

Ceyhun SARI

M.Sc. Thesis

AGU 2025

DESIGN AND IDENTIFICATION OF NOVEL CANDIDATES AGAINST THE TYROSINE KINASE DOMAIN OF ALK BY COMPREHENSIVE IN SILICO APPROACHES

M.Sc. THESIS

SUBMITTED TO THE DEPARTMENT OF BIOENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND
SCIENCE OF ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By

Ceyhun SARI

August 2025

DESIGN AND IDENTIFICATION OF NOVEL
CANDIDATES AGAINST THE TYROSINE
KINASE DOMAIN OF ALK BY
COMPREHENSIVE IN SILICO APPROACHES

A THESIS
SUBMITTED TO THE DEPARTMENT OF BIOENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF
ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Ceyhun SARI
August 2025

SCIENTIFIC ETHICS COMPLIANCE

I hereby declare that all information in this document has been obtained in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name-Surname: Ceyhun SARI

Signature:

X X X X

REGULATORY COMPLIANCE

M.Sc. thesis titled Identification of Novel Candidates Against the Tyrosine Kinase Domain of ALK by Comprehensive In Silico Approaches has been prepared in accordance with the Thesis Writing Guidelines of the Abdullah Gül University, Graduate School of Engineering & Science.

Prepared By
Ceyhun SARI
Signature

Advisor
Asst. Prof. İsmail AKÇOK
Signature

Head of the Bioengineering Program
Asst. Prof. İsmail AKÇOK
Signature

ACCEPTANCE AND APPROVAL

M.Sc. titled Identification of Novel Candidates Against the Tyrosine Kinase Domain of ALK by Comprehensive In Silico Approaches and prepared by Ceyhun SARI has been accepted by the jury in the Bioengineering Graduate Program at Abdullah Gül University, Graduate School of Engineering & Science.

08 /08 / 2025

JURY:

Advisor: Asst. Prof. İsmail AKÇOK

Member: Asst. Prof. Özkan FİDAN

Member: Asst. Prof. Şerife AYAZ GÜNER

APPROVAL:

The acceptance of this M.Sc. thesis has been approved by the decision of the Abdullah Gül University, Graduate School of Engineering & Science, Executive Board dated /..... / and numbered

..... / /

(Date)

Graduate School Dean
Prof. İrfan ALAN

IDENTIFICATION OF NOVEL CANDIDATES AGAINST THE TYROSINE KINASE DOMAIN OF ALK BY COMPREHENSIVE IN SILICO APPROACHES

Ceyhun SARI
MSc in Bioengineering
Advisor: Asst. Prof. İsmail AKÇOK
August 2025

After the first description of Anaplastic Lymphoma Kinase (ALK) in an anaplastic large cell lymphoma cell line as a nucleophosmin (NPM) fusion partner, ALK and its various fusion partners have been implicated in numerous cancers such as non-small cell lung cancer (NSCLC), anaplastic large cell lymphoma (ALCL), neuroblastoma, rhabdomyosarcoma. In the last decade, several compounds targeting ALK have been developed and approved by the Food and Drug Administration (FDA). Despite the advances of generations of ALK inhibitors, a recent study highlighted that around half of the ALK-positive NSCLC patients will go through disease progression in response to first-line Alectinib and Lorlatinib, which is a second-generation and third generation ALK inhibitors, respectively. Given these points, this study focused on two distinct paths to discover and develop novel compounds that could target tyrosine kinase domain of the ALK. Firstly, we designed more than 200 α -carboline derivatives and investigated their binding properties against ALK tyrosine kinase by using in silico protocols consisting of molecular docking studies, molecular dynamics simulations, and MM/PBSA binding free energy calculation. As a second approach, we virtually screened a large natural product database to a novel drug candidate for possibly inhibiting ALK TK and investigated their binding properties by similar ways utilised in the first approach. Considering the obtained results, we developed three promising candidates, compounds 208, 209 and CNP106316.1 with -9.08, -9.80, and -11.6 kcal/mol and binding energies respectively, which demonstrated improved binding profiles over their respective MD simulations, 300ns for carbolines, 500ns for the NP.

Keywords: Virtual Screening, Molecular Docking, Molecular Dynamics, Drug Discovery, Drug Design

ALK'NİN TİROZİN KİNAZ BÖLGESİNİ HEDEFLEYEN YENİ İLAÇ ADAYLARININ KAPSAMLI IN SILICO YAKLAŞLARIMLA TANIMLANMASI

Ceyhun SARI
Biyomühendislik Anabilim Dalı Yüksek Lisans
Tez Danışmanı: Dr. Öğr. Üyesi İsmail AKÇOK
Ağustos 2025

Anaplastik büyük hücreli lenfoma hücre hatlarında füzyon protein ortağı olarak keşif edilen Anaplastik Lenfoma Kinaz'ın (ALK), keşfinden bu yana ALK çeşitli füzyon ortakları ile çok sayıda kanserde rol oynadığı ortaya çıkmıştır. Rol oynadıkları kanser şu şekilde sıralanabilir: küçük hücreli olmayan akciğer kanseri (NSCLC); anaplastik büyük hücreli lenfoma (ALCL); nöroblastom; rabdomiyosarkom; vb. Son yılda, ABD Gıda ve İlaç Dairesi (FDA) ALK'yi hedefleyen birçok bileşik onay almıştır. Bu gelişmelere rağmen, yakın zamanda yapılan bir çalışma ALK pozitif NSCLC hastalarının yaklaşık yarısının hastalık ilerlemesi yaşanacağını vurgulanmıştır, başlangıç tedavisi olan Alectinib, ikinci nesil ALK inhibitörü ve üçüncü nesil ALK inhibitörü Lorlatinib rağmen. Bu noktaları göze alarak, bu çalışma ALK'nin tirozin kinaz alanını hedefleyebilecek yeni bileşikler keşfetmek ve geliştirmek için iki farklı yola odaklanmıştır. İlk yaklaşım olarak 200'den fazla α -carboline türevi tasarladık. Devamında moleküler yanalştırma (Docking), moleküler dinamik (MD) simülasyonlarını MM/PBSA ile serbest bağlanma enerjisi hesaplamalarından oluşan in silico protokolleri kullanarak tasarılarımızın hedefimize karşı bağlanma özelliklerini araştırdık. İkinci yaklaşım olarak büyük bir doğal ürün veritabanını aynı amaca yönelik yeni bir ilaç adayı keşfetme adına sanal olarak taradık. Devamında bağlanma özelliklerini ilk yaklaşımda kullanılan yöntemlerle inceledik. Elde edilen bütün sonuçları göz önünde bulundurarak, sonuçlar takip eden şekilde özetlenebilir. Üç umut verici ilaç adayı aralarından yükselmiştir test edilen bileşikler arasında, bileşik 208, 209 ve CNP0106316.1. Serbest bağlanma enerjileri ise sırasıyla -9.08, -9.80 ve -11.6 kcal/mol olarak bulunmuştur. Ek olarak, ismi geçen bileşikler ilgili MD simülasyonlarında stabil bağlanma profilleri göstermişlerdir.

Anahtar Kelimeler: Sanal Tarama, Moleküler Yanalştırma, Moleküler Dinamik simülasyonları, İlaç Keşfi, İlaç Tasarımı

Acknowledgements

I would like to begin by expressing my sincere gratitude to Dr. İsmail Akçok for his unwavering academic support, insightful feedback, and invaluable guidance throughout this thesis. I am especially grateful for the time he generously dedicated to reviewing my work and helping me stay on track. This thesis would not have been possible without his patience and consistent encouragement.

I would also like to extend my gratitude to Dr. Emel Başak Gencer Akçok for her expertise in protein biochemistry and for the critical role she played throughout my research. Her support, thoughtful input, and willingness to almost act as an unofficial co-advisor greatly enriched the quality and depth of this thesis.

I would like to acknowledge the institutional and technical support that contributed to this research. I am especially thankful to Dr. İsmail Akçok not only for his academic guidance but also for maintaining access to the Life and Natural Sciences Department's workstation, codenamed "Togo." This resource was vital to running computational tasks central to the project. I am also grateful to TRUBA (Turkish Academic Network and Information Center) for providing high-performance computing infrastructure, which significantly supported the essential computational tasks.

I would like to express my appreciation to Lecturer Hüseyin Güner for his initial support and encouragement at the beginning of this thesis, particularly in helping me establish the research environment and get off to a strong start. I am also grateful to my lab mate, Muhsin Samet Yücel, for sharing this thesis journey alongside me. Working on our projects in parallel provided both motivation and a sense of fellowship.

I would like to express my deepest gratitude to my family for their unwavering support, patience, and encouragement throughout this journey. I am also thankful to Enes Çakmaklı, my dear friend, for his persistent and insightful layman questions about my research. His curiosity challenged me to clarify my understanding and communicate my work effectively to a broader audience.

TABLE OF CONTENTS

1. INTRODUCTION	1
1.1 THE DRUG DISCOVERY AND DEVELOPMENT	1
1.1.1 <i>Drug Discovery Methods</i>	2
1.2 COMPUTER-AIDED DRUG DESIGN APPROACHES	4
1.2.1 <i>Structure-Based Drug Design</i>	4
1.3 DRUG DISCOVERY IN PROTEIN KINASES.....	7
1.3.1 <i>Anaplastic Lymphoma Kinase (ALK)</i>	8
1.3.2 <i>ALK in disease</i>	11
1.3.3 <i>ALK inhibitors</i>	12
1.3.4 <i>Objectives & Scope of the Thesis</i>	16
2. EXPERIMENTAL PROCEDURE	18
2.1 DESIGN AND EVALUATION OF A-CARBOLINE DERIVATIVES FOR ALK TYROSINE KINASE DOMAIN USING COMPREHENSIVE IN SILICO METHODS	18
2.1.1 <i>Target Protein and Ligand Preparation</i>	18
2.1.2 <i>Molecular Docking</i>	18
2.1.3 <i>MD Simulation</i>	19
2.1.4 <i>MM-PBSA Calculations</i>	20
2.1.5 <i>Comparative Essential Dynamics and Clustering</i>	20
2.2 IN SILICO DISCOVERY OF NOVEL ALK INHIBITORS FROM A NATURAL PRODUCT DATABASE.....	20
2.2.1 <i>Pre-Docking Procedures - Protein Structure Equilibration</i>	20
2.2.2 <i>Ligand Acquisition and Preparation</i>	21
2.2.3 <i>Molecular Docking</i>	22
2.2.4 <i>MD Simulation</i>	23
2.2.5 <i>MM-PBSA Calculations</i>	23
3. RESULTS & DISCUSSION.....	24
3.1 DESIGN AND EVALUATION OF A-CARBOLINE DERIVATIVES FOR ALK TYROSINE KINASE DOMAIN USING COMPREHENSIVE IN SILICO METHODS	24
3.1.1 <i>Molecular Docking Studies of the Designed Ligands</i>	25
3.1.2 <i>Molecular Dynamics Simulation of Selected Ligands Against ALK Kinase Domain</i>	31
3.1.3 <i>MM/PBSA Average Binding Energy Calculation</i>	41
3.1.4 <i>Towards the Design of More Stable A-Carboline ALK Kinase Inhibitors</i>	42
3.2 IN SILICO DISCOVERY OF NOVEL ALK INHIBITORS FROM A NATURAL PRODUCT DATABASE.....	52
3.2.1 <i>Molecular Docking of the COCONUT NP Database against ALK tyrosine kinase domain</i>	52
3.2.2 <i>MD simulation of the selected top 5 NPs against tyrosine kinase domain of the ALK</i>	57
3.2.3 <i>MM/PBSA Average Binding Energy Calculation Between NPs and ALK TK over the Course of MD Simulation</i>	68
4. CONCLUSIONS AND FUTURE PROSPECTS.....	70

4.1 CONCLUSIONS	70
4.2 SOCIETAL IMPACT AND CONTRIBUTION TO GLOBAL SUSTAINABILITY	71
4.3 FUTURE PROSPECTS.....	72



LIST OF FIGURES

Figure 1.1 Small molecule kinase inhibitors discovered using CADD [15].	6
Figure 1.2 Timeline of approved kinase inhibitors. Drugs approved for non-oncology indicated with blue coloured boxes, while nomenclature of kinase inhibitors and monoclonal antibodies were provided in the legend [38]	8
Figure 1.3 A) Representation of general structure of human physiological ALK receptor protein-tyrosine kinase, light-blue section represents extracellular segment (residues 19 -- 1038), orange section represents protein-tyrosine kinase domain (residues 1116 – 1392) B) 3D structure of ALK kinase domain (PDB ID 8ARJ) and the amino acids that are highlighted with different colours signify various regions as follows: red: Glycine-rich loop's start, green: α C-glutamate, blue: Hinge residues, yellow: Gatekeeper Residue, magenta: First Aspartate of K/D/D catalytic signature [49]	10
Figure 1.4 Structures of ALK tyrosine kinase inhibitors approved by the FDA and in various stages of clinical trial in last 5 years. (A-F: Approved by FDA, G-O: various stages of clinical trial.) A) Crizotinib, B) Ceritinib, C) Alectinib, D) Brigatinib, E) Ensartinib, F) Lorlatinib, G) XZP-3621, H) NVL-655, I) WX-0593, J) TGRX-326, K) Gilteritinib, L) CT-707, M) SGR-1505, N) TPX-0005, O) BMS-986369	13
Figure 1.5 Structures of ALK tyrosine kinase inhibitors that has been validated in vivo test in last 5 years.	15
Figure 1.6 Structure of ALK tyrosine kinase inhibitors that has been validated only in vitro test so far in last 5 years.	16
Figure 2.1 Clustering analysis of 8ARJ equilibrium MD and two prominent clusters' alignment to 8ARJ. White protein structures represents 8ARJ while red represents cluster 1 and magenta represents cluster 5. Orange circle highlights the "closed" structure of cluster 5 compared to original. A) Histogram plot of number of members across all clusters. B) A linear projection of clusters across the trajectory. C) Root mean square deviation (RMSD), ($\text{\AA}$) matrix between clusters. D) Cluster 1-8ARJ overlap. E) Cluster 5-8ARJ overlap.	21
Figure 3.1 RMSD plots from their MD simulation. Blue line represents the backbone atoms' RMSD, the orange line represents the protein atoms' RMSD, and the green	

line represents the RMSD values between the compound and the backbone atoms.	32
Figure 3.2 RMSF analysis of the selected compounds against ALK kinase domain.	34
Figure 3.3 Rg analysis of the selected compounds against ALK kinase domain.	35
Figure 3.4 Average distance (in Å) heatmaps of 14 compounds, the lowest binding energy from each category during 100 ns simulation.	37
Figure 3.5 RMSD analysis of 6 short-listed compounds from triplicate MD simulations.	39
Figure 3.6 RMSF analysis of 6 short-listed compounds from triplicate MD simulations.	39
Figure 3.7 Rg analysis of 6 short-listed compounds from triplicate MD simulations....	40
Figure 3.8 Number of hydrogen bonds between the various ligands and the ALK kinase domain over the course of the simulation.....	41
Figure 3.9 MM/PBSA average binding free energy graphs of 6 shortlisted compounds from triplicate 100ns MD simulations.....	42
Figure 3.10 Structure of compound 193, as an example of chalcone, substituted carboline, and the representation of rotation on sigma bond between two phenyl groups.....	43
Figure 3.11 MD simulation analysis of compounds 208 and 209. A, B, and C sections represent the compound 208; D, E, and F sections represent the compound 209. A, D) RMSD plot of ALK kinase domain with compounds over 300 ns MD simulation. B, E) Average distance heatmaps. C, F) Number of hydrogen bonds between the ligands and the protein over the course of the simulation.	46
Figure 3.12 Principal component analysis of the compounds 208 and 209 – ALK TK complexes, and only ALK TK. Percentages signify how much of variance that PC account for.....	47
Figure 3.15 RMSD-Rg based FEL analysis of only tyrosine kinase domain, 208-ALK TK complex, and 209-ALK TK complex.	50
Figure 3.16 MM/PBSA average binding free energy graph of modified compounds from triplicate 300ns MD simulation	51
Figure 3.17 RMSD values between protein's backbone and NP atoms across 500ns MD simulation time for each selected NP	59
Figure 3.18 Backbone atoms' RMSF analysis of selected top 5 NP-ALK TK complexes	61

Figure 3.19 Protein atoms' Rg analysis of selected top 5 NP-ALK TK complexes.....	63
Figure 3.20 Average distance heatmaps of selected top 5 NPs.....	65
Figure 3.21 Number of hydrogen bonds between NPs and ALK TK domain over the course of the simulation.	67
Figure 3.22 MM/PBSA average binding free energies of the selected top 5 compounds and Lorlatinib.....	68



LIST OF TABLES

Table 3.1 The list of selected compounds from designed α -Carbolines with binding affinity towards kinase domain of ALK	25
Table 3.2 2D and 3D interactions docking results of the selected designed compounds (C. no: A = 5, B = 23, C = 49, D = 60, E = 72, F = 81, G = 101, H = 110, I = 127, J = 154, K = 164, L = 180, M = 183, N = 196)	27
Table 3.3 Reproducibility analysis of 6 short-listed compounds from triplicate MD simulations.....	38
Table 3.4 Molecular Docking Evaluations and 2D interactions of the modified compounds.....	44
Table 3.5 The top 5 highest binding affinity of COCONUT NPs that also have at least one hydrogen bond with hinge residues	53
Table 3.6 2D and 3D interactions between top 5 NPs and ALK tyrosine kinase domain.	54
Appendix Table 1. The list of designed Carbolines with binding affinity towards kinase domain of ALK	84
Appendix Table 2. The top lowest 109 of COCONUT docking results against ALK tyrosine domain.....	90

LIST OF ABBREVIATIONS

ADME	Absorption, Distribution, Metabolism, Excretion
ALCL	Anaplastic large cell lymphoma
ALK	Anaplastic Lymphoma Kinase
ATP	Adenosine Triphosphate
CADD	Computer-Aided drug design
CGenFF	CHARMM General Force Field
CLC	Chalcone
COCONUT	Collection of open natural products
COM	Centre of Mass
EML4	Echinoderm microtubule-associated protein-like 4
FEL	Free energy Landscape
FDA	Food and Drug Administration
HTS	High-Throughput Screening
JAK-STAT	Janus kinase/signal transducer and activator of transcription
LALa	Low-density lipoprotein class a
MAPK	Mitogen Activated Protein Kinase
MD	Molecular Dynamics
MM/PBSA	Molecular mechanics Poisson-Boltzmann surface area
NP	Natural Product
NPM	Nucleophosmin
NPT	Fixed number of particles, pressure, temperature
NSCLC	Non-Small Cell Lung Cancer
NVT	Fixed number of particles, volume, temperature
PCA	Principal component analysis
PFS	Progression free survival

PI3K-AKT	PhosphoInositide 3-Kinase/Akt
PRP	Propenone
Rg	Radius of Gyration
RMSD	root mean square deviation
RMSF	Root Mean Square Fluctuation
SDG	Sustainable Developments Goals
SMKIs	Small Molecule Kinase Inhibitors
TK	Tyrosine Kinase
TKIs	Tyrosine Kinase inhibitors
TRUBA	Turkish National e-Science e-Infrastructure

Chapter 1

Introduction

A drug refers to a product that contains active and/or inactivate molecules, which is targeted to treat, prevent or alleviate the symptoms of disease [1]. Contextually, ideal drug should have specific properties to have intended effect without causing unintended consequences. Some of the crucial properties can be listed as follows: having specific action while being non-toxic and with little off target effects as possible; being soluble in water not to face problems of precipitation while in blood flow; also, being soluble in lipids as well to bypass lipid bilayer membranes [2], [3]. Drugs exert their effect of treating, preventing and easing the symptoms of disease by interacting specific targets in an organism, which is human body normally. In order to have these effects on the human body, society applies the various discovery and development methods on and to find the possible drug candidates.

1.1 The Drug Discovery and Development

The process drug discovery and development can be divided into two major stages: discovery and development [4]. Furthermore, drug discovery could also be divided into three substages as well: target discovery, lead discovery, and lead optimization [4]. Target selection part refers to, as the name suggest, the selecting target that is associated with the disease state for to be used in singling out a compound that has potential to be clinically used [4]. While in lead discovery stage large number of compounds are biologically tested against the target to eventually identify a structure motif that have intended effect on the target. Finally in the lead optimization stage, analogues of previously identified motif are designed and biologically tested against the target to improve the efficiency of the drug, lessen the off-target effects [4]. Following the lead optimization substage of drug discovery, the development stage begins with the selection of a candidate compound [4]. In this stage, the compound undergoes pharmacokinetic and pharmacodynamic testing in animals to demonstrate efficacy and acceptable toxicity levels. If successful, it proceeds

to human pharmacokinetic and pharmacodynamic studies, known as clinical trials, with the aim of obtaining regulatory approval for market authorization. [4]. Despite appearing simple and risk-free when summarized, drug discovery and development faces a high attrition rate, with about 70% of projects being cancelled for a variety of reasons [5]. The main factor that leads to cancellation is the lack of efficiency against in-market drug for the targeted disease such as even though animal studies were promisingly effective comparatively, clinical trials could result in absent or negligible intended effects relative to in market drug [6]. Historically, pharmacokinetic limitations, including poor bioavailability, toxic metabolites, and unsuitable half-lives, accounted for up to 39% of project cancellations. With the advent of advanced ADME studies and in silico prediction tools, this rate has now declined to about 1% [5], [7]. Other factors could be listed as animal toxicity, off-target reactions, commercial issues, etc [5].

There are various number of methods that could be applied in each stage of the drug discovery and development. Hence, the following section focuses on the methods of drug discovery.

1.1.1 Drug Discovery Methods

Among the many approaches to drug discovery, the earliest was serendipity, where therapeutic compounds were identified either by chance observations or through rudimentary trial-and-error strategies. There are many examples of these throughout the history of pharmacology [8]. In the anticancer drug discovery scene, serendipity and coincidence have played and still playing huge role in it. A substantial number of widely used anticancer agents were initially developed for unrelated conditions. For instance, artemisinin was introduced as an antimalarial, aspirin as an antirheumatic, etoposide as an emetic, metformin as an antidiabetic, and vinblastine as an antimicrobial. [9].

Another angle that has been used for drug discovery, is the chemically altering the existing drugs or natural products. For example, aspirin that still being used was discovered by the chemical modification of a natural product, physotigmine was modified to improve chemical stability in vivo and as a result modified versions have lower adverse effects while improving intended effect comparatively [10], [11].

Along the same vein of approaching drug discovery, screening of databases, containing synthetic or natural products, has been proven effective. Screening seeks to detect interactions between biological targets and molecules, either synthetic or natural, through biochemical assays. High-throughput screening (HTS) applies this principle to

very large compound libraries. Initially, HTS tested only a few hundred molecules per week. Today, it can evaluate more than 100,000 molecules per day [12]. Additionally, virtual screening has been developed as completing tool to HTS as an in silico HTS and involves virtually placing (docking) collection of molecules, usually from databases against biological target followed by assessing the interaction strength [13]. Virtual Screening significantly reduces the number of compounds that need to be tested for lead identification, thereby saving time and cost. However, there are still requirements for the virtual screening, main requirements can be listed as follows: synthetic/natural product database; structure of target; suitable scoring (docking) scheme [13]. In this case frequently used databases of small molecules can be listed as follows: Zinc (<https://zinc.docking.org/>, accessed on 11.06.2025); ChEMBL (<https://www.ebi.ac.uk/chembl/> accessed on 11.06.2025); PubChem (<https://pubchem.ncbi.nlm.nih.gov/> accessed on 11.06.2025)

The state-of-the-art method for discovery of drugs, is rational drug design that consists of approaching the discovery with a design and iteratively modifying the compound to enhance or lower the targeted aspects, such as binding strength towards target, selectivity, ADME properties, etc. [14]. However several prerequisites are required for this approach: targetable biomolecule should exists with significant ties to target disease/condition; 3D representation of the target should exists either from crystallography or AlphaFold system, which should be considered to be used with a grain of salt; a method to asses virtual interactions between the compound and target structure must be available [14].

Rational drug design and virtual screening can be included under big umbrella term of computer-aided drug design, which will be delved into more detail in the following section.

1.2 Computer-Aided Drug Design Approaches

Thanks to advancements in the computer-aided drug design approaches (CADD) cost of research and development of a new drug from scratch to market has been reduced by altering time and money needed for specific stages such as target identification, validation, lead discovery, lead optimization and pre-clinical test [15]. CADD had this effect on the drug research and development as a result of the development of variety of methods. A few examples can be listed as follows: quantum mechanics [16]; molecular docking [17]; molecular dynamics simulations [18]; homologous modelling [19]; multiple sequence alignment [20]; quantitative structure-activity relationship [21]; and so on. Computational drug design utilizes two principal strategies. The first, structure-based drug design, includes virtual HTS and rational design guided by target structures. The second, ligand-based drug design, focuses on known compounds and applies pharmacophore modelling combined with QSAR for systematic improvement of drug candidates [15].

Due to scope of thesis focus on applying the structure-based drug design method for targeting anaplastic lymphoma kinase's tyrosine kinase domain, structure-based drug design and its tools will be explained more in-depth.

1.2.1 Structure-Based Drug Design

As previously explained, in order to apply rationality to drug design, specifically to utilize the structure-based method, it is crucial that fully comprehend 3D structures of target molecules, mostly proteins, enzymes, or basically biomolecules. Achieving this goal can be approached in two ways. The first method uses molecules that have already been isolated and structurally characterized through experimental techniques. These molecules are catalogued in databases, most notably the Protein Data Bank (PDB, <https://www.rcsb.org/>, accessed at 11.06.2025). As of it now, PDB holds more than 237 thousand experimental PDB structures. Second method is using prediction model of the target, if no experimentally available structure of desired target exists. Orthodox method of doing this process was homology modelling using targets sequence [19] and followed by molecular dynamics simulations [18]. With recent developments in protein structure predictions by AI there are 999 thousand structure predictions in the AlphaFold, which is AI based protein prediction tool [22].

Structure-based drug design involves three primary steps. First, ligands are either designed according to the target's structure and characteristics or selected from databases; these may include synthetic molecules, natural-product analogues, or natural products. Second, molecular simulations, including docking and molecular dynamics, are performed to evaluate ligand-target interactions and stability. Third, the predicted interactions are assessed by binding energy calculations, ligand properties, and subsequent in vitro biological assays to validate the top candidate ligands [15].

To elucidate the importance of molecular docking in both structure-based drug design and CADD, few examples of research achievements utilizing it with its basic working principles and names of tools will be given in the following paragraphs.

Molecular docking tools predict the binding pose and energy of small-molecule ligands at the molecular level. These predictions are performed using specific algorithms and scoring functions. Each software may employ different algorithms and scoring functions; for example, AutoDock uses a genetic algorithm and an environmental matching score, which summarizes the energies of hydrogen bonding, hydrophobic interactions, steric hindrance, repulsion, electrostatic interactions, and internal ligand torsional constraints [23]. More examples include: "Dock" with fragment growth algorithm combined with force field, surface matching and environmental matching scoring function [24]; "GOLD" with genetic algorithm with empirical scoring function [25]; "Z-Dock" with geometric matching and molecular dynamics algorithm with CAPRI+ scoring function [26]; and various other algorithms and scoring functions.

Although molecular docking plays a critical role in CADD, solvent effects are not considered in the identification of binding poses or in energy calculations. When this limitation is combined with the significant role of solvent in drug-target interactions, additional methods are required to evaluate solvent effects on ligand binding. Molecular dynamics (MD) simulations are commonly applied for this purpose. MD simulation method utilizes Newton's laws of motion on systems' atoms that their parameters are defined with force-fields, which is mathematical model characterizes the forces between specific atoms that is verified quantum calculations and experimental data [27], [28]. By applying Newton's law of motions over certain time period, behaviour of the ligand-protein complex, binding poses' stability in a solvent environment can be assessed [28]. MD simulation usually used in two aspects of molecular dynamics, in pre-stages of molecular docking to ensure protein stability and binding pose improvement, and in post-molecular docking to evaluate binding energy calculation across a certain time interval

and stability of the initial binding pose over the same time as well [27]. There is several software for these purposes such as Amber [29], GROMACS [30], NAMD [31], Charmm [32], etc. that specialized in the biological and chemical systems while Lammmps[33] that specialized in material and solid-state physical systems.

Structure-based drug design method has been utilized in the development of drugs for various disease areas. For example, it has been used in developing anti-cancer drugs [34], in anti-virus [35], anti-bacterial [36], etc. Specifically, in oncology, CADD and structure-based drug design have been extensively employed, resulting in number of drugs on the market that target specific kinases. (See Figure 1.1.)

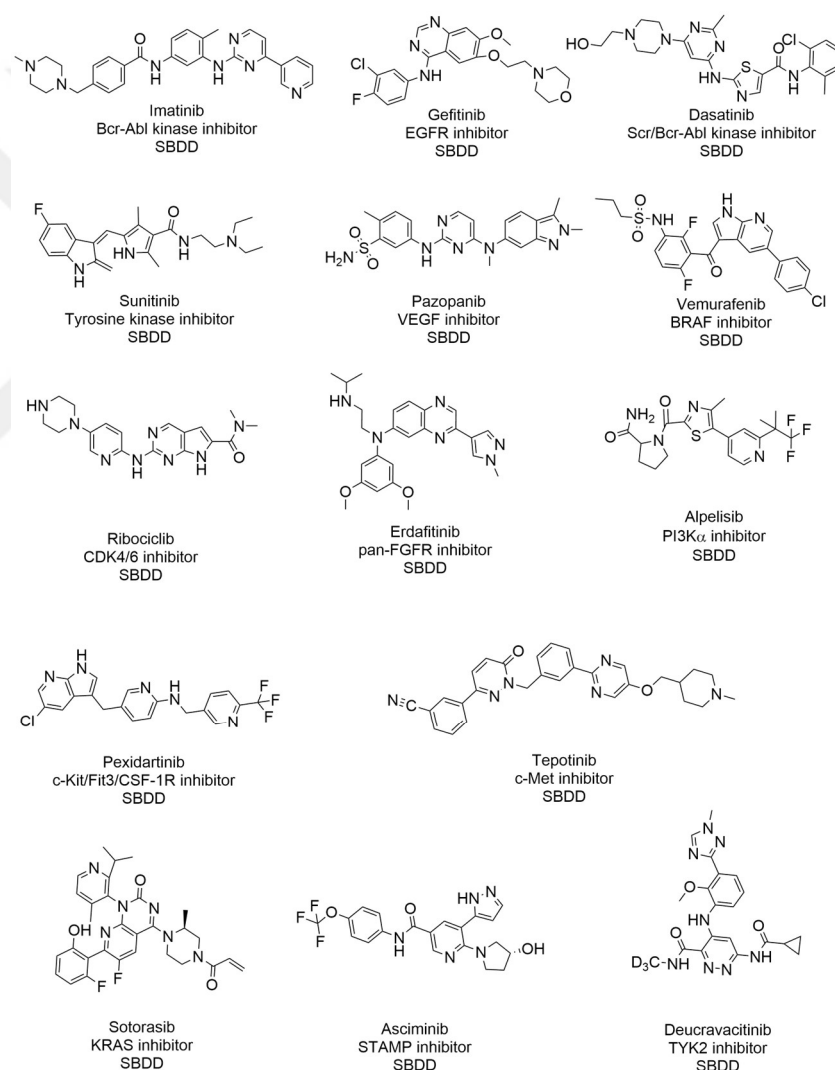


Figure 1.1 Small molecule kinase inhibitors discovered using CADD [15].

The success of structure-based drug design in small molecule kinase inhibitors demonstrates the important role of protein kinases as drug targets. Hence, the next section

will delve into the significance of protein kinases in drug discovery, and especially regarding this thesis's target anaplastic lymphoma kinase.

1.3 Drug Discovery in Protein Kinases

Protein phosphorylation is one of the most common post-translational modifications [37]. This phosphorylation controlled by kinases, which adds phosphate (PO_4) to the side chains of various amino acids, and phosphatases which removes the phosphate, has major role in regulation of cellular processes for example cell proliferation, subcellular translocation, metabolism, etc. [37], [38]. Since 1980s, protein kinases have been identified as potential targets for various diseases, and in 1995 Fasudil has been approved for inhibiting Rho-associated coiled-coil-containing protein kinases 1 and 2 for treatment of cerebral vasospasm [38]. Since its approval number of approved kinase inhibitors has dramatically increased to 98 and 71 of them are small-molecule kinase inhibitors (SMKIs)[38](See Figure 1.2.). Additionally, these 71 SMKIs target 21 distinct kinase families belong in five of the eight major protein kinase groups (tyrosine kinases, tyrosine kinase like, homologues sterile 7, cyclin-dependent kinases, and protein kinase A,G and C) with 49 of approved drugs target 30 proteins that belong in the tyrosine kinase family [38], [39], [40]. The 19 of the approved SMKIs have been for non-small-cell lung cancer (NSCLC) which could be demonstrated meaningful when considered the fact that lung cancer being front runner in the cancer-related deaths with 80-85% of them belonging in NSCLC [38], [41], [42].

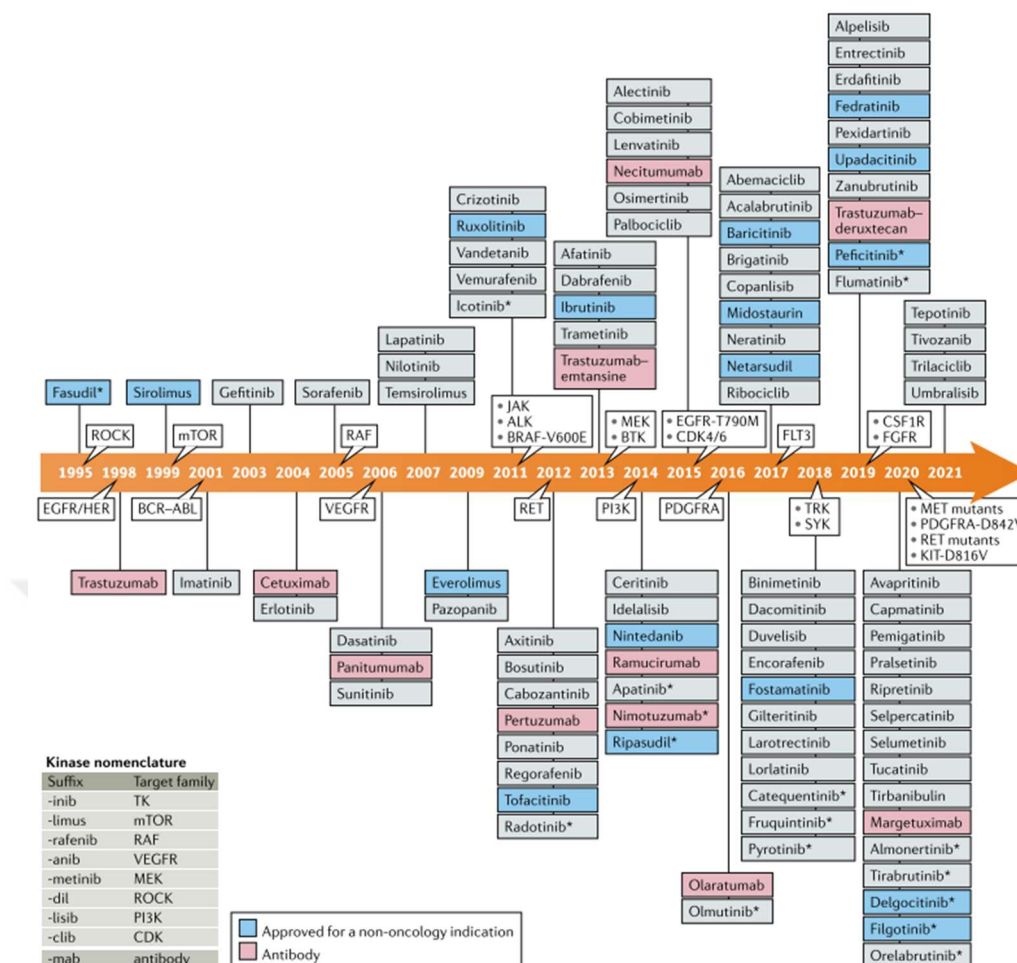


Figure 1.2 Timeline of approved kinase inhibitors. Drugs approved for non-oncology indicated with blue coloured boxes, while nomenclature of kinase inhibitors and monoclonal antibodies were provided in the legend [38]

Instead of trying to cover each kinase that has been targeted and could have been targeted, only the main target of the thesis is focused extensively, which the details could be found the next section.

1.3.1 Anaplastic Lymphoma Kinase (ALK)

Anaplastic lymphoma kinase (ALK) was first described in 1994 in an anaplastic large cell lymphoma cell line as a nucleophosmin (NPM) fusion partner [43]. Since its discovery, ALK and its various fusion partners have been implicated in numerous cancers such as NSCLC, anaplastic large cell lymphoma (ALCL), neuroblastoma, rhabdomyosarcoma and many more [44], [45], [46]. In addition, physiological ALK is activated like other receptor tyrosine kinases. ALK activation occurs when the endogenous ligand of ALK binds the extracellular region of the protein, and thus,

dimerization and autophosphorylation eventually results in the activation of downstream signalling pathways such as phosphoinositide 3-kinase/Akt (PI3K-AKT), mitogen-activated protein kinase (MAPK), and Janus kinase/signal transducer and activator of transcription (JAK-STAT) which all could be defined substantial in cell proliferation, survival, and differentiation [47], [48].

The physiological ALK is a 220 kDa (with its N-linked glycosylation) transmembrane protein, which belongs to the insulin receptor protein-tyrosine kinase superfamily, that has 3 sections: N-terminal extracellular region, which consists of two MAM segments (meprin, A5 protein, and receptor protein tyrosine phosphatase μ), one LDLa segment (low-density lipoprotein class A) and a glycine-rich segment; juxtamembrane region; an intracellular region that houses a kinase domain, which has two domains labelled as small amino-terminal lobe and large carboxy-terminal lobe [49]. Those two lobes have additional structures that further facilitate the Adenosine triphosphate (ATP) binding and one example of this is the hinge region (residues 1197-1201, numbering according to human physiological ALK), which directs the ATP to the binding pocket, while other structures effect the kinase's flexibility as well as overall structure such as changing the overall conformation between active and inactive forms [50]. Furthermore, most of the approved kinase inhibitors competitively bind the hinge region amino acids via mostly hydrogen bonding, hence successfully imitating ATP binding [51]. In the small amino-terminal lobe, there is a conserved salt bridge between residues K1150 and E1167 and without this salt bridge, it has been proven that the kinase is rendered functionally inactive [52]. Roskoski (2013) highlights that on the large carboxy-terminal lobe, there are conserved amino acid residues that form the catalytic core of the ALK. One of these regions are catalytic K/D/D (1247-1249) with catalytic loop lysine (1267). Roskoski (2013) describes another region of the large carboxyterminal lobe that is crucial for the ALK kinase domain's activity as the activation segments, which in a nutshell its function can be summarized as guiding the phosphates of ATP to their spot for catalytic activity, these residues are activation K/D/D region (1270-1272),

residues that specify the tyrosine phosphorylation sites (1278, 1282, 1283). (See Figure 1.3.)

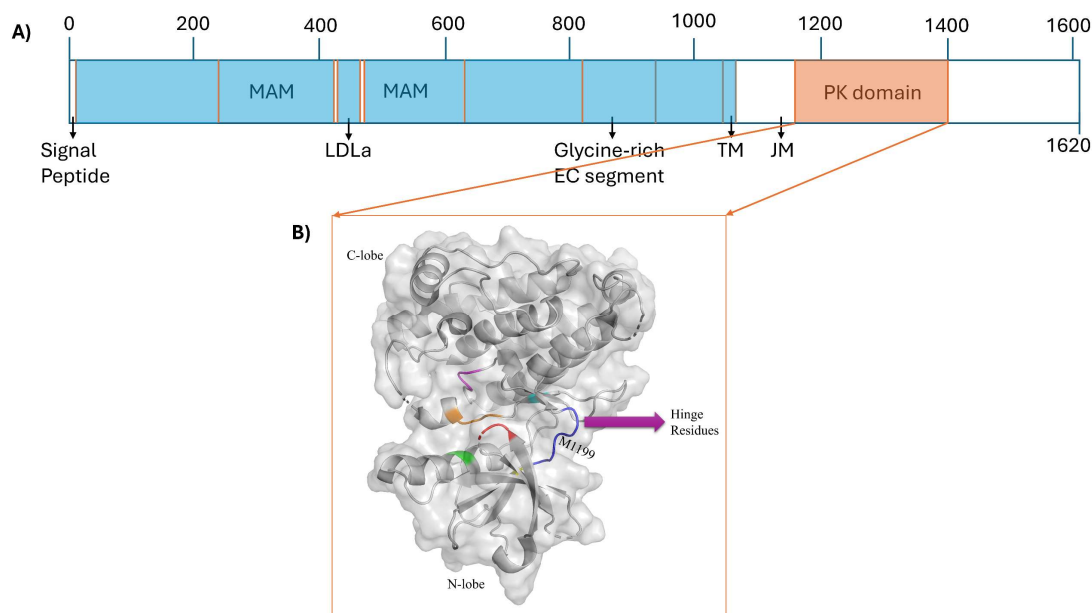


Figure 1.3 A) Representation of general structure of human physiological ALK receptor protein-tyrosine kinase, light-blue section represents extracellular segment (residues 19 -- 1038), orange section represents protein-tyrosine kinase domain (residues 1116 – 1392) B) 3D structure of ALK kinase domain (PDB ID 8ARJ) and the amino acids that are highlighted with different colours signify various regions as follows: red: Glycine-rich loop's start, green: α C-glutamate, blue: Hinge residues, yellow: Gatekeeper Residue, magenta: First Aspartate of K/D/D catalytic signature [49]

Rearrangements that include chromosome 2 that carries the region of the ALK gene result in ALK's tyrosine kinase domain expression where normally would not occur with continuous activation with its efficiency and localization depending on its fusion partner [48]. Furthermore, its fusion partner leads to ligand-independent activation state of ALK protein kinase domain, which is the dimerized state [49]. Scheider and colleagues also remark on the fact that some of the ALK-positive cancers display ALK dependency such as lung cancer and hence they are vulnerable to ALK inhibitors. It should be also noted that ALK fusions, in which Echinoderm microtubule-associated protein-like 4 (EML4)-ALK is the most common fusion in NSCLC, can be reported as being infrequent at best, with an occurrence rate of around 2-7% of NSCLC. However, for ALCL, around 90% of pediatric patients and around 50% of adult patients are ALK-positive [53], [54]. More detail in regarding ALK's role in various diseases is introduced at following portion.

1.3.2 ALK in disease

Since its discovery, ALK and its various fusion partners have been implicated in numerous cancers such as NSCLC, ALCL, neuroblastoma, rhabdomyosarcoma and many more [44], [45], [46]. Two of more common and targeted *ALK*⁺ cancers is introduced below.

Anaplastic Large Cell Lymphoma (ALCLs)

ALCLs are characterized by their proliferation of large lymphoid cells and their typical invasion into lymph nodes sinuses while also presenting high levels CD30 antigen [55]. Additionally, ALCL are further categorised into *ALK*⁺ and *ALK*⁻ based on dependency on ALK activated signaling pathways, which caused by the chromosome 2 abnormality that result in the constantly activating protein kinase domain of the ALK [49], [55]. It also important to realize that *ALK*⁺ was reported to be comprise 75-80% of all ALCLs cases [56]. Most commonly used method for treatment procedure for *ALK*⁺ALCL is the abbreviated as CHOP, which include cyclophosphamide, doxorubicin, hydrochloride, vincristine sulphate, and prednisone, with this regimen overall survival rate after five years was found to be 70% while progression-free survival rate at the same interval, was found to be 60% [55].

Non-small cell lung cancer (NSCLC)

As stated earlier, lung cancer is the number one cause of cancer-related deaths with five-year survival rate being less than 20%, while 80-85% of lung cancer cases categorised under NSCLC [41], [42]. About 5% of NSCLCs contain fusion protein that results in intracellular region of ALK with fused with *EML4*. According to 2021 article fusion products containing various exons of *EML4*, 16 different variants has been found in *EML4-ALK*⁺ NSLCL [57]. Even though it has been found that there is different fusion product that contain different fusion partner than the *EML4*, *EML4-ALK* comprise 85% of *ALK*⁺NSCLC fusion products [57]. Even with existence of targeted chemotherapy, such as a regimen that depends on the driver mutation and dependency, with the fact of five-year survival rate being low as reported, proves the need for developing alternative therapies that can results better prognosis. With only 5% of the NSCLC shows ALK dependency while ALCL shows around 75% in mind, one could assume that it is naturally to focus on the ALCL for targeting the ALK, but due to large numbers of cases of NSCLC and its mortality rate most of the research and approval for ALK targeted kinase inhibitors focuses on the NSCLC.

1.3.3 ALK inhibitors

Thanks to developments on clarifying the importance of ALK in the aforementioned cancer types, number of generations ALK tyrosine-kinase inhibitors have been developed for the betterment of mentioned diseases. (See Figure 1.4.) The first inhibitor to be approved by FDA was Crizotinib, now categorised as first generation ALK tyrosine kinase inhibitor, that demonstrated landslide superiority over first-line cytotoxic treatment in clinical trials phase 1,2, and 3[58]. However, Heigener and Reck also highlighted the fact after Crizotinib treatment that the median of progression free survival (PFS) was discovered to be around 8-11 months, mostly can be tied to ALK tyrosine kinase mutations that end up causing resistance.

Consequently, this prompted new generation of drug candidates titled as second generation, which 3 of them ended up approved by the FDA (Ceritinib, Alectinib, Brigatinib), and one of them approved by the China (Ensartinib) [48]. The second generation of ALK tyrosine kinase inhibitors (TKIs) proved to be overcome resistance to Crizotinib while also presenting better inhibitory effectiveness even in non-resistance *ALK*+NSCLC [59]. Within the second generation ALK tyrosine kinase inhibitors, Alectinib were verified as being significantly better than Crizotinib in phase 3 clinical trial with its PFS discovered to 34.8 while Crizotinib's was 10.9 [60] However, unique to second generation resistance mutations to Alectinib also has been reported [59], [60].

As a direct consequence of resistances, and third generation ALK TKI has been developed and approved by the FDA, called as Lorlatinib, which could be reported as one example of macrocyclization by first generation ALK TKI [61]. Followingly, the phase 3 clinical trial comparing Lorlatinib and Crizotinib as first in line treatment for ALK positive NSCLC featured that Lorlatinib has longer PFS than Crizotinib [62]. In line with those result, Lorlatinib become first-line treatment for untreated *ALK*+NSCLC in 2021 [63]. However, Lorlatinib specific mutations to TKI has been recorded, which highlights the need for discovering and developing new candidates for ALK TKIs [38].

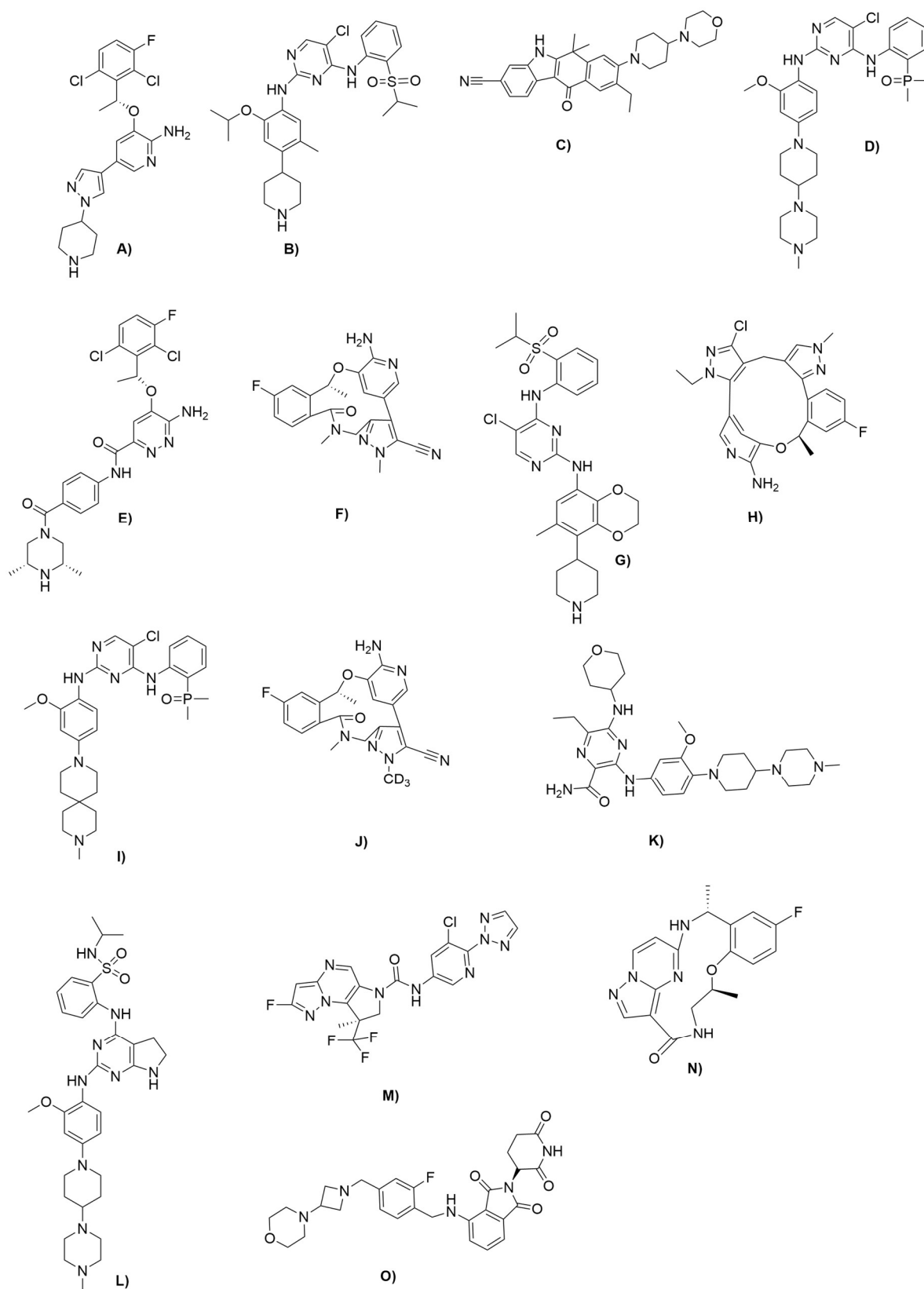


Figure 1.4 Structures of ALK tyrosine kinase inhibitors approved by the FDA and in various stages of clinical trial in last 5 years. (A-F: Approved by FDA, G-O: various stages of clinical trial.) A) Crizotinib, B) Ceritinib, C) Alectinib, D) Brigatinib, E) Ensartinib, F) Lorlatinib, G) XZP-3621, H) NVL-655, I) WX-0593, J) TGRX-326, K) Gilteritinib, L) CT-707, M) SGR-1505, N) TPX-0005, O) BMS-986369

This thesis is not only one that noticed the need to discover and develop drug that selectively target ALK positive NSCLC. There are more than several drugs that are in different stages of drug discovery and development for ALK dependent cancer types. This section will provide examples of drug candidates that selectively target ALK from the clinical stages of the development to in vitro testing stages of drug discovery in the last 5 years.

There are more than 4 drug candidates that are in the phase 3 of the clinical trials. (NVL-655, WX-0593, XZP-3621, TGRX-326) As of the time of access (16/06/2025), one of them titled as NVL-655 was actively recruiting participants [64], while two of them, WX-0593, TGRX-326, were on-going with 292 and 321 participants already enrolled, respectively [65], [66]. Lastly, in XZP-3621's phase 3 trial, was reported as not yet recruiting with their 238 planned enrolment estimates [67]. (See Figure 1.4. G-J)

At the same time, there are more than 5 drug candidates that were either in phase I/II or just phase I clinical trials. Two of them were phase I and II combined, which the candidates are titled as TPX-005 and BMS-986369 with both were actively recruiting [68], [69], [70]. Within the rest of the phase I and phase II drug candidates, all were in the phase I stage, with two of them, Gilteritinib and SGR-1505, were actively recruiting at time of access [71], [72]. Lastly, CT-707's phase I trial was last updated in 2017, which was in recruitment stage as well, but still there is no update on it until now [73], [74]. (See Figure 1.4. K-O)

Seeing that there are number of ongoing clinical trials for ALK signaling dependent diseases, it is not surprising that there are also studies that are in the earlier stages of drug discovery and development. Following will provide overview of the ALK targeting compounds that are validated to inhibit ALK in the last 5 years either in vivo or in vitro. (See figure 1.5.& Figure 1.6) In the last 5 years, there are number of compounds that have been discovered and validated to the target ALK in vivo, 10 compounds to be exact that target solely ALK. (Figure 1.5) Within 10 compounds, 3 of them, labelled in their original article as 4B, Compound 17, Compound B3, was a proteolysis targeting chimera (PROTAC) instead traditional inhibitor, which basically instead of inhibiting the ALK, PROTACs initiate selective proteolysis of the ALK [75], [76], [77]. Rest of the 7 compounds could be reported as traditional inhibitors, labelled in their original article as Compound 3b, XMU-MP-5, ZX-29, DA-0157, Ld-10, Compound 9, Compound 10k [78], [79], [80], [81], [82], [83], [84].

In the same time interval, there also total of 5 distinct compounds that have been designed and validated to inhibit solely ALK in vitro test, labelled in their original article as O-10, H-80, Compound 19, Compound 11k, Compound 149 [85], [86], [87], [88], [89]. (See Figure 1.6)

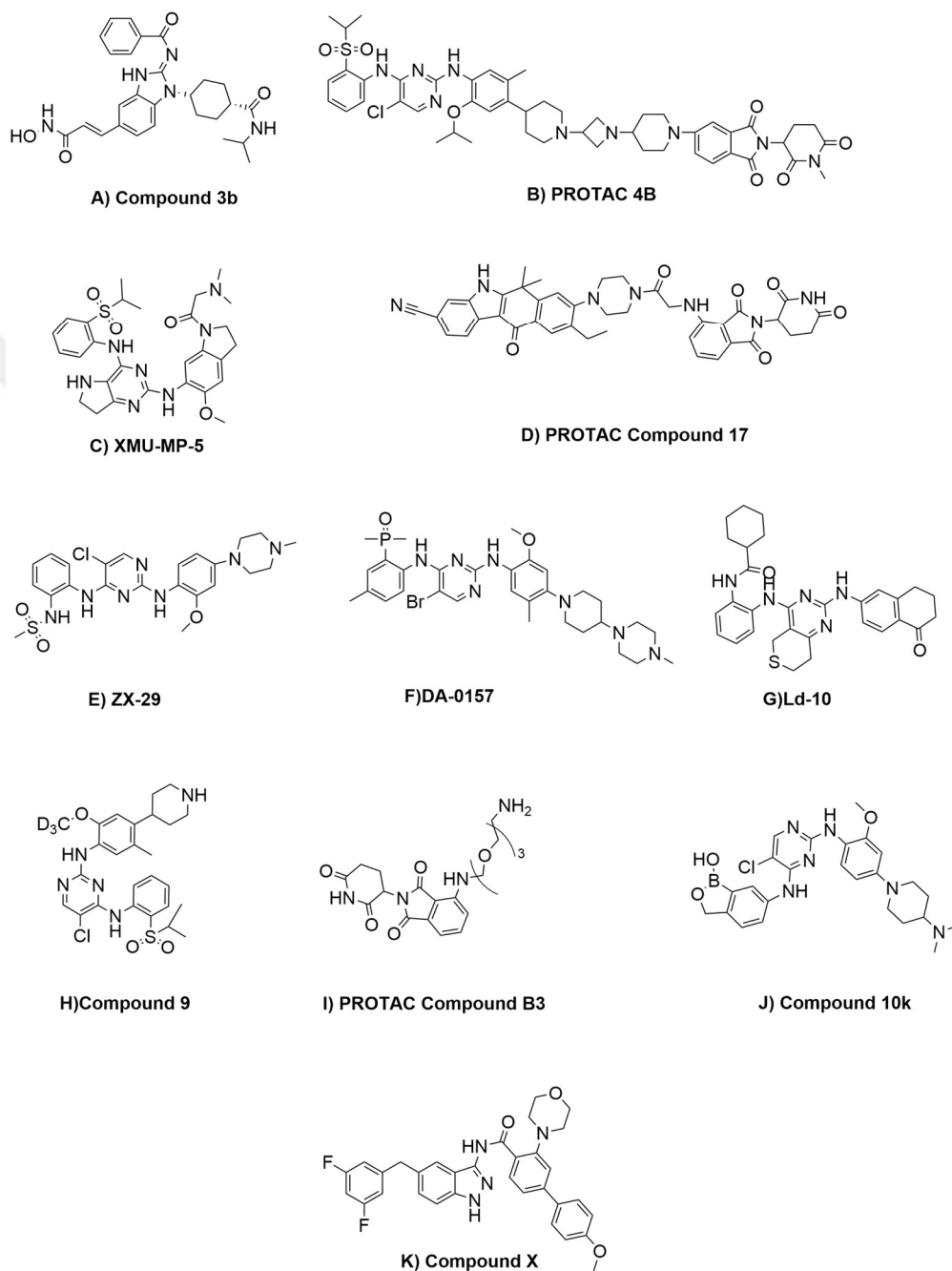


Figure 1.5 Structures of ALK tyrosine kinase inhibitors that has been validated in vivo test in last 5 years.

Additionally, there also considerable number of drug candidates, that focuses on the being dual inhibitor that have been validated in the either or both in vivo and in vitro, but due to dual inhibitors being out of scope of this thesis, not all dual inhibitors were reported ALK inhibitors. Instead, 1 compound that has been reported, which was labelled as compound X4 in the original article, and it had been validated inhibit both ROS1 and ALK both in vitro and in vivo evaluations [90].

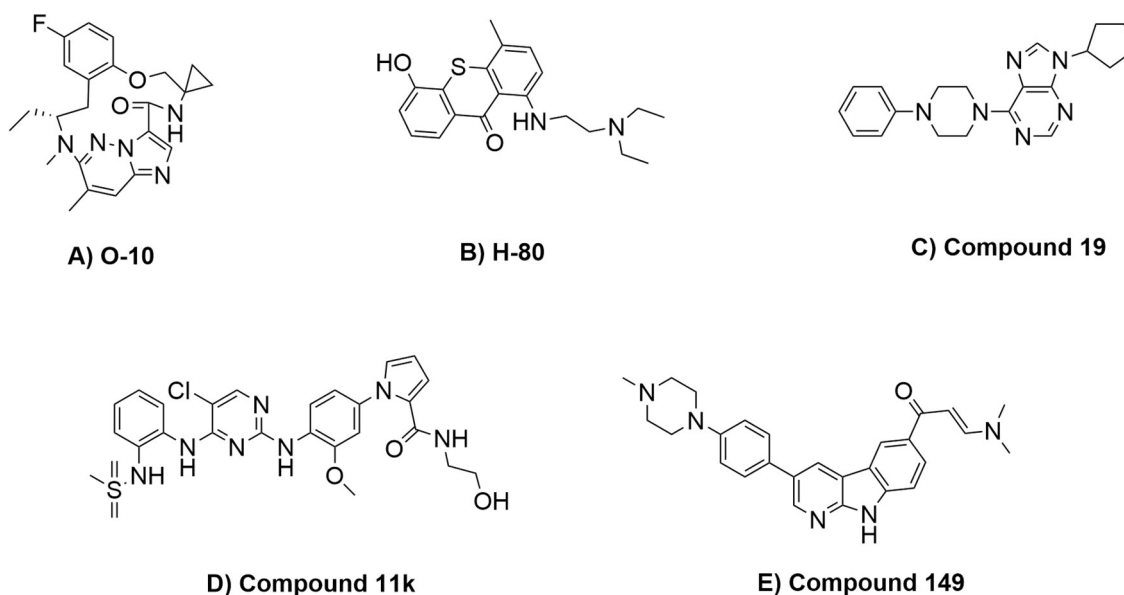


Figure 1.6 Structure of ALK tyrosine kinase inhibitors that has been validated only in vitro test so far in last 5 years.

1.3.4 Objectives & Scope of the Thesis

Given these developments in the ALK targeting drug candidates, this thesis focused on two distinct avenues to discover and develop new and novel drug compounds and motifs that could target tyrosine kinase domain of the ALK. First avenue selected carboline motif in the Figure1.6.'s structure labelled with E and as compound 149 in the original article, tried to significantly improve the binding affinity towards to ALK tyrosine kinase [89]. Carbolines, a carbazole derivative, have gained attraction in the last two decades in various areas such as anti-tumour, anti-Alzheimer's disease, and neuroprotection functions [91]. The various activities against different disorders might be due to carboline's structural diversity. They are named by their position of nitrogen atom relative to indole ring as α -, β -, γ -, or δ - carbolines. Specifically, α -carboline, which has the pyridine ring that is fused with an indole ring, attracted relatively more interest in

recent years due to having varied biological activities such as anti-tumour activity against leukaemia, colorectal cancer, glioblastoma, anti-neuroinflammatory effect for neuroprotection [92], [93], [94], [95]. Concurrently, there was relatively recent two studies focused on the ALK and carboline, which is published by the same first author, that proved by in-vitro experiments, specifically by utilizing enzyme-linked immunosorbent assay based kinase assay, α -carboline derivatives has significant potential to inhibit ALK kinase [89], [96].

As first path, this study focused on rationally designing a novel α -carboline compound for ALK tyrosine kinase to achieve distinct activity profiles against various cancer types in which ALK fusion proteins may be involved. In this regard, a series of α -carboline derivatives is designed and their binding properties against ALK tyrosine kinase is investigated by using in silico protocols consisting of molecular docking studies, MD simulations, and molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) binding free energy calculation. As a second path, this study focused on screening a large natural product database to find novel drug candidate to inhibit ALK tyrosine kinase domain.

Chapter 2

Experimental Procedure

2.1 Design and Evaluation of α -Carboline Derivatives for ALK Tyrosine Kinase Domain Using Comprehensive in Silico Methods

2.1.1 Target Protein and Ligand Preparation

For the target protein, Anaplastic Lymphoma Kinase with a carboline inhibitor with 1.65 angstrom ($\text{\AA}$) resolution was acquired from the protein databank (PDB ID: 8ARJ). Inhibitor location was used as reference for coordinates in following docking procedures. For adding the missing side chains, SwissPDB-Viewer has been utilized, while UCSF Chimera has been used for deleting heteroatoms [97], [98]. Consequently, missing hydrogen atoms and charges according to AMBER ff14SB were added by using of UCSF Chimera's DockPrep, [99].

For ligand preparation, molecules were drawn using Marvin from ChemAxon, Marvin 23.04.0, 2023 (<http://www.chemaxon.com/>), minimized by utilizing MMFF94 force, 3D coordinates generated, and converted to ready-to-dock format by using OpenBabel [100], [101]. The ligand topology and parameters' files of designed compounds for use in the MD simulations were generated using SwissParam web service [102].

2.1.2 Molecular Docking

Molecular docking was done to evaluate initial binding affinities of the designed compounds against the target section of the protein and to generate initial coordinates that will be required in MD simulation by use of AutoDock Vina[103], [104]. The energy

calculation of receptor-ligand was based on focused docking with coordinates of grid's centre XYZ: 33.45, 52.39, 11.72, dimensions XYZ: 24.60, 23.83, 27.80. A maximum of ten docked conformation and exhaustiveness of 8 with the rest of the parameters as default from AutoDock Vina per compound were utilized. The scoring function of AutoDock Vina can be highlighted as a summary of energies of hydrogen bonding, hydrophobic interactions, steric hindrance, repulsion, electrostatic interactions, and internal ligand torsional constraints. The docked poses for each compound were selected based on the binding affinity and from each category of the designed compound. Interactions between docked ligand and receptor were visualized by BIOVIA, Dassault Systèmes, Discovery Studio Visualizer, v21.1.0.20298, San Diego: Dassault Systèmes, 2021 and PyMOL, The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC. Furthermore, one of them from each category of designed compound according to their residue interactions and as well as binding affinity was selected for MD analysis.

2.1.3 MD Simulation

Molecular dynamics simulations were conducted to assess the compounds' stability in the active site of the target protein over simulation time using GROMACS version 2023.1 [105], [106]. Additionally, only 8ARJ without any compound with same parameters as follows also conducted to have valid comparison analysis in later stages. The topology, which defines the particle types, bond types, and force field parameters, and the coordinates file, which specifies the 3D positions of all atoms within the system dimensions, were generated using the CHARMM27 force field embedded in GROMACS. [106], [107]. The protein-ligand complexes were confined centric manner in a triclinic simulation box that contains SPCE water model containing around 11000 water molecules with appropriate NaCl ions at 0.1M to neutralize the system. Consequently, the box generations were trailed with energy minimization steep adopting steepest descent minimization with 50000 maximum step number. Afterward, 100ps NVT (fixed number of particles, volume, and temperature) with a modified Berendsen thermostat, V-rescale, at 300K ensemble followed [108]. Subsequently to the NVT ensemble, for NPT (fixed number of particles, pressure, and temperature) Berendsen pressure coupling has been used [109]. The processed systems were then subjected to a 100 ns production MD simulation, except for the final compound iteration, for which the simulation was extended to 300 ns. A time step of 2 fs was used, and energies and coordinates were saved every 10.0ps in the trajectory. Root Mean Square deviation (RMSD) plots were drawn

utilizing the MDAnalysis package [110], [111], [112], [113]. Average distance heatmaps were drawn utilizing Gromacs' distance function combined with the Python package seaborn [114], [115].

2.1.4 MM-PBSA Calculations

To clearly assess the binding of the selected compounds with the target section of the protein, MM-PBSA calculations were done using the `g_mmpbsa` package of GROMACS [116]. Binding averages of the whole trajectory have been utilized for each compound at every 1000 ps, resulting in 100 frames in total or 300 frames for the extended MD simulations.

2.1.5 Comparative Essential Dynamics and Clustering

To further gain inferences about the protein stability and as well as conformational information in our MD analysis, some of essential dynamics methods were applied to 8ARJ-compound complexes and as well as only 8ARJ (without a compound). Such as clustering analysis utilising ward clustering method with kmeans with elbow algorithm for discovering optimal number of cluster by using TTClust [117]. PCA on binding pocket, which was selected using previously stated information, using MDAnalysis package[50,51], and lastly RG-RMSD based free energy landscape analysis (FEL) inspired from a 2024 study [118].

2.2 In silico Discovery of Novel ALK inhibitors from a Natural Product Database

2.2.1 Pre-Docking Procedures - Protein Structure Equilibration

To ensure target structure is adequately calibrated for use in the large-scale docking procedure that was planned for this section, it was decided that more equilibration was needed. For this purpose, the same ALK tyrosine kinase domain that initial preparation has been done from the same PDB, 8ARJ, was subjected to 500ns MD simulation with the same parameters as reported in 2.1.3. The resulting trajectory were corrected for periodic boundary conditions and subjected to clustering analysis by utilizing TTClust python program [117] with “ward” clustering method with kmeans clustering with elbow

algorithm for discovering optimal number of clusters. After analysing the clusters of various possible conformations (See Figure 2.1 A-C), two prominent conformations was seen, cluster 5, and 1. To evaluate conformations structure, PDB that represents each cluster were selected from trajectory using TtClust. These two conformations were aligned to original structure of 8ARJ. (See Figure 2.1 D & E)

In the cluster 1's alignment average RMSD was found to be 1.683Å, while cluster 5's was 2.019Å. More importantly, the active site in the conformation of cluster 5 was found to be sterically hindered. In Figure 2.1B, the orange circle indicates the region responsible for this steric hindrance of the active site. it. From this, cluster 1's representative PDB was chosen to be continued with docking procedure the same procedure that has been done in the 2.1.1. for target structure preparation has been also utilized for the cluster 1.

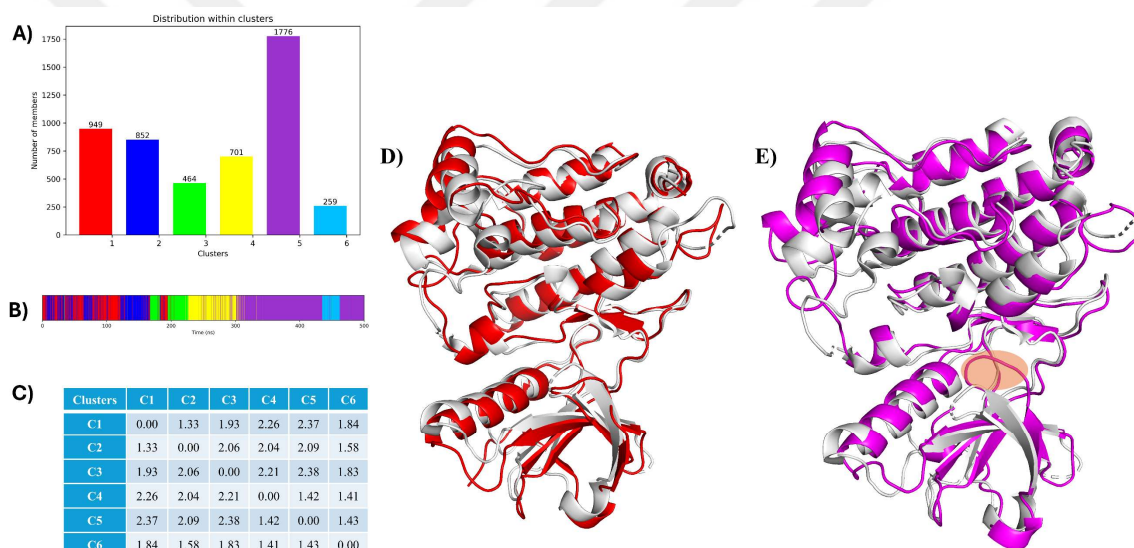


Figure 2.1 Clustering analysis of 8ARJ equilibrium MD and two prominent clusters' alignment to 8ARJ. White protein structures represents 8ARJ while red represents cluster 1 and magenta represents cluster 5. Orange circle highlights the “closed” structure of cluster 5 compared to original. A) Histogram plot of number of members across all clusters. B) A linear projection of clusters across the trajectory. C) Root mean square deviation (RMSD), (Å) matrix between clusters. D) Cluster 1-8ARJ overlap. E) Cluster 5-8ARJ overlap.

2.2.2 Ligand Acquisition and Preparation

Ligands were obtained from COLleCtion of Open Natural ProdUcTs (COCONUT) database version 2 (published in 2024 November) which had more than 695k distinct molecules [119]. Lipinski's rule of five has been applied to database to filter-out

compounds that has less potential to be drug candidate. Lipinski's rule outlines that most compounds reaching the market share similar properties. These properties include a molecular weight of less than 500 Da; a logP, the logarithm of the octanol-water partition coefficient measuring the drug candidate's affinity for lipophilic versus aqueous environments, below 5; fewer than 5 hydrogen bond donors; and fewer than 10 rotatable bonds [4]. After filtering the COCONUT database, 361732 molecules were shortlisted based on Lipinski's rule of five.

The shortlisted molecules were prepared for docking by converting each of the 361,732 natural products into a unique, ready-to-dock format with 3D coordinates representing the energetically most favorable conformation. This process was performed using OpenBabel version 3.1.1. [100]. Furthermore, for MD simulation of selected natural products, required CHARMM force field parameters for each select natural product were composed using CHARMM General Force Field (CGenFF) tool version 4.6. [120].

As a reference for comparison with natural products, the 3D coordinates of Lorlatinib, a first-line treatment for ALK-dependent NSCLC, were obtained from PubChem (CID 71731823). The same preparation steps applied to the COCONUT molecules were also applied to Lorlatinib.

2.2.3 Molecular Docking

Similar docking procedure was applied to each molecule that has been prepared in the previous segment to judge their binding affinity towards the tyrosine kinase domain of the ALK. Similarly, Smina, a fork of AutoDock Vina version 1.1.2, was used. Its default scoring function, identical to Vina's, was retained for this study. [103], [121]. As docking grid box XYZ coordinates of 30.58, 37.38, and 30.03, and sizes of 25.08, 22.51, and 30.92 has been used, respectively. Additionally, top 10 binding modes has been saved for each ligand as output of the docking procedure with minimum energy range set as 4, exhaustiveness, which determines amount of computational effort that Smina/AutoDock Vina will focus on each molecules with default of 8, setting higher values produces more consistent docking, set at 16, and lastly seed set at 44. For Lorlatinib typical settings was utilised expect for exhaustiveness, which was increased 32 for most consistent docking possible for it.

From ten conformation of each molecule, first pose of each molecule was selected and all COCONUT molecules first poses were ordered from lowest to highest binding energies. After analysis of the results, top 5 of molecules that has hydrogen bond with

hinge residues was selected for followings studies. While 3D interaction figures were rendered using the same as before, 2D figures was rendered utilizing LigPlot+ [122].

2.2.4 MD Simulation

Similar MD procedure to procedure numbered as 2.1.3. was applied, however, some of parameters were changed. Changed parameters can be described by the following sentences. Topology and coordinates files were generated using CHARMM36 force field[123]. TIP3P water model with dodecahedron box was used [124]. While NVT parameters were left the same as before, but in NPT, C-rescale pressure coupling has been used instead of Berendsen pressure coupling [125]. Production MD run parameters were kept same, expect for the length of simulation, which was increased to 500ns. Average distance heatmaps were drawn same way as previously described.

2.2.5 MM-PBSA Calculations

The same MM-PBSA method from procedure 2.1.4 was applied. Specifically, snapshots were extracted every 1 ns from the 500 ns trajectory to calculate the average binding affinity over the entire trajectory.

Chapter 3

Results & Discussion

3.1 Design and Evaluation of α -Carboline Derivatives for ALK Tyrosine Kinase Domain Using Comprehensive In Silico Methods

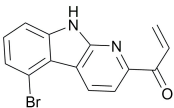
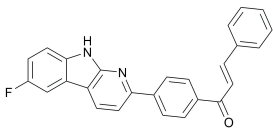
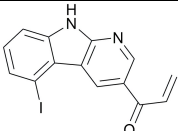
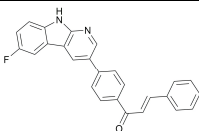
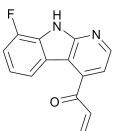
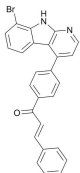
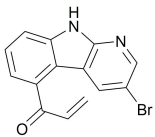
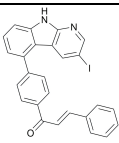
To discover applicable lead compound to be continued as a pharmacophore, an extensive literature review was done. Among those, α -carboline (1-azacarbazole) has been suspected to hold the most potential to be further developed mainly due to its potential to have hydrogen binding due to its nitrogen atoms and the fact that it can imitate the ATP's adenine ring, which can result in binding in the hinge region of the kinase, also helped to see it as an attractive candidate [126].

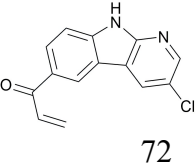
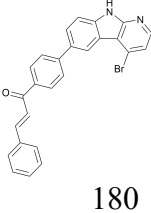
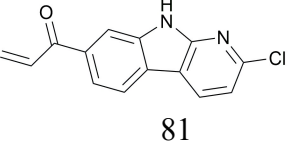
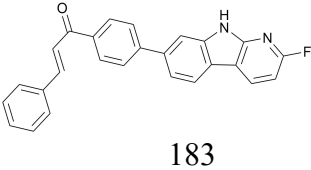
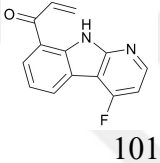
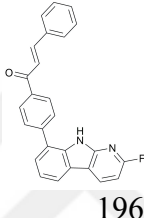
In addition to this, the literature review also turned up evidence of more favourable modifications. Mologni et al. performed a structure-activity relationship study about ALK and α -carboline derivatives, and reported low-micromolar range activity due to the presence of α,β -unsaturated carbonyl substitution [89]. This evidence inspired us to further design the various possible α,β -unsaturated carbonyl compound substitutions that were not tested in the related study, such as propenone and chalcone at every possible substitution point in the α -carboline structure with various halogenations at the opposite ring of the α,β -unsaturated substitution. Propenone was chosen as relatively small-size substituent, on the other hand chalcone structure was chosen as relatively large-size substituent for α,β -unsaturated carbonyl moiety.

3.1.1 Molecular Docking Studies of the Designed Ligands

Molecular docking study of designed ligands was carried out making use of the AutoDock Vina program to get the lowest energy poses of our ligands at the active site of the ALK kinase domain. The results were tabulated according to their substitution type and position (Table 3.1 & Appendix (A) Table 1). In this manner, compound 1 is the α -carboline itself without any modification, while compounds from 2 to 104 are substituted with propenone and halogenations at various positions in the α -carboline. In line with propenone substitution, compounds from 105 to 207 are substituted with chalcone and halogenations at various positions. Essentially, the table follows the pattern of one ring having either propenone or chalcone modification, while the other ring has either no modification or F, Cl, Br, or I halogenation at every position of the corresponding ring.

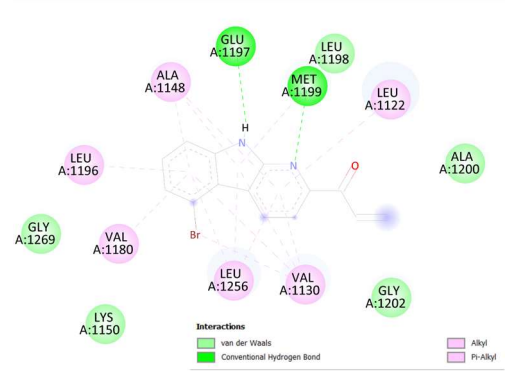
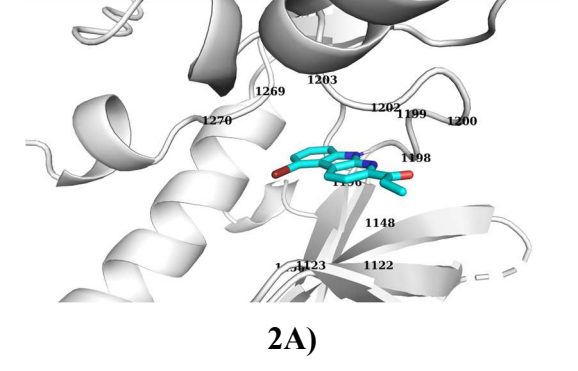
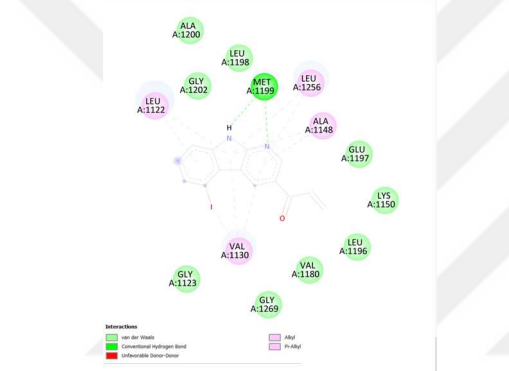
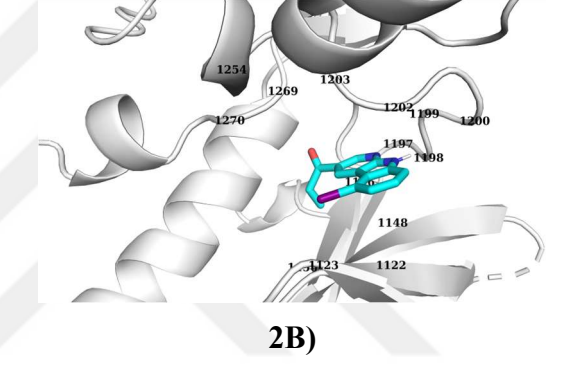
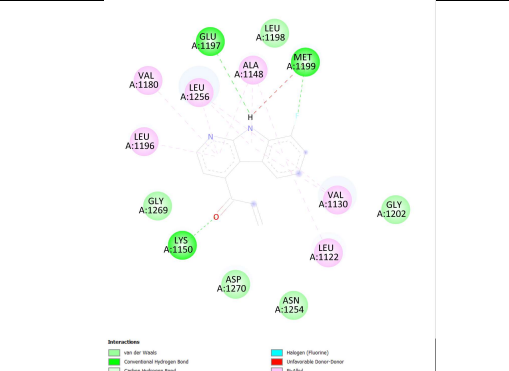
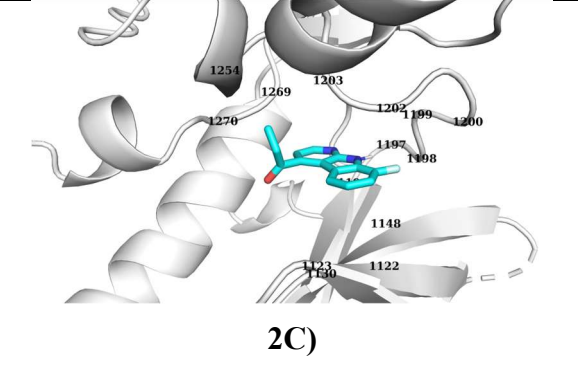
Table 3.1 The list of selected compounds from designed α -Carbolines with binding affinity towards kinase domain of ALK

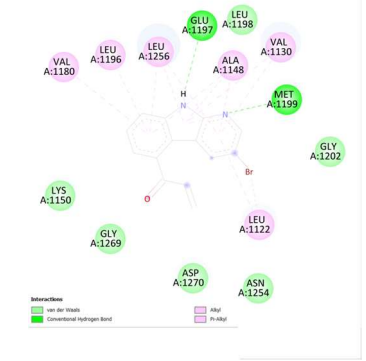
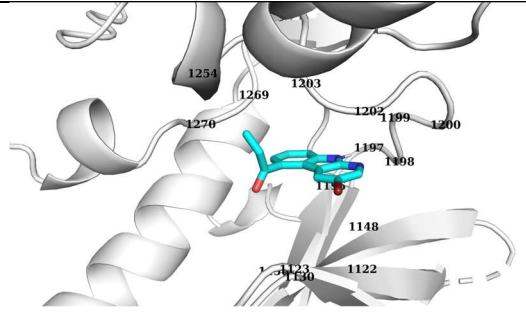
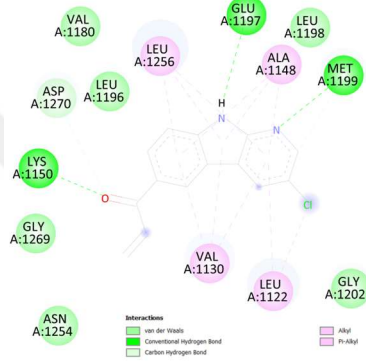
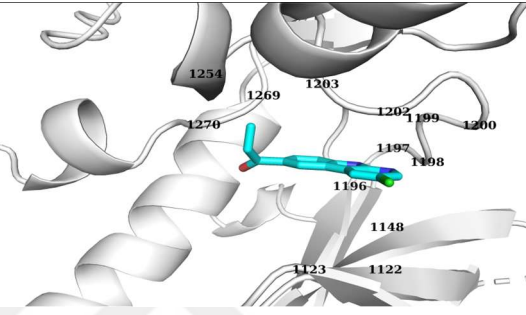
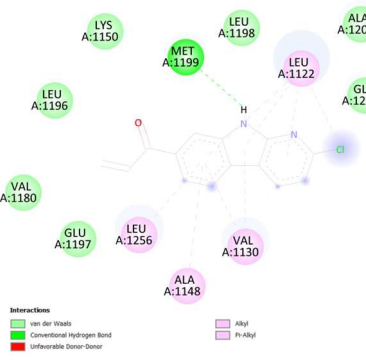
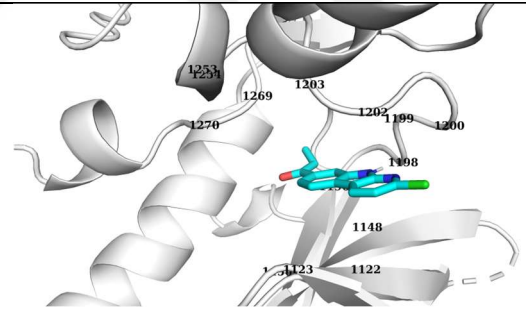
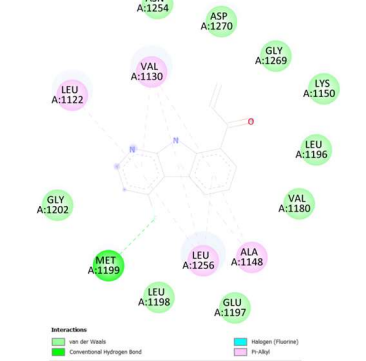
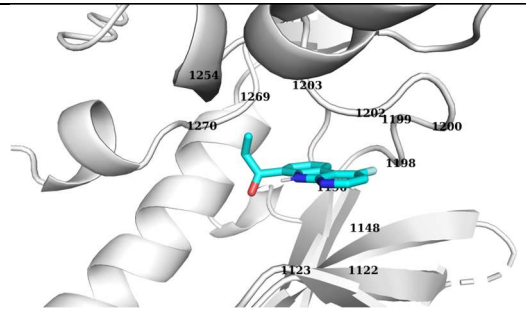
Structure C. Number	& Binding Affinity (kcal/mol)	Structure & S. Number	Binding Affinity (kcal/mol)
 5	-7.74	 110	-10.16
 23	-8.21	 127	-10.63
 49	-8.30	 154	-10.37
 60	-8.30	 164	-10.60

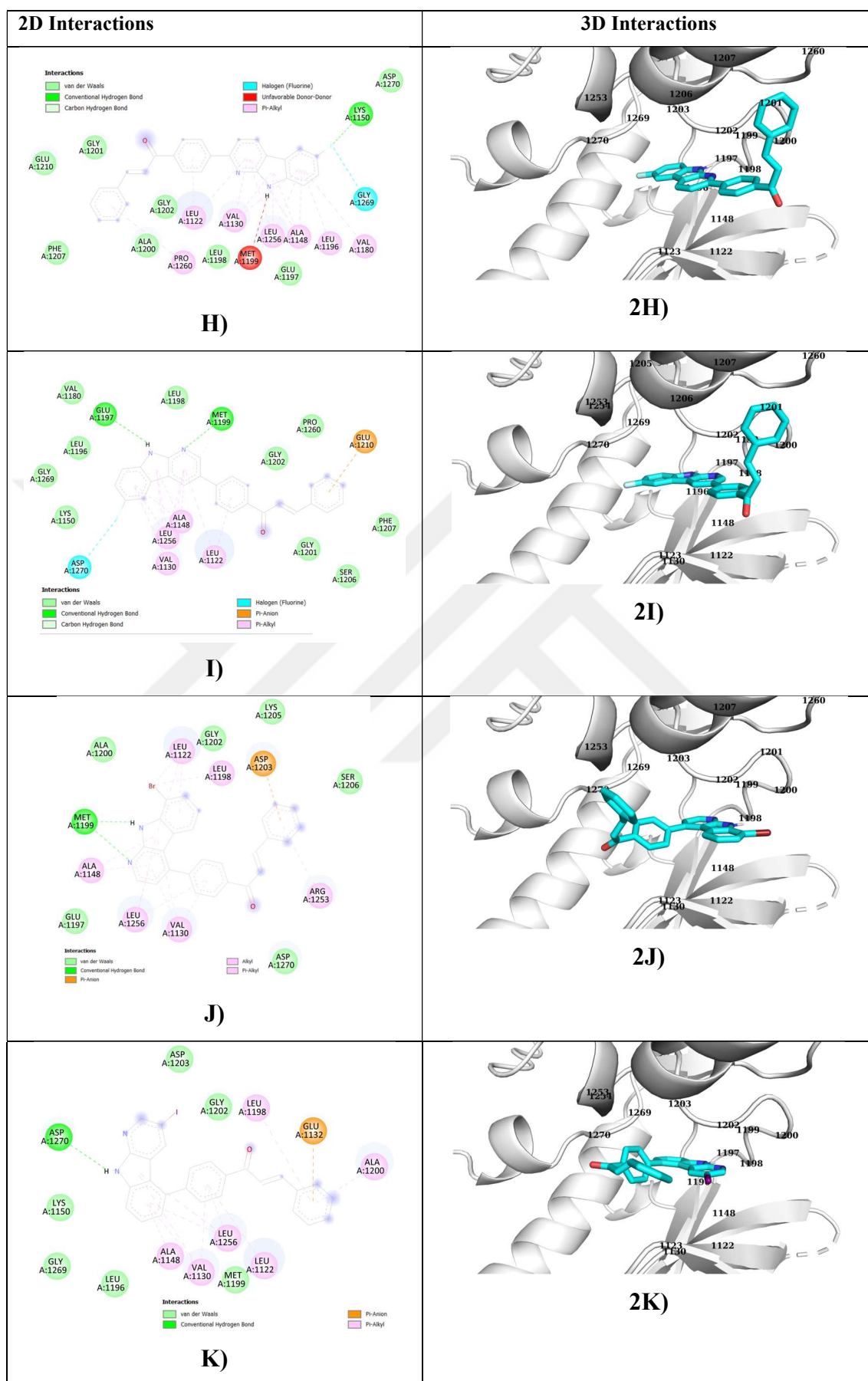
Structure C. Number	&	Binding Affinity (kcal/mol)	Structure & S. Number	Binding Affinity (kcal/mol)
 72		-8.08	 180	-10.13
 81		-7.81	 183	-10.20
 101		-8.21	 196	-9.70

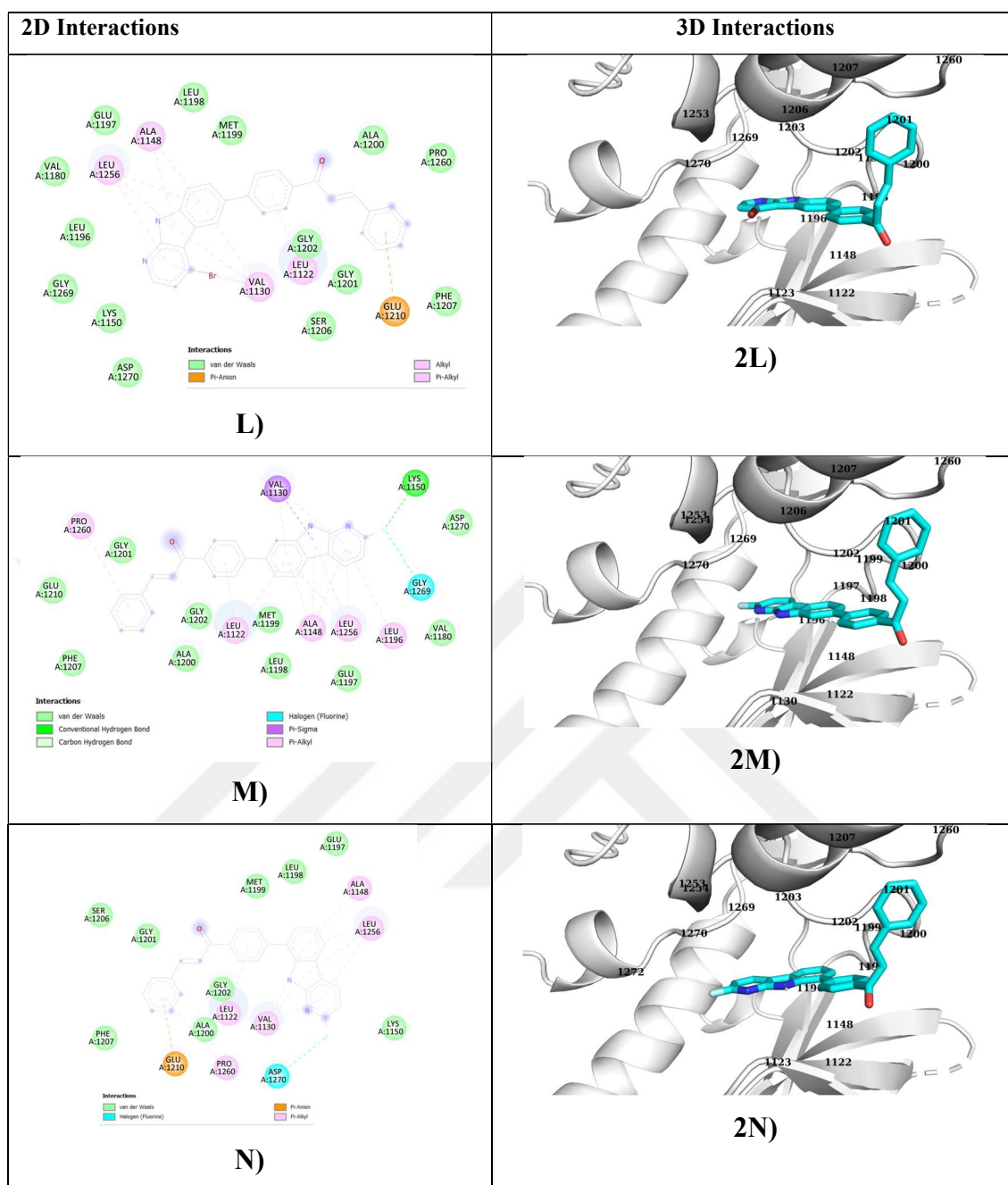
In this vein, some apparent clusters in the table formed, such as compounds 2 to 18 are substituted with propenone at R1 position and have F, Cl, Br, and I halogenations at R4 to R7 positions. The same cluster pattern is shown throughout the table. For example, compounds 19 to 35 have propenone at the R2 position with the same halogenations at R4 to R7 locations. In table 3.2, each cluster's lowest binding affinity compound's 2D and 3D interaction figures to the ALK kinase domain are shown (A-N, 2A-2N sections). 2D and 3D interaction maps show that most of the analysed compounds have hydrogen bonding with the M1199 residue or with the rest of the hinge region residues, from 1197 to 1201 [49].

Table 3.2 2D and 3D interactions docking results of the selected designed compounds (C. no: A = 5, B = 23, C = 49, D = 60, E = 72, F = 81, G = 101, H = 110, I = 127, J = 154, K = 164, L = 180, M = 183, N = 196)

2D Interactions	3D Interactions
 A)	 2A)
 B)	 2B)
 C)	 2C)

2D Interactions	3D Interactions
 D)	 2D)
 E)	 2E)
 F)	 2F)
 G)	 2G)





Overall chalcone substituted α -carboline ligands have all lower binding energies than the propenone substituted ones. While in the chalcone substituted ligands, chalcone with halogen substituted ones mostly have lower energies than only chalcone substituted ligands. Moreover, the ligand internally labelled as 127, chalcone substituted at the 3' position with fluorination at the 6' position has the lowest binding affinity with -10.63 kcal/mol amongst all designed ligands. The 2D interaction of compound 127, illustrated possible two hydrogen bond with fluorine, Pi-Anion, and various other interactions.

When we investigate the interactions of the lowest affinity compound of each ligand category, it is clearly seen that all of them make some degree of interaction with the active

site of ALK kinase domain. However, some of them have the potential to be more stable due to the presence of hydrogen bonds with key residues as it was highlighted in the introduction, such as M1199, in which most of the approved kinase inhibitors inhibit the kinase activity by binding this specific residue. These compounds are 5, 23, 49, 60, 72, 81, 101, 127, 154. Consequently, the rest of each cluster's lowest binding affinity compounds were selected to further study their potential stability via subjugating them to all-atom 100 ns MD simulation via GROMACS.

3.1.2 Molecular Dynamics Simulation of Selected Ligands Against ALK Kinase Domain

To realistically infer the stability of the ligands in their docked location, utilizing MD simulation is the preferred course of action in literature. Hence, all-atom MD simulation for selected compounds was executed by using GROMACS. To have an appropriate understanding of the stability of the protein-ligand complex, a 100 ns all-atom MD simulation was performed.

The calculated trajectory of the atoms through MD simulation can be further analysed to produce data about specific aspects of the simulation such as the average distance between two selected elements, the centre of mass (COM) of ligand and COM of important residues in the kinase domain, RMSD value, number of hydrogen bonds between two selected elements, etc. As a result, we first focused on the RMSD values over simulation time, which all can be located at the Fig 3.1. RMSD values calculate the distance between two points. In our figures, the green line represents RMSD values of the specific compound that was tested, between the ligand and protein's backbone atoms, while the protein and backbone refer to the distance difference between their initial position and their current position. Hence, from this information alone one can interpret the graph as the lesser the extent of fluctuation in the graph means better the stability of the protein-ligand complex. From the RMSD figures, some compounds' RMSD values do fluctuate more than others while also even exceeding the 10 Å distance from protein (Fig 3.1). These observations strongly suggest that compounds that did not fit the aforementioned criteria are highly likely to be unstable while considering the environment in the simulation. One of the interesting observations is the fact that most of the chalcone-substituted compounds, regardless of their substitution position, seemed to fit our potentially unstable criteria, which was having a constant RMSD value over 10 Å.

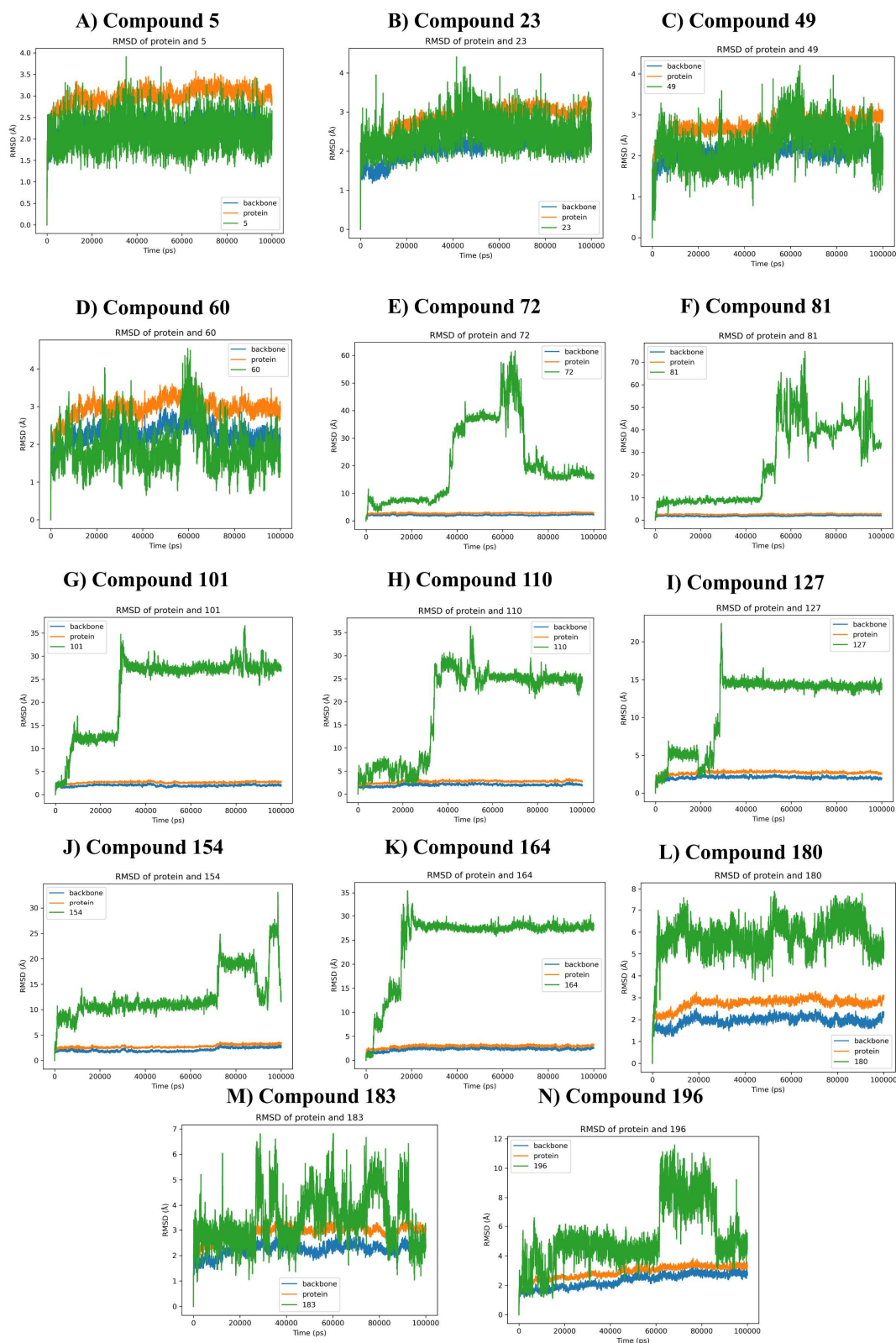


Figure 3.1 RMSD plots from their MD simulation. Blue line represents the backbone atoms' RMSD, the orange line represents the protein atoms' RMSD, and the green line represents the RMSD values between the compound and the backbone atoms.

The root mean square fluctuation (RMSF) analysis measures the flexibility and stability of the residues of the protein by calculating the positions of the residues in the trajectory and comparing them to the initial structure. When the RMSF values of the tested compounds are evaluated, it could be easily seen that the compounds that had higher RMSD values had higher number of RMSF value peaks, which we can define as RMSF values higher than 2 Å. Compound 5 has the lowest number of RMFS peaks, defined as the same as before, indicating that ligand does not significantly affect the stability/flexibility of the residues and hence, the ALK kinase domain. This observation of the figure could be interpreted as our protein-ligand complex is showing hints of stability. RMSF graphs of all the test compounds are given in the Figure 3-2.

The Radius of gyration (Rg) analysis measures the compactness of a protein by comparing the RMSD values of the residues to the centre of gravity of the protein. Hitherto, if the ligand binding affects the conformation/structure of the protein, large differences in Rg should be observed if one presumes that overall structure change results in the increasing distance from the centre of gravity to the residues. In this manner, the Rg graph of compound 5 shows no significant differences compared to the initial value. Moreover, all the test compounds are in the range of 1.95 (nm) to 2.05 (nm) range, hence all could be reported as relatively stable over the course of the simulations. (Fig. 3.3)

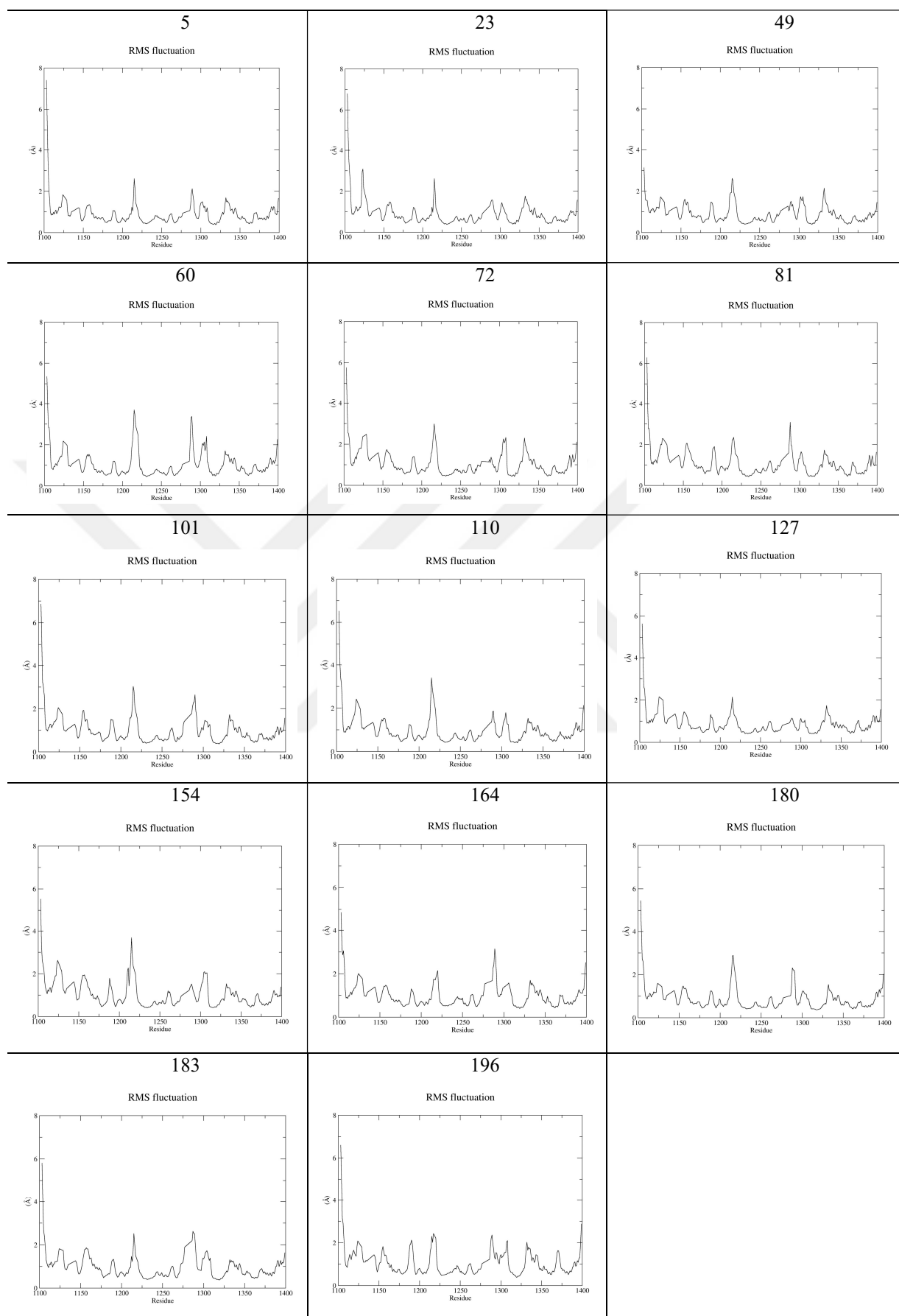


Figure 3.2 RMSF analysis of the selected compounds against ALK kinase domain.

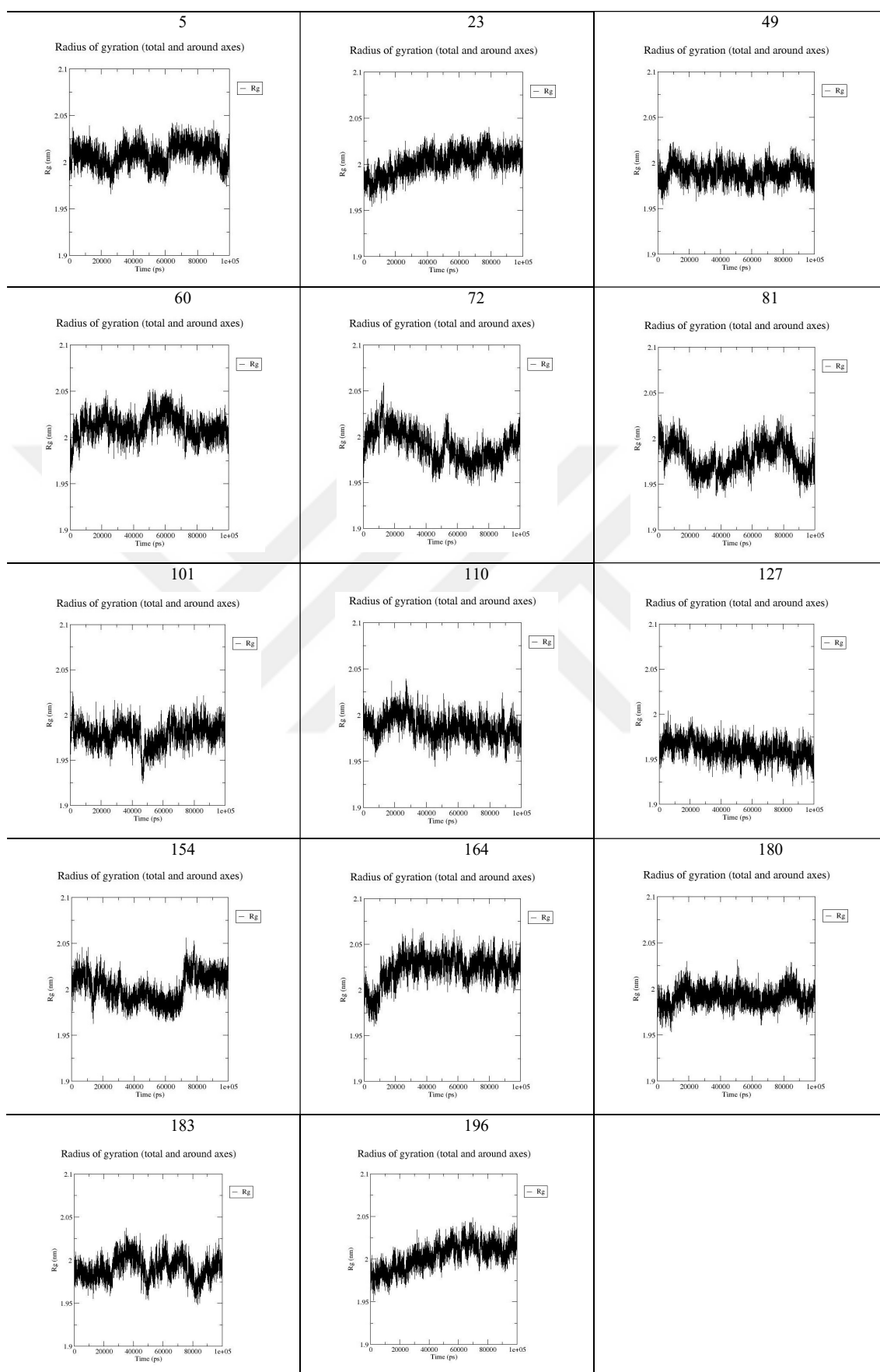


Figure 3.3 Rg analysis of the selected compounds against ALK kinase domain.

However, the RMSD value alone does not necessarily prove that compounds are in fact unstable. To elucidate this fact more clearly, we performed the average distance between two selections by utilizing the MDAnalysis Python package [110], [111]. In Fig. 3.4, average distance is presented between the COM of the compound that is being tested, and the COM of selected residues. On the Y-axis, each line represents a residue selected from the binding site based on literature, while the other selection represents the compounds of interest. The X-axis represents simulation time in picoseconds. From Fig. 3.4, the proximity of each compound to the specified residues can be interpreted: when a compound is near a residue, the average distance decreases and the heatmap appears darker; as the compound moves away, the distance increases and may exceed 10 Å, the ceiling of the heatmap.

As a result of both RMSD values and heatmaps, we can more strongly interpret the location and stability of the selected ligands. In our cases, 6 of the compounds seem more promising for further analysis. These are compounds labelled as 5, 23, 49, 60, 180, and 183. The main reason behind this selection was based on the RMSD values as well as mostly their suggested location in the active site via analysing the heatmaps, dark heatmaps signify better average distance over simulations, which greatly suggest the location and stability of the compound over the course of the simulation, while RMSD alone cannot achieve this level of specific information.

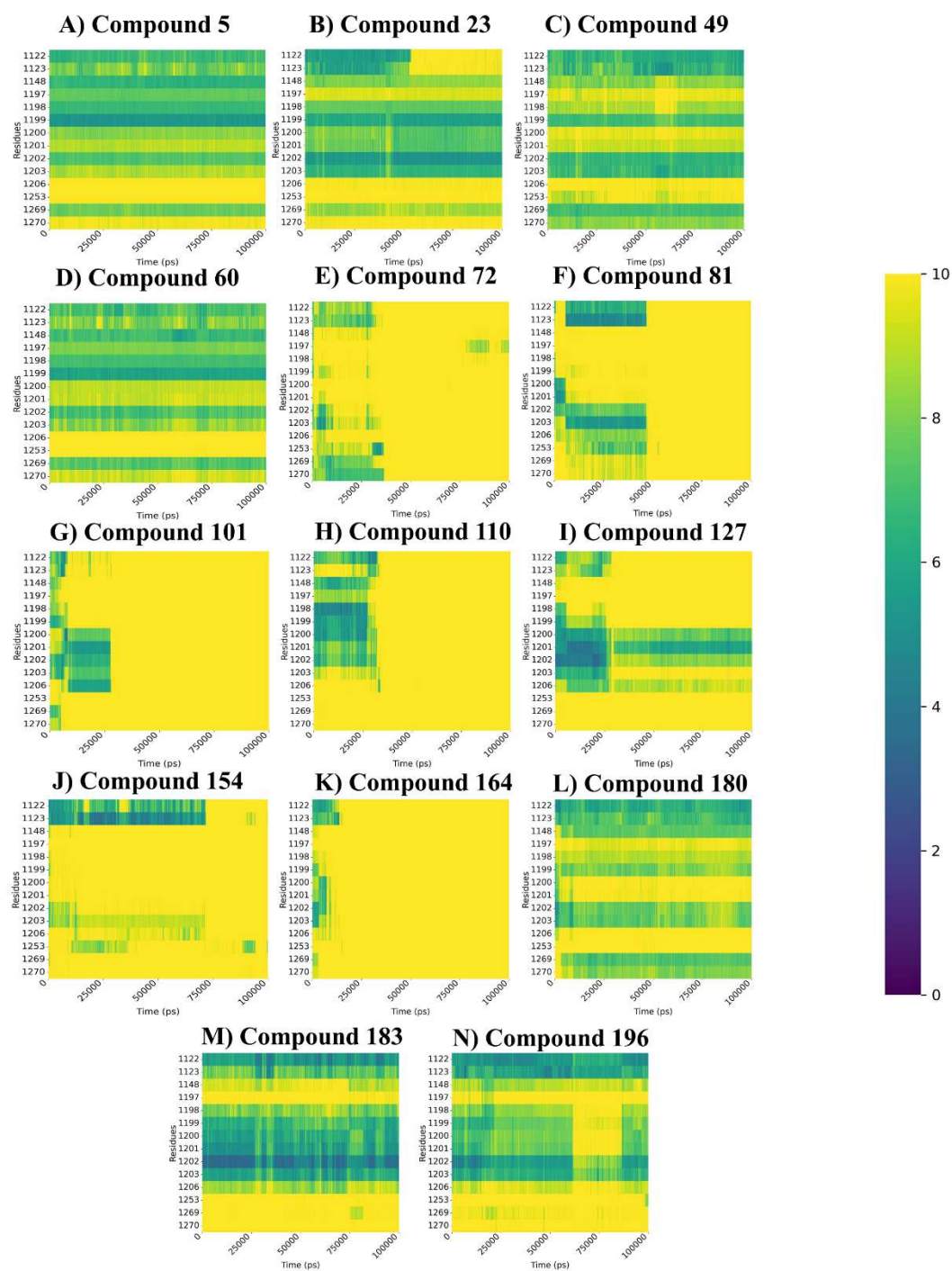


Figure 3.4 Average distance (in Å) heatmaps of 14 compounds, the lowest binding energy from each category during 100 ns simulation.

To further ensure the reproducibility of the 6 short-listed compounds stability, they were subjected to 2 more replication of MD simulations with same conditions that are reported in the appropriate method section with different initial seeds. Each run was in the acceptable range of their initial run, which can be understood from the table 3.3. Especially from Standard Deviation (SD), coefficient of variation values (CV), max pairwise difference (Max Δ). However, one of reruns of MD simulations of compound 49 seemed diverge from their initial run by quite a bit, but their rest of two of runs were similar to each which could be observed from the figures 3.5, 3.6, and 3.7. Altogether, this suggest that even though reproducing exact trajectory down to its every atom position on a system is not possible with our standard methodology, capturing same trend with different seeds to our simulation which was randomly assigned by the GROMACS was observed thoroughly in these replicates.

Table 3.3 Reproducibility analysis of 6 short-listed compounds from triplicate MD simulations.

C. No	Metric	Mean + SD (Å)	CV (%)	Max Δ (Å)
5	RMSD	1.92 $\pm$ 0.13	6.88	0.26
	RMSF	0.91 $\pm$ 0.02	2.55	0.04
	Rg	20.05 $\pm$ 0.05	0.23	0.08
23	RMSD	1.95 $\pm$ 0.10	4.89	0.17
	RMSF	0.94 $\pm$ 0.08	8.80	0.15
	Rg	20.05 $\pm$ 0.17	0.87	0.34
49	RMSD	4.53 $\pm$ 4.68	103.42	8.24
	RMSF	0.97 $\pm$ 0.10	9.85	0.18
	Rg	20.00 $\pm$ 0.11	0.55	0.22
60	RMSD	1.59 $\pm$ 0.14	8.92	0.28
	RMSF	1.00 $\pm$ 0.04	4.43	0.09
	Rg	20.06 $\pm$ 0.17	0.84	0.32
180	RMSD	5.36 $\pm$ 0.11	2.11	0.21
	RMSF	0.91 $\pm$ 0.04	4.01	0.07
	Rg	19.91 $\pm$ 0.00	0.02	0.01
183	RMSD	3.05 $\pm$ 0.19	6.13	0.37
	RMSF	0.93 $\pm$ 0.10	10.48	0.19
	Rg	19.89 $\pm$ 0.04	0.18	0.07

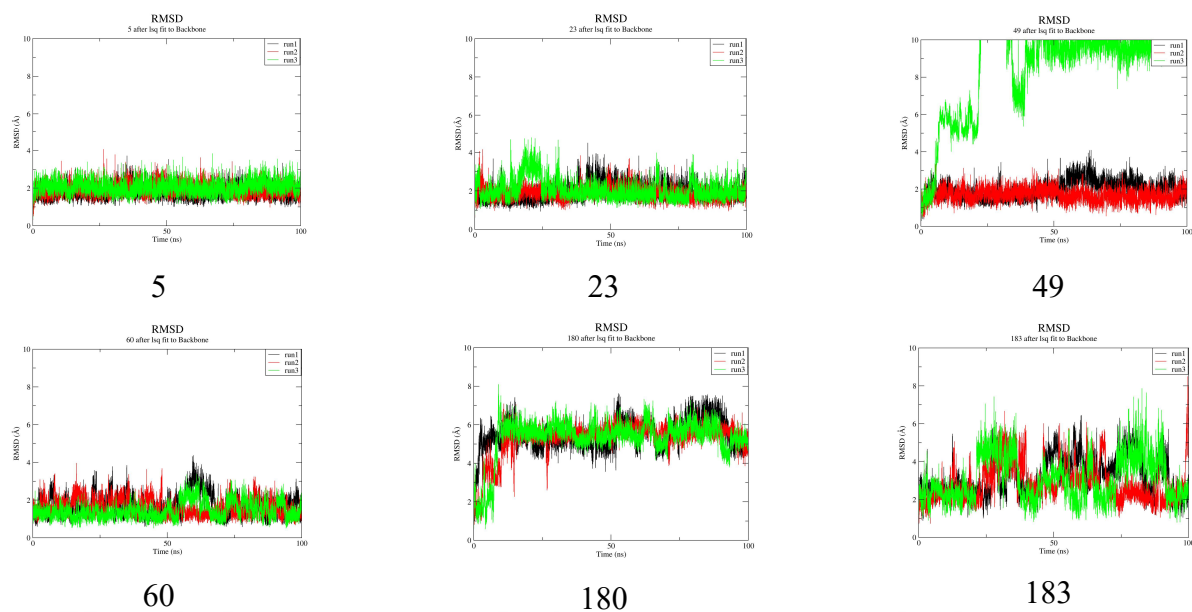


Figure 3.5 RMSD analysis of 6 short-listed compounds from triplicate MD simulations.

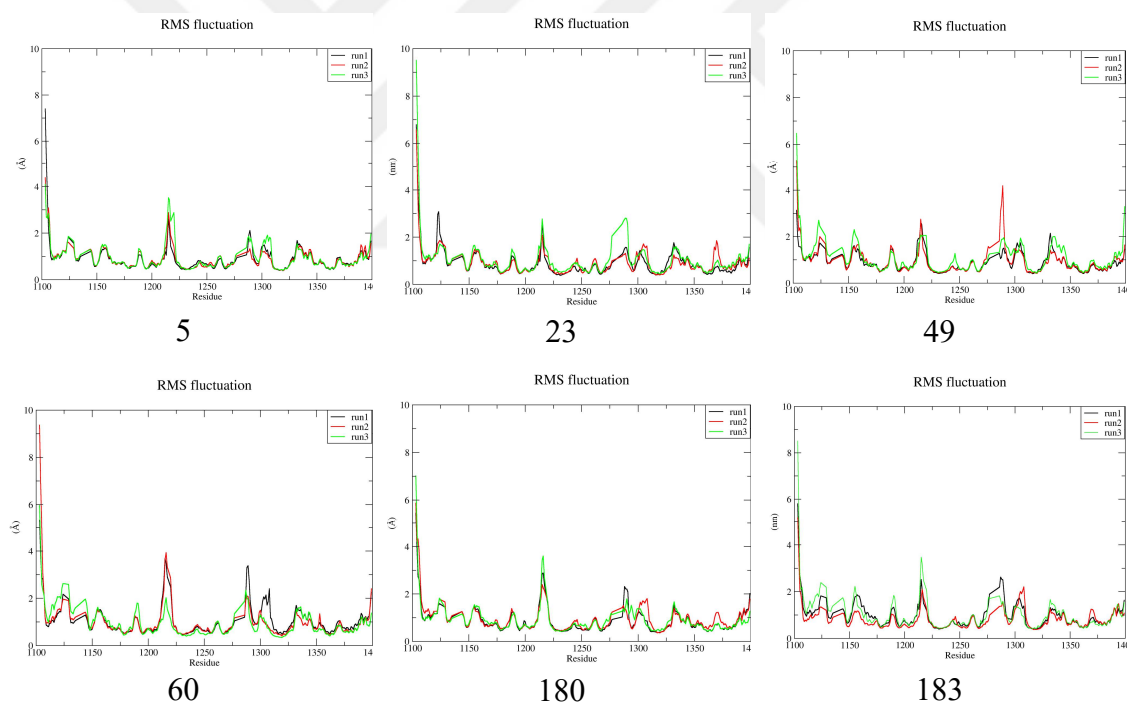


Figure 3.6 RMSF analysis of 6 short-listed compounds from triplicate MD simulations.

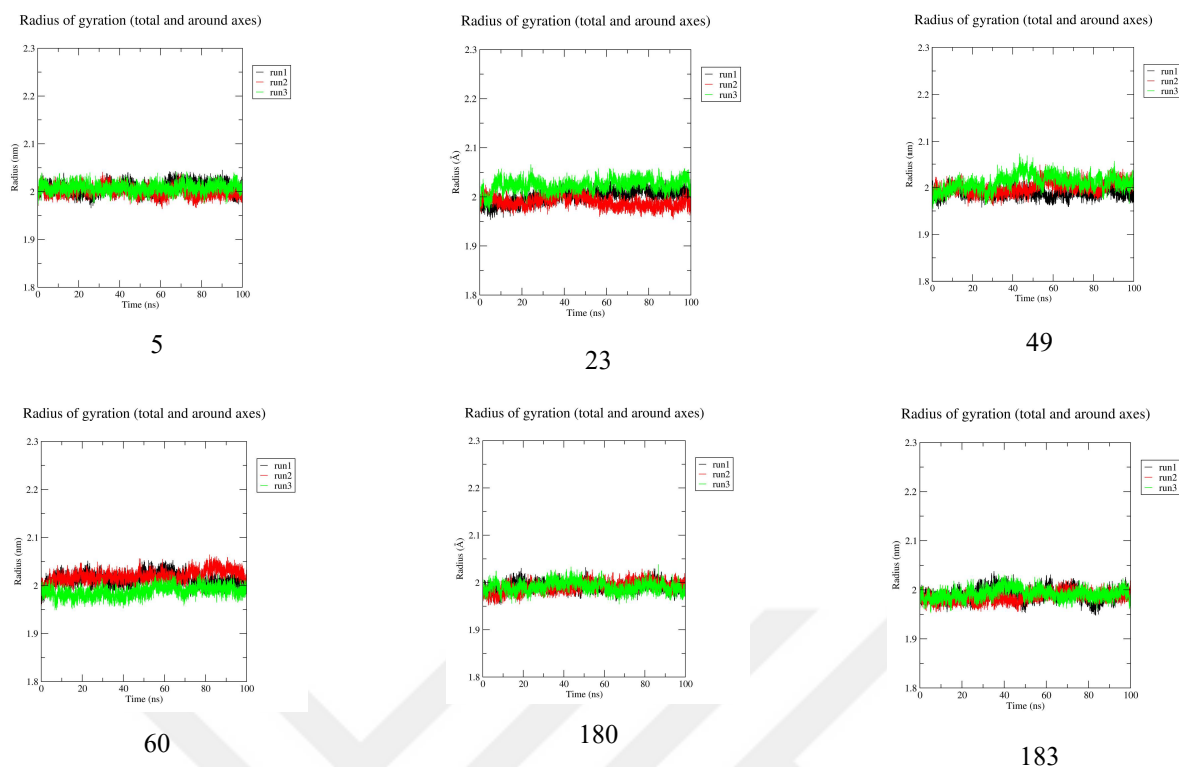


Figure 3.7 Rg analysis of 6 short-listed compounds from triplicate MD simulations.

Furthermore, the number of hydrogen bonds over the course of the simulations from triplicates was analysed and listed in Fig. 5. As suspected from the stability suggestion via average distance and RMSD analysis, these seven compounds retain their hydrogen bonding in the specified poses to an extent while also fluctuating in regard to number of hydrogen bonds, to an extent. Specifically, 5 seems to keep its two original hydrogen bonds that have been identified in the docking (see table 3.2, Fig. 3.1'A) over the whole course of the simulations in every replicate, which is an immense indication of stability..

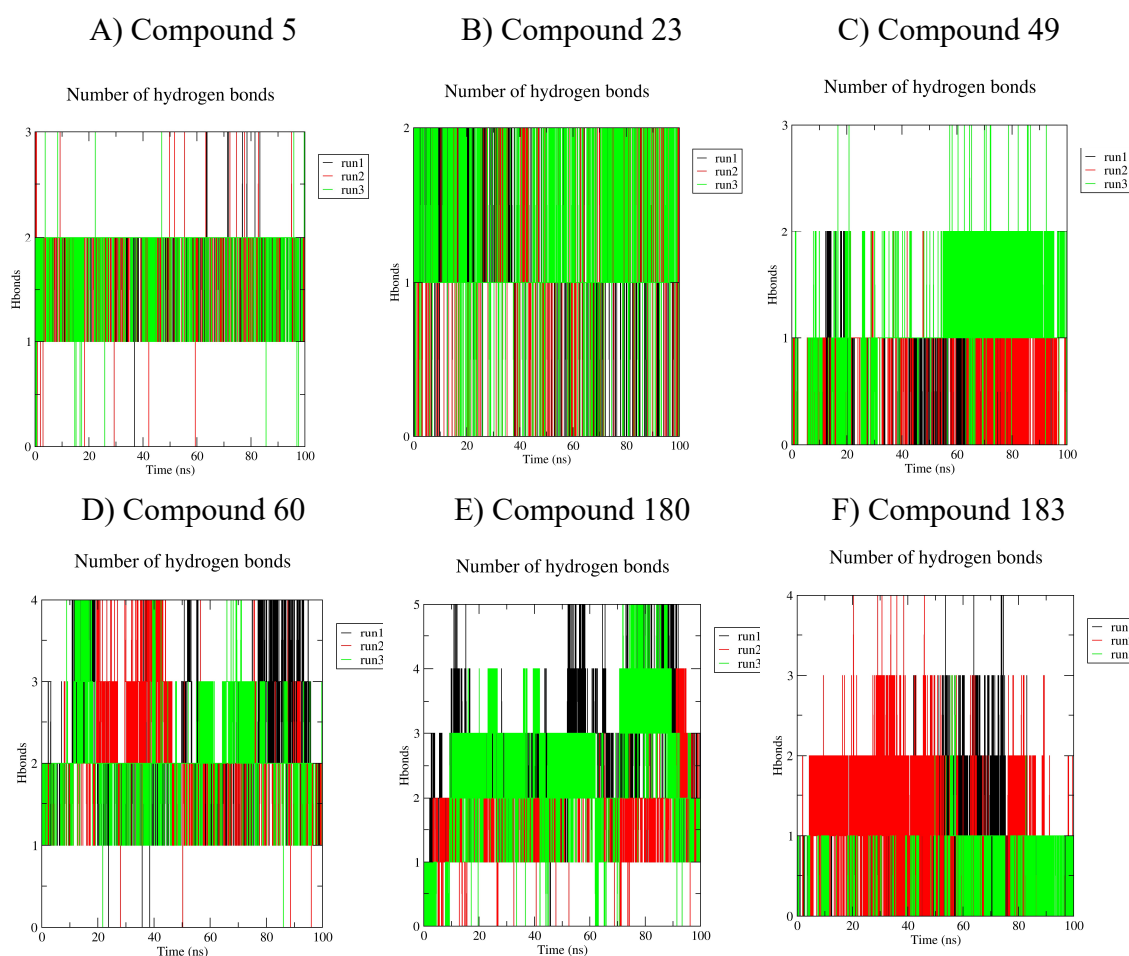


Figure 3.8 Number of hydrogen bonds between the various ligands and the ALK kinase domain over the course of the simulation

3.1.3 MM/PBSA Average Binding Energy Calculation

Calculation of binding free energies was performed via MM/PBSA analysis by `g_mmpbsa` package. In Fig 3.9, the binding free energies of six short-listed compounds have been illustrated in the bar graph. As hypothesized from the hydrogen bond analysis of compound 5 has shown the lowest binding free energies in our interested bunch with average almost -18 kcal/mol. This fact combined with the 2D interaction of the compound 5 displaying 2 hydrogen bonds with the fact that it does indeed have had 2 hydrogen bonds between the compound and protein over the course of the simulation, indicates that compound 5 can stably keep the original binding mode. Similarly, the graph of compound 23 with -13.8 kcal/mol indicates closest energy values to compound 5 between 6 short-listed compounds, combined with the fact that in, combined with the fact that in figure 3.8's B, which signify either 1 or 2 hydrogen bonding over the course of simulation and two hydrogen bonds in the docking position, respectively, makes the compound 23

attractive choice as compound 5. Therefore, compounds 5 and 23 hold the most promise of all the designed ligands, and they are both propenone substituted at either R1 and R2 positions, respectively.

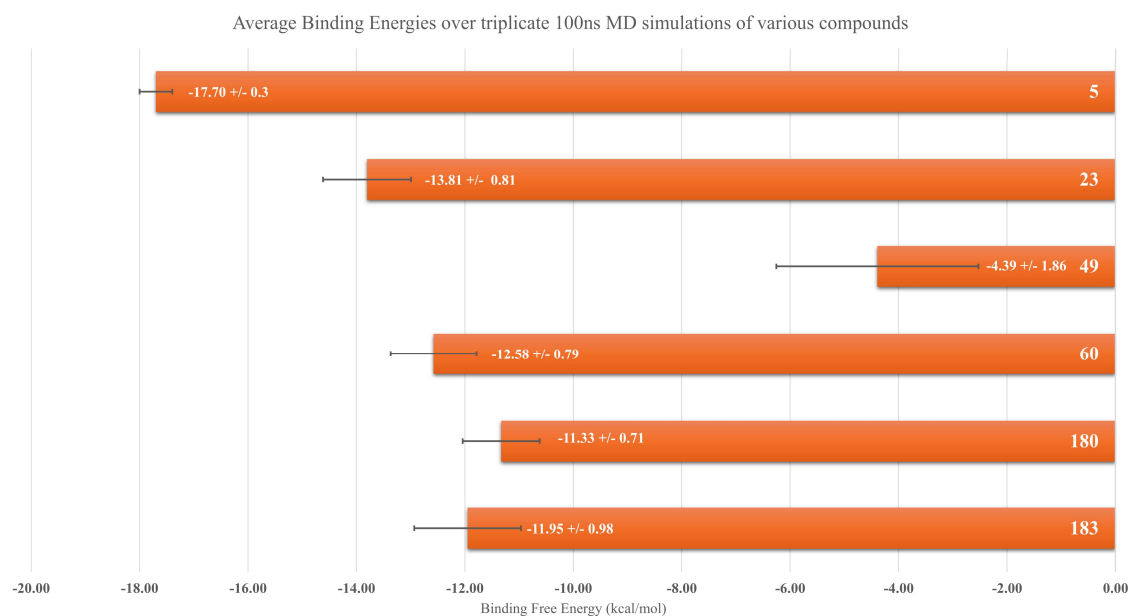


Figure 3.9 MM/PBSA average binding free energy graphs of 6 shortlisted compounds from triplicate 100ns MD simulations.

3.1.4 Towards the Design of More Stable A-Carboline ALK Kinase Inhibitors

When the molecular docking and MD results were analysed, even though chalcone substituted α -carboline analogues hold the most potential according to the lowest binding affinity in the AutoDock Vina, they have been discovered as unstable and resulted in not being able to hold onto the original docking position in the MD simulations. When the MD simulation movies of the chalcone substituted compounds are analysed, a reason behind the observed instability was discovered as the carbon-carbon sigma bond between the phenyl group (Ring A) of the chalcone and substituted position of the α -carboline (see the MD movies) due to its ability to rotate freely at the axis, which ultimately results in not being able to stick to the active site of the ALK's kinase domain. (Fig 3.10)

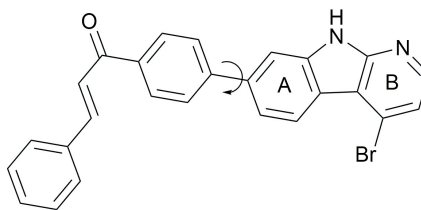
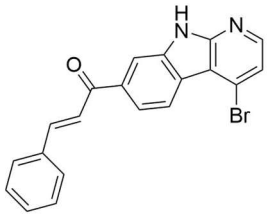
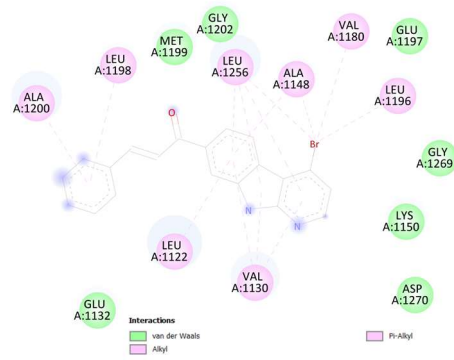
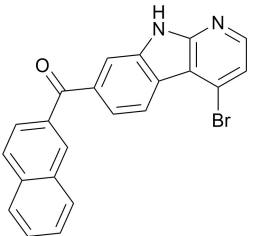
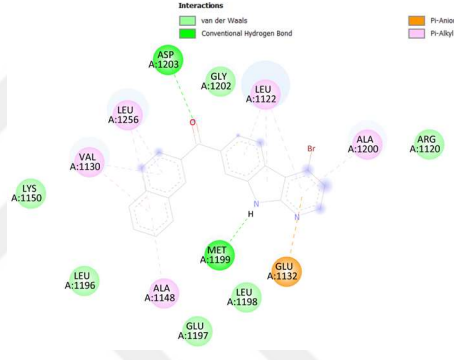


Figure 3.10 Structure of compound 193, as an example of chalcone, substituted carboline, and the representation of rotation on sigma bond between two phenyl groups.

Hence, from these observations, we concluded that a modification at the carbon-carbon sigma bond will result in a more stable compound while also saving the Pi interactions that other phenyl group (Ring B) deals with that significantly contributes to lower binding affinity. In this manner, we design a new compound that basically lacks the phenyl group (Ring A) which is directly bonded to the carboline structure. In the new design, α -carboline directly links with the carbonyl carbon of the cinnamoyl group (compound 208; Table 3.4). After the evaluation of the molecular docking studies of compound 208, a restricted analogue of the compound 208 was designed (compound 209; Table 3.1). The effect of the modification on compound 193 (-9.99kcal/mol binding affinity) is demonstrated in Table 3.4 that new compounds are relatively close to this value specifically compound 209 with the -9.80 kcal/mol binding affinity.

Table 3.4 Molecular Docking Evaluations and 2D interactions of the modified compounds.

Compound No. & Structure	Binding Affinity (kcal/mol)	2D interactions
208 	-9.05	
209 	-9.80	

In a similar manner to previous modifications, all-atom 300 ns MD simulations with the same setting as previous, were applied to both compounds that are designed in this section in a triplicate manner. The same analysis to simulations was also conducted with additional clustering analysis, PCA to binding region included specifically hinge residues and FEL analysis from protein atoms' RMSD and protein Rg was analysed. (Fig 3.11, 3.12 and 3.13-15) When the results of simulations from triplicates were compared to template compound 183, which is the closest compound that was simulated while having the chalcone at the same position with R1 fluorination instead of R3 bromination, it could be easily seen that overall, all of the RMSD values of the compound 208 is significantly lower than the 183. In fact, it could be reported as not stabilised over the course of the simulation due to its fluctuation continuing even in the last 10 ns jumping from the 3 Å to 7 Å range. Moreover, compound 208 had lower RMSD values than 183 and it is much more stable than the 183, only fluctuating between 1-3. When average distance heatmaps of the lowest RMSD average of the course of the simulations of the new compounds compared to 183, in the 183's simulation average distance between COM of the 183 and

key residue 1199 is around 8 Å and around 9+ Å with the rest of the hinge region residues, which indicates that it has not able to hold onto target site of the active site while in the new compounds' especially compound 208 average distance is being in the 5-6 Å between amino acid 1199 and relatively rest of the close that amino acid over the length of the simulation. However, the same could not be reported for compound 209, it is in the ~9 Å range in the amino acid 1199, and farther away from the rest of the amino acids. Interestingly, the same trend of 208 indicating more stableness than 209 was not followed in the number of hydrogen bonds during the simulation. In this analysis, 209 was observed to form more hydrogen bonds than 208, which could be interpreted as 209's conformation forming more stable complex than the 208's. Supporting this interpretation, same stableness was observed in the RMSD graph in run 1 and 3, which did not significantly fluctuate after 25ns simulation in both runs. Even though run 2 MD simulation of the 209 followed the similar pattern until ~150ns mark, after this time it stabilised in different region around 20 angstrom. Due to this, run 1 was considered for the following analysis on the 209 while run 2 was selected for the 208. On the other hand, this stableness of the conformation of the 209's after around 25ns simulation can be construed indicative of as not as effective binding as the 208, from the observation of the how close to one of the key residue of ALK kinase's hinge region 1199, which floated around ~9-10 Å. Overall, this could be depicted as the original 209's conformation was not preserved and new conformation and site in the region kinase could be less effective than the 208's due to 208 being in the binding site location of the kinase domain was thoroughly preserved which can observed from the average distance between COM of residues and COM of 208 and as well as the MD movies.

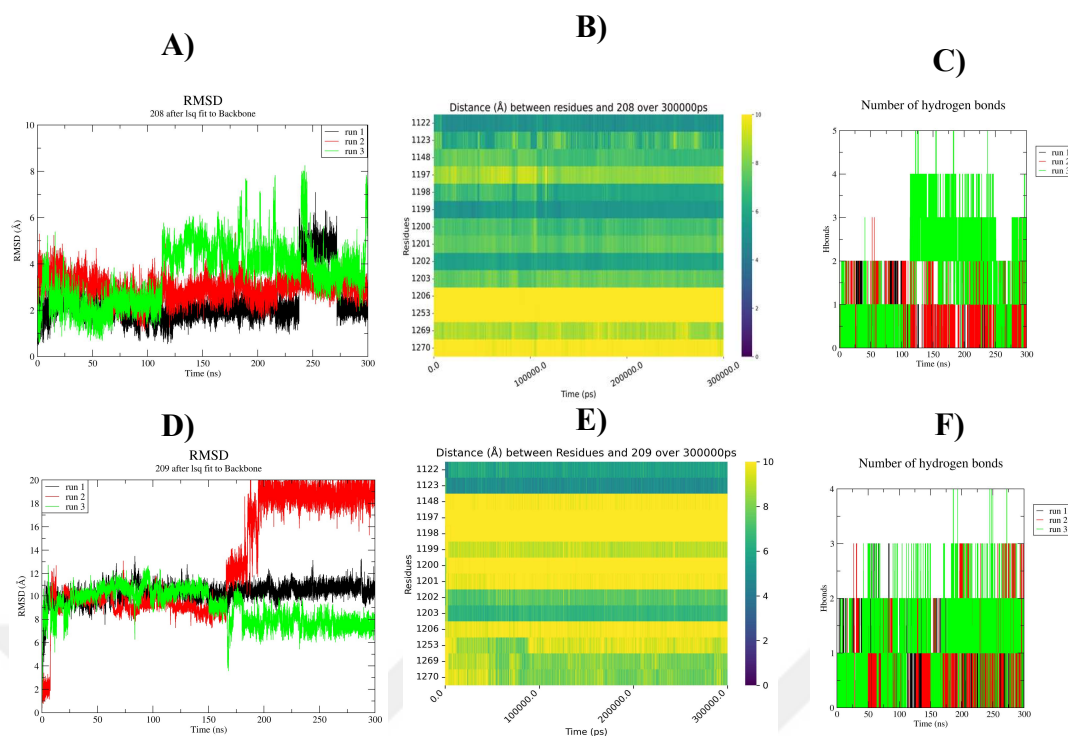


Figure 3.11 MD simulation analysis of compounds 208 and 209. A, B, and C sections represent the compound 208; D, E, and F sections represent the compound 209. A, D) RMSD plot of ALK kinase domain with compounds over 300 ns MD simulation. B, E) Average distance heatmaps. C, F) Number of hydrogen bonds between the ligands and the protein over the course of the simulation.

To gain deeper insight into binding pocket movements, PCA was done on the binding pocket of only ALK tyrosine kinase (TK) domain, 208-ALK TK complex's MD simulation titled as run 2, 209-ALK TK complex's MD simulation titled as run 1. Binding pocket was selected by selecting following residue IDs 1196, 1198, 1199, 1200, 1202, 1203, 1120, 1122, 1123, 1253, 1254, 1256, 1130, 1132, 1269, 1270, 1146, 1148, 1150. After aligning the trajectory to backbone atoms of the ALK TK, binding pocket was defined as previously stated, cartesian coordinates of the defined and selected pocket atoms was extracted. Afterwards, rest of PCA commenced. Via using result of the PCA, trajectories were projected on the two PCs, which we plotted in 2D representations. (Figure 3.12)

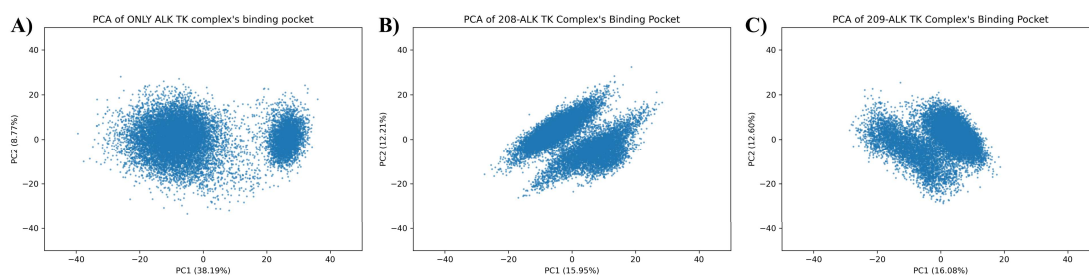


Figure 3.12 Principal component analysis of the compounds 208 and 209 – ALK TK complexes, and only ALK TK. Percentages signify how much of variance that PC account for.

As it is clear in the figure 3.12, only ALK TK binding pocket shows so much more movement than the rest, which could be reported as expected. Due to the fact there is no molecule/compound the restrict the movement of the atoms, which ends up binding pocket atoms freely exploring available space. While 208-ALK TK complex and 209-ALK TK cases, much more restricted variance is observed, especially in 208. 208-ALK TK binding pocket seemed to be more restricted than 209 and only ALK TK, due to its seemingly tighter clustering than both. From this it could be inferred that, less movement on the binding pocket, could be a result of stable binding of the compounds to the binding pocket, which one would expect to significantly hinder its ability to freely explore the available space.

To also methodically study the movements of the backbone atoms of compound-ALK TK complexes clustering analysis on the backbone atoms to discover prominent conformations were applied. (Figure 3.13-14) In both cases, 4 prominent clusters were discovered and both complexes' cluster's relative distance each other was similar while 209's clusters being more closely related than the 208's. This fact could be inferred from the dendrogram section of clustering analysis, which shows that in 208's case there is almost 60 atomic unit (AU) distance between cluster four and both cluster three and two. While these clusters distance between them discovered to be around 35 AU. Interestingly cluster one's distance to others was more than 80 AU. In slight difference to 208, 209-ALK TK complex's clusters showed similar relative differences to 208 with 209's clusters' spread radius being smaller than the 208's which indicate within clusters there is less variance to each other than 208. But one examines clustering dendrogram of the 209-ALK TK domain cluster and compared them to 208, stark difference could be observed. The clusters are more far away from each other than 208, cluster four and three

have more than 40AU distance between them while clusters two and one, have even more than that. On top of that, distance between clusters 1-2 and 3-4 are almost 100 AU.



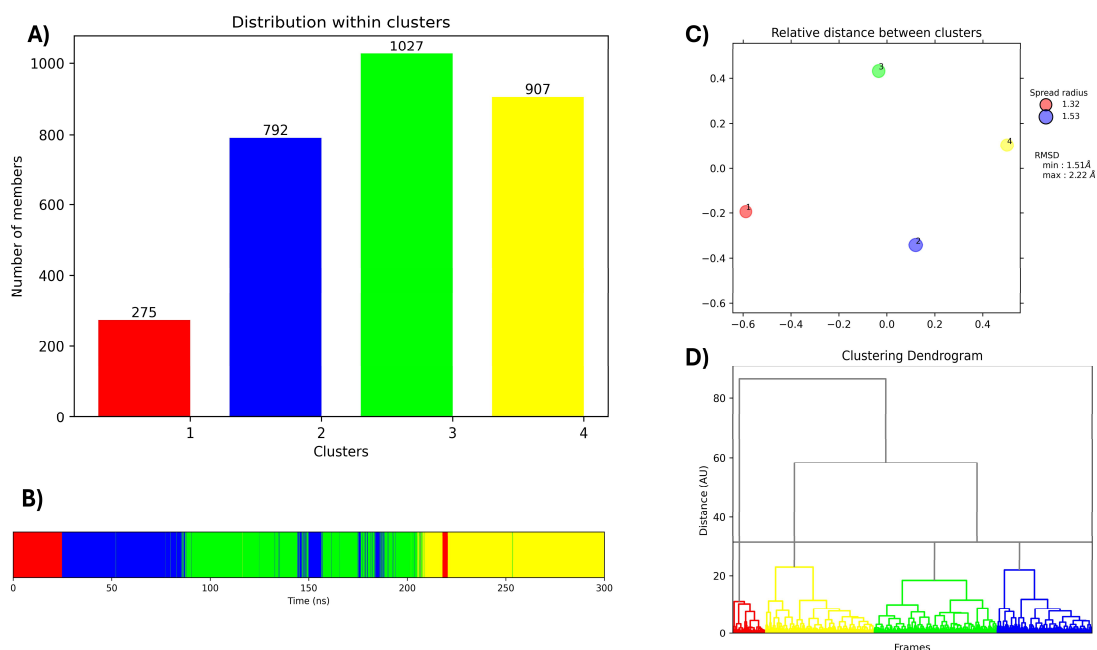


Figure 3.13 Clustering analysis of 208-ALK tyrosine kinase complex. A) Histogram plot of number of members across all clusters. B) A linear projection of clusters across the trajectory C) Distribution of clusters relative each other. D) Clustering dendrogram.

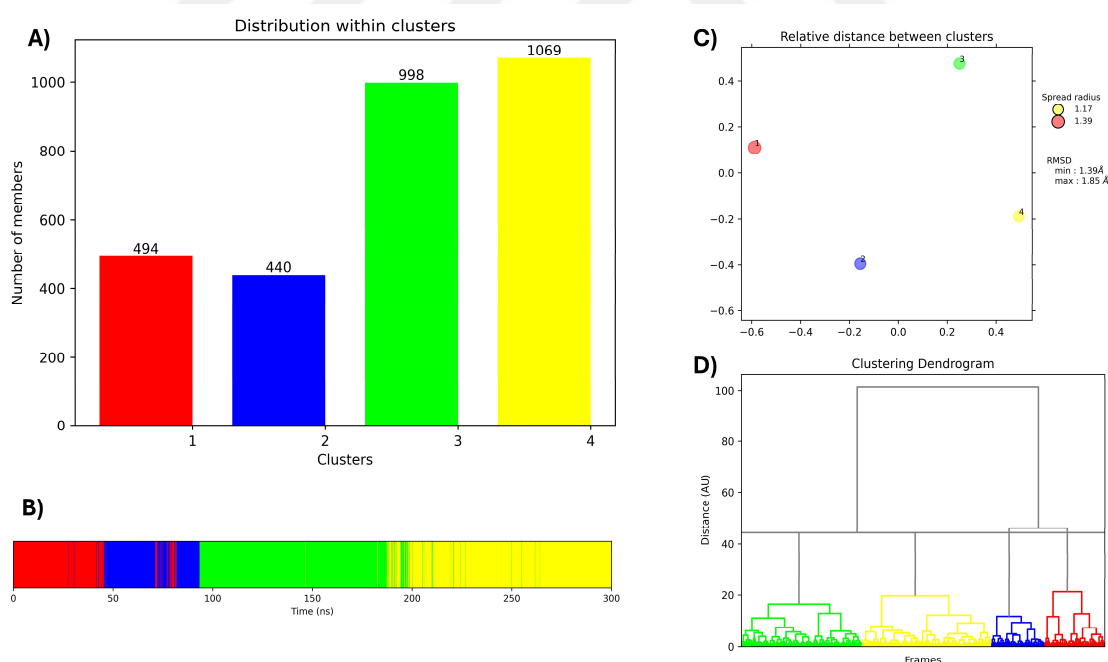


Figure 3.14 Clustering analysis of 209-ALK tyrosine kinase complex. A) Histogram plot of number of members across all clusters. B) A linear projection of clusters across the trajectory C) Distribution of clusters relative each other. D) Clustering dendrogram.

Concurrently, to clarify the fact that 208-ALK TK complex's entire backbone and its binding pocket seemingly being more stable than the 209-ALK TK complex, Rg-RMSD (both protein atoms) based FEL analysis was conducted. (Fig 3.15) In figure 3.15 as control only ALK TK domain's FEL was also constructed, which helped to compare both compound complexes not to just each other and to also compare to control as well. To this end, in control two strong basins are observed, while around 20Å Rg and RMSD at 2.75Å being more prominent than the other one, around 19.6Å and 2.75 Å. Between these two basins, relatively low free energy location around 8kJ/mol area exists, which means that in order to pass from one basin's conformation to other one only around 8kJ/mol barriers needs to be passed. In this case, we can see that protein atoms explored large stage of the available space regardless of the changes in the Rg values. One could label this as protein atoms moving unimpededly, while in both 208 and 209 complexes with ALK TK, much more restricted movement of the protein atoms are exhibited, especially in 208's. In both cases only one energetically prominent basin is discovered, around 20Å Rg and 3Å RMSD for 208 and around 20.1Å Rg and 3.1 Å RMSD. However, 208's protein atoms, shows slightly more restricted movement compared to 209, which supports the previous analyses.

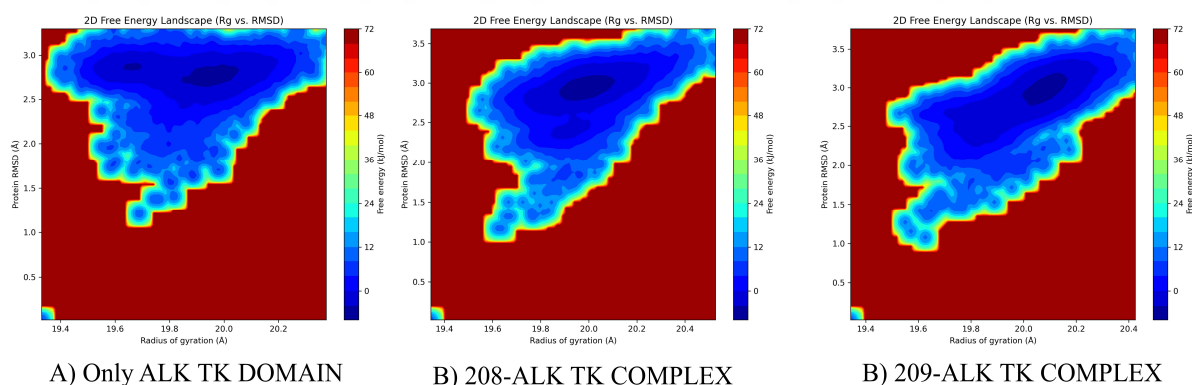


Figure 3.15 RMSD-Rg based FEL analysis of only tyrosine kinase domain, 208-ALK TK complex, and 209-ALK TK complex.

To further analyse the stability and binding affinity of the newly designed compounds 208 and 209, the same *g_mmpbsa* package was applied to the compounds on triplicate MD simulations. Compound 208's average binding energy over triplicate MD simulations was discovered as being in the similar range to the previously tested batch's best compound, ~ -14.5 kcal/mol, which differs from the compound 209's results that had the average binding energy of around -11.9 kcal/mol over the course of the

simulations(Fig 3.16), Therefore, this result indicates similar binding affinity towards to the ALK kinase domain as previously tested batch's best candidate, compound 5. When the MM/PBSA results are combined with RMSD, average distance results, PCA on binding region, clustering analysis , and Rg-RMSD based FEL analysis, it can be concluded that compound 208 has the potential to successfully inhibit the ALK kinase domain, while also compound 209 has almost similar potential, but their average distance to various key residues could be inferred as the not as effective as others due to interacting with different residues.

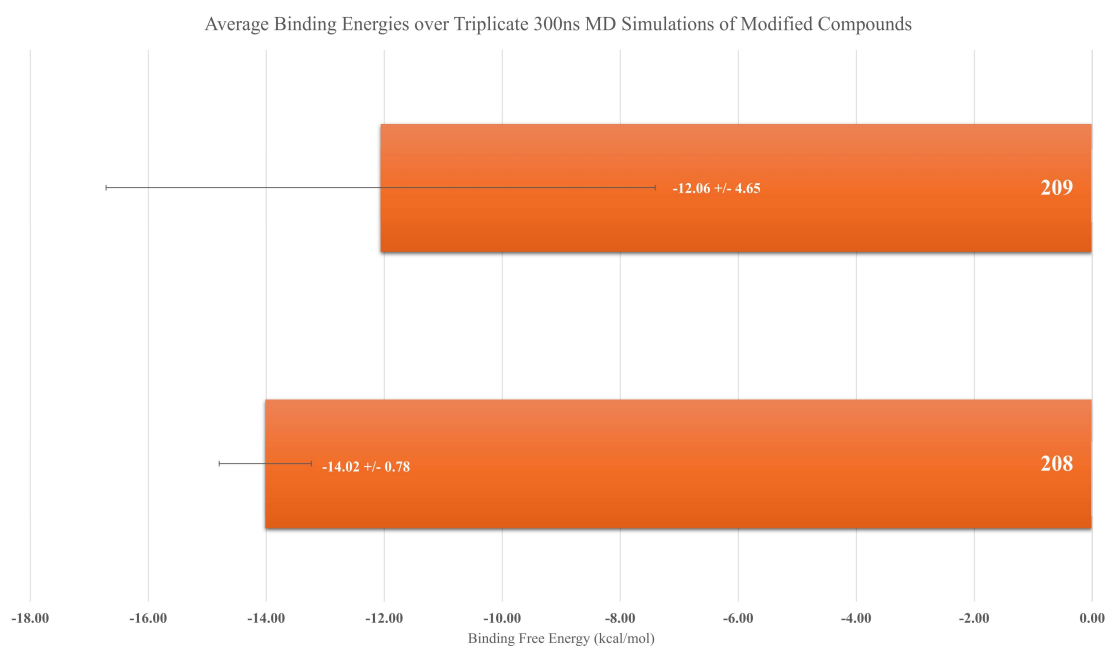


Figure 3.16 MM/PBSA average binding free energy graph of modified compounds from triplicate 300ns MD simulation

3.2 In silico Discovery of Novel ALK inhibitors from a Natural Product Database

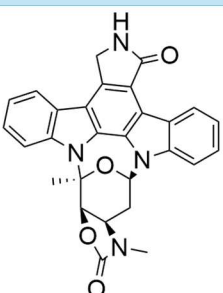
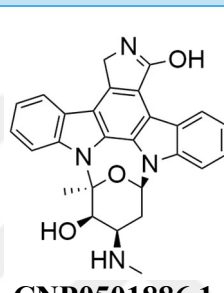
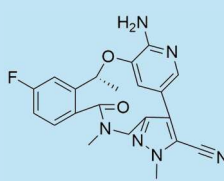
To ensure properly predict potential ALK inhibitors from a Natural Product (NP) Database, pre-docking equilibrium and analysis of it was performed. Subsequently after the selection of proper conformation from equilibrium MD, NPs from COCONUT database, which had more than 695k distinct NPs and with filtration according Lipinski's rule of five that lowered to count to 361732 molecules, were also prepared as reported in 2.2.2. From this preparation Docking studies had been commenced as outlined in 2.2.3.

3.2.1 Molecular Docking of the COCONUT NP Database against ALK tyrosine kinase domain

All the docking results from Smina, has been tabulated by selecting each molecule's their first binding mode, which are ordered by lowest binding affinity to highest in their allowed 10, and ordered by their lowest binding affinity to highest amongst all NPs. Lowest top 108 has been reported in the appendix table 2, which overall binding affinity has been found as lowest as -11.6 to highest -10.8 kcal/mol. The reason behind selecting top 108 instead of 100, was the fact that -10.8 kcal/mol binding energy starts from 80. NP with CNP0074120.0 COCONUT ID and ends with 108. NP with CNP0488686.1 COCONUT ID. Hence, one could understand that we wanted to include all of the last binding energy that is present in the top 100, which resulted in including 8 NP molecules to top 100.

After identifying the 108 natural products (NPs) with the lowest binding energy (i.e., highest binding affinity) for the ALK tyrosine kinase domain, subjecting all of them to further in silico studies, such as MD simulations and binding energy calculations, is impractical and resource intensive. Therefore, further selection was performed based on interaction analysis of the docking poses. The top five NPs were chosen if they formed at least one hydrogen bond with the hinge residues of the ALK tyrosine kinase domain, corresponding to residues 1197–1201. Details of the active region structure are provided in Section 1.3.1. Lastly, as reference molecule to compare the result Lorlatinib's results were also reported.

Table 3.5 The top 5 highest binding affinity of COCONUT NPs that also have at least one hydrogen bond with hinge residues

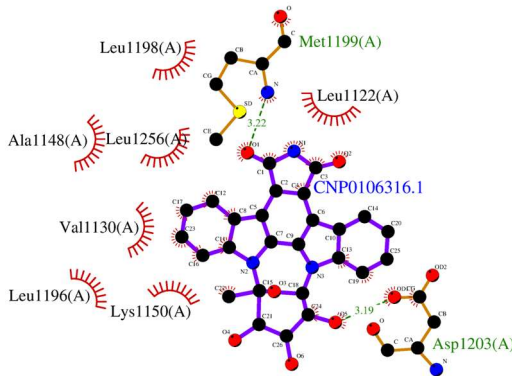
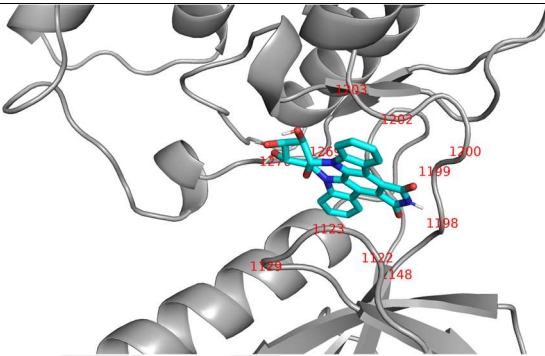
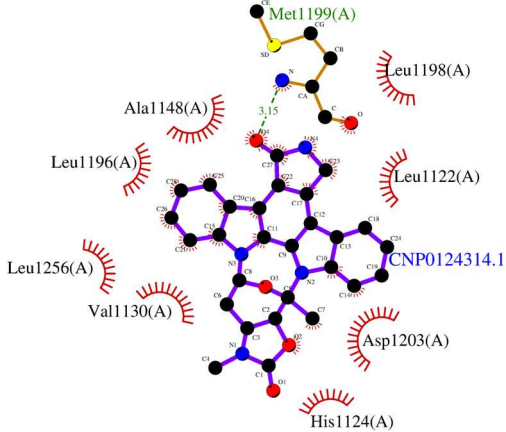
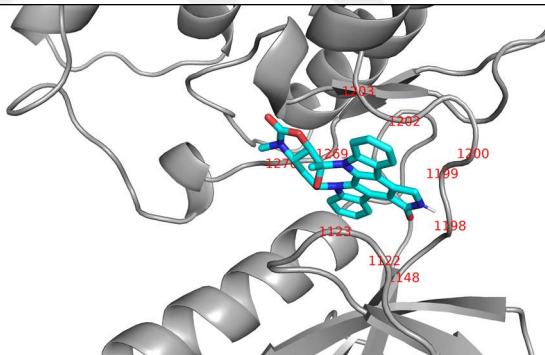
COCONUT NP ID & Structure	Binding Affinity (kcal/mol)	COCONUT NP ID & Structure	Binding Affinity (kcal/mol)
 CNP0106316.1	-11.6	 CNP0273044.2	-11.4
 CNP0124314.1	-11.4	 CNP0501886.1	-11.4
 CNP0273044.1	-11.4	 Lorlatinib	-8.6

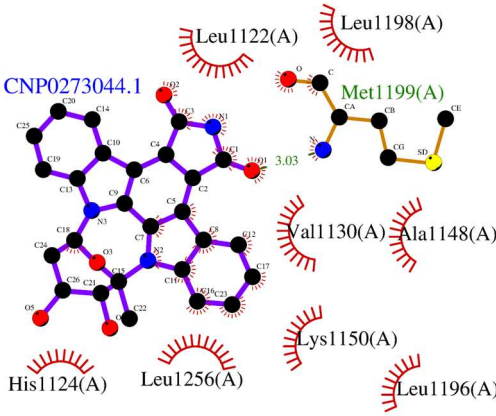
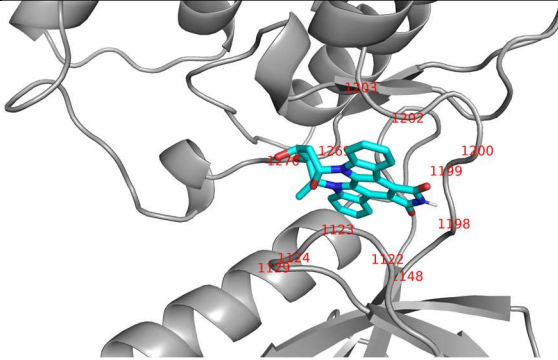
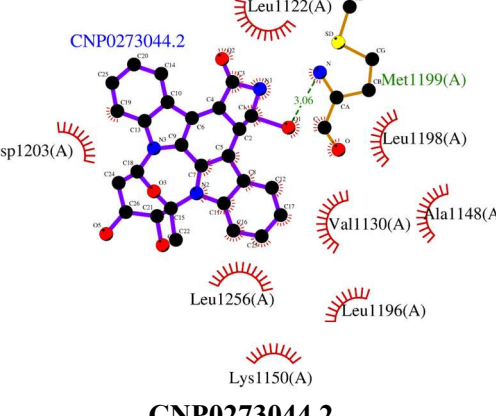
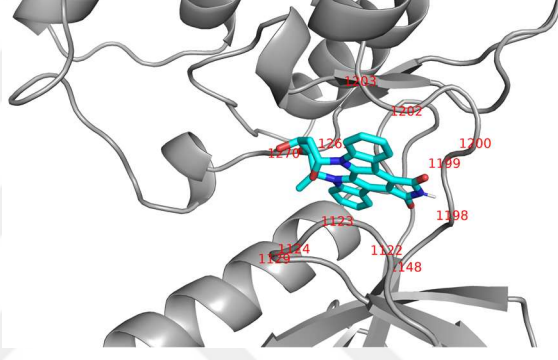
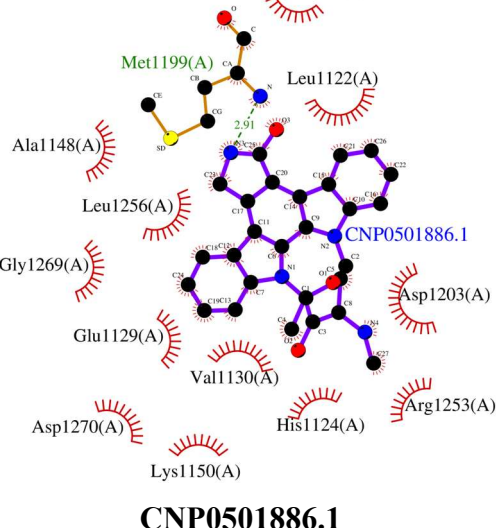
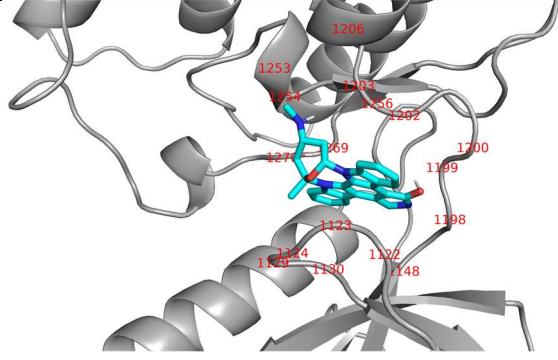
Top 5 NPslowest binding energy NPs that has at least 1 hydrogen bond with hinge residues were found as CNP0106316.1, CNP0124314.1, CNP0273044.1, CNP0273044.2, CNP0501886.1, with -11.6 kcal/mol binding energy for CNP0106316.1 and for the rest -11.4 kcal/mol (See Table 3.5). These NPs can be found in top 10 in the original tabulation that is reported in the appendix table 2. From structures of these NPs, one can easily see that they are closely related. In some of them, only difference of that is observed was the chirality, such as CNP0273044.1 and CNP0273044.2. At the same time, all five seems to be derivatives of bisindolylmaleimide, however some of the does not necessarily follow the maleimide structure, such as CNP0124314.1 and CNP510886.1

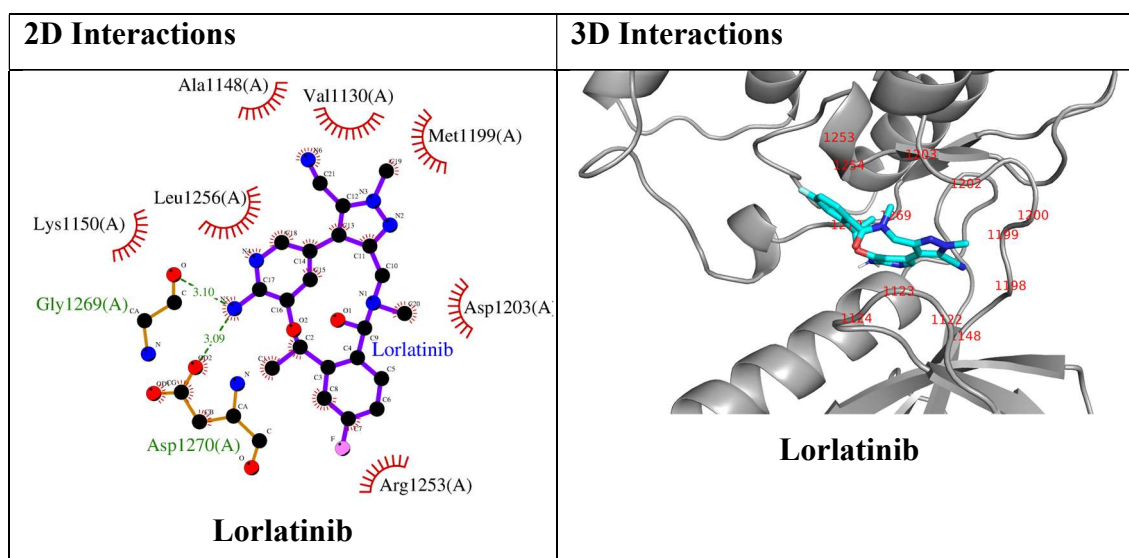
Additionally, reference molecule's binding energy was found to be -8.6. Lorlatinib's binding energy was found to be higher than what we have found from our docking results

by using environmental scoring function of the Smina. However, from this alone, one cannot report all the above compounds as superior inhibitor to Lorlatinib or even as an ALK inhibitor without further analysis and experiments. Hence, from docking poses binding interactions to the amino acids has been studied, and 2D and 3D interaction figures could be found at the following table. (See Table 3.6)

Table 3.6 2D and 3D interactions between top 5 NPs and ALK tyrosine kinase domain.

2D Interactions	3D Interactions
 CNP0106316.1	 CNP0106316.1
 CNP0124314.1	 CNP0124314.1

2D Interactions	3D Interactions
 CNP0273044.1	 CNP0273044.1
 CNP0273044.2	 CNP0273044.2
 CNP0501886.1	 CNP0501886.1



Footnote: In the 2D figures, dotted lines between residues' and the ligand's atoms indicate hydrogen bonds with respective lengths in angstrom while residues with the vertical red lines indicate the Van Der Waals interactions between the residue and the ligand. In the 3D figures, all amino acids that are in 5 angstroms of the centre of mass of each NP was labelled with their residue number in red.

As a first step, our analysis of binding poses started with the reference molecule, Lorlatinib. Our analysis predicted that Lorlatinib makes two hydrogen bonds with one of the activation segment's amino acids, DFG motif's aspartic acid numbered 1270, and with Glycine numbered 1269. However, when the literature was checked both Lorlatinib's original discover group's study with crystallography experiment that co-crystallized Lorlatinib with ALK tyrosine kinase domain and related study in 2021, Lorlatinib demonstrated binding poses that had hydrogen bonds with hinge residues, with binding the tyrosine kinase domain of the ALK while overall conformation of the domain as in the DFG-in state [127], [128].

This discrepancy could be attributed number of reasons, but we have narrowed it down to three possibilities: 1) selected post-equilibrium DFG conformation being slightly different, which in both studies Lorlatinib was discovered to bind DFG in state; 2) bias due to applying rigid docking; 3) inherent scoring function bias. First possibility can easily be checked by comparing our conformation DFG distance to hinge residue to the of known the distance of DFG-in state's conformation. For this end, 4CLI's the distance between phenylalanine's phenol and hinge residues, 1197, 1198 and 1199, was measured as 12.3Å, 14.6Å, and 13.8Å, respectively. When these distances compared to our conformation's distances, which was measured at 12.9Å, 14.8Å, and 14.1Å, respectively, one can easily see that they are very close to each other. The observed closeness validates our chosen conformation, ruling out the factor responsible for the previous discrepancy.

Second and third possibilities most likely reasons behind the discrepancy, however this thesis will compare the literature findings as well as our findings to our NP docking results to eliminate these possible biases in our experimental protocol.

Without delay due to hinge region selection, our top 5 compounds show one hydrogen bond with hinge residues, especially M1199. In all NPs this bond is detected however in the CNP0106316.1's another hydrogen bond with D1203 also present, which could mean the CNP0106316.1 is more stable than rest in this specific binding pose. Concurrently, these hydrogen bond with hinge region residues could also confirm our NPs being inhibitors due to interacting with same type of residues that orthodox tyrosine kinase inhibitors interact with, such as our literature findings of Lorlatinib, however in our case it was different and the reason behind is explained in the previous paragraph. In our pose of Lorlatinib, it shows interactions with activation segment, which is still indicative of an inhibition on an activity of tyrosine kinase domain, and combined with the fact that it shows two hydrogen bonding, it could be taken as immense indicative of stability of our specific binding pose. Regarding the rest of the NPs, similar Van der Waals interactions can be observed in all the top 5 NPs' interactions and the 3D figure of the docking poses at the tyrosine kinase domain of the ALK also supports this. Specifically, all the maleimide moieties of the top 5 NPs faces the same region of the domain, towards hinge region of the tyrosine kinase domain. When interpreted all together with the fact our top 5 compounds showed remarkable similarity to each other, it is not that surprising that the Van der Waals interactions show similar in all of them.

Ultimately from this analysis, one can say that CNP0106316.1 and Lorlatinib shows the most promise in all the tested NPs, but without further experiments this cannot be said confidently. Based on this observation, all of the NPs and as well as Lorlatinib, was subjected whole atom MD simulation of 500ns.

3.2.2 MD simulation of the selected top 5 NPs against tyrosine kinase domain of the ALK

To elucidate the inherent problems of the docking procedure, such as rigid docking, lack of water and ions atoms that always present at normal conditions of the protein, and as well as and forces between atoms changing the pose of both protein and the molecule across a time interval, MD simulation has been utilised. By analysing the calculated trajectories of the protein-NP complex in specified time interval, 500ns in this case, one can infer about potential stabilities of all of NPs, which allows for testing our hypotheses

from docking results, which two compounds, CNP0106316.1 and reference compound Lorlatinib, seemed more likely to stable than the rest.

Therefore, after the MD simulation, trajectories of each NP, has been corrected for periodic boundary conditions and first analysis initiated which was calculation of RMSD values between NP atoms and backbone atoms of protein. This analysis provides basic information of distances between two selections, in our case NP and backbone atoms, due to fact RMSD equation that is been utilised in the GROMACS can be summarized in a nutshell as square root of mean of squared distances between selections. Hence from this one can deduce that larger distances between atoms higher the RMSD values will be. Ultimately, one can use this information gather information about stability of the NP in the trajectory. Such as, if the docked pose of the NP is in fact stable, i.e. energetically most favourable in the docking box that was provided, NP will not be found more energetically favourable pose, the distance between atoms will not change across time. However, due to aforementioned points in the previous paragraph, docking poses is not always that accurate. For this reason, our analysis is begun with interpreting the RMSD figure. (Figure 3.17)

From the figure 3.17, one can readily interpret that even though all our NPs share similarity with each other, their stability at respective docking poses, despite sharing same relative position at the tyrosine kinase domain, differed significantly from one another. One of our selected NP, CNP0106316.1, demonstrates higher stability than the rest. This can be observed from the RMSD values across its 500ns MD simulation mostly not exceeding 4Å with mainly fluctuating between 2-4Å, see figure 3.17's part A. One can interpret this as CNP0106316.1 did not significantly relocate even with the presence of water and ion atoms, which strongly indicates stability in its docking position at the targeted tyrosine kinase. When this observation compared to Lorlatinib's RMSD plot from its MD simulation, CNP0106316.1's RMSD values seems to be lower than the Lorlatinib, which seems to fluctuate between 4-6Å. Hence, this can be considered as indicating that the CNP0106316.1's stability is more than Lorlatinib while on the tyrosine kinase domain of the ALK.

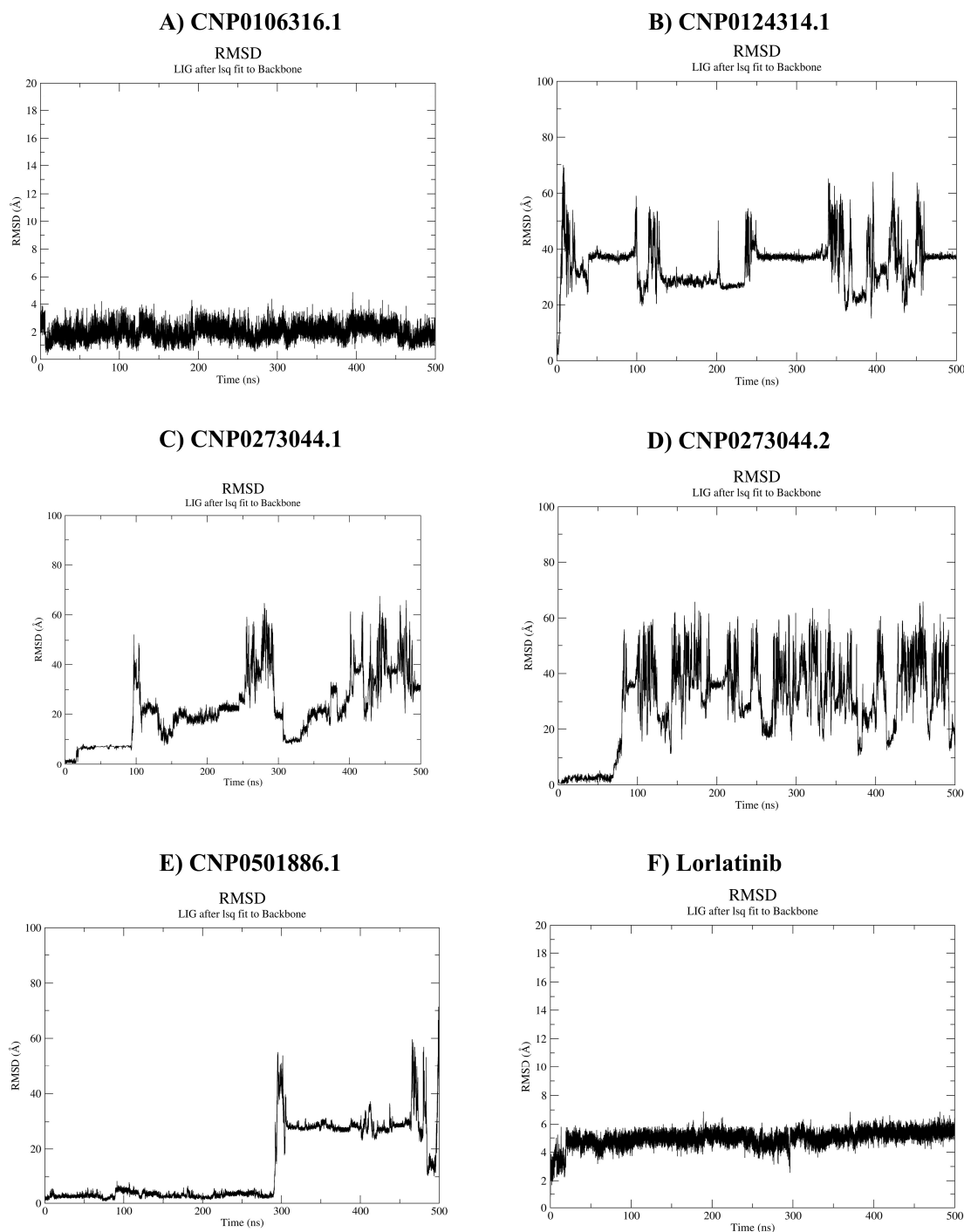


Figure 3.17 RMSD values between protein's backbone and NP atoms across 500ns MD simulation time for each selected NP

In the meantime, CNP0124314.1 shows quite unstability across whole 500ns MD simulation fluctuating between 20-70Å, starting from first 10ns never reaching equilibrium at any stage across of the simulation. Additionally, when the size of targeted active region of the tyrosine kinase considered, distances more than 10Å, can be easily

considered too high be considered “good” binding pose. However, since RMSD is the square root of average distances between selections, reaching the previous conclusion from RMSD analysis alone may be insufficient. Therefore, this conclusion requires more direct evidence based on fewer assumptions. Ultimately, one can achieve this goal by assessing the individual distances between amino acids, more specifically key ones, and the NP.

Next two NPs, CNP0273044.1 and CNP0273044.2, shows comparable trend in their RMSD plots. In the first 15-20% of the simulation time, they show rather low RMSD values, lower than 10Å that we can considered as a cut-off point due to explained reasons. But after a specific time, around 100ns for CNP0273044.1 and around 75ns for CNP0273044.2, they start to exhibit alternation in its location, across the ALK tyrosine kinase (TK) and NP MD simulation, and even with relatively long simulation, they never reach equilibrium again, which can be observed with continuing fluctuations across the whole simulation. Lastly, CNP0501886.1, is an interesting case, because for more than 250ns simulation, it illustrates rather promising RMSD values. However, as it can be seen from Fig. 3.17’s part E, after around 275ns mark, it displays fluctuation, even at 500ns mark it still did not reach equilibrium again.

Given these points, our predictions of higher stability of CNP0106316.1 from docking poses, has been supported by the plots of RMSD values between each NP and its ALK TK’s backbone atoms across the course of the MD simulation. Also, CNP0106316.1 featured lower RMSD values than the Lorlatinib. Before moving the assessing the specific distances between NP and key amino acids in the ALK TK domain, for each NP-ALK TK complex’s protein stability and possible conformation change in the protein, can be analysed. This analysis allows further supporting the stability of the complex.

To this end, RMSF analysis was conducted to backbone atoms of the ALK TK. (see Figure 3.18)

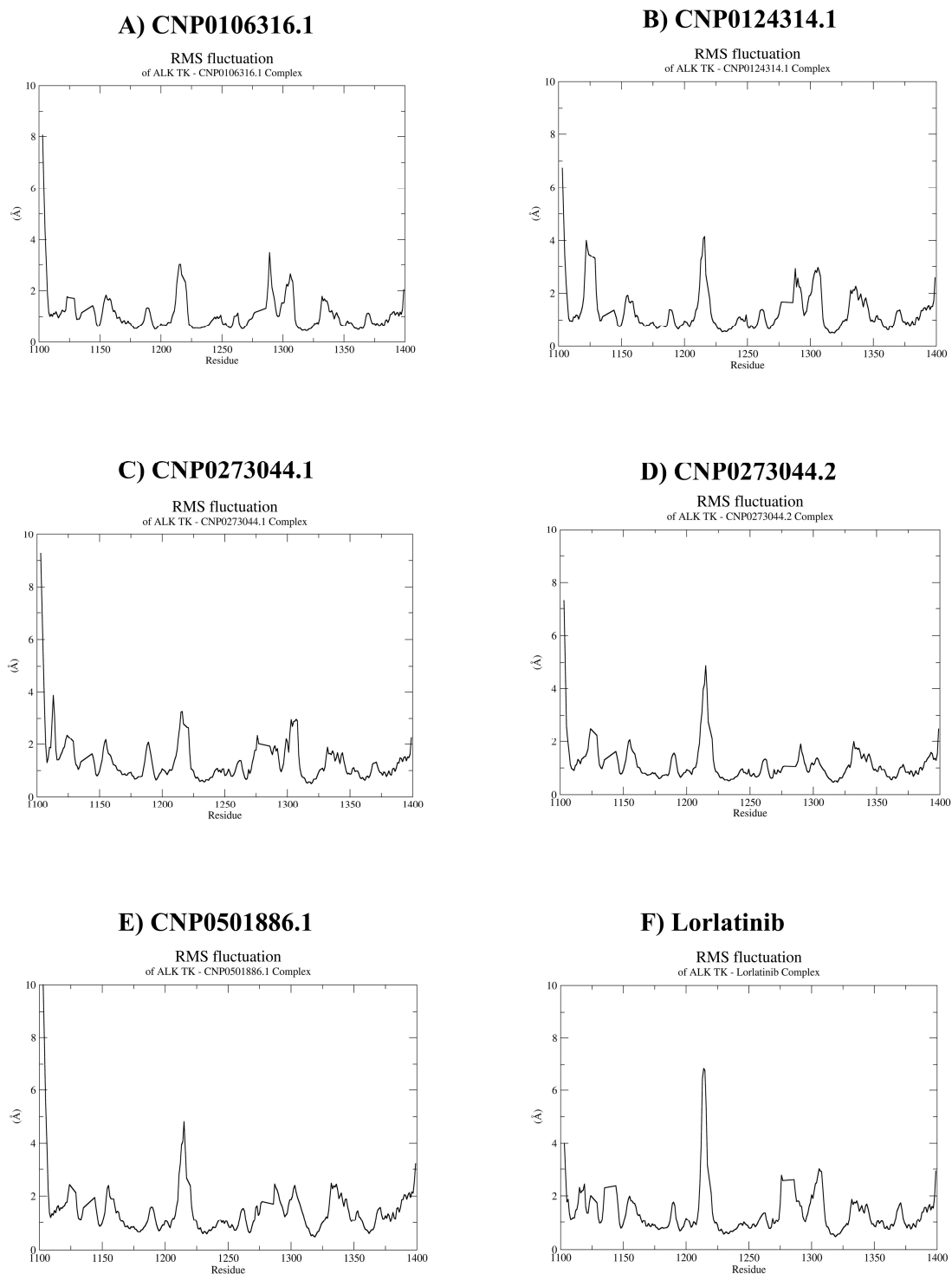


Figure 3.18 Backbone atoms' RMSF analysis of selected top 5 NP-ALK TK complexes

Due to our complex only consisting of TK domain of the ALK, X-axis start from 1100, C-terminal of protein is naturally fluctuating or layman terms moving around more, due to being at the one end of protein while similar trend observed in the N terminal residue as well, but with less magnitude of increase. However, an increase compared to

previous residues to N terminal is observed, hence the expected and observed trend. Granted to stability inferences from RMSD values, one could hypothesize that reason behind the unstableness of NPs, that caused alternation from their equilibrium poses could be emanated from possible conformation change in the ALK TK domain. If this is indeed the case, large fluctuations should be observed across many of the residues, that corresponds the changing the conformation of the domain. However, this observation was not detected in our analysis. There were no significant differences between all the complexes. Almost all, of the peaks and troughs in the RMSF plots were in the same residues, and when this observation combined with the fact that even magnitudes of fluctuations are all within 1-2Å each other and mostly less than 1Å. On that account, one cannot confidently report the reason of observed unstableness of the four of the NP was the conformation change in the ALK TK domain.

To further test the hypothesis from another angle, Rg analysis was conducted to measure changes in overall, meaning from all axes, compactness of TK domain. (see Figure 3.19) Hence if there is significant change in the conformation, one would expect to see change in the Rg value over course of the simulation.

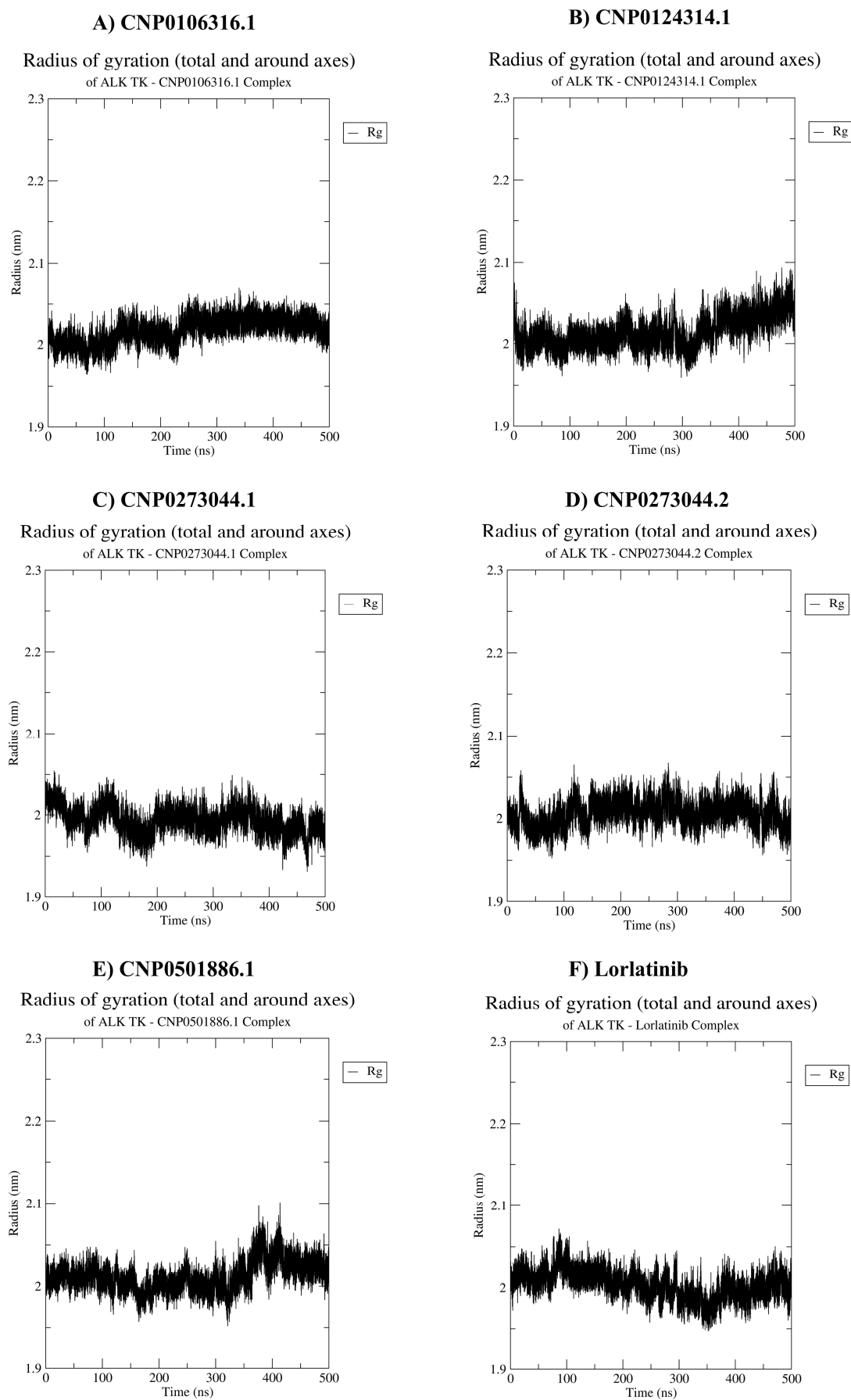


Figure 3.19 Protein atoms' Rg analysis of selected top 5 NP-ALK TK complexes.

All the tested NPs shows comparable Rg values across the course of the simulation. The values rested changed between 1.95 to 2.05 nm. Only two of NPs, CNP0124314.1 and CNP0501886.1, exceed 1.95-2.05 interval, which in both cases Rg value nears 2.1nm at some point of the MD simulation, around 500ns and 400ns marks, respectively. But at these marks, there is no sudden change in the RMSD value between backbone and NP atoms, which could be inferred as the similarly as RMSF conclusion. While rest of the plots all stay within the mentioned the interval, without drastic changes. But as a side note, CNP0106316.1 illustrating less change the rest of the Rg values of NP and as well as Lorlatinib. Ultimately, both analyses, RMSF and Rg, did not support the hypothesis of unstableness of NPs caused by the conformation change in ALK TK domain.

To further highlight the stableness of the CNP0106316.1, distance analysis between COM of NP and COM of previously selected residues has been done. (See figure 3.20) This analysis allows the make inferences about specific location of the NP, by comparing the distance to the key residues in the targeted site. Most compelling factor is that one can observe changes in the whole course of the simulation, such as if the NP is locating another site of the ALK TK, distance will change, and in the heatmap will change colour depending on the distances between key residues' COM and NPs' COM.

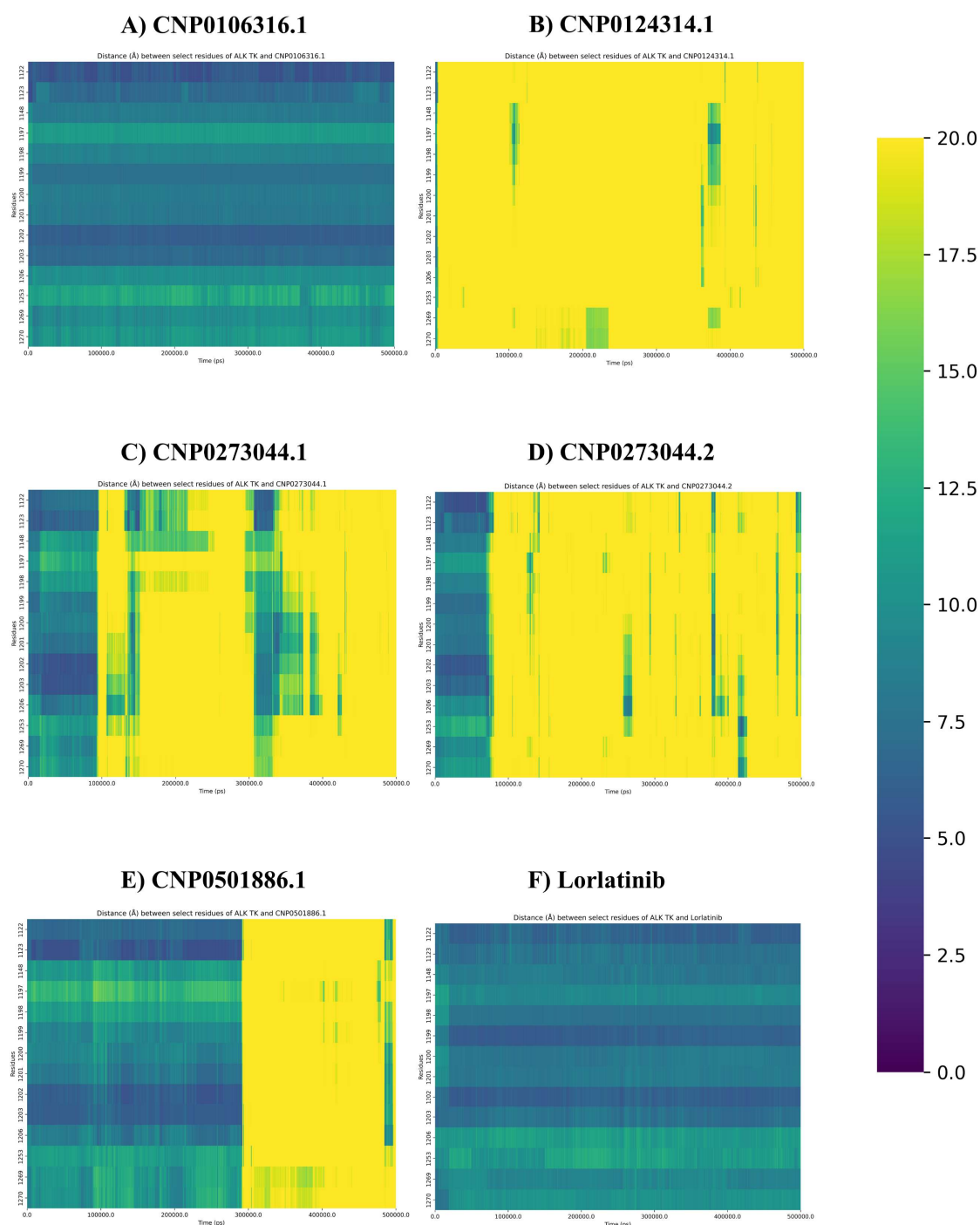


Figure 3.20 Average distance heatmaps of selected top 5 NPs.

With data represented at the Figure 3.20 combined with RMSD plots from Figure 3.17, one can infer about the whether the NP is indeed the docked location. The heatmaps are configured to change colour depending on the distance at the specific time. From the colour gradient legend, it can be observed that from 10Å to 0Å it will change colour from gradient of cyan colours to dark purple, and from 10Å to 20Å to will go to green to yellow while +20Å values will be rendered as yellow. In this case, heatmaps strongly correlate

the observations from RMSD figure. All the possibly unstable NPs, CNP0124314.1, CNP0273044.1, CNP0273044.2, and CNP0501886.1, exceeds 20Å at the similar time mark as the RMSD figure, and as it suggested in the discussion of the RMSD values, they did not reach equilibrium again. This observation is eminent in the heatmaps, which can be observed by the fact that sometimes NPs going closer to key residues and right after going back to +20Å again. However, in the case of CNP0106316.1's case, observed distances between COMs of between key residues and NP, supports the stability. The supporting observations are to hinge residues, CNP0106316.1's distances do not change drastically from start to finish. Specifically, stays around 10Å at the whole simulation. This seemingly "high" number can be explained by the fact that we are comparing COM, centre of mass, between two selections, residues and NP, and due to not comparing distances between first selection's closest atom to other selection's closest atom, instead we are comparing COM of a molecule to COM of another molecule, which adequately explains the "high" distance values. To prevent this issue in the analysis, examining the change in the distances, meaning comparing within itself instead directly analysing distance value itself, makes the most sense. Lastly, CNP0106316.1's heatmaps also shows similarities between Lorlatinib, specifically at the hinge region. Especially after the around 25ns mark of its MD simulation, which can be considered that it stays the same to finish of the simulation. This 25ns marks change also observed in the related RMSD plots as well, by fluctuating the same RMSD interval to end after 25 ns. In short, it could be reported that Lorlatinib ALK TK complex MD simulation predicts that our reported binding pose changes into different one, which more closely resembles to literature reported ones inferred by the fact that is closer to hinge region amino acids than what is started with it. Altogether, up until now our analysis of MD simulations, strongly suggests that CNP0106316's is the most promising candidate in our selected top 5 NPs.

Another important point the stability of the binding poses is the hydrogen bonding, as reported in docking poses, all selected NPs have at least 1 hydrogen bond with hinge region while CNP0106316.1 having two hydrogen bonds even though one of them is not to hinge region amino acids. By looking at number of hydrogen bonds between ALK TK and NP, one can further ascertain whether the NPs were able to keep their hydrogen bonds in their original pose. (see Figure 3.21) However, by looking at the previous analyses, one can readily report that almost all of them could not hold on their hydrogen bonds which can be inferred from distance between NP and key residues alone.

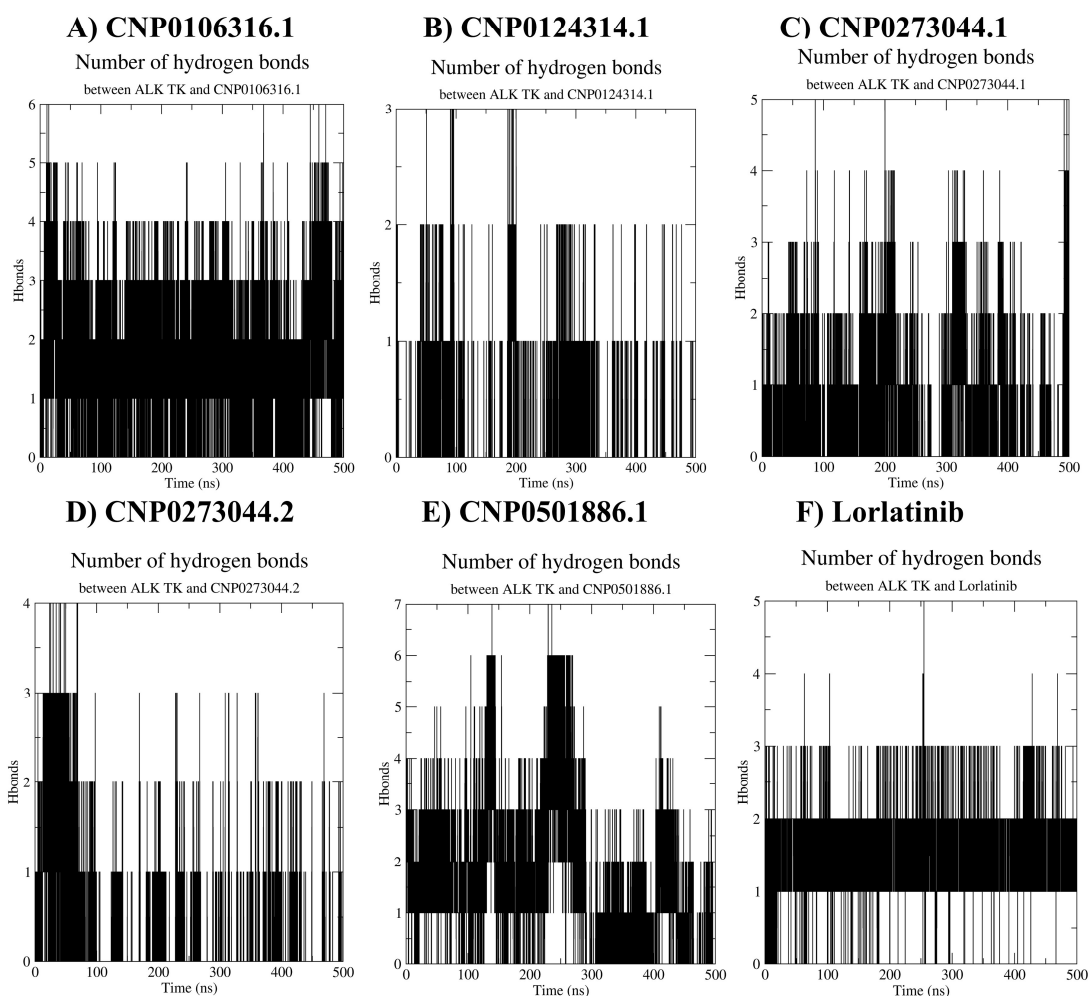


Figure 3.21 Number of hydrogen bonds between NPs and ALK TK domain over the course of the simulation.

As suspected, four of the NPs, all except for CNP0106316, were not able consistently hold on the hydrogen bonds, which could explain the observed unstableness of the NPs. In the case CNP0106316.1, in all of its simulation it demonstrates at least one hydrogen bond while at most of the time its two and with less frequently it jumps to 2-3. When observation is compared to Lorlatinib's case, it could explain the observed differences in the RMSD and very comparable distance heatmap results. Number of hydrogen bonds between Lorlatinib and ALK TK domain almost stays at least two most simulation, while less frequently to lowering one or going +2.

Therefore, as suspected from earlier analyses of MD and docking poses, CNP0106316.1 is the most promising NP in selected top 5 NPs.

3.2.3 MM/PBSA Average Binding Energy Calculation Between NPs and ALK TK over the Course of MD Simulation

In Figure 3.22, the average binding free energies across the whole simulation have been reported.

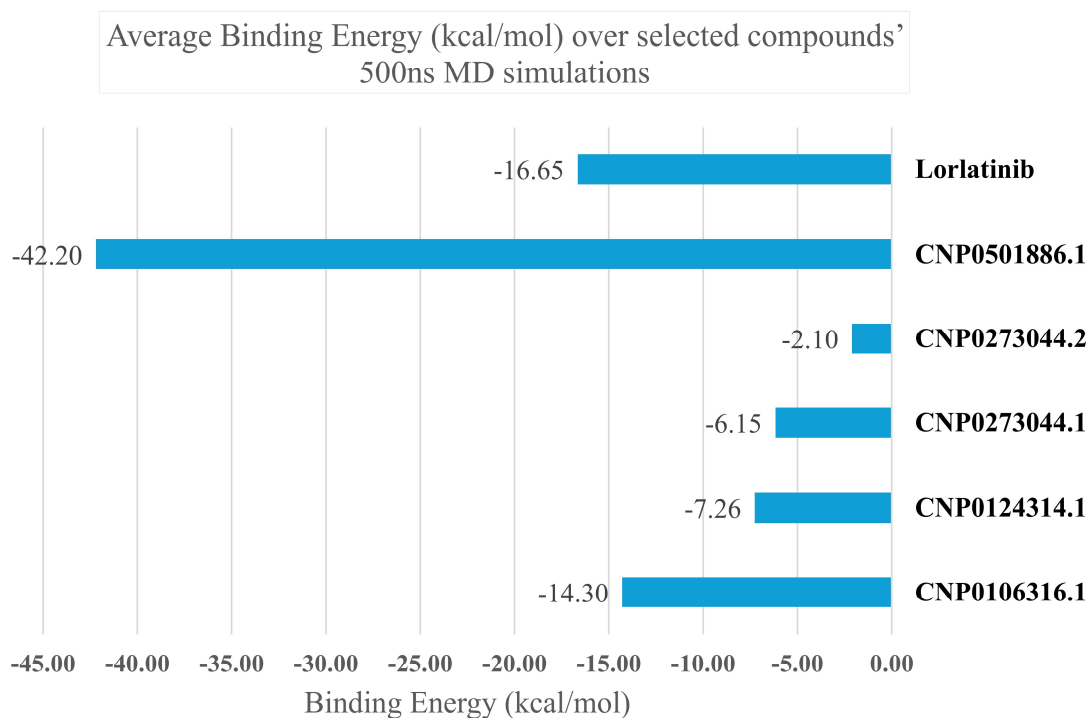


Figure 3.22 MM/PBSA average binding free energies of the selected top 5 compounds and Lorlatinib.

The average binding energy of CNP0106316.1 and Lorlatinib is very close, -14.30 and -16.65, respectively. Combined with previous analyses, this indicates that CNP0106316.1 have the most potential to be ALK TK kinase inhibitor, maybe same as Lorlatinib maybe better potency, but without further experiment, especially in vitro experiment, such as kinase assays, etc. Reporting a promise is the best one can do. Besides this observation, in CNP0501886.1's binding energy is significantly lower than any other test NP and Lorlatinib with -42.20. But due to its unstableness that has been observed at the target location, one cannot report CNP0501886.1's as promising candidate for ALK inhibition. However, as reported earlier, without in vitro and in vivo validations, all of this are just predictions or in more harsh terms can almost reported as pure speculations. But purpose of the in silico methods of drug discovery is reducing cost of time money that is needed for discovering and developing drugs, and in this case, by applying number of in silico methods, filtration according Lipinski's rule of five, and logic based selections,

we can strongly recommend one NP that has the most potential to inhibit ALK TK domain from our COCONUT NP database.



Chapter 4

Conclusions and Future Prospects

4.1 Conclusions

Cancer is leading cause of death in the 57 countries [129]. Moreover, lung cancer is the leading cause of cancer related mortality [130]. ALK positive lung cancer takes up 5-7 percent of all lung cancer and there are several ALK dependent cancers alongside with the lung cancer. To combat this, FDA approved number of drugs for ALK positive lung cancers. Despite this, a recent study highlighted that around half of the ALK-positive NSCLC patients is going through disease progression in response to first-line Alectinib, which is a second-generation ALK inhibitor and containing carbazole structure, as a predictable result of cancer's disposition and typical appearance of resistance. Therefore, this thesis primarily tried two distinct methods to discover and develop new and novel drug compounds and motifs to target ALK positive cancers.

Firstly, we designed novel α -carboline structures for targeting ALK kinase domain to be effective for number of ALK positive cancers, and tested these compounds via utilizing in silico approach, comprising of molecular docking, molecular dynamics simulations, MM/PBSA binding free energy calculation, and essential dynamics analysis. As results of the analysis two promising candidate emerged compound 208 and 209, when all the results were considered holistically from AutoDock Vina's binding affinities to MD simulation results. Compound 208 showed the best binding free energy during 300 ns molecular dynamics simulation. Moreover, compound 208 had more stable RMSD values with 2-3 Angstrom and it had significantly less distance to one of the hinge residues, M1199, compared to compound 209. Considering the ease of synthesis of the target compounds, we may conclude that compound 208 could be considered as a potential ALK inhibitor.

Secondly, we had gathered more than 695k distinct NP from COCONUT NP database and around 361732 NP have met the Lipinski's rule of five filtration criteria. By

utilizing docking via Smina, MD simulation via GROMACS, and MM/PBSA calculations via *g_mmpbsa*. We have screened filtered NPs against ALK TK domain, 8ARJ. Among docking results, top 108 were tabulated by ordering their predicted binding affinity from lowest to highest. From these 108, top five were further selected by carefully evaluating their interactions in their binding pose the docked location. In effect, selections encompassed basically selecting NPs that had at least one hydrogen bond with hinge residues of ALK TK. Furthermore, selected top five's stability at the docked location, further evaluated by MD simulation by analysing RMSD, RSMF, Rg, average distances between NP and key residues and number of hydrogen bonds across the simulation. Concurrently, each NPs binding free energy towards ALK TK across the simulation had been calculated by MM/PBSA method. From inferences of those results only one NP from five candidates, has emerged as strong candidate as ALK TK inhibitor, numbered as CNP0106316.1.

4.2 Societal Impact and Contribution to Global Sustainability

The thesis makes use of in silico approaches to identify novel compounds that target ALK TK domain using in silico approaches. Seeing this, it can be reported that research output of this study could result in direct societal benefits. It could achieve this by accelerating time necessary and reducing resource intensity that normally is required for developing effective cancer therapies with less side effects if possible. In fact, this goal we are trying to accomplish directly supports Sustainable Development Goals 3 (SDG), good health and well-being. By accelerating early-stage drug discovery, this research could improve access to potential therapies in low- and middle-income countries. These regions often lack the infrastructure for long-term drug development cycles. Therefore, a computationally guided, lower-cost strategy contributes to reducing health inequities and expands global access to cancer treatment, which ends up further reinforcing the health equity aspects of SDG 3. Besides these, our research goal is setup to lower the costs and time of drug discovery and development, by reducing the need for depending on HTS both in vitro and in vivo even though it still necessary for validating in silico results, overall resource needed significantly lower than what one would normally require for HTS. Furthermore, the computational workflow developed in this thesis is designed to be

reproducible and transferable, enabling other researchers, including those in resource-limited academic or clinical settings, to apply these methodologies without extensive laboratory infrastructure. This promotes wider scientific collaboration, encourages open research practices, and facilitates global knowledge sharing. In effect, our goal is aligned with responsible consumption and production goal of SDGs (SDG 12). In addition, by creating a cost-effective and adaptable pipeline, the research contributes to more equitable participation in drug discovery, allowing scientists worldwide to engage in early-stage development of cancer therapies. This can accelerate the translation of computational discoveries into experimental validation, thereby improving efficiency and inclusivity in the global scientific community.

4.3 Future Prospects

Ultimately, to fully report the compounds 208, 209, or CNP0106316.1 as potent and selective ALK TK inhibitor, in vitro and in vivo validations are required. In summary, this thesis could help to pave the way for discovering or designing and developing new generation of ALK targeting inhibitors.

BIBLIOGRAPHY

- [1] 'National Cancer Institute. NCI Thesarus'. Accessed: May 27, 2025. [Online]. Available: <https://evsexplore.semantics.cancer.gov>
- [2] D. C. Young, *Properties that make a molecule a good drug*. Hoboken, New Jersey: John Wiley & Sons, Inc. , 2009.
- [3] I. Doytchinova, 'Drug Design—Past, Present, Future', *Molecules*, vol. 27, no. 5, p. 1496, Feb. 2022, doi: 10.3390/molecules27051496.
- [4] B. E. Blass, *Basic Principles of Drug Discovery and Development*, 2nd ed. Academic Press, 2021.
- [5] M. J. Waring *et al.*, 'An analysis of the attrition of drug candidates from four major pharmaceutical companies', *Nat Rev Drug Discov*, vol. 14, no. 7, pp. 475–486, Jul. 2015, doi: 10.1038/NRD4609.
- [6] D. B. Fogel, 'Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review', *Contemp Clin Trials Commun*, vol. 11, pp. 156–164, Sep. 2018, doi: 10.1016/J.CONCTC.2018.08.001.
- [7] R. Prentis, Y. Lis, and S. Walker, 'Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964-1985)', *Br J Clin Pharmacol*, vol. 25, no. 3, p. 387, 1988, doi: 10.1111/J.1365-2125.1988.TB03318.X.
- [8] T. A. Ban, 'The role of serendipity in drug discovery', *Dialogues Clin Neurosci*, vol. 8, no. 3, pp. 335–344, 2006, doi: 10.31887/DCNS.2006.8.3/TBAN.
- [9] S. Prasad, S. C. Gupta, and B. B. Aggarwal, 'Serendipity in Cancer Drug Discovery: Rational or Coincidence?', *Trends Pharmacol Sci*, vol. 37, no. 6, pp. 435–450, Jun. 2016, doi: 10.1016/J.TIPS.2016.03.004.
- [10] M. J. R. Desborough and D. M. Keeling, 'The aspirin story - from willow to wonder drug', *Br J Haematol*, vol. 177, no. 5, pp. 674–683, Jun. 2017, doi: 10.1111/BJH.14520.
- [11] Z. Guo, 'The modification of natural products for medical use', *Acta Pharm Sin B*, vol. 7, no. 2, pp. 119–136, Mar. 2017, doi: 10.1016/J.APSB.2016.06.003.
- [12] M. Entzeroth, H. Flotow, and P. Condron, 'Overview of high-throughput screening', *Curr Protoc Pharmacol*, vol. Chapter 9, no. SUPPL. 44, 2009, doi: 10.1002/0471141755.PH0904S44.
- [13] M. Kontoyianni, 'Docking and Virtual Screening in Drug Discovery', *Methods Mol Biol*, vol. 1647, pp. 255–266, 2017, doi: 10.1007/978-1-4939-7201-2_18.
- [14] P. Van Antwerpen and K. Zouaoui Boudjeltia, 'Rational drug design applied to myeloperoxidase inhibition', *Free Radic Res*, vol. 49, no. 6, pp. 711–720, Jun. 2015, doi: 10.3109/10715762.2015.1027201.
- [15] Y. An *et al.*, 'Updated Review on Developing Small Molecule Kinase Inhibitors Using Computer-Aided Drug Design Approaches', *Int. J. Mol. Sci*, p. 24, 2023, doi: 10.3390/ijms241813953.
- [16] P. U. Kulkarni, H. Shah, and V. K. Vyas, 'Hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) Simulation: A Tool for Structure-Based Drug Design and

- Discovery', *Mini Rev Med Chem*, vol. 22, no. 8, pp. 1096–1107, Oct. 2022, doi: 10.2174/1389557521666211007115250.
- [17] L. G. Ferreira, R. N. Dos Santos, G. Oliva, and A. D. Andricopulo, 'molecules Molecular Docking and Structure-Based Drug Design Strategies', *Molecules*, vol. 20, pp. 13384–13421, 2015, doi: 10.3390/molecules200713384.
 - [18] A. Parise, S. Cresca, and A. Magistrato, 'Molecular dynamics simulations for the structure-based drug design: targeting small-GTPases proteins', *Expert Opin Drug Discov*, vol. 19, no. 10, 2024, doi: 10.1080/17460441.2024.2387856.
 - [19] A. Waterhouse *et al.*, 'SWISS-MODEL: homology modelling of protein structures and complexes', *Nucleic Acids Res*, vol. 46, no. W1, pp. W296–W303, Jul. 2018, doi: 10.1093/NAR/GKY427.
 - [20] A. R. Amorim, G. F. D. Zafalon, A. de G. Contessoto, C. R. Valêncio, and L. M. Sato, 'Metaheuristics for multiple sequence alignment: A systematic review', *Comput Biol Chem*, vol. 94, Oct. 2021, doi: 10.1016/J.COMPBIOLCHEM.2021.107563.
 - [21] T. Wang, M. Bin Wu, J. P. Lin, and L. R. Yang, 'Quantitative structure-activity relationship: promising advances in drug discovery platforms', *Expert Opin Drug Discov*, vol. 10, no. 12, pp. 1283–1300, Dec. 2015, doi: 10.1517/17460441.2015.1083006.
 - [22] A. David, S. Islam, E. Tankhilevich, and M. J. E. Sternberg, 'The AlphaFold Database of Protein Structures: A Biologist's Guide', *J Mol Biol*, vol. 434, no. 2, Jan. 2022, doi: 10.1016/J.JMB.2021.167336.
 - [23] J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli, 'AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings', *J Chem Inf Model*, vol. 61, no. 8, pp. 3891–3898, Aug. 2021, doi: 10.1021/ACS.JCIM.1C00203/SUPPL_FILE/CI1C00203_SI_002.ZIP.
 - [24] P. T. Lang *et al.*, 'DOCK 6: combining techniques to model RNA-small molecule complexes', *RNA*, vol. 15, no. 6, pp. 1219–1230, Jun. 2009, doi: 10.1261/RNA.1563609.
 - [25] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor, 'Improved protein-ligand docking using GOLD', *Proteins*, vol. 52, no. 4, pp. 609–623, Sep. 2003, doi: 10.1002/PROT.10465.
 - [26] R. Chen, L. Li, and Z. Weng, 'ZDOCK: an initial-stage protein-docking algorithm', *Proteins*, vol. 52, no. 1, pp. 80–87, Jul. 2003, doi: 10.1002/PROT.10389.
 - [27] D. B. Singh, Ed., *Computer-Aided Drug Design*. Singapore: Springer Singapore, 2020. doi: 10.1007/978-981-15-6815-2.
 - [28] X. Wu, L. Y. Xu, E. M. Li, and G. Dong, 'Application of molecular dynamics simulation in biomedicine', *Chem Biol Drug Des*, vol. 99, no. 5, pp. 789–800, May 2022, doi: 10.1111/CBDD.14038.
 - [29] D. A. Case *et al.*, 'AmberTools', *J Chem Inf Model*, vol. 63, no. 20, pp. 6183–6191, Oct. 2023, doi: 10.1021/ACS.JCIM.3C01153/ASSET/IMAGES/LARGE/CI3C01153_0002.JPEG.

- [30] M. J. Abraham *et al.*, ‘GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers’, *SoftwareX*, vol. 1–2, pp. 19–25, Sep. 2015, doi: 10.1016/J.SOFTX.2015.06.001.
- [31] M. T. Nelson *et al.*, ‘NAMD: a Parallel, Object-Oriented Molecular Dynamics Program’, *The International Journal of Supercomputer Applications and High Performance Computing*, vol. 10, no. 4, pp. 251–268, 1996, doi: 10.1177/109434209601000401.
- [32] B. R. Brooks *et al.*, ‘CHARMM: the biomolecular simulation program’, *J Comput Chem*, vol. 30, no. 10, pp. 1545–1614, Jul. 2009, doi: 10.1002/JCC.21287.
- [33] A. P. Thompson *et al.*, ‘LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales’, *Comput Phys Commun*, vol. 271, p. 108171, Feb. 2022, doi: 10.1016/J.CPC.2021.108171.
- [34] X. Qiao, X. Wu, S. Chen, M. M. Niu, H. Hua, and Y. Zhang, ‘Discovery of novel and potent dual-targeting AXL/HDAC2 inhibitors for colorectal cancer treatment via structure-based pharmacophore modelling, virtual screening, and molecular docking, molecular dynamics simulation studies, and biological evaluation’, *J Enzyme Inhib Med Chem*, vol. 39, no. 1, 2024, doi: 10.1080/14756366.2023.2295241.
- [35] C. Zhang *et al.*, ‘Design, Synthesis, Molecular Docking Analysis and Biological Evaluations of 4-[(Quinolin-4-yl)amino]benzamide Derivatives as Novel Anti-Influenza Virus Agents’, *Int J Mol Sci*, vol. 23, no. 11, Jun. 2022, doi: 10.3390/IJMS23116307.
- [36] S. Eshtiaghi, R. Nazari, and M. Fasihi-Ramandi, ‘Molecular Docking, Anti-Biofilm & Antibacterial Activities and Therapeutic Index of mCM11 Peptide on *Acinetobacter baumannii* Strains’, *Curr Microbiol*, vol. 80, no. 6, Jun. 2023, doi: 10.1007/S00284-023-03217-Z.
- [37] F. Ardito, M. Giuliani, D. Perrone, G. Troiano, and L. Lo Muzio, ‘The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review)’, *Int J Mol Med*, vol. 40, no. 2, pp. 271–280, Aug. 2017, doi: 10.3892/IJMM.2017.3036.
- [38] M. M. Attwood, D. Fabbro, A. V. Sokolov, S. Knapp, and H. B. Schiöth, ‘Trends in kinase drug discovery: targets, indications and inhibitor design’, *Nat Rev Drug Discov*, vol. 20, no. 11, pp. 839–861, Nov. 2021, doi: 10.1038/s41573-021-00252-y.
- [39] R. Roskoski, ‘Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes’, *Pharmacol Res*, vol. 103, pp. 26–48, Jan. 2016, doi: 10.1016/J.PHRS.2015.10.021.
- [40] R. Roskoski, ‘A historical overview of protein kinases and their targeted small molecule inhibitors’, *Pharmacol Res*, vol. 100, pp. 1–23, Oct. 2015, doi: 10.1016/J.PHRS.2015.07.010.
- [41] F. Bray, M. Laversanne, E. Weiderpass, and I. Soerjomataram, ‘The ever-increasing importance of cancer as a leading cause of premature death worldwide’, *Cancer*, vol. 127, no. 16, pp. 3029–3030, Aug. 2021, doi: 10.1002/CNCR.33587.

- [42] H. Sung *et al.*, ‘Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries’, *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/CAAC.21660.
- [43] S. W. Morris *et al.*, ‘Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin’s lymphoma’, *Science*, vol. 263, no. 5151, pp. 1281–1284, Mar. 1994, doi: 10.1126/SCIENCE.8122112.
- [44] A. E. M. van Erp, M. H. S. Hillebrandt-Roeffen, L. van Houdt, E. D. G. Fleuren, W. T. A. van der Graaf, and Y. M. H. Versleijen-Jonkers, ‘Targeting Anaplastic Lymphoma Kinase (ALK) in Rhabdomyosarcoma (RMS) with the Second-Generation ALK Inhibitor Ceritinib’, *Target Oncol*, vol. 12, no. 6, pp. 815–826, Dec. 2017, doi: 10.1007/s11523-017-0528-z.
- [45] N. Aygun, ‘Biological and Genetic Features of Neuroblastoma and Their Clinical Importance’, *Curr Pediatr Rev*, vol. 14, no. 2, pp. 73–90, Jan. 2018, doi: 10.2174/1573396314666180129101627.
- [46] B. Golding, A. Luu, R. Jones, and A. M. Vilorio-Petit, ‘The function and therapeutic targeting of anaplastic lymphoma kinase (ALK) in non-small cell lung cancer (NSCLC)’, Feb. 19, 2018, *BioMed Central Ltd*. doi: 10.1186/s12943-018-0810-4.
- [47] A. V. Reshetnyak *et al.*, ‘Mechanism for the activation of the anaplastic lymphoma kinase receptor’, *Nature*, vol. 600, no. 7887, pp. 153–157, Dec. 2021, doi: 10.1038/s41586-021-04140-8.
- [48] J. L. Schneider, J. J. Lin, and A. T. Shaw, ‘ALK-positive lung cancer: a moving target’, *Nat Cancer*, vol. 4, no. 3, pp. 330–343, Feb. 2023, doi: 10.1038/s43018-023-00515-0.
- [49] R. Roskoski, ‘Anaplastic lymphoma kinase (ALK): Structure, oncogenic activation, and pharmacological inhibition’, Feb. 2013. doi: 10.1016/j.phrs.2012.11.007.
- [50] B. Hallberg and R. H. Palmer, ‘The role of the ALK receptor in cancer biology’, *Annals of Oncology*, vol. 27, pp. iii4–iii15, Sep. 2016, doi: 10.1093/ANNONC/MDW301.
- [51] J. Zhang, P. L. Yang, and N. S. Gray, ‘Targeting cancer with small molecule kinase inhibitors’, Jan. 2009. doi: 10.1038/nrc2559.
- [52] S. S. Taylor and A. P. Kornev, ‘Protein Kinases: Evolution of Dynamic Regulatory Proteins’, *Trends Biochem Sci*, vol. 36, no. 2, p. 65, Feb. 2011, doi: 10.1016/J.TIBS.2010.09.006.
- [53] E. L. Kwak *et al.*, ‘Anaplastic Lymphoma Kinase Inhibition in Non-Small-Cell Lung Cancer’, *New England Journal of Medicine*, vol. 363, no. 18, pp. 1693–1703, Oct. 2010, doi: 10.1056/nejmoa1006448.
- [54] L. Wang and W. Wang, ‘Safety and efficacy of anaplastic lymphoma kinase tyrosine kinase inhibitors in non-small cell lung cancer (Review)’, *Oncol Rep*, vol. 45, no. 1, Jan. 2020, doi: 10.3892/or.2020.7851.
- [55] X. R. Zhang, P. N. Chien, S. Y. Nam, and C. Y. Heo, ‘Anaplastic Large Cell Lymphoma: Molecular Pathogenesis and Treatment’, *Cancers (Basel)*, vol. 14, no. 7, Apr. 2022, doi: 10.3390/CANCERS14071650.

- [56] J. Duyster, R.-Y. Bai, and S. W. Morris, 'Translocations involving anaplastic lymphoma kinase (ALK)', Accessed: May 30, 2025. [Online]. Available: www.nature.com/onc
- [57] S. S. Zhang, M. Nagasaka, V. W. Zhu, and S.-H. I. Ou, 'Going beneath the tip of the iceberg. Identifying and understanding EML4-ALK variants and TP53 mutations to optimize treatment of ALK fusion positive (ALK+) NSCLC', *Lung Cancer*, vol. 158, pp. 126–136, 2021, doi: 10.1016/j.lungcan.2021.06.012.
- [58] D. F. Heigener and M. Reck, 'Crizotinib', *Recent Results Cancer Res*, vol. 211, pp. 57–65, 2018, doi: 10.1007/978-3-319-91442-8_4.
- [59] J. Jiang *et al.*, 'ALK inhibitors in ALK-rearranged non-small cell lung cancer with and without brain metastases: systematic review and network meta-analysis', *BMJ Open*, vol. 12, p. 60782, 2022, doi: 10.1136/bmjopen-2022-060782.
- [60] T. Mok *et al.*, 'Updated overall survival and final progression-free survival data for patients with treatment-naïve advanced ALK-positive non-small-cell lung cancer in the ALEX study', *Ann Oncol*, vol. 31, no. 8, pp. 1056–1064, Aug. 2020, doi: 10.1016/J.ANNONC.2020.04.478.
- [61] A. T. Shaw *et al.*, 'Lorlatinib in non-small-cell lung cancer with ALK or ROS1 rearrangement: an international, multicentre, open-label, single-arm first-in-man phase 1 trial', *Lancet Oncol*, vol. 18, no. 12, pp. 1590–1599, Dec. 2017, doi: 10.1016/S1470-2045(17)30680-0.
- [62] A. T. Shaw *et al.*, 'First-Line Lorlatinib or Crizotinib in Advanced ALK-Positive Lung Cancer', *N Engl J Med*, vol. 383, no. 21, pp. 2018–2029, Nov. 2020, doi: 10.1056/NEJM0A2027187.
- [63] B. J. Solomon *et al.*, 'Efficacy and safety of first-line lorlatinib versus crizotinib in patients with advanced, ALK-positive non-small-cell lung cancer: updated analysis of data from the phase 3, randomised, open-label CROWN study', *Lancet Respir Med*, vol. 11, no. 4, pp. 354–366, Apr. 2023, doi: 10.1016/S2213-2600(22)00437-4.
- [64] Nuvalent Inc., 'Neladalkib (NVL-655) for TKI-naïve Patients With Advanced ALK-Positive NSCLC Identifier NCT06765109'. Accessed: Jun. 16, 2025. [Online]. Available: <https://clinicaltrials.gov/study/NCT06765109>
- [65] Qilu Pharmaceutical Co. Ltd., 'Study Comparing WX-0593 to Crizotinib in ALK Positive Non-Small Cell Lung Cancer (NSCLC) Patients Identifier NCT04632758'. Accessed: Jun. 16, 2025. [Online]. Available: <https://clinicaltrials.gov/study/NCT04632758>
- [66] Inc. Shenzhen TargertRx, 'TGRX-326 Chinese Phase III for Advanced Non-small Cell Lung Cancer (NSCLC) Identifier NCT06082635'. Accessed: Jun. 16, 2025. [Online]. Available: <https://clinicaltrials.gov/study/NCT06082635?intr=TGRX-326&aggFilters=phase:3&rank=1&tab=table>
- [67] Xuanzhu Biopharmaceutical Co. Ltd., 'A Study to Evaluate and Compare the Efficacy and Safety of XZP-3621 Versus Crizotinib Identifier NCT05204628'. Accessed: Jun. 16, 2025. [Online]. Available: <https://clinicaltrials.gov/study/NCT05204628>
- [68] Inc. Turning Point Therapeutics, 'A Study of Repotrectinib in Pediatric and Young Adult Subjects Harboring ALK, ROS1, OR NTRK1-3 Alterations Identifier

- NCT04094610'. Accessed: Jun. 16, 2025. [Online]. Available: <https://clinicaltrials.gov/study/NCT04094610>
- [69] F. Wachter *et al.*, 'Molecular Characterization of Inflammatory Tumors Facilitates Initiation of Effective Therapy', *Pediatrics*, vol. 148, no. 6, Dec. 2021, doi: 10.1542/PEDS.2021-050990.
- [70] Bristol-Myers Squibb, 'A Study to Assess the Safety, Tolerability, Efficacy, and Drug Levels of BMS-986369 (Golcadomide) in Participants With Relapsed or Refractory T-cell Lymphomas in Japan (GOLSEEK-3) Identifier NCT06035497'. Accessed: Jun. 16, 2025. [Online]. Available: <https://clinicaltrials.gov/study/NCT06035497>
- [71] University of Michigan Rogel Cancer Center, 'Gilteritinib for the Treatment of ALK NSCLC Identifier NCT06225427'. Accessed: Jun. 16, 2025. [Online]. Available: <https://clinicaltrials.gov/study/NCT06225427>
- [72] Schrödinger Inc., 'Study of SGR-1505 in Mature B-Cell Neoplasms Identifier NCT05544019 '. Accessed: Jun. 16, 2025. [Online]. Available: <https://clinicaltrials.gov/study/NCT05544019>
- [73] Centaurus Biopharma Co. Ltd, 'Study of Safety, Efficacy and Pharmacokinetics of CT-707 in Patients With ALK-positive Non-small Cell Lung Cancer Identifier NCT02695550'. Accessed: Jun. 16, 2025. [Online]. Available: <https://clinicaltrials.gov/study/NCT02695550>
- [74] C. Liang *et al.*, 'CT-707 Overcomes Resistance of Crizotinib through Activating PDPK1- AKT1 Pathway by Targeting FAK', *Curr Cancer Drug Targets*, vol. 19, no. 8, pp. 655–665, Nov. 2018, doi: 10.2174/1568009618666181031152140.
- [75] H. Zhou *et al.*, 'Discovery of orally bioavailable ALK PROTACs based ceritinib against ALK positive cancers', *Eur J Med Chem*, vol. 279, p. 116827, Dec. 2024, doi: 10.1016/J.EJMECH.2024.116827.
- [76] S. Xie *et al.*, 'Development of Alectinib-Based PROTACs as Novel Potent Degradors of Anaplastic Lymphoma Kinase (ALK)', *J Med Chem*, vol. 64, no. 13, pp. 9120–9140, Jul. 2021, doi: 10.1021/ACS.JMEDCHEM.1C00270/SUPPL_FILE/JM1C00270_SI_002.CSV.
- [77] G. Yan *et al.*, 'Discovery of a PROTAC targeting ALK with in vivo activity', *Eur J Med Chem*, vol. 212, p. 113150, Feb. 2021, doi: 10.1016/J.EJMECH.2020.113150.
- [78] K. L. Wang *et al.*, 'Discovery of novel anaplastic lymphoma kinase (ALK) and histone deacetylase (HDAC) dual inhibitors exhibiting antiproliferative activity against non-small cell lung cancer', *J Enzyme Inhib Med Chem*, vol. 39, no. 1, 2024, doi: 10.1080/14756366.2024.2318645.
- [79] Y. Lu *et al.*, 'A new ALK inhibitor overcomes resistance to first- and second-generation inhibitors in NSCLC', *EMBO Mol Med*, vol. 14, no. 1, Jan. 2022, doi: 10.15252/EMMM.202114296/SUPPL_FILE/EMMM202114296-SUP-0005-SDATAEV.ZIP.
- [80] W. Gou *et al.*, 'ZX-29, a novel ALK inhibitor, induces apoptosis via ER stress in ALK rearrangement NSCLC cells and overcomes cell resistance caused by an ALK mutation', *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1867, no. 7, p. 118712, Jul. 2020, doi: 10.1016/J.BBAMCR.2020.118712.

- [81] P. He *et al.*, 'Discovery and preclinical evaluations of drug candidate DA-0157 capable of overcoming EGFR drug-resistant mutation C797S and EGFR/ALK co-mutations', *Eur J Med Chem*, vol. 287, p. 117323, Apr. 2025, doi: 10.1016/J.EJMECH.2025.117323.
- [82] R. Wang *et al.*, 'Novel ALK inhibitors containing DAAP fragments: Rational drug design and anti-tumor activity research', *Bioorg Chem*, vol. 160, p. 108416, Jun. 2025, doi: 10.1016/J.BIOORG.2025.108416.
- [83] D. Das *et al.*, 'Discovery and efficacy study of an ALK inhibitor AMX6001 in anaplastic large cell lymphoma Karpas299 mice models', *European Journal of Medicinal Chemistry Reports*, vol. 12, p. 100209, Dec. 2024, doi: 10.1016/J.EJMCR.2024.100209.
- [84] J. Ren *et al.*, 'Design and synthesis of boron-containing ALK inhibitor with favorable in vivo efficacy', *Bioorg Med Chem*, vol. 75, p. 117071, Dec. 2022, doi: 10.1016/J.BMC.2022.117071.
- [85] X. Xiao *et al.*, 'Discovery of imidazo[1,2-b]pyridazine macrocyclic derivatives as novel ALK inhibitors capable of combating multiple resistant mutants', *Bioorg Med Chem Lett*, vol. 89, p. 129309, Jun. 2023, doi: 10.1016/J.BMCL.2023.129309.
- [86] I. El-Jundi, S. Daoud, and M. O. Taha, 'Discovery of novel chemotype inhibitors targeting Anaplastic Lymphoma Kinase receptor through ligand-based pharmacophore modelling', *SAR QSAR Environ Res*, Sep. 2024, doi: 10.1080/1062936X.2024.2406398.
- [87] E. Bilget Guven *et al.*, 'Newly synthesized 6-substituted piperazine/phenyl-9-cyclopentyl containing purine nucleobase analogs act as potent anticancer agents and induce apoptosis via inhibiting Src in hepatocellular carcinoma cells', *RSC Med Chem*, vol. 14, no. 12, pp. 2658–2676, Dec. 2023, doi: 10.1039/D3MD00440F.
- [88] M. Cao, Y. Chen, T. Zhao, S. Wei, M. Guo, and X. Zhai, 'Pyrroformyl-containing 2,4-diaminopyrimidine derivatives as a new optimization strategy of ALK inhibitors combating mutations', *Bioorg Med Chem*, vol. 28, no. 20, p. 115715, Oct. 2020, doi: 10.1016/J.BMC.2020.115715.
- [89] L. Mologni *et al.*, 'Discovery of Novel α -Carboline Inhibitors of the Anaplastic Lymphoma Kinase', *ACS Omega*, 2022, doi: 10.1021/acsomega.2c00507.
- [90] X. Huang *et al.*, 'Revealing 5-(3,5-difluorobenzyl)-1H-indazole as the active pharmacophore of ALK/ROS1 dual inhibitors through theoretical calculations and biological activity evaluations', *Bioorg Chem*, vol. 154, p. 108014, Jan. 2025, doi: 10.1016/J.BIOORG.2024.108014.
- [91] D. Li, R. Yang, J. Wu, B. Zhong, and Y. Li, 'Comprehensive review of α -carboline alkaloids: Natural products, updated synthesis, and biological activities', 2022, doi: 10.3389/fchem.2022.988327.
- [92] J. Y. Tsai, Y. C. Lin, M. H. Hsu, S. C. Kuo, and L. J. Huang, 'Synthesis and cytotoxicity of 1,6,8,9-substituted α -carboline derivatives', *Kaohsiung Journal of Medical Sciences*, vol. 26, no. 11, pp. 593–602, Nov. 2010, doi: 10.1016/S1607-551X(10)70091-7.

- [93] Y. C. Lin *et al.*, 'The novel synthetic compound 6-acetyl-9-(3,4,5-trimethoxybenzyl)-9H-pyrido[2,3-b]indole induces mitotic arrest and apoptosis in human COLO 205 cells', *Int J Oncol*, vol. 43, no. 5, pp. 1596–1606, Nov. 2013, doi: 10.3892/ijo.2013.2069.
- [94] H. C. Huang *et al.*, ' α -Carboline derivative TJY-16 inhibits tumor growth by inducing G2/M cell cycle arrest in glioma cells', *J Biomed Sci*, vol. 23, no. 1, Jan. 2016, doi: 10.1186/s12929-016-0222-y.
- [95] W. Dan *et al.*, 'Novel N1 or N9 modified α -carboline analogues as potential ligands in Alzheimer's disease therapy: Synthesis and neurobiological activity evaluation', *Bioorg Chem*, vol. 133, Apr. 2023, doi: 10.1016/j.bioorg.2023.106378.
- [96] L. Mologni *et al.*, 'Identification of non-ATP-competitive α -carboline inhibitors of the anaplastic lymphoma kinase', *Eur J Med Chem*, vol. 238, p. 114488, Aug. 2022, doi: 10.1016/J.EJMECH.2022.114488.
- [97] N. Guex and M. C. Peitsch, 'SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling', *Electrophoresis*, vol. 18, no. 15, pp. 2714–2723, Dec. 1997, doi: 10.1002/ELPS.1150181505.
- [98] E. F. Pettersen *et al.*, 'UCSF Chimera--a visualization system for exploratory research and analysis', *J Comput Chem*, vol. 25, no. 13, pp. 1605–1612, Oct. 2004, doi: 10.1002/JCC.20084.
- [99] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, 'ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB', *J Chem Theory Comput*, vol. 11, no. 8, pp. 3696–3713, Jul. 2015, doi: 10.1021/ACS.JCTC.5B00255/SUPPL_FILE/CT5B00255_SI_001.PDF.
- [100] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, 'Open Babel: An Open chemical toolbox', *J Cheminform*, vol. 3, no. 10, pp. 1–14, Oct. 2011, doi: 10.1186/1758-2946-3-33/TABLES/2.
- [101] Thomas A. Halgren, 'Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94', *J Comput Chem*, vol. 17, pp. 490–519, Apr. 1996.
- [102] V. Zoete, M. A. Cuendet, A. Grosdidier, and O. Michielin, 'SwissParam: A fast force field generation tool for small organic molecules', *J Comput Chem*, vol. 32, no. 11, pp. 2359–2368, Aug. 2011, doi: 10.1002/JCC.21816.
- [103] O. Trott and A. J. Olson, 'AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading', *J Comput Chem*, vol. 31, no. 2, p. 455, Jan. 2010, doi: 10.1002/JCC.21334.
- [104] J. Eberhardt, D. Santos-Martins, A. F. Tillack, and S. Forli, 'AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings', *J Chem Inf Model*, vol. 61, no. 8, pp. 3891–3898, Aug. 2021, doi: 10.1021/ACS.JCIM.1C00203/SUPPL_FILE/CI1C00203_SI_002.ZIP.
- [105] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, 'GROMACS: A message-passing parallel molecular dynamics implementation', *Comput Phys Commun*, vol. 91, no. 1–3, pp. 43–56, Sep. 1995, doi: 10.1016/0010-4655(95)00042-E.

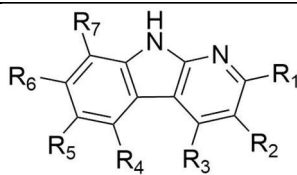
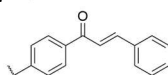
- [106] M. J. Abraham *et al.*, ‘GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers’, *SoftwareX*, vol. 1–2, pp. 19–25, Sep. 2015, doi: 10.1016/J.SOFTX.2015.06.001.
- [107] E. Lindahl, P. Bjelkmar, P. Larsson, M. A. Cuendet, and B. Hess, ‘Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models’, *J Chem Theory Comput*, vol. 6, no. 2, pp. 459–466, Feb. 2010, doi: 10.1021/CT900549R.
- [108] G. Bussi, D. Donadio, and M. Parrinello, ‘Canonical sampling through velocity rescaling’, *Journal of Chemical Physics*, vol. 126, no. 1, Jan. 2007, doi: 10.1063/1.2408420/186581.
- [109] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, ‘Molecular dynamics with coupling to an external bath’, *J Chem Phys*, vol. 81, no. 8, pp. 3684–3690, Oct. 1984, doi: 10.1063/1.448118.
- [110] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, ‘MDAnalysis: A toolkit for the analysis of molecular dynamics simulations’, *J Comput Chem*, vol. 32, no. 10, pp. 2319–2327, Jul. 2011, doi: 10.1002/JCC.21787.
- [111] R. J. Gowers *et al.*, ‘MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations’, *Proceedings of the 15th Python in Science Conference*, pp. 98–105, 2016, doi: 10.25080/MAJORA-629E541A-00E.
- [112] D. L. Theobald, ‘Rapid calculation of RMSDs using a quaternion-based characteristic polynomial’, *Acta Crystallogr A*, vol. 61, no. Pt 4, pp. 478–480, 2005, doi: 10.1107/S0108767305015266.
- [113] P. Liu, D. K. Agrafiotis, and D. L. Theobald, ‘Fast determination of the optimal rotational matrix for macromolecular superpositions’, *J Comput Chem*, vol. 31, no. 7, pp. 1561–1563, May 2010, doi: 10.1002/JCC.21439.
- [114] M. L. Waskom, ‘seaborn: statistical data visualization’, *J Open Source Softw*, vol. 6, no. 60, p. 3021, Apr. 2021, doi: 10.21105/JOSS.03021.
- [115] S. Van Der Walt, S. C. Colbert, and G. Varoquaux, ‘The NumPy array: A structure for efficient numerical computation’, *Comput Sci Eng*, vol. 13, no. 2, pp. 22–30, Mar. 2011, doi: 10.1109/MCSE.2011.37.
- [116] R. Kumari, R. Kumar, and A. Lynn, ‘G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations’, *J Chem Inf Model*, vol. 54, no. 7, pp. 1951–1962, Jul. 2014, doi: 10.1021/CI500020M/SUPPL_FILE/CI500020M_SI_001.PDF.
- [117] T. Tubiana, J. C. Carvaille, 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/SUPPL_FILE/CI8B00512_SI_001.PDF.
- [118] A. S. Abouzied *et al.*, ‘Structural and free energy landscape analysis for the discovery of antiviral compounds targeting the cap-binding domain of influenza polymerase PB2’, *Scientific Reports 2024 14:1*, vol. 14, no. 1, pp. 1–14, Oct. 2024, doi: 10.1038/s41598-024-69816-3.
- [119] V. Chandrasekhar *et al.*, ‘COCONUT 2.0: a comprehensive overhaul and curation of the collection of open natural products database’, *Nucleic Acids Res*, vol. 53, no. D1, pp. D634–D643, Jan. 2025, doi: 10.1093/NAR/GKAE1063.

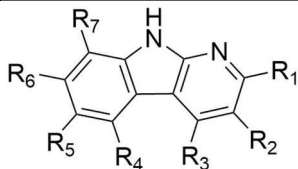
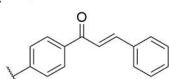
- [120] K. Vanommeslaeghe *et al.*, ‘CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields’, *J Comput Chem*, vol. 31, no. 4, pp. 671–690, Mar. 2010, doi: 10.1002/JCC.21367.
- [121] D. R. Koes, M. P. Baumgartner, and C. J. Camacho, ‘Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise’, *J Chem Inf Model*, vol. 53, no. 8, pp. 1893–1904, Aug. 2013, doi: 10.1021/CI300604Z/SUPPL_FILE/CI300604Z_SI_002.PDF.
- [122] R. A. Laskowski and M. B. Swindells, ‘LigPlot+: multiple ligand-protein interaction diagrams for drug discovery’, *J Chem Inf Model*, vol. 51, no. 10, pp. 2778–2786, Oct. 2011, doi: 10.1021/CI200227U.
- [123] J. Huang *et al.*, ‘CHARMM36m: an improved force field for folded and intrinsically disordered proteins’, *Nature Methods* 2016 14:1, vol. 14, no. 1, pp. 71–73, Nov. 2016, doi: 10.1038/nmeth.4067.
- [124] P. Mark and L. Nilsson, ‘Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K’, *Journal of Physical Chemistry A*, vol. 105, no. 43, pp. 9954–9960, Nov. 2001, doi: 10.1021/JP003020W/ASSET/IMAGES/LARGE/JP003020WF00004.JPEG.
- [125] M. Bernetti and G. Bussi, ‘Pressure control using stochastic cell rescaling’, *Journal of Chemical Physics*, vol. 153, no. 11, p. 114107, Sep. 2020, doi: 10.1063/5.0020514/199610.
- [126] M. Krug *et al.*, ‘Discovery of 4-Benzylamino-Substituted α -Carbolines as a Novel Class of Receptor Tyrosine Kinase Inhibitors’, *ChemMedChem*, vol. 6, no. 1, pp. 63–72, Jan. 2011, doi: 10.1002/cmdc.201000384.
- [127] S. Liang *et al.*, ‘Deciphering the Mechanism of Gilteritinib Overcoming Lorlatinib Resistance to the Double Mutant I1171N/F1174I in Anaplastic Lymphoma Kinase’, *Front Cell Dev Biol*, vol. 9, p. 808864, Dec. 2021, doi: 10.3389/FCELL.2021.808864/FULL.
- [128] T. W. Johnson *et al.*, ‘Discovery of (10 R)-7-Amino-12-fluoro-2,10,16-trimethyl-15-oxo-10,15,16,17- tetrahydro- 2H -8,4-(metheno)pyrazolo[4,3- h][2,5,11]-benzoxadiazacyclotetradecine-3-carbonitrile (PF-06463922), a macrocyclic inhibitor of anaplastic lymphoma kinase (ALK) and c-ros oncogene 1 (ROS1) with preclinical brain exposure and broad-spectrum potency against ALK-resistant mutations’, *J Med Chem*, vol. 57, no. 11, pp. 4720–4744, Jun. 2014, doi: 10.1021/JM500261Q/SUPPL_FILE/JM500261Q_SI_001.PDF.
- [129] F. Bray, M. Laversanne, E. Weiderpass, and I. Soerjomataram, ‘The ever-increasing importance of cancer as a leading cause of premature death worldwide’, *Cancer*, vol. 127, no. 16, pp. 3029–3030, Aug. 2021, doi: 10.1002/CNCR.33587.
- [130] H. Sung *et al.*, ‘Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries’, *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/CAAC.21660.

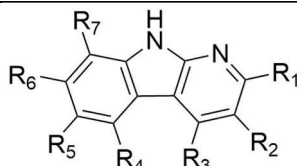
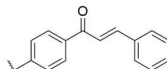


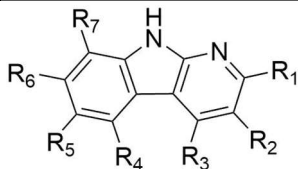
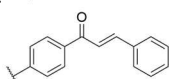
APPENDIX

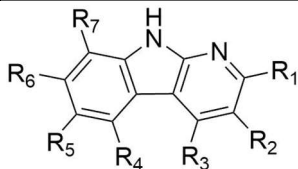
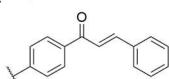
Appendix Table 1. The list of designed Carbolines with binding affinity towards kinase domain of ALK

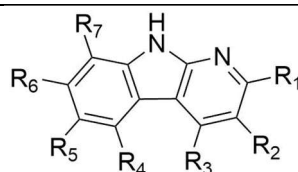
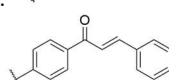
Chemical Abbreviations of the R groups:								
				Propenone, PRP: 				
				Chalcone, CLC: 				
C. No	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	Binding Affinity (kcal/mol)
1	-H	-H	-H	-H	-H	-H	-H	-7.23
2	-PRP	-H	-H	-H	-H	-H	-H	-7.47
3	-PRP	-H	-H	-F	-H	-H	-H	-7.55
4	-PRP	-H	-H	-Cl	-H	-H	-H	-7.71
5	-PRP	-H	-H	-Br	-H	-H	-H	-7.74
6	-PRP	-H	-H	-I	-H	-H	-H	-7.64
7	-PRP	-H	-H	-H	-F	-H	-H	-7.73
8	-PRP	-H	-H	-H	-Cl	-H	-H	-7.62
9	-PRP	-H	-H	-H	-Br	-H	-H	-7.41
10	-PRP	-H	-H	-H	-I	-H	-H	-7.44
11	-PRP	-H	-H	-H	-H	-F	-H	-7.67
12	-PRP	-H	-H	-H	-H	-Cl	-H	-7.56
13	-PRP	-H	-H	-H	-H	-Br	-H	-7.49
14	-PRP	-H	-H	-H	-H	-I	-H	-7.50
15	-PRP	-H	-H	-H	-H	-H	-F	-7.64
16	-PRP	-H	-H	-H	-H	-H	-Cl	-7.54
17	-PRP	-H	-H	-H	-H	-H	-Br	-7.61
18	-PRP	-H	-H	-H	-H	-H	-I	-7.69
19	-H	-PRP	-H	-H	-H	-H	-H	-7.92
20	-H	-PRP	-H	-F	-H	-H	-H	-8.15
21	-H	-PRP	-H	-Cl	-H	-H	-H	-8.13
22	-H	-PRP	-H	-Br	-H	-H	-H	-8.20
23	-H	-PRP	-H	-I	-H	-H	-H	-8.21
24	-H	-PRP	-H	-H	-F	-H	-H	-8.11
25	-H	-PRP	-H	-H	-Cl	-H	-H	-8.03
26	-H	-PRP	-H	-H	-Br	-H	-H	-7.86
27	-H	-PRP	-H	-H	-I	-H	-H	-7.55
28	-H	-PRP	-H	-H	-H	-F	-H	-8.04
29	-H	-PRP	-H	-H	-H	-Cl	-H	-8.08
30	-H	-PRP	-H	-H	-H	-Br	-H	-8.17

Chemical Abbreviations of the R groups:								
				Propenone, PRP:  Chalcone, CLC: 				
C. No	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	Binding Affinity (kcal/mol)
31	-H	-PRP	-H	-H	-H	-I	-H	-8.20
32	-H	-PRP	-H	-H	-H	-H	-F	-8.06
33	-H	-PRP	-H	-H	-H	-H	-Cl	-8.20
34	-H	-PRP	-H	-H	-H	-H	-Br	-8.08
35	-H	-PRP	-H	-H	-H	-H	-I	-8.03
36	-H	-H	-PRP	-H	-H	-H	-H	-7.51
37	-H	-H	-PRP	-F	-H	-H	-H	-7.83
38	-H	-H	-PRP	-Cl	-H	-H	-H	-7.72
39	-H	-H	-PRP	-Br	-H	-H	-H	-7.51
40	-H	-H	-PRP	-I	-H	-H	-H	-7.25
41	-H	-H	-PRP	-H	-F	-H	-H	-7.89
42	-H	-H	-PRP	-H	-Cl	-H	-H	-7.70
43	-H	-H	-PRP	-H	-Br	-H	-H	-7.34
44	-H	-H	-PRP	-H	-I	-H	-H	-7.80
45	-H	-H	-PRP	-H	-H	-F	-H	-7.95
46	-H	-H	-PRP	-H	-H	-Cl	-H	-7.75
47	-H	-H	-PRP	-H	-H	-Br	-H	-7.99
48	-H	-H	-PRP	-H	-H	-I	-H	-7.67
49	-H	-H	-PRP	-H	-H	-H	-F	-8.30
50	-H	-H	-PRP	-H	-H	-H	-Cl	-7.50
51	-H	-H	-PRP	-H	-H	-H	-Br	-7.62
52	-H	-H	-PRP	-H	-H	-H	-I	-8.07
53	-H	-H	-H	-PRP	-H	-H	-H	-7.65
54	-F	-H	-H	-PRP	-H	-H	-H	-8.29
55	-Cl	-H	-H	-PRP	-H	-H	-H	-8.02
56	-Br	-H	-H	-PRP	-H	-H	-H	-7.82
57	-I	-H	-H	-PRP	-H	-H	-H	-7.63
58	-H	-F	-H	-PRP	-H	-H	-H	-7.81
59	-H	-Cl	-H	-PRP	-H	-H	-H	-8.06
60	-H	-Br	-H	-PRP	-H	-H	-H	-8.30
61	-H	-I	-H	-PRP	-H	-H	-H	-7.86
62	-H	-H	-F	-PRP	-H	-H	-H	-8.03
63	-H	-H	-Cl	-PRP	-H	-H	-H	-7.97
64	-H	-H	-Br	-PRP	-H	-H	-H	-7.68
65	-H	-H	-I	-PRP	-H	-H	-H	-7.31
66	-H	-H	-H	-H	-PRP	-H	-H	-7.79

Chemical Abbreviations of the R groups:								
				Propenone, PRP:  Chalcone, CLC: 				
C. No	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	Binding Affinity (kcal/mol)
67	-F	-H	-H	-H	-PRP	-H	-H	-8.13
68	-Cl	-H	-H	-H	-PRP	-H	-H	-8.02
69	-Br	-H	-H	-H	-PRP	-H	-H	-8.04
70	-I	-H	-H	-H	-PRP	-H	-H	-8.05
71	-H	-F	-H	-H	-PRP	-H	-H	-8.04
72	-H	-Cl	-H	-H	-PRP	-H	-H	-8.08
73	-H	-Br	-H	-H	-PRP	-H	-H	-8.04
74	-H	-I	-H	-H	-PRP	-H	-H	-7.88
75	-H	-H	-F	-H	-PRP	-H	-H	-7.88
76	-H	-H	-Cl	-H	-PRP	-H	-H	-7.80
77	-H	-H	-Br	-H	-PRP	-H	-H	-7.96
78	-H	-H	-I	-H	-PRP	-H	-H	-8.04
79	-H	-H	-H	-H	-H	-PRP	-H	-7.56
80	-F	-H	-H	-H	-H	-PRP	-H	-7.79
81	-Cl	-H	-H	-H	-H	-PRP	-H	-7.81
82	-Br	-H	-H	-H	-H	-PRP	-H	-7.81
83	-I	-H	-H	-H	-H	-PRP	-H	-7.76
84	-H	-F	-H	-H	-H	-PRP	-H	-7.62
85	-H	-Cl	-H	-H	-H	-PRP	-H	-7.66
86	-H	-Br	-H	-H	-H	-PRP	-H	-7.69
87	-H	-I	-H	-H	-H	-PRP	-H	-7.65
88	-H	-H	-F	-H	-H	-PRP	-H	-7.71
89	-H	-H	-Cl	-H	-H	-PRP	-H	-7.69
90	-H	-H	-Br	-H	-H	-PRP	-H	-7.67
91	-H	-H	-I	-H	-H	-PRP	-H	-7.50
92	-H	-H	-H	-H	-H	-H	-PRP	-7.46
93	-F	-H	-H	-H	-H	-H	-PRP	-7.82
94	-Cl	-H	-H	-H	-H	-H	-PRP	-7.80
95	-Br	-H	-H	-H	-H	-H	-PRP	-8.0
96	-I	-H	-H	-H	-H	-H	-PRP	-7.84
97	-H	-F	-H	-H	-H	-H	-PRP	-8.09
98	-H	-Cl	-H	-H	-H	-H	-PRP	-7.63
99	-H	-Br	-H	-H	-H	-H	-PRP	-7.53
100	-H	-I	-H	-H	-H	-H	-PRP	-7.53
101	-H	-H	-F	-H	-H	-H	-PRP	-8.21
102	-H	-H	-Cl	-H	-H	-H	-PRP	-7.70

 Chemical Abbreviations of the R groups: Propenone, PRP:  Chalcone, CLC: 								
C. No	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	Binding Affinity (kcal/mol)
103	-H	-H	-Br	-H	-H	-H	-PRP	-7.58
104	-H	-H	-I	-H	-H	-H	-PRP	-7.55
105	-CLC	-H	-H	-H	-H	-H	-H	-9.73
106	-CLC	-H	-H	-F	-H	-H	-H	-9.95
107	-CLC	-H	-H	-Cl	-H	-H	-H	-9.99
108	-CLC	-H	-H	-Br	-H	-H	-H	-9.94
109	-CLC	-H	-H	-I	-H	-H	-H	-10.08
110	-CLC	-H	-H	-H	-F	-H	-H	-10.16
111	-CLC	-H	-H	-H	-Cl	-H	-H	-10.10
112	-CLC	-H	-H	-H	-Br	-H	-H	-9.98
113	-CLC	-H	-H	-H	-I	-H	-H	-9.31
114	-CLC	-H	-H	-H	-H	-F	-H	-9.96
115	-CLC	-H	-H	-H	-H	-Cl	-H	-9.51
116	-CLC	-H	-H	-H	-H	-Br	-H	-8.91
117	-CLC	-H	-H	-H	-H	-I	-H	-9.04
118	-CLC	-H	-H	-H	-H	-H	-F	-9.95
119	-CLC	-H	-H	-H	-H	-H	-Cl	-9.89
120	-CLC	-H	-H	-H	-H	-H	-Br	-10.01
121	-CLC	-H	-H	-H	-H	-H	-I	-9.90
122	-H	-CLC	-H	-H	-H	-H	-H	-9.92
123	-H	-CLC	-H	-F	-H	-H	-H	-10.17
124	-H	-CLC	-H	-Cl	-H	-H	-H	-10.32
125	-H	-CLC	-H	-Br	-H	-H	-H	-10.26
126	-H	-CLC	-H	-I	-H	-H	-H	-10.17
127	-H	-CLC	-H	-H	-F	-H	-H	-10.63
128	-H	-CLC	-H	-H	-Cl	-H	-H	-10.36
129	-H	-CLC	-H	-H	-Br	-H	-H	-9.87
130	-H	-CLC	-H	-H	-I	-H	-H	-9.63
131	-H	-CLC	-H	-H	-H	-F	-H	-10.00
132	-H	-CLC	-H	-H	-H	-Cl	-H	-9.52
133	-H	-CLC	-H	-H	-H	-Br	-H	-9.16
134	-H	-CLC	-H	-H	-H	-I	-H	-9.32
135	-H	-CLC	-H	-H	-H	-H	-F	-9.95
136	-H	-CLC	-H	-H	-H	-H	-Cl	-9.96
137	-H	-CLC	-H	-H	-H	-H	-Br	-9.63
138	-H	-CLC	-H	-H	-H	-H	-I	-9.40

Chemical Abbreviations of the R groups:								
				Propenone, PRP:  Chalcone, CLC: 				
C. No	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	Binding Affinity (kcal/mol)
139	-H	-H	-CLC	-H	-H	-H	-H	-10.25
140	-H	-H	-CLC	-F	-H	-H	-H	-10.09
141	-H	-H	-CLC	-Cl	-H	-H	-H	-9.48
142	-H	-H	-CLC	-Br	-H	-H	-H	-9.30
143	-H	-H	-CLC	-I	-H	-H	-H	-9.28
144	-H	-H	-CLC	-H	-F	-H	-H	-10.08
145	-H	-H	-CLC	-H	-Cl	-H	-H	-9.80
146	-H	-H	-CLC	-H	-Br	-H	-H	-9.52
147	-H	-H	-CLC	-H	-I	-H	-H	-9.83
148	-H	-H	-CLC	-H	-H	-F	-H	-10.18
149	-H	-H	-CLC	-H	-H	-Cl	-H	-10.07
150	-H	-H	-CLC	-H	-H	-Br	-H	-9.91
151	-H	-H	-CLC	-H	-H	-I	-H	-10.29
152	-H	-H	-CLC	-H	-H	-H	-F	-10.10
153	-H	-H	-CLC	-H	-H	-H	-Cl	-9.80
154	-H	-H	-CLC	-H	-H	-H	-Br	-10.37
155	-H	-H	-CLC	-H	-H	-H	-I	-9.32
156	-H	-H	-H	-CLC	-H	-H	-H	-9.85
157	-F	-H	-H	-CLC	-H	-H	-H	-10.14
158	-Cl	-H	-H	-CLC	-H	-H	-H	-10.17
159	-Br	-H	-H	-CLC	-H	-H	-H	-10.17
160	-I	-H	-H	-CLC	-H	-H	-H	-10.02
161	-H	-F	-H	-CLC	-H	-H	-H	-9.97
162	-H	-Cl	-H	-CLC	-H	-H	-H	-9.99
163	-H	-Br	-H	-CLC	-H	-H	-H	-10.02
164	-H	-I	-H	-CLC	-H	-H	-H	-10.60
165	-H	-H	-F	-CLC	-H	-H	-H	-10.19
166	-H	-H	-Cl	-CLC	-H	-H	-H	-9.66
167	-H	-H	-Br	-CLC	-H	-H	-H	-9.58
168	-H	-H	-I	-CLC	-H	-H	-H	-9.74
169	-H	-H	-H	-H	-CLC	-H	-H	-9.69
170	-F	-H	-H	-H	-CLC	-H	-H	-9.97
171	-Cl	-H	-H	-H	-CLC	-H	-H	-9.50
172	-Br	-H	-H	-H	-CLC	-H	-H	-9.23
173	-I	-H	-H	-H	-CLC	-H	-H	-9.17
174	-H	-F	-H	-H	-CLC	-H	-H	-9.94

 Chemical Abbreviations of the R groups: Propenone, PRP:  Chalcone, CLC: 								
C. No	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	Binding Affinity (kcal/mol)
175	-H	-Cl	-H	-H	-CLC	-H	-H	-9.82
176	-H	-Br	-H	-H	-CLC	-H	-H	-9.70
177	-H	-I	-H	-H	-CLC	-H	-H	-9.62
178	-H	-H	-F	-H	-CLC	-H	-H	-9.85
179	-H	-H	-Cl	-H	-CLC	-H	-H	-9.94
180	-H	-H	-Br	-H	-CLC	-H	-H	-10.13
181	-H	-H	-I	-H	-CLC	-H	-H	-9.93
182	-H	-H	-H	-H	-H	-CLC	-H	-9.79
183	-F	-H	-H	-H	-H	-CLC	-H	-10.20
184	-Cl	-H	-H	-H	-H	-CLC	-H	-9.57
185	-Br	-H	-H	-H	-H	-CLC	-H	-9.06
186	-I	-H	-H	-H	-H	-CLC	-H	-8.95
187	-H	-F	-H	-H	-H	-CLC	-H	-10.04
188	-H	-Cl	-H	-H	-H	-CLC	-H	-9.81
189	-H	-Br	-H	-H	-H	-CLC	-H	-9.20
190	-H	-I	-H	-H	-H	-CLC	-H	-9.20
191	-H	-H	-F	-H	-H	-CLC	-H	-10.13
192	-H	-H	-Cl	-H	-H	-CLC	-H	-10.00
193	-H	-H	-Br	-H	-H	-CLC	-H	-9.99
194	-H	-H	-I	-H	-H	-CLC	-H	-9.83
195	-H	-H	-H	-H	-H	-H	-CLC	-9.23
196	-F	-H	-H	-H	-H	-H	-CLC	-9.70
197	-Cl	-H	-H	-H	-H	-H	-CLC	-9.43
198	-Br	-H	-H	-H	-H	-H	-CLC	-9.65
199	-I	-H	-H	-H	-H	-H	-CLC	-9.33
200	-H	-F	-H	-H	-H	-H	-CLC	-9.52
201	-H	-Cl	-H	-H	-H	-H	-CLC	-9.51
202	-H	-Br	-H	-H	-H	-H	-CLC	-9.26
203	-H	-I	-H	-H	-H	-H	-CLC	-9.36
204	-H	-H	-F	-H	-H	-H	-CLC	-9.40
205	-H	-H	-Cl	-H	-H	-H	-CLC	-9.55
206	-H	-H	-Br	-H	-H	-H	-CLC	-9.50
207	-H	-H	-I	-H	-H	-H	-CLC	-9.53

Appendix Table 2. The top lowest 109 of COCONUT docking results against ALK tyrosine domain.

No	COCONUT ID	Binding Affinity (kcal/mol)
1	CNP0106316.1	-11.6
2	CNP0264000.1	-11.6
3	CNP0311003.1	-11.6
4	CNP0289278.1	-11.5
5	CNP0124314.1	-11.4
6	CNP0169424.1	-11.4
7	CNP0273044.1	-11.4
8	CNP0273044.2	-11.4
9	CNP0501886.1	-11.4
10	CNP0573612.1	-11.4
11	CNP0073860.1	-11.3
12	CNP0180079.1	-11.3
13	CNP0212191.1	-11.3
14	CNP0214469.1	-11.3
15	CNP0241884.1	-11.3
16	CNP0254622.2	-11.3
17	CNP0290931.0	-11.3
18	CNP0431234.1	-11.3
19	CNP0538300.1	-11.3
20	CNP0541655.4	-11.3
21	CNP0554764.1	-11.3
22	CNP0582844.1	-11.3
23	CNP0144079.0	-11.2
24	CNP0241884.2	-11.2

No	COCONUT ID	Binding Affinity (kcal/mol)
25	CNP0248990.2	-11.2
26	CNP0254622.1	-11.2
27	CNP0264442.2	-11.2
28	CNP0427126.1	-11.2
29	CNP0431234.2	-11.2
30	CNP0463073.1	-11.2
31	CNP0088709.1	-11.1
32	CNP0132508.0	-11.1
33	CNP0214469.2	-11.1
34	CNP0283500.1	-11.1
35	CNP0342738.2	-11.1
36	CNP0376544.1	-11.1
37	CNP0393527.1	-11.1
38	CNP0484299.1	-11.1
39	CNP0506945.2	-11.1
40	CNP0521909.1	-11.1
41	CNP0523393.0	-11.1
42	CNP0541655.3	-11.1
43	CNP0572176.0	-11.1
44	CNP0072683.1	-11
45	CNP0086971.3	-11
46	CNP0104925.1	-11
47	CNP0112873.1	-11
48	CNP0138873.0	-11
49	CNP0214122.0	-11
50	CNP0219868.0	-11

No	COCONUT ID	Binding Affinity (kcal/mol)
51	CNP0280499.0	-11
52	CNP0303417.1	-11
53	CNP0303417.2	-11
54	CNP0316072.0	-11
55	CNP0342738.1	-11
56	CNP0343202.1	-11
57	CNP0354949.2	-11
58	CNP0363593.0	-11
59	CNP0378659.1	-11
60	CNP0521909.2	-11
61	CNP0586714.2	-11
62	CNP0048261.0	-10.9
63	CNP0074122.0	-10.9
64	CNP0086971.2	-10.9
65	CNP0192112.1	-10.9
66	CNP0192112.2	-10.9
67	CNP0192112.3	-10.9
68	CNP0192112.4	-10.9
69	CNP0192112.7	-10.9
70	CNP0312998.1	-10.9
71	CNP0324958.2	-10.9
72	CNP0340497.2	-10.9
73	CNP0351785.1	-10.9
74	CNP0354566.0	-10.9
75	CNP0362893.0	-10.9
76	CNP0403407.0	-10.9

No	COCONUT ID	Binding Affinity (kcal/mol)
77	CNP0506945.1	-10.9
78	CNP0513178.1	-10.9
79	CNP0539369.1	-10.9
80	CNP0074120.0	-10.8
81	CNP0089194.1	-10.8
82	CNP0134183.1	-10.8
83	CNP0149599.1	-10.8
84	CNP0149599.2	-10.8
85	CNP0149599.3	-10.8
86	CNP0156931.1	-10.8
87	CNP0166805.0	-10.8
88	CNP0191856.1	-10.8
89	CNP0193779.0	-10.8
90	CNP0229913.11	-10.8
91	CNP0229913.13	-10.8
92	CNP0229913.17	-10.8
93	CNP0229913.19	-10.8
94	CNP0229913.22	-10.8
95	CNP0229913.2	-10.8
96	CNP0229913.32	-10.8
97	CNP0229913.8	-10.8
98	CNP0231286.0	-10.8
99	CNP0257215.0	-10.8
100	CNP0264788.5	-10.8
101	CNP0294386.0	-10.8
102	CNP0300369.0	-10.8

No	COCONUT ID	Binding Affinity (kcal/mol)
103	CNP0306932.3	-10.8
104	CNP0318435.1	-10.8
105	CNP0324958.1	-10.8

No	COCONUT ID	Binding Affinity (kcal/mol)
106	CNP0338372.0	-10.8
107	CNP0372341.0	-10.8
108	CNP0488686.1	-10.8



CURRICULUM VITAE

2023 - Present	Bioengineering M.Sc. student, Akçok Lab, Abdullah Gül University, Kayseri, Türkiye
04/2024 - Present	Research Assistant, Bioengineering, Abdullah Gül University, Kayseri, Türkiye
05/2022 – 06/2022	Intern, DETA-GEN Genetic Diseases Diagnostic Centre, Kayseri, Türkiye
2017 - 2023	B.Sc., Molecular Biology and Genetics, Abdullah Gül University, Kayseri, Türkiye

SELECTED PUBLICATIONS AND PRESENTATIONS

- J1)** Kurt, H., Sever Kaya, D., Akçok, İ., Sarı, C., Albayrak, E., Velioğlu, H. M., Şamlı, H. E., Özdüven, M. L., & Sürmeli, Y. (2025). Discovery and in silico characterization of Anatolian water buffalo rumen-derived bacterial thermostable xylanases: A sequence-based metagenomic approach. *ACS Omega*, 10(12), 12679–12698. <https://doi.org/10.1021/acsomega.5c00965>
- J2)** Şansaçar, M., Sarı, C., Yücel, M. S., Gencer Akçok, E. B., & Akçok, İ. (2025). Discovery of new candidates targeting the SH2 domains of spleen tyrosine kinase (Syk) through in silico studies. *ChemistrySelect*, 10(22). <https://doi.org/10.1002/slct.202500547>
- C1)** C. Sarı, İ. Akçok, A comprehensive in silico approach towards the design of new α -carboline derivatives against Anaplastic lymphoma kinase (ALK) in 13. International Drug Chemistry Conference (February, 2025)