UNIVERSITE YAHIA FARES DE MEDEA

FACULTE DES SCIENCES

DEPARTEMENT DES SCIENCES DE LA MATIERE

Mémoire

Présenté pour obtenir le diplôme de **Master**

En Chimie

Spécialité :

Chimie Pharmaceutique

Par :

KERMEZLI Oumaima

Thème :

Molecular Simulation of the Interaction Between Natural Anticancer Compounds and Their Biological Targets

Soutenue publiquement le : 15 /06 /2026 ; Devant le jury composé de :

Pr. BELHADJ Abdelmouneim	Grade :	Professeur	Président
Dr.EULDJI Amel	Grade :	MCB	Examinateur
Pr.BITAM Said	Grade :	Professeur	Encadreur

Année universitaire 2025/2026

Dedication

Dedication

To myself, for every effort made, every challenge faced, and every moment I chose to keep going despite the difficulties. This achievement is a reminder of my perseverance, growth, and belief in my dreams.

To my beloved parents, for their endless love, prayers, sacrifices, and support.

To my dear grandmother, whose love, kindness, and prayers have always held a special place in my heart.

To my siblings, for their love, support, and constant encouragement.

To all my family, from the youngest to the eldest, for their encouragement, care, and presence throughout my journey.

To my friends, who made this path more beautiful with their support, laughter, and unforgettable memories.

To all the people I love, and those I will love in the future.

Your presence in my life, in one way or another, has left a meaningful mark on my heart.

With all my love and gratitude, I dedicate this achievement ***to you.***

Acknowledgements

Acknowledgements

First and foremost, I would like to thank Allah, the Almighty, for giving me the strength, patience, courage, and determination to complete this work. Without His guidance and blessings, this achievement would not have been possible.

I would also like to thank myself for never giving up, for staying strong through difficult moments, and for continuing to believe in this journey even when it was not easy. This work represents not only an academic achievement, but also personal growth, effort, and perseverance.

I would like to express my sincere gratitude to my supervisor, ***Mr. Bitam,*** *for his guidance, support, valuable advice, and patience throughout the realization of this final year project. His encouragement and supervision have been truly appreciated.*

My deepest thanks go to all the teachers who have accompanied me throughout my educational journey, from my very first years of school until this second year of Master's degree at the university. Each teacher has contributed in their own way to my knowledge, my development, and the person I am today.

A very special and heartfelt thank you goes to my beloved parents, whose love, sacrifices, prayers, and endless support have always been my greatest source of strength. I am deeply grateful to my family for their encouragement, trust, and presence in every step of my life.

To everyone who believed in me, supported me, and stood by my side, I offer my sincere gratitude from the bottom of my heart.

Abstract

Abstract

Cancer remains a major public health challenge due to its high mortality rate, complex biological mechanisms, and the limitations of conventional therapies, including toxicity, low selectivity, and drug resistance. In this context, natural compounds represent an important source of potential anticancer agents because of their structural diversity and ability to modulate several cancer-related pathways. This study focuses on the molecular simulation of the interaction between selected natural anticancer compounds and key biological targets involved in tumor progression, namely VEGFR2, androgen receptor, BCL-2, PARP1, and CDK4.

The objective of this work was to identify promising natural molecules with potential multi-target anticancer activity using computational drug discovery approaches. Protein structures were obtained from the Protein Data Bank, while natural compounds were collected from chemical databases and literature sources. The selected ligands were subjected to ADME/Tox prediction in order to evaluate their pharmacokinetic properties, drug-likeness, and potential toxicity. Molecular docking and cross-docking studies were then performed to estimate ligand–target binding affinities. Post-docking analysis was used to identify the main interactions involved in ligand binding, including hydrogen bonds, hydrophobic interactions, π–π stacking, and key amino acid residues.

The results of this in silico study provide valuable insights into the binding behavior of natural compounds toward several validated anticancer targets. This work highlights the usefulness of molecular simulation as a rapid and cost-effective tool in the early stages of anticancer drug discovery. The identified promising compounds may serve as starting points for further molecular dynamics simulations, structural optimization, and experimental validation through in vitro and in vivo studies.

Keywords: Cancer, natural compounds, molecular docking, ADME/Tox, anticancer targets, molecular simulation.

Résumé

Résumé

Le cancer demeure l'un des principaux problèmes de santé publique dans le monde en raison de sa forte mortalité, de la complexité de ses mécanismes biologiques et des limites des traitements conventionnels, notamment la toxicité, le manque de sélectivité et l'apparition de résistances thérapeutiques. Dans ce contexte, les composés naturels représentent une source importante de molécules bioactives à potentiel anticancéreux, grâce à leur diversité structurale et à leur capacité à moduler plusieurs voies biologiques impliquées dans la progression tumorale.

Ce travail porte sur la simulation moléculaire de l'interaction entre des composés naturels anticancéreux sélectionnés et plusieurs cibles biologiques impliquées dans le cancer, notamment VEGFR2, le récepteur aux androgènes, BCL-2, PARP1 et CDK4. L'objectif principal de cette étude est d'identifier des molécules naturelles prometteuses présentant une activité anticancéreuse potentielle multi-cible.

La méthodologie adoptée repose sur des approches de conception de médicaments assistée par ordinateur. Les structures protéiques ont été obtenues à partir de la Protein Data Bank, tandis que les composés naturels ont été collectés à partir de bases de données chimiques et de sources bibliographiques. Une analyse ADME/Tox a été réalisée afin d'évaluer les propriétés pharmacocinétiques, la drug-likeness et la toxicité potentielle des ligands sélectionnés. Des études de docking moléculaire et de cross-docking ont ensuite été effectuées pour prédire l'affinité de liaison des composés envers les cibles biologiques choisies. Enfin, une analyse post-docking a permis d'identifier les principales interactions moléculaires impliquées dans la stabilisation des complexes ligand–protéine.

Les résultats de cette étude in silico fournissent des informations importantes sur le comportement d'interaction des composés naturels avec plusieurs cibles anticancéreuses validées. Ce travail met en évidence l'intérêt de la simulation moléculaire comme outil rapide et économique dans les premières étapes de la découverte de nouveaux agents anticancéreux. Les composés les plus prometteurs pourraient faire l'objet d'études complémentaires, notamment par dynamique moléculaire, optimisation structurale et validation expérimentale in vitro et in vivo.

Mots-clés : Cancer, composés naturels, simulation moléculaire, docking moléculaire, ADME/Tox, cibles anticancéreuses.

مُلَخَّص

مُلَخَّص

يُعدّ السرطان من أهم مشكلات الصحة العامة في العالم، نظراً لارتفاع معدل الوفيات المرتبط به وتعقّد آلياته البيولوجية، إضافةً إلى محدودية العلاجات التقليدية مثل السمية، ضعف الانتقائية، وظهور مقاومة للأدوية. وفي هذا السياق، تمثل المركبات الطبيعية مصدراً مهماً للجزيئات النشطة بيولوجياً ذات التأثير المحتمل المضاد للسرطان، وذلك بفضل تنوعها البنيوي وقدرتها على التأثير في عدة مسارات خلوية مرتبطة بتطور الأورام.

يركز هذا العمل على المحاكاة الجزيئية للتفاعل بين مركبات طبيعية مختارة ذات نشاط مضاد للسرطان وعدة أهداف بيولوجية مرتبطة بتطور المرض، من بينهاVEGFR2 ، مستقبل الأندروجين، BCL-2، PARP1 و .CDK4 ويهدف هذا البحث إلى تحديد مركبات طبيعية واعدة يمكن أن تمتلك نشاطاً مضاداً للسرطان متعدد الأهداف.

اعتمدت هذه الدراسة على مناهج تصميم الأدوية بمساعدة الحاسوب، حيث تم الحصول على البنى البروتينية من قاعدة بيانات البروتينات، كما تم جمع المركبات الطبيعية من قواعد بيانات كيميائية ومصادر علمية مختلفة. بعد ذلك، أُجري تحليل ADME/Tox من أجل تقييم الخصائص الحركية الدوائية، قابلية المركبات لأن تكون أدوية، وسميتها المحتملة. كما أُنجزت دراسات الالتحام الجزيئي والالتحام المتقاطع من أجل التنبؤ بقوة ارتباط كل مركب بالأهداف البيولوجية المختارة. وأتاحت مرحلة تحليل ما بعد الالتحام تحديد أهم التفاعلات الجزيئية المسؤولة عن استقرار معقدات الليغاند–البروتين، مثل الروابط الهيدروجينية، التفاعلات الكارهة للماء، وتفاعلات.π–π .

تقدم نتائج هذه الدراسة الحاسوبية معلومات مهمة حول طريقة تفاعل المركبات الطبيعية مع عدة أهداف مضادة للسرطان. كما تبرز أهمية المحاكاة الجزيئية كأداة سريعة واقتصادية في المراحل الأولى لاكتشاف أدوية جديدة مضادة للسرطان. ويمكن أن تشكل المركبات الواعدة نقطة انطلاق لدراسات لاحقة تشمل المحاكاة بالديناميكا الجزيئية، التحسين البنيوي، والتحقق التجريبي من النشاط المضاد للسرطان من خلال دراسات in vitro و in vivo.

الكلمات المفتاحية: السرطان، المركبات الطبيعية، المحاكاة الجزيئية، الالتحام الجزيئي، ADME/Tox، الأهداف المضادة للسرطان..

..

List of Abbreviations

List of Abbreviations

Abbreviation(s)	Full Term
ADME	Absorption, Distribution, Metabolism, and Excretion
ADME/Tox	Absorption, Distribution, Metabolism, Excretion, and Toxicity
ADMET	Absorption, Distribution, Metabolism, Excretion, and Toxicity
ADP	Adenosine Diphosphate
AKT	Protein Kinase B
Ala	Alanine
AR	Androgen Receptor
Arg	Arginine
Asn	Asparagine
Asp	Aspartate / Aspartic Acid
ATP	Adenosine Triphosphate
BA Score	Bioavailability Score
Bax	Bcl-2-Associated X Protein
BBB	Blood–Brain Barrier
Bcl-2 / BCL-2 / BCL2	B-cell Lymphoma 2
BCL-201 / BCL201	Reference BCL-2 inhibitor used in the study
Bcl-xL	B-cell Lymphoma Extra Large
BH3	BCL-2 Homology Domain 3
BIOVIA DS	BIOVIA Discovery Studio
BOILED-Egg	Brain Or IntestinaL EstimateD permeation model
Cardio / Cardiotox.	Cardiotoxicity
Carcino.	Carcinogenicity
CDK	Cyclin-Dependent Kinase
CDKs	Cyclin-Dependent Kinases
CDK4	Cyclin-Dependent Kinase 4
CDK4/6	Cyclin-Dependent Kinases 4 and 6
ChEMBL	Chemical European Molecular Biology Laboratory Database
Clinical Tox.	Clinical Toxicity
CNS	Central Nervous System
COCONUT	COlleCtion of Open NatUral prodUcTs
Cpd	Compound
Cys	Cysteine
CYP	Cytochrome P450
CYP Inh.	Cytochrome P450 Inhibition
CYP450	Cytochrome P450

List of Abbreviations

Abbreviation(s)	Full Term
CYP1A2	Cytochrome P450 Family 1 Subfamily A Member 2
CYP2C19	Cytochrome P450 Family 2 Subfamily C Member 19
CYP2C9	Cytochrome P450 Family 2 Subfamily C Member 9
CYP2D6	Cytochrome P450 Family 2 Subfamily D Member 6
CYP3A4	Cytochrome P450 Family 3 Subfamily A Member 4
DNA	Deoxyribonucleic Acid
E_score1	First Docking Score, London dG
E_score2	Second Docking Score, GBVI/WSA dG
ECFP	Extended-Connectivity Fingerprint
EGCG	Epigallocatechin Gallate
EGFR	Epidermal Growth Factor Receptor
ER	Estrogen Receptor
ESOL	Estimated Solubility
G0/G1	Gap 0 / Gap 1 Phases
G2/M	Gap 2 / Mitosis Phase
GBVI/WSA	Generalized Born Volume Integral / Weighted Surface Area
GI	Gastrointestinal
GI Abs.	Gastrointestinal Absorption
Gln	Glutamine
Glu	Glutamate / Glutamic Acid
Gly	Glycine
HA	Hyaluronic Acid
HBA	Hydrogen Bond Acceptor
HBD	Hydrogen Bond Donor
Hepato / Hepatotox.	Hepatotoxicity
HER2	Human Epidermal Growth Factor Receptor 2
HIF-1α	Hypoxia-Inducible Factor-1 Alpha
His	Histidine
Ile	Isoleucine
Immuno.	Immunotoxicity
irAEs	Immune-Related Adverse Events
kcal/mol	Kilocalories per Mole
LD50	Median Lethal Dose
Leu	Leucine
Lip.	Lipinski
Lip. Viol.	Lipinski Violations

List of Abbreviations

Abbreviation(s)	Full Term
LogP / logP	Logarithm of the Partition Coefficient; lipophilicity descriptor
Lys	Lysine
MACCS	Molecular ACCess System structural keys/fingerprints
MAPK	Mitogen-Activated Protein Kinase
Met	Methionine
MMP	Matrix Metalloproteinase
MOE	Molecular Operating Environment
Muta.	Mutagenicity
MW	Molecular Weight
NAD^+	Nicotinamide Adenine Dinucleotide, oxidized form
NC	Natural Compound
Neuro / Neurotox.	Neurotoxicity
NF-κB	Nuclear Factor Kappa B
Nutritional Tox.	Nutritional Toxicity
PARP / PARP1	Poly(ADP-ribose) Polymerase / Poly(ADP-ribose) Polymerase 1
PDB	Protein Data Bank
P-gp / Pgp	P-glycoprotein
Phe	Phenylalanine
PI3K	Phosphoinositide 3-Kinase
PK	Pharmacokinetic
PR	Progesterone Receptor
Pro	Proline
PubChem	Public Chemical Database, NCBI
π–π	Pi–pi stacking interaction
RAS	Rat Sarcoma Viral Proto-oncogene
RB1	Retinoblastoma Gene 1
RCSB	Research Collaboratory for Structural Bioinformatics
RDKit	Open-source cheminformatics toolkit
Ref.	Reference
Resp.	Respiratory Toxicity
RMSD	Root Mean Square Deviation
RMSDrefine	Refinement Root Mean Square Deviation
ROS	Reactive Oxygen Species
RTK	Receptor Tyrosine Kinase
S	Docking Score, usually GBVI/WSA dG in kcal/mol
SA Score	Synthetic Accessibility Score

List of Abbreviations

Abbreviation(s)	Full Term
SAR	Structure–Activity Relationship
Ser	Serine
SMILES	Simplified Molecular Input Line Entry System
STAT3	Signal Transducer and Activator of Transcription 3
Sub.	Substrate
Tanimoto	Tanimoto Similarity Coefficient
Thr	Threonine
Tox	Toxicity
TP53	Tumour Protein 53
TPSA	Topological Polar Surface Area
TRAIL	TNF-Related Apoptosis-Inducing Ligand
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
VEGF	Vascular Endothelial Growth Factor
VEGFR / VEGFR2	Vascular Endothelial Growth Factor Receptor / Vascular Endothelial Growth Factor Receptor 2
WlogP	Wildman–Crippen logP, lipophilicity descriptor
Wnt	Wingless-related Integration Site

List of Figures

List of Figures

CHAPTER I

Overview of Cancer and Anticancer Drug Discovery Strategies

Figure I.1 Hallmarks of Cancer and Their Biological Significance (adapted from[1]) 4
Figure I.2 Schematic representation of the role of a HA-rich microenvironment in cancer progression. 5
Figure I.3: Comparison between single-target activity of synthetic drugs and multi-target (pleiotropic) activity of natural compounds (adapted from [9 10
Figure I.4 General chemical structure of Alkaloids 11
Figure I.5: Chemical structure of paclitaxel 13
Figure I.6: Taxus species(source of paclitaxel 15
Figure I.7: Catharanthus roseus(source of vincristine 15
Figure I.8: Ecteinascidia turbinata, marine tunicate producing anticancer compounds 16
Figure I.9 Streptomyces species 17
Figure I.10: Chemical structure of curcumin 23
Figure I.11: Chemical structure of Resveratrol 24
Figure I.12: Chemical structure of EGCG 24

CHAPTER II

Materials And Methods

Figure II. 1 Three-dimensional structure of the VEGFR receptor. 30
Figure II. 2 Three-dimensional structure of the AR receptor. 31
Figure II. 3 Three-dimensional structure of the BCL-2 receptor. 32
Figure II. 4 Three-dimensional structure of the PARP1 receptor. 33
Figure II. 5 Three-dimensional structure of the CDK4 receptor. 34
Figure II. 6 RCSB Protein Data Bank Homepage. 35
Figure II.7 Homepage of the COCONUT (COlleCtion of Open NatUral prodUcTs) database used for the construction of the MPA13 ligand library (Source: COCONUT database, https://coconut.naturalproducts.net). 36
Figure II. 8 Chemical Structure of Sorafenib. 39
Figure II. 9 Chemical Structure of Enzalutamide 40
Figure II. 10 Chemical Structure of BCL-201. 40
Figure II. 11 Chemical Structure of Rucaparib. 40
Figure II. 12 Chemical Structure of Palbociclib. 41
Figure II. 13 SwissADME Web Home Page. 47
Figure II. 14 Tox-Prediction Web Home Page. 48
Figure II. 15 MOE Docking Application Logo. 54
Figure II. 17 Binding site definition using MOE Site Finder. 56

List of Figures

Figure II. 18 Docking setup of selected ligands in MOE. 57
Figure II. 19 Docking score results generated by MOE for the selected ligands..... 58
Figure II. 20 Visualization of ligand–receptor interactions using BIOVIA Discovery Studio. 59

CHAPTER III

RESULTS AND DISCUSSION

Figure III. 1 VEGFR similarity scores of top compound 64
Figure III. 2 AR similarity scores of top compounds. 66
Figure III. 3 BCL similarity scores of top compounds. 67
Figure III. 4 PARP similarity scores of top compounds. 68
Figure III. 5 Top CDK Similar Compounds. 69
Figure III. 6 Maximum Similarity Scores Across Targets. 72
Figure III. 7 Toxicity heatmap of AR candidate compounds. 74
Figure III. 8 BCL2 Candidates Toxicity Heatmap. 80
Figure III. 9 Cross-target comparison of mean TPSA and consensus LogP across all five compound libraries..... 81
Figure III. 10 Methylxanthin Compounds 82
Figure III. 11 Heatmap of predicted active toxicity endpoints among toxicity-evaluated candidates. 83
Figure III. 12 Multi-Parameter Physicochemical Profile Candidate Compounds..... 85
Figure III. 13 MOE Docking Scores of AR Candidates vs.Enzalutamide. 90
Figure III. 14 MOE Docking Scores of BCL2 Candidates vs.BCL201. 90
Figure III. 15 MOE Docking Scores of CDK4 Candidates vs.Palbociclib 91
Figure III. 16 MOE Docking Scores of VEGFR2 Candidates vs.Sorafenib. 91
Figure III. 17 MOE Docking Scores of PARP1 Candidates vs.Rucaparib. 92
Figure III. 18 Compound targets occurrence matrix distribution of recurrent phytochemicals 93
Figure III. 19 Binding Interaction Maps of Selected MPA13Compounds with AR 94
Figure III. 20 Interaction Types in AR Ligand Complexes. 96
Figure III. 21 Observed Amino Acid Residues in AR Ligand Complexes. 97
Figure III. 22 Interaction Types in BCL2 Ligand Complexes. 99
Figure III. 23 Binding Interaction Maps of Selected MPA13 Compounds with BCL2. 100
Figure III. 24 Observed Amino Acid Residues in BCL2 Ligand Complexes. 101
Figure III. 25 Observed Intreraction Types in CDK4 Ligand Interaction. 102
Figure III. 26 Binding Interaction Maps of Selected MPA13 Compounds with CDK4. 103
Figure III. 27 Observed Amino Acid Residues in CDK4 Ligand Interaction. 104
Figure III. 28 Observed Interaction Types in PARP1 Ligand Complexes. 106
Figure III. 29 Binding Interaction Maps of Selected MPA13 Compounds with PARP1. 106
Figure III. 30 Observed Amino Acid Residues in PARP1 Ligand Complexes. 108

List of Figures

Figure III. 31 Observed Interaction Types in VEGFR2 Ligand Complexes. 110
Figure III. 32 Binding Interaction Maps of Selected MPA13 Compounds with VEGFR2…….110
Figure III. 33 Observed Amino Acid Residues in VEGFR2 Ligand Complexes.....113

List of Tables

Table of content

CHAPTER I

Overview of Cancer and Anticancer Drug Discovery Strategies

Table I.1 Summary of Main Anticancer Therapies and Their Limitations..... 6
Table I.2 Examples of Anticancer Drugs Derived from Natural Products 9

CHAPTER III

RESULTS AND DISCUSSION

Table III. 1 Number of compounds retained after TanimotoMACCS-based similarity filtering. 62
Table III. 2 Final ranking of MPA13 compounds based on Tanimoto MACCS similarity against sorafenib for VEGFR-2. 63
Table III. 3Representative of some multitarget compounds identified during similarity screening. 71
Table III. 4 Physicochemical and pharmacokinetic properties of AR candidate compounds. 73
Table III. 5 CYP isoform inhibition profiles and solubility classifications for AR candidates. 73
Table III. 6 Predicted toxicity endpoints for AR candidates. 74
Table III. 7 Descriptive physicochemical statistics for the CDK4 candidate library. 75
Table III. 8 Drug-likeness and absorption filter outcomes for the CDK4 candidate library 75
Table III. 9 Selected CDK4 candidates with optimal ADME profiles 76
Table III. 10 Toxicity prediction results for nine CDK-evaluated candidates. 77
Table III. 11 ADME properties of selected PARP candidate compounds 78
Table III. 12 Toxicity predictions for PARP1 candidates with dedicated toxicity assessment. 78
Table III. 13 ADME profiles of VEGFR candidate compounds versus sorafenib. 79
Table III. 14 Toxicity predictions for VEGFR-evaluated candidates. 79
Table III. 15 Descriptive ADME statistics for the BCL candidate library (n = 170) . 80
Table III. 16 Cross-target ADME and drug-likeness comparison summary. 81
Table III. 17 Top five pharmacokinetically prioritised candidates based on combined ADMET criteria 84
Table III. 18 Docking Scores Summary Across All Targets 86
Table III. 19 Cross-target summary and global priority ranking of screened natural compounds. 88

Table of content

Table of content

CHAPTER I

Overview of Cancer and Anticancer Drug Discovery Strategies

Part 1. Cancer and Anticancer Drug Discovery: General Background 4

1.1. Overview of Cancer Biology 4

1.2. Molecular Mechanisms of Cancer Development 5

1.3. Conventional Anticancer Therapies and Their Limitations..... 6

1.4. Role of Natural Products in Anticancer Drug Discovery 8

1.5. Advantages of Natural Compounds as Anticancer Agents..... 9

Part 2. Natural Anticancer Compounds 11

2.1. Classification of Natural Anticancer Compounds 11

2.1.1. Alkaloids..... 11

2.1.2. Polyphenols and Flavonoids 12

2.1.3. Terpenoids..... 12

2.1.4. Other Bioactive Natural Compounds..... 13

2.2. Sources of Natural Anticancer Compounds..... 14

2.2.1. Plants..... 14

2.2.2. Marine Organisms 16

2.2.3. Microorganisms 17

2.3. Mechanisms of Anticancer Activity..... 18

2.3.1. Induction of Apoptosis..... 18

2.3.2. Cell Cycle Arrest..... 18

2.3.3. Inhibition of Angiogenesis and Metastasis..... 19

Part 3. Anticancer Targets and Computational Approaches in Drug Discovery

3.1. Anticancer Targets..... 20

3.1.1. Kinases..... 20

3.1.2. Hormone Receptors 21

3.1.3. Anti-apoptotic Proteins (Bcl-2 Family) 21

3.1.4. DNA Repair Enzymes..... 22

3.1.5. Cell Cycle Regulators..... 22

3.2. Natural Compounds with Anticancer Potential..... 23

3.3. Computational Drug Design Concepts..... 25

3.3.1. ADME/Tox In Silico 25

3.3.2. Molecular Docking..... 25

Table of content

3.3.3. Cross-Docking and Multi-Target Screening 26

CHAPTER II

Materials And Methods

1. Dataset Preparation 28

1.1 Receptor Selection and Preparation 29

1.2 Construction of the MPA13 Ligand Dataset 35

1.3 Ligand Preparation and Optimization 37

2. Molecular Similarity Analysis 38

2.1 Rationale of the Similarity Analysis 10

2.2 Molecular Fingerprint Generation 41

2.3 Similarity Calculation Using the Tanimoto Coefficient 42

2.4 Python Script and RDKit Implementation 43

2.5 Similarity-Based Comparison with Reference Drugs 44

2.6 Integration with the Global Screening Workflow 45

3. ADME/Tox *In Silico* Analysis 45

3.1 Tools Used for ADME/Tox Prediction 46

3.2 Pharmacokinetics Evaluation 48

3.3 Toxicity Prediction 50

3.4 Comparison with Reference Drugs 51

4. Molecular Docking 52

4.1 Software Used 53

4.2 Preparation of Receptor and Ligand Files 54

4.3 Binding Site Definition 54

4.4 Docking Procedure and Interaction Analysis 56

5. Post-Docking Analysis: 59

Table of content

CHAPTER III

RESULTS AND DISCUSSION

1.Similarity Analysis Results 62

1.1 VEGFR-2 similarity analysis against sorafenib..... 63

1.2 AR similarity analysis against Enzalutamide 65

1.3 BCL-2 similarity analysis against BCL-201..... 66

1.4 PARP-1 similarity analysis against Rucaparib..... 67

1.5 CDK-4similarity analysis against pablociclib..... 68

1.6 Multitarget Similarity Profile..... 69

2. ADME/Tox Results: 73

2.1 ADME/TOX Evaluation of AR Candidates 73

2.2 ADME/TOX Evaluation of CDK4 Candidates..... 74

2.3 ADME/TOX Evaluation of PARP1 Candidates 77

2.4 ADME/TOX Evaluation of VEGFR2 Candidates..... 78

2.5 ADME/TOX Evaluation of BCL2 Candidates 79

2.6 Cross-Target Comparative Analysis..... 80

2.7 Discussion..... 81

3. Docking Results..... 84

3.1 Docking Score Matrix..... 85

3.2 Ranking of Compounds..... 87

3.3 Comparison with Known Inhibitors 88

3.4 Recurrence Analysis 91

4. Interaction Analysis 93

4.1 Interaction Analysis of AR–Ligand Complexes..... 93

4.1.1 Description of binding modes 93

4.1.2 Key amino acids involved 96

4.1.3 Structure–activity relationship elements 98

4.2 Interaction Analysis of BCL-2 –Ligand Complexes..... 99

4.2.1 Description of binding modes..... 99

4.2.2 Key amino acids involved 101

4.2.3 Structure–activity relationship elements 101

4.3 Interaction Analysis of CDK4 –Ligand Complexes.....102
4.3.1 Description of binding modes.....103
4.3.2 Key amino acids involved.....104
4.3.3 Structure–activity relationship elements.....104
4.4 Interaction Analysis of PARP1 –Ligand Complexes.....105
4.4.1 Description of binding modes.....107
4.4.2 Key amino acids involved.....107
4.4.3 Structure–activity relationship elements.....108
4.5 Interaction Analysis of VEGFR2–Ligand Complexes.....109
4.5.1 Description of binding modes.....111
4.5.2 Key amino acids involved.....112
4.5.3 Structure-activity relationship elements.....113

Introduction

Cancer remains one of the leading causes of mortality worldwide and represents a major challenge in modern medicine. It is characterized by uncontrolled cell proliferation, resistance to apoptosis, and the ability to invade and metastasize. Despite significant progress in cancer therapy, conventional treatments such as chemotherapy often suffer from limitations including toxicity, lack of selectivity, and the emergence of drug resistance. These challenges highlight the urgent need for the discovery of safer and more effective anticancer agents.

In this context, natural compounds have gained increasing interest due to their structural diversity and wide range of biological activities. Many plant-derived molecules can modulate key cancer-related pathways such as cell signaling, apoptosis, DNA repair, and cell cycle regulation. Several clinically used anticancer drugs are either derived from or inspired by natural products, emphasizing their importance in drug discovery.

At the molecular level, cancer progression is driven by several validated therapeutic targets, including VEGFR2, androgen receptor (AR), BCL-2, PARP1, and CDK4. These proteins are involved in essential biological processes such as angiogenesis, hormone signaling, apoptosis control, DNA repair, and cell cycle progression. Their dysregulation contributes significantly to tumor development, making them key targets for anticancer drug design.

Computational approaches have become essential tools in modern drug discovery, allowing the prediction of drug–target interactions in a fast and cost-effective manner. Techniques such as ADME/Tox prediction, molecular docking, and cross-docking provide valuable insights into the pharmacokinetic properties and binding affinities of bioactive compounds before experimental validation.

The present study aims to investigate the interactions between selected natural anticancer compounds and multiple cancer-related targets (VEGFR2, AR, BCL-2, PARP1, and CDK4) using molecular simulation approaches. The study includes ADME/Tox evaluation, molecular docking, cross-docking, and post-docking interaction analysis to identify promising multi-target inhibitors.

To achieve these objectives, this dissertation is organized into three chapters: Chapter 1 presents the literature review on cancer biology, anticancer targets, natural compounds, and computational drug design; Chapter 2 describes the materials and methods used; and Chapter 3 presents and discusses the results obtained from similarity, ADME/Tox, docking, and interaction analyses.

CHAPTER I

Overview of Cancer and Anticancer Drug Discovery Strategies

Part 1. Cancer and Anticancer Drug Discovery: General Background

1.1. Overview of Cancer Biology

Cancer represents a heterogeneous group of diseases characterized by uncontrolled cell proliferation, the ability to invade surrounding tissues, and, in many cases, the capacity to metastasize to distant organs. At its core, cancer arises from the accumulation of genetic and epigenetic alterations that disrupt normal cellular homeostasis. These alterations affect key regulatory pathways governing cell growth, differentiation, and programmed cell death.

Normal cells maintain a balance between proliferation and apoptosis through tightly regulated signaling networks. In contrast, cancer cells acquire a set of functional capabilities commonly referred to as the "hallmarks of cancer"(Figure I.1), including sustained proliferative signaling, resistance to cell death, replicative immortality, angiogenesis, and immune evasion [1]. These hallmarks are not independent; rather, they interact dynamically within the tumor microenvironment, which includes stromal cells, immune cells, extracellular matrix components, and signaling molecules.

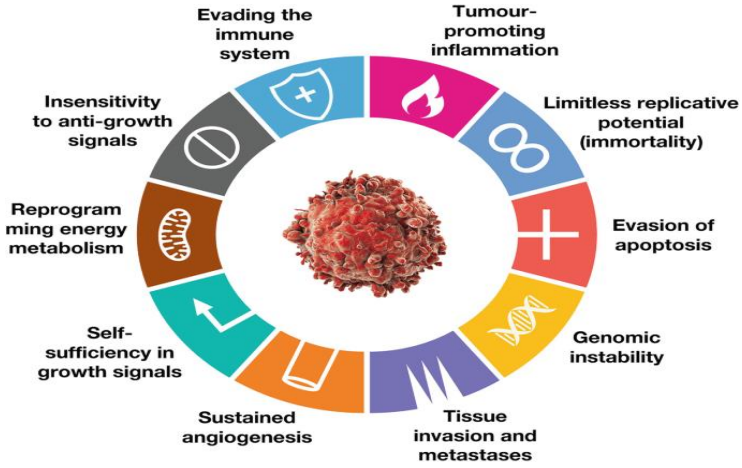


Figure 0I.1: Hallmarks of Cancer and Their Biological Significance (adapted from [1]).

The tumor microenvironment, particularly when enriched with components such as Hyaluronic Acid (HA) (Figure I.2), plays a critical role in cancer progression. It not only supports tumor growth but also contributes to therapeutic resistance. Increasing evidence highlights that cancer should not be viewed solely as a cell-autonomous

disease but as a complex, evolving system influenced by both intrinsic genetic changes and extrinsic environmental factors [2].

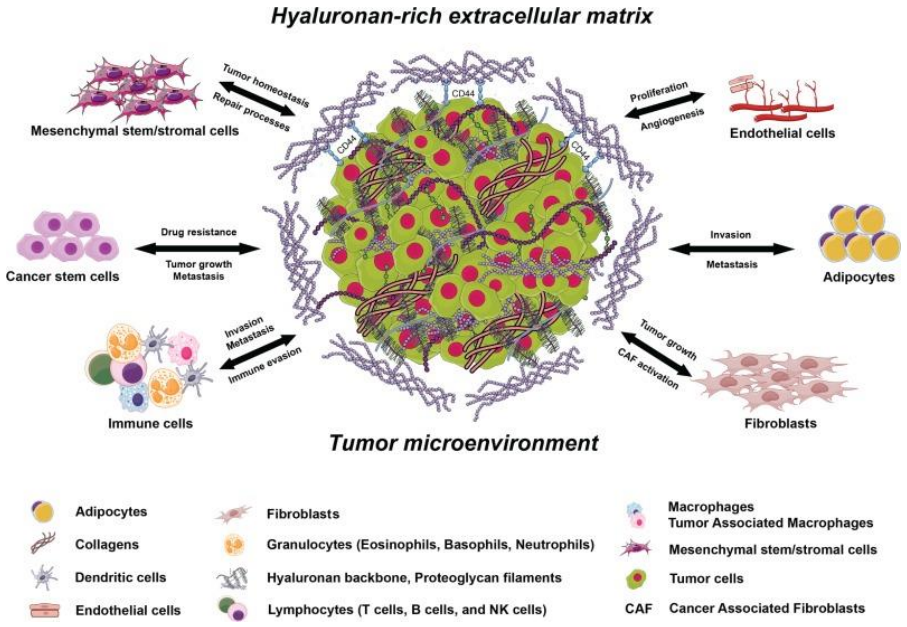


Figure I.2: Schematic representation of the role of a HA-rich microenvironment in cancer progression.

1.2. Molecular Mechanisms of Cancer Development

Cancer development is a multistep process involving the progressive accumulation of molecular alterations. These changes primarily affect three major classes of genes: oncogenes, tumor suppressor genes, and genes involved in DNA repair.

Oncogenes arise from the activation of proto-oncogenes, leading to excessive cell proliferation. Examples include mutations in signaling pathways such as RAS and PI3K/AKT, which drive continuous growth signals independent of external stimuli. Conversely, tumor suppressor genes, such as TP53 and RB1, normally act to inhibit cell cycle progression or induce apoptosis. Their inactivation removes critical regulatory checkpoints, allowing damaged cells to survive and proliferate [3].

Another crucial mechanism involves genomic instability, which accelerates mutation rates and facilitates tumor evolution. Defects in DNA repair pathways, such as mismatch repair or homologous recombination, contribute significantly to this instability. In addition to genetic mutations, epigenetic modifications—including DNA

methylation and histone modification—alter gene expression without changing the DNA sequence, further promoting tumorigenesis [4].

Furthermore, dysregulation of cell signaling pathways plays a central role in cancer progression. Pathways such as MAPK, Wnt/β-catenin, and NF-κB are frequently altered, leading to enhanced proliferation, survival, and metastasis. The interplay between these pathways underscores the complexity of cancer as a systemic disease rather than a single-pathway disorder.

1.3. Conventional Anticancer Therapies and Their Limitations

Conventional cancer treatments include surgery, chemotherapy, and radiotherapy, which remain the cornerstone of clinical oncology. While these approaches have significantly improved patient outcomes, they are often associated with substantial limitations.

The principal conventional anticancer therapies and their major limitations are summarized in TableI.1.

Table II.1: Summary of Main Anticancer Therapies and Their Limitations

Therapy Type	Mechanism of Action	Main Limitations
Surgery	Physical removal of tumor mass	Invasive, limited to accessible/localized tumors, risk
Chemotherapy	Kills rapidly dividing cells	Systemic toxicity, lack of selectivity, drug resistance.
Radiotherapy	Induces DNA damage via ionizing	Damage to adjacent healthy tissues, limited to localized
Targeted Therapy	Blocks specific molecular targets	Development of resistance, effective only for specific
Immunotherapy	Enhances host immune response	Autoimmunity (irAEs), variable response rates, high costs.

Chemotherapy relies on cytotoxic agents that target rapidly dividing cells. However, its lack of selectivity leads to damage of healthy tissues, resulting in adverse effects such as myelosuppression, gastrointestinal toxicity, and hair loss. Moreover, cancer cells can develop resistance through multiple mechanisms, including drug efflux, mutation of drug targets, and activation of survival pathways [5].

Radiotherapy, which uses ionizing radiation to induce DNA damage, is effective in localized tumors but may also harm surrounding normal tissues. Similarly, surgical intervention is limited to accessible tumors and may not prevent recurrence or metastasis.

In recent years, targeted therapies and immunotherapies have emerged as promising alternatives. Targeted therapies aim at specific molecular abnormalities, while immunotherapies harness the immune system to recognize and eliminate cancer

cells. Despite their success in certain cancers, these approaches are not universally effective and can still lead to resistance and high treatment costs [6].

Overall, the limitations of conventional therapies highlight the urgent need for novel, safer, and more effective anticancer agents.

1.4. Role of Natural Products in Anticancer Drug Discovery

Natural products have historically played a pivotal role in drug discovery, particularly in oncology. A significant proportion of clinically used anticancer drugs are either natural compounds or their derivatives. Classic examples include paclitaxel, derived from the Pacific yew tree, and vincristine, isolated from *Catharanthus roseus* [7].

The structural diversity and biological activity of natural compounds make them an invaluable source for identifying new therapeutic agents. Unlike synthetic libraries, natural products often possess complex three-dimensional architectures that enable specific interactions with biological targets. These interactions can modulate multiple signaling pathways simultaneously, which is particularly advantageous in a multifactorial disease such as cancer.

In recent years, advances in analytical chemistry, high-throughput screening, and computational approaches—such as molecular docking and molecular dynamics simulations—have significantly enhanced the exploration of natural compounds. These technologies allow for the identification and optimization of bioactive molecules with improved efficacy and reduced toxicity [8].

Furthermore, natural products frequently serve as lead compounds for the development of semi-synthetic derivatives with enhanced pharmacological properties. Their continued relevance in modern drug discovery underscores their importance in the search for novel anticancer therapies.

Several clinically used anticancer drugs are derived from natural products, demonstrating the potential of these molecules to modulate critical cellular processes in cancer. Examples of such natural compounds and their mechanisms of action are summarized in Table **I**.2.

Table I.2: Examples of Anticancer Drugs Derived from Natural Products

Drug	Natural Source	Main Mechanism
Paclitaxel	Plant	Microtubule stabilization
Doxorubicin	Microorganism	DNA intercalation
Vincristine	Plant	Mitotic arrest
Camptothecin derivatives	Plant	Topoisomerase inhibition

1.5. Advantages of Natural Compounds as Anticancer Agents

Natural compounds offer several advantages over conventional synthetic drugs in the context of cancer treatment. One of their key strengths lies in their ability to target multiple pathways simultaneously. This pleiotropicmulti-target (Figure I.3) activity is particularly beneficial in overcoming the complexity and heterogeneity of cancer.

In addition, many natural compounds exhibit relatively low toxicity toward normal cells compared to traditional chemotherapeutic agents. This selective cytotoxicity can reduce adverse side effects and improve patient quality of life. Compounds such as curcumin, resveratrol, and epigallocatechin-3-gallate (EGCG)have demonstrated promising anticancer properties with favorable safety profiles [9].

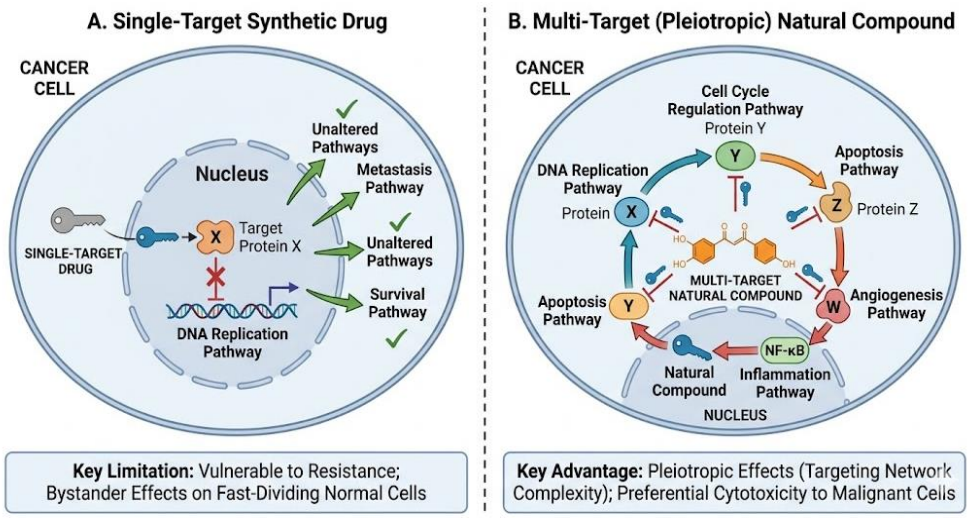


Figure I.3: Comparison between single-target activity of synthetic drugs and multi-target (pleiotropic) activity of natural compounds (Adapted from [9]).

Another advantage is their potential to circumventdrug resistance. By modulating various molecular pathways, natural compounds can sensitize cancer cells to existing therapies or inhibit resistance mechanisms. This makes them attractive candidates for combination therapies.

Moreover, the increasing integration of computational methods, including molecular simulation techniques, has accelerated the identification of natural compounds with high binding affinity to specific biological targets. These approaches provide valuable insights into molecular interactions and facilitate rational drug design.

Despite these advantages, challenges remain, including issues related to bioavailability, stability, and standardization. Nevertheless, ongoing research continues to address these limitations, reinforcing the potential of natural products as a cornerstone of future anticancer drug discovery.

Part 2. Natural Anticancer Compounds

2.1. Classification of Natural Anticancer Compounds

Natural anticancer compounds comprise a broad and chemically diverse group of bioactive molecules derived from living organisms. Their classification is generally based on chemical structure, biosynthetic origin, and biological activity. Among the most studied groups are alkaloids, polyphenols and flavonoids, terpenoids, and several other classes of secondary metabolites with recognized antitumor potential. This diversity is one of the principal reasons why natural products continue to occupy a central place in anticancer drug discovery. Their structural complexity often enables selective interactions with proteins involved in cell survival, proliferation, angiogenesis, and metastatic dissemination [7, 8, 10].

2.1.1. Alkaloids

Alkaloids are nitrogen-containing natural compounds that are mainly produced by plants, although some are also obtained from microorganisms and marine organisms (Figure I.4). They represent one of the most important classes of natural products in oncology because several clinically used anticancer agents belong to this family. Vinca alkaloids, such as vincristine and vinblastine, are classical examples that act by disrupting microtubule dynamics and inhibiting mitosis. Other alkaloids, including camptothecin derivatives, exert their cytotoxic effects through inhibition of topoisomerase I, thereby interfering with DNA replication and transcription [7, 10].

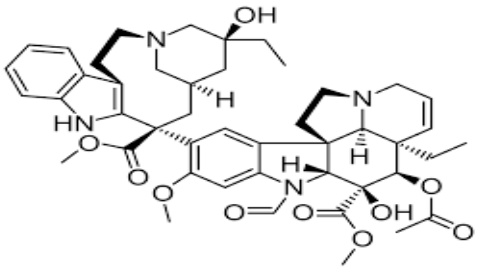


Figure I.4 General chemical structure of Alkaloids

Beyond these established agents, many alkaloids under preclinical investigation exhibit broader mechanisms of action, including induction of apoptosis, modulation of

oxidative stress, suppression of inflammatory signaling, and interference with metastatic pathways. Their pharmacological relevance is closely linked to their ability to affect essential cellular machinery with high potency. However, issues such as toxicity, poor water solubility, and narrow therapeutic windows is consistent with limit direct clinical application, which is why many alkaloids serve as templates for structural optimization and semisynthetic development [10, 11].

2.1.2. Polyphenols and Flavonoids

Polyphenols constitute a large family of plant-derived compounds characterized by the presence of multiple phenolic groups. Flavonoids are a major subclass of polyphenols and include compounds such as quercetin, luteolin, kaempferol, apigenin, and epigallocatechin gallate. These molecules are widely distributed in fruits, vegetables, tea, cocoa, and medicinal plants, and they have attracted considerable attention because of their antioxidant, anti-inflammatory, and anticancer properties [8, 9].

Their anticancer activity is not limited to a single mechanism. Polyphenols and flavonoids may modulate intracellular signaling pathways involved in proliferation, apoptosis, autophagy, angiogenesis, and invasion. In many experimental models, they regulate transcription factors such as NF-κB, STAT3, and HIF-1α, as well as kinase-driven pathways including PI3K/AKT and MAPK. This multi-target behavior is especially relevant in cancer, where redundant signaling networks often contribute to treatment failure [8, 12].

Although numerous polyphenols show promising in vitro and in vivo effects, their clinical translation remains constrained by low bioavailability, rapid metabolism, and limited systemic stability. Even so, these limitations have encouraged the development of improved formulations and delivery systems, reinforcing their importance as lead compounds in oncology research [12, 13].

2.1.3. Terpenoids

Terpenoids, also referred to as isoprenoids, form one of the largest classes of natural products. They are biosynthesized from isoprene units and include

monoterpenes, sesquiterpenes, diterpenes, triterpenes, and tetraterpenes. Several terpenoids have shown remarkable anticancer potential, either as approved drugs or as promising leads. Paclitaxel (Figure I.5), one of the most significant natural-product-based anticancer drugs, is a diterpenoid that stabilizes microtubules and blocks cell division [7].

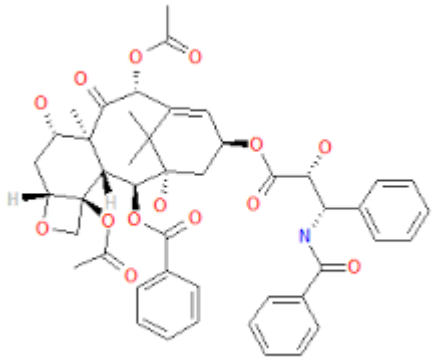


Figure I.5 Chemical structure of Paclitaxel

Other terpenoids, such as betulinic acid, artemisinin derivatives, celastrol, and withaferin-like compounds, have demonstrated the capacity to inhibit tumor growth through induction of apoptosis, generation of oxidative stress, mitochondrial dysfunction, and suppression of metastatic signaling. Their structural variability allows interaction with numerous molecular targets, which may be advantageous in tumors characterized by complex pathway dysregulation [11, 14].

Despite their biological promise, some terpenoids face pharmaceutical limitations comparable to those of other natural compounds, particularly poor aqueous solubility and variable pharmacokinetics. Nonetheless, their broad biological spectrum continues to make them attractive candidates for anticancer drug development [8, 14].

2.1.4. Other Bioactive Natural Compounds

In addition to the major classes described above, several other groups of natural compounds contribute to the anticancer repertoire. These include saponins, quinones, lignans, coumarins, peptides, polyketides, and sulfur-containing compounds. Many of these metabolites have shown strong cytotoxicity or pathway-specific anticancer effects in preclinical studies [10, 15].

Polyketides, for example, are of particular interest because numerous clinically relevant drugs originate from microbial polyketide biosynthesis. Likewise, certain natural peptides and depsipeptides isolated from marine organisms and microorganisms have demonstrated potent antiproliferative activity. The importance of these "other" classes lies not only in their pharmacological activity but also in their capacity to expand the chemical space available for drug discovery. In many cases, their mechanisms differ from those of standard chemotherapeutics, making them especially valuable in the search for compounds capable of overcoming resistance [7,15 ,16].

2.2. Sources of Natural Anticancer Compounds

Natural anticancer compounds are obtained from a wide range of biological sources. Historically, terrestrial plants provided the earliest and most extensively studied molecules. However, marine biodiversity and microbial metabolism have become increasingly important in recent decades. Each source offers distinct chemical scaffolds, biosynthetic pathways, and pharmacological opportunities [7, 8, 15].

2.2.1. Plants

Plants remain the most recognized and widely exploited source of natural anticancer agents. Their secondary metabolites are produced as part of defense, signaling, and adaptation mechanisms, and many of these molecules exhibit potent effects on mammalian cells. The discovery of paclitaxel from *Taxus* species (Figure I.6) and vincristine from *Catharanthus roseus* (Figure I.7) clearly illustrates the contribution of plants to modern cancer therapy [7].



Figure I.6 Taxus species (source of paclitaxel).



Figure I.7 Catharanthus roseus (source of vincristine).

In addition to approved drugs, a vast number of plant-derived molecules such as curcumin, resveratrol, quercetin, berberine, and genistein have shown promising anticancer properties in experimental models. Their actions include suppression of proliferation, induction of apoptosis, modulation of inflammation, and inhibition of migration and invasion. Because plants synthesize an exceptional diversity of alkaloids, flavonoids, terpenoids, and phenolic compounds, they continue to be a major reservoir for identifying new lead structures [8, 12, 13].

The increasing interest in plant-derived anticancer agents is also supported by advances in extraction techniques, phytochemical profiling, and computational

screening. These tools have made it possible to study plant metabolites more systematically and to relate their chemical properties to specific biological targets [8, 15].

2.2.2. Marine Organisms

Marine organisms have emerged as a highly promising source of anticancer compounds because the marine environment drives the production of structurally unique metabolites that are rarely found in terrestrial organisms. Sponges, tunicates, algae, mollusks, and marine microorganisms produce compounds with unusual ring systems, halogenated structures, and potent biological activities (Figure I.8). This chemical originality has made the marine environment a valuable frontier in oncology drug discovery [15, 16].



Figure I.8: Ecteinascidia turbinata, marine tunicate producing anticancer compounds.

Several marine-derived compounds have already reached clinical application or advanced developmental stages, demonstrating that marine metabolites are not merely chemically interesting but also pharmacologically relevant. These compounds may act through diverse mechanisms, including microtubule interference, DNA damage induction, apoptosis activation, and inhibition of angiogenic or metastatic signaling [15].

Recent research has expanded beyond isolated compounds to integrated strategies involving metabolomics, genome mining, and synthetic biology, which are helping to identify and optimize marine bioactives more efficiently. As a result, marine organisms

are increasingly recognized as an indispensable source of novel anticancer scaffolds [15, 16].

2.2.3. Microorganisms

Microorganisms, particularly bacteria and fungi, represent another major source of anticancer natural products. Their biosynthetic capacity is extraordinary, and many clinically important drugs have originated from microbial metabolites or their derivatives. Actinomycetes, especially *Streptomyces* species (Figure I.9), are notable producers of structurally diverse secondary metabolites with strong cytotoxic and antitumor properties [16, 17].

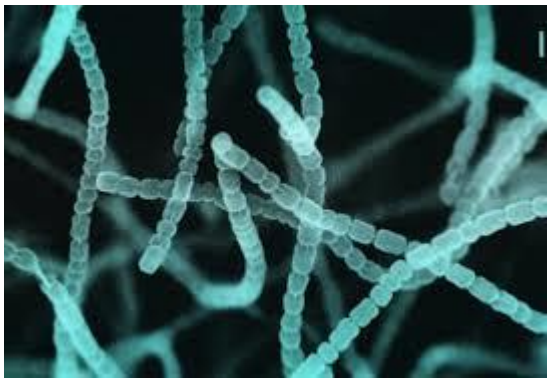


Figure I.9: *Streptomyces* species

Fungal metabolites also provide a rich reservoir of compounds with antiproliferative, proapoptotic, and anti-invasive effects. One of the key advantages of microorganisms is that they can be cultured, genetically manipulated, and explored using genome-based approaches. This makes them highly attractive for scalable drug discovery. Recent technological advances, including metagenomics, biosynthetic gene cluster analysis, and pathway engineering, have further expanded access to previously silent or cryptic microbial metabolites [16-18].

Because microbial compounds often possess mechanisms distinct from conventional cytotoxic agents, they are particularly relevant in the context of multidrug resistance and target diversification. Their contribution to anticancer drug discovery is therefore both historical and forward-looking [16-18].

2.3. Mechanisms of Anticancer Activity

Natural anticancer compounds exert their effects through multiple molecular and cellular mechanisms. Unlike many single-target synthetic drugs, natural molecules frequently act on interconnected signaling pathways, which is consistent with enhance therapeutic efficacy in heterogeneous tumors. Among the most important mechanisms are the induction of apoptosis, arrest of the cell cycle, and inhibition of angiogenesis and metastasis [8, 12, 19].

2.3.1. Induction of Apoptosis

Apoptosis is a tightly regulated form of programmed cell death essential for tissue homeostasis. One of the defining features of cancer is the ability of malignant cells to evade apoptosis, thereby surviving despite genetic damage and adverse microenvironmental conditions. Restoring apoptotic signaling is therefore a major objective in anticancer therapy [1, 19].

Many natural compounds promote apoptosis through intrinsic and extrinsic pathways. The intrinsic pathway is generally associated with mitochondrial dysfunction, increased Bax/Bcl-2 ratio, cytochrome c release, and activation of caspase-9 and caspase-3. The extrinsic pathway involves death receptors such as Fas and TRAIL receptors, leading to caspase-8 activation. Several alkaloids, flavonoids, and terpenoids have been shown to trigger one or both pathways depending on cell type and molecular context [10, 12, 14, 19].

In addition, some natural compounds induce apoptosis indirectly by increasing reactive oxygen species, causing DNA damage, or inhibiting survival pathways such as PI3K/AKT and NF-κB. This broad mechanistic spectrum explains why apoptosis induction remains one of the most frequently reported anticancer effects of natural metabolites [12, 19].

2.3.2. Cell Cycle Arrest

Uncontrolled proliferation is a hallmark of cancer, and disruption of cell cycle progression is one of the most effective ways to limit tumor growth. Natural compounds

can interfere with the cell cycle at different checkpoints, notably G0/G1, S, and G2/M phases, depending on their molecular targets [1, 12].

Alkaloids such as vinca derivatives and terpenoids such as paclitaxel classically affect mitotic spindle function, producing arrest at the G2/M phase. Other natural compounds regulate cyclins, cyclin-dependent kinases, and checkpoint proteins including p21, p27, and p53. Through these actions, they prevent DNA replication, inhibit mitotic progression, and reduce cancer cell proliferation [7, 10, 14].

Cell cycle arrest is often closely linked to apoptosis. When checkpoint activation is prolonged or when DNA damage cannot be repaired, arrested cells may subsequently undergo programmed death. This functional connection makes cell cycle modulation a particularly valuable anticancer strategy for natural compounds acting on proliferative signaling [12, 19].

2.3.3. Inhibition of Angiogenesis and Metastasis

Tumor progression depends not only on uncontrolled proliferation but also on the ability to establish a vascular supply and to disseminate beyond the primary site. Angiogenesis provides oxygen and nutrients to growing tumors, while metastasis is responsible for the majority of cancer-related deaths. Natural compounds have shown significant potential in targeting both processes [1, 20].

Several phytochemicals suppress angiogenesis by downregulating proangiogenic mediators such as vascular endothelial growth factor (VEGF), hypoxia-inducible factor-1 alpha (HIF-1α), and matrix metalloproteinases. Others interfere with endothelial cell migration and tube formation, thereby limiting neovascularization. These effects can reduce tumor growth and may also improve responsiveness to conventional therapy [12, 20].

Natural compounds may inhibit metastasis through multiple mechanisms, including reversal of epithelial-mesenchymal transition, suppression of cell adhesion and motility, inhibition of extracellular matrix degradation, and modulation of inflammatory and survival signaling. Because invasion and metastasis involve

numerous coordinated events, the multitarget nature of many natural products is especially advantageous in this context [12, 20].

Taken together, these mechanisms demonstrate that natural anticancer compounds do not act merely as general cytotoxins. Rather, they influence several critical stages of tumor development and progression, which explains their enduring relevance in anticancer drug discovery and supports their investigation through molecular simulation approaches.

Part3. Anticancer Targets and Computational Approaches in Drug Discovery.

3.1. Anticancer Targets

The identification of appropriate molecular targets is a central step in anticancer drug discovery. Cancer progression is driven by dysregulated signaling networks, and targeting key proteins involved in these pathways can effectively inhibit tumor growth and survival. Among the most extensively studied anticancer targets are kinases, hormone receptors, anti-apoptotic proteins, DNA repair enzymes, and cell cycle regulators.

3.1.1. Kinases

Protein kinases are enzymes that regulate cellular processes through phosphorylation of target proteins. They play a crucial role in signal transduction pathways controlling proliferation, survival, differentiation, and metabolism. In cancer, aberrant activation of kinases—often due to mutations or overexpression—leads to sustained proliferative signaling, one of the fundamental hallmarks of malignancy [1].

Key oncogenic kinases include receptor tyrosine kinases (RTKs) such as EGFR and HER2, as well as intracellular kinases involved in pathways like PI3K/AKT and MAPK. Targeting these kinases has led to the development of numerous small-molecule inhibitors. However, resistance mechanisms, including secondary mutations and pathway redundancy, remain major challenges. Natural compounds have shown

potential to modulate kinase activity either directly or indirectly, often acting on multiple nodes within signaling networks [8, 13].

3.1.2. Hormone Receptors

Hormone-dependent cancers, such as breast and prostate cancer, rely on signaling mediated by hormone receptors, including estrogen receptors (ER), progesterone receptors (PR), and androgen receptors (AR). These receptors function as transcription factors that regulate gene expression in response to hormonal stimuli.

Targeting hormone receptors has been a successful therapeutic strategy, as seen with selective estrogen receptor modulators and androgen deprivation therapies. Nevertheless, resistance frequently develops through receptor mutations, altered co-regulator expression, or activation of alternative signaling pathways [6].

Certain natural compounds, particularly phytoestrogens such as genistein and resveratrol, can interact with hormone receptors and modulate their activity. These compounds may act as agonists or antagonists depending on cellular context, highlighting their potential as modulators of hormone-driven cancers [12, 13].

3.1.3. Anti-apoptotic Proteins (Bcl-2 Family)

The Bcl-2 family of proteins plays a central role in regulating apoptosis, particularly through control of mitochondrial membrane permeability. Anti-apoptotic members, such as Bcl-2 and Bcl-xL, are often overexpressed in cancer cells, allowing them to evade programmed cell death.

Targeting these proteins has become an important therapeutic approach. Inhibition of Bcl-2 can restore apoptotic signaling and sensitize cancer cells to treatment. Several natural compounds have demonstrated the ability to modulate Bcl-2 family proteins by increasing pro-apoptotic factors (e.g., Bax) and decreasing anti-apoptotic expression [10, 19].

This modulation contributes significantly to the pro-apoptotic effects discussed in Part 2 and reinforces the importance of apoptosis-related targets in anticancer strategies.

3.1.4. DNA Repair Enzymes

DNA repair systems are essential for maintaining genomic stability. In cancer cells, defects in repair pathways can both promote tumorigenesis and create therapeutic vulnerabilities. Enzymes involved in pathways such as base excision repair, nucleotide excision repair, and homologous recombination are therefore important drug targets.

Inhibitors of DNA repair enzymes, such as PARP inhibitors, exploit the concept of synthetic lethality by selectively targeting cancer cells with pre-existing repair deficiencies. Natural compounds have also been shown to interfere with DNA repair mechanisms, either by directly inhibiting repair enzymes or by inducing DNA damage that overwhelms repair capacity [3, 13].

This dual role of DNA repair—as both a protective and exploitable system—makes it a particularly attractive target in precision oncology.

3.1.5. Cell Cycle Regulators

Cell cycle progression is tightly controlled by cyclins, cyclin-dependent kinases (CDKs), and checkpoint proteins. Dysregulation of these components leads to uncontrolled cell division, a defining characteristic of cancer [1].

Targeting CDKs and associated regulatory proteins has become an effective strategy in cancer therapy. Natural compounds have been shown to influence cell cycle progression by modulating the expression or activity of cyclins, CDKs, and inhibitors such as p21 and p27. These effects can result in arrest at specific phases of the cell cycle, ultimately reducing tumor proliferation [12, 19].

The ability of natural compounds to simultaneously affect cell cycle regulators and apoptosis-related proteins highlights their potential as multitarget anticancer agents.

3.2. Natural Compounds with Anticancer Potential

A wide range of natural compounds have demonstrated significant anticancer activity through diverse molecular mechanisms. These compounds often act on multiple targets simultaneously, making them particularly effective against complex and heterogeneous tumors.

Curcumin (Figure I.10), a polyphenolic compound derived from *Curcuma longa*, is one of the most extensively studied natural anticancer agents. It modulates numerous signaling pathways, including NF-κB, PI3K/AKT, and MAPK, and induces apoptosis through mitochondrial pathways. Curcumin also inhibits angiogenesis and metastasis, highlighting its broad spectrum of activity [9, 12].

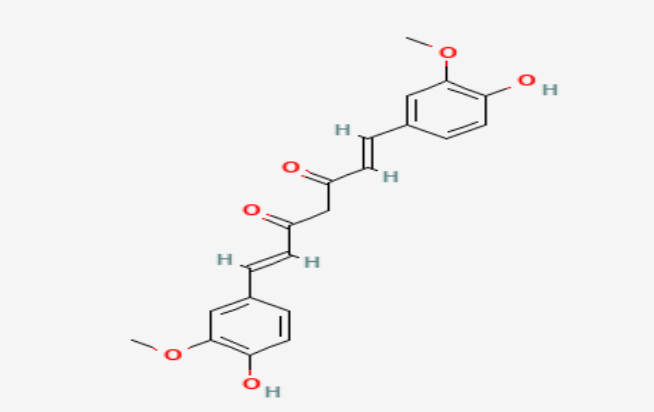


Figure I.10Chemical structure of curcumin.

Resveratrol (**Figure I.11)**, a stilbene found in grapes and berries, exhibits anticancer effects by regulating cell cycle progression, promoting apoptosis, and inhibiting inflammation and oxidative stress. It has also been shown to interact with hormone receptors and influence epigenetic regulation [12, 13].

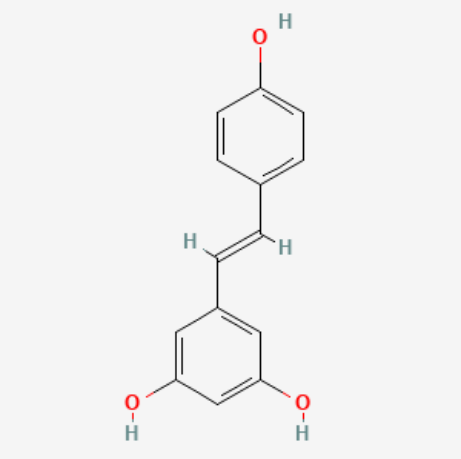


Figure I.11 Chemical structure of Resveratrol .

Epigallocatechin gallate (EGCG) (**Figure I.12**), a major component of green tea, demonstrates strong antioxidant and anticancer properties. It can inhibit tumor growth by modulating signaling pathways, suppressing angiogenesis, and inducing apoptosis [9, 12].

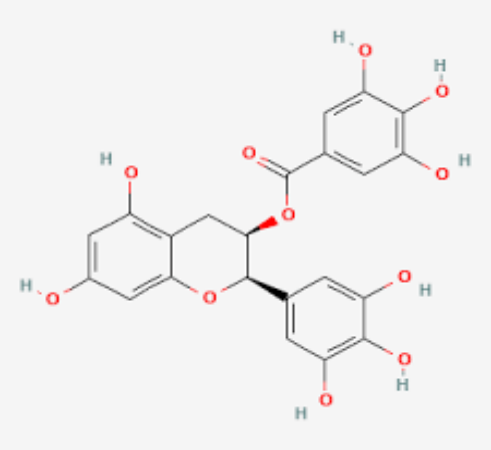


Figure I.12 Chemical structure of EGCG .

Alkaloids such as berberine and camptothecin derivatives exert their anticancer effects through DNA interaction, inhibition of topoisomerases, and activation of apoptotic pathways. Similarly, terpenoids such as paclitaxel and betulinic acid disrupt microtubule function and mitochondrial integrity, respectively [7, 10, 14].

These examples illustrate that natural compounds can target a wide range of biological processes, including proliferation, survival, angiogenesis, and metastasis.

Their structural diversity and multitarget activity make them valuable candidates for both standalone therapies and combination strategies [8,13].

3.3. Computational Drug Design Concepts

Advances in computational methods have transformed the field of drug discovery by enabling the rapid identification and optimization of bioactive compounds. In the context of natural products, computational approaches are particularly valuable for screening large compound libraries, predicting biological activity, and understanding molecular interactions at the atomic level.

3.3.1. ADME/Tox In Silico

The success of a drug candidate depends not only on its biological activity but also on its pharmacokinetic and toxicity profiles. ADME (Absorption, Distribution, Metabolism, and Excretion) and toxicity (Tox) properties are critical determinants of drug efficacy and safety.

In silico ADME/Tox prediction tools allow early evaluation of these properties, reducing the likelihood of late-stage failure. These models estimate parameters such as oral bioavailability, membrane permeability, metabolic stability, and potential toxicity. For natural compounds, which often suffer from poor bioavailability, these predictions are particularly important for guiding structural optimization and formulation strategies [21].

3.3.2. Molecular Docking

Molecular docking is a computational technique used to predict the binding mode and affinity of a ligand toward a target protein. It involves positioning the ligand within the binding site and evaluating interactions based on geometric and energetic criteria.

Docking algorithms generate multiple possible conformations, and scoring functions are used to estimate binding affinity. These scoring functions typically consider factors such as hydrogen bonding, hydrophobic interactions, electrostatic forces, and van der Waals interactions. Although docking does not provide exact

binding energies, it offers valuable insights into ligand–target interactions and helps prioritize compounds for further study [21, 22].

In the context of natural products, docking is widely used to identify potential targets, validate experimental findings, and explore structure–activity relationships.

3.3.3. Cross-Docking and Multi-Target Screening

Cancer is a multifactorial disease involving multiple dysregulated pathways. Consequently, the concept of targeting a single protein is often insufficient. Cross-docking and multi-target screening approaches address this limitation by evaluating the interaction of a compound with multiple targets.

Cross-docking involves testing a ligand against several protein structures to assess binding versatility, while multi-target screening aims to identify compounds capable of modulating multiple pathways simultaneously. These approaches are particularly relevant for natural compounds, which frequently exhibit multitarget activity.

By integrating computational screening with experimental validation, researchers can identify compounds with improved efficacy and reduced resistance potential. This strategy aligns well with the complex biology of cancer and supports the development of more effective therapeutic agents [8, 21, 22].

CHAPTER II

Materials And Methods

1. Dataset Preparation

In structure-based drug discovery, dataset preparation represents a critical foundation rather than a secondary technical step, as the reliability of molecular docking outcomes is highly dependent on the quality of the input data. Inaccurate selection or improper preparation of protein targets and ligand structures may lead to misleading binding poses and consequently incorrect biological interpretations. Cancer biology is characterized by high genomic complexity and dynamic molecular alterations, which further increases the importance of selecting biologically relevant targets [1,3]. In addition, cancer progression is strongly associated with adaptive mechanisms of resistance that can affect drug–target interactions and therapeutic efficacy [5,6].

For this reason, careful curation of both receptor and ligand datasets is essential prior to virtual screening. The ligand library must be chemically consistent, structurally reliable, and suitable for computational processing to ensure meaningful docking results. In this context, molecular docking remains a key computational tool in modern drug discovery, enabling the prediction of ligand–target interactions and supporting rational drug design strategies [22]. Moreover, recent advances in computational approaches, particularly those involving natural product-based drug discovery, have further expanded the diversity and potential of anticancer lead identification [13].

The strategy adopted in this work was based on the idea that cancer is not a single-pathway disease. Instead, tumor progression results from the interaction of multiple altered biological processes, including sustained proliferative signaling, resistance to apoptosis, genomic instability, treatment resistance, and abnormal communication with the tumor microenvironment [1–6]. Consequently, limiting the screening campaign to only one target would not fully reflect the complexity of cancer biology. To address this point, a diversified panel of receptors was selected, representing different mechanistic axes involved in cancer development and therapeutic response. At the same time, the ligand dataset was constructed from a natural product-oriented source in order to explore the chemical diversity that natural compounds continue to offer in anticancer drug discovery [7–10,13].

1.1 Receptor Selection and Preparation

The first step in receptor dataset preparation consisted of identifying protein targets that are not only therapeutically relevant, but also structurally appropriate for docking studies. In this context, five receptors were selected: VEGFR2, AR, BCL-2, PARP1, and CDK4. These targets were chosen because they represent major processes that support cancer cell survival and progression, namely angiogenesis, hormone-dependent signaling, evasion of apoptosis, DNA damage repair, and cell-cycle dysregulation. The choice of this receptor panel was therefore intended to provide a broader biological framework for evaluating the anticancer potential of the screened compounds, rather than restricting the study to a single signaling pathway [1,3,5,6].

Among the selected targets, VEGFR2 (vascular endothelial growth factor receptor 2)(Figure II. 1) was retained because of its central role in tumor angiogenesis. In solid tumors, continued growth depends on the capacity of cancer cells to secure oxygen and nutrient supply through the formation of new blood vessels. VEGFR2 is one of the main mediators of this angiogenic response, and its activation promotes endothelial proliferation, vascular permeability, and neovascularization. From a therapeutic point of view, inhibition of VEGFR2 has long been recognized as an effective strategy to restrict tumor nourishment and slow cancer progression. This makes VEGFR2 particularly relevant in screening studies focused on identifying compounds with antiangiogenic potential [1,20]. Its selection in the present work also contributes to the biological diversity of the target panel by representing the vascular dimension of tumor development.

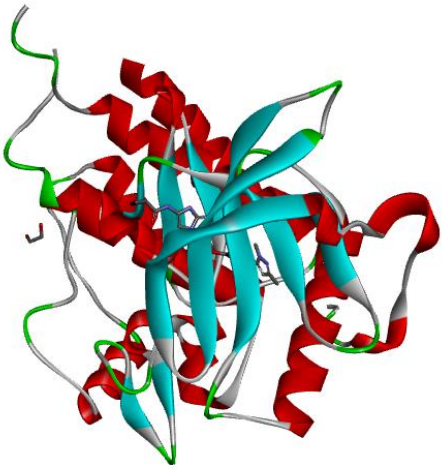


Figure II. 1Three-dimensional structure of the VEGFR receptor.

The androgen receptor (AR) (**Figure II. 2)** was selected because hormone signaling remains a key driver of prostate cancer progression. AR functions as a nuclear receptor that regulates the transcription of genes involved in cellular survival, differentiation, and proliferation. Although androgen deprivation therapy often produces an initial clinical benefit, tumor cells can progressively adapt and reactivate AR signaling through several resistance mechanisms, making this receptor a persistent therapeutic target even in advanced disease stages [5,6]. The inclusion of AR in this study is therefore important because it introduces the endocrine component of cancer biology into the screening strategy. In other words, it allows the virtual screening workflow to address not only general proliferative mechanisms, but also ligand-dependent transcriptional pathways that are especially relevant in hormone-sensitive malignancies.



Figure II. 2 Three-dimensional structure of the AR receptor.

A third receptor, BCL-2 (B-cell lymphoma-2) (**Figure II. 3)**, was included because the ability of cancer cells to evade programmed cell death is one of the best-established hallmarks of cancer [1]. BCL-2 is an anti-apoptotic protein that preserves mitochondrial integrity and protects malignant cells from apoptosis, thereby facilitating long-term survival and resistance to treatment. In many cancers, especially hematological malignancies, high BCL-2 activity contributes to the persistence of tumor cells despite therapeutic pressure. For this reason, BCL-2 is considered a meaningful target in studies seeking compounds capable of restoring apoptotic sensitivity [19]. Its inclusion strengthens the mechanistic balance of the receptor panel, since it addresses not the proliferative side of cancer alone, but also the survival strategies that allow malignant cells to escape elimination.

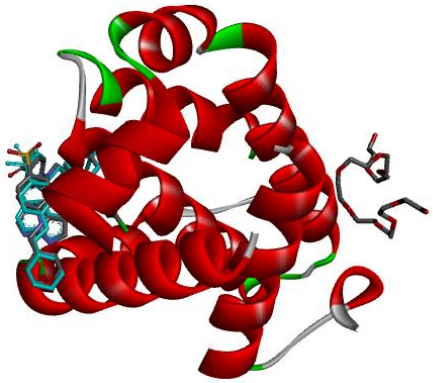


Figure II. 3 Three-dimensional structure of the **BCL-2** receptor.

The receptor PARP1 (poly[ADP-ribose] polymerase 1) (**Figure II. 4)** was selected because it is deeply involved in DNA repair pathways. Cancer cells often rely on compensatory repair mechanisms to tolerate genomic instability and continue proliferating despite accumulating DNA damage [3]. PARP1 plays a crucial role in the repair of single-strand DNA breaks, and inhibition of this enzyme can lead to the collapse of DNA repair capacity, particularly in tumors already impaired in homologous recombination. This therapeutic concept has become highly relevant in ovarian and breast cancers, where PARP inhibition has shown considerable value. The inclusion of PARP1 in the present study therefore introduces the dimension of genome maintenance and repair vulnerability into the screening design, which is highly relevant in modern anticancer research.

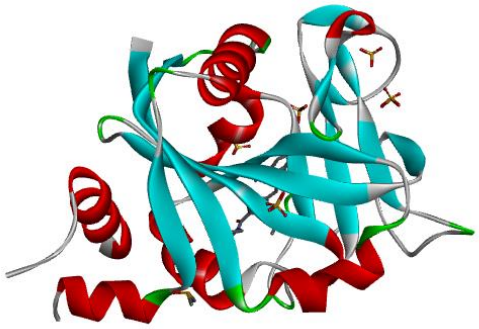


Figure II. 4 Three-dimensional structure of the **PARP1** receptor**.**

Finally, CDK4 (cyclin-dependent kinase 4) (**Figure II. 5)** was retained because deregulation of the cell cycle is one of the most fundamental events in tumor progression. CDK4 contributes to the transition from the G1 phase to the S phase of the cell cycle and, when abnormally activated, promotes uncontrolled proliferation. In several cancers, especially breast cancer, the cyclin D–CDK4 axis is closely linked to sustained cell division and poor growth control. For this reason, CDK4 has become an important therapeutic target in oncology. Including CDK4 in the receptor set provides a direct way to evaluate whether the screened compounds may interfere with one of the core proliferative engines of malignant cells [1]. Taken together, the five selected receptors create a coherent yet biologically diverse panel, allowing the study to explore multiple vulnerable points in cancer progression rather than approaching the disease from a single perspective.

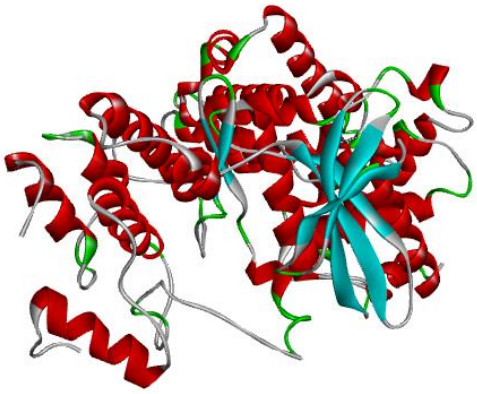


Figure II. 5 Three-dimensional structure of the CDK4 receptor.

After defining the biological relevance of the receptor set, the next step consisted of selecting suitable three-dimensional protein structures. For this purpose, receptor coordinates were retrieved from the RCSB Protein Data Bank (PDB) (**Figure II. 6)**, which is the principal international repository for experimentally determined biomolecular structures [23]. Structure selection was not random. Whenever possible, preference was given to human proteins, X-ray crystallographic structures, acceptable resolution, and well-defined active sites, since these characteristics increase confidence in docking experiments. When several PDB entries were available for the same target, priority was given to structures that best represented the active binding region and, when feasible, to those associated with a co-crystallized ligand, because such entries often provide clearer information about the geometry of the binding pocket.

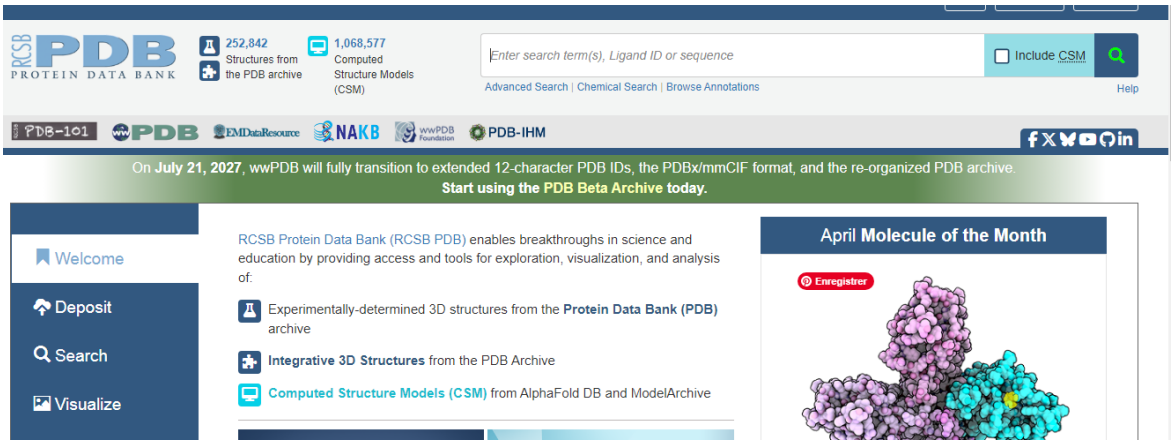


Figure II. 6 RCSB Protein Data Bank Homepage.

Once selected, the receptor structures were prepared before docking in order to remove elements that were not directly relevant to ligand-binding analysis. This preparation usually involves cleaning the structure, excluding unnecessary molecules such as excess solvent or unrelated heteroatoms, checking the integrity of the protein, and keeping the biologically relevant binding region. Such steps are essential because docking does not depend only on the scoring function, but also on the structural quality of the receptor model itself [22]. In this study, receptor preparation was therefore considered an integral part of methodological rigor rather than a routine formal step.

1.2 Construction of the MPA13 Ligand Dataset

In parallel with receptor preparation, the ligand dataset used in this work was constructed with the same level of care. To support the virtual screening stage, an in-house phytochemical library named MPA13 was established as the principal ligand dataset of the study. MPA13 was derived from the COCONUT (COlleCtion of Open NatUral prodUcTs) database (Figure II.7), which is an open-access resource specifically developed for natural product research [24]. The use of COCONUT as the starting point for ligand collection was especially appropriate in the present study because natural products remain one of the richest sources of chemically diverse bioactive molecules and continue to contribute significantly to anticancer drug discovery [7–10,13].

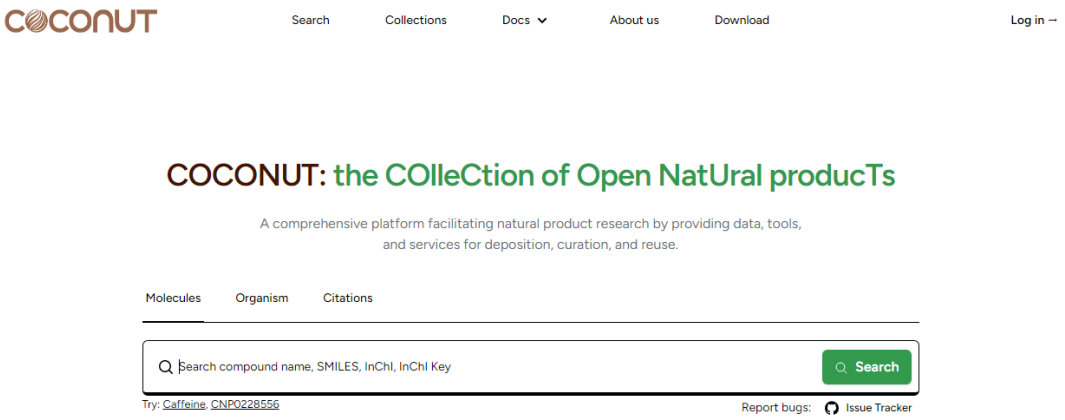


Figure II.7 Homepage of the COCONUT (COlleCtion of Open NatUral prodUcTs) database used for the construction of the MPA13 ligand library (Source: COCONUT database, https://coconut.naturalproducts.net).

The decision to build a dedicated in-house dataset rather than relying on a raw download from a large public source was methodologically important. Large natural product databases provide extensive chemical diversity, but the compounds they contain are not always immediately ready for virtual screening. In many cases, records may include duplicates, heterogeneous annotations, or structures that require standardization before computational use. For this reason, the MPA13 dataset was not treated as a simple collection of molecules, but as a curated working library prepared specifically for the objectives of this study. This approach improves not only the technical quality of the screening process, but also the scientific coherence of the entire methodology.

The construction of MPA13 began with compound retrieval from COCONUT, followed by a curation process designed to improve the consistency and usability of the dataset. Duplicate entries were removed, structural records were standardized, and molecules were examined to ensure that the retained compounds were suitable for downstream in silico analysis. This step was particularly important because even small inconsistencies in molecular representation can influence docking behavior, physicochemical calculations, and the interpretation of biological relevance. By applying this curation phase early in the workflow, the final ligand dataset became more robust and more reliable as a screening resource.

To reinforce the quality of the dataset, additional annotation was verified using PubChem and ChEMBL, two widely recognized databases in chemical and

pharmacological research [25,26]. PubChem was useful for confirming chemical identifiers, names, and structural representations, while ChEMBL provided an additional layer of information related to compound bioactivity and drug discovery relevance. The combined use of these resources did not replace COCONUT, but rather complemented it. In other words, COCONUT served as the principal source of natural product compounds, whereas PubChem and ChEMBL were used to support chemical validation and annotation. This three-resource strategy gave MPA13 a stronger methodological basis and reduced the risk of working with incompletely characterized entries.

Another advantage of building MPA13 as a named in-house library is that it gives the study a clearer identity and a more organized workflow. Instead of repeatedly referring to "retrieved compounds" in a generic way, the methodology can clearly state that all screened ligands belonged to the curated MPA13 dataset. This improves readability, especially in later sections such as virtual screening, hit selection, and discussion of results. From an academic writing perspective, it also helps the reader follow the progression of the study more easily, since the ligand source remains consistent and explicitly defined throughout the chapter.

1.3 Ligand Preparation and Optimization

After the construction of the MPA13 dateset, the selected ligands were subjected to a preparation phase before docking. This step is essential in computational studies because the predictive value of molecular docking depends heavily on the chemical correctness of ligand structures. Even when compounds are retrieved from reputable databases, they still require preprocessing to ensure compatibility with the docking platform and to minimize structural artifacts. Therefore, ligand preparation in the present study included standardization of chemical structures, hydrogen addition, checking of valence consistency when needed, and geometry optimization before the screening stage [21,22].

The importance of this step goes beyond technical formatting. Proper ligand preparation directly affects the way a molecule fits into the receptor binding site, the orientation it adopts during docking, and the type of interactions predicted with surrounding amino acid residues. An inadequately prepared ligand may lead to unstable

poses, unrealistic contacts, or biased docking scores. For this reason, the preparation phase was considered a necessary bridge between ligand collection and biological interpretation, ensuring that the compounds entering the screening workflow were represented in a chemically meaningful manner.

In addition to structural optimization, it is often useful to evaluate basic physicochemical and pharmacokinetic properties at an early stage of screening. Computational tools such as SwissADME can support this process by providing preliminary information about molecular descriptors, drug-likeness, lipophilicity, and absorption-related parameters [21]. Although such predictions do not replace experimental validation, they are helpful for prioritizing compounds and identifying molecules that combine promising receptor affinity with acceptable drug-like characteristics. In this way, ligand preparation was not limited to geometry correction alone, but formed part of a broader effort to build a credible early-stage screening pipeline.

Overall, the dataset preparation strategy used in this study was based on a clear integration of biological relevance, structural quality, and chemical consistency. The receptor dataset was selected to represent major cancer-related pathways, while the ligand side of the workflow was built around MPA13, an in-house curated natural product dataset derived from COCONUT and supported by PubChem and ChEMBL annotation. By carefully preparing both receptors and ligands before docking, the study established a stronger methodological foundation for virtual screening and improved the reliability of the subsequent computational results [13,21,22].

2. Molecular Similarity Analysis

After the construction and preparation of the MPA13 ligand dataset, a molecular similarity analysis was performed using a Python-based script. This step was introduced as an intermediate stage between dataset preparation and ADME/Tox evaluation. Its main purpose was to compare the phytochemical compounds included in MPA13 with the selected reference anticancer drugs, namely sorafenib (Figure II.8), enzalutamide(Figure II.9), BCL201(Figure II.10), rucaparib(Figure II.11), and palbociclib(Figure II.12). The reference drugs associated with the selected targets were

retrieved from **TargetValidation.org**, which provides curated information on clinically relevant molecules and their therapeutic targets. This comparison helped identify natural compounds that may share important structural features with known therapeutic agents before moving to pharmacokinetic, toxicity, and docking analyses.

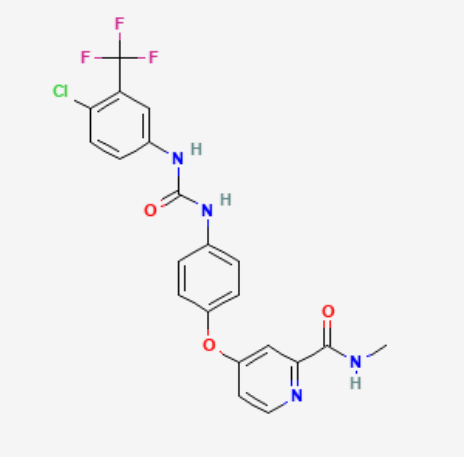


Figure II. 8 Chemical Structure of Sorafenib.

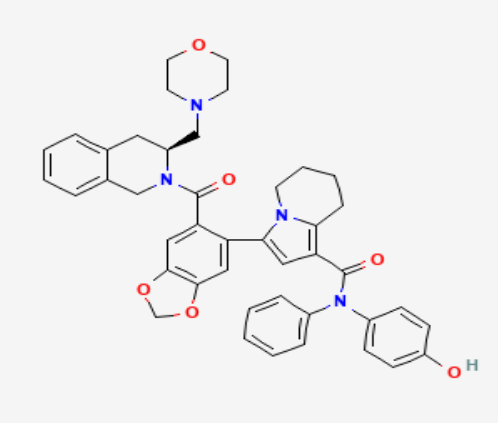


Figure II. 9 Chemical Structure of Enzalutamide .

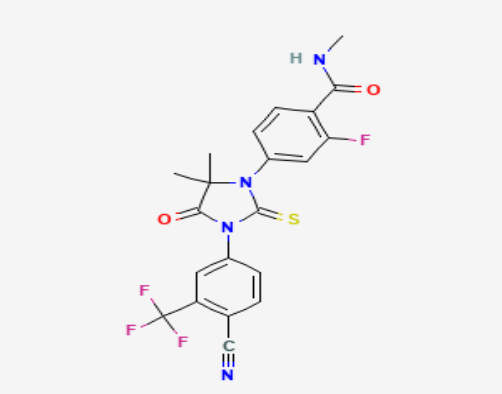


Figure II. 10 Chemical Structure of BCL-201.

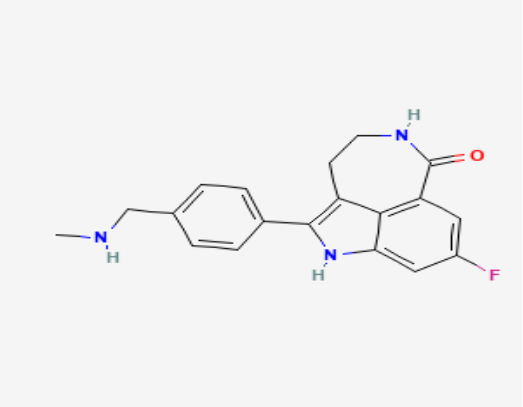


Figure II. 11 Chemical Structure of Rucaparib.

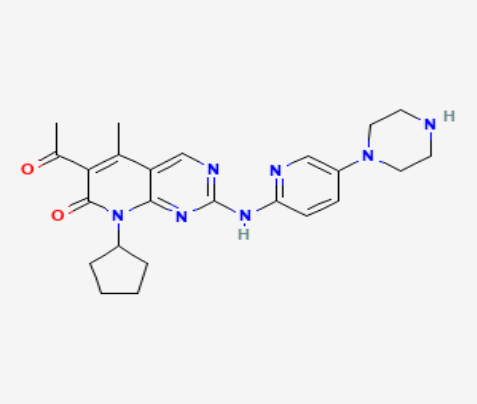


Figure II. 12 Chemical Structure of Palbociclib.

Molecular similarity analysis is based on the general principle that compounds with similar chemical structures may sometimes present related biological profiles. However, this principle was interpreted with caution in the present study. A high similarity value does not prove that a compound has the same biological activity as a

reference drug, and a low similarity value does not necessarily mean that the compound is inactive. A compound may show structural resemblance to a known inhibitor but still have poor pharmacokinetic behavior or weak binding affinity. Instead, similarity analysis was used as a rational prioritization tool to guide the selection of promising compounds from a chemically diverse natural product dataset.

2.1 Rationale of the Similarity Analysis

The MPA13 dataest was derived from natural products, which are known for their structural diversity and chemical complexity. This diversity is valuable in anticancer drug discovery, but it also creates a large chemical space that needs to be explored in an organized way. For this reason, molecular similarity analysis was used to provide an early chemical comparison between the screened phytochemicals and the reference drugs associated with the selected cancer-related targets.

The reference drugs were chosen according to the receptors included in the study. Sorafenib was used as the reference compound for VEGFR2, enzalutamide for the androgen receptor, BCL201 for BCL-2, rucaparib for PARP1, and palbociclib for CDK4. By comparing the MPA13 compounds with these drugs, the analysis aimed to identify molecules that may share chemical patterns with already established anticancer agents. This step therefore helped connect the ligand dataset with the biological objectives of the study.

Nevertheless, similarity analysis was not considered a final selection criterion. It was used as a supportive filter, together with ADME/Tox prediction and molecular docking. This approach is important because natural compounds may act through original scaffolds that are different from synthetic drugs, while still showing relevant biological activity. Therefore, the similarity results were interpreted as a chemical guide rather than as direct evidence of anticancer activity.

2.2 Molecular Fingerprint Generation

To perform the similarity study, the molecular structures of both MPA13 compounds and reference drugs were first represented in a format suitable for cheminformatics analysis, such as SMILES notation or molecular structure files. Invalid

or unreadable structures were excluded before calculation in order to avoid errors during the analysis.

Each molecule was then converted into a molecular fingerprint. Molecular fingerprints are numerical representations of chemical structures that encode the presence of specific atom environments and substructural patterns. In this study, circular fingerprints were used because they are widely applied in ligand-based virtual screening and molecular similarity studies. Extended-connectivity fingerprints are especially useful because they describe the local environment around each atom and allow molecules to be compared according to their structural features [27].

This fingerprint-based representation made it possible to compare compounds from the MPA13 dataset with the selected reference drugs in a standardized and reproducible way. Instead of comparing molecules visually, the Python script transformed each structure into a computational representation that could be used for quantitative similarity calculation.

2.3 Similarity Calculation Using the Tanimoto Coefficient

Several molecular similarity descriptors were generated, including Morgan-based Tanimoto, Morgan-based Dice, MACCS-based Tanimoto, and MACCS-based Dice values. However, the final compound ranking was based on the Tanimoto coefficient calculated using MACCS structural keys. MACCS keys were selected because they describe molecules through predefined structural fragments, making them useful for comparing small molecules according to shared substructural features.

The Tanimoto coefficient is widely used in molecular similarity analysis, particularly for binary molecular fingerprints. It produces values between 0 and 1, where values closer to 1 indicate greater structural similarity and values closer to 0 indicate lower similarity [28,29].

The Tanimoto coefficient can be expressed as follows:

$$Tanimoto = \frac{c}{(a + b - c)}$$

where:

- **a** represents the number of fingerprint features present in the first molecule,
- **b** represents the number of fingerprint features present in the second molecule,
- **c** represents the number of common features shared by both molecules.

Although other similarity metrics and fingerprint types were considered, MACCS-based Tanimoto similarity was selected as the principal ranking criterion because of its robustness, interpretability, and suitability for comparing chemically diverse natural products with reference drugs [27–29]. Morgan fingerprints, such as ECFP, encode detailed circular atomic environments and may therefore penalize structural differences more strongly. In contrast, MACCS keys capture broader substructural patterns, which may be more appropriate for natural-product-based screening, where phytochemicals can differ from synthetic drugs while still retaining relevant functional groups or scaffolds associated with biological activity [7,8,13,27,28].

In this study, each compound from the MPA13 dataset was compared with five reference drugs using MACCS-Tanimoto similarity. Compounds with Tanimoto_MACCS values equal to or greater than 0.50 were retained, allowing the selection of phytochemicals with at least moderate structural similarity while preserving natural products with original chemical scaffolds. The resulting ranked compounds were then prioritized for further ADME/Tox evaluation and molecular docking analysis.

2.4 Python Script and RDKit Implementation

The similarity analysis was carried out using a Python script based on RDKit, an open-source cheminformatics toolkit widely used for molecular handling, fingerprint generation, and similarity calculations [29]. RDKit was suitable for this step because it allows chemical structures to be read, standardized, converted into fingerprints, and compared using similarity metrics within a reproducible computational workflow.

In practice, the script imported the molecular structures of MPA13 compounds and reference drugs, generated circular fingerprints for each molecule, and calculated pairwise Tanimoto similarity scores. The output was organized as a ranked table showing the similarity score of each MPA13 compound against each reference drug. This made it possible to identify, for each therapeutic target, the compounds that were chemically closest to the corresponding reference drug.

The use of Python also made the workflow flexible and reproducible. Since the analysis was script-based, the same procedure could be repeated if new compounds were added to the MPA13 dataset or if additional reference drugs were included later. This reproducibility is important in computational drug discovery because it allows the same criteria to be applied consistently across the entire dataset.

2.5 Similarity-Based Comparison with Reference Drugs

The similarity results were interpreted separately for each reference drug. Compounds showing higher similarity to sorafenib were considered chemically relevant for further evaluation against VEGFR2, since sorafenib is associated with inhibition of angiogenesis-related kinase signaling. In the same way, compounds with higher similarity to enzalutamide were considered interesting for androgen receptor analysis, while those resembling BCL201 were examined in relation to BCL-2.

For PARP1, the comparison with rucaparib was useful because it helped identify compounds that may share structural features with known PARP inhibitors. Similarly, comparison with palbociclib helped highlight compounds that may present chemical characteristics related to CDK4/6 inhibition. This target-based interpretation allowed the similarity analysis to remain connected to the biological design of the study, instead of being treated as a purely numerical comparison.

However, compounds were not selected only because they had the highest similarity scores. A balanced interpretation was adopted. Highly similar compounds were considered important because they may share structural motifs with known drugs, but moderately similar compounds were also kept under consideration when they presented interesting natural-product scaffolds. This was especially relevant because natural products often contain complex ring systems, oxygenated groups, and

stereochemical features that may differ from approved drugs while still supporting biological activity.

2.6 Integration with the Global Screening Workflow

The molecular similarity analysis played an important role in connecting the MPA13 dataset preparation with the next computational steps. After this analysis, the most relevant compounds were advanced to ADME/Tox in silico evaluation, where their pharmacokinetic and toxicity profiles were assessed. Compounds showing acceptable similarity patterns, favorable ADME/Tox properties, and promising docking results were then considered stronger candidates for further interpretation.

Thus, the workflow followed a progressive logic. First, the MPA13 dataset was constructed and curated. Second, the compounds were compared with reference anticancer drugs using Python-based molecular similarity analysis. Third, their pharmacokinetic and toxicity properties were evaluated. Finally, their binding behavior toward selected cancer-related receptors was studied by molecular docking.

This organization strengthened the methodology because compound prioritization was not based on one single parameter. Instead, the study combined chemical similarity, predicted pharmacokinetic behavior, toxicity profile, and receptor binding analysis. In this way, the similarity study contributed to a more rational and structured selection of promising phytochemicals from the MPA13 dataset.

3. ADME/Tox *In Silico* Analysis

Following the preparation of both receptor structures and the MPA13 ligand dataset, an additional filtering stage was necessary in order to evaluate whether the screened compounds could present acceptable pharmacokinetic and toxicological characteristics. Indeed, a strong docking score alone is not sufficient to consider a compound as a promising drug candidate, because many molecules with favorable binding affinity may still fail later due to poor absorption, inadequate bioavailability, rapid metabolism, or undesirable toxicity. For this reason, the compounds selected from the virtual screening workflow were further examined through *in silico* ADME/Tox analysis. This step was introduced as a complementary stage to molecular docking in

order to move from simple receptor–ligand affinity prediction toward a more realistic estimation of the drug development potential of the studied molecules [13,21,22].

In the context of anticancer drug discovery, this evaluation is particularly important. Many anticancer agents are biologically active but remain limited by poor pharmacokinetic behavior or considerable toxicity, which can restrict their clinical usefulness [5,6]. Therefore, integrating ADME/Tox prediction into the workflow makes it possible to identify compounds that are not only active at the molecular level, but also more likely to display a balanced pharmacological profile. In the present study, this stage was performed using two widely used web-based computational tools: **SwissADME** for pharmacokinetic and drug-likeness assessment, and **ProTox-II** for toxicity prediction [21,30].

3.1 Tools Used for ADME/Tox Prediction

The first platform used in this work was SwissADME (Figure II.13), which is a freely accessible tool developed to estimate the physicochemical properties, pharmacokinetic behavior, drug-likeness, and medicinal chemistry friendliness of small molecules [21]. One of the main advantages of SwissADME is that it provides, in a single interface, several descriptors that are especially useful in early-stage screening. These include molecular weight, lipophilicity, polarity, solubility, flexibility, hydrogen-bonding capacity, and other features commonly considered when evaluating the oral suitability of candidate compounds. In addition, SwissADME includes practical predictive models such as the Bioavailability Radar and the BOILED-Egg**,** which support the evaluation of drug-likeness, passive gastrointestinal absorption, and blood–brain barrier permeation. The platform also estimates whether a compound may behave as a P-glycoprotein substrateand whether it is likely to inhibit major cytochrome P450 isoforms, both of which are highly relevant in pharmacokinetic profiling and possible drug–drug interactions [21].



Figure II. 13 SwissADME Web Home Page.

The second platform employed was **ProTox-II (** Figure II.14), a webserver designed for the *in silico* prediction of chemical toxicity [30]. This tool was selected because it allows a rapid preliminary assessment of safety-related endpoints that are particularly useful during early hit prioritization. ProTox-II provides predictions related to **acute oral toxicity**, expressed through estimated **LD50 values** and **toxicity classes**, and it also predicts several biological toxicity endpoints, including **hepatotoxicity, carcinogenicity, mutagenicity, immunotoxicity, and cytotoxicity**. The use of ProTox-II in the present study added an important safety dimension to the screening pipeline by allowing potentially harmful compounds to be recognized at an early stage, even before any experimental validation [30].

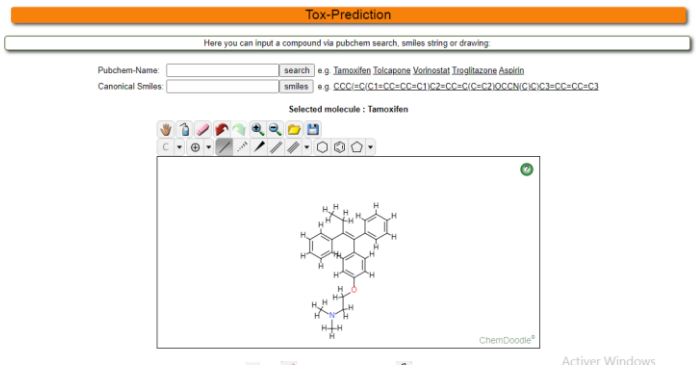


Figure II. 14 Tox-Prediction Web Home Page.

The combined use of SwissADME and ProTox-II was therefore considered appropriate for this work, as it enabled a more integrated and predictive screening strategy. While molecular docking provides insight into the binding affinity and

interaction modes of the selected compounds toward cancer-related targets, ADMET and toxicity predictions allow for the evaluation of their pharmacokinetic profiles and potential safety issues, which is essential in anticancer drug development [21,22,30].

In this context, these computational tools do not replace molecular docking but rather complement it by adding an additional translational layer, improving the biological relevance and interpretability of virtual screening results [22,21].

3.2 Pharmacokinetics Evaluation

The pharmacokinetic analysis was carried out in order to estimate whether the selected compounds possessed molecular characteristics compatible with potential drug development. At this stage, attention was given not only to the classical descriptors of small-molecule behavior, but also to the overall balance between structural complexity and predicted oral suitability. This point is particularly relevant for natural products, since such compounds often display interesting biological activity but may also present high molecular complexity, excessive polarity, or limited solubility, which can reduce their pharmacokinetic performance [7,8,13].

Using SwissADME, the compounds were evaluated according to several physicochemical and pharmacokinetic parameters. These included molecular weight, number of hydrogen bond donors and acceptors, topological polar surface area, number of rotatable bonds, lipophilicity (WLOGP), and water solubility [21]. Together, these descriptors provide a first overview of whether a molecule lies within a chemical space generally considered favorable for oral bioavailability. In addition to these parameters, SwissADME also allows the application of common drug-likeness filters, such as **Lipinski's rule of five**, which remains one of the most frequently used empirical frameworks for estimating oral drug-likeness in the early phases of drug discovery [21]. To further assess oral drug-likeness, **Lipinski's Rule of Five** was applied. This widely used empirical rule proposes that, for a compound to have a high probability of good oral absorption, it should meet the following criteria:

1. **Molecular weight < 500 Da**

2. **Number of hydrogen bond donors ≤ 5**
3. **Number of hydrogen bond acceptors ≤ 10**
4. **logP (lipophilicity) ≤ 5 [21].**

Compounds violating more than one of these rules are generally considered at higher risk of poor oral bioavailability. Lipophilicity, often expressed as WlogP, is a critical component because it reflects the balance between a molecule's lipid and aqueous solubility, which directly affects absorption, distribution, and cellular permeability [21].

Beyond these general descriptors, the pharmacokinetic evaluation also considered **gastrointestinal absorption**, **blood–brain barrier permeability**, **P-glycoprotein interaction**, and possible inhibition of major **CYP450 enzymes**. These parameters are important because they provide a broader view of how a molecule might behave after administration. For example, a compound with good receptor affinity but poor gastrointestinal absorption may have limited usefulness in oral therapy. Likewise, strong inhibition of cytochrome enzymes may indicate a higher risk of metabolic interactions, while P-glycoprotein substrate behavior may affect membrane transport and efflux [21]. In cancer treatment, these properties are especially meaningful because drug resistance and variable exposure are often influenced by transport and metabolism mechanisms [5,6].

For this reason, the pharmacokinetic analysis in the present study was not interpreted as a rigid pass/fail step, but rather as a rational prioritization tool. Compounds showing balanced physicochemical profiles, acceptable absorption-related predictions, and fewer concerning metabolism-related alerts were considered more favorable for further consideration. On the other hand, molecules presenting multiple unfavorable pharmacokinetic features were interpreted with more caution, even when their docking scores were promising. This approach is important because successful lead identification depends not only on molecular recognition of the target, but also on the possibility that a compound may later behave as a developable therapeutic agent [13,21,22].

3.3 Toxicity Prediction

After pharmacokinetic evaluation, toxicity prediction was performed in order to examine the safety profile of the selected compounds at a preliminary computational level. This stage was particularly important because one of the major challenges in anticancer drug development is the narrow therapeutic window often associated with cytotoxic or highly bioactive compounds [5,6]. A molecule may show strong affinity toward an oncogenic target and still be unsuitable if it presents substantial predicted toxicity. Therefore, toxicity prediction was integrated into the present workflow to support a more realistic interpretation of the screened hits.

Using **ProTox-II**, each selected compound was evaluated for estimated **acute oral toxicity**, including the prediction of **LD50** and assignment to a corresponding **toxicity class** [30]. This provides an initial indication of general toxic risk and makes it possible to differentiate between compounds with relatively safer predicted profiles and those that may deserve early exclusion or careful interpretation. In addition to acute toxicity, the platform was used to predict several biologically relevant endpoints, including **hepatotoxicity, mutagenicity, carcinogenicity, immunotoxicity, and cytotoxicity** [30]. These endpoints are highly useful in early computational studies because they offer a first warning regarding possible safety liabilities before moving toward more advanced stages of research.

The interpretation of toxicity prediction was approached cautiously. It was recognized that *in silico* toxicity models do not provide definitive toxicological proof and cannot substitute for experimental assays. However, they remain valuable in the early filtering of chemical libraries because they allow the rapid identification of compounds that may carry structural alerts or toxicity-related concerns. In this study, compounds predicted to have lower toxicity classes and fewer unfavorable toxicity endpoints were considered more attractive for further prioritization, particularly when these properties were combined with acceptable docking and pharmacokinetic results. Conversely, compounds associated with multiple toxicity warnings were regarded as less favorable, even if their target-binding profile appeared strong [13,22,30].

This toxicity filtering step was especially meaningful in the context of natural-product-based screening. Natural compounds are often perceived as inherently safer, but

such an assumption is not always justified. Many natural molecules possess potent biological activity and may also carry toxicological risks depending on their structure, dose, and mechanism of action. For this reason, the use of an independent toxicity prediction stage strengthened the methodological rigor of the study and reduced the risk of overestimating the value of compounds based solely on docking results [8,10,13,30].

3.4 Comparison with Reference Drugs

To place the ADME/Tox predictions into a more meaningful pharmacological context, the selected compounds from the **MPA13** dataset were compared with the **reference drugs** previously associated with the chosen targets, namely **sorafenib**, **enzalutamide**, **BCL201**, **rucaparib**, and **palbociclib**. This comparison was not intended to suggest that a newly screened natural compound must outperform an approved anticancer drug in every parameter. Rather, the purpose was to use clinically relevant molecules as benchmarks in order to better understand whether the predicted properties of the selected hits fell within a plausible and pharmacologically acceptable range.

This comparative approach is important because approved anticancer drugs themselves do not necessarily present ideal pharmacokinetic or toxicological properties. Many of them remain in clinical use despite limitations related to solubility, metabolism, resistance, or adverse effects, because their therapeutic benefit is considered sufficient in a specific clinical context [5,6]. Therefore, when comparing the screened compounds with reference drugs, the objective was not perfection, but **relative positioning**. If a natural candidate displayed a docking profile comparable to that of a reference ligand while also showing acceptable pharmacokinetic behavior and lower predicted toxicity, this would strengthen its value as a promising hit for further investigation.

From a methodological point of view, the comparison focused on several shared indicators. On the pharmacokinetic side, the candidate compounds and reference drugs could be examined in terms of drug-likeness descriptors, gastrointestinal absorption, lipophilicity, and predicted CYP-related interactions. On the toxicological side, their predicted toxicity classes and major toxicity endpoints could be considered side by side. This made it possible to identify whether the screened natural compounds merely

showed theoretical binding affinity, or whether they also possessed a profile that could be viewed as competitive when judged against established therapeutic agents [21,30].

In this way, the ADME/Tox analysis served as a bridge between molecular docking and pharmacological interpretation. After receptor selection, ligand curation, and docking evaluation, this step helped refine the pool of compounds by highlighting those that combined biological relevance with more balanced predicted behavior. As a result, the integration of SwissADME and ProTox-II into the workflow improved the depth of compound prioritization and provided a more complete basis for selecting the most promising candidates from the MPA13 dataset for subsequent analysis [13,21,22,30].

4. Molecular Docking

After the preliminary pharmacokinetic and toxicological profiling, the compounds that showed the most acceptable **ADME/Tox** characteristics were advanced to the molecular docking stage. This step was included to examine, at the atomic level, how the selected phytochemicals from the **MPA13** dataset could interact with the previously chosen cancer-related receptors. In this way, molecular docking served as a bridge between the chemical filtering stage and the biological interpretation of ligand–target recognition. While the ADME/Tox analysis helped identify compounds with potentially favorable pharmacological behavior, docking made it possible to evaluate whether those compounds could adopt stable and meaningful binding orientations within the active sites of the selected proteins [13,21,22].

More broadly, molecular docking remains one of the most widely used computational approaches in structure-based drug discovery because it allows the prediction of ligand binding modes and the estimation of the relative strength of receptor–ligand interactions before experimental testing. Its usefulness is particularly evident in natural product research, where large numbers of structurally diverse compounds can be screened in a rational and relatively rapid way. Nevertheless, the interpretation of docking outcomes requires careful methodology, since docking quality

depends strongly on receptor preparation, ligand preparation, accurate definition of the binding region, and appropriate protocol validation [22,31].

4.1 Software Used

Molecular docking simulations in the present study were carried out using **Molecular Operating Environment (MOE)** software **(**Figure II.15), which is an integrated platform widely used in molecular modeling, structure-based design, and virtual screening studies. MOE offers several functions relevant to docking workflows, including protein and ligand preparation, active-site detection, small-molecule docking, induced-fit exploration, and interaction analysis. Its integrated environment is particularly advantageous in academic studies because it allows multiple stages of receptor–ligand modeling to be performed within the same software framework, thereby reducing methodological fragmentation and improving workflow consistency [22,32].



Figure II. 2MOE Docking Application Logo.

In the context of the present work, MOE was chosen because it provides a practical platform for handling both the structural preparation phase and the docking phase of small molecules against protein targets. It also facilitates the visualization of the binding pocket, the inspection of predicted interactions, and the comparison of docked poses with reference ligands. These aspects are especially important when the objective is not only to obtain docking scores, but also to interpret the molecular basis of possible anticancer activity through hydrogen bonding, hydrophobic contacts, aromatic interactions, and spatial complementarity within the active site [22,32].

4.2 Preparation of Receptor and Ligand Files

Before starting the docking calculations, both receptor and ligand files were carefully prepared in order to ensure that the final complexes would be generated from structurally reliable inputs. The receptor structures were first retrieved from the **RCSB Protein Data Bank (PDB),** with preference given to experimentally determined human proteins presenting suitable structural quality. Prior to docking, each protein structure was cleaned by removing unnecessary molecules not directly involved in the binding event, such as irrelevant solvent molecules or unrelated crystallographic components. The integrity of the receptor structure and the accessibility of the active pocket were also checked before proceeding to docking [22,23].

The ligand structures, on the other hand, were taken from the curated **MPA13** dataset described previously. These compounds had already undergone curation and standardization, but further preparation was required before docking. In this phase, hydrogen atoms were added, protonation states were adjusted when necessary, and geometry optimization was performed to obtain energetically reasonable conformations. This ligand preparation stage was essential because the success of docking depends not only on the receptor structure, but also on the chemical correctness of the ligand input. Poorly prepared ligands can generate unrealistic poses or unreliable interaction patterns, even when their theoretical biological potential is high [21,22,31].

To maintain methodological consistency, the same preparation logic was also applied to the **reference drugs** associated with each receptor, namely sorafenib, enzalutamide, BCL201, rucaparib, and palbociclib. Docking the test compounds and the reference ligands under comparable conditions allowed a more meaningful comparison of binding behavior and helped place the results of the MPA13 compounds within a pharmacologically relevant context. This comparative strategy is important because docking scores alone are rarely sufficient; their value becomes stronger when interpreted alongside known active ligands occupying the same receptor environment [22,31].

4.3 Binding Site Definition

The binding site was defined separately for each of the five selected protein targets. When a co-crystallized ligand was available in the protein structure, its position was used as the main reference for locating the active pocket. This approach was adopted because the native ligand provides structural information about the experimentally observed binding region of the receptor [22,31].

After the receptor preparation step, the active pocket was inspected in MOE. The native ligand was used to identify the region where ligand binding occurs, then it was deleted before docking the selected compounds (Figure II. 3). The Site Finder tool in MOE(Figure II. 17) was then used to detect the corresponding pocket. Dummy atoms were generated in the selected site in order to define the docking region and guide the placement of ligands during the docking calculation.

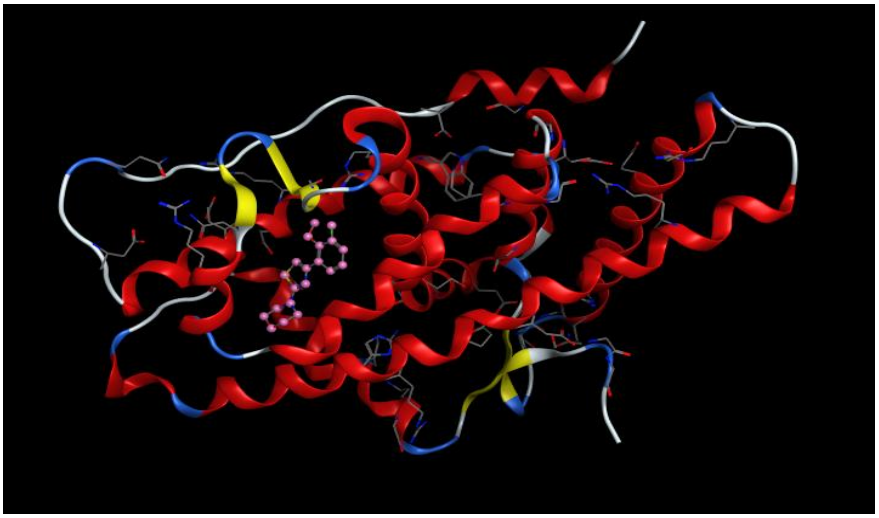


Figure II. 16 Position of the active ligand inside the binding pocket before docking site definition.

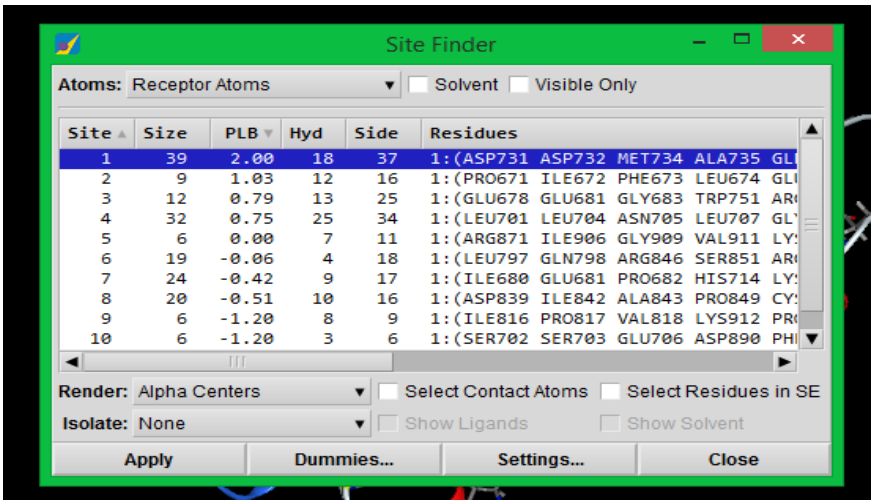


Figure II. 4 Binding site definition using MOE Site Finder.

The same binding-site definition strategy was applied to all selected targets. This ensured that the docking procedure was focused on biologically relevant regions rather than random surface cavities. The selected site was required to be large enough to accommodate the tested compounds while remaining centered on the active pocket of the receptor. This step was important because inappropriate site selection can affect the predicted binding poses, docking scores, and interaction interpretation [22,31,32].

4.4 Docking Procedure and Interaction Analysis

After ligand preparation, receptor preparation, and binding-site definition, molecular docking was performed in MOE using the prepared ligand dataset and the prepared protein targets. The selected compounds were docked into the active pocket of each receptor, and the docking program generated possible binding poses for each ligand(Figure II. 5). The poses were ranked according to their docking scores, which were used as an initial indicator of the predicted stability of the ligand–receptor complexes [22,31,32].

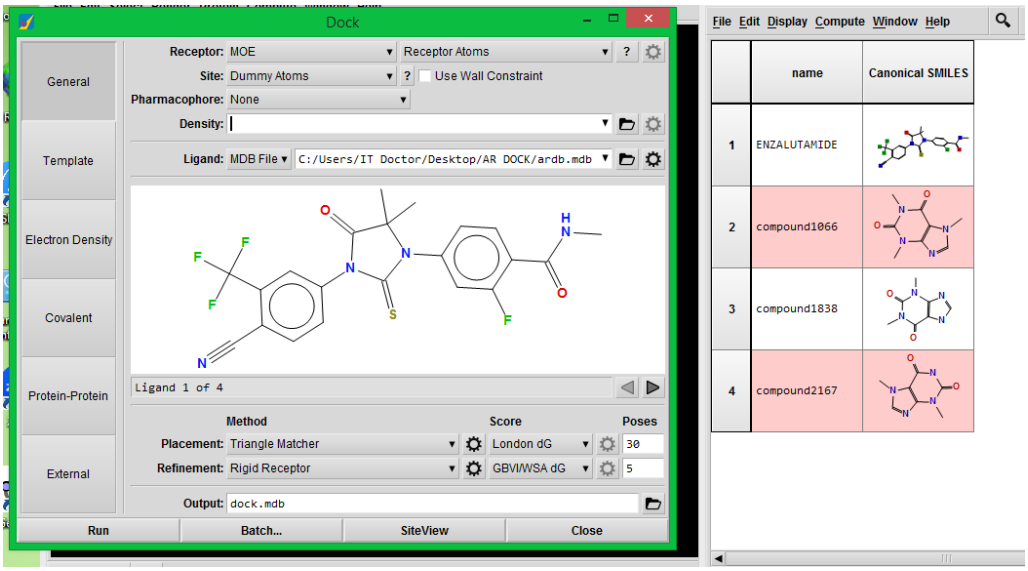


Figure II. 18 Docking setup of selected ligands in MOE.

For each target, the best-ranked poses were inspected visually in order to verify their orientation inside the binding pocket(Figure II. 19). The analysis did not rely only on the numerical docking score, but also considered the interaction pattern between the ligand and the amino acid residues of the active site. This interpretation is important because a favorable score may be more meaningful when the ligand also forms relevant interactions within the receptor pocket [22,31].

For each target, the co-crystallised reference inhibitor defined the active site geometry and served as a positive control. All candidate compounds were scored using the London dG function (E_score1), followed by GBVI/WSA dG rescoring (E_score2, denoted S), which accounts for solvation effects and provides a more physically grounded estimate of binding free energy [22]. The final docking score S (kcal/mol) was used as the primary ranking criterion; more negative values correspond to greater predicted binding affinity and more stable receptor–ligand complexes. Pose geometry was assessed by the refinement RMSD (RMSDrefine), which quantifies conformational displacement during energy minimisation of the bound pose.

Database Viewer : $HOME/dock.mdb

File Edit Display Compute Window Help SVL DBV MOE

	mol	rseq	mseq	S	rm
1	Oc1ccc(cc1)N(C(	1	1	-7.8731	
2	Oc1ccc(cc1)N(C(	1	1	-7.5515	
3	Oc1ccc(cc1)N(C(	1	1	-7.5272	
4	Oc1ccc(cc1)N(C(	1	1	-7.2994	
5	Oc1ccc(cc1)N(C(	1	1	-7.2387	
6	O[C@H]1CCN2C1=N	1	2	-5.0358	
7	O[C@H]1CCN2C1=N	1	2	-5.0150	
8	O[C@H]1CCN2C1=N	1	2	-4.7099	
9	O[C@H]1CCN2C1=N	1	2	-4.6544	
10	O[C@H]1CCN2C1=N	1	2	-4.5653	
11	OC[C@H](c1ccccc	1	3	-5.9624	
12	OC[C@H](c1ccccc	1	3	-5.9018	
13	OC[C@H](c1ccccc	1	3	-5.7823	
14	OC[C@H](c1ccccc	1	3	-5.7033	
15	OC[C@H](c1ccccc	1	3	-5.6880	
16	OCc1ccc(n1CCc1c	1	4	-5.2804	
17	OCc1ccc(n1CCc1c	1	4	-5.0422	
18	OCc1ccc(n1CCc1c	1	4	-5.0130	
19	OCc1ccc(n1CCc1c	1	4	-4.9651	
20	OCc1ccc(n1CCc1c	1	4	-4.8983	

20 entries, 0 selected, all visible. 15 fields, 0 selected, all visible.

Figure II. 19Docking score results generated by MOE for the selected ligands.

The selected docked complexes were then saved in PDB format and opened in BIOVIA Discovery Studio for further visualization. Two-dimensional interaction diagrams were generated to present the binding interactions in a clearer way and to facilitate comparison between the selected phytochemicals and the reference drugs. These diagrams were used to support the interpretation of the docking results and to identify the residues involved in ligand recognition (Figure II. 20).

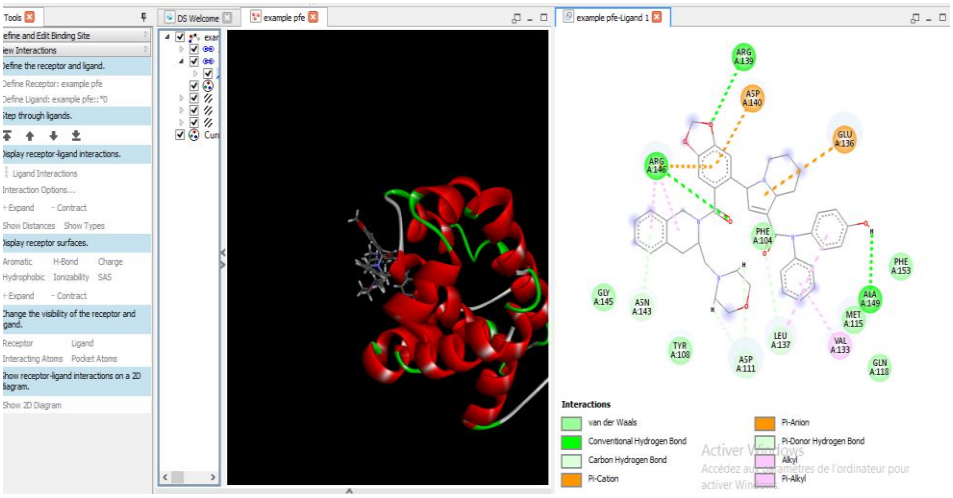


Figure II. 20 Visualization of ligand–receptor interactions using BIOVIA Discovery Studio.

Overall, the molecular docking step allowed the compounds retained after ADME/Tox screening to be evaluated against the five selected cancer-related targets. By combining docking scores with interaction analysis, this stage provided useful information about the possible binding behavior of the selected phytochemicals and helped prioritize compounds with potentially relevant anticancer activity.

5. Post-Docking Analysis:

A major objective of the post-docking phase was the identification of the main intermolecular forces contributing to ligand binding. Among the various contact types that can occur at the ligand–protein interface, particular attention was given to hydrogen bonds, hydrophobic interactions, and π-π stacking because these interactions are commonly associated with binding stability, specificity, and molecular recognition in docking studies [33,34]. Protein–ligand interaction profilers are specifically designed to detect and classify such contacts, including hydrogen bonds, hydrophobic contacts, and π-stacking interactions, which further supports their relevance in structure-based analysis [34].

Hydrogen bonds were considered particularly important because they are directional interactions that can help anchor the ligand within the active site and contribute to binding specificity. When a ligand forms hydrogen bonds with residues known to be critical in the receptor pocket, this often supports the plausibility of the predicted pose. In the present study, hydrogen-bond analysis was therefore used to identify compounds capable of establishing stable polar interactions with key amino acids surrounding the active site. Such contacts were interpreted as favorable when they involved reasonable donor–acceptor geometry and contributed to a coherent ligand orientation within the pocket [22,31].

At the same time, hydrophobic interactions were examined because many active-site regions are partially nonpolar and rely on hydrophobic complementarity to stabilize bound ligands. These contacts do not possess the same directionality as hydrogen bonds, but they often contribute significantly to ligand accommodation, especially when aromatic or aliphatic groups occupy lipophilic subpockets. Large-scale analyses of protein–ligand complexes have shown that hydrophobic contacts are highly frequent

and often enriched in efficient ligands, which explains why they were carefully considered in this work [34].

Finally, π-π stacking interactions were investigated in complexes containing aromatic systems. These contacts occur when aromatic rings from the ligand align favorably with aromatic residues in the receptor, such as phenylalanine, tyrosine, or tryptophan. Although not every ligand forms such interactions, π-π stacking can make an important contribution to binding affinity and pose stabilization, particularly in receptors with aromatic regions in or near the active site. Their identification was therefore useful when interpreting complexes involving polyaromatic phytochemicals or ligands bearing conjugated ring systems [33].

Taken together, the analysis of these interaction types made it possible to identify not only which compounds bound more favorably according to docking scores, but also why certain ligands appeared more promising than others. A compound capable of combining a good docking score with well-positioned hydrogen bonds, favorable hydrophobic complementarity, and, when applicable, aromatic stacking interactions was considered more convincing than a compound selected on score alone. This integrative interpretation was especially important for comparing the MPA13 compounds with the reference drugs, since it helped determine whether the natural compounds were reproducing key recognition features expected for biologically relevant binding [22,31,33].

Overall, the post-docking analysis provided a crucial interpretative layer in the present study. After docking had suggested which compounds could potentially fit within the active sites of the selected cancer-related receptors, post-docking analysis helped refine this information by integrating energetic ranking, interaction mapping, and visual structural inspection. As a result, the final prioritization of compounds was based not only on predicted affinity, but also on the quality and consistency of their interaction profiles, thereby providing a more reliable basis for selecting the most promising phytochemicals for further investigation [22,31–33].

CHAPTER III

RESULTS AND DISCUSSION

1.Similarity Analysis Results

As described in Chapter 2 (2.Molecular Similarity Analysis), the structural comparison between the MPA13 dataset compounds and the five reference anticancer drugs was carried out using a Python-based script implemented with RDKit [29], employing MACCS fingerprints combined with the Tanimoto coefficient as the primary similarity metric [27,28]. The present section reports and discusses the results obtained from this analysis for each of the five therapeutic targets: VEGFR2, BCL-2, AR, PARP1, and CDK4.The number of retained compounds differed clearly between the five reference drugs(Table III.1).

Table III. 1 Number of compounds retained after TanimotoMACCS-based similarity filtering.

Target	Reference drug	Selection criterion	Number of retained compounds	General observation
BCL-2	BCL201	Tanimoto_MACCS ≥ 0.50	**170**	Highest number of retained compounds
CDK4	Palbociclib	Tanimoto_MACCS ≥ 0.50	**102**	Second highest similarity group
PARP1	Rucaparib	Tanimoto_MACCS ≥ 0.50	**22**	Moderate number of retained compounds
VEGFR2	Sorafenib	Tanimoto_MACCS ≥ 0.50	**16**	Limited but relevant similarity group
AR	Enzalutamide	Tanimoto_MACCS ≥ 0.50	**3**	Most selective similarity group

1.1 VEGFR-2 similarity analysis against Sorafenib

Sorafenib showed a Tanimoto MACCS value of **1.000**, which is expected because it was used as the reference compound. Therefore, sorafenib was not included in the ranking of MPA13 compounds. Among the screened compounds, the Tanimoto MACCS values ranged from **0.500 to 0.614**, indicating moderate structural similarity to sorafenib based on shared MACCS substructural features (Table *III.2)*.

Table III. 2 Final ranking of MPA13 compounds based on Tanimoto MACCS similarity against sorafenib for VEGFR-2.

Rank	Compound	Tanimoto MACCS
Reference	Sorafenib	1.000
1	NC94	0.614
2	NC1887	0.604
3	NC645	0.577
4	NC2604	0.549
5	NC1759	0.547
6	NC1838	0.541
7	NC2167	0.541
8	NC471	0.519
9	NC1710	0.518
10	NC303	0.517
11	NC1066	0.516

Rank	Compound	Tanimoto MACCS
12	NC838	0.508
13	NC112	0.506
14	NC2470	0.500
15	NC373	0.500
16	NC160	0.500

Based on the final ranking, NC94 showed the highest similarity to sorafenib, with a Tanimoto MACCS value of 0.614. This was followed by NC1887 with a score of 0.604 and NC645 with a score of 0.577. These three compounds showed the strongest structural resemblance to sorafenib among the screened MPA13 compounds according to the selected similarity criterion. This similarity indicates possible engagement with the ATP-binding site, consistent with reported kinase inhibition mechanisms [7,10].

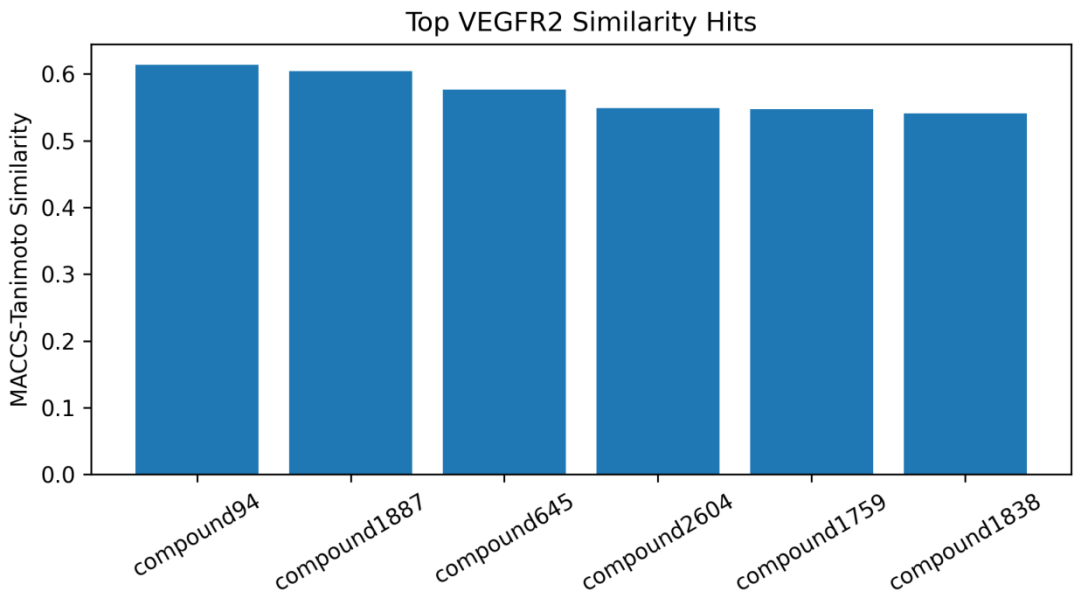


Figure III. 1VEGFR similarity scores of top compound

Compounds 2604 and 1759 also showed relatively good similarity values, with Tanimoto MACCS scores of 0.549 and 0.547, respectively. Although these scores are

lower than those of NC94 and NC1887, they still indicate the presence of shared structural features with sorafenib. Therefore, these compounds were also considered relevant hits for further evaluation.

The remaining compounds showed Tanimoto MACCS values between 0.500 and 0.541. These values indicate lower but still measurable structural similarity to the reference inhibitor. Since the similarity values were relatively close across several compounds, additional filters such as ADME/Tox properties, docking scores, and protein–ligand interaction patterns are required to determine which compounds have the most promising overall profile.

The moderate similarity values observed overall (maximum Tanimoto–MACCS = 0.614) confirm that the MPA13 compounds occupy a distinct but not entirely unrelated chemical space relative to sorafenib. This is pharmacologically significant: sorafenib's VEGFR2 inhibitory activity is primarily mediated by its urea linker and aromatic hinge-binding moieties, and even partial structural overlap in these subregions — as captured by MACCS keys — may indicate shared pharmacophoric potential [20,22].

1.2 AR similarity analysis againstEnzalutamide

For the androgen receptor, using Enzalutamide as the reference compound, only **3 compounds** satisfied the similarity criterion. These were **NC1066**, **NC1838** and **NC2167**, each with a Tanimoto_MACCS value of approximately **0.578**. This limited number suggests that enzalutamide has a more specific structural profile compared with most of the selected natural compounds.

This observation may reflect the more specific chemical architecture of Enzalutamide. Nevertheless, the retained compounds are important because they also appeared in other similarity subsets, especially in relation to Sorafenib and Palbociclib. Their recurrence suggests that these molecules may contain common substructural motifs relevant to more than one anticancer reference compound.

The relatively restricted number of similar compounds identified against Enzalutamide — compared to the other reference drugs — is consistent with the structural uniqueness of its thiohydantoin scaffold, which is rarely encountered in natural product chemical space. This finding aligns with the broader observation in

natural product drug discovery that synthetic AR antagonists tend to occupy a pharmacophoric niche that is poorly represented in plant-derived compound libraries [8,13].

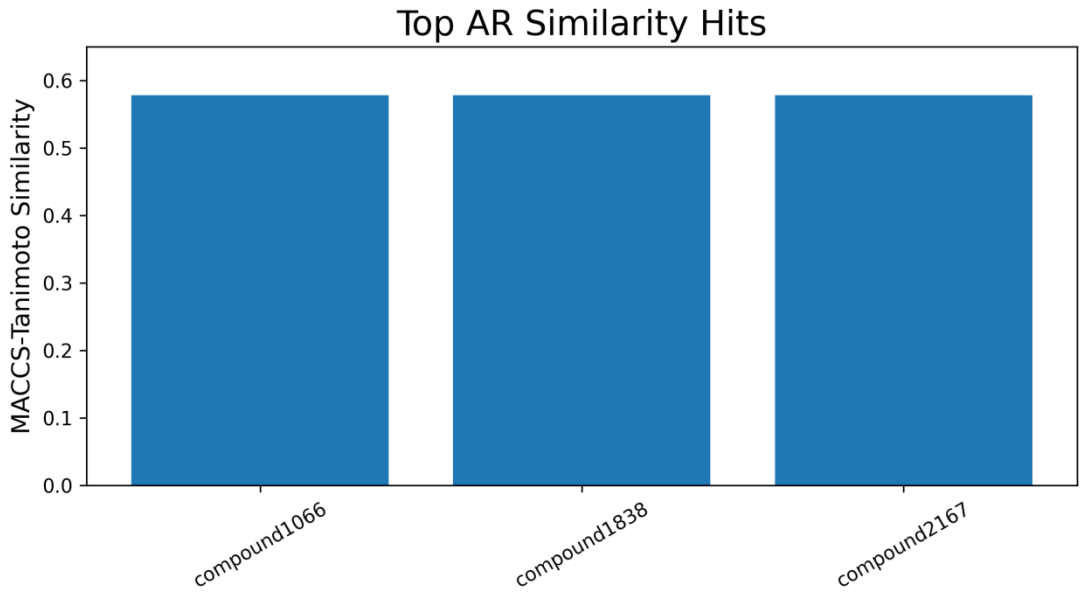


Figure III. 2 AR similarity scores of top compounds.

1.3 BCL-2 similarity analysis against BCL-201

For BCL-2, using BCL201 as the reference molecule, **170 compounds** were retained, making this the richest similarity group in the screening. The highest similarity was observed for **NC2598,** with a Tanimoto_MACCS value of approximately **0.6875**. Several other compounds also showed moderate to good similarity values, indicating that the MPA13 dataset contained many compounds with structural features potentially relevant to BCL-2-focused analysis.

BCL-201 is a BH3-mimetic compound designed to occupy the hydrophobic BH3-binding groove of BCL-2, displacing pro-apoptotic proteins [1,5]. The relatively high number of similar compounds identified suggests that the MPA13 library contains multiple scaffolds sharing substructural features relevant to BH3-groove binding — a finding consistent with the reported capacity of various natural polyphenols and terpenoids to modulate BCL-2 family proteins [19]. The broader distribution of MACCS similarity scores (0.50–0.6875) compared to the other targets also suggests

greater chemical complementarity between natural product scaffolds and the BCL-201 reference pharmacophore.

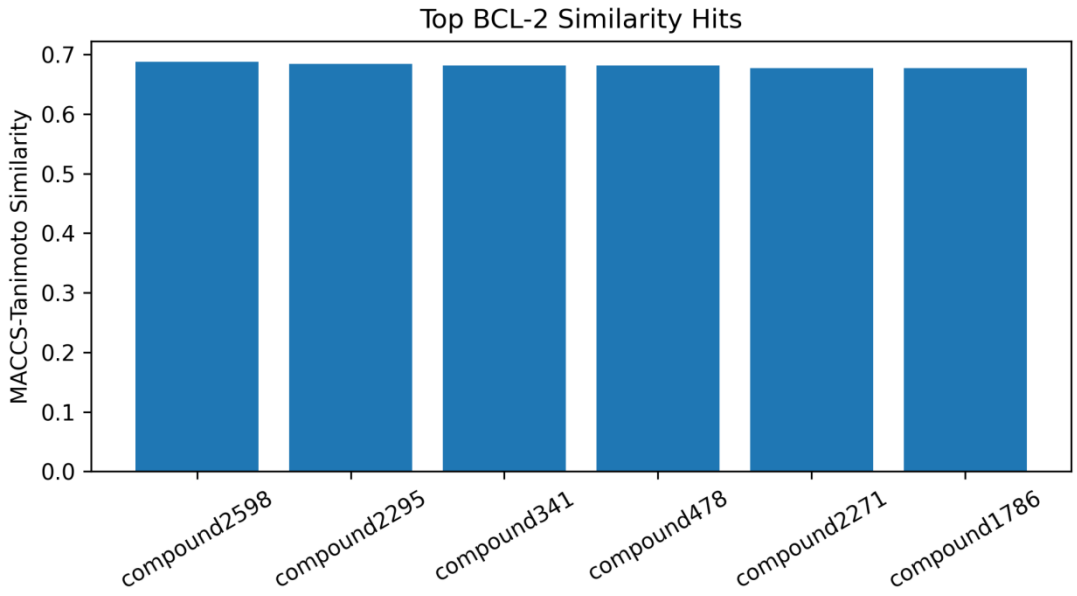


Figure III. 3 BCL similarity scores of top compounds.

1.4 PARP-1 similarity analysis against Rucaparib

For PARP-1, using Rucaparib as the reference inhibitor, **22 compounds** were retained. The highest similarity values were observed for **NC1001**, **NC91** , **NC1034**, **NC676** and **NC848** with Tanimoto_MACCS values close to **0.644**, **0.643** ,**0.642**, **0.567** and **0.566** respectively. These results indicate that a smaller but still relevant group of MPA13 compounds shared structural features with Rucaparib.

Rucaparib contains a rigid tricyclic indole/azepinoindolone scaffold that mimics the nicotinamide portion of NAD^+ and binds within the PARP1 catalytic NAD^+-binding pocket. Its core is stabilized by conserved interactions with Gly863, Ser904 and Tyr907, including π–π stacking mainly with Tyr907 [22]. Natural compounds sharing fused aromatic systems — as suggested by their MACCS similarity — may therefore interact with PARP1 through geometrically analogous binding modes, a hypothesis to be evaluated in the docking analysis.

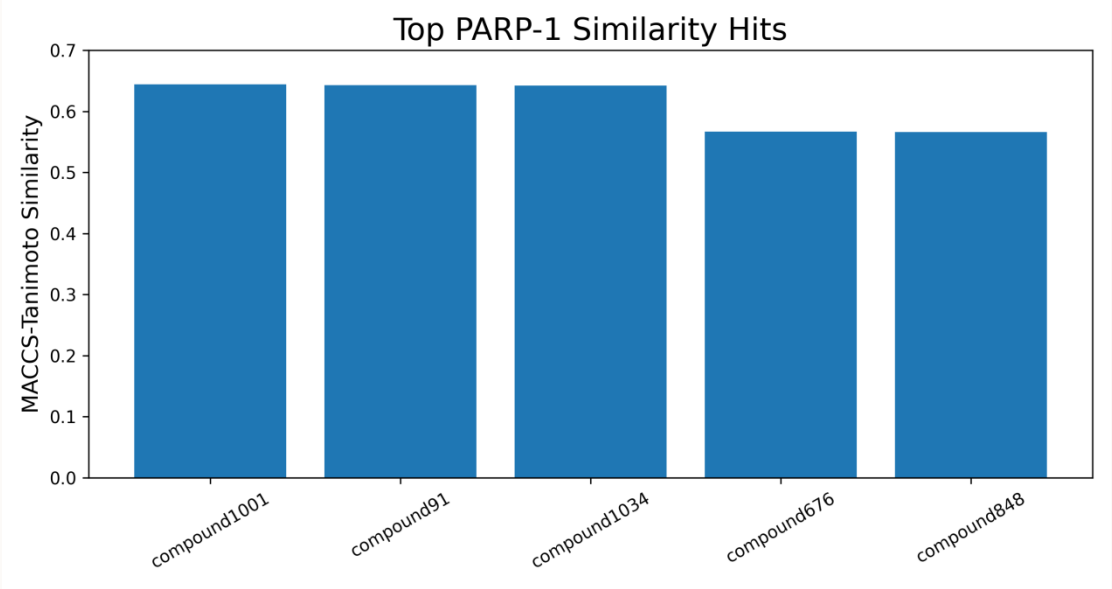


Figure III. 4 PARP similarity scores of top compounds.

1.5 CDK-4 similarity analysis against Palbociclib

For CDK4, using palbociclib as the reference compound, **102 compounds** passed the similarity threshold. The highest value was observed for **NC936**, with a Tanimoto_MACCS value of approximately **0.703**, this value was clearly higher than those observed in the other target groups, suggesting strong substructural similarity between NC936 and CDK-4, followed by **NC112** with a value close to **0.697** , **NC331** with a value of approximately **0.676**,**NC2295** with a value of approximately **0.671**, and **NC1838** with a value o approximately **0.646** also ranked among the top entries. These compounds were considered the most structurally similar to palbociclib.

Palbociclib is a pyridopyrimidine-based CDK4 inhibitor whose pharmacophore includes a central bicyclic pyrimidine ring engaging the hinge region of CDK4, flanked by a piperazine moiety and an acetyl-amino-pyridyl group [3,22]. The high MACCS similarity scores observed for several MPA13 compounds may reflect the presence of pyrimidine-like aromatic nitrogen heterocycles or structurally analogous fused ring systems within these natural scaffolds, which represent substructural features that are not uncommon among alkaloid-type phytochemicals [11,13].

The large number of compounds identified in this target category, combined with the highest Tanimoto–MACCS values in the entire study, suggests that the CDK4 target may represent the most accessible pharmacophoric space within the MPA13 dataset

from a similarity perspective. This preliminary observation is consistent with published reports identifying plant-derived flavonoids, alkaloids, and terpenoids as structurally compatible with CDK inhibitor pharmacophores [13,14].

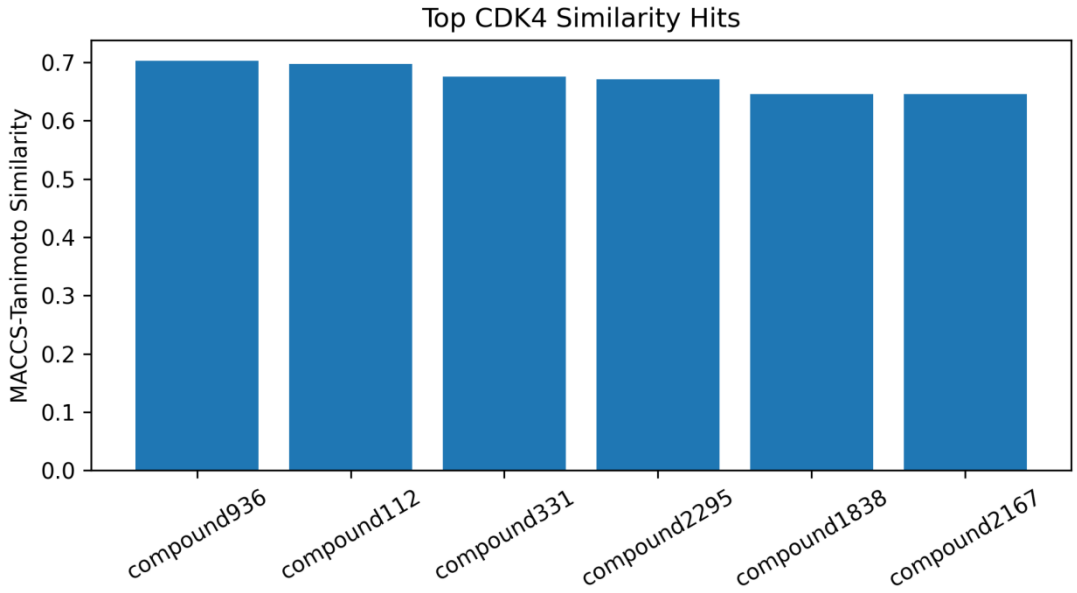


Figure III. 5 Top CDK Similar Compounds.

1.6 Multitarget Similarity Profile

An important observation from the similarity results was the recurrence of some compounds across more than one reference-drug subset. This point is particularly relevant because the anticancer strategy adopted in this study was not limited to a single target. As discussed in Chapter I, cancer progression depends on multiple interconnected biological processes, and natural compounds are often attractive because of their structural diversity and possible multi-target behavior [7,8,13]. Similarly, Chapter II described the use of a diversified receptor panel including VEGFR2, AR, BCL-2, PARP1, and CDK4 to represent angiogenesis, hormone signaling, apoptosis evasion, DNA repair, and cell-cycle dysregulation.

Among the recurrent compounds, NC1066, NC1838, and NC2167 were particularly notable because they appeared in the Enzalutamide-like subset and also showed similarity to other reference drugs. Their recurrence suggests that these compounds may contain shared substructural motifs relevant to different anticancer

drug classes. Additional recurrent hits, such as NC1710, NC676, NC483, NC2607, NC2533, NC1984, NC1249, and NC1001, may also be useful for cross-target comparison in the next stages.

It should be noted that recurrence across similarity groups may indicate the presence of common structural fragments, but it does not prove that the compound can bind effectively to all corresponding receptors.

Table III. 3Representative of some multitarget compounds identified during similarity screening.

Compound	Associated Targets
NC1838	VEGFR2, AR, CDK4
NC1066	VEGFR2, AR, CDK4, BCL-2
NC1613	BCL-2, PARP1, CDK4
NC2278	BCL-2, PARP1, CDK4
NC2295	BCL-2, CDK4
NC2167	VEGFR2, AR, CDK4
NC1710	VEGFR2, PARP1, CDK4
NC676	BCL-2, PARP1, CDK4
NC483	BCL-2, PARP1, CDK4
NC2607	BCL-2, PARP1, CDK4
NC2533	BCL-2, PARP1, CDK4
NC1984	BCL-2, PARP1, CDK4
NC1249	BCL-2, PARP1, CDK4
NC1001	BCL-2, PARP1, CDK4

Overall, the similarity screening successfully reduced the chemical space of the MPA13 dataset to a prioritized subset of compounds showing moderate or higher structural similarity to known anticancer reference drugs. The filtered compounds identified in this section were advanced to the next stage of the study, namely

ADME/Tox prediction, in order to determine whether their structural similarity to known anticancer drugs is accompanied by acceptable drug-like and safety profiles.

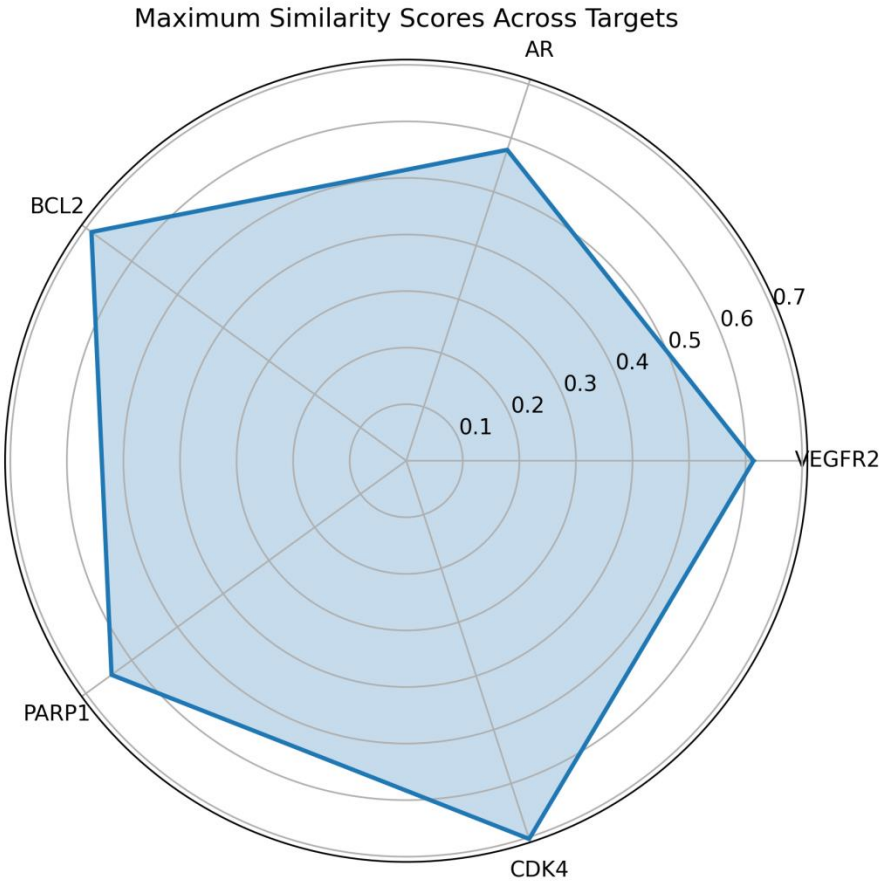


Figure III. 6 Maximum Similarity Scores Across Targets.

2. ADME/Tox Results:

The number of compounds retained varied between targets depending on the initial library size and chemical diversity. The following sections present the ADME and toxicity results for each target, followed by a cross-target comparative analysis.

2.1 ADME/TOX Evaluation of AR Candidates

Four molecules were evaluated in the AR dataset. The three test compounds (Compounds 1066, 1838, and 2167) share a xanthine scaffold and present markedly lower molecular weights relative to the reference inhibitor, enzalutamide.

Table III. 4 Physicochemical and pharmacokinetic properties of AR candidate compounds.

Compound	MW (g/mol)	TPSA (Å²)	LogP	HBD	HBA	GI Abs.	BBB	Pgp Sub.	Lip. Viol.	SA Score
Enzalutamide (Ref.)	464.44	108.53	3.85	1	7	High	No	No	0	3.17
NC 1066	194.19	61.82	−0.08	0	3	High	No	No	0	2.03
NC 1838	180.16	72.68	−0.27	1	3	High	No	No	0	1.87
NC 2167	180.16	72.68	−0.12	1	3	High	No	No	0	1.93

All four compounds fulfilled Lipinski's rule of five without violation. The three test compounds present MW values well below 200 g/mol, TPSA values between 61.8 and 72.7 **Å²**, and negative consensus LogP values indicating high hydrophilicity. All exhibit high GI absorption with no BBB permeability or Pgp substrate liability. Their SA scores (1.87–2.03) indicate straightforward synthetic accessibility compared to enzalutamide (3.17). Solubility predictions classify all three test compounds as "Very soluble", contrasting with enzalutamide's "Moderately soluble" profile.

Table III. 5 CYP isoform inhibition profiles and solubility classifications for AR candidates.

Compound	CYP1A2	CYP2C19	CYP2C9	CYP2D6	CYP3A4	ESOL Solubility Class	BA Score
Enzalutamide (Ref.)	No	Yes	Yes	No	Yes	Moderately soluble	0.55
NC 1066	No	No	No	No	No	Very soluble	0.55
NC 1838	No	No	No	No	No	Very soluble	0.55
NC 2167	No	No	No	No	No	Very soluble	0.55

Table III. 6 Predicted toxicity endpoints for AR candidates.

Compound	Hepato	Neuro	Resp.	Cardio	Carcino	BBB Tox	Immuno	Muta	Clinical
Enzalutamide (Ref.)	Active	Active	Inactive	Inactive	Inactive	Active	Inactive	Inactive	Active
NC 1066	Inactive	Active	Active	Inactive	Inactive	Active	Inactive	Inactive	Active
NC 1838	Inactive	Active	Active	Inactive	Inactive	Active	Inactive	Inactive	Active
NC 2167	Inactive	Active	Active	Inactive	Inactive	Active	Inactive	Inactive	Active

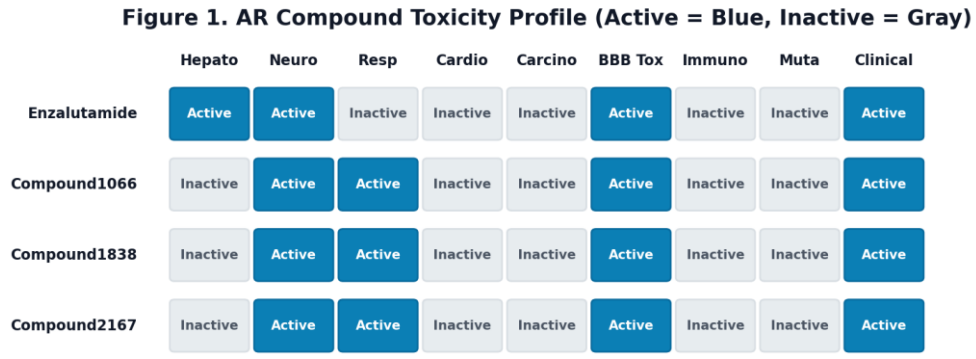


Figure III. 7 Toxicity heatmap of AR candidate compounds.

2.2 ADME/TOX Evaluation of CDK4 Candidates

The CDK4 library comprises 103 compounds, with palbociclib as the approved reference. Library-level descriptive statistics and drug-likeness filter outcomes are presented in Tables *III* .7 and *III* .8, respectively. Nine compounds were further evaluated for toxicity endpoint prediction. The full ADME dataset reveals notable structural diversity.

Table III. 7 Descriptive physicochemical statistics for the CDK4 candidate library.

Parameter	Minimum	Maximum	Mean	Palbociclib (Ref.)
MW (g/mol)	152.1	1,098.4	374.6	447.53
TPSA (Å²)	15.3	346.9	85.5	105.04
Consensus LogP	−3.69	5.54	2.05	2.31
HBD	0	8	1.5	2
HBA	1	18	4.6	6
SA Score	1.71	9.45	4.01	3.57

Table III. 8 Drug-likeness and absorption filter outcomes for the CDK4 candidate library .

Drug-likeness / PK Criterion	Compliant (n)	Non-compliant (n)	Rate (%)
Lipinski rule of five	80	23	77.7%
High GI absorption	84	19	81.6%
BBB permeant	48	55	46.6%
No Pgp substrate	45	58	43.7%
Bioavailability Score ≥ 0.55	91	12	88.34%

Table III. 9 Selected CDK4 candidates with optimal ADME profiles .

Compound	MW	LogP	TPSA	GI Abs.	BBB	Lip. Viol.	SA Score
Palbociclib (Ref.)	447.53	2.31	105.04	**High**	No	0	3.57
NC1066	194.19	−0.08	61.82	**High**	No	0	2.03
NC1838	180.16	−0.27	72.68	**High**	No	0	1.87
NC1326	245.27	1.32	62.46	**High**	**Yes**	0	1.88
NC1830	188.23	1.23	35.83	**High**	No	0	3.36
NC1812	317.34	1.96	65.07	**High**	**Yes**	0	3.18
NC2167	180.16	−0.12	72.68	**High**	No	0	1.93
NC2212	242.23	−0.61	104.55	**High**	No	0	3.64
NC2322	251.24	−0.68	119.31	**High**	No	0	3.6
NC2224	218.21	0.78	66.48	**High**	No	0	1.62

Table III. 10 Toxicity prediction results for nine CDK-evaluated candidates.

Compound	Hepatotox.	Neurotox.	Resp.	Cardiotox.	Carcino.	BBB Tox.	Immuno.	Muta.	Clinical Tox.
NC1066	Inactive	**Active**	**Active**	Inactive	Inactive	**Active**	Inactive	Inactive	**Active**
NC1326	Inactive	**Active**	**Active**	Inactive	Inactive	**Active**	Inactive	Inactive	**Active**
NC1838	Inactive	**Active**	**Active**	Inactive	Inactive	**Active**	Inactive	Inactive	**Active**
NC1830	Inactive	**Active**	**Active**	Inactive	Inactive	**Active**	Inactive	Inactive	**Active**
NC1812	Inactive	**Active**	**Active**	Inactive	Inactive	**Active**	**Active**	Inactive	**Active**
NC2167	Inactive	**Active**	**Active**	Inactive	Inactive	**Active**	Inactive	Inactive	**Active**
NC2212	Inactive	**Active**	**Active**	Inactive	Inactive	**Active**	Inactive	Inactive	**Active**
NC2224	Inactive	**Active**	**Active**	Inactive	Inactive	**Active**	Inactive	Inactive	**Active**
NC2322	**Active**	**Active**	**Active**	Inactive	Inactive	**Active**	Inactive	Inactive	Inactive

2.3ADME/TOX Evaluation of PARP1 Candidates

Twenty-two compounds were evaluated in the PARP1 dataset. Three compounds (12%) satisfy Lipinski's rule of five. Rucaparib serves as the approved reference inhibitor.

Table III. 11 ADME properties of selected PARP candidate compounds .

Compound	MW	LogP	TPSA	HBD	HBA	GI Abs.	BBB	Lip. Viol.	SA Score
Rucaparib (Ref.)	323.36	3.02	56.92	3	3	High	Yes	0	2.81
NC 848	176.22	1.33	62.04	3	2	High	Yes	0	1.42
NC 1034	190.24	1.72	48.05	3	2	High	Yes	0	1.52
NC 1249	244.33	1.93	25.24	0	2	High	Yes	0	3.77

Table III. 12 Toxicity predictions for PARP1 candidates with dedicated toxicity assessment.

Compound	Hepatotox.	Neurotox.	Resp.	Cardiotox.	Carcino.	BBB Tox.	Clinical Tox.	Nutritional Tox.
NC848	Inactive	**Active**	**Active**	Inactive	Inactive	**Active**	Inactive	Inactive
NC1249	Inactive	**Active**	**Active**	Inactive	Inactive	**Active**	**Active**	Inactive
NC1034	Inactive	**Active**	**Active**	Inactive	Inactive	**Active**	Inactive	Inactive

2.4.ADME/TOX Evaluation of VEGFR2 Candidates

Seventeen molecules constitute the VEGFR2 dataset, with Sorafenib as the clinical reference. Lipinski compliance reaches 23.52% (4/17 compounds), the highest observed across all datasets.

Table III. 13 ADME profiles of VEGFR candidate compounds versus sorafenib.

Compound	MW	LogP	TPSA	GI Abs.	BBB	CYP1A2	CYP2C9	CYP3A4	Lip. Viol.
Sorafenib (Ref.)	464.82	4.10	92.35	Low	No	Yes	Yes	Yes	0
NC 1759	219.19	1.20	79.39	High	No	No	No	No	0
NC 2167	180.16	−0.12	72.68	High	No	No	No	No	0
NC 1066	194.19	−0.08	61.82	High	No	No	No	No	0
NC 1838	180.16	−0.27	72.68	High	No	No	No	No	0

Table III. 14 Toxicity predictions for VEGFR-evaluated candidates.

Compound	Hepatotox.	Neurotox.	Resp.	Cardiotox.	BBB Tox.	Clinical Tox.
NC1759	Inactive	Inactive	Inactive	Inactive	**Active**	**Active**
NC1066	Inactive	**Active**	**Active**	Inactive	**Active**	**Active**
NC1838	Inactive	**Active**	**Active**	Inactive	**Active**	**Active**
NC2167	Inactive	**Active**	**Active**	Inactive	**Active**	**Active**

2.5ADME/TOX Evaluation of BCL2 Candidates

The BCL dataset is the most structurally diverse in the present study, comprising 170 compounds across a broad MW range (180.2–1,098.4 g/mol). 10.6% (n = 18) are Lipinski-compliant.

Table III. 15 Descriptive ADME statistics for the BCL candidate library (n = 170) .

Parameter	Minimum	Maximum	BCL201 (Ref.)
MW (g/mol)	180.2	1,098.4	710.82
TPSA (Å²)	21.7	346.9	96.71
Consensus LogP	−3.69	5.54	5.17
SA Score	1.00	9.45	5.70
HBA	2	18	High
HBD	0	8	No

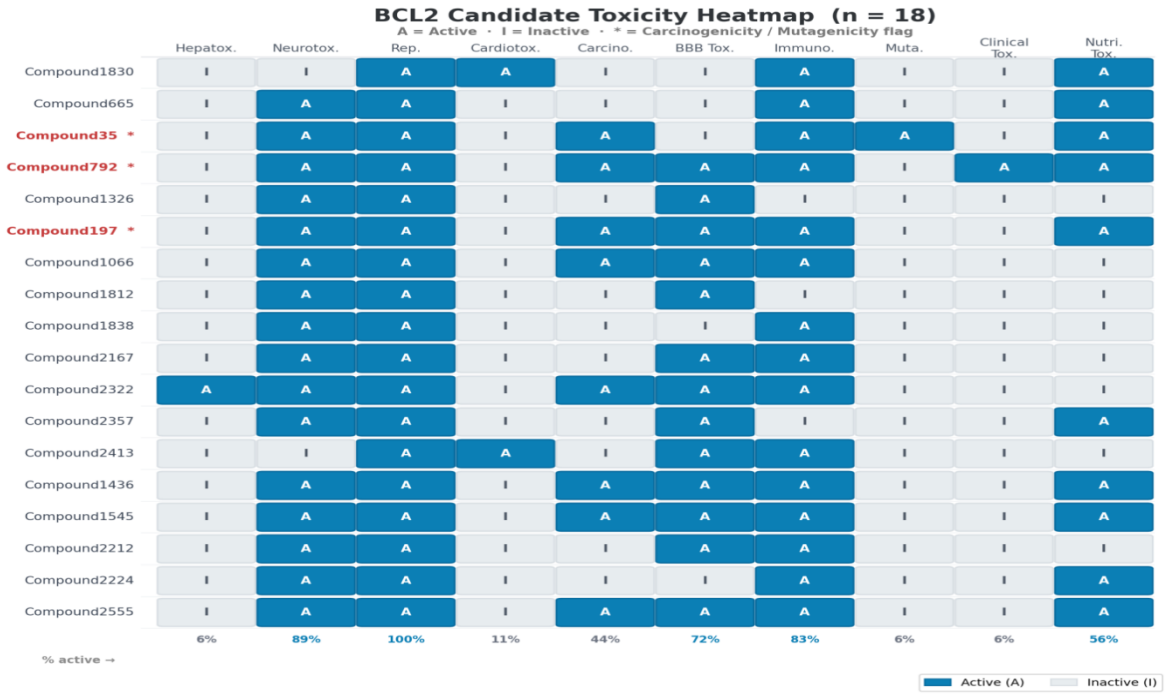


Figure III. 8 BCL2 Candidates Toxicity Heatmap.

2.6 Cross-Target Comparative Analysis

Table III. 16 Cross-target ADME and drug-likeness comparison summary.

Target	n	Mean MW	Mean LogP	Mean TPSA	Lip.-compliant (%)	High GI (%)	BBB perm. (%)	Best candidate
AR	4	254.7	0.84	78.9	100%	100%	0%	Cpd 1066
CDK	103	374.6	2.05	85.5	8.73%	81.6%	46.6%	Cpd 2224
PARP	25	406.2	2.49	97.9	12%	68.0%	60.0%	Cpd 848
VEGFR	17	249.6	0.80	88.4	23.52%	76.5%	23.5%	Cpd 1759
BCL	170	361.5	2.18	74.4	10.58%	87.1%	60.6%	Cpd 1830

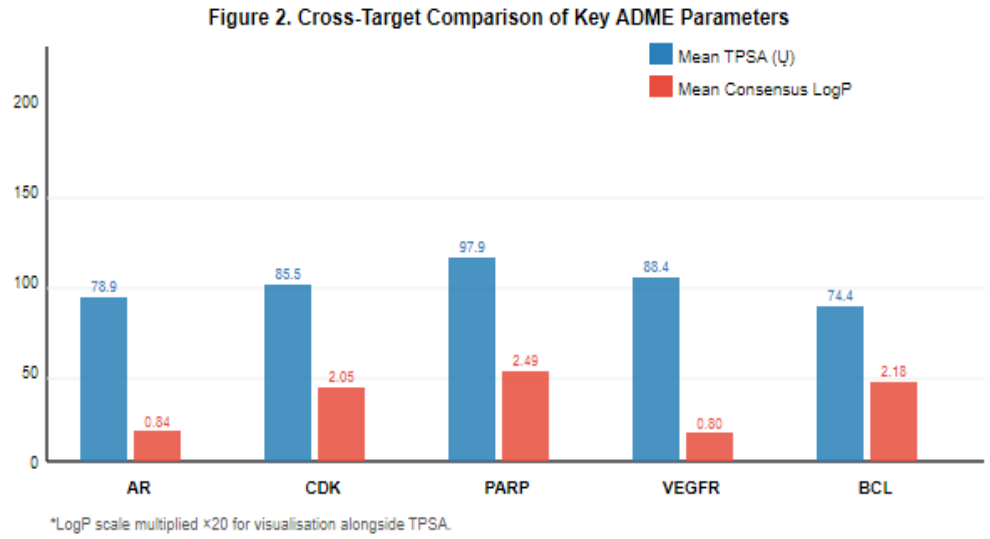


Figure III. 9 Cross-target comparison of mean TPSA and consensus LogP across all five compound libraries.

2.7 Discussion

- ✓ **Drug-likeness and Physicochemical Properties**

Compliance rates ranging from 8.73% (CDK) to 100% (AR) in the present study are broadly consistent with those reported for natural product-derived libraries, which

inherently include scaffold classes — such as cyclic peptides and alkaloid dimers — that exceed conventional MW and polarity limits [8]. The xanthine-based compounds recurrently retained across four target datasets (Compounds 1066, 1838, 2167) (Figure III.10) present MW values below 200 g/mol, negative LogP, and TPSA values well within the 60–140 Å² range associated with high oral absorption, consistent with established computational drug-likeness criteria. [21]

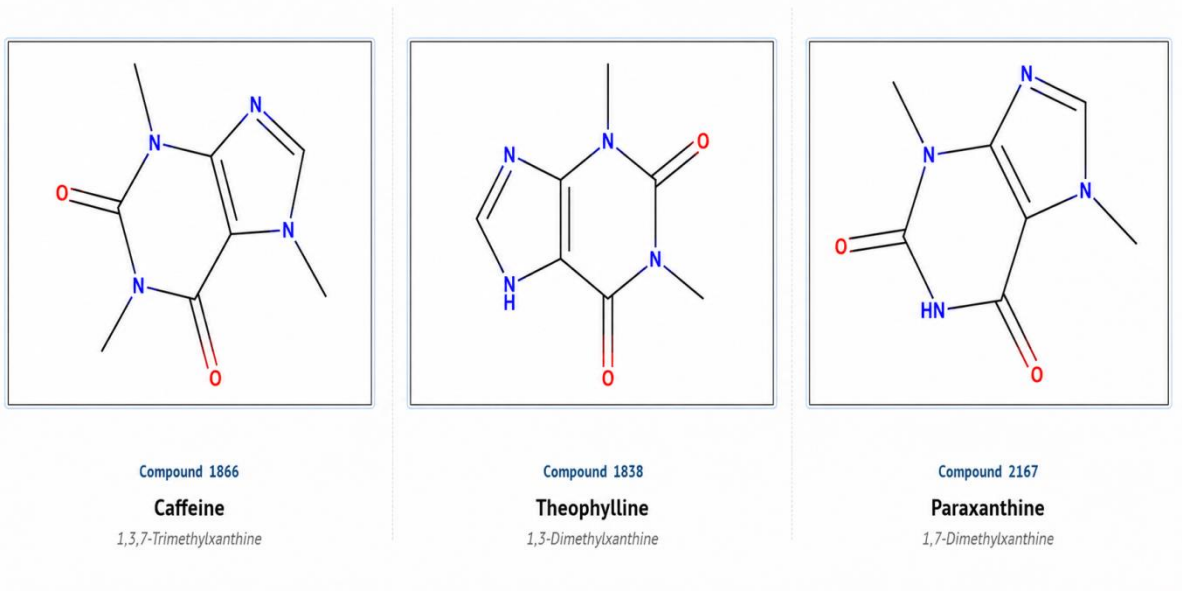


Figure III. 10 Methylxanthin Compounds

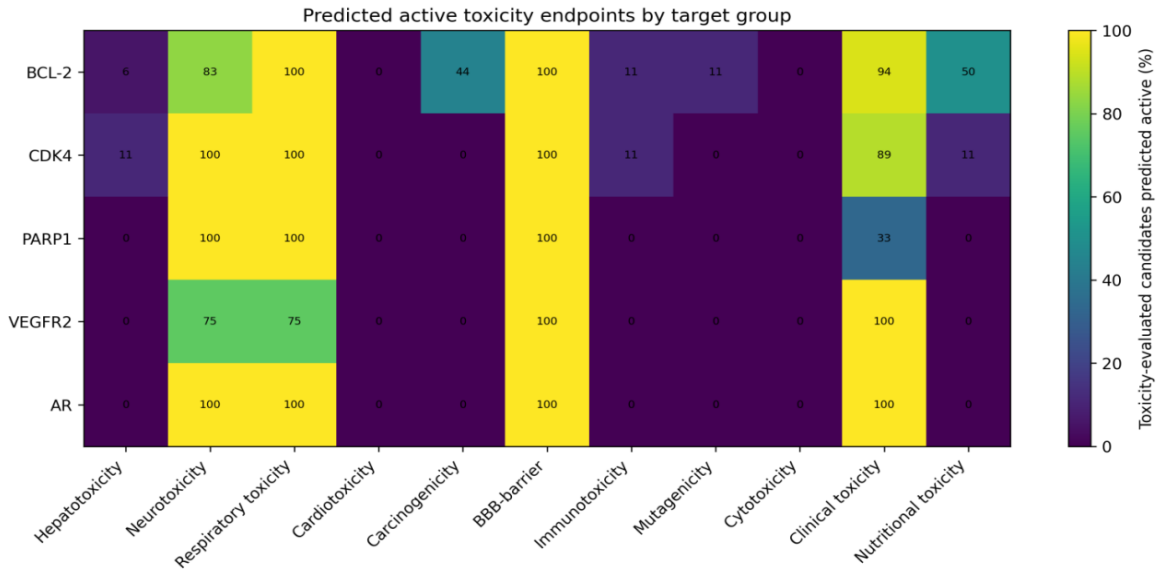
✓ **Absorption and Cytochrome P450 Interactions**

Gastrointestinal absorption, predicted as high for 68–100% of compounds across datasets, reflects the predominantly Lipinski-compliant physicochemical character of the libraries. High BBB permeability in the BCL (60.6%) and PARP (60%) datasets warrants consideration with respect to CNS toxicity risk. The reference inhibitors enzalutamide and sorafenib both exhibit documented CYP inhibition liabilities, [5] which are absent from the majority of test compounds. The uniformly clean CYP profile of the xanthine scaffold series represents a pharmacokinetic advantage relative to structurally more complex approved agents, reducing the probability of clinically significant drug–drug interactions.

✓ **Toxicity Profile and Safety Assessment**

The toxicity predictions showed that most candidates were not predicted to be active for hepatotoxicity, cardiotoxicity, mutagenicity, or cytotoxicity. This is a positive observation because these endpoints are important during early drug discovery prioritisation. However, neurotoxicity, respiratory toxicity, predicted BBB-barrier activity, and clinical toxicity appeared frequently in several target groups. These alerts do not confirm real toxicity, but they indicate that the corresponding compounds should be interpreted with caution and may require additional validation.

For the VEGFR2 subset, NC1759 showed the lowest number of predicted active toxicity endpoints among the selected candidates, with two active alerts. For PARP1, NC848 and NC1034 showed favourable profiles compared with the other evaluated



candidates. In the AR subset, all three candidates showed a similar toxicity pattern, mainly associated with neurotoxicity, respiratory toxicity, BBB-barrier activity, and clinical toxicity. For BCL-2 and CDK4, NC1326 appeared as one of the best-balanced candidates when ADME suitability and available toxicity predictions were considered together.

Figure III. 11 Heatmap of predicted active toxicity endpoints among toxicity-evaluated candidates.

✓ **Candidate Prioritisation: Best Pharmacokinetic Profiles**

On the basis of integrated ADMET criteria — Lipinski compliance, high GI absorption, solubility, absence of CYP inhibition, minimal toxicity endpoint burden,

and synthetic accessibility — the following candidates emerge as the most favourable for advancement:

Table III. 17 Top five pharmacokinetically prioritised candidates based on combined ADMET criteria

Compound	Target	MW	LogP	TPSA	GI Abs.	CYP Inh.	Hepatotox.	Cardiotox.	SA Score
NC 1759	VEGFR	219.19	1.20	79.39	High	None	Inactive	Inactive	1.84
NC 848	PARP	176.22	1.33	62.04	High	Partial	Inactive	Inactive	1.42
NC 1066	Multi	194.19	−0.08	61.82	High	None	Inactive	Inactive	2.03
NC 2224	CDK	218.21	0.78	66.48	High	None	Inactive	Inactive	1.62
NC 1838	Multi	180.16	−0.27	72.68	High	None	Inactive	Inactive	1.87

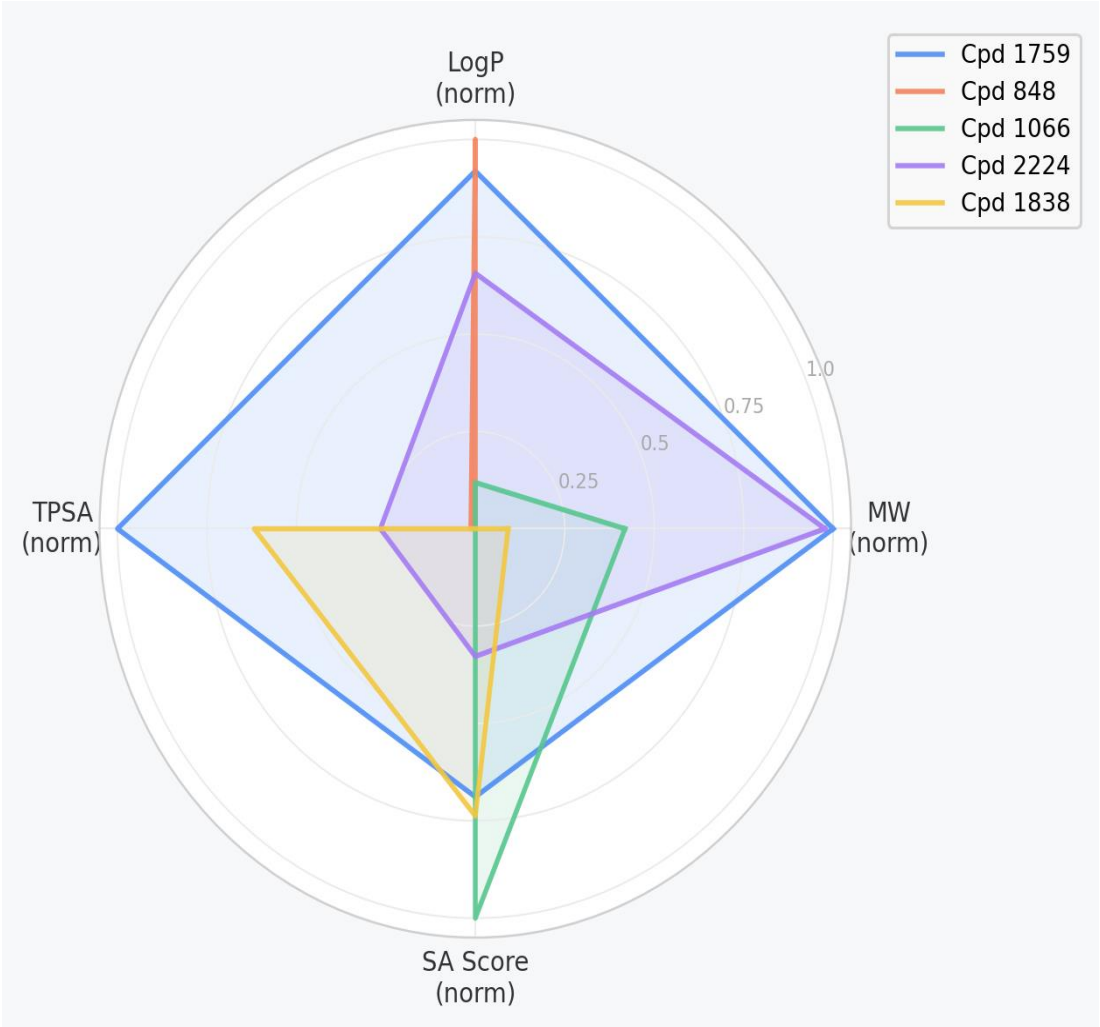


Figure III. 12*Multi-Parameter Physicochemical Profile Candidate Compounds*

3. Docking Results

The obtained results revealed significant variations among the tested phytochemicals depending on both ligand structure and receptor type, indicating differences in molecular recognition and binding-site compatibility.

3.1 Docking Score Matrix

The docking score matrix obtained for the selected compounds against the five investigated receptors is presented in Table III.18. The scores demonstrated that several

phytochemicals exhibited favorable predicted affinities toward multiple targets, suggesting possible multitarget anticancer potential.

Table III. 18 Docking Scores Summary Across All Targets .

Compound	AR (Ref: Enzalutamide)	BCL-2 (Ref: BCL201)	CDK4 (Ref: Palbociclib)	PARP-1 (Ref: Rucaparib)	VEGFR-2 (Ref: Sorafenib)
Ref	+1.45	−6.36	−6.08	−4.78	−9.62
NC1066	−6.19 1st	−4.61 5th	−4.38 3rd	—	−5.59 4th
NC1326	—	−4.72 2nd	−4.99 1st	—	—
NC1759	—	—	—	—	−5.76 2nd
NC1830	—	−4.07 7th	−4.22 5th	—	—
NC1838	−5.73 2nd	−4.11 6th	−4.04 7th	—	−5.89 1st
NC2167	−5.67 3rd	−4.05 8th	−4.19 6th	—	−5.62 3rd
NC2212	—	−4.67 3rd	−4.87 2nd	—	—
NC2224	—	−4.61 4th	−4.37 4th	—	—
NC665	—	−5.14 1st	—	—	—
NC848	—	—	—	−4.10 2nd	—
NC1034	—	—	—	−4.07 3rd	—
NC1249	—	—	—	−4.39 1st	—

❖ **Per-receptor analysis:**

- **AR**: This receptor produced the lowest docking scores among all investigated receptors, indicating stronger predicted ligand accommodation within its active pocket. The best-performing compound was NC1066 with −6.19 kcal/mol (rank 1), followed by NC1838 with −5.73 kcal/mol (rank 2) and NC2167 with −5.67 kcal/mol (rank 3). The reference score (Enzalutamide, +1.45) is atypical and reflects a different scoring convention for this receptor.
- **BCL-2**: The best score among the phytochemicals was obtained by NC665 with −5.14 kcal/mol (rank 1), followed by NC1326 (−4.72, rank 2), NC2212 (−4.67, rank 3), NC2224 (−4.61, rank 4), NC1066 (−4.61, rank 5), NC1838 (−4.11, rank 6), NC1830 (−4.07, rank 7), and NC2167 (−4.05, rank 8). The reference inhibitor BCL201 scored −6.36 kcal/mol.
- **CDK4**: NC1326 achieved the best score with −4.99 kcal/mol (rank 1), followed by NC2212 (−4.87, rank 2), NC1066 (−4.38, rank 3), NC2224 (−4.37, rank 4), NC1830 (−4.22, rank 5), NC2167 (−4.19, rank 6), and NC1838 (−4.04, rank 7). The reference Palbociclib scored −6.08 kcal/mol.
- **PARP-1**: Only three compounds were tested against this target. NC1249 achieved the best score with −4.39 kcal/mol (rank 1), followed by NC848 (−4.10, rank 2) and NC1034 (−4.07, rank 3). The reference Rucaparib scored −4.78 kcal/mol, a value not reached by the phytochemicals but closely approached by NC1249.
- **VEGFR-2**: NC1838 achieved the best score with −5.89 kcal/mol (rank 1), followed by NC1759 (−5.76, rank 2), NC2167 (−5.62, rank 3), and NC1066 (−5.59, rank 4). The reference Sorafenib scored −9.62 kcal/mol, a value substantially higher than those of the tested phytochemicals.

These variations across receptors may reflect differences in pocket geometry, residue composition, hydrophobicity, and steric accessibility of the receptor active sites [22,31].

3.2 Ranking of Compounds

To facilitate interpretation of the docking results, the compounds were ranked according to their average docking affinity across the investigated receptors. The

ranking procedure allowed the identification of phytochemicals with the most promising global binding behavior.

A cross-target comparison of docking performance allows the identification of priority compounds based on both score magnitude and target breadth. Table 7 summarises the global ranking of all natural compounds by their best score across any single target, together with the number of targets for which they demonstrated competitive binding.

Table III. 19 Cross-target summary and global priority ranking of screened natural compounds.

Global Rank	Compound	Best S (kcal/mol)	Best Target	Targets Hit	Priority Class
1	NC 1066	−6.19	AR	4 (AR, BCL-2, VEGFR-2, CDK4)	Multi-target hit
2	NC 1838	−5.89	VEGFR-2	4 (AR,BCL-2 , VEGFR-2, CDK4)	Multi-target hit
3	NC 1759	**−5.76**	VEGFR-2	1	Selective VEGFR-2 hit
4	NC 2167	−5.67	AR	4 (AR,BCL-2 , VEGFR-2, CDK4)	Multi-target hit
5	NC 665	−5.14	BCL-2	1	BCL-2 hit
6	NC 1326	−4.99	CDK4	2 (BCL-2, CDK4)	Dual BCL-2 / CDK hit
7	NC 2212	−4.87	CDK4	2 (BCL-2,CDK4)	Dual BCL-2 / CDK hit

8	NC 2224	-4.61	BCL-2	2 (BCL-2,CDK4)	Dual BCL-2 / CDK hit
9	NC 1249	−4.39	PARP-1	1	PARP-1 hit
10	NC 1830	−4.22	CDK4	2 (BCL-2,CDK4)	Dual BCL-2 / CDK hit
11	NC 848	−4.10	PARP-1	1	Selective PARP-1 hit
12	NC 1034	−4.07	PARP-1	1	PARP-1 hit

Among all investigated compounds, NC1066 consistently exhibited the best docking performance across most receptors, particularly against VEGFR and AR, where docking scores reached -5.59 kcal/mol and -6.19 kcal/mol, respectively. These values suggest the formation of relatively stable ligand–receptor complexes and indicate favorable structural complementarity between the ligand and the binding pocket.

3.3 Comparison with Known Inhibitors

The comparison showed that the phytochemicals did not systematically reach the reference values, but the gap remained limited in several cases.

- On **AR**, the reference score of Enzalutamide (+1.45) reflects a different scoring convention, making direct numerical comparison non-applicable. The phytochemicals all produced negative scores, indicating favorable predicted affinity within the MOE scoring framework.

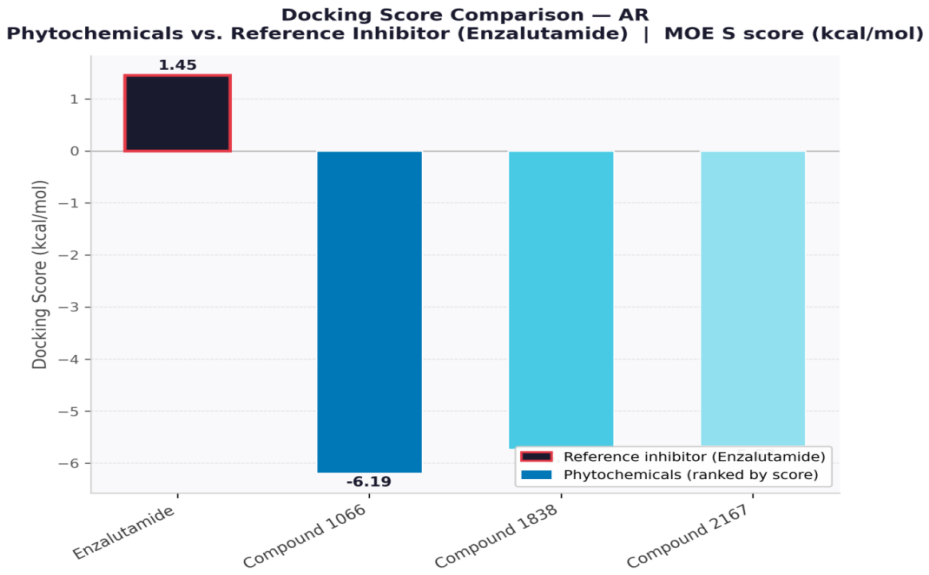


Figure III. 13 MOE Docking Scores of AR Candidates vs.Enzalutamide.

- On **BCL-2**, NC665 with −5.14 kcal/mol was the closest to BCL201 (−6.36), with a gap of 1.22 kcal/mol.

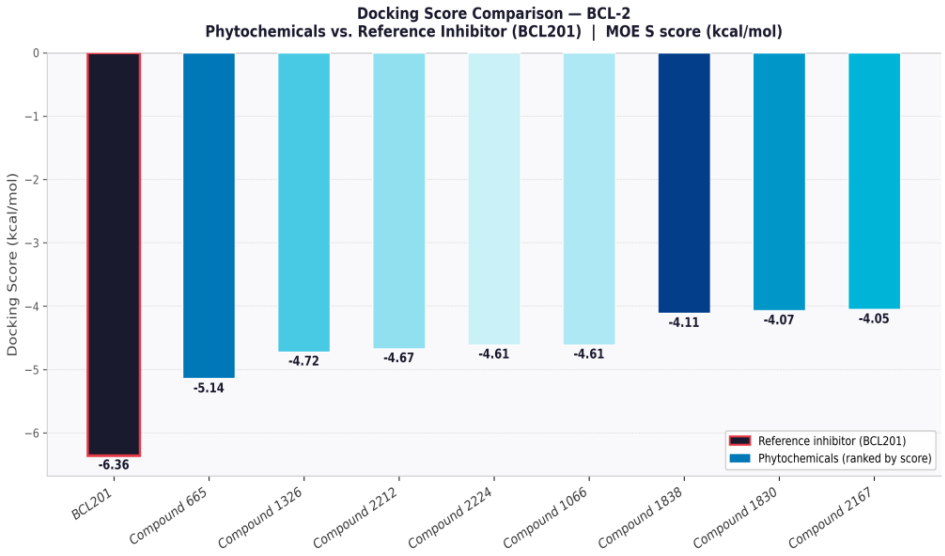


Figure III. 14 MOE Docking Scores of BCL2 Candidates vs.BCL201.

- On **CDK4**, NC1326 with −4.99 kcal/mol approaches Palbociclib (−6.08) reasonably closely, with a gap of 1.09 kcal/mol. NC2212 (−4.87) and NC1066 (−4.38) also present relevant values against this target.

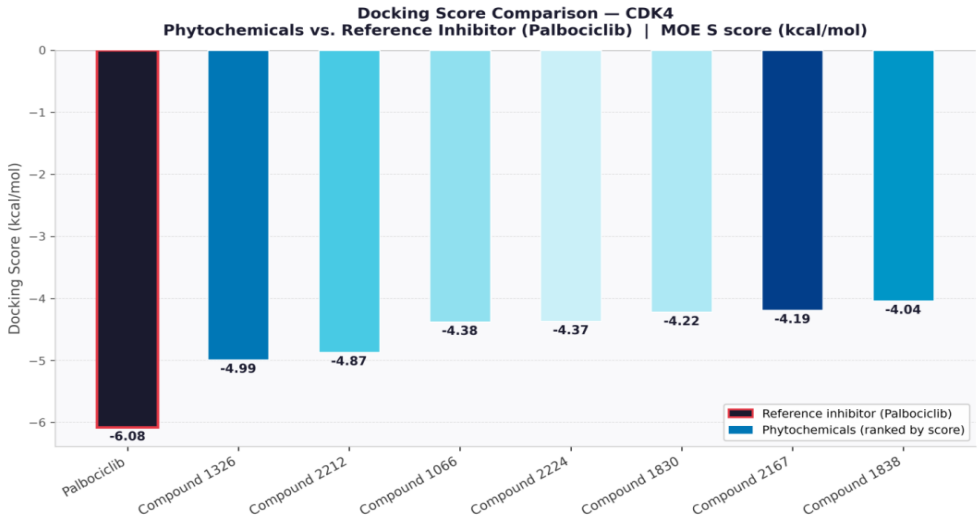


Figure III. 15 MOE Docking Scores of CDK4 Candidates vs.Palbociclib .

- On **VEGFR-2**, the gap between the phytochemicals (best score: −5.89, NC1838) and Sorafenib (−9.62) is the largest observed across all targets, approximately 3.73 kcal/mol. However, the compounds still maintained significant negative scores.

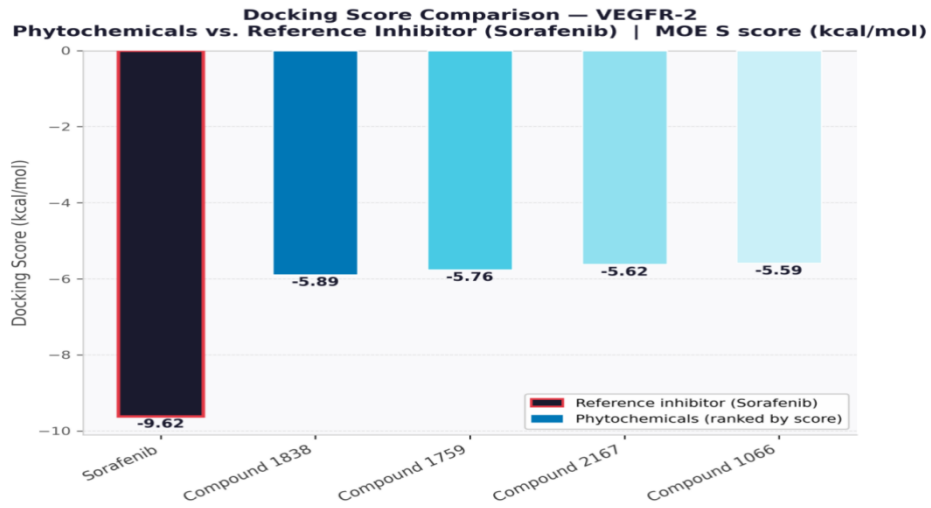


Figure III. 16 MOE Docking Scores of VEGFR2 Candidates vs.Sorafenib.

- On **PARP-1**, NC1249 (−4.39) is the closest to Rucaparib (−4.78), with a gap of only 0.39 kcal/mol — the smallest gap observed relative to any reference inhibitor across the entire dataset.

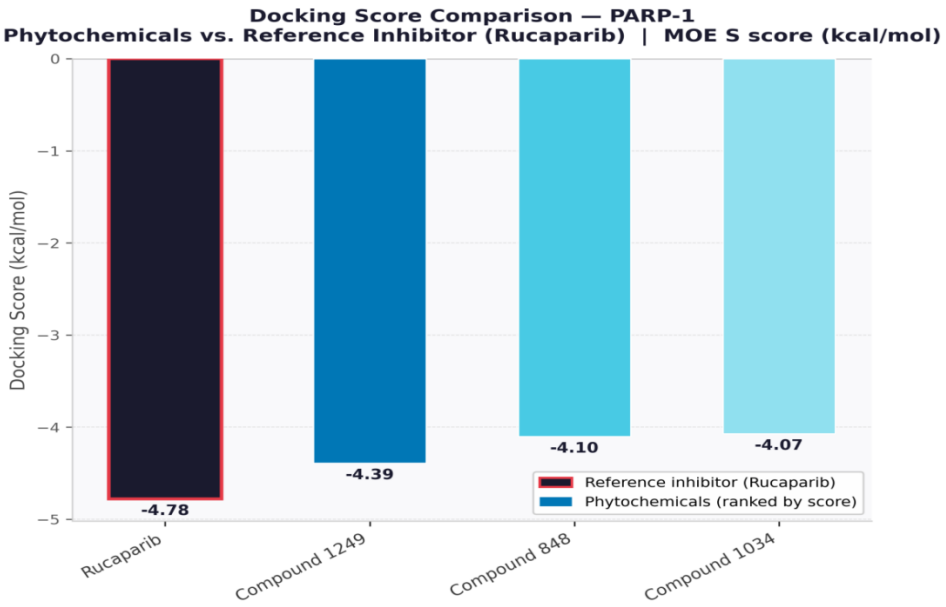


Figure III. 17 MOE Docking Scores of PARP1 Candidates vs.Rucaparib.

Although the reference drugs generally retained slightly stronger predicted affinity, the difference remained relatively limited in some cases, suggesting that the selected natural compounds may possess relevant structural features for receptor recognition.

3.4 Recurrence Analysis

✓ The recurrence analysis identified compounds appearing repeatedly among the best-ranked ligands across the different receptors.

✓ NC1066 appeared consistently among the best-ranked ligands across AR, BCL-2, VEGFR-2, and CDK4, making it the most recurrent compound in the dataset. This repeated presence may indicate broad structural adaptability and possible multitarget anticancer potential.

✓ NC1838 and NC2167 also demonstrated repeated occurrence across AR, VEGFR-2, BCL-2, and CDK4, confirming their multitarget profile observed in the global ranking.

✓ Compounds 1830, 1326, 2212, and 2224 were mainly associated with BCL-2 and CDK4 targets. Such recurrence patterns may reflect similarities in active-site compatibility and favorable interaction capabilities within receptors involved in apoptosis regulation and cell-cycle progression [1,5,22].

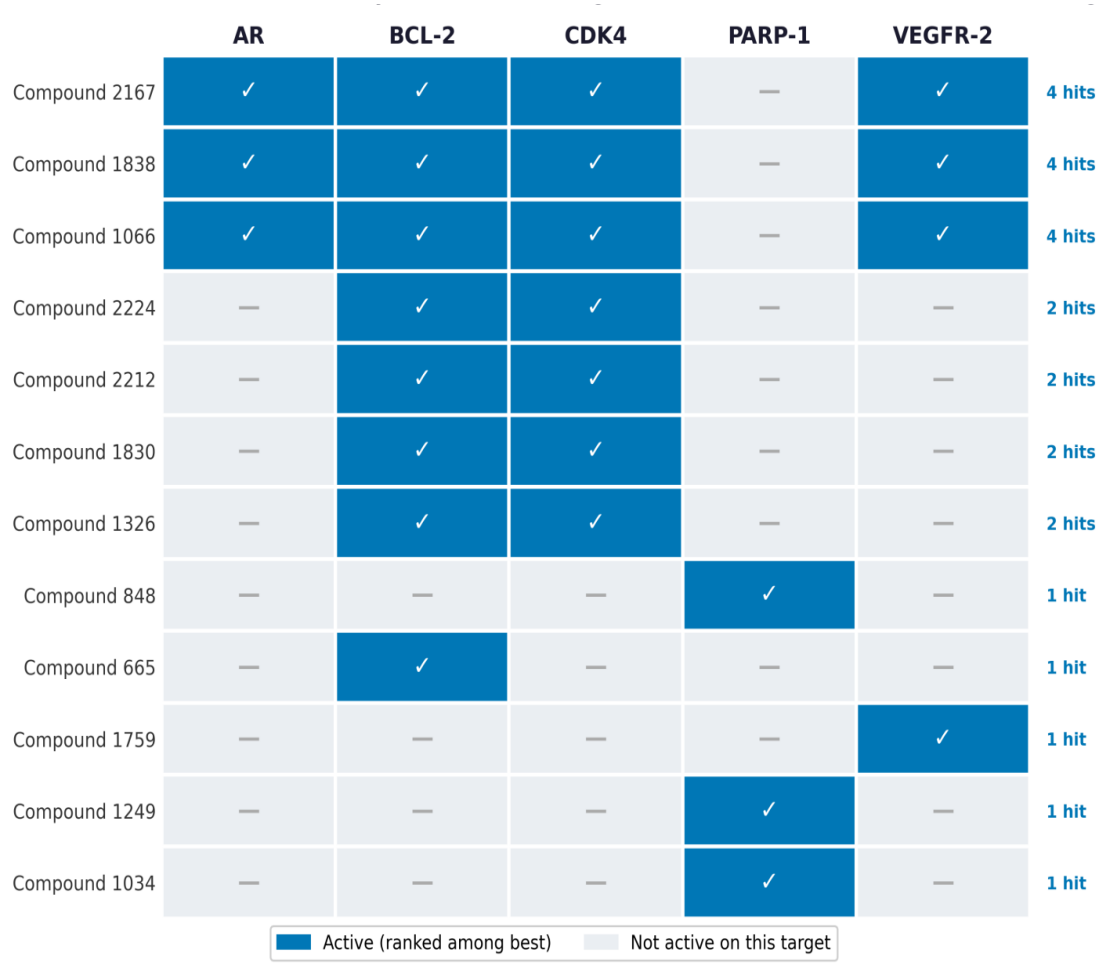


Figure III. 18 Compound targets occurrence matrix distribution of recurrent phytochemicals .

4. Interaction Analysis

4.1 Interaction Analysis of AR–Ligand Complexes

The interaction analysis of the AR complexes showed that NC 1066, NC 1838, and NC 2167 interact mainly through hydrophobic contacts, including van der Waals, alkyl, π-alkyl, and π-sulfur interactions. The most recurrent residues were Leu704, Asn705, Met745, Met780, Leu873, Phe876, Ala877, Leu880, Met895, and Phe891. These residues appear to play an important role in the stabilization of the ligands within the AR binding pocket.

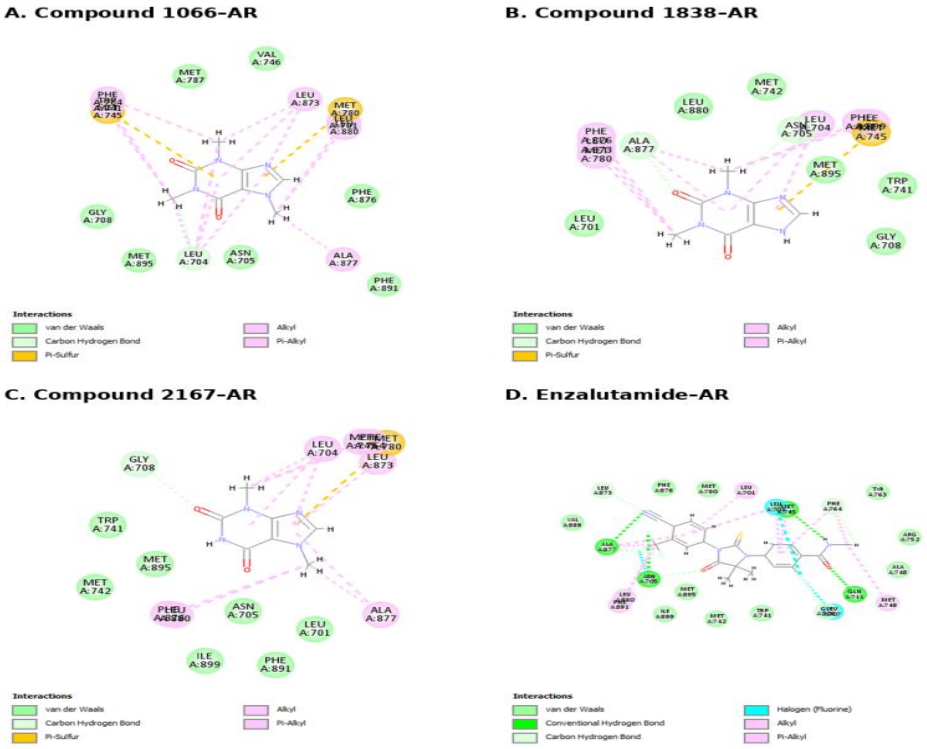


*Figure III. 19 :**Binding Interaction Maps of Selected MPA13 Compounds with AR .***

Based only on the 2D interaction maps, NC 2167 appears to be the most promising among the three compounds because it shows a broader hydrophobic interaction network. Compared with enzalutamide, the selected compounds show promising hydrophobic binding behavior, but they present fewer directional interactions such as conventional hydrogen bonds and halogen contacts.

4.1.1 Description of binding modes

The three selected compounds, NC 1066, NC 1838, and NC 2167, displayed binding modes mainly dominated by hydrophobic interactions. This suggests that these ligands are inserted into a hydrophobic region of the AR binding pocket, where residues such as leucine, phenylalanine, alanine, and methionine contribute strongly to ligand stabilization.

NC 1066 showed a binding profile characterized by several van der Waals, alkyl, π-alkyl, carbon hydrogen bond, and π-sulfur interactions. The main residues involved included Leu704, Asn705, Gly708, Val746, Met745, Met780, Met787, Leu873, Phe876, Ala877, Leu880, Met895, and Phe891. The presence of π-alkyl interactions

with aromatic and aliphatic residues indicates that the ligand is stabilized by its hydrophobic surface and heterocyclic scaffold. In addition, π-sulfur interactions involving methionine residues, particularly Met745 and/or Met780, may contribute to the stabilization of the ligand inside the AR pocket.

NC 1838 presented a similar hydrophobic binding pattern. It interacted with residues such as Leu701, Leu704, Asn705, Gly708, Trp741, Met742, Met745, Phe764, Phe876, Ala877, Leu880, and Met895. The ligand formed several alkyl and π-alkyl contacts, suggesting favorable insertion into the non-polar region of the binding site. The presence of carbon hydrogen bonds also indicates weak polar contributions, although the interaction profile remains mainly hydrophobic. This type of profile suggests that NC 1838 may be stabilized by both its heterocyclic core and hydrophobic contacts with surrounding AR residues.

NC 2167 also showed an extended hydrophobic interaction network. The residues involved included Leu701, Leu704, Asn705, Gly708, Trp741, Met742, Met745, Met780, Leu873, Ala877, Leu880, Met895, Phe891, and Ile899. Compared with compounds 1066 and 1838, NC 2167 appears to interact with a broader set of hydrophobic residues, particularly methionine, leucine, alanine, and phenylalanine residues. This suggests a good occupation of the AR binding pocket. The presence of π-sulfur and π-alkyl interactions may further reinforce its anchoring within the receptor.

In comparison, the reference ligand enzalutamide showed a more diversified interaction pattern. In addition to hydrophobic interactions such as alkyl and π-alkyl contacts, enzalutamide also formed conventional hydrogen bonds, carbon hydrogen bonds, and halogen interactions, especially involving fluorine atoms. The main residues observed in the enzalutamide complex included Asn705, Gln711, Leu704, Met745, Phe764, Met749, Met780, Ala877, Leu880, Met895, Phe891, and Ile899. This richer interaction profile suggests that enzalutamide benefits not only from hydrophobic stabilization but also from more directional interactions, which may improve its binding orientation and stability.

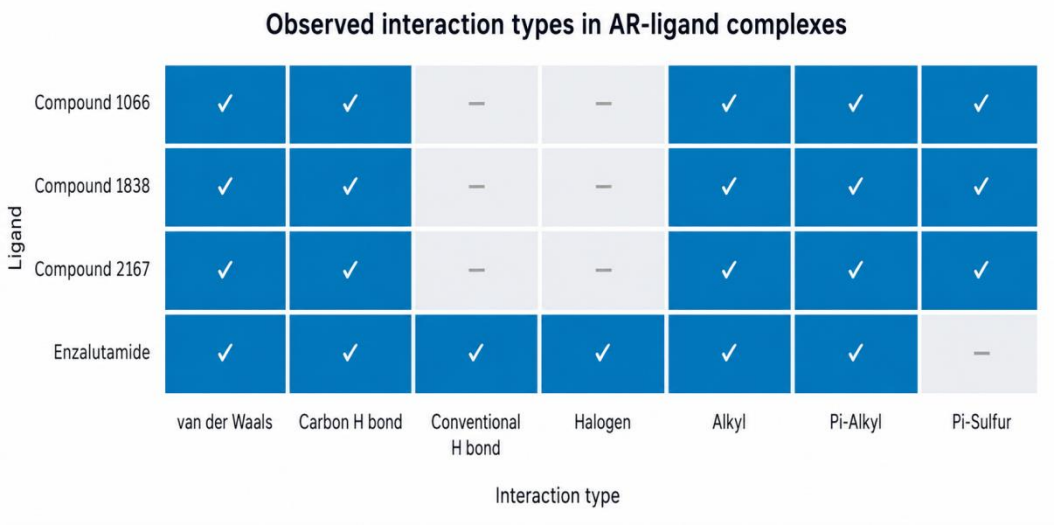


Figure III. 20 :Interaction Types in AR Ligand Complexes.

4.1.2 Key amino acids involved

The comparison of the four AR complexes revealed several recurrent amino acid residues. These residues can be considered important for ligand stabilization within the AR binding pocket.

Overall, Leu704, Asn705, Met745, Met780, Ala877, Leu880, Met895, and Phe891 appear to be among the most relevant residues in the AR interaction profiles. The recurrence of these residues suggests that the selected compounds bind in a region comparable to that occupied by enzalutamide.

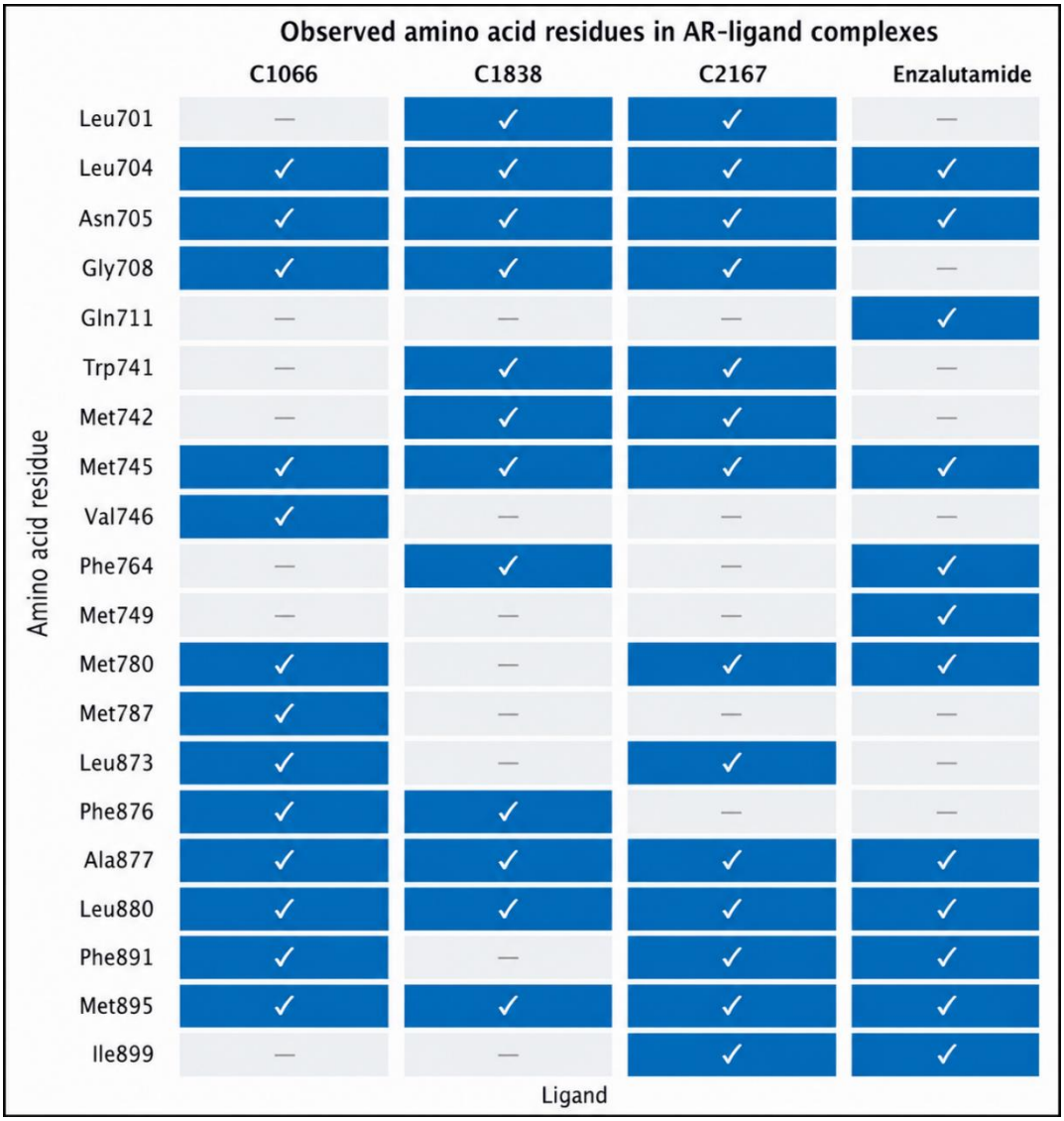


Figure III. 21 Observed Amino Acid Residues in AR Ligand Complexes.

4.1.3 Structure–activity relationship elements

From a SAR perspective, the three selected compounds appear to rely mainly on their hydrophobic and heterocyclic structural features to interact with the AR binding pocket. Their scaffolds allow multiple interactions with non-polar residues such as Leu704, Leu873, Leu880, Ala877, Phe876, Phe891, Met745, Met780, and Met895.

The presence of π-alkyl interactions suggests that aromatic or pseudo-aromatic regions of the ligands play an important role in stabilizing the complexes. Such interactions are important in ligand–protein interfaces and can influence the quality of docking poses and ligand recognition [33]. In addition, π-sulfur interactions with methionine residues may represent a favorable SAR element, especially for compounds containing heterocyclic or aromatic systems positioned close to sulfur-containing residues.

However, compared with enzalutamide, compounds 1066, 1838, and 2167 showed fewer strong polar interactions. Enzalutamide formed conventional hydrogen bonds and halogen interactions, which are more directional than simple hydrophobic contacts. This difference suggests that the selected compounds reproduce an important part of the hydrophobic binding profile of enzalutamide, but they may require further optimization to improve specific polar interactions.

Therefore, future structural optimization could focus on two main aspects. First, the hydrophobic groups responsible for interactions with Leu, Phe, Ala, and Met residues should be conserved, because they contribute strongly to pocket occupation. Second, the introduction or repositioning of hydrogen bond donor or acceptor groups could improve interactions with polar residues such as Asn705 and Gln711. The controlled addition of halogenated substituents may also help mimic some features of enzalutamide, but such modifications should be evaluated carefully to maintain acceptable ADME and drug-likeness properties [21].

4.2 Interaction Analysis of BCL-2–Ligand Complexes

As shown in Figure III .22 and Figure III .23, all analyzed ligands displayed van der Waals interactions, indicating that non-specific hydrophobic contacts contribute to

ligand stabilization within the BCL-2 pocket.Most compounds also formed conventional hydrogen bonds and/or carbon hydrogen bonds, suggesting additional polar contributions to ligand orientation. Several ligands showed π-mediated interactions, including π-alkyl, π-anion, π-cation, π–π stacked, π–π T-shaped, and amide–π stacked interactions.

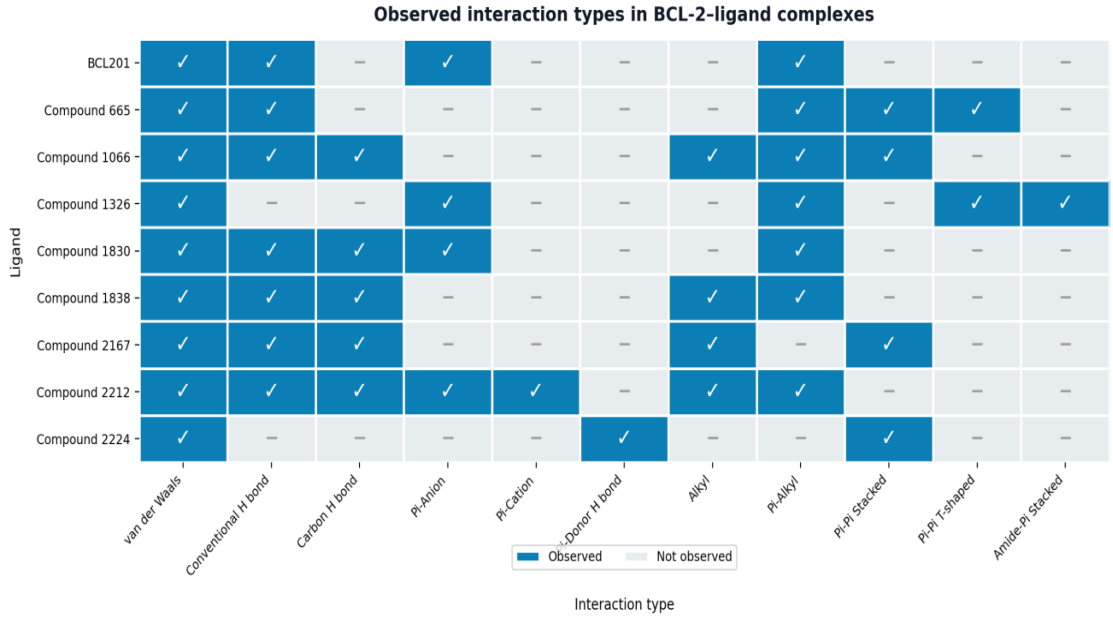


Figure III. 22 Interaction Types in BCL2 Ligand Complexes.

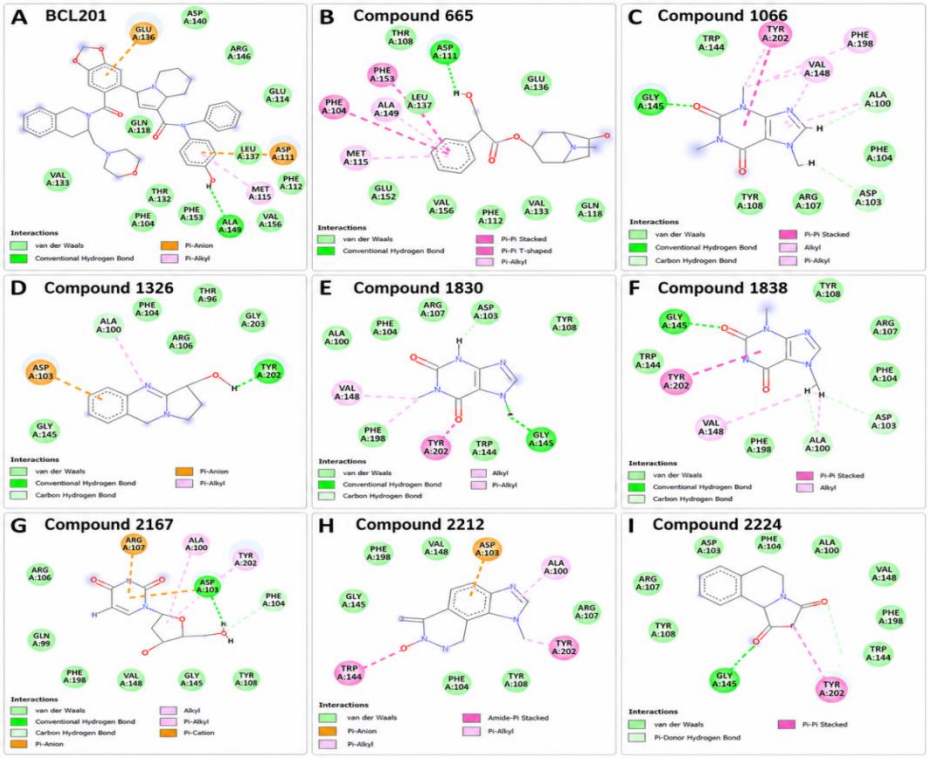


Figure III. 23Binding Interaction Maps of Selected MPA13 Compounds with BCL2.

4.2.1 Description of binding modes

The 2D interaction maps indicate that the BCL-2 complexes are mainly stabilized by a combination of van der Waals contacts, hydrogen-bond interactions, and π-mediated interactions. The universal presence of van der Waals contacts across all complexes, shown in Figure III .22, suggests that hydrophobic surface complementarity represents a common stabilizing factor for ligand accommodation in the BCL-2 binding pocket.

Several ligands, including BCL201, NC 665, NC 1066, NC 1830, NC 1838, NC 2167, and NC 2212, formed conventional hydrogen bonds. These interactions may help orient the ligands within the binding pocket and improve binding specificity. In particular, residues such as Asp103, Asp111, Gly145, and Tyr202 were involved in hydrogen-bond-related contacts depending on the ligand.

The interaction profile was also strongly influenced by aromatic and π-related contacts. Compounds 665, 1066, 2167, and 2224 showed π–π stacked interactions, while NC 1326 showed π–π T-shaped and amide–π stacked interactions. BCL201, NC 1326, NC 1830, and NC 2212 displayed π-anion interactions, and NC 2212 additionally

showed a π-cation interaction. These observations suggest that aromatic moieties play an important role in ligand stabilization within the BCL-2 binding pocket, which is consistent with the recognized contribution of aromatic interactions in ligand–protein interfaces [33].

4.2.2 Key amino acids involved

As presented in Figure III .24, several amino acid residues were repeatedly observed across the BCL-2 complexes. The most recurrent residue was Phe104, which appeared in all analyzed complexes. This suggests that Phe104 may represent a central contact point within the BCL-2 binding site.

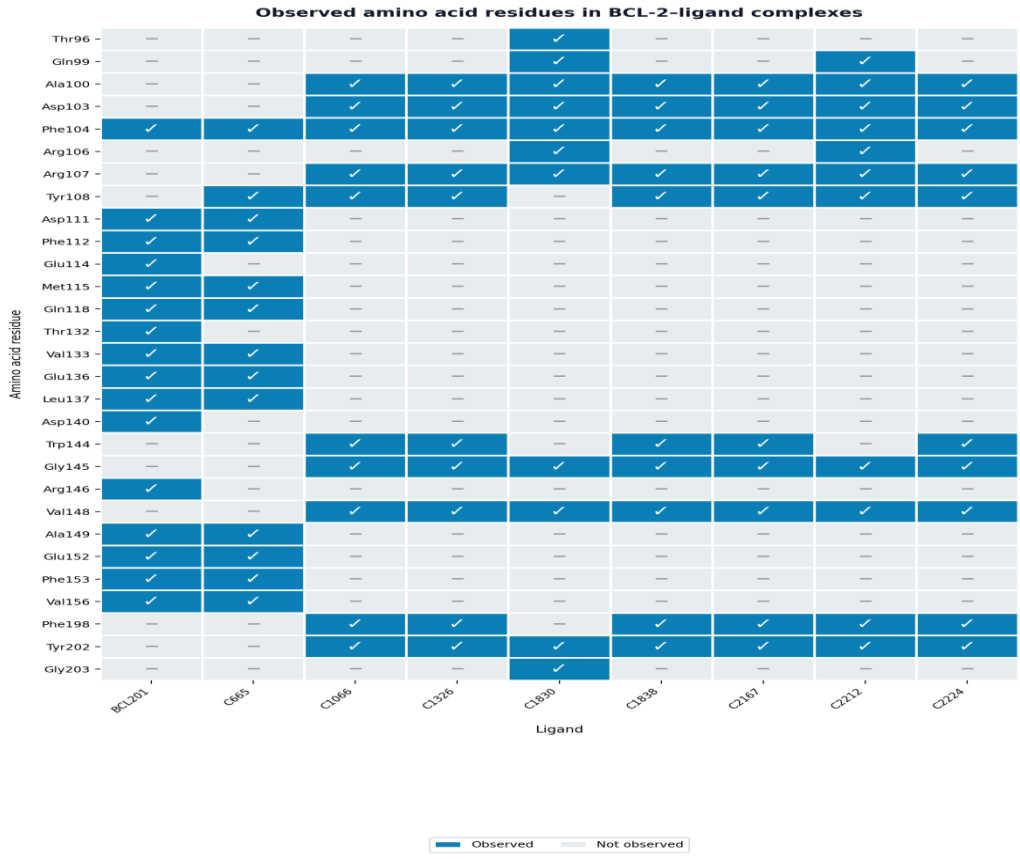


Figure III. 24 Observed Amino Acid Residues in BCL2 Ligand Complexes.

4.2.3 Structure–activity relationship elements

From a SAR perspective, the BCL-2 interaction profiles suggest that the selected compounds rely on two major structural features: polar functional groups capable of hydrogen bonding and aromatic or hydrophobic regions capable of π-related interactions.

Compounds that displayed conventional hydrogen bonds, such as 1066, 1838, 2167, and 2212, may benefit from improved orientation within the BCL-2 pocket. In contrast, compounds rich in aromatic contacts, such as 665, 1326, 2167, and 2224, appear to exploit π–π or π-alkyl interactions with aromatic residues such as Phe104, Phe198, and Tyr202.

Among the analyzed compounds, NC 2212 showed the most diversified interaction profile, combining van der Waals contacts, hydrogen bonds, carbon hydrogen bonds, π-anion, π-cation, alkyl, and π-alkyl interactions. This diversity may indicate a more complex stabilization pattern within the BCL-2 pocket. However, this interpretation remains based only on interaction maps .

4.3 Interaction Analysis of CDK4–Ligand Complexes

As shown in Figure III.25 and Figure III.26, all CDK4 complexes displayed van der Waals interactions, indicating that non-specific hydrophobic contacts contribute to ligand stabilization within the CDK4 binding pocket.

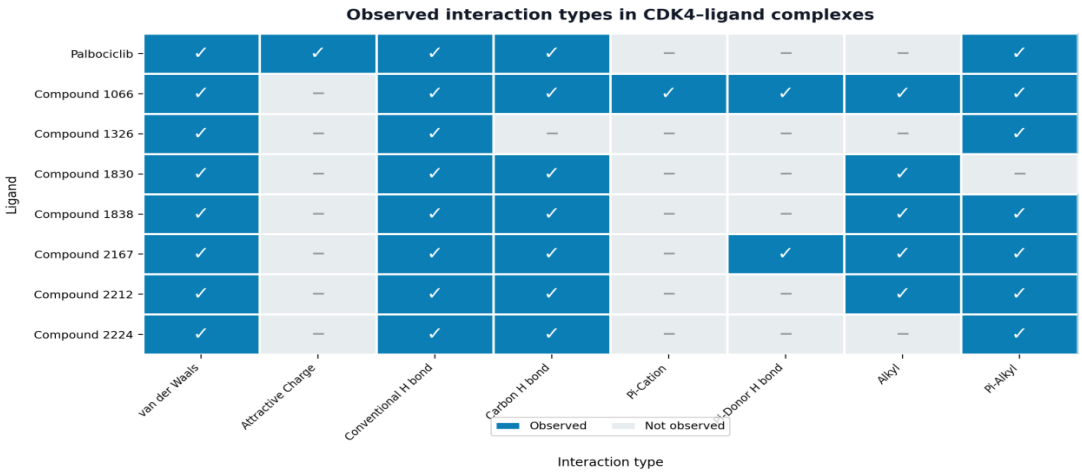


Figure III. 25 Observed Intreraction Types in CDK4 Ligand Interaction.

Most ligands also showed conventional hydrogen bonds and carbon hydrogen bonds, suggesting that polar interactions contribute to ligand orientation. π-mediated interactions, especially π-alkyl interactions, were also frequently observed.

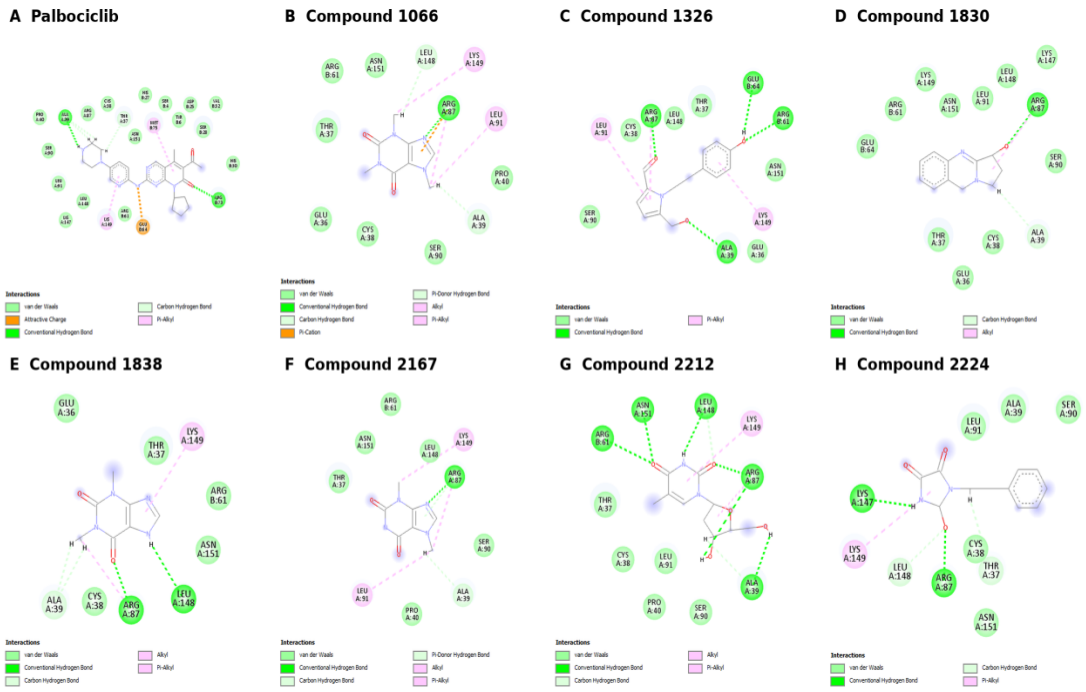


Figure III. 26 Binding Interaction Maps of Selected MPA13 Compounds with CDK4.

4.3.1 Description of binding modes

The CDK4 interaction maps indicate that ligand stabilization is mainly governed by a combination of van der Waals contacts, hydrogen bonds, carbon hydrogen bonds, and π-alkyl interactions. The universal presence of van der Waals contacts in Figure III.25 suggests that hydrophobic complementarity is a common feature of ligand accommodation within the CDK4 binding pocket.

The reference ligand palbociclib displayed a diversified interaction profile, including conventional hydrogen bonds, carbon hydrogen bonds, attractive charge interactions, and π-alkyl contacts. This combination suggests that palbociclib is stabilized by both polar/electrostatic interactions and hydrophobic contacts.

Among the selected compounds, NC 1066 showed the most diverse interaction profile, with conventional hydrogen bonds, carbon hydrogen bonds, π-cation, π-donor hydrogen bond, alkyl, and π-alkyl interactions. NC 2167 also presented a relatively rich

profile, combining hydrogen-bond interactions with alkyl, π-alkyl, and π-donor hydrogen bond contacts. Compounds 1838, 2212, and 2224 showed balanced profiles dominated by hydrogen-bond and π-alkyl interactions, while NC 1830 was mainly stabilized by hydrogen-bond and alkyl contacts.

4.3.2 Key amino acids involved

As shown in Figure III.27, several residues were repeatedly observed across the CDK4 complexes. The frequency analysis identified Ala39, Thr37, Cys38, Arg87, Leu148, Lys149, and Asn151 as the most recurrent residues, appearing in all analyzed complexes. These residues may therefore represent key contact points within the CDK4 binding site.

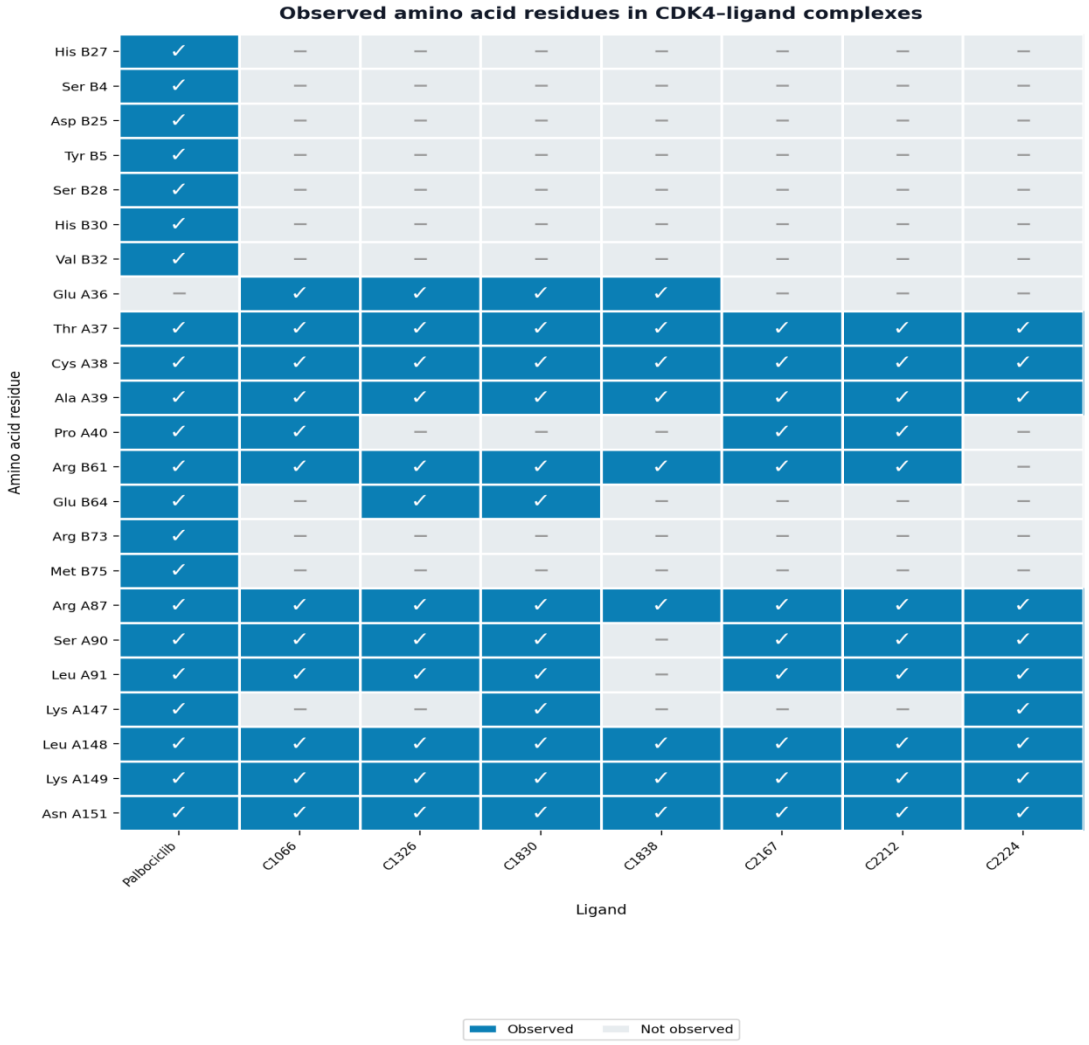


Figure III. 27 Observed Amino Acid Residues in CDK4 Ligand Interaction.

Other residues, including Leu91, Ser90, and Arg61, were also frequently observed. Their recurrence suggests that they may contribute to ligand stabilization through hydrophobic or polar contacts, depending on the ligand structure.

4.3.3 Structure–activity relationship elements

From a SAR perspective, the CDK4 interaction profiles suggest that the selected compounds rely on two main structural elements: polar groups capable of hydrogen bonding and hydrophobic/aromatic regions capable of alkyl or π-alkyl interactions.

Compounds such as 1066, 2167, 2212, and 2224 showed interaction profiles involving both hydrogen-bond contacts and π-related interactions. This suggests that their scaffolds may allow simultaneous polar anchoring and hydrophobic stabilization within the CDK4 pocket.

The repeated involvement of Arg87, Ala39, Leu148, Lys149, and Asn151 indicates that structural modifications maintaining interactions with these residues could be important for preserving CDK4 binding. In particular, groups able to maintain hydrogen bonds with Arg87 or hydrophobic/π-alkyl contacts with Lys149 and Leu148 may be favorable for ligand stabilization.

4.4 Interaction Analysis of PARP1–Ligand Complexes

As shown in Figure III.28 and Figure III.29, all PARP1 complexes displayed van der Waals interactions, suggesting that hydrophobic contacts contribute to ligand accommodation within the PARP1 binding site. NC 848, NC 1249, and rucaparib formed conventional hydrogen bonds, while NC 1034 and rucaparib showed carbon hydrogen bonds. π-mediated interactions were also observed, including π-alkyl interactions for NC 848, π–π stacked interactions for NC 1034 and rucaparib, and a π-cation interaction for rucaparib.

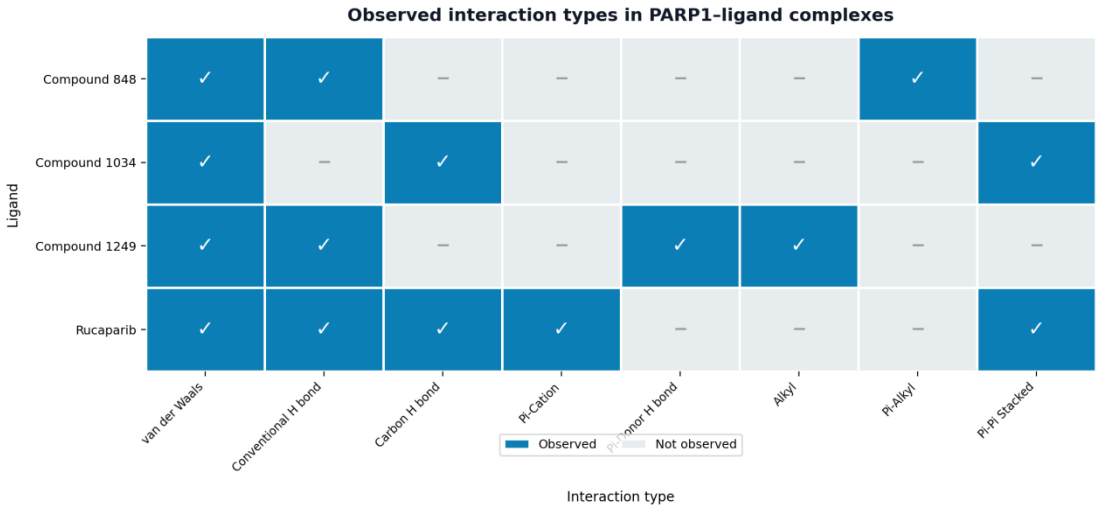


Figure III. 28 Observed Interaction Types in PARP1 Ligand Complexes.

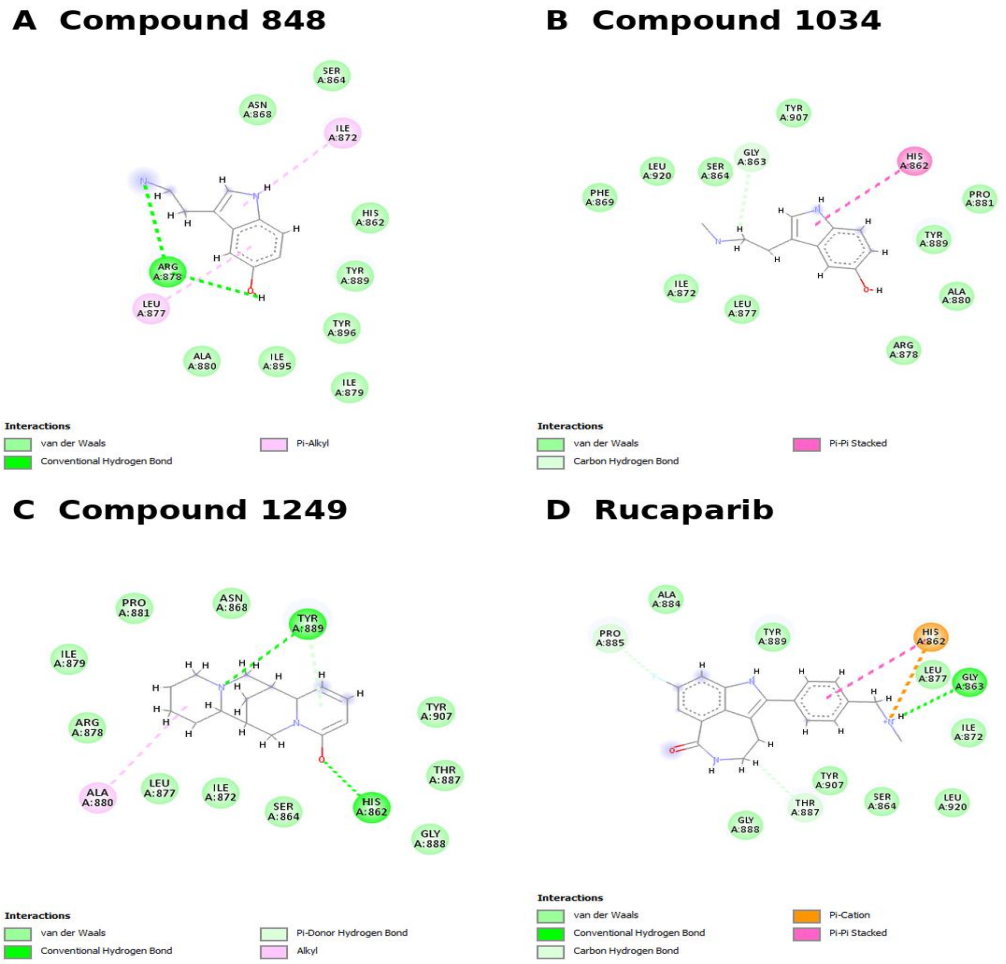


Figure III. 29 Binding Interaction Maps of Selected MPA13 Compounds with PARP1.

4.4.1 Description of binding modes

The PARP1 interaction maps indicate that ligand stabilization is mainly governed by van der Waals contacts, hydrogen-bond interactions, and π-mediated interactions. The universal presence of van der Waals contacts. suggests that hydrophobic complementarity is a common factor in ligand accommodation within the PARP1 binding pocket.

NC 848 formed conventional hydrogen bonds and π-alkyl interactions. The interaction with Arg878, together with hydrophobic contacts involving Ile872, Leu877, Ala880, and Ile895, suggests that this compound is stabilized by both polar anchoring and hydrophobic contacts.

NC 1034 showed a different profile, with carbon hydrogen bonding and π–π stacked interaction. The presence of His862, Tyr889, and Tyr907 around the ligand indicates that aromatic residues may contribute to ligand stabilization through π-mediated contacts.

NC 1249 displayed conventional hydrogen bonds, a π-donor hydrogen bond, and alkyl interactions. This interaction profile suggests a relatively balanced binding mode involving both polar and hydrophobic stabilization. The involvement of His862, Tyr889, Arg878, and Ala880 supports the role of both hydrogen bonding and hydrophobic contacts in the PARP1 pocket.

The reference ligand rucaparib presented the most diversified interaction profile among the analyzed PARP1 complexes. It combined conventional hydrogen bonding, carbon hydrogen bonding, π-cation interaction, and π–π stacking. This broader interaction pattern suggests that rucaparib benefits from both polar anchoring and aromatic/electrostatic stabilization within the PARP1 binding pocket.

4.4.2 Key amino acids involved

As shown in Figure III.30, several residues were repeatedly observed across the PARP1 complexes. The frequency analysis identified His862, Ser864, Ile872, Leu877, and Tyr889 as the most recurrent residues, appearing in all analyzed complexes. These residues may therefore represent major contact points within the PARP1 binding site.

Other residues, including Arg878, Ala880, and Tyr907, were observed in three complexes, suggesting that they may also contribute significantly to ligand stabilization. Residues such as Gly863, Asn868, Ile879, Pro881, Thr887, Gly888, and Leu920 were observed in two complexes and may participate in ligand accommodation depending on the ligand scaffold.

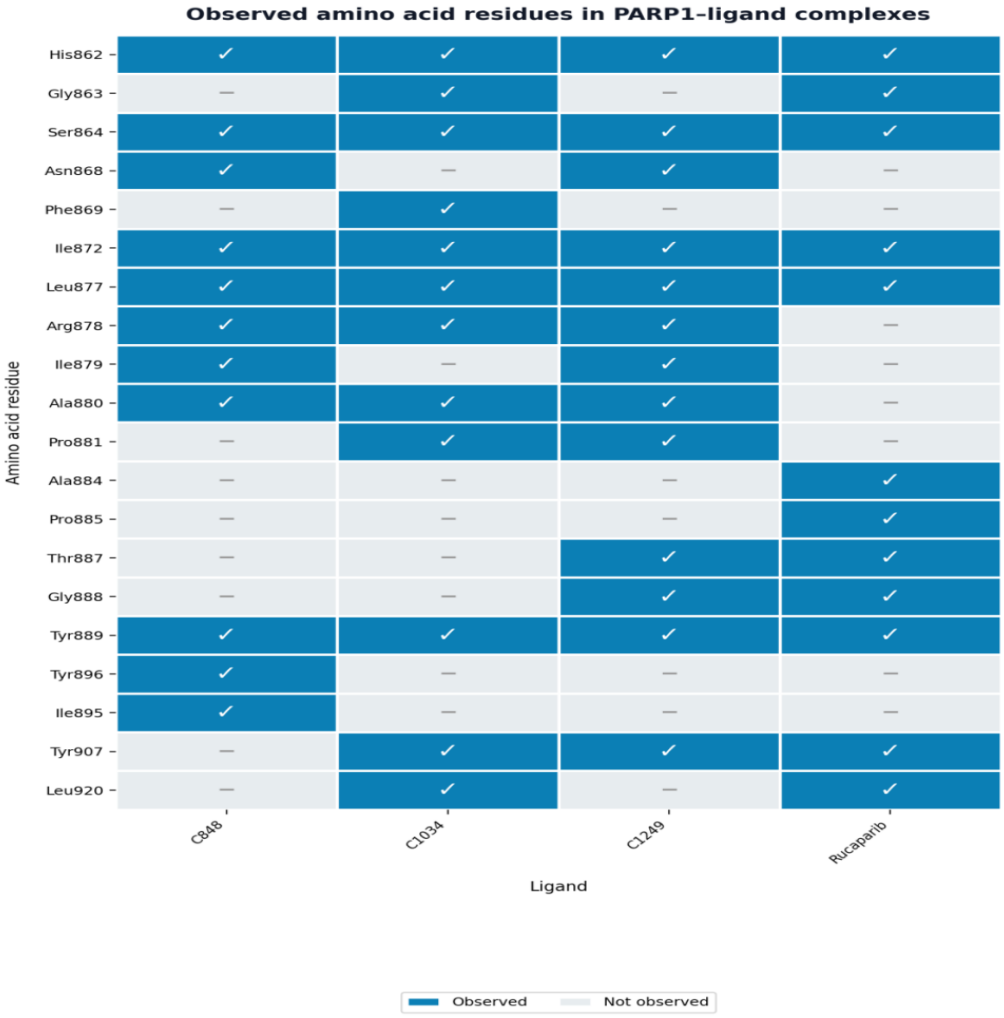


Figure III. 30 Observed Amino Acid Residues in PARP1 Ligand Complexes.

4.4.3 Structure–activity relationship elements

From a SAR perspective, the PARP1 interaction profiles suggest that efficient ligand stabilization depends on the presence of polar groups capable of hydrogen bonding and aromatic/hydrophobic regions capable of π-mediated interactions.

Compounds 848 and 1249 formed conventional hydrogen bonds, suggesting that their polar functional groups are favorably oriented toward polar residues of the PARP1 binding site. NC 1034 relied more on carbon hydrogen bonding and π–π stacking, indicating that aromatic stacking may contribute significantly to its binding mode. Rucaparib displayed the most diverse interaction profile, combining hydrogen bonding, carbon hydrogen bonding, π-cation interaction, and π–π stacking, which may explain its stable positioning as a reference ligand.

The repeated involvement of His862, Leu877, Ile872, and Tyr889 suggests that structural features maintaining hydrophobic and aromatic interactions with these residues may be important for PARP1 binding. In addition, maintaining polar interactions with residues such as Arg878, Ser864, or Gly863 could help improve ligand orientation and specificity.

4.5 Interaction Analysis of VEGFR2–Ligand Complexes

As shown in Figure III.31 and Figure III.32, all VEGFR2 complexes displayed van der Waals interactions, conventional hydrogen bonds, carbon hydrogen bonds, and π-alkyl interactions. This indicates that the analyzed ligands share a common stabilization pattern involving both polar and hydrophobic contacts. In addition, π-anion interactions were observed for NC 1066, NC 1759, and sorafenib, whereas π-cation interactions were observed for NC 1838 and NC 2167. Compounds 1838 and 2167 also showed π-sulfur interactions, suggesting additional stabilization involving sulfur-containing residues such as Cys1045.

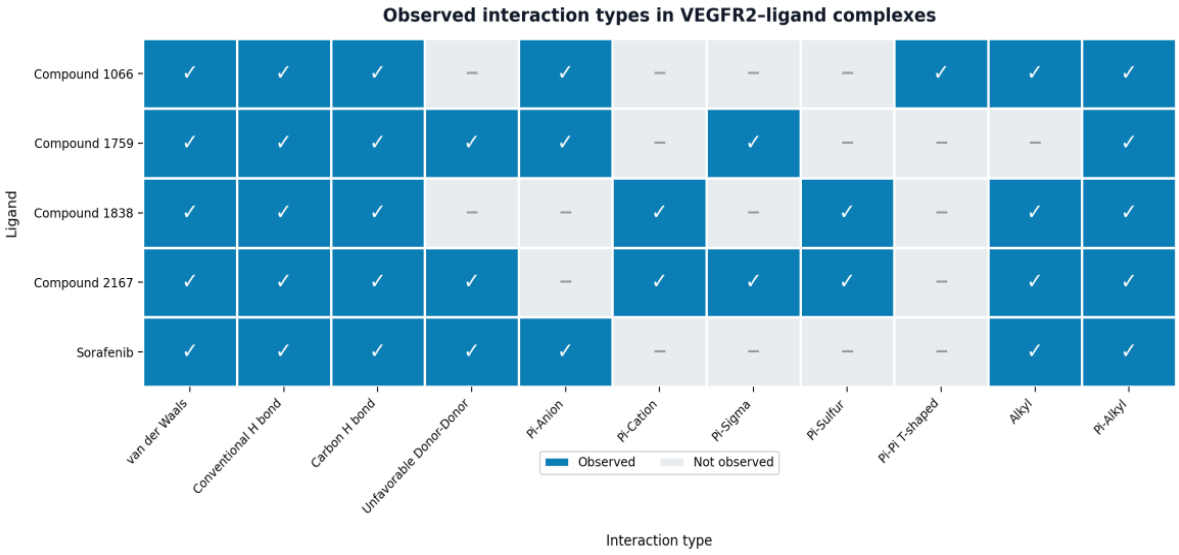


Figure III. 31 Observed Interaction Types in VEGFR2 Ligand Complexes.

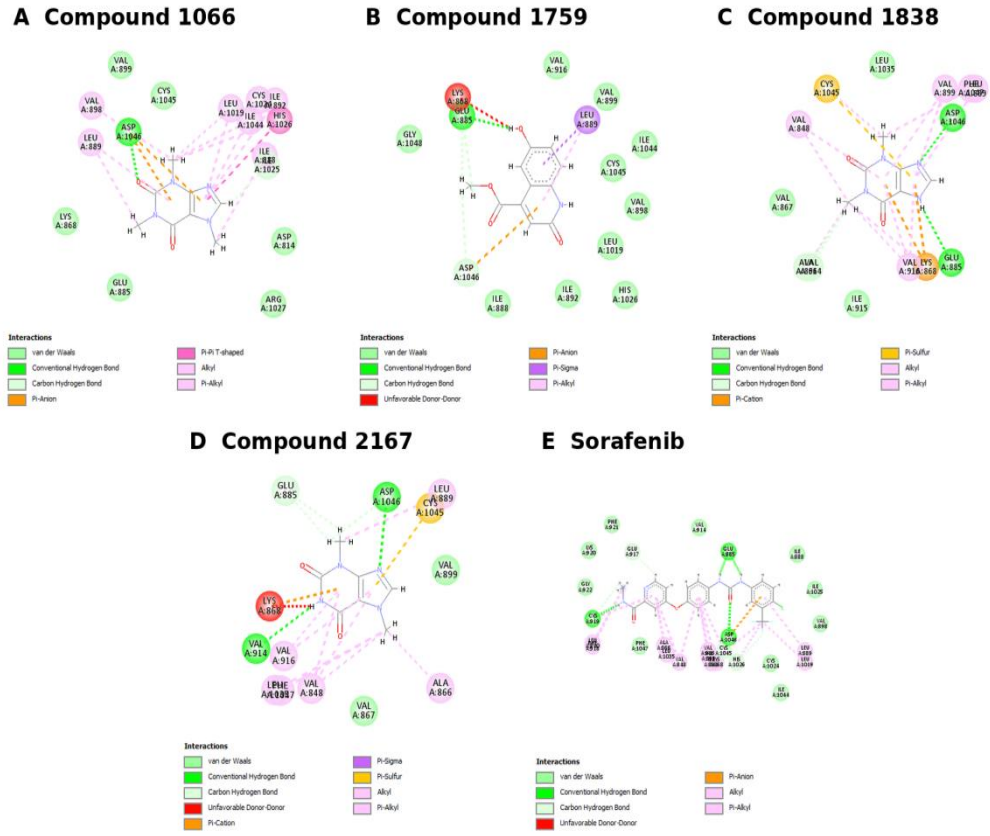


Figure III. 32 Binding Interaction Maps of Selected MPA13 Compounds with VEGFR2.

4.5.1 Description of binding modes

The VEGFR2 interaction maps indicate that all analyzed ligands are stabilized by a combination of hydrogen-bond interactions, carbon hydrogen bonds, hydrophobic contacts, and π-mediated interactions. The universal presence of van der Waals contacts suggests that hydrophobic complementarity is a common factor in ligand accommodation within the VEGFR2 binding pocket.

NC 1066 exhibited a diversified binding profile involving conventional hydrogen bonding, carbon hydrogen bonding, π-anion, π–π T-shaped, alkyl, and π-alkyl interactions. Its interactions with Asp1046, Lys868, Glu885, Leu889, and Val899 suggest a binding mode supported by both polar anchoring and hydrophobic stabilization.

NC 1759 displayed hydrogen bonds, carbon hydrogen bonds, π-anion, π-sigma, and π-alkyl interactions. However, the presence of an unfavorable donor–donor interaction suggests that part of its binding orientation may include a less favorable polar contact, which could reduce binding stability despite the presence of multiple stabilizing interactions.

NC 1838 showed conventional hydrogen bonds, carbon hydrogen bonds, π-cation, π-sulfur, alkyl, and π-alkyl interactions. The presence of a π-sulfur contact involving Cys1045 and hydrophobic contacts with Val899, Val916, and Phe1047 suggests favorable insertion within the VEGFR2 binding region.

NC 2167 presented the broadest interaction diversity among the selected compounds, including π-cation, π-sigma, π-sulfur, alkyl, and π-alkyl contacts. This indicates that NC 2167 may engage both polar and hydrophobic regions of the VEGFR2 pocket. Nevertheless, the observed unfavorable donor–donor interaction should be considered when interpreting its binding quality.

The reference ligand sorafenib showed a broad interaction pattern involving conventional hydrogen bonds, carbon hydrogen bonds, π-anion, alkyl, and π-alkyl interactions. Its interaction profile overlaps with several residues observed for the

selected compounds, supporting the relevance of the common VEGFR2 binding-site residues identified in this analysis.

4.5.2 Key amino acids involved

As shown in Figure III.33, several residues were repeatedly observed across the VEGFR2 complexes. The frequency analysis identified Asp1046, Val899, and Glu885 as the most recurrent residues, appearing in all analyzed complexes. These residues may therefore represent major contact points within the VEGFR2 binding site.

Other frequently observed residues included Lys868, Cys1045, Leu889, and Val916, which appeared in four complexes. Residues such as Val848, Ala866, Val898, Leu1019, His1026, Ile1044, and Phe1047 were also observed in several complexes, indicating their contribution to ligand accommodation depending on the ligand scaffold.

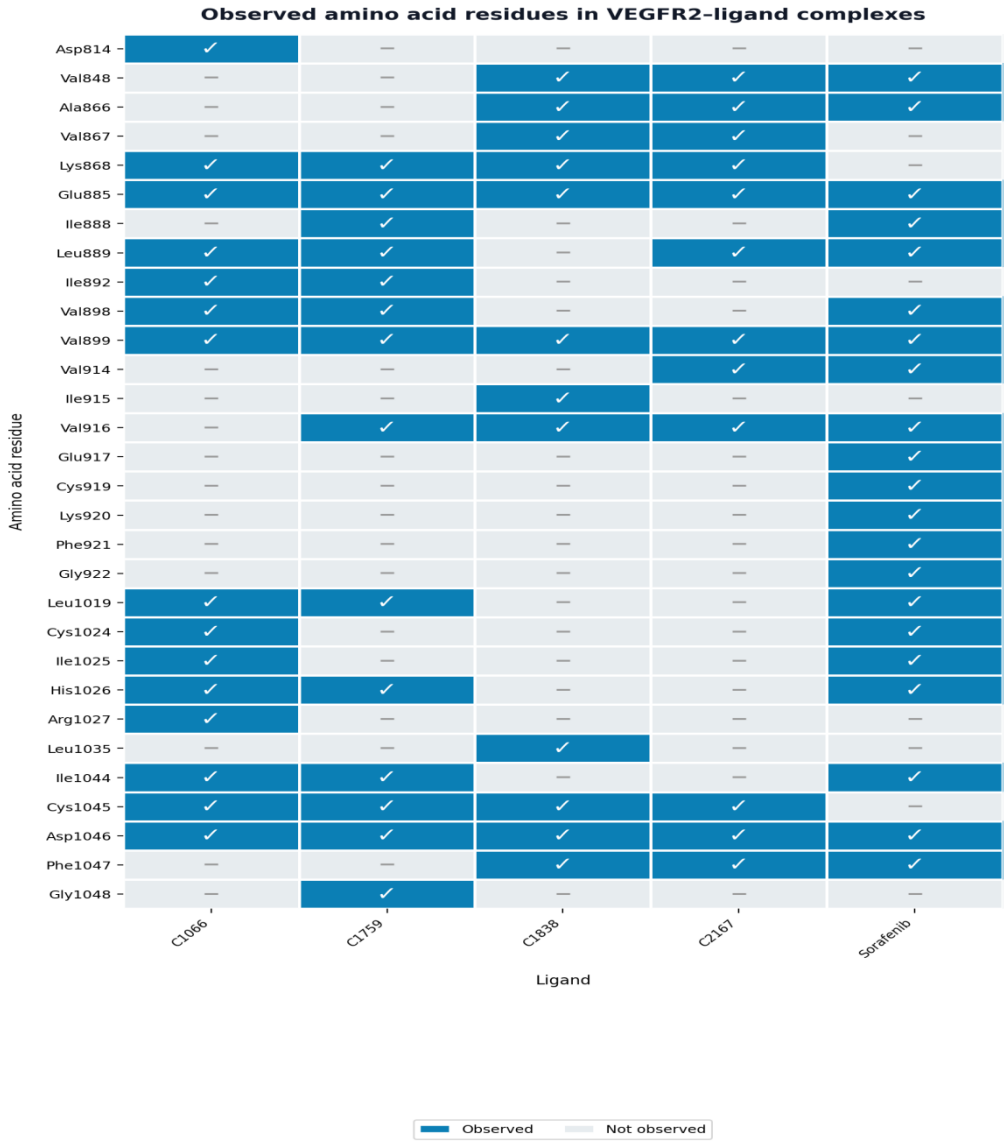


Figure III. 33 Observed Amino Acid Residues in VEGFR2 Ligand Complexes.

4.5.3 Structure–activity relationship elements

From a SAR perspective, the VEGFR2 interaction profiles suggest that ligand stabilization depends on the combination of polar functional groups capable of hydrogen bonding and hydrophobic or aromatic regions capable of alkyl, π-alkyl, and other π-mediated interactions.

The repeated involvement of Asp1046 and Glu885 indicates that polar groups capable of forming hydrogen bonds with acidic residues may support ligand anchoring within the binding pocket. In parallel, the frequent presence of Val899, Leu889, Val916, and Phe1047 suggests that hydrophobic/aromatic surfaces are important for maintaining ligand accommodation.

Compounds 1838 and 2167 showed π-sulfur interactions, suggesting that structural motifs positioned near Cys1045 may contribute additional stabilization. NC 2167 displayed the widest interaction diversity, but its unfavorable donor–donor contact indicates that optimization may be needed to reduce unfavorable polar interactions while preserving beneficial hydrophobic and π-mediated contacts.

Conclusion

This work focused on the **molecular simulation of the interaction between natural anticancer compounds and their biological targets**. The study began with the selection of important cancer-related molecular targets, including kinases, hormone receptors, anti-apoptotic proteins, DNA repair enzymes, and cell cycle regulators. These targets are strongly involved in cancer progression because they control essential mechanisms such as cell proliferation, survival, apoptosis, DNA repair, and cell cycle regulation.

The work also included the preparation of protein structures obtained from the Protein Data Bank and the collection and optimization of natural compounds from chemical databases and literature sources. In silico ADME/Tox analysis was performed to evaluate the pharmacokinetic behavior, drug-likeness, and potential toxicity of the selected compounds. Molecular docking and cross-docking studies were then carried out to predict the binding affinity of each ligand toward the selected biological targets. Finally, post-docking analysis allowed the identification of important molecular interactions, including hydrogen bonds, hydrophobic interactions, π–π stacking, and key amino acid residues involved in ligand binding.

The scientific contribution of this study lies in the use of computational drug design methods to screen natural compounds against several anticancer targets. This approach helps to identify promising molecules with potential multi-target activity and acceptable pharmacokinetic and toxicity profiles. It also provides a rational basis for selecting compounds that could be further investigated experimentally.

However, this study has some limitations. Molecular docking and ADME/Tox predictions are computational approaches and do not fully reproduce the complexity of biological systems. Docking scores are only predictive and may not always reflect the real biological activity of a compound. In addition, protein flexibility, solvent effects, cellular metabolism, and experimental toxicity were not completely evaluated in this in silico study.

Future work should therefore include molecular dynamics simulations to study the stability of the best ligand–target complexes over time. The most promising compounds should also be validated through in vitro and in vivo experiments to confirm their anticancer activity, toxicity, and mechanism of action. Further structural optimization

may also be performed to improve their binding affinity, selectivity, pharmacokinetic properties, and therapeutic potential.

In conclusion, this project highlights the importance of natural compounds as valuable sources of anticancer agents and demonstrates the usefulness of molecular simulation in the early stages of drug discovery. The results obtained may serve as a starting point for further experimental studies and for the development of new anticncer therapeutic strategies.

References

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-674. doi : 10.1016/j.cell.2011.02.013.
2. Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression. *Cancer Res*. 2019;79(18):4557-4566. doi : 10.1158/0008-5472.CAN-18-3962.
3. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science*. 2013;339(6127):1546-1558. doi : 10.1126/science.1235122.
4. Baylin SB, Jones PA. Epigenetic determinants of cancer. *Cold Spring Harb Perspect Biol*. 2016;8(9):a019505. doi: 10.1101/cshperspect.a019505.
5. Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer*. 2013;13(10):714-726. doi:10.1038/nrc3599.
6. Vasan N, Baselga J, Hyman DM. A view on drug resistance in cancer. *Nature*. 2019;575(7782):299-309. doi :10.1038/s41586-019-1730-1.
7. Newman DJ, Cragg GM. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J Nat Prod*. 2020;83(3):770-803. doi:10.1021/acs.jnatprod.9b01285.
8. Atanasov AG, Zotchev SB, Dirsch VM, Supuran CT. Natural products in drug discovery: advances and opportunities. *Nat Rev Drug Discov*. 2021;20(3):200-216. doi:10.1038/s41573-020-00114-z.
9. Aggarwal BB, Danda D, Gupta S, Gehlot P. Models for prevention and treatment of cancer: problems vs promises. *Biochem Pharmacol*. 2009;78(9):1083-1094. doi:10.1016/j.bcp.2009.05.027.
10. Izzo AA, Teixeira M, Alexander SPH, Cirino G, Docherty JR, George CH, et al. Natural products and cancer: from drug discovery to prevention and therapy. *Br J Pharmacol*. 2025;182:2069-2074. doi:10.1111/bph.70014.
11. Chidananda C, Thakur G, Datta D, Popat K. A comprehensive review of alkaloids in cancer therapy: focusing on molecular mechanisms and synergistic potential of piperine in colorectal cancer. *3 Biotech*. 2025;15:403. doi:10.1007/s13205-025-04535-8.

12. Muttiah B, Abdullah NAH. Natural compounds in ovarian cancer: mechanistic insights and therapeutic potential. *Front Pharmacol*. 2026;16:1687805. doi:10.3389/fphar.2025.1687805.
13. Chunarkar-Patil P, Kaleem M, Mishra R, Ray S, Ahmad A, Verma D, et al. Anticancer drug discovery based on natural products: from computational approaches to clinical studies. *Biomedicines*.2024;12(1):201. doi:10.3390/biomedicines12010201.
14. Yang L-J, Han T, Liu R-N, et al. Plant-derived natural compounds: A new frontier in inducing immunogenic cell death for cancer treatment. Biomed Pharmacother. 2024;177:117099. doi: 10.1016/j.biopha.2024.117099.
15. El-Seedi HR, Refaey MS, Elias N, et al. Marine natural products as a source of novel anticancer drugs: an updated review (2019–2023). *Nat Prod Bioprospect.* 2025;15(1):13. doi: 10.1007/s13659-024-00493-5.
16. Giurini EF, Godla A, Gupta KH. Redefining bioactive small molecules from microbial metabolites as revolutionary anticancer agents. *Cancer Gene Ther*. 2024;31(2):187-206. doi : 10.1038/s41417-023-00715-x.
17. Fang L, Xu L, Kader M, Ding T, Lu S, Wang D, et al. Salt-adapted microorganisms: a promising resource for novel anti-cancer drug discovery. *Mar Drugs*. 2025;23(8):296. doi: 10.3390/md23080296.
18. García-Estrada C, Barreiro C, Martín JF. Anticancer secondary metabolites produced by fungi: potential and representative compounds. *Int J Mol Sci*. 2025;27(1):101. doi : 10.3390/ijms27010101.
19. Li Q, Tong Y, Chen J, Xie T. Targeting programmed cell death via active ingredients from natural plants: a promising approach to cancer therapy. *Front Pharmacol*. 2024;15:1491802. doi:10.3389/fphar.2024.1491802.
20. Azevedo T, Ferreira T, Peña-Corona SI, Cortes H, Silva-Reis R, da Costa RMG, et al. Natural products-based antiangiogenic agents: new frontiers in cancer therapy. *Food Front*. 2024:1-44. doi :10.1002/fft2.466.
21. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep.* 2017;7:42717. doi:10.1038/srep42717.
22. Pinzi L, Rastelli G. Molecular docking: shifting paradigms in drug discovery. *Int J Mol Sci*. 2019;20(18):4331. doi:10.3390/ijms20184331

23. Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, et al. RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Res. 2019;47(D1):D464-D474. doi:10.1093/nar/gky1004.
24. Sorokina M, Merseburger P, Rajan K, Yirik MA, Steinbeck C. COCONUT online: collection of open natural products database. J Cheminform. 2021;13(1):2. doi:10.1186/s13321-020-00478-9.
25. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S , et al. PubChem 2023 update. Nucleic Acids Res. 2023;51(D1):D1373-D1380. doi: 10.1093/nar/gkac956.
26. Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S, et al. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Res. 2024;52(D1):D1180-D1192. doi:10.1093/nar/gkad1004.
27. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010;50(5):742-754. doi :10.1021/ci100050t.
28. Bajusz D, Rácz A, Héberger K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? J Cheminform. 2015;7:20. doi:10.1186/s13321-015-0069-3.
29. RDKit: Open-source cheminformatics software [Internet]. RDKit contributors; [cited 2026 Apr 26].
30. Banerjee P, Eckert AO, Schrey AK, Preissner R. ProTox-II: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Res. 2018;46(W1):W257-W263. doi:10.1093/nar/gky318.
31. Forli S. Charting a path to success in virtual screening. Molecules. 2015;20(10):18732-18758. doi:10.3390/molecules201018732.
32. Molecular Operating Environment (MOE), version 2019.0102. Montreal (QC, Canada): Chemical Computing Group ULC; 2019.
33. Brylinski M. Aromatic interactions at the ligand-protein interface: Implications for the development of docking scoring functions. Chem Biol Drug Des. 2018;91(2):380-390. doi :10.1111/cbdd.13084.
34. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. Journal of Chemical Information and Computer Sciences. 2002;42(6):1273–1280.doi:10.1021/ci010132r.