

**IDENTIFICATION OF ALLOSTERIC CONTROL  
IN PROTEINS USING COMPUTATIONAL  
METHODS**

**A Thesis Submitted to  
the Graduate School for Engineering and Sciences of  
İzmir Institute of Technology  
in Partial Fulfillment of the Requirements for the Degree of**

**MASTER OF SCIENCE**

**in Polymer Science and Engineering**

**by  
Sude GÜNEŞ**

**July 2024  
İZMİR**

We approve the thesis of **Sude GÜNEŞ**

---

**Asst. Prof. Dr. Arzu UYAR**

Department of Bioengineering, İzmir Institute of Technology

---

**Assoc. Prof. Dr. Beste BAYRAMOĞLU**

Department of Food Engineering, İzmir Institute of Technology

---

**Asst. Prof. Dr. Cihangir YANDIM**

Department of Genetics and Bioengineering, İzmir University of Economics

**11 July 2024**

---

**Asst. Prof. Dr. Arzu UYAR**

Supervisor, Department of Bioengineering  
İzmir Institute of Technology

---

**Asst. Prof. Dr. Hümevra**

**TAŞKENT SEZGİN**

Co-Supervisor, Department of  
Bioengineering  
İzmir Institute of Technology

---

**Prof. Dr. Ekrem ÖZDEMİR**

Head of the Department of Polymer Science and  
Engineering

---

**Prof. Dr. Mehtap EANES**

Dean of the Graduate School  
of Engineering and Sciences

## ACKNOWLEDGMENTS

I would like to thank my supervisor Asst. Prof. Dr. Arzu UYAR who always shared her knowledge and supported me starting from April 2022. I am grateful for every information, and suggestion that she shared and for her encouragement to pursue my dreams. It is my chance to be a student at UBioModel Group under her supervision.

I would like to thank my co-supervisor Asst. Prof. Dr. Hümeýra TAŞKENT SEZGİN, who provided me insights on proteins from my undergraduate to graduate education. I am grateful to her for every information she shared and her support. I am very lucky to have their guidance throughout my education.

I would like to thank my thesis committee members Asst. Prof. Dr. Arzu UYAR, Assoc. Prof. Dr. Beste BAYRAMOĞLU, and Asst. Prof. Dr. Cihangir YANDIM for their contributions to my thesis.

Special thanks to Prof. Dr. Wyatt W. Yue and Dr. Sabrina R. Mackinnon from Newcastle University for their collaboration on the galactokinase 1 enzyme. I am grateful for their suggestions and contributions to my thesis.

I would like to acknowledge the IZTECH Scientific Research Projects Coordinator Unit (project no: AUDP-2022-İYTE-3-0041 and 2023IYTE-1-0055) for financial support. In addition, the numerical calculations reported in this paper were fully performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources), and Google Colab.

I would like to thank Ekin Kestevur Doğru who supported me throughout my undergraduate and graduate education and made me realize my interest in doing computational research on proteins. I would also like to thank Mehmet Can for his contributions to ESSA scripts and all UBioModel Group members who always supported me. I am very lucky to have such friends and thankful for scientific discussions.

I would like to thank Mehmet Emin Aygen, Tuğba Bayer, Ceren Elmacı, Nurşah Halisdemir, and Dilara Demirel for their support and great friendship.

I am grateful to my mother, father, and brother, who always supported me and believed in me for pursuing my dreams.

# ABSTRACT

## IDENTIFICATION OF ALLOSTERIC CONTROL IN PROTEINS USING COMPUTATIONAL METHODS

The prediction of orthosteric, allosteric, and cryptic sites through computational approaches is significant for drug discovery studies. To this extent, a ligand binding site prediction method called Essential Site Scanning Analysis (ESSA) was used to investigate allosteric and cryptic sites of TEM-1 beta-lactamase, cell division protein kinase 2, and galactokinase by applying various cutoff values (7 Å, 10 Å, 11 Å, and 13 Å) and combining 10/20 modes in this study. On the other hand, molecular dynamics (MD) and ClustENMD were performed to investigate *hGALK1* enzyme dynamics. In addition, Principal Component Analysis (PCA) was applied to investigate collective motions in the presence and the absence of an allosteric inhibitor in *hGALK1*. The ligand binding sites of *hGALK1* were investigated using frames obtained from MD and ClustENMD. According to the findings, the application of combined 10/20 modes improved the success of the ESSA method. Moreover, binding site prediction success demonstrated variation in the presence of different cutoff values and conformations. Different cutoff values demonstrated success in the prediction of allosteric and orthosteric sites highlighting the role of contacting atoms in the Gaussian Network Model (GNM) calculations. To this extent, the combination of results from various cutoff values has the potential to improve the allosteric site prediction success of the ESSA method.

# ÖZET

## PROTEİNLERDE ALOSTERİK KONTROLÜN HESAPLAMALI YÖNTEMLER KULLANILARAK TANIMLANMASI

Ortosterik, alosterik, ve kriptik bölgelerin hesaplamalı yöntemlerle tahmini ilaç keşif çalışmaları için oldukça önemlidir. Bu doğrultuda, ligand bağlanma bölgesi tespitine yönelik Temel Bölge Tarama Analizi (ESSA) yöntemi, TEM-1 beta-laktamaz, hücre bölünmesi proteini kinaz 2, ve galaktokinazdaki (GALK1) alosterik ve kriptik bölgelerin araştırılması için çeşitli yarıçap değerleri kullanılarak (7 Å, 10 Å, 11 Å ve 13 Å) ve 10/20 modun birleştirilmesi şeklinde uygulanarak bu çalışmada kullanılmıştır. Diğer yandan, moleküler dinamik (MD) ve ClustENMD hesapları *hGALK1* enziminin dinamiğinin araştırılması için gerçekleştirilmiştir. Ek olarak Temel Bileşen Analizi (PCA) *hGALK1*'de alosterik inhibitör varlığında ve yokluğundaki kolektif hareketleri araştırmak üzere uygulanmıştır. MD ve ClustENMD'den elde edilen konformasyonlar kullanılarak *hGALK1*'e ait ligand bağlanma bölgeleri araştırılmıştır. Bulgulara göre kombine 10/20 modun uygulanması ESSA metodunun başarısını artırmıştır. Dahası, bağlanma bölge tespitinin başarısı farklı yarıçap değerlerinin ve konformasyonların varlığında çeşitlilik göstermiştir. Farklı yarıçap değerleri Gaussian Ağ-yapı Modeli (GNM) hesaplamalarında bağlantıdaki atomların rolünü vurgulayarak alosterik ve ortosterik bölgelerin tahmininde başarı göstermiştir. Bu doğrultuda, çeşitli yarıçap değerlerinden elde edilen sonuçların kombinasyonu ESSA metodunun alosterik bölge tahminindeki başarısını artırma potansiyeline sahiptir.

# TABLE OF CONTENTS

LIST OF FIGURES .....	viii
LIST OF TABLES .....	xi
ABBREVIATIONS .....	xiii
CHAPTER 1. INTRODUCTION .....	1
1.1. Allostery.....	1
1.1.1. The Detection of Allosteric Sites by Computational Methods .....	1
1.2. Cryptic Pockets .....	3
1.2.1. The Investigation of Cryptic Pockets Using Computational Methods .....	4
1.3. The Investigation of Protein Dynamics Using Computational Methods .....	5
1.4. Galactosemia and Classical Galactosemia.....	6
1.4.1. Galactokinase 1 (GALK1) .....	8
1.4.2. The Investigation of <i>h</i> GALK1 Dynamics Using Computational Methods .....	10
1.5. Aim of The Study .....	11
CHAPTER 2. MATERIALS AND METHODS .....	13
2.1. Materials .....	13
2.2. Methods .....	20
2.2.1. Elastic Network Model .....	20
2.2.2. The Prediction of Binding Sites by Essential Site Scanning Analysis (ESSA)..	21
2.2.3. Molecular Dynamics .....	22
2.2.3.1. The Investigation of GALK1 Dynamics by MD Simulations .....	24
2.2.3.2. The Evaluation of Collective Motions by Principal Component Analysis (PCA) Machine Learning Method.....	25
2.2.4. ClustENMD (for GALK1).....	26
CHAPTER 3. RESULTS AND DISCUSSION.....	28
3.1. Allosteric and Cryptic Site Detection .....	28
3.1.1. ESSA Results .....	28
3.2. MD Results for GALK1.....	61
3.2.1. The RMSD Profiles of <i>h</i> GALK1 (alpha-carbon protein RMSD, substrate RMSD, fragment RMSD).....	61
3.2.2. The RMSF Profiles of <i>h</i> GALK1 (alpha-carbon protein).....	64
3.2.3. The Radius of Gyration ( $R_g$ ) Profiles of <i>h</i> GALK1 .....	66

3.2.4. Principal Component Analysis (PCA) Results .....	67
3.3. The Application of ESSA to The Selected MD Frames of <i>hGALK1</i> .....	71
3.4. The Application of ClustENMD to GALK1 .....	83
3.5. The Structural Comparison between the frames obtained from Molecular Dynamics (MD) and ClustENMD and the Crystal Structure (PDB ID: 6Q91) .....	93
CHAPTER 4. CONCLUSIONS AND RECOMMENDATIONS .....	100
REFERENCES .....	102
APPENDICES .....	
APPENDIX A.....	111

# LIST OF FIGURES

<b><u>Figure</u></b>	<b><u>Page</u></b>
Figure 1. The Leloir pathway. ....	6
Figure 2. The crystal structure of <i>hGALK1</i> (PDB ID: 6Q91). ....	10
Figure 3. The alignment of the holo form of TEM-1 beta-lactamase (PDB ID: 1PZO) and the apo form of TEM-1 beta-lactamase (PDB ID: 1ZG4). ....	18
Figure 4. The alignment of the holo form of CDK2 (PDB ID: 3PXZ) and the apo form of CDK2 (PDB ID: 4EK3). ....	19
Figure 5. The visualization of the representative connections in the elastic network model of the GALK1 enzyme. ....	20
Figure 6. The scheme for molecular dynamics (MD). ....	23
Figure 7. The workflow for ClustENMD. ....	27
Figure 8. Z-score distributions of residues for TEM-1 beta-lactamase ....	30
Figure 9. Z-score distributions of residues for CDK2 ....	31
Figure 10. Z-score variation in the holo TEM-1 beta-lactamase residues between different cutoff values (1PZO) ....	32
Figure 11. Z-score variation in the holo CDK2 residues between different cutoff values (3PXZ) ....	33
Figure 12. The ESSA results for the holo form of TEM-1 beta-lactamase (1PZO). ....	35
Figure 13. The ESSA results for the apo form of TEM-1 beta-lactamase (1ZG4). ....	38
Figure 14. The ESSA results for the holo form of the CDK2 (3PXZ). ....	40
Figure 15. The ESSA results for the apo form of CDK2 (4EK3). ....	42
Figure 16. The ESSA results in the presence of cutoff 7 Å averaged over 10 modes for 6Q91 ( <i>hGALK1</i> ). ....	47
Figure 17. The ESSA results in the presence of cutoff 7 Å combined 20 modes for 6Q91 ( <i>hGALK1</i> ). ....	49
Figure 18. The ESSA results in the presence of cutoff 10 Å averaged over 10 modes for 6Q91 ( <i>hGALK1</i> ). ....	50
Figure 19. The ESSA results in the presence of cutoff 10 Å combined 20 modes for 6Q91 ( <i>hGALK1</i> ). ....	51



<b><u>Figure</u></b>	<b><u>Page</u></b>
Figure 20. The ESSA results in the presence of cutoff 11 Å averaged over 10 modes for 6Q91 ( <i>hGALK1</i> ).....	52
Figure 21. The ESSA results in the presence of cutoff 11 Å combined 20 modes for 6Q91 ( <i>hGALK1</i> ).....	53
Figure 22. The ESSA results in the presence of cutoff 13 Å averaged over 10 modes for 6Q91 ( <i>hGALK1</i> ).....	54
Figure 23. The ESSA results in the presence of cutoff 13 Å combined 20 modes for 6Q91 ( <i>hGALK1</i> ).....	55
Figure 24. PASSer ensemble mode result for the orthosteric site in the holo form of GALK1 (PDB ID: 6Q91).....	57
Figure 25. PASSer ensemble mode result for the allosteric fragment e binding site in GALK1 (PDB ID: 6Q91).....	57
Figure 26. PocketMiner result for GALK1 (PDB ID: 6Q91).....	60
Figure 27. The RMSD profiles of <i>hGALK1</i> .....	62
Figure 28. The RMSD profiles of the ligands in the <i>hGALK1</i> structure .....	63
Figure 29. The RMSF profiles of <i>hGALK1</i> . .....	65
Figure 30. The radius of gyration ( $R_g$ ) profiles of <i>hGALK1</i> .....	67
Figure 31. The representation of principal component (PC) modes obtained from PCA analysis of <i>hGALK1</i> (front view).....	69
Figure 32. The representation of principal component (PC) modes obtained from PCA analysis of <i>hGALK1</i> .....	70
Figure 33. The ESSA results for frame 5143 ( <i>hGALK1</i> ).....	73
Figure 34. The ESSA results for frame 5198 ( <i>hGALK1</i> ).....	76
Figure 35. The ESSA results for frame 6700 ( <i>hGALK1</i> ).....	78
Figure 36. The ESSA results for frame 7116 ( <i>hGALK1</i> ).....	80
Figure 37. The Backbone RMSD Profile of the <i>hGALK1</i> protein (Overall). .....	83
Figure 38. The backbone RMSD profile of the residues around 5 Å of the fragment in <i>hGALK1</i> .....	84
Figure 39. The ESSA results for frame 176 ( <i>hGALK1</i> ).....	86
Figure 40. The ESSA results for frame 211 ( <i>hGALK1</i> ).....	89
Figure 41. The ESSA results for frame 212 ( <i>hGALK1</i> ).....	91
Figure 42. The alignment of ClustENMD frames to 6Q91 .....	95

<b><u>Figure</u></b>	<b><u>Page</u></b>
Figure 43. The demonstration of the fragment binding sites of <i>h</i> GALK1 in ClustENMD frames.....	95
Figure 44. The alignment of <i>h</i> GALK1 MD frames to 6Q91 .....	96
Figure 45. The demonstration of the fragment binding sites of <i>h</i> GALK1 in molecular dynamics (MD) frames .....	97
Figure 46. The backbone RMSD profile of GALK1_1.....	121
Figure 47. The backbone RMSD profile of GALK1_2.....	121
Figure 48. The backbone RMSD profile of GALK1_3.....	122
Figure 49. The backbone RMSD profile of GALK1_F_1.....	122
Figure 50. The backbone RMSD profile of GALK1_F_2.....	123
Figure 51. The backbone RMSD profile of GALK1_F_3.....	123

## LIST OF TABLES

<b><u>Table</u></b>	<b><u>Page</u></b>
Table 1. Some of the accessible GALK1 crystal structures and their bound ligands in the Protein Data Bank (PDB) .....	9
Table 2. The list and properties of structures used in the analysis .....	14
Table 3. The list and residues around 5 Å of orthosteric and allosteric ligands (Alphabetically ordered according to PDB IDs) .....	15
Table 4. The structures of ligands with ligand IDs (Atoms are colored according to element in PyMOL (C: purple, O: red, N: blue, S: green)) (Alphabetically ordered according to PDB IDs) .....	16
Table 5. The essential residues obtained from ESSA analysis for the holo form of TEM-1 beta-lactamase enzyme with the PDB ID of 1PZO .....	36
Table 6. The essential residues obtained from ESSA analysis for the apo form of TEM-1 beta-lactamase enzyme with the PDB ID of 1ZG4. ....	37
Table 7. The essential residues obtained from ESSA analysis for the holo form of CDK2 enzyme with the PDB ID of 3PXZ .....	39
Table 8. The essential residues obtained from ESSA analysis for the apo form of CDK2 enzyme with the PDB ID of 4EK3 .....	41
Table 9. The radius of gyration values of the studied structures. ....	43
Table 10. The essential residues obtained from ESSA analysis for the holo form of GALK1 enzyme with the PDB ID of 6Q91 .....	48
Table 11. The residues in the pockets found by PASSer.....	58
Table 12. A list of MD frames with high RMSD values with respect to the initial galactokinase 1 (GALK1) structure .....	71
Table 13. The essential residues obtained from ESSA analysis for the GALK1 molecular dynamics (MD) frame 5143.....	74
Table 14. The essential residues obtained from ESSA analysis for the GALK1 molecular dynamics (MD) frame 5198.....	75
Table 15. The essential residues obtained from ESSA analysis for the GALK1 molecular dynamics (MD) frame 6700 .....	77

<b><u>Table</u></b>	<b><u>Page</u></b>
Table 16. The essential residues obtained from ESSA analysis for the GALK1 molecular dynamics (MD) frame 7116. ....	79
Table 17. A list of <i>h</i> GALK1 frames obtained from ClustENMD with high RMSD values for the residues around 5 Å of the fragment.....	84
Table 18. ESSA result for ClustENMD frame 176 ( <i>h</i> GALK1).....	87
Table 19. ESSA result for ClustENMD frame 211 ( <i>h</i> GALK1).....	88
Table 20. ESSA result for ClustENMD frame 212 ( <i>h</i> GALK1).....	90
Table 21. The RMSD comparison of selected <i>h</i> GALK1 frames for ESSA .....	94
Table 22. The radius of gyration values of <i>h</i> GALK1 frames and its crystal structure..	98
Table 23. 1AXB-FOS PDBsum interaction.....	111
Table 24. 1PZO-CBT300 PDBsum interaction. ....	112
Table 25. 1PZO-CBT301 PDBsum interaction. ....	113
Table 26. 3PXZ-JWS PDBsum interaction. ....	114
Table 27. 3PXZ-2AN299 PDBsum interaction. ....	115
Table 28. 3PXZ-2AN300 PDBsum interaction. ....	116
Table 29. 6Q91-GAL PDBsum interaction. ....	117
Table 30. 6Q91-HFK PDBsum interaction.....	118
Table 31. 6Q91-HR8 PDBsum interaction .....	119
Table 32. Residues with RMSF above or equal to 0.2 nm .....	120

## ABBREVIATIONS

2AN:	8-anilino-1-naphthalene sulfonate
4QI:	(4R)-2-[(1,3-benzoxazol-2-yl)amino]-4-(4-chloro-1H-pyrazol-5-yl)- 4,6,7,8 tetrahydroquinazolin-5(1H)-one
ADP:	Adenosine-5'-Diphosphate
ANP:	Phosphoaminophosphonic acid-adenylate ester
CBT:	N,N-bis(4-chlorobenzyl)-1h-1,2,3,4-tetraazol-5-amine
FOS:	[[N-(benzyloxycarbonyl)amino]methyl]phosphate
GAL:	Beta-D-galactopyranose
HFK:	2-(1,3-benzoxazol-2-ylamino)spiro[1,6,7,8-tetrahydroquinazoline-4,1'- cyclohexane]-5-one
HGE:	Ethyl 1-pyrazin-2-ylpiperidine-4-carboxylate
HR5:	~{N}-(cyclobutylmethyl)-1,5-dimethyl-pyrazole-4-carboxamid
HR8:	5-chloranyl-2-methoxy-~{N}-(2-methylpropyl)benzamide
J4Q:	4-[(2-methylsulfonylimidazol-1-yl)methyl]-1,3-thiazole
JHJ:	N-(4-methoxyphenyl)-N'-pyridin-4-ylurea
JWS:	2-(4,6-diamino-1,3,5-triazin-2-yl)-4-methoxyphenol
QKZ:	2-(4-chlorophenyl)-~{N}-pyrimidin-2-yl-ethanamide
WNP:	N-(3-chlorophenyl)-2,2,2-trifluoroacetamide

# CHAPTER 1

## INTRODUCTION

### 1.1. Allostery

Allosteric regulation is the change in active site dynamics in the presence of ligand binding to a region far away from the protein's active site (orthosteric site) (Other factors can also trigger this change). To this extent, the ligand binding may have an effector or inhibitor role in the protein's dynamics. In this scope, the binding of the substrate can be inhibited with the inhibitory role of the allosteric ligand. Opening/closing of the active site, large or small conformational changes, or some variation in the electrostatic properties of the structure were reported to occur due to allosteric regulation (Laskowski, Gerick, and Thornton 2009). The first drug that demonstrated an allosteric effect was confirmed in 2004 by the United States of America Food and Drug Administration ("Drug Approval Package: Sensipar (Cinacalcet HCl) NDA #021688" 2024). In this scope, more selective and effective drugs for a target can be designed when the allostery is better understood. In addition, allosteric sites can be preferred as targets instead of orthosteric regions due to their potential for being species-specific from an evolutionary perspective (Laskowski, Gerick, and Thornton 2009). Also, the determination of key amino acids in the allosteric pathway is significant to understanding the protein's relation from structure to function (Xie et al. 2023).

#### 1.1.1. The Detection of Allosteric Sites by Computational Methods

Large or small conformational changes can be clearly observed in some of the proteins as a result of allosteric binding when apo and holo (bound/complex) crystal

structures are compared, but these changes are not visible in some crystal structures (Laskowski, Gerick, and Thornton 2009). Protein dynamics should be considered while investigating allostery in the scope of conformational changes. In this scope, the review by Nerín-Fonz and Cournia provides insights into the latest allosteric site detection tools (Nerín-Fonz and Cournia 2024). These methods are significant for the development of drug candidates that target allosteric sites for multiple diseases in the scope of proteins' function (Abdel-Magid 2015; Nussinov and Tsai 2013).

Notably, ALLO, Allosite, AlloPred, and PASSer were reported as the most successful tools in estimating allosteric regions using machine learning (W. Huang et al. 2013; Greener and Sternberg 2015; Tian et al. 2023; Akbar and Helms 2018). Among these servers, the highest accuracy (97.4) was reported for PASSer (Nerín-Fonz and Cournia 2024).

PASSer provides three different models for the prediction of the allosteric regions. Among these models, ensemble learning and learning-to-rank models give the output in less amount of time (a few seconds) compared to the other model called automated machine learning (Tian et al. 2023). In addition, Fpocket, a pocket detection tool, is also included in the PASSer, and used to extract the properties of the pockets that are used in the training (Le Guilloux, Schmidtke, and Tuffery 2009). On the other hand, GRaSP also estimates ligand binding sites using machine learning. However, the server demonstrates variation from other indicated servers since it uses a graph-based approach. To this extent, interacting residues are determined for each amino acid in the structure by calculating several properties such as solvent relative accessibility. After the determination of the environment around residues (which can be denoted as the first and second shells) and properties, feature vectors are formed for all residues in the structure (Santana et al. 2020).

These above methods do not consider any information related to protein dynamics. However, the examination of the movements of atoms and how they are correlated are significant for comprehending the biological function (Ichiye and Karplus 1991), therefore, the conformational landscape of the protein. The Elastic Network Model (ENM) investigates collective motions in proteins using single structure information and samples the conformational landscape (Haliloglu, Bahar, and Erman 1997). The collective motion is the motion of residues within different domains of a protein toward specific directions together (Ichiye and Karplus 1991). Essential Site Scanning Analysis (ESSA) is an ENM-based method that calculates the significant essential residues that might cause significant changes in motions (modes) during the protein dynamics when

the structure is hypothetically bound to a ligand. ESSA is used for the prediction of allosteric/cryptic and orthosteric sites. With this method, the presence of ligands is mimicked by considering the side chain of each amino acid in the protein one by one, and the ligand binding sites can be predicted in a short period (less than a few minutes for a protein with a mid-size) (Kaynak, Bahar, and Doruker 2020). A single crystal or modeled structure is used as input in ESSA. No ligand information or training/testing procedure is needed.

## **1.2. Cryptic Pockets**

In protein structures, ligand binding sites can be predicted using various computational methods. However, the regions called “cryptic pockets” may not be observed in the apo form of a protein and can be only observed in the holo (ligand-bound) form (Oleinikovas et al. 2016). However, it was indicated that these pockets can also be formed in the absence of ligand-binding (Sun et al. 2020). The identification of cryptic pockets was considered significant since several proteins were thought to be undruggable in the past, which limited the drug discovery studies (Oleinikovas et al. 2016).

According to Kuzmanic et al., the research about cryptic pockets continues to elucidate how cryptic pockets are formed and what types of motions dominate the appearance of cryptic pockets. The findings were reported to demonstrate variation between protein structures (Kuzmanic et al. 2020). One of the investigations was related to the formation of cryptic pockets, and three possible types were reported, which were conformational selection, induced fit, and the mixture of these two types (Oleinikovas et al. 2016). To this extent, the presence of a ligand triggers conformational changes in the protein structure and results in the formation of a cryptic pocket. For instance, TEM-1 beta-lactamase was indicated as one of the protein structures that require the presence of a ligand to observe the cryptic pocket. However, this cryptic pocket can also be observed through conformational selection in TEM-1 beta-lactamase (Kuzmanic et al. 2020).



### **1.2.1. The Investigation of Cryptic Pockets Using Computational Methods**

With the increase in computational power and the development of artificial intelligence studies, computational approaches were found to be promising for identifying cryptic sites. In this scope, various computational tools/methodologies were developed for the prediction of cryptic pockets, including PocketMiner, CryptoSite, SWISH, CrypticScout, and FTMove (Meller et al. 2023; Cimermancic et al. 2016; Oleinikovas et al. 2016; Martinez-Rosell et al. 2020; Egbert et al. 2022).

Recently, Meller et al. reported PocketMiner, which uses a neural network to estimate cryptic pockets in protein structures. One of the advantages of this method was indicated as being time efficient and the usage of only one conformation as input structure (Meller et al. 2023). On the other hand, CryptoSite, which was developed by Cimermancic et al., includes the application of machine learning to differentiate residues from each other using properties from known cryptic pockets. In this scope, the cryptic pockets in TEM-1 beta-lactamase (apo form) can be detected with 60% accuracy by CryptoSite (Cimermancic et al. 2016).

Cryptic pockets can also be predicted through molecular dynamics (MD) simulations. However, the detection of cryptic sites requires long MD simulations, which is not advantageous (Cimermancic et al. 2016). Advantageous approaches for the prediction of cryptic pockets include the development of collective variables or the usage of enhanced sampling approaches that are independent of collective variables (Oleinikovas et al. 2016). In this scope, Oleinikovas et al. reported the method called Sampling Water Interfaces through Scaled Hamiltonians (SWISH), which detects cryptic pockets using Hamiltonian Replica Exchange. In this approach, interactions were considered for scaling to provide the opening of the cryptic site. The presence of water molecules was considered to mimic ligands, and the properties of water changed to this extent. The solvent-carbon and sulfur atoms of the protein were considered for interactions in this method. In addition, the method combined with the inclusion of fragments in the simulation, and the results were assessed through occupancy of the fragments. The combinatory approach was reported to increase the success of the cryptic pocket prediction (Oleinikovas et al. 2016).

Recently, ESSA was also reported to be successful in predicting cryptic pockets for 4 protein structures. However, pocket prediction information through Fpocket was included in the ESSA analysis. In this thesis, ESSA alone (no information from Fpocket) was applied to the selected structures having cryptic pockets, and the results were compared to those of PASSer, PocketMiner, and GRaSP.

### **1.3. The Investigation of Protein Dynamics Using Computational Methods**

*In silico* approaches are significant in the research ecosystem in terms of decreasing the cost for drug discovery and accelerating the process by predicting potential outcomes (Shaker et al. 2021). To this extent, MD simulations provide information about how protein dynamics change over a time scale using Newton's Law of Motion (Hollingsworth and Dror, 2018). The investigation of conformational changes in a protein and the variation in the protein dynamics in the presence and absence of a ligand on a time scale are significant for drug discovery studies. The effect of ligand binding on protein dynamics, ligand stability, and allostery can be investigated by MD simulations. More detailed information and application areas of MD simulations can be found in the review written by Hollingsworth and Dror (Hollingsworth and Dror, 2018). On the other hand, cryptic pockets can also be explored by performing MD simulations (Oleinikovas et al. 2016). In addition, coarse-grained methods are alternatives to MD simulations and can be used to explore conformational landscapes of proteins. ClustENMD, which is an ENM-based conformational sampling method (see Section 2.2.9 for more detail), generates all-atom structures of a protein in a very short time (a few minutes to hours depending on the size of the protein) by applying collective mode deformations to the structure (Kaynak et al. 2021).

In the scope of this thesis, galactokinase 1 (GALK1), a significant enzyme for classic galactosemia, was selected as a case study to investigate the protein dynamics and allostery by performing classical MD simulations and ClustENMD and predict allosteric and cryptic pockets in GALK1 using the ESSA method.

## 1.4. Galactosemia and Classical Galactosemia

A rare disease is defined as a disease that is observed in less than five people out of 10,000, according to Europe (Gimenez-Lozano et al. 2022). These values demonstrate variation between Europe and the United States of America (USA) (Danese and Lippi 2018). In the literature, approximately 7000 rare diseases were defined (Gimenez-Lozano et al. 2022). Most of the described rare diseases are genetics-related, and more than 300 million were diagnosed with a rare disease, according to the report (Editorial 2022).

Galactosemia is a rare disease related to disruptions in the galactose metabolism and is observed in approximately 1 in 40,000 people globally. In Türkiye, it is detected in 1 in 23,775 people (Atik et al. 2016).

In the galactose metabolism, the Leloir pathway is very significant in terms of including the conversion of galactose to glucose-1-phosphate as well as being used as an energy source. Leloir pathway includes galactose mutarotase (GALM), galactokinase 1 (GALK1), galactose-1-phosphate-uridyl-transferase (GALT), and uridine diphosphate-galactose-4' epimerase (GALE) enzymes (McCorvie and Timson 2020). The Leloir pathway and the enzymes involved in this pathway were demonstrated in Figure 1.

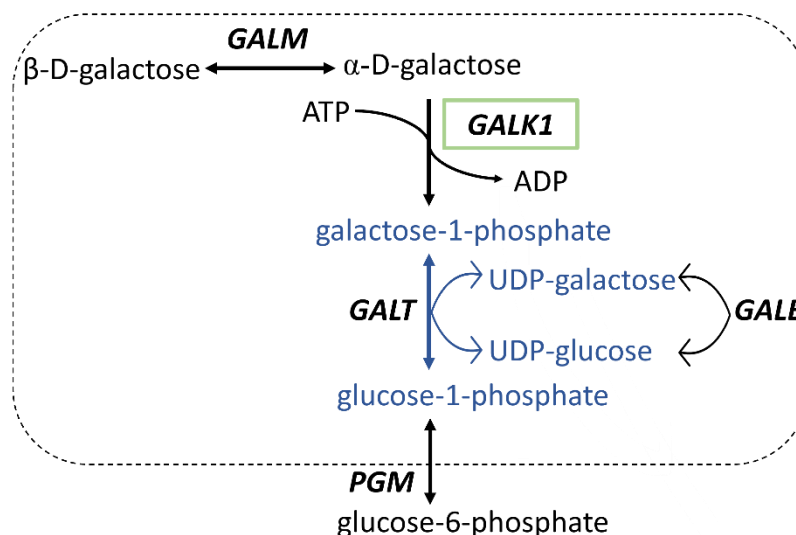


Figure 1. The Leloir pathway. The figure was adapted from (Mackinnon et al. 2019).

Galactosemia is categorized into four categories: type 1, 2, 3, and 4 galactosemia according to the molecular properties. The variation in the categories is based on the presence of different disease-related genes in the galactose metabolism. For instance, type 1 galactosemia (classic galactosemia) is related to the mutations in the GALT gene, while type 2 galactosemia is related to the GALK1 variants. On the other hand, type 3 and 4 galactosemia occur due to mutations in the GALE and GALM genes, respectively (Succoio et al. 2022). The most common type of galactosemia is related to GALT. To this extent, Q188R, K285N, and L195P mutations were reported for GALT. The prevalence of Q188R mutation was higher compared to the other mutations in patients with classic galactosemia (Atik et al. 2016). While mutations in the GALT gene decrease the enzyme activity, this results in the accumulation of galactose-1-phosphate. The accumulation of galactose-1-phosphate damages the liver, kidney, and brain. In addition, it triggers the formation of toxic metabolites (galactitol and D-galactonate) by causing the activation of the other pathways (Succoio et al. 2022). For instance, the increase in galactitol causes cataracts (Rubio-Gozalbo et al. 2020).

The effects of galactosemia are primarily observed in the first months of the newborn due to the consumption of milk (Mackinnon et al. 2021). According to the studies in Türkiye, the most common symptom was indicated as jaundice. In addition, classic galactosemia results in cataracts, hypotonia, and sepsis. The current approaches are based on the prevention of consuming lactose-rich food. However, these approaches were not observed as adequate due to the disruption of speech and motor functions. Also, the effects on the reproductive system are involved in the concerns about classic galactosemia (Garcia et al. 2016). For instance, hypergonadotropic hypogonadism and early ovarian insufficiency were observed in 85% of the classic galactosemia patients (Lai et al. 2008). Early ovarian insufficiency results in infertility and autoimmune diseases (Jankowska 2017). On the other hand, the potential effects of galactosemia on skin aging were reported due to the role of D-galactose (Umbayev et al. 2020).

The following part focuses on the structure and allosteric sites of galactokinase 1 (GALK1) enzyme, which is significant in the Leloir pathway in terms of the prevention of galactose-1-phosphate accumulation.

### 1.4.1. Galactokinase 1 (GALK1)

GALK1 enzyme is involved in the transfer of the phosphate of adenosine triphosphate (ATP) to alpha-D-galactose as a catalyst, and adenosine diphosphate (ADP) and alpha-D-galactose-1-phosphate forms with the transfer. Galactokinase is included in the GHMP superfamily. GHMP indicates galactokinase, homoserine kinase, mevalonate kinase, and phospho mevalonate kinase (Rubio-Gozalbo et al. 2020). Pro-X-X-X-Gly-Leu-X-Ser-Ser-Ala motif, which was significant for the nucleotide binding, was observed to be conserved in this superfamily (Hu et al. 2019).

According to the BRENDA database, GALK1 is classified as EC 2.7.1.6 according to the enzyme classification. This classification indicates transferases, the transfer of phosphor-including groups, acceptor phosphotransferases with alcohol groups, and galactokinases, respectively (Chang et al. 2021). Even though classic galactosemia occurs due to mutations in the *GALT* gene, one of the therapeutic approaches involves the inhibition of GALK1 to prevent the accumulation of galactose-1-phosphate (Delnoy, Coelho, and Rubio-Gozalbo 2021). For this reason, the detailed examination of GALK1 dynamics has significant potential to lead to further drug discovery studies. The single mutation in GALK1 can decrease the enzyme activity, and mutations may demonstrate variations from population to population. For instance, the A198V variant was mostly observed in Japan and Korea (Rubio-Gozalbo et al. 2020). On the other hand, P28T and A198V mutations were reported to trigger cataracts (Thoden et al. 2005).

Twenty crystal structures belonging to the human GALK1 were reported and solved by X-ray diffraction in the UniProt. The resolution of these structures varies from 1.86 to 2.50 Å (Bateman et al. 2023). Some of the accessible crystal structures of GALK1 were demonstrated in Table 1, and the positions of ligands on the structure were given in Figure 2.

Table 1. Some of the accessible GALK1 crystal structures and their bound ligands in the Protein Data Bank (PDB). The table is alphabetically ordered according to PDB IDs (\*HFK: It was found in the ATP binding site and the position of HFK was not shown in the table. Also, the substrate was only indicated for the structure with the PDB ID of 6Q91 in the table).

<b>PDB ID</b>	<b>Ligand ID</b>	<b>Ligand Position</b>	<b>References</b>
1WUU	ANP	ATP binding site	(Thoden et al. 2005)
6Q3W	HGE and HFK*	Active site entrance	(Bezerra et al. 2019)
6Q8Z	HR5 and HFK*	C-terminal groove	(Bezerra et al. 2019)
6Q90	JHJ and HFK*	Binding hotspot	(Bezerra et al. 2019)
6Q91	HR8, GAL, and HFK*	Binding hotspot, Active site	(Bezerra et al. 2019)
6QJE	J4Q and HFK*	Active site entrance	(Bezerra et al. 2019)
6ZGV	QKZ and HFK*	Binding hotspot	(Mackinnon et al. 2021)
6ZH0	WNP and HFK*	Binding hotspot	(Mackinnon et al. 2021)
7RCL	ADP	ATP binding site	(Liu et al. 2021)
7RCM	ADP	ATP binding site	(Liu et al. 2021)
7S49	4QI	ATP binding site	(Liu et al. 2021)

The *h*GALK1 structure with a PDB ID of 6Q91 is composed of 392 amino acids and has a molecular weight of 42 kiloDalton (kDa). It consists of 12.8% Ala and %12.0 Leu amino acids according to the residue composition of GALK1. It contains 12 negatively-charged amino acids in total, while this value is 37 for positively-charged amino acids. In addition, the aliphatic index was calculated as 91.40 according to the ExPASy ProtParam tool (Gasteiger et al. 2005). In Figure 2, the GALK1 structure with a

PDB ID of 6Q91 was demonstrated, which involves bound-new candidate GALK1 inhibitor (Ligand ID: HR8, 5-chloranyl-2-methoxy-~{N}-(2-methylpropyl) benzamide ( $C_{12}H_{16}ClNO_2$ )) that was discovered by the Yue Laboratory at Newcastle University through fragment screening (Mackinnon et al. 2021; Bezerra et al. 2019).

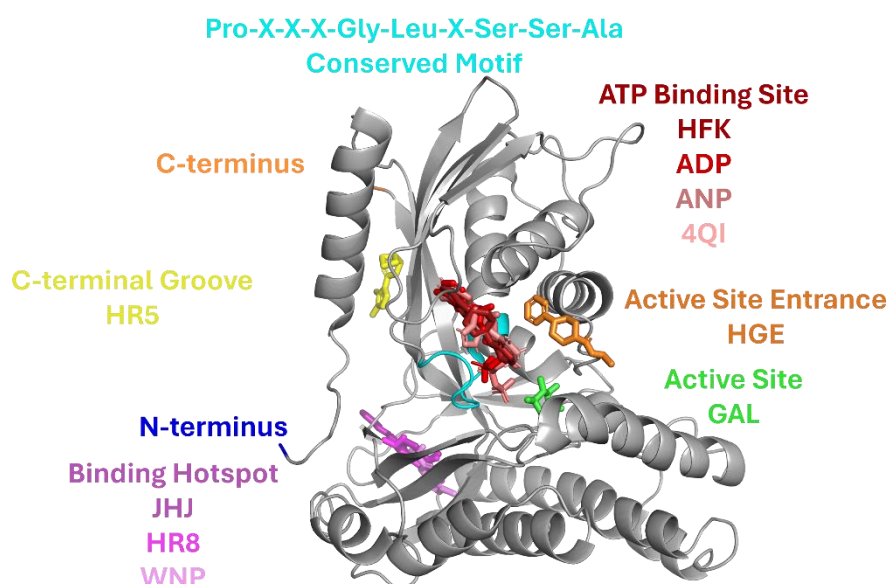


Figure 2. The crystal structure of *h*GALK1 (PDB ID: 6Q91). Important sites and codes of ligands that are bound to the other GALK1 structures were denoted on the structure (GALK1: Gray, N-terminus: Dark blue, C-terminus: Orange, Conserved Motif: Pro-X-X-X-Gly-Leu-X-Ser-Ser-Ala: Cyan, G136, G138, S140, S141: ATP binding site).

#### 1.4.2. The Investigation of *h*GALK1 Dynamics Using Computational Methods

In the literature, McAuley, Huang, and Timson (2017) investigated the roles of GALK1 residues Arg105, Glu174, and Arg228 by performing MD simulations. In the scope of the study, the necessity of the negative charge of Glu174, the interactions of Arg228 with the substrate, and the importance of Arg105 for ATP binding were detected.

Also, the effect of mutating residues around the active site and far away from the active site on GALK1 dynamics was investigated. To this extent, Met60Val mutation was observed to change the dynamics in the active site, and Arg366Gln mutation was observed to make the Glu257 and Val258, including the loop region, more flexible (McAuley et al. 2018). On the other hand, virtual screening studies provide the discovery and development of drug candidates. In this scope, Hu et al. determined the GALK1 inhibitors that bind to the ATP binding site by performing structure-based virtual screening (SB-VS) (Hu et al. 2019). In addition, Sneha et al. performed 50 nanoseconds (ns) MD simulations to investigate the effects of mutations on GALK1 dynamics and applied Principal Component Analysis (PCA) (Sneha et al. 2018). On the other hand, Huang et al. examined the phosphorylation reaction mechanism of GALK1 by MD simulations and molecular mechanics calculations (Huang et al. 2013). In addition, Chiappori et al. analyzed how GALK1 interacts with the four inhibitors that bind to the ATP binding site as well as the interaction of the substrate with GALK1 through MD simulations (Chiappori et al. 2013). However, it was observed that the GALK1 dynamics in the presence of allosteric inhibitors has not been studied computationally, yet.

## **1.5. Aim of The Study**

The prediction of ligand binding sites in protein structures is significant for drug discovery. In this scope, Fpocket is a tool widely used for ligand binding site prediction and applied with other tools to increase binding site prediction success. For instance, the integration of Fpocket with ESSA was reported to increase the prediction of allosteric sites in some proteins compared to performing ESSA without including Fpocket. In this scope, the allosteric site prediction success was aimed to be increased using ESSA independent from Fpocket. The allosteric/cryptic and orthosteric binding site prediction success was aimed to be improved by changing cutoff values in GNM calculations and applying different number of modes. To this extent, two structures which are TEM-1 beta-lactamase and cell division protein kinase 2 (CDK2) were investigated due to involving cryptic sites. On the other hand, human galactokinase 1 (*hGALK1*) as selected as a case



study due to its significance in classic galactosemia as a target, and the ineffectiveness of default ESSA parameters in *hGALK1* allosteric site prediction.

Therefore, the improved ESSA method was developed and tested to predict the orthosteric, allosteric, and cryptic binding sites of TEM-1 beta-lactamase, cell division protein kinase 2, and *hGALK1* in this thesis (see Chapter 2 for the studied crystal structures). To this extent, the effects of various cutoff values and different modes on the prediction of allosteric/cryptic and orthosteric sites were investigated, and the success of ligand binding site prediction was aimed to be improved for the ESSA method. In addition, the allostery and effect of a known allosteric inhibitor on *hGALK1* were aimed to be investigated using computational methods. In this scope, the effect of a newly developed candidate *hGALK1* inhibitor (Ligand ID: HR8, 5-chloranyl-2-methoxy- $\sim$ {N}-(2-methylpropyl) benzamide (C<sub>12</sub> H<sub>16</sub>ClNO<sub>2</sub>)), which was developed by Prof. Dr. Wyatt W. Yue's laboratory at Newcastle University, on *hGALK1* dynamics was examined by classical MD simulations and the Principal Component Analysis (PCA) machine learning method. PCA was applied to analyze the collective motions of both the inhibitor-bound and inhibitor-removed forms of *hGALK1*. On the other hand, the conformational changes in *hGALK1* were also explored by applying ClustENMD, which provides unbiased sampling using slow modes and short MD simulations. The obtained various conformations from both MD simulations and ClustENMD were used to predict ligand binding sites of *hGALK1* computationally using ESSA. Also, even though alternative approaches are known for the treatment of galactosemia, there is no treatment for classic galactosemia that is effective in the long term. Current approaches are limited to the arrangement of diet to prevent the accumulation of galactose-1-phosphate (Succoio et al. 2022). In this scope, the understanding of *hGALK1* dynamics and allosteric effect is thought to contribute to developing potential drug candidates for classic galactosemia.

## CHAPTER 2

### MATERIALS AND METHODS

#### 2.1. Materials

In the first part of the thesis, the investigated proteins for allosteric, orthosteric, and cryptic pocket prediction are TEM-1 beta-lactamase (PDB ID: 1PZO) and cell division protein kinase 2 (PDB ID: 3PXZ) (Horn and Shoichet 2004; Betzi et al. 2011). On the other hand, apo forms of the proteins that are TEM-1 beta-lactamase (PDB ID: 1ZG4), and CDK2 (PDB ID: 4EK3) were also used in the analysis (Stec et al. 2005). The proteins were determined according to the presence of prior knowledge on the bound allosteric and orthosteric ligands and the good resolution of the structures (1.34 -1.90 Å). In addition to their better resolution values and prior knowledge of bound ligands, the structures were selected according to the preference of monomer as global stoichiometry. On the other hand, these structures were selected by considering the structures with allosteric sites that are cryptic in the article reported by Kaynak et al. which is about the ESSA method (Kaynak et al. 2020).

In the second part of the thesis, the *hGALK1* structure with the PDB ID of 6Q91 was analyzed as a case study in terms of binding site prediction and dynamics. The information about the protein structures, the residues around 5 Å of ligands, and the structures of ligands were demonstrated in Table 2 – Table 4, respectively, and explanations about the structures were given in the following part.

Table 2. The list and properties of structures used in the analysis. All structures are monomers (Alphabetically ordered according to PDB IDs).

Protein and PDB ID	Resolution (Å)	Total # of Residues	Ligand IDs and positions		Mutations	References
			Orthostheric	Allosteric		
TEM-1 beta-lactamase (holo form) 1PZO	1.90	263	-	CBT (A300, A301)	M182T	(Horn and Shoichet 2004)
TEM-1 beta-lactamase (apo form) 1ZG4	1.55	286	-	-	V84I and A184V	(Stec et al. 2005)
Cell division protein kinase 2 (holo form) 3PXZ	1.70	306	JWS (A301)	2AN (A299, A300)	-	(Betzi et al. 2011)
Cell division protein kinase 2 (apo form) 4EK3	1.34	299	-	-	Acetylation is present in the first position (Not included)	To be published (Kang and Stuckey, n.d.)
GALK1 (holo form) 6Q91	2.40	392	GAL	HR8	K252A, E253A	(Mackinnon et al. 2021)

Table 3. The list and residues around 5 Å of orthosteric and allosteric ligands (Alphabetically ordered according to PDB IDs).



PDB ID	Ligand ID (Orthosteric/Allosteric) and Ligand position		The Minimum Distance Between Allosteric and Orthosteric Ligand (Å)	The Residues Around 5 Å of Orthosteric Ligand	The Residues Around 5 Å of Allosteric Ligand
	Orthosteric	Allosteric			
1PZO	FOS (The information was obtained using the PDB ID of 1AXB which is the other structure of TEM-1 beta-lactamase in orthosteric ligand-bound form.	CBT (A300, A301)	FOS`291/O1P-CBT`300/N20: 4.3  FOS`291/O1P-CBT`301/CL3: 12.1	M69, S70, K73, S130, N132, E166, P167, N170, K234, S235, G236, A237, G238, E240	<b>A300:</b> S130, V216, A217, L220, S235, G236, A237, R244, G245, I246, I263, N276, I279; <b>A301:</b> A217, L220, L221, A224, L225, R244, L250, V261, I263, N276, I279, A280, G283, A284, L286
3PXZ	JWS (A301)	2AN (A299, A300)	JWS`301/C17-2AN`300/O3: 2.9. 2AN`299/C6-JWS`301/C17: 10.2	I10, G11, E12, V18, A31, K33, V64, F80, E81, F82, L83, H84, Q85, D86, Q131, L134, A144, D145	<b>A299:</b> I35, L37, V44, I49, I52, K56, V69, H71, L76, L78, V154; <b>A300:</b> Y15, K33, I35, I52, L55, K56, I63, V64, L66, L78, F80, A144, D145, F146, L148, A149, V154, V156

(cont. on next page)

**Table 3. (cont.).**

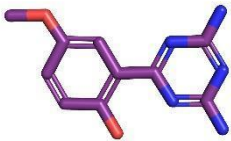
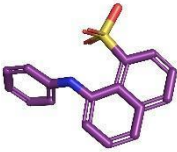
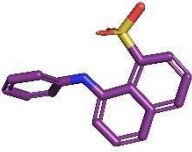

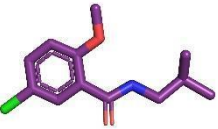
6Q91	GAL	HR8	HR8`403/C1- GAL`401/O6: 16.2	R37, N39, G42, E43, H44, T45, D46, Y47, M180, C182, G183, I184, M185, D186, Y236, G345, G346	L40, M55, L213, S214, D215, L218, A219, V220, A294, L295, G298, D299, Y300, F303, L355, S381, Q382, A383
------	-----	-----	---------------------------------	---	--

Table 4. The structures of ligands with ligand IDs (Atoms are colored according to element in PyMOL: C in purple, O in red, N in blue, S in yellow, and Cl in green). The ligand structure figures were prepared using PyMOL.

PDB IDs	Ligand IDs		Ligand Structures		
	Orthosteric	Allosteric	Orthosteric	Allosteric	
1PZO	-	CBT	-	CBT (A300)	CBT (A301)
					

**(cont. on next page)**

**Table 4. (cont.).**

				2AN (A299)	2AN (A300)
3PXZ	JWS	2AN			
6Q91	GAL	HR8			

TEM-1 beta-lactamase was reported as a hydrolase to have a role in resistance to penicillin which is an antibiotic that includes a beta-lactam ring in its structure (Salverda, de Visser, and Barlow 2010). Beta-lactamases are categorized into four groups including classes A, B, C, and D. TEM-1 beta-lactamase was indicated to be in class A. The residues E166 and N170 were considered to be significant for the catalysis and located in the active site of the enzyme (Brown et al. 2009). The structure of TEM-1 beta-lactamase with the PDB ID of 1PZO involves a bound-allosteric inhibitor and the global stoichiometry of the structure is a monomer (Horn and Shoichet 2004). On the other hand, the apo form of the TEM-1 beta-lactamase has a PDB ID of 1ZG4. The RMSD value between the structures with the PDB IDs of 1PZO and 1ZG4 was calculated as 1.2 Å. The alignment of 1PZO and 1ZG4 was demonstrated in Figure 3. All structure figures in the thesis were prepared using PyMOL.

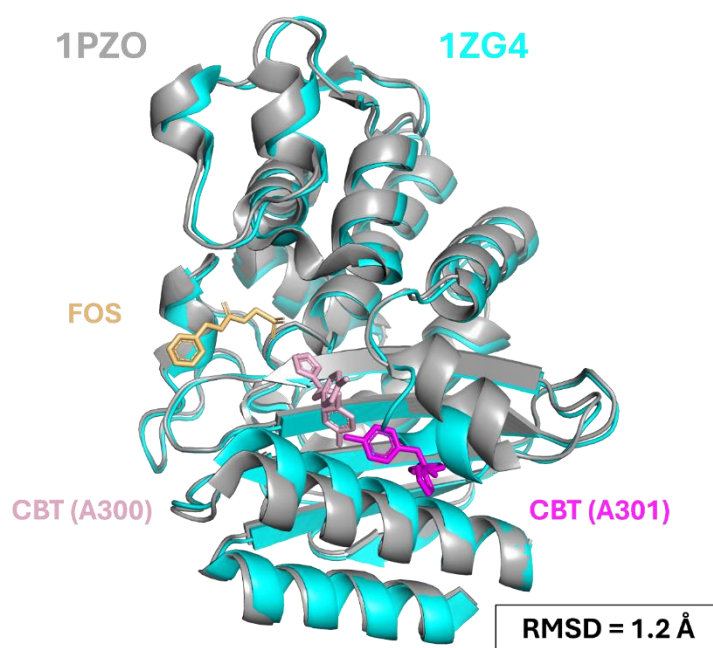


Figure 3. The alignment of the holo form of TEM-1 beta-lactamase (PDB ID: 1PZO) and apo form of TEM-1 beta-lactamase (PDB ID: 1ZG4). The proteins 1PZO (gray) and 1ZG4 (cyan) were shown in cartoon representation. The substrate (FOS), allosteric ligands CBT (A300) and CBT (A301) were demonstrated in light orange, light pink, and magenta respectively (FOS: [[N-(benzyloxycarbonyl)amino]methyl]phosphate, CBT: N,N-bis(4-chlorobenzyl)-1h-1,2,3,4-tetraazol-5-amine).

Cyclin-dependent kinase (CDK) was indicated as a significant enzyme for cell cycle progression. The interaction with cyclin is required for the activity of the enzyme. In this scope, several inhibitors were reported for cyclin-dependent kinase. However, most of the inhibitors were reported to bind to the ATP binding site and the discovery of allosteric sites was indicated to be significant for target selectivity (Betzi et al. 2011). The structure with the PDB ID of 3PXZ includes CDK2 in complex with an allosteric ligand named 2AN which was used as a probe in fluorescence spectroscopy to explore binding sites. On the other hand, the apo form of the enzyme has a PDB ID of 4EK3. The RMSD value between the structures with the PDB ID of 3PXZ and 4EK3 was calculated as 2.7 Å. The alignment of 3PXZ and 4EK3 was demonstrated in Figure 4.

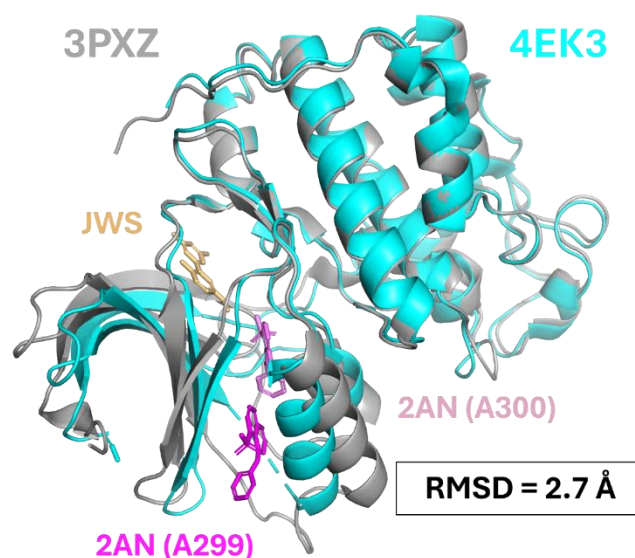


Figure 4. The alignment of the holo form of CDK2 (PDB ID: 3PXZ) and the apo form of CDK2 (PDB ID: 4EK3). The proteins 3PXZ (gray) and 4EK3 (cyan) were shown in cartoon representation. The substrate (JWS), and allosteric ligands 2AN (A300) and 2AN (A299) were demonstrated in light orange, light pink, and magenta respectively (JWS: 2-(4,6-diamino-1,3,5-triazin-2-yl)-4-methoxyphenol, 2AN: 8-anilino-1-naphthalene sulfonate).

The structure with the PDB ID of 6Q91 belongs to the galactokinase 1 enzyme. The structure includes bound-inhibitor (5-Chloro-N-isobutyl-2-methoxybenzamide), ATP competitive inhibitor (2-(1,3-benzoxazol-2-ylamino)spiro[1,6,7,8-tetrahydroquinazoline-4,1'-cyclohexane]-5-one), and beta-D-galactopyranose. The global stoichiometry of the structure was reported as monomer (Mackinnon et al. 2021). The enzyme is involved in the Leloir pathway which is significant for galactose metabolism and the inhibition of the GALK1 is promising for the prevention of the disease called classic galactosemia (Mackinnon et al. 2021). On the other hand, the interactions of the ligands with the corresponding protein structures were analyzed in the PDBsum and demonstrated in Appendix A (Table 23 - Table 31) (Laskowski et al. 2018).



## 2.2. Methods

### 2.2.1. Elastic Network Model

The elastic network model (ENM) is a coarse-grained approach to model protein dynamics characteristics. In the scope of the ENM, the connectivity of carbon-alpha atoms of the protein is determined in a certain cutoff value with the harmonic spring that has the same force constant. Harmonic mode shapes (eigenvectors) and frequencies (eigenvalues) that provide an understanding of the collective motions of proteins are obtained by solving potential energy function for the protein system (Haliloglu, Bahar, and Erman 1997). To this extent, representative elastic network model connections for the GALK1 protein were demonstrated in Figure 5.

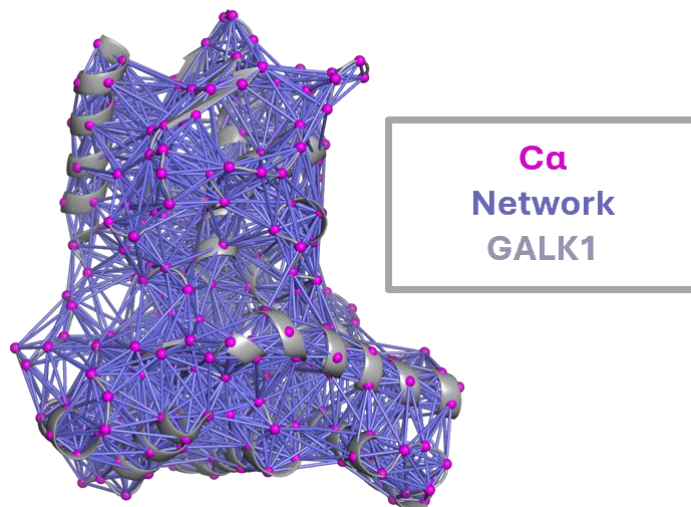


Figure 5. The visualization of the representative connections in the elastic network model of the GALK1 enzyme.

Elastic network models are composed of two categories including the Gaussian Network Model (GNM) and the Anisotropic Network Model (ANM). Haliloglu, Bahar, and Erman determined that residues demonstrate fluctuations with Gaussian distributions

in the proposed Gaussian Network Model (Haliloglu, Bahar, and Erman 1997). The potential function ( $V$ ) in GNM was demonstrated in equation (1), where  $\gamma$  is the spring constant between nodes  $i$  and  $j$ ,  $h(x)$  is the Heaviside step function,  $R_c$  is cutoff (10 Å),  $R_{ij}$  is the instantaneous distance vector, and  $\Delta R$  is the fluctuation in the position vector of residue.

$$V_{GNM} = \frac{\gamma}{2} \left[ \sum_i \sum_{j,j>i} h(R_c - R_{ij}) [\Delta R_j - \Delta R_i]^2 \right] \quad (1)$$

GNM only gives frequency information. In addition, GNM provides  $N-1$  modes, and  $N$  denotes the number of nodes or residues in the system (Doruker, Atilgan, and Bahar 2000).

### **2.2.2. The Prediction of Binding Sites by Essential Site Scanning Analysis (ESSA)**

Kaynak, Bahar, and Doruker reported a new method called Essential Site Scanning Analysis (ESSA), which is an elastic network model-based approach and open-source program embedded in ProDy (Kaynak, Bahar, and Doruker 2020; Bakan, Meireles, and Bahar 2011). The allosteric and orthosteric sites can be predicted using ESSA. This method uses GNM as the default for the mode extraction (or generation). In this scope, in addition to alpha-carbons, the side chain of each amino acid is involved in the ESSA calculation while solving the potential energy function for the protein system. The shifts in the frequencies are found by calculating the difference between cases when both alpha-carbons and side chains are considered and when only alpha-carbons are used in ENM. For instance, a protein with 100 residues containing 100 alpha-carbons would yield 99 different modes in the GNM. Among these modes, the first 10 modes were considered as the reference frequency because they are the most relevant modes to the function of the proteins. On the other hand, when side chains of each residue were added until 100<sup>th</sup> carbon alpha one by one in separate groups,  $N-1$  modes were also generated for each group, and  $N$  varies depending on the number of atoms in the side chain which

would give 100 groups in total. Among these modes, the modes that overlap with the first 10 modes of the reference frequency were taken by calculating mode overlap. In equation (2), the percent shift in the frequency ( $\Delta\lambda_k^{(i)}$  (%)) is calculated by considering the difference between frequency in the presence of side chain as well as carbon alpha ( $\lambda_k^{(i)}$ ) and the frequency in the presence of only carbon alpha ( $\lambda_k$ ). The overall shift value for the average of ten modes is converted into a z-score value as given in equation (3), where  $\mu$  is the mean and  $\sigma$  is the standard deviation.

$$\Delta\lambda_k^{(i)}(\%) = \frac{\lambda_k^{(i)} - \lambda_k}{\lambda_k} \times 100 \quad (2)$$

$$z_i = \frac{\langle \Delta\lambda_{1-10}^{(i)} \rangle - \mu}{\sigma} \quad (3)$$

In this scope, the observed regions with higher shifts emphasize the most probable regions for ligand binding. To this extent, the residues with higher z-scores were demonstrated in red, while residues with lower z-scores were shown in dark blue according to the color scale. As a result, this approach provides a mimicry for the ligand binding effect by including side chains in the calculation, and the ligand binding sites are predicted without the need for a ligand structure.

According to this information, the effect of different mode values (cumulative 1 to 20 modes for *hGALK1*, and 1 to 10 modes for the proteins in the first part of the thesis) and cutoff values (7 Å, 10 Å, and 13 Å) on the prediction success of the ESSA method were investigated in terms of cryptic sites for the proteins indicated in the Materials section.

### 2.2.3. Molecular Dynamics

MD simulations provide information about how the position of atoms changes with respect to time using Newton's Law of Motion. To this extent, the partial derivative of potential energy gives the force. The consideration of the force on the atoms in the system and the generation of the new positions for each atom within the force field results in the formation of MD simulation output, which is called a trajectory. The trajectory is

composed of frames, in other words, snapshots of the simulated system (Hollingsworth and Dror 2018). The scheme for molecular dynamics (MD) simulation was demonstrated in Figure 6.

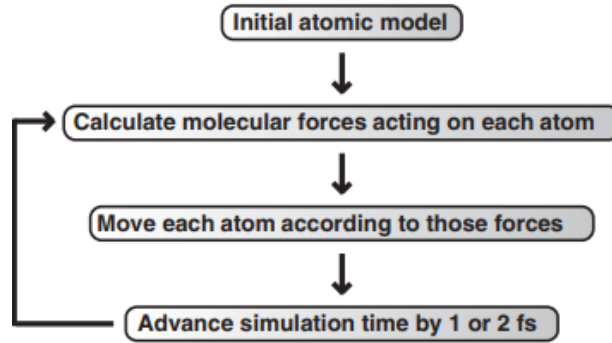


Figure 6. The scheme for molecular dynamics (MD). Copyright © 2011, Durrant and McCammon; licensee BioMed Central Ltd. (Source: Durrant and McCammon (2011)).

The force field includes bond distance, bond angle, dihedral angle, electrostatic interactions, and van der Waals (vdW) interactions. The equation (4) for the force field which includes bond stretching, bending, torsion, vdW, and electrostatic interactions, was given below (Bahar, Jernigan, and Dill 2017).

$$\begin{aligned}
 U(r_1, r_2, \dots, r_N) = & \\
 & \sum_{i=2}^N \frac{k_l}{2} (l_i - l_i^0)^2 + \sum_{i=2}^{N-1} \frac{k_\theta}{2} (\theta_i - \theta_i^0)^2 + \sum_{i=3}^{N-1} \left\{ \sum_{n=1}^3 k_\phi [1 + \cos(n\Phi_i - \Phi_i^0)] \right\} \\
 & + \sum_{i,j} \frac{a_{ij}}{r_{ij}^{12}} - \frac{b_{ij}}{r_{ij}^6} + \frac{332q_i q_j}{\epsilon r_{ij}} \quad (4)
 \end{aligned}$$

In the equation, U is the energy of the conformation, variables with superscript *i* indicate conformation while superscript 0 indicates constant experimental value,  $l_i$  and  $l_i^0$  are length,  $\theta_i$  and  $\theta_i^0$  are angles,  $r_{ij}$  denotes the distance between atoms *i* and *j*, n is an

index,  $a_{ij}$  and  $b_{ij}$  are the Lennard-Jones repulsion coefficient and the Lennard-Jones attraction coefficient, respectively,  $\epsilon$  is the effective dielectric constant, 332 is a conversion factor,  $q_i$  and  $q_j$  are charges.

### **2.2.3.1. The Investigation of GALK1 Dynamics by MD Simulations**

In the scope of the thesis, *h*GALK1 protein dynamics was investigated in the presence and absence of the fragment inhibitor (ligand ID: HR8) by performing classical MD simulations. The protein was placed inside a box containing a solution (water molecules and ions) to mimic the environment inside the cell. CHARMM-GUI was used to prepare a simulation box with periodic boundary conditions for the protein-solution system (Jo et al. 2008). The simulation box preparation parameters were determined as 12 Å edge distance for box boundaries and 303.15 Kelvin for the temperature. The simulation box type and crystal type were rectangular and cubic, respectively, while dimensions were (94.0, 94.0, 94.0, 90.0, 90.0, 90.0). 0.15 M potassium chloride (KCl) was used for the neutralization of the system, and the Monte Carlo method was used to replace the ions. On the other hand, TIP3 water molecules were used in the simulation box. The default options were used in protonation states (HSD for histidine). NVT ensemble was used for equilibration input generation and NPT ensemble was used for dynamics input generation. Monte Carlo Barostat type is isotropic and pressure coupling frequency was 100 steps. The total number of atoms in the system was 77,772 for the simulation box for substrate-bound *h*GALK1. On the other hand, PME was used as electrostatic cut-off method and Ewald error tolerance was 0.0005. The force-switch was used as vdW cut-off method. In this scope, switch-on and switch-off distances were 1.0 nm and 1.2 nm, respectively. In addition, HBonds was the constraints method. The positional restraint force constant for the protein backbone was 400.0 kJ/mol/nm<sup>2</sup> while this value was 40.0 kJ/mol/nm<sup>2</sup> for the protein side chain. In addition, the dihedral restraint force constant was 4.0 kJ/mol/rad<sup>2</sup> for carbohydrates. The simulations were performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources). OpenMM 7.6 program with CHARMM36m force field (McAuley, Huang, and Timson 2017; Eastman et al. 2017) was used for MD simulations. Before

starting simulations, 2 nanoseconds (ns) work was run with 2 femtoseconds (fs) timesteps to equilibrate the system. MD simulations were performed in three replicates for each substrate-bound and substrate-inhibitor-bound *hGALK1* system (Each simulation was 200 ns). The parameters of the inhibitor molecule were determined by the CGenFF program (Vanommeslaeghe et al. 2010; Yu et al. 2012). Visual Molecular Dynamics (VMD), UCSF Chimera, and PyMOL were used for the visualization (Humphrey, Dalke, and Schulten 1996; Pettersen et al. 2004; Schrödinger LLC 2022).

The analysis of MD results is significant for evaluating the conformational changes in the protein and the stability of the ligand. The performance of MD simulations provided 10,000 frames for each simulation, resulting in the production of 60,000 frames in total.

In this scope, root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), and radius of gyration ( $R_g$ ) were analyzed to investigate the protein dynamics. MDTraj Python library was used to analyze MD trajectories (McGibbon et al. 2015). Both TRUBA and Google Colab resources were used for the analysis of the simulations. The analysis was performed, RMSD and RMSF profiles were determined after the alignment of all frames in the VMD program (Humphrey, Dalke, and Schulten 1996).

### **2.2.3.2. The Evaluation of Collective Motions by Principal Component Analysis (PCA) Machine Learning Method**

The Principal Component Analysis (PCA) is an unsupervised dimensionality reduction method, and it was observed to decrease dimension by reaching variance to a maximum level without losing information in large data sets (Jolliffe and Cadima 2016). Otherwise, the calculations take more time due to the usage of a large data set. It is possible to find the directions (principal components-PCs) that maximize variance in the system by PCA.

The first principal component (PC1) is defined as the line with the highest total square distance from data points to the origin, while the second principal component (PC2) is perpendicular to this line. The ratio in the scope of linear combination in PC1 provides information about what is significant according to the distribution of the data.

The eigenvector for PC1 is defined when PC1 is scaled as a unit vector. Eigenvalue calculation can be done through total square distance information for PC1 and PC2.

In the literature, the PCA machine learning method was used to examine biomolecular systems, and Stein et al. provided examples of the applications of the PCA method to the protein systems in the written book chapter (Stein et al. 2006). In addition, (A. Uyar, Karamyan, and Dickson 2018) performed full-atom MD simulations for the neurolysin enzyme and used machine learning methods while investigating enzyme dynamics. To this extent, PCA was applied to MD trajectories, and principal components (modes) were determined. Therefore, PCA provides principal components that include eigenvectors, which define collective motions for the biomolecular systems (apo or ligand-bound form of the protein). More detailed information was given in the references (Cossio-Pérez, Palma, and Pierdominici-Sottile 2017; Uyar and Dickson 2021). In this thesis, 3-dimensional (3D) conformations of *hGALK1* from MD simulations were used for PCA calculations.

#### **2.2.4 ClustENMD (for GALK1)**

The ClustENMD method is used to produce different conformations of a structure. ClustENMD is an ENM-based method, where a single structure of a protein is deformed in the direction of the randomly weighted normal modes obtained from ENM, and new conformations are produced (Kaynak et al. 2021). New conformations are then subjected to a short MD simulation to minimize their energy. The final conformers contain all atoms of the structure. In this thesis, 10 normal modes were used in ClustENMD analysis. ClustENMD is accessible in the ProDy package, and it can only be run on a GPU platform (Bakan, Meireles, and Bahar 2011). The workflow is given below in Figure 7.

In the workflow, it can be seen that ClustENMD uses ANM, which was proposed by Atilgan et al., and provides eigenvalues and eigenvectors that represent frequency and mode shapes, respectively (Atilgan et al. 2001). ANM differentiates from GNM since ANM is in the  $3N \times 3N$  dimension ( $N$  indicates the number of nodes or residues). In addition, ANM gives  $3N-6$  modes that are represented by eigenvectors and eigenvalues,

which neglects rotation and translation motions when x, y, and z coordinates are considered (Doruker, Atilgan, and Bahar 2000).

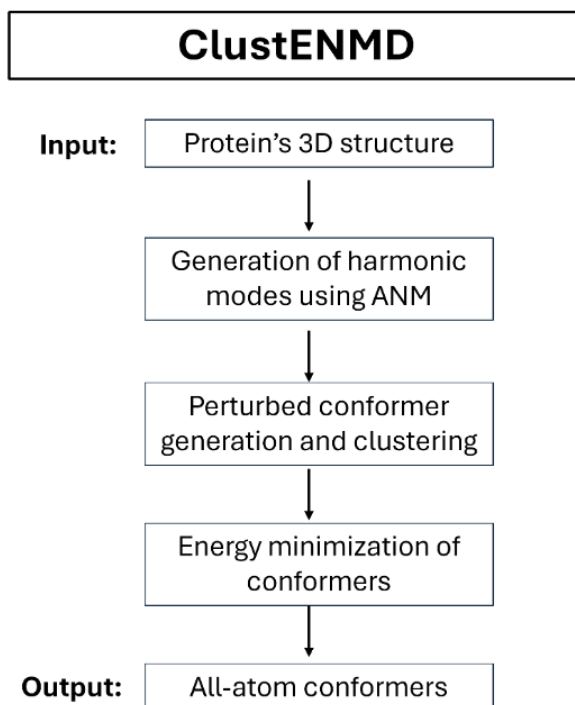


Figure 7. The workflow for ClustENMD.



## CHAPTER 3

### RESULTS AND DISCUSSION

#### 3.1. Allosteric and Cryptic Site Detection

##### 3.1.1. ESSA Results

The ESSA method uses a cutoff distance of 10 Å and averaged over 10 modes for the calculation default. In this thesis, the significant residues in the ligand binding sites were aimed to be predicted by changing the cutoff distance to 7, 11, and 13 Å in addition to the default cutoff distance value of 10 Å (The cutoff value of 11 Å was only applied for the GALK1 structure). In this scope, the orthosteric, allosteric, and cryptic sites for TEM-1 beta-lactamase (holo form: 1PZO and apo form: 1ZG4), CDK2 (holo form: 3PXZ and apo form: 4EK3), and GALK1 (holo form: 6Q91) were predicted using ESSA with various cutoff values without changing the total number of modes used in the calculation. In addition, the ESSA analysis was performed using the ProDy package. On the other hand, the binding sites were also predicted by including the contribution of each mode to the calculation of the average of the modes and merging the z-score results obtained from each successive mode set, such as mode 1, modes 1-to-2, modes 1-to-3, etc., which from here on is defined as “combined 10 modes” if the total number of modes is 10. On the other hand, the binding sites were also determined by considering averaged over 10 modes, which is the default parameter in ESSA. Only the GALK1 structure was investigated with the combined 20-mode sets. In the evaluations, the residue indexes (resid) in the RCSB Protein Data Bank (PDB) were taken as reference residue numbers, and plots were generated according to this information. Firstly, z-score distributions of residues were investigated in the presence of cutoff 7 Å, 10 Å, and 13 Å combined 10 modes for TEM-1 beta-lactamase (holo form: 1PZO) and CDK2 (holo form: 3PXZ)

(Figure 8 and Figure 9). According to these findings, very few residues in the holo form (1PZO) of TEM-1 beta-lactamase demonstrated z-score values above 5 in cutoff 7 Å. On the other hand, z-score distributions demonstrated similar profiles between cutoff 10 Å and 13 Å. However, the high z-score values above 5 were observed in higher counts compared to that of the cutoff 7 Å. To this extent, cutoff 10 Å and 13 Å can be considered more successful in predicting residues with higher z-scores compared to cutoff 7 Å (Figure 8). In the case of the CDK2 holo form (3PXZ), cutoffs 10 Å and 13 Å demonstrate similar patterns with slight differences in the counts of especially the residues with the z-score value of 5. On the other hand, cutoff 7 Å mostly differentiated from cutoff 10 Å and 13 Å in the counts of the residues with the z-score values of 1, 2, and 4.

Candlestick charts are widely used in finance for the demonstration of decreases and increases in prices. In these charts, the opening, the maximum, the minimum, and the closing groups should be defined. Within these considerations, candlestick charts were used to investigate z-score variation in the residues between cutoff 10 Å and 7 Å combined 10 modes as well as between cutoff 10 Å and 13 Å combined 10 modes. To this extent, the opening value was determined as z-scores obtained from cutoff 10 Å, the maximum group with z-scores determined as 7 for all residues (because it is the maximum value in the range), the minimum group with z-scores determined as 1 for all residues (because it is the minimum value in the range) and the closing group was determined as z-scores obtained from cutoff 7 Å or cutoff 13 Å depending on the comparison groups. The candlestick charts were demonstrated by excluding the minimum and maximum groups for clear depiction and generated for structures of 1PZO and 3PXZ in Figure 10 and Figure 11, respectively.

In the candlestick charts, the blue-colored vertical line represents an increase, and the salmon-colored vertical line represents a decrease in the z-score value in the corresponding residue number. The body parts of the candlesticks provide information about the change in the z-score value in the corresponding residue number. For instance, the upper part of the vertical salmon-colored candlestick represents the z-score value of the corresponding residue in the presence of a cutoff 10 Å since the opening group is cutoff 10 Å and the lower part of the candlestick indicates the z-score value in the presence of cutoff 7 Å in Figure 10a and Figure 11a. The reverse position explanation is valid for the vertical blue candlestick. To this extent, the constant horizontal salmon-colored line represents no change in the z-score value of the corresponding residue.

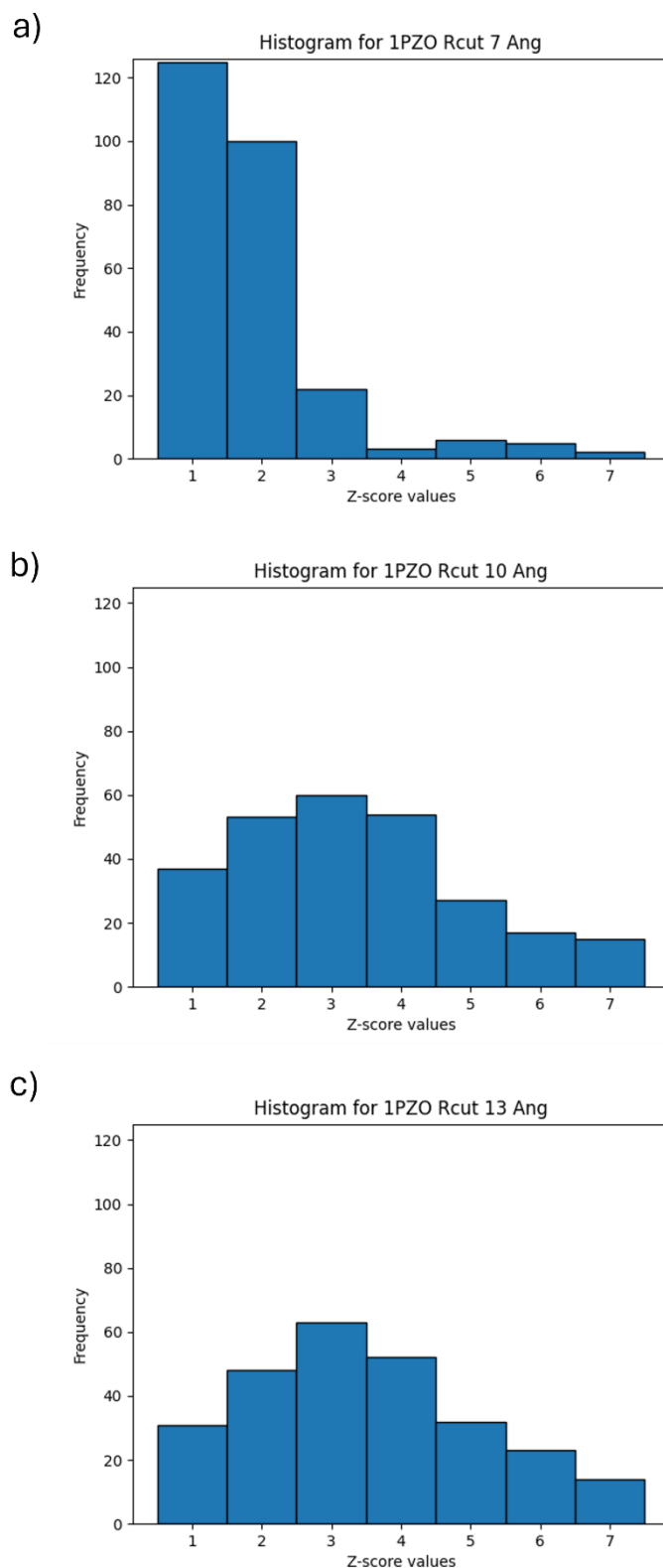


Figure 8. Z-score distributions of residues for TEM-1 beta-lactamase (holo form: 1PZO). a) in the presence of cutoff 7 Å, b) in the presence of cutoff 10 Å, c) in the presence of cutoff 13 Å.

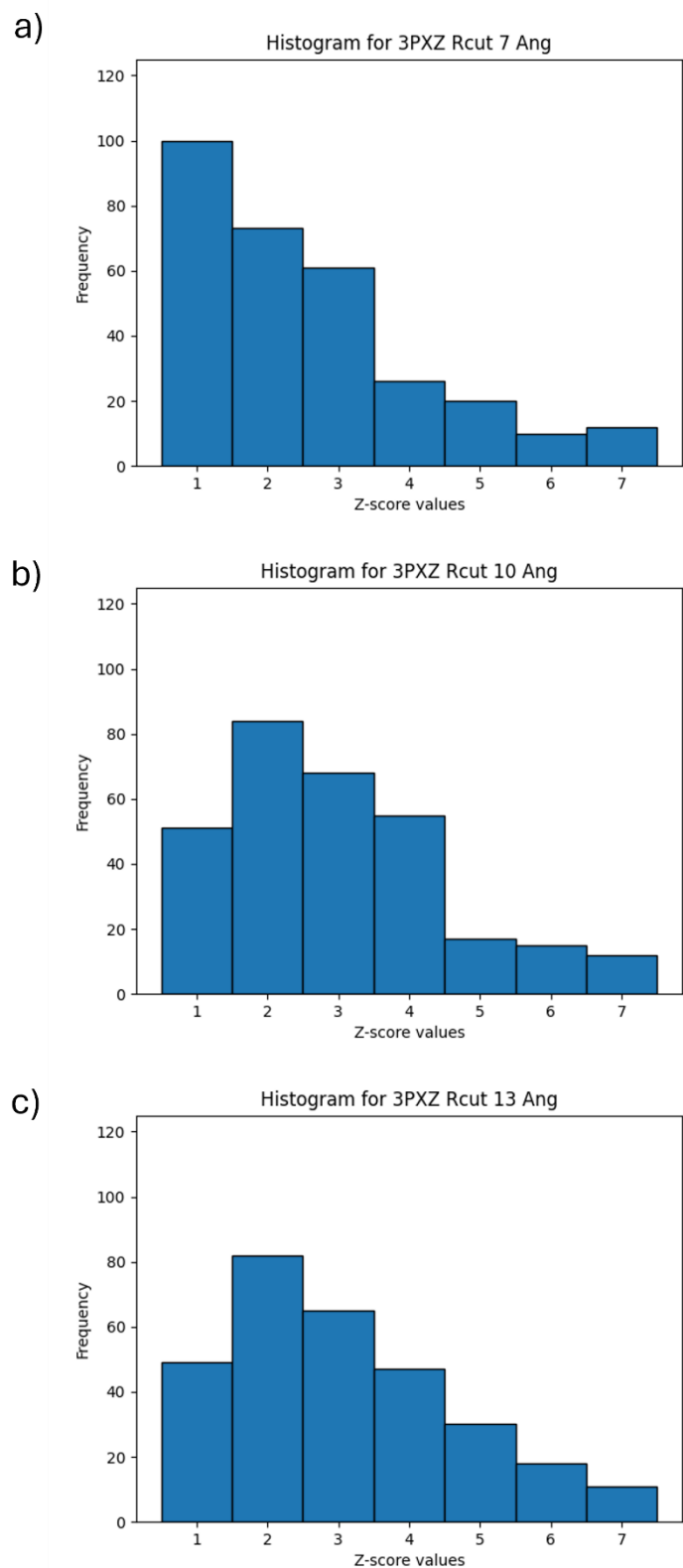


Figure 9. Z-score distributions of residues for CDK2 (holo form: 3PXZ). a) in the presence of cutoff 7 Å, b) in the presence of cutoff 10 Å, c) in the presence of cutoff 13 Å.

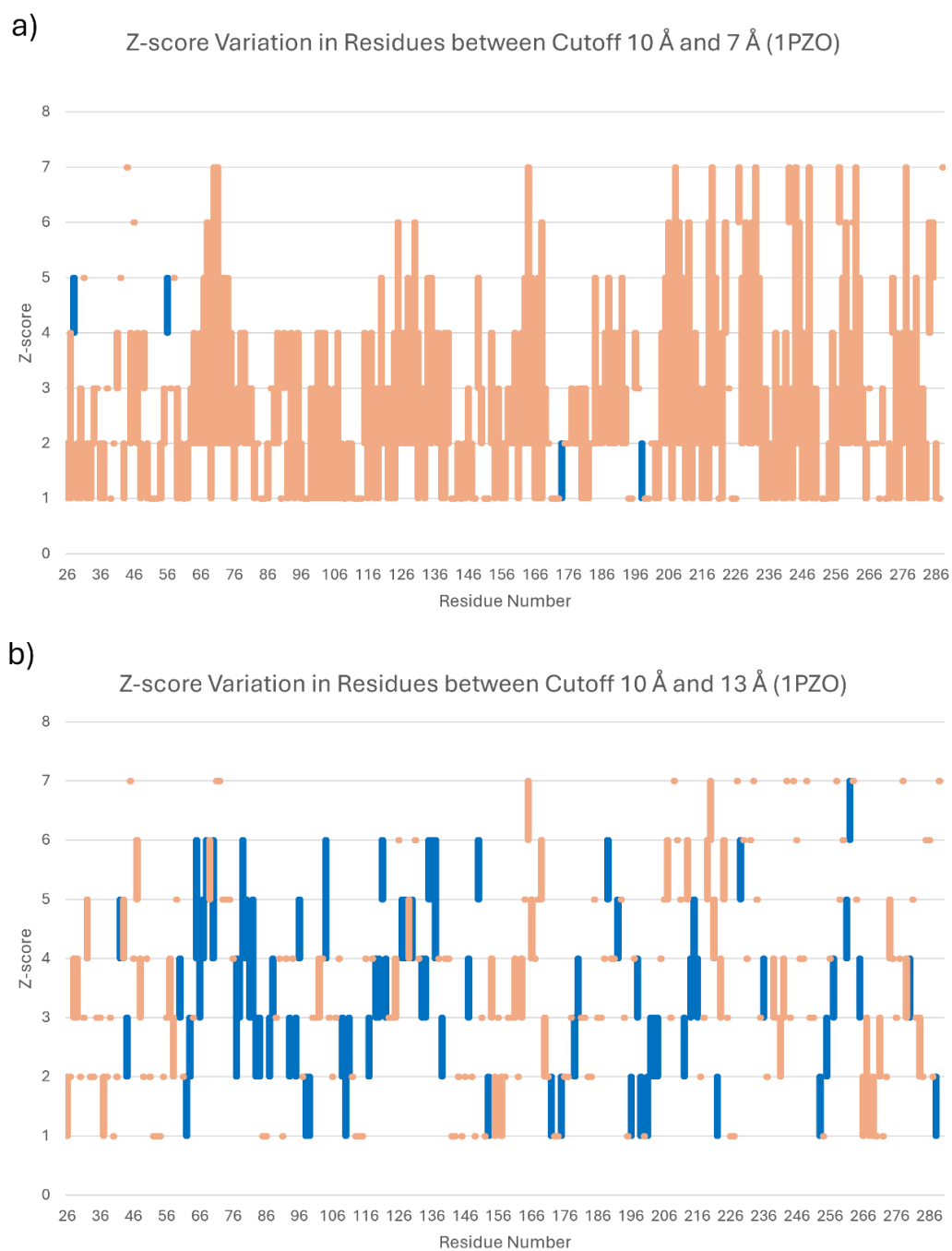


Figure 10. Z-score variation in the holo TEM-1 beta-lactamase residues between different cutoff values (1PZO). a) between cutoff 10 Å and 7 Å, b) between cutoff 10 Å and 13 Å. Dark blue: increase in z-score; light orange: decrease in z-score.

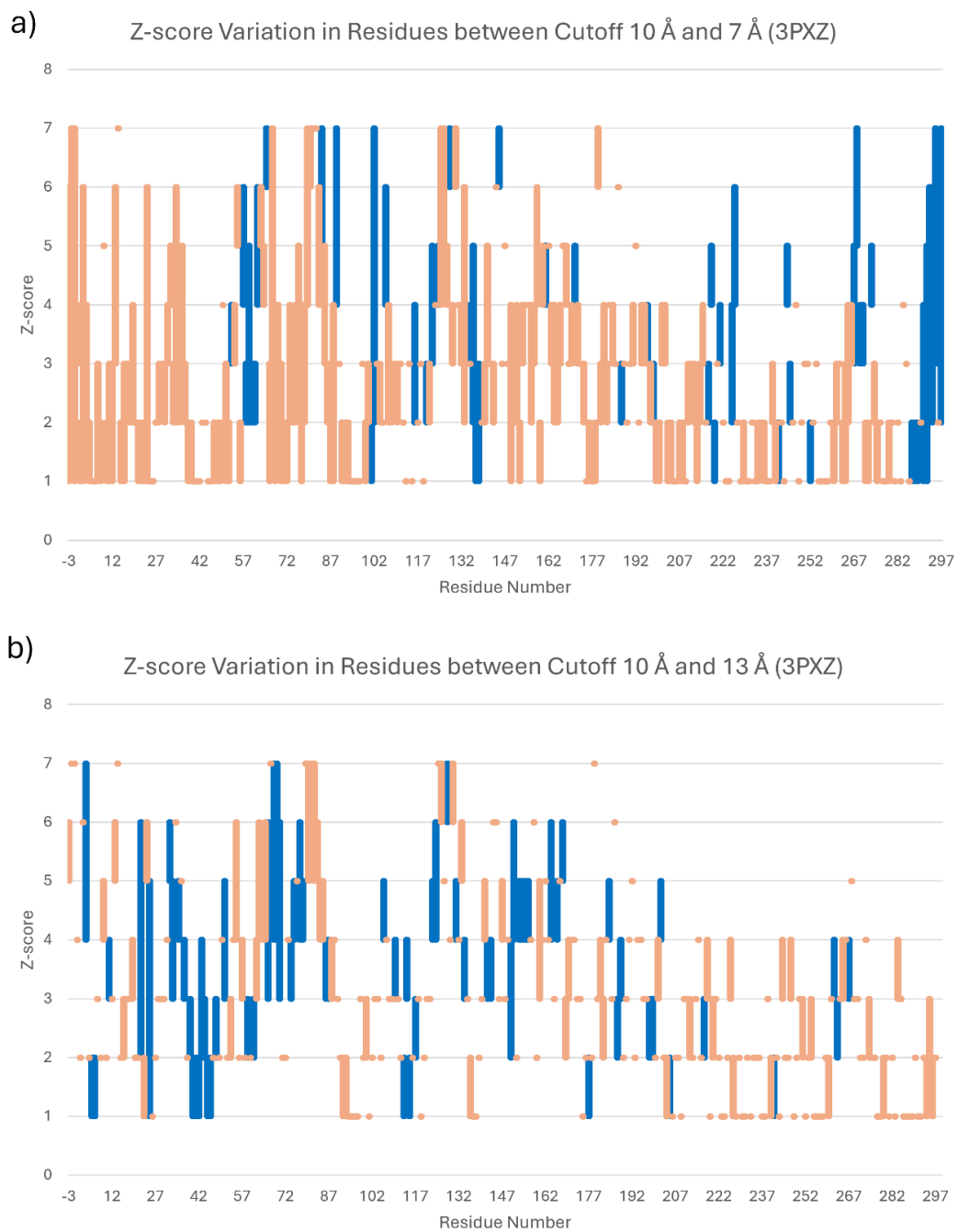


Figure 11. Z-score variation in the holo CDK2 residues between different cutoff values (3PXZ). a) between cutoff 10 Å and 7 Å, b) between cutoff 10 Å and 13 Å.

According to these results, the number of significant residues (residues with higher z-scores) decreased due to the cutoff change from 10 Å to 7 Å in 1PZO (Figure 10a). On the other hand, more improvements were observed from the cutoff change from 10 Å to 13 Å. However, several declines were also detected (Figure 10b). In the scope of 3PXZ results, the decline in the significant residues was observed from cutoff 10 Å to 7 Å. However, more increments were also observed than the results obtained from 1PZO in Figure 10a. On the other hand, when the residue numbers up to 87 were considered, the number of significant residues increased in the presence of 13 Å while it decreased in the presence of 7 Å (Figure 11a and Figure 11b). In addition to the histograms and candlestick charts for the structures of 1PZO and 3PXZ, tables were generated for each structure that includes the residues with high z-score values of 6 and 7 in the groups with the cutoff values of 7 Å, 10 Å, 11 Å (only for 6Q91) and 13 Å as well as combined 10 modes or averaged over 10 modes. In the tables, significant residues were highlighted with specific colors that indicate the presence of the residue around 5 Å of the corresponding ligand. In addition, the highlighted residues were labeled with the corresponding number of stars according to the involvement of the residue in hydrogen bonds or non-bonded interactions provided by PDBsum. The below part of each table includes the total and ratio values that represent the number of residues with z-score values above 5 and the number of residues that are around 5 Å of the ligands with z-score values above 5 compared to the total number of residues in the corresponding group with z-score values above 5, respectively. The results for 1PZO, 1ZG4, 3PXZ, and 4EK3 were given in Table 5 - Table 8, respectively. In addition, the surface representations of allosteric and orthosteric sites of the proteins colored according to z-score values were demonstrated in Figure 12 – Figure 15, respectively.





Table 5. The essential residues obtained from ESSA analysis for the holo form of TEM-1 beta-lactamase enzyme with the PDB ID of 1PZO (Bold: Present in 1PZO absent in 1ZG4 in the corresponding group (for colored residues), \*: PDBsum hydrogen bond, \*\*: PDBsum nonbonded, \*\*\*: PDBsum both hydrogen bond and nonbonded).

Cutoff 7 Å combined 10 modes	Cutoff 7 Å averaged over 10 modes	Cutoff 10 Å combined 10 modes	Cutoff 10 Å averaged over 10 modes	Cutoff 13 Å combined 10 modes	Cutoff 13 Å averaged over 10 modes
TYR, 46	TYR, 46	TYR, 46	TYR, 46	TYR, 46	TYR, 46
GLU, 48	GLU, 48	GLU, 48	TRP, 229	PHE, 66	PHE, 72
TRP, 229	TRP, 229	<b>SER, 70***</b>	<b>ARG, 244**</b>	<b>MET, 69**</b>	LEU, 137
<b>ARG, 244**</b>	ARG, 259	PHE, 72	LEU, 250	THR, 71	PHE, 151
ARG, 259	TRP, 290	<b>LYS, 73</b>	ARG, 259	PHE, 72	TRP, 210
<b>ILE, 263**</b>		ILE, 127	TRP, 290	<b>LYS, 73</b>	TRP, 229
TRP, 290		<b>ASN, 132***</b>		VAL, 80	PHE, 230
		<b>GLU, 166</b>		TYR, 105	<b>ARG, 244**</b>
		<b>ASN, 170**</b>		LEU, 122	<b>LEU, 250</b>
		ILE, 208		ILE, 127	ARG, 259
		TRP, 210		<b>ASN, 132***</b>	<b>ILE, 279**</b>
		MET, 211		ASN, 136	TRP, 290
		ASP, 214		LEU, 137	
		<b>LEU, 220**</b>		LEU, 138	
		LEU, 221**		PHE, 151	
		<b>LEU, 225</b>		<b>GLU, 166</b>	
		TRP, 229		LEU, 190	
		ILE, 231		TRP, 210	
		ASP, 233		MET, 211	
		<b>LYS, 234</b>		<b>LEU, 221**</b>	
		<b>ARG, 244**</b>		TRP, 229	
		<b>ILE, 246**</b>		PHE, 230	
		ILE, 247		ILE, 231	
		LEU, 250		ASP, 233	
		ARG, 259		<b>LYS, 234</b>	
		VAL, 261		<b>ARG, 244**</b>	
		<b>ILE, 263**</b>		<b>ILE, 246**</b>	
		TYR, 264		ILE, 247	
		<b>ILE, 279**</b>		LEU, 250	
		LEU, 286		ARG, 259	
		ILE, 287		<b>VAL, 261</b>	
		TRP, 290		<b>ILE, 263**</b>	
				TYR, 264	
				<b>ILE, 279**</b>	
				<b>LEU, 286</b>	
				ILE, 287	
				TRP, 290	
Total: 7	Total: 5	Total: 32	Total: 6	Total: 37	Total: 12
Ratio: 2/7	Ratio: 0/5	Ratio: 16/32	Ratio: 2/6	Ratio: 13/37	Ratio: 3/12
<b>FOS</b>	<b>CBT_A300</b>	<b>CBT_A301</b>	<b>BOTH CBT_A300 AND CBT_A301</b>		

Table 6. The essential residues obtained from ESSA analysis for the apo form of TEM-1 beta-lactamase enzyme with the PDB ID of 1ZG4 (Bold: Present in 1ZG4 absent in 1PZO in the corresponding group (for colored residues), \*: PDBsum hydrogen bond, \*\*: PDBsum nonbonded, \*\*\*: PDBsum both hydrogen bond and nonbonded).

Cutoff 7 Å combined 10 modes	Cutoff 7 Å averaged over 10 modes	Cutoff 10 Å combined 10 modes	Cutoff 10 Å averaged over 10 modes	Cutoff 13 Å combined 10 modes	Cutoff 13 Å averaged over 10 modes
TYR, 46	TYR, 46	TYR, 46	TYR, 46	PHE, 72	TRP, 290
<b>VAL, 216**</b>		GLU, 48	TYR, 97	<b>LYS, 73</b>	
<b>LEU, 220**</b>		PHE, 72	LEU, 137	LEU, 75	
ARG, 222		<b>LYS, 73</b>	PHE, 151	LEU, 76	
ASP, 233		LEU, 75	TRP, 210	TYR, 105	
<b>ARG, 244**</b>		TYR, 97	TRP, 229	LEU, 122	
<b>ILE, 263**</b>		ILE, 127	PHE, 230	ILE, 127	
<b>ILE, 279**</b>		<b>SER, 130***</b>	<b>ARG, 244**</b>	ASN, 136	
		<b>ASN, 132***</b>	LEU, 250	<b>GLU, 166</b>	
		LEU, 137	ARG, 259	TRP, 210	
		PHE, 151	TRP, 290	MET, 211	
		<b>GLU, 166</b>		GLU, 212	
		MET, 186		ARG, 222	
		LEU, 190		TRP, 229	
		LEU, 193		PHE, 230	
		LEU, 194		ILE, 231	
		ILE, 208		ASP, 233	
		TRP, 210		<b>LYS, 234</b>	
		MET, 211		ILE, 247	
		LEU, 221**		LEU, 250	
		ARG, 222		ARG, 259	
		TRP, 229		TYR, 264	
		PHE, 230		TRP, 290	
		ILE, 231			
		ASP, 233			
		<b>LYS, 234</b>			
		<b>ARG, 244**</b>			
		ILE, 246**			
		ILE, 247			
		LEU, 250			
		ARG, 259			
		ILE, 260			
		VAL, 261			
		<b>ILE, 263**</b>			
		TYR, 264			
		<b>ILE, 279**</b>			
		LEU, 286			
		TRP, 290			
Total: 8	Total: 1	Total: 38	Total: 11	Total: 23	Total: 1
Ratio: 5/8	Ratio: 0/1	Ratio: 13/38	Ratio: 2/11	Ratio: 4/23	Ratio: 0/1
FOS	CBT_A300	CBT_A301	<b>BOTH CBT_A300 AND CBT_A301</b>	<b>BOTH CBT_A300 AND FOS</b>	

### 1ZG4 – Combined 10 modes

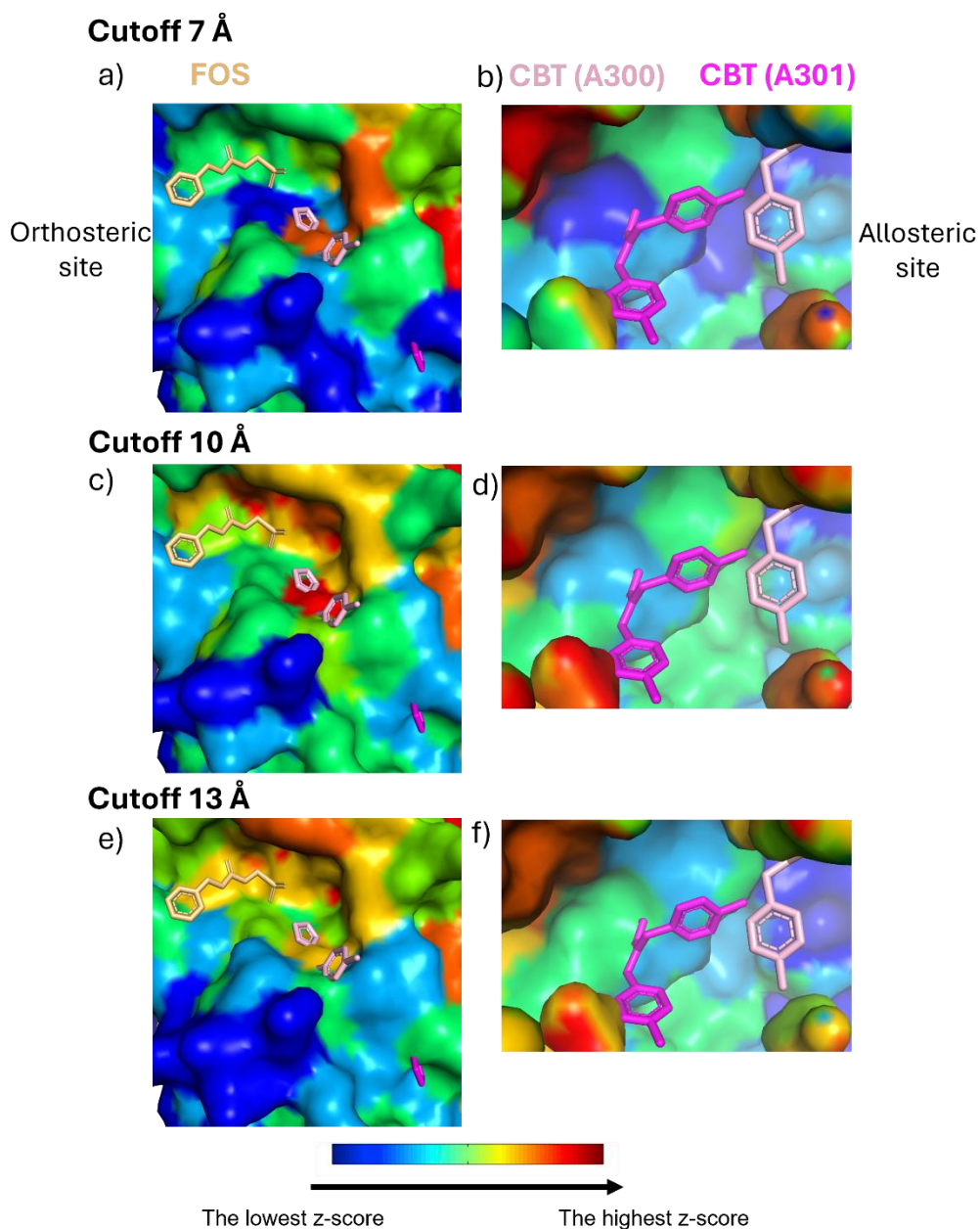


Figure 13. The ESSA results for the apo form of TEM-1 beta-lactamase (1ZG4). The results were shown in the presence of combined 10 modes where a) and b) cutoff 7 Å; c) and d) cutoff 10 Å; e) and f) cutoff 13 Å. The orthosteric (left column) and allosteric ligands (right column) were demonstrated in sticks with corresponding colors and protein in surface representation colored according to z-score values in the color spectrum.

Table 7. The essential residues obtained from ESSA analysis for the holo form of CDK2 enzyme with the PDB ID of 3PXZ (Bold: Present in 3PXZ absent in 4EK3 in the corresponding group (for colored residues), \*: PDBsum hydrogen bond, \*\*: PDBsum nonbonded, \*\*\*: PDBsum both hydrogen bond and nonbonded).

Cutoff 7 Å combined 10 modes	Cutoff 7 Å averaged over 10 modes	Cutoff 10 Å combined 10 modes	Cutoff 10 Å averaged over 10 modes	Cutoff 13 Å combined 10 modes	Cutoff 13 Å averaged over 10 modes
TYR, 15**	<b>GLN, 85**</b>	PRO, -2	GLU, -1	GLU, -1	GLU, -1
LEU, 58	PHE, 90	GLU, -1	PHE, 0	PHE, 0	PHE, 0
<b>ILE, 63</b>	TYR, 269	PHE, 0	ASN, 3	ASN, 3	ASN, 3
<b>LEU, 66**</b>	LEU, 296	ASN, 3	<b>TYR, 15**</b>	PHE, 4	PHE, 4
<b>GLU, 81**</b>	LEU, 298	THR, 14	LEU, 25	<b>TYR, 15**</b>	<b>TYR, 15**</b>
PHE, 82		<b>TYR, 15**</b>	ASP, 68	ASN, 23	ASP, 68
<b>LEU, 83***</b>		LEU, 25	<b>PHE, 80**</b>	<b>LYS, 33***</b>	<b>VAL, 69**</b>
<b>GLN, 85**</b>		<b>ILE, 35**</b>	<b>PHE, 82</b>	<b>ILE, 35**</b>	ILE, 70
PHE, 90		<b>LYS, 56***</b>	ARG, 126	LEU, 67	
LEU, 103		<b>VAL, 64**</b>	TRP, 187	ASP, 68	
TYR, 107		<b>LEU, 66**</b>		<b>VAL, 69**</b>	
LYS, 129		ASP, 68		ILE, 70	
<b>GLN, 131</b>		PHE, 80**		<b>HIS, 71***</b>	
<b>ASP, 145***</b>		<b>GLU, 81**</b>		<b>LEU, 78**</b>	
<b>PHE, 146***</b>		PHE, 82		PHE, 80**	
TYR, 180		<b>LEU, 83***</b>		<b>PHE, 82</b>	
TRP, 187		<b>HIS, 84**</b>		HIS, 125	
TRP, 227		ARG, 126		ARG, 126	
TYR, 269		ASP, 127		ASP, 127	
PRO, 294		LYS, 129		LYS, 129	
LEU, 296		<b>GLN, 131</b>		<b>GLN, 131</b>	
LEU, 298		<b>LEU, 134**</b>		ASP, 145***	
		ASP, 145***		<b>PHE, 146***</b>	
		<b>PHE, 146***</b>		PHE, 152	
		TYR, 159		TYR, 159	
		TYR, 180		THR, 165	
		TRP, 187		ARG, 169	
				TYR, 180	
				TRP, 187	
Total: 22	Total: 5	Total: 27	Total: 10	Total: 29	Total: 8
Ratio: 10/22	Ratio: 1/5	Ratio: 14/27	Ratio: 3/10	Ratio: 11/29	Ratio: 2/8
JWS	A299_2AN	A300_2AN	Both A299_2AN and A300_2AN		Both A300_2AN and JWS

### 3PXZ – Combined 10 modes

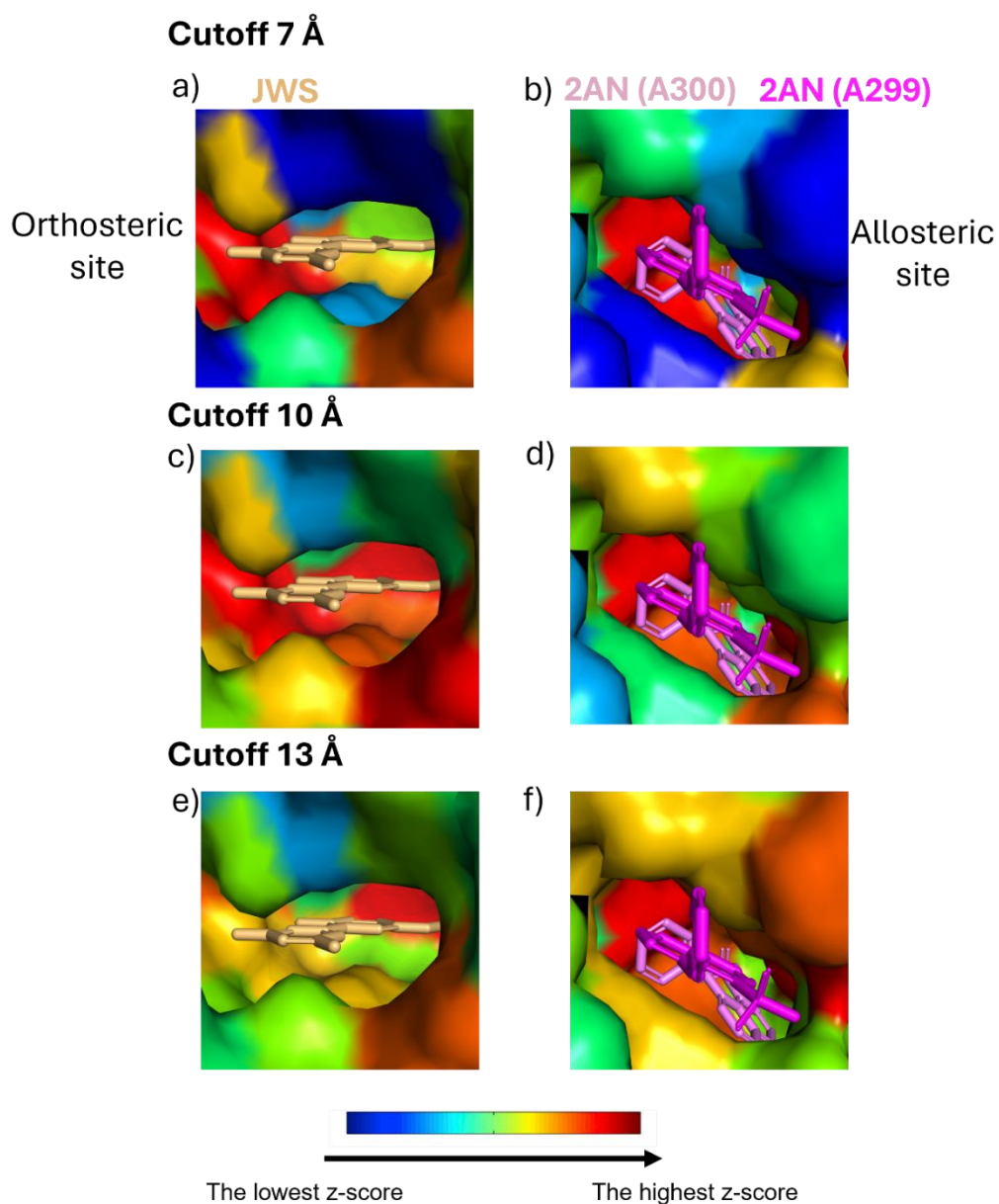


Figure 14. The ESSA results for the holo form of the CDK2 (3PXZ). The results were shown in the presence of combined 10 modes where a) and b) cutoff 7 Å; c) and d) cutoff 10 Å; e and f) cutoff 13 Å. The orthosteric (left column) and allosteric ligands (right column) were demonstrated in sticks with corresponding colors and protein in surface representation colored according to z-score values in the color spectrum.

Table 8. The essential residues obtained from ESSA analysis for the apo form of CDK2 enzyme with the PDB ID of 4EK3 (Bold: Present in 4EK3 absent in 3PXZ in the corresponding group (for colored residues), \*: PDBsum hydrogen bond, \*\*: PDBsum nonbonded, \*\*\*: PDBsum both hydrogen bond and nonbonded).

Cutoff 7 Å combined 10 modes	Cutoff 7 Å averaged over 10 modes	Cutoff 10 Å combined 10 modes	Cutoff 10 Å averaged over 10 modes	Cutoff 13 Å combined 10 modes	Cutoff 13 Å averaged over 10 modes
TYR, 15**	<b>TYR, 15**</b>	MET, 1	MET, 1	MET, 1	MET, 1
<b>LYS, 33***</b>	<b>VAL, 69**</b>	TYR, 15**	TYR, 15**	PHE, 4	LYS, 34
<b>VAL, 64**</b>	<b>LEU, 76**</b>	LEU, 32	<b>ILE, 35**</b>	TYR, 15**	<b>ILE, 35**</b>
<b>VAL, 69**</b>	PHE, 152	<b>LYS, 33***</b>	ILE, 70	LEU, 32	ILE, 70
<b>LEU, 76**</b>	TYR, 269	<b>ILE, 35**</b>	<b>HIS, 71***</b>	<b>LYS, 33***</b>	TYR, 77
<b>PHE, 80**</b>		ILE, 70	<b>LEU, 76**</b>	LYS, 34	
<b>PHE, 82</b>		<b>HIS, 71***</b>	TYR, 77	<b>ILE, 35**</b>	
ASP, 127		<b>LEU, 76**</b>	PHE, 152	LEU, 67	
LYS, 129		TYR, 77		ASP, 68	
PHE, 152		<b>LEU, 78**</b>		ILE, 70	
VAL, 164		PHE, 80**		TYR, 77	
TYR, 168		<b>GLU, 81**</b>		<b>LEU, 78**</b>	
ARG, 169		PHE, 82		VAL, 79	
TYR, 269		<b>LEU, 83***</b>		PHE, 80**	
		<b>HIS, 84**</b>		<b>GLU, 81</b>	
		ARG, 126		PHE, 82	
		ASP, 127		<b>LEU, 83</b>	
		LYS, 129		<b>HIS, 84</b>	
		<b>GLN, 131</b>		HIS, 125	
		ASP, 145***		ARG, 126	
		<b>LEU, 148</b>		LYS, 129	
		PHE, 152		<b>GLN, 131</b>	
		VAL, 164		<b>LEU, 134**</b>	
		TYR, 168		ASP, 145***	
		ARG, 169		<b>PHE, 146***</b>	
		TYR, 180		ARG, 150	
		TRP, 187		PHE, 152	
				TRP, 167	
				TYR, 168	
				ARG, 169	
				TYR, 180	
				TRP, 187	
				PHE, 203	
				TYR, 269	
Total: 14	Total: 5	Total: 27	Total: 8	Total: 34	Total: 5
Ratio: 7/14	Ratio: 3/5	Ratio: 14/27	Ratio: 4/8	Ratio: 13/34	Ratio: 1/5
JWS	A299_2AN	A300_2AN	Both A299_2AN and A300_2AN	Both A300_2AN and JWS	

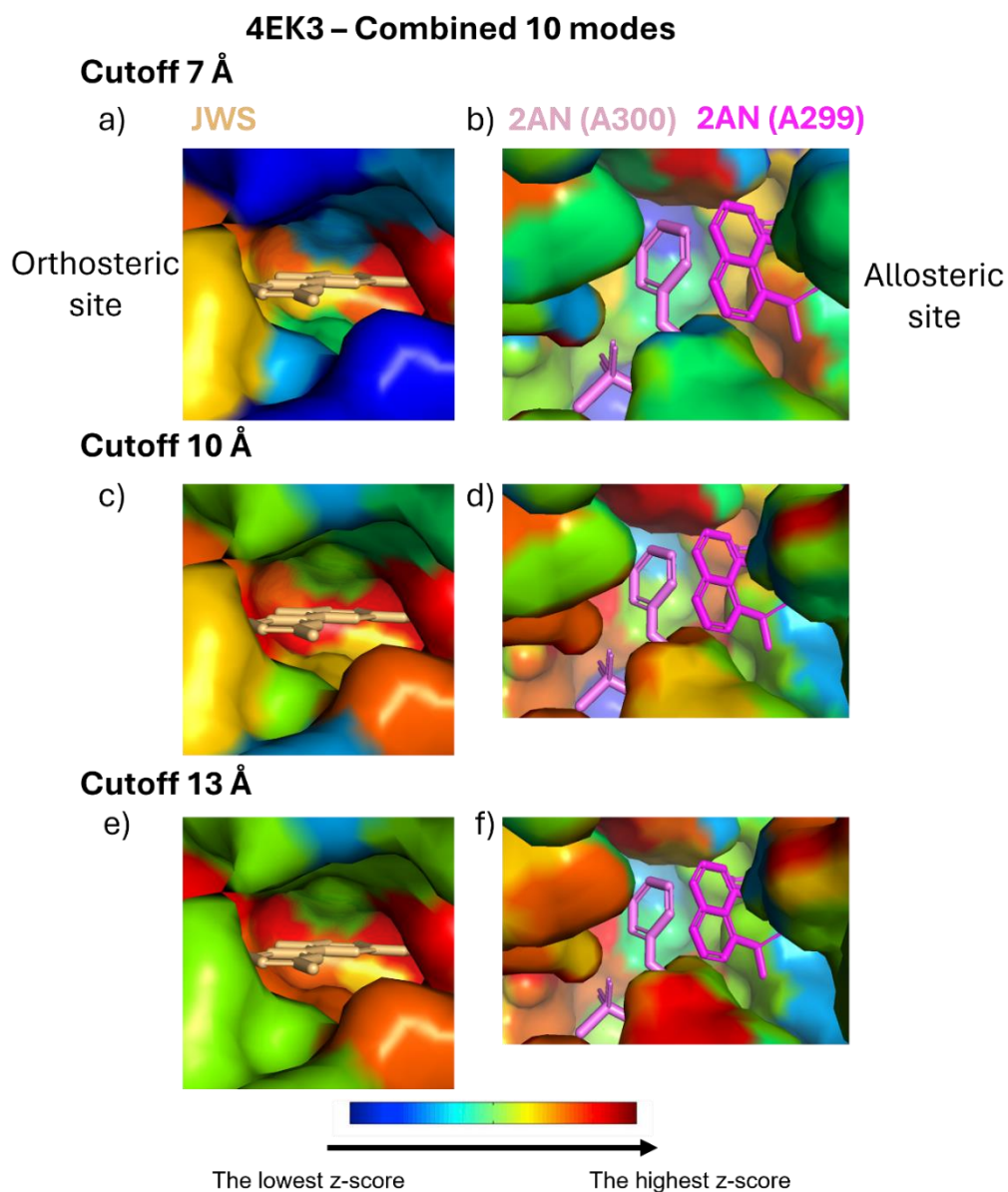


Figure 15. The ESSA results for the apo form of CDK2 (4EK3). The results were shown in the presence of combined 10 modes where a) and b) cutoff 7 Å; c) and d) cutoff 10 Å; e and f) cutoff 13 Å. The orthosteric (left column) and allosteric ligands (right column) were demonstrated in sticks with corresponding colors and protein in surface representation colored according to z-score values in the color spectrum.

On the other hand, the radius of gyration was calculated using “gyradius” command (PyMOL extension known as psico) (“Psico - PyMOLWiki” 2024) for TEM-1

beta-lactamase (holo form: 1PZO and apo form: 1ZG4), CDK2 (holo form: 3PXZ and apo form: 4EK3) (Table 9).

Table 9. The radius of gyration values of the studied structures.

<b>PDB ID of The Structures</b>	<b>Radius of Gyration (Å)</b>
1PZO	17.78
1ZG4	17.72
3PXZ	20.18
4EK3	19.37

The radius of gyration was observed to be highest in 3PXZ with a value of 20.18 Å while the lowest in 1ZG4 with a value of 17.72 Å. The radius of gyration values of the apo and holo forms of the TEM-1 beta-lactamase did not demonstrate significant change, while slight variation was observed between the apo and holo forms of the CDK2.

According to the ESSA results for the holo form of TEM-1 beta-lactamase structure with the PDB ID of 1PZO, the contributions of combining 10 modes in the cutoff values of 7 Å, 10 Å, and 13 Å were observed compared to the averaged over 10 modes for the corresponding groups in terms of the observation of higher number of residues with the z-scores values above 5. These results demonstrate the significance of each mode in a combined manner on the prediction of possible ligand binding sites for this structure. On the other hand, the highest number, 37, of essential residues where the z-score is above 5, was obtained in the presence of the cutoff value of 13 Å in combined 10 modes while these values are 7 and 32 in cutoffs 7 Å and 10 Å in combined 10 modes. The advantage of cutoff 13 Å can also be observed with the higher number of residues with z-scores above 5 when the results obtained from averaged over 10 modes with the other cutoff values were compared for corresponding groups. On the other hand, several residues around 5 Å of allosteric and orthosteric ligands can be predicted by the usage of cutoff 10 and cutoff 13 Å in combined 10 modes for this structure, while it was not observed in both groups of cutoffs 7 Å, 10 Å, and 13 Å averaged over 10 modes for orthosteric ligand



binding site. It should be emphasized that the crystal structure of the holo form (1PZO) does not involve an orthosteric ligand, and the interaction information with the orthosteric ligand was obtained from another crystal structure of TEM-1 beta-lactamase with the PDB ID of 1AXB. Cutoff 10 Å combined over 10 modes is more successful when the intersection was considered for the residues around 5 Å of ligands and the residues with z-scores above 5. To this extent, 16 out of 32 predicted residues are located in the 5 Å of allosteric or orthosteric ligands for cutoff 10 Å combined 10 modes, while this value was 13/37 for cutoff 13 Å combined 10 modes. On the other hand, Leu221, Leu250, Val261, and Leu286, which are around 5 Å of CBT (resid: A301), were predicted in cutoffs 10 Å and 13 Å combined 10 modes. In addition, Ile246, which interacts with CBT (resid: A300) through non-bonded interactions, and Arg244, Ile263, and Ile279, which are around 5 Å of both CBT (resid: A300 and A301) were also found in the same groups. On the other hand, Leu250 was also observed in cutoffs 10 Å and 13 Å averaged over 10 modes.

On the other hand, the results obtained for the apo form of TEM-1 beta-lactamase structure with the PDB ID of 1ZG4 demonstrated the highest number of residues (the total value of 38) with z-scores above 5 from cutoff 10 Å combined 10 modes, while a single residue was predicted for both groups of cutoff 7 Å and cutoff 13 Å averaged over 10 modes. On the other hand, the residues around 5 Å of the orthosteric ligand (which can also be referred to as “orthosteric site” for simplicity) were predicted in the presence of cutoff 10 Å and cutoff 13 Å combined 10 modes only. To this extent, Lys73, Glu166, and Lys234 were predicted in both cutoff groups. In addition, the allosteric ligand binding site was not predicted successfully in the cutoff 7 Å and 13 Å averaged over 10 modes, while Arg244 and Leu250 were observed in cutoff 10 Å averaged over 10 modes. On the other hand, Leu220, Arg244, Ile263, and Ile279, which are around 5 Å of both CBT (resid: A300 and A301) were predicted in cutoff 7 Å combined 10 modes. In addition, the highest ratio was observed in the same group with the ratio of 5/8 in the scope of predicting residues around 5 Å of the ligands with z-scores 6 and 7 in total. When the predicted residues that are around 5 Å of ligands were compared between 1PZO (holo) and 1ZG4 (apo) for their corresponding groups, Ser70, Asn170, Leu220, and Leu225 were found in cutoff 10 Å combined 10 modes in 1PZO (holo) while these residues were not observed in 1ZG4 (apo). Met69, Asn132, Leu221, Arg244, Ile246, Val261, Ile263, Ile279, and Leu286 were present in cutoff 13 Å combined 10 modes in 1PZO (holo) but absent in 1ZG4. Similarly, Arg244, Leu250, and Ile279 were observed in cutoff 13 Å averaged over 10 modes in 1PZO only. On the other hand, Val216, Leu220, and Ile279

were predicted in cutoff 7 Å and Ser130 in cutoff 10 Å combined 10 modes in the apo form (1ZG4), while these residues were not found in the holo form (1PZO). According to these findings, cutoff 13 Å combined 10 modes was considered successful in predicting orthosteric and allosteric sites and finding different critical residues in these sites in the ligand-bound (holo) form (1PZO) compared to the apo form.

According to the ESSA analysis for the holo form of CDK2 enzyme (PDB ID: 3PXZ), all groups except cutoff 7 Å averaged over 10 modes predicted several residues in the allosteric ligand binding site. On the other hand, the highest number of residues with z-scores above 5 were found in cutoff 13 Å combined 10 modes with the total value of 29 in total while this value was 5 in cutoff 7 Å averaged over 10 modes. Also, Glu81, Phe82, Leu83, Gln131, and Asp145 that are around 5 Å of the orthosteric ligand were found in both cutoff 7 Å and 10 Å combined 10 modes. In addition, Tyr15 and Phe146 which are around 5 Å of the ligand 2AN (resid: A300) were predicted in both cutoff 7 Å, 10 Å, and 13 Å combined 10 modes. Moreover, the highest number of residues in the orthosteric site was observed in cutoff 10 Å combined 10 modes with a value of 9. These residues include Val64, Phe80, Glu81, Phe82, Leu83, His84, Gln131, Leu134, and Asp145. In addition, 14 of the predicted 27 residues are around 5 Å of the ligands in cutoff 10 Å combined 10 modes while these values were 10/22 and 11/29 in cutoff 7 Å and 13 Å combined 10 modes, respectively.

According to the ESSA result for the apo form of CDK2 enzyme (PDB ID: 4EK3), the highest number of residues with z-scores above 5 were found in cutoff 13 Å combined 10 modes similar to that of results for the holo form of CDK2 (PDB ID: 3PXZ). However, the total number of residues was found to be 34 in 4EK3 (apo), which is higher than that of the holo form (29). Interestingly, Glu81, Phe82, Leu83, Gln131, and Asp145, which are around 5 Å of the orthosteric ligand were observed in cutoffs 10 Å and 13 Å combined 10 modes while all these residues were observed in cutoffs 7 Å and 10 Å combined 10 modes groups in 3PXZ (holo). The highest number of residues in the orthosteric site, that is 9, was observed in the cutoff 13 Å combined 10 modes compared to the 3PXZ which was in the cutoff 10 Å combined 10 modes. However, 8 residues in the orthosteric site were also predicted as essential in cutoff 10 Å combined 10 modes in 4EK3 (apo). Also, cutoff 7 Å averaged over 10 modes in 4EK3 (apo) was considered more successful for binding site prediction compared to the corresponding group in 3PXZ (holo) since Tyr15, Val69, and Leu76 that are around 5 Å of the allosteric ligand-2AN (resid: A300 and A299, respectively) were found. Meanwhile, only Gln85 that interacts with the substrate through

non-bonded interaction can be found in 3PXZ for the same group. In the scope of 4EK3 (apo) results, averaged over 10 modes is not successful in predicting orthosteric site for all cutoff values while cutoffs 7 Å and 10 Å averaged over 10 modes can find at least a single residue in the orthosteric site when 3PXZ (holo) results were considered.

When the predicted residues that are around 5 Å of ligands were compared between 3PXZ (holo) and 4EK3 (apo) for their corresponding groups in ESSA, Ile63, Leu66, Glu81, Leu83, Gln85, Gln131, Asp145, and Phe146 were observed in cutoff 7 Å combined 10 modes in 3PXZ, while these residues were not found in 4EK3. Similarly, Gln85 in cutoff 7 Å averaged over 10 modes, Lys56, Val64, Leu66, Leu134, Phe146 in cutoff 10 Å combined 10 modes, Phe80 and Phe82 in cutoff 10 Å averaged over 10 modes, Val69 and His71 in cutoff 13 Å combined 10 modes, Tyr15 and Val69 in cutoff 13 Å averaged over 10 modes were predicted in 3PXZ (holo) but not found in 4EK3 (apo). On the other hand, Lys33, Val64, Val69, Leu76, and Phe80 (residues around the allosteric and orthosteric ligands) were predicted in cutoff 7 Å combined 10 modes, Tyr15, Val69, and Leu76 in cutoff 7 Å averaged over 10 modes, Lys33, His71, Leu76, Leu78, and Leu148 in cutoff 10 Å combined 10 modes, Ile35, His71, Leu76 in cutoff 10 Å averaged over 10 modes, Glu81, Leu83, His84, and Leu134 combined 10 modes, Ile35 in averaged over 10 modes in cutoff 13 Å in 4EK3 only. According to these findings, more variation between findings in the corresponding groups was in cutoff 7 Å combined 10 modes between 3PXZ (holo) and 4EK3 (apo). On the other hand, the ESSA analysis was performed for the holo form of GALK1 enzyme with the PDB ID of 6Q91. The findings were demonstrated in Table 10 in terms of the essential residues and surface representations of whole protein as well as allosteric and orthosteric sites of the protein colored according to z-score values in various cutoff values were shown in Figure 16- Figure 23.

### 6Q91 (*hGALK1*) - Cutoff 7 Å averaged over 10 modes

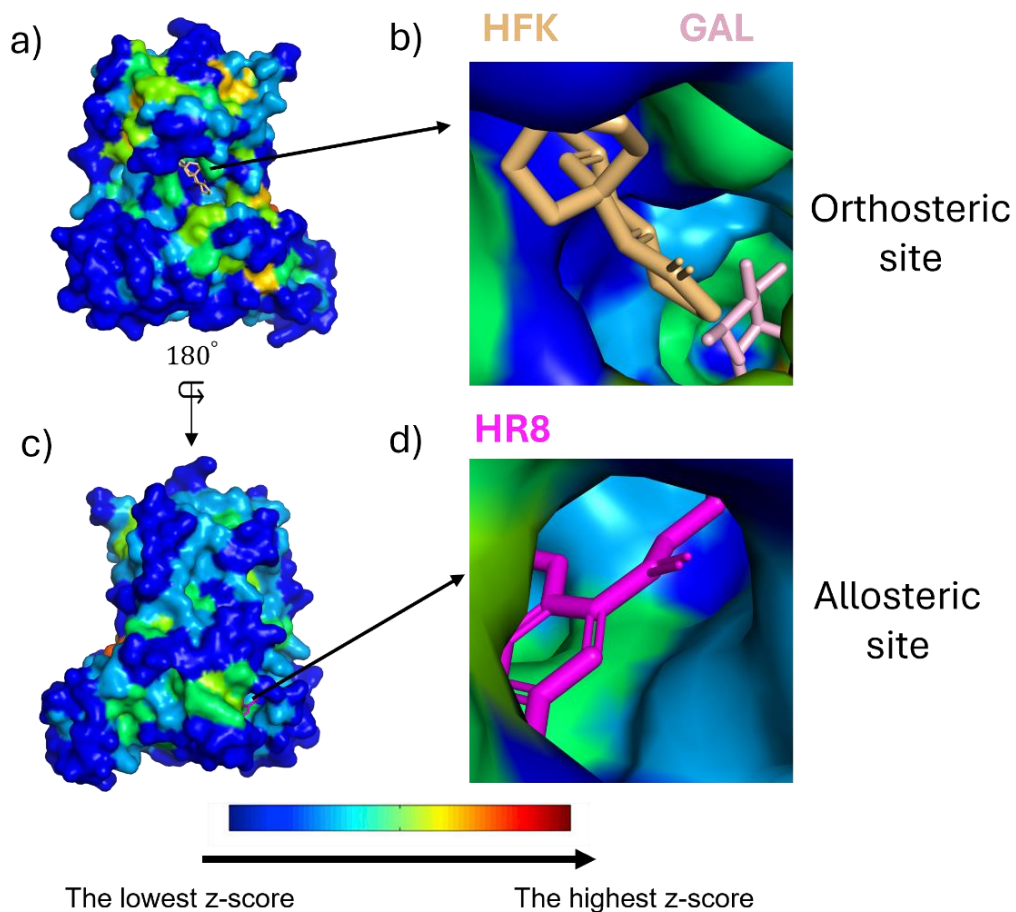


Figure 16. The ESSA results in the presence of cutoff 7 Å averaged over 10 modes for 6Q91 (*hGALK1*). a) The protein was demonstrated in surface representation and colored according to the z-score values according to the front view, b) The focus on the orthosteric site and demonstration of the ligand positions, c) The rotation of the protein by 180°, d) The focus on the allosteric site and the demonstration of the ligand HR8.

Table 10. The essential residues obtained from ESSA analysis for the holo form of GALK1 enzyme with the PDB ID of 6Q91 (\*: PDBsum hydrogen bond, \*\*: PDBsum nonbonded, \*\*\*: PDBsum both hydrogen bond and nonbonded).

Cutoff 7 Å combined 20 modes	Cutoff 7 Å averaged over 10 modes	Cutoff 10 Å combined 20 modes	Cutoff 10 Å averaged over 10 modes	Cutoff 11 Å combined 20 modes	Cutoff 11 Å averaged over 10 modes	Cutoff 13 Å combined 20 modes	Cutoff 13 Å averaged over 10 modes
LEU , 12	TYR , 47**	ARG , 37**	LEU , 92	ARG , 37**	LEU , 92	ARG , 37**	LEU , 92
PHE , 20	TYR , 236***	HIS , 44***	PRO , 93	GLU , 43***	PRO , 93	HIS , 44***	LEU , 99
ASP , 46***	ARG , 240	TYR , 47**	GLN , 115	HIS , 44***	LEU , 99	ASP , 46***	GLN , 115
TYR , 47**		ASN , 48		TYR , 47**	GLN , 115	TYR , 47**	
LEU , 92		LEU , 59		LEU , 57		PRO , 91	
TYR , 150		PHE , 90		LEU , 59		LEU , 92	
MET , 180		PRO , 91		PHE , 90		PRO , 93	
PRO , 181		LEU , 92		LEU , 92		THR , 94	
PHE , 188		PRO , 93		PRO , 93		ARG , 97	
TYR , 236***		ARG , 97		LEU , 99		SER , 98	
ARG , 239		SER , 98		TYR , 109**		LEU , 99	
ARG , 240		LEU , 99		ILE , 114		GLU , 100	
TYR , 318		LYS , 111		GLN , 115		TYR , 109**	
PHE , 348		ILE , 114		TYR , 117		ILE , 114	
		GLN , 115		PHE , 125		GLN , 115	
		PHE , 125		PRO , 134		TYR , 116	
		PRO , 134		LEU , 135**		LEU , 135**	
		LEU , 135**		LEU , 139		GLN , 171	
		LEU , 139		TYR , 150		GLU , 174	
		PHE , 152		PHE , 152		HIS , 175	
		GLN , 171		LEU , 153		MET , 180	
		MET , 180		MET , 180		PRO , 181	
		PRO , 181		PRO , 181		CYS , 182**	
		CYS , 182**		CYS , 182**		ILE , 184	
		ILE , 184		ILE , 184		MET , 185**	
		MET , 185**		MET , 185**		ASP , 186***	
		ASP , 186***		ASP , 186***		GLN , 187	
		PHE , 188		PHE , 188		PHE , 188	
		TYR , 236***		TYR , 236***		TYR , 236***	
		ARG , 239		ARG , 239		ARG , 239	
		ARG , 240		ARG , 240		ARG , 240	
		ARG , 256		ARG , 277		ARG , 277	
		ARG , 277		HIS , 280		HIS , 280	
		HIS , 280		ARG , 287		GLU , 284	
		TYR , 318		PHE , 303**		TYR , 318	
		TYR , 379		HIS , 311			
				TYR , 318			
				ARG , 342			
				TYR , 379			
Total: 14	Total: 3	Total: 36	Total: 3	Total: 39	Total: 4	Total: 35	Total: 3
Ratio: 4/14	Ratio: 2/3	Ratio: 10/36	Ratio: 0/3	Ratio: 13/39	Ratio: 0/4	Ratio: 12/35	Ratio: 0/3
GAL	HFK	HR8					

### 6Q91 (*hGALK1*) – Cutoff 7 Å combined 20 modes

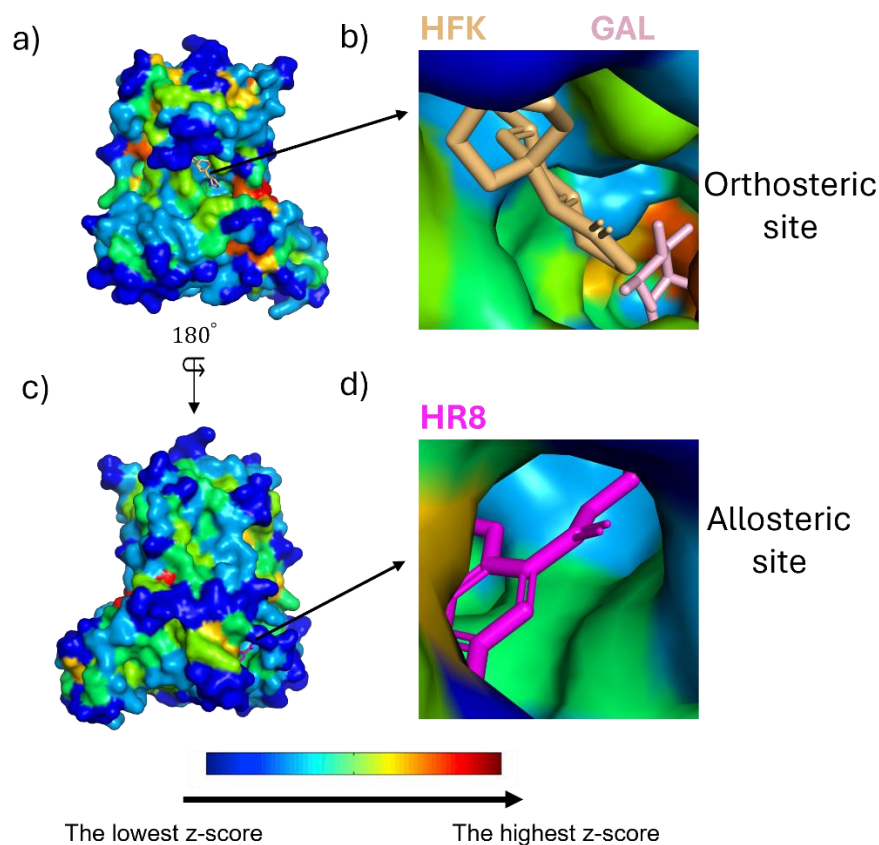


Figure 17. The ESSA results in the presence of cutoff 7 Å combined 20 modes for 6Q91 (*hGALK1*). a) The protein was demonstrated in surface representation and colored according to the z-score values according to the front view, b) The focus on the orthosteric site and demonstration of the ligand positions, c) The rotation of the protein by 180°, d) The focus on the allosteric site and the demonstration of the ligand HR8.

6Q91 (*hGALK1*) – Cutoff 10 Å averaged over 10 modes

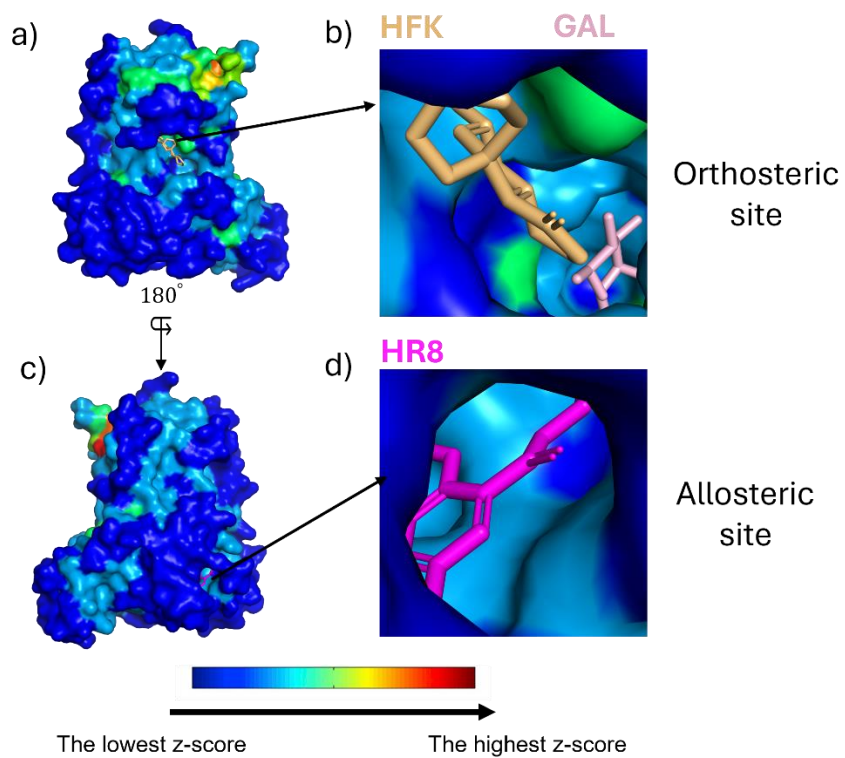


Figure 18. The ESSA results in the presence of cutoff 10 Å averaged over 10 modes for 6Q91 (*hGALK1*). a) The protein was demonstrated in surface representation and colored according to the z-score values according to the front view, b) The focus on the orthosteric site and demonstration of the ligand positions, c) The rotation of the protein by 180°, d) The focus on the allosteric site and the demonstration of the ligand HR8.

### 6Q91 (*hGALK1*) – Cutoff 10 Å combined 20 modes

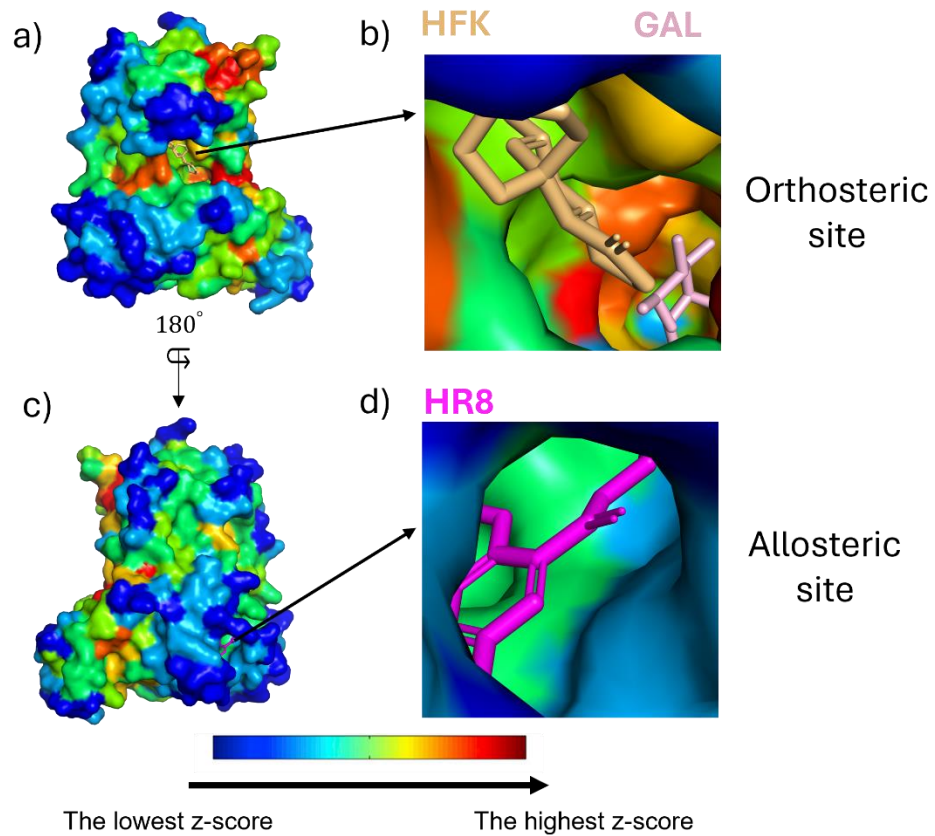


Figure 19. The ESSA results in the presence of cutoff 10 Å combined 20 modes for 6Q91 (*hGALK1*). a) The protein was demonstrated in surface representation and colored according to the z-score values according to the front view, b) The focus on the orthosteric site and demonstration of the ligand positions, c) The rotation of the protein by 180°, d) The focus on the allosteric site and the demonstration of the ligand HR8.



### 6Q91 (*hGALK1*) – Cutoff 11 Å averaged over 10 modes

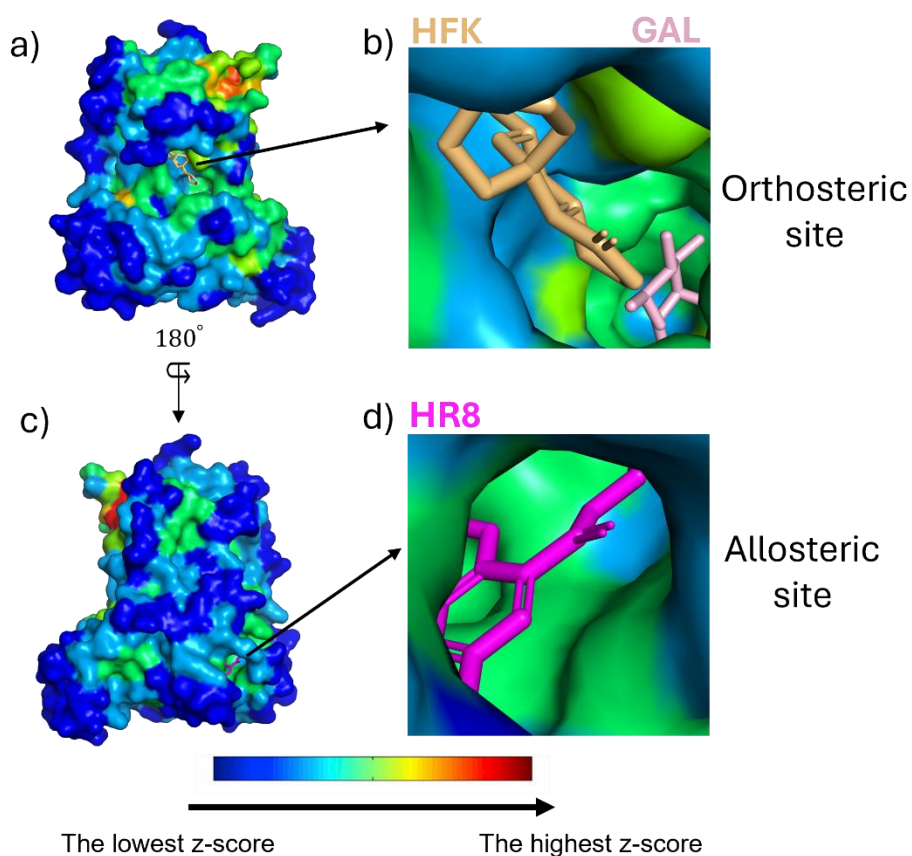


Figure 20. The ESSA results in the presence of cutoff 11 Å averaged over 10 modes for 6Q91 (*hGALK1*). a) The protein was demonstrated in surface representation and colored according to the z-score values according to the front view, b) The focus on the orthosteric site and demonstration of the ligand positions, c) The rotation of the protein by 180°, d) The focus on the allosteric site and the demonstration of the ligand HR8.

6Q91 (*hGALK1*) – Cutoff 11 Å combined 20 modes

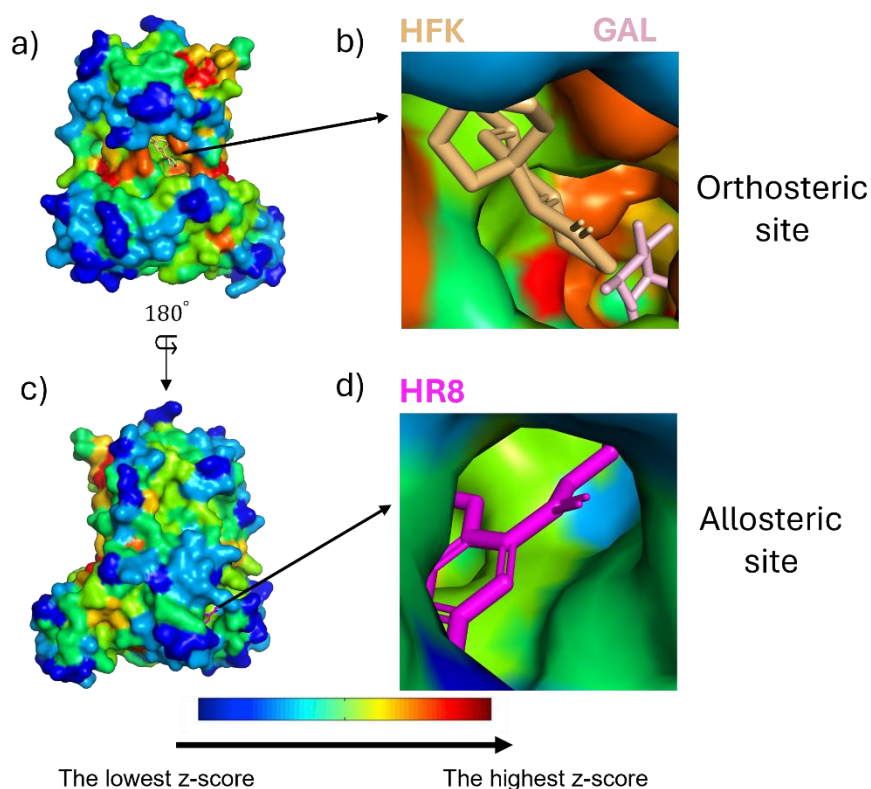


Figure 21. The ESSA results in the presence of cutoff 11 Å combined 20 modes for 6Q91 (*hGALK1*). a) The protein was demonstrated in surface representation and colored according to the z-score values according to the front view, b) The focus on the orthosteric site and demonstration of the ligand positions, c) The rotation of the protein by 180°, d) The focus on the allosteric site and the demonstration of the ligand HR8.

**6Q91 (*hGALK1*) – Cutoff 13 Å averaged over 10 modes**

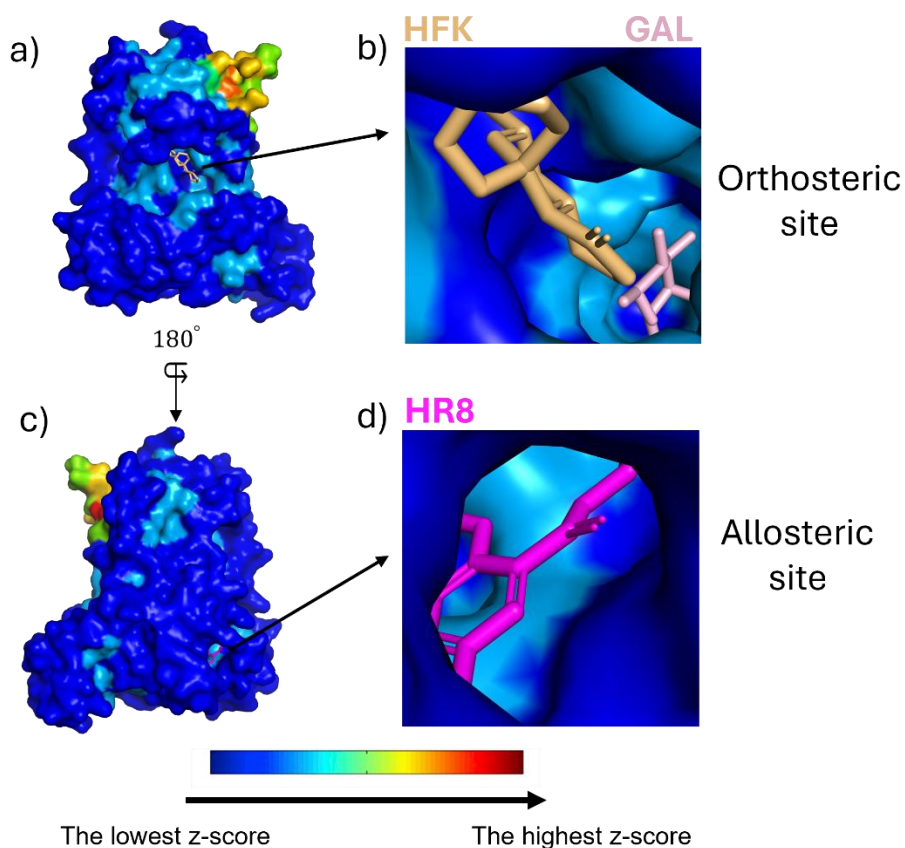


Figure 22. The ESSA results in the presence of cutoff 13 Å averaged over 10 modes for 6Q91 (*hGALK1*). a) The protein was demonstrated in surface representation and colored according to the z-score values according to the front view, b) The focus on the orthosteric site and demonstration of the ligand positions, c) The rotation of the protein by 180°, d) The focus on the allosteric site and the demonstration of the ligand HR8.

### 6Q91 (*hGALK1*) – Cutoff 13 Å combined 20 modes

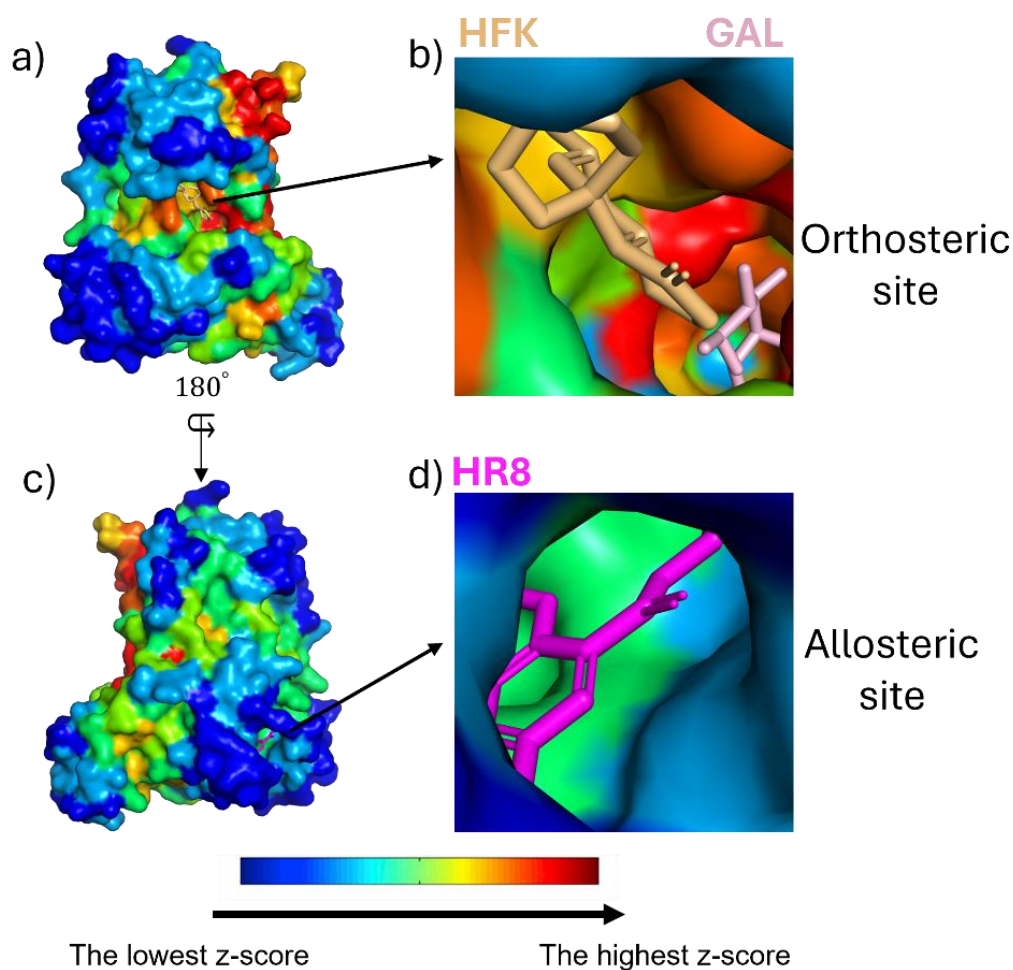


Figure 23. The ESSA results in the presence of cutoff 13 Å combined 20 modes for 6Q91 (*hGALK1*). a) The protein was demonstrated in surface representation and colored according to the z-score values (dark blue: the lowest z-score, red: the highest z-score) according to the front view, b) The focus on the orthosteric site and demonstration of the ligand positions, c) The rotation of the protein by 180°, d) The focus on the allosteric site and the demonstration of the ligand HR8.

According to the ESSA result for the holo form of GALK1 enzyme (PDB ID: 6Q91), averaged over 10 modes was not successful in predicting the residues around 5 Å of the ligands except the cutoff value of 7 Å. On the other hand, the residues around 5 Å of the fragment HR8 were only found in cutoff 11 Å combined over 20 modes. In this

scope, Phe303, which is involved in non-bonded interactions with the fragment HR8, was predicted as essential. On the other hand, several residues around 5 Å of the HFK were predicted in cutoffs 10 Å, 11 Å, and 13 Å combined 20 modes groups. To this extent, Tyr109 and Leu135, which are involved in non-bonded interactions with HFK, were noted in cutoff 11 Å and 13 Å combined 20 modes, while Tyr109 was not observed in cutoff 10 Å combined 20 modes. On the other hand, His44, Asp186, and Tyr236, which are involved in both hydrogen bond and non-bonded interactions with the substrate, were predicted in cutoffs 10 Å, 11 Å, and 13 Å combined 20 modes. Also, the highest number of residues with the z-scores 6 and 7 were observed in the cutoff 11 Å combined 20 modes group with a value of 39. Among these residues, 13 residues in total were located around 5 Å of the ligands. On the other hand, three residues were found with the z-scores above 5 in cutoff 7 Å averaged over 10 modes, while two of them Tyr47 and Tyr236 are known to interact with the substrate.

The ESSA method was used for the orthosteric, allosteric, and cryptic site prediction throughout the thesis, applying various cutoff values and combined or averaged modes. In order to compare the findings, two different servers, namely PASSer and GRaSP, which provide the prediction of ligand binding sites, were applied to the GALK1 structure with the PDB ID of 6Q91 (Tian et al. 2023; Santana et al. 2020). PASSer is used to determine potential allosteric regions using various machine-learning models (Tian et al. 2023). To this extent, the ensemble learning model was selected to predict allosteric sites of GALK1, the structure with the PDB ID of 6Q91, by only giving the PDB file without ligands in the input file. The selected model is time-efficient (1-2 seconds for a single job).

The web interface of the PASSer provides three pocket representations with the highest probabilities of being an allosteric region and residue information. To this extent, 46.36%, 32.83%, and 31.07% probabilities were obtained for three pockets, respectively. The three pockets and the position of ligands were demonstrated in Figure 24 and Figure 25.

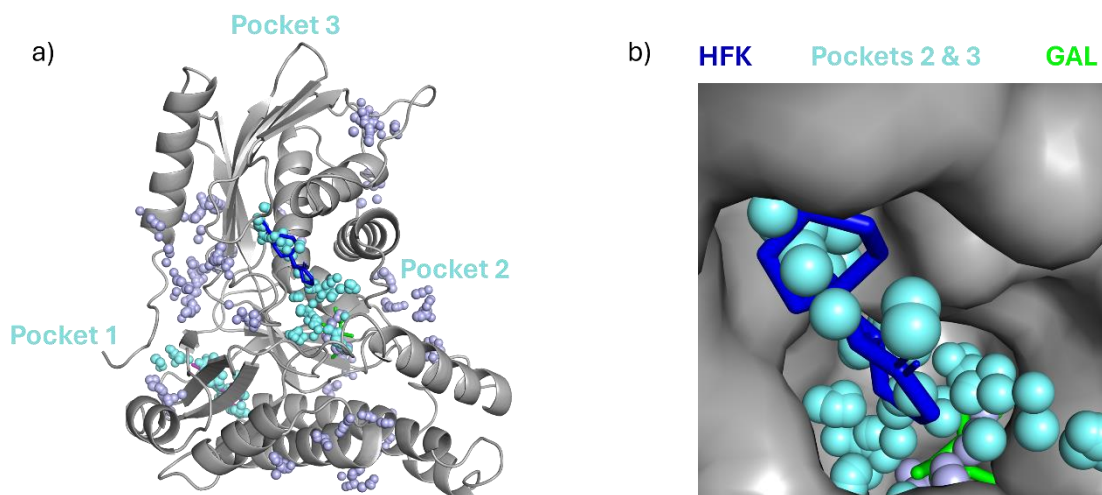


Figure 24. PASSer ensemble mode result for the orthosteric site in the holo form of GALK1 (PDB ID: 6Q91). A) Three pockets obtained from PASSer, and B) orthosteric site.

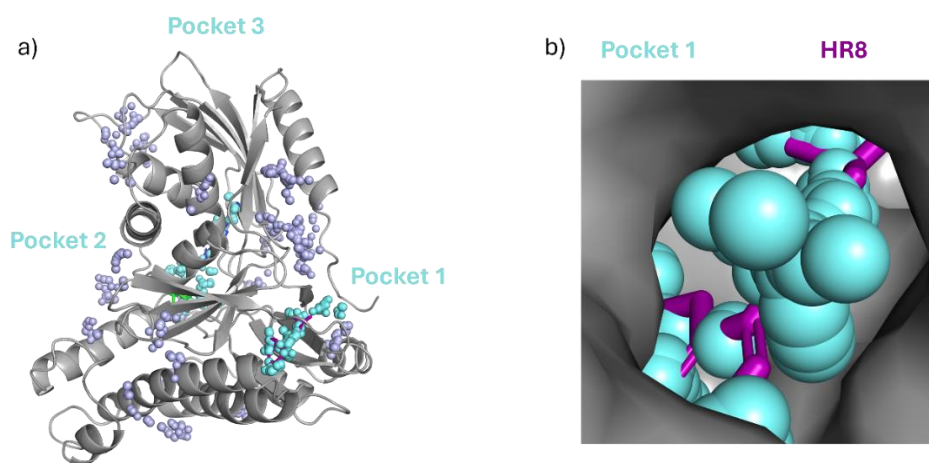


Figure 25. PASSer ensemble mode result for the allosteric fragment e binding site in GALK1 (PDB ID: 6Q91). A) Three pockets obtained from PASSer, and B) allosteric fragment e (HR8) binding site.

The probabilities in PASSer provide information about whether the region is allosteric or not (Tian et al. 2023). The residues in pocket 1, 2, and 3 that PASSer predicted were given in Table 11.

Table 11. The residues in the pockets found by PASSer.

The Residues in The Pocket 1	The Residues in The Pocket 2	The Residues in The Pocket 3
Ala358	Asp186	Tyr109
Asp299	Ser141	Asp83
Ala219	Met180	Ser141
Leu380	Ser142	Gly136
Pro216	Gly138	Thr61
Val353	Cys182	Gly81
Leu355	Ala143	Ser142
Ser381	Ser140	Leu145
Met307	Gly349	Arg105
Pro212	Glu174	Ser131
Gln382	Gly179	Ser79
Leu40	Gly347	Ala82
Phe303	Tyr109	Leu135
Asp215	Gly136	Val129
Leu218	Ala232	Thr77
Arg296	Gly346	Trp106
Ser214	Phe348	
Val220	Tyr236	
Met55	Arg37	
Ala383	Ser233	
Leu295	Ala178	
Leu213	Gly350	
Tyr300	Arg228	
Gly298		

According to these findings, PASSer can find the binding sites for both the substrate, HFK, and the fragment HR8. In addition, the output file was investigated for the rest of the predicted pockets. It should be emphasized that PASSer uses Fpocket, which is a pocket detection tool that provides the prediction of ligand binding sites in protein structures (Tian et al. 2023; Le Guilloux, Schmidtke, and Tuffery 2009). Kaynak et al. also reported the usage of Fpocket with ESSA to increase the ligand binding site prediction success (Kaynak et al. 2020). However, in the scope of this thesis, ESSA was aimed to be used without a requirement for an additional tool by changing cutoff values or combined mode options to increase the success of ligand binding site prediction.

As a second comparison tool, GRaSP was applied to the GALK1 structure with the PDB ID of 6Q91 to predict binding site residues. The time required for the prediction was calculated by the server as 18.99 seconds, which is slower than that of PASSer. In addition to the prediction of binding site residues, the server provided the potential ligands from templates as NG1 (Template: 2A2C), ANP (Template: 2A2D), ANP (Template: 1WUU), ANP (Template: 2AJ4). Among these structures, templates 2A2C and 2A2D belong to the N-acetylgalactosamine kinase while 1WUU and 2AJ4 belong to the galactokinase. In the predicted binding site residues, the residues with the highest confidence value (equal to or above 80% confidence) were determined as Arg37 (93%), Asp46 (84%), Ser141 (88%), Ser142 (84%), Asp186 (92%), Gly345 (80%), and Gly346 (90%). Among these residues, Arg37, Asp46, and Asp186 are known to interact with the substrate (GAL). In addition, Gly345 and Gly346 are around 5 Å of the substrate, and Gly346 interacts with the substrate through non-bonded interactions. Also, Ser141 is involved in hydrogen bond and non-bonded interactions with HFK, while Ser142 only interacts through non-bonded interactions with HFK according to PDBsum. However, GRaSP was not successful in predicting the allosteric fragment binding site. On the other hand, potential cryptic pockets in the GALK1 were also investigated using PocketMiner. According to the color scale that represents the probability of being a cryptic site that ranges from 0 to 1, the highest probability of being a cryptic site was demonstrated in red while the lowest probability was shown in dark blue for each residue in the structure (Meller et al. 2023). The obtained result for the GALK1 was given in Figure 26.



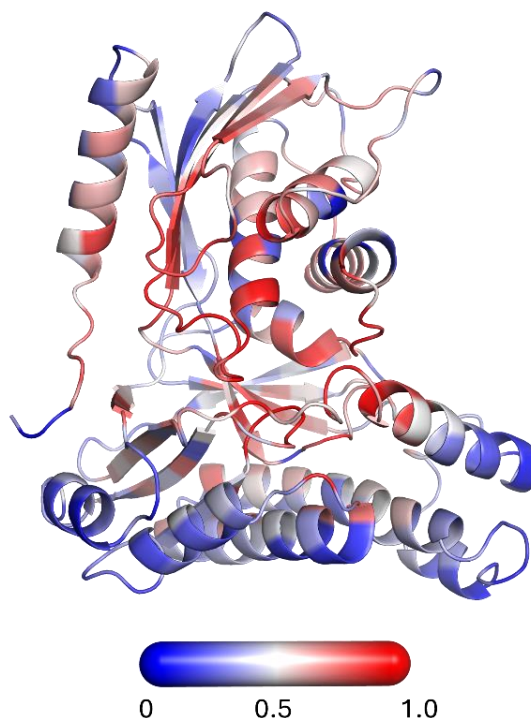


Figure 26. PocketMiner result for GALK1 (PDB ID: 6Q91). The color code demonstrates probability.

According to findings, PocketMiner can successfully predict most of the residues in the HFK and substrate binding site for the structure with the PDB ID of 6Q91. On the other hand, Leu295 in the fragment binding site was also detected by PocketMiner.

Among these results, PASSer was the most successful in predicting the fragment binding site (PocketMiner predicted a single residue in this site), while both tools were successful in predicting HFK and the substrate binding site. In addition, GRaSP also predicted residues around HFK and substrate. However, it was not successful in detecting the fragment binding site. In the scope of time comparison, 1-to-10 modes can be calculated in 5 minutes for the 6Q91 structure in the ESSA method (It should be emphasized that the required time depends on the computational power. PASSer uses high-performance computing in the calculations) while the other tools require less time. When the outputs of ESSA and other tools (PocketMiner and GRaSP) were considered, especially for the fragment binding site, one of the questions that arose was the involvement of protein dynamics in the binding site prediction since both evaluations

were performed using a single crystal structure. To this extent, only a single residue (Phe303) in the fragment binding site (allosteric site) was predicted in the 6Q91 structure using ESSA in the presence of cutoff 11 Å combined 20 modes.

The potential of increasing success in the binding site prediction with the ESSA method was considered by involving dynamics and performing analysis in the presence of various cutoff values and combined modes. In this scope, GALK1 dynamics was further investigated by performing classical all-atom MD simulations as well as ClustENMD to increase conformational sampling and investigate allosteric changes in GALK1. The further sections present MD results and the application of ESSA to a set of distinct frames from MD and ClustENMD.

## **3.2. MD Results for GALK1**

Classical (conventional) MD simulations (200 ns x 3) were separately performed for the apo and holo forms of GALK1, and 30000 frames in total were collected for each form. The RMSD, RMSF, and  $R_g$  profiles for the MD simulations of substrate and substrate-inhibitor (fragment)-bound *h*GALK1 enzyme were provided in this section.

### **3.2.1. The RMSD Profiles of *h*GALK1 (Alpha-carbon Protein RMSD, Substrate RMSD, Fragment RMSD)**

The RMSD profiles for the protein part in the *h*GALK1 three different 200 ns MD trajectories were calculated using  $\alpha$ -carbon atoms after aligning each frame into the initial structure and provided in Figure 27. According to the RMSD values of the substrate and substrate-inhibitor-bound *h*GALK1 groups, the equilibration of the system around 0.25 nanometers (nm) was observed after 80 ns (vertical dashed line) except the first substrate-bound simulation (galk1\_1) (Figure 27), which was fluctuating around 0.35 nm. Therefore, the first 80 ns of each trajectory was taken as the equilibration part, and RMSF, and  $R_g$ , were performed by including the last 120 ns of the trajectories.

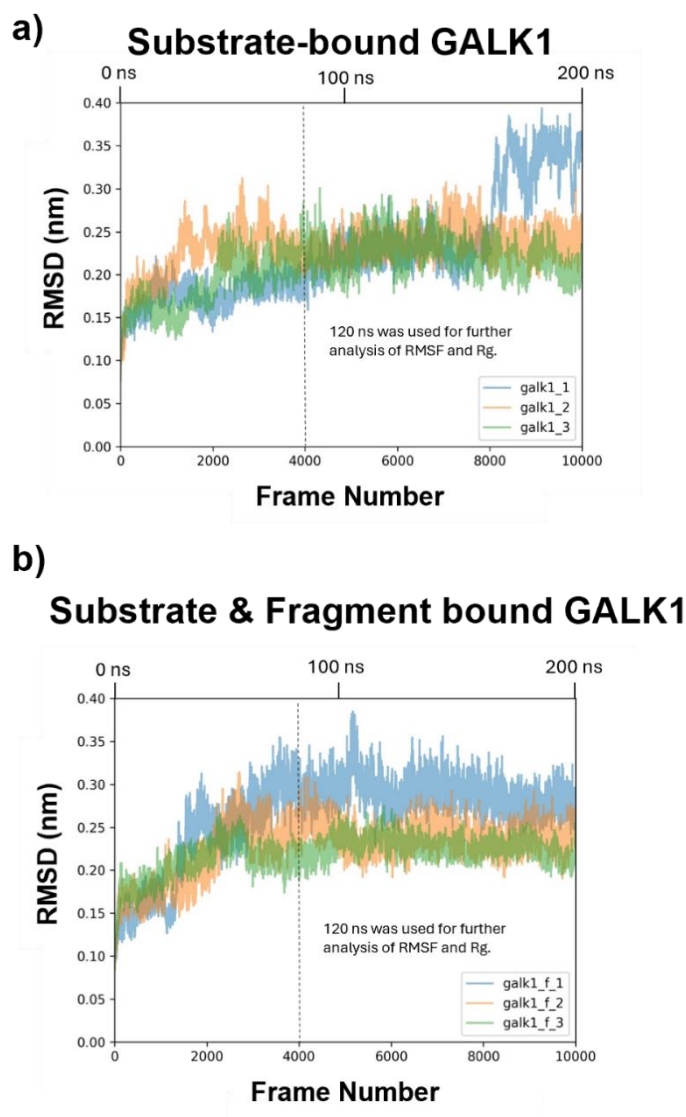


Figure 27. The RMSD profiles of *h*GALK1. The RMSD comparison according to the  $\alpha$ -carbon of a) Substrate-bound and b) Substrate-fragment-bound *h*GALK1 MD simulation groups (galk1\_*i*): Substrate-bound *h*GALK1 simulation; galk1\_f\_*i*): substrate and fragment-bound *h*GALK1 simulation; *i* values indicate the simulation number as 1, 2, and 3).

In addition, ligand RMSD plots were demonstrated in Figure 28 for both the substrate and fragment. According to the results, the substrate was observed stable in both substrate and substrate-fragment-bound MD simulation groups. On the other hand, fragment RMSD plots demonstrated several frequent fluctuations around 0.2 to 1.2 Å overall while fewer fluctuations were observed in the third run of the substrate-fragment-

bound MD simulation group (galk1\_f\_3). Thus, it was concluded that the fragment at the allosteric site was not as stable as the substrate, and the complex structure of the *hGALK1* with the compound(s) developed using the fragment should be also investigated using MD simulations.

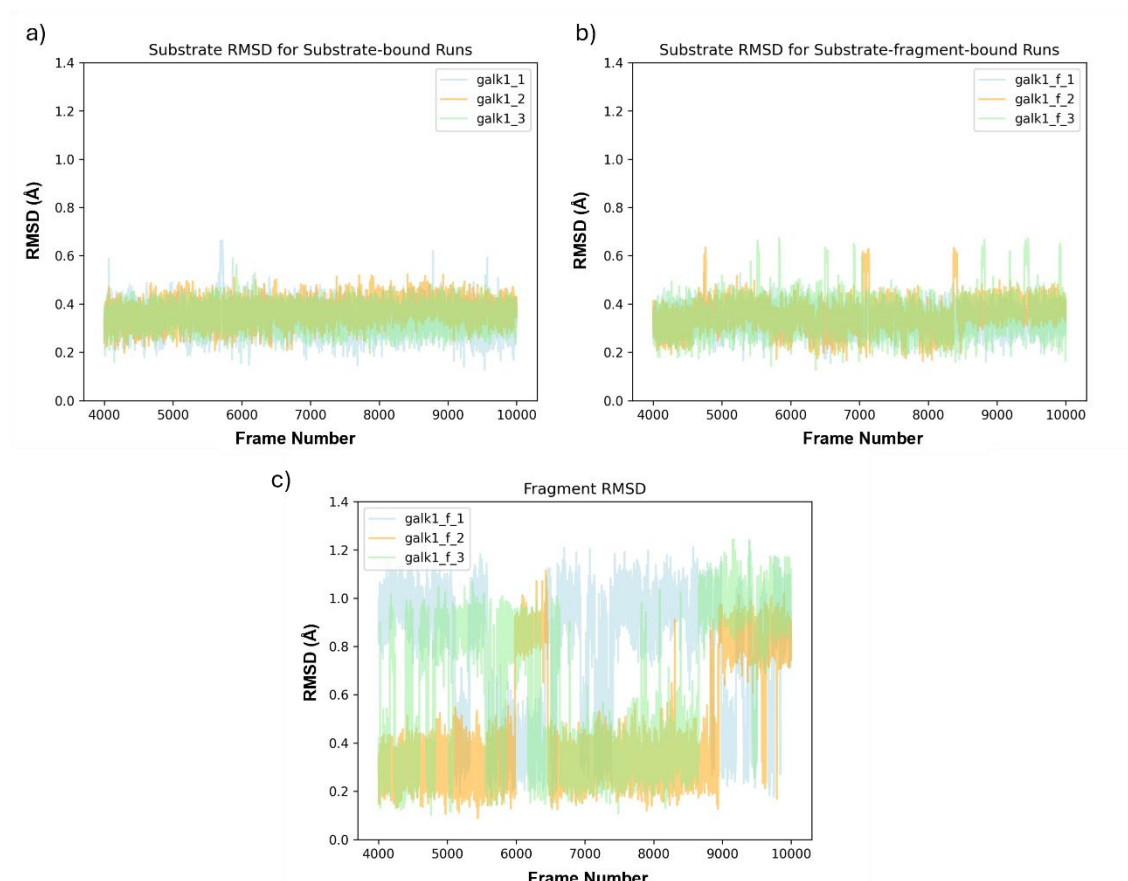


Figure 28. The RMSD profiles of the ligands in the *hGALK1* structure. The substrate RMSD comparison according to the a) Substrate-bound and b) Substrate-fragment-bound *hGALK1* MD simulation groups. c) The fragment RMSD comparison according to the substrate-fragment-bound *hGALK1* MD simulation groups (galk1\_*i*): Substrate-bound *hGALK1* simulation; galk1\_f\_*i*): substrate and fragment-bound *hGALK1* simulation; *i* values indicate the simulation number as 1, 2, and 3).

### 3.2.2. The RMSF Profiles of *h*GALK1 (Alpha-carbon Protein)

The RMSF values according to the  $\alpha$ -carbon atoms were demonstrated for the substrate-bound and substrate-inhibitor-bound *h*GALK1 simulations in Figure 29a and Figure 29b, respectively. As expected, the N- and C-terminus showed high flexibility in all simulations. It was observed that the RMSF profiles in only substrate-bound simulations do not differ much (Figure 29a). However, there are slight differences between substrate-inhibitor (fragment)-bound simulations as seen in Figure 29b. Specifically, the residues 228, 229, 232, and 233 in the second run (*galk1\_f\_2*) have higher RMSF values that are 0.34 nm for both 228 and 229, 0.36 and 0.32 nm for residues 232 and 233, respectively, compared to the other simulations. Among these residues, Arg228 was considered significant since this residue is located around 5 Å of the ligand HFk and involved in non-bonded interaction with HFk according to the PDBsum. On the other hand, Arg228 and Ser233 were observed with z-scores above 5 in ESSA result of frame 212 obtained from ClustENMD in the presence of cutoff 7 Å combined 20 modes, which was also shown in Section 3.4. When the substrate-bound and substrate-inhibitor-bound simulations were compared in terms of RMSF profiles, the RMSF of 0.2 nm was taken as a threshold as shown in the figure with a thick black line. Residues with RMSF above or equal to 0.2 nm were given as a list in Appendix A (Table 32) for both substrate-bound and substrate-inhibitor-bound *h*GALK1 MD simulation groups. The RMSF value of the residue M180 that is in the active site was less than 0.2 nm in all substrate-inhibitor-bound *h*GALK1 simulations (Figure 29b and c), whereas the residue was much more flexible in the substrate-bound simulations. The decrease in the flexibility of the residue in the active site might be due to the binding of the allosteric inhibitor. On the other hand, the residues with  $\text{RMSF} \geq 0.2$  nm in the second substrate-inhibitor-bound *h*GALK1 were shown on the 3D structure of *h*GALK1 in Figure 29d.

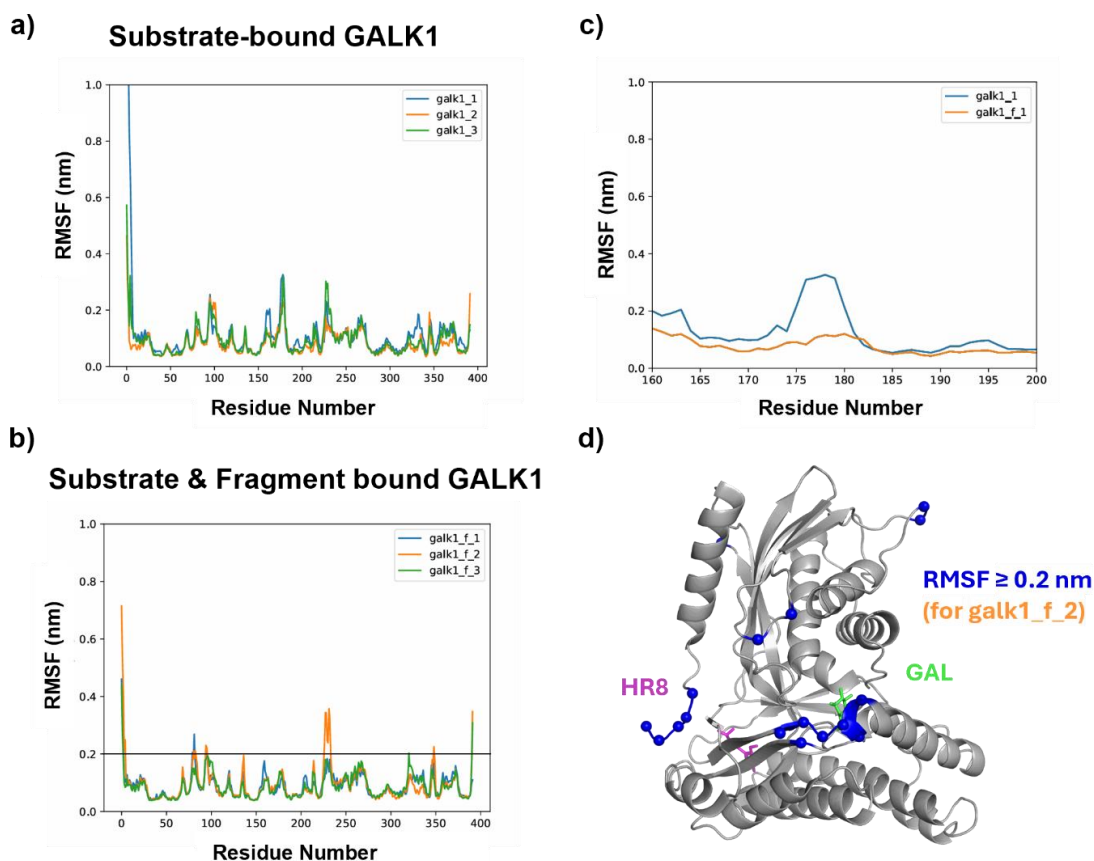


Figure 29. The RMSF profiles of *hGALK1*. The comparison was done according to the  $\alpha$ -carbon of a) Substrate-bound b) Substrate-inhibitor (fragment)-bound *hGALK1* MD simulation groups and c) The RMSF comparison of the first substrate and substrate-inhibitor-bound MD simulations for the residues in the interval 160-200. d) The residues with  $\text{RMSF} \geq 0.2$  nm in the second substrate-inhibitor-bound *hGALK1* were shown in dark blue on the protein structure. The substrate and inhibitor were demonstrated in green and purple, respectively, in the same structure (galk1\_*i*): Substrate-bound *hGALK1* Simulation; galk1\_f\_*i*): Substrate and inhibitor-bound *hGALK1* Simulation; *i* values indicate the simulation number as 1, 2, and 3).

However, there are residues that showed high flexibility only in the fragment-bound simulations. The RMSF value higher than 0.2 nm was observed in the first simulation of substrate-inhibitor-bound *hGALK1* for residue A82, which is located around 5 Å of HFK. Similarly, the residue D83, which is involved in non-bonded

interaction with HFK according to the PDBsum, had high flexibility in the second simulation of the substrate-inhibitor-bound *hGALK1*. The RMSF values were lower than 0.2 nm for both A82 and D83 in the substrate-bound *hGALK1* MD simulation groups (Table 32).

### 3.2.3. The Radius of Gyration ( $R_g$ ) Profiles of *hGALK1*

The radius of gyration values of the substrate-bound and substrate-inhibitor-bound *hGALK1* trajectories were provided as histogram plots in Figure 30a and Figure 30b, respectively. The analysis was done by considering three simulations together in each group. The  $R_g$  values varied from 2.10 to 2.18 nm in the presence of the substrate, while these values were observed around 2.10 to 2.21 nm in the presence of both the substrate and inhibitor. Interestingly, there was a variation in the  $R_g$  range when the substrate-inhibitor-bound simulations were compared to each other. To this extent, 2.14-2.21 nm, 2.11-2.18 nm, and 2.10-2.16 nm of  $R_g$  ranges were observed for the first, second, and third substrate-inhibitor-bound simulations, respectively. On the other hand, 2.11-2.17 nm, 2.10-2.17 nm, and 2.11-2.18 nm of  $R_g$  ranges were found for the first, second, and third substrate-bound simulations, respectively. According to these findings, the binding of an inhibitor did not significantly change the radius of gyration values of *hGALK1*. However, there is more variation in  $R_g$  between the three replicas when the inhibitor is bound compared to values obtained when the inhibitor is removed. The  $R_g$  coverage of the second and the third runs with the inhibitor was found to be similar to the  $R_g$  of the inhibitor-removed runs. Only the  $R_g$  of the inhibitor-bound first run was found above the  $R_g$  of the inhibitor-removed runs. This might be due to inadequate ligand parameterization, or the fragment binding might affect the stability of the enzyme.

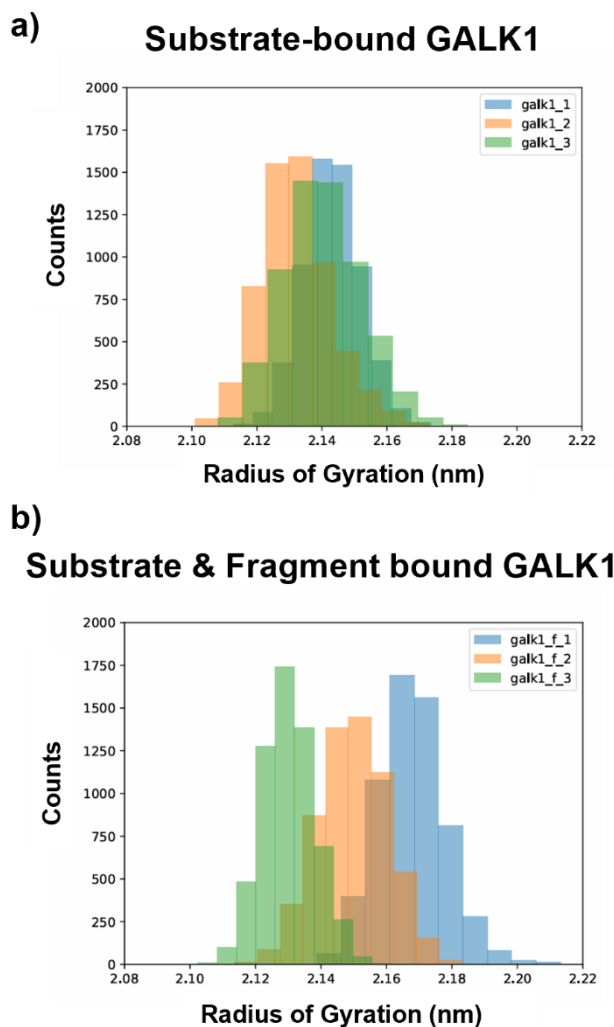


Figure 30. The radius of gyration ( $R_g$ ) profiles of *hGALK1*. The comparison was done according to the  $\alpha$ -carbon of a) Substrate-bound b) Substrate-inhibitor-bound *hGALK1* MD simulation groups (galk1\_*i*): Substrate-bound *hGALK1* simulation; galk1\_f\_*i*): Substrate and inhibitor-bound *hGALK1* simulation; *i* values indicate the simulation number as 1, 2, and 3).

### 3.2.4. Principal Component Analysis (PCA) Results

Principal Component Analysis (PCA), an unsupervised machine learning method, was applied to *hGALK1* MD simulation trajectories to extract modes (principal components-PCs) that most describe the *hGALK1* dynamics. The first five modes were



used to compare the MD data of substrate- and substrate-inhibitor-bound *h*GALK1 groups.

The PCA analysis was performed by combining three substrate-bound simulation data that include frames after the equilibration (3999<sup>th</sup>-10,000<sup>th</sup> frames). In addition, the alignment was performed through backbone atoms, and only the carbon-alpha atoms were used in the PCA calculations. The same analysis was done also for the substrate-fragment-bound dataset. In addition, PCA analysis was performed using MDAAnalysis and MDTraj packages, and each frame was transformed by projecting through principal components. The vector representations of modes that were obtained from PCA analysis were generated using the “modevectors.py” script using PyMOL. The vector set that forms each principal component was amplified by multiplying with various coefficient values to clearly observe the direction of the movement.

According to the findings, a collective motion was only observed in PC1 through PC3 among these five principal components of both substrate-bound and substrate-fragment-bound-GALK1, but not observed in PC4 and PC5. The vector representation was demonstrated from the front view only for PC1, PC2, and PC3 for both substrate-bound and substrate-fragment-bound-GALK1 groups in Figure 31, while the side and bottom views of GALK1 protein were shown for PC2 and PC3, respectively, in Figure 32.

According to the obtained results, substrate binding/unbinding movement was observed in PC1 in both groups (Figure 31). The movement in the bottom part of the protein (substrate and fragment binding region) was observed as compatible with each other in two groups, while the compatibility was not observed between the two groups for the movement in the upper part of the protein in the scope of PC1. On the other hand, the movement in the ATP binding site was not observed in PC1 when the inhibitor was bound. However, this movement was observed in the only substrate-bound group. Also, the collective rotational motion was observed especially in the bottom part of the protein when PC2 modes were compared, and protein was analyzed from a side view in only the substrate-bound group. However, this rotational motion was observed to form in a less collective manner in the substrate-fragment-bound group. At the same time, when the strong reverse rotational movement of the bottom and upper parts of the protein with respect to each other in PC3 of the substrate-bound group was considered, this reverse rotational movement was observed in lesser amount in the substrate-fragment-bound group (Figure 32). These kinds of variations in the modes belong to the substrate-bound

and substrate-fragment-bound groups, or the disappearance of a mode might be considered to be due to the allosteric effect of the inhibitor.

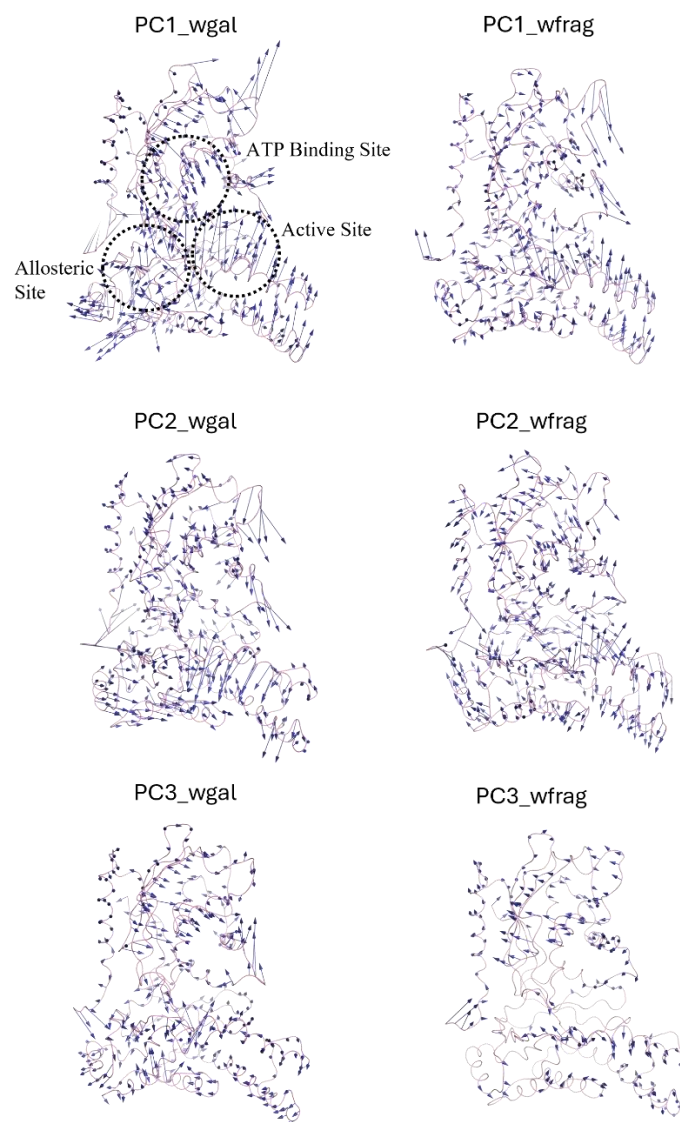


Figure 31. The representation of principal component (PC) modes obtained from PCA analysis of *hGALK1* (front view). The protein was demonstrated in light pink color as a cartoon. The dark blue arrows represent the movement direction of alpha-carbon atoms in the corresponding principal component (PC1: Principal component 1, PC2: Principal component 2, PC3: Principal component 3, wgal: Substrate-bound *hGALK1*, wfrag: Substrate-inhibitor (fragment)-bound *hGALK1*).

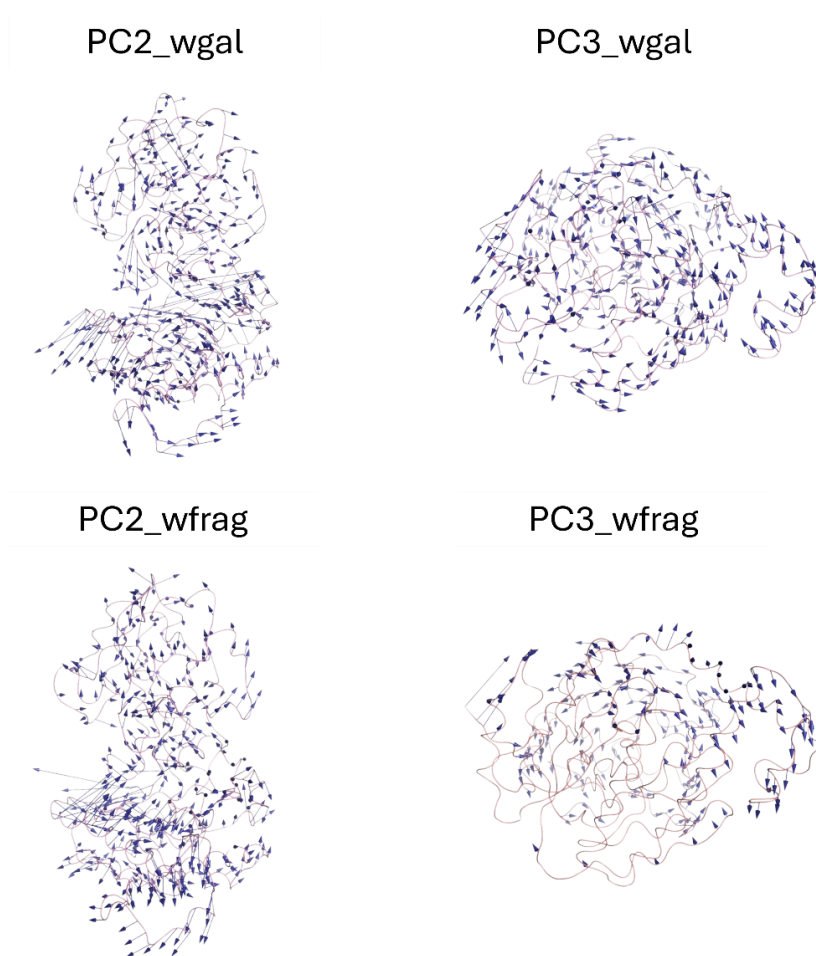


Figure 32. The representation of principal component (PC) modes obtained from PCA analysis of *hGALK1*. PC2 was demonstrated from the side view and PC3 was shown from the bottom view. The protein was demonstrated in light pink color as a cartoon. The dark blue arrows represent the movement direction of alpha-carbon atoms in that principal component (PC2: Principal component 2, PC3: Principal component 3, wgal: Substrate-bound *hGALK1*, wfrag: Substrate-inhibitor (fragment)-bound *hGALK1*).

In conclusion, these results demonstrate the significant differences between the substrate-bound and substrate-fragment-bound *hGALK1* dynamics through MD simulations and PCA analysis. These findings provide insights into the effect mechanism of an allosteric inhibitor bound to the *GALK1*.

### 3.3. The Application of ESSA to The Selected MD Frames of *h*GALK1

As shown in section 3.1.2, the ESSA analysis for the X-ray crystal structure of *h*GALK1 detected the substrate and HFK binding site, especially in cutoff 10 Å, 11 Å, and 13 Å combined 20 modes. However, ESSA, even using different parameter values, was not successful in detecting the fragment binding site (Only a single residue (Phe303) can be detected in cutoff 11 Å combined 20 modes). The possible effect of the protein conformation on ESSA results was investigated by applying ESSA to different frames selected from the trajectories. For this aim, first, the residues around 5 Å of the fragment were determined in all MD simulation groups, and then their backbone RMSD profiles were plotted concerning the initial structure as given in Figure 46 through Figure 51 in the Appendix. The MD frames were selected within the frames obtained after the equilibration that have high RMSD values. Frames with RMSD > 2 Å were listed in Table 12.

Table 12. A list of MD frames with high RMSD values with respect to the initial galactokinase 1 (GALK1) structure. The RMSD profiles for residues around 5 Å of the fragment were considered (The calculations were done in VMD).

Group	Frame Number	RMSD <sub>binding site</sub> Value (Å)
GALK1_1	3574	2.33
GALK1_2	1817	2.32
GALK1_3	<b>6700</b>	<b>2.20</b>
GALK1_F_1	<b>5143</b>	<b>2.78</b>
	<b>5198</b>	<b>2.82</b>
	<b>7116</b>	<b>2.72</b>
GALK1_F_2	4204	2.22
GALK1_F_3	8139	2.32

To this extent, the 5143<sup>rd</sup>, 5198<sup>th</sup>, and 7116<sup>th</sup> frames were selected with the RMSD values of 2.78 Å, 2.82 Å, and 2.72 Å, respectively, from the first MD run of substrate-fragment-bound *h*GALK1. These three frames were selected since higher RMSD values were observed in the fragment binding site compared to the other frames in the list. In addition, frame number 6700 with the RMSD value of 2.20 Å was selected from the third run of substrate-bound GALK1 in order to compare both substrate-bound and substrate-fragment-bound groups. These MD frames were used for the ligand binding site prediction using ESSA in the presence of various cutoff values including 7 Å, 10 Å, 11 Å, and 13 Å as well as combined 20 modes and averaged over 10 modes. Similarly to the investigation of the ligand binding site in the crystal structure of GALK1 (PDB ID: 6Q91), the essential residues with high z-score values of 6 and 7 were determined for each group. The residues were analyzed according to their involvement around the 5 Å of the ligands and in the interactions provided by PDBsum. In addition, the essential residues around the 5 Å of the ligands with z-score values of 6 and 7 were compared with the corresponding group in the ESSA results obtained from 6Q91. The different residues obtained from 6Q91 were shown in bold in each group. The ESSA results for MD frames were demonstrated in Table 13 - Table 16.

The findings were also demonstrated in the surface representation of the orthosteric and allosteric sites of the protein and stick representations of the ligands in the presence of various cutoff values. The obtained ESSA results for the combined 20 modes were shown in Figure 33 - Figure 36 for frames 5143, 5198, 6700, and 7116, respectively.

Frame 5143 (*hGALK1*) – Combined 20 modes

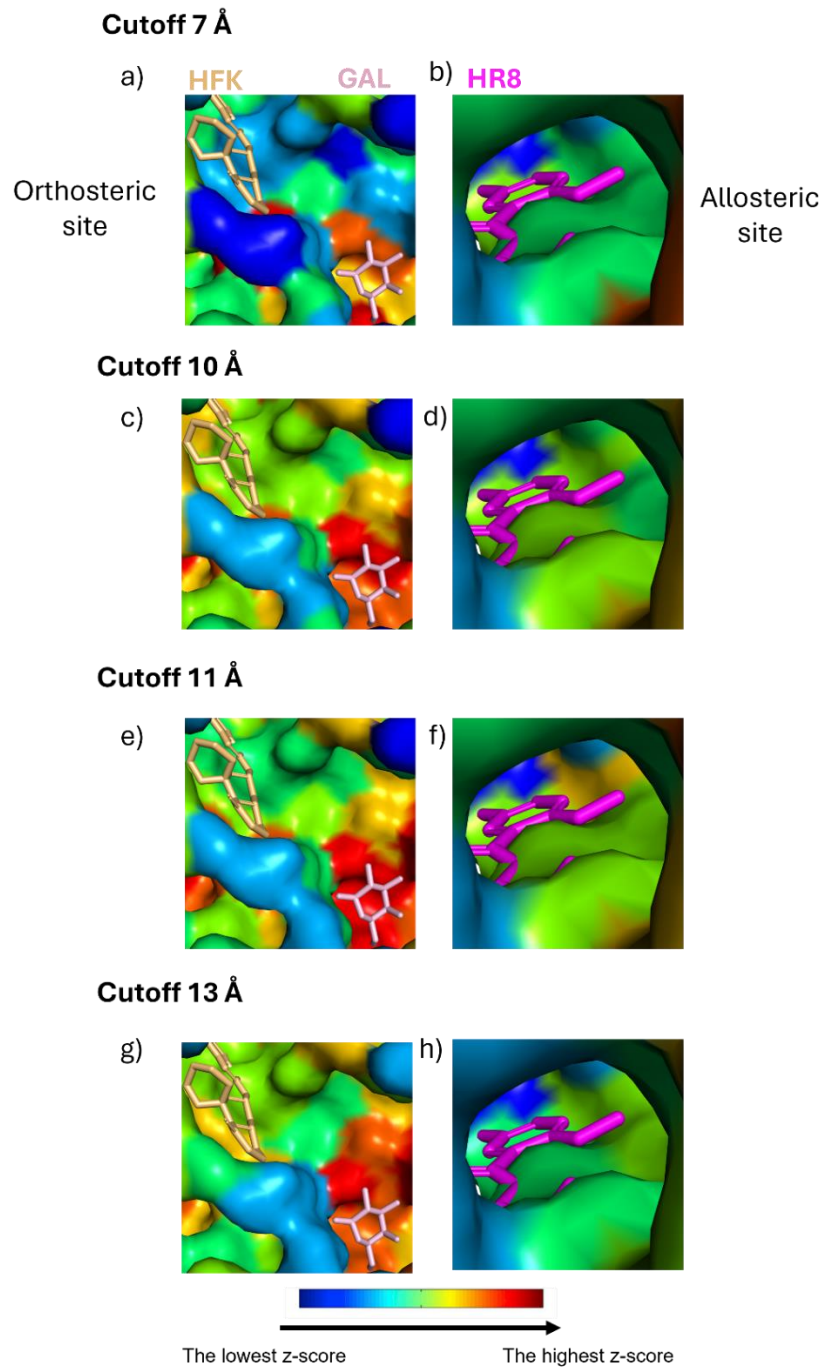


Figure 33. The ESSA results for frame 5143 (*hGALK1*). The results were shown in the presence of combined 20 modes where a) and b) cutoff 7 Å; c) and d) cutoff 10 Å; e and f) cutoff 11 Å, g and h) cutoff 13 Å. The orthosteric (left column) and allosteric ligands (right column) were demonstrated in sticks with corresponding colors and protein in surface representation colored according to z-score values in the color spectrum.

Table 13. The essential residues obtained from ESSA analysis for the GALK1 molecular dynamics (MD) frame 5143 (Bold: Not found in the 6Q91 result for the corresponding group, \*: PDBsum hydrogen bond, \*\*: PDBsum nonbonded, \*\*\*: PDBsum both hydrogen bond and nonbonded).

Cutoff 7 Å combined 20 modes	Cutoff 7 Å averaged over 10 modes	Cutoff 10 Å combined 20 modes	Cutoff 10 Å averaged over 10 modes	Cutoff 11 Å combined 20 modes	Cutoff 11 Å averaged over 10 modes	Cutoff 13 Å combined 20 modes	Cutoff 13 Å averaged over 10 modes
PRO, 7	LEU, 12	PRO, 7	<b>TRP, 106**</b>	LEU, 4	LEU, 92	ARG, 37**	LEU, 92
VAL, 9	PHE, 20	ARG, 37**	GLN, 115	ARG, 5	PRO, 93	ASN, 39	LEU, 99
LEU, 12	PHE, 24	VAL, 38	PHE, 188	ARG, 37**	LEU, 99	GLU, 43***	GLN, 115
LEU, 13	PRO, 28	ASN, 39	ILE, 201	VAL, 38	<b>TRP, 106**</b>	HIS, 44***	
ARG, 17	VAL, 62	LEU, 40**		ASN, 39	GLN, 115	ASP, 46***	
PHE, 20	VAL, 64	GLU, 43***		ILE, 41	PHE, 188	PHE, 90	
PHE, 24	LEU, 76	HIS, 44***		GLU, 43***		PRO, 91	
PRO, 28	ARG, 87	VAL, 52		HIS, 44***		LEU, 92	
ALA, 31	<b>TRP, 106**</b>	LEU, 57		VAL, 52		PRO, 93	
ARG, 37**	VAL, 128	LEU, 59		LEU, 57		ARG, 97	
VAL, 38	<b>VAL, 130</b>	LEU, 63		GLU, 58		LEU, 99	
GLU, 43***		LEU, 75		LEU, 59		<b>TRP, 106**</b>	
PRO, 54		PHE, 90		PHE, 90		ILE, 114	
LEU, 57		LEU, 92		LEU, 92		GLN, 115	
GLU, 58		PRO, 93		PRO, 93		TYR, 116	
LEU, 59		<b>TRP, 106**</b>		THR, 94		PHE, 125	
MET, 60		VAL, 110		SER, 98		TYR, 150	
VAL, 62		LYS, 111		LEU, 99		PHE, 152	
VAL, 64		ILE, 114		<b>TRP, 106**</b>		ILE, 184	
LEU, 76		GLN, 115		LYS, 111		MET, 185**	
ARG, 87		PHE, 125		ILE, 114		ASP, 186***	
LEU, 92		LEU, 139		GLN, 115		PHE, 188	
LEU, 99		<b>LEU, 145</b>		PRO, 134		TYR, 236***	
<b>TRP, 106**</b>		TYR, 150		LEU, 139		ARG, 279	
GLN, 115		PHE, 152		ILE, 184		HIS, 280	
VAL, 128		ILE, 184		MET, 185**		GLU, 284	
<b>VAL, 130</b>		MET, 185**		PHE, 188		ARG, 287	
LEU, 139		PHE, 188		ILE, 222		TYR, 318	
<b>MET, 185**</b>		ILE, 201		HIS, 280		ARG, 342	
PHE, 188		ARG, 239		GLU, 284		TYR, 379	
ILE, 189		GLU, 284		ARG, 287			
ILE, 201		HIS, 311		HIS, 311			
HIS, 311		ARG, 315		TYR, 318			
ARG, 342		TYR, 318		ARG, 342			
PHE, 348		ARG, 342		PHE, 378			
<b>GLN, 382</b>		TYR, 379		TYR, 379			
LEU, 390							
LEU, 392							
Total: 38	Total: 11	Total: 36	Total: 4	Total: 36	Total: 6	Total: 30	Total: 3
Ratio: 6/38	Ratio: 2/11	Ratio: 9/36	Ratio: 1/4	Ratio: 7/36	Ratio: 1/6	Ratio: 10/30	Ratio: 0/3
<b>GAL</b>	<b>HFK</b>	<b>HR8</b>					

Table 14. The essential residues obtained from ESSA analysis for the GALK1 molecular dynamics (MD) frame 5198 (Bold: Not found in the 6Q91 result for the corresponding group, \*: PDBsum hydrogen bond, \*\*: PDBsum nonbonded, \*\*\*: PDBsum both hydrogen bond and nonbonded).

Cutoff7 Å combined 20 modes	Cutoff7 Å averaged over 10 modes	Cutoff10 Å combined 20 modes	Cutoff10 Å averaged over 10 modes	Cutoff11 Å combined 20 modes	Cutoff11 Å averaged over 10 modes	Cutoff13 Å combined 20 modes	Cutoff13 Å averaged over 10 modes
LEU, 4	LEU, 75	ARG, 5	PHE, 90	ARG, 37**	PRO, 93	LEU, 4	ARG, 5
PRO, 7	LEU, 99	ARG, 37**	LEU, 92	ASN, 39	GLN, 115	ARG, 5	GLU, 80
PHE, 24	TRP, 106**	VAL, 38	PRO, 93	ILE, 41		ARG, 37**	GLN, 115
ARG, 37**	PHE, 152	ASN, 39	LEU, 99	GLU, 43***		GLU, 43***	PHE, 378
GLU, 43***	PHE, 177	GLU, 43***	LYS, 111	HIS, 44***		HIS, 44***	TYR, 379
HIS, 44***	TYR, 318	HIS, 44***	ILE, 114	ASP, 46***		ASP, 46***	
GLU, 58	ARG, 342	VAL, 52	GLN, 115	LEU, 59		LEU, 59	
LEU, 59	PHE, 348	PRO, 54	PHE, 125	PHE, 90		GLU, 80	
LEU, 63	PHE, 378	LEU, 57	TYR, 150	PRO, 91		PHE, 90	
LEU, 75		LEU, 59	PHE, 152	LEU, 92		PRO, 91	
LEU, 76		PHE, 90	PHE, 188	PRO, 93		LEU, 92	
ARG, 87		LEU, 92	TYR, 318	THR, 94		PRO, 93	
PHE, 90		PRO, 93		ARG, 97		THR, 94	
LEU, 99		LEU, 99		SER, 98		ARG, 97	
TRP, 106**		TRP, 106**		LEU, 99		SER, 98	
TYR, 109**		TYR, 109**		TRP, 106**		LEU, 99	
LYS, 111		LYS, 111		ILE, 114		GLN, 115	
PHE, 125		ILE, 114		GLN, 115		VAL, 133	
VAL, 128		GLN, 115		PHE, 125		PRO, 134	
VAL, 130		PHE, 125		LEU, 139		ILE, 184	
LEU, 139		LEU, 139		TYR, 150		MET, 185**	
LEU, 145		TYR, 150		ILE, 184		ASP, 186***	
GLU, 146		PHE, 152		MET, 185**		PHE, 188	
TYR, 150		ILE, 184		PHE, 188		ASN, 224	
PHE, 152		MET, 185**		ILE, 201		TYR, 236***	
PHE, 177		PHE, 188		TYR, 236***		ARG, 256	
MET, 185**		ILE, 201		HIS, 280		ARG, 279	
GLN, 187		ARG, 239		GLU, 284		HIS, 280	
PHE, 188		HIS, 280		ARG, 287		GLU, 284	
ILE, 189		HIS, 311		HIS, 311		ARG, 287	
LEU, 199		TYR, 318		LEU, 314		LEU, 314	
ILE, 201		ARG, 342		TYR, 318		TYR, 318	
ARG, 239		TYR, 379		ARG, 342		ARG, 342	
TYR, 300**				TYR, 379		THR, 377	
HIS, 311						PHE, 378	
TYR, 318						TYR, 379	
ARG, 342							
MET, 343							
PHE, 348							
MET, 365							
GLN, 369							
PHE, 378							
TYR, 379							
LEU, 380							
GLN, 382							
Total: 45	Total: 9	Total: 33	Total: 12	Total: 34	Total: 2	Total: 36	Total: 5
Ratio: 10/45	Ratio: 1/9	Ratio: 8/33	Ratio: 0/12	Ratio: 9/34	Ratio: 0/2	Ratio: 8/36	Ratio: 0/5
GAL	HFK	HR8					

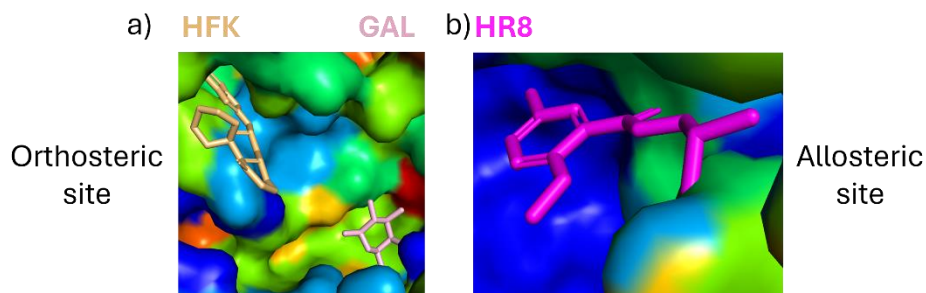




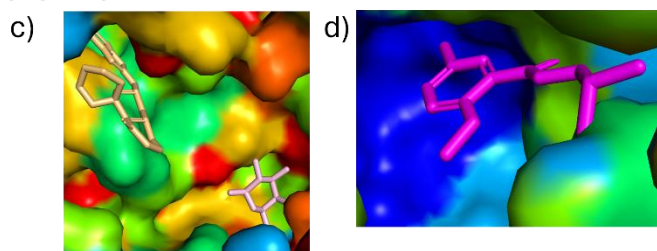


Frame 6700 (hGALK1) – Combined 20 modes

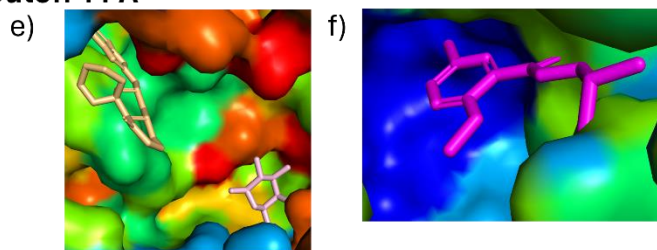
Cutoff 7 Å



Cutoff 10 Å



Cutoff 11 Å



Cutoff 13 Å

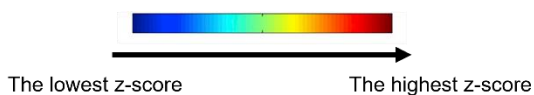
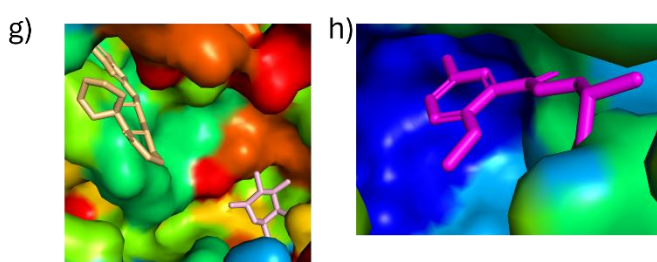


Figure 35. The ESSA results for frame 6700 (hGALK1). The results were shown in the presence of combined 20 modes where a) and b) cutoff 7 Å; c) and d) cutoff 10 Å; e and f) cutoff 11 Å, g and h) cutoff 13 Å. The orthosteric (left column) and allosteric ligands (right column) were demonstrated in sticks with corresponding colors and protein in surface representation colored according to z-score values in the color spectrum.

Table 16. The essential residues obtained from ESSA analysis for the GALK1 molecular dynamics (MD) frame 7116 (Bold: Not found in the 6Q91 result for the corresponding group, \*: PDBsum hydrogen bond, \*\*: PDBsum nonbonded, \*\*\*: PDBsum both hydrogen bond and nonbonded).

Cutoff 7 Å combined 20 modes	Cutoff 7 Å averaged over 10 modes	Cutoff 10 Å combined 20 modes	Cutoff 10 Å averaged over 10 modes	Cutoff 11 Å combined 20 modes	Cutoff 11 Å averaged over 10 modes	Cutoff 13 Å combined 20 modes	Cutoff 13 Å averaged over 10 modes
ALA, 2	LEU, 4	HIS, 1	HIS, 1	LEU, 4	LEU, 92	ARG, 37**	PRO, 93
ALA, 3	MET, 365	ALA, 2	ALA, 2	ARG, 5	PRO, 93	GLU, 43***	ARG, 97
LEU, 4	GLN, 369	ALA, 3	ARG, 5	ARG, 37**	GLN, 115	HIS, 44***	LEU, 99
ARG, 5	PHE, 378	LEU, 4	MET, 365	VAL, 38		PRO, 91	GLN, 115
GLN, 6		ARG, 5	ARG, 366	ASN, 39		LEU, 92	
PRO, 7		GLN, 6	GLN, 369	GLU, 43***		PRO, 93	
LEU, 12		ARG, 37**		HIS, 44***		THR, 94	
LEU, 59		HIS, 44***		ASP, 46***		ARG, 97	
MET, 60		ASP, 46***		LEU, 57		SER, 98	
VAL, 62		LEU, 57		LEU, 59		LEU, 99	
VAL, 64		GLU, 58		PHE, 90		GLU, 100	
<b>VAL, 130</b>		LEU, 59		LEU, 92		GLN, 115	
<b>ILE, 184</b>		LEU, 139		PRO, 93		<b>ILE, 184</b>	
ILE, 201		<b>ILE, 184</b>		THR, 94		MET, 185**	
<b>TYR, 236***</b>		MET, 185**		ARG, 97		<b>ASP, 186***</b>	
HIS, 311		PHE, 188		SER, 98		PHE, 188	
ARG, 342		ARG, 256		LEU, 99		HIS, 280	
MET, 365		HIS, 280		ILE, 114		GLU, 284	
ARG, 366		ARG, 286		GLN, 115		TYR, 318	
GLN, 369		HIS, 311		PHE, 125		ARG, 342	
ALA, 376		TYR, 318		PRO, 134		TYR, 379	
PHE, 378		ARG, 342		LEU, 139			
TYR, 379		MET, 365		<b>ILE, 184</b>			
LEU, 380		ARG, 366		MET, 185**			
		GLN, 369		PHE, 188			
		ALA, 376		<b>TYR, 236***</b>			
		TYR, 379		ARG, 239			
				ARG, 256			
				HIS, 280			
				GLU, 284			
				ILE, 285			
				ARG, 286			
				HIS, 311			
				TYR, 318			
				ARG, 342			
				MET, 365			
				ARG, 366			
				GLN, 369			
				TYR, 379			
Total: 24	Total: 4	Total: 27	Total: 6	Total: 39	Total: 3	Total: 21	Total: 4
Ratio: 3/24	Ratio: 0/4	Ratio: 5/27	Ratio: 0/6	Ratio: 8/39	Ratio: 0/3	Ratio: 6/21	Ratio: 0/4
<b>GAL</b>	<b>HFK</b>	<b>HR8</b>					

Frame 7116 (*hGALK1*) – Combined 20 modes

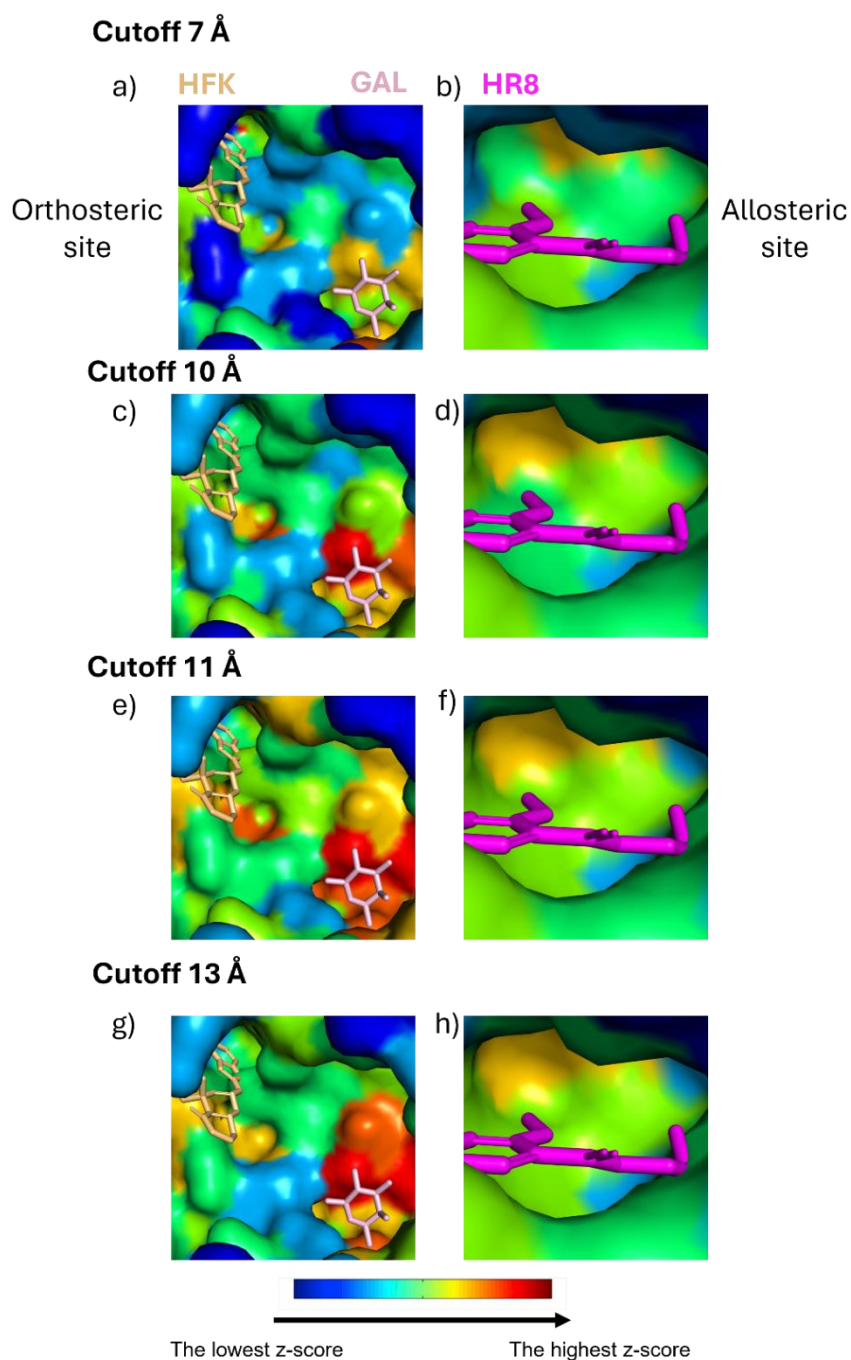


Figure 36. The ESSA results for frame 7116 (*hGALK1*). The results were shown in the presence of combined 20 modes where a) and b) cutoff 7 Å; c) and d) cutoff 10 Å; e and f) cutoff 11 Å, g and h) cutoff 13 Å. The orthosteric (left column) and allosteric ligands (right column) were demonstrated in sticks with corresponding colors and protein in surface representation colored according to z-score values in the color spectrum.

According to the ESSA result for MD frame 5143 in Table 13, the higher number of essential residues with z-scores 6 and 7 were detected in the presence of combined 20 modes in all cutoff values compared to the averaged over 10 modes. The observed total number of essential residues was 38 for cutoff 7 Å, 36 for both cutoff 10 and 11 Å, and 30 for cutoff 13 Å. In addition, one-third of the essential residues can be detected as ligand-interacting residues when the cutoff distance value is either 10 or 13 Å and combined 20 modes was used or the cutoff is 10 Å and averaged over 10 modes was used. These groups were considered significant in terms of their proportion to include the potential ligand-interacting residues. On the other hand, Trp106, which is involved in non-bonded interaction with HFK can be detected in both groups except cutoff 13 Å averaged over 10 modes. In addition, Arg37, Glu43, and Met185, which are in the orthosteric site and involved in interactions with the substrate, were monitored in all groups of combined 20 modes, while these residues were not noted in all cutoff values averaged over 10 modes. On the other hand, the allosteric fragment HR8-interacting residues Gln382 and Leu40 were only found in cutoff 7 Å combined 20 modes and cutoff 10 Å combined 20 modes for residues, respectively. When the findings from the crystal structure with the PDB ID of 6Q91 and MD frame 5143 were compared, it was observed that all groups except cutoff 13 Å averaged over 10 modes were able to find ligand-interacting residues that are not observed in the 6Q91 structure. For instance, Arg37, Glu43, Trp106, Val130, Met185, and Gln382 can be detected in the cutoff 7 Å combined 20 modes, while these residues were not identified in the same group for the 6Q91 structure. Also, Asn39, Leu40, Trp106, and Leu145 in cutoff 10 Å combined 20 modes were not viewed in the equivalent group for the 6Q91 structure.

According to the ESSA result for MD frame 5198, the highest number of essential residues were observed in the combined 20 modes in all cutoff values similar to that of MD frame 5143. Moreover, His44, which is involved in both hydrogen bond and non-bonded interactions with the substrate, was observed in all combined 20 modes groups in addition to the detected residues Arg37, Glu43, and Met185 also in the MD frame 5143. On the other hand, averaged over 10 modes were not successful for all cutoff values except cutoff 7 Å to detect the residues around 5 Å of the ligands, while only Trp106 can be found with cutoff 7 Å. The comparison of ESSA results for MD frame 5198 with 6Q91 demonstrated that Arg37, Glu43, His44, Trp106, Tyr109, Val130, Leu145, Met185, Tyr300, Gln382 can be recognized in the cutoff 7 Å combined 20 modes in the MD frame 5198, while these residues were missing in the ESSA result of 6Q91 for that group. In

addition, Asn39, Gln43, Trp106, and Tyr109 can be found in cutoff 10 Å combined 20 modes, while the condition is not valid for the same group in the ESSA result of the 6Q91. On the other hand, only cutoff 7 Å combined 20 modes is successful in finding residues Tyr300 and Gln382 that are around 5 Å of the fragment. Among these residues, Tyr300 is involved in non-bonded interaction with the fragment HR8. When the calculated ratios were compared, the highest ratios were obtained from cutoff 11 and 10 Å for combined 20 modes with ratios 9/34 and 8/33, respectively, compared to the other groups.

Interestingly, a greater number of essential residues were noted in the averaged over 10 modes groups in the MD frame 6700 compared to the other MD frames. To this extent, 22, 26, 24, and 23 residues in total were found to have z-scores 6 and 7 in cutoff 7, 10, 11, and 13 Å averaged over 10 modes, respectively, in the MD frame 6700. On the other hand, these values were 33 for both cutoff 7 and 11 Å, 42 for cutoff 10, and 40 for 13 Å. However, the residues around 5 Å of the fragment HR8 were not found in all groups even though several residues around 5 Å of the substrate and HFK can be observed in all groups for MD frame 6700. Also, cutoff 11 Å was considered more successful in both combined and averaged over 10 modes in terms of detecting residues around 5 Å of the ligands (according to the ratios of 11/33 and 8/24, respectively) compared to the other groups in MD frame 6700. In addition, Arg37, Tyr47, Arg105, Trp106, Tyr109, Ile184, Met185, and Tyr236 were noted to be found in cutoff 11 Å averaged over 10 modes, while these residues were not predicted in the ESSA result for 6Q91 in the same group. Among these residues, Arg105, Trp106, and Tyr109 were known to interact with HFK through non-bonded interactions, while the other indicated residues are around 5 Å of the substrate and demonstrate non-bonded interactions except Ile184.

The ESSA results for MD frame 7116 demonstrated a smaller number of essential residues for combined 20 modes groups in various cutoff values compared to the combined 20 modes groups in other MD frames except the group cutoff 11 Å with the 39 residues in total. According to the findings, the residues around 5 Å of the fragment HR8 were not detected in all groups for MD frame 7116. In addition, Val130, which is around 5 Å of HFK, was only found in the presence of cutoff 7 Å combined 20 modes. On the other hand, Arg37, His44, Ile184, and Met185, which are around 5 Å of the substrate, were noted in the presence of cutoff 10, 11, and 13 Å combined 20 modes, while the residues around 5 Å of ligands were not detected in averaged over 10 modes in all cutoff values. Also, cutoff 11 Å combined 20 modes were considered better for the prediction of residues that are around 5 Å of ligands when the ratios were compared in the scope of

MD frame 7116 (The ratio is 8/39 for cutoff 11 Å combined 20 modes). In addition, Asn39 and Asp46, which are around 5 Å of the substrate, can be observed in this group, while these residues were absent in the same group of 6Q91 results.

To sum up, it was observed that the conformation of the GALK1 protein had some effect on the ESSA results. Therefore, a faster method compared to MD, namely ClustENMD, was also applied to the GALK1 structure to increase the conformational sampling, and the results were provided in the following section.

### 3.4. The Application of ClustENMD to GALK1

The substrate-bound *h*GALK1 initial structure prepared using CHARMM-GUI was cleaned from ligand, ion, and water molecules. Then, ClustENMD was applied to the cleaned structure, and a total of 300 conformers were generated. The backbone RMSD profile of the conformers was determined after the backbone alignment of all frames to the input file given to the ClustENMD and provided in Figure 37. Most of the conformers have an RMSD value of less than 2 Å. However, it was observed that a few conformers deviated from the initial structure by 2.5 Å up to 4 Å.

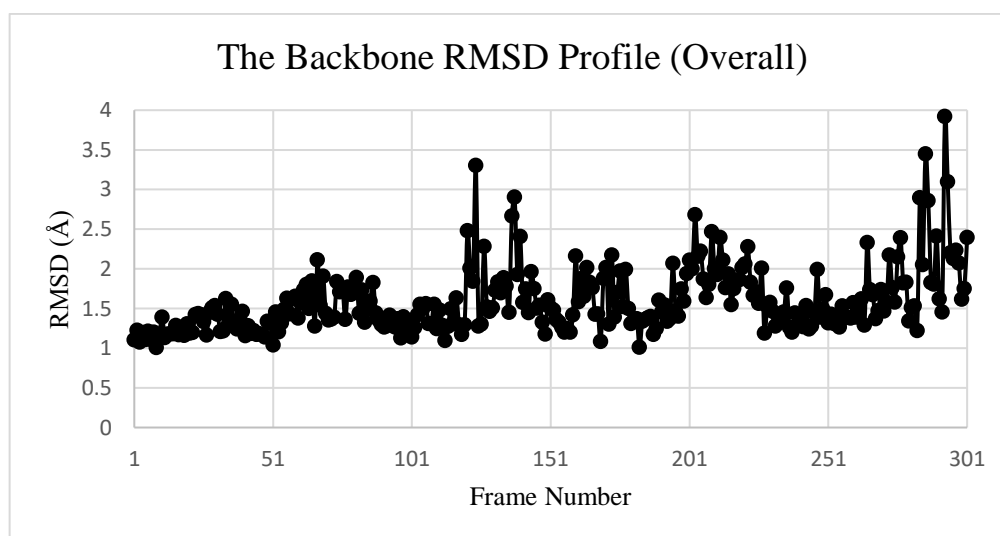


Figure 37. The Backbone RMSD Profile of the *h*GALK1 protein (Overall).



The backbone RMSD profile was also calculated for the residues around 5 Å of fragment (HR8) for 301 frames obtained from ClustENMD. The backbone RMSD profile of the residues around 5 Å of the fragment was demonstrated in Figure 38.

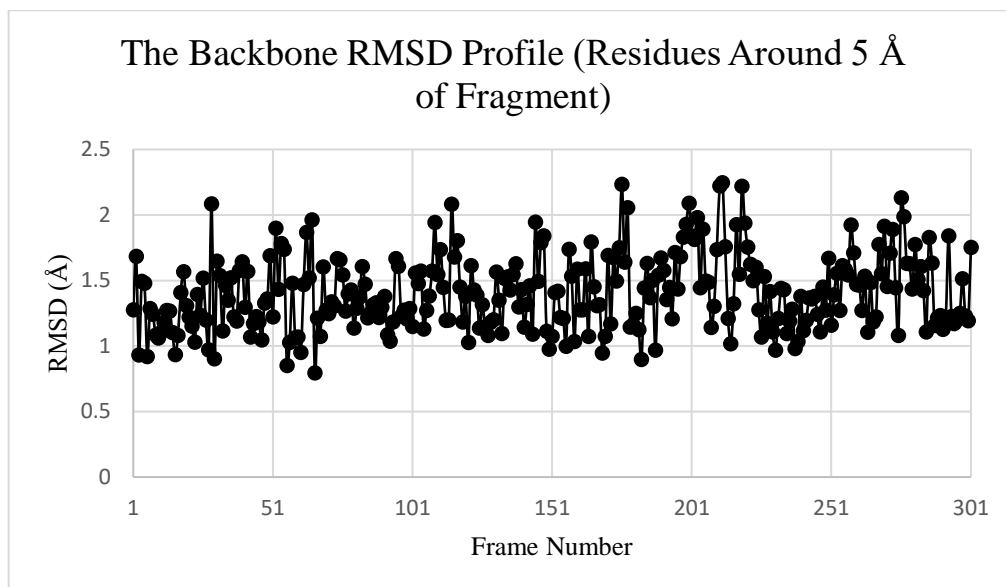


Figure 38. The backbone RMSD profile of the residues around 5 Å of the fragment in *hGALK1*.

The frames with higher RMSD values than 2 Å were determined within 301 frames for the residues around 5 Å of the fragment and listed in Table 17.

Table 17. A list of *hGALK1* frames obtained from ClustENMD with high RMSD values for the residues around 5 Å of the fragment. The selected frames with high RMSD values were demonstrated in bold.

<b>Frame Number</b>	<b>RMSD Value (Å)</b>
29	2.09

(cont. on next page)

**Table 17. (cont.).**

115	2.08
<b>176</b>	<b>2.23</b>
178	2.06
200	2.09
<b>211</b>	<b>2.22</b>
<b>212</b>	<b>2.25</b>
219	2.22
276	2.13

ESSA was applied to three out of nine frames (frames 176, 211, and 212) that have higher RMSD values to observe whether the fragment binding site in *hGALK1* can be predicted through ESSA. The potential effects of four different cutoff distance values (7, 10, 11, and 13 Å) and how modes are included in the ESSA calculation on the binding site prediction were tested for each frame. The essential residues obtained from ESSA for ClustENMD frames were demonstrated in Table 18 - Table 20. Here, the substrate-interacting, HFK-interacting, and fragment-interacting residues are highlighted as green, blue, and light orange, respectively.

The findings were also demonstrated in the surface representation of the orthosteric and allosteric sites of the protein. In addition, the position of the orthosteric and allosteric ligands was also included. The obtained ESSA results for combined 20 modes and cutoff values including 7 Å, 10 Å, 11 Å, and 13 Å were given for frames 176, 211, and 212 in Figure 39 - Figure 41, respectively.

Frame 176 (*hGALK1*) – Combined 20 modes

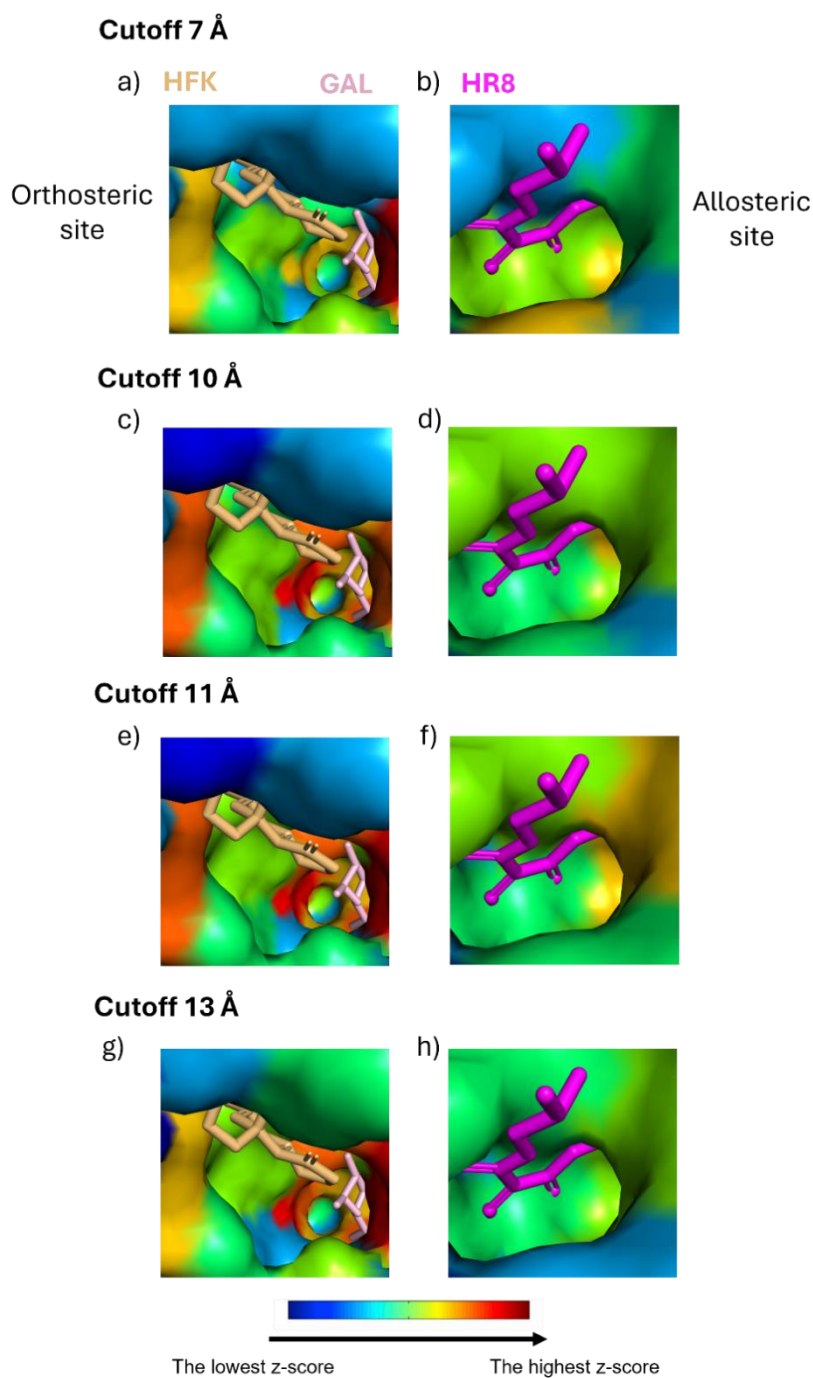


Figure 39. The ESSA results for frame 176 (*hGALK1*). The results were shown in the presence of combined 20 modes where a) and b) cutoff 7 Å; c) and d) cutoff 10 Å; e and f) cutoff 11 Å, g and h) cutoff 13 Å. The orthosteric (left column) and allosteric ligands (right column) were demonstrated in sticks with corresponding colors and protein in surface representation colored according to z-score values in the color spectrum.

Table 18. ESSA result for ClustENMD frame 176 (*hGALK1*) (Bold: Not found in the 6Q91 result for the corresponding group, \*: PDBsum hydrogen bond, \*\*: PDBsum nonbonded, \*\*\*: PDBsum both hydrogen bond and nonbonded).

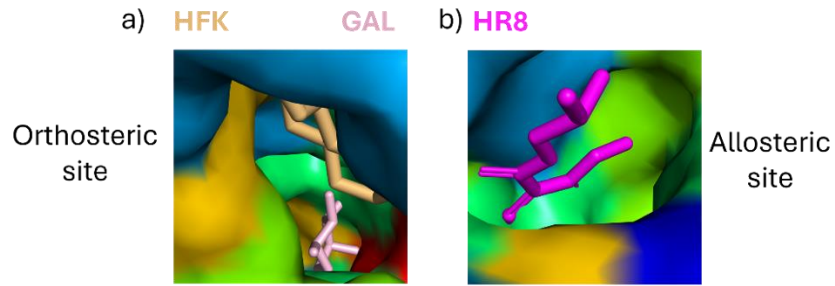
Cutoff 7 Å combined 20 modes	Cutoff 7 Å averaged over 10 modes	Cutoff 10 Å combined 20 modes	Cutoff 10 Å averaged over 10 modes	Cutoff 11 Å combined 20 modes	Cutoff 11 Å averaged over 10 modes	Cutoff 13 Å combined 20 modes	Cutoff 13 Å averaged over 10 modes
PHE, 20	PHE, 20	ARG, 37**	LEU, 92	LEU, 4	LEU, 4	ARG, 5	LEU, 92
PHE, 24	PHE, 24	HIS, 44***	PRO, 93	ARG, 5	ARG, 5	ARG, 37**	PRO, 93
TYR, 47**	TYR, 47**	TYR, 47**	LEU, 99	GLN, 6	GLN, 6	HIS, 44***	ARG, 97
ASN, 48	PHE, 90	ASN, 48	ILE, 114	ARG, 37**	LEU, 92	TYR, 47**	LEU, 99
LEU, 63	LEU, 92	PHE, 90	GLN, 115	HIS, 44***	PRO, 93	PRO, 91	GLN, 115
PHE, 90	PRO, 93	LEU, 92		TYR, 47**	LEU, 99	LEU, 92	TYR, 116
LEU, 92	LEU, 99	PRO, 93		ASN, 48	GLN, 115	PRO, 93	TYR, 379
PRO, 93	<b>TRP, 106**</b>	LEU, 99		GLN, 49	MET, 365	ARG, 97	
LEU, 99	GLN, 115	LYS, 111		PHE, 90	GLN, 369	LEU, 99	
<b>TRP, 106**</b>	PHE, 125	ILE, 114		LEU, 92	PHE, 378	GLU, 100	
GLN, 115	PHE, 152	GLN, 115		PRO, 93		TYR, 109**	
PHE, 125	TYR, 236***	PHE, 125		LEU, 99		GLN, 115	
VAL, 128	LEU, 255	LEU, 135**		GLN, 115		TYR, 116	
LEU, 139	PHE, 275	LEU, 139		PHE, 125		GLN, 171	
PHE, 152	<b>TYR, 300**</b>	PHE, 152		PRO, 134		GLU, 174	
MET, 180	TYR, 318	PHE, 177		LEU, 135**		HIS, 175	
PRO, 181	ARG, 240	MET, 180		LEU, 139		PHE, 177	
PHE, 188	ARG, 342	PRO, 181		PHE, 152		MET, 180	
TYR, 236***		CYS, 182**		GLU, 174		PRO, 181	
PRO, 237		ILE, 184		HIS, 175		CYS, 182**	
ARG, 240		MET, 185**		MET, 180		ILE, 184	
VAL, 246		ASP, 186***		PRO, 181		MET, 185**	
LEU, 250		PHE, 188		CYS, 182**		ASP, 186***	
LEU, 255		TYR, 236***		ILE, 184		PHE, 188	
PHE, 275		ARG, 239		MET, 185**		TYR, 236***	
<b>TYR, 300**</b>		ARG, 240		ASP, 186***		ARG, 240	
MET, 307		LEU, 255		PHE, 188		HIS, 280	
TYR, 318		ARG, 256		ILE, 222		TYR, 318	
ARG, 342		ARG, 277		TYR, 236***		MET, 365	
GLN, 369		TYR, 318		ARG, 239		PHE, 378	
PHE, 378		ARG, 342		ARG, 240		TYR, 379	
TYR, 379		PHE, 378		ARG, 256		LEU, 380	
		TYR, 379		ARG, 277			
				HIS, 280			
				VAL, 282			
				TYR, 318			
				ARG, 342			
				MET, 365			
				GLN, 369			
				PHE, 378			
				TYR, 379			
Total: 32	Total: 18	Total: 33	Total: 5	Total: 41	Total: 10	Total: 32	Total: 7
Ratio: 5/32	Ratio: 4/18	Ratio: 10/33	Ratio: 0/5	Ratio: 10/41	Ratio: 0/10	Ratio: 10/32	Ratio: 0/7
GAL	HFK	HR8					

Table 19. ESSA result for ClustENMD frame 211 (*hGALK1*) (Bold: Not found in the 6Q91 result for the corresponding group, \*: PDBsum hydrogen bond, \*\*: PDBsum nonbonded, \*\*\*: PDBsum both hydrogen bond and nonbonded).

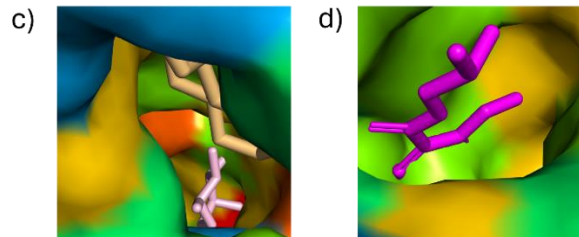
Cutoff 7 Å combined 20 modes	Cutoff 7 Å averaged over 10 modes	Cutoff 10 Å combined 20 modes	Cutoff 10 Å averaged over 10 modes	Cutoff 11 Å combined 20 modes	Cutoff 11 Å averaged over 10 modes	Cutoff 13 Å combined 20 modes	Cutoff 13 Å averaged over 10 modes
PHE, 20	PHE, 90	ARG, 37**	PHE, 90	ARG, 37**	PHE, 90	ARG, 37**	LEU, 92
PHE, 24	PRO, 93	HIS, 44***	LEU, 92	GLU, 43***	LEU, 92	<b>GLU, 43***</b>	LEU, 99
LEU, 30	LEU, 99	TYR, 47**	PRO, 93	HIS, 44***	PRO, 93	HIS, 44***	ILE, 114
<b>TYR, 47**</b>	GLN, 115	PHE, 90	LEU, 99	TYR, 47**	LEU, 99	ASP, 46***	GLN, 115
LEU, 63	PHE, 152	LEU, 92	LYS, 111	PHE, 90	ILE, 114	TYR, 47**	TYR, 116
LEU, 75	<b>TYR, 236***</b>	PRO, 93	ILE, 114	LEU, 92	GLN, 115	PRO, 91	
PHE, 90	LEU, 255	LEU, 99	GLN, 115	PRO, 93	TYR, 116	LEU, 92	
LEU, 92	LEU, 263	LYS, 111	TYR, 117	LEU, 99	TYR, 117	PRO, 93	
PRO, 93	TYR, 318	ILE, 114	PHE, 125	ILE, 114	PHE, 125	THR, 94	
LEU, 99		GLN, 115	PHE, 152	GLN, 115	TYR, 150	LEU, 99	
<b>TRP, 106**</b>		TYR, 116		TYR, 116	PHE, 152	<b>TYR, 109**</b>	
ILE, 114		TYR, 117		TYR, 117		ILE, 114	
GLN, 115		PHE, 125		PHE, 125		GLN, 115	
PHE, 125		LEU, 139		PRO, 134		TYR, 116	
VAL, 128		TYR, 150		<b>LEU, 135**</b>		TYR, 117	
PHE, 152		PHE, 152		LEU, 139		PHE, 125	
<b>MET, 180</b>		LEU, 153		TYR, 150		<b>LEU, 135**</b>	
PRO, 181		<b>MET, 180</b>		PHE, 152		TYR, 150	
<b>ILE, 184</b>		PRO, 181		LEU, 153		PHE, 152	
<b>TYR, 236***</b>		<b>ILE, 184</b>		GLU, 174		LEU, 153	
ARG, 239		<b>MET, 185**</b>		HIS, 175		GLN, 171	
ARG, 240		<b>ASP, 186***</b>		<b>MET, 180</b>		GLU, 174	
VAL, 246		PHE, 188		PRO, 181		HIS, 175	
LEU, 250		ARG, 239		<b>CYS, 182**</b>		<b>MET, 180</b>	
LEU, 255		LEU, 250		ILE, 184		PRO, 181	
VAL, 258		LEU, 263		<b>MET, 185**</b>		<b>CYS, 182**</b>	
LEU, 263		PHE, 275		<b>ASP, 186***</b>		<b>ILE, 184</b>	
PHE, 275		ARG, 277		PHE, 188		<b>MET, 185**</b>	
TYR, 318		HIS, 280		<b>TYR, 236***</b>		<b>ASP, 186***</b>	
ARG, 342		TYR, 318		ARG, 239		GLN, 187	
TYR, 379		ARG, 342		ARG, 240		PHE, 188	
		TYR, 379		ARG, 256		<b>TYR, 236***</b>	
				ARG, 277		ARG, 240	
				HIS, 280		HIS, 280	
				TYR, 318		TYR, 318	
				ARG, 342		ARG, 342	
				TYR, 379		TYR, 379	
Total: 31	Total: 9	Total: 32	Total: 10	Total: 37	Total: 11	Total: 37	Total: 5
Ratio: 5/31	Ratio: 1/9	Ratio: 7/32	Ratio: 0/10	Ratio: 11/37	Ratio: 0/11	Ratio: 13/37	Ratio: 0/5
<b>GAL</b>	<b>HFK</b>	<b>HR8</b>					

Frame 211 (*hGALK1*) – Combined 20 modes

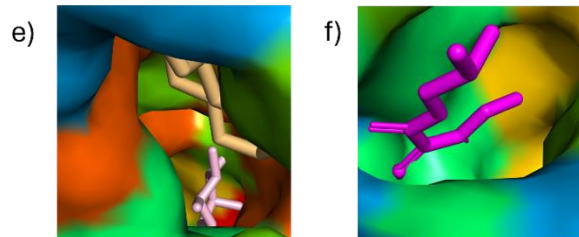
Cutoff 7 Å



Cutoff 10 Å



Cutoff 11 Å



Cutoff 13 Å

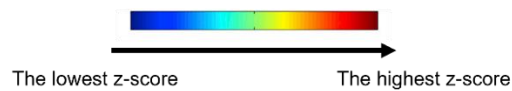
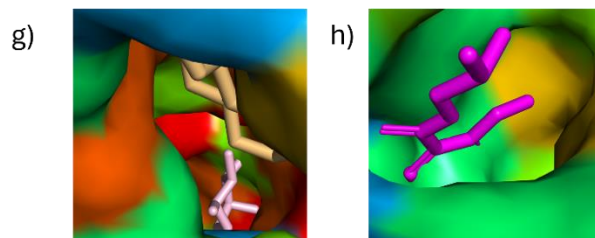


Figure 40. The ESSA results for frame 211 (*hGALK1*). The results were shown in the presence of combined 20 modes where a) and b) cutoff 7 Å; c) and d) cutoff 10 Å; e and f) cutoff 11 Å, g and h) cutoff 13 Å. The orthosteric (left column) and allosteric ligands (right column) were demonstrated in sticks with corresponding colors and protein in surface representation colored according to z-score values in the color spectrum.

Table 20. ESSA result for ClustENMD frame 212 (*hGALK1*) (Bold: Not found in the 6Q91 result for the corresponding group, \*: PDBsum hydrogen bond, \*\*: PDBsum nonbonded, \*\*\*: PDBsum both hydrogen bond and nonbonded).

Cutoff 7 Å combined 20 modes	Cutoff 7 Å averaged over 10 modes	Cutoff 10 Å combined 20 modes	Cutoff 10 Å averaged over 10 modes	Cutoff 11 Å combined 20 modes	Cutoff 11 Å averaged over 10 modes	Cutoff 13 Å combined 20 modes	Cutoff 13 Å averaged over 10 modes
ARG, 5	TYR, 47**	ARG, 37**	PRO, 93	ARG, 5	LEU, 92	ARG, 37**	PRO, 91
GLN, 6	PHE, 90	GLU, 43***		ARG, 37**	PRO, 93	GLU, 43***	LEU, 92
PRO, 7	LEU, 92	HIS, 44***		VAL, 38	LEU, 99	HIS, 44***	PRO, 93
GLN, 8	PRO, 93	TYR, 47**		ASN, 39	GLN, 115	TYR, 47**	THR, 94
LEU, 12	LEU, 99	PHE, 90		LEU, 40**		PRO, 91	ARG, 97
ARG, 37**	LEU, 218**	PRO, 91		ILE, 41		LEU, 92	LEU, 99
ILE, 41	TYR, 300**	LEU, 92		GLU, 43***		PRO, 93	GLN, 115
ASP, 46***	TYR, 318	PRO, 93		HIS, 44***		THR, 94	
TYR, 47**	ARG, 342	ARG, 97		TYR, 47**		GLN, 96	
ASN, 48		SER, 98		LEU, 59		ARG, 97	
LEU, 59		LEU, 99		PHE, 90		SER, 98	
VAL, 62		LYS, 111		LEU, 92		LEU, 99	
PHE, 90		GLN, 115		PRO, 93		GLU, 100	
LEU, 92		PRO, 134		LEU, 99		GLN, 115	
PRO, 93		LEU, 139		LYS, 111		MET, 180	
LEU, 99		MET, 180		ILE, 114		PRO, 181	
GLN, 115		PRO, 181		GLN, 115		CYS, 182**	
VAL, 130		MET, 185**		PHE, 125		ILE, 184	
LEU, 139		PHE, 188		PRO, 134		MET, 185**	
PHE, 152		TYR, 236***		LEU, 139		ASP, 186***	
MET, 180		ARG, 279		MET, 180		PHE, 188	
MET, 185**		HIS, 280		PRO, 181		TYR, 236***	
VAL, 211		GLU, 284		CYS, 182**		ARG, 239	
ASP, 215**		ARG, 287		MET, 185**		ARG, 279	
LEU, 218**		HIS, 311		PHE, 188		HIS, 280	
ARG, 228**		LEU, 314		TYR, 236***		GLU, 284	
SER, 233		TYR, 318		ARG, 239		ARG, 287	
TYR, 236***		ARG, 342		HIS, 280		TYR, 318	
ARG, 239		TYR, 379		GLU, 284		ARG, 342	
LEU, 295**				ARG, 287			
TYR, 300**				LEU, 314			
TYR, 318				TYR, 318			
ARG, 342				ARG, 342			
PHE, 348				TYR, 379			
MET, 365							
GLN, 369							
PHE, 378							
TYR, 379							
Total: 38	Total: 9	Total: 29	Total: 1	Total: 34	Total: 4	Total: 29	Total: 7
Ratio: 12/38	Ratio: 3/9	Ratio: 7/29	Ratio: 0/1	Ratio: 10/34	Ratio: 0/4	Ratio: 10/29	Ratio: 0/7
GAL	HFK	HR8					

Frame 212 (*hGALK1*) – Combined 20 modes

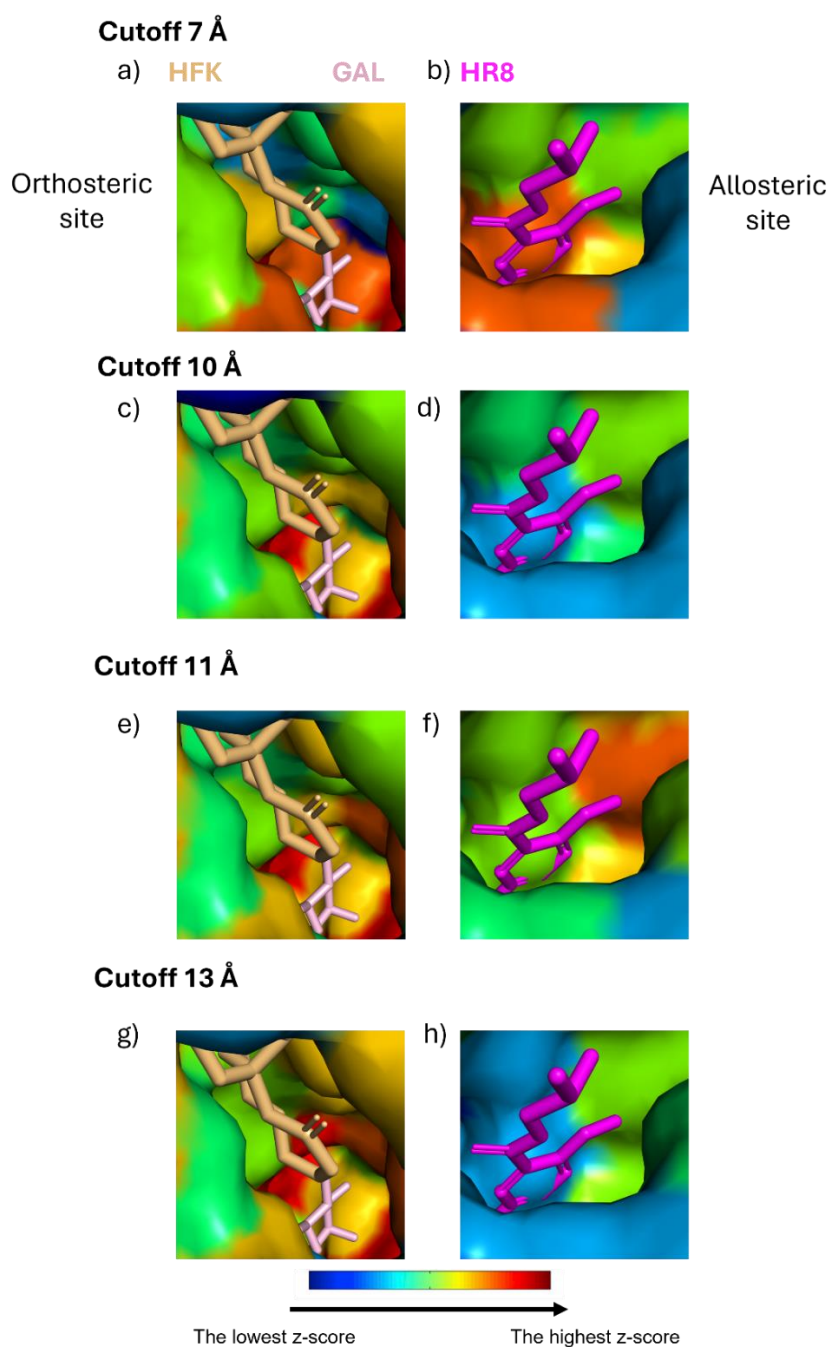


Figure 41. The ESSA results for frame 212 (*hGALK1*). The results were shown in the presence of combined 20 modes where a) and b) cutoff 7 Å; c) and d) cutoff 10 Å; e) and f) cutoff 11 Å, g) and h) cutoff 13 Å. The orthosteric (left column) and allosteric ligands (right column) were demonstrated in sticks with corresponding colors and protein in surface representation colored according to z-score values in the color spectrum.



According to the ESSA result for ClustENMD frame 176, averaged over 10 modes groups were not successful in predicting the residues around 5 Å of all ligands in cutoff 10, 11, and 13 Å, while cutoff 7 Å can find several residues around 5 Å of both allosteric and orthosteric ligands. To this extent, Tyr300, which is involved in non-bonded interaction with fragment HR8, and Trp106, which interacts with HFK through non-bonded interactions, were only observed in cutoff 7 Å in both combined 20 modes and averaged over 10 modes. In addition, these residues were not detected as essential in the 6Q91 results for the same groups. Therefore, cutoff 7 Å can be considered successful in terms of predicting the fragment binding site in frame 176 compared to other cutoff values. On the other hand, Tyr47, Met180, and Tyr236, which are around 5 Å of the substrate, can be found in cutoffs 7, 10, 11, and 13 Å combined 20 modes. Among these residues, Tyr236 is involved in hydrogen bond and non-bonded interactions with the substrate. When the ratios were compared, cutoff 13 Å combined 20 modes was considered more successful in frame 176 to predict the residues around 5 Å of ligands with the ratio of 10/32.

In the scope of the ESSA result for ClustENMD frame 211, the residues around 5 Å of the ligands were not observed in averaged over 10 modes except cutoff 7 Å which includes only the prediction of Tyr236 among the residues near the substrate. On the other hand, Tyr47, Met180, and Ile184 which are around 5 Å of the substrate were predicted in combined 20 modes for all cutoff values. However, key residues in the fragment binding site were not found in all groups. On the other hand, Trp106 which is around 5 Å of the HFK, and Ile184 which is around 5 Å of the substrate were predicted in cutoff 7 Å combined 20 modes in frame 211 while these residues were not observed in the same group of 6Q91 results. Similarly, Glu43, which is involved in both hydrogen bond and non-bonded interactions with the substrate, was found in cutoff 13 Å combined 20 modes while this residue is absent in the same group of 6Q91 results. In addition, the analysis with cutoff 13 Å combined 20 modes was considered more successful for frame 211 in predicting the residues around 5 Å of the ligands due to the ligand-interacting *vs.* essential residue ratio of 13/37.

According to the ESSA result for ClustENMD frame 212, the residues that are around 5 Å of the ligands were not found in averaged over 10 modes except cutoff 7 Å. In addition, Leu218 and Tyr300, which are involved in non-bonded interactions with the fragment HR8, were predicted in cutoff 7 Å in both combined 20 modes and averaged over 10 modes. Additionally, Asp215 and Leu295, which also interact with the fragment

through non-bonded interactions, were found in cutoff 7 Å combined 20 modes. On the other hand, Leu40, which is in the fragment binding site, was found in cutoff 11 Å combined 20 modes. These findings demonstrate the significance of both groups in cutoff 7 Å and combined 20 modes in cutoff 11 Å for the prediction of the fragment binding site compared to the other groups in frame 212 results. In addition, Val130 and Arg228, which are around 5 Å of HFK, were only observed in cutoff 7 Å combined 20 modes. Moreover, Arg37, Val130, Met185, Asp215, Leu218, Arg228, Leu295, and Tyr300 were not predicted in the cutoff 7 Å combined 20 modes of 6Q91 while these residues were found in the corresponding group of ClustENMD frame212. On the other hand, the residues that are around 5 Å of the substrate, Arg37, Tyr47, Met180, Met185, and Tyr236 were predicted in all combined 20 modes groups. Also, cutoff 13 Å combined 20 modes were considered more successful in predicting the residues around 5 Å of the ligands in frame 212 according to the ratio of 10/29.

### **3.5. The Structural Comparison Between The Frames Obtained From Molecular Dynamics (MD) and ClustENMD and The Crystal Structure (PDB ID: 6Q91)**

The selected frames for ESSA from MD, ClustENMD, and the GALK1 structure with the PDB ID of 6Q91 were compared in terms of RMSD and the differences in the radius of gyration values. The RMSD comparison was demonstrated in Table 21, and the alignments of selected frames obtained from ClustENMD and MD for ESSA to the crystal structure with the PDB ID of 6Q91 were shown in overall and zoomed fragment binding site forms in Figure 42 through Figure 45.

According to the obtained results, the highest RMSD value was found for MD frames 5143 and 5198 with an RMSD value of 3.89 Å in the scope of their alignment to 6Q91. On the other hand, the lowest RMSD value was observed for ClustENMD frame 211 with an RMSD value of 2.08 Å in the scope of its alignment to 6Q91. When the zoomed alignment figures in the allosteric site were considered, the conformational change was observed around the beta-sheet and loop regions (Met192-Val211) in all frames with respect to 6Q91. On the other hand, the conformation of the loop region

(Leu213-Leu218) also varied, the loop region was located in a more distant position (away from the fragment) in frames 5198 and 7116 compared to 6Q91. In addition,  $\alpha$ -helices around Arg286-Leu295 and Arg301-Ser310 demonstrated slight conformational differences. The variations were more in frames 211, 212, and 7116, especially in the beta-sheet region (Leu206-Val211). On the other hand, independent of the fragment binding site, all frames demonstrated a conformational change in the loop region (Leu92-Pro104), located in the upper right corner of the structure in Figure 42 and Figure 44.

Table 21. The RMSD comparison of selected *h*GALK1 frames for ESSA. The alignment was performed in PyMOL.

<b>Groups &amp; RMSD Values (Å)</b>	<b>CMD_176</b>	<b>CMD_211</b>	<b>CMD_212</b>	<b>6Q91</b>	<b>MD_5143</b>	<b>MD_5198</b>	<b>MD_6700</b>	<b>MD_7116</b>
<b>CMD_176</b>	-	1.94	3.35	2.09	5.48	5.39	4.61	5.33
<b>CMD_211</b>	-	-	2.71	2.08	5.18	5.13	4.35	5.00
<b>CMD_212</b>	-	-	-	2.53	4.35	4.27	4.01	4.28
<b>6Q91</b>	-	-	-	-	3.89	3.89	2.92	3.76
<b>MD_5143</b>	-	-	-	-	-	2.18	3.18	2.16
<b>MD_5198</b>	-	-	-	-	-	-	3.34	2.69
<b>MD_6700</b>	-	-	-	-	-	-	-	3.13
<b>MD_7116</b>	-	-	-	-	-	-	-	-

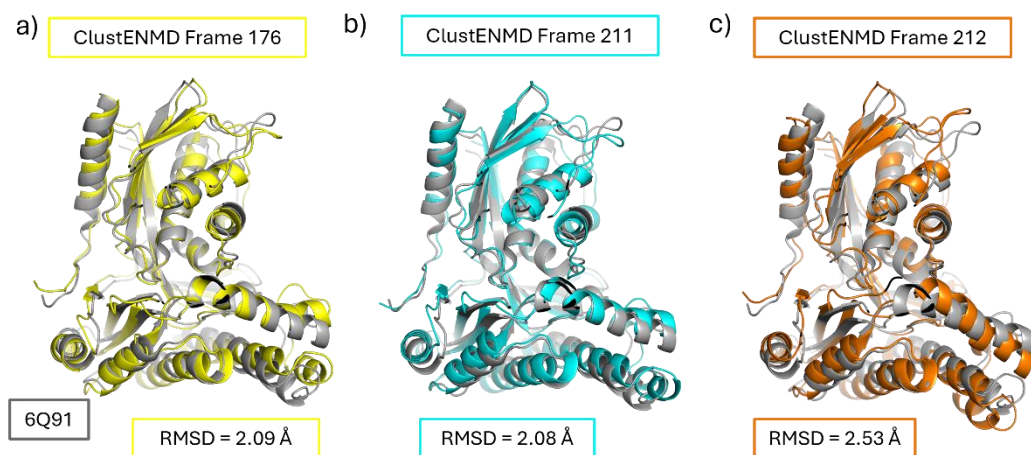


Figure 42. The alignment of ClustENMD frames to 6Q91. a) Frame 176, b) Frame 211, c) Frame 212.

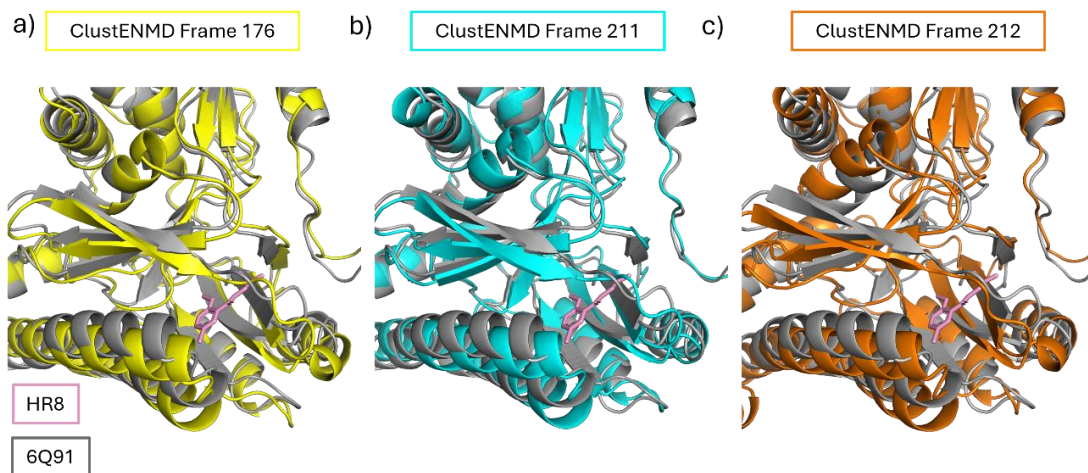


Figure 43. The demonstration of the fragment binding sites of *hGALK1* in ClustENMD frames. The structure was rotated 180° on the y-axis and zoomed in on the fragment binding site. a) Frame 176, b) Frame 211, c) Frame 212.

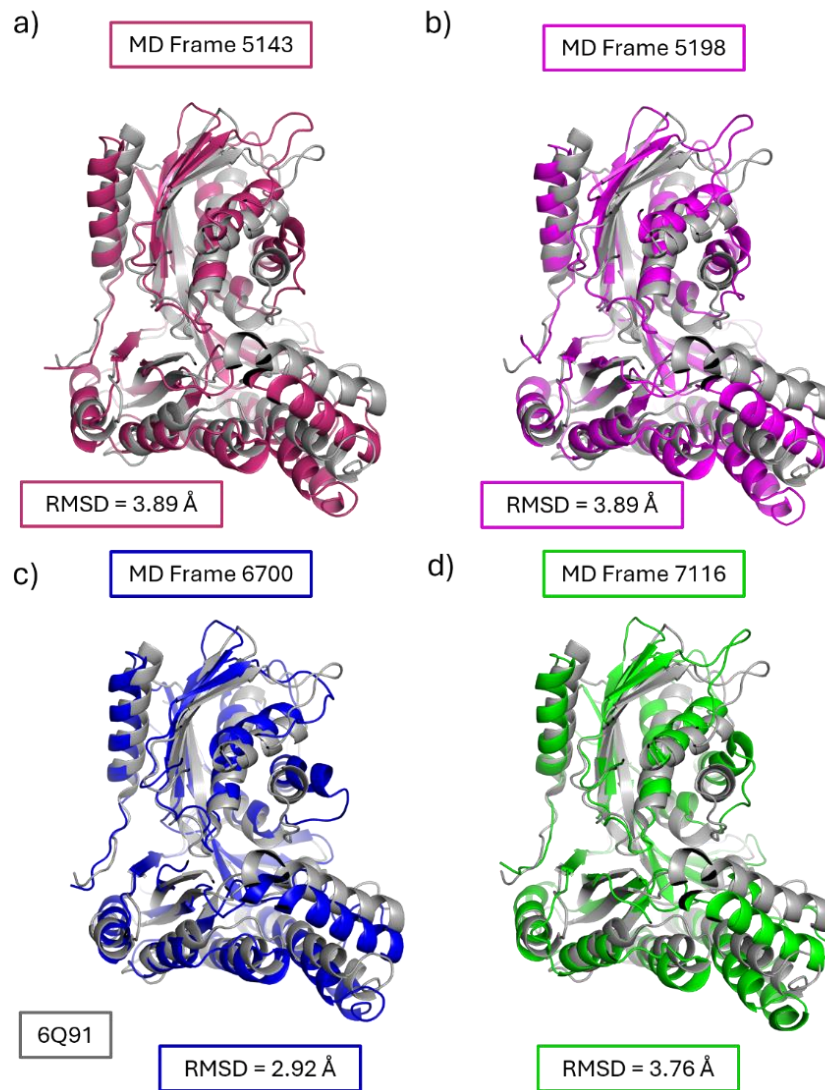


Figure 44. The alignment of *h*GALK1 MD frames to 6Q91. a) Frame 5143, b) Frame 5198, c) Frame 6700, d) Frame 7116.

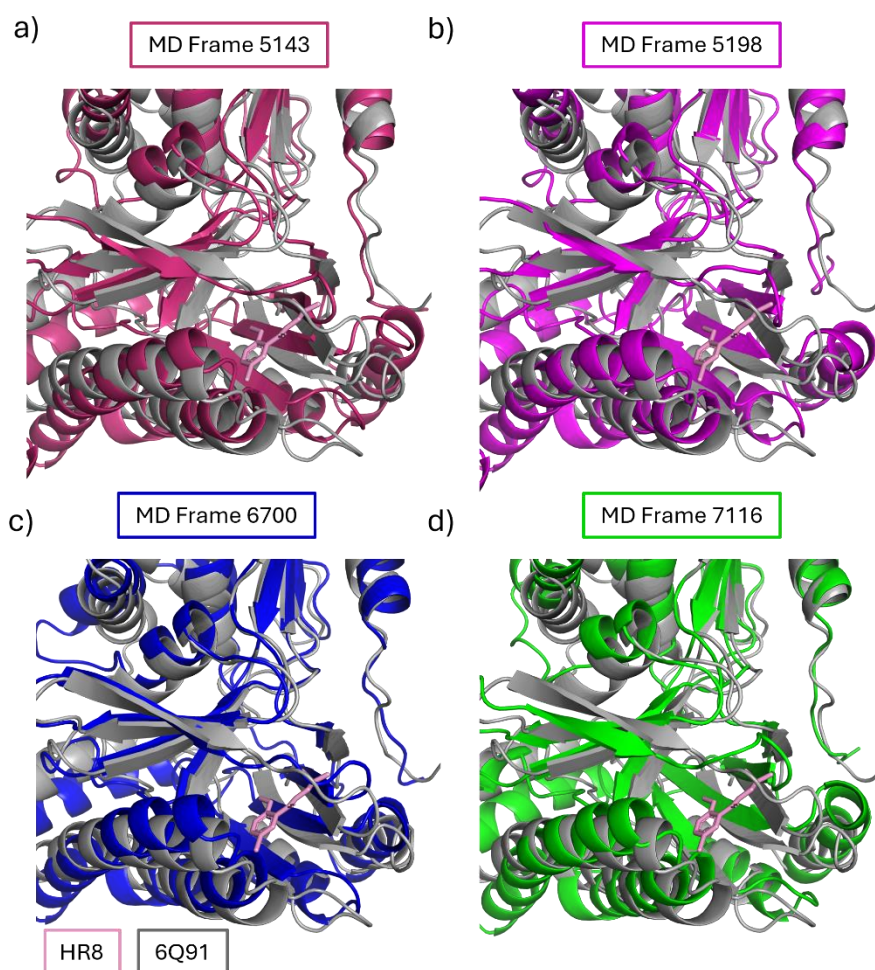


Figure 45. The demonstration of the fragment binding sites of *hGALK1* in molecular dynamics (MD) frames. The structure was rotated 180° on the y-axis and zoomed in on the fragment binding site. a) Frame 5143, b) Frame 5198, c) Frame 6700, d) Frame 7116.

However, the variations in this site were observed more in frames 212, 5143, 5198, and 7116. Among these frames, frame 212 obtained from ClustENMD was the most successful case for the prediction of fragment binding site (allosteric site) in the presence of cutoff 7 Å combined 20 modes. On the other hand, the radius of gyration was calculated for the crystal structure of 6Q91, ClustENMD, and MD frames (Table 22).

Table 22. The radius of gyration values of *h*GALK1 frames and its crystal structure.

The Structures	Radius of Gyration (nm)
6Q91	2.08
ClustENMD Frame 176	2.16
ClustENMD Frame 211	2.15
ClustENMD Frame 212	2.18
MD Frame 5143	2.24
MD Frame 5198	2.24
MD Frame 6700	2.18
MD Frame 7116	2.23

According to the findings, the lowest radius of gyration value was obtained from 6Q91 with a value of 2.08 nm while the highest radius of gyration value was calculated from frame 5143 from MD simulations with a value of 2.24 nm. The frames obtained from ClustENMD have a lower radius of gyration values compared to MD frames except the frame 6700 from MD simulations. In the ESSA results, the success of the cutoff values demonstrates variation between different frames in terms of allosteric and orthosteric site prediction. For instance, 6Q91 can only find a single residue in the fragment binding site in the presence of cutoff 7 Å. On the other hand, the same cutoff value is more successful for both substrate and fragment binding site prediction in frame 212 from ClustENMD. To conclude, cutoff 7 Å can be considered more successful for frame 5198 in terms of the prediction of the fragment binding site while cutoffs 10 Å, 11 Å, and 13 Å can be considered better for the prediction of the orthosteric site for this frame.

The variation in these results may be explained in terms of the determination of contacting atoms in the GNM calculations since a higher number of residues were assumed to be in contact when the cutoff value increased, which decreases fluctuations. On the other hand, fluctuations increase for a decrease in the cutoff value since a smaller number of residues were assumed to be in contact. The concept was also explained

through the first coordination shell and the effect of constraints. To this extent, 4 Å was not considered rational for a cutoff value (Erman 2022). Therefore, the optimization of the cutoff value was also considered to be significant from these perspectives.



## CHAPTER 4

### CONCLUSIONS AND RECOMMENDATIONS

In the first part of the thesis, the ESSA method was applied to predict the ligand binding sites including orthosteric, allosteric, and cryptic sites in apo and holo form of the TEM-1 beta-lactamase and CDK2. In addition, ESSA was applied to the holo form of *hGALK1* enzyme. To this extent, various cutoff parameters and the combination of modes were tried, and different ligand binding site prediction successes were observed. On the other hand, *hGALK1* dynamics were investigated through MD simulations in the presence and absence of an allosteric fragment. In addition, ClustENMD was performed for *hGALK1* to increase conformational sampling. The selected frames from MD and ClustENMD were used to predict allosteric and cryptic sites using ESSA. It was observed that the combination of modes improves the prediction of essential residues in the ESSA method. To this extent, ESSA can predict essential residues without depending on Fpocket by changing default parameters in ESSA. On the other hand, the allosteric site of *hGALK1* was predicted more successfully in several frames including frame 212 from ClustENMD in the presence of cutoff 7 Å combined 20 modes (also successful for the orthosteric site prediction). Also, the cutoff 11 Å combined 20 modes group detected an essential residue at the fragment binding site in the crystal structure 6Q91. In addition, cutoff 7 Å combined 20 modes is the only successful group for the prediction of fragment binding site in frame 176. However, the other cutoff values in a combined 20 modes were more successful in detecting the orthosteric site. Therefore, the outputs from different cutoff values can be combined to predict orthosteric, allosteric, and cryptic sites in future work to increase the success of the prediction. However, it was unclear why the most successful result was observed in *hGALK1* frame 212 from ClustENMD in the presence of cutoff 7 Å combined 20 modes. In this scope, graph-based approaches may be used to analyze protein network through centrality. For instance, Franke and Peter investigated frames obtained from MD simulations through residue interaction networks. To this extent, closeness centrality and dimensionality reduction were used. Closeness centrality can be considered as a descriptor for residue interaction networks (Franke and Peter

2023). In the scope, such approaches can be used to understand residue interaction networks obtained from both the crystal structure (PDB ID: 6Q91), MD, and ClustENMD frames of *hGALK1*. The graph-based approach evaluates residues as nodes and interactions as edges. On the other hand, the cutoff in GNM calculations determines the connectivity of atoms in the specified system. To this extent, the residue interaction networks may be compared in different conformations obtained from ClustENMD and MD overall and by considering the substrate binding site, ATP binding site, and allosteric site of *hGALK1*. The residues with the highest closeness centrality may be compared with those around 5 Å of orthosteric and allosteric ligands, and evaluated in terms of overlap in the determined residues. It was reported that the graph-based approaches are not only limited to distances but different variables such as molecular interactions can be also used to compare structures (Franke and Peter 2023). Such analysis may provide more insights into the cutoff and mode effect on allosteric site prediction success. In addition, residue-residue (RR) distance maps of various frames can be compared to evaluate the changes in distances between residues (Chen, Huang, and Ferrin 2015).

On the other hand, even though the thesis mostly focused on the intersection between residues around 5 Å of the known orthosteric and allosteric ligands and residues with z-score values above 5, the other sites with z-scores above 5 can be investigated further, and virtual screening can be performed for these sites for drug discovery studies. Besides binding site prediction, PCA was applied to the *hGALK1* structure to investigate collective motions in both substrate-bound *hGALK1* and substrate-fragment-bound *hGALK1*. PCA results demonstrated motion variations between these groups in PC2 and PC3 modes which provided insights on allostery in addition to differences observed from MD simulations. In future work, linear discriminant analysis (LDA) can be implemented to analyze the changes between substrate-bound and substrate-fragment-bound *hGALK1* structure.

## REFERENCES

- Abdel-Magid, Ahmed F. 2015. "Allosteric Modulators: An Emerging Concept in Drug Discovery." *ACS Medicinal Chemistry Letters* 6 (2): 104–7. doi:10.1021/M15005365.
- Akbar, Rahmad, and Volkhard Helms. 2018. "ALLO: A Tool to Discriminate and Prioritize Allosteric Pockets." *Chemical Biology & Drug Design* 91 (4): 845–53. doi:10.1111/Cbdd.13161.
- Atik, Sezen Ugan, Semra Gürsoy, Tuba Koçkar, Hasan Önal, and Servet Erdal Adal. 2016. "Clinical, Molecular, and Genetic Evaluation of Galactosemia in Turkish Children." *Turkish Archives of Pediatrics/Türk Pediatri Arşivi* 51 (4): 204. doi:10.5152/Turkpediatriars.2016.3759.
- Atilgan, Ali R., S. R. Durell, Robert L. Jernigan, Melik C. Demirel, Ozlem Keskin, and Ivet Bahar. 2001. "Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model." *Biophysical Journal* 80 (1): 505. doi:10.1016/S0006-3495(01)76033-x.
- Bahar, Ivet, Robert L. Jernigan, and Ken Dill. 2017. *Protein Actions: Principles and Modeling*. CRC Press LLC.
- Bakan, Ahmet, Lidio M. Meireles, and Ivet Bahar. 2011. "ProDy: Protein Dynamics Inferred from Theory and Experiments." *Bioinformatics* 27 (11): 1575. doi:10.1093/Bioinformatics/btr168.
- Bateman, Alex, Maria Jesus Martin, Sandra Orchard, Michele Magrane, Shadab Ahmad, Emanuele Alpi, Emily H. Bowler-Barnett, et al. 2023. "UniProt: The Universal Protein Knowledgebase in 2023." *Nucleic Acids Research* 51 (1): 523. doi:10.1093/Nar/gkac1052.
- Betzi, Stephane, Riazul Alam, Mathew Martin, Donna J. Lubbers, Huijong Han, Sudhakar R. Jakkaraj, Gunda I. Georg, and Ernst Schönbrunn. 2011. "Discovery of a Potential Allosteric Ligand Binding Site in CDK2." *ACS Chemical Biology* 6 (5): 492–501. doi:10.1021/Cb100410m.
- Bezerra, Gustavo Arruda, Sabrina MacKinnon, Thomas McCorvie, Minghao Zhang, Manshu Tang, William Foster, Michael Fairhead, et al. "Human Galatose-1-

Phosphate Uridyltransferase (GALT), Galactokinase 1 (GALK1); A Target Enabling Package.” Zenodo, 29 May 2019. doi:10.5281/Zenodo.3250158.

- Brown, Nicholas G., Sreejesh Shanker, B. V. Venkataram Prasad, and Timothy Palzkill. 2009. “Structural and Biochemical Evidence That a TEM-1  $\beta$ -Lactamase N170G Active Site Mutant Acts via Substrate-Assisted Catalysis.” *The Journal of Biological Chemistry* 284 (48): 33703. doi:10.1074/Jbc.m109.053819.
- Chang, Antje, Lisa Jeske, Sandra Ulbrich, Julia Hofmann, Julia Koblitz, Ida Schomburg, Meina Neumann-Schaal, Dieter Jahn, and Dietmar Schomburg. 2021. “BRENDA, the ELIXIR Core Data Resource in 2021: New Developments and Updates.” *Nucleic Acids Research* 49 (1): 498. doi:10.1093/Nar/gkaa1025.
- Chen, Jonathan E., Conrad C. Huang, Thomas E. Ferrin. 2015. “RRDistMaps: a UCSF Chimera tool for viewing and comparing protein distance maps.” *Bioinformatics* 31(9): 1484. doi:10.1093/Bioinformatics/btu841.
- Chiappori, Federica, Ivan Merelli, Luciano Milanese, and Anna Marabotti. 2013. “Static and Dynamic Interactions between GALK Enzyme and Known Inhibitors: Guidelines to Design New Drugs for Galactosemic Patients.” *European Journal of Medicinal Chemistry* 63 (5): 423–34. doi:10.1016/J.ejmech.2013.02.021.
- Cimermancic, Peter, Patrick Weinkam, T. Justin Rettenmaier, Leon Bichmann, Daniel A. Keedy, Rahel A. Woldeyes, Dina Schneidman-Duhovny, et al. 2016. “CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites.” *Journal of Molecular Biology* 428 (4): 709–19. doi:10.1016/J.jmb.2016.01.029.
- Cossio-Pérez, Rodrigo, Juliana Palma, and Gustavo Pierdominici-Sottile. 2017. “Consistent Principal Component Modes from Molecular Dynamics Simulations of Proteins.” *Journal of Chemical Information and Modeling* 57 (4): 826–34. doi:10.1021/Acs.jcim.6b00646.
- Danese, Elisa, and Giuseppe Lippi. 2018. “Rare Diseases: The Paradox of an Emerging Challenge.” *Annals of Translational Medicine* 6 (17): 329–329. doi:10.21037/Atm.2018.09.04.
- Delnoy, Britt, Ana I. Coelho, and Maria Estela Rubio-Gozalbo. 2021. “Current and Future Treatments for Classic Galactosemia.” *Journal of Personalized Medicine* 11 (2): 75. doi:10.3390/Jpm11020075.

- Doruker, Pemra, Ali Rana Atilgan, and Ivet Bahar. 2000. "Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to- $\alpha$ -Amylase Inhibitor." *Proteins* 40(3): 512-524. .doi:10.1002/1097-0134(20000815)40:3<512::aid-prot180>3.0.co;2-m.
- "Drug Approval Package: Sensipar (Cinacalcet HCl) NDA #021688." 2024. Accessed June 29. [https://www.accessdata.fda.gov/drugsatfda\\_docs/nda/2004/21-688\\_Sensipar.cfm](https://www.accessdata.fda.gov/drugsatfda_docs/nda/2004/21-688_Sensipar.cfm).
- Durrant, Jacob D., and J. Andrew McCammon. 2011. "Molecular Dynamics Simulations and Drug Discovery." *BMC Biology* 9 (1): 1–9. doi:10.1186/1741-7007-9-71.
- Eastman, Peter, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee Ping Wang, et al. 2017. "OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics." *PLoS Computational Biology* 13 (7): 1005659. doi:10.1371/Journal.pcbi.1005659.
- Editorial. 2022. "Rare Diseases, Common Challenges." *Nature Genetics* 54 (3): 215. doi:10.1038/s41588-022-01037-8.
- Egbert, Megan, George Jones, Matthew R. Collins, Dima Kozakov, and Sandor Vajda. 2022. "FTMove: A Web Server for Detection and Analysis of Cryptic and Allosteric Binding Sites by Mapping Multiple Protein Structures." *Journal of Molecular Biology* 434 (11): 167587. doi:10.1016/J.jmb.2022.167587.
- Erman, Burak. 2022. "Gaussian Network Model Revisited: Effects of Mutation and Ligand Binding on Protein Behavior." *Physical Biology* 19 (2): 026005. doi:10.1088/1478-3975/Ac50ba.
- Franke, Leon, Peter, Christine. 2023. "Visualizing the Residue Interaction Landscape of Proteins by Temporal Network Embedding". *Journal of Chemical Theory and Computation* 19(10): 2985–2995. <https://doi.org/10.1021/Acs.jctc.2c01228>.
- Garcia, Daniel F., José S. Camelo, Greice A. Molfetta, Marlene Turcato, Carolina F.M. Souza, Gilda Porta, Carlos E. Steiner, and Wilson A. Silva. 2016. "Clinical Profile and Molecular Characterization of Galactosemia in Brazil: Identification of Seven Novel Mutations." *BMC Medical Genetics* 17 (1): 39. doi:10.1186/S12881-016-0300-8.
- Gasteiger, Elisabeth, Christine Hoogland, Alexandre Gattiker, S'everine Duvaud, Marc R. Wilkins, Ron D. Appel, and Amos Bairoch. 2005. "Protein Identification and

Analysis Tools on the ExPASy Server.” *The Proteomics Protocols Handbook*. Humana Press, 571–607. doi:10.1385/1-59259-890-0:571.

- Gimenez-Lozano, Cristina, Lucía Páramo-Rodríguez, Clara Cavero-Carbonell, Francisca Corpas-Burgos, Aurora López-Maside, Sandra Guardiola-Villarroy, and Oscar Zurriaga. 2022. “Rare Diseases: Needs and Impact for Patients and Families: A Cross-Sectional Study in the Valencian Region, Spain.” *International Journal of Environmental Research and Public Health* 19 (16): 1. doi:10.3390/Ijerp191610366.
- Greener, Joe G., and Michael J.E. Sternberg. 2015. “AlloPred: Prediction of Allosteric Pockets on Proteins Using Normal Mode Perturbation Analysis.” *BMC Bioinformatics* 16 (335): 1–7. doi:10.1186/S12859-015-0771-1.
- Guilloux, Vincent Le, Peter Schmidtke, and Pierre Tuffery. 2009. “Fpocket: An Open Source Platform for Ligand Pocket Detection.” *BMC Bioinformatics* 10 (1): 1–11. doi:10.1186/1471-2105-10-168.
- Haliloglu, Turkan, Ivet Bahar, and Burak Erman. 1997. “Gaussian Dynamics of Folded Proteins.” *Physical Review Letters* 79 (16): 3090. doi:10.1103/Physrevlett.79.3090.
- Hollingsworth, Scott A., and Ron O. Dror. 2018. “Molecular Dynamics Simulation for All.” *Neuron* 99 (6): 1129. doi:10.1016/J.neuron.2018.08.011.
- Horn, James R., and Brian K. Shoichet. 2004. “Allosteric Inhibition Through Core Disruption.” *Journal of Molecular Biology* 336 (5): 1283–91. doi:10.1016/J.jmb.2003.12.068.
- Hu, Xin, Ya Qin Zhang, Olivia W. Lee, Li Liu, Manshu Tang, Kent Lai, Matthew B. Boxer, Matthew D. Hall, and Min Shen. 2019. “Discovery of Novel Inhibitors of Human Galactokinase by Virtual Screening.” *Journal of Computer-Aided Molecular Design* 33 (4): 405. doi:10.1007/S10822-019-00190-3.
- Huang, Meilan, Xiaozhou Li, Jian Wei Zou, and David J. Timson. 2013. “Role of Arg228 in the Phosphorylation of Galactokinase: The Mechanism of GHMP Kinases by Quantum Mechanics/Molecular Mechanics Studies.” *Biochemistry* 52 (28): 4858–68. doi:10.1021/Bi400228e.

- Huang, Wenkang, Shaoyong Lu, Zhimin Huang, Xinyi Liu, Linkai Mou, Yu Luo, Yanlong Zhao, et al. 2013. "Allosite: A Method for Predicting Allosteric Sites." *Bioinformatics* 29 (18): 2357–59. doi:10.1093/Bioinformatics/btt399.
- Humphrey, William, Andrew Dalke, and Klaus Schulten. 1996. "VMD: Visual Molecular Dynamics." *Journal of Molecular Graphics* 14 (1): 33–38. doi:10.1016/0263-7855(96)00018-5.
- Ichiye, Toshiko, and Martin Karplus. 1991. "Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations." *Proteins* 11 (3): 205–17. doi:10.1002/Prot.340110305.
- Jankowska, Katarzyna. 2017. "Premature Ovarian Failure." *Przegląd Menopauzalny = Menopause Review* 16 (2): 51. doi:10.5114/Pm.2017.68592.
- Jo, Sunhwan, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. 2008. "CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM." *Journal of Computational Chemistry* 29 (11): 1859–65. doi:10.1002/Jcc.20945.
- Jolliffe, Ian T., and Jorge Cadima. 2016. "Principal Component Analysis: A Review and Recent Developments." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2065): 1. doi:10.1098/Rsta.2015.0202.
- Kaynak, Burak T., Ivet Bahar, and Pemra Doruker. 2020. "Essential Site Scanning Analysis: A New Approach for Detecting Sites That Modulate the Dispersion of Protein Global Motions." *Computational and Structural Biotechnology Journal* 18 (1): 1577–86. doi:10.1016/J.csbj.2020.06.020.
- Kaynak, Burak T., She Zhang, Ivet Bahar, and Pemra Doruker. 2021. "ClustENMD: Efficient Sampling of Biomolecular Conformational Space at Atomic Resolution." *Bioinformatics* 37 (21): 3956–58. doi:10.1093/BIOINFORMATICS/Btab496.
- Kuzmanic, Antonija, Gregory R. Bowman, Jordi Juarez-Jimenez, Julien Michel, and Francesco L. Gervasio. 2020. "Investigating Cryptic Binding Sites by Molecular Dynamics Simulations." *Accounts of Chemical Research* 53 (3): 654–61. doi:10.1021/Acs.accounts.9b00613.

- Lai, Kent, Manshu Tang, X. Yin, Helene Klapper, Klaas J. Wierenga, and Louis. J. Elsas. 2008. "ARHI: A New Target of Galactose Toxicity in Classic Galactosemia." *Bioscience Hypotheses* 1 (5): 263–71. doi:10.1016/J.bihy.2008.06.011.
- Laskowski, Roman A., Fabian Gerick, and Janet M. Thornton. 2009. "The Structural Basis of Allosteric Regulation in Proteins." *FEBS Letters* 583 (11): 1692–98. doi:10.1016/J.febslet.2009.03.019.
- Laskowski, Roman A., Jagoda Jabłońska, Lukáš Pravda, Radka Svobodová Vařeková, and Janet M. Thornton. 2018. "PDBsum: Structural Summaries of PDB Entries." *Protein Science: A Publication of the Protein Society* 27 (1): 129. doi:10.1002/Pro.3289.
- Liu, Li, Manshu Tang, Rajan Pragani, Frank G. Whitby, Ya Qin Zhang, Bijina Balakrishnan, Yuhong Fang, et al. 2021. "Structure-Based Optimization of Small Molecule Human Galactokinase Inhibitors." *Journal of Medicinal Chemistry* 64 (18): 13551–71. doi:10.1021/Acs.jmedchem.1c00945.
- Mackinnon, Sabrina R., Tobias Krojer, William R. Foster, Laura Diaz-Saez, Manshu Tang, Kilian V.M. Huber, Frank Von Delft, et al. 2021. "Fragment Screening Reveals Starting Points for Rational Design of Galactokinase 1 Inhibitors to Treat Classic Galactosemia." *ACS Chemical Biology* 16 (4): 586–95. doi:10.1021/Acschembio.0c00498.
- Martinez-Rosell, Gerard, Silvia Lovera, Zara A. Sands, and Gianni De Fabritiis. 2020. "PlayMolecule CrypticScout: Predicting Protein Cryptic Sites Using Mixed-Solvent Molecular Simulations." *Journal of Chemical Information and Modeling* 60 (4): 2314–24. doi:10.1021/Acs.jcim.9b01209.
- McAuley, Margaret, Meilan Huang, and David J. Timson. 2017. "Insight into the Mechanism of Galactokinase: Role of a Critical Glutamate Residue and Helix/Coil Transitions." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1865 (3): 321–28. doi:10.1016/J.bbapap.2016.10.012.
- McAuley, Margaret, Noel Mesa-Torres, Aisling McFall, Sarah Morris, Meilan Huang, Angel L. Pey, and David J. Timson. 2018. "Improving the Activity and Stability of Human Galactokinase for Therapeutic and Biotechnological Applications." *ChemBioChem* 19 (10): 1088–95. doi:10.1002/Cbic.201800025.
- McCorvie, Thomas J., and David J. Timson. 2020. "Galactosemia: Opportunities for Novel Therapies." *Protein Homeostasis Diseases: Mechanisms and Novel Therapies*, 9(1): 221–45. doi:10.1016/B978-0-12-819132-3.00011-7.



- McGibbon, Robert T., Kyle A. Beauchamp, Matthew P. Harrigan, Christoph Klein, Jason M. Swails, Carlos X. Hernández, Christian R. Schwantes, Lee Ping Wang, Thomas J. Lane, and Vijay S. Pande. 2015. “MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories.” *Biophysical Journal* 109 (8): 1528–32. doi:10.1016/J.bpj.2015.08.015.
- Meller, Artur, Michael Ward, Jonathan Borowsky, Meghana Kshirsagar, Jeffrey M. Lotthammer, Felipe Oviedo, Juan Lavista Ferres, and Gregory R. Bowman. 2023. “Predicting Locations of Cryptic Pockets from Single Protein Structures Using the PocketMiner Graph Neural Network.” *Nature Communications* 14 (1): 1–15. doi:10.1038/S41467-023-36699-3.
- Nerín-Fonz, Francho, and Zoe Cournia. 2024. “Machine Learning Approaches in Predicting Allosteric Sites.” *Current Opinion in Structural Biology* 85 (4): 102774. doi:10.1016/J.sbi.2024.102774.
- Nussinov, Ruth, and Chung Jung Tsai. 2013. “Allostery in Disease and in Drug Discovery.” *Cell* 153 (2): 293–305. doi:10.1016/J.cell.2013.03.034.
- Oleinikovas, Vladimiras, Giorgio Saladino, Benjamin P. Cossins, and Francesco L. Gervasio. 2016. “Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations.” *Journal of the American Chemical Society* 138 (43): 14257–63. doi:10.1021/Jacs.6b05425.
- Pettersen, Eric F., Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. 2004. “UCSF Chimera - A Visualization System for Exploratory Research and Analysis.” *Journal of Computational Chemistry* 25 (13): 1605–12. doi:10.1002/Jcc.20084.
- “Psico - PyMOLWiki.” 2024. Accessed June 30. <https://pymolwiki.org/index.php/Psico>.
- Rubio-Gozalbo, M. Estela, Britt Derks, Anibh Martin Das, Uta Meyer, Dorothea Möslinger, M. Luz Couce, Aurélie Empain, et al. 2020. “Galactokinase Deficiency: Lessons from the GalNet Registry.” *Genetics in Medicine* 23 (1): 202–10. doi:10.1038/S41436-020-00942-9.
- Salverda, Merijn L.M., J. Arjan G.M. de Visser, and Miriam Barlow. 2010. “Natural Evolution of TEM-1  $\beta$ -Lactamase: Experimental Reconstruction and Clinical Relevance.” *FEMS Microbiology Reviews* 34 (6): 1015–36. doi:10.1111/J.1574-6976.2010.00222.x.

- Santana, Charles A., Sabrina A. de Silveira, João P.A. Moraes, Sandro C. Izidoro, Raquel C. de Melo-Minardi, António J.M. Ribeiro, Jonathan D. Tyzack, Neera Borkakoti, and Janet M. Thornton. 2020. "GRaSP: A Graph-Based Residue Neighborhood Strategy to Predict Binding Sites." *Bioinformatics* 36(2): 726–34. doi:10.1093/Bioinformatics/btaa805.
- Schrödinger LLC. 2022. "The PyMOL Molecular Graphics System, Version 2.5.4."
- Shaker, Bilal, Sajjad Ahmad, Jingyu Lee, Chanjin Jung, and Dokyun Na. 2021. "In Silico Methods and Tools for Drug Discovery." *Computers in Biology and Medicine* 137 (10): 104851. doi:10.1016/J.combiomed.2021.104851.
- P, Sneha, Elaheh Ahmad Ebrahimi, Sara Ahmed Ghazala, D. Thirumal Kumar, R. Siva, George Priya Doss C, and Hatem Zayed. 2018. "Structural Analysis of Missense Mutations in Galactokinase 1 (GALK1) Leading to Galactosemia Type-2." *Journal of Cellular Biochemistry* 119 (9): 7585–98. doi:10.1002/Jcb.27097.
- Stec, Boguslaw, Kathleen M. Holtz, Cheryl L. Wojciechowski, and Evan R. Kantrowitz. 2005. "Structure of the Wild-Type TEM-1  $\beta$ -Lactamase at 1.55 Å and the Mutant Enzyme Ser70Ala at 2.1 Å Suggest the Mode of Noncovalent Catalysis for the Mutant Enzyme." *International Union of Crystallography* 61 (8): 1072–79. doi:10.1107/S0907444905014356.
- Stein, Sarah A. Mueller, Anne E. Loccisano, Steven M. Firestine, and Jeffrey D. Evanseck. 2006. "Chapter 13 Principal Components Analysis: A Review of Its Application on Molecular Dynamics Data." *Annual Reports in Computational Chemistry* 2 (1): 233–61. doi:10.1016/S1574-1400(06)02013-5.
- Succoio, Mariangela, Rosa Sacchetti, Alessandro Rossi, Giancarlo Parenti, and Margherita Ruoppolo. 2022. "Galactosemia: Biochemistry, Molecular Genetics, Newborn Screening, and Treatment." *Biomolecules* 12 (7): 968. doi:10.3390/Biom12070968.
- Sun, Zhuyezhi, Amanda Elizabeth Wakefield, Istvan Kolossvary, Dmitri Beglov, and Sandor Vajda. 2020. "Structure-Based Analysis of Cryptic-Site Opening." *Structure* 28 (2): 223–235. doi:10.1016/J.str.2019.11.007.
- Thoden, James B., David J. Timson, Richard J. Reece, and Hazel M. Holden. 2005. "Molecular Structure of Human Galactokinase." *Journal of Biological Chemistry* 280 (10): 9662–70. doi:10.1074/Jbc.m412916200.

- Tian, Hao, Sian Xiao, Xi Jiang, and Peng Tao. 2023. "PASSer: Fast and Accurate Prediction of Protein Allosteric Sites." *Nucleic Acids Research* 51 (1): 427–31. doi:10.1093/Nar/gkad303.
- Umbayev, Bauyrzhan, Sholpan Askarova, Aigul Almabayeva, Timur Saliev, Abdul Razak Masoud, and Denis Bulanin. 2020. "Galactose-Induced Skin Aging: The Role of Oxidative Stress." *Oxidative Medicine and Cellular Longevity* 4 (1): 7145656. doi:10.1155/2020/7145656.
- Uyar, Arzu, V. T. Karamyan, and Alex Dickson. 2018. "Long-Range Changes in Neurolysin Dynamics Upon Inhibitor Binding." *Journal of Chemical Theory and Computation* 14 (1): 444–52. doi:10.1021/Acs.jctc.7b00944.
- Uyar, Arzu, and Alex Dickson. 2021. "Perturbation of ACE2 Structural Ensembles by SARS-CoV-2 Spike Protein Binding." *Journal of Chemical Theory and Computation* 17 (9): 5896–5906. doi:10.1021/Acs.jctc.1c00325.
- Vanommeslaeghe, K., E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, et al. 2010. "CHARMM General Force Field (CGenFF): A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields." *Journal of Computational Chemistry* 31 (4): 671. doi:10.1002/Jcc.21367.
- Xie, Juan, Weilin Zhang, Xiaolei Zhu, Minghua Deng, and Luhua Lai. 2023. "Coevolution-Based Prediction of Key Allosteric Residues for Protein Function Regulation." *eLife* 12(1): 9. doi:10.7554/Elife.81850.
- Yu, Wenbo, Xibing He, Kenno Vanommeslaeghe, and Alexander D. MacKerell. 2012. "Extension of the CHARMM General Force Field to Sulfonyl-Containing Compounds and Its Utility in Biomolecular Simulations." *Journal of Computational Chemistry* 33 (31): 2451. doi:10.1002/Jcc.23067.

## APPENDIX A

Table 23. 1AXB-FOS PDBsum interaction.

Hydrogen Bonds											
	Atom	Atom	Res	Res	Chain	Atom	Atom	Res	Res	Chain	Distance (Å)
	no.	name	name	no.		no.	name	name	no.		
1	355	N	SER	70	A	2046	O2P	FOS	291	A	2.87
2	823	OG	SER	130	A	2045	O1P	FOS	291	A	2.75
3	839	ND2	ASN	132	A	2041	O'	FOS	291	A	3.01
4	1641	N	ALA	237	A	2046	O2P	FOS	291	A	2.92
5	1644	O	ALA	237	A	2042	N	FOS	291	A	3.02
6	1644	O	ALA	237	A	2046	O2P	FOS	291	A	3.09
Non-bonded Contacts											
	Atom	Atom	Res	Res	Chain	Atom	Atom	Res	Res	Chain	Distance (Å)
	no.	name	name	no.		no.	name	name	no.		
1	351	CB	MET	69	A	2046	O2P	FOS	291	A	3.62
2	355	N	SER	70	A	2044	P	FOS	291	A	3.52
3	355	N	SER	70	A	2046	O2P	FOS	291	A	2.87
4	356	CA	SER	70	A	2044	P	FOS	291	A	3.57
5	356	CA	SER	70	A	2046	O2P	FOS	291	A	3.45
6	359	CB	SER	70	A	2044	P	FOS	291	A	2.61
7	359	CB	SER	70	A	2045	O1P	FOS	291	A	3.21
8	359	CB	SER	70	A	2046	O2P	FOS	291	A	2.9
9	360	OG	SER	70	A	2043	CM	FOS	291	A	2.7
10	360	OG	SER	70	A	2044	P	FOS	291	A	1.61
11	360	OG	SER	70	A	2045	O1P	FOS	291	A	2.47
12	360	OG	SER	70	A	2046	O2P	FOS	291	A	2.5
13	822	CB	SER	130	A	2045	O1P	FOS	291	A	3.76
14	823	OG	SER	130	A	2044	P	FOS	291	A	3.43
15	823	OG	SER	130	A	2045	O1P	FOS	291	A	2.75
16	838	OD1	ASN	132	A	2043	CM	FOS	291	A	3.43
17	839	ND2	ASN	132	A	2041	O'	FOS	291	A	3.01
18	1135	CG	ASN	170	A	2040	C'	FOS	291	A	3.69
19	1135	CG	ASN	170	A	2042	N	FOS	291	A	3.71
20	1136	OD1	ASN	170	A	2040	C'	FOS	291	A	3.46
21	1136	OD1	ASN	170	A	2041	O'	FOS	291	A	3.32
22	1136	OD1	ASN	170	A	2042	N	FOS	291	A	3.71
23	1137	ND2	ASN	170	A	2042	N	FOS	291	A	3.69
24	1137	ND2	ASN	170	A	2043	CM	FOS	291	A	3.9
25	1638	CA	GLY	236	A	2046	O2P	FOS	291	A	3.31
26	1639	C	GLY	236	A	2046	O2P	FOS	291	A	3.6
27	1641	N	ALA	237	A	2046	O2P	FOS	291	A	2.92
28	1643	C	ALA	237	A	2046	O2P	FOS	291	A	3.86
29	1644	O	ALA	237	A	2039	O	FOS	291	A	3.18
30	1644	O	ALA	237	A	2040	C'	FOS	291	A	3.6
31	1644	O	ALA	237	A	2042	N	FOS	291	A	3.02
32	1644	O	ALA	237	A	2046	O2P	FOS	291	A	3.09
33	1648	C	GLY	238	A	2036	C5	FOS	291	A	3.54
34	1648	C	GLY	238	A	2037	C6	FOS	291	A	3.8
35	1649	O	GLY	238	A	2036	C5	FOS	291	A	3.33
36	1649	O	GLY	238	A	2037	C6	FOS	291	A	3.88
37	1655	CG	GLU	240	A	2034	C3	FOS	291	A	3.39
38	1655	CG	GLU	240	A	2035	C4	FOS	291	A	3.7
Number of hydrogen bonds: 6											
Number of non-bonded contacts: 38											

Table 24. 1PZO-CBT300 PDBsum interaction.

Hydrogen Bonds											
	Atom no.	Atom name	Res name	Res no.	Chain	Atom no.	Atom name	Res name	Res no.	Chain	Distance (Å)
1	1632	OG	SER	235	A	2041	N19	CBT	300	A	2.6
Non-bonded Contacts											
	Atom no.	Atom name	Res name	Res no.	Chain	Atom no.	Atom name	Res name	Res no.	Chain	Distance (Å)
1	1492	CG1	VAL	216	A	2040	N21	CBT	300	A	3.86
2	1492	CG1	VAL	216	A	2041	N19	CBT	300	A	3.76
3	1516	CD1	LEU	220	A	2046	C7	CBT	300	A	3.53
4	1516	CD1	LEU	220	A	2047	CL8	CBT	300	A	3.72
5	1516	CD1	LEU	220	A	2048	C6	CBT	300	A	3.83
6	1631	CB	SER	235	A	2031	C26	CBT	300	A	3.81
7	1631	CB	SER	235	A	2034	C22	CBT	300	A	3.68
8	1631	CB	SER	235	A	2035	C17	CBT	300	A	3.82
9	1631	CB	SER	235	A	2041	N19	CBT	300	A	3.29
10	1632	OG	SER	235	A	2035	C17	CBT	300	A	3.55
11	1632	OG	SER	235	A	2037	C16	CBT	300	A	3.7
12	1632	OG	SER	235	A	2040	N21	CBT	300	A	3.34
13	1632	OG	SER	235	A	2041	N19	CBT	300	A	2.6
14	1633	N	GLY	236	A	2040	N21	CBT	300	A	3.74
15	1633	N	GLY	236	A	2041	N19	CBT	300	A	3.76
16	1634	CA	GLY	236	A	2040	N21	CBT	300	A	3.69
17	1635	C	GLY	236	A	2040	N21	CBT	300	A	3.73
18	1636	O	GLY	236	A	2031	C26	CBT	300	A	3.83
19	1641	CB	ALA	237	A	2038	N18	CBT	300	A	3.78
20	1641	CB	ALA	237	A	2039	N20	CBT	300	A	3.56
21	1680	CB	ARG	244	A	2030	C28	CBT	300	A	3.74
22	1681	CG	ARG	244	A	2030	C28	CBT	300	A	3.7
23	1682	CD	ARG	244	A	2030	C28	CBT	300	A	3.36
24	1682	CD	ARG	244	A	2031	C26	CBT	300	A	3.82
25	1685	NH1	ARG	244	A	2042	C1	CBT	300	A	3.84
26	1687	N	GLY	245	A	2030	C28	CBT	300	A	3.43
27	1688	CA	GLY	245	A	2028	CL3	CBT	300	A	3.81
28	1688	CA	GLY	245	A	2030	C28	CBT	300	A	3.8
29	1689	C	GLY	245	A	2028	CL3	CBT	300	A	3.69
30	1689	C	GLY	245	A	2029	C29	CBT	300	A	3.73
31	1689	C	GLY	245	A	2030	C28	CBT	300	A	3.55
32	1690	O	GLY	245	A	2029	C29	CBT	300	A	3.63
33	1690	O	GLY	245	A	2030	C28	CBT	300	A	3.18
34	1690	O	GLY	245	A	2031	C26	CBT	300	A	3.78
35	1691	N	ILE	246	A	2028	CL3	CBT	300	A	3.81
36	1695	CB	ILE	246	A	2028	CL3	CBT	300	A	3.75
37	1695	CB	ILE	246	A	2029	C29	CBT	300	A	3.67
38	1695	CB	ILE	246	A	2032	C27	CBT	300	A	3.74
39	1697	CG2	ILE	246	A	2028	CL3	CBT	300	A	3.39
40	1697	CG2	ILE	246	A	2029	C29	CBT	300	A	3.85
41	1697	CG2	ILE	246	A	2032	C27	CBT	300	A	3.9
42	1698	CD1	ILE	246	A	2032	C27	CBT	300	A	3.83
43	1909	CG	ASN	276	A	2049	C4	CBT	300	A	3.83
44	1910	OD1	ASN	276	A	2049	C4	CBT	300	A	3.59
45	1911	ND2	ASN	276	A	2048	C6	CBT	300	A	3.74
46	1911	ND2	ASN	276	A	2049	C4	CBT	300	A	3.54
47	1937	CG1	ILE	279	A	2028	CL3	CBT	300	A	3.82
48	1938	CG2	ILE	279	A	2032	C27	CBT	300	A	3.87
49	1939	CD1	ILE	279	A	2028	CL3	CBT	300	A	3.57
Number of hydrogen bonds: 1											
Number of non-bonded contacts: 49											

Table 25. 1PZO-CBT301 PDBsum interaction.

Non-bonded Contacts											
	Atom no.	Atom name	Res name	Res no.	Chain	Atom no.	Atom name	Res name	Res no.	Chain	Distance (Å)
1	1513	O	LEU	220	A	2057	C17	CBT	301	A	3.87
2	1514	CB	LEU	220	A	2055	C25	CBT	301	A	3.35
3	1516	CD1	LEU	220	A	2054	C27	CBT	301	A	3.64
4	1525	CD2	LEU	221	A	2065	C2	CBT	301	A	3.72
5	1525	CD2	LEU	221	A	2068	C7	CBT	301	A	3.83
6	1525	CD2	LEU	221	A	2070	C6	CBT	301	A	3.46
7	1525	CD2	LEU	221	A	2071	C4	CBT	301	A	3.35
8	1547	CB	ALA	224	A	2058	N13	CBT	301	A	3.64
9	1547	CB	ALA	224	A	2059	C16	CBT	301	A	3.41
10	1547	CB	ALA	224	A	2060	N18	CBT	301	A	3.54
11	1547	CB	ALA	224	A	2063	N19	CBT	301	A	3.86
12	1547	CB	ALA	224	A	2064	C1	CBT	301	A	3.86
13	1685	NH1	ARG	244	A	2050	CL3	CBT	301	A	3.4
14	1810	CD1	ILE	263	A	2069	CL8	CBT	301	A	3.79
15	1907	O	ASN	276	A	2052	C28	CBT	301	A	3.43
16	1908	CB	ASN	276	A	2050	CL3	CBT	301	A	3.81
17	1909	CG	ASN	276	A	2050	CL3	CBT	301	A	3.5
18	1910	OD1	ASN	276	A	2050	CL3	CBT	301	A	3.07
19	1935	O	ILE	279	A	2066	C3	CBT	301	A	3.59
20	1935	O	ILE	279	A	2067	C5	CBT	301	A	3.88
21	1936	CB	ILE	279	A	2052	C28	CBT	301	A	3.56
22	1938	CG2	ILE	279	A	2052	C28	CBT	301	A	3.56
23	1938	CG2	ILE	279	A	2053	C26	CBT	301	A	3.83
24	1938	CG2	ILE	279	A	2066	C3	CBT	301	A	3.84
25	1938	CG2	ILE	279	A	2067	C5	CBT	301	A	3.74
26	1940	N	ALA	280	A	2053	C26	CBT	301	A	3.8
27	1944	CB	ALA	280	A	2059	C16	CBT	301	A	3.76
28	1944	CB	ALA	280	A	2060	N18	CBT	301	A	3.44
29	1944	CB	ALA	280	A	2061	N20	CBT	301	A	3.42
30	1944	CB	ALA	280	A	2062	N21	CBT	301	A	3.74
31	1963	CA	GLY	283	A	2065	C2	CBT	301	A	3.69
32	1963	CA	GLY	283	A	2066	C3	CBT	301	A	3.67
33	1963	CA	GLY	283	A	2067	C5	CBT	301	A	3.72
34	1963	CA	GLY	283	A	2068	C7	CBT	301	A	3.8
35	1963	CA	GLY	283	A	2070	C6	CBT	301	A	3.81
36	1963	CA	GLY	283	A	2071	C4	CBT	301	A	3.76

Number of non-bonded contacts: 36

Table 26. 3PXZ-JWS PDBsum interaction.

Hydrogen Bonds											
	Atom	Atom	Res	Res	Chain	Atom	Atom	Res	Res	Chain	Distance (Å)
	no.	name	name	no.		no.	name	name	no.		
1	694	N	LEU	83	A	2482	O8	JWS	301	A	2.99
2	697	O	LEU	83	A	2482	O8	JWS	301	A	3.14
3	697	O	LEU	83	A	2490	N16	JWS	301	A	2.85
4	728	OD2	ASP	86	A	2477	N3	JWS	301	A	3.29
Non-bonded Contacts											
	Atom	Atom	Res	Res	Chain	Atom	Atom	Res	Res	Chain	Distance (Å)
	no.	name	name	no.		no.	name	name	no.		
1	116	O	ILE	10	A	2475	N1	JWS	301	A	3.21
2	116	O	ILE	10	A	2476	C2	JWS	301	A	3.75
3	116	O	ILE	10	A	2477	N3	JWS	301	A	3.86
4	116	O	ILE	10	A	2489	C15	JWS	301	A	3.89
5	117	CB	ILE	10	A	2490	N16	JWS	301	A	3.83
6	120	CD1	ILE	10	A	2482	O8	JWS	301	A	3.83
7	120	CD1	ILE	10	A	2488	N14	JWS	301	A	3.45
8	120	CD1	ILE	10	A	2489	C15	JWS	301	A	3.8
9	275	CB	ALA	31	A	2481	C7	JWS	301	A	3.6
10	275	CB	ALA	31	A	2482	O8	JWS	301	A	3.62
11	275	CB	ALA	31	A	2483	C9	JWS	301	A	3.37
12	291	CE	LYS	33	A	2486	O12	JWS	301	A	3.54
13	291	CE	LYS	33	A	2491	C17	JWS	301	A	3.6
14	292	NZ	LYS	33	A	2486	O12	JWS	301	A	3.11
15	292	NZ	LYS	33	A	2491	C17	JWS	301	A	3.12
16	535	CG1	VAL	64	A	2491	C17	JWS	301	A	3.84
17	536	CG2	VAL	64	A	2484	C10	JWS	301	A	3.73
18	667	CB	PHE	80	A	2484	C10	JWS	301	A	3.7
19	668	CG	PHE	80	A	2484	C10	JWS	301	A	3.77
20	668	CG	PHE	80	A	2491	C17	JWS	301	A	3.68
21	669	CD1	PHE	80	A	2491	C17	JWS	301	A	3.67
22	670	CD2	PHE	80	A	2491	C17	JWS	301	A	3.5
23	671	CE1	PHE	80	A	2491	C17	JWS	301	A	3.47
24	672	CE2	PHE	80	A	2491	C17	JWS	301	A	3.29
25	673	CZ	PHE	80	A	2491	C17	JWS	301	A	3.27
26	677	O	GLU	81	A	2483	C9	JWS	301	A	3.36
27	694	N	LEU	83	A	2482	O8	JWS	301	A	2.99
28	695	CA	LEU	83	A	2482	O8	JWS	301	A	3.64
29	696	C	LEU	83	A	2482	O8	JWS	301	A	3.81
30	696	C	LEU	83	A	2490	N16	JWS	301	A	3.9
31	697	O	LEU	83	A	2482	O8	JWS	301	A	3.14
32	697	O	LEU	83	A	2488	N14	JWS	301	A	3.24
33	697	O	LEU	83	A	2489	C15	JWS	301	A	3.46
34	697	O	LEU	83	A	2490	N16	JWS	301	A	2.85
35	698	CB	LEU	83	A	2482	O8	JWS	301	A	3.57
36	704	C	HIS	84	A	2490	N16	JWS	301	A	3.19
37	705	O	HIS	84	A	2490	N16	JWS	301	A	2.96
38	712	N	GLN	85	A	2490	N16	JWS	301	A	3.47
39	713	CA	GLN	85	A	2490	N16	JWS	301	A	3.63
40	728	OD2	ASP	86	A	2477	N3	JWS	301	A	3.29
41	1113	CD1	LEU	134	A	2480	C6	JWS	301	A	3.56
42	1113	CD1	LEU	134	A	2481	C7	JWS	301	A	3.7
43	1113	CD1	LEU	134	A	2483	C9	JWS	301	A	3.81
44	1113	CD1	LEU	134	A	2484	C10	JWS	301	A	3.79
45	1113	CD1	LEU	134	A	2485	C11	JWS	301	A	3.67
46	1113	CD1	LEU	134	A	2487	C13	JWS	301	A	3.54
47	1114	CD2	LEU	134	A	2478	N4	JWS	301	A	3.5
48	1114	CD2	LEU	134	A	2479	C5	JWS	301	A	3.52
49	1114	CD2	LEU	134	A	2480	C6	JWS	301	A	3.76
Number of hydrogen bonds: 4											
Number of non-bonded contacts: 49											

Table 27. 3PXZ-2AN299 PDBsum interaction.

Hydrogen Bonds											
	Atom	Atom	Res	Res	Chain	Atom	Atom	Res	Res	Chain	Distance (Å)
	no.	name	name	no.		no.	name	name	no.		
1	471	NZ	LYS	56	A	2453	O3	2AN	299	A	2.63
2	594	NE2	HIS	71	A	2452	O2	2AN	299	A	2.91
Non-bonded Contacts											
	Atom	Atom	Res	Res	Chain	Atom	Atom	Res	Res	Chain	Distance (Å)
	no.	name	name	no.		no.	name	name	no.		
1	309	CD1	ILE	35	A	2438	C6	2AN	299	A	3.79
2	328	CD2	LEU	37	A	2446	C12	2AN	299	A	3.57
3	328	CD2	LEU	37	A	2447	C13	2AN	299	A	3.68
4	440	CD1	ILE	52	A	2434	C2	2AN	299	A	3.89
5	440	CD1	ILE	52	A	2435	C3	2AN	299	A	3.62
6	440	CD1	ILE	52	A	2436	C4	2AN	299	A	3.6
7	440	CD1	ILE	52	A	2437	C5	2AN	299	A	3.84
8	469	CD	LYS	56	A	2453	O3	2AN	299	A	3.62
9	470	CE	LYS	56	A	2453	O3	2AN	299	A	3.27
10	471	NZ	LYS	56	A	2453	O3	2AN	299	A	2.63
11	576	CG2	VAL	69	A	2453	O3	2AN	299	A	3.82
12	592	CD2	HIS	71	A	2447	C13	2AN	299	A	3.76
13	593	CE1	HIS	71	A	2452	O2	2AN	299	A	3.27
14	594	NE2	HIS	71	A	2446	C12	2AN	299	A	3.86
15	594	NE2	HIS	71	A	2447	C13	2AN	299	A	3.89
16	594	NE2	HIS	71	A	2452	O2	2AN	299	A	2.91
17	634	CD1	LEU	76	A	2446	C12	2AN	299	A	3.8
18	634	CD1	LEU	76	A	2447	C13	2AN	299	A	3.89
19	635	CD2	LEU	76	A	2433	C1	2AN	299	A	3.71
20	635	CD2	LEU	76	A	2437	C5	2AN	299	A	3.86
21	635	CD2	LEU	76	A	2443	C9	2AN	299	A	3.8
22	635	CD2	LEU	76	A	2444	C10	2AN	299	A	3.48
23	635	CD2	LEU	76	A	2446	C12	2AN	299	A	3.56
24	635	CD2	LEU	76	A	2452	O2	2AN	299	A	3.71
25	655	CD2	LEU	78	A	2439	C7	2AN	299	A	3.67
Number of hydrogen bonds: 2											
Number of non-bonded contacts: 25											



Table 28. 3PXZ-2AN300 PDBsum interaction.

Hydrogen Bonds											
	Atom no.	Atom name	Res name	Res no.	Chain	Atom no.	Atom name	Res name	Res no.	Chain	Distance (Å)
1	292	NZ	LYS	33	A	2472	O1	2AN	300	A	3.01
2	1186	N	ASP	145	A	2474	O3	2AN	300	A	2.63
3	1194	N	PHE	146	A	2473	O2	2AN	300	A	2.88
Non-bonded Contacts											
	Atom no.	Atom name	Res name	Res no.	Chain	Atom no.	Atom name	Res name	Res no.	Chain	Distance (Å)
1	153	CE1	TYR	15	A	2468	C13	2AN	300	A	3.65
2	153	CE1	TYR	15	A	2469	C14	2AN	300	A	3.87
3	154	CE2	TYR	15	A	2467	C12	2AN	300	A	3.58
4	154	CE2	TYR	15	A	2468	C13	2AN	300	A	3.79
5	155	CZ	TYR	15	A	2467	C12	2AN	300	A	3.72
6	155	CZ	TYR	15	A	2468	C13	2AN	300	A	3.44
7	155	CZ	TYR	15	A	2469	C14	2AN	300	A	3.67
8	156	OH	TYR	15	A	2468	C13	2AN	300	A	3.7
9	156	OH	TYR	15	A	2469	C14	2AN	300	A	3.45
10	156	OH	TYR	15	A	2470	C15	2AN	300	A	3.49
11	156	OH	TYR	15	A	2471	C16	2AN	300	A	3.77
12	290	CD	LYS	33	A	2472	O1	2AN	300	A	3.79
13	291	CE	LYS	33	A	2472	O1	2AN	300	A	3.88
14	292	NZ	LYS	33	A	2463	S	2AN	300	A	3.73
15	292	NZ	LYS	33	A	2472	O1	2AN	300	A	3.01
16	292	NZ	LYS	33	A	2474	O3	2AN	300	A	3.3
17	309	CD1	ILE	35	A	2467	C12	2AN	300	A	3.78
18	309	CD1	ILE	35	A	2468	C13	2AN	300	A	3.57
19	460	CG	LEU	55	A	2457	C4	2AN	300	A	3.47
20	460	CG	LEU	55	A	2458	C5	2AN	300	A	3.48
21	460	CG	LEU	55	A	2459	C6	2AN	300	A	3.82
22	462	CD2	LEU	55	A	2454	C1	2AN	300	A	3.71
23	462	CD2	LEU	55	A	2457	C4	2AN	300	A	3.81
24	462	CD2	LEU	55	A	2458	C5	2AN	300	A	3.6
25	462	CD2	LEU	55	A	2465	C10	2AN	300	A	3.54
26	468	CG	LYS	56	A	2456	C3	2AN	300	A	3.57
27	468	CG	LYS	56	A	2457	C4	2AN	300	A	3.8
28	533	O	VAL	64	A	2460	C7	2AN	300	A	3.51
29	535	CG1	VAL	64	A	2460	C7	2AN	300	A	3.74
30	535	CG1	VAL	64	A	2461	C8	2AN	300	A	3.56
31	553	CD2	LEU	66	A	2457	C4	2AN	300	A	3.83
32	654	CD1	LEU	78	A	2467	C12	2AN	300	A	3.68
33	671	CE1	PHE	80	A	2460	C7	2AN	300	A	3.75
34	671	CE1	PHE	80	A	2461	C8	2AN	300	A	3.54
35	671	CE1	PHE	80	A	2464	C9	2AN	300	A	3.83
36	673	CZ	PHE	80	A	2464	C9	2AN	300	A	3.72
37	673	CZ	PHE	80	A	2472	O1	2AN	300	A	3.63
38	1182	CA	ALA	144	A	2474	O3	2AN	300	A	3.59
39	1183	C	ALA	144	A	2474	O3	2AN	300	A	3.55
40	1185	CB	ALA	144	A	2474	O3	2AN	300	A	3.79
41	1186	N	ASP	145	A	2463	S	2AN	300	A	3.61
42	1186	N	ASP	145	A	2473	O2	2AN	300	A	3.32
43	1186	N	ASP	145	A	2474	O3	2AN	300	A	2.63
44	1187	CA	ASP	145	A	2473	O2	2AN	300	A	3.36
45	1187	CA	ASP	145	A	2474	O3	2AN	300	A	3.48
46	1188	C	ASP	145	A	2473	O2	2AN	300	A	3.54
47	1194	N	PHE	146	A	2473	O2	2AN	300	A	2.88
48	1195	CA	PHE	146	A	2473	O2	2AN	300	A	3.85
49	1198	CB	PHE	146	A	2473	O2	2AN	300	A	3.77
50	1200	CD1	PHE	146	A	2473	O2	2AN	300	A	3.81
Number of hydrogen bonds: 3											
Number of non-bonded contacts: 50											

Table 29. 6Q91-GAL PDBsum interaction.

Hydrogen Bonds											
	Atom	Atom	Res	Res	Chain	Atom	Atom	Res	Res	Chain	Distance (Å)
	no.	name	name	no.		no.	name	name	no.		
1	300	OE1	GLU	43	A	11029	O6	GAL	401	A	2.55
2	302	N	HIS	44	A	11029	O6	GAL	401	A	2.61
3	325	OD1	ASP	46	A	11027	O4	GAL	401	A	2.75
4	326	OD2	ASP	46	A	11026	O3	GAL	401	A	2.69
5	1318	N	GLY	183	A	11026	O3	GAL	401	A	2.9
6	1345	OD2	ASP	186	A	11025	O2	GAL	401	A	2.6
7	1721	OH	TYR	236	A	11027	O4	GAL	401	A	2.47
8	1721	OH	TYR	236	A	11027	O4	GAL	401	A	2.47
Non-bonded Contacts											
	Atom	Atom	Res	Res	Chain	Atom	Atom	Res	Res	Chain	Distance (Å)
	no.	name	name	no.		no.	name	name	no.		
1	257	NH2	ARG	37	A	11024	O1	GAL	401	A	3.65
2	294	CA	GLU	43	A	11029	O6	GAL	401	A	3.17
3	295	C	GLU	43	A	11029	O6	GAL	401	A	3.33
4	297	CB	GLU	43	A	11029	O6	GAL	401	A	3.81
5	298	CG	GLU	43	A	11029	O6	GAL	401	A	3.39
6	299	CD	GLU	43	A	11029	O6	GAL	401	A	3.35
7	300	OE1	GLU	43	A	11023	C6	GAL	401	A	3.59
8	300	OE1	GLU	43	A	11029	O6	GAL	401	A	2.55
9	302	N	HIS	44	A	11023	C6	GAL	401	A	3.23
10	302	N	HIS	44	A	11029	O6	GAL	401	A	2.61
11	303	CA	HIS	44	A	11023	C6	GAL	401	A	3.81
12	303	CA	HIS	44	A	11029	O6	GAL	401	A	3.66
13	307	CG	HIS	44	A	11023	C6	GAL	401	A	3.81
14	308	ND1	HIS	44	A	11023	C6	GAL	401	A	3.8
15	309	CD2	HIS	44	A	11023	C6	GAL	401	A	3.83
16	310	CE1	HIS	44	A	11023	C6	GAL	401	A	3.85
17	311	NE2	HIS	44	A	11023	C6	GAL	401	A	3.87
18	324	CG	ASP	46	A	11021	C4	GAL	401	A	3.45
19	324	CG	ASP	46	A	11026	O3	GAL	401	A	3.53
20	324	CG	ASP	46	A	11027	O4	GAL	401	A	3.51
21	325	OD1	ASP	46	A	11021	C4	GAL	401	A	3.15
22	325	OD1	ASP	46	A	11026	O3	GAL	401	A	3.64
23	325	OD1	ASP	46	A	11027	O4	GAL	401	A	2.75
24	326	OD2	ASP	46	A	11020	C3	GAL	401	A	3.26
25	326	OD2	ASP	46	A	11021	C4	GAL	401	A	3.33
26	326	OD2	ASP	46	A	11026	O3	GAL	401	A	2.69
27	326	OD2	ASP	46	A	11027	O4	GAL	401	A	3.68
28	334	CD2	TYR	47	A	11027	O4	GAL	401	A	3.45
29	336	CE2	TYR	47	A	11027	O4	GAL	401	A	3.43
30	1314	C	CYS	182	A	11026	O3	GAL	401	A	3.82
31	1316	CB	CYS	182	A	11025	O2	GAL	401	A	3.59
32	1316	CB	CYS	182	A	11026	O3	GAL	401	A	3.62
33	1318	N	GLY	183	A	11026	O3	GAL	401	A	2.9
34	1319	CA	GLY	183	A	11026	O3	GAL	401	A	3.48
35	1320	C	GLY	183	A	11026	O3	GAL	401	A	3.55
36	1321	O	GLY	183	A	11026	O3	GAL	401	A	3.42
37	1337	CE	MET	185	A	11021	C4	GAL	401	A	3.68
38	1337	CE	MET	185	A	11022	C5	GAL	401	A	3.84
39	1337	CE	MET	185	A	11029	O6	GAL	401	A	3.83
40	1343	CG	ASP	186	A	11025	O2	GAL	401	A	3.29
41	1344	OD1	ASP	186	A	11020	C3	GAL	401	A	3.81
42	1344	OD1	ASP	186	A	11025	O2	GAL	401	A	3.63
43	1345	OD2	ASP	186	A	11019	C2	GAL	401	A	3.69
44	1345	OD2	ASP	186	A	11025	O2	GAL	401	A	2.6
45	1719	CE2	TYR	236	A	11027	O4	GAL	401	A	3.58
46	1719	CE2	TYR	236	A	11028	O5	GAL	401	A	3.47
47	1720	CZ	TYR	236	A	11027	O4	GAL	401	A	3.46
48	1721	OH	TYR	236	A	11019	C2	GAL	401	A	3.59
49	1721	OH	TYR	236	A	11020	C3	GAL	401	A	3.69
50	1721	OH	TYR	236	A	11021	C4	GAL	401	A	3.58
51	1721	OH	TYR	236	A	11026	O3	GAL	401	A	3.5
52	1721	OH	TYR	236	A	11027	O4	GAL	401	A	2.47
53	2552	N	GLY	346	A	11024	O1	GAL	401	A	3.23
54	2552	N	GLY	346	A	11028	O5	GAL	401	A	3.63
55	2553	CA	GLY	346	A	11024	O1	GAL	401	A	3.61
56	2553	CA	GLY	346	A	11028	O5	GAL	401	A	3.58
Number of hydrogen bonds: 8											
Number of non-bonded contacts: 56											

Table 30. 6Q91-HFK PDBsum interaction.

Hydrogen Bonds											
	Atom	Atom	Res	Res	Chain	Atom	Atom	Res	Res	Chain	Distance (Å)
	no.	name	name	no.		no.	name	name	no.		
1	1015	OG	SER	141	A	11052	N17	HFK	402	A	2.97
Non-bonded Contacts											
	Atom	Atom	Res	Res	Chain	Atom	Atom	Res	Res	Chain	Distance (Å)
	no.	name	name	no.		no.	name	name	no.		
1	439	OG1	THR	61	A	11036	C26	HFK	402	A	3.71
2	553	OG1	THR	77	A	11035	C24	HFK	402	A	3.74
3	567	OG	SER	79	A	11035	C24	HFK	402	A	3.62
4	567	OG	SER	79	A	11049	C25	HFK	402	A	3.79
5	588	OD1	ASP	83	A	11045	C11	HFK	402	A	3.46
6	588	OD1	ASP	83	A	11046	C12	HFK	402	A	3.78
7	752	CB	ARG	105	A	11046	C12	HFK	402	A	3.79
8	756	CZ	ARG	105	A	11054	O01	HFK	402	A	3.28
9	757	N	TRP	106	A	11046	C12	HFK	402	A	3.63
10	758	CA	TRP	106	A	11046	C12	HFK	402	A	3.81
11	761	CB	TRP	106	A	11048	C23	HFK	402	A	3.66
12	793	CE2	TYR	109	A	11032	C15	HFK	402	A	3.52
13	793	CE2	TYR	109	A	11052	N17	HFK	402	A	3.88
14	794	CZ	TYR	109	A	11032	C15	HFK	402	A	3.84
15	795	OH	TYR	109	A	11032	C15	HFK	402	A	3.56
16	795	OH	TYR	109	A	11041	C06	HFK	402	A	3.49
17	795	OH	TYR	109	A	11051	N16	HFK	402	A	3.84
18	941	CG1	VAL	129	A	11049	C25	HFK	402	A	3.6
19	954	CB	SER	131	A	11036	C26	HFK	402	A	3.8
20	954	CB	SER	131	A	11049	C25	HFK	402	A	3.8
21	981	CG	LEU	135	A	11034	C21	HFK	402	A	3.54
22	981	CG	LEU	135	A	11047	C18	HFK	402	A	3.78
23	981	CG	LEU	135	A	11055	O22	HFK	402	A	3.46
24	982	CD1	LEU	135	A	11033	C20	HFK	402	A	3.62
25	982	CD1	LEU	135	A	11034	C21	HFK	402	A	3.7
26	982	CD1	LEU	135	A	11053	N19	HFK	402	A	3.85
27	983	CD2	LEU	135	A	11034	C21	HFK	402	A	3.81
28	983	CD2	LEU	135	A	11048	C23	HFK	402	A	3.68
29	984	N	GLY	136	A	11051	N16	HFK	402	A	3.57
30	985	CA	GLY	136	A	11040	C05	HFK	402	A	3.68
31	985	CA	GLY	136	A	11051	N16	HFK	402	A	3.66
32	1011	CA	SER	141	A	11053	N19	HFK	402	A	3.86
33	1012	C	SER	141	A	11053	N19	HFK	402	A	3.38
34	1013	O	SER	141	A	11053	N19	HFK	402	A	3.46
35	1014	CB	SER	141	A	11047	C18	HFK	402	A	3.79
36	1014	CB	SER	141	A	11052	N17	HFK	402	A	3.66
37	1014	CB	SER	141	A	11053	N19	HFK	402	A	3.27
38	1015	OG	SER	141	A	11032	C15	HFK	402	A	3.79
39	1015	OG	SER	141	A	11047	C18	HFK	402	A	3.59
40	1015	OG	SER	141	A	11051	N16	HFK	402	A	3.46
41	1015	OG	SER	141	A	11052	N17	HFK	402	A	2.97
42	1015	OG	SER	141	A	11053	N19	HFK	402	A	3.58
43	1016	N	SER	142	A	11052	N17	HFK	402	A	3.56
44	1016	N	SER	142	A	11053	N19	HFK	402	A	3.73
45	1017	CA	SER	142	A	11052	N17	HFK	402	A	3.76
46	1657	CZ	ARG	228	A	11039	C04	HFK	402	A	3.82
47	1659	NH2	ARG	228	A	11039	C04	HFK	402	A	3.76
Number of hydrogen bonds: 1											
Number of non-bonded contacts: 47											

Table 31. 6Q91-HR8 PDBsum interaction.

Non-bonded Contacts											
	Atom	Atom	Res	Res	Chain	Atom	Atom	Res	Res	Chain	Distance (Å)
	no.	name	name	no.		no.	name	name	no.		
1	279	CD1	LEU	40	A	11063	C1	HR8	403	A	3.72
2	1544	CA	LEU	213	A	11068	O2	HR8	403	A	3.84
3	1545	C	LEU	213	A	11068	O2	HR8	403	A	3.38
4	1546	O	LEU	213	A	11062	C10	HR8	403	A	3.36
5	1546	O	LEU	213	A	11069	C9	HR8	403	A	3.53
6	1546	O	LEU	213	A	11071	C12	HR8	403	A	3.08
7	1548	CG	LEU	213	A	11064	O1	HR8	403	A	3.78
8	1548	CG	LEU	213	A	11071	C12	HR8	403	A	3.53
9	1549	CD1	LEU	213	A	11063	C1	HR8	403	A	3.85
10	1549	CD1	LEU	213	A	11064	O1	HR8	403	A	3.6
11	1549	CD1	LEU	213	A	11065	C2	HR8	403	A	3.88
12	1550	CD2	LEU	213	A	11071	C12	HR8	403	A	3.52
13	1551	N	SER	214	A	11068	O2	HR8	403	A	3.16
14	1552	CA	SER	214	A	11068	O2	HR8	403	A	3.88
15	1553	C	SER	214	A	11068	O2	HR8	403	A	3.71
16	1557	N	ASP	215	A	11068	O2	HR8	403	A	3.4
17	1557	N	ASP	215	A	11069	C9	HR8	403	A	3.69
18	1560	O	ASP	215	A	11069	C9	HR8	403	A	3.65
19	1561	CB	ASP	215	A	11068	O2	HR8	403	A	3.71
20	1579	C	LEU	218	A	11070	C11	HR8	403	A	3.83
21	1580	O	LEU	218	A	11070	C11	HR8	403	A	3.45
22	1581	CB	LEU	218	A	11056	N1	HR8	403	A	3.73
23	1583	CD1	LEU	218	A	11059	C6	HR8	403	A	3.35
24	1583	CD1	LEU	218	A	11060	C7	HR8	403	A	3.62
25	1583	CD1	LEU	218	A	11061	C8	HR8	403	A	3.69
26	1586	CA	ALA	219	A	11070	C11	HR8	403	A	3.62
27	2152	CA	LEU	295	A	11067	CL1	HR8	403	A	3.75
28	2154	O	LEU	295	A	11067	CL1	HR8	403	A	3.7
29	2157	CD1	LEU	295	A	11057	C4	HR8	403	A	3.41
30	2157	CD1	LEU	295	A	11058	C5	HR8	403	A	3.22
31	2157	CD1	LEU	295	A	11059	C6	HR8	403	A	3.37
32	2157	CD1	LEU	295	A	11060	C7	HR8	403	A	3.72
33	2157	CD1	LEU	295	A	11065	C2	HR8	403	A	3.87
34	2157	CD1	LEU	295	A	11066	C3	HR8	403	A	3.74
35	2157	CD1	LEU	295	A	11067	CL1	HR8	403	A	3.89
36	2182	CA	GLY	298	A	11067	CL1	HR8	403	A	3.45
37	2183	C	GLY	298	A	11067	CL1	HR8	403	A	3.27
38	2184	O	GLY	298	A	11067	CL1	HR8	403	A	3.63
39	2185	N	ASP	299	A	11067	CL1	HR8	403	A	3.46
40	2187	C	ASP	299	A	11067	CL1	HR8	403	A	3.74
41	2193	N	TYR	300	A	11067	CL1	HR8	403	A	3.4
42	2194	CA	TYR	300	A	11067	CL1	HR8	403	A	3.8
43	2226	CG	PHE	303	A	11057	C4	HR8	403	A	3.7
44	2228	CD2	PHE	303	A	11057	C4	HR8	403	A	3.8
45	2612	CD1	LEU	355	A	11063	C1	HR8	403	A	3.84
46	2612	CD1	LEU	355	A	11064	O1	HR8	403	A	3.69
47	2612	CD1	LEU	355	A	11065	C2	HR8	403	A	3.76
48	2613	CD2	LEU	355	A	11066	C3	HR8	403	A	3.84

Number of non-bonded contacts: 48

Table 32. Residues with RMSF above or equal to 0.2 nm (GALK1\_: indicates the simulations with substrate-bound GALK1, GALK1\_F: indicates the simulations with substrate- and fragment-bound GALK1. The numbers in groups denote the simulation replicas (the first, second, and third simulations).

<b>Groups</b>	<b>Residues</b>	<b>RMSF Values (nm)</b>
GALK1_1	1-7, 95-97, 164, 176-181, 228, 229	1.41, 1.17, 1.09, 0.85, 0.71, 0.52, 0.34, 0.22, 0.26, 0.21, 0.20, 0.21, 0.30, 0.31, 0.32, 0.31, 0.21, 0.23, 0.22
GALK1_2	1-3, 96-102, 178-180, 392	0.46, 0.31, 0.23, 0.25, 0.21, 0.22, 0.22, 0.20, 0.22, 0.22, 0.21, 0.22, 0.25
GALK1_3	1-3, 5, 96-98, 179, 180, 228-231, 233	0.57, 0.39, 0.23, 0.32, 0.23, 0.22, 0.21, 0.32, 0.28, 0.30, 0.28, 0.29, 0.21, 0.21
GALK1_F_1	1-3, 80, 82, 95, 96, 349	0.46, 0.32, 0.21, 0.21, 0.27, 0.20, 0.21, 0.20
GALK1_F_2	1-5, 81, 83, 95, 96, 227-233, 349, 392	0.72, 0.53, 0.40, 0.21, 0.25, 0.21, 0.21, 0.23, 0.22, 0.22, 0.34, 0.34, 0.25, 0.24, 0.36, 0.32, 0.23, 0.35,
GALK1_F_3	1, 2, 321, 392	0.44, 0.27, 0.20, 0.31

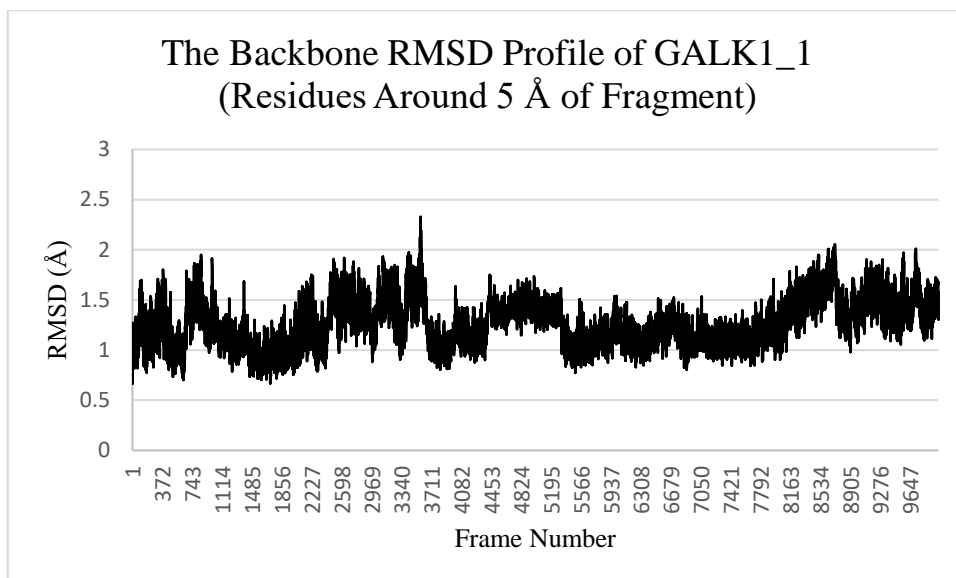


Figure 46. The backbone RMSD profile of GALK1\_1.

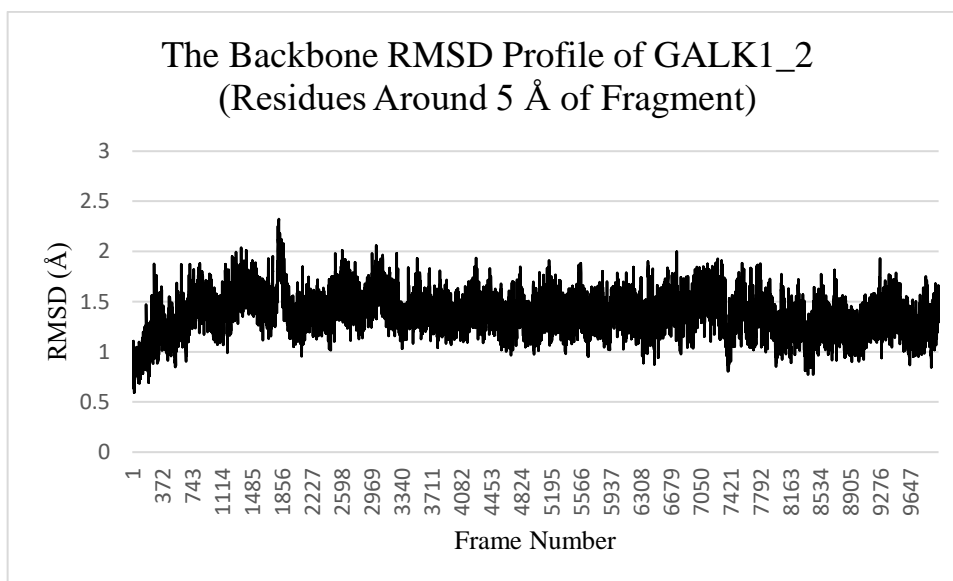


Figure 47. The backbone RMSD profile of GALK1\_2.

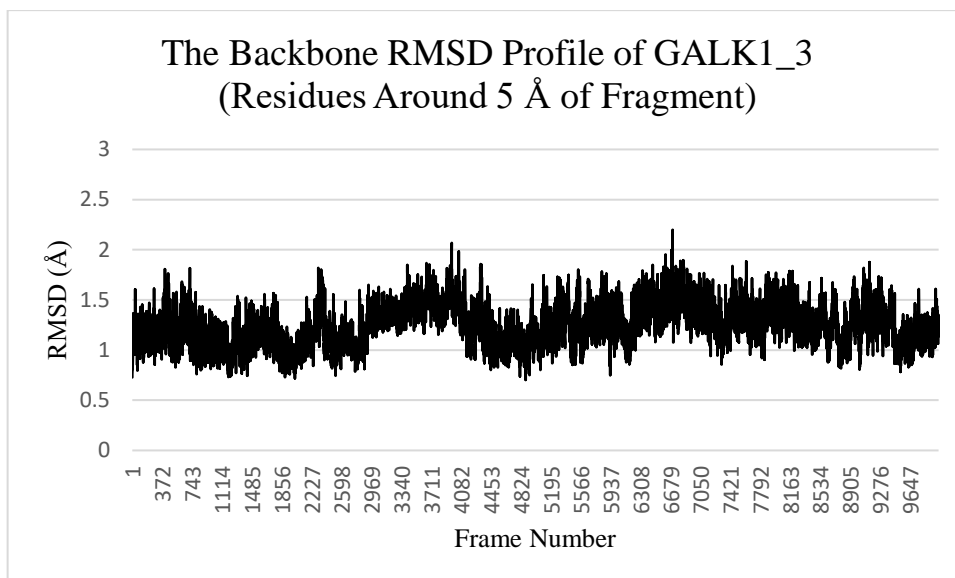


Figure 48. The backbone RMSD profile of GALK1\_3.

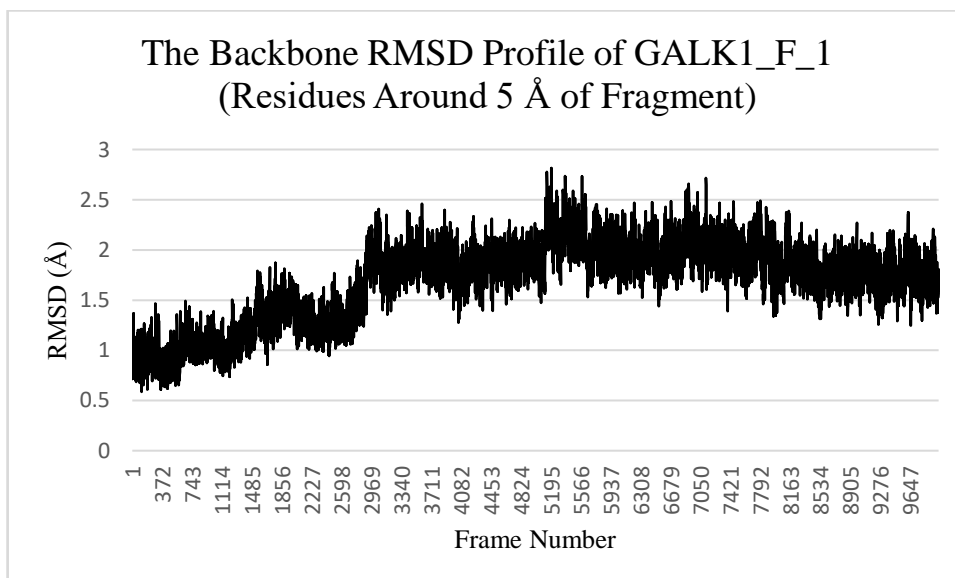


Figure 49. The backbone RMSD profile of GALK1\_F\_1.

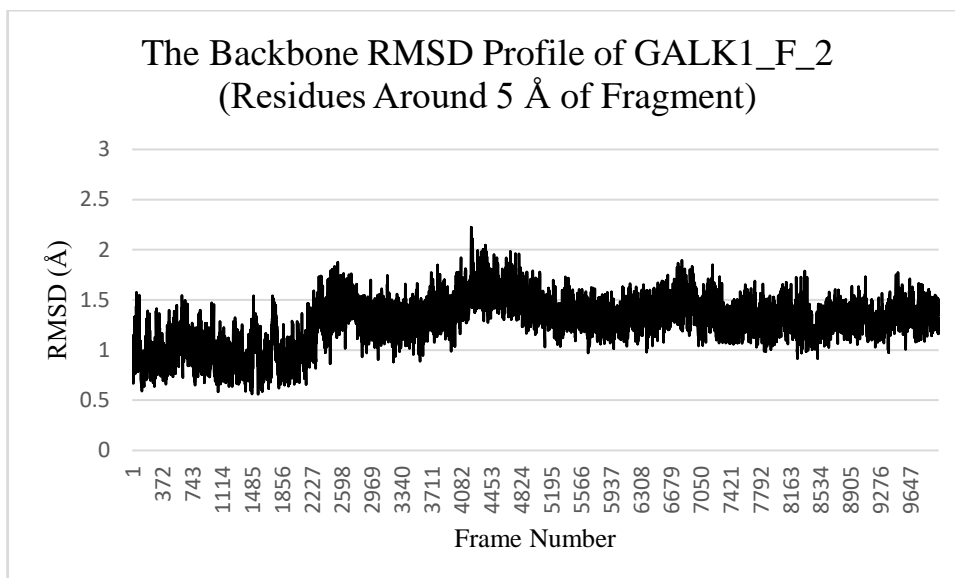


Figure 50. The backbone RMSD profile of GALK1\_F\_2.

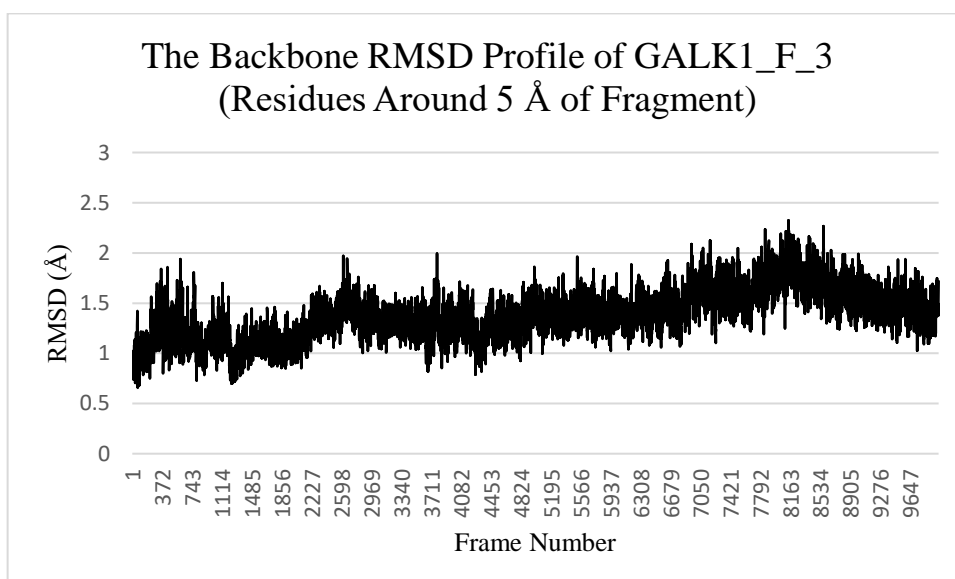


Figure 51. The backbone RMSD profile of GALK1\_F\_3.