KADIR HAS UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

CENTRAL TUNNEL AS AN ALTERNATIVE ALLOSTERIC SITE IN GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE FOR POTENTIAL USE AS SPECIES-SPECIFIC DRUG TARGET

SERKAN CELIKER

MASTER'S THESIS

ISTANBUL June 2020

SERKAN CELIKER

M.Sc. Thesis

2020

CENTRAL TUNNEL AS AN ALTERNATIVE ALLOSTERIC SITE IN GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE FOR POTENTIAL USE AS SPECIES-SPECIFIC DRUG TARGET

SERKAN CELIKER

Submitted to the Graduate School of Science and Engineering
in partial fulfilment of the requirements for the degree of
Master of Science
in
Computational Biology and Bioinformatics

KADIR HAS UNIVERSITY

June 2020

DECLARATION OF RESEARCH ETHICS /
METHODS OF DISSEMINATION

I, SERKAN CELIKER, hereby declare that;

- this Master's Thesis/Project/PhD Thesis is my own original work and that due references have been appropriately provided on all supporting literature and resources;

- this Master's Thesis/Project/PhD Thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;

- I have followed "Kadir Has University Academic Ethics Principles" prepared in accordance with the "The Council of Higher Education's Ethical Conduct Principles"

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

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

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

SERKAN CELIKER

_____

DATE AND SIGNATURE

KADIR HAS UNIVERSITY

SCHOOL OF GRADUATE STUDIES

ACCEPTANCE AND APPROVAL

This work entitled CENTRAL TUNNEL AS AN ALTERNATIVE ALLOSTERIC SITE IN GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE FOR POTENTIAL USE AS SPECIES-SPECIFIC DRUG TARGET **THESIS TITLE** prepared by **SERKAN CELIKER** has been judged to be successful at the defense exam held on _____ and accepted by our jury as **MASTER THESIS**.

APPROVED BY:

Prof. Ebru Demet AKTEN AKDOGAN (Advisor)   _____
Kadir Has University

Prof. Kemal YELEKÇİ   _____
Kadir Has University

Asst. Prof. Özge KÜRKÇÜOĞLU   _____
Istanbul Technical University

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

_____
Prof. Dr. Sinem AKGÜL AÇIKMEŞE
Dean of School of Graduate Studies
DATE OF APPROVAL: (Day/Month/Year)

TABLE OF CONTENT

CENTRAL TUNNEL AS AN ALTERNATIVE ALLOSTERIC SITE IN GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE FOR POTENTIAL USE AS SPECIES-SPECIFIC DRUG TARGET

ABSTRACT

One of the newest approaches in drug discovery studies is drug design for potential allosteric regions in protein. These regions are based on the design of inhibitors that can act remotely on the active site, rather than those that directly target the active sites of the enzymes. The rationale behind this approach is that the active site is evolutionarily more conserved than other parts of the enzyme which prevent the design of selective inhibitors. In this study, the allosteric enzyme *glyceraldehyde-3-phosphate dehydrogenase* (GADPH) that functions in the glycolytic pathway was targeted as the glycolytic pathway is vital for energy production in almost all organisms. The aim of this study was to propose potential allosteric regions in the receptor of the infected organism, which will be the parasite *Tripanosoma Cruzi (T.cruzi)* and the bacteria *Staphylcoccus Aureus (S.aureus)*. Computational solvent mapping was used to identify all binding sites on *T.cruzi, S.aureus* and *H.sapiens* receptors. The clusters obtained as a result of computational solvent mapping were observed to be at the interface areas for all three species. Then, a theoretical model with an elastic network model was used to extract regions that most affect the global dynamics of the receptor. Sequence similarities between *H.Sapiens-T.cruzi* and *H.Sapiens-S.aureus* structures were investigated. Finally, the designated tunnel region was suggested to be the potential allosteric target region. Tunnel region identified in these infected organisms were docked with molecules approved by "FDA" and "World-not-FDA ". A total of 7 and 25 potential molecules were detected for "FDA" and "World-not-FDA", which are thought to inhibit the infecting bacteria, respectively.

Keywords: glyceraldehyde-3-phosphate dehydrogenase, elastic network model, allostery, species-specific drug design, solvent mapping, docking.

CENTRAL TUNNEL AS AN ALTERNATIVE ALLOSTERIC SITE IN GLYCERALDEHYDE-3-PHOSPHATE-DEHYDROGENASE FOR POTENTIAL USE AS SPECIES SPECIFIC DRUG TARGETS

ÖZET

İlaç keşif çalışmalarında en yeni yaklaşımlardan biri proteindeki potansiyel allosterik bölgeler için ilaç tasarımıdır. Bu bölgeler enzimlerin aktif bölgelerini doğrudan hedefleyen inhibitörler yerine aktif bölge üzerine uzaktan bir etki yapabilen inhibitörlerin tasarımına dayanır. Bu yaklaşımın arkasındaki ana sebep, aktif bölgenin, seçici inhibitörlerin tasarımını önleyen enzimin diğer kısımlarından evrimsel olarak daha fazla korunmuş olmasıdır. Bu çalışmada, glikolitik yolakta işlev gören enzimler, glikolitik yolak hemen hemen tüm organizmalarda enerji üretimi için hayati önem taşıdığından hedeflenmiştir. Bu çalışmada amaç, Tripanosoma Cruzi (*T.cruzi*) ve Staphylcoccus Aureus (*S.aureus*) olan enfekte organizmanın reseptöründeki allosterik bölgeleri belirlemektir.

Çalışılan reseptör, glikolitik yolda bir allosterik enzim olan "gliseraldehid-3-fosfat dehidrojenaz" dır (GADPH). İlk olarak, *T.cruzi*, *S.aureus* ve *H.sapiens* reseptörlerindeki tüm bağlanma bölgelerini belirlemek için hesaplamalı çözücü haritalaması kullanıldı. Daha sonra, alıcının küresel dinamiklerini en çok etkileyen bölgeleri çıkarmak için elastik ağ modeli olan teorik bir model kullanılmıştır. *H.sapiens-T.cruzi* ve *H.sapiens-S.aureus* arasında hizalama araçları kullanılarak tespit edilen bölgelerin türe özgü ne kadar spesifik olduğu tespit edilmiştir.

Son olarak, tüm sonuçlar birleştirildi ve tünel bölgesinin potansiyel allosterik bölge olduğuna karar verildi. Bu bölge için daha sonra belirlenen moleküller uygulanarak enfekte olan bu organizmaları inhibe edecek "FDA" için 7, "World-not-FDA" için 25 potansiyel molekülün varlığı tespit edildi.

Anahtar Sözcükler: Gliseraldehid-3-fosfat dehidrojenaz, Kaba taneli elastik ağ modeli, Türe özgü ilaç tasarımı, çözücü haritalaması, doking işlemi.

ACKNOWLEDGEMENTS

To my wife

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS/ABBREVIATIONS

ACE	Analytic Continuum Electrostatic
ANM	Anisotropic network model
ASP	Astex Statistical Potential
ATP	Adenosine triphosphate
Cα	Carbon alpha
CS	Crosscluster
ENM	Elastic network model
FDA	U.S Food and Drug Adminitration
FFT	Fourier transfor
GNM	Gaussian network model
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GLY,G	Glycine
G3P	Glyceraldehyde-3-phosphate
*h*GADPH	*H.sapiens* glyceraldehyde-3-phosphate dehydrogenase
kDa	KiloDalton
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NMA	Normal mode analysis
NMR	Nuclear Magnetic Resonance
PLP	Piecewise Linear Potential
PDB	Protein Data Bank
PDB ID	Protein Data Bank identification code
rSASA	Relative Solvent Accessible Surface Area
RMSD	Root Mean Square Deviation
*Sa*GADPH	*S.aureus* glyceraldehyde-3-phosphate dehydrogenase
SER, S	Serine
TcGADPH	*T.cruzi* glyceraldehyde-3-phosphate dehydrogenase
TYR,Y	Tyrosine
vdW	van der Waal

LIST OF SYMBOLS

Å	Angstrom
λ_i	Eigenvalue of Mode i
λ_i^0	Eigenvalue of Reference Mode i
$\%s_i$	Shift for Mode i
Δ	Difference operator
R_c	Cutoff value
ΔG	Gibbs free energy

1 INTRODUCTION

1.1 Glycolytic Pathway Plays a Vital Role in Energy Production

Glycolysis is a metabolic pathway that allows glucose to be separated into pyruvate and hydrogen ions and to transfer the high free energy released during this destruction to molecules such as adenosine triphosphate (ATP) and NADH. At this stage, as a result of the reaction that started with the introduction of two molecules of ATP to break down the 6-Carbon glucose, the conversion of two molecules to 3-Carbon smaller molecules with ATP gain is achieved (Barnett 2003; Meyerhof and Junowicz-Kocholaty 1943). The fact that glycolysis can occur in the absence (anaerobic) or presence (aerobic) of oxygen makes glycolysis indispensable for every living cell in terms of energy production. Successive reactions in the glycolytic pathway are carried out with ten different enzymes as illustrated in Figure 1.1, of which phosphofructokinase, glyceraldehyde-3-phosphate dehydrogenase and pyruvate kinase are functioning allosterically.
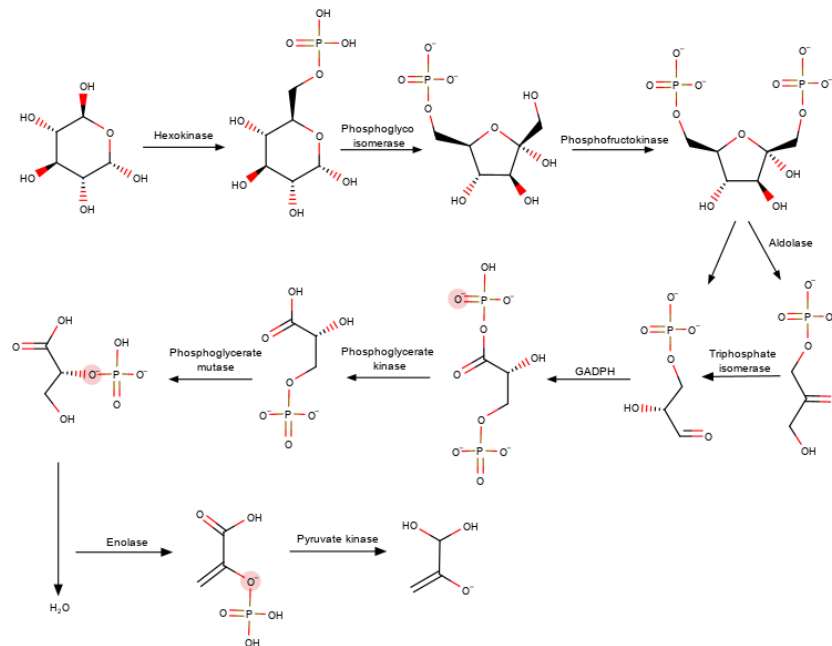
Figure 1.1 Glycolytic pathway and the corresponding enzymes catalysing seach reaction.

1.2 Overview of Allostery

Allostery can be generally defined as increasing (positive) or decreasing (negative) activity as a result of a strong communication between a catalytic region and an effector region located at a distant site. This occurs by connecting a ligand or an effector to a region that will trigger global dynamics, making the change in the conformation of the catalytic region either through facilitating or preventing the access of the substrate to the catalytic region. Therefore, allostery is of great importance in regulating protein activity (J. Monod, Changeux, and Jacob 1963; Koshland and Hamadani 2002; Perutz 1989). The first allosteric model (MWC model) was introduced in the 1960s. In this model, the structures of allosteric proteins with identical protomers which can be found in at least two conformational states (T and R), were described (Jacque Monod, Wyman, and Changeux 1965). Subsequently, a powerful concept was introduced by Weber regarding the allosteric mechanism (Weber 1972). It was claimed that the ligand binding procedure changed the population of the conformational states existing in the dynamic ensemble. Briefly, allostery was defined as population shift or re-distribution in conformational states.

Gunasekaran and his coworkers stated that non-allosteric proteins can display allosteric properties by introducing point mutations, appropriate drugs and various external conditions (Gunasekaran, Ma, and Nussinov 2004). Studies on the effect of distal mutations on global dynamics of protein support the above view (Pan, Lee, and Hilser 2000).

Allostery as an intrinsic feature of protein played an important role in the emergence of alternative methods in drug design effort (Christopoulos 2002; Ellis 1997). Rather than the catalytic region, the focus is on alternative allosteric regions called effectors. The low degree of sequence conservation displayed by these effector regions relative to the catalytic region, indicates a high interspecific sequence variability, thus can be used as target regions in species-specific drug design studies.

Various methods are used to investigate alternative allosteric regions. While some of these are evaluating the static data obtained from NMR or X-ray crystallography.There exist theoretical algorithms such as normal mode analysis (NMA) using coarse-grained

elastic network model (Bahar and Rader 2005; Tama and Brooks 2006) which takes into account large scale motions. Another alternative method is molecular dynamic simulations (Lou and Cukier 2006; Hornak et al. 2006). It is important to identify the slowest frequeny modes to understand large-scale motions which are intrinsic features of the protein (Tobi and Bahar 2005). It is known that local disturbances in conformation, represented by high frequency modes, also have an effect on disrupting the transmission between distant connections in protein (Hammes-Schiffer and Benkovic 2006; Hawkins and McLeish 2006).

1.3 General Structure of Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) and Allostery

GADPH is an allosteric enzyme catalyzing the sixth step of the glycolytic pathway (See Figure 1.1). It converts glyceraldehyde-3-phosphate (G3P) through oxidative phosphorylation into a high-energy compound, 1,3 bisphosphoglycerate, with the help of cofactors such as NAD or NADP. GADPH is formed by a combination of four identical 37 kDa subunits, each of which binds nicotinamide adenine dinucleotide (NAD^+) (See Figure 1.2).

It is known that binding of NAD to subunits in GAPDH allosterically affects the structure of the protein and changes the global dynamics of the protein (Kirschner, Eigen, Bittman, & Voigt, 1966). In another widely known feature of GADPH, the presence of flexible S-loop is known to be important in cofactor binding and allosteric activation of the structure (Biesecker, Ieuan Harris, Thierry, Walker, & Wonacott, 1977) (See Figure 1.2). However, this study does not provide an explanation of how flexible S-loop regulates the structure. In a recent study on Toxoplasma GAPDH 1, some inter- and intra-phosphorylations in the S-loop revealed the effect of structure on allosteric activation, cofactor binding and oligomerization (Dubey et al., 2017). As a conclusion, in this study, *H.Sapiens, S.Aureus* and *T.Cruzi* were determined in GADPH to investigate the presence of potential allosteric sites.
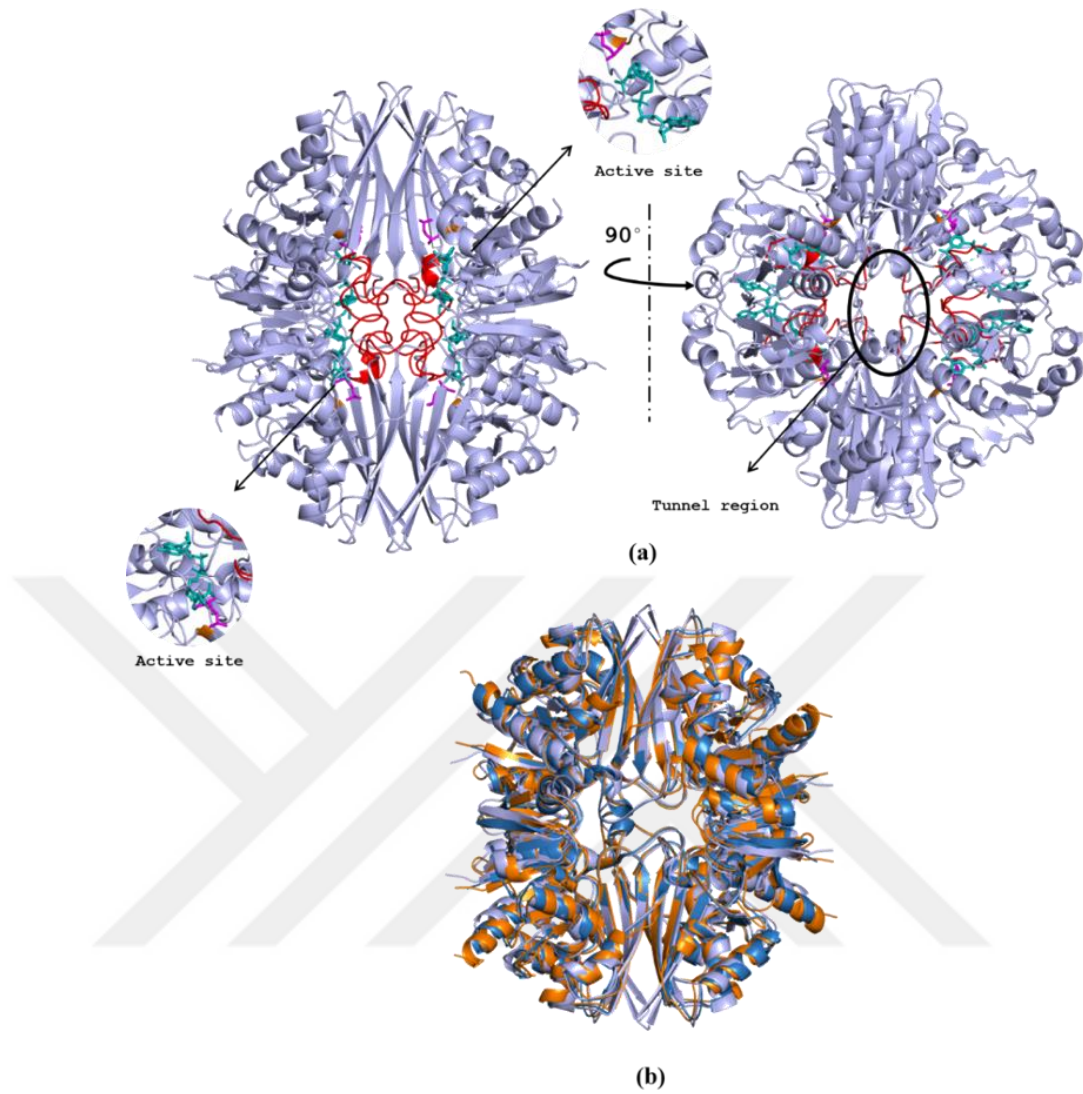
Figure 1.2 a) General structure of GADPH including S-loops, G3P and NAD. b) Structural similarity of three species

2 METHOD

2.1 System Preparation

All X-ray crystallographic structures of GADPH were extracted from PDB database (Berman 2000) for selected species of human (*H.sapiens*), bacteria (*S.aureus*) and parasite (*T.cruzi*) and listed in Table 2.1.

Table 2.1 X-ray structures of GADPH for species from Protein Data Bank (www.rcsb.org).

GADPH	Human (*H.sapiens*)	Bacterium (*S.aureus*)	Parasite (*T.cruzi*)
Monomer	N/A	N/A	N/A
Dimer	N/A	3VAZ	4LSM
Tetramer	5C7L, 5C70, 4WNC, 4WNI, 3PFW, 3H9E, 1U8F, 1ZNQ, 3GPD	5T73, 3K9Q, 3K73, 3KSD, 3KSZ, 3KV3, 3L4S, 3L60, 3LC1, 3LC2, 3LC7, 3LVF, 3HQ4	3IDS, 3DMT, 1QXS, 1ML3, 1K3T

It is known that oligomeric enzymes consisting of identical subunits are commonly arranged in dimer and tetramer structures. The quaternary and dimer structure of the protein contributes to the biological activity and stability especially in quaternary oligomeric enzymes. It is seen that the biological activity in the specified species for GADPH, which is an oligomeric enzyme, occurs in the dimer and tetramer structure. As a result, the study focused on the tetramer structures of the species. The first step in investigating the similarity of protein structures between human and bacteria/parasite was to determine a structure to represent each species. (See Table 2.1) Therefore, the structural similarities and differences of nine homotetrameric structures mentioned for H. sapiens were firstly considered, and root mean squared deviation (RMSD) values of each of them

were calculated with the *super* module of PyMOL graphics visualization tool (Chakraborty et al. 2014). The same process was applied to parasite including five homotetrameric structures and bacteria including thirteen homotetrameric structures and RMSD values were recorded as a matrix (See Appendix for Tables A.1, A.2 and A.3). As a result, heat map and cluster dendrograms were created for each species using R free software (R Development Core Team 2017).

Considering the RMSD values in the matrix, it was observed that the values are quite close to each other. For this reason, the selection of the species representing each species was arbitrary performed. Heatmap results were added to the appendix as extra information (See Figure A.1 in Appendix). As a result for the GAPDH enzyme, *T.cruzi* (PDB id:3DMT) (Guido et al. 2009), *S.aureus* (PDB id:3HQ4) (Mukherjee et al. 2010a) and *H. sapiens* (PDB id:4WNI) (White et al. 2015) were selected from the x-ray structures shown in table 1.

2.2 Sequence and Structural Alignment

After selection of structures representing human, bacterial and parasite species, commonly known sequence and structural based alignment tools were used to identify similarities and differences between human and the infecting organism which is either bacteria or parasite. Based on the Needleman-Wunsch global sequence alignment algorithm (Needleman and Wunsch 1970), the EMBOSS-Needle (Rice, Longden, and Bleasby 2000) web server was used to reveal evolutionarily conserved and nonconserved regions in the aligned sequences of two species. Needleman-Wunsch algorithm was performed for global alignment with default settings of EMBOSS-Needle (Scoring matrix = Blosum62; Gap Open = 10; Gap Extend = 0.5).

Furthermore, structural alignment was performed by PyMOL (Schrödinger 2015) *Super* module. *Super* module performs a residue-based structural superposition regardless of the amino acid identity, using backbone α-Carbon atoms only. Module generates a number of refinement cycles to remove unfit structural segments and minimizes RMSD between the two aligned structures. Calculated RMSD value before and after refinement were recorded to observe the difference. When the two species overlap, they are colored to

show more clearly the distance of the residues to each other. Another PyMOL module, called *Color By RMSD*, was used for coloring, which allows comparing the RMSD values calculated for the α-Carbon backbone of the atomic pair of residues aligned in the two proteins as a β factor.

2.3 Computational Solvent Mapping

FTMap aims to reveal hot spot regions by imitating experimental NMR or X-ray crystallographic studies trying to solve binding sites in protein structure using various organic solvents. It provides the identification of low energy regions called "hot spots" on the surface of the protein. In this study, a widely used solvent mapping tool FTMap (Brenke et al. 2009) (Kozakov et al. 2015) was used. The use of FTMap as a popular in the identification of potential druggable sites on the protein surface is due to the high consistency of FTMap with the results obtained from experimental studies (Wakefield et al. 2019). The method is mainly based on the Fourier transform (FFT) correlation approach, which can be used to exemplify billions of probe poses and to calculate their energy based on an energy function (Brenke et al. 2009). The steps of the principle of FTMap algorithm implementation can be simplified as follows.

- The ligand and water molecules of the protein taken from the PDB database are removed. This protein is then docked with small 16 organic molecules (*acetaldehyde, benzaldehyde, urea, dimethyl ether, acetonitrile, isopropanol, ethane, acetamide, phenol, benzene, methylamine, cyclohexane, ethanol, N, N-dimethylformamide, isotanol and acetanol*), called probes. 2000 poses for each ligand are generated.
- The energy minimization for the complex structure consisting of 2000 poses for each probe is performed with CHARMM potential with the Analytic Continuum Electrostatic (ACE) model (Brenke et al. 2009).
- Minimized probe conformations in the previous stage are made in clusters using a greedy algorithm. After selecting the lowest energy structure, structures within the 3 A˚ RMSD value are collected in a cluster. The creation of the second cluster continues by selecting the second structure with the lowest energy.

- The "hot spot", which is formed by the overlap of different types of probe clusters that overlap at the last stage, has been assembled to the consensus sites.

After removing all water and ligand molecules, the free target receptor was submitted to FTMap server in two different ways. First was the mapping of the overall tetramer. Next attempt was the chain-by-chain mapping in which each monomeric chain was mapped individually. The main objective here was to uncover the existence of all possible hot spot regions (See Figure 2.1)
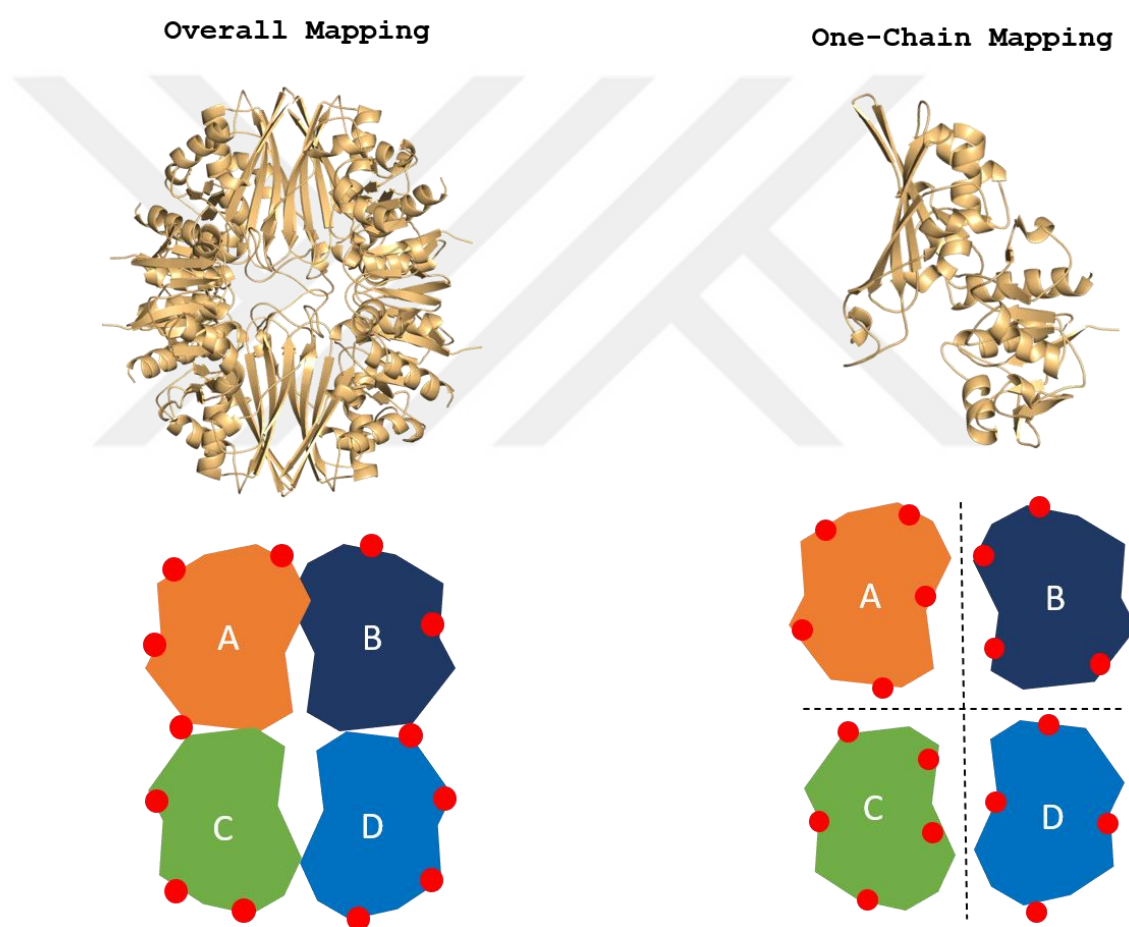
Figure 2.1 General solvent mapping strategy with FTMap.

Probe clusters obtained from monomeric structure were represented on the homotetrameric structure. However, during this representation, some clusters were unlikely to be seen when representing the homotetramer structure. The existence of these clusters, which are defined as inaccessible clusters, was also investigated.

2.4 ENM-Based Residue Scanning

The elastic network model (ENM) is an approach that takes the biological macromolecule as an elastic mass and spring network to predict basic dynamics, i.e. to clarify the internal motions of the protein based on harmonic potential (Tirion 1996; Haliloglu, Bahar, and Erman 1997; Atilgan et al. 2001; Doruker, Atilgan, and Bahar 2000). In the macromolecule, the alpha Carbon atoms of each residue are defined as nodes which are connected to each other by fluctuating bonds known as Hookean springs to form spring networks.The cutoff distance is used as a parameter to define the placement of the springs and connections between the node pairs, and the node pairs within defined cutoff distance are considered to be connected. This optimum cutoff distance is generally taken as 7-8 Å for Gaussian Network Model (GNM) (Kundu et al. 2002) and 13-15Å for ANM (Atilgan et al. 2001). The basic approach that forms the dynamics of the protein is actually the sum of the harmonic potential between these pairs of atoms in a 3D environment.

In this study, a different approach based on standard.ENM method was used. In this method backbone Cα atoms, of residue is represented as a single node. In addition to alpha Carbon atoms, heavy atoms of the side chains of a particular residue are taken up as extra nodes in order to mimic the presence of a ligand (Kurkcuoglu et al. 2015). Then, for each scanned residue, the percentage shift in the ith collective mode's eigenvalue λ_i is calculated (See Equation 2.1).

$$\% \, shift \, for \, mode \, i \, (\%s_i) = \frac{\lambda_i(modified) - \lambda_i(original)}{\lambda_i^0(original)} \times 100 \tag{2.1}$$

λ_i (original) represents the collective mode's eigenvalue calculated according to the standard ENM. The percentage shift value for each residue was calculated as an average over the slowest 20 modes ($\sum_{i=1}^{20}(\%s_i)/20$) and represented by using a color scale (blue-white-red from lowest to highest) β-factor column in PDB file.

The positive values with high percentage shifts obtained for residues represent the positive contribution of residue to protein-ligand collective dynamics, while negative values represent the negative contribution of residue to collective dynamics.As a result, the collective dynamics of the protein was mapped with the frequency shift values

calculated for each residue. However, given the slowest mode average values obtained for each residue, it was observed that the average frequency shifts of certain residues were extremly high or low, and they were considered as the code's artifacts. When preparing the color spectrum during visualization, the maximum and minimum percentage shift values were required to determine the upper and lower limits, and the values of the residues identified as code's artifacts were not taken into account when determining these lower and upper limits.

2.5 Merging FTMap and ENM Based Residue Scanning Results

The cross-clusters obtained from FTMap were investigated to identify proximal residues closer than 5 Å to the relative solvent molecules present in the clusters. The sum of the average frequency shift value of all proximal residues was then divided by the total number of proximal residues for that cluster to determine an average frequency shift for that cluster $\%\hat{S}_i = \sum_{i=1}^{n}(\%s_i)/n$. The %s value determined for each cluster is important for predicting ligand cluster interactions. Clusters with a low %s value naturally have a low impact on the global dynamics of the receptor. For this reason, clusters with %s value below 50% were ignored. The threshold value has been reduced to 25% in case the number of solutions is insufficient.

2.6 Determination of Interface Regions of Protein with Relative Solvent Accessible Surface Area (rASA)

Evolutionarily, the ability to define the relationship of proteins to one another requires a clear examination of the structural regions of proteins. Protein interface residues are a determinant of the specificity and stability of protein-protein interaction. The contribution of such hotspot residues to free energy of binding is known to be essential (Bogan and Thorn 1998; Kortemme and Baker 2002; Reš and Lichtarge 2005). Relative solvent accessibility (RSA) is a widely used system used to identify buried and exposed residues in the protein. In this study, a method that enables the determination of the interior, surface and interface regions in the protein was used to identify interface residues based on relative solvent accessibility surface area (rSASA). Calculation of rASA is generally based on a normalization process obtained by dividing the solvent accessibility (ASA) of

a residue by the maximum ASA value of the residue (Rose et al. 1985). Maximum ASA values should be determined in order to calculate the rASA values of each residue. The maximum ASA of each X residue is determined as the highest ASA obtained from theoretically calculated Gly-X-Gly tripeptides and physically possible conformational change (Tien et al. 2013). This computational model also divides the interface residues into three different regions as core, rim and support. However, this study focused on the detection of interface regions in general. The separation of the interface and surface regions depends on the specific threshold (25%). Residues below the threshold are defined as the interface region (Levy 2010).

A code written according to the above mentioned criteria was used to identify interface, surface and interior residues for selected human, bacteria and parasite. The interface, interior or surface residues of the proteins were stored as β-factor value on PDB file and coloration was made depending on β-factor. Here, however, interior residues were ignored to highlight the interface residues.

2.7   Comparsion of ENM Based Residue Scanning and DoGSiteScorer

At this stage, after the results obtained by the methods mentioned in the previous stages were combined, the presence of the detected regions was compared with DoGSiteScorer. DoGSiteScorer as a web tool (dogsite.zbh.uni-hamburg.de) was used to identify druggable regions and pockets. In this method, the presence of pockets splitting into multiple sub-pockets on the protein surface is based on the grid-based prediction using difference of gaussian filter. The global properties of each potential druggable pocket including the shape, size and chemical properties are calculated to determine the simple scores of these pockets based on some combination, such as, hydrophobicity, enclosure and volume (Volkamer et al. 2012). As a result, pockets covering these combinations are represented by a scoring between 0 and 1. The most likely binding pockets are represented by close to 1.

2.8 Observing Result of the Allosteric Effect with AlloSigMA on the Active Sites

AlloSigMA is a web server tool which is used to estimate allosteric binding site regions based on the change in free energy of binding as a result of ligand binding and/or mutation, at a specific site or residue (Guarnera et al. 2017).At this stage, the effects of the potential allosteric region on the active areas were investigated. GADPH uses G3P as substrate and NAD as a cofactor.As a result, a remote effect of the proposed potential allosteric region on the G3P and NAD binding sites was observed. *Sa*GADPH (PDB id: 3HQ4) contained NAD in its structure but not G3P. Therefore, the substrate was added to observe whether it would have an effect on the substrate or not. It was taken from another structure (PDB id: 3LC2 (Mukherjee et al. 2010b)) containing G3P of *Sa*GADPH with structural alignment. A similar situation has been achieved for *Tc*GADPH. However, for *Tc*GADPH, a structure with substrate could not be seen in the structures in PBD databank, therefore G3P was not taken.

2.9 High-Throughput Virtual Screening via Docking

The docking for virtual screening was carried out with GOLD software tool (Jones et al. 1997)has four different scoring functions: ChemPLP (Korb, Stützle, and Exner 2009), ChemScore (Verdonk et al. 2003) (Eldridge et al. 1997). ASP (Astex Statistical Potential) (Mooij and Verdonk 2005), and GoldScore (Jones et al. 1997). Goldscore is a method of calculating hydrogen bonding, van der Waals (vdW) energies as well as metal interactions and intra-ligand bond energy. ChemScore has an empirical scoring function obtained by parameterizing binding affinity from a given number of complexes. ASP is a knowledge-based method based on atom-atom distances using the structures of known protein ligand complexes, unlike the force field method used in the other three types of scoring methods. ChemPLP is a scoring method that takes into account hydrogen and vdW interactions similar to GoldScore. In this context, some of the previous studies have shown that ChemPLP is a fast and accurate scoring function for both pose prediction and virtual screening (Liebeschuetz, Cole, and Korb 2012). Therefore, ChemPLP was selected as the scoring function.
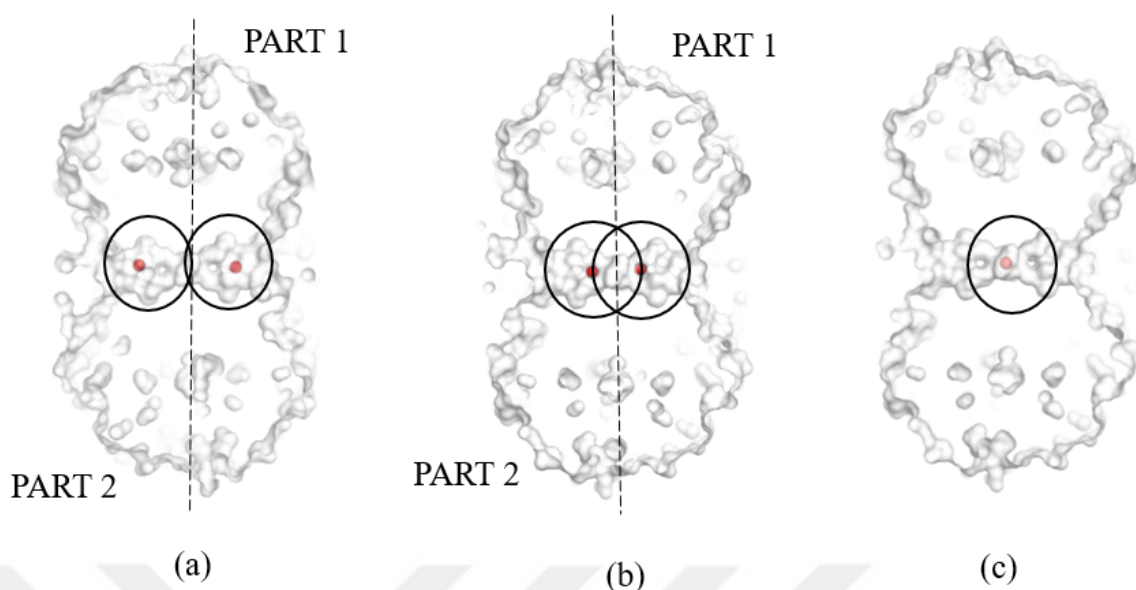
Figure 2.2 Locations of the gridboxes for docking process. a) *End* b) *Middle* c) *Center* region

The tunnel area has a narrowand long structure. This situation may cause the possibility of binding of drug molecules to decrease towards the center. For this reason, 3 different coordinates have been determined for docking in the tunnel region. In the first one, the center of the tunnel region (*Center*) was targeted, and in the second, a region that could be defined as the entrance of the tunnel region (*End*) was determined.The last third region was defined as the region (*Middle)* which corresponds between the center and the region closer to the entrance of the tunnel region (See Figure 2.2).In addition, the results obtained for the tunnel region were observed symmetrically. Therefore, since there are similar allosteric regions on both sides of the tunnel, the docking process was applied to both regions (defined as part1 and part2) that divide the tunnel region into two (See Figure 2.2).

The screened library of chemical compounds was taken from ZINC15 (https://zinc.docking.org), a database that provides free access to commercially available compounds for virtual screening (Sterling and Irwin 2015). Approved Worldnot-FDA" and "FDA" compounds under the subset catalog "Bioactive and Drugs" on the website were taken consideration for docking process.As a result, due to the symmetrical findings obtained for both exits in the tunnel region, a detailed docking operation involving the equivalent of these three regions (*Center*, *Middle*, *End*) in symmetrical regions (part 1

and part 2) was performed. The flowchart below gives an overview of the docking procedure and the compounds used (See Figure 2.3).
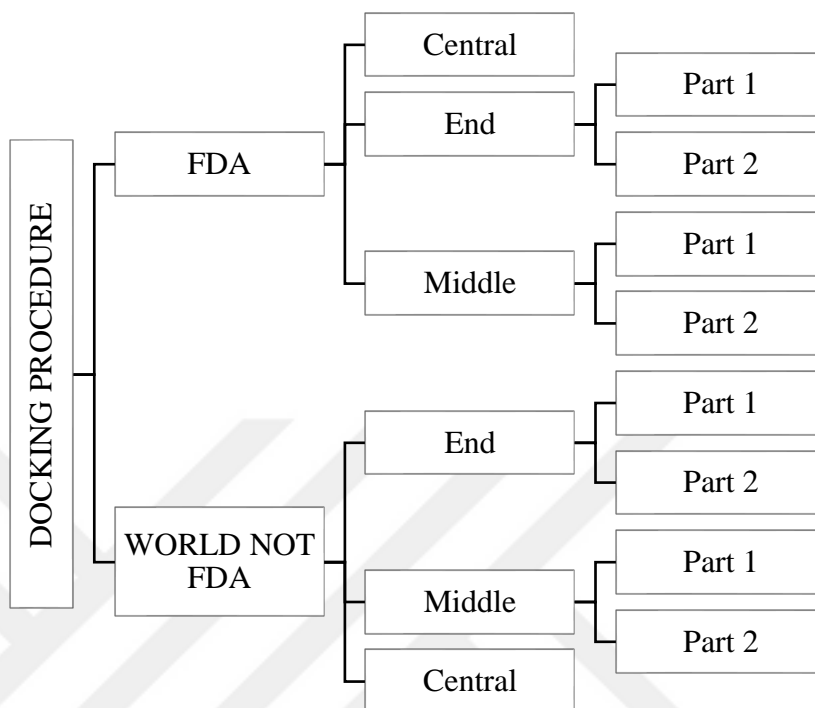

Figure 2.3 The flowchart of docking process for proposed regions.

3 RESULTS AND DISCUSSION

3.1 Criteria for Selecting X-ray Structures for Species

The structure to be determined was randomly selected to represent each species from the X-Ray structures determined for *H.sapiens, S.aureus* and *T.cruzi*. However, the most important criterion in random selection stage was the structural alignment of the crystal structures. The maximum and minimum RMSD values obtained from the structural alignment were found to be 1.97 and 0.44 for human GAPDH, 2.61 and 0.41 for bacterial GADPH and 1.34 and 0.77 for *T.cruzi* respectively (See Tables A.1, A.2 and A.3 in Appendix). Furthermore, heatmap which was based RMSD values obtained from X-ray structures were added to Appendix as a supplementary information. Consequently, due to low RMSD values, one ofthe structures was chosen arbitrarily for each species as a representative and used for further analysis.

3.2 Computational Solvent Mapping

In uncovering potential binding regions for the specified species, regions known as consensus site, also known as hot spots, were determined through tetramer crystal structures. In addition, tetramer structures were decomposed into its subunits and given as chain by chain to reveal the presence of alternative potential binding sites on the structure (See Figure 2.1).

It was observed that some CSs obtained from the structure mapped as chain by chain coincided with CS obtained from overall mapping. In addition, some monomeric clusters have been observed to be inaccessible when representing monomeric clusters in the tetramer structure. As a result, overlapping and inaccesible clusters were removed (See Table 3.1).

The total number of CSs found for *H.sapiens* GADPH (*h*GADPH) was found to be 39. The total consensus sites remaining after removing the inaccessible and overlapping ones was reduced to 37. In other words, 2 CS were eliminated as a result of overlapping and inaccessible clusters. A total of 47 consensus sites was found for *T.cruzi* GADPH (*Tc*GADPH), but 10 consensus sites were eliminated because they were inacccesible or overlapping with tetramer cluster results. Finally, the total amount of CS determined for *S.aureus* GADPH (*Sa*GADPH) was 44. After the overlapping and inaccesible ones were removed, the remaining amount of CS was determined to be 31. After ENM filtering (frequency shift > 25% / 50%), remaining clusters for *h*GADPH was detected 21 and 16 respectively. While remaining clusters for *Tc*GADPH was 31 and 19, for *Sa*GADPH was 31 and 29 respectively (see Table 3.1). The lowest amount of CS sites for tetramer mapping was observed for human GAPDH, while the maximum amount was 15 for *T.cruzi*. It has been observed that the amount of CS site in the structures mapped as chain-by-chain was very close and comparable especially for tetramer human GADPH.

Table 3.1 Number of cluster after filtering for three species (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
GADPH	*H.Sapiens*	8	7	7	9	8	39
		8	6	7	9	7	37
		8 / 8	3 / 3	3 / 2	4 / 2	3 / 1	21 / 16
	S.Aureus	14	6	9	8	7	44
		12	3	6	6	4	31
		12 / 12	3 / 3	6 / 6	6 / 5	4 / 3	31 / 29
	T.Cruzi	15	7	9	7	9	47
		15	5	5	5	7	37
		15 / 2	5 / 5	5 / 2	5 / 2	7 / 7	37 / 18

3.3 ENM-Based Residue Scanning

In the previous stage, inaccessible and overlapping CS sites have been eliminated. The remaining CS sites were evaluated based on their average frequency shift results (%s_i) as mentioned in the Methods section. Before proceeding to this stage, the results predicting the regions that will potentially affect the global dynamics in the structure were

highlighted with a color spectrum according to the average frequency shift values calculated for each species which varies from negative (blue), zero (white) to positive (red).

### 3.3.1 Some residues defined as outlier because of extremely high eigenvalues

Considering the calculated average frequency shift values, it was observed that the average frequency shift values of some residues in x-ray structures were quite high or low, unlike the general trend. This situation announced the need to examine the locations of the residues of these frequency shift values observed at extreme points. When the locations of the residues on the structure are examined, the residues mentioned were observed in regions which are considered to be more filexible than other regions in the relatively structure defined as a loop on the x-ray structure. In addition, when the general trend of the region in which it is located (density of colors) shows that the residues act on their own rather than affecting a global dynamic. For this reason, residues which have extreme frequency shift values were ignored for the minimum and maximum determined for each structure.

The average frequency values calculated in *Tc*GADPH, the total number of residues in the *Tc*GADPH structure was 1436 and only 29 of 1436 resiudes were defined as artifact of code which is observed on the graph roughly above $+5$ and below -5. (See Figure 3.1a) Percentage of the outlier residues was found to be nearly 2%. A similar situation was observed in *Sa*GADPH residues calculated on average frequency shift values. As shown in Figure 3.1b, for *Sa*GADPH, 8 out of 1336 residues in total ($\sim 0.6\%$) were determined as outlier.

In *h*GADPH, only 2 out of 1332 residues ($\sim 0.15\%$) were determined as outlier in the average frequency changes calculated for residues (See Figure 3.1c). In summary, considering the percentages of the residues determined as outlier, neglecting these residues did not create any bias for the analysis, and provided more specific expression and representation of the regions that are considered to have an impact on the global dynamics. In other words, it provided a depiction of regions or residues that are thought to have a significant impact on global dynamics on the crystal structures.
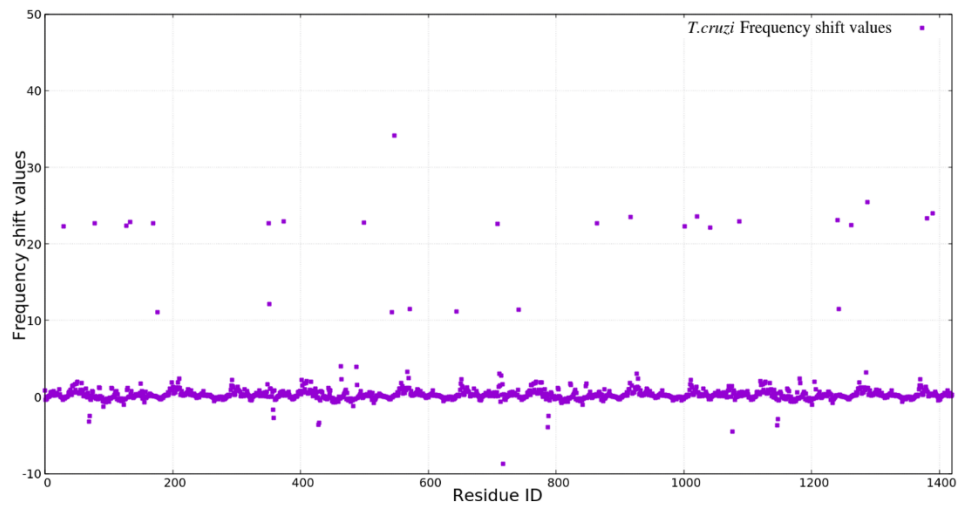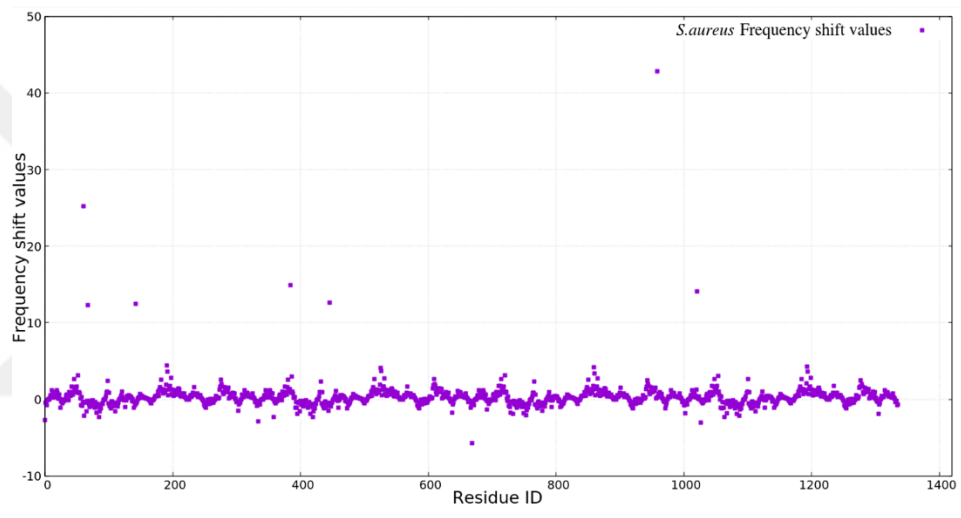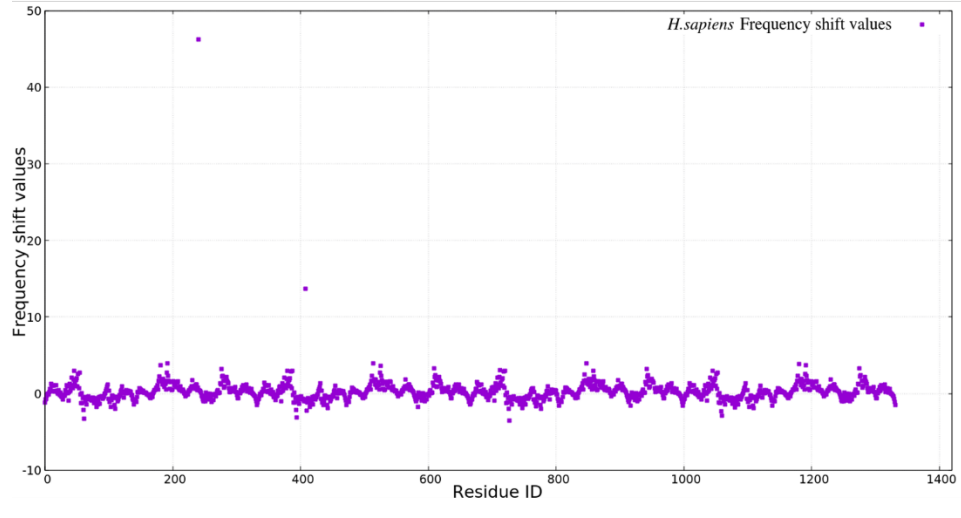
Figure 3.1 Frequency shift value residues of all three species. a) *Tc*GADPH b) *Sa*GADPH c) *h*GADPH

After elimination, the minimum and maximum average frequency shift values obtained for hGADPH, SaGADPH and TcGAPDH were defined as [-3.55:3.92], [-3.0383:4.45] and [-4.53:4.03], respectively.

Certain regions, which are thought to affect the global dynamics of the receptor in all three species, were investigated. The color contrast between the presence of red-colored potential regions that are thought to affect the collective mode frequenciesand blue regions that reduce global dynamics, in other words, have negative effects on receptor dynamics was clearly visible in hGADPH and SaGADPH (See Figure 3.2a and 3.2b).

When the TcGADPH structure was examined with color-based coding, the presence of white areas with a neutral effect, which we do not consider to have any effect, appears to be dominant. (See Figure 3.2c). Comparing hGADPH, SaGADPH and TcGAPDH, in active sites where catalytic reactions occurred, the presence of red regions that positively contributed to the dynamics of the protein was observed for all three species. In addition, the presence of red zones in the tunnel region, which contributed significantly to the dynamics, was determined. As a result, according to the color-based screening method, the tunnel region was found to contribute to protein dynamics as much as active sites.

In addition, the structures were compared among each other. Different range was used for an inclusive comparison; the lowest minimum frequency shift value between all species was defined as the optimum minimum (-4.53) point for all species and the highest maximum value was likewise determined as the optimum maximum point (4.45) for all species. As a result, the higher color density of active sites in hGADPH and SaGADPH showed that the dynamic of the active site was stronger in these structures than TcGADPH. Similar situation was clearly seen in tunnel areas. In this case, the positive contribution of the SaGADPH tunnel region to the dynamics was stronger than TcGADPH, which makes SaGADPH more preferable in species-specific drug design (see Figure 3.3).
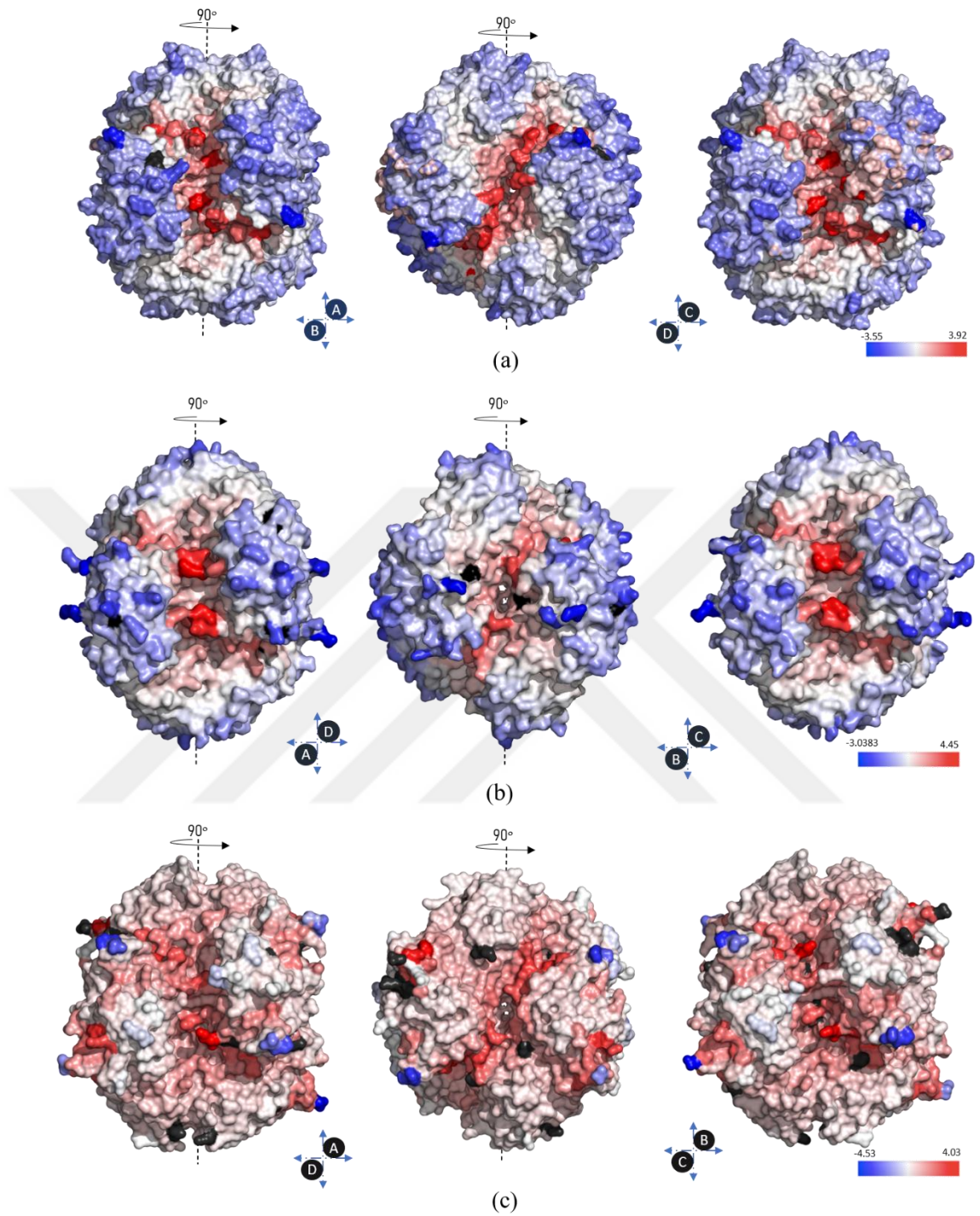
Figure 3.2 Frequency shift values calculated for GADPH. a) *h*GADPH b) *Sa*GADPH c) *Tc*GADPH.
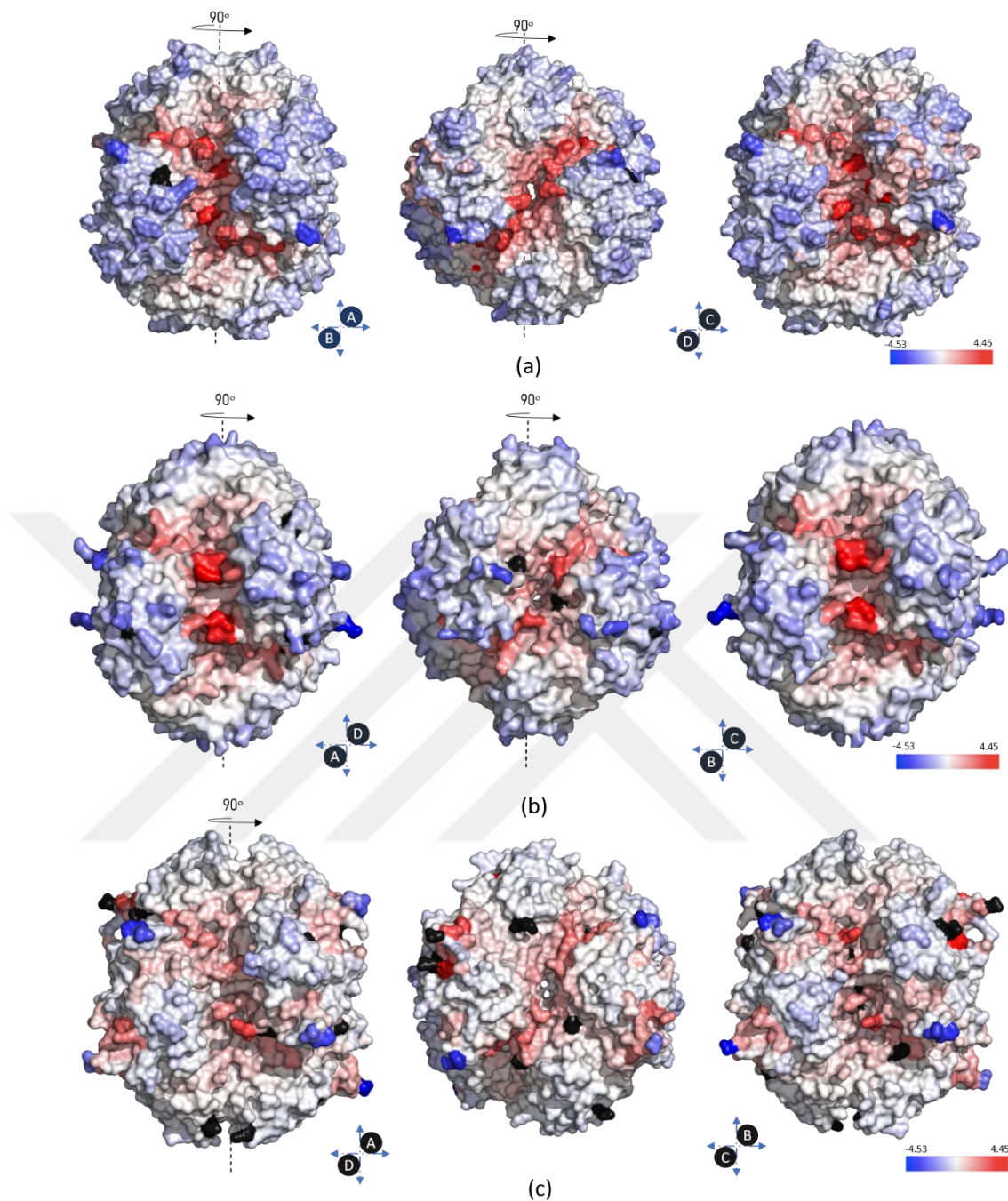
Figure 3.3 Comparison of the the species with common range. a) *h*GADPH b) *Sa*GADPH c) *Tc*GADPH

3.3.2 Combining ENM results with solvent mapping results and eliminating some CSs

As mentioned in the method, clusters obtained from FTMap and residues located at 5 Å near the bound solvent molecules in these clusters were determined and their frequency shift values were divided by the total number of residues, thus an average frequency shift

value for each CS was determined. The inaccessible and overlapping CSs were eliminated. After elimination, frequency shift values per residue were calculated for remaining CSs in each species. As a result, a total of 5 druggable hotspot regions were observed for all three species.When 50% threshold value was applied, it was seen that one of the hotspot regions that are completely within the interface region determined in TcGADPH was explicitly eliminated. However, as a result of the results obtained, it was thought that the existence of such a significant hotspot region consisting of more than one CS should not be ignored. For this reason, the existing threshold has been reduced from 50% to 25%. In this case, the presence of hotspot regions determined in *Sa*GADPH, hGADPH and *Tc*GADPH is the same for all. In addition, lowering the threshold strengthened the consensus regions determined for *h*GADPH (See Table 3.2). As a result, CSs with average frequency shift values below 25% were eliminated. The mean frequency shifts values, standard deviations of the remaining CSs and general information about the total number of residues around the CSs were given in Tables A4, A.5 and A.6 in Appendix. The remaining CSs in the five drugabble hotspot regions were also shown on structures which is represented by the color spectrum (See Figure A.3 in Appendix).

Table 3.2 Druggable hotspot regions and CSs obtained from one chain and overall mapping.

Druggable Site ID	*S.aureus* (PDB id: 3HQ4)	*T.cruzi* (PDB id: 3DMT)	*H. sapiens* (PDB id: 4WNI)
1	1A-2A-5A-7A-8A-9A-2	2A-3A-4A-5A-7A-12	2A-3A*-4A-5A*-6A**-7A*- 7
2	2B-3B-4B-6B**-7B-8B-5	1B-3B**-6B-7B-8B**-9**-15**	1B*-2B-3B*-4B-5B*-6B**-7B*
3	3C-4C-6C-7C**-7-8-12	2C-4C-5C**-6C**-7C**-13**-14**	1C*-2C-3C**-4C*-5C-6C*-7C*-8C**-9C*
4	3D-4D-5D-3-14	1D-3D-4D-6D-7D-8D**-9D-3	2D-3D*-4D*-5D**-6D*-7D*-8D**-8
5	1-4-6-10-11	1**-2**-4-5**-6**-7**-8**-10**-11**	1-2-3-4-5-6

 * Represent CS which is below 25%
** Represents between 25% and 50% threshold

3.4 Some Druggable Hotspot Regions Found on the Interface Regions of the Structure

The interface regions determined for each species were colored in blue. CSs from one chain and overall mapping were represented in magenta and green, respectively and remaining CSs corresponding to those with frequency shift between 25% and 50% (see Figure 3.4). The proximity of most CS in the designated hotspot regions to the interface regions is a remarkable point. Most of the CSs in the hotspot regions identified in the catalytic pockets were observed on the interface regions in all three species. In addition, it was observed that all the alternative region CSs located in a tunnel region, which is clearly accessible from both sides towards the center of the structure, were completely surrounded by interface residues (see Figure 3.4).
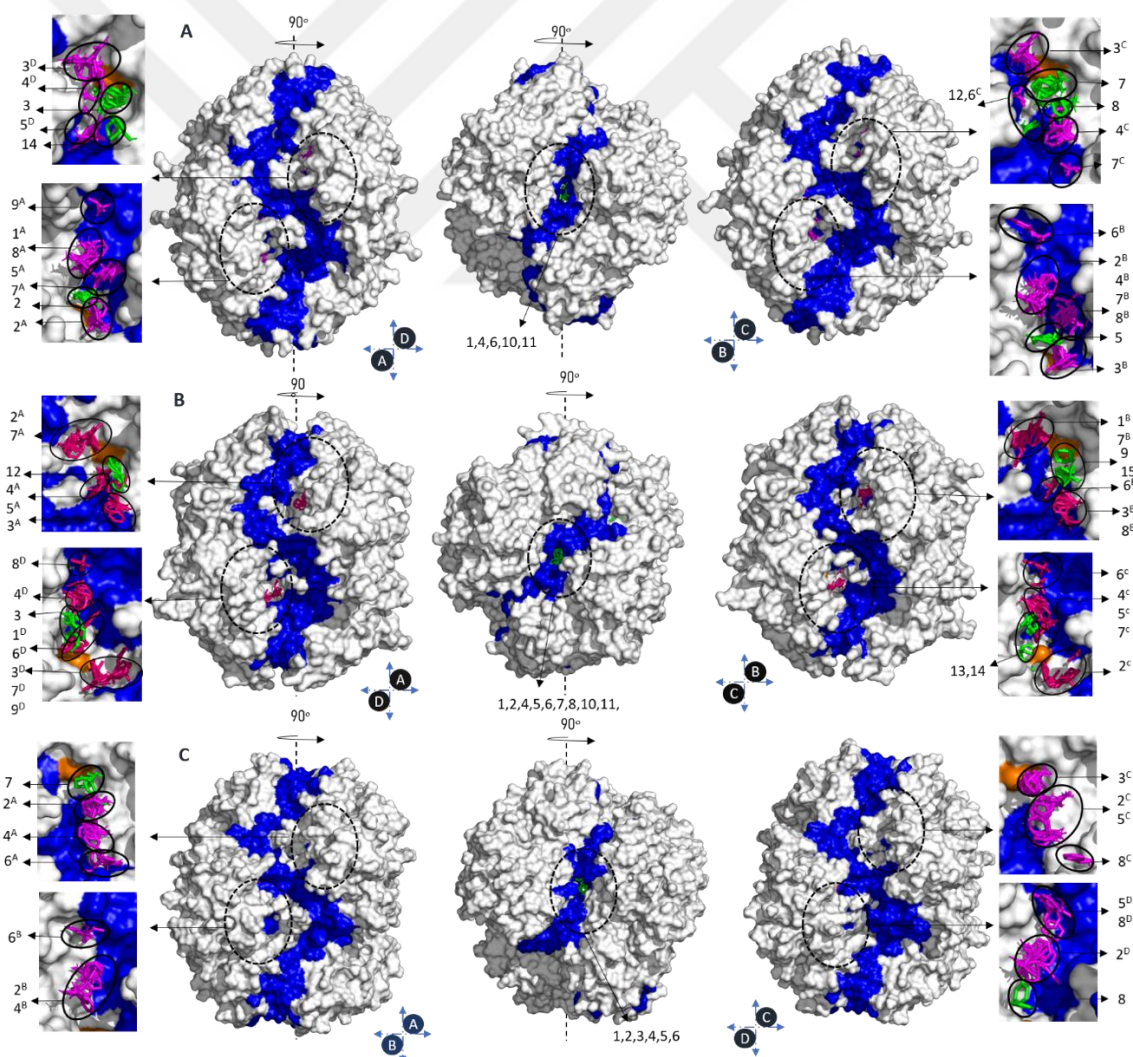

Figure 3.4 Representation of interface residues in *Sa*GADPH, *Tc*GADPH and *h*GADPH respectively.

3.5 Tunnel Region on the GAPDH Can Be a High Candidate for Possible Effector site

For *Sa*GADPH, *Tc*GADPH and *h*GADPH structures, the presence of hotspots, potentially druggable regions, was detected in the previous stage. However, the presence of these identified hotspot regions has been identified in pockets containing most catalytic regions and therefore cannot be proposed as an allosteric site. But it can be presented as a good proof of the accuracy of the method applied. Total druggable hotspot region in all GADPH was determined as five. However, one of the hotspot regions has been observed in a corridor that can be defined as tunnel extending towards the homotetrameric center. In *Sa*GADPH and *h*GADPH, this tunnel region was able to exceed 50% criterion, while in *Tc*GADPH the CSs ranged from 25% to 50%.

The significance of the tunnel region in GADPH has been reported in several experimental studies (Carlile et al. 2000; Dubey et al. 2017). The tunnel region coincides with some critical residues in the S-loop region, which plays an important role in protein regulation. Contribution of phosphorylation of Ser50, Ser203 and Tyr41 residues to the NAD binding pockets and the dynamic movement of the S-loop in regulating the oligomer assembly, is well reported experimentally. It was also predicted that S-loop interferes with oligomerization assembly and allosteric activation in enzyme activity in GADPH (Dubey et al. 2017). During the dephosphorylation of Ser50, S287 helps to stabilize the S-loop by contributing to homodimerization by forming hydrogen bonds along the dimer interface. In addition, the dephosphorylation of Tyr 41 stabilizes the adjacent S-loop with the hydrogen bonds formed. S203 dephosphorylation establishes atomic interactions with other neighboring S-loops, reconfigure the S-loop to fit back into the neighboring NAD binding pocket.

The presence of S50 and S287 from these three residues in the tunnel area was determined. In Figure 3.5 for *Sa*GADPH, the resiudes representing S50 and S287 were colored with red and cyan respectively. In addition, the indicated S-loop region was shown in yellow. The same result for *Tc*GADPH is attached to Appendix (See Figure A.2 in Appendix)
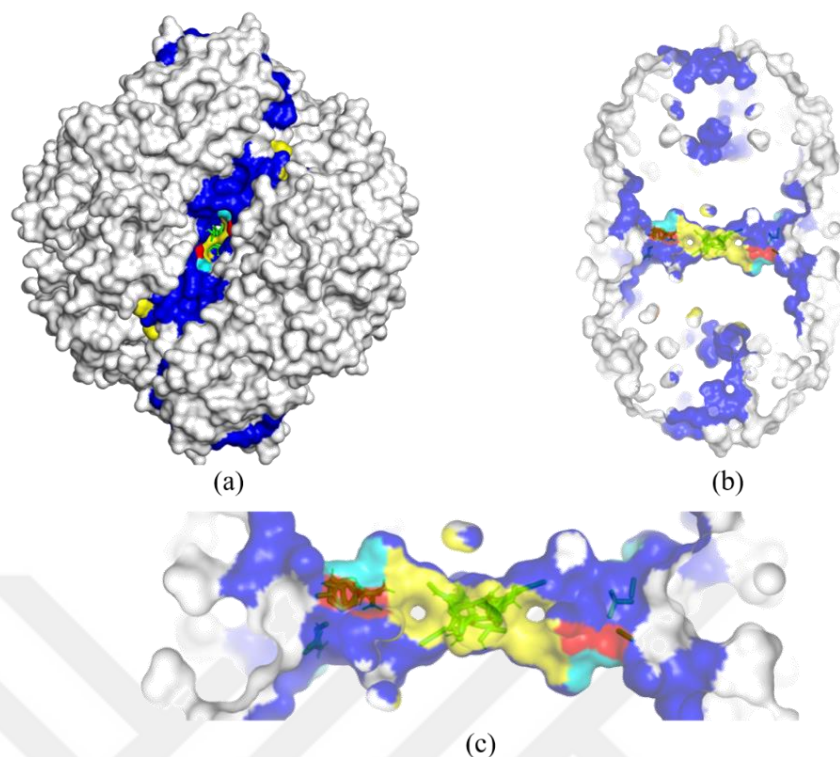
Figure 3.5 The corridor-like tunnel region in SaGADPH and S50 and S287 residues (Ayyildiz *et al*. 2020).

3.6 Detection of Species-Specific Differences in The Tunnel Region by Sequence Alignment

In order to talk about species-specific drug design, it is an important point to observe the differences between the species and to reveal the presence of a satisfactory amount of interspecific mutations. For this reason, the similarities of the existing tunnel regions were compared with an amino acid-based sequence alignment between human and bacteria/parasite. The residues containing the representation of the enzyme from two different points of tunnel view and the tunnel region residues were indicated by the red frame. The identical residues were shown in blue and similars were orange (see Figure 3.6 & 3.7). The total number of residues in the tunnel region in all three species was determined to be 18. Similarity and identity percentage of the tunnel region for *h*GADPH-*Sa*GADPH and *h*GADPH-*Tc*GADPH were determined as ~55% and `~ 44%, respectively (see Figure 3.6 & 3.7). As a result, the low similarity and identity percentage of the tunnel region with *h*GADPH make the tunnel regions of *Sa*GADPH and *Tc*GADPH an effective target in species-specific drug design.
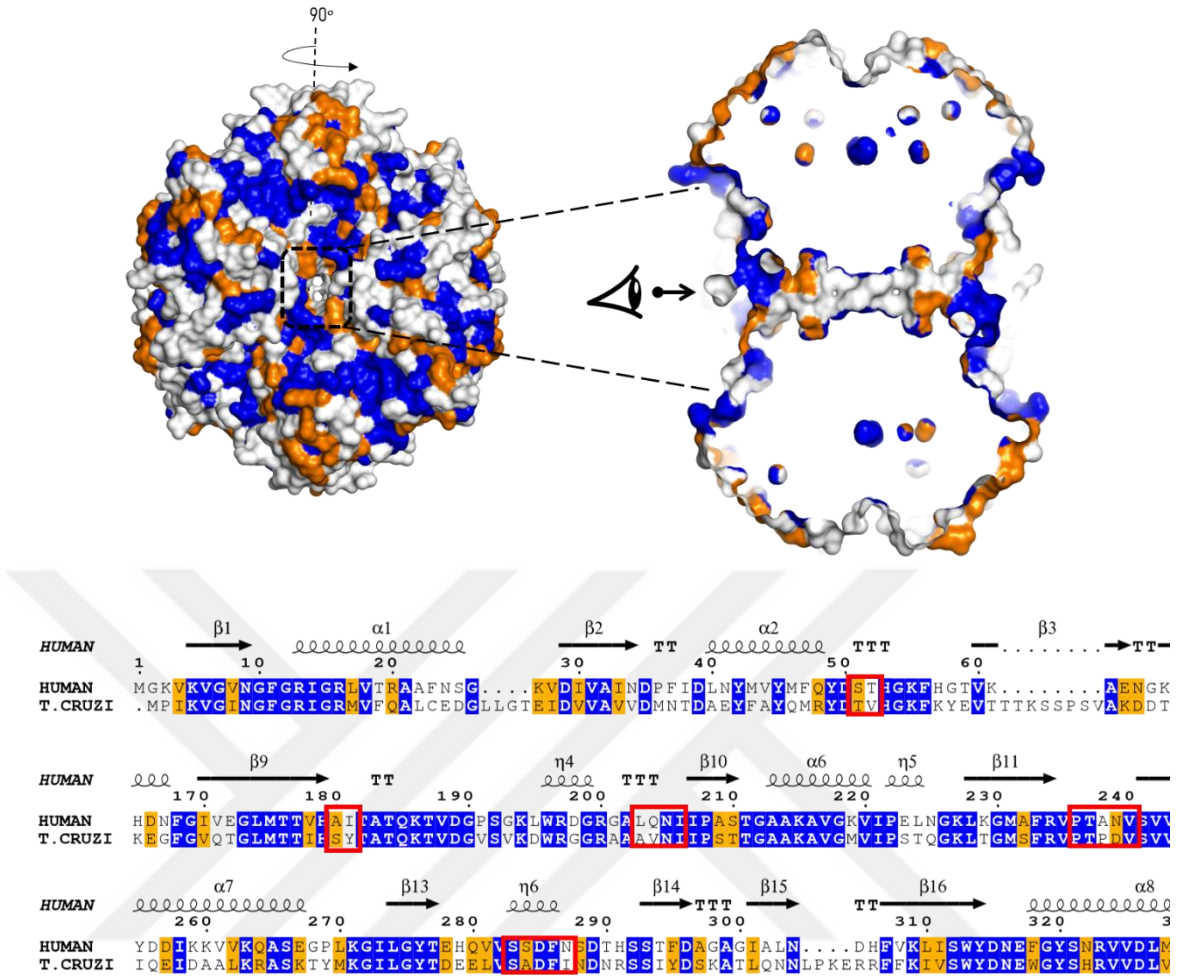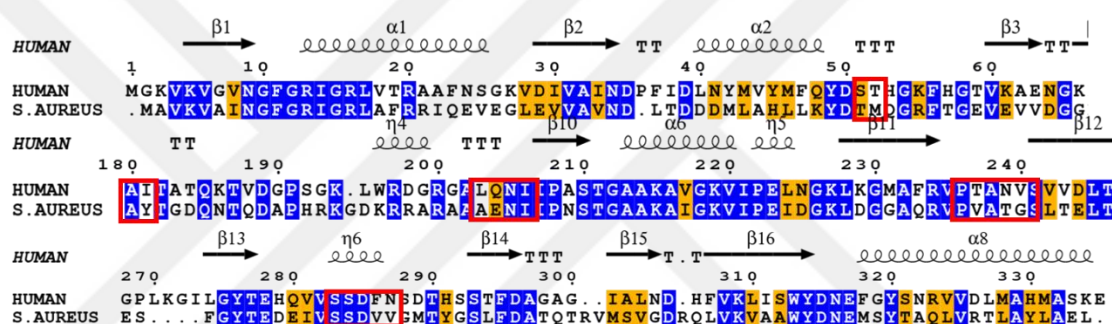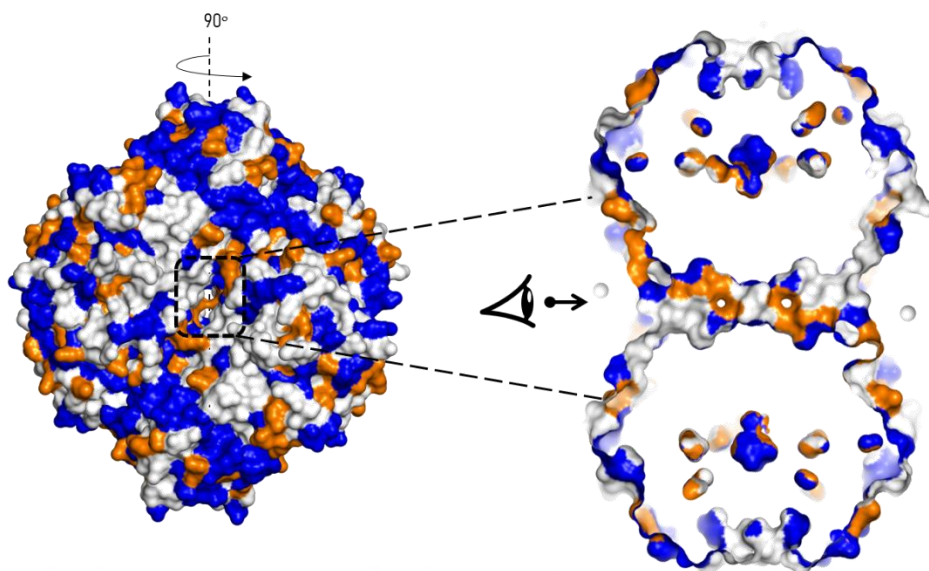
Figure 3.6 Sequence similarities between *h*GADPH and *Sa*GADPH (Ayyildiz *et al*. 2020).

3.7 DoGSiteScorer Supports Tunnel Region on The Species

Finally, it is important to assess the druggability of the proposed regions with another tool which was DoGSiteScorer. The pockets obtained with DoGSiteScorer with high score values coincided well with all of the hot spot regions proposed by our methodology. The drug scores of the pockets determined by DoGSiteScorer corresponding to the catalytic regions and tunnel region were given in Table 3.3. Considering the drug scores of DoGSiteScorer in SaGADPH, the drug scores in the catalytic region were equal to those of the proposed tunnel region. This emphasizes that the tunnel area is an important drugable region as well as the catalytic site When the scores obtained for *Tc*GADPH were analyzed, it was clear that the highest value is in the catalytic pocket in the C chain. In addition, the numerical closeness of the proposed tunnel region's drug score to the pocket with the highest score determined can be clearly seen (see Figure 3.8).
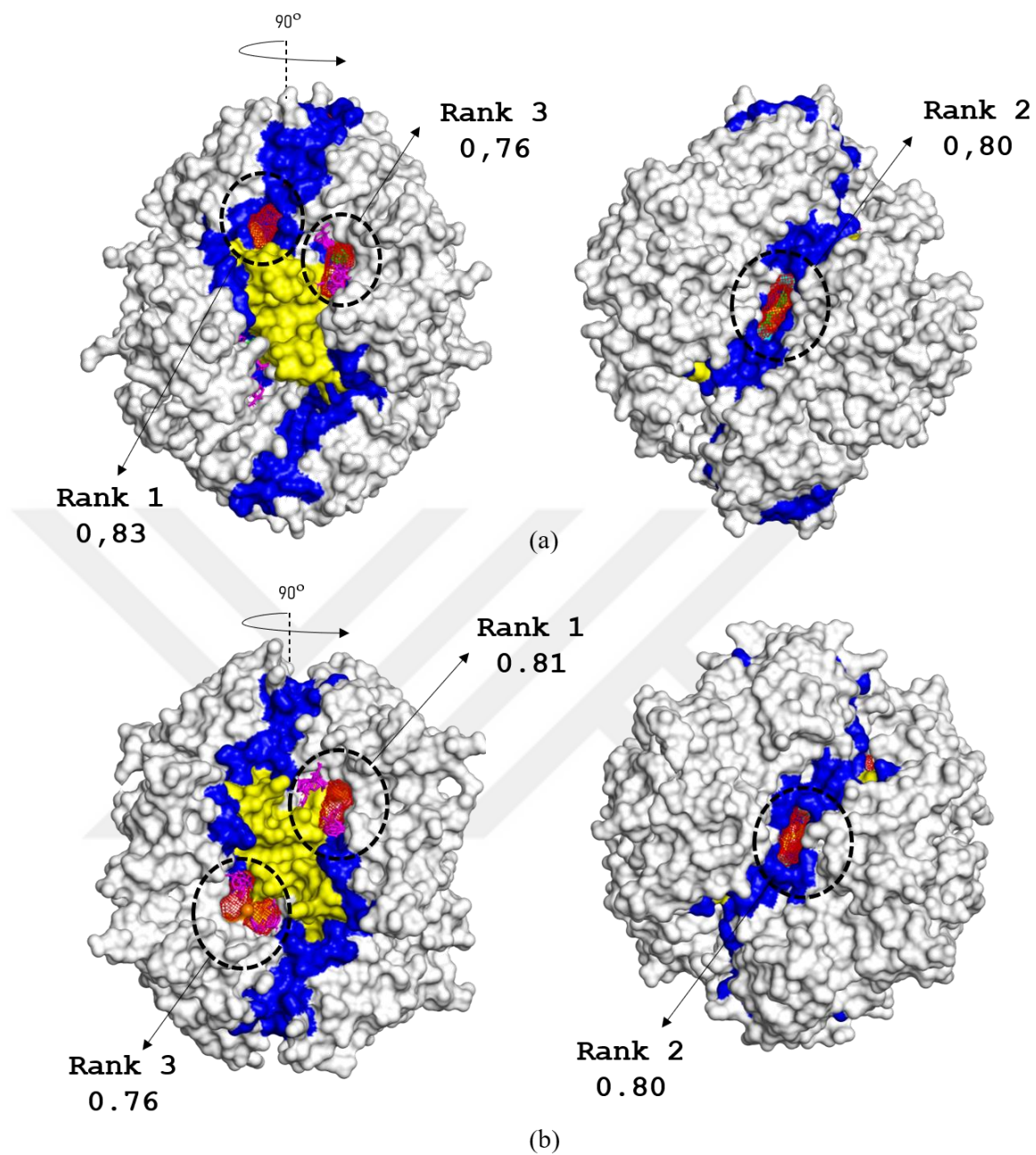
Figure 3.7 DoGSite pockets, scores and ranks and overlapping CSs. a) *Sa*GADPH b) *Tc*GADPH
(Ayyildiz *et al*. 2020).

Table 3.3 DogSite scores and ranks for pockets. Bold-highlighted letters represent the highest scores.

Enzyme	*S.aureus*	Region	Score*/ Rank	*T.cruzi*	Score/ Rank	Region
GADPH	**1A-2A-5A-7A-8A-9A-2**	Catalytic site	0.80 / 2	1D-3D-4D-6D-7D-9D-3	0.48 / 6	Catalytic site
	2B-3B-4B-7B-8B-5	Catalytic site	0.76 / 3	2A-3A-4A-5A-7A-12	0.44 / 8	Catalytic site
	3C-4C-6C-7-8-12	Catalytic site	0.73 / 4	1B-6B-7B	0.72 / 3	Catalytic site
	3D-4D-5D-3-14	Catalytic site	0.64 / 8	**2C-4C**	0.81 / 1	Catalytic site
	1-4-6-10-11	Allosteric site	0.80 / 2	1-2-4-5-6-7-8-10-11	0.80/2	Allosteric site

*Minimum-maximum range for score values:
For *S.aureus* GADPH: [0.14-0.83]
For *T.cruzi* GADPH:[0.15-0.81]

3.8   Allosteric Effect of Tunnel Region on Catalytic Regions determined via AlloSigMA

AlloSigMA was used to investigate the allosteric effect of the tunnel region on the catalytic regions as a result of a ligand binding. The gibbs free energy calculated for the residues with the effect of the tunnel region on each residue was represented on the crystal structure by giving in the beta columns of the structures. Here, negative values are defined as the places where protein stabilization (increase of the dynamic) increases, while blue positive values are regions representing destablization (increase of the dynamic) in protein. Accordingly, minimum ΔG value of -3.76 and maximum value of 1.76 were observed for SaGADPH. As a result, the presence of stabilization originating from the tunnel region can be seen in the active site pockets in *Sa*GADPH clearly (See Figure 3.9). The maximum and minimum ΔG values found for *Tc*GADPH were -1.35 and 0.65, respectively. The average ΔG values of the reisudues around NAD and G3P in the catalytic region around 5 Å were examined. The average values of the residues located around the G3P in the catalytic region in *Sa*GADPH were determined as -0.66 and -0.40, respectively. The same values were determined as -0.29 and 0.052 for NAD binding pocket. It has been seen that the tunnel region in *Tc*GADPH caused an increase in stabilization in the catalytic sites (See Figure 3.10). The ΔG of the residues around the NAD for TcGADPH was -0.24 and -0.11, respectively.
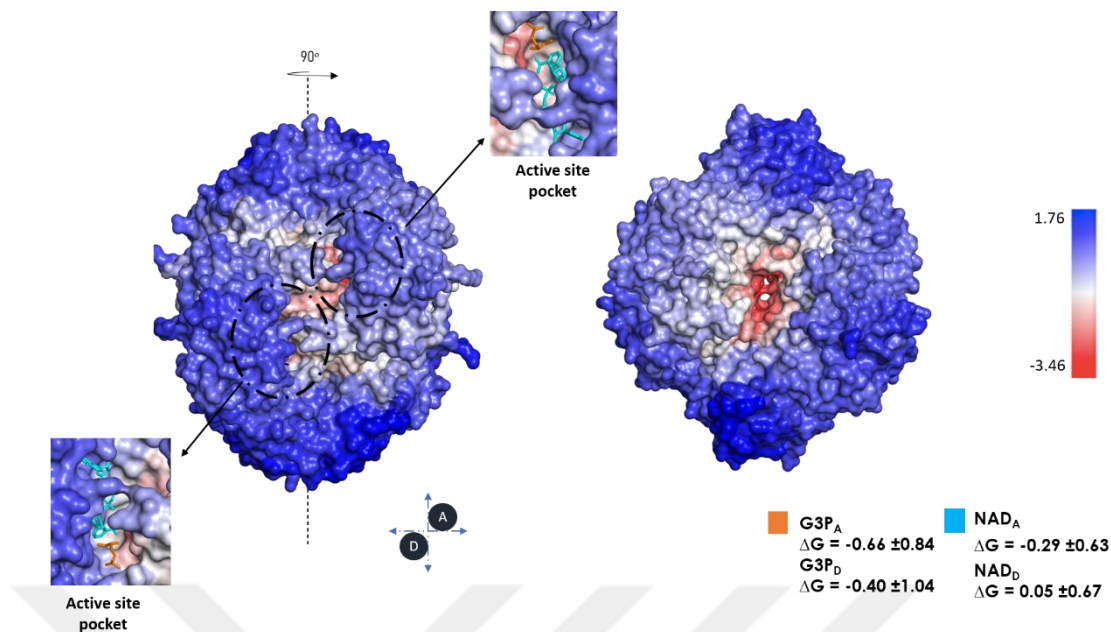
Figure 3.8 Allosteric free energy changes calculated for each residue in *Sa*GADPH.
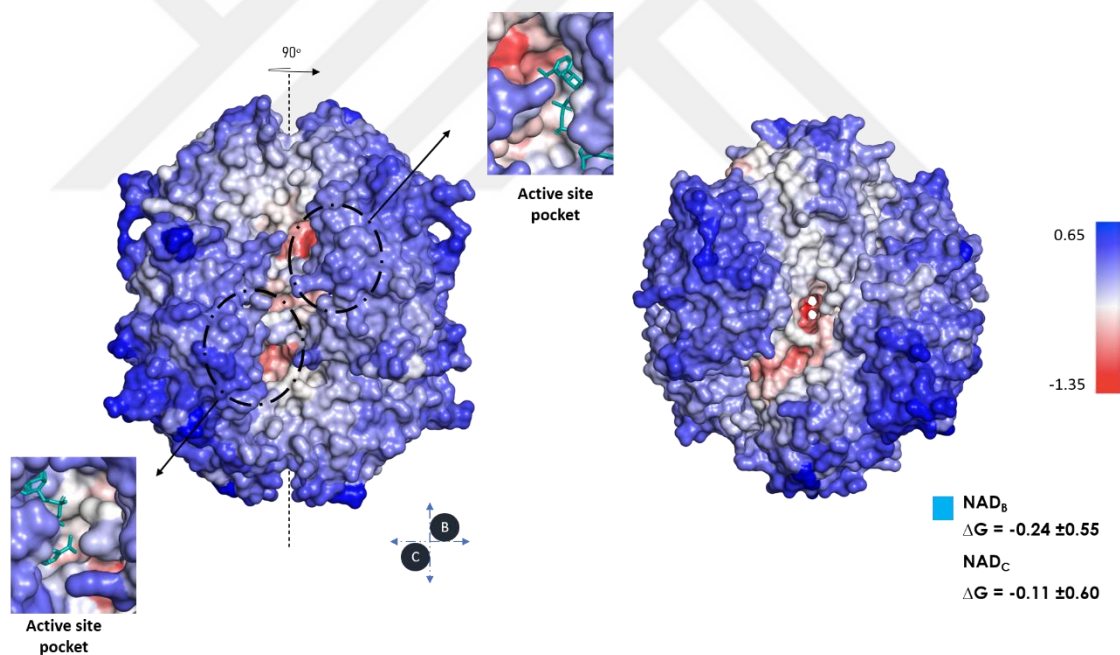

Figure 3.9 Allosteric free energy changes calculated for each residue in *Tc*GADPH.

The point of interest here was whether a ligand which bind to the tunnel region would allosterically affect active site residues or not. As a result, residues in the active region were focused. As an additional information here, it was stated in the method that G3P was not in the *Sa*GADPH structure and it was taken from another structure (PDB

id:3LC2) belonging to *Sa*GADPH with structural alignment. As a result of structural alignment, RMSD value was determined as ~0.3. Since the value was quite small, it was not objectionable to use it. As a result, residues of G3P and NAD molecules near 5 Å were determined and the average allosteric free energy values of these residues were determined (See Table 3.4 & 3.5).

Table 3.4 Gibbs free energy values of residues close to 5 Å to G3P and NAD molecules for *Sa*GADPH.

Name	Sum of ΔG values	Number of residues	Average
G3P(Glyceraldehyde 3-phosphate)			
Chain A	-13.27	20	-0.66 ± 0.84
Chain B	-8.49	19	-0.44 ± 0.95
Chain C	-4.59	19	-0.24 ± 1.10
Chain D	-8.82	22	-0.40 ± 1.04
NAD (Nicotinamide adenine dinucleotide)			
Chain A	-9.43	32	-0.29 ± 0.63
Chain B	-6.22	31	-0.20 ± 0.55
Chain C	2.91	30	0.09 ± 0.60
Chain D	1.58	30	0.05 ± 0.67

Table 3.5 Gibbs free energy values of residues close to 5 Å to NAD molecules for *Tc*GADPH.

Name	Sum of ΔG values	Number of residues	Avarage
NAD (Nicotinamide adenine dinucleotide)			
Chain A	-2.14	32	-0.06 ± 0.63
Chain B	-7.72	31	-0.24± 0.55
Chain C	-3.80	34	-0.11 ± 0.60
Chain D	-4.53	32	-0.14 ± 0.67

As a conclusion, increased stabilization will cause restrictions on the binding of the substrate to the active site. Similar situation has been observed partially for NAD. Considering the ΔG values of the residues determined for NAD, stabilization increased only in two active regions (See Table 3.4). However, the NAD moleculedisplays some redness increasing towards the center in all active regions. Even the NAD being affected partially can affect the stability of the substrat in the active region. Consequently, considering the triggering role of NAD as a cofactor for the substrate, it showed strong results for the inhibition of the active regions where negative ΔG values determined for both G3P and NAD are observed (See Table 3.4).

In *Tc*GADPH, the ΔG values of all four regions that connect NAD were negative. This clearly shows the effect of the tunnel region on the cofactor NAD. As with *Sa*GADPH, in *Tc*GADPH, the dominance of negative residues increases at the end of the NAD molecule approaching the active region. This may hinder their interaction with the substrate (See Table 3.5).

3.9   Detection of Common Molecules for *End* and *Middle* Tunnel Region via Docking.

At this stage, the application of docking process targeting the potential tunnel regions determined for *Tc*GADPH and *Sa*GADPH types was performed. At this point, the docking process was applied only for *Sa*GADPH, and the docking study for *Tc*GADPH was left as the focus of another study. As shown in Figure 2.2, 1416 FDA approved and 2922 World-not-FDA approved molecules were docked in the region. As a result of virtual screening, the conformation of the molecule with the highest ChemPLP score was taken from 10 different docking process applied for each molecule

Docking experiments for GADPH were applied to three different locations of the tunnel area proposed as shown in Figure 2.2 in Methods section. Two of these locations were taken into consideration as "*End Tunnel*" and "*Middle Tunnel*". When docking was carried out in the third region which was the exact center of the tunnel, the restricted area caused a high steric overlap between the receptor and the drug molecule, so it was not possible to fit the the ligand molecule inside the center. As a result, negative score values were mostly found.

Since our main goal here was to identify molecules which are sensitiveto bacteria, the docking scores obtained for the bacteria are ranked from highest to lowest. The results of this ranking for FDA molecules were given in Figure 3.16a and 3.16b. ChemPLP scores were given in the first *y* axis, while differences were given in the second *y*-axis indicated on the right side. A positive difference means that the score of the docked pose in bacteria is higher than that in human, which can be interpreted as a higher binding affinity.

In addition, an elimination process was carried out for Part 1 and Part 2. Although the enzyme has a homotetrameric structure, in order to increase the number of solutions, the

protein was divided into two symmetric regions as Part1 and Part 2, and docking was applied to the *Middle* and *End* tunnel regions for both parts. It has been shown that Part1 and Part 2 docking results have high similarities due to their symmetrical coordinates (see Figure 3.11).

As a result, *End* tunnel docking results for Part 1 and Part 2 were compared. End of the comparison, it was observed that the Part1 and Part2 scores were quite similar to each other, so the next step was continued with the arbitrarily chosen Part 2 (See Figure 3.12a & 3.12b).
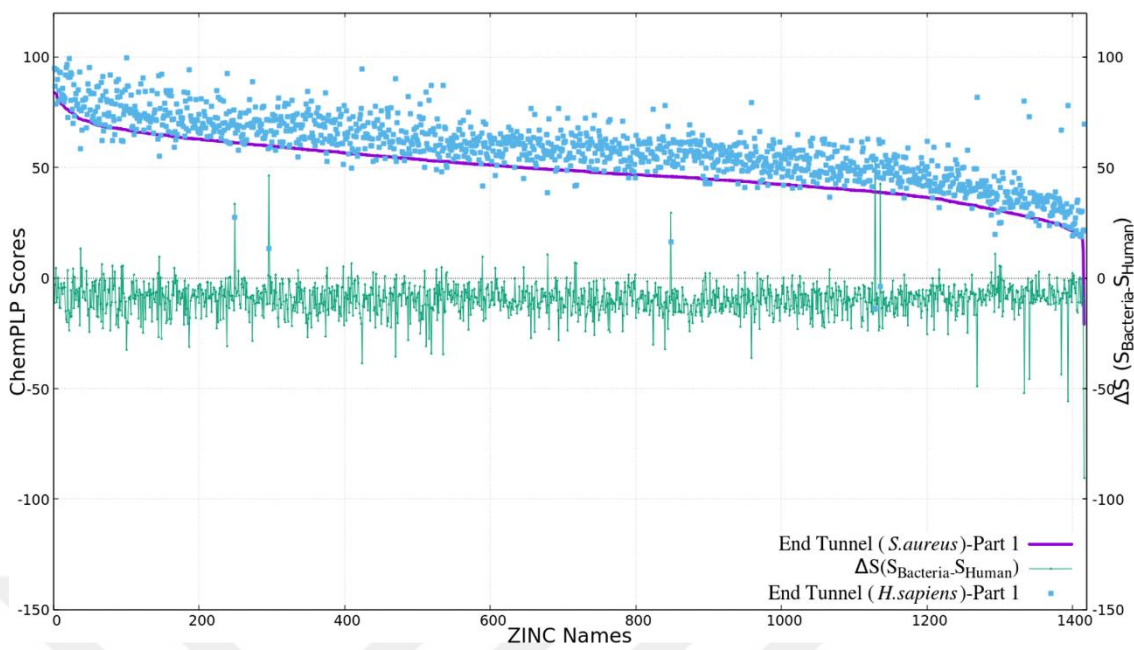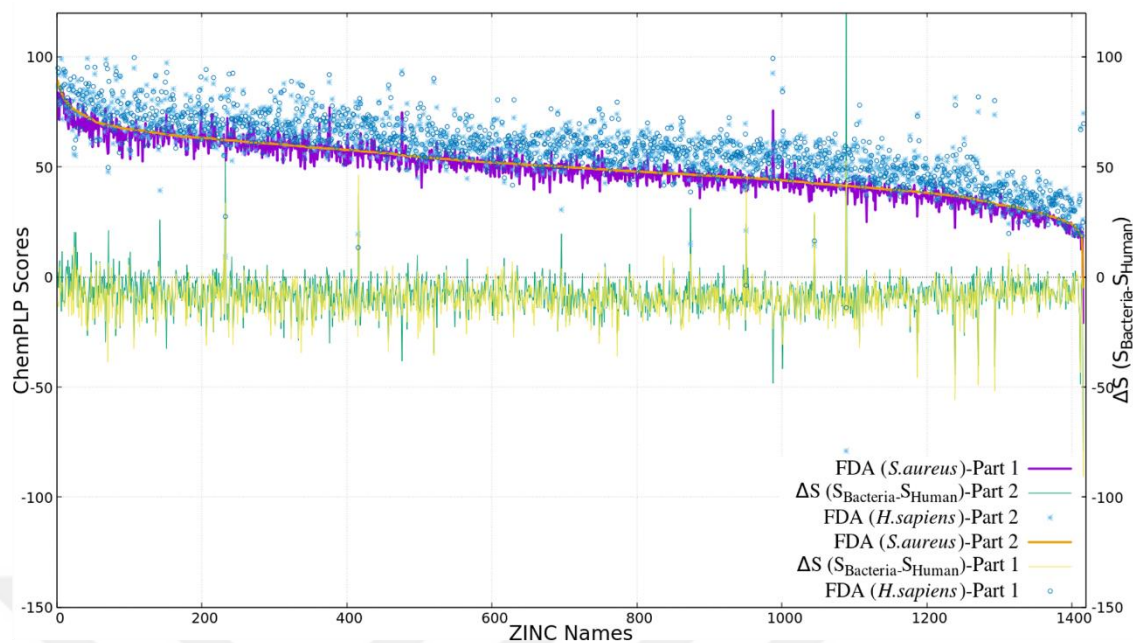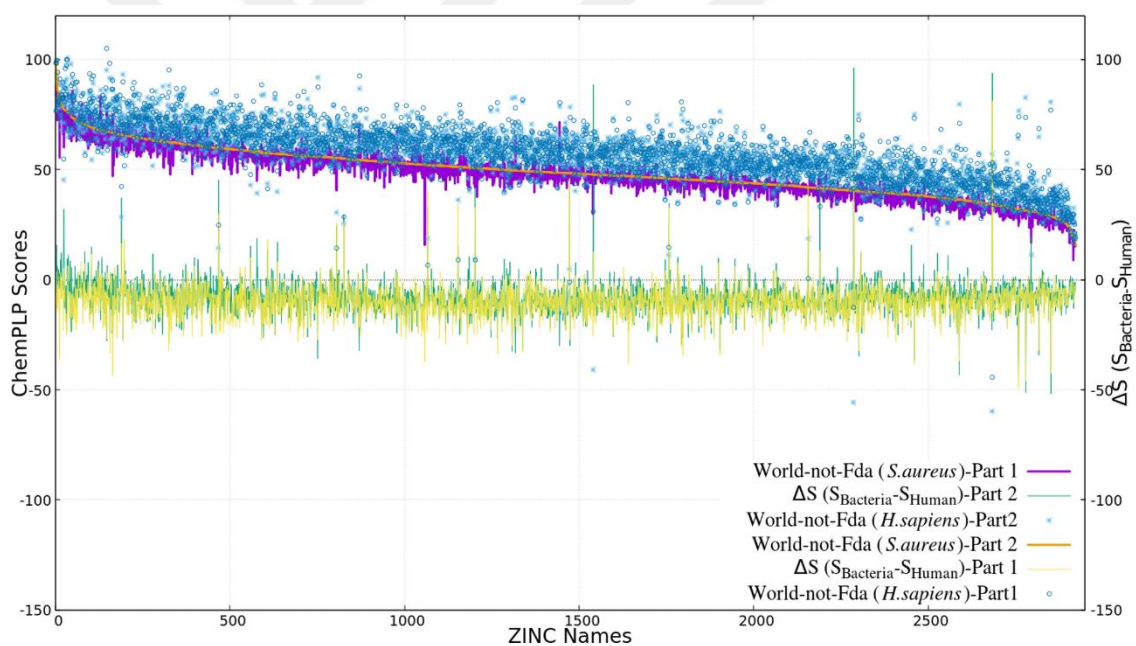
(a)
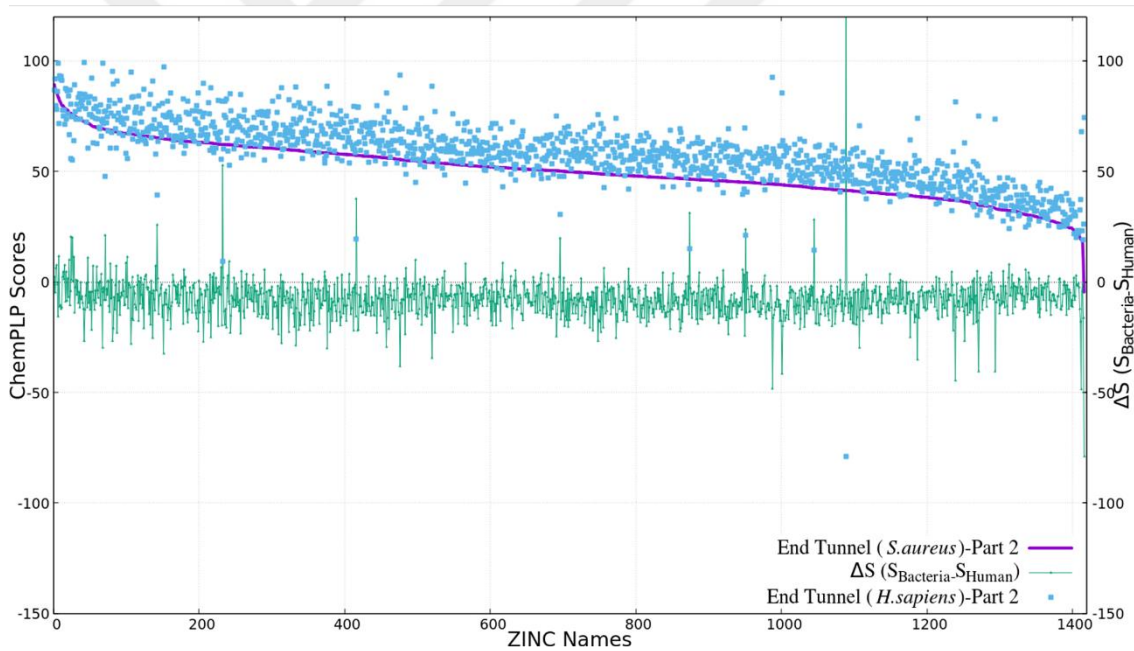
(b)

Figure 3.10 Docking result for end tunnel FDA molecules A) Part 1 B) Part 2 results

Figure 3.11 Comparison of *End* tunnel Part 1 and Part 2 a) FDA b) World-not-FDA molecules.

Molecules with docking scores calculated for *Sa*GADPH were sorted after, the molecules in the top 2% were selected. For FDA molecules and World-not-FDA molecules, the number of molecules at top 2% was found as 28 and 58, respectively. After the processes were applied for both regions (*End* and *Middle*), common molecules that exist in both

regions were determined.The number of remain molecules was determined as 7 and 25 for FDA and World-not-FDA molecules, respectively. The properties of the common molecules determined for the *End* and *Middle* Tunnel regions are given in detail in Table A.7 & A.8 in Appendix.. The whole docking process was summarized in the flowchart below (See Figure 3.13).
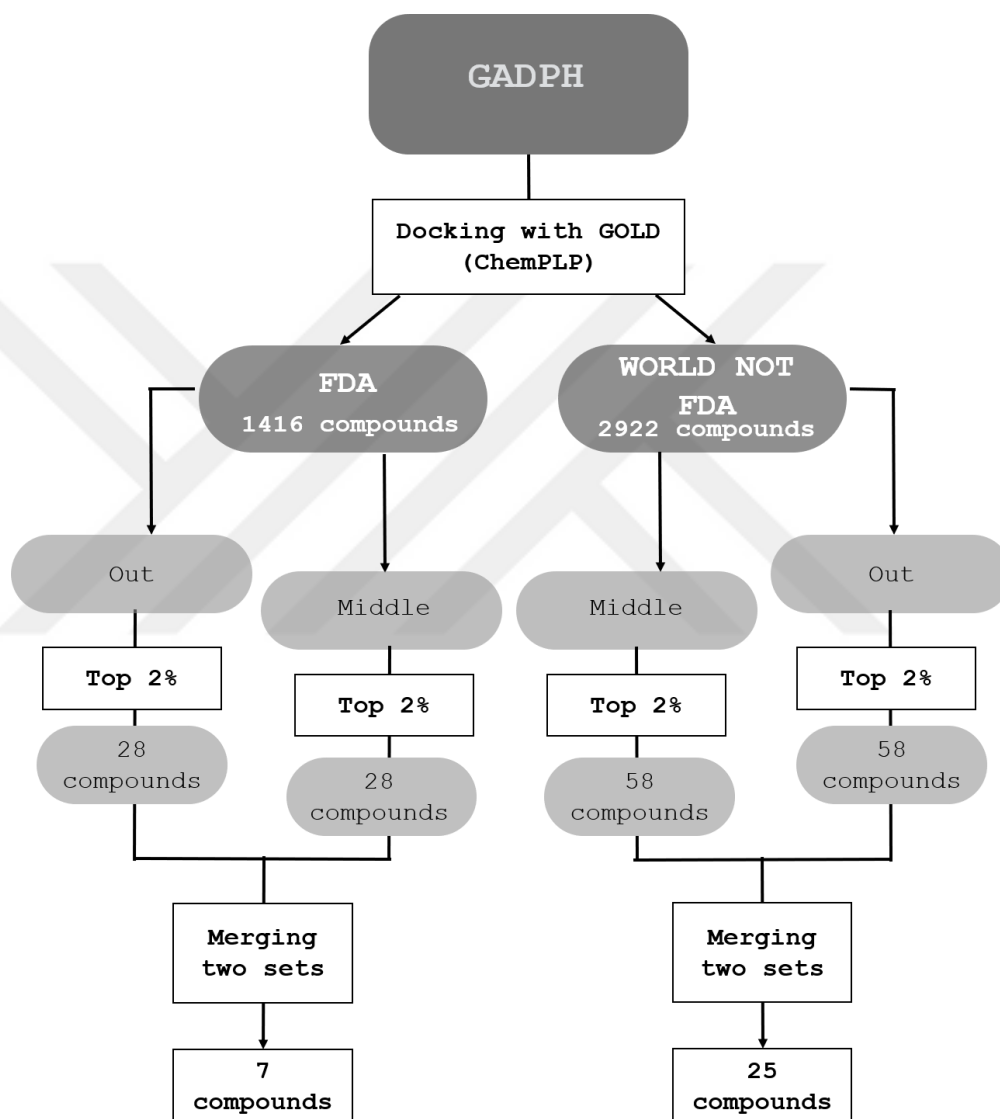

**Figure 3.12** Flowchart of the docking process.

When comparing the *End* tunnel and *Middle* tunnel regions determined in the *Sa*GADPH enzyme (See Figure 3.14a & 3.14b), it can be clearly observed that the End tunnel region provides a species-specific environment than the *Middle* tunnel region. In the *Middle* tunnel (See Figure 3.14b), this difference was generally negative in all regions. Briefly,

the molecule was bind to the *h*GADPH structure with a higher score than *Sa*GADPH structure. In addition, a remerkable negative difference was observed on the right side of the Figure 3.19b. The positive difference distribution for the *End* tunnel area showed a homogeneous distribution in all regions. Also, the number of molecules which is more selective to the bacteria was significantly higher than the *Middle* tunnel region (See Figure 3.14a). It is clearly seen that similar trends are observed for World-not-FDA molecules. (See Figure 3.15a & 3.15b) For this reason, it was estimated that the *End* Tunnel region among the specified regions will be a more suitable region for *Sa*GADPH for species-specific drug target.
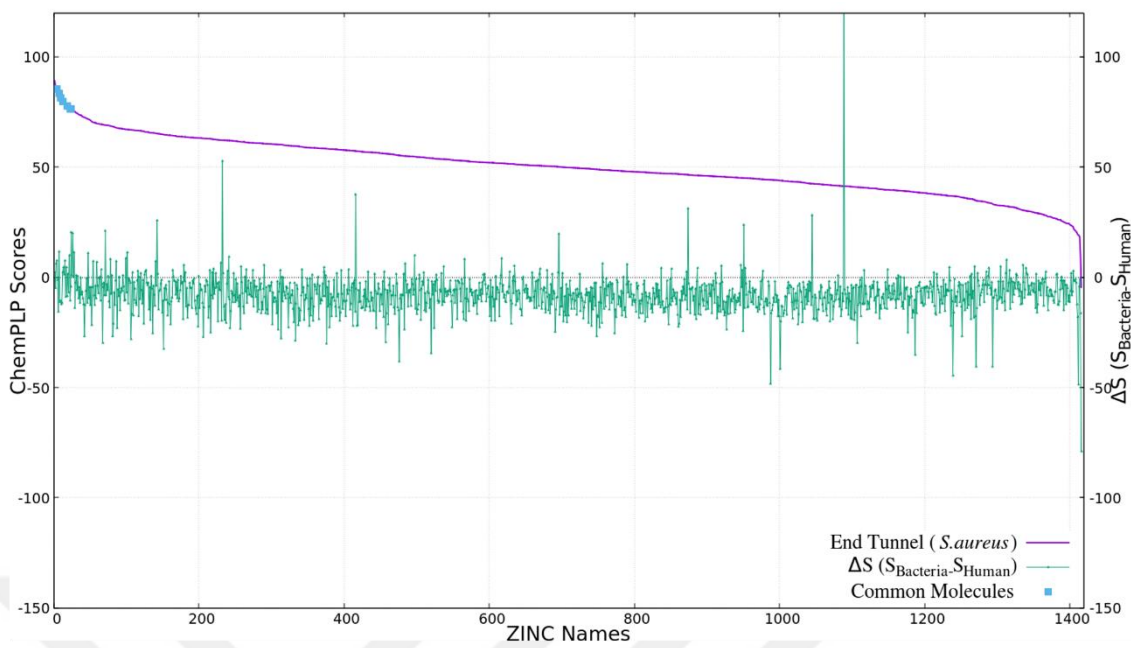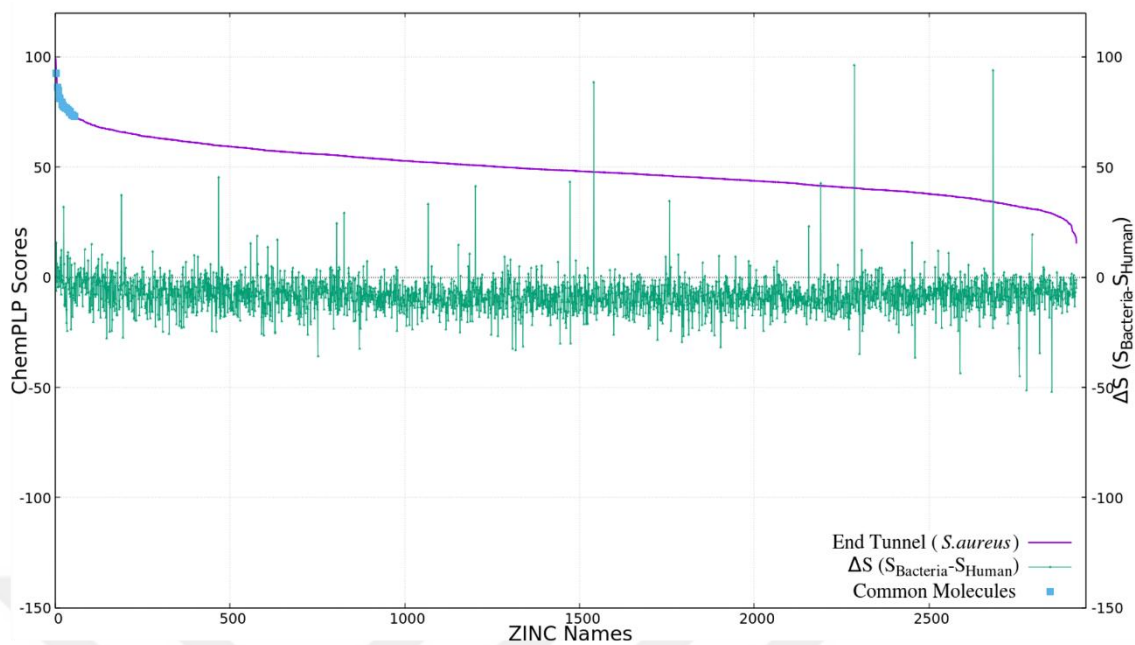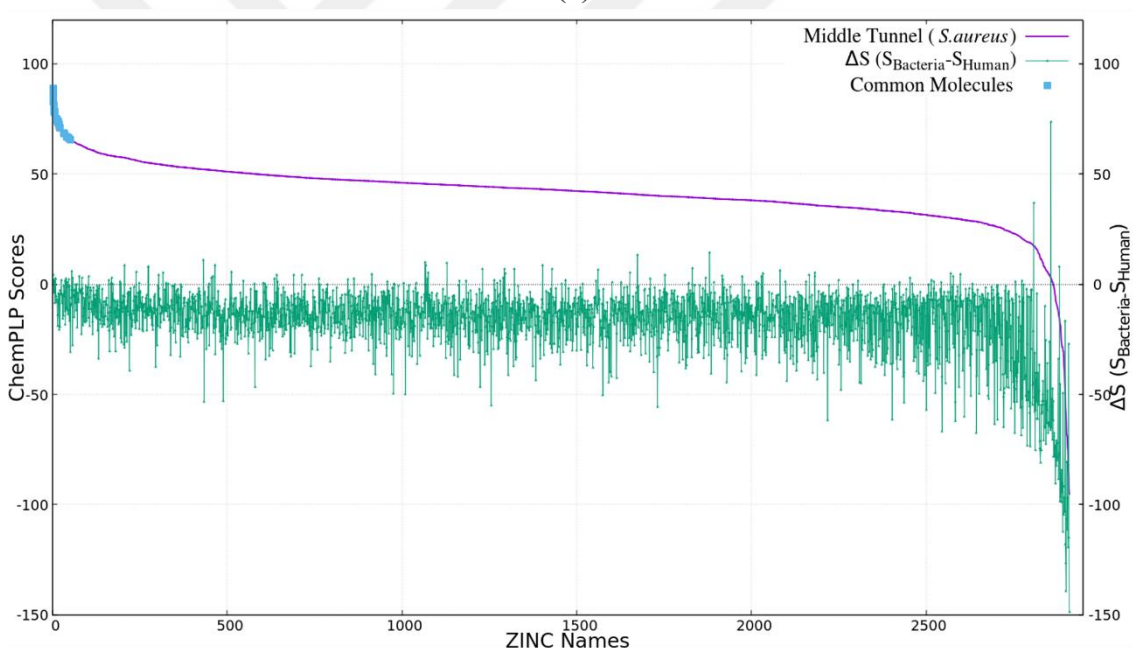
(a)
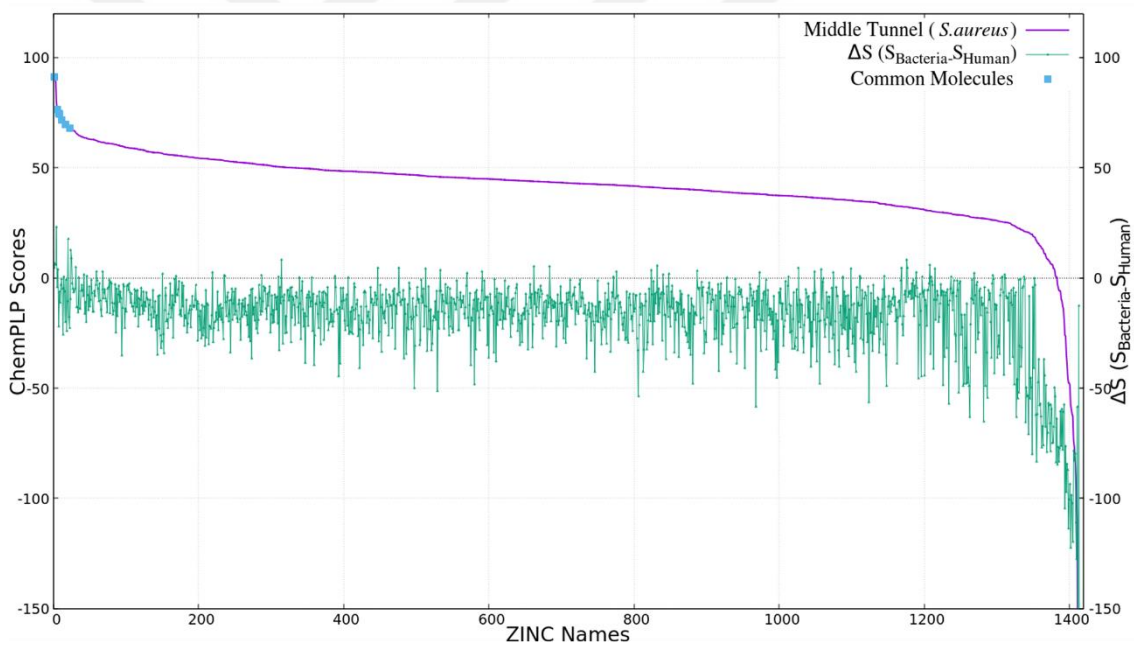
(b)

Figure 3.13 FDA molecules for *Sa*GADPH enzyme. a) *End* tunnel b) *Middle* tunnel.

(a)

(b)

Figure 3.14 World-not-FDA molecules for *Sa*GADPH enzyme. a) *End* tunnel b) *Middle* tunnel.

4   CONCLUSIONS

Combination of elastic network model, solvent mapping, structural and sequence alignment and docking played an important role in the discovery of alternative allosteric binding sites in GADPH, one of the key allosteric enzymes in the glycolytic pathway. The combination of all methods has revealed the presence of a potential tunnel region that will contribute to species-specific drug design in combination with the presence of catalytic regions previously described in the literature. The tunnel region contains the S-loop, which is thought to have a critical role in regulating its global enzymatic activity by phosphorylation. The ligand, which binds to two symmetrical sites in the tunnel region in the S-loop, contributed to the decrease in its activity in four catalytic sites, which played an important role in receptor activity.In addition, the fact that the tunnel region is one of the high druggable score pockets in DogSiteScore and caused significant changes in free energy of catalytic regionswas in accordance with the results of the applied methodology. Eventually, although the common potential tunnel region was observed in all three species, comparative sequence alignments of human/bacteria and human/parasite have shown that the tunnel region has remarkable differences in species-specific drug design. Our findings represent one of the allosteric enzymes in the glycolytic pathway. However, the present method is an approach that can be used for the detection of potential allosteric regions in other enzymes in the glycolytic pathway. Considering that all proteins has intrinsically allosteric, there is always the possibility of encountering new regions to be proposed as the target region for drug discovery.

Supportive results were obtained regarding the species specificity of potential allosteric regions determined by docking processes performed with "FDA" and "World-not-FDA" molecules in the Tunnel region. The findings showed that 7 molecules for the "FDA" and 25 molecules for the "World-not-FDA" were more strongly bound to the bacterial tunnel region than the human tunnel region. Briefly, the presence of some potential molecules to inhibit specifically bacteria and parasite more strongly than the host organism has been

determined for further experimental studies. The results obtained suggested that another criterion should be taken into consideration. These are molecules that bind to both the bacteria and the human tunnel region with high scores. World-not-FDA and FDA molecules are used in the treatment of various diseases in the world. The side effects of these molecules are acceptable for human health. Therefore, a molecule that binds to bacteria and human with a high score may not cause a serious effect on the human while there is a possibility of inhibiting the bacteria. Briefly, these molecules can be an effective method of inhibiting the bacteria on the host organism.

For further studies, the stabilization effect of "FDA" and "World-not-FDA" molecules, which have higher scores in bacteria than human, on the receptor's dynamics can be investigated. The ideal approach for this would be Molecular Dynamics (MD) simulation studies. The results from MD simulation will be an important key data on the necessity of testing the specified molecules at the experimental stage.

REFERENCES

Atilgan, A. R., S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar. 2001. "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model." *Biophysical Journal* 80 (1): 505–15. https://doi.org/10.1016/S0006-3495(01)76033-X.

Ayyildiz, M., S. Celiker, F. Ozhelvaci, and E.D. Akten. 2020. "Identification of Alternative Allosteric Sites in Glycolytic Enzymes for Potential Use as Species-Specific Drug Targets." *Frontiers in Molecular Biosciences*. https://doi.org/doi:10.3389/fmolb.2020.00088.

Bahar, Ivet, and A. J. Rader. 2005. "Coarse-Grained Normal Mode Analysis in Structural Biology." *Current Opinion in Structural Biology*. https://doi.org/10.1016/j.sbi.2005.08.007.

Berman, H M. 2000. "The Protein Data Bank / Biopython." *Presentation* 28 (1): 235–42. https://doi.org/10.1093/nar/28.1.235.

Bogan, Andrew A., and Kurt S. Thorn. 1998. "Anatomy of Hot Spots in Protein Interfaces." *Journal of Molecular Biology* 280 (1): 1–9. https://doi.org/10.1006/jmbi.1998.1843.

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

Carlile, Graeme W., Ruth M.E. Chalmers-Redman, Nadine A. Tatton, Amanda Pong, Katherine E. Borden, and William G. Tatton. 2000. "Reduced Apoptosis after Nerve Growth Factor and Serum Withdrawal: Conversion of Tetrameric Glyceraldehyde-3-Phosphate Dehydrogenase to a Dimer." *Molecular Pharmacology* 57 (1): 2–12.

Chakraborty, Apurba, Saptarsi Ghosh, Partha Mukhopadhyay, Syed Mukulika Dinara,

Ankush Bag, Mihir K Mahata, Rahul Kumar, et al. 2014. "Trapping Effect Analysis of AlGaN/InGaN/GaN Heterostructure by Conductance Frequency Measurement." In *MRS Proceedings*, XXXIII:81–87. https://doi.org/10.1007/s13398-014-0173-7.2.

Christopoulos, Arthur. 2002. "Allosteric Binding Sites on Cell-Surface Receptors: Novel Targets for Drug Discovery." *Nature Reviews Drug Discovery*. https://doi.org/10.1038/nrd746.

Doruker, Pemra, Ali Rana Atilgan, and Ivet Bahar. 2000. "Dynamics of Proteins Predicted by Molecular Simulations and Analytical Approaches: Application to α-Amylase Inhibitor." *Proteins: Structure, Function and Genetics* 40 (3): 512–24. https://doi.org/10.1002/1097-0134(20000815)40:3<512::AID-PROT180>3.0.CO;2-M.

Dubey, Rashmi, Bart L. Staker, Ian T. Foe, Matthew Bogyo, Peter J. Myler, Huân M. Ngô, and Marc Jan Gubbels. 2017. "Membrane Skeletal Association and Post-Translational Allosteric Regulation of Toxoplasma Gondii GAPDH1." *Molecular Microbiology*. https://doi.org/10.1111/mmi.13577.

Eldridge, Matthew D., Christopher W. Murray, Timothy R. Auton, Gaia V. Paolini, and Roger P. Mee. 1997. "Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes." *Journal of Computer-Aided Molecular Design*. https://doi.org/10.1023/A:1007996124545.

Ellis, John. 1997. "Allosteric Binding Sites on Muscarinic Receptors." *Drug Development Research*. https://doi.org/10.1002/(SICI)1098-2299(199702)40:2<193::AID-DDR9>3.0.CO;2-G.

Guarnera, Enrico, Zhen Wah Tan, Zejun Zheng, and Igor N. Berezovsky. 2017. "AlloSigMA: Allosteric Signaling and Mutation Analysis Server." *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btx430.

Guido, Rafael, Tatiane Balliano, Adriano Andricopulo, and Glaucius Oliva. 2009. "Kinetic and Crystallographic Studies on Glyceraldehyde-3-Phosphate Dehydrogenase from Trypanosoma Cruzi in Complex with Iodoacetate." *Letters in Drug Design & Discovery* 6 (3): 210–14. https://doi.org/10.2174/157018009787847774.

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

Haliloglu, Turkan, Ivet Bahar, and Burak Erman. 1997. "Gaussian Dynamics of Folded Proteins." *Physical Review Letters* 79 (16): 3090–93. https://doi.org/10.1103/PhysRevLett.79.3090.

Hammes-Schiffer, Sharon, and Stephen J. Benkovic. 2006. "Relating Protein Motion to Catalysis." *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev.biochem.75.103004.142800.

Hawkins, Rhoda J., and Tom C.B. McLeish. 2006. "Coupling of Global and Local Vibrational Modes in Dynamic Allostery of Proteins." *Biophysical Journal*. https://doi.org/10.1529/biophysj.106.082180.

Hornak, Viktor, Asim Okur, Robert C. Rizzo, and Carlos Simmerling. 2006. "HIV-1 Protease Flaps Spontaneously Close to the Correct Structure in Simulations Following Manual Placement of an Inhibitor into the Open State." *Journal of the American Chemical Society*. https://doi.org/10.1021/ja058211x.

Jones, Gareth, Peter Willett, Robert C. Glen, Andrew R. Leach, and Robin Taylor. 1997. "Development and Validation of a Genetic Algorithm for Flexible Docking." *Journal of Molecular Biology* 267 (3): 727–48. https://doi.org/10.1006/jmbi.1996.0897.

Korb, Oliver, Thomas Stützle, and Thomas E. Exner. 2009. "Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS." *Journal of Chemical Information and Modeling* 49 (1): 84–96. https://doi.org/10.1021/ci800298z.

Kortemme, Tanja, and David Baker. 2002. "A Simple Physical Model for Binding Energy Hot Spots in Protein-Protein Complexes." *Proceedings of the National Academy of Sciences of the United States of America* 99 (22): 14116–21. https://doi.org/10.1073/pnas.202485799.

Koshland, Daniel E., and Kambiz Hamadani. 2002. "Proteomics and Models for Enzyme Cooperativity." *Journal of Biological Chemistry*. https://doi.org/10.1074/jbc.R200014200.

Kozakov, Dima, Laurie E. Grove, David R. Hall, Tanggis Bohnuud, Scott E. Mottarella,

Lingqi Luo, Bing Xia, Dmitri Beglov, and Sandor Vajda. 2015. "The FTMap Family of Web Servers for Determining and Characterizing Ligand-Binding Hot Spots of Proteins." *Nature Protocols* 10 (5): 733–55. https://doi.org/10.1038/nprot.2015.043.

Kundu, Sibsankar, Julia S. Melton, Dan C. Sorensen, and George N. Phillips. 2002. "Dynamics of Proteins in Crystals: Comparison of Experiment with Simple Models." *Biophysical Journal*. https://doi.org/10.1016/S0006-3495(02)75203-X.

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

Levy, Emmanuel D. 2010. "A Simple Definition of Structural Regions in Proteins and Its Use in Analyzing Interface Evolution." *Journal of Molecular Biology* 403 (4): 660–70. https://doi.org/10.1016/j.jmb.2010.09.028.

Liebeschuetz, John W., Jason C. Cole, and Oliver Korb. 2012. "Pose Prediction and Virtual Screening Performance of GOLD Scoring Functions in a Standardized Test." *Journal of Computer-Aided Molecular Design*. https://doi.org/10.1007/s10822-012-9551-4.

Lou, Hongfeng, and Robert I. Cukier. 2006. "Molecular Dynamics of Apo-Adenylate Kinase: A Distance Replica Exchange Method for the Free Energy of Conformational Fluctuations." *Journal of Physical Chemistry B*. https://doi.org/10.1021/jp064303c.

Monod, J., J. P. Changeux, and F. Jacob. 1963. "Allosteric Proteins and Cellular Control Systems." *Journal of Molecular Biology* 6 (4): 306–29. https://doi.org/10.1016/S0022-2836(63)80091-1.

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

Mooij, Wijnand T.M., and Marcel L. Verdonk. 2005. "General and Targeted Statistical Potentials for Protein-Ligand Interactions." *Proteins: Structure, Function and Genetics*. https://doi.org/10.1002/prot.20588.

Mukherjee, Somnath, Debajyoti Dutta, Baisakhee Saha, and Amit Kumar Das. 2010a. "Crystal Structure of Glyceraldehyde-3-Phosphate Dehydrogenase 1 from

Methicillin-Resistant Staphylococcus Aureus MRSA252 Provides Novel Insights into Substrate Binding and Catalytic Mechanism." *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2010.07.002.

———. 2010b. "Crystal Structure of Glyceraldehyde-3-Phosphate Dehydrogenase 1 from Methicillin-Resistant Staphylococcus Aureus MRSA252 Provides Novel Insights into Substrate Binding and Catalytic Mechanism." *Journal of Molecular Biology*. https://doi.org/10.1016/j.jmb.2010.07.002.

Needleman, S. B., and C. D. Wunsch. 1970. "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins." *Journal of Molecular Biology* 48 (3): 443–53. https://doi.org/10.1016/0022-2836(70)90057-4.

Pan, H., J. C. Lee, and V.J. Hilser. 2000. "Binding Sites in Escherichia Coli Dihydrofolate Reductase Communicate by Modulating the Conformational Ensemble." *Proceedings of the National Academy of Sciences of the United States of America* 97 (22): 12020–25. https://doi.org/10.1073/pnas.220240297.

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

R Development Core Team. 2017. "R: A Language and Environment for Statistical Computing." *Vienna, Austria*. https://doi.org/R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reš, Ivica, and Olivier Lichtarge. 2005. "Character and Evolution of Protein-Protein Interfaces." *Physical Biology*. https://doi.org/10.1088/1478-3975/2/2/S04.

Rice, P., L. Longden, and A. Bleasby. 2000. "EMBOSS: The European Molecular Biology Open Software Suite." *Trends in Genetics* 16 (6): 276–77. https://doi.org/10.1016/S0168-9525(00)02024-2.

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

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

Sterling, Teague, and John J. Irwin. 2015. "ZINC 15 - Ligand Discovery for Everyone."

Journal of Chemical Information and Modeling 55 (11): 2324–37.
https://doi.org/10.1021/acs.jcim.5b00559.

Tama, Fl., and C.L. Brooks. 2006. "Symmetry, Form, and Shape: Guiding Principles for
Robustness in Macromolecular Machines." *Annual Review of Biophysics and
Biomolecular Structure* 35: 115–33.
https://doi.org/10.1146/annurev.biophys.35.040405.102010.

Tien, M.Z., A.G. Meyer, D. K. Sydykova, S. J. Spielman, and C.O. Wilke. 2013.
"Maximum Allowed Solvent Accessibilites of Residues in Proteins." *PLoS ONE* 8
(11): e80635. https://doi.org/10.1371/journal.pone.0080635.

Tirion, M. M. 1996. "Large Amplitude Elastic Motions in Proteins from a Single-
Parameter, Atomic Analysis." *Physical Review Letters* 77 (9): 1905–8.
https://doi.org/10.1103/PhysRevLett.77.1905.

Tobi, D., and I. Bahar. 2005. "Structural Changes Involved in Protein Binding Correlate
with Intrinsic Motions of Proteins in the Unbound State." *Proceedings of the
National Academy of Sciences of the United States of America* 102 (52): 18908–13.
https://doi.org/10.1073/pnas.0507603102.

Verdonk, Marcel L., Jason C. Cole, Michael J. Hartshorn, Christopher W. Murray, and
Richard D. Taylor. 2003. "Improved Protein-Ligand Docking Using GOLD."
Proteins: Structure, Function and Genetics. https://doi.org/10.1002/prot.10465.

Volkamer, A., D. Kuhn, F. Rippmann, and M. Rarey. 2012. "Dogsitescorer: A Web
Server for Automatic Binding Site Prediction, Analysis and Druggability
Assessment." *Bioinformatics* 28 (15): 2074–75.
https://doi.org/10.1093/bioinformatics/bts310.

Weber, G. 1972. "Ligand Binding and Internal Equilibria in Proteinst." *Biochemistry* 11
(5): 864–78. https://doi.org/10.1021/bi00755a028.

White, Michael R., Mohd M. Khan, Daniel Deredge, Christina R. Ross, Royston
Quintyn, Beth E. Zucconi, Vicki H. Wysocki, Patrick L. Wintrode, Gerald M.
Wilson, and Elsa D. Garcin. 2015. "A Dimer Interface Mutation in
Glyceraldehyde-3-Phosphate Dehydrogenase Regulates Its Binding to AU-Rich
RNA." *Journal of Biological Chemistry* 290 (3): 1770–85.
https://doi.org/10.1074/jbc.M114.618165.

CURRICULUM VITAE

Personal Information
Name Surname   : Serkan CELIKER
Place and Date of Birth   : Sivas 08/09/1988

Education
Undergraduate Education   : B.Sc Faculty of Education Department of Secondary School Science and Mathematics Fields Education Secondary Biology Education-19 Mayıs University(2006-2001)

Graduate Education   :M.Sc Graduate School for Science and Engineering Computational Biology and Bioinformatics (2018-2020)

Foreign Language Skills   : English

Contact:
Telephone   : 05454731402
E-mail Address   : serkan.ceeliker@gmail.com

APPENDIX
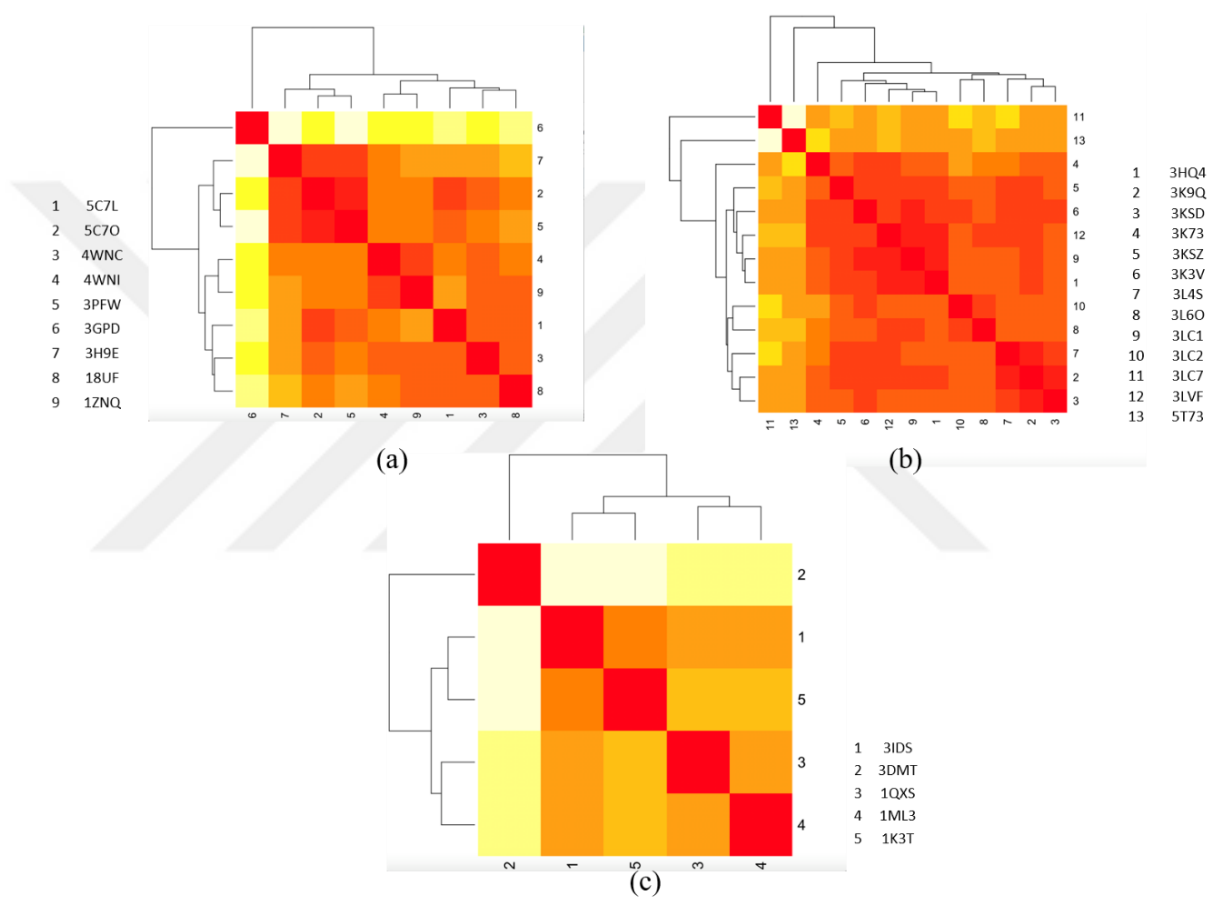
Figure A.1 Heatmap results of RMSD for three species. A) *H.sapiens* B) *S.aureus* C) *T.cruzi*
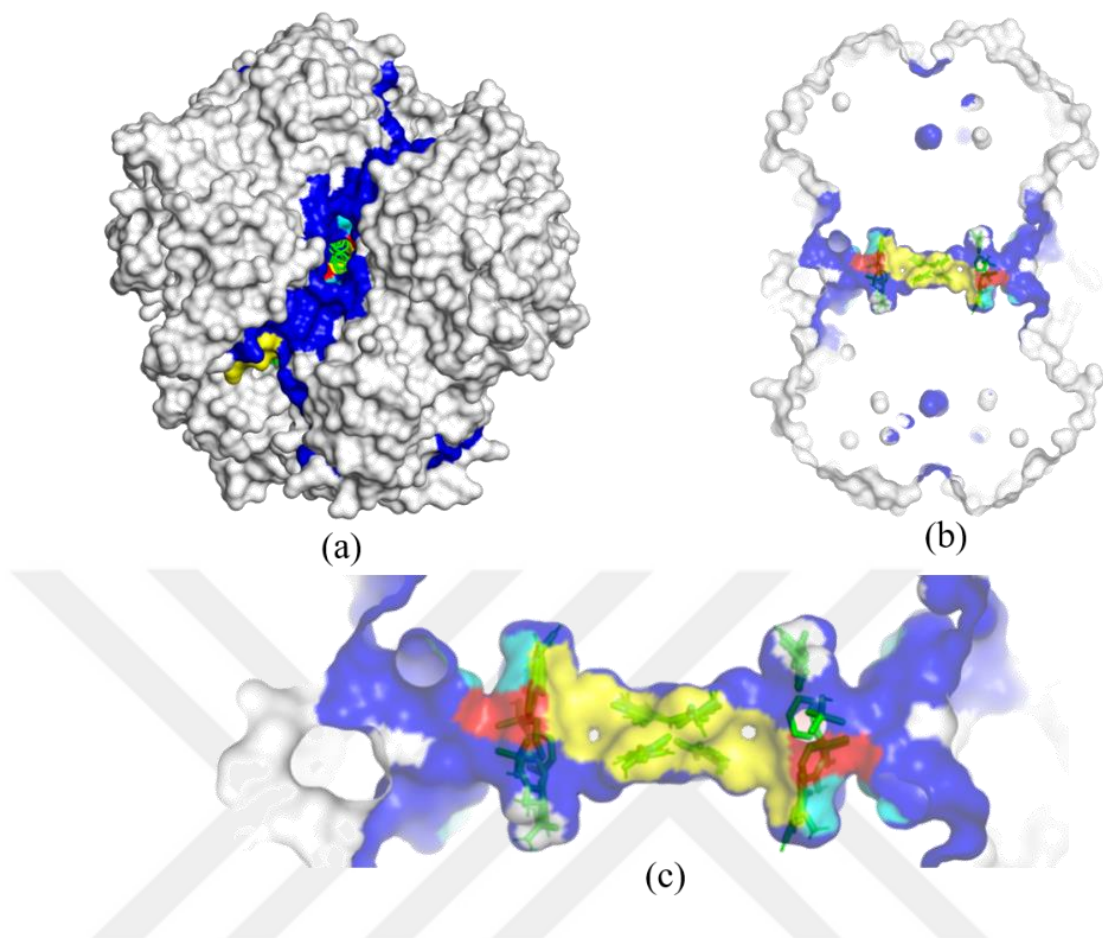
Figure A.2 The corridor-like tunnel region in *Tc*GADPH and S50 and S287 residues.
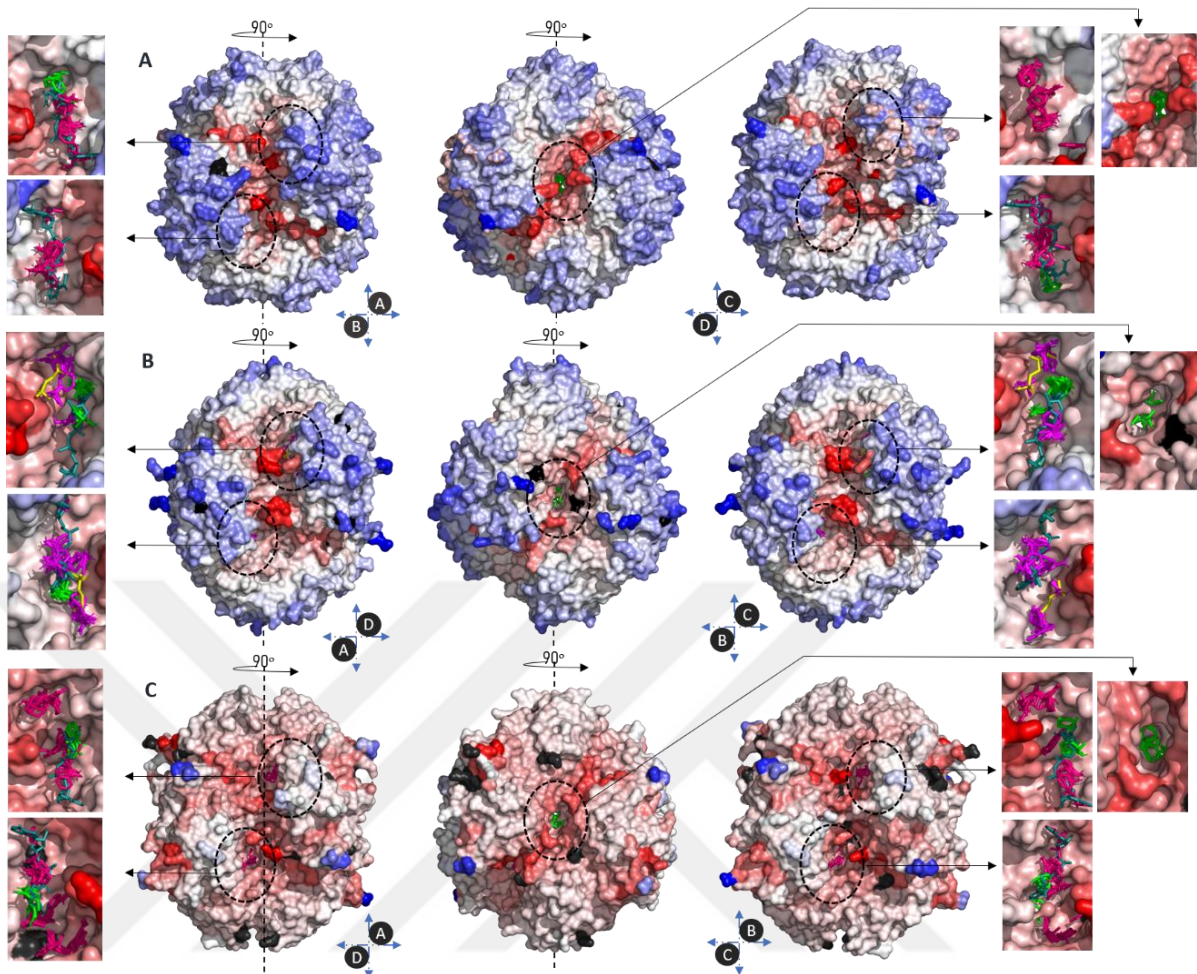
Figure A.3 Combination of ENM results and FTMap clusters.

Table A.1 RMSD values for *S.aureus* obtained from structural alignment.

S. AUREUS	3HQ4	3K9Q	3K73	3KSD	3KSZ	3K3V	3L4S	3L6O	3LC1	3LC2	3LC7	3LVF	5T73
3HQ4	0												
3K9Q	0.75	0											
3K73	0.81	0.41	0										
3KSD	0.87	0.9	0.88	0									
3KSZ	0.79	0.75	0.73	0.85	0								
3K3V	0.64	0.61	0.67	0.84	0.72	0							
3L4S	0.77	0.65	0.71	0.94	0.76	0.58	0						
3L6O	0.97	0.94	0.94	1.01	0.87	0.88	0.83	0					
3LC1	0.58	0.69	0.73	0.86	0.74	0.62	0.77	0.96	0				
3LC2	0.89	0.96	0.99	1.07	0.88	0.74	0.84	0.82	0.9	0			
3LC7	1.48	1.34	1.3	1.37	1.4	1.4	1.44	1.53	1.44	1.67	0		
3LVF	0.63	0.69	0.71	0.86	0.76	0.68	0.74	0.95	0.43	0.95	1.41	0	
5T73	1.91	1.99	2.06	0.938	1.92	1.89	1.82	1.83	1.96	1.5	2.61	1.99	0

Table A.2 RMSD values for *H.sapiens* obtained from structural alignment.

H.SAPIENS	5C7L	5C7O	4WNC	4WNI	3PFW	3GPD	3H9E	1U8F	1ZNQ,
5C7L	0								
5C7O	0.73	0							
4WNC	1.05	1.02	0						
4WNI	1.09	1.18	0.84	0					
3PFW	0.8	0.44	0.95	1	0				
3GPD	1.81	1.86	1.94	1.92	1.87	0			
3H9E	1	0.63	1.03	1.14	0.53	1.87	0		
1U8F	0.85	0.91	0.69	0.82	1	1.97	1.11	0	
1ZNQ,	0.99	0.91	0.82	0.88	0.97	1.89	1.03	0.89	0

Table A.3 RMSD value for *T.cruzi* obtained from structural alignment.

T.CRUZI	3IDS	3DMT	1QXS	1ML3	1K3T
3IDS	0				
3DMT	1.2	0			
1QXS	0.89	1.07	0		
1ML3	0.92	1.22	0.77	0	
1K3T	0.91	1.34	1	0.99	0

Table A.4 Summary of all CSs that exceed the 25% threshold for *h*GADPH with their average frequency shift value.

CLUSTERS	MEAN FREQUENCY SHIFT		SUM OF FREQUENCY SHIFT VALUE	TOTAL NUMBER OF RESİDUES
		4WNI CHAIN A		
CLUSTER 2	0.96 ± 0.89		13.44	14
CLUSTER 4	0.57 ± 1.06		8.05	14
CLUSTER 6	0.35 ± 1.17		5.28	15
		4WNI CHAIN B		
CLUSTER 2	0.73 ± 0.83		14.74	20
CLUSTER 4	0.66 ± 0.92		9.32	14
CLUSTER 6	0.28 ± 1.15		4.62	16
		4WNI CHAIN C		
CLUSTER 2	0.62 ± 0.76		9.97	16
CLUSTER 3	0.31 ± 0.38		4.1	13
CLUSTER 5	1.07 ± 0.71		11.81	11
CLUSTER 8	0.28 ± 1.17		2.89	10
		4WNI CHAIN D		
CLUSTER 2	0.72 ± 0.83		15.17	21
CLUSTER 5	0.37 ± 1.23		4.86	13
CLUSTER 8	0.29 ± 1.20		3.24	11
		4WNI OVERALL		
CLUSTER 1	1.15 ± 0.35		20.77	18
CLUSTER 2	1.27 ± 0.46		24.29	19
CLUSTER 3	1.25 ± 0.43		20.05	16
CLUSTER 4	1.26 ± 0.42		24.06	19
CLUSTER 5	1.13 ± 0.35		20.41	18
CLUSTER 6	1.19 ± 0.39		16.73	14
CLUSTER 7	0.69 ± 0.49		11.1	16
CLUSTER 8	0.78 ± 0.50		10.17	13

Table A.5 Summary of all CSs that exceed the 25% threshold for *Sa*GADPH wih their average frequency shift values.

CLUSTERS	MEAN FREQUENCY SHIFT	SUM OF FREQUENCY SHIFT VALUE	TOTAL NUMBER OF RESIDUES
		3HQ4 CHAIN R	
CLUSTER 3	0.88 ± 0.42	14.12	16
CLUSTER 4	1.04 ± 0.53	13.58	13
CLUSTER 5	1.47 ± 0.69	7.35	5
		3HQ4 CHAIN O	
CLUSTER 1	0.51 ± 0.33	7.79	15
CLUSTER 2	0.84 ± 0.39	14.33	17
CLUSTER 5	0.98 ± 0.31	8.84	9
CLUSTER 7	1.40 ± 0.63	8.43	6
CLUSTER 8	0.51 ± 0.35	4.65	9
CLUSTER 9	0.57 ± 0.64	5.79	10
		3HQ4 CHAIN P	
CLUSTER 2	0.58 ± 0.55	9.42	16
CLUSTER 3	0.78 ± 0.36	10.96	14
CLUSTER 4	0.86 ± 0.31	11.25	13
CLUSTER 6	0.28 ± 0.78	3.74	13
CLUSTER 7	1.66 ± 0.58	6.64	4
CLUSTER 8	1.38 ± 0.61	8.33	6
		3HQ4 CHAIN Q	
CLUSTER 3	0.81 ± 0.39	12.28	15
CLUSTER 4	0.57 ± 0.31	6.34	11
CLUSTER 6	1.78 ± 0.59	7.12	4
CLUSTER 7	0.43 ± 0.27	3.46	8
		3HQ4 OVERALL	
CLUSTER 1	0.82 ± 0.26	17.29	21
CLUSTER 2	0.78 ± 0.39	11.81	15
CLUSTER 3	0.70 ± 0.35	9.16	13
CLUSTER 4	1.16 ± 0.23	11.6	10
CLUSTER 5	0.76 ± 0.40	11.52	15
CLUSTER 6	0.84 ± 0.29	13.52	16
CLUSTER 7	1.66 ± 3.47	23.3	14
CLUSTER 8	0.52 ± 034	5.24	10
CLUSTER 10	1.39 ± 0.59	15.3	11
CLUSTER 11	1.23 ± 0.27	18.46	15
CLUSTER 12	2.17 ± 3.79	23.94	11
CLUSTER 14	0.53 ± 0.32	4.79	9

Table A.6 Summary of all CSs that exceed the 25% threshold for *Tc*GADPH wih their average frequency shift values.

CLUSTERS	MEAN FREQUENCY SHIFT	SUM OF FREQUENCY SHIFT VALUE	TOTAL NUMBER OF RESIDUES
		3DMT CHAIN A	
CLUSTER 2	0.59 ± 0.29	10.75	18
CLUSTER 3	2.05 ± 5.76	28.7	14
CLUSTER 4	1.90 ± 5.40	30.52	16
CLUSTER 5	0.57 ± 0.25	5.14	9
CLUSTER 7	1.34 ± 0.04	4.04	3
		3DMT CHAIN B	
CLUSTER 1	0.58 ± 0.31	11.74	20
CLUSTER 3	0.45 ± 0.27	5.97	13
CLUSTER 6	0.59 ± 0.46	7.74	13
CLUSTER 7	0.76 ± 0.42	9.94	13
CLUSTER 8	0.51 ± 0.20	4.63	9
		3DMT CHAIN C	
CLUSTER 2	1.83 ± 4.98	36.67	20
CLUSTER 4	0.54 ± 0.39	10.29	19
CLUSTER 5	0.42 ± 0.24	5.57	13
CLUSTER 6	0.25 ± 0.38	3.01	12
CLUSTER 7	0.28 ± 0.17	3.38	12
		3DMT CHAIN D	
CLUSTER 1	2.72 ± 6.10	40.83	15
CLUSTER 3	2.42 ± 5.60	43.69	18
CLUSTER 4	1.86 ± 5.44	29.79	16
CLUSTER 6	1.52 ± 3.15	16.76	11
CLUSTER 7	2.82 ± 6.76	28.25	10
CLUSTER 8	0.34 ± 0.25	3.13	9
CLUSTER 9	0.84 ± 0.48	6.78	8
		3DMT OVERALL	
CLUSTER 1	0.35 ± 0.15	5.29	15
CLUSTER 2	0.38 ± 0.19	7.01	12
CLUSTER 4	0.25 ± 0.10	2.58	10
CLUSTER 5	0.43 ± 0.23	8.78	20
CLUSTER 6	0.27 ± 0.12	3.6	13
CLUSTER 7	0.43 ± 0.22	6.91	16
CLUSTER 8	0.25 ± 0.10	3.27	13
CLUSTER 9	0.39 ± 0.19	4.34	11
CLUSTER 10	0.25 ± 0.99	3.25	13
CLUSTER 11	0.36 ± 0.22	4.76	13
CLUSTER 12	2.11 ± 5.97	27.48	16
CLUSTER 13	0.42 ± 0.24	5.09	12
CLUSTER 14	0.40 ± 0.20	4.07	10
CLUSTER 15	0.35 ± 0.14	3.56	10

Table A.7 Molecular features of 7 FDA molecules with ChemPLP score values for *Middle* and *End*.
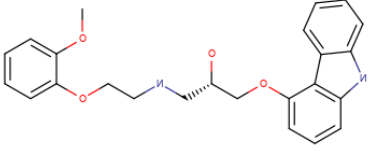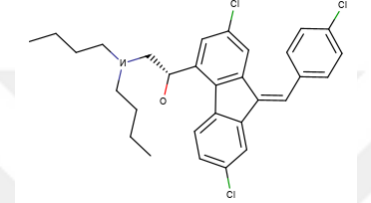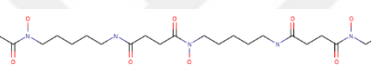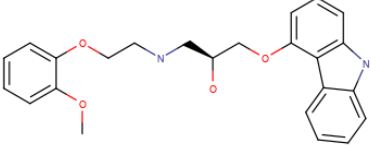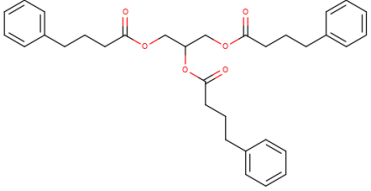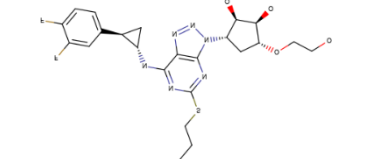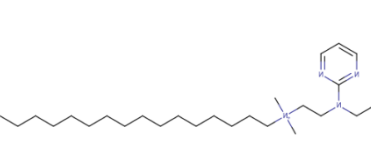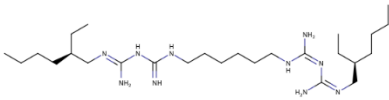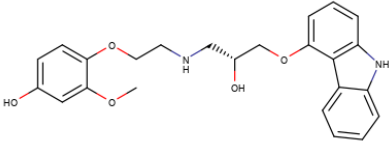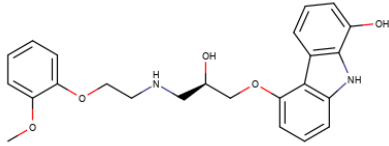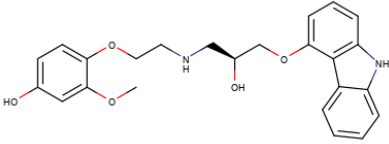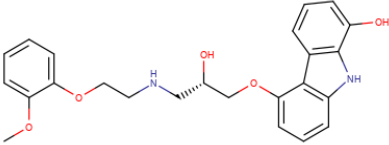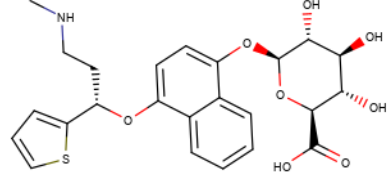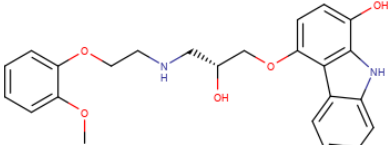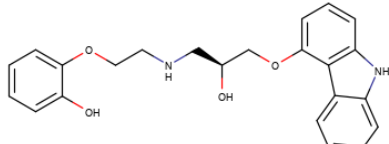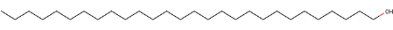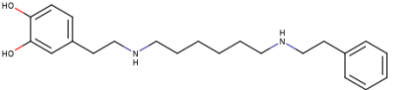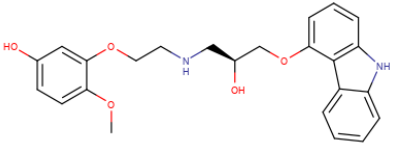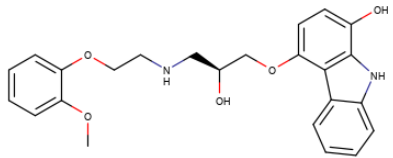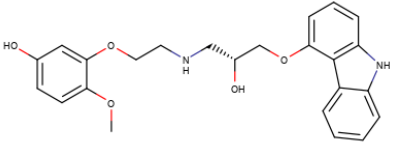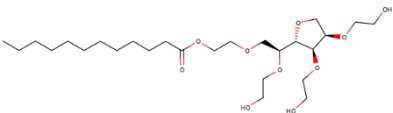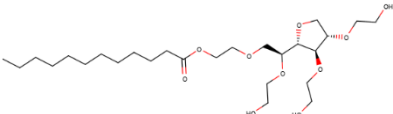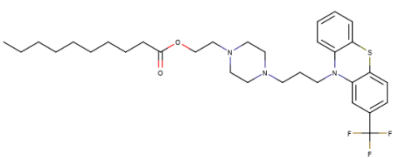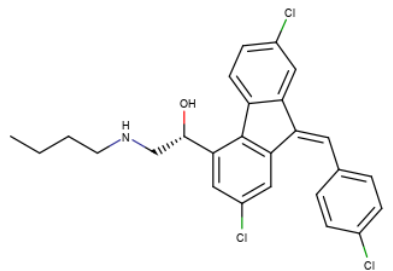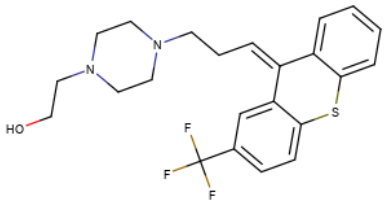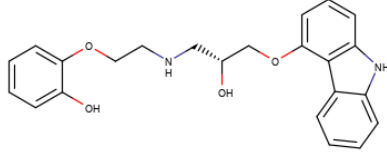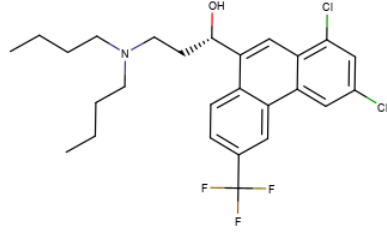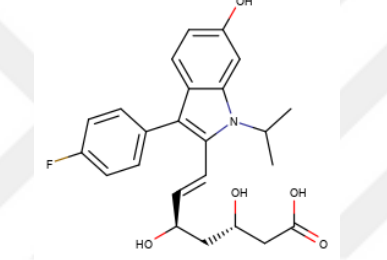
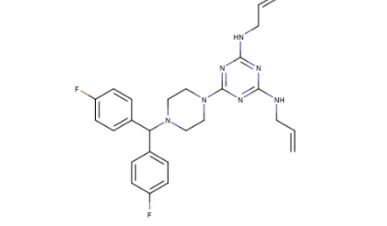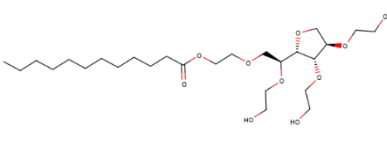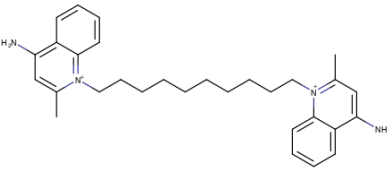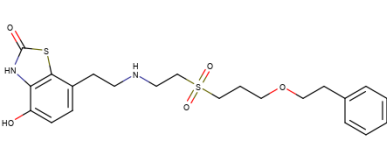| ZINC ID	MOL. NAME	MOLECULE STRUCTURE	MOL. WEIGHT	SChemPLP (*S.Aureus*)		SChemPLP (*H.Sapiens*)		ΔS (*S.Aureus - H.Sapiens*)	
				End	**Middle**	**End**	**Middle**	**End**	**Middle**
ZINC000 0015305 80	Coreq		406.482	85.46	74.41	77.97	80.60	7.49	-6.19
ZINC000 1000043 43	Coartem		528.951	83.34	71.63	71.65	72.12	11.69	-0.49
ZINC000 0038306 35	Dfo		560.693	81.49	74.70	92.07	96.72	-10.58	-22.02
ZINC000 0015305 79	Coreq		406.482	79.84	76.24	77.72	80.34	2.12	-4.1
ZINC000 0389456 66	Ravicti		530.661	77.85	69.70	86.41	89.69	-8.56	-19.99
ZINC000 0289574 44	Brilinta		522.578	76.28	67.87	86.98	90.63	-10.7	-22.76
ZINC000 0037852 68	Salmeter ol		415.574	76.27	91.07	80.48	86.51	-4.21	4.56

GLYCERALDEHYDE 3 PHOSPHATE DEHYDROGENASE(GADPH)-FDA

Table A.8 Molecular features of 25 FDA molecules with ChemPLP score values for *Middle* and *End*.

				GLYCERALDEHYDE 3 PHOSPHATE DEHYDROGENASE(GADPH)-Worl not FDA						

ZINC ID	MOL. NAME	MOLECULE STRUCTURE	MOL. WEIGHT	SChemPLP (*S.Aureus*)		SChemPLP (*H.Sapiens*)		ΔS (*S.Aureus - H.Sapiens*)	
				End	Middle	End	Middle	End	Middle
ZINC00003 8144598	Alexidine		508.804	92.51	78.47	76.75	81.26	15.76	-2.79
ZINC00000 2526388	4-Hydroxycar vedilol		422.481	86.25	82.95	82.04	82.04	4.21	0.91
ZINC00009 5617650	8-Hydroxycar vedilol		422.481	84.67	73.66	89.04	81.79	-4.37	-8.13
ZINC00000 2528509	4-Hydroxycar vedilol		422.481	82.73	77.61	78.08	86.23	4.65	-8.62
ZINC00009 5617649	8-Hydroxycar vedilol		422.481	81.49	74.94	82.24	82.44	-0.75	-7.5
ZINC00002 2060640	4-Hydroxy Duloxetine Glucuronid e		489.546	81.31	72.27	78.49	79.76	2.82	-7.49
ZINC00000 1540640	1-Hydroxycar vedilol		422.481	81.12	66.74	83.06	80.36	-1.94	-13.6
ZINC00000 2528486	O-Desmethylc arvedilol		392.455	79.54	80.96	78.82	84.73	0.72	3.77

ZINC00000 8221075	Octacosyl		410.771	78.58	75.11	80.92	80.62	-2.34	-5.51
ZINC00000 3798745	Dopexamine		356.51	78.42	88.91	79.47	84.67	-1.05	4.24
ZINC00000 2528510	5-Hydroxycar vedilol		422.481	78.23	74.55	79.73	82.87	-1.5	-8.32
ZINC00000 6019305	1-Hydroxycar vedilol		422.481	77.49	74.75	80.73	81.56	-3.24	-6.81
ZINC00000 2526389	5-Hydroxycar vedilol		422.481	7706	73.75	81.39	87.61	-4.33	-13.06
ZINC00000 8445713	None		522.676	76.85	71.10	87.32	90.38	-10.47	-19.28
ZINC00021 5581164	None		522.676	76.83	65.60	79.47	96.23	-2.64	-30.63
ZINC00002 2816725	Fluphenazi ne Decanoate		591.784	76.76	72.02	91.38	92.77	-14.62	-20.75
ZINC00010 0054862	Desbutyl-Lumefantri ne		472.843	75.99	68.26	64.62	71.84	11.37	-3.58

ZINC ID	Name	Structure							
ZINC000029489118	Flupentixol		434.527	75.42	73.09	76.06	77.39	-0.64	-4.3
ZINC000002509755	O-Desmethylcarvedilol		392.455	75.21	8617	82.87	84.40	-7.66	1.77
ZINC000001530866	Halofantrine		500.432	75.04	67.08	66.42	77.61	8.62	-10.53
ZINC000095617672	6-Hydroxyfluvastatin		427.472	74.71	68.50	83.41	82.80	-8.7	-14.3
ZINC000004214772	Almitrine		477.563	73.70	71.72	78.64	76.12	-4.94	-4.4
ZINC000008552163	E432		522.676	7362	73.88	78.61	92.81	-4.99	-18.93
ZINC000001655706	Dequalinium		456.678	73.47	66.23	82.86	78.06	-9.39	-11.83
ZINC000036268680	Sibenadet		464.609	73.00	78.55	83.59	77.06	-10.59	1.49