T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL

COMPUTATIONAL INVESTIGATION OF POTENTIAL ALLOSTERIC ERK5 KINASE INHIBITORS

İLAYDA ERDOĞAN

A THESIS OF MASTER OF SCIENCE
DEPARTMENT OF BIOENGINEERING
BIOINFORMATICS AND SYSTEMS BIOLOGY PROGRAM

ADVISOR: ASST. PROF. DR. ONUR SERÇİNOĞLU

MARCH 2024

T.R.

GEBZE TECHNICAL UNIVERSITY

GRADUATE SCHOOL

COMPUTATIONAL INVESTIGATION OF POTENTIAL ALLOSTERIC ERK5 KINASE INHIBITORS

İLAYDA ERDOĞAN

A THESIS OF MASTER OF SCIENCE

DEPARTMENT OF BIOENGINEERING

BIOINFORMATICS AND SYSTEMS BIOLOGY PROGRAM

ADVISOR: ASST. PROF. DR. ONUR SERÇİNOĞLU

2. ADVISOR: PROF. DR. ASUMAN DEMİROĞLU ZERGEROĞLU

MARCH 2024

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

POTANSİYEL ALLOSTERİK ERK5 KİNAZ İNHİBİTÖRLERİNİN HESAPLAMALI İNCELENMESİ

İLAYDA ERDOĞAN

YÜKSEK LİSANS TEZİ
BİYOMÜHENDİSLİK ANABİLİM DALI
BİYOİNFORMATİK VE SİSTEM BİYOLOJİSİ PROGRAMI

DANIŞMAN: DR. ÖĞR. ÜYESİ ONUR SERÇİNOĞLU
2. DANIŞMAN: PROF. DR. ASUMAN DEMİROĞLU ZERGEROĞLU

MART 2024

MASTER of SCIENCE JURY APPROVAL FORM

A thesis submitted by İlayda ERDOĞAN, defended on 05/03/2024 before the jury formed with the 12/02/2024 date and 2024/13 numbered decision of the GTU Graduate Administration Board, has been accepted as a MASTER of SCIENCE thesis in the Department of Bioengineering, Bioinformatics and Systems Biology Program.

JURY

JURY MEMBERS

(THESIS ADVISOR) : Dr. Öğr. Üyesi Onur SERÇİNOĞLU

MEMBER : Doç.Dr. Pemra ÖZBEK SARICA

MEMBER : Prof. Dr. Tunahan ÇAKIR

APPROVAL

Gebze Technical University Graduate Administrative Board

…..…/…..…/..…… date and ……/…….. numbered decision.

SIGNATURE/SEAL

ABSTRACT

Protein kinases are enzymes that regulate signaling pathways in mammalian cells by catalyzing the phosphorylation of various protein substrates. ERK5 is a member of the MAPK family of kinases that has a prominent role in cancer-related cellular functions. Recent studies have shown that the structure of ERK5 is unique, featuring a trans-activation domain (TAD) in addition to the kinase domain. Due to the role of ERK5 in cancer development, there is an urgent need for effective and selective ERK5 inhibitors. However, the majority of ERK5 inhibitors identified are Type 1 kinase inhibitors and inhibit competitively with ATP. Current orthosteric ERK5 kinase inhibitors either suffer from selectivity issues or cause paradoxical activation of the TAD. Allosteric protein kinase inhibitors, on the other hand, bind to pockets located in alternative regions instead of the ATP binding region, which is conserved in kinases, and therefore have high selectivity for the kinase they target. The aim of this thesis work is to identify potential allosteric and selective ERK5 inhibitors using structure-based drug discovery methods in a drug repositioning context. First, ERK5 structures available from the Protein Data Bank were subjected to a coarse-grained biomolecular simulation procedure using the CABS-Flex algorithm to generate an ensemble that can be used to extract conformations to be used as starting structures in atomistic molecular dynamics (MD) simulations. A clustering analysis of obtained ensemble yielded five representative clusters, which were subjected to one microsecond-long MD simulations using gromacs 2023.1 and CHARMM36 force-field. Potentially druggable binding pockets in ERK5 kinase conformations obtained from these simulations were then identified using the DogSiteScorer algorithm. Hierarchical clustering was then applied to identify stable druggable and allosteric potential ligand-binding pockets. Molecular docking simulations of FDA-approved drugs into identified pockets were performed using Smina, LeDock and GOLD. In order to validate the docking results, 100 nanosecond-long simulations were performed by gromacs for selected molecules, and MM/PBSA binding free energies were computed. As a result, tafluprost (-37.1 kcal/mol), tirofiban (-31.38 kcal/mol), panobinostat (-29.27 kcal/mol), nebivolol (-29.29 kcal/mol), dexamethasone phosphate (-30.08 kcal/mol) and pitavastatin (-29.38 kcal/mol) were found to show the highest binding affinities to allosteric binding pockets on ERK5 kinase domain structure. These drugs thus have the potential to be repurposed as allosteric ERK5 inhibitors.

Keywords: Molecular Dynamics Simulations, ERK5, Kinase Inhibitors, Drug Repurposing, Molecular Docking

ÖZET

Protein kinazlar, çeşitli protein substratlarının fosforilasyonunu katalize ederek memeli hücrelerinde sinyal yolaklarını düzenleyen enzimlerdir. ERK5, kanserle ilişkili hücresel işlevlerde önemli bir role sahip olan MAPK kinaz ailesinin bir üyesidir. Son çalışmalar, ERK5'in yapısının benzersiz olduğunu, kinaz alanına ek olarak bir trans-aktivasyon alanı (TAD) içerdiğini göstermiştir. ERK5'in kanser gelişimindeki rolü nedeniyle etkili ve seçici ERK5 inhibitörlerine ihtiyaç vardır. Ancak tanımlanan ERK5 inhibitörlerinin çoğu Tip 1 kinaz inhibitörüdür ve ATP ile yarışmalı olarak inhibe ederler. Mevcut ortosterik ERK5 kinaz inhibitörleri ya seçicilik sorunlarından muzdariptir ya da TAD alanının paradoksal aktivasyonuna neden olur. Allosterik protein kinaz inhibitörleri ise kinazlarda korunan ATP bağlanma bölgesi yerine alternatif bölgelerde bulunan ceplere bağlanırlar ve bu nedenle hedefledikleri kinaz için yüksek seçiciliğe sahiptirler. Bu tez çalışmasının amacı, ilaç yeniden konumlandırma bağlamında yapı bazlı ilaç keşif yöntemlerini kullanarak potansiyel allosterik ve seçici ERK5 inhibitörlerini tanımlamaktır. İlk olarak, Protein Veri Bankasından temin edilebilen ERK5 yapıları, atomistik moleküler dinamik (MD) simülasyonlarında başlangıç yapıları olarak kullanılacak konformasyonları çıkarmak için kullanılabilecek bir örneklem oluşturmak üzere CABS-Flex algoritması kullanılarak kaba taneli bir biyomoleküler simülasyon prosedürüne tabi tutuldu. Elde edilen topluluğun kümeleme analizi, gromacs 2023.1 ve CHARMM36 kuvvet alanı kullanılarak bir mikrosaniye uzunluğundaki MD simülasyonlarına tabi tutulmak üzere beş temsili başlangıç yapısı ortaya çıkardı. Bu simülasyonlardan elde edilen ERK5 kinaz konformasyonlarındaki potansiyel olarak ilaçlanabilir bağlanma cepleri daha sonra DogSiteScorer algoritması kullanılarak tanımlandı. Daha sonra stabil, ilaçlanabilir ve allosterik potansiyel ligand bağlama ceplerini tanımlamak için hiyerarşik kümeleme uygulandı. FDA onaylı ilaçların belirlenen ceplere moleküler yanaştırma simülasyonları Smina, LeDock ve GOLD kullanılarak gerçekleştirildi. Sonuçları doğrulamak amacıyla, seçilen moleküller için gromacs tarafından 100 nanosaniye uzunluğunda simülasyonlara tabi tutuldu ve MM/PBSA bağlanma serbest enerjileri hesaplandı. Sonuç olarak tafluprost (-37,1 kcal/mol), tirofiban (-31,38 kcal/mol), panobinostat (-29,27 kcal/mol), nebivolol (-29,29 kcal/mol), deksametazon fosfat (-30,08 kcal/mol) ve pitavastatin (-29.38 kcal/mol) moleküllerinin ERK5 kinaz alanı yapısındaki allosterik bağlanma ceplerine en yüksek bağlanma afinitesini gösterdiği bulunmuştur. Dolayısıyla bu ilaçların allosterik ERK5 inhibitörleri olarak yeniden kullanılma potansiyeli vardır.

Anahtar Kelimeler: Moleküler Dinamik Simülasyonu, ERK5, Kinaz İnhibitörü, İlaç Yeniden Konumlandırma, Moleküler Yanaştırma

ACKNOWLEDGEMENTS

I have started my master's degree in bioinformatics and systems biology at Gebze Technical University in 2021. I would like to thank my professors who helped me discover my interest in this field throughout my undergraduate life and all the professors with whom I took courses during my master's degree, especially my advisor Onur Serçinoğlu, for his great guidance and support. I would like to thank my family and friends who have supported me in everything I wanted to do throughout my life. I would also like to thank Health Institutes of Turkiye (TUSEB) for providing project grant. (Project No: 16877)

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

MAPK	:	Mitogen-activated protein kinase
ERK5	:	Extracellular signal-related kinase 5
TAD	:	Transactivation Domain
PDB	:	Protein Data Bank
RMSD	:	Root-Mean-Square-Deviation
RMSF	:	Root-Mean-Square-Fluctuation
MD	:	Molecular Dynamics
MMPBSA	:	Molecular Mechanics Poisson-Boltzmann Surface Area
AAMD	:	All-atom Molecular Dynamics

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Cancer

Cancer is a highly-devastating disease characterized by a wide range of distinct clinical features. Cancer results in millions of fatalities worldwide annually [1]. The main characteristic of cancer is the uncontrolled division of cells, making the identification of compounds that can inhibit cell proliferation and induce tumor regression the primary focus of therapeutic research [2]. Cancer is often associated with the accumulation of multiple genetic mutations that interfere with crucial cellular communication pathways. These pathways are organized as interconnected networks that allow for continuous signaling. The components of these pathways interact in a binary manner, where the interaction between two proteins can either directly activate or inhibit the next factor in the pathway. The abnormal functioning of various signaling pathways in cancer is caused by changes in genes, gene expression patterns, and epigenetic modifications. These alterations affect various cellular processes and outcomes [3].

1.2. Kinases

Protein kinases are enzymes that catalyze the phosphorylation of protein substrates and thereby function as regulators of signaling cascades in biological cells. Slightly more than 2% of all human genes are known to encode for kinases, which makes roughly 500 protein kinase genes [4]. The structure of a kinase basically consists of two subdomains: a small, primarily β-sheet N-terminal subdomain, and a larger, primarily helical C-terminal subdomain. The cleft between the two subdomains serves as a binding site for both ATP and metal ions [5]. Mutations and interactions far from the substrate binding active sites of kinases can also alter their activity [6]. The Protein Data Bank (PDB) presently contains more than 5000 X-ray structures of complexes formed between kinases and small molecules. These structures encompass a minimum of 3000 inhibitors and involve 300 distinct kinases [7]. Taken together, these structures

provide a rich resource for exploring characteristics of kinase-inhibitor interactions, and a knowledgebase for developing novel kinase inhibitors.

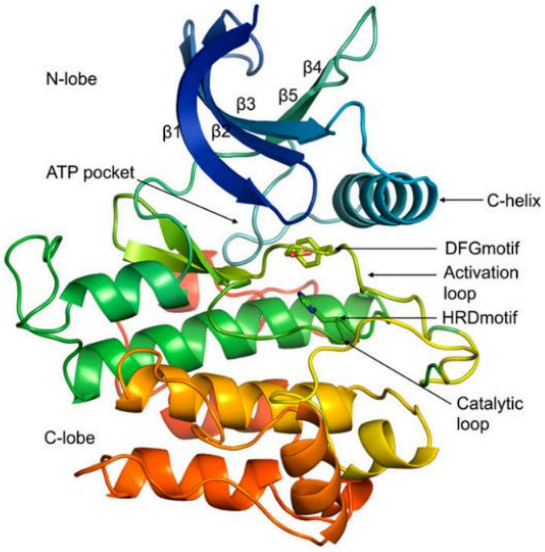


Figure 1.1. Structure of a typical protein kinase domain displaying the ATP binding site and conserved elements [8].

1.3. Kinases in Cancer

Phosphorylation plays a vital role in the regulation of various cellular processes, including proliferation, cell cycle control, apoptosis, motility, growth, and differentiation [9]. Kinases can contribute to the development of cancer through various mechanisms, including dysregulated expression and amplification, abnormal phosphorylation, mutations, chromosomal translocation, and epigenetic regulation. In some instances, phosphorylation of these kinases serves as a more reliable indicator than the expression level of the kinase itself. Also, mutations in genes, including kinases, are recognized as one of the most significant routes towards the development and progression of cancer. These mutations can cause kinases to remain permanently active, leading to a range of abnormal cellular behaviors that play a role in initiating or advancing cancer [9].

The Ras/RAF/MEK/ERK pathway, also known as the MAPK signaling pathway, is a key pathway involved in the regulation of cellular processes such as proliferation, differentiation, and survival [10]. In certain types of cancer, the prognosis can be linked to the phosphorylation of specific kinases, including EGFR, ERBB2, ERK [11], AURKA, p38, and AKT. Abnormal activation of receptor tyrosine kinases or specific mutations in the RAS or RAF genes often lead to changes in the RAS-RAF-ERK-MEK-MAPK pathway, which have been frequently observed in human cancer. Consequently, this pathway has emerged as a promising target for cancer treatment. In recent times, there have been significant advancements in the development of small-molecule inhibitors that specifically target this pathway. These inhibitors are currently undergoing evaluation in clinical trials [12].

Table 1.1. Numbers of FDA-approved protein kinase inhibitors and their respective targets [13].

Kinase family	Class of Kinase	US FDA approved
EGFR/ErbB	RY	10
JAK	NRY	9
VEGFR	RY	9
BCR-Abl	NRY	6
ALK	RY	5
FGFR	RY	5
CDK4/6	S/T	4
MEK1/2	Y/T	4
BTK	NRY	4
BRAF	S/T	3
FKBP	S/T	3
Flt3	RY	3
MET	RY	3
RET	RY	2
ROCK	S/T	2
TRKA	RY	2
CSF1	RY	1
Kit	RY	1
PDGFR	RY	1
ROS1	RY	1
SYK	RY	1
TYK2	NRY	1
Total		80

1.4. Kinase Inhibitors

The primary emphasis in cancer treatment revolves around targeting kinases, due to their critical roles in regulating protein activities. The development of kinase inhibitors for the treatment of human tumors began in the 1970s, with the first oncogene being identified as a protein kinase in 1978 [14]. Since then, a diverse array of kinase inhibitors, derived from both synthetic and natural sources, have been employed in the treatment of cancer [15]. By inhibiting kinase activity in treated patients, these inhibitors activate anti-proliferative mechanisms, ultimately resulting in the remission of cancer. Targeting kinase signaling pathways in cancer treatment allows lower toxicity towards non-cancerous cells, compared to other alternatives [16]. Several kinase inhibitors have demonstrated remarkable efficacy in cancer treatment, especially against specific mutations that play a key role in the development of tumors. [17]. The majority of these inhibitors act by binding to the ATP-binding site of kinases [18], while others bind to allosteric sites [19].

According to their mechanism of action and the site on the kinase structure to which they bind, kinase inhibitors are classified into several types. Type I kinase inhibitors are described as "small molecules that bind to the active conformation of a kinase in the ATP pocket, type II inhibitors are "small molecules that bind to an inactive conformation of a kinase," and type III and IV inhibitors are known as non-ATP competitive and allosteric inhibitors [20].

The field of kinase inhibitor drug discovery has evolved since imatinib, the first protein kinase inhibitor approved for therapeutic use. The use of kinase inhibitors such as imatinib has led to improved survival rates among patients with gastrointestinal stromal tumors (GIST) and myeloid leukemia (CML). As a result of their enhanced clinical effectiveness, there is now a wide range of small-molecule kinase inhibitors readily accessible for treatment.

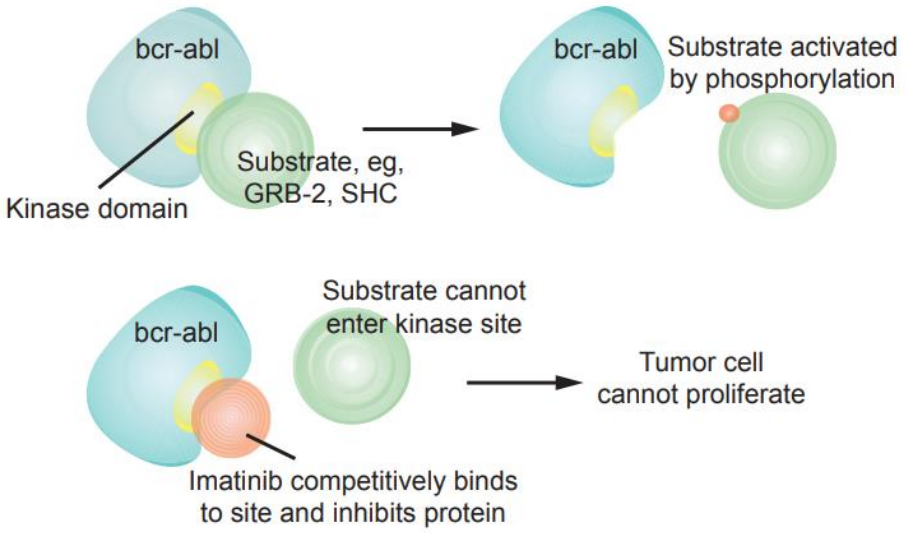


Figure 1.2. Mechanism of action of imatinib [21].

As of 2023, 80 kinase inhibitors have been approved by the US Food and Drug Administration (FDA) [13]. Utilizing drugs that inhibit kinase signaling through allosteric mechanisms provides a promising strategy for targeted therapies. The majority of allosteric kinase inhibitors developed thus far belong to type III and IV, demonstrating significant selectivity within the kinome (collection of protein kinases encoded by the human genome) but often lacking subtype selectivity within the same kinase family. However, recent advancements have led to the development of allosteric inhibitors that exhibit exceptional selectivity for specific kinase subtypes and bind outside the catalytic kinase domain [22].

Table 1.2. FDA-approved small molecule protein kinase inhibitors, their protein kinase targets, and therapeutic indications [13].

Drug	Code	Company	Trade name	Year approved	Primary targets
Abemaciclib	LY2835219	Lilly	Verzenio	2017	CDK4/6
Abrocitinib	PF-04965842	Pfizer	Cibinqo	2022	JAK1
Acalabrutinib	ACP-196	Acerta Pharma	Calquence	2017	BTK
Afatinib	BIBW 2992	Boehringer Ingelheim	Tovok	2013	ErbB1/2/4
Alectinib	CH5424802	Roche	Alecensa	2015	ALK, RET
Asciminib	ABL001	Novartis	Scemblix	2021	BCR-Abl
Avapritinib	BLU285	Blueprint Medicines	Ayvakit	2020	PDGFRα
Axitinib	AG-013736	Pfizer	Inlyta	2012	VEGFR1/2/3
Baricitinib	LY 3009104	Lilly	Olumiant	2018	JAK1/2
Belumosudil	KD025	Kadmon Pharma	Rezurock	2021	ROCK2
Binimetinib	MEK162	Array BioPharma	Mektovi	2018	MEK1/2
Bosutinib	SKI-606	Pfizer	Bosulif	2012	BCR-Abl
Brigatinib	AP 26113	Ariad Pharm	Alunbrig	2017	ALK
Cabozantinib	BMS-907351	Exelixis	Cometriq & Cabometyx	2012	RET, VEGFR2
Capmatinib	INC-280	Novartis	Tabrecta	2020	MET (HGFR)
Ceritinib	LDK378	Novartis	Zykadia	2014	ALK
Cobimetinib	GDC-0973	Genentech	Cotellic	2015	MEK1/2
Crizotinib	PF 2341066	Pfizer	Xalkori	2011	ALK, ROS1
Dabrafenib	GSK2118436	GSK	Tafinlar	2013	B-Raf
Dacomitinib	PF-00299804	Pfizer	Visimpro	2018	EGFR
Dasatinib	BMS-354825	Bristol Myers Squibb	Sprycel	2006	BCR-Abl
Deucravacitinib	BMS-986165	Bristol Myers Squibb	Sotyktu	2022	TYK2
Encorafenib	LGX818	Array BioPharma	Braftovi	2018	B-Raf
Entrectinib	RXDX-101	Genentech	Rozlytrek	2019	TRKA/B/C, ROS1
Erdafitinib	JNJ-42756493	Jansen Pharm	Balversa	2019	FGFR1/2/3/4
Erlotinib	OSI-774	Genentech	Tarceva	2004	EGFR
Everolimus	RAD001	Novartis	Afinitor	2009	FKBP12/mTOR
Fedratinib	TG101348	Celgene	Inrebic	2019	JAK2
Fostamatinib	R788	Rigel Pharma.	Tavalisse	2018	Syk
Futibatinib	TAS_120	Tiaho Pharma	Lytgobi	2022	FGFR2
Gefitinib	ZD1839	AstraZeneca	Iressa	2003	EGFR

		Table 1.2. Continued.			
Gilteritinib	ASP2215	Astellas Pharma	Xospata	2018	Flt3
Ibrutinib	PCI-32765	Johnson & Johnson	Imbruvica	2013	BTK
Imatinib	STI571	Novartis	Gleevec	2001	BCR-Abl
Infigratinib	BGJ 398	QED Therapeutics	Truseltiq	2021	FGFR2
Lapatinib	GW572016	GSK	Tykerb	2007	EGFR, ErbB2/HER2
Larotrectinib	LOXO-101	Bayer	Vitrakvi	2018	TRKA/B/C
Lenvatinib	AK175809	Easai Co.	Lenvima	2015	VEGFR, RET
Lorlatinib	PF-06463922	Pfizer	Lorbrena	2018	ALK
Midostaurin	CPG 41251	Novartis	Rydapt	2017	Flt3
Mobocertinib	TAK-788	Takeda Pharm.	Exkivity	2021	EGFR
Neratinib	HKI-272	Puma Biotech	Nerlynx	2017	ErbB2/HER2
Netarsudil	AR11324	Aerie Pharma	Rhopressa	2018	ROCK1/2
Nilotinib	AMN107	Novartis	Tasigna	2007	BCR-Abl
Nintedanib	BIBF-1120	Boehringer Ingelheim	Vargatef	2014	FGFR1/2/3
Osimertinib	AZD-9292	AstraZeneca	Tagrisso	2015	EGFR
Pacritinib	SB1518	CTI BioPharma	Vonjo	2022	JAK2
Palbociclib	PD-0332991	Parke-Davis	Ibrance	2015	CDK4/6
Pazopanib	GW786034	GSK	Votrient	2009	VEGFR1/2/3
Pemigatinib	INCB054828	Incyte Corp.	Pemazyre	2020	FGFR2
Pexidartinib	PLX3397	Plexxikon Inc	Turalio	2019	CSF1R
Pirtobrutinib	LOXO-305	Lilly	Jaypirca	2023	BTK
Ponatinib	AP 24534	Ariad Pharm	Iclusig	2012	BCR-Abl
Pralsetinib	Blu-667	Blueprint Medicines	Gavreto	2020	RET
Regorafenib	BAY 73–4506	Bayer	Stivarga	2012	VEGFR1/2/3
R406 active metabolite of fostamatinib		Rigel Pharma.		2018	Syk
Ribociclib	LEE011	Novartis	Kisqali	2017	CDK4/6
Ripretinib	DCC-2618	Decipera Pharma.	Qinlock	2020	Kit, PDGFRα
Ruxolitinib	INCB-018424	Incyte Corp.	Jakafi	2011	JAK1/2/3,Tyk
Selpercatinib	CEGM9YBNG	Lilly	Retevmo	2020	RET
Selumetinib	AZD6224	AstraZeneca	Koselugo	2020	MEK1/2
Sirolimus	AY 22989	Wyeth, LLC	Rapamycin	1999	FKBP12/mTOR
Sorafenib	BAY 43–9006	Bayer	Nexavar	2005	VEGFR1/2/3
Sunitinib	SU11248	Pfizer	Sutent	2006	VEGFR2

		Table 1.2. Continued.			
Temsirolimus	CCI-779	Wyeth, LLC	Torisel	2007	FKBP12/mTOR
Tepotinib	EMD 1214063	EMD SeronoInc.	Tepmetko	2021	MET (HGFR)
Tivozanib	AV951	AVEO Pharma	Fotvida	2021	VEGFR2
Tofacitinib	CP-690550	Pfizer	Tasocitinib	2012	JAK3
Trametinib	GSK1120212	GSK	Mekinist	2013	MEK1/2
Trilaciclib	G1T28	G1 Therapeutics	Cosela	2021	CDK4/6
Tucatinib	ONT-380	Seattle Genetics	Tukysa	2020	ErbB2/HER2
Upadacitinib	ABT-494	AbbVie	Rinvoq	2019	JAK1
Vandetanib	ZD6474	Sanofi	Zactima	2011	VEGFR2
Vemurafenib	PLX-4032	Genentech	Zelboraf	2011	B-Raf
Zanubrutinib	BGB3111	BeiGene	Brukinsa	2019	BTK

1.5. ERK5 Kinase

Extracellular signal-related kinase 5 (ERK5) is a member of the mitogen-activated protein kinase (MAPK) family. A wide range of stimuli, including mitogens, cytokines, stress, and UV radiation, activate MAPK pathways [23]. The human ERK5 is expressed by the *MAPK7* gene. The protein includes 816 amino acids and is made up of an N-terminal kinase domain (78-406 aa) and a distinct C-terminal tail (410-816 aa) that has an autoinhibitory effect. A myocyte enhancer factor 2 (MEF-2)-interacting region (440-501 aa), a nuclear localization signal (NLS) (505-539 aa), and a transcriptional activation domain (TAD) (664-789 aa) are also found near the C-terminus [24].

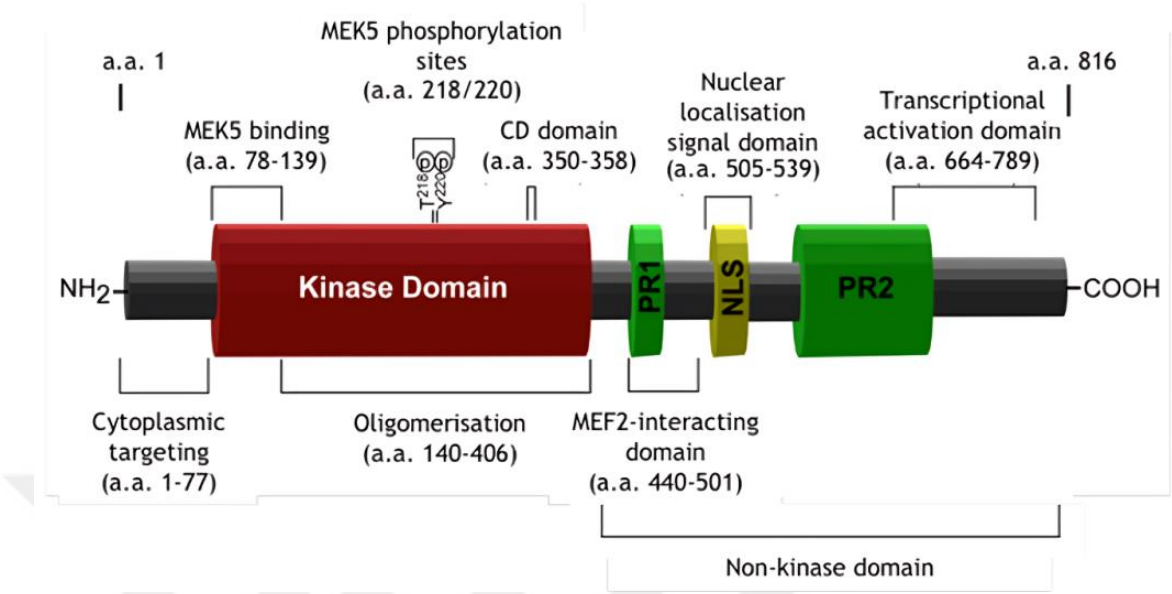


Figure 1.3. Structure of ERK5 and functional domains [25].

Dual phosphorylation of threonine and tyrosine residues within a TEY motif in the activation loop of the kinase domain is necessary for ERK5 activation [24], [26]. The importance of ERK5 signaling in some disorders, most notably inflammation and cancer, has become more widely recognized. In human endothelial cells and monocytes, for example, ERK5 has a pro-inflammatory function, while inhibiting ERK5 has an anti-inflammatory effect. ERK5 has also been linked to cancer caused by inflammation [27]. Because of its wide and distinct C-terminus, ERK5 is also known as big MAP kinase 1 (BMK1) [28]. This extended C-terminus is responsible for ERK5's unique size, which is almost twice the molecular weight of other MAPKs. The extended C-terminus serves a regulatory role. Unlike ERK1/2, which is typically activated by growth factors, and SAPKs, which is largely activated by stress factors, ERK5 may be activated by both mitogenic stimuli, such as EGF and FCS, and stress factors, such as high osmolarity and fluid shear stress [23]. Finally, several investigations have found that ERK5 signaling enhances cell proliferation, survival, motility, and invasion [29].

1.6. Paradoxical Activity and Specificity of ERK5 Kinase Inhibitors

ERK5 is associated with a wide range of cellular functions, including cellular migration, angiogenesis, proliferation, and survival [30]. Components of the ERK5 pathway are ubiquitously expressed in adult tissues. The MEK5/ERK5 MAPK cascade is a distinctive signaling pathway that is activated by both mitogens and stress-inducing factors, such as cytokines, fluid shear stress, high osmolarity, and oxidative stress. Its main role is to serve as a mechanoreceptive pathway in the endothelium, where it facilitates the transmission of various vasoprotective effects associated with the presence of laminar blood flow [23]. MEK5 and ERK5 are necessary for the formation of the heart and blood vessels, and ERK5 controls the pluripotency of mouse embryonic stem cells during development. Targeting the MEK5-ERK5 pathway in diseases, specifically cancer and inflammation has therapeutic potential [31]. However, previously-identified ERK5 kinase inhibitors were found to induce the activation of ERK5 transcriptional activity as well. In other words, even though inhibitors block ERK5 phosphorylation, they also potentiate transcriptional activity. This phenomenon is termed "paradoxical activation of ERK5", and currently is a major issue that needs to be addressed to enable the development of effective ERK5 inhibitors that do not block TAD activity. Therefore, it is important to consider both ERK5 kinase and TAD activities when evaluating the function of ERK5 and the efficacy of anti-ERK5 treatments [31].

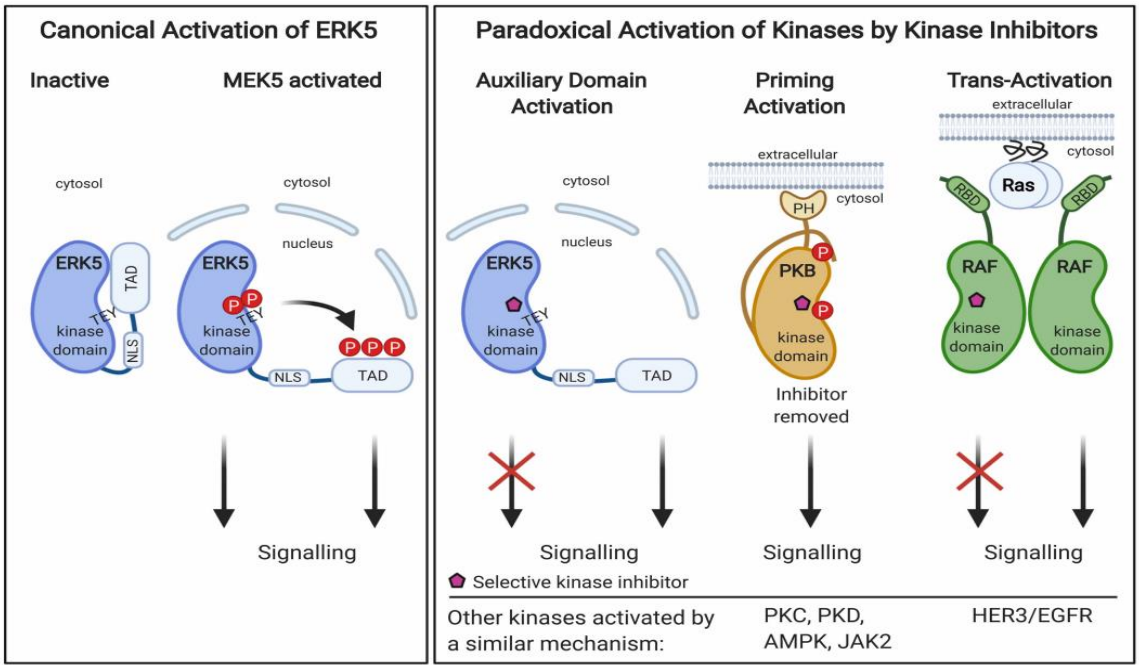


Figure 1.4. Canonical Activation of ERK5 and Paradoxical Activation of ERK5 (PKB and RAF by kinase inhibitors. TAD, transactivation domain; NLS, nuclear localisation signal; PH, pleckstrin homology domain; RBD, RAS binding domain.) [32].

1.7. Drug Repurposing

The development of new drug candidates typically involves a series of stages, which encompass the identification of a disease state or target, validation of the target, identification and optimization of a lead compound, preclinical studies, toxicity assessments, formulation, clinical trials, the approval process, and the subsequent marketing of the drugs. Additionally, post-marketing surveillance or safety monitoring is conducted to ensure ongoing drug safety [33]. Within the conventional drug discovery process, the path from concept to approved therapy is long, expensive, and filled with uncertainties. The failure rate is substantial, and the financial and time commitments are substantial. Drug repurposing presents an alternative approach that can help mitigate these challenges by identifying new applications for drugs that are already approved or being researched. This strategy enables a reduction in both the time and costs associated with pharmaceutical research. Drug repurposing, which involves investigating alternative uses for drugs that have already been approved or

progressing drugs that have been previously studied but not approved, is a fundamental strategy in the field of drug development [34]. It aims to accelerate the process and provide faster solutions. Drug repurposing involves the identification of existing drugs, evaluating their effectiveness using preclinical models, and advancing to phase II clinical trials. The identification of potential drug candidates can be achieved through computational and experimental methods, often utilizing publicly available drug databases [35].

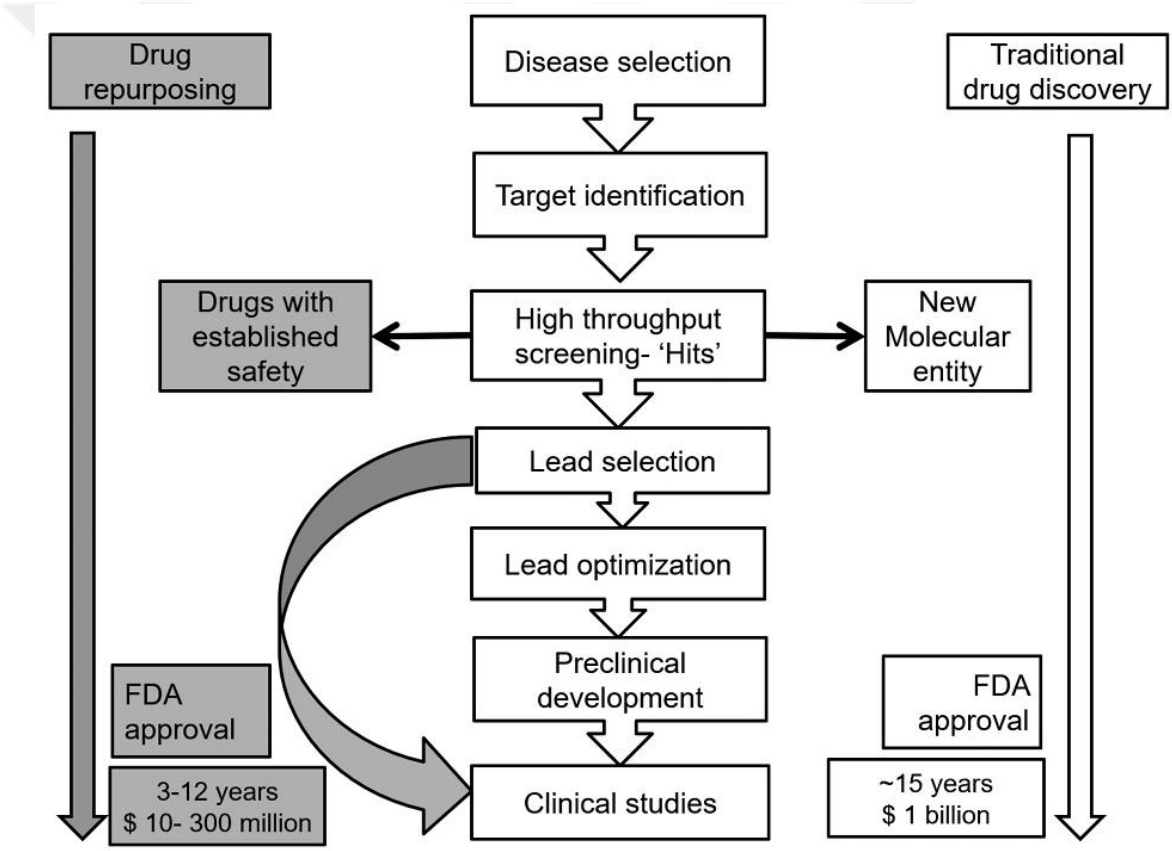


Figure 1.5. The concept of drug repurposing in the field of drug development [36].

1.8. Aim of the Study

The aim of this thesis is to discover potential allosteric inhibitors of ERK5 kinase using structure-based drug discovery methods. For this purpose, first molecular dynamics simulations were conducted to identify potential allosteric binding pockets on the ERK5 kinase structure. Then, a drug repurposing approach was used by employing consensus molecular docking simulations of FDA-approved drugs into the identified allosteric pockets. Finally, potential ERK5 inhibitors were identified through validation MD simulations and MM/PBSA binding free energy calculations.

2. METHODS

The experimental procedures employed in this study involved a series of computational techniques for the identification of potential allosteric ERK5 inhibitors. Initially, protein structures were sourced from publicly available databases, ensuring the inclusion of a diverse set of structures. To generate a representative sample, Monte Carlo simulations were performed, which allowed for the exploration of conformational space and the generation of structurally diverse protein models. The generated sample was then subjected to clustering analysis, wherein clusters were formed based on structural similarity. Representative protein structures were selected from each cluster to capture the essential structural diversity within the dataset.

Subsequently, molecular dynamics simulations were conducted on the selected representative structures. This computational technique enabled the investigation of the dynamic behavior and stability of the protein structures over time. Following the molecular dynamics simulations, prediction of the ligand-binding pockets on the protein structures was carried out. This prediction step aimed to identify regions within the proteins where small molecules, known as ligands, could potentially bind and form stable interactions.

To evaluate the binding affinity and suitability of various ligands, a library of molecules was docked into the predicted ligand-binding pockets. Consensus docking, a widely accepted approach, was employed to assess the ligand-protein interactions. Binding scores were obtained as a measure of the ligands' propensity to bind to the protein targets.

Ligands exhibiting high docking scores and consistent results across multiple docking simulations were considered as potential candidates for further analysis. To validate the initial findings, the selected ligands underwent additional validation simulations, which aimed to confirm the accuracy and reliability of the predicted binding interactions.

The overall workflow used is shown in **Figure 2.1.** Each step is explained in detail in the following sections.

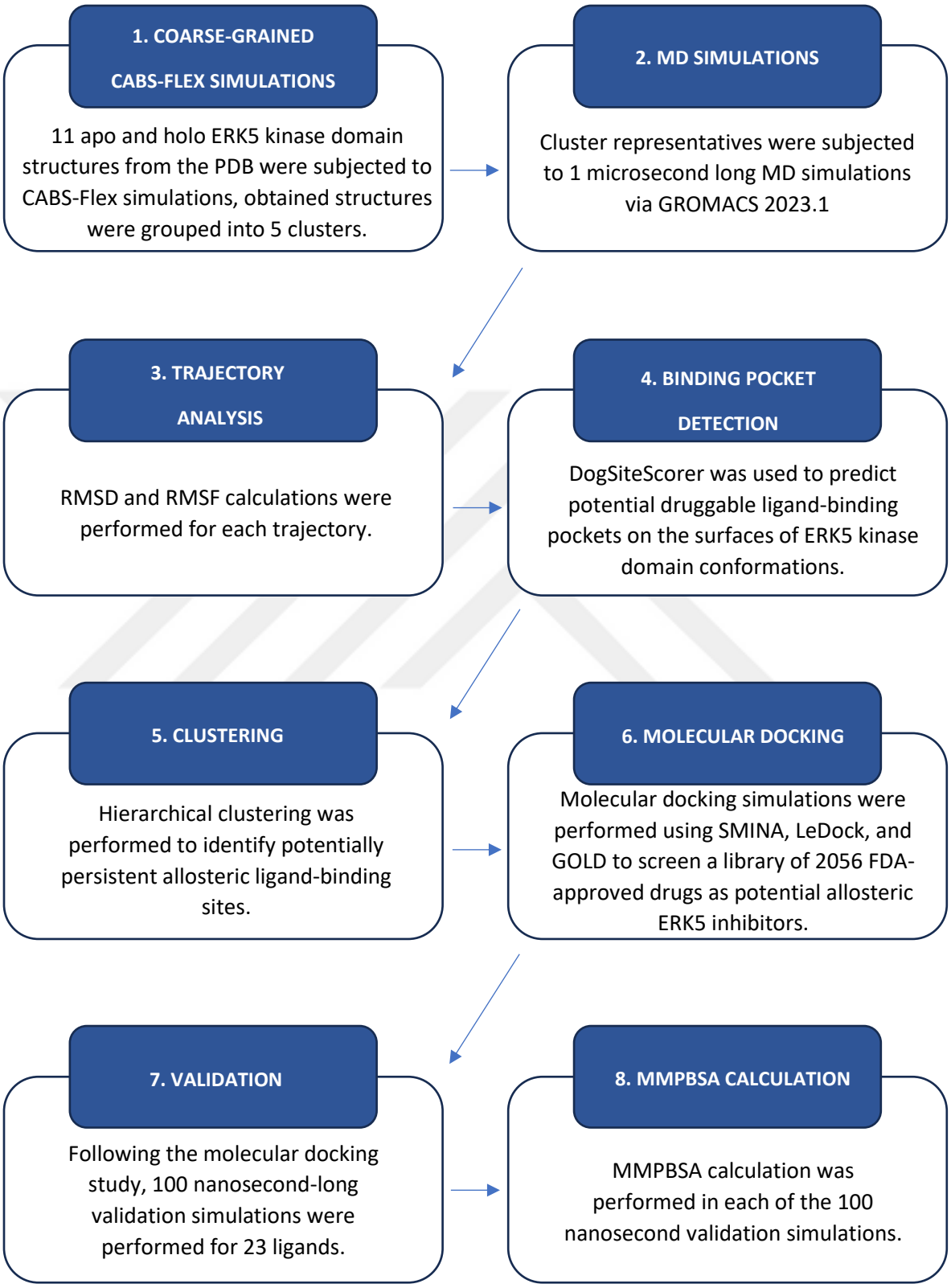


Figure 2.1. Methodology of the study.

2.1. Collection of ERK5 Kinase Structures from Protein Data Bank (PDB)

First, structures of human ERK5 kinase were identified from UniProt (ID: Q13164) [37]. A total of 11 apo and holo protein structures (PDB IDs: 4BYY, 4IC7, 4IC8, 4ZSG,4ZSJ, 4ZSL, 5BYY, 5BYZ, 5O7I, 6HKN, 6HKM) belonging to the ERK5 kinase domain were then obtained from the Protein Data Bank (PDB) [38].

Table 2.1. Details of ERK5 kinase structures, properties and ligands bound to holo structures.

PDB CODE	METHOD	RESOLUTION	CHAIN	POSITIONS	LIGAND
4B99	X-Ray	2.80 Å	A	1-397	11-cyclopentyl-2-[[2-methoxy-4-[4-(4-methylpiperazin-1-yl)piperidin-1-yl]carbonyl-phenyl]amino]-5-methyl-pyrimido[4,5-b][1,4]benzodiazepin-6-one
4IC7	X-Ray	2.60 Å	A/D	1-431	Phosphoamino phosphonic acid-adenylate ester
4IC8	X-Ray	2.80 Å	A/B	1-431	-
4ZSG	X-Ray	1.79 Å	A	47-393	3-amino-5-[(4-chlorophenyl)amino]-N-(propan-2-yl)-1H-1,2,4-triazole-1-carboxamide
4ZSJ	X-Ray	2.48 Å	A	50-393	3-amino-5-[(4-chloro-3-methylphenyl)amino]-N-(propan-2-yl)-1H-1,2,4-triazole-1-carboxamide

Table 2.1. Continued.

4ZSL	X-Ray	2.25 Å	A	53-393	3-amino-5-[(4-chlorophenyl)amino]-N-[(1S)-1-phenylethyl]-1H-1,2,4-triazole-1-carboxamide
5BYY	X-Ray	2.79 Å	A	49-394	2-{[2-ethoxy-4-(4-hydroxypiperidin-1-yl)phenyl]amino}-5,11-dimethyl-5,11-dihydro-6H-pyrimido[4,5-b][1,4]benzodiazepin-6-one
5BYZ	X-Ray	1.65 Å	A	48-395	4-({5-fluoro-4-[2-methyl-1-(propan-2-yl)-1H-imidazol-5-yl]pyrimidin-2-yl}amino)-N-[2-(piperidin-1-yl)ethyl]benzamide
5O7I	X-Ray	2.38 Å	A	46-402	4-(2-bromanyl-6-fluoranyl-phenyl)carbonyl-~{N}-pyridin-3-yl-1~{H}-pyrrole-2-carboxamide
6HKM	X-Ray	2.47 Å	A	49-395	[4-(6,7-dimethoxyquinazolin-4-yl)piperidin-1-yl]-[4-(trifluoromethyloxy)phenyl]methanone
6HKN	X-Ray	2.33 Å	A	54-393	[2-azanyl-4-(trifluoromethyloxy)phenyl]-[4-(7-methoxyquinazolin-4-yl)piperidin-1-yl]methanone

2.2 Completing Missing Residues

Missing amino acids in experimental protein structures were completed using the Modeller [39] program. Modeller is a program used for comparative protein structure modeling. Comparative structure modeling is the modeling of the three-dimensional structure of the target protein sequence by predicting it according to a previously determined template structure. In this study, the missing amino acids in the protein structures taken from PDB were completed by the Modeller program. In order not to change the conformation of the entire structure, only the regions with missing amino acids were given as input to the Modeller program. For this purpose, the ERK5 kinase region with PDB code 4ZSG, which does not contain any missing amino acids in its structure, was used as a template. An example illustration of this process for the ERK5 kinase domain with PDB code 4B99 is given in **Figure 2.2.** While completing the PDB structures, amino acids between 47 and 393 were modeled for all of them. This is because the three-dimensional structure of the first 47 amino acids has not been experimentally obtained before and the Modeller program was not able to predict this region. The study proceeded by focusing on the amino acid range of 47-393, which displayed more identifiable structural features. This region allowed for a more comprehensive investigation of the protein's structure-function relationship and facilitated subsequent analyses, such as dynamics and ligand-binding studies.

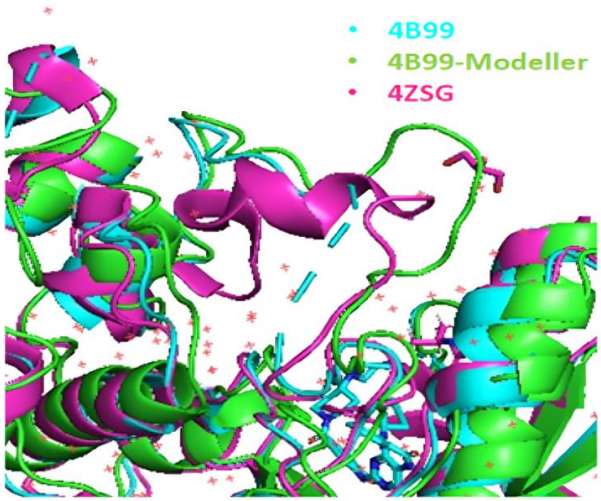


Figure 2.2. Modeller's working principle. PDB structure of 4B99 with the missing amino acids (blue), 4ZSG structure used as a template (pink), and 4B99 structure completed using the template (green).

2.3. CABS-Flex Simulations

After all structures were prepared for the molecular simulation process by completing the missing amino acids, the ligands in the structures were removed, and the resulting ERK5 kinase domain structures were subjected to CABS-Flex simulations using the CABS-Flex web server. This tool performs biomolecular conformation sampling using the Monte Carlo-based CABS-Flex algorithm [40]. The goal of a Monte Carlo (MC) simulation is to generate a collection of representative configurations for a complex macromolecular system, under specific thermodynamic conditions [41]. The Monte Carlo simulation follows a specific procedure: starting from an initial configuration of particles in a system, a Monte Carlo move is attempted to alter the particle configuration. The acceptance or rejection of this move is determined by an acceptance criterion, ensuring that the simulation samples configurations from a statistical mechanics ensemble distribution with the correct weight. After accepting or rejecting a move, the value of a property of interest is calculated. By repeating this process numerous times, an accurate average value of the property can be obtained. Unlike

molecular dynamics simulations, Monte Carlo simulations are not bound by the requirement of solving Newton's equations of motion. This flexibility allows for the clever design of moves that generate trial configurations within the desired statistical mechanics ensemble. However, since Newton's equations of motion are not solved, traditional Monte Carlo simulations do not provide dynamic information [42].

The CABS-flex pipeline utilizes the CABS simulation protocol to achieve near-native dynamics of globular proteins. The trajectory obtained from this simulation is then examined and grouped into a set of protein models that accurately capture the flexibility of the original structure. One notable characteristic of these protein models generated by the CABS model is their spatial resolution, which enables the reconstruction of all-atom physically realistic models. In the final stage of the CABS-flex, the representative set of predicted models is transformed into an all-atom representation, visualized, and converted into data formats that facilitate easy analysis and utilization [43].

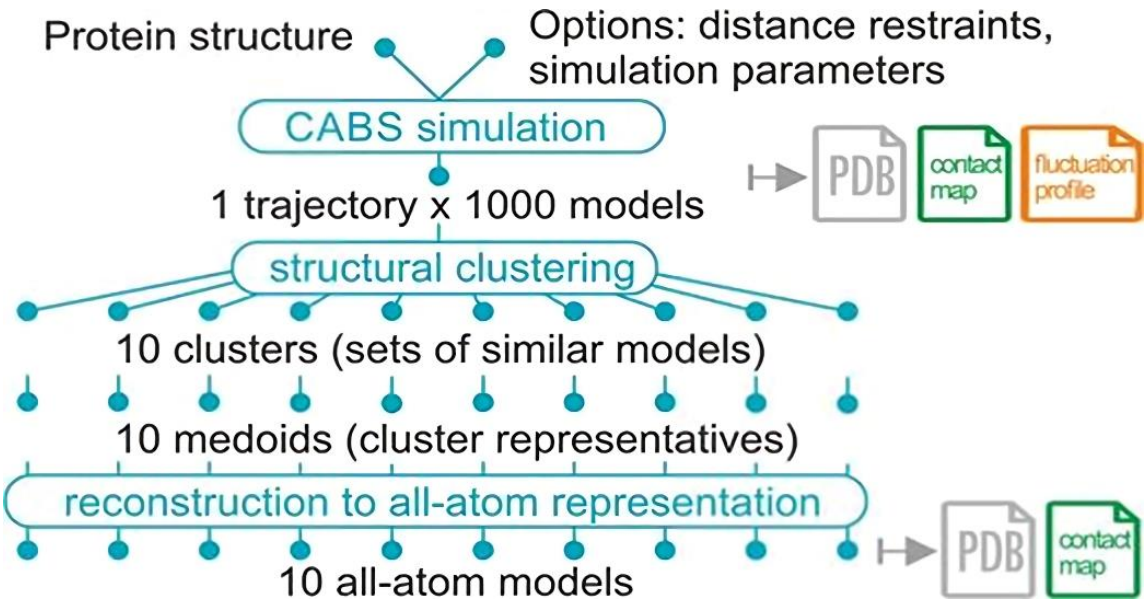


Figure 2.3. CABS-Flex pipeline [44].

The purpose of this process is to obtain an additional set of conformations that the ligand-free ERK5 kinase domain can visit, in addition to the ERK5 kinase conformations observed in the crystal structures in the PDB database for molecular dynamics simulations. CABS-Flex web server was used individually for all 11 ERK5 kinase domain structures with default parameters. These obtained models were then subjected to TTClust [45] clustering algorithm to obtain a set of representative structures for subsequent atomistic MD simulations.

2.4 Molecular Dynamics Simulations

Molecular dynamics simulation is a computational technique that integrates Newton's laws of motion to gather insights on the behavior of molecular systems. In the context of drug discovery and design, MD simulations helps refine the structure around potential ligand/drug binding sites, and also offers valuable information that aids in the comprehension and comparison of various interactions in the context of protein-ligand interaction dynamics [46]. In the atomistic "all-atom" molecular dynamics (AAMD) approach, the system being studied is made up of particles represented as atoms. These atoms represent both the solute and solvent components that interact with each other. The particles are placed in a simulation box that is large enough to accommodate them, and their movements are governed by Newton's laws of motion. To track the evolution of the system over time, a suitable algorithm such as Velocity-Verlet or Leap-frog is used to update the state of the model system through multiple time steps [47]. It is important to mention that different force fields have been created to handle different types of molecules. Several software packages for molecular dynamics simulations of biomolecules are listed in **Table 2.2.**

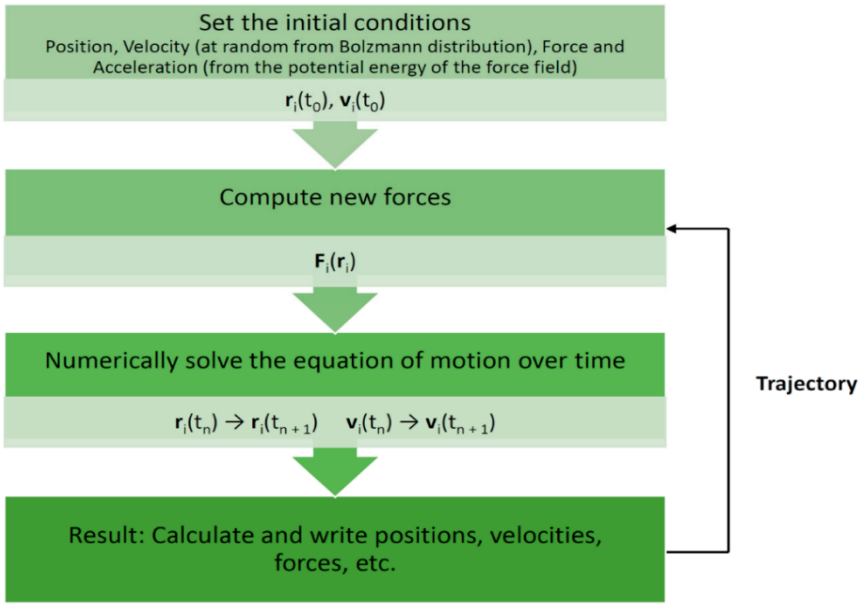


Figure 2.4. Basic steps in molecular dynamics simulation. [47].

Table 2.2. The most common software programs and force fields used in MD simulations [48].

Software	Web/reference	Processing Units	Force Field employed
Abalone	http://www.biomolecularmodeling.com/Abalone/index.html	GPU	AMBER, OPLS
ACEMD	https://www.acellera.com/products/molecular-dynamics-software-GPU-acemd	GPU	CHARM, AMBER, OPLS
AMBER	http://ambermd.org/	CPU/GPU	AMBER ff99SBildn LIPID14
NAMD	http://www.ks.uiuc.edu/Research/namd/	CPU/GPU	CHARMM, AMBER
CHARMM	http://yuri.harvard.edu	CPU/GPU	CHARMM
GROMACS	http://www.gromacs.org/	CPU/GPU	GROMOS 43A1-S3 CPU MBER99SB-ILDN
MOLDY	http://www.tchpc.tcd.ie/projects/completed/moldy	CPU	CHARMM
TINKER	http://dasher.wustl.edu/tinker/	GPU	AMBER, CHARMM, OPLS, MMFF
DESMOND	https://www.deshawresearch.com/downloads/download_desmond.cgi/	GPU	AMBER, CHARMM, OPLS
LAMMPS	http://lammps.sandia.gov/	GPU	CHARMM, AMBER, OPLS, GROMACS

Accordingly, it was decided to conduct MD simulations to screen for possible allosteric binding pockets on the ERK5 kinase surface, using a PDB structure representing each cluster as a starting point. Each of the clusters obtained as a result of the analysis has similar protein structures. Instead of all 11 structures taken from the Protein Data Bank, representatives from each cluster were selected to reduce the number of simulations. The structures 4IC8 (apo state), 4B99, 4IC8, 4ZSG, 4ZSJ, and 6HKN, along with their respective ligands, were collectively subjected to simulation. The reason for this is to increase the number of conformations obtained. Liganded structures were also used, as it was predicted that the structure in the apo state would

show similar folding throughout the simulation. Ligands were removed from the protein structure for subsequent analysis. Accordingly, a total of 10 simulation trajectories, each 1 microsecond long, were obtained with Gromacs 2023.1 [49]. Testing of the ERK5 kinase domain was performed with CHARMM36 [50] under NPT conditions at 1 bar pressure and 310 K temperature. SPC216 was used as the water model. The system was minimized for 10 ns for all simulations. Following this, the systems were subjected to equilibration steps of 10 ns each time in the NVT and NPT ensemble. The systems subjected to MD simulation are listed in **Table 2.3.**

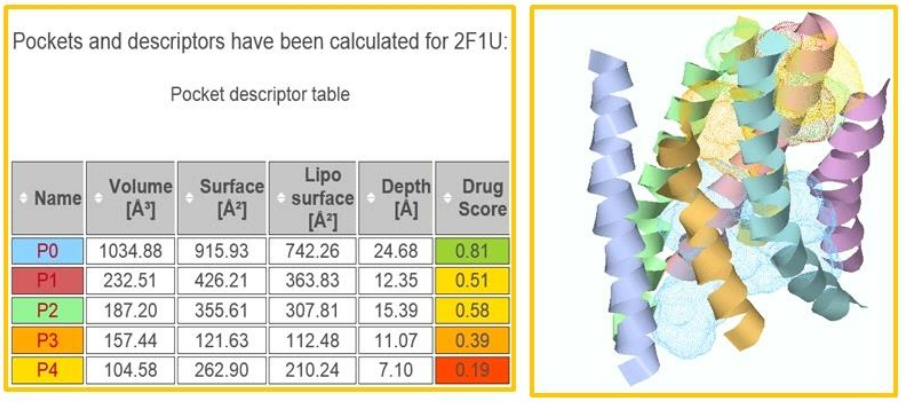
Table 2.3. Representation of the protein structure selected from each cluster, how long it was simulated, and the number of simulations.

PDB ID	Simulation Time	Number of Simulations
4B99	1 microsecond	1
4IC7	1 microsecond	2
4IC8	1 microsecond	3
4ZSG	1 microsecond	1
4ZSJ	1 microsecond	1
6HKN	1 microsecond	2

Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuations (RMSF) profiles were obtained in order to characterize the conformational stability of the ERK5 kinase structure ensembles obtained as a result of the simulations. The RMSD time profile is an indication of the change of protein structure relative to the initial structure from the beginning to the end of the simulation [51]. Accordingly, low level fluctuations seen in RMSD values indicate that the structure vibrates around a certain equilibrium structure, and higher level fluctuations indicate a more mobile structure. RMSF, or Root Mean Square Fluctuation, quantifies the average displacement of a specific atom or group of atoms in relation to a reference structure. It accounts for the collective movement of atoms and provides a measure of their overall fluctuation [52]. Elevated RMSDs (Root Mean Square Deviations) or RMSFs (Root Mean Square Fluctuations) can suggest either overall structural fluctuations or significant displacements within a specific subset of the protein structure. In both calculations, only the coordinates of alpha carbons, which represent each amino acid in the structures, were considered.

2.5. Binding Site Detection via DogSiteScorer

Proteins typically have multiple potential binding sites on their surface. The process of binding site detection involves identifying and locating all the potential binding sites within a protein molecule [53]. Over the past few years, several techniques have been developed to detect binding sites. These methods utilize various properties to analyze the pockets, including volume, hydrophobicity, hydrogen bonding, energy potential, solvent accessibility, desolvation energy, or residue propensity. The choice of descriptor directly impacts the accuracy of the prediction. Therefore, it is crucial to explore the incorporation of different features in representing proteins in order to accurately predict binding sites.



Pockets and descriptors have been calculated for 2F1U:

Pocket descriptor table

Name	Volume [Å³]	Surface [Å²]	Lipo surface [Å²]	Depth [Å]	Drug Score
P0	1034.88	915.93	742.26	24.68	0.81
P1	232.51	426.21	363.83	12.35	0.51
P2	187.20	355.61	307.81	15.39	0.58
P3	157.44	121.63	112.48	11.07	0.39
P4	104.58	262.90	210.24	7.10	0.19

Figure 2.5. DogSiteScorer result from an example study. Predicted pocket volume, surface, lipo surface, depth and drug score values for each pocket. [54]

Following the completion of the MD simulations, potential small molecule binding pockets observed on the structures obtained from the simulation trajectories were determined along with the druggability score (DrugScore) of the pockets using the DogSiteScorer [55] method. DogSiteScorer is a pocket detection tool that enables the identification of potential binding pockets and subpockets in protein structures. The analysis features a support vector machine (SVM) to estimate their druggability. [56]. Accordingly, each of the conformations obtained from the simulation runs was analyzed using DoGSiteScorer 2.0, and only pockets with a DrugScore value above 0.5 were listed. Each pocket was then encoded with an N-length vector, each

representing amino acid positions in the ERK5 sequence. Vectors contain the value 1 if the relevant amino acid is contained in the relevant pocket, and 0 if it is not. These vectors were then stacked on top of each other to obtain an MxN matrix. Here, M is the number of pockets detected during the relevant simulation. In the final stage, the resulting matrix was subjected to a hierarchical clustering analysis and the pockets that were similar to each other in terms of the amino acids they contained were grouped in the same cluster. Finally, the pockets representing the clusters obtained from each simulation were further subjected to a clustering analysis, and pocket clusters representing all the pockets obtained from all simulations were obtained.

2.6 Molecular Docking Simulations

Molecular docking is a method that facilitates the screening and discovery of inhibitors and candidate molecules capable of interacting with specific targets. Through virtual docking experiments, the binding interactions between small molecule compounds or potential drugs and target proteins are simulated. By examining the binding affinities and patterns of interaction, scientists can identify compounds that exhibit strong binding affinity and possess desirable drug-like characteristics [57].

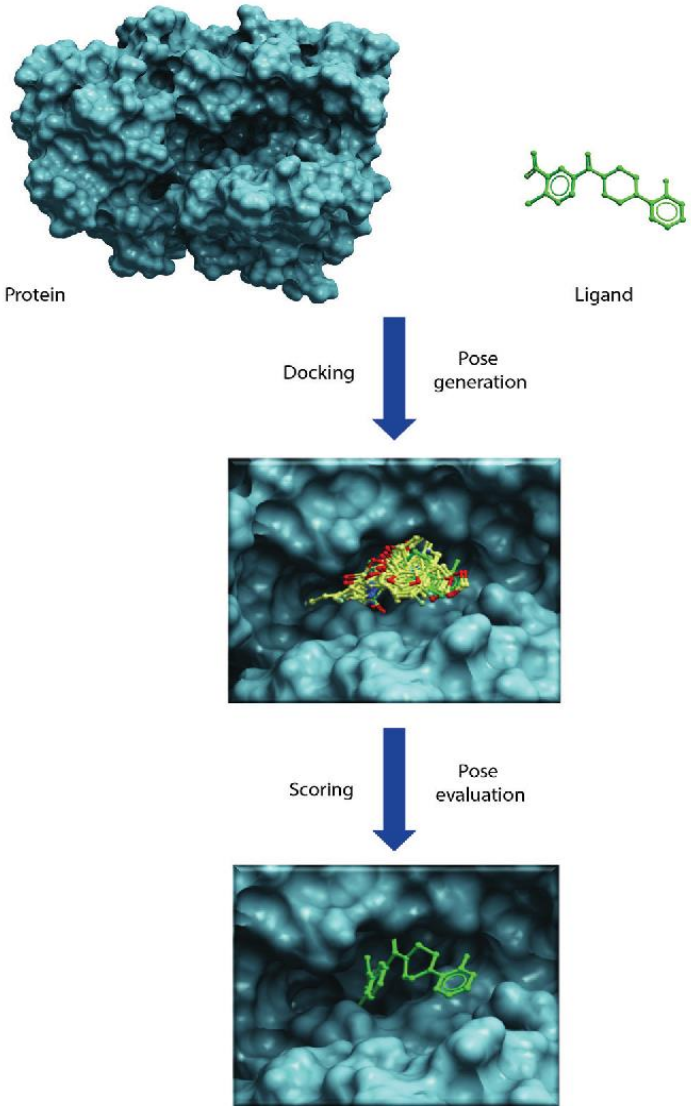


Figure 2.6. Working principle of molecular docking [58].

In order to apply a consensus molecular docking method instead of a single molecular docking method, AutoDock Vina (via SMINA) [59], LeDock [60] and GOLD [61] methods were used together.

Consensus docking and scoring strategies aim to enhance the quality of datasets in virtual screening, irrespective of the specific target, docking programs, or molecular libraries utilized [62]. By employing a consensus approach, which combines results from multiple molecular docking algorithms, the accuracy, confidence, and

comprehensiveness of molecular docking predictions can be improved [62]. Since different docking tools employ distinct search and scoring methods, focusing on the intersection of their predictions can help compensate for the limitations of individual algorithms [63]. By considering the consensus prediction from multiple docking algorithms, the likelihood of encountering false positives or false negatives is reduced, thereby increasing confidence in the precision of the outcomes.

In conclusion, consensus docking and scoring strategies offer an effective approach to improve the accuracy, confidence, and comprehensiveness of molecular docking predictions. By integrating results from multiple docking algorithms, the limitations of individual algorithms can be compensated for, leading to enhanced outcomes in drug discovery, virtual screening, and structure-based drug design applications.

Thus, a total of 2056 FDA-approved drugs in the eDrug3D library [64] were placed in each of the pockets with three different molecular docking programs, and the best scores and poses reported by each program were obtained for each molecule. The binding pockets obtained from DogSiteScorer were used as targets in docking applications. Exhaustiveness level for AutoDock Vina has been set to 16. Default parameters were used for the LeDock and GOLD algorithms. For each targeted pocket, 3 repetitions were made with each method and the average of the results was used. All algorithms used were automated in Python. Python scripts and files used in docking steps and analysis are given in appendix section. Then, the obtained results were subjected to a three-stage filtering process and candidate molecules to be subjected to MD simulations and validation simulations were determined. First, (1) a score was determined for each molecular docking and molecules above this score were selected. In the next step (2), the binding scores were compared with the molecules in the pocket containing the active site, and the molecules with good scores from the active pocket remained, and the molecules previously in the Protein Kinase Inhibitor Database (PKIDB) were removed. In the final step (3), the poses of each of the resulting molecules obtained by the docking programs were visually compared and the ligands with consistent pose were selected.

2.7 Calculating Binding Affinities with MM/PBSA

The Molecular Mechanics-Poisson Boltzmann Surface Area (MM/PBSA) method is widely used to estimate the binding free energy of a complex formed by a ligand and a protein [65]. The MM/PBSA approach is also frequently used in virtual screening studies to identify potential new hits. This method offers more reliable assessments of ligand-binding affinities compared to the simple scoring functions used in molecular docking algorithms [66].

In MM/PBSA method, the difference between the free energy of the complex and the individual free energies of the unbound receptor and ligand are computed. Since the MD simulations from which the conformations used in this calculation is obtained a constant temperature and pressure, this difference is the Gibbs Free Energy of binding of the ligand to a protein macromolecule. The thermodynamics-based definition of Gibbs free energy is:

$$\Delta G = \Delta H - T\Delta S$$

where ΔG is the change in Gibbs free energy, ΔH is the change in enthalpy, T is temperature, and ΔS is the change in entropy.

For the purpose of computing the free energy of ligand binding, MM/PBSA method decomposes the enthalpic component into gas-phase molecular mechanics interaction energy between the ligand and the protein (ΔE_{MM}), and the solvation free energy (ΔG_{solv}):

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S$$

Here, $T\Delta S$ represents the change in conformational entropy associated with ligand binding.

ΔE_{MM} consists of change in internal energy (ΔE_{int} corresponding to bond, angle, and dihedral energies), electrostatic energy (ΔE_{ele}), and the van der Waals energies (ΔE_{vdW}):

$$\Delta E_{MM} = \Delta E_{int} + \Delta E_{ele} + \Delta E_{vdW}$$

ΔG_{solv} is the sum of polar contribution to solvation (electrostatic solvation energy ΔG_{PB}) and a nonpolar contribution (ΔG_{SA}) estimated through the solvent-accessible surface area between the solute and solvent:

$$\Delta G_{solv} = \Delta G_{PB} + \Delta G_{SA}$$

Here, the ΔG_{PB} is computed using the Poisson-Boltzmann equation.

Binding free energy calculations serve various purposes. When combined with reliable docking studies and sufficiently long molecular dynamics (MD) simulations, MM-PBSA and similar approaches are useful in comprehensive pose prediction studies. These studies aim to determine the most accurate arrangement of an active compound within its target's binding site, especially in cases where the bioactive conformation of the compound is unknown. MD trajectories representing different conformations of the ligand and protein are used to calculate the binding free energies of the complexes. The final binding energy value is obtained by averaging the energies estimated for these different snapshots [67]. However, due to the significant computational time required, the evaluation of binding free energy is typically limited to only those molecules that meet certain selection criteria. As a result, MM/PBSA is often considered a complementary strategy employed after initial filtering calculations [68].

MM/PBSA binding free energy calculations were performed using the gmx_MMPBSA tool [69] for each frame obtained from MD simulations of ligand-ERK5 kinase complexes. The MM/PBSA calculation was applied to all 100 nanosecond trajectories.

3. RESULTS

3.1. CABS-Flex Simulations

As a result of the CABS-Flex web server used to obtain different conformations, 110 models for 11 protein structures were obtained. These models were subjected to the TTClust algorithm and divided into classes according to the similarity of the structures. After TTClust analysis, the structures obtained from the CABS-Flex program were divided into five clusters. CABS-Flex simulations show that even when different PDB structures are used as a starting point, the visited conformations cannot be included in the conformation sample around another PDB. On the other hand, it is understood that the apo 4IC8 structure differs conformationally from other ligand-containing structures. Structure 4B99 is also in a cluster by itself. 6HKM and 6HKN structures, 5B77 and 4IC7, and 5O7I, 5BYZ, 4ZSJ, 4ZSL, 4ZSG structures are structurally similar to each other.

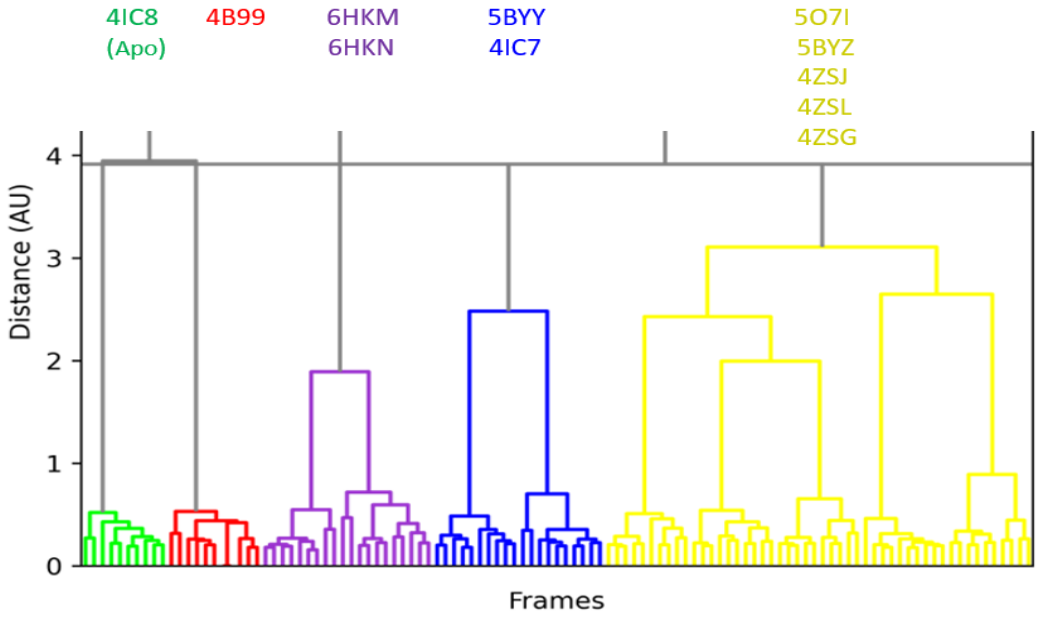


Figure 3.1. TTClust analysis results. The 11 ERK5 kinase domain structures were divided into 5 clusters in total according to the similarities of the conformations in which they can be found.

3.2. Molecular Dynamics Simulations

After completing a 1 microsecond long simulation for each of the 10 structures selected from the clusters, RMSD and RMSF analyzes were performed for each simulation. High RMSD values may indicate fluctuation in the entire structure or may reflect large displacements of only a small subset within the protein structure. As a result of the RMSF calculation, the movement of each amino acid is shown. To calculate these two values, the alpha carbons of the amino acids in the structures were considered. Similar RMSF values were obtained as a result of the simulations of the systems shown in **Figure 3.2**.

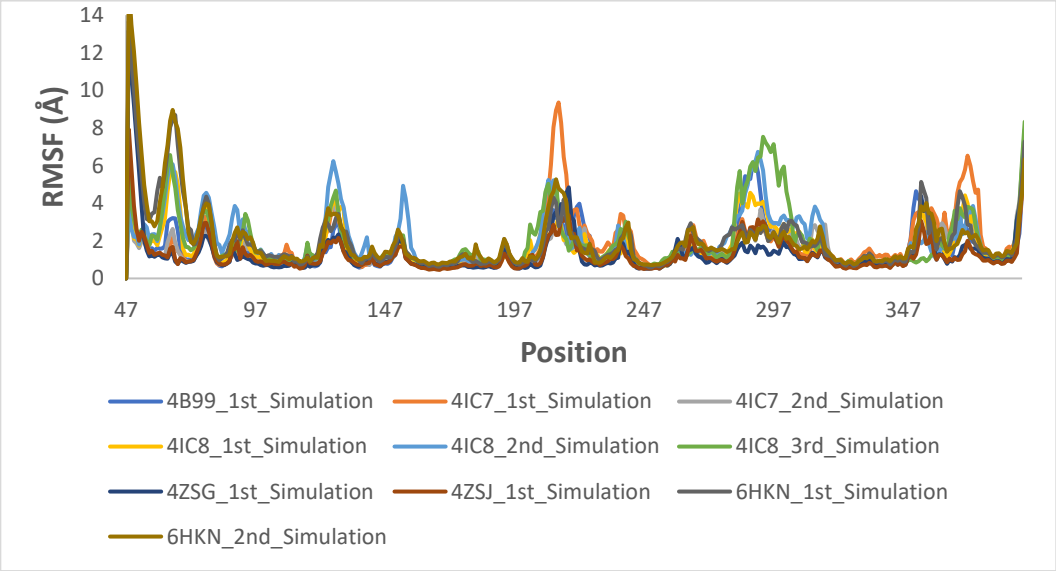


Figure 3.2. Calculated RMSF data for 10 simulations.

RMSF analysis was performed to analyze the individual movement of residues, and it was determined that the fluctuating amino acids were in loops and not in important regions affecting protein function. The areas where differences are observed are the loop regions that do not have a secondary structure in the protein structure, and the natural properties of these structures are also prone to fluctuation. Loop regions are already highly fluctuating regions in the protein structure, and the results were as expected.

To determine the stability of the structures, RMSD calculations were made to show how they changed over time. RMSD results shown in **Figure 3.3** also show that the structures reached equilibrium in all simulations. No unusual fluctuations were observed in the functional parts of the protein and the structures reached equilibrium state.

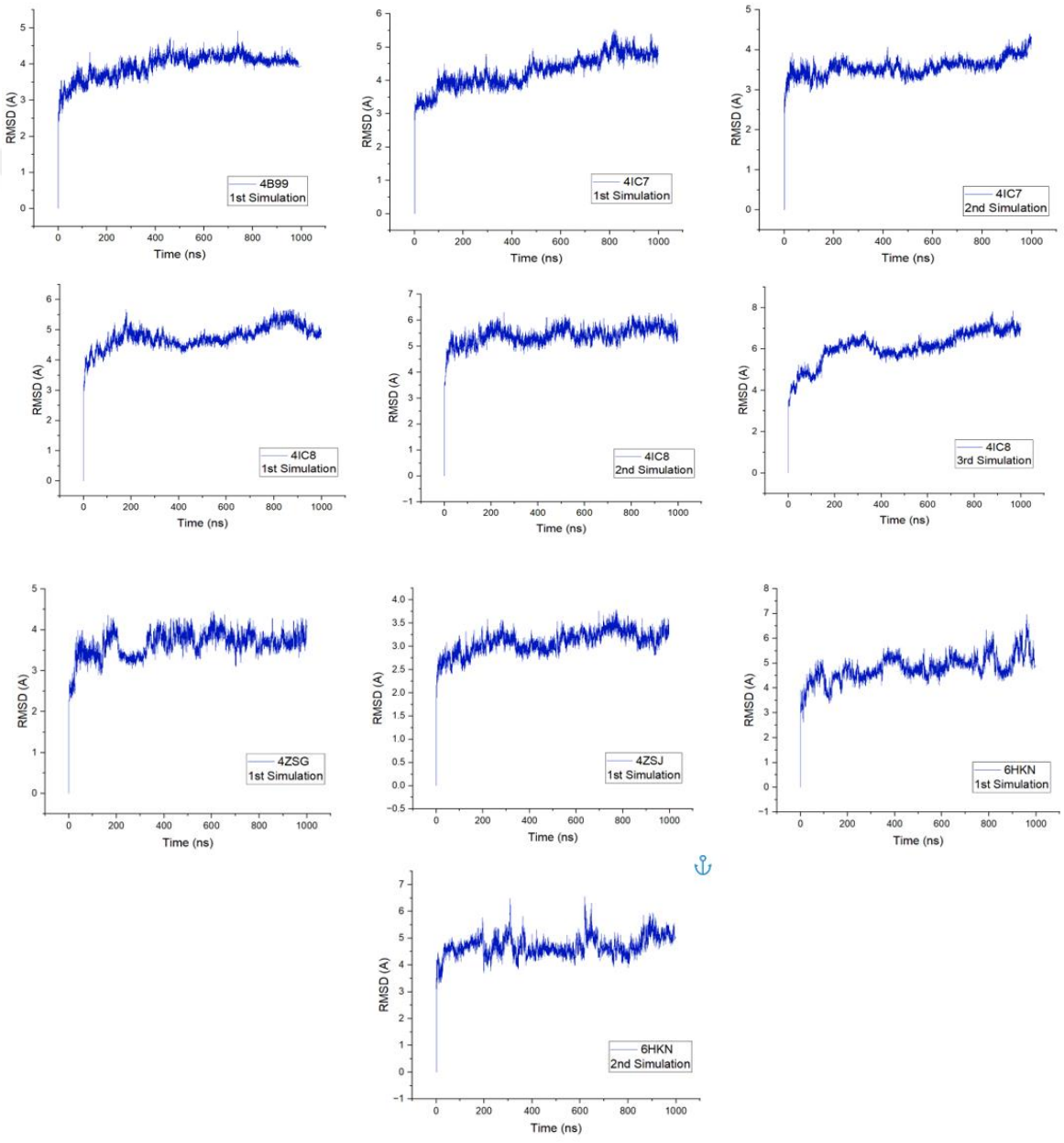


Figure 3.3. RMSD results of each simuation.

3.3. Binding Site Detection

As a result of pocket detection using DogSiteScorer, pockets suitable for ligand binding on the ERK5 kinase domain were determined and the resulting subpockets were filtered to have a drug score value greater than 0.5. The resulting subpockets and the amino acids in these pockets were subjected to hierarchical clustering, and the pockets with the highest DrugScore value in each cluster were determined. At the end of this analysis, in addition to a pocket containing the active site, three more pockets outside the active site that may have allosteric ligand binding potential were identified. Additionally, the list of residues forming each pocket is given in **Table 3.1.**

Table 3.1. The list of residues forming each pocket.

Pockets	**Amino acids**
1 (active region)	I61 G62 N63 G64 A65 Y66 G67 V69 A82 K84 I86 P90 T99 E102 L103 L106 I115 I117 I120 V133 Y134 V135 L137 D138 L139 M140 E141 S142 D143 Q146 L189 G199 D200 P201 I381 I385 P388
2	I86 P87 N88 A89 T94 N95 K97 R98 T99 R101 E102 I105 Q177 V178 I179 H180 R181 L184 N187 D200 P201 G202 M203 A204 R205 G206 L207 C208 T209 P217 M218
3	D53 Y66 K84 I86 P87 F90 N95 R98 T99 E102 L103 L106 I115 I117 I120 L121 R122 P123 T124 V125 F130 S132 V133 Y134 V135 L137 I381 V382 I385
4	R225 W226 Y227 R228 A229 P230 M233 W246 C250 P262 P263 Q270 L271 L273 I274 M275 M276 V277 L278 G279 T280 P281 S282 V285 W309 V312 Y313 L328 R329 P330

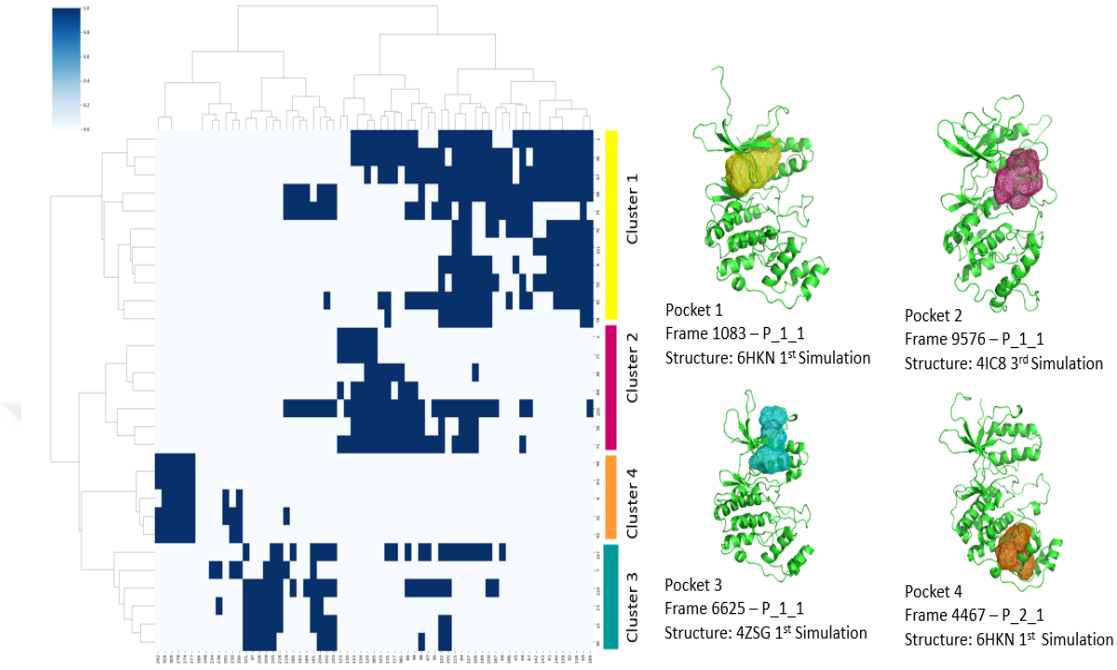


Figure 3.4. Clustermap representation (right). Each row on the clustermap represents a pocket, and each column represents an amino acid. Amino acids found in the pocket are shown in dark blue, and amino acids not found in the pocket are shown in light blue. Representation of pockets on the protein structure (left). The matching pocket and cluster are presented in the same color. The location of each pocket on the protein structure is shown. Pocket 1 is located in the active site, pockets 2, 3 and 4 are in the so-called allosteric regions.

A clustermap was obtained by hierarchical clustering according to the amino acids they contained. The pockets within the clusters seen on **Figure 3.4** are similar to each other. When the highest scoring pockets were selected, 1 active pocket and 3 allosteric binding pockets were obtained. 2056 drugs in the eDrug3D library were tested in these pockets using the molecular docking method.

3.4. Comparative Evaluation of Molecular Docking Software

3.4.1 Re-docking Performances of Autodock Vina, GOLD, and LeDock for Allosteric Inhibitors of MAPK Family

To assess the reliability of the molecular docking algorithms employed in this study and to evaluate their effectiveness in predicting the binding of allosteric inhibitors to kinase structures, a redocking analysis was conducted. A curated list of MAPK structures containing known allosteric inhibitors was compiled, and their three-dimensional structures were retrieved from the Protein Data Bank. **Table A1** in **Appendix** provides detailed information about the ligands present in these structures, as well as the structural characteristics and the kinase family to which they belong. Redocking experiments were performed using the AutoDock Vina, LeDock, and GOLD molecular docking software. The docking scores obtained from these algorithms for each structure are presented in **Table A2**. These scores serve as indicators of the binding affinity between the inhibitors and the respective kinase structures.

Furthermore, RMSD (Root Mean Square Deviation) calculations were conducted to evaluate the similarity between the predicted poses obtained from redocking and the native ligands already present in the structures. For this purpose, the DockRMSD method [68] ,which considers potential symmetry in the docked molecule for correct calculation of RMSD. The RMSD values, which quantify the deviation between the predicted and native ligand conformations, are provided in **Table A3**. This analysis helps assess the consistency and accuracy of the redocking predictions. **Figure A1** illustrates the representations of the estimated poses resulting from the redocking experiments. These visualizations offer insights into the spatial arrangement and interaction patterns between the predicted poses and the binding sites of the kinase structures.

Table A2 shows that the re-docking scores obtained from Autodock Vina, LeDock and GOLD were ranged from -6 to -12.7 kcal/mol, -3.63 to -14.29 kcal/mol, and 52.58 to 116.49, respectively. **Table A3** also shows that at least 10 out of 12 allosteric

MAPK7 ligands were successfully (RMSD < 2 Å) docked into their respective binding pockets by each tool used. **Figure A1** further demonstrates that the predicted binding modes were mostly consistent across the docking software as well. These results indicate that the chosen molecular docking software may be reliably used to predicted modes of interactions for discovery of novel allosteric inhibitors of kinases, including ERK5.

3.4.2 Consistency Between Rankings Obtained from Autodock Vina, GOLD, and LeDock

To assess the consistency of the algorithms used in this study, the distributions of the scores obtained from Autodock Vina, LeDock, and GOLD were compared. The scores obtained from each algorithm were plotted to visualize their respective distributions.

The distribution of scores obtained from Autodock Vina was examined, followed by the distribution of scores obtained from LeDock and GOLD, respectively. By analyzing these distributions, it was possible to gain insights into the range and spread of the scores generated by each algorithm.

Additionally, a comparative analysis was conducted to evaluate the relationships between the scores obtained from different algorithms. Firstly, the scores obtained from Autodock Vina were compared with those obtained from GOLD, enabling a direct comparison between these two algorithms. Next, the scores obtained from Autodock Vina were compared with those obtained from LeDock. Lastly, the scores obtained from LeDock were compared with those obtained from GOLD. The results are shown in **Figure 12.**

By comparing the scores obtained from different algorithms, it was observed that there exists a correlation between these scores. This correlation implies that the algorithms produce consistent results in terms of ranking the ligands based on their predicted binding affinity to the kinase structure.

This comparative analysis provides valuable information regarding the agreement and consistency between the scores generated by different molecular docking algorithms.

The observed correlation between the scores strengthens the reliability and robustness of the predictions made by these algorithms in the context of this study.

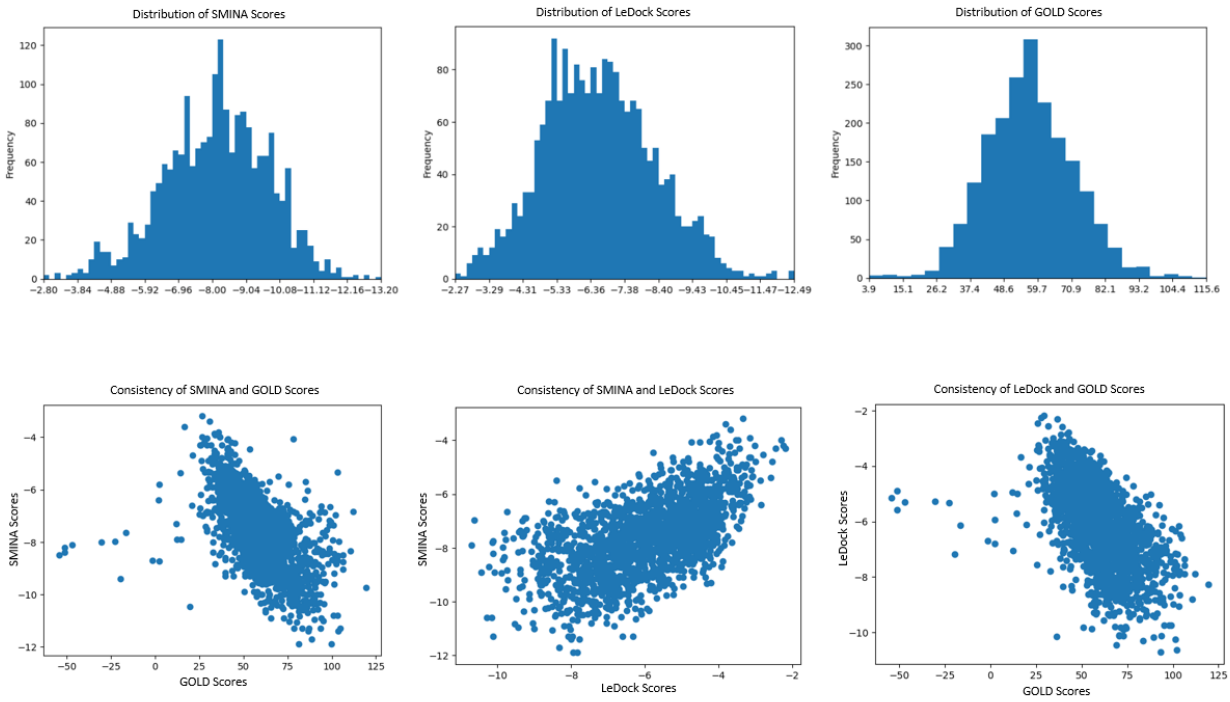


Figure 3.5. Distribution of Smina, Ledock and GOLD scores (top row). Scatter plots showing correlations between obtained scores (bottom row).

3.5. Molecular Docking Simulations and MM/PBSA calculations

Instead of using a single method, Autodock Vina, LeDock and GOLD methods were used together as a consensus docking approach. The molecular docking simulations for each pocket was repeated 3 times and average values were used. Different filtering methods were used when evaluating the results. Accordingly, first, a minimum score was determined for each molecular docking simulation (-9 kcal/mol for SMINA/Autodock Vina, -7 kcal/mol for LeDock, 65 for GOLD) and only molecules posing above these scores were listed. In the next step, molecules with binding scores higher than those obtained in Pocket 1 containing the active site and not previously identified as kinase inhibitors in the Protein Kinase Inhibitor Database (PKIDB) were identified. In the final step, the best binding poses reported by each program for each

of these molecules were visually compared. A total of 32 allosteric inhibitors were obtained, passing through these three filters. In the next stage, they were subjected to short MD simulations of 100 ns each in order to verify the binding poses determined as a result of molecular docking simulations and to determine the binding free energies of the confirmed poses using the MM/PBSA method. These simulations were carried out with the same procedure as the simulations carried out within the scope of MD simulations, and CHARMM General Force-Field (CGENFF) was used for the parameterization of the ligands. As a result of running MD simulations, MM/PBSA values could not be obtained for droperidol due to a technical error caused by gmx_MMPBSA. In the MD simulation trajectories of the vilazodone and glycerol phenylbutyrate-bound ERK5 kinase domain, the ligand binding stability could not be evaluated accurately, since the correction of the periodic boundary conditions (during the unwrapping process) could not be successfully achieved. Simulations of the ERK5 kinase domain bound to bictegravir, sulfoxone, and cariprazine could not be carried out successfully due to errors due to parameterization. In the simulations of capreomycin, neomycin, and gentamicin-bound ERK5 kinase domain, it was observed that all of the bonds of the ligand with the protein were broken within 100 ns and left the binding pocket, therefore the predicted binding pose could not be confirmed. Other inhibitor candidates were observed to remain in the binding pocket, although significant changes were observed in their conformations throughout the simulations**.**

A total of 23 (8 ligands for pocket 2, 15 ligands for pocket 4) FDA-approved drug molecules, whose binding was reported consistently by at least two molecular docking programs and which had not previously been reported as kinase inhibitors, were identified as allosteric ERK5 inhibitor candidates. Docking scores and MM/PBSA binding free energies of these molecules are given in **Table 3.2.**

Table 3.2. Molecular docking and MM/PBSA results. Those shown in red represent drugs that we consider to be potential allosteric ERK5 kinase inhibitors.

Pocket 4	Molecular Docking Scores			MM/PBSA results	
Ligand	**Vina (kcal/mol)**	**LeDock (kcal/mol)**	**GOLD**	**ΔΔG (kcal/mol)**	**SD**
Dolutegravir	-11.7	-8.32	88.5	-28.57	4.43
Ziprasidone	-11.03	-8.03	76.02	-20.22	4.20
Panobinostat	-10.3	-7.8	86.01	-29.27	4.05
Nebivolol	-10.56	-7.05	92.19	-29.29	3.33
Folic Acid	-9.2	-9.26	75.46	-21.8	9.10
Cabotegravir	-10.5	-7.69	81.45	-27.75	3.82
Fosinoprilat	-9.5	-7.05	96.17	-28.4	6.71
M-Travoprost	-10.2	-7.07	89.03	-35.82	8.12
Rilpivirine	-9.9	-7.88	80.03	-23.17	10.39
Tafluprost	-9.43	-7.7	89.17	-37.15	6.41
M-Pimavanserin	-9.3	-7.68	75.98	-30.44	3.28
Sulfinpyrazone	-9.36	-7.4	76.15	-29.75	7.01
Acetohexamide	-9.4	-7.13	70.08	-27.95	7.53
Tirofiban	-8.53	-7.56	75.93	-31.38	7.46
M-Acetohexamide	-9.46	-7.09	68.97	-29.74	4.58
Vilazodone	-11	-9	95.2	-	-
Glycerol Phenylbutyrate	-9.73	-8.28	119.26	-	-
Droperidol	-9.93	-7.76	83.77	-	-
M-Cariprazine	-9.56	-7.58	70.71	-	-
Bictegravir	-11.13	-8.07	83.32	-	-
Sulfoxone	-8.93	-9.9	73	-	-
Pocket 2					
Odevixibat	-9.36	-11.71	87.64	-20.16	7.48
Idarubicin	-11.9	-8.48	73.67	-17.49	6.33
Avatrombopag	-10.4	-9.74	78.73	-15.55	11.46
Dexamethasone Phosphate	-11.5	-8.15	70.56	-30.08	4.88
Topotecan	-11.23	-7.95	74.92	-18.25	7.07
Pralatrexate	-9.53	-9.58	79.78	-19.17	7.82
Rosuvastatin	-9.43	-8.99	80.16	-25.97	6.70
Pitavastatin	-9.53	-7.98	79.04	-29.38	7.35
Capreomycin	-10.26	-12.39	82.54	-	-
Neomycin	-10.93	-11.16	74.8	-	-
Gentamicin	-9.8	-8.71	67.77	-	-

It was observed that tafluprost, tirofiban, panobinostat, nebivolol, dexamethasone phosphate and pitavastatin had relatively better MM/PBSA binding free energies, and the poses obtained from different docking software were consistent with each other. In making this selection, in addition to the MM/PBSA score, the RMSD values of each ligand, their placement within the pocket, and the consistency of the predicted poses were also taken into account. For this reason, some drugs with high MM/PBSA scores were eliminated. Therefore, these drugs may act as allosteric ERK5 kinase inhibitors.

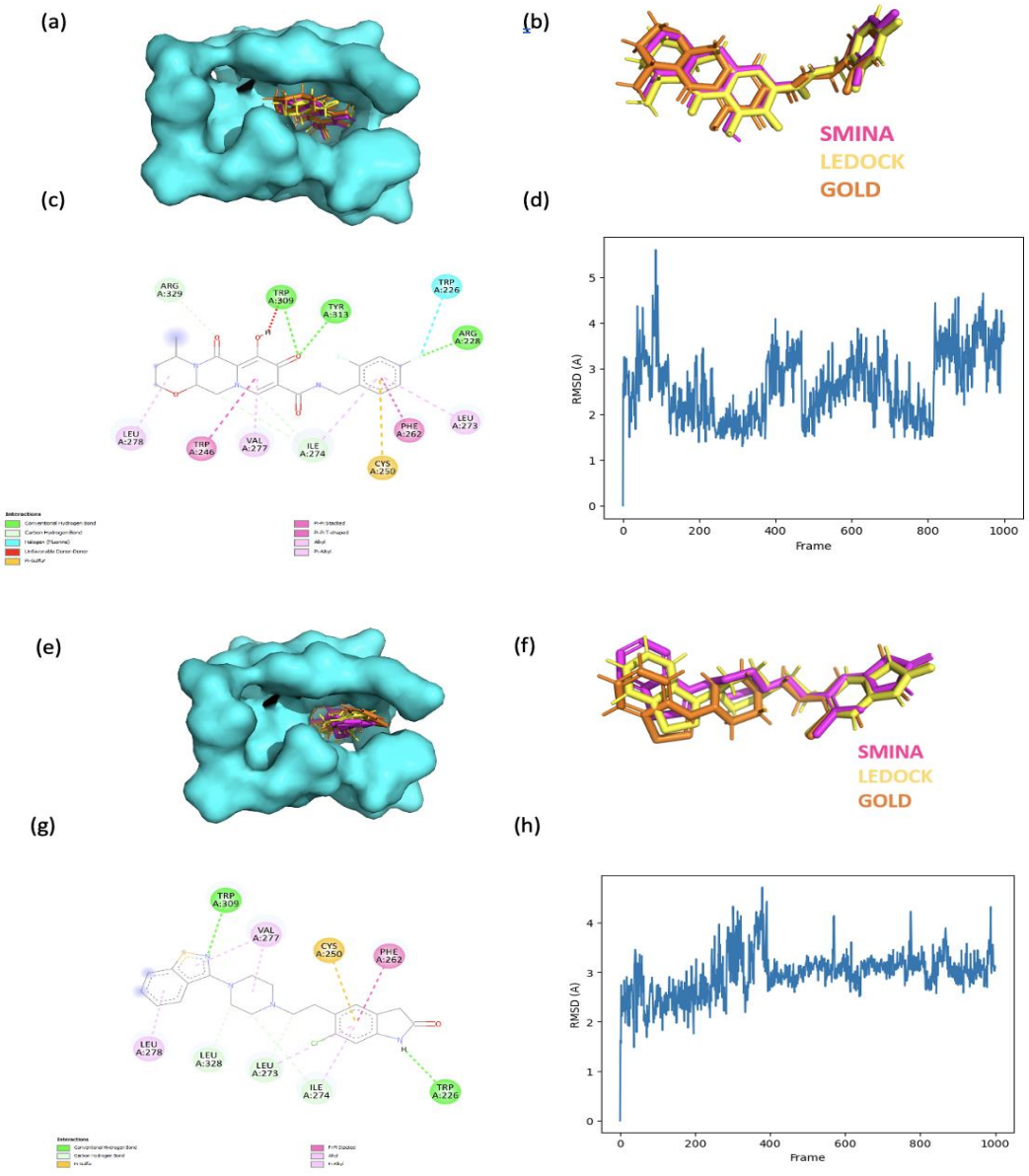


Figure 3.6. Predicted poses of Dolutegravir in pocket 4 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Dolutegravir in pocket 4 amino acids (c), calculated RMSD plot of Dolutegravir over 100 ns simulation (d), predicted poses of Ziprasidone in pocket 4 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of Ziprasidone in pocket 4 amino acids (g), calculated RMSD plot of Ziprasidone over 100 ns simulation (h).

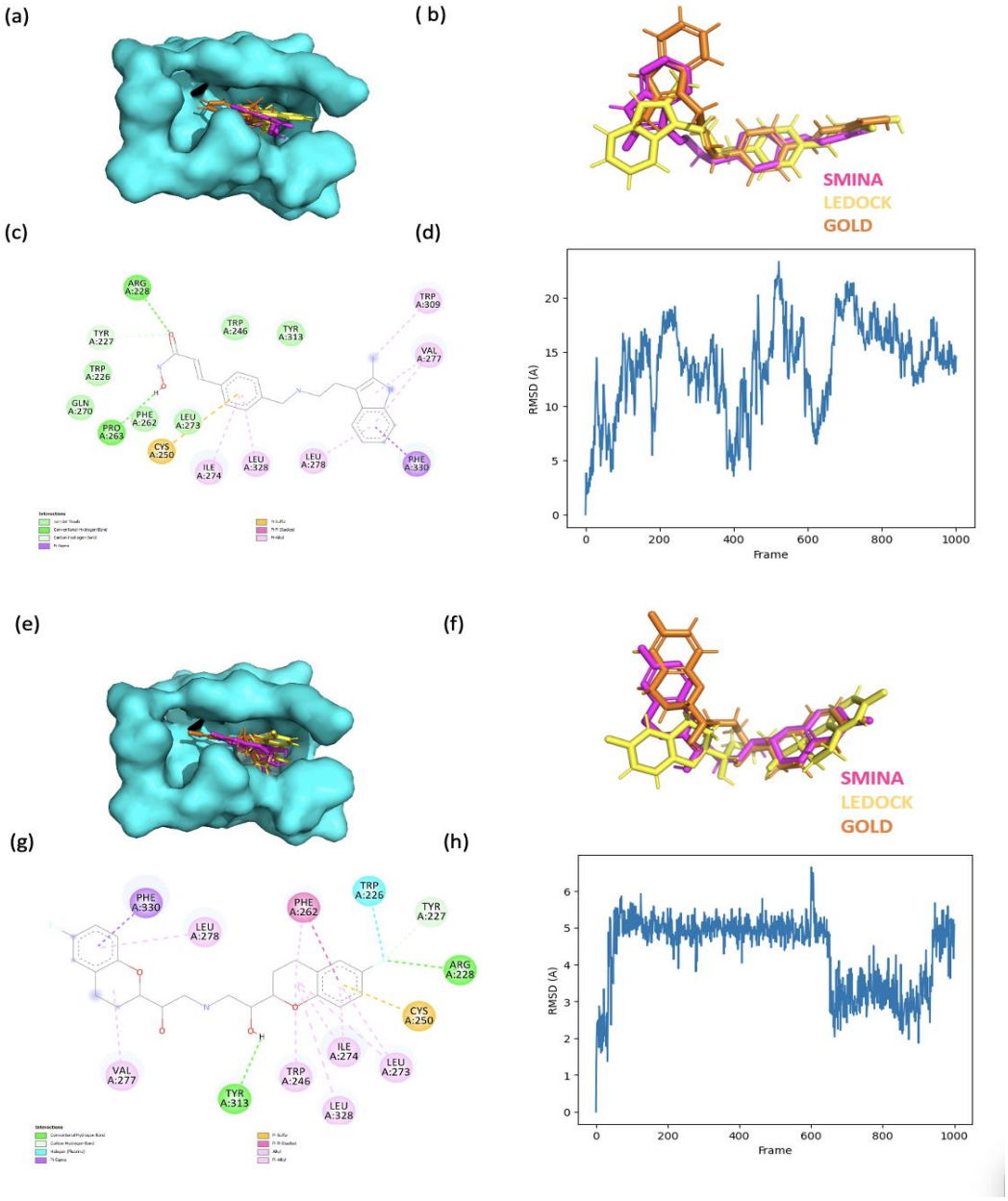


Figure 3.7. Predicted poses of Panobinostat in pocket 4 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Panobinostat in pocket 4 amino acids (c), calculated RMSD plot of Panobinostat over 100 ns simulation (d), predicted poses of Nebivolol in pocket 4 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of Nebivolol in pocket 4 amino acids (g), calculated RMSD plot of Nebivolol over 100 ns simulation (h).

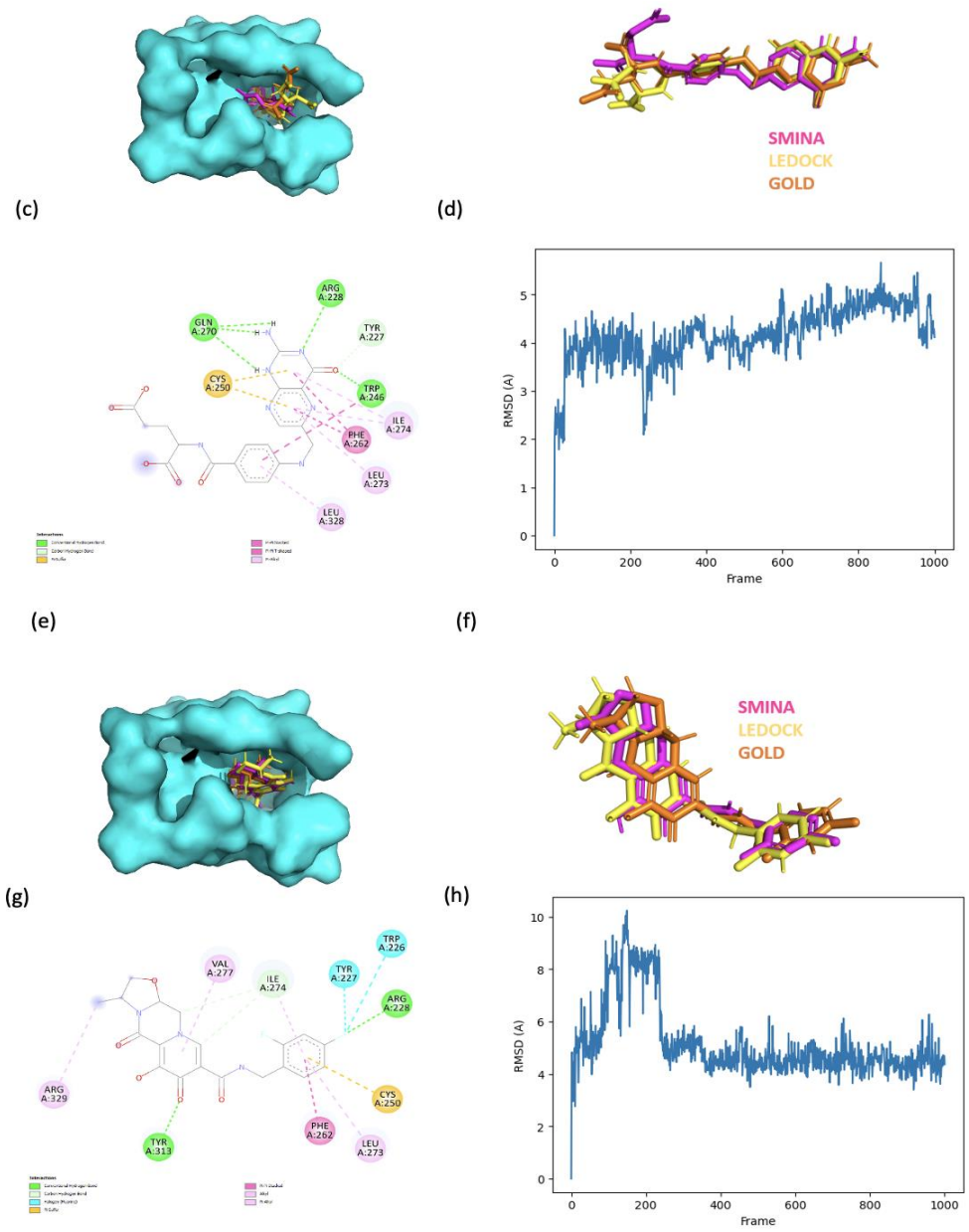


Figure 3.8. Predicted poses of Folic Acid in pocket 4 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Folic Acid in pocket 4 amino acids (c), calculated RMSD plot of Folic Acid stat over 100 ns simulation (d), predicted poses of Cabotegravir in pocket 4 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of Cabotegravir in pocket 4 amino acids (g), calculated RMSD plot of Cabotegravir over 100 ns simulation (h).

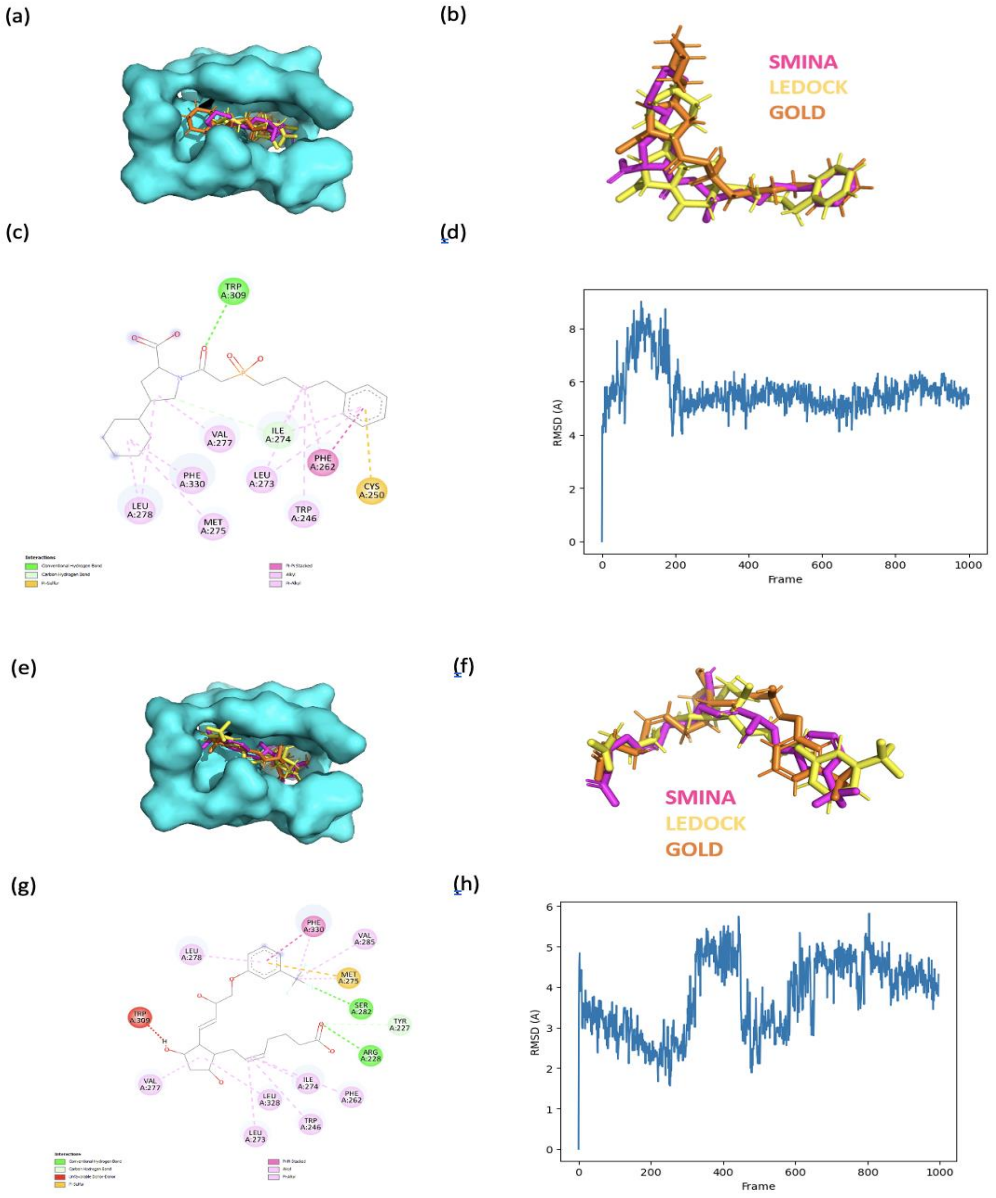


Figure 3.9. Predicted poses of Fosinoprilat in pocket 4 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Fosinoprilat in pocket 4 amino acids (c), calculated RMSD plot of Fosinoprilat stat over 100 ns simulation (d), predicted poses of M-travoprost in pocket 4 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of M-travoprost in pocket 4 amino acids (g), calculated RMSD plot of M-travoprost over 100 ns simulation (h).

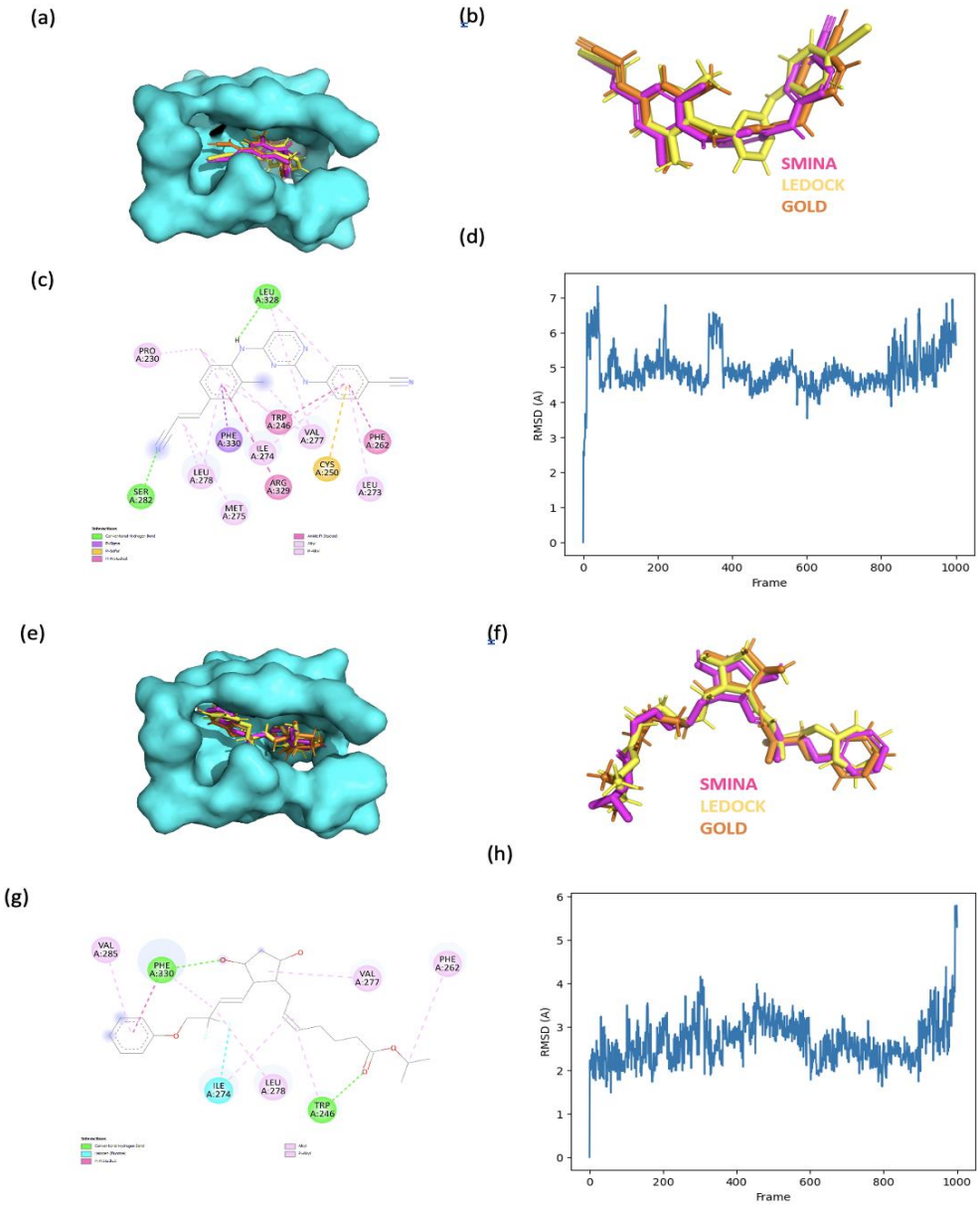


Figure 3.10. Predicted poses of Rilpivirine in pocket 4 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Rilpivirine in pocket 4 amino acids (c), calculated RMSD plot of Rilpivirine stat over 100 ns simulation (d), predicted poses of Tafluprost in pocket 4 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of Tafluprost in pocket 4 amino acids (g), calculated RMSD plot of Tafluprost over 100 ns simulation (h).

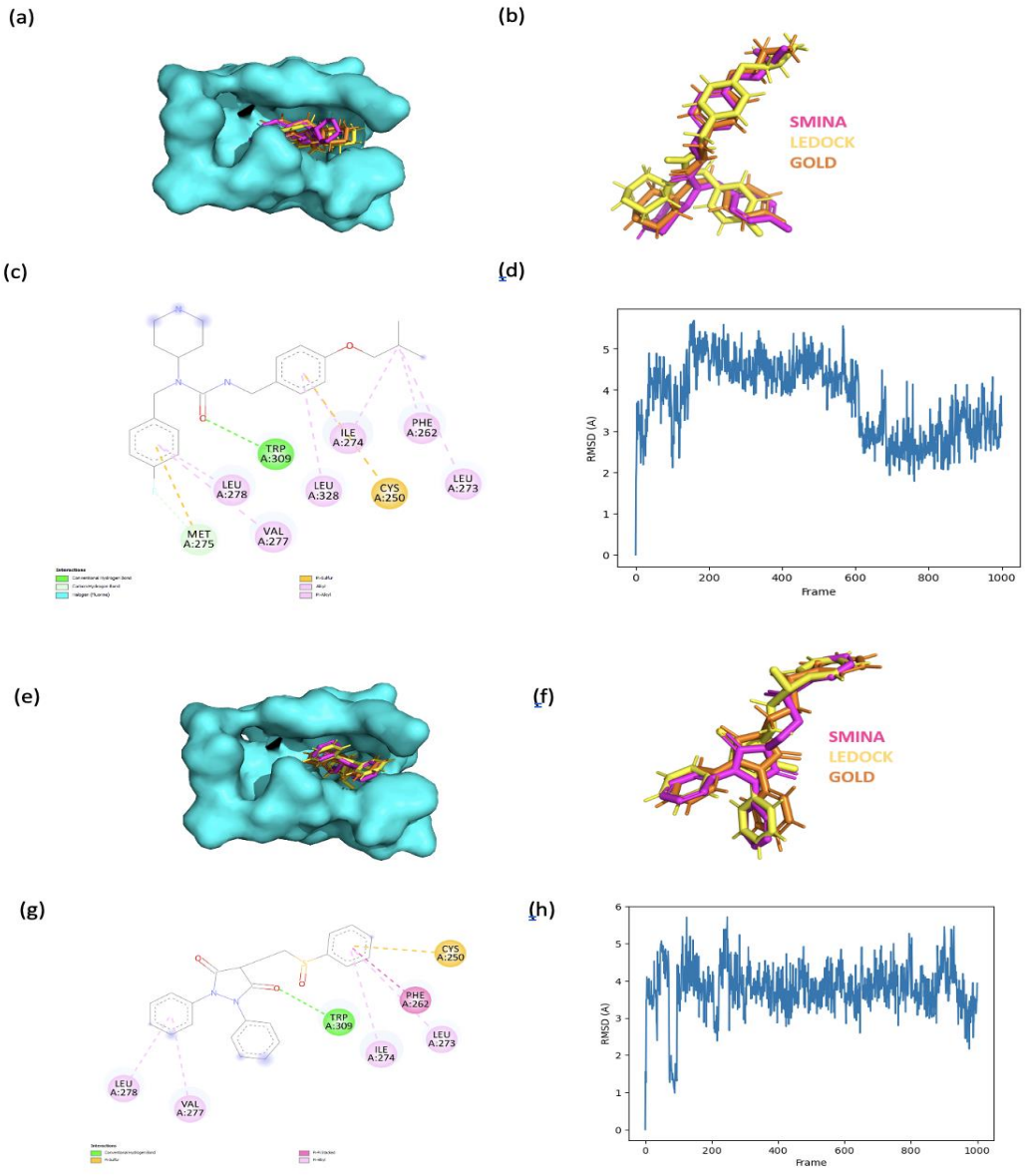


Figure 3.11. Predicted poses of M-pimavanserin in pocket 4 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of M-pimavanserin in pocket 4 amino acids (c), calculated RMSD plot of M-pimavanserin stat over 100 ns simulation (d), predicted poses of Sulfinpyrozone in pocket 4 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of Sulfinpyrozone in pocket 4 amino acids (g), calculated RMSD plot of Sulfinpyrozone over 100 ns simulation (h).

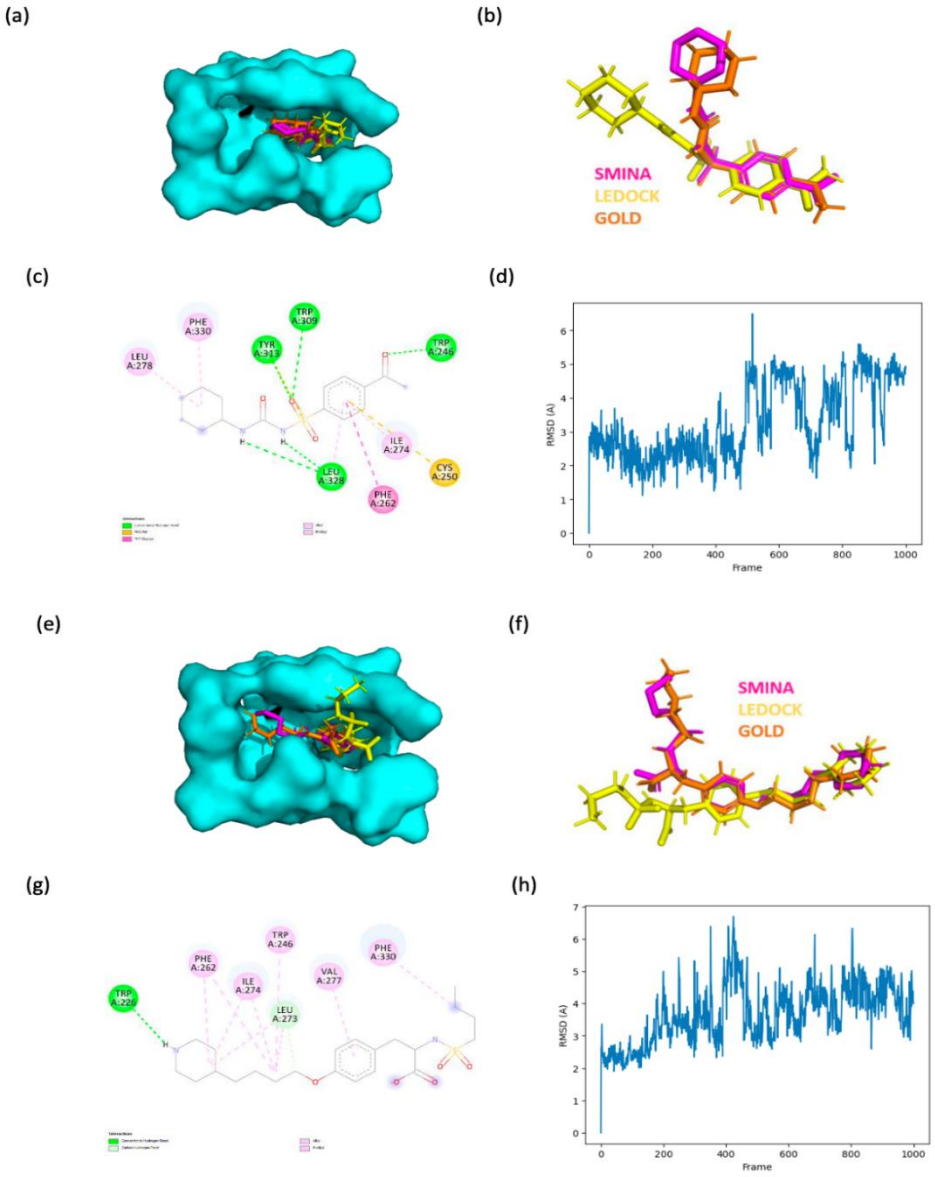


Figure 3.12. Predicted poses of Acetohexamide in pocket 4 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Acetohexamide in pocket 4 amino acids (c), calculated RMSD plot of Acetohexamide stat over 100 ns simulation (d), predicted poses of Tirofibanin pocket 4 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of Tirofibanin pocket 4 amino acids (g), calculated RMSD plot of Tirofiban over 100 ns simulation (h).

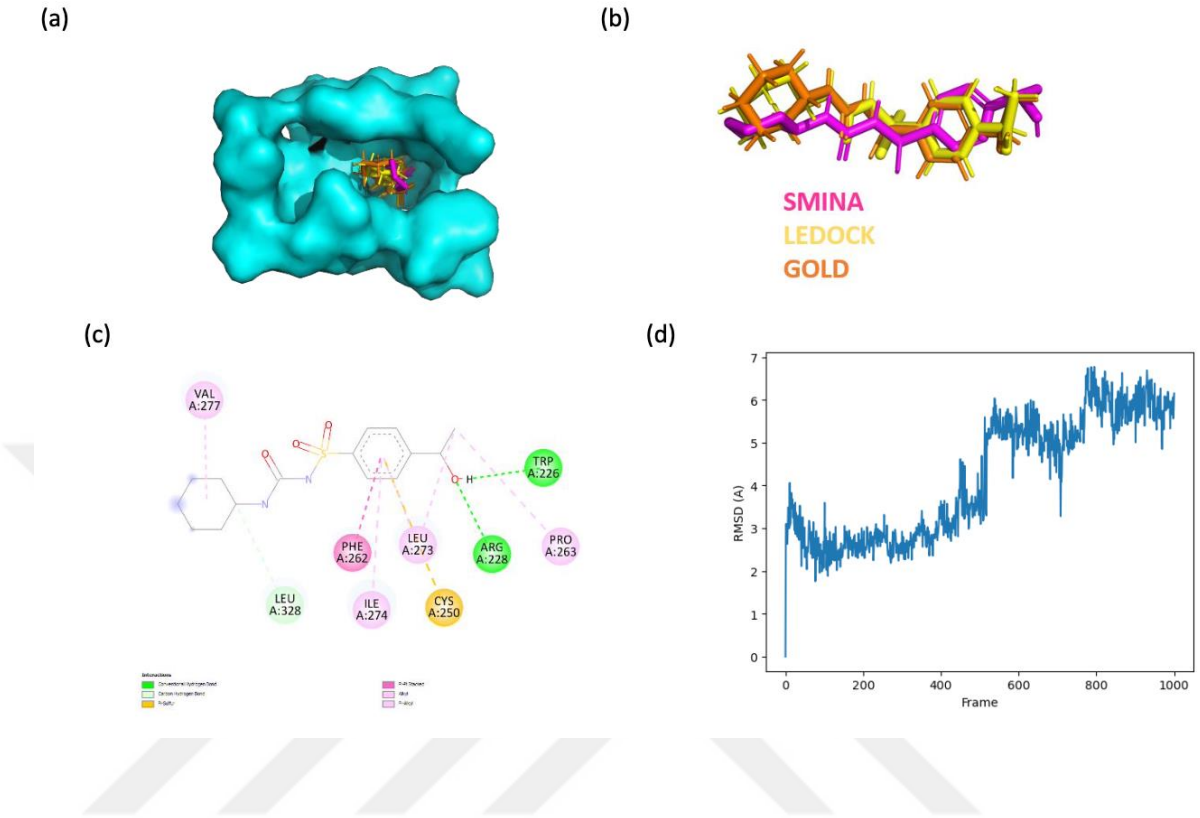


Figure 3.13. Predicted poses of M-acetohexamide in pocket 4 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of M-acetohexamide in pocket 4 amino acids (c), calculated RMSD plot of M-acetohexamide over 100 ns simulation (d).

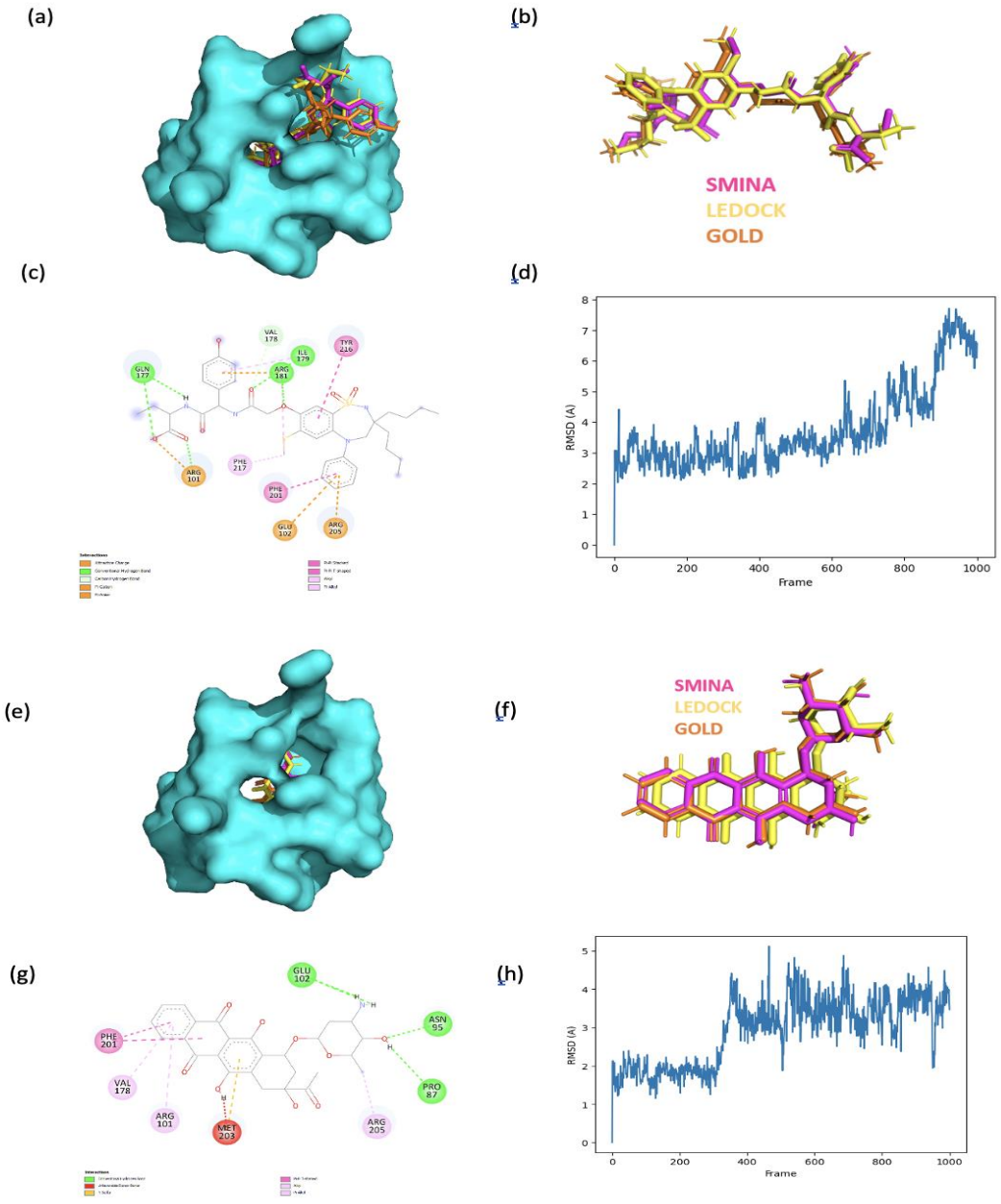


Figure 3.14. Predicted poses of Odevixibat in pocket 2 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Odevixibat in pocket 2 amino acids (c), calculated RMSD plot of Odevixibat stat over 100 ns simulation (d), predicted poses of Idarubicin pocket 2 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of Idarubicin pocket 2 amino acids (g), calculated RMSD plot of Idarubicin over 100 ns simulation (h).

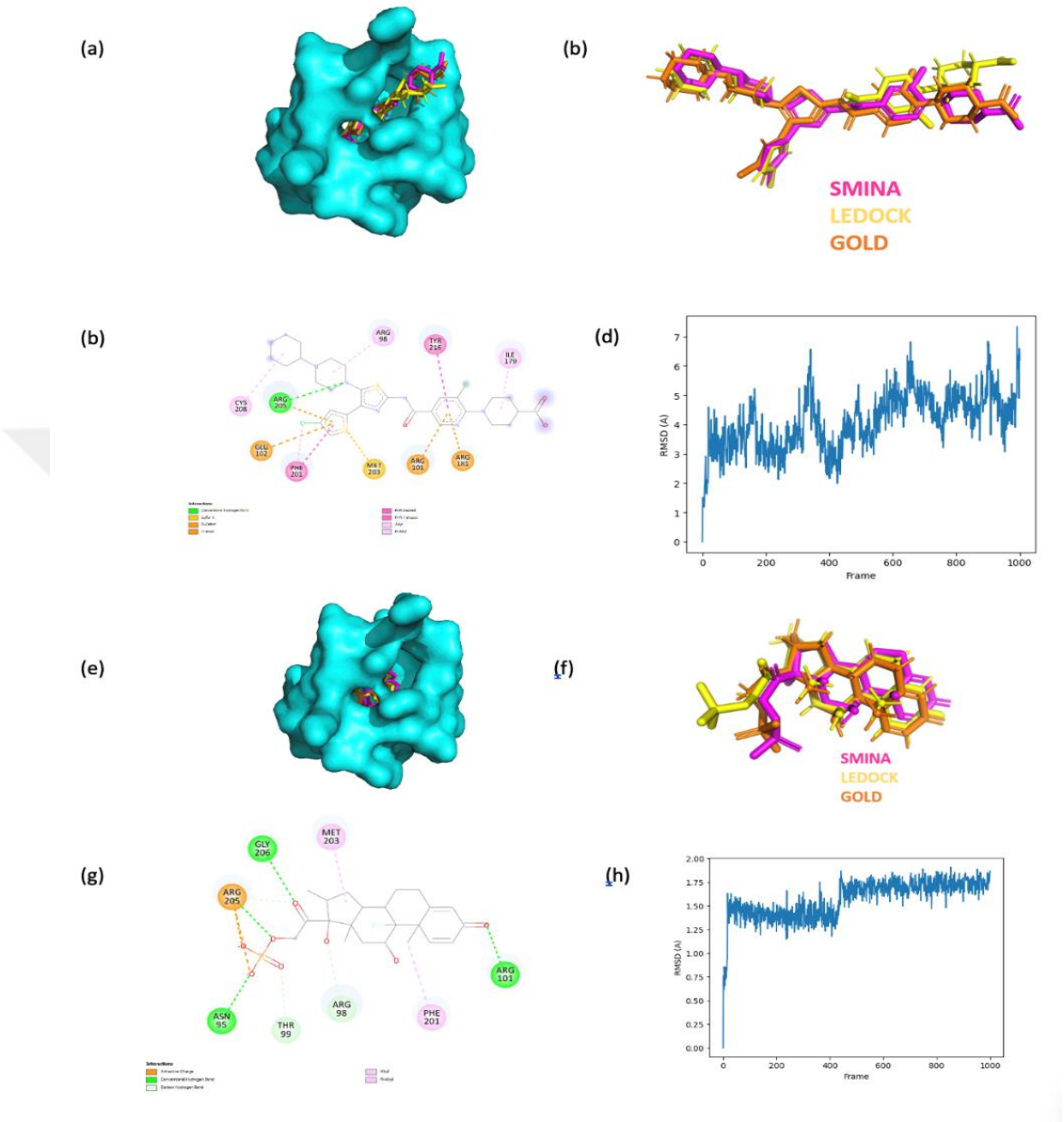


Figure 3.15. Predicted poses of Avatrombopag in pocket 2 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Avatrombopag in pocket 2 amino acids (c), calculated RMSD plot of Avatrombopag stat over 100 ns simulation (d), predicted poses of Dexamethasone phosphate pocket 2 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of Dexamethasone phosphate pocket 2 amino acids (g), calculated RMSD plot of Dexamethasone phosphate over 100 ns simulation (h).

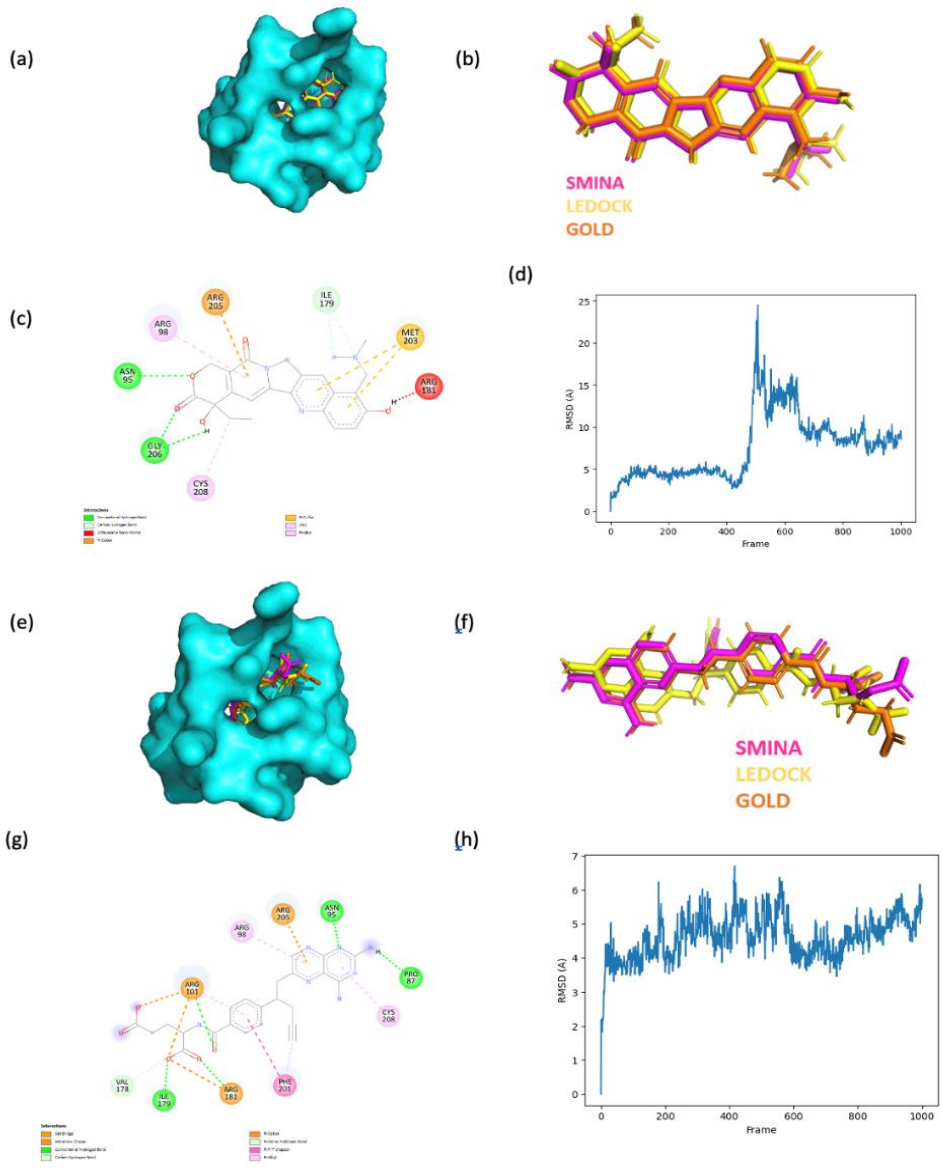


Figure 3.16. Predicted poses of Topotecan in pocket 2 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Topotecan in pocket 2 amino acids (c), calculated RMSD plot of Topotecan stat over 100 ns simulation (d), predicted poses of Pralatrexate pocket 2 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of Pralatrexate pocket 2 amino acids (g), calculated RMSD plot of Pralatrexate over 100 ns simulation (h).

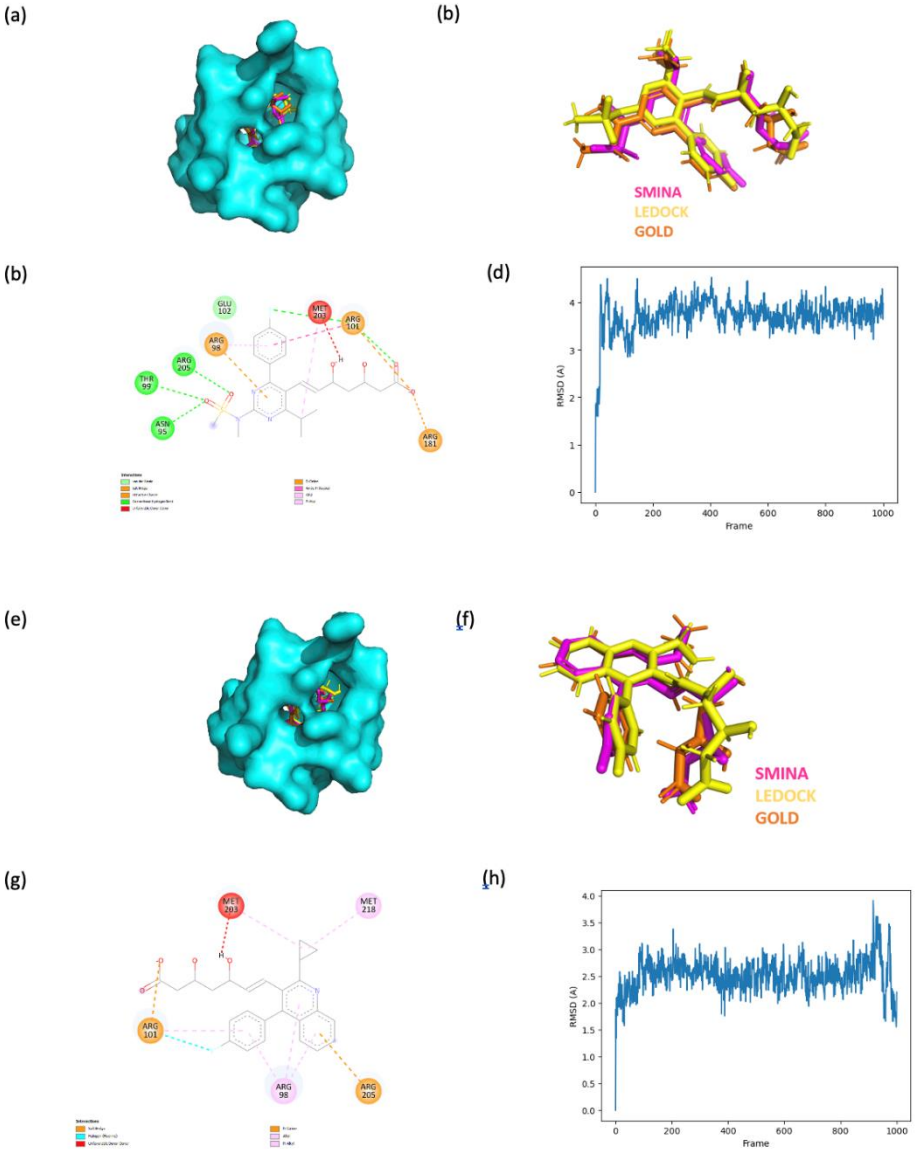


Figure 3.17. Predicted poses of Rosuvastatin in pocket 2 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Rosuvastatin in pocket 2 amino acids (c), calculated RMSD plot of Rosuvastatin over 100 ns simulation (d), predicted poses of Pitavastatin pocket 2 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of Pitavastatin pocket 2 amino acids (g), calculated RMSD plot of Pitavastatin over 100 ns simulation (h).

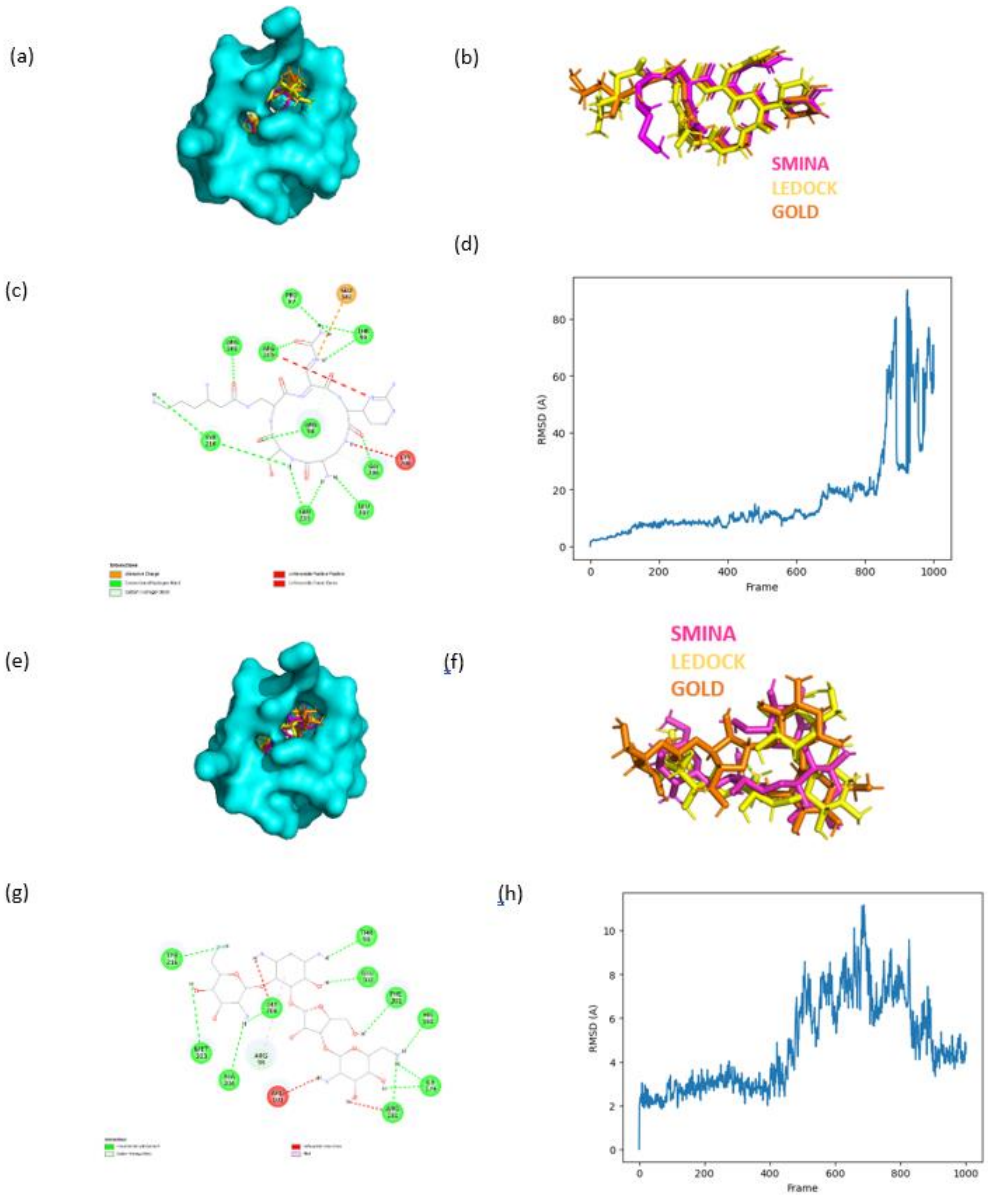


Figure 3.18. Predicted poses of Capreomycin in pocket 2 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Capreomycin in pocket 2 amino acids (c), calculated RMSD plot of Capreomycin over 100 ns simulation (d), predicted poses of Neomycin pocket 2 as a result of SMINA, LeDock and GOLD analyzes (e), comparison of predictions to see the consistency of predicted poses (f), interaction of Neomycin pocket 2 amino acids (g), calculated RMSD plot of Neomycin over 100 ns simulation (h).

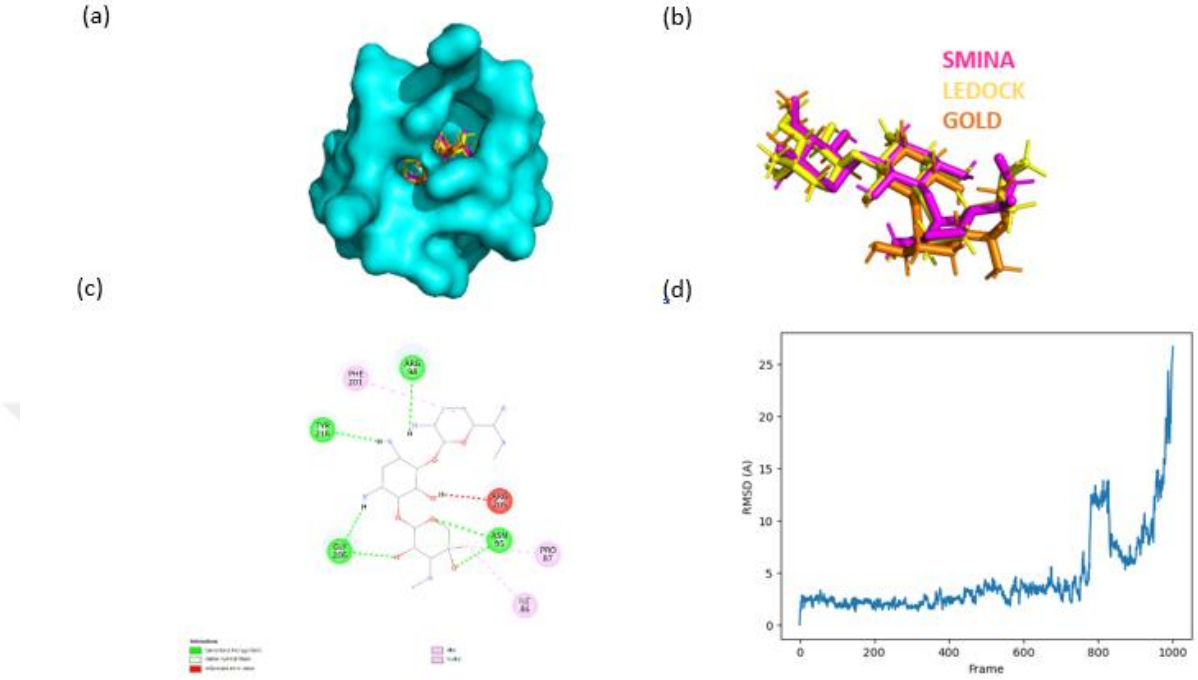


Figure 3.19. Predicted poses of Gentamicin in pocket 2 as a result of SMINA, LeDock and GOLD analysis (a), comparison of predictions to see the consistency of predicted poses (b), interaction of Gentamicin in pocket 2 amino acids (c), calculated RMSD plot of Gentamicin over 100 ns simulation (d).

4. DISCUSSION

In this thesis study, it was aimed to find potential allosteric ERK5 kinase inhibitors with a drug repurposing approach. While large-scale small compound libraries can be used to screen for potential active molecules against a protein target, a drug repurposing approach was reasoned to yield candidate molecules that can more swiftly be developed as novel selective ERK5 inhibitors. Drug repurposing/repositioning offers substantial benefits in terms of accelerating development timelines and reducing costs, while also maximizing revenue and minimizing risk. Numerous databases exist for the purpose of drug repurposing, enabling researchers to identify previously discovered drugs that target specific entities, without the need for extensive experimentation. These databases encompass a range of resources, including proteomic databases like UniProt, genomics databases such as Entrez-Gene, and pharmaceutical databases like DrugBank or PubChem. Because the compounds used have already been proven safe in humans, it shortens the drug development timeline and does not require Phase 1 clinical trials.

Here, first a sample of protein structures was created using the Monte Carlo simulation method, and then these structures were subjected to clustering. Molecular dynamics simulations were performed by selecting representatives from each cluster and binding pocket detection was performed on the obtained trajectories. Molecular docking studies were carried out using Autodock Vina, LeDock and GOLD within the scope of consensus docking for the predicted pockets. The molecules obtained from the docking results were filtered based on their scores and the consistency of the predicted poses, and MMPBSA was calculated for the remaining molecules. As a result, it was determined that tafluprost, tirofiban, panobinostat, nebivolol, dexamethasone phosphate and pitavastatin molecules gave relatively good and consistent scores and could be potential ERK5 kinase inhibitors. The docking scores obtained for these drugs were found to be within the range of scores obtained from the re-docking analysis of known allosteric inhibitors of the MAPK family, highlighting their potential for allosteric ERK5 inhibition.

Nevertheless, the study has limitations. Although the obtained results indicated certain potential for inhibition of ERK5 kinase activity, the inclusion of a larger ligand library could have led to the identification of ligands with higher scores. Moreover each of the identified ligands should be experimentally tested to determine its effect on ERK5 function, their specificities, and moreover whether they induce a potential paradoxical TAD activity increase. The focus of this study was limited to the ERK5 kinase domain due to the inability to experimentally model the entire structure of ERK5, including TAD. An elucidation of the full ERK5 structure experimentally would contribute significantly to future drug discovery efforts in this regard.

Another limitation is the incomplete observation of the effect of predicted molecules on protein dynamics. This is particularly relevant to observe whether an allosteric effect is produced on the active site of the ERK5. The 100 nanosecond-long validation simulations conducted only indicate whether the small molecule remains stable in the predicted binding pocket and conformational changes in protein structure, and may insufficiently capture longer time scale global protein dynamics. To this end, conducting longer simulations in this regard would yield more informative results.

5. CONCLUSION

This study was conducted using molecular dynamics simulations, molecular docking, and ligand binding pocket detection algorithms to find allosteric ERK5 kinase inhibitors. The primary objective of this study was to address the specificity and paradoxical activity of ERK5 kinase and discover a more specific ligands. Through the simulations and docking analyses, it was observed that certain drug molecules, such as tafluprost, tirofiban, panobinostat, nebivolol, dexamethasone phosphate, and pitavastatin, exhibited relatively favorable MMPBSA scores. Furthermore, the docking results consistently revealed conformations of these drugs that were compatible with each other. Consequently, it was inferred that these drugs possess the potential to act as allosteric inhibitors of ERK5.

REFERENCES

[1] A. Upadhyay, "Cancer: An unknown territory; rethinking before going ahead," *Genes and Diseases*, vol. 8, no. 5. Chongqing University, pp. 655–661, Sep. 01, 2021. doi: 10.1016/j.gendis.2020.09.002.

[2] S. Kumar, "Drug Targets for Cancer Treatment: An Overview," *Med Chem (Los Angeles)*, vol. 5, no. 3, 2015, doi: 10.4172/2161-0444.1000252.

[3] C. Braicu *et al.*, "A comprehensive review on MAPK: A promising therapeutic target in cancer," *Cancers*, vol. 11, no. 10. MDPI AG, Oct. 01, 2019. doi: 10.3390/cancers11101618.

[4] P. Theivendren *et al.*, "Importance of Protein Kinase and Its Inhibitor: A Review." [Online]. Available: www.intechopen.com

[5] E. D. Scheeff and P. E. Bourne, "Structural Evolution of the Protein Kinase–Like Superfamily," *PLoS Comput Biol*, vol. 1, no. 5, p. e49, Oct. 2005, doi: 10.1371/journal.pcbi.0010049.

[6] C. L. McClendon, A. P. Kornev, M. K. Gilson, and S. S. Taylora, "Dynamic architecture of a protein kinase," *Proc Natl Acad Sci U S A*, vol. 111, no. 43, pp. E4623–E4631, Oct. 2014, doi: 10.1073/pnas.1418402111.

[7] O. Laufkötter, H. Hu, F. Miljković, and J. Bajorath, "Structure- and Similarity-Based Survey of Allosteric Kinase Inhibitors, Activators, and Closely Related Compounds," *J Med Chem*, vol. 65, no. 2, pp. 922–934, Jan. 2022, doi: 10.1021/acs.jmedchem.0c02076.

[8] V. Modi and R. L. Dunbrack, "Defining a new nomenclature for the structures of active and inactive kinases," *Proc Natl Acad Sci U S A*, vol. 116, no. 14, pp. 6818–6827, Apr. 2019, doi: 10.1073/pnas.1814279116.

[9] J. Cicenas, E. Zalyte, A. Bairoch, and P. Gaudet, "Kinases and cancer," *Cancers*, vol. 10, no. 3. MDPI AG, Mar. 01, 2018. doi: 10.3390/cancers10030063.

[10] J. Yuan, X. Dong, J. Yap, and J. Hu, "The MAPK and AMPK signalings: Interplay and implication in targeted cancer therapy," *Journal of Hematology and Oncology*, vol. 13, no. 1. BioMed Central Ltd, Aug. 17, 2020. doi: 10.1186/s13045-020-00949-4.

[11] K. Milde-Langosch *et al.*, "Expression and prognostic relevance of activated extracellular-regulated kinases (ERK 1/2) in breast cancer," *Br J Cancer*, vol. 92, no. 12, pp. 2206–2215, Jun. 2005, doi: 10.1038/sj.bjc.6602655.

[12] L. Santarpia, S. M. Lippman, and A. K. El-Naggar, "Targeting the MAPKRASRAF signaling pathway in cancer therapy," *Expert Opinion on Therapeutic Targets*, vol. 16, no. 1. pp. 103–119, Jan. 2012. doi: 10.1517/14728222.2011.645805.

[13] R. Roskoski, “Properties of FDA-approved small molecule protein kinase inhibitors: A 2024 update,” *Pharmacol Res*, vol. 200, p. 107059, Feb. 2024, doi: 10.1016/j.phrs.2024.107059.

[14] K. S. Bhullar *et al.*, “Kinase-targeted cancer therapies: Progress, challenges and future directions,” *Molecular Cancer*, vol. 17, no. 1. BioMed Central Ltd., Feb. 19, 2018. doi: 10.1186/s12943-018-0804-2.

[15] M. Singha *et al.*, “Unlocking the Potential of Kinase Targets in Cancer: Insights from CancerOmicsNet, an AI-Driven Approach to Drug Response Prediction in Cancer,” *Cancers (Basel)*, vol. 15, no. 16, Aug. 2023, doi: 10.3390/cancers15164050.

[16] S. P. Davies, H. Reddy, M. Caivano, and P. Cohen, “Specificity and mechanism of action of some commonly used protein kinase inhibitors,” 2000.

[17] K. S. Bhullar *et al.*, “Kinase-targeted cancer therapies: Progress, challenges and future directions,” *Molecular Cancer*, vol. 17, no. 1. BioMed Central Ltd., Feb. 19, 2018. doi: 10.1186/s12943-018-0804-2.

[18] S. Shukla, R. W. Robey, S. E. Bates, and S. V. Ambudkar, “Sunitinib (Sutent, SU11248), a small-molecule receptor tyrosine kinase inhibitor, blocks function of the ATP-binding cassette (ABC) transporters P-glycoprotein (ABCB1) and ABCG2,” *Drug Metabolism and Disposition*, vol. 37, no. 2, pp. 359–365, Feb. 2009, doi: 10.1124/dmd.108.024612.

[19] L. O. Kirkland and C. McInnes, “Non-ATP competitive protein kinase inhibitors as anti-tumor therapeutics,” *Biochemical Pharmacology*, vol. 77, no. 10. Elsevier Inc., pp. 1561–1571, May 15, 2009. doi: 10.1016/j.bcp.2008.12.022.

[20] A. C. Dar and K. M. Shokat, “The evolution of protein kinase inhibitors from antagonists to agonists of cellular signaling,” *Annu Rev Biochem*, vol. 80, pp. 769–795, Jul. 2011, doi: 10.1146/annurev-biochem-090308-173656.

[21] J. T. Vetto, “OTT_4740_role-of-imatinib-in-the-management-of-early,” 2009. [Online]. Available: www.dovepress.com

[22] X. Lu, J. B. Smaill, and K. Ding, “New Promise and Opportunities for Allosteric Kinase Inhibitors,” *Angewandte Chemie - International Edition*, vol. 59, no. 33. Wiley-VCH Verlag, pp. 13764–13776, Aug. 10, 2020. doi: 10.1002/anie.201914525.

[23] R. Paudel, L. Fusi, and M. Schmidt, “The mek5/erk5 pathway in health and disease,” *International Journal of Molecular Sciences*, vol. 22, no. 14. MDPI, Jul. 02, 2021. doi: 10.3390/ijms22147594.

[24] B. Stecca and E. Rovida, “Impact of ERK5 on the hallmarks of cancer,” *International Journal of Molecular Sciences*, vol. 20, no. 6. MDPI AG, Mar. 02, 2019. doi: 10.3390/ijms20061426.

[25] G. N. Nithianandarajah-Jones, B. Wilm, C. E. P. Goldring, J. Müller, and M. J. Cross, "ERK5: Structure, regulation and function," *Cellular Signalling*, vol. 24, no. 11. pp. 2187–2196, Nov. 2012. doi: 10.1016/j.cellsig.2012.07.007.

[26] S. Nishimoto and E. Nishida, "MAPK signalling: ERK5 versus ERK1/2," *EMBO Reports*, vol. 7, no. 8. pp. 782–786, Aug. 2006. doi: 10.1038/sj.embor.7400755.

[27] P. A. Lochhead *et al.*, "Paradoxical activation of the protein kinase-transcription factor ERK5 by ERK5 kinase inhibitors," *Nat Commun*, vol. 11, no. 1, Dec. 2020, doi: 10.1038/s41467-020-15031-3.

[28] D. Dinev *et al.*, "Extracellular signal regulated kinase 5 (ERK5) is required for the differentiation of muscle cells," 2001.

[29] P. A. Lochhead *et al.*, "Paradoxical activation of the protein kinase-transcription factor ERK5 by ERK5 kinase inhibitors," *Nat Commun*, vol. 11, no. 1, Dec. 2020, doi: 10.1038/s41467-020-15031-3.

[30] G. N. Nithianandarajah-Jones, B. Wilm, C. E. P. Goldring, J. Müller, and M. J. Cross, "ERK5: Structure, regulation and function," *Cellular Signalling*, vol. 24, no. 11. pp. 2187–2196, Nov. 2012. doi: 10.1016/j.cellsig.2012.07.007.

[31] P. A. Lochhead *et al.*, "Paradoxical activation of the protein kinase-transcription factor ERK5 by ERK5 kinase inhibitors," *Nat Commun*, vol. 11, no. 1, Dec. 2020, doi: 10.1038/s41467-020-15031-3.

[32] S. J. Cook, J. A. Tucker, and P. A. Lochhead, "Small molecule ERK5 kinase inhibitors paradoxically activate ERK5 signalling: be careful what you wish for…," *Biochem Soc Trans*, vol. 48, no. 5, pp. 1859–1875, Oct. 2020, doi: 10.1042/BST20190338.

[33] B. M. Sahoo, B. V. V. Ravi Kumar, J. Sruti, M. K. Mahapatra, B. K. Banik, and P. Borah, "Drug Repurposing Strategy (DRS): Emerging Approach to Identify Potential Therapeutics for Treatment of Novel Coronavirus Infection," *Frontiers in Molecular Biosciences*, vol. 8. Frontiers Media S.A., Feb. 26, 2021. doi: 10.3389/fmolb.2021.628144.

[34] N. Krishnamurthy, A. A. Grimshaw, S. A. Axson, S. H. Choe, and J. E. Miller, "Drug repurposing: a systematic review on root causes, barriers and facilitators," *BMC Health Serv Res*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12913-022-08272-z.

[35] V. S. Kulkarni, V. Alagarsamy, V. R. Solomon, P. A. Jose, and S. Murugesan, "Drug Repurposing: An Effective Tool in Modern Drug Discovery," *Russ J Bioorg Chem*, vol. 49, no. 2, pp. 157–166, Apr. 2023, doi: 10.1134/S1068162023020139.

[36] P. Ayyar and U. Subramanian, "Repurposing – second life for drugs," *Pharmacia*, vol. 69, no. 1, pp. 51–59, 2022, doi: 10.3897/PHARMACIA.69.E72548.

[37] A. Bateman *et al.*, "UniProt: The universal protein knowledgebase," *Nucleic Acids Res*, vol. 45, no. D1, pp. D158–D169, Jan. 2017, doi: 10.1093/nar/gkw1099.

[38] H. M. Berman *et al.*, "The Protein Data Bank," 2000. [Online]. Available: http://www.rcsb.org/pdb/status.html

[39] B. Webb and A. Sali, "Comparative protein structure modeling using MODELLER," *Curr Protoc Bioinformatics*, vol. 2016, pp. 5.6.1-5.6.37, 2016, doi: 10.1002/cpbi.3.

[40] M. Jamroz, A. Kolinski, and S. Kmiecik, "CABS-flex: Server for fast simulation of protein structure fluctuations.," *Nucleic Acids Res*, vol. 41, no. Web Server issue, 2013, doi: 10.1093/nar/gkt332.

[41] E. Paquet and H. L. Viktor, "Molecular dynamics, monte carlo simulations, and langevin dynamics: A computational review," *BioMed Research International*, vol. 2015. Hindawi Publishing Corporation, 2015. doi: 10.1155/2015/183918.

[42] D. J. Earl and M. W. Deem, "Monte Carlo Simulations."

[43] M. Jamroz, A. Kolinski, and S. Kmiecik, "CABS-flex: Server for fast simulation of protein structure fluctuations.," *Nucleic Acids Res*, vol. 41, no. Web Server issue, 2013, doi: 10.1093/nar/gkt332.

[44] A. Kuriata *et al.*, "CABS-flex 2.0: A web server for fast simulations of flexibility of protein structures," *Nucleic Acids Res*, vol. 46, no. W1, pp. W338–W343, Jul. 2018, doi: 10.1093/nar/gky356.

[45] T. Tubiana, J. C. Carvaillo, Y. Boulard, and S. Bressanelli, "TTClust: A Versatile Molecular Simulation Trajectory Clustering Program with Graphical Summaries," *J Chem Inf Model*, vol. 58, no. 11, pp. 2178–2182, Nov. 2018, doi: 10.1021/acs.jcim.8b00512.

[46] M. R. Challapa-Mamani *et al.*, "Molecular Docking and Molecular Dynamics Simulations in Related to Leishmania donovani: An Update and Literature Review," *Tropical Medicine and Infectious Disease*, vol. 8, no. 10. Multidisciplinary Digital Publishing Institute (MDPI), Oct. 01, 2023. doi: 10.3390/tropicalmed8100457.

[47] O. M. H. Salo-Ahen *et al.*, "Molecular dynamics simulations in drug discovery and pharmaceutical development," *Processes*, vol. 9, no. 1. MDPI AG, pp. 1–63, 2021. doi: 10.3390/pr9010071.

[48] M. Hernández-Rodríguez, M. C. Rosales-Hernández, J. E. Mendieta-Wejebe, M. Martínez-Archundia, and J. Correa Basurto, "Current Tools and Methods in Molecular Dynamics (MD) Simulations for Drug Design," *Curr Med Chem*, vol. 23, no. 34, pp. 3909–3924, Jun. 2016, doi: 10.2174/0929867323666160530144742.

[49] S. Pronk *et al.*, "GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, no. 7, pp. 845–854, Apr. 2013, doi: 10.1093/bioinformatics/btt055.

[50] K. Vanommeslaeghe *et al.*, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields," *J Comput Chem*, vol. 31, no. 4, pp. 671–690, Mar. 2010, doi: 10.1002/jcc.21367.

[51] I. Aier, P. K. Varadwaj, and U. Raj, "Structural insights into conformational stability of both wild-type and mutant EZH2 receptor," *Sci Rep*, vol. 6, Oct. 2016, doi: 10.1038/srep34984.

[52] L. Martínez, "Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis," *PLoS One*, vol. 10, no. 3, Mar. 2015, doi: 10.1371/journal.pone.0119264.

[53] Y. Yuan, J. Pei, and L. Lai, "Send Orders of Reprints at reprints@benthamscience.net Binding Site Detection and Druggability Prediction of Protein Targets for Structure-Based Drug Design," 2013. [Online]. Available: http://fpocket.sourceforge.net/dcd

[54] A. Kumar, R. Puri, K. Gupta, and A. Pal, "To Predict Human Biomarker for the Obesity Using Mouse Homologous Expression Data at Different Theiler Stages," *International Letters of Natural Sciences*, vol. 45, pp. 9–17, Aug. 2015, doi: 10.56431/p-5t9c53.

[55] A. Volkamer, D. Kuhn, F. Rippmann, and M. Rarey, "Dogsitescorer: A web server for automatic binding site prediction, analysis and druggability assessment," *Bioinformatics*, vol. 28, no. 15, pp. 2074–2075, Aug. 2012, doi: 10.1093/bioinformatics/bts310.

[56] A. Volkamer, D. Kuhn, F. Rippmann, and M. Rarey, "Dogsitescorer: A web server for automatic binding site prediction, analysis and druggability assessment," *Bioinformatics*, vol. 28, no. 15, pp. 2074–2075, Aug. 2012, doi: 10.1093/bioinformatics/bts310.

[57] L. Pinzi and G. Rastelli, "Molecular docking: Shifting paradigms in drug discovery," *International Journal of Molecular Sciences*, vol. 20, no. 18. MDPI AG, Sep. 01, 2019. doi: 10.3390/ijms20184331.

[58] A. Tripathi and V. A. Bankaitis, "Molecular Docking: From Lock and Key to Combination Lock."

[59] L. Masters, S. Eagon, and M. Heying, "Evaluation of consensus scoring methods for AutoDock Vina, smina and idock," *J Mol Graph Model*, vol. 96, May 2020, doi: 10.1016/j.jmgm.2020.107532.

[60] N. Liu and Z. Xu, "Using LeDock as a docking tool for computational drug design," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Feb. 2019. doi: 10.1088/1755-1315/218/1/012143.

[61] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor, "Improved Protein-Ligand Docking Using GOLD," 2003.

[62] R. Ochoa, K. Palacio-Rodriguez, C. M. Clemente, and N. S. Adler, “dockECR: Open consensus docking and ranking protocol for virtual screening of small molecules,” *J Mol Graph Model*, vol. 109, Dec. 2021, doi: 10.1016/j.jmgm.2021.108023.

[63] B. Dankwa, E. Broni, K. S. Enninful, S. K. Kwofie, and M. D. Wilson, “Consensus docking and MM-PBSA computations identify putative furin protease inhibitors for developing potential therapeutics against COVID-19,” *Struct Chem*, vol. 33, no. 6, pp. 2221–2241, Dec. 2022, doi: 10.1007/s11224-022-02056-1.

[64] E. Pihan, L. Colliandre, J. F. Guichou, and D. Douguet, “E-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design,” *Bioinformatics*, vol. 28, no. 11, pp. 1540–1541, Jun. 2012, doi: 10.1093/bioinformatics/bts186.

[65] C. Wang, D. Greene, L. Xiao, R. Qi, and R. Luo, “Recent developments and applications of the MMPBSA method,” *Frontiers in Molecular Biosciences*, vol. 4, no. JAN. Frontiers Media S.A., Jan. 10, 2018. doi: 10.3389/fmolb.2017.00087.

[66] T. Tuccinardi, “What is the current value of MM/PBSA and MM/GBSA methods in drug discovery?,” *Expert Opinion on Drug Discovery*, vol. 16, no. 11. Taylor and Francis Ltd., pp. 1233–1237, 2021. doi: 10.1080/17460441.2021.1942836.

[67] G. Poli, C. Granchi, F. Rizzolio, and T. Tuccinardi, “Application of MM-PBSA methods in virtual screening,” *Molecules*, vol. 25, no. 8. MDPI AG, Apr. 01, 2020. doi: 10.3390/molecules25081971.

[68] M. S. Valdés-Tresanco, M. E. Valdés-Tresanco, P. A. Valiente, and E. Moreno, “Gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS,” *J Chem Theory Comput*, vol. 17, no. 10, pp. 6281–6291, Oct. 2021, doi: 10.1021/acs.jctc.1c00645.

[69] E. W. Bell and Y. Zhang, “DockRMSD: An open-source tool for atom mapping and RMSD calculation of symmetric molecules through graph isomorphism,” *J Cheminform*, vol. 11, no. 1, 2019, doi: 10.1186/s13321-019-0362-7.

BIOGRAPHY

I graduated from Biruni University, Department of Molecular Biology and Genetics in 2020. At the same time, I completed my minor in Biomedical Engineering. I started working as a molecular biologist in 2021. I am currently continuing my master's degree in Bioinformatics and Systems biology at Gebze Technical University and to work as a molecular biologist at Gen Era Diagnostics.

PUBLICATIONS AND PRESENTATIONS FROM THE THESIS

İlayda Erdoğan, Gülseren Turhal, Asuman Demiroğlu-Zergeroğlu, Onur Serçinoğlu, "Structure-based Discovery of Potential Allosteric Inhibitors of ERK5 Kinase". 16th The International Symposium on Health Informatics and Bioinformatics, Poster Presentation, Ankara, 2023

İlayda Erdoğan, Gülseren Turhal, Asuman Demiroğlu-Zergeroğlu, Onur Serçinoğlu, "Repurposing of FDA-approved Drugs as Potential Allosteric Inhibitors of ERK5 Kinase". Gebze Technical University, 7th GTU Graduate Research Symposium, Poster Presentation, Gebze/Kocaeli, 2023.

APPENDIX

Table A1. Allosteric ligand-containing kinase structures used to test the consistency of docking applications.

PDB ID	METHOD	RESOLUTION	FAMILY	LIGAND
3HL7	X-Ray	1.88 Å	MAPK14	2-fluoro-4-[4-(4-fluorophenyl)-1H-pyrazol-3-yl]pyridine
3NEW	X-Ray	2.51 Å	MAPK14	4-(trifluoromethyl)-3-[3-(trifluoromethyl)phenyl]-1,7-dihydro-6H-pyrazolo[3,4-b]pyridin-6-one
3ROC	X-Ray	1.70 Å	MAPK14	4-[4-(4-fluorophenyl)-1H-pyrazol-3-yl]pyridine
4DLJ	X-Ray	2.60 Å	MAPK14	2-phenyl-N~4~-(2-phenylethyl)quinazoline-4,7-diamine
4E6C	X-Ray	2.39 Å	MAPK14	(1,1-dimethylpiperidin-1-ium-4-yl) octadecyl hydrogen phosphate
4EH2	X-Ray	2.00 Å	MAPK14	N~4~-cyclopropyl-2-phenylquinazoline-4,7-diamine
4F9Y	X-Ray	1.85 Å	MAPK14	4-[3-(4-FLUOROPHENYL)-1H-PYRAZOL-4-YL]PYRIDINE
5N64	X-Ray	2.40 Å	MAPK14	2-phenyl-~{N}4-(thiophen-2-ylmethyl)quinazoline-4,7-diamine
5N66	X-Ray	2.40 Å	MAPK14	~{N}4-[[4-(cyclopropylmethyl)furan-2-yl]methyl]-2-phenyl-quinazoline-4,7-diamine
5N67	X-Ray	1.90 Å	MAPK14	1-[4-[4-[7-azanyl-4-(2-phenylethylamino)quinazolin-2-yl]phenyl]piperazin-1-yl]ethanone
5N68	X-Ray	1.85 Å	MAPK14	2-(4-morpholin-4-ylphenyl)-~{N}4-(2-phenylethyl)quinazoline-4,7-diamine
5O8V	X-Ray	2.00 Å	MAPK14	~{N}-[3-[7-azanyl-4-(2-phenylethylamino)quinazolin-2-yl]phenyl]propanamide

Table A2. Results obtained from the redocking application to test the reliability of docking applications.

PDB ID	SMINA Score (kcal/mol)	LeDock Score (kcal/mol)	GOLD Score
3HL7	-10.9	-3.87	80.24
3NEW	-12.1	-4.65	84.14
3ROC	-10.8	-3.7	75.29
4DLJ	-11.9	-5.36	101.15
4E6C	-6	-14.29	52.58
4EH2	-10.5	-4.66	84.6
4F9Y	-10.5	-3.63	73.96
5N64	-10.9	-5.26	91.9
5N66	-11.7	-5.54	107.85
5N67	-12.7	-6.27	116.49
5N68	-12.3	-7.5	108.16
5O8V	-11.7	-5.8	100.97

Table A3. RMSD calculation results made to define the consistency of redocking results between each other.

PDB ID	SMINA (Å)	LeDock (Å)	GOLD (Å)
3HL7	0.58	3.31	0.58
3NEW	0.81	0.23	0.62
3ROC	0.32	0.27	0.47
4DLJ	0.61	0.79	1.31
4E6C	4.29	7.07	4.19
4EH2	0.85	0.92	0.63
4F9Y	0.44	0.26	0.48
5N64	6.03	1.49	0.97
5N66	0.58	0.79	0.51
5N67	1.82	0.81	0.48
5N68	1.87	1.14	0.54
5O8V	1.19	1.03	1.12

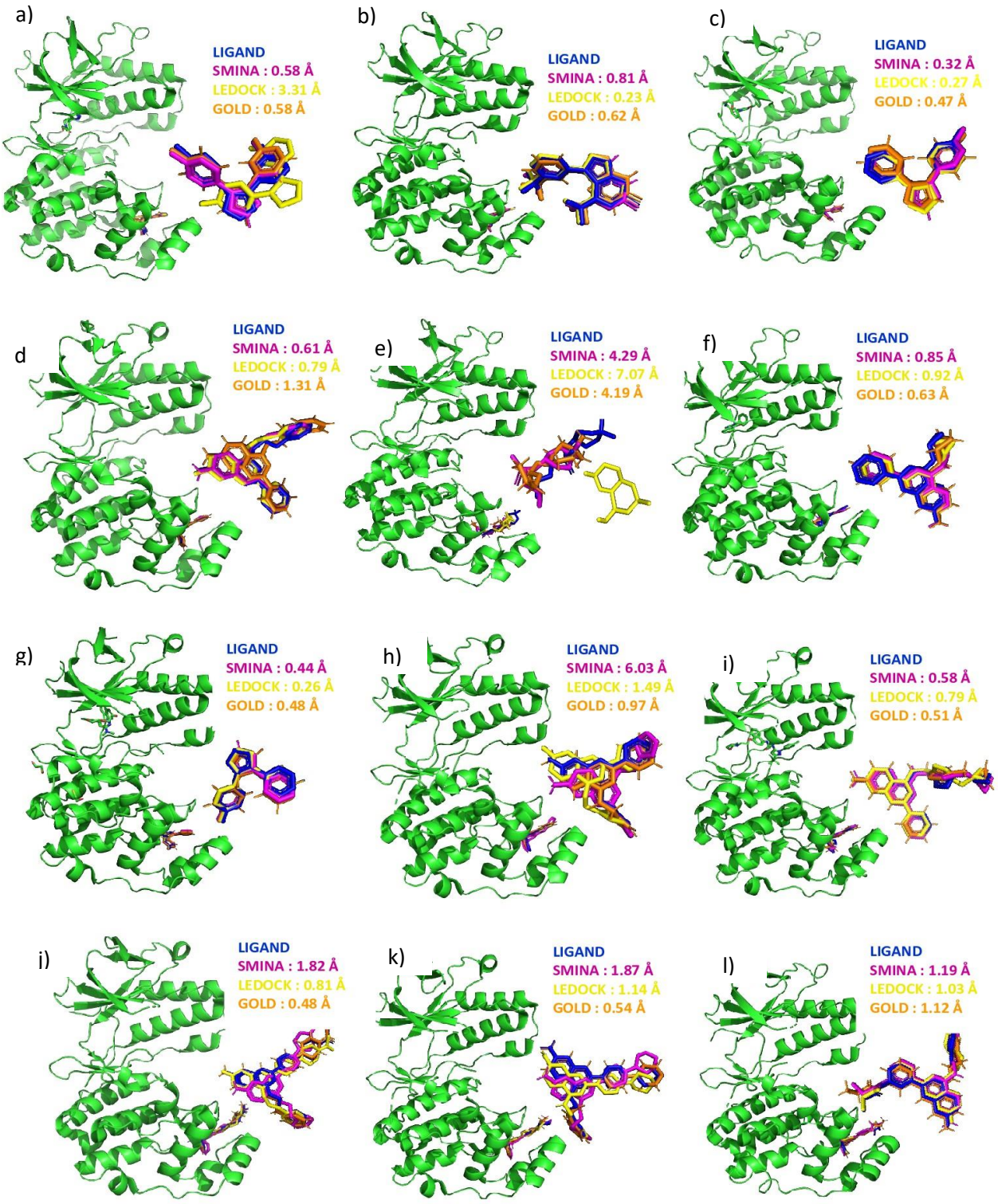


Figure A1. Representation of redocking poses on the MAPK structure and comparison of poses obtained from algorithms. 3HL7 (a), 3NEW (b), 3ROC (c), 4DLJ (d), 4E6C (e), 4EH2 (f), 4F9Y (g), 5N64 (h), 5N66 (i), 5N67 (j), 5N68(k), 5O8V(l).

OTHER APPENDICES (CD)

- Python scripts used in the study.

- LeDock input files.

- GOLD input files.