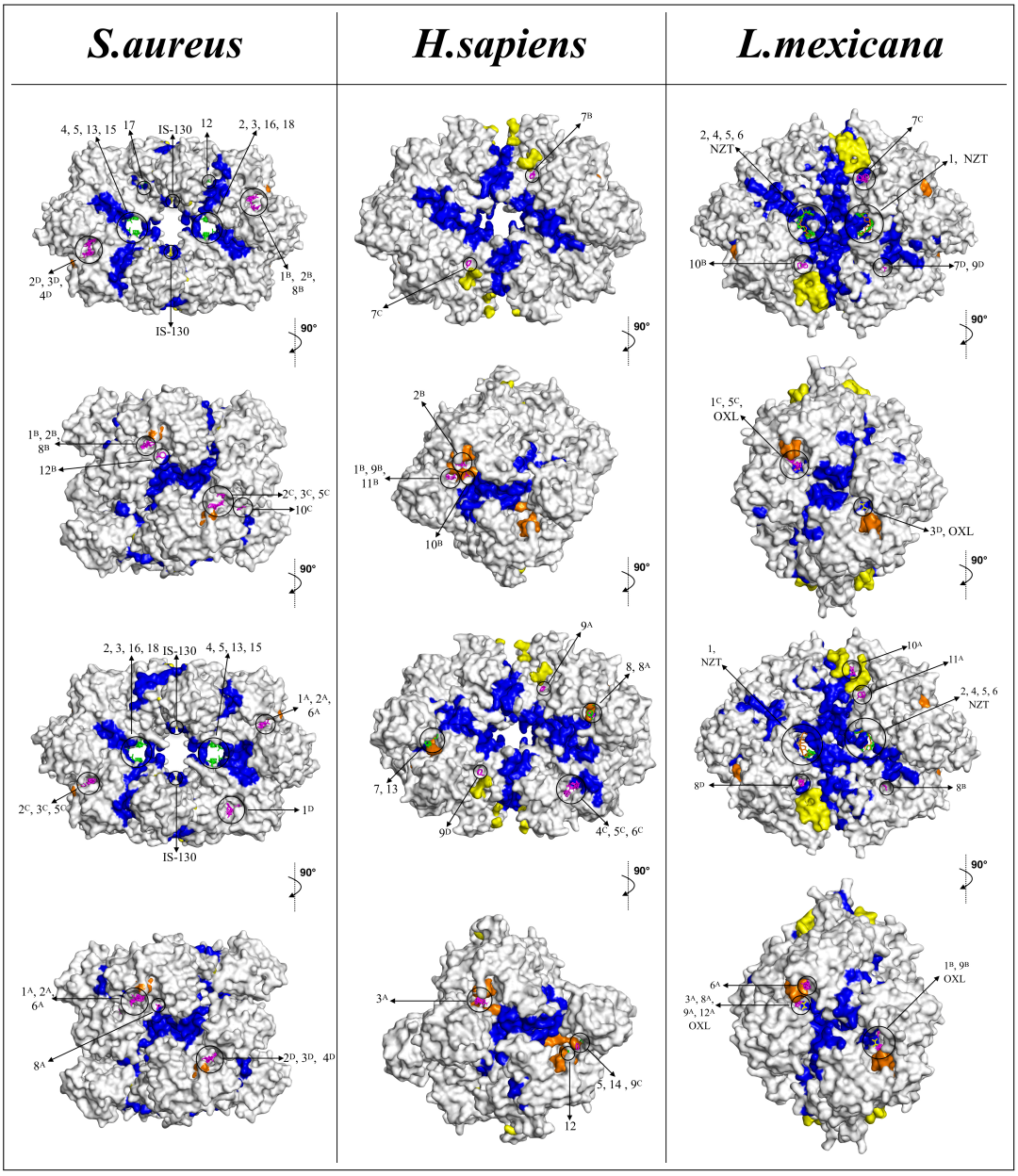


Figure A.5. Representation of all consensus sites above 50% frequency shift value for Pyruvate Kinase in *S.Aureus*, *L.mexicana* and *H.Sapiens*. Whole structure CSs results are shown in green. Chain by chain structure results are shown in magenta color