DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

MOHAMED KHIDER UNIVERSITY OF BISKRA

FACULTY OF EXACT SCIENCES, SCIENCES OF NATURE AND LIFE

DEPARTMENT OF MATTER SCIENCES

DISSERTATION

Presented by

BELKADI Ahlem

To obtain the degree of

Doctorate in Chemistry

Option:
Pharmaceutical and computational chemistry

Entitled:

Investigation of cytotoxic properties of some heterocyclic derivatives by molecular modeling approaches

Publicly defended on: 23th June 2022

In front of the jury composed of:

Mr. BELAIDI Salah	Prof.	University of Biskra	President
Mr. MELKEMI Nadjib	Prof.	University of Biskra	Supervisor
Mr. KENOUCHE Samir	MCA	University of Biskra	CO-Supervisor
Mr. MESSAOUDI Abdelatif	Prof.	University of Batna 1	Examiner
Mr. DAOUD Ismail	MCA	University of Biskra	Examiner

This dissertation is dedicated to my beloved parents

To my dear brother and sister

To my precious uncles and aunts

To all those who are dear to me

Acknowledgments

"Progress occurs when courageous, skilful leaders seize the opportunity to change things for the better." Harry S. Truman

Abstract

Currently, many technologies have been adopted to boost the efficiency of drug development and overcome obstacles in the drug discovery pipeline. The application of these approaches spans a wide range, from bioactivity predictions, de novo compound synthesis, target identification to hit discovery, and lead optimization. This dissertation comprises two studies. First, we proposed an original approach based on statistical consideration dedicated to k-means clustering analysis in order to define a set of rules for structural features that would help in designing novel anti-cancer drug candidates. It has been applied successfully to classify 500 cytotoxic compounds with 21 molecular descriptors into distinct clusters. The percentage of molecules in each cluster is 50%, 24.88%, and 25.12% for cluster 1, cluster 2, and cluster 3, respectively. Each cluster groups a homogeneous class of molecules with respect to their molecular descriptors. Silhouette analysis, used as a cluster validation approach proves that the molecules are very well clustered, and there are no molecules placed in the wrong cluster. In silico screening of pharmacological properties ADME and evaluation of drug-likeness were performed for all molecules. The quantitative analysis of molecular electrostatic potential was performed to identify the nucleophilic and electrophilic sites in the representative molecule of each cluster. In addition, a molecular docking study was carried out to investigate the interactions of the paragon molecules with the active binding sites of six different targets. Our findings provide a guide to assist the chemist in selecting and testing only the potential molecules with good pharmacokinetic profiles to improve the clinical outcomes of drug therapies.

Second, a simulation-based investigation was conducted to examine the CHK1 inhibitory activity of cytotoxic xanthone derivatives using a hierarchical workflow for molecular docking, MD simulation, ADME-TOX prediction, and MEP analysis. A molecular docking study was conducted for the forty-three xanthone derivatives along with standard Prexasertib into the selected CHK1 protein structures 7AKM and 7AKO. Furthermore, MD studies support molecular docking results and validate the stability of studied complexes in physiological conditions. Moreover, *in silico* ADME-TOX studies are used to predict the pharmacokinetic, pharmacodynamic, and toxicological properties of the selected eight xanthones and the standard Prexasertib. The quantitative analysis of electrostatic potential was performed for the lead compound L36 to identify the reactive sites and possible non-covalent interactions.

Our study provides new unexplored insights into xanthones as CHK1 inhibitors and identified L36 as a potential drug candidate that could undergo further *in vivo* assays and optimization, laying a solid foundation for the development of CHK1 inhibitors and cancer drug discovery. To the best of our knowledge, this is the first time such a study was conducted for the xanthones with CHK1 by using a computational based approach.

Keywords: k-means clustering, cytotoxic activity, statistical analysis, ADME, drug-likeness, Molecular docking, CHK1, Xanthones, MD simulation, ADME-TOX, MEP analysis, Prexasertib.

الملخص

حاليًا تم اعتماد العديد من التقنيات لتعزيز كفاءة تطوير الأدوية والتغلب على العقبات في عملية اكتشاف الأدوية. يمتد تطبيق هذه الأساليب على نطاق واسع من تنبؤات النشاط الحيوي ، والتوليف المركب de novo ، وتحديد الهدف للوصول إلى الاكتشاف ، وتحسين المركبات الرائدة. تتكون هذه الرسالة من دراستين. أولاً اقترحنا نهجًا أصليًا يعتمد على اعتبارات إحصائية مخصصة للخوارزمية التصنيفية k-means من أجل تحديد مجموعة من القواعد للسمات الهيكلية التي من شأنها أن تساعد في تصميم أدوية جديدة مضادة للسرطان. تم تطبيقه بنجاح لتصنيف 500 مركب سام للخلايا مع 21 واصف جزيئي الى مجموعات مختلفة. النسبة المئوية للجزيئات في كل مجموعة هي 50٪ ، 24.88٪ و 25.12٪ للمجموعة 1 و المجموعة 2 والمجموعة 3 على التوالي. كل مجموعة تجمع فئة متجانسة من الجزيئات فيما يتعلق بأوصافها الجزيئية. أثبت تحليل silhouette المستخدم كنهج للتحقق من نتائج التصنيف أن الجزيئات متجمعة جيدًا ولا توجد جزيئات موضوعة في المجموعة الخاطئة. تم إجراء فحص السيليكو للخصائص الدوائية ADME وتقييم تشابه الدواء لجميع الجزيئات. قمنا بإجراء التحليل الكمي للقدرة الكهروستاتيكية الجزيئية لتحديد المواقع النيوكليوفيلية والالكتروفيلية في الجزيء الممثل لكل مجموعة. و كذا أجريت دراسة الالتحام الجزيئي للتحقيق في تفاعلات جزيئات الباراجونات مع مواقع الارتباط النشطة لستة أهداف مختلفة. توفر النتائج التي توصلنا إليها دليلًا لمساعدة الكيميائي في اختيار واختبار الجزيئات الفعالة فقط مع خصائص حركية دوائية جيدة لتحسين النتائج السريرية للعلاجات الدوائية.

ثانيًا تم إجراء تحقيق قائم على المحاكاة لفحص النشاط المثبط لـ CHK1 لمشتقات الزانثون السامة للخلايا باستخدام سير عمل هرمي للالتحام الجزيئي ، ومحاكاة MD ، وتنبؤ ADME-TOX ، وتحليل MEP. تم إجراء دراسة الالتحام الجزيئي لمركبات الزانثون الثلاثة والأربعين جنبًا إلى جنب مع Prexasertib المرجعي نحو هياكل بروتين CHK1 المختارة 7AKM و 7AKO. علاوة على ذلك ، تدعم دراسات MD نتائج الالتحام الجزيئي وتأكد استقرار المعقدات المدروسة في الظروف الفيزيولوجية. أيضا تم استخدام دراسات السيليكو ADME-TOX للتنبؤ بالحركية الدوائية ، والخصائص السمية للزانثونات الثمانية المختارة و Prexasertib المرجعي. تم إجراء التحليل الكمي للقدرة الكهروستاتيكية للمركب الرئيسي L36 لتحديد المواقع التفاعلية المهمة والروابط غير التساهمية المحتملة.

توفر دراستنا رؤى جديدة غير مستكشفة حول الزانثونات كمثبطات CHK1 وحددنا L36 كمرشح دواء قوي يمكن أن يخضع لمزيد من الاختبارات والتحسين في الجسم الحي ، مما يضع أساسًا متينًا لتطوير مثبطات CHK1 واكتشاف أدوية السرطان. على حد علمنا، هذه هي المرة الأولى التي يتم فيها إجراء مثل هذه الدراسة للزانثونات مع CHK1 باستخدام نهج قائم على الحسابات.

الكلمات المفتاحية: k-means ، النشاط السام للخلايا ، التحليل الإحصائي ، ADME ، تشابه الدواء ، الالتحام الجزيئي ، CHK1، الزانثونات ، محاكاة MD، ADME-TOX، تحليل MEP، Prexasertib.

Contents

CHAPTER Ⅰ : Biological background

CHAPTER Ⅱ : Computational methods for drug discovery and development

CHAPTER Ⅲ: K-means clustering analysis, ADME/Pharmacokinetic Prediction, MEP and Molecular docking studies of potential cytotoxic agents

CHAPTER Ⅳ: Molecular Docking/Dynamic simulations, MEP, ADME-TOX based analysis of xanthone derivatives as CHK1 inhibitors

List of works

International publications:

1. **Belkadi A**, Kenouche S, Melkemi N, et al (2021) K-means clustering analysis, ADME/pharmacokinetic prediction, MEP, and molecular docking studies of potential cytotoxic agents. Struct Chem 32:2235–2249.   https://doi.org/10.1007/s11224-021-01796-w

2. **Belkadi A**, Kenouche S, Melkemi N, et al (2022) Molecular docking/dynamic simulations, MEP, ADME-TOX-based analysis of xanthone derivatives as CHK1 inhibitors. Struct Chem. https://doi.org/10.1007/s11224-022-01898-z

3. Kenouche S, **Belkadi A**, Djebaili R, Melkemi N (2021) High regioselectivity in the amination reaction of isoquinolinequinone derivatives using conceptual DFT and NCI analysis. Journal of Molecular Graphics and Modelling 104:107828. https://doi.org/10.1016/j.jmgm.2020.107828

International conferences:

1. **Belkadi A**, Harkat H, Gueraiche R, Préparation des dérivés dihydropyrimidinones par catalyse heterogene USY-H, Le Cinquième Colloque International de chimie (CIC-5), 4-6 december 2018, Batna, Algeria.

2. **Belkadi A**, Melkemi N, Kenouche S, Fellousi S, QSAR study of Wogonin derivatives as cytotoxic agents, the 5th International Workshop on the thermodynamics of metallic alloys (WITAM), 13-14 november 2019, Batna, Algeria.

3. **Belkadi A**, Melkemi N, Kenouche S, A Relevant Tool for K-means Clustering Analysis, $13^{èmes}$ Journées internationales de Chimie Théorique et Computationnelle (JCTC'13), 2-4 february 2020, Biskra, Algeria.

National conferences:

1. **Belkadi A**, Melkemi N, Kenouche S, Fellousi S, Étude quantitative de la relation structure-activité (QSAR) des propriétés cytotoxiques d'une série de dérivés de 5,7-Dihydroxy-8-methoxy-2-phenyl-4*H*-chromen-4-one, Les 1ères Journées d'Etude sur la Chimie et ses Applications (JECA-1-2019), 27 november 2019, Batna, Algeria.

2. **Belkadi A**, Melkemi N, Kenouche S, Daoud I, Molecular docking studies of potent anticancer agents, Le 2ème Séminaire National Sciences d'Interfaces Chimie-Biologie, 20 february 2021, Souk-Ahras, Algeria.

List of abbreviations

WHO: World Health Organization

DNA: Deoxyribonucleic Acid

FDA: Food and Drug Administration

X-ray: Electromagnetic Radiation

MRI: Magnetic Resonance Imaging

CT: Computed Tomography

RNA: Ribonucleic Acid

ATP: Adenosine Triphosphate

MOA: Mechanism Of Action

FUCCI: Fluorescence Ubiquitination-based Cell Cycle Indicators

CHK1: Checkpoint kinase 1

DDR: DNA Damage Response

EGFR: Epidermal Growth Factor Receptor

PARP: Poly ADP-ribose Polymerase

MAPK: Mitogen-activated Protein Kinase

VEGF: Vascular Endothelial Growth Factor

STAT3: Signal Transducer and Activator of Transcription 3

CDKs : Cyclin-dependent kinases

ATR: Ataxia Telangiectasia and Rad3-related protein

PCA: Principal Component Analysis

HCA: Hierarchical Cluster Analysis

WSS: Within-cluster Sum of Squares

MEP: Molecular Electrostatic Potential

ADME: Absorption, Distribution, Metabolism, and Elimination

ADME-TOX: Absorption, Distribution, Metabolism, Elimination, and Toxicity

HIA: Human Intestinal Absorption

PK: PharmacoKinetic

BBB: Blood-Brain Barrier

CNS: Central Nervous System

P-gp: Permeability Glycoprotein

SVM: Support Vector Machine

CL: Clearance

DILI: Drug-induced liver injury

hERG: human Ether-à-go-go-Related Gene

LQTS: Long QT Syndrome

TdP: Torsade de Pointes

HTS: High Throughput Screening

PDB: Protein Data Bank

NMR: Nuclear Magnetic Resonance

RMSD: Root Mean Square Deviation

EM: Energy Minimization

MD: Molecular Dynamics

VDW: Van Der Waals

TPSA: Topological Polar Surface Area

DFT: Density Functional Theory

ESP: Electrostatic Potential

IC$_{50}$: Half Maximal Inhibitory Concentration

PAINS: Pan Assay Interference Compounds

List of figures

List of tables

General introduction

Cancer has been a great threat to humankind for thousands of years and despite the significant efforts of chemists in developing cancer treatments, but unfortunately it continues to be a scientific challenge to researchers worldwide. Until now, there is no effective agent that can strongly control the unlimited cell growth and selectively target cancer without damaging the healthy cells, causing severe side effects such as bone marrow suppression, GI tract lesions, hair loss, nausea, and the rapid development of clinical resistance [1]. The side effects vary from one person to another. Some of them disappear after a few days, but it usually takes a few months to recover from chemotherapy. The first practical chemotherapeutic agent was discovered accidentally during World War II, nitrogen mustard which had antileukemic properties, a powerful alkylating agent that damage DNA, leading to cell death [2-3]. Another important development in DNA-reactive drugs came with the discovery of cisplatin, which was also discovered by chance, provided a major advance in the treatment of testicular and ovarian carcinomas [4-5]. The success of these initial agents led to the development and synthesis of a large series of clinically useful derivatives, including Paclitaxel, Doxorubicin [6-7] and the kinase inhibitors; Imatinib, Gefitinib and Erlotinib [8-10].

Drug discovery and development is a long and expensive process that involves in vitro and in vivo experiments to evaluate the efficacy of compounds. Furthermore, pharmacokinetics PK and ADME studies are recently incorporated to improve the success rate of drug candidates, thereby get approval from the United States Food and Drug Administration FDA to reach the consumer market [11]. *In silico* medicinal chemistry is a well-established interdisciplinary field that encompasses chemoinformatics, molecular modeling, and computational chemistry. These techniques are applied effectively in the design of new drugs to reduce the costs, speed up the drug discovery cycle, and provide fruitful insights into the field of cancer [12].

This dissertation is divided into four chapters:

The first chapter is devoted to a biological background related to basic concepts of cancer, risk factors for cancer, diagnosis and treatments, cancer cell characteristics, cell death mechanisms, classification of chemotherapeutic drugs, drug discovery and development, CHK1 inhibition in cancer therapy, and the biological activity of xanthones.

The second chapter focuses on the basic methodologies and techniques used in our research, including the classification methods; K-means, PCA, hierarchical clustering and validation approaches; silhouette analysis, and elbow method. It also contains molecular modeling techniques; molecular docking and dynamics, as well as ADME-TOX descriptors, and MEP analysis. In the other chapters, we will present and discuss the results obtained from our studies.

In chapter Ⅲ, the k-means algorithm was applied on 500 cytotoxic agents to classify them into distinct clusters and the test statistic Z_k was effectively used to characterize the compounds by descriptors with the aim of identifying a set of rules for structural features that would help in designing novel anti-cancer drug candidates.

In chapter Ⅳ, a hierarchical *in silico* approach was conducted on xanthone compounds to identify and characterize new potential CHK1 inhibitors beneficial for cancer therapy.

Finally, a general conclusion will end this manuscript by exposing the perspectives and the improvements that can be made to consolidate the results obtained through this work.

References

1. Jaracz S, Chen J, Kuznetsova LV, Ojima I (2005) Recent advances in tumor-targeting anticancer drug conjugates. Bioorganic & Medicinal Chemistry 13:5043–5054. https://doi.org/10.1016/j.bmc.2005.04.084

2. Thurston DE (2007) Chemistry and pharmacology of anticancer drugs. CRC Press/Taylor & Francis, Boca Raton

3. Kohn KW, (1996) Beyond DNA Cross-Linking: History and Prospects of DNA-targeted Cancer Treatment Fifteenth Bruce F. Cain Memorial Award Lecture. American Association for Cancer Research 56:5533-5546

4. Baguley BC, Kerr DJ (2002) Anticancer drug development. Academic Press, San Diego

5. Nussbaumer S, Bonnabry P, Veuthey J-L, Fleury-Souverain S (2011) Analysis of anticancer drugs: A review. Talanta 85:2265–2289. https://doi.org/10.1016/j.talanta.2011.08.034

6. Spencer CM, Faulds D (1994) Paclitaxel: A Review of its Pharmacodynamic and Pharmacokinetic Properties and Therapeutic Potential in the Treatment of Cancer. Drugs 48:794–847. https://doi.org/10.2165/00003495-199448050-00009

7. Singal PK, Iliskovic N (1998) Doxorubicin-Induced Cardiomyopathy. The New England Journal of Medicine 339:900-905

8. Moen MD, McKeage K, Plosker GL, Siddiqui MAA (2007) Imatinib: A Review of its Use in Chronic Myeloid Leukaemia. Drugs 67:299–320. https://doi.org/10.2165/00003495-200767020-00010

9. Cohen MH, Williams GA, Sridhara R, et al (2003) FDA Drug Approval Summary: Gefitinib (ZD1839) (Iressa [®]) Tablets. The Oncologist 8:303–306. https://doi.org/10.1634/theoncologist.8-4-303

10. Cohen MH, Johnson JR, Chen Y, et al (2005) FDA Drug Approval Summary: Erlotinib (Tarceva ®) Tablets. The Oncologist 10:461–466. https://doi.org/10.1634/theoncologist.10-7-461

11. Zhang D, Luo G, Ding X, Lu C (2012) Preclinical experimental models of drug metabolism and disposition in drug discovery and development. Acta Pharmaceutica Sinica B 2:549–561. https://doi.org/10.1016/j.apsb.2012.10.004

12. Lin X, Li X, Lin X (2020) A Review on Applications of Computational Methods in Drug Screening and Design. Molecules 25:1375. https://doi.org/10.3390/molecules25061375

CHAPTER Ⅰ: Biological background

I.1 What is cancer?

Cancer is a term referring to a group of genetic diseases characterized by unlimited cell growth. There are more than 200 types of cancers, categorized based on the organ, tissue, and cell type in which cancer begins, each type of cancer requires a specific form of treatment [1]. Lung cancer is one of the most common types of cancer worldwide, with 2,206,771 cases recorded in 2020, as illustrated in (Figure Ⅰ. 1), where the number of deaths approaches the number of incidences. Likewise, liver cancer is another cancer type with very close incidence and mortality rates. According to another study at WHO, 70% of all cancer deaths occur in

Estimated number of incident cases and deaths worldwide, both sexes, all ages
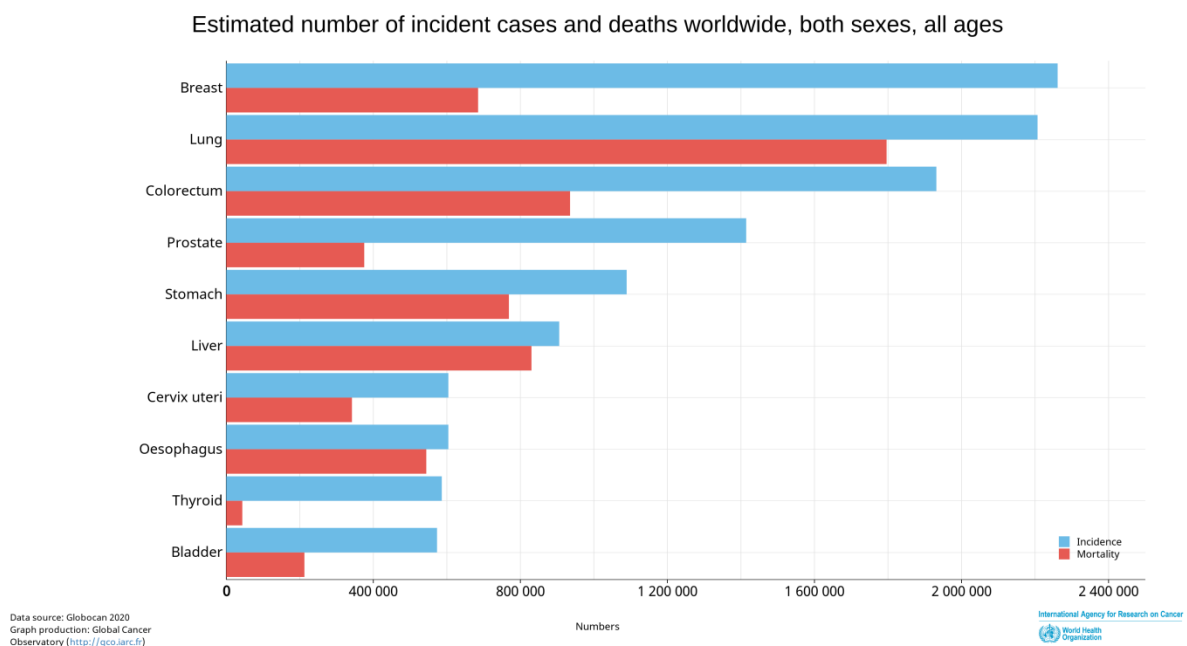
low- and middle-income countries [2].

Figure Ⅰ. 1 Estimated number of incident cases and deaths worldwide in 2020, both sexes, all ages, adapted from [3]

Normal cells divide only when receiving a chemical signal controlled by the nucleus. The DNA within the nucleus contains the genetic information to trigger division into two identical daughter cells through a process called mitosis [4]. If the DNA gets damaged, it either repairs itself or sends instructions to program cell death [5]. The DNA aberration that controls cell proliferation leads to abnormal cellular division even without receiving the appropriate signals. Consequently, cancer cells continue to grow randomly and multiply

indefinitely, resulting in the accumulation of cells that form a mass, called a tumor [6]. Some tumors have poor invasive potential and do not spread into other tissues, known as benign tumors, and usually do not cause serious health risks [7]. In contrast, malignant tumors can invade adjacent organs, intravasate into the nearest blood vessel, and subsequently spread to other parts of the body where they can form new secondary tumors, this process is known as metastasis, and it can be life threatening [8]. In fact, about 90 % of cancer deaths are attributed to metastasis, revealing the failure to manage the disease once it diffuses in the body [9].

All cancers are assumed to be caused by changes in DNA sequences that alter critical genes and change the behavior of the affected cells. An alteration in a gene is called a mutation. DNA mutations can occur during the replication process itself called spontaneous mutations [10]. A mutation may also result from environmental factors, representing 90%-95% of all mutations, as displayed in (Figure Ⅰ.2). These factors include poor diet, tobacco smoke, infections, obesity, alcohol, and others involving UV radiation, pollution, and certain chemicals. Much more rarely, people can inherit genetic mutations that make them more susceptible to developing cancer.
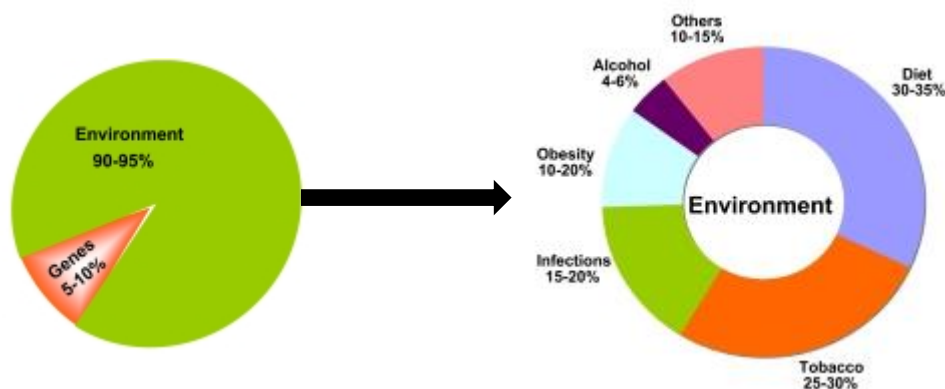
Figure Ⅰ.2 The percentage of environmental and genetic factors contributing to cancer, modified from [11]

Therefore, it is of vital importance to find new ways to detect cancers earlier, while they are still curable and before it has the chance to spread. Various investigations are used to diagnose cancer including blood tests [12], imaging tests such as X-ray [13], MRI [14], CT [15], ultrasound [16], and other genetic tests that screen for mutation in DNA [17]. The choice of the test depends on the type of cancer that is suspected based on the present symptoms. If

something suspicious appears on the scan, a biopsy test is needed, where a tissue sample is examined microscopically by a pathologist [18]. These tests help the oncologist to determine the type, size, and stage of cancer and also decide about the most appropriate treatment for patients.

There are three main therapies for cancer:

- Surgery: where the solid tumor is physically removed from the affected part of the body. More often, chemotherapy is used as a neoadjuvant to shrink the tumor before surgery.

- Chemotherapy: systemic treatment that uses chemotherapeutic medications to destroy and kill cancer cells. It may be used alone or in combination with other treatments to improve clinical outcomes.

- Radiotherapy: local treatment that aims at a certain area in the body by using targeted radiation (high-energy, X-rays, or electrons) to kill cancer cells.

Other categories of cancer therapy include hormone therapy, immunotherapy, gene therapy, antibodies therapy, and targeted therapy. Cancer treatment aims to cure the disease by removing the tumor cells completely or improving symptom control and thereby extending life and improving quality of life in the most serious cases [19]. Despite the remarkable advances in cancer therapy, it is often difficult to know whether cancer cells have been totally removed, or some remain in the body, and subsequently, cancer might return, months or years later, termed as recurrence [20]. Therefore, further research efforts are urgently needed to find new and selective drugs that hold great promise for monitoring cancer progression and improving treatment efficacy.

I.2 Healthy cells vs. cancer cells

Normal cells are programmed to die and be replaced by new ones. This requires signals delivered by ligands to the cell surface receptors. Physiologically, when a ligand (such as growth factor) specifically binds to the extracellular domain of the receptor and initiates a conformational change in the receptor, a signal is transmitted to the intracellular domain that in turn activates its associated receptor protein kinases. The activated kinases then add phosphate groups to serine, threonine, or tyrosine residues on other proteins and generate a cascade of signals to the nucleus that tells the cell to grow and divide or stop growing [21], as displayed in (Figure Ⅰ . 3).
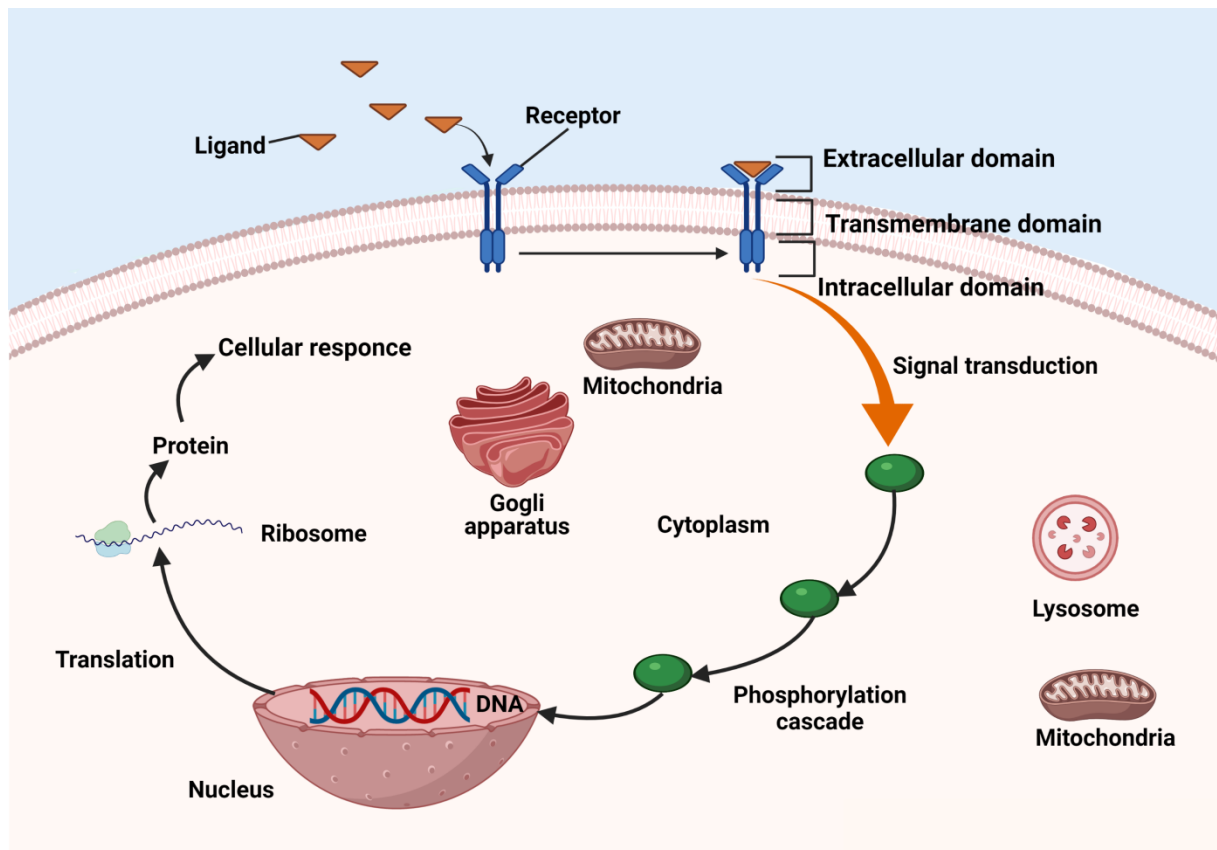
Figure Ⅰ.3 Intracellular view of receptor-ligand pharmacology, created with
BioRender.com

However, the mutation in genes disrupts this well-defined regulation. Mutations include two types of genes that regulate cell growth and division. Firstly, the proto-oncogenes give commands for the cell to grow and divide. In tumor cells, these genes are often mutated, known as oncogenes. Secondly, the tumor suppressor genes that stop the cellular growth, these genes are more involved in the development of cancer than oncogenes. An alteration in these two genes allows a cell to grow and divide indefinitely, and it may also stimulate cancerous cells to produce their own signals independently [22, 23].

I.3 Cell cycle phases

The cell cycle is a cascade of four coordinated stages, during which a cell duplicates all of its genetic material and divides into two daughter cells. The frequency of cell division depends on cell types, some cells divide continuously such as blood cells and skin cells, while bone and liver cells only divide under certain conditions (in response to injury). However,

nerve cells, as well as muscle, heart, and lens cells, do not divide at all in mature humans. Cell division is tightly regulated by small signaling molecules in the cytoplasm such as cyclins and protein kinases to ensure the production of two genetically identical cells. In eukaryotic cells, the cell cycle is usually divided into two main phases. The phase in which the cell divides is called the M phase, or mitosis. While the interphase is the period spent not dividing and cells remain 95% of their time in this phase, growing, duplicating DNA, and preparing for cell division. The interphase consists of three subphases, G1, S, and G2 [24, 25].

- G1 phase: or Gap 1 phase, involves cell growth, production of organelles and RNA, and preparation for S phase. After mitosis, new daughter cells enter G1 again to either undergo another round of cell division or leave the cell cycle, and enter the resting state, referred to as G0.
- S phase: the cell synthesizes two identical copies of DNA.
- G2 phase: Cell growth continues, mitochondrial density increases and proteins and RNA are synthesized in preparation for mitosis.
- M phase: involves nuclear division (mitosis) followed by cytoplasmic division (cytokinesis) that results in two daughter cells, and the cell cycle can start again (Figure Ⅰ. 4).
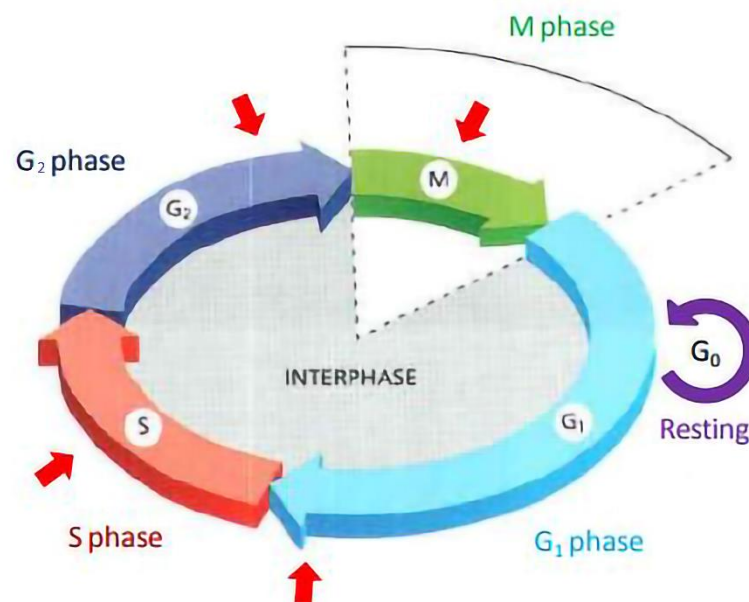
Figure Ⅰ. 4 Phases of cell cycle with checkpoints marked as red arrows, adapted from [26]

Cells use checkpoint systems at specific points on the cell cycle to ensure proper propagation of the cell division process and prevent the accumulation of genetic aberrations.

The first checkpoint during the G1 phase is called the restriction point that determines whether the cell is ready to divide and move forward through the cell cycle, remain in G1 phase, or enter the G0 phase. The next checkpoint occurs during the S phase to ensure that DNA is properly replicated. Other checkpoints happen at the end of G2 phase to confirm that cells won't proceed to mitosis until damaged DNA is repaired and during M phase (spindle checkpoint) to assure that the chromosomes are properly attached to the spindle fibers in order to achieve proper segregation. If DNA damage is detected at any stage of the cell cycle, checkpoints delay the cycle progression until defect repair is completed, otherwise, the cell is forced to undergo programmed cell death [27].

I.4 Cell death mechanisms

All cells in a multi-cellular organism undergo growth and death. When a mutation occurs such that cell activity threatens the survival of the entire organism, these defective cells must be eliminated through different molecular mechanisms [28]. There are two major forms of cell death, apoptosis, and necrosis. Apoptosis or programmed cell death is an important cellular process by which damaged cells are eliminated. In an average human adult, 50 to 70 billion cells undergo apoptosis per day [29]. In apoptosis, the cell is broken down from within through the production of proteins called caspases. Apoptosis could be triggered by two distinct pathways. The first is the extrinsic pathway, where the initial signal comes from outside the cell, such as some cytokines released by other cells and degradation of cytoplasmic material into apoptotic bodies. Then, these apoptotic bodies are completely engulfed by phagocytes without activating the immune response. The second is the intrinsic pathway (mitochondrial pathway), initiated by signals from within the cell, such as DNA damage and cytotoxic drugs [30]. Loss of apoptosis in cancer cells leads to uncontrolled cell proliferation and growth, resulting in malignant tumors. Indeed, apoptosis is an important hallmark of cancer that allows us to design targeted therapies based on better regulation of apoptosis. Alternately, necrosis is a process of cellular death that occurs when cells are exposed to extreme conditions such as toxins, infection, radiation, and chemical agents. During necrosis, this process leads to cell swelling and disintegration of plasma membrane inducing the release of cellular contents into extracellular space, and damage reaching the surrounding tissues that eventually causes an inflammatory response [31, 32]. TableⅠ. 1 summarizes the main differences between apoptosis and necrosis.

Table Ⅰ.1 Differences between apoptosis and necrosis

Apoptotic	Necrosis
• Programmed process	• Unplanned process
• Physiological or pathological	• Exclusively pathological
• Cellular contents are shrinked and safely removed by phagocytes.	• Uncontrolled release of cellular contents, damaging surrounding tissues.
• No inflammation	• Inflammation occurs
• Active (ATP dependent)	• Passive (ATP independent)
• Both external and internal factors	• External factors only
• single cell is affected	• Many cells are affected

I.5 Classification of chemotherapeutic drugs

Cytotoxicity is defined as the ability of certain chemotherapeutics to alter the basic cellular functions that ultimately damage and kill living cells [33]. Chemotherapy is a systemic treatment that distributes via the bloodstream to the whole body affecting all cells that grow and divide at a fast rate. These drugs do not distinguish between normal cells and cancer cells and attack other frequently dividing cells like skin cells, follicle cells, and gastrointestinal tract cells resulting in several symptoms such as alopecia (hair loss), nausea, painful ulceration, vomiting, and weight loss. The most serious side effect is bone marrow suppression which decreases the production of blood cells that leads to anemia, fatigue, and dizziness. Usually, this life-threatening effect is the dose-limiting factor in chemotherapy. These drugs are mainly administered intravenously or orally in combination with other drugs that act at different phases of the cell cycle and exhibit different toxicity profiles to avoid overlapping side effects and thus increase the tolerability of treatments. The combinations are often cytotoxic to various mutations within cancerous cells and consequently overcome drug resistance [34-37]. Chemotherapeutic drugs can be classified by their origin and type of damages, as summarized in (Table Ⅰ.2).

Table Ⅰ.2 Classification of chemotherapeutic drugs according to tissue damage after extravasation, adapted from [38]

Type of damages	Principal categories	Drugs
Vesicants	**DNA-binding compounds**	
	Anthracyclines	Amrubicin, Daunorubicin, Doxorubicin, Epirubicin, Idarubicin, Mitoxantrone, Pirarubicin
	Alkylating agents	Bendamustine, Busulfan, Carmustine, Melphalan, Nimustine, Ranimustine, Streptozocin
	Antitumor antibiotic	Actinomycin D
	Other	Trabectedin
	Non-DNA-binding compounds	
	Taxanes	Docetaxel, Paclitaxel, Nab-paclitaxel
	Vinka alkaloids	Vinblastine, Vincristine, Vindesine, Vinorelbine
	Others	
	Antibody-drug conjugate	Gemtuzumab ozogamicin
	Antitumor antibiotic	Mitomycin C
Irritants	Anthracyclines	Aclarubicin, Liposomal doxorubicin
	Alkylating agents	Dacarbazine, Cyclophosphamide, Ifosfamide, Temozolomide
	Antibody-drug conjugate	Trastuzumab emtansine
	Antimetabolites	Azacitidine, Gemcitabine, Fluorouracil, Tegafur
	Antitumor antibiotics	Bleomycin, Peplomycin
	Platinum salts	Carboplatin, Cisplatin Oxaliplatin, Nedaplatin, Miriplatin
	Proteasome inhibitor	Bortezomib
	Taxane	Cabazitaxel
	Topoisomerase I inhibitors	Irinotecan, Topotecan
	Topoisomerase II inhibitor	Etoposide
	Others	Arsenic Trioxide, Nelarabine, Picibanil, Porfimer

Nonvesicants/ Nonirritants	Antimetabolites	Cladribine, Clofarabine, Cytarabine, Enocitabine, Fludarabine, Methotrexate, Pemetrexed
	Antibody-drug conjugates	Brentuximab vedotin, Ibritumomab tiuxctan
	Monoclonal antibodies	Alemtuzumab, Bevacizumab, Cetuximab, Ipilimumab, Mogamulizumab, Ofatumumab, Panitumumab, Pertuzumab, Ramucirumab, Rituximab, Trastuzumab
	Others	BCG, Calcium folinate, Celmoleukin, Dexrazoxane, Eribulin, Interferon, L-asparaginase, Levofolinate, Octreotide, Pentostatin, Talaporfin sodium, Teceleukin

I.6 Mechanism of action of anticancer drugs

Anticancer drugs affect malignant cells in three main ways: blocking receptors, inhibiting signaling compounds and proteins synthesis, or interfering with DNA synthesis. In particular, the main goal of cytotoxic chemotherapies is to selectively kill as many cancer cells as possible while sparing healthy cells that in turn minimize the severity of side effects. In order to accomplish this, a more profound understanding of cancer cell biology and the molecular mechanisms of conventional anticancer drugs at the cellular level will significantly improve patient survival and quality of life. Cytotoxic anticancer drugs can also be classified according to their mechanism of action (MOA) that involves the biochemical interactions of drugs with DNA. These drugs exert their action in different ways, such as inhibition of purine, pyrimidine, or nucleotides synthesis, inhibition of DNA, RNA, or proteins synthesis, and block the function of microtubules and subsequently inhibit cellular mitosis, as depicted in (FigureⅠ.5).
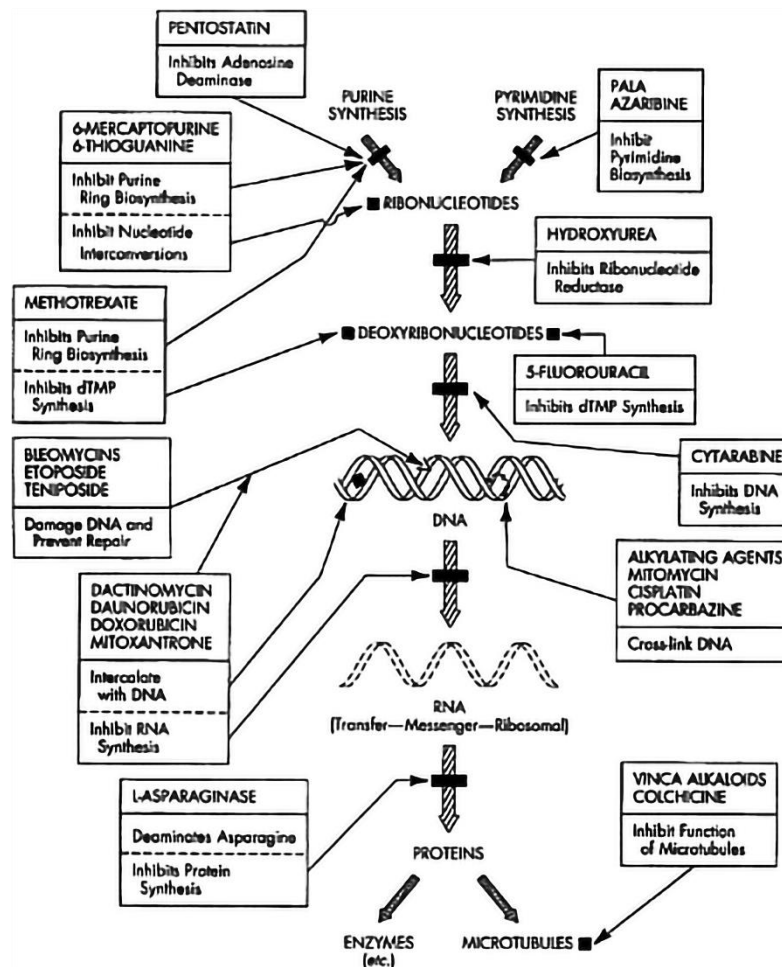
Figure Ⅰ.5 Mechanisms and sites of action of anticancer drugs, extended from [39]

Chemotherapy resistance is a persistent challenge in oncology that mainly occurs when quiescent cancer cells evade the effects of conventional cytotoxic drugs. Accordingly, shuya et al, used intravital FUCCI imaging that color-codes the phases of the cell cycle in real-time and thereby demonstrated that solid tumors contain a large number of quiescent cells (G0/G1) in the center area of tumors and a minor proportion of proliferative cells were located at the surface area nearby blood vessels that are deficient in oxygen and nutrition which is required to restart the cell cycle. Therefore, these cells can survive from chemotherapy that only targets dividing cells [40]. To address these limitations, most research focuses on developing cell cycle specific drugs to improve anticancer therapy of solid tumors. Chemotherapeutic drugs can be generalized into two categories, cell cycle phase specific drugs that act on a specific phase of the cell cycle and cell cycle phase non-specific drugs that affect all phases of the cell cycle as illustrated in (Figure Ⅰ.6).
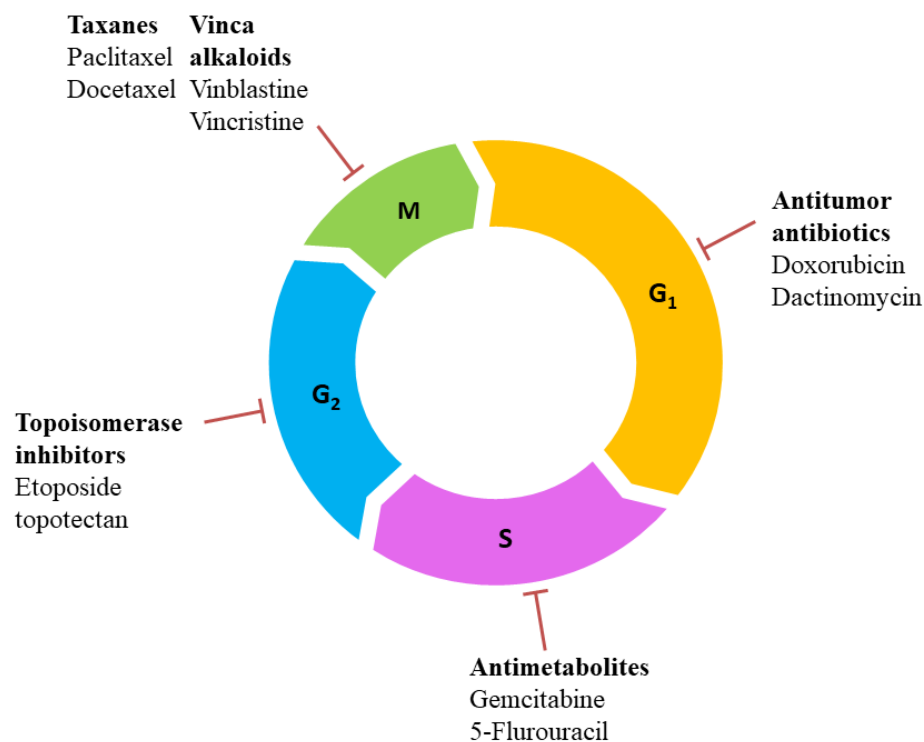
Figure Ⅰ. 6 Diagram of cell cycle specific anticancer drugs

I.7 Drug discovery and development

Drug discovery and development is a highly regulated process designed to ensure that innovative new medications are effective, safe, and have the required pharmacokinetic properties such as appropriate absorption and metabolism by the human body. Discovering a new drug and bringing it to the market typically takes an average of 14 years of research and clinical development and costs around 1-2 billion USD. Out of 10.000 or more hits tested in early drug discovery, only one may ultimately lead to a drug that reaches the market [41, 42].

Drug discovery involves target discovery, hit generation (chemical entities that have a promising affinity with the target), lead identification and optimization to increase their selectivity and efficacy before they move forward drug development stage. Subsequently, the candidate is evaluated for safety, and pharmacology and toxicology studies are diligently performed to establish the maximum safe doses in animals prior to clinical trials. The latter is clinical development wherein the safety and efficacy of drug candidates are assessed in humans and thus obtains approval from the United States Food and Drug Administration (FDA) to reach the consumer market and the safety of drugs will be continuously monitored during the years of its distribution (Figure Ⅰ. 7).
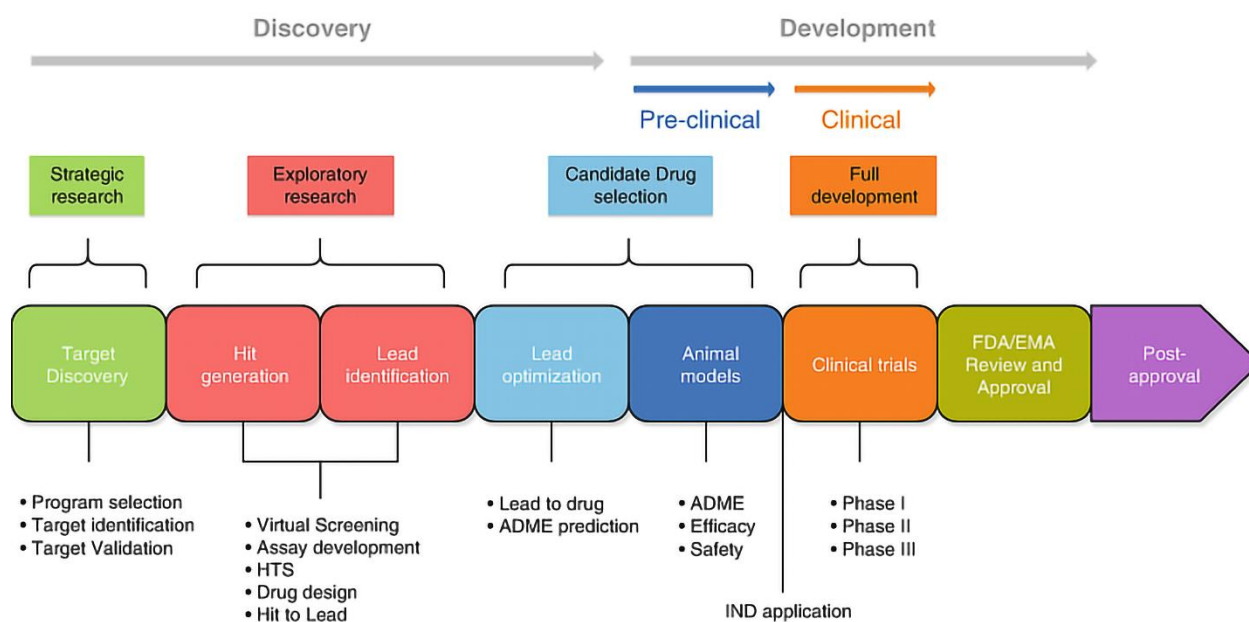
Figure Ⅰ. 7 Drug discovery and development pipeline, adapted from [43]

I.7.1   Preclinical trials

A research stage that includes all studies conducted on the drug in the laboratory to safely precede the drug into clinical trials. Among them, *in vitro*, *in vivo,* and *in silico* models that determine the biological efficacy on living organisms, toxicity profiles, identify the lead candidate from multiple hits and estimate the appropriate starting dose for human trials [44, 45]. In the next section, we will further discuss *in vitro* and *in vivo* methods and how these studies are conducted.

I.7.1.1   *In vitro* assays

In Vitro is the Latin word for in-glass, refers to any procedure or experiment performed outside the living organisms. These rapid and cost-effective studies are critical for preliminary research, which is conducted on cells or microorganisms within test tubes or Petri dishes to mimic and predict biological responses of tissue of interest prior to actual dosing in animals. However, one of the main concerns of *in vitro* testing remains its predictive ability that may not represent physiological conditions in a robust and reproducible manner. Therefore, the use of these models must be rigorously evaluated for each application, and standardization of cultivation conditions, sample collection, and protocols is essential to accurately portray *in vivo* biology. Particularly, *in vitro* studies used in cancer drug discovery are cell-based assays that determine the effects of compounds evaluated on cell proliferation and their cytotoxic

effects that ultimately lead to cell death. These assays require incubation of colorimetric tetrazolium reagents with a population of living cells to generate a fluorescent signal that can be detected using a plate reader to estimate the number of viable cells remaining at the end of the experiment. The most commonly used compounds include MTT, MTS, XTT, and WST-1. IC_{50} is the most widely used metric of cytotoxicity in cancer research, determined by *in vitro* assays representing the concentration of substance required to inhibit 50% of a biological process, which is pivotal for analyzing the efficacy of new drugs [46-48].

- *Cancer cell lines*

Cancer cell lines are routinely used by researchers as simple cellular models for *in vitro* assays as they retain most of the cellular and molecular alterations of cancer from which they originated (Table Ⅰ.3). Each cell line is derived from a different human tumor and immortalized through the use of viral oncogenes capable of evading the normal constraints, allowing the cell cycle to proceed and cells continue to proliferate indefinitely, resulting in a consistent sample ready for scientific research at any time [49]. Cell culture is the process of growing cells in controlled conditions outside their native environment. To keep the cells alive, it is important to keep the conditions as close to the physiological conditions as possible, therefore appropriate oxygen level is maintained and supplements such as amino acids, inorganic salts, vitamins, lipids, and growth factors are added to the culture media. In Addition, the culture media contains a buffering system to maintain the proper pH such as sodium bicarbonate to keep the pH of the media between 7.2 and 7.4 with 5% gaseous carbon dioxide. Moreover, cells are kept in a sterile environment to avoid contamination, incubated at 37° C for optimal growth, and a pH indicator (phenol red) is routinely added to the media to visually monitor changes in pH. For long-term storage, these cell lines are often stored in liquid nitrogen to avoid genomic instability [50, 51].

Table Ⅰ.3 Summary of the most prominently studied human cancer cell lines, extended from [52]

Cell line	Organ of Origin	Disease
MCF7	Breast	Adenocarcinoma
MDA-MB-231	Breast	Adenocarcinoma
T-47D	Breast	Ductal Carcinoma
SNB-75	CNS	Astrocytoma
U251	CNS	Glioblastoma
HCT-116	Colon	Carcinoma

HCT-15	Colon	Colorectal adenocarcinoma
HT29	Colon	Colorectal adenocarcinoma
SW620	Colon	Adenocarcinoma
HL-60	Leukemia	Acute promyelocytic leukemia
K562	Leukemia	Chronic myelogenous leukemia
A549	Lung	Adenocarcinoma
DU145	Prostate	Carcinoma
PC-3	Prostate	Grade IV, adenocarcinoma

I.7.1.2   *In vivo* assays

In this phase, drug candidates must undergo extensive animal studies to further evaluate their toxicity, distribution, metabolism, and overall efficacy to safely proceed the drug into human testing. *In vivo* assays describe any procedure or test performed within a living organism such as an animal, plant, or human being. *In vivo* studies include a wide variety of laboratory animals such as rodents (rats, mice, and hamsters), rabbits, dogs, cats, monkeys, zebrafish [53-57]. Mice are the most commonly used due to their low cost, small size, ease of maintenance, and similarity to humans genetically (85% of genomes are identical) [58].

I.7.2   Clinical trials

Once a drug candidate has demonstrated sufficient efficacy and safety during cell and animal studies, human studies may begin which involve people who volunteer to test a new drug, find the right dose, evaluate its safety, and look for side effects. Clinical trials are conducted to find out if new drugs are more effective than standard treatments and/or have less severe side effects than currently available treatments. Clinical trials involving new drugs are usually divided into four phases.

I.7.2.1   Stage Ⅰ

At this stage, the safety of the drug, tolerability, pharmacokinetics, and pharmacodynamics effects are determined. This phase is conducted on a small number of participants (20-80), during which escalating single doses of the drug are administered to accurately calculate the optimal safe dose. This stage may last for several months.

I.7.2.2   Stage Ⅱ

About 70% of the drugs tested in phase I move on to phase II studies. Phase II studies are designed to determine efficacy and further evaluate the drug's safety in a larger group of

volunteers (100-300). This stage lasts from several months to two years.

I.7.2.3 Stage Ⅲ

If a drug or treatment is found to be safe in phase I and effective in phase II, it will then enter a phase III trial. This stage is performed on a large-scale population (1000-3000) to confirm drug efficacy for a wide variety of people, check for side effects, and compare it with conventional treatments. It occurs over a longer period of time (1-4 years). About 30% of tested drugs complete phase III successfully and eventually the FDA consider approving the drug for marketing.

I.7.2.4 Stage Ⅳ

Once the drug is marketed, it continues to be closely scrutinized for adverse effects as negative effects may not appear for a long time. Phase Ⅳ is often referred to as a post-marketing surveillance trial that allows drug companies to gather additional information about risks, benefits, optimal use, and side effects that might occur over time. This phase continues for as long as the drug remains on the market [59].

I.8 Serine/threonine protein kinase

Protein kinases are enzymes that transfer a phosphate group onto a target protein, thus changing protein function from inactive to active. Protein kinases usually use adenosine triphosphate (ATP), as a source of phosphate. ATP is a molecule composed of adenosine and three phosphate groups (called alpha, beta, and gamma phosphates). Kinases catalyze the transfer of gamma phosphate from ATP to the hydroxyl function in the side chain of particular amino acids. It is called phosphorylation, a reversible process mediated by protein kinases and phosphoprotein phosphatases, in which proteins transmit chemical signals to one another (Figure Ⅰ.8). Out of the twenty different amino acids in the body, only three can get phosphorylated; serine, threonine, and tyrosine. When a protein gets phosphorylated, the phosphate at serine, threonine, or tyrosine residue acts as a docking site for other proteins to interact with resulting in a conformational change whereby the signal can be passed on all the way through the cell. Ultimately, this can lead to a multitude of cellular processes, including proliferation, apoptosis, metabolism, and differentiation. At the same time, faulty hyperactivated kinases can transmit too many phosphorylation signals, and this can lead to cancer [60].
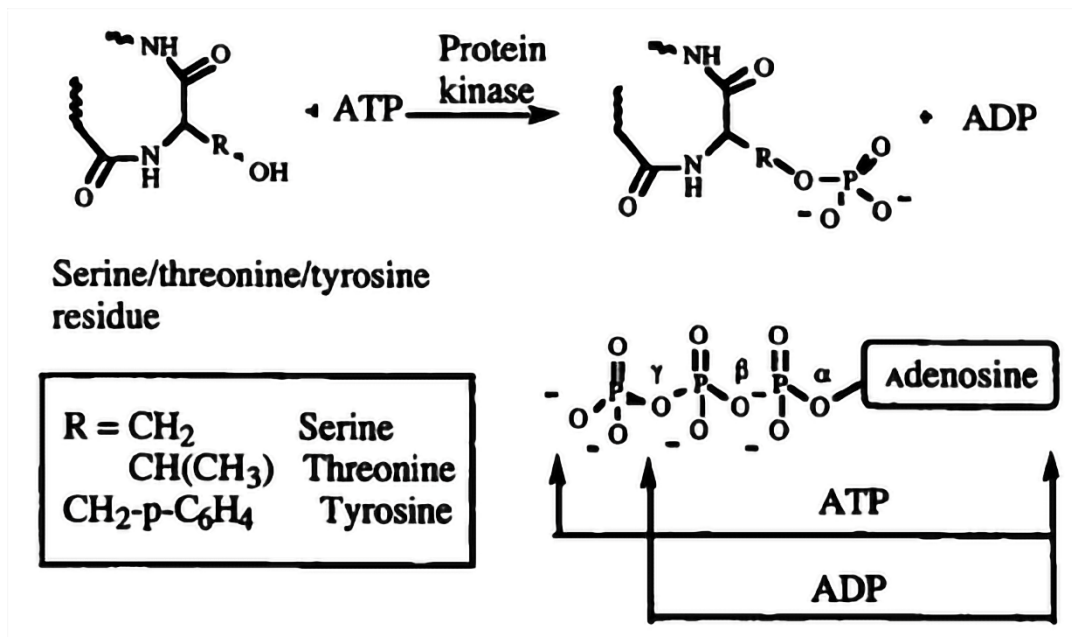
Figure Ⅰ. 8 The mechanism of phosphorylation catalysed by protein kinase, adapted from

[61]

Kinases are broadly classified according to the amino acid residues they phosphorylate. However, kinases capable of transferring phosphate groups to serine residues are also able to do so for threonine residues. Therefore, the kinases are often grouped into two families; protein tyrosine kinases and protein serine/threonine kinases. The human genome contains about 500 protein kinase genes, constituting about 2% of all human genes and at least 125 of which are serine/threonine kinases [62, 63].

I.8.1   Checkpoint kinase 1

CHK1 is a promising molecular target that gained immense attention recently for the development of cancer therapeutics. The checkpoint kinase 1, commonly known as CHK1 or CHEK1, is a member of the serine/threonine kinase family of enzymes that mediate DNA damage response (DDR) and cell cycle progression (Figure Ⅰ. 9). It is well established that CHK1 overexpressed in numerous tumors, including breast, colon, liver, gastric and nasopharyngeal carcinoma [64, 65]. Moreover, many cancers rely on CHK1 mediated cell cycle arrest, especially if they are deficient in p53 [66]. Therefore, about 50% of all human cancers contain p53 mutations indicating that many tumors are dependent on CHK1 signaling pathway [67].
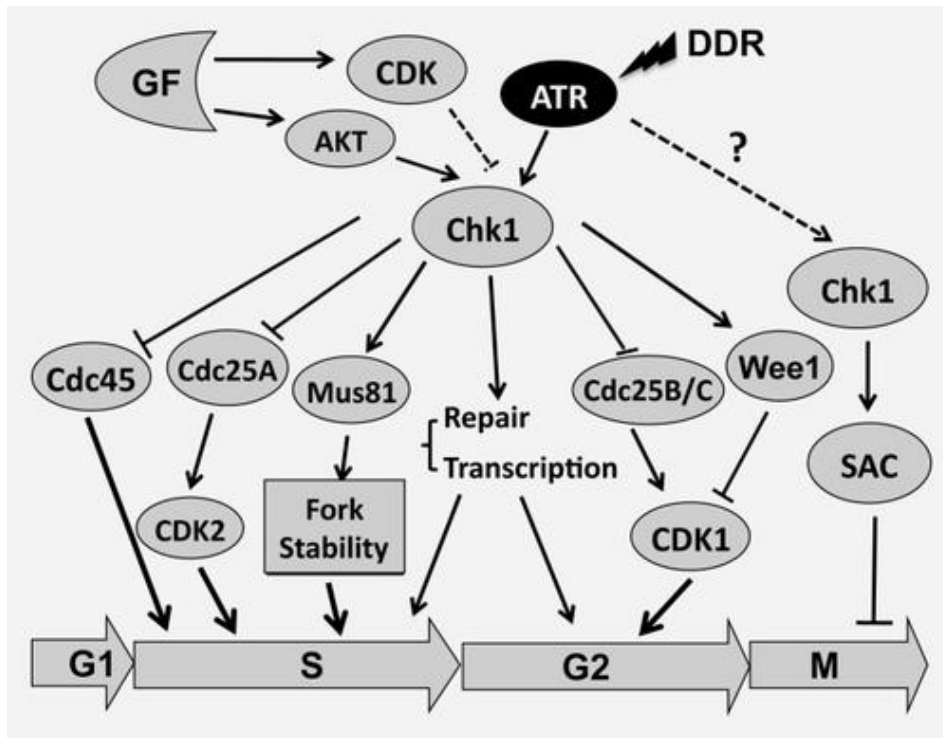
Figure Ⅰ.9 The roles of CHK1 in the DNA damage response during the S, G2, and M phases of the cell cycle, dashed lines represent as-yet confirmed signalling, adapted from [65]

I.8.2 CHK1 inhibition in cancer therapy

CHK1 inhibition is an effective targeting strategy in cancer therapy that inhibits cancer cells proliferation, increases genomic instability, stops cell cycle arrest, and may also reverse tumor cell resistance to chemotherapies. Hence, cancer cells lose their ability to repair damaged DNA upon CHK1 inhibition, which improves the effectiveness of chemotherapy and/or radiotherapy, thereby promoting targeted cell death and providing synthetic lethality. Numerous studies, both *in vivo* and *in vitro,* attempt to identify and develop specific CHK1 inhibitors that would synergize with chemotherapy [68]. One of the most studied agents in this class is prexasertib, a clinical candidate inhibitor that selectively binds to CHK1, blocking DNA repair leading to the accumulation of damaged DNA that eventually triggers apoptosis in cancer cells. Indeed, prexasertib showed promising results for phase I and II trials as a monotherapy in patients with advanced solid tumors (squamous cell carcinomas) and in combination with cytotoxic DNA damaging agents, including Cisplatin [69], the EGFR approved inhibitor Cetuximab (Erbitux) [70] or with the poly ADP-ribose polymerase (PARP) inhibitors such as Olaparib [71, 72].

I.9 Xanthones

Xanthones represent an interesting class of phenolic compounds with oxygenated tricyclic scaffold (9H-xanthen-9-one) containing eight substitution positions that could lead to a broad range of chemical entities with numerous pharmacological activities [73]. These compounds are widely distributed in nature and can be isolated from the pericarp of (Garcinia mangostana L.) fruit, including α-mangostin, β-mangostin, γ-mangostin, neosmitilbin, mangostanol, desoxymorellin, garcinone B, and gambogic acid [74]. It was reported that xanthones possess significant biological activities such as antioxidant [75], anti-inflammatory [76], antidiabetic [77], antiviral [78], antimicrobial [79], antifungal [80], anti-Alzheimer [81], and anti-allergy [82] activities, as presented in (FigureⅠ.10).
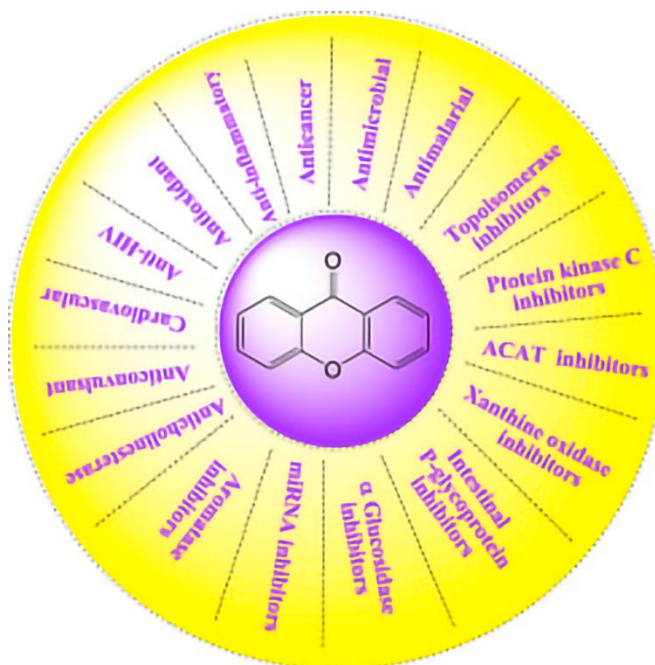
FigureⅠ.10 Biological activities and molecular targets of xanthones, extended from [83]

Furthermore, multiple *in vitro* and *in vivo* studies demonstrated that xanthones could bind to numerous target proteins, including topoisomerases, cyclooxygenases, protein kinases, DNA polymerases, and many others resulting in the inhibition of several human cancer cells [84]. The mechanisms of action of xanthones commonly involve cell cycle arrest, induction of apoptosis in malignant cells, inhibition of proliferation, prevention of invasion and metastasis of tumor cells, interruption of angiogenesis, etc. Therefore, Kritsanawong *et al* [85], reported that α-mangostin inhibits cell proliferation and induces apoptosis in HER2/PI3K/Akt and

MAPK signaling pathways in human breast carcinoma cells (T47D). Some other studies demonstrate that α-mangostin is involved in downstream of vascular endothelial growth factor (VEGF) [86], STAT3 signaling [87], cyclin-dependent kinases (CDKs) [88], as well as promoting cellular apoptosis through inhibition of the ERK1/2 pathway [89].

In a study performed by Lili *et al* [90], gambogic acid inhibited the proliferation by inducing G1 phase cell-cycle arrest of hepatocellular carcinoma (HCC) cell lines *in vitro* and *in vivo* while sparing normal hepatocytes. It is also involved in the downregulation of cyclin-CDK complex activity and GSK3β/β-catenin pathway as well as suppression of Hep3B tumor growth in a dose-dependent manner *in vivo*. In 2016, Liu *et al* [91], synthesized a series of twenty-one xanthone derivatives with microwave-assisted technique and revealed that the compound XD-1 stimulated HepG2 cell apoptosis through the caspase-dependent mitochondrial pathway and cell cycle block by regulating G0/G1 checkpoint proteins. Rong *et al* [92], demonstrated the ability of xanthone derivatives to modulate p53/p21 through the ATR/CHK1 signaling pathway in HepG2 and A549 cancer cell lines.

I.10 References

1. Service RF (2005) Nanotechnology Takes Aim at Cancer. Science 310:1132–1134. https://doi.org/10.1126/science.310.5751.1132

2. Dalton M, Holzman E, Erwin E, et al (2019) Patient navigation services for cancer care in low-and middle-income countries: A scoping review. PLoS ONE 14:e0223537. https://doi.org/10.1371/journal.pone.0223537

3. https://gco.iarc.fr/today/online-analysis-multi-bars. Accessed 3 Jan 2022

4. McIntosh JR, Molodtsov MI, Ataullakhanov FI (2012) Biophysics of mitosis. Quart Rev Biophys 45:147–207. https://doi.org/10.1017/S0033583512000017

5. Gorbunova V, Seluanov A, Mao Z, Hine C (2007) Changes in DNA repair during aging. Nucleic Acids Research 35:7466–7474. https://doi.org/10.1093/nar/gkm756

6. Thurston DE (2007) Chemistry and pharmacology of anticancer drugs. CRC Press/Taylor & Francis, Boca Raton

7. Samah AA, Fauzi MFA, Mansor S (2017) Classification of benign and malignant tumors in histopathology images. In: 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA). IEEE, Kuching, pp 102–106

8. Duffy M, McGowan P, Gallagher W (2008) Cancer invasion and metastasis: changing views. J Pathol 214:283–293. https://doi.org/10.1002/path.2282

9. Hanahan D, Weinberg RA (2000) The Hallmarks of Cancer. Cell 100:57–70. https://doi.org/10.1016/S0092-8674(00)81683-9

10. Maki H (2002) Origins of Spontaneous Mutations: Specificity and Directionality of Base-Substitution, Frameshift, and Sequence-Substitution Mutageneses. Annu Rev Genet 36:279–303. https://doi.org/10.1146/annurev.genet.36.042602.094806

11. Anand P, Kunnumakara AB, Sundaram C, et al (2008) Cancer is a Preventable Disease that Requires Major Lifestyle Changes. Pharm Res 25:2097–2116. https://doi.org/10.1007/s11095-008-9661-9

12. Hanash SM, Baik CS, Kallioniemi O (2011) Emerging molecular biomarkers—blood-based strategies to detect and monitor cancer. Nat Rev Clin Oncol 8:142–150. https://doi.org/10.1038/nrclinonc.2010.220

13. Lu L, Sun M, Lu Q, et al (2021) High energy X-ray radiation sensitive scintillating materials for medical imaging, cancer diagnosis and therapy. Nano Energy 79:105437. https://doi.org/10.1016/j.nanoen.2020.105437

14. Morrow M, Waters J, Morris E (2011) MRI for breast cancer screening, diagnosis, and treatment. The Lancet 378:1804–1811. https://doi.org/10.1016/S0140-6736(11)61350-0

15. Hasegawa S, Yoshikawa T, Shirai J, et al (2013) A Prospective Validation Study to Diagnose Serosal Invasion and Nodal Metastases of Gastric Cancer by Multidetector-row CT. Ann Surg Oncol 20:2016–2022. https://doi.org/10.1245/s10434-012-2817-1

16. Meys EMJ, Kaijser J, Kruitwagen RFPM, et al (2016) Subjective assessment versus ultrasound models to diagnose ovarian cancer: A systematic review and meta-analysis. European Journal of Cancer 58:17–29. https://doi.org/10.1016/j.ejca.2016.01.007

17. Kaz AM, Brentnall TA (2006) Genetic testing for colon cancer. Nat Rev Gastroenterol Hepatol 3:670–679. https://doi.org/10.1038/ncpgasthep0663

18. Brown MV, McDunn JE, Gunst PR, et al (2012) Cancer detection and biopsy classification using concurrent histopathological and metabolomic analysis of core biopsies. Genome Med 4:33. https://doi.org/10.1186/gm332

19. Aigner KR, Stephens FO (2009) Basics of Oncology. Springer Berlin Heidelberg, Berlin, Heidelberg

20. Taylor C, Richardson A, Cowley S (2011) Surviving cancer treatment: An investigation of the experience of fear about, and monitoring for, recurrence in patients following treatment for colorectal cancer. European Journal of Oncology Nursing 15:243–249. https://doi.org/10.1016/j.ejon.2011.03.010

21. Tabernero J (2007) The Role of VEGF and EGFR Inhibition: Implications for Combining Anti–VEGF and Anti–EGFR Agents. Mol Cancer Res 5:203–220. https://doi.org/10.1158/1541-7786.MCR-06-0404

22. Kontomanolis EN, Koutras A, Syllaios A, et al (2020) Role of Oncogenes and Tumor-suppressor Genes in Carcinogenesis: A Review. Anticancer Res 40:6009–6015. https://doi.org/10.21873/anticanres.14622

23. Aktipis CA, Nesse RM (2013) Evolutionary foundations for cancer biology. Evol Appl 6:144–159. https://doi.org/10.1111/eva.12034

24. Cooper GM (2000) The cell: a molecular approach, 2. ed. ASM Press [u.a.], Washington, DC

25. Mills CC, Kolb Ea, Sampson VB (2018) Development of Chemotherapy with Cell-Cycle Inhibitors for Adult and Pediatric Cancer Therapy. Cancer Res 78:320–325. https://doi.org/10.1158/0008-5472.CAN-17-2782

26. Alberts B (2008) Molecular biology of the cell, 5th ed. Garland Science, New York

27. DeVita VT (2001) Cancer: principles & practice of oncology, 6. ed. Lippincott Williams & Wilkins, Philadelphia

28. Galluzzi L, Vitale I, Aaronson SA, et al (2018) Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. Cell Death Differ 25:486–541. https://doi.org/10.1038/s41418-017-0012-4

29. Reed JC (1999) Dysregulation of Apoptosis in Cancer. JCO 17:2941–2941. https://doi.org/10.1200/JCO.1999.17.9.2941

30. Reed JC (2000) Mechanisms of Apoptosis. The American Journal of Pathology 157:1415–1430. https://doi.org/10.1016/S0002-9440(10)64779-7

31. Golstein P, Kroemer G (2007) Cell death by necrosis: towards a molecular definition. Trends in Biochemical Sciences 32:37–43. https://doi.org/10.1016/j.tibs.2006.11.001

32. Syntichaki P, Tavernarakis N (2002) Death by necrosis: Uncontrollable catastrophe, or is there order behind the chaos? EMBO Rep 3:604–609. https://doi.org/10.1093/embo-reports/kvf138

33. Prager I, Watzl C (2019) Mechanisms of natural killer cell-mediated cellular cytotoxicity. J Leukoc Biol 105:1319–1329. https://doi.org/10.1002/JLB.MR0718-269R

34. Kerbel RS, Kamen BA (2004) The anti-angiogenic basis of metronomic chemotherapy. Nat Rev Cancer 4:423–436. https://doi.org/10.1038/nrc1369

35. Damen MPF, Rheenen J, Scheele CLGJ (2021) Targeting dormant tumor cells to prevent cancer recurrence. FEBS J 288:6286–6303. https://doi.org/10.1111/febs.15626

36. Feng L, Huang Q, Huang Z, et al (2016) Optimized Animal Model of Cyclophosphamide-induced Bone Marrow Suppression. Basic Clin Pharmacol Toxicol 119:428–435. https://doi.org/10.1111/bcpt.12600

37. McKnight JA (2003) Principles of chemotherapy. Clinical Techniques in Small Animal Practice 18:67–72. https://doi.org/10.1053/svms.2003.36617

38. Sugita Y, Takao K, Uesawa Y, et al (2020) Development of Newly Synthesized Chromone Derivatives with High Tumor Specificity against Human Oral Squamous Cell Carcinoma. Medicines 7:50. https://doi.org/10.3390/medicines7090050

39. Rauf S, Gooding JJ, Akhtar K, et al (2005) Electrochemical approach of anticancer drugs–DNA interaction. Journal of Pharmaceutical and Biomedical Analysis 37:205–217. https://doi.org/10.1016/j.jpba.2004.10.037

40. Yano S, Tazawa H, Kagawa S, et al (2020) FUCCI Real-Time Cell-Cycle Imaging as a Guide for Designing Improved Cancer Therapy: A Review of Innovative Strategies to Target Quiescent Chemo-Resistant Cancer Cells. Cancers 12:2655. https://doi.org/10.3390/cancers12092655

41. Abrusán G, Marsh JA (2019) Ligands and Receptors with Broad Binding Capabilities Have Common Structural Characteristics: An Antibiotic Design Perspective. J Med Chem 62:9357–9374. https://doi.org/10.1021/acs.jmedchem.9b00220

42. Gassmann O, Zedtwitz M von, Reepmeyer G (2004) Leading pharmaceutical innovation: trends and drivers for growth in the pharmaceutical industry

43. Matei F, Zirra D (2019) Introduction to Biotech Entrepreneurship: From Idea to Business: A European Perspective. Springer International Publishing, Cham

44. Steinmetz KL, Spack EG (2009) The basics of preclinical drug development for neurodegenerative disease indications. BMC Neurol 9:S2. https://doi.org/10.1186/1471-2377-9-S1-S2

45. Gad SC (2008) Preclinical development handbook. Wiley-Interscience, Hoboken, NJ

46. Markossian S, Grossman A, Brimacombe K, et al (2004) Assay Guidance Manual. Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda (MD)

47. Ediriweera MK, Tennekoon KH, Samarakoon SR (2019) In vitro assays and techniques utilized in anticancer drug discovery: In vitro assays in anticancer drug development. J Appl Toxicol 39:38–71. https://doi.org/10.1002/jat.3658

48. Aykul S, Martinez-Hackert E (2016) Determination of half-maximal inhibitory concentration using biosensor-based protein interaction analysis. Analytical Biochemistry 508:97–103. https://doi.org/10.1016/j.ab.2016.06.025

49. Linder S, Marshall H (1990) Immortalization of primary cells by DNA tumor viruses. Experimental Cell Research 191:1–7. https://doi.org/10.1016/0014-4827(90)90027-8

50. Michl J, Park KC, Swietach P (2019) Evidence-based guidelines for controlling pH in mammalian live-cell culture systems. Commun Biol 2:144. https://doi.org/10.1038/s42003-019-0393-7

51. Fazekas J, Grunt TW, Jensen-Jarolim E, Singer J (2017) Long term storage in liquid nitrogen leads to only minor phenotypic and gene expression changes in the mammary carcinoma model cell line BT474. Oncotarget 8:35076–35087. https://doi.org/10.18632/oncotarget.16623

52. Abaan OD, Polley EC, Davis SR, et al (2013) The Exomes of the NCI-60 Panel: A Genomic Resource for Cancer Biology and Systems Pharmacology. Cancer Res 73:4372–4382. https://doi.org/10.1158/0008-5472.CAN-12-3342

53. lto N, Tamano S, Shirai T (2003) A medium-term rat liver bioassay for rapid in vivo detection of carcinogenic potential of chemicals. Cancer Science 94:3–8. https://doi.org/10.1111/j.1349-7006.2003.tb01343.x

54. Yue EW, Douty B, Wayland B, et al (2009) Discovery of Potent Competitive Inhibitors of Indoleamine 2,3-Dioxygenase with in Vivo Pharmacodynamic Activity and Efficacy in a Mouse Melanoma Model. J Med Chem 52:7364–7367. https://doi.org/10.1021/jm900518f

55. Salzwedel AO, Han J, LaRocca CJ, et al (2018) Combination of interferon-expressing oncolytic adenovirus with chemotherapy and radiation is highly synergistic in hamster model of pancreatic cancer. Oncotarget 9:18041–18052. https://doi.org/10.18632/oncotarget.24710

56. Zhong L, Huang Y, He J, et al (2021) Generation of in situ CRISPR-mediated primary and metastatic cancer from monkey liver. Sig Transduct Target Ther 6:411. https://doi.org/10.1038/s41392-021-00799-7

57. Asokan N, Daetwyler S, Bernas SN, et al (2020) Long-term in vivo imaging reveals tumor-specific dissemination and captures host tumor interaction in zebrafish xenografts. Sci Rep 10:13254. https://doi.org/10.1038/s41598-020-69956-2

58. Coleman O, Henry M, O'Neill F, et al (2018) A Comparative Quantitative LC-MS/MS Profiling Analysis of Human Pancreatic Adenocarcinoma, Adjacent-Normal Tissue, and Patient-Derived Tumour Xenografts. Proteomes 6:45. https://doi.org/10.3390/proteomes6040045

59. Pecorino L (2021) Molecular biology of cancer 5e: mechanisms, targets, and therapeutics, 5th ed. AC Higher Education: ACHE, New York

60. Ardito F, Giuliani M, Perrone D, et al (2017) The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). International Journal of Molecular Medicine 40:271–280. https://doi.org/10.3892/ijmm.2017.3036

61. Tong L (2003) The synthesis of novel anticancer drugs. University of Glasgow

62. Jacob T, Van den Broeke C, Favoreel HW (2011) Viral Serine/Threonine Protein Kinases. Journal of Virology 85:1158–1173. https://doi.org/10.1128/JVI.01369-10

63. Capra M, Nuciforo PG, Confalonieri S, et al (2006) Frequent Alterations in the Expression of Serine/Threonine Kinases in Human Cancers. Cancer Res 66:8147–8154. https://doi.org/10.1158/0008-5472.CAN-05-3489

64. Schuler F, Weiss JG, Lindner SE, et al (2017) Checkpoint kinase 1 is essential for normal B cell development and lymphomagenesis. Nat Commun 8:1697. https://doi.org/10.1038/s41467-017-01850-4

65. Zhang Y, Hunter T (2014) Roles of Chk1 in cell biology and cancer therapy. Int J Cancer 134:1013–1023. https://doi.org/10.1002/ijc.28226

66. Chen Z, Xiao Z, Gu W, et al (2006) Selective Chk1 inhibitors differentially sensitize p53-deficient cancer cells to cancer therapeutics. Int J Cancer 119:2784–2794. https://doi.org/10.1002/ijc.22198

67. Tao Z-F, Lin N-H (2006) Chk1 Inhibitors for Novel Cancer Treatment. ACAMC 6:377–388. https://doi.org/10.2174/187152006777698132

68. Lim S, Kim Y, Lee S-B, et al (2020) Inhibition of Chk1 by miR-320c increases oxaliplatin responsiveness in triple-negative breast cancer. Oncogenesis 9:91. https://doi.org/10.1038/s41389-020-00275-x

69. Sen T, Tong P, Stewart CA, et al (2017) CHK1 Inhibition in Small-Cell Lung Cancer Produces Single-Agent Activity in Biomarker-Defined Disease Subsets and Combination Activity with Cisplatin or Olaparib. Cancer Res 77:3870–3884. https://doi.org/10.1158/0008-5472.CAN-16-3409

70. Zeng L, Beggs RR, Cooper TS, et al (2017) Combining Chk1/2 Inhibition with Cetuximab and Radiation Enhances *In Vitro* and *In Vivo* Cytotoxicity in Head and Neck Squamous Cell Carcinoma. Mol Cancer Ther 16:591–600. https://doi.org/10.1158/1535-7163.MCT-16-0352

71. Brill E, Yokoyama T, Nair J, et al (2017) Prexasertib, a cell cycle checkpoint kinases 1 and 2 inhibitor, increases *in vitro* toxicity of PARP inhibition by preventing Rad51 foci formation in *BRCA* wild type high-grade serous ovarian cancer. Oncotarget 8:111026–111040. https://doi.org/10.18632/oncotarget.22195

72. Jagtap P, Szabó C (2005) Poly(ADP-ribose) polymerase and the therapeutic effects of its inhibitors. Nat Rev Drug Discov 4:421–440. https://doi.org/10.1038/nrd1718

73. Nourisefat M, Salehi N, Yousefinejad S, et al (2019) Biological evaluation of 9-(1H-Indol-3-yl) xanthen-4-(9H)-ones derivatives as noncompetitive α-glucosidase inhibitors:

kinetics and molecular mechanisms. Struct Chem 30:703–714. https://doi.org/10.1007/s11224-018-1218-x

74. Kurniawan YS, Priyangga KTA, Jumina, et al (2021) An Update on the Anticancer Activity of Xanthone Derivatives: A Review. Pharmaceuticals 14:1144. https://doi.org/10.3390/ph14111144

75. Yang J, Liu RH, Halim L (2009) Antioxidant and antiproliferative activities of common edible nut seeds. LWT - Food Science and Technology 42:1–8. https://doi.org/10.1016/j.lwt.2008.07.007

76. Shan T, Ma Q, Guo K, et al (2011) Xanthones from Mangosteen Extracts as Natural Chemopreventive Agents: Potential Anticancer Drugs. CMM 11:666–677. https://doi.org/10.2174/156652411797536679

77. Santos CMM, Freitas M, Fernandes E (2018) A comprehensive review on xanthone derivatives as α-glucosidase inhibitors. European Journal of Medicinal Chemistry 157:1460–1479. https://doi.org/10.1016/j.ejmech.2018.07.073

78. Kang H-H, Zhang H-B, Zhong M-J, et al (2018) Potential Antiviral Xanthones from a Coastal Saline Soil Fungus Aspergillus iizukae. Marine Drugs 16:449. https://doi.org/10.3390/md16110449

79. Araújo J, Fernandes C, Pinto M, Tiritan M (2019) Chiral Derivatives of Xanthones with Antimicrobial Activity. Molecules 24:314. https://doi.org/10.3390/molecules24020314

80. Peng X, Sun F, Li G, et al (2021) New Xanthones with Antiagricultural Fungal Pathogen Activities from the Endophytic Fungus *Diaporthe goulteri* L17. J Agric Food Chem 69:11216–11224. https://doi.org/10.1021/acs.jafc.1c03513

81. Cruz MI, Cidade H, Pinto M (2017) Dual/multitargeted xanthone derivatives for Alzheimer's disease: where do we stand? Future Medicinal Chemistry 9:1611–1630. https://doi.org/10.4155/fmc-2017-0086

82. Aye A, Song Y-J, Jeon Y-D, Jin J-S (2020) Xanthone suppresses allergic contact dermatitis in vitro and in vivo. International Immunopharmacology 78:106061. https://doi.org/10.1016/j.intimp.2019.106061

83. Shagufta, Ahmad I (2016) Recent insight into the biological activities of synthetic xanthone derivatives. European Journal of Medicinal Chemistry 116:267–280. https://doi.org/10.1016/j.ejmech.2016.03.058

84. Sun J, Chu Y-F, Wu X, Liu RH (2002) Antioxidant and antiproliferative activities of common fruits. J Agric Food Chem 50:7449–7454. https://doi.org/10.1021/jf0207530

85. Kritsanawong S, Innajak S, Imoto M, Watanapokasin R (2016) Antiproliferative and apoptosis induction of α-mangostin in T47D breast cancer cells. International Journal of Oncology 48:2155–2165. https://doi.org/10.3892/ijo.2016.3399

86. Jittiporn K, Suwanpradid J, Patel C, et al (2014) Anti-angiogenic actions of the mangosteen polyphenolic xanthone derivative α-mangostin. Microvascular Research 93:72–79. https://doi.org/10.1016/j.mvr.2014.03.005

87. Zhang H, Tan Y, Zhao L, et al (2020) Anticancer activity of dietary xanthone α-mangostin against hepatocellular carcinoma by inhibition of STAT3 signaling via stabilization of SHP1. Cell Death Dis 11:63. https://doi.org/10.1038/s41419-020-2227-4

88. Zhang K, Gu Q, Yang K, et al (2016) Anticarcinogenic Effects of α-Mangostin: A Review. Planta Med 83:188–202. https://doi.org/10.1055/s-0042-119651

89. Lee HN, Jang HY, Kim HJ, et al (2016) Antitumor and apoptosis-inducing effects of α-mangostin extracted from the pericarp of the mangosteen fruit (Garcinia mangostana L.) in YD-15 tongue mucoepidermoid carcinoma cells. International Journal of Molecular Medicine 37:939–948. https://doi.org/10.3892/ijmm.2016.2517

90. Wu L, Guo H, Sun H, et al (2016) UNC119 mediates gambogic acid-induced cell-cycle dysregulation through the Gsk3β/β-catenin pathway in hepatocellular carcinoma cells. Anti-Cancer Drugs 27:988–1000. https://doi.org/10.1097/CAD.0000000000000416

91. Liu J, Bao H, Wang H, et al (2019) Synthesis of xanthone derivatives and anti-hepatocellular carcinoma potency evaluation: induced apoptosis. RSC Adv 9:40781–40791. https://doi.org/10.1039/C9RA06408G

92. Rong J-J, Hu R, Song X-M, et al (2010) Gambogic acid triggers DNA damage signaling that induces p53/p21Waf1/CIP1 activation through the ATR-Chk1 pathway. Cancer Letters 296:55–64. https://doi.org/10.1016/j.canlet.2010.03.016

CHAPTER Ⅱ : Computational methods for drug discovery and development

II.1   Clustering analysis

Clustering is a common unsupervised machine-learning technique that partitions a set of data points into groups (clusters) in such a manner that the individuals in the same cluster are similar to each other (internally homogeneous) and dissimilar to those in other clusters (externally heterogeneous). Clustering allows the analysis of multivariate data sets to gain insights and extract some features and patterns that may be hidden in the data. It can be achieved by some characteristics that create a distinction between observations and some sort of similarity metrics to group them into homogenous clusters. Clustering is widely used in various fields such as bioinformatics, market analysis, psychology, economics, information retrieval, and computer graphics. Furthermore, clustering algorithms have been successfully implemented into medical image segmentation to enhance the diagnosis and detection of tumor cells in MRI, CT scans, and ultrasound images [1-3].

II.2   Type of clustering algorithm

In fact, there are more than 100 clustering algorithms known and each clustering methodology follows a different set of rules to assign data points into distinct clusters. Two main categories of clustering algorithms have generally been used in machine learning, namely hierarchical and partitional approaches, as depicted in (Figure Ⅱ. 1).
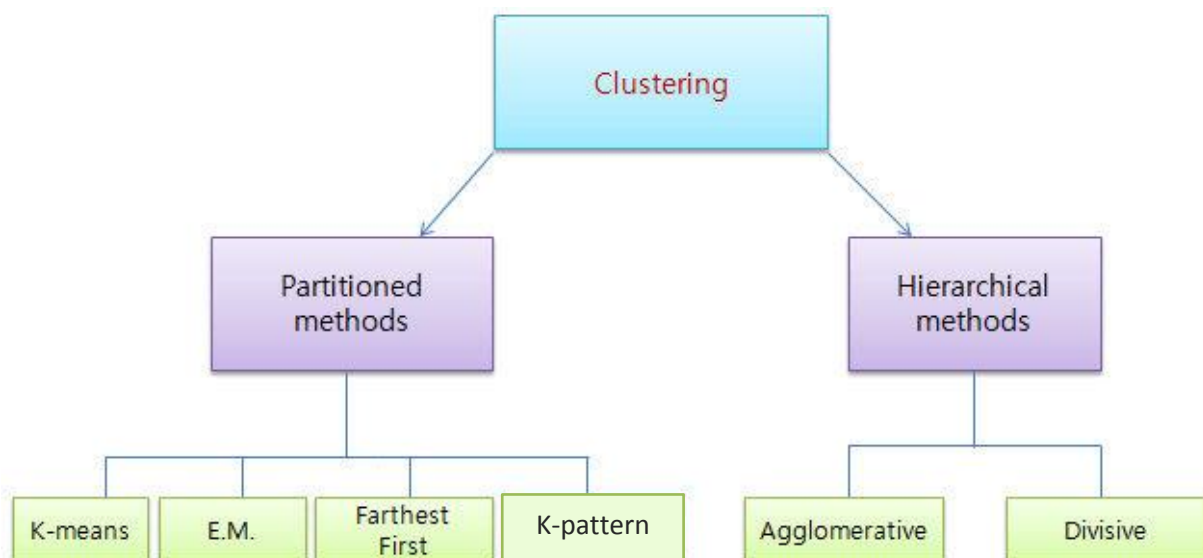
Figure Ⅱ. 1 Taxonomy of clustering algorithms, adapted from [4]

In the case of hierarchical clustering procedures, a sequence of partitions of data observations is generated. It proceeds successively by either merging smaller clusters into larger ones or splitting larger clusters. The results of hierarchical clustering are usually represented as dendrograms or tree-like graphs. On the other hand, the partitional clustering algorithms use iterative operations to divide a set of variables into K clusters based on their proximity to the cluster centroid.

II.3   Principal component analysis

Principal component analysis (PCA) is an exploratory data analysis tool that represents a multivariate dataset as a low dimensional plane while preserving as much variance as possible. It attempts to project a large set of possibly correlated variables into a reduced space defined by a smaller set of new orthogonal variables called principal components presented as a linear combination of the original variables [5]. The first principal component has the largest projected variance in the data. The second principal component is perpendicular to the first component associated with the second largest variance (Figure Ⅱ.2).
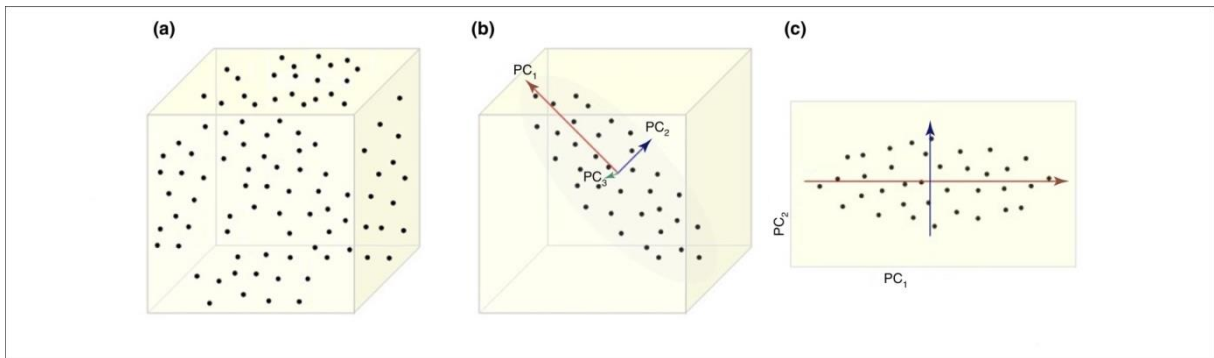
Figure Ⅱ.2 PCA dimensionality reduction, Panel (a) represents a 3D data cloud, Panel (b) shows the three principal components of the database, in Panel (c) 2D data projection of variables into the space of two major components, modified from [6]

Suppose that we have X, (n, p) data matrix of n independent observations on p variables. The first step in PCA involves calculating the mean and covariance matrix between the initial vectors. The sample covariance S between the jth and kth variables is defined by [7]:

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(\tilde{x}_{ij} - \bar{x}_j)(\tilde{x}_{ik} - \bar{x}_k) \qquad (1)$$

Where the samples mean of jth variable is:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_{ij}, \qquad j = 1, 2, \dots, p \qquad (2)$$

Then Singular Vector Decomposition (SVD) of S is implemented to extract principal components. The covariance matrix S is can therefore be written as:

$$S = ULA' \qquad (3)$$

Where U is $(n \times r)$ matrix of left singular vectors, A is $(p \times r)$ matrix of right singular vectors with $U'U = I_r$, $A'A = I_r$, and L is the diagonal matrix of singular values.

The kth PC is given by [7]:

$$z_{ik} = u_{ik} l^{1/2} \qquad i = 1, 2, \dots, n, \quad k = 1, 2, \dots, r \qquad (4)$$

Where u_{ik} is the (i, k)th element of U, and $l^{1/2}$ is the kth diagonal element of L. The kth eigenvector corresponding to its largest eigenvalue is the first principal component, PC1, which explains the most variance in the data. The second principal component PC2 associated with the second largest eigenvalue will be orthogonal to the first one, and so on. After defining the principal components, the data matrix X can be projected onto the subspace spanned by the first and second PCs.

II.4   Hierarchical cluster analysis

Hierarchical cluster analysis or HCA is a stepwise clustering technique that creates a hierarchy of clusters based on the proximity matrix that refers to the distance between data points and repeats this scaling procedure until a dendrogram is constructed. The hierarchical algorithms are divided into agglomerative and divisive clustering. The agglomerative or the bottom-up approach considers each individual as a cluster and then merges the nearest pair of clusters and iteratively proceeds until only a single cluster is left. Many linkage criteria can be used to measure the similarity between clusters, for example, the complete linkage (maximum dissimilarity), single linkage (minimum dissimilarity), average linkage (average dissimilarity), centroid linkage, or the ward's method [8]. Contrary to the agglomerative, the divisive Clustering or also called the top-down approach starts with all data points in one cluster and then splits them recursively to generate more clusters (Figure II. 3). The main advantage of hierarchical clustering is that it does not require a predefined number of clusters, unlike non-hierarchical methods such as k-means.

Types of Hierarchical Clustering
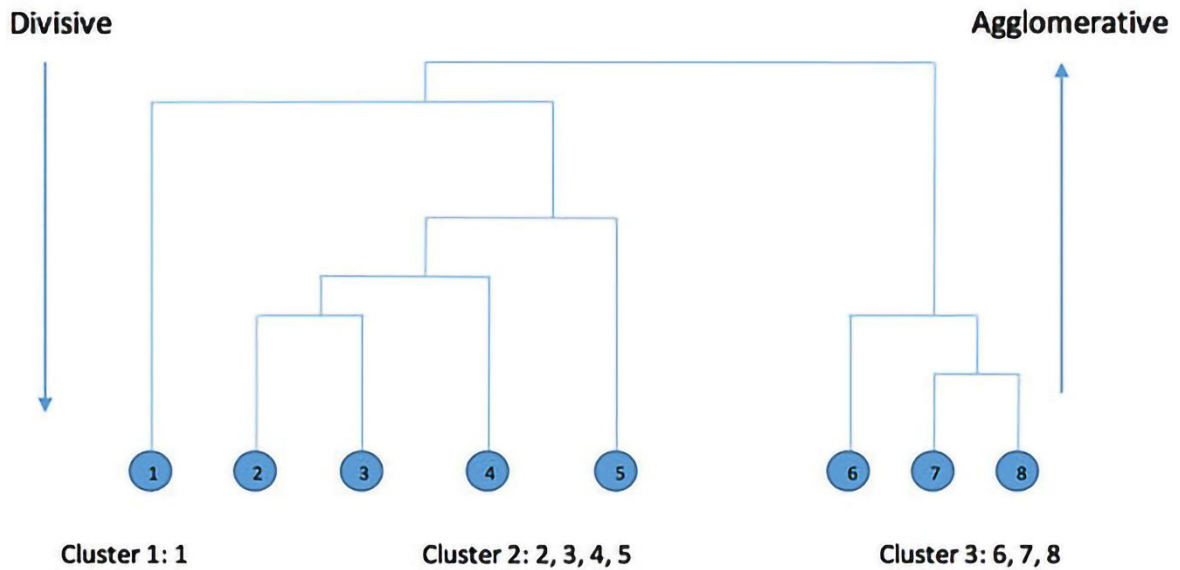
Figure Ⅱ.3 A dendrogram represents both types of hierarchical clustering, agglomerative and divisive, extended from [9]

II.5   K-means clustering

K-means is a partitional clustering algorithm that iteratively segregates data items into K clusters such that individuals in the same cluster are similar to each other and continues until some defined convergence criterion is met. The algorithm starts by randomly choosing the initial centroids K, where a cluster centroid is the mean of all data points within a cluster. Afterward, each point in the dataset is assigned to the nearest center by minimizing the distance between data points and the corresponding centroids. After the clusters are formed, the algorithm iterates this process until no point changes from clusters [10]. The total within-cluster variation function is defined as [11]:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \qquad (5)$$

Where k is the number of clusters, n is the number of data points, x_i is a data point in the cluster j, and c_j is the centroid of cluster j.

In partitional clustering algorithms, the random selection of initial cluster centers is crucial to the result that may depend on this step; a wrong estimation of K can lead to misclassification of observations. Several approaches have been proposed to solve this

problem such as the thumb rule, elbow method, Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), cross-validation method, and silhouette analysis.

II.6   Choosing the number of clusters

Determining the optimal number of clusters in a dataset is a fundamental issue in cluster partitioning, particularly for k-means clustering, which requires the user to specify the number of clusters k as an input parameter. However, there are no valid criteria for choosing the ideal number of clusters in data science. The optimal number of clusters is somewhat subjective and depends on the method used to measure similarities and the parameters used for partitioning [12].

II.6.1   The elbow method

The elbow method is expressed by the total within-cluster (WSS) sum of squares from data points to the closest center as a function of the number of clusters. The total WSS will decrease significantly in the graph toward the inflection point or elbow, and this point corresponds to the right number of clusters with a minimal value of WSS.

II.6.2   The silhouette analysis

An alternative approach to the elbow method is the silhouette analysis that calculates the difference between the average of within-cluster and between-cluster distances and the mean of distances between each point to the other clusters (inter-cluster) [13]. This cluster validation approach is used to determine the appropriate number of clusters and confirm that there is no misallocation or overlapping clusters. The silhouette $S(i)$ can be computed as follows:

$$S(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \qquad (6)$$

Where i is an observation in the dataset assigned to cluster A, $a(i)$ is the average dissimilarity of i to all other observations of A, and $b(i)$ is the minimum average dissimilarity of i to all other observations in different clusters [14].

The quality of clustering can be quantified by calculating the average silhouette index. This method is an internal cluster validation that uses the internal information of the clustering process to evaluate the clustering performance. A high average silhouette width indicates a good clustering. The idea is to be able to establish a robust clustering without drastically reducing the size of the database. In other words, the challenge is to find the best

possible balance between the size of the database and the quality of clustering. This objective is achieved by maximizing the silhouette index.

II.7 Similarity and dissimilarity metrics

The determination of similarity or proximity between each pair of data points is essential for clustering. Distance measurements determine the distance between two observations within the same cluster. The choice of distance measure is often arbitrary but may depend on the type of data and the scaling of variables. Moreover, for most commonly used statistical packages, the default distance metric is the Euclidean distance (Table $\mathrm{II}.1$). On the other hand, the linkage criteria determine how far apart two observations are in different clusters. In hierarchical agglomerative algorithms, the linkage function determines the merging criterion for individuals or clusters at each iteration step, based on the chosen similarity mesure (Table $\mathrm{II}.2$).

Table $\mathrm{II}.1$ Basic distance measures, extended from [15]

Names	Equations
Euclidean distance	$\left\|x_i - x_j\right\|_2 = \sqrt{\left(a_i - a_j\right)^2 + \left(b_i - b_j\right)^2}$
Manhattan distance	$\left\|x_i - x_j\right\|_1 = \left\|a_i - a_j\right\| + \left\|b_i - b_j\right\|$
Maximum distance	$\left\|x_i - x_j\right\|_\infty = max\left\{\left\|a_i - a_j\right\|, \left\|b_i - b_j\right\|\right\}$
Mahalanobis distance	$\sqrt{\left(x_i - x_j\right)^T S^{-1}\left(x_i - x_j\right)}$, where S is the covariance matrix and x_i and x_j are variable vectors

X_i and X_j are the ith and jth observations, where i and j are indices. a and b are feature variables.

Table Ⅱ. 2 Overview of the most commonly used linkage criteria, adapted from [9]

Linkage criteria	
Single	Measures distance between the closest elements in clusters
Centroid	Measures distance between the centroids of each cluster
Ward's method	Measures the distance between clusters as the ANOVA sum of squares, i.e. combining information from all cluster observations
Complete	Measures the distance between observations of groups that are farthest from each other.

II.8 Hopkins statistic for validating cluster tendency

One of the major shortcomings of unsupervised machine learning is that clustering algorithms will return clusters even if the data does not contain clusters. Therefore, it is important to assess whether the datasets include significant clusters, i.e., non-random structures, or not, also known as cluster tendency. In this case, the Hopkins statistic is used to evaluate the feasibility of the clustering analysis and test the null hypothesis of the spatial randomness of the data.

The Hopkins statistic is defined as [16]:

$$H = \frac{\sum_{j=1}^{m} u_j^d}{\sum_{j=1}^{m} u_j^d + \sum_{j=1}^{m} w_j^d} \qquad (7)$$

Where X is a dataset of n data points that have d dimensions, Y is a set of random data points following the uniform distribution; y_j ($j = 1\ to\ m$, $m \ll n$) and x_i ($i = 1\ to\ m$) is a sample of chosen data points from X. u_j is the Euclidean distance from y_j to its nearest neighbor in X and w_j is also the Euclidean distance from xi to its closest neighbor in X.

If the value of the Hopkins statistic is close to 1, the null hypothesis is rejected, indicating the presence of meaningful clusters in the dataset X. A value of H=0.5 denotes that

the data is randomly distributed. The alternative hypothesis is rejected when H<0.5, indicating that the dataset is regularly spaced and there is no cluster structure at X [17].

II.9 Molecular Electrostatic Potential

The Molecular Electrostatic Potential (MEP) is an important reactivity index that allows the visualization of the charge distributions of molecules three-dimensionally. MEP is frequently used to describe non-covalent interactions and design bioactive compounds with improved structural features. It is usually calculated and displayed on the van der Waals surface to generate a molecular electrostatic potential map with the extreme MEP values related to the electronic density and some coded colors to distinguish between electron excess zones and electron deficiency zones.

If a molecule has an electronic density $\rho(r')$ at a point r', then its electrostatic potential at an adjacent point r is evaluated by the following equation [18]:

$$V(r) = \sum_A \frac{Z_A}{|R_{A-}r|} - \int \frac{\rho(r')dr'}{|r' - r|} \qquad (8)$$

Where Z_A is the charge on nucleus A, located at R_A, and A is the number of nuclei in the system. The average number of electrons in each volume element dr' is expressed by the electronic density function $\rho(r')$, a physical observable that can be obtained either experimentally or computationally. The positive first term, represents the contribution of the nuclei; the second term is negative and gives the contribution of the electrons. $V(r)$ can be positive or negative in any given region, depending upon whether the effect of the nuclei or the electrons is dominant there. Overall, Positive $V(r)$ areas will interact favorably with nucleophiles, whereas negative $V(r)$ regions will interact with electrophiles.

II.10 ADME-TOX prediction

Accurate prediction of pharmacokinetics, pharmacodynamics, and toxicological properties plays a key/critical role in the drug development process. Previous research has indicated that approximately 40 % of all drug failures in clinical trials are attributed to poor ADME characteristics and low bioavailability [19]. ADME-TOX screening at the early stages of discovery assists in the recognition and withdrawal of problematic candidates [20]. Therefore, we can test and promote only the most likely to succeed drugs providing good

ADME profile and bioavailability [21]. Thus, numerous software and web tools are hence developed for the prediction of ADME-TOX properties such as SwissADME, pkCSM, and ADMETlab 2.0. These *in silico* models are generated by machine learning methods, such as support vector machine (SVM) [22], artificial neural networks (ANN) [23], recursive partitioning (RP) [24], and naïve Bayesian (NB) [25] together with molecular fingerprints that represent the molecular structures. In the next section, we will provide a set of pharmacokinetic and toxicological definitions in detail that will help to understand and interpret ADME-TOX prediction results.

II.10.1 The gastrointestinal absorption

Gastrointestinal absorption (GI) involves the absorption of drugs through intestinal epithelium by different mechanisms (via the lipophilic transcellular or hydrophilic paracellular) [26]. The human intestinal absorption (HIA) of a substance is usually quantified as the fraction of the given administered dose that reaches the portal vein [27]:

$$HIA = \frac{D_{Blood}}{D_{Oral}} \qquad (9)$$

Where D_{Blood} is the amount of a substance that reaches the portal vein and D_{Oral} is the total amount of the substance administered orally. Most ADME-TOX predictors display the results in a binary form (high/low) or as a percentage value.

II.10.2 The volume of distribution

The volume of distribution (V_d), also known as the apparent volume of distribution, is another critical PK parameter that relates the administered amount of drug to the plasma concentration at time zero C_0. It can be defined mathematically as:

$$V_d = \frac{Dose}{C_0} \qquad (10)$$

The accurate estimation of this descriptor is useful in determining dosage requirements to achieve the required therapeutic plasma concentration. Hydrophilic drugs exhibit a high plasma affinity that decreases the apparent volume of distribution, while lipophilic drugs with a high value of V_d denote higher drug distribution in tissues than plasma [28].

II.10.3   The blood-brain barrier permeability

The blood-brain barrier (BBB) is a biochemical barrier that Limits the distribution of harmful substances into the central nervous system (CNS) [29]. This descriptor is used to predict the permeable compounds in a binary form (yes/no) or as a value that describe the probability of being permeable or not.

II.10.4   Permeability glycoprotein substrates and inhibitors

P-gp (permeability glycoprotein) is an efflux-transporter that pumps xenobiotic compounds across the cellular membranes for excretion. It is highly expressed in the intestine and also occurs in the liver, brain, and kidney cells. The drug molecules transported out of the cells by this efflux protein are termed "p-gp substrates" [30]. These predictive models apply the SVM algorithm on large datasets of known substrates/non-substrates or inhibitors/non-inhibitors for binary classification that returns "Yes" or "No" if the investigated drug has a higher probability to be P-gp substrate/inhibitor or not.

II.10.5   Cytochrome P450 substrates and inhibitors

Cytochrome P450 (CYP450) is the principal enzyme system involved in phase I of drugs biotransformation via oxidation. It is mainly expressed in the liver and also occurs in the small intestine, kidney, and lung [31]. Moreover, out of 57 human CYP450 enzymes the CYP1A2, CYP2C9, CYP2C19, CYP2D6, CYP3A4, and CYP2E1 metabolize 90 % of drugs [32]. P450 isoenzymes interact with molecules differently, it can be inhibited which decreases the clearance rate of drugs, or induced to enhance the activity of CYP450. Indeed, it is well established that interaction with CYP cytochromes may lead to Drug-drug interactions (DDI) [33-36].

Before the drug molecule reaches systemic circulation, it must overcome several biological barriers. The absorbed fraction of oral drug through the gastrointestinal (GI) tract can be excreted by p-gp, metabolized by intestinal CYP enzymes, or transferred to the liver via the portal vein [37]. In the liver, the drug proportion that escapes the hepatic p-gp transporter and CYP metabolism will reach the systemic circulation and may then be delivered to its target site (Figure Ⅱ.4). These models implement the SVM algorithm to predict whether the studied compound is expected to be a CYP substrate and/or inhibitor or not.
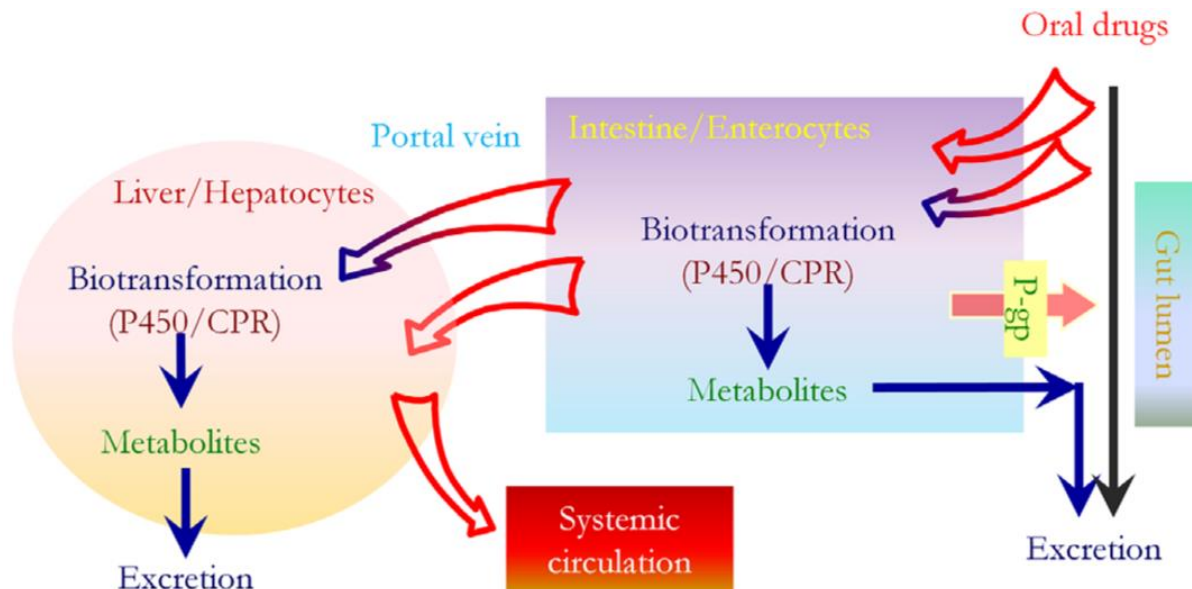
Figure Ⅱ.4 Overview of the intestinal and hepatic P450 enzymes action on orally administered drugs, adapted from [37]

II.10.6 Clearance

Clearance (CL) is the most key determinant of the dosing regimen and the frequency of administration of drugs. Drugs can be excreted from the body through numerous pathways. However, it occurs primarily as a combination of renal, hepatic, and biliary clearance. Drug clearance is defined as the volume of plasma from which xenobiotics would be totally removed per unit time [38]. Mathematically, the clearance can be expressed as:

$$Clearance = V_d \times K_e \qquad (11)$$

Where K_e is the elimination rate constant, and V_d is the apparent volume of distribution for the drug. A high clearance value indicates that the xenobiotic is cleared rapidly from the body, while a low clearance value denotes a slower clearance rate.

II.10.7 AMES toxicity

The Ames test is one of the most common drug safety assessments designed to determine the mutagenic potential of compounds that cause mutations in the DNA of the bacterium *Salmonella typhimurium* (test organism). This test is an efficient alternative to predict the carcinogenic potential of drug candidates because standard carcinogenicity assays on mice and rats are time-consuming (up to three years to complete) and expensive. A

compound is classified as Ames positive, which indicates that the compound is mutagenic and thus may act as a carcinogen [39].

II.10.8 Hepatotoxicity

Drug-induced liver injury (DILI) is a major safety concern for drug development and a primary cause of failure in preclinical and clinical trials and withdrawals of approved drugs. Currently, more than 50 drugs have been implicated in causing severe liver injury and hence withdrawn from the market worldwide, such as Troglitazone, Bromfenac, and Ticrynafen [40]. Therefore, eliminating hepatotoxic drug candidates in advance is considered a cost-effective strategy to reduce the rate of attrition in drug discovery. In silico models generated using different learning algorithms on diverse large-scale datasets can provide an accurate estimation of hepatotoxicity and the result is displayed in a binary form (yes/no).

II.10.9 HERG inhibition

The human Ether-à-go-go-Related Gene (hERG) encodes the cardiac potassium ion channel involved in regulating the exchange of cardiac action potential and resting potential by transporting potassium ($K+$) ions out of the cardiac myocytes. The inhibition of hERG results in slowed repolarisation and therefore, a prolongation of the QT interval (LQTS) in the electrocardiogram (ECG), leading to cardiac arrhythmia called Torsade de Pointes (TdP). The withdrawal of many drugs from late-stage clinical trials due to these cardiotoxic effects emphasizes the importance of early detection of hERG inhibitors in the drug discovery process [41]. This descriptor is used to predict if a given compound is likely to be a hERG I/II inhibitor in a categorical form (yes/no).

II.11 Molecular docking

Molecular docking is one of the most commonly used virtual screening approaches that plays a pivotal role in structure-based rational drug design, particularly for the prediction of intermolecular interactions. In silico molecular docking is designed to predict the position, preferred orientation, affinity, and interaction of a small molecule (ligand) that binds with a biological target of therapeutic interest (usually a protein). The conventional concept of docking methods was based on the lock-and-key theory proposed by Fischer [42], where both the ligand and the receptor can be treated as rigid systems to find the correct orientation for the "key" to open up the "lock" (Figure Ⅱ.5). Later, Koshland suggested the "induced-fit" theory, in which the ligand and receptor are considered flexible during docking and thus

increase the accuracy of prediction [43]. The details of target-ligand interactions at the molecular level are therefore of great interest to many diseases to design a ligand that either inhibits or activates the target protein that significantly affects fundamental biochemical processes such as signal transduction, transcription, and enzymatic catalysis. Molecular docking techniques represent an attractive and cost-effective alternative to experimental methods, including high throughput screening (HTS), which requires a significant investment in drug discovery [44].
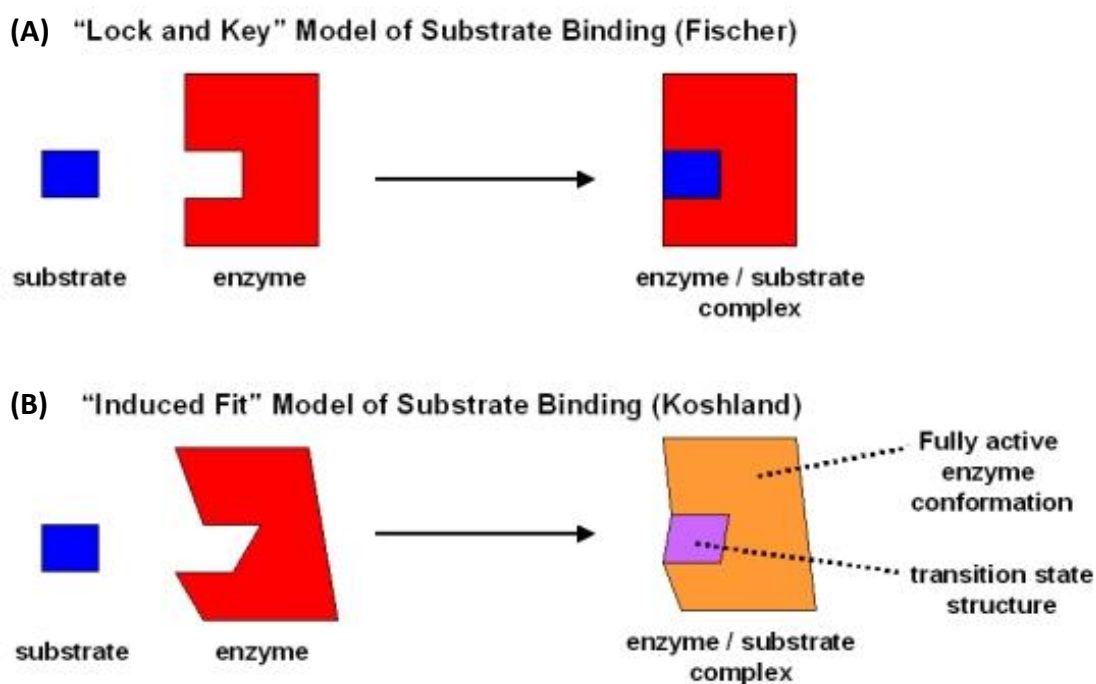
Figure Ⅱ.5 Schematic diagram of two binding models for enzyme-substrate interaction. (A) The lock-and-key model, (B) induced-fit model describing the conformational changes in the active site, adapted from [45]

II.11.1 Types of molecular docking

II.11.1.1 Rigid docking

This method implements the simplest docking algorithm that ignores the flexibility of molecules and assumes that the receptor and the ligand are rigid entities, in which the bond angles, bond lengths, and torsion angles of the components do not change at any step of complex formation. It reduces the number of degrees of freedom from several thousand to only six (three translational and three rotational) and thus limits the search space for the ligand to fit in the binding site [46]. Rigid docking is often used to examine large systems,

such as protein-protein interactions. DOCK is a well-known program that applies this algorithm to find molecules matching by shape complementarity [47-50].

II.11.1.2 Semi-flexible docking

It represents the most common approach in docking programs that incorporates the flexibility of ligands while the proteins remain rigid during the docking process in order to describe possible torsion angles and bond rotation. It is suitable for docking small ligand molecules and macromolecules, such as proteins or nucleic acids [51]. In semi-flexible docking, six degrees of translational and rotational freedom are sampled, as well as the conformational degrees of freedom of the ligand, however, it needs more computational time. For this category, MOE [52], FlexX [53], Glide [54], and AutoDock [55] are widely used to perform semi-flexible docking.

II.11.1.3 Flexible-flexible docking

Recently, it has become increasingly clear that side-chain flexibility plays a crucial role in the molecular recognition of protein-ligand complexes as targets can adopt different conformations in the unbound and bound states. The protein main chain torsion angles are also frequently altered from their unbound conformations, which the previous methods fail to detect. Modern docking tools address these issues by allowing the dynamic conformational change of both the target and the ligand, thus mimicking the nature of biological molecules to describe intermolecular binding interactions more accurately. Fully-flexible docking takes longer than rigid-body docking due to the large degrees of freedom, and consequently a larger search space is used to examine the binding modes of protein-ligand complexes [56]. The most common docking programs used in this class primarily include MOE [52], GOLD [57], ICM [58], Surflex [59], and AutoDock Vina [60].

II.11.2 Molecular docking procedure

II.11.2.1 Target preparation

In this phase, the 3D structures of biological macromolecules are downloaded as PDB files from the RCSB Protein Data Bank (https://www.rcsb.org/). A huge number of proteins, lipids, carbohydrates, and nucleic acids are stored in this database, both as single structures and as complexes. However, structures with low resolution and a co-crystalized ligand bound in the binding site should be preferred. These structures are determined experimentally using X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy, or electron

microscopy. The preparation of protein structure includes protonation, charge addition, cofactors removal, residue complementation, and other operations [61].

II.11.2.2 Ligand preparation

Ligand preparation for docking simulations is highly important for improving the accuracy of results, during which the potential energy is minimized to obtain a low energy conformation for each ligand. The ligand structures can be retrieved from scientific literature, patents, or public databases, including the ZINC database [62], ChEMBL [63], IUPHAR/BPS [64], SuperNatural [65], and PubChem [66]. These online databases provide a large number of compounds to select the most interesting and promising molecules for a particular molecular target. Filters of physicochemical properties are frequently used to reduce the number of molecules to be docked, such as the molecular weight, log P, polar surface area, number of rotatable bonds, net charge, and synthetic accessibility [67].

II.11.2.3 Active site detection

At this point, the position of the co-crystalized ligand within the protein is identified as the binding pocket that enables the user to know the location of the active site prior to docking. In the absence of native ligands, cavity detection tools, such as SurfNet [68, 69], POCKET [70], GRID [71, 72], PASS [73] and MMC [74] can be used to determine the binding sites within proteins.

II.11.2.4 Docking

During this stage, several copies (conformations) of a ligand are randomly distributed on the protein binding cavity. Each resulting pose is then optimized simultaneously by several minimization cycles. MOE allows the performance of semi-flexible docking where the protein is kept rigid while the ligand is flexible. If two poses converge at the same position, only that of better interaction energy (lowest energy) is conserved. This whole process can be repeated during a user-defined number of iterations. At the end of docking, all generated poses are associated with interaction energy and ranked according to certain scoring systems from the most favorable energy score to the least favorable one.

II.11.2.5 Evaluating Docking Results

The prediction accuracy of protein-ligand complexes assessed by docking simulations can be evaluated by calculating the Root Mean Square Deviation (RMSD) between the generated structure (top scored pose) and the co-crystalized ligand (experimental

conformation) of the protein. This similarity metric is defined by the Euclidean distance between the coordinates of the atoms of the two superimposed structures, as expressed:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N} d_i^2} \qquad (12)$$

Where N is the total number of atoms in the molecule and d_i is the distance between coordinates of atom i in the two structures [75]. RMSD is always positive, and a value of 0 (rarely achieved in practice) indicates a perfect fit for the data. A good performance is usually considered when the RMSD is less than 2Å.

II.11.3 Docking theory

Molecular docking software aims to accurately predict the best binding mode (orientation, position, and conformation) between two biochemical entities based on two main components: the search algorithm and the scoring function (Figure II.6). The search algorithm is able to search and generate all possible conformations of the small organic molecule in the binding cavity of a therapeutic macromolecule and therefore generate possible modes of interaction. While the scoring function is used to predict the best orientation that represents the true structure of the complex and estimate the free energy of binding.
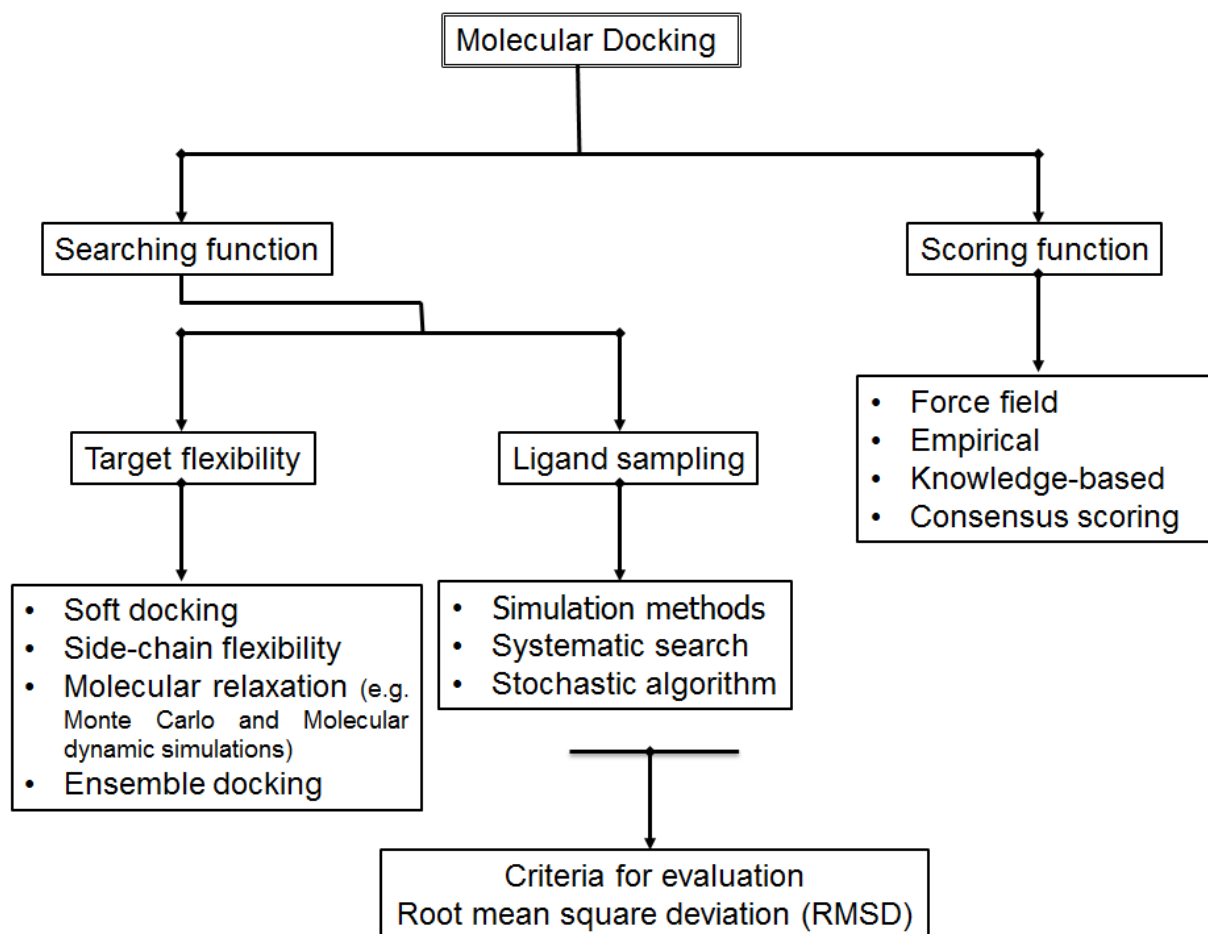
Figure Ⅱ. 6 Classification of algorithms for protein-ligand docking, modified from
[76]

II.11.3.1   Search algorithm

Search algorithms go through all generated binding modes across two molecules to ensure that the binding site is explored to the maximum. Common Ligand flexibility search algorithms include simulation methods, systematic search methods, and stochastic search methods. The simulation methods start with a known initial state up to the state of lower energy. Energy minimization (EM), molecular dynamics (MD), and simulated annealing are examples of simulation methods. In the systematic search, the ligand conformation is recurrently approached until the minimum energy conformation is identified. It includes three subclasses exhaustive search algorithms, fragment based algorithms, and conformational ensemble [77]. In stochastic algorithms, ligand conformations are sampled by making random changes to the degrees of freedom of the ligands. Monte Carlo (MC) [78], genetic algorithms (GAs) [79], Tabu Search (TS) [80], and Swarm Optimisation (SO) [81] algorithms fall into this category.

II.11.3.2 Scoring Functions

Sampling algorithms are further complemented by scoring functions that adopt various assumptions and simplifications to rank the generated conformations according to their estimated binding affinity and recognize the most relevant pose among the many others created by the search algorithm. Therefore, numerous scoring functions have been developed, in this regard to precisely characterize the behavior of protein-ligand complexes [82-85]. They can be classified into four basic categories: force-field, empirical, knowledge-based, and consensus scoring.

- *Force-field based scoring functions*

Force field–based scoring functions express the binding free energy of complex as a sum of several uncorrelated terms of molecular mechanics, including van der Waals interactions, and electrostatic interactions between and within docked molecules. This type of function usually neglects most solvent effects and solute entropies. One typical force field scoring function used in molecular docking through DOCK is composed of two energy components from the Amber force fields, Lennard-Jones VDW, and an electrostatic term. It can be expressed by equation (13):

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right) \qquad (13)$$

Where r_{ij} stands for the distance between protein atom i and ligand atom j, A_{ij} and B_{ij} are the VDW parameters, and q_i and q_j are the atomic charges. $\varepsilon(r_{ij})$ in the Coulombic term is the distance dependent dielectric constant that is usually set to $4r_{ij}$, representing the effect of solvent implicitly [77].

- *Empirical scoring functions*

The empirical scoring functions estimate the binding affinity of a complex by a sum of individual energy terms, including hydrogen bond, ionic interaction, hydrophobic effect, binding entropy, and the number of rotatable bonds in the ligand. The coefficients associated with the functional terms are obtained from regression analysis using experimentally determined binding energies and x-ray structures. The calculation of binding score using empirical scoring functions is much faster than force field scoring functions, due to their simple energy terms. Experimental binding free energies and protein-ligand crystal structures

for 45 complexes were used to build the first empirical scoring system for predicting binding free energies, which was incorporated in LUDI as follow: shown in equation (14):

$$\Delta G_{binding} = \Delta G_0 + \Delta G_{hb} \sum_{h-bonds} f(\Delta R, \Delta \alpha) + \Delta G_{ionic} \sum_{ionic\ int.} f(\Delta R, \Delta \alpha)$$
$$+ \Delta G_{lipo}|A_{lipo}| + \Delta G_{rot}\ NROT \qquad (14)$$

Where ΔG_0 is a contribution to the binding energy that is independent of any protein interactions, ΔG_{hb} describes the contribution from an ideal hydrogen bond, ΔG_{ionic} indicates the contribution from unperturbed ionic interactions, ΔG_{lipo} denotes the contribution from lipophilic interactions, while A_{lipo} is the lipophilic contact surface between the protein and the ligand, ΔG_{rot} considers the loss of binding energy due to freezing of internal degrees of freedom in the ligand, $NROT$ represents the number of rotatable bonds and $f(\Delta R, \Delta \alpha)$ is a penalty function that accounts for large deviations from ideal hydrogen bond geometry [86].

- ***Knowledge-based scoring functions***

In knowledge-based functions, the frequency of atom pairs interactions observed in protein-ligand complexes is determined and then converted into pairwise atomic potentials, describing the preferred geometries of the protein-ligand pairwise atoms. Popular implementations of such functions include Potential of Mean Force (PMF) and DrugScore of FLEX program. The use of these functions provides a compromise between accuracy and speed compared to force field and empirical scoring functions. The overall score is calculated by equation (15):

$$W(r) = -k_B\ T\ \ln[g(r)], g(r) = \rho(r)/\rho^*(r) \qquad (15)$$

Where k_B is the Boltzmann constant, T is the absolute temperature of the system, $\rho(r)$ is the number density of the protein-ligand atom pair at distance r, $\rho^*(r)$ is the pair density in a reference state where the interatomic interactions are zero, and $g(r)$ is the pair distribution function [87].

- ***Consensus scoring functions***

Consensus scoring is a recent scoring strategy that evaluates the docked conformations by combining multiple scoring functions that significantly reduce the number of false positives identified by individuals scoring functions, thereby enhancing the ability to discriminate between active and inactive enzyme inhibitors, which in turn provides notable

improvements in docking accuracy [88-91]. Examples of consensus scoring include MultiScore [92], GFscore [93], X-Cscore [94], SeleX-CS [95], and SCS [96].

II.12   Molecular dynamics

Molecular dynamics is the most popular simulation approach in the field of molecular modeling that explores the conformational space and incorporates the flexibility of both ligand and receptor more effectively than other algorithms. MD simulations consider the thermodynamic state variables of a system at the microscopic scale, such as temperature and pressure to further investigate the binding affinity and stability of complexes. The trajectory of particles in the system is usually captured using Newton's equations of motion, and the potential energy is determined using force fields. In a system of N interacting atoms $=$ $1, 2, ... N$, the equation of motion for atom i is a second order differential equation according to Newton's second law [97]:

$$ F_i = m_i \frac{d^2 r_i}{dt^2} \qquad (16) $$

Where m_i is the mass of atom i, r_i denotes the position of atom i at time t, and F_i is the net force exerted on the atom i.

II.12.1   MD simulation applications

MD approaches are routinely used to evaluate the protein and ligand in their bound and unbound states in order to approximate the free energy change that occurred in each stage of the transition between the two states. Such methods model all the degrees of freedom in the protein-ligand complex and explicitly include the solvent effects more naturally, providing a more realistic and detailed picture of the binding mode [98, 99]. One of the most common uses of MD simulations has focused on refining and rescoring the docking poses to investigate the basis of protein structure and function [100-102]. Currently, molecular dynamics simulations are widely used to refine molecular 3D structures of proteins obtained by X-ray [103], NMR [104], or homology modeling [105].

The drawbacks of this technique are related to the molecule geometry that changes in a series of small steps, requiring many motions and thus making it difficult to cross the high energy conformational barriers within a reasonable period of simulation time, therefore can get easily trapped in local minima of the energy surface. Additionally, simulation methodologies are slow and require high computational expenses. There have been many

approaches to address these limitations, including simulated annealing, soft potentials, and the use of an initial set of ligand conformations [106-111]. Several programs have been used to perform MD simulations, including GROMACS [112], AMBER [113], NAMD [114], CHARMM [115], OpenMM [116], and LAMMPS [117].

II.13 References

1. Kumar DM, Satyanarayana D, Prasad MNG (2021) MRI brain tumor detection using optimal possibilistic fuzzy C-means clustering algorithm and adaptive k-nearest neighbor classifier. J Ambient Intell Human Comput 12:2867–2880. https://doi.org/10.1007/s12652-020-02444-7

2. Morozov SP, Gombolevskiy VA, Elizarov AB, et al (2021) A simplified cluster model and a tool adapted for collaborative labeling of lung cancer CT scans. Computer Methods and Programs in Biomedicine 206:106111. https://doi.org/10.1016/j.cmpb.2021.106111

3. Lei Y, He X, Yao J, et al (2021) Breast tumor segmentation in 3D automatic breast ultrasound using Mask scoring R-CNN. Med Phys 48:204–214. https://doi.org/10.1002/mp.14569

4. Bourobou S, Yoo Y (2015) User Activity Recognition in Smart Homes Using Pattern Clustering Applied to Temporal ANN Algorithm. Sensors 15:11953–11971. https://doi.org/10.3390/s150511953

5. Balakin KV (2010) Pharmaceutical data mining: approaches and applications for drug discovery

6. Giuliani A (2017) The application of principal component analysis to drug discovery and biomedical data. Drug Discovery Today 22:1069–1076. https://doi.org/10.1016/j.drudis.2017.01.005

7. Jolliffe IT (2002) Principal component analysis, 2nd ed. Springer, New York

8. Greenacre M, Primicerio R (2013) Multivariate analysis of ecological data. Fundación BBVA, Bilbao

9. Deliu M, Sperrin M, Belgrave D, Custovic A (2016) Identification of Asthma Subtypes Using Clustering Methodologies. Pulm Ther 2:19–41. https://doi.org/10.1007/s41030-016-0017-z

10. Aggarwal CC, Reddy CK (2014) Data clustering: algorithms and applications. Chapman and Hall/CRC, Boca Raton

11. Soni K, Patel A (2017) Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data. 13:899–906

12. Borcard D, Gillet F, Legendre P (2018) Numerical Ecology with R. Springer International Publishing, Cham

13. Kassambara A (2017) Practical guide to cluster analysis in R: unsupervised machine learning, Edition 1. STHDA, Erscheinungsort nicht ermittelbar

14. Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7

15. Zhang Z, Murtagh F, Van Poucke S, et al (2017) Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R. Ann Transl Med 5:75–75. https://doi.org/10.21037/atm.2017.02.05

16. Mahalanabis AK, International Federation of Automatic Control (1982) Theory and application of digital control: proceedings of the IFAC Symposium, New Delhi, India, 5-7 January 1982, 1st ed. Published for the International Federation of Automatic Control by Pergamon Press, Oxford, [Oxfordshire] ; New York

17. Theodoridis S, Koutroumbas K (2009) Pattern recognition, 4th ed. Academic Press, Burlington, MA London

18. Murray JS, Politzer P (2011) The electrostatic potential: an overview. WIREs Comput Mol Sci 1:153–163. https://doi.org/10.1002/wcms.19

19. Bocci G, Carosati E, Vayer P, et al (2017) ADME-Space: a new tool for medicinal chemists to explore ADME properties. Sci Rep 7:6359. https://doi.org/10.1038/s41598-017-06692-0

20. Reddy P, Reddy MBM, Reddy R, et al (2020) Molecular docking, PKPD, and assessment of toxicity of few chalcone analogues as EGFR inhibitor in search of anticancer agents. Struct Chem 31:2249–2255. https://doi.org/10.1007/s11224-020-01571-3

21. Verma SK, Ratre P, Jain AK, et al (2021) De novo designing, assessment of target affinity and binding interactions against aromatase: Discovery of novel leads as anti-breast cancer agents. Struct Chem 32:847–858. https://doi.org/10.1007/s11224-020-01673-y

22. Li Q, Bender A, Pei J, Lai L (2007) A Large Descriptor Set and a Probabilistic Kernel-Based Classifier Significantly Improve Druglikeness Classification. J Chem Inf Model 47:1776–1786. https://doi.org/10.1021/ci700107y

23. Sadowski J, Kubinyi H (1998) A Scoring Scheme for Discriminating between Drugs and Nondrugs. J Med Chem 41:3325–3329. https://doi.org/10.1021/jm9706776

24. Walters WP, Murcko MA (2002) Prediction of 'drug-likeness.' Advanced Drug Delivery Reviews 54:255–271. https://doi.org/10.1016/S0169-409X(02)00003-0

25. Tian S, Wang J, Li Y, et al (2012) Drug-likeness Analysis of Traditional Chinese Medicines: Prediction of Drug-likeness Using Machine Learning Approaches. Mol Pharmaceutics 9:2875–2886. https://doi.org/10.1021/mp300198d

26. Ozawa T, Takayama K, Okamoto R, et al (2015) Generation of enterocyte-like cells from human induced pluripotent stem cells for drug absorption and metabolism studies in human small intestine. Sci Rep 5:16479. https://doi.org/10.1038/srep16479

27. Durán-Iturbide NA, Díaz-Eufracio BI, Medina-Franco JL (2020) *In Silico* ADME/Tox Profiling of Natural Products: A Focus on BIOFACQUIM. ACS Omega 5:16076–16084. https://doi.org/10.1021/acsomega.0c01581

28. Diasio RB (2012) Principles of Drug Therapy. In: Goldman's Cecil Medicine. Elsevier, pp 124–132

29. Cecchelli R, Berezowski V, Lundquist S, et al (2007) Modelling of the blood–brain barrier in drug discovery and development. Nat Rev Drug Discov 6:650–661. https://doi.org/10.1038/nrd2368

30. Hugger ED, Novak BL, Burton PS, et al (2002) A comparison of commonly used polyethoxylated pharmaceutical excipients on their ability to inhibit P-glycoprotein activity in vitro. Journal of Pharmaceutical Sciences 91:1991–2002. https://doi.org/10.1002/jps.10176

31. Ding X, Kaminsky LS (2003) Human Extrahepatic Cytochromes P450: Function in Xenobiotic Metabolism and Tissue-Selective Chemical Toxicity in the Respiratory and Gastrointestinal Tracts. Annu Rev Pharmacol Toxicol 43:149–173. https://doi.org/10.1146/annurev.pharmtox.43.100901.140251

32. Stavropoulou E, Pircalabioru GG, Bezirtzoglou E (2018) The Role of Cytochromes P450 in Infection. Front Immunol 9:89. https://doi.org/10.3389/fimmu.2018.00089

33. Tornio A, Filppula AM, Niemi M, Backman JT (2019) Clinical Studies on Drug–Drug Interactions Involving Metabolism and Transport: Methodology, Pitfalls, and Interpretation. Clin Pharmacol Ther 105:1345–1361. https://doi.org/10.1002/cpt.1435

34. Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7:42717. https://doi.org/10.1038/srep42717

35. Yu J, Petrie ID, Levy RH, Ragueneau-Majlessi I (2019) Mechanisms and Clinical Significance of Pharmacokinetic-Based Drug-Drug Interactions with Drugs Approved by the U.S. Food and Drug Administration in 2017. Drug Metab Dispos 47:135–144. https://doi.org/10.1124/dmd.118.084905

36. Filppula AM, Parvizi R, Mateus A, et al (2019) Improved predictions of time-dependent drug-drug interactions by determination of cytosolic drug concentrations. Sci Rep 9:5850. https://doi.org/10.1038/s41598-019-42051-x

37. Xie F, Ding X, Zhang Q-Y (2016) An update on the role of intestinal cytochrome P450 enzymes in drug disposition. Acta Pharmaceutica Sinica B 6:374–383. https://doi.org/10.1016/j.apsb.2016.07.012

38. Hinderliter P, Saghir SA (2014) Pharmacokinetics. In: Encyclopedia of Toxicology. Elsevier, pp 849–855

39. Hillebrecht A, Muster W, Brigo A, et al (2011) Comparative Evaluation of *in Silico* Systems for Ames Test Mutagenicity Prediction: Scope and Limitations. Chem Res Toxicol 24:843–854. https://doi.org/10.1021/tx2000398

40. Liu L, Fu L, Zhang J-W, et al (2019) Three-Level Hepatotoxicity Prediction System Based on Adverse Hepatic Effects. Mol Pharmaceutics 16:393–408. https://doi.org/10.1021/acs.molpharmaceut.8b01048

41. Garrido A, Lepailleur A, Mignani SM, et al (2020) hERG toxicity assessment: Useful guidelines for drug design. European Journal of Medicinal Chemistry 195:112290. https://doi.org/10.1016/j.ejmech.2020.112290

42. Fischer E (1894) Einfluss der Configuration auf die Wirkung der Enzyme. Ber Dtsch Chem Ges 27:2985–2993. https://doi.org/10.1002/cber.18940270364

43. Koshland DE (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis. Proc Natl Acad Sci USA 44:98–104. https://doi.org/10.1073/pnas.44.2.98

44. Kortagere S (2013) In Silico Models for Drug Discovery. Humana Press, Totowa, NJ

45. http://www.mikeblaber.org/oldwine/BCH4053/Lecture26/Lecture26.htm

46. Gschwend DA, Good AC, Kuntz ID (1996) Molecular docking towards drug discovery. J Mol Recognit 9:175–186. https://doi.org/10.1002/(SICI)1099-1352(199603)9:2<175::AID-JMR260>3.0.CO;2-D

47. Gschwend DA, Kuntz ID (1996) Orientational sampling and rigid-body minimization in molecular docking revisited: On-the-fly optimization and degeneracy removal. J Computer-Aided Mol Des 10:123–132. https://doi.org/10.1007/BF00402820

48. Otyepka M, Kryštof V, Havlíček L, et al (2000) Docking-Based Development of Purine-like Inhibitors of Cyclin-Dependent Kinase-2. J Med Chem 43:2506–2513. https://doi.org/10.1021/jm990506w

49. Sauton N, Lagorce D, Villoutreix BO, Miteva MA (2008) MS-DOCK: Accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. BMC Bioinformatics 9:184. https://doi.org/10.1186/1471-2105-9-184

50. Reedijk J (2014) Reference module in chemistry, molecular sciences and chemical engineering. Elsevier, Oxford

51. Jiang W, Li W, Hong Y, et al (2016) Cloning, Expression, Mutagenesis Library Construction of Glycerol Dehydratase, and Binding Mode Simulation of Its Reactivase

with Ligands. Appl Biochem Biotechnol 178:739–752. https://doi.org/10.1007/s12010-015-1906-6

52. Molecular Operating Environment (MOE) (2014) Chemical Computing Group Inc, 1010 Sherbrooke St. West, Suite # 910, Montreal QC, Canada H3A 2R7

53. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A Fast Flexible Docking Method using an Incremental Construction Algorithm. Journal of Molecular Biology 261:470–489. https://doi.org/10.1006/jmbi.1996.0477

54. Friesner RA, Banks JL, Murphy RB, et al (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 47:1739–1749. https://doi.org/10.1021/jm0306430

55. Morris GM, Goodsell DS, Halliday RS, et al (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 19:1639–1662. https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

56. Sotriffer C (2011) Virtual screening: principles, challenges, and practical guidelines. Wiley-VCH, Weinheim, Germany

57. Jones G, Willett P, Glen RC, et al (1997) Development and validation of a genetic algorithm for flexible docking 1 1Edited by F. E. Cohen. Journal of Molecular Biology 267:727–748. https://doi.org/10.1006/jmbi.1996.0897

58. Schapira M, Abagyan R, Totrov M (2003) Nuclear Hormone Receptor Targeted Virtual Screening. J Med Chem 46:3045–3059. https://doi.org/10.1021/jm0300173

59. Jain AN (2003) Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine. J Med Chem 46:499–511. https://doi.org/10.1021/jm020406h

60. Eberhardt J, Santos-Martins D, Tillack AF, Forli S (2021) AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model 61:3891–3898. https://doi.org/10.1021/acs.jcim.1c00203

61. Pagadala NS, Syed K, Tuszynski J (2017) Software for molecular docking: a review. Biophys Rev 9:91–102. https://doi.org/10.1007/s12551-016-0247-1

62. Irwin JJ, Sterling T, Mysinger MM, et al (2012) ZINC: A Free Tool to Discover Chemistry for Biology. J Chem Inf Model 52:1757–1768. https://doi.org/10.1021/ci3001277

63. Gaulton A, Bellis LJ, Bento AP, et al (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Research 40:D1100–D1107. https://doi.org/10.1093/nar/gkr777

64. Armstrong JF, Faccenda E, Harding SD, et al (2019) The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. Nucleic Acids Research gkz951. https://doi.org/10.1093/nar/gkz951

65. Dunkel M (2006) SuperNatural: a searchable database of available natural compounds. Nucleic Acids Research 34:D678–D683. https://doi.org/10.1093/nar/gkj132

66. Bolton EE, Wang Y, Thiessen PA, Bryant SH (2008) PubChem: Integrated Platform of Small Molecules and Biological Activities. In: Annual Reports in Computational Chemistry. Elsevier, pp 217–241

67. Guedes IA, de Magalhães CS, Dardenne LE (2014) Receptor–ligand molecular docking. Biophys Rev 6:75–87. https://doi.org/10.1007/s12551-013-0130-2

68. Laskowski RA (1995) SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. Journal of Molecular Graphics 13:323–330. https://doi.org/10.1016/0263-7855(95)00073-9

69. Glaser F, Morris RJ, Najmanovich RJ, et al (2005) A method for localizing ligand binding pockets in protein structures. Proteins 62:479–488. https://doi.org/10.1002/prot.20769

70. Levitt DG, Banaszak LJ (1992) POCKET: A computer graphies method for identifying and displaying protein cavities and their surrounding amino acids. Journal of Molecular Graphics 10:229–234. https://doi.org/10.1016/0263-7855(92)80074-N

71. Goodford PJ (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 28:849–857. https://doi.org/10.1021/jm00145a002

72. Kastenholz MA, Pastor M, Cruciani G, et al (2000) GRID/CPCA: A New Computational Tool To Design Selective Ligands. J Med Chem 43:3033–3044. https://doi.org/10.1021/jm000934y

73. Brady Jr. GP, Stouten PFW (2000) Fast prediction and visualization of protein binding pockets with PASS. Journal of Computer-Aided Molecular Design 14:383–401. https://doi.org/10.1023/A:1008124202956

74. Mezei M (2003) A new method for mapping macromolecular topography. Journal of Molecular Graphics and Modelling 21:463–472. https://doi.org/10.1016/S1093-3263(02)00203-6

75. Bhat J, Chatterjee S (2016) Skeleton selectivity in complexation of chelerythrine and chelerythrine-like natural plant alkaloids with the G-quadruplex formed at the promoter of c-MYC oncogene: in silico exploration. RSC Adv 6:36667–36680. https://doi.org/10.1039/C6RA04671A

76. Hernndez-Santoyo A, Yair A, Altuzar V, et al (2013) Protein-Protein and Protein-Ligand Docking. In: Ogawa T (ed) Protein Engineering - Technology and Application. InTech

77. Huang S-Y, Zou X (2010) Advances and Challenges in Protein-Ligand Docking. IJMS 11:3016–3034. https://doi.org/10.3390/ijms11083016

78. Liu M, Wang S (1999) MCDOCK: A Monte Carlo simulation approach to the molecular docking problem. Journal of Computer-Aided Molecular Design 13:435–451. https://doi.org/10.1023/A:1008005918983

79. Thormann M, Pons M (2001) Massive docking of flexible ligands using environmental niches in parallelized genetic algorithms. J Comput Chem 22:1971–1982. https://doi.org/10.1002/jcc.1146

80. Baxter CA, Murray CW, Clark DE, et al (1998) Flexible docking using tabu search and an empirical estimate of binding affinity. Proteins 33:367–382.

https://doi.org/10.1002/(SICI)1097-0134(19981115)33:3<367::AID-PROT6>3.0.CO;2-W

81. Tai HK, Jusoh SA, Siu SWI (2018) Chaos-embedded particle swarm optimization approach for protein-ligand docking and virtual screening. J Cheminform 10:62. https://doi.org/10.1186/s13321-018-0320-9

82. Muryshev AE, Tarasov DN, Butygin AV, et al (2003) A novel scoring function for molecular docking. J Comput Aided Mol Des 17:597–605. https://doi.org/10.1023/B:JCAM.0000005766.95985.7e

83. Tanchuk VY, Tanin VO, Vovk AI, Poda G (2016) A New, Improved Hybrid Scoring Function for Molecular Docking and Scoring Based on AutoDock and AutoDock Vina. Chem Biol Drug Des 87:618–625. https://doi.org/10.1111/cbdd.12697

84. Thomsen R, Christensen MH (2006) MolDock: A New Technique for High-Accuracy Molecular Docking. J Med Chem 49:3315–3321. https://doi.org/10.1021/jm051197e

85. Macari G, Toti D, Pasquadibisceglie A, Polticelli F (2020) DockingApp RF: A State-of-the-Art Novel Scoring Function for Molecular Docking in a User-Friendly Interface to AutoDock Vina. IJMS 21:9548. https://doi.org/10.3390/ijms21249548

86. Böhm H-J (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Computer-Aided Mol Des 8:243–256. https://doi.org/10.1007/BF00126743

87. Huang S-Y, Grinter SZ, Zou X (2010) Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. Phys Chem Chem Phys 12:12899. https://doi.org/10.1039/c0cp00151a

88. Charifson PS, Corkery JJ, Murcko MA, Walters WP (1999) Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. J Med Chem 42:5100–5109. https://doi.org/10.1021/jm990352k

89. Houston DR, Walkinshaw MD (2013) Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context. J Chem Inf Model 53:384–390. https://doi.org/10.1021/ci300399w

90. Kukol A (2011) Consensus virtual screening approaches to predict protein ligands. European Journal of Medicinal Chemistry 46:4661–4664. https://doi.org/10.1016/j.ejmech.2011.05.026

91. Chang MW, Ayeni C, Breuer S, Torbett BE (2010) Virtual Screening for HIV Protease Inhibitors: A Comparison of AutoDock 4 and Vina. PLoS ONE 5:e11955. https://doi.org/10.1371/journal.pone.0011955

92. Terp GE, Johansen BN, Christensen IT, Jørgensen FS (2001) A New Concept for Multidimensional Selection of Ligand Conformations (MultiSelect) and Multidimensional Scoring (MultiScore) of Protein−Ligand Binding Affinities. J Med Chem 44:2333–2343. https://doi.org/10.1021/jm001090l

93. Betzi S, Suhre K, Chétrit B, et al (2006) GFscore: A General Nonlinear Consensus Scoring Function for High-Throughput Docking. J Chem Inf Model 46:1704–1712. https://doi.org/10.1021/ci0600758

94. Wang R, Lai L, Wang S (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. Journal of Computer-Aided Molecular Design 16:11–26. https://doi.org/10.1023/A:1016357811882

95. Bar-Haim S, Aharon A, Ben-Moshe T, et al (2009) SeleX-CS: A New Consensus Scoring Algorithm for Hit Discovery and Lead Optimization. J Chem Inf Model 49:623–633. https://doi.org/10.1021/ci800335j

96. Teramoto R, Fukunishi H (2008) Structure-Based Virtual Screening with Supervised Consensus Scoring: Evaluation of Pose Prediction and Enrichment Factors. J Chem Inf Model 48:747–754. https://doi.org/10.1021/ci700464x

97. González MA (2011) Force fields and molecular dynamics simulations. JDN 12:169–200. https://doi.org/10.1051/sfn/201112009

98. Duan Y, Kollman PA (1998) Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution. Science 282:740–744. https://doi.org/10.1126/science.282.5389.740

99. Salmaso V, Moro S (2018) Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. Front Pharmacol 9:923. https://doi.org/10.3389/fphar.2018.00923

100. Wang J, Kollman PA, Kuntz ID (1999) Flexible ligand docking: A multistep strategy approach. Proteins 36:1–19. https://doi.org/10.1002/(SICI)1097-0134(19990701)36:1<1::AID-PROT1>3.0.CO;2-T

101. Panneerselvam S, Yesudhas D, Durai P, et al (2015) A Combined Molecular Docking/Dynamics Approach to Probe the Binding Mode of Cancer Drugs with Cytochrome P450 3A4. Molecules 20:14915–14935. https://doi.org/10.3390/molecules200814915

102. Shahbaaz M, Qari SH, Abdellattif MH, Hussien MA (2020) Structural analyses and classification of novel isoniazid resistance coupled mutational landscapes in *Mycobacterium tuberculosis* : a combined molecular docking and MD simulation study. Journal of Biomolecular Structure and Dynamics 1–10. https://doi.org/10.1080/07391102.2020.1861986

103. Brunger AT, Adams PD (2002) Molecular Dynamics Applied to X-ray Structure Refinement. Acc Chem Res 35:404–412. https://doi.org/10.1021/ar010034r

104. Linge JP, Williams MA, Spronk CAEM, et al (2003) Refinement of protein structures in explicit solvent. Proteins 50:496–506. https://doi.org/10.1002/prot.10299

105. Fan H (2004) Refinement of homology-based protein structures by molecular dynamics simulation techniques. Protein Science 13:211–220. https://doi.org/10.1110/ps.03381404

106. Goodsell DS, Olson AJ (1990) Automated docking of substrates to proteins by simulated annealing. Proteins 8:195–202. https://doi.org/10.1002/prot.340080302

107. Kmiecik S, Gront D, Kolinski M, et al (2016) Coarse-Grained Protein Models and Their Applications. Chem Rev 116:7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163

108. Wu G, Robertson DH, Brooks CL, Vieth M (2003) Detailed analysis of grid-based molecular docking: A case study of CDOCKER?A CHARMm-based MD docking algorithm. J Comput Chem 24:1549–1562. https://doi.org/10.1002/jcc.10306

109. Khrapak S, Klumov B, Couëdel L (2017) Collective modes in simple melts: Transition from soft spheres to the hard sphere limit. Sci Rep 7:7985. https://doi.org/10.1038/s41598-017-08429-5

110. Luitz MP, Zacharias M (2014) Protein–Ligand Docking Using Hamiltonian Replica Exchange Simulations with Soft Core Potentials. J Chem Inf Model 54:1669–1675. https://doi.org/10.1021/ci500296f

111. Doerr S, De Fabritiis G (2014) On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. J Chem Theory Comput 10:2064–2069. https://doi.org/10.1021/ct400919u

112. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4:435–447. https://doi.org/10.1021/ct700301q

113. Salomon-Ferrer R, Case DA, Walker RC (2013) An overview of the Amber biomolecular simulation package: Amber biomolecular simulation package. WIREs Comput Mol Sci 3:198–210. https://doi.org/10.1002/wcms.1121

114. Phillips JC, Braun R, Wang W, et al (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802. https://doi.org/10.1002/jcc.20289

115. Brooks BR, Brooks CL, Mackerell AD, et al (2009) CHARMM: The biomolecular simulation program. J Comput Chem 30:1545–1614. https://doi.org/10.1002/jcc.21287

116. Eastman P, Swails J, Chodera JD, et al (2017) OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput Biol 13:e1005659. https://doi.org/10.1371/journal.pcbi.1005659

117. Grindon C, Harris S, Evans T, et al (2004) Large-scale molecular dynamics simulation of DNA: implementation and validation of the AMBER98 force field in LAMMPS. Philosophical Transactions of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences 362:1373–1386. https://doi.org/10.1098/rsta.2004.1381

CHAPTER Ⅲ: K-means clustering analysis, ADME/Pharmacokinetic Prediction, MEP and Molecular docking studies of potential cytotoxic agents

III.1 Introduction

Computational and statistical tools are essential to analyze and manage big chemical databases with biological activity to extract new knowledge and make better decisions that guide chemists to synthesize and test only the most potent molecules. The application of virtual screenings and machine learning methods may be a promising strategy to identify new anticancer drug candidates.

In this study, we classified 500 potent cytotoxic agents according to their different scaffolds with the aim of identifying a specific class that has the best ADME profile to define a set of rules for structural features that would help in designing novel anticancer drug candidates. For this purpose, the k-means algorithm was applied on 500 cytotoxic agents to assign them into distinct clusters and determine the similarities and the dissimilarities within and between clusters. The test statistic Z_k was effectively used to characterize the compounds by descriptors and identify the subset of relevant descriptors characterizing each cluster. The silhouette analysis is used as cluster validation approach and a statistical analysis was performed to characterize each cluster by a subset of molecular descriptors. Additionally, the pharmacokinetic properties ADME and drug-likeness parameters were calculated for all molecules to evaluate their profile. The quantitative analysis of the molecular electrostatic potential (MEP) on van der Waals surface was performed to identify the nucleophilic and electrophilic sites in the paragon molecules. Finally, a molecular docking study was performed on the paragon molecules toward six different targets to investigate the molecular binding models for these compounds interacting with the active site of biological receptors.

III.2   Materials and methods

III.2.1   Biological database

A total of 500 cytotoxic molecules were collected for the first time from the previous studies [1-62] to create a database for analysis. All compounds have shown potent cytotoxicity toward a broad spectrum of human cancer cell lines *in vitro*.

III.2.2   Molecular descriptors generation

The compounds were initially optimized using the MM+ force field of molecular mechanics in HyperChem 8.08 software [63]. The descriptors mass, volume, log P, and surface grid were calculated by HyperChem 8.08. The software Molecular Operating Environment(MOE) [64] was used to determine the number of hydrogen donor, number of hydrogen acceptor, number of hydrogen donor and acceptor, the number of oxygen atoms, number of nitrogen atoms, number of sulfur atoms, the number of heteroatoms (oxygen, nitrogen, and sulfur), the number of chiral centers, dipolar moment, flexibility, number of rotatable bonds, topological polar surface area (TPSA), number of double bonds, and the number of rings. The polarizability, number of heterocyclic rings, and the number of aromatic rings were obtained by ChemAxon Marvin software [65].

III.2.3   ADME and drug-likeness prediction

In silico screening of pharmacological properties ADME and evaluation of drug-likeness were performed by the web tool SwissADME [66]. The pharmacokinetics parameters predicted in this study are: gastrointestinal absorption, P-glycoprotein, blood-brain barrier, and cytochrome enzymes were reported in this study. In addition, the drug-likeness prediction is based on several rules such as Lipinski, Ghose, Veber, Egan, and Muegge rules.

III.2.4   Molecular electrostatic potential

Molecular geometries of the paragon molecules were optimized by the Gaussian 09 program using the density functional theory (DFT) at B3LYP/6-31+G(d,p) level of theory [67]. For all stationary points, there is no imaginary frequency at the optimized molecular geometries indicating that the optimized structures are in the minimum on the potential energy surface. The multifunctional wavefunction analyzer program, Multiwfn 3.7 [68], was used for the quantitative analysis of molecular electrostatic potential on van der Waals surface combined with the VMD 1.9.1 software [69].

III.2.5 Molecular docking study

III.2.5.1 Compounds preparation

In this study, we selected the paragon molecules (P1, P2, and P3) that displayed potent cytotoxic activity against six cancer cell lines [44, 32, 14] for docking analysis. These molecules were used to create a database input in the MOE software.

III.2.5.2 Targets preparation

The X-ray crystal structures of enzymes, A-549 (PDB ID: 2XP2) [70], MCF-7 (PDB ID: 1UOM) [71], Bel7402 (PDB ID: 1CP3) [72], SGC7901 (PDB ID: 2HY8) [73], Hela (PDB ID: 3F81) [74], and HepG2 (PDB ID: 4ASD) [75] were downloaded from the Protein Data Bank (http://www.rcsb.org/pdb/) to study the Interactions of compounds with the active site of receptors. Additionally, the water molecules, ions, and cofactors were removed from the structures to simplify the simulation. Afterward, the protein structures were corrected for missing atoms and the hydrogen atoms were added [76].

III.2.6 Data clustering

K-means is a simple and efficient clustering algorithm widely used in machine learning to partition a set of objects into K clusters. The algorithm starts by choosing randomly K initial centroids, where a cluster centroid is the mean of all data points within a cluster. Afterward, each point in the dataset is assigned to the nearest center by minimizing the distance between data points and the corresponding centroids. After the clusters are formed, the algorithm iterates this process until no point changes clusters [77]. In clustering algorithms, the random selection of initial cluster centers is crucial to the result that may depend on this step; a wrong estimation of K can lead to misclassification of observations. In this study, the Hartigan–Wong algorithm is used [78]. This algorithm generally does a better job than either of MacQueen [79], Lloyd [80], and Forgy algorithms [81], but trying several random starts is often recommended [82]. As the k-means clustering algorithm is sensitive to the random starting assignments, we specify 50 (nstart = 50) different random starting assignments, in order to have a more stable result. The maximum number of iterations allowed is used by default (equal to 10). Euclidean distances are used as a metric for calculating dissimilarities between observations.

For this purpose, determining the optimal number of clusters was performed using the average silhouette method [83]. The principle of this approach involves calculating the

average silhouette for each k partition; the optimal number of clusters corresponds to the maximum of the average silhouette. In this study, principal component analysis was performed as a preprocessing to provide more stable clusters. Furthermore, the silhouette analysis is used as a cluster validation approach in order to confirm that there are no misallocations or overlapping clusters. We used the FactoMineR R package [84] to compute principal components and the cluster R package [85] was used to perform the K-means algorithm.The Hopkins statistic [86] has been carried out using the Clustertend R package for evaluating the clustering tendency of our database. More the value of this statistic is close to 1 means that the dataset contains statistically significant clusters [86].

III.2.7 Clusters characterization by descriptors and molecules

In order to characterize each cluster by a subset of molecular descriptors, the mean of each molecular descriptor belonging to cluster k is computed and compared to the overall mean in the dataset according to statistics [87]:

$$Z_k = \frac{\bar{x}_k - \bar{x}}{\sqrt{\frac{s_k^2}{n_p}\left(\frac{n - n_p}{n - 1}\right)}} \qquad (1)$$

Where $\bar{x}_k$ is the mean of descriptor x in cluster k, $\bar{x}$ (s_k) is the mean (standard deviation) of descriptor x in the dataset and n_p is the cardinal of cluster k. The p-value or the critical probability is computed as follows:

$$p.value = 1 - P\left(|Z| < |Z_k|\right) = 1 - \Phi(|Z_k|), \quad Z \sim N(0,1) \qquad (2)$$

Where $\Phi(|Z_k|)$ is the cumulative distribution function of the standard normal distribution. The p.value is computed under the null hypothesis$(H_0 : \bar{x}_k - \bar{x} = 0)$. This hypothesis is satisfied if $1 - \Phi(|Z_k|)$ is strictly greater than the significance level of 0.05. On the other hand, the alternative hypothesis $(H_1 : \bar{x}_k - \bar{x} \neq 0)$ postulates that there is a significant difference between $\bar{x}_k$ and $\bar{x}$. In this case, the null hypothesis is rejected at the 5% level of significance if$1 - \Phi(|Z_k|) \leq 0.05$, which means that the corresponding molecular descriptor characterizes the kth cluster. More details related to this test statistic can be found at [88].

Moreover, the cluster centroid is a fictional mean observation, to understand the average behavior of molecules in each cluster, it is better to use a real observation. The closest observation to the barycenter of each cluster is called paragon and defined according to [89] :

$$I_k \in \min_{i \in k} dist(I_i, \mu_k) \qquad (\mathbf{3})$$

Where μ_k is the barycenter of cluster k and I_i is the ith observation. Therefore, we can describe the overall behavior of the molecules in each cluster by their respective paragons.

III.3 Results and discussion

III.3.1 Clusters analysis

We have identified three distinct and compact clusters using the K-means algorithm. The number of molecules in each cluster is 209, 104, and 105 for cluster 1, cluster 2, and cluster 3, respectively, as depicted in (Figure III. 1). As shown in (Table III. 1), the optimal number of clusters identified using the average silhouette method is K = 3. The values of the Hopkins statistic obtained are 0.637 and 0.922 for the dataset with and without the outliers, respectively. As can be seen, the dataset without outliers is highly clusterable. The presence of outliers (even in small numbers) tends to significantly reduce the tendency of the dataset to form clusters.
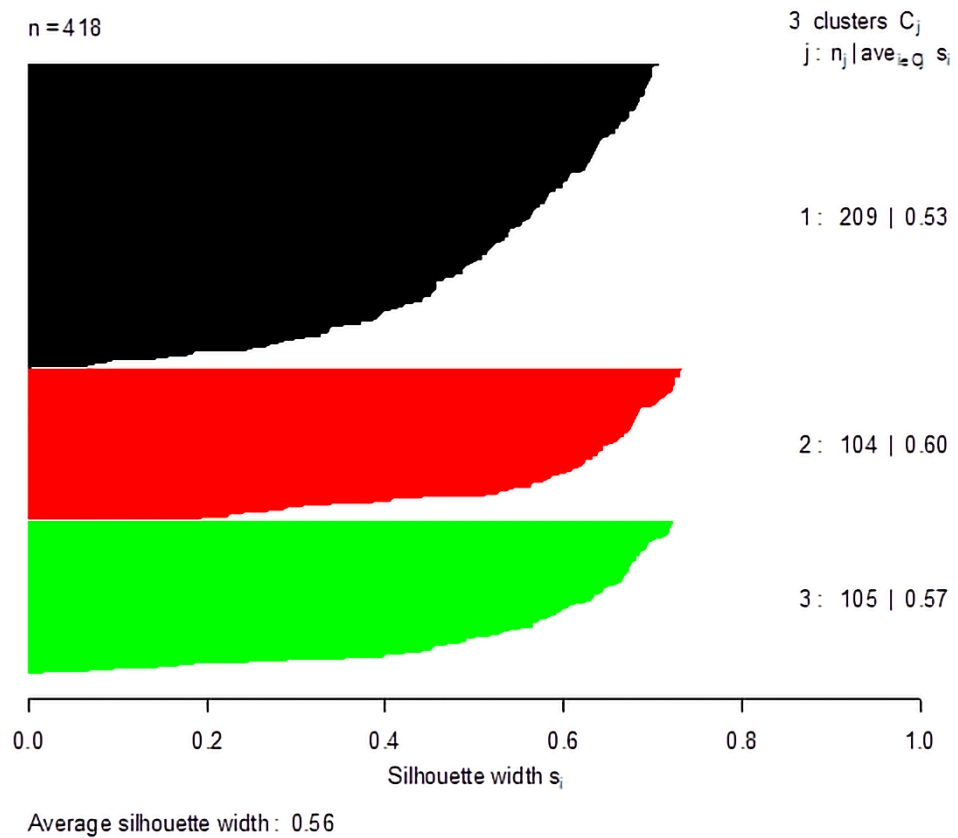
Figure III. 1 Clusters silhouette plot

Silhouette averagewidth	K
0.4954795	2
0.565419	3
0.4522308	4
0.4429059	5
0.446656	6

Indeed, adding another cluster decreases the average silhouette. Figure Ⅲ. 2 shows the partition on the principal component map. As can be seen from this figure, there is a clear separation between them and there is no overlap. The percentage of molecules in each cluster is 50%, 24.88% and 25.12% for the black, red, and green clusters, respectively.
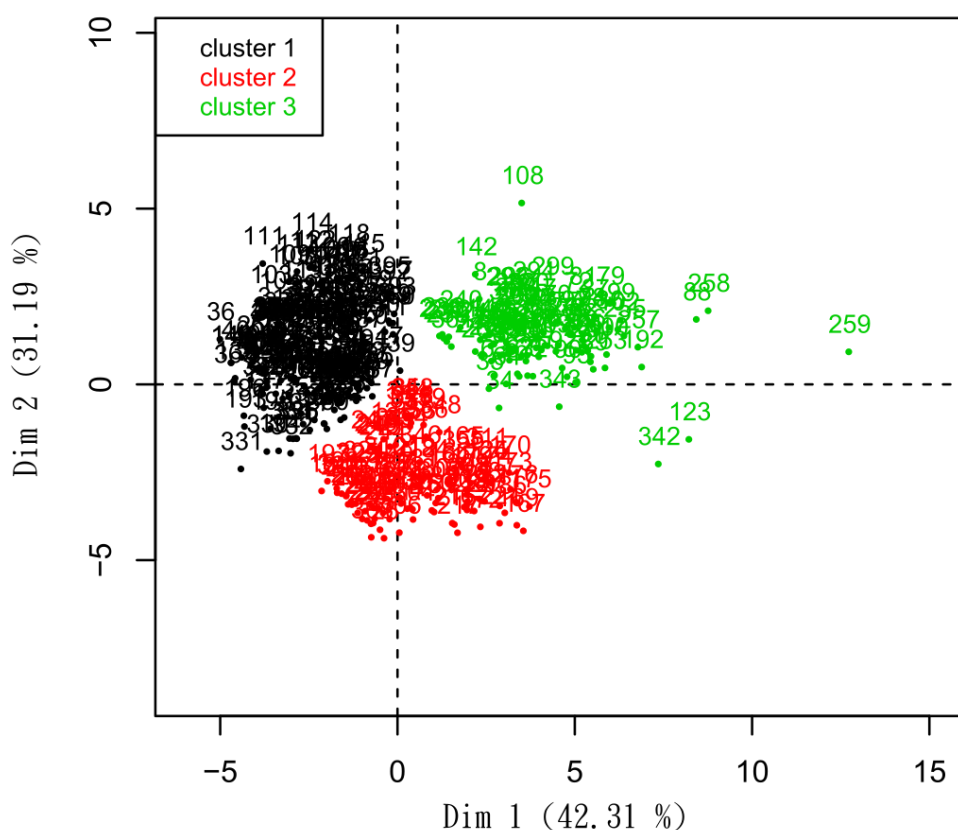
Figure Ⅲ. 2 Partition on the principal component map

In addition, the quality of K-means clustering has been validated using silhouette analysis. As shown in (Figure Ⅲ. 1), the total of molecules is 418 because we removed 82 molecules considered as outliers. Silhouette average widths of these 82 molecules are

negative. The positive values of silhouette width in each cluster are all higher than 0.5 (0.53 for cluster 1, 0.60 for cluster 2 and 0.57 for cluster 3), as well as the average width with a value of 0.56, implies that the molecules are very well clustered and there are no misallocations or overlapping clusters.

The statistical results for all clusters are listed from Table A1 to Table A3 (see Appendix A). In order to analyze and understand why the descriptors were divided into three clusters, we have classified them into three categories based on their descriptive power (Table III. 2).

Table III. 2 z_k-statistic values obtained for the three clusters

Category	Descriptor	Cluster 1	Cluster 2	Cluster 3
First category	Oxygen atoms	-11.920292	/	13.360962
	Chiral centers	-12.756367	13.927741	/
	aromatic rings	/	-12.165403	9.824948
	Nitrogen atoms	/	-9.449154	10.914702
	Sulfur atoms	/	-6.259531	6.626399
	Hydrogen donor	-3.143361	/	4.513423
	Dipolar moment	/	-2.231093	3.365505
Second category	Rotatable bonds	-12.628486	2.139250	12.478667
	Number of rings	-12.129848	3.279142	10.769558
	Hydrogen acceptor	-11.266461	-3.225036	16.232926
	Hydrogen donor + acceptor	-10.244272	-2.925194	14.752944
	TPSA	-9.473377	-3.601449	14.533308
	Double bonds	-8.385037	6.259117	3.478373
Third category	Volume	-17.358715	9.468625	10.667217
	Polarizability	-17.029683	8.585525	11.164061
	Hetero-atoms N+O+S	-9.783489	-5.674194	16.951175
	Log p	-6.551781	16.731131	-9.045471
	Mass	-16.565628	6.125284	13.071631
	s-grid	-16.519684	8.516432	10.642925
	Flexibility	-15.714565	8.812227	9.417991
	heterocyclic rings	-4.348880	-7.219752	12.201912

The first category of molecular descriptors includes those contributing differently to the three clusters. For instance, the descriptor number of oxygen atoms contributes negatively to the first cluster which means, it shows values less than the overall mean of oxygen atoms. On the other hand, this descriptor contributes positively for the third cluster indicating higher values than the overall mean of all other molecules. Concerning the second cluster, the descriptor exhibits a value very close or equal to the overall mean. Therefore, the descriptor number of oxygen atoms does not distinguish between the molecules belonging to the second cluster. The descriptors classified in the second category are those contributing positively or negatively with a high magnitude to two clusters and exhibit a value very close to the overall mean for the remaining cluster. For example, the descriptor rotatable bonds contribute negatively to the first cluster, which means that it has negative values that are significantly lower than the overall mean for all molecules. It contributes positively for the third cluster indicating values significantly higher than the overall mean for all other molecules. This descriptor characterize very slightly the second cluster because it shows a very close value to the overall mean, this category is quite similar to the previous one.

Using the same approach, the third category contains the descriptors that differentiate one cluster from the other two clusters such as the lipophilicity log P, contributing positively to the second cluster and negatively for the other two clusters. This result suggests that the molecules belonging to the second cluster are more lipophilic than those of cluster 1 and 3. As can be seen in(Table Ⅲ. 1), it emerges that the molecular descriptors contribute negatively for the first cluster, positively for the third except for Log P and both positively and negatively for the second cluster. These findings confirm the agreement between the results derived from the statistical analysis and those obtained from the K-means algorithm. Furthermore, for a given descriptor, the higher its z_k value, the higher is its influence on the cluster k molecules so we can select the most relevant descriptor that describes each cluster. As a result, cluster 1 molecules have a small volume and mass. They have very few oxygen atoms and consequently low value of polarizability and topological polar surface area (TPSA). For cluster 2, the lipophilicity is significantly high, as well as the number of chiral centers. A large number of heteroatoms in cluster 3 would necessarily increase the number of hydrogen donors and acceptors, polarizability, and TPSA.

Interestingly, the representative (paragon) molecules of clusters shown in (Figure Ⅲ . 3) have confirmed the reliability of z_k values since there are a clear resemblance and

compatibility between them. Paragon 1 (P1) has a small volume and few of oxygen atoms, while paragon 2 (P2) has high lipophilicity, a lot of chiral centers and contains few heteroatoms. Paragon 3 (P3) has a large number of heteroatoms and subsequently a lot of hydrogen donors and acceptors. These results emphasize the validity of our test statistic and provide evidence that the molecules are very well clustered.
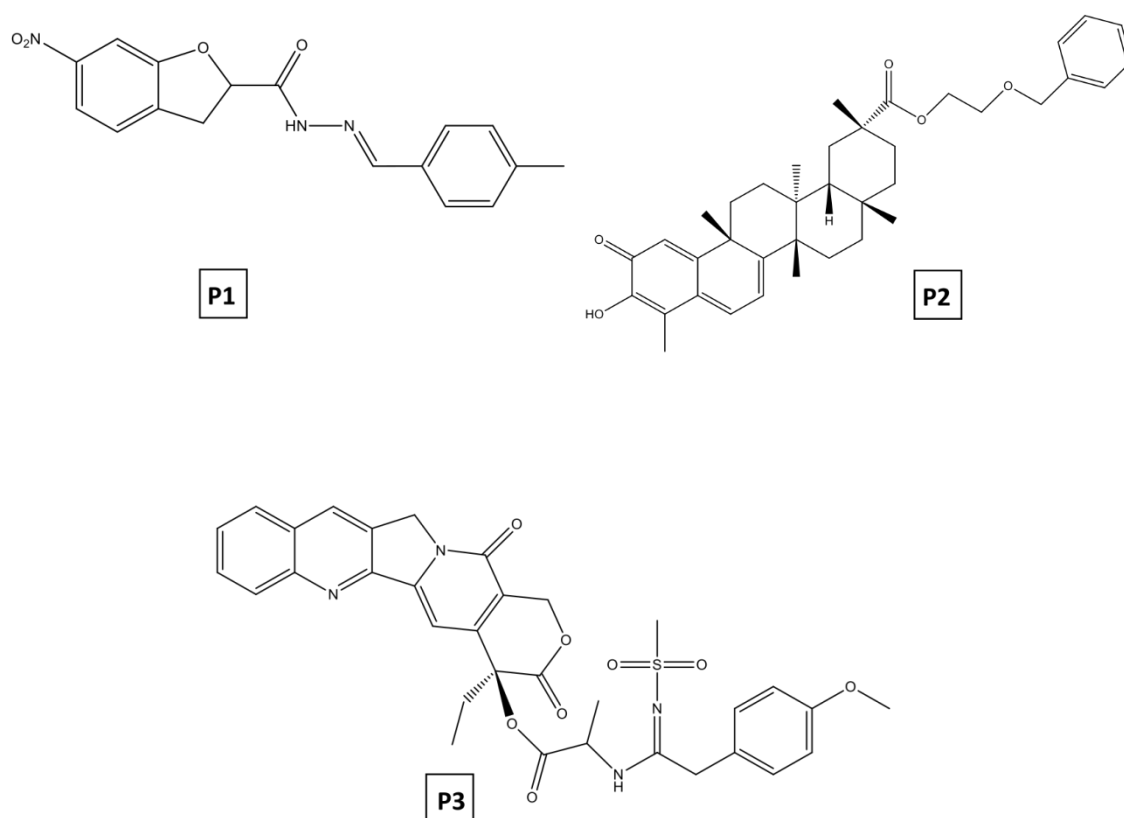
Figure Ⅲ. 3 Core structures of clusters

III.3.2 ADME properties and drug-likeness evaluation

In silico screening of pharmacological properties ADME and evaluation of drug-likeness of 418 cytotoxic molecules are summarized in (Table Ⅲ. 3). As can be seen in (Table Ⅲ. 3), almost all cluster 1 molecules are estimated to have high gastrointestinal absorption and almost half of them are capable of crossing the blood-brain barrier. In contrast, the other cluster molecules have a low GI absorption, therefore poor bioavailability, and they cannot penetrate the BBB at all. It can be explained by the large size of molecules and the high polarity of cluster 3 molecules as previously mentioned in Z_k-test results. The findings are

directly in line with previous research showing that hydrophilic and large molecules are unable to cross the lipid cell membranes [90, 91]. Thus, a good balance between lipophilic and hydrophilic characters in drug molecules is indispensable for intestinal absorption. These molecules are expected to be eliminated before reaching the target site.

Table Ⅲ. 3 ADME properties and drug-likeness of cytotoxic compounds

Properties		cluster 1	cluster 2	cluster 3
Pharmacokinetics	GI absorption	85.17% High	76.92% Low	80.95% Low
	BBB permeant	52.63% No	100% No	99.05% No
	P-gp substrate	71.29% No	70.19% Yes	59.05% Yes
	CYP1A2 inhibitor	69.38% Yes	98.08% No	68.57% No
	CYP2C19 inhibitor	63.16% Yes	93.27% No	78.10% No
	CYP2C9 inhibitor	71.77% Yes	79.81% No	61.90% Yes
	CYP2D6 inhibitor	71.29% No	98.08% No	81.90% No
	CYP3A4 inhibitor	55.98% Yes	61.5% No	50.48% Yes
Drug likeness	Lipinski	92.82% Yes	66.35% No	65.71% No
	Ghose	73.21% Yes	98.08% No	98.10% No
	Veber	93.78% Yes	79.81% Yes	84.76% No
	Egan	85.65% Yes	86.50% No	76.19% No
	Muegge	74.16% Yes	87.50% No	91.43% No

Almost all cluster 2 and 3 molecules are estimated to be actively effluxed by P-glycoprotein transporter, thus limiting the intestinal absorption and consequently low bioavailability. Nevertheless, cluster 1 molecules are not affected by p-gp efflux pump that helps to overcome multidrug resistance and effectively deliver the anticancer drugs into the tumors [92, 93]. Moreover, nearly all molecules of cluster1 are estimated to inhibit the different CYP isoforms except for CYP2D6 and CYP3A4. These two isoenzymes may be responsible for the metabolic clearance or activation of this fraction of molecules. Accordingly, it has been shown that CYP2D6 and CYP3A4 converse the anticancer drug tamoxifen into more potent metabolites. The metabolite endoxifen is primarily formed by CYP2D6 has approximately 100 times better affinity to the estrogen receptor than the parent drug tamoxifen [94]. On the other hand, almost all compounds of other clusters are estimated to act as non-inhibitors toward CYP isoenzymes, except for CYP2C9 and CYP3A4 in cluster 3. The second cluster molecules showed no selectivity in the interaction with CYP cytochromes compared to the other clusters. This finding supports previous studies in which the high lipophilicity was associated with poor selectivity of CYP [95]. Besides, nearly all compounds of cluster 1 proved to comply with the Lipinski rules and successfully passed the Veber, Egan, Ghose, and Muegge filters. Meanwhile, the molecules of other clusters showed poor drug-likeness as they violate the five filter rules. It is interesting to note that only 50 % of cytotoxic agents (cluster 1 molecules) reported in the last decade showed good ADME profile and drug-likeness. This finding reveals the need to improve the ADME behaviors and consequently bioavailability of anticancer drug candidates to avoid clinical failures.

For further investigation, the percentages of 62 cancer cell lines in each cluster were calculated as depicted in (Figure Ⅲ. 4). The highest percentages of database molecules showed cytotoxic activity toward A549 (lung cancer) and MCF-7 (breast cancer). A total of 44.74 % of compounds are cytotoxic against A549, appearing in all clusters with 15.79 %, 9.09 %, and 19.86 % for cluster 1, 2, and 3, respectively, and 33.73 % are active against MCF-7 with 20.81 %, 5.98 %, and 6.94 % for cluster 1, 2, and 3, respectively. Therefore, lung and breast cancers considered the most common types of cancer, with each contributing about 2.09 million cases in 2018, according to estimates from the World Health Organization (WHO) [96]. Besides, HepG2, Hela, and KB showed good ratios, as 22 % of molecules exhibit cytotoxicity toward HepG2; 11.24 % from cluster 1 and 10.53 % from cluster 2. In total, 19.38 % of molecules are cytotoxic toward Hela with 7.42%, 7.66%, and 4.31 % for

cluster 1, 2 and 3, respectively and 19.38 % against KB and the highest percentage observed in cluster 3 by 16.27 %.
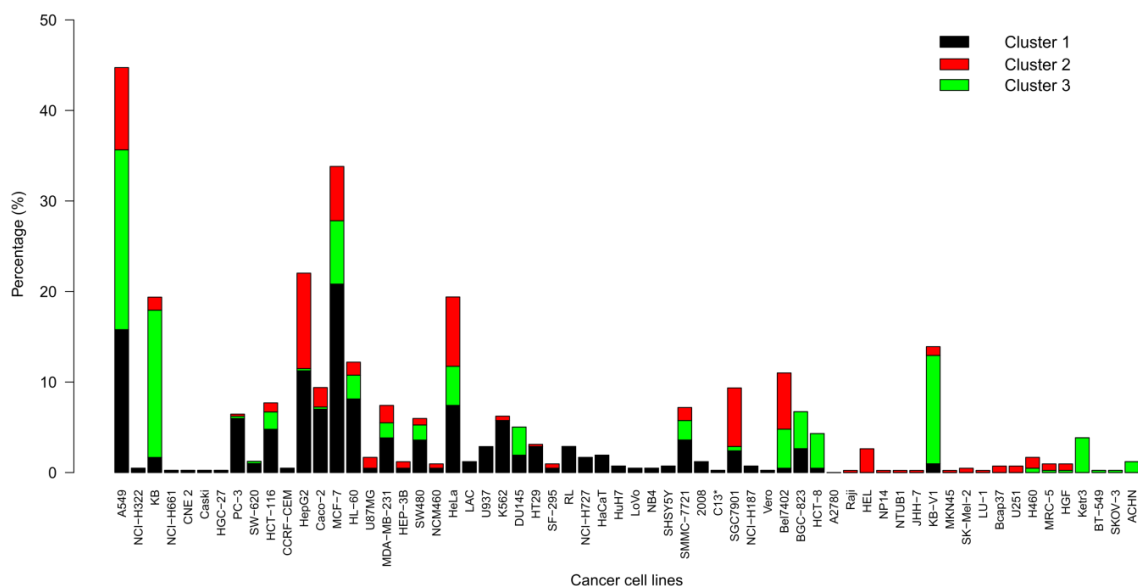
Figure Ⅲ. 4 Types of cancer cell lines in each cluster

Overall, these results indicate that recent research is oriented toward the five cancer cell lines A549, MCF-7, HepG2, Hela and KB as they represent the most common cancer types worldwide.

III.3.3 Quantitative MEP analysis of paragon molecules

The map of ESP (Electrostatic Potential) van der Waals surfaces of the paragon molecules (P1, P2, and P3) is shown in (Figure Ⅲ. 5). By examining these surfaces, it is clear that the three compounds have large positive surfaces which occupy 63 %, 71 %, and 64% respectively of the overall surface. Thus, the electrophilic character is higher in them, particularly for paragon 2. These electron-deficient regions are susceptible to be attacked by nucleophilic reagents. As shown in (Figure Ⅲ. 5), the highest charge density fluctuation was recorded in paragon 3 which represents the overall variance (OV) of electrostatic potential on van der Waals surface, with a value of 321.59 $[kcal/mol]^2$. For paragon 1, the global surface minimum is observed around the nitro substituent indicates an electron-rich site, vulnerable for electrophilic attack. Meanwhile, the global surface maximum near the nitrogen's hydrogen reflects the low electronic density in this area justified by the withdrawing effect of amide carbonyl and the nitrobenzofuran system.
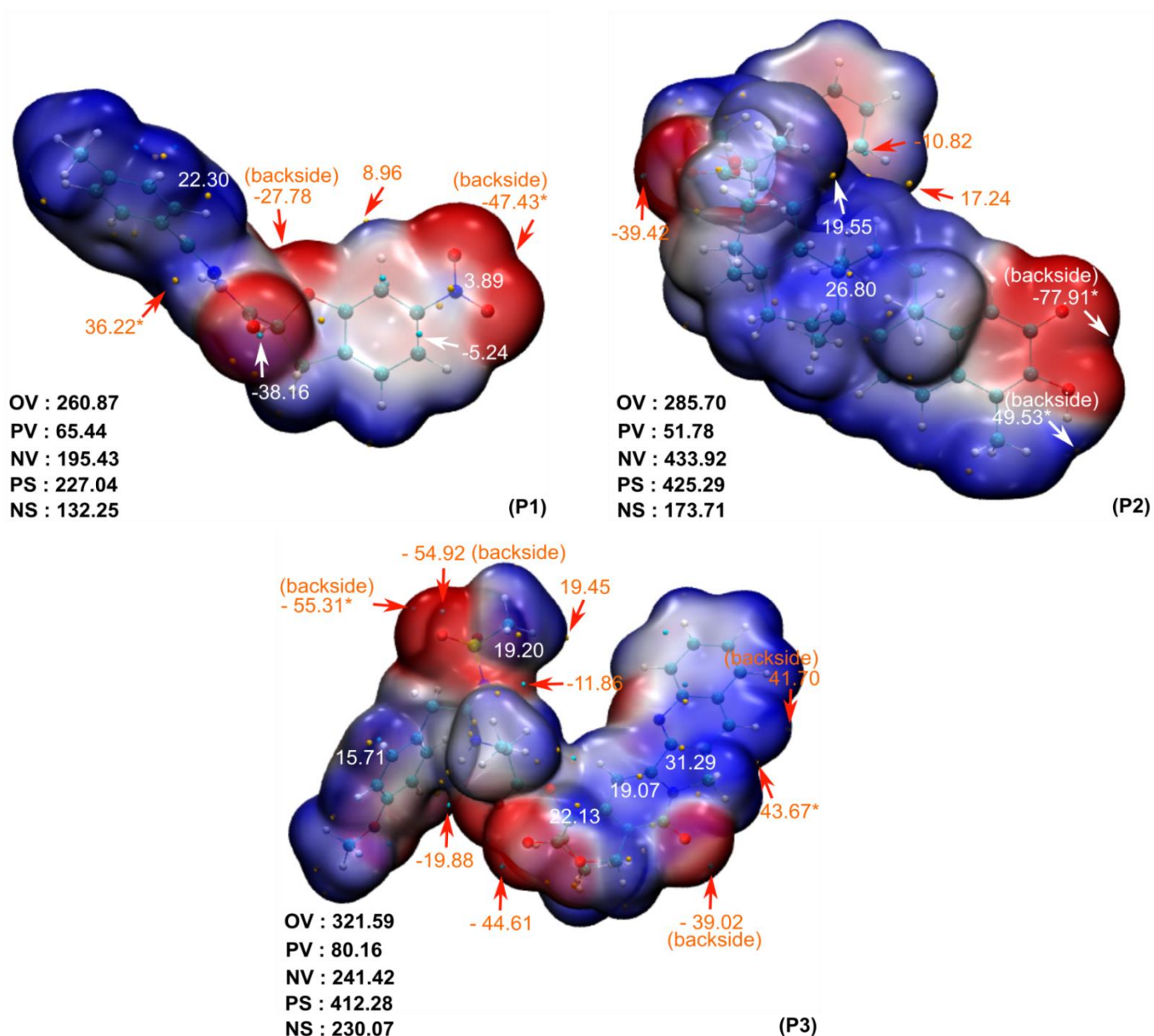
Figure Ⅲ. 5 ESP-mapped van der Waals surfaces (kcal/mol) of P1, P2 and P3 molecules, using a color scale ranging from red (negative ESP) through white (neutral ESP) to blue (positive ESP). The blue regions indicate a vulnerable site for nucleophilic attack, and the red regions are sites for electrophilic attack. All the iso-surface maps were rendered by VMD software based on the surface analysis result of Multiwfn program. The grid spacings were set to 0.2 Bohr and the van der Waals surface denotes the iso-surface of $\rho = 0.001\ e/bohr^3 \equiv a.u.$ Value with a star indicates the global extrema. The bold numbers in the bottom left-hand corner are the overall ESP variance (OV), positive variance (PV), negative variance (NV), positive surface area (PS), and negative surface area (NS) whose units are [Kcal/mol]^2, Å^2, respectively

The lone pairs of carbonyl and hydroxyl oxygen atoms result in recording the most negative surface potential in paragon 2. However, the global positive ESP is associated with the hydrogen of hydroxyl function, indicating potential hydrogen bond donor. The minimum potential surface surrounding the carbonyl oxygen of the ester group reveals the high electronic density in this region. The global minimum in paragon 3 is located around the carbonyl oxygen atoms of the sulfonyl group. On the other hand, the global maximum is observed nearby H-4 of quinoline moiety. For all compounds, the nucleophilic and electrophilic reactions in the white-colored regions (neutral ESP value) are unfavorable, as compared with the red and blue areas.

III.3.4 Molecular docking simulation

According to Clément *et al* [97], a resolution of protein structures between 1.5 and 2.5 Å is considered a good quality for docking studies; thus, all targets resolution belong to this interval. We started with the validation of enzyme by re-docking the native ligand of each target into the active site and calculate the root-mean-square deviation (RMSD) and the binding free energy (S score) for each complex (Table Ⅲ. 4). This allows the comparison of affinity between the compounds and the targets.

Table Ⅲ. 4 Properties of enzymes, energy score and RMSD values

Enzymes ID	Amino acids number	Resolution (Å)	S score (kcal/mol)	RMSD (Å)
2XP2	327	1.90 Å	-7.5839	2.8304
1UOM	254	2.28 Å	-9.9366	0.6792
1CP3	277	2.30 Å	-9.3879	0.8779
2HY8	297	2.00 Å	-8.0702	0.5713
3F81	183	1.90 Å	-4.6782	2.8403
4ASD	353	2.03 Å	-9.9430	0.8780

Afterward, we docked three compounds into the active site cavity of six enzymes by using the Molecular Operating Environment software (MOE). Finally, ten top conformations were obtained for each compound and the optimal geometry of compound was selected based on their energy score. The results of docking calculation of energy scores, half maximal

inhibitory concentration (IC_{50}) values, type of interactions, and distances between (P1, P2, and P3) compounds and six targets: A-549 (PDB ID: 2XP2), MCF-7 (PDB ID: 1UOM), Bel7402 (PDB ID: 1CP3), SGC7901 (PDB ID: 2HY8), Hela (PDB ID: 3F81), HepG2 (PDB ID: 4ASD), which are summarised in (Table Ⅲ. 5). As shown in (Table Ⅲ. 5), all compounds form interactions with 2XP2 with a score from -6.11 to-7.58 kcal/mol. P3 has the best energy score of -7.09 kcal/mol and the closest to the native ligand. Imberty *et al* [98], have shown that the distances of hydrogen bonds between 2.5 and 3.1 Å are considered strong interaction and 3.1 to 3.55 Å are weak interaction. P3 establishes a weak hydrogen acceptor bond with the residues of active site ALA 1200 (3.33 Å) as illustrated in Figure A2 (see Appendix A). While P2 shows a strong hydrogen acceptor bond with ARG 1253 (2.89 Å) and P1 have a weak hydrogen donor bond with ASP 1270 (3.26 Å).

Table Ⅲ. 5 Docking score and interactions between compounds and the active site residues of six targets

Compounds	S (Kcal/mol)	IC_{50} (μM)	Compound atoms	Receptor atoms	Receptor residues	Type of bond interaction	Distance (Å)	Energies (Kcal/mol)
A-549 (PDB ID: **2XP2**)							
Native ligand	-7.58	/	N22 23	O	GLU 1197	H-donor	3.32	-1.5
			N23 26	N	MET 1199	H-acceptor	3.23	-3.8
P1	-6.11	/	O 1 6-ring 6-ring	OE2	GLU 353	H-donor	2.94	-5.6
P2	-6.88	3.87	O 23	NE	ARG 1253	H-acceptor	2.89	-3.3
P3	-7.09	0.061	O 33	CA	ALA 1200	H-acceptor	3.33	-0.7
			6-ring	CD2	LEU 1122	pi-H	4.45	-0.7
MCF-7 (PDB ID: **1UOM**)							
Compounds	**S (Kcal/mol)**	**IC_{50} (μM)**	**Compound atoms**	**Receptor atoms**	**Receptor residues**	**Type of bond interaction**	**Distance (Å)**	**Energies (Kcal/mol)**
Native ligand	-9.94	/	N1 10	OD1	ASP 351	H-donor	2.86	-13
			O53 53	OE1	GLU 353	H-donor	3.12	-0.6
			O53 53	OE2	GLU 353	H-donor	3.02	-3.9
			N1 10	OD1	ASP 351	Ionic	2.86	-5.5
			N1 10	OD2	ASP 351	Ionic	3.85	-0.8
			C60 60	5-ring	HIS 524	H-pi	4.23	-0.6
P1	-6.43	7.5	O 1	OE2	GLU 353	H-donor	2.94	-5.6
P2	-6.88	/	/	/	/	/	/	/
P3	-5.37	/	/	/	/	/	/	/

Bel7402 (PDB ID: 1CP3)							
Compounds	**S (Kcal/mol)**	**IC$_{50}$ (μM)**	**Compound atoms**	**Receptor atoms**	**Receptor residues**	**Type of bond interaction**	**Distance (Å)**	**Energies (Kcal/mol)**
Native ligand	-9.39	/	N 7	O	PHE 250	H-donor	3.15	-3.6
			N 19	O	ARG 207	H-donor	2.88	-5.9
			N 45	O	SER 205	H-donor	3.40	-1.8
			OD1 17	ND2	ASN 208	H-acceptor	3.35	-1.1
			OD1 17	NE1	TRP 214	H-acceptor	3.08	-3
			OD2 18	N	PHE 250	H-acceptor	3.08	-5.6
			O 24	N	ARG 207	H-acceptor	3.14	-1.1
			O 50	N	GLY 122	H-acceptor	3.48	-1.1
			OD1 55	NE	ARG 64	H-acceptor	2.88	-2.8
			OD1 55	NE	ARG 207	H-acceptor	2.91	-6.3
			OD1 55	NH2	ARG 207	H-acceptor	2.83	-7.3
			OD2 56	NH2	ARG 64	H-acceptor	3.15	-5.6
			OD2 56	NE2	GLN 161	H-acceptor	2.89	-6.9
			OD1 55	NE	ARG 64	Ionic	2.88	-5.3
			OD1 55	NH2	ARG 64	Ionic	3.37	-2.4
			OD1 55	NE	ARG 207	Ionic	2.91	-5.1
			OD1 55	NH2	ARG 207	Ionic	2.83	-5.7
			OD2 56	NE	ARG 64	Ionic	3.91	-0.7
			OD2 56	NH2	ARG 64	Ionic	3.15	-3.5
P1	-5.51	/	/	/	/	/	/	/
P2	-7.09	0.51	6-ring	CE1	TYR 204	pi-H	4.32	-0.8
P3	-7.35	/	O 22	N	PHE 252	H-acceptor	3.3	-0.8
SGC7901 (PDB ID: 2HY8)							
Compounds	**S (Kcal/mol)**	**IC$_{50}$ (μM)**	**Compound atoms**	**Receptor atoms**	**Receptor residues**	**Type of bond interaction**	**Distance (Å)**	**Energies (Kcal/mol)**
Native ligand	-8.07	/	N19 26	O	GLU 345	H-donor	3.14	-3
			O30 45	N	LEU 347	H-acceptor	3.02	-4.1
			5-ring	CG2	VAL 284	pi-H	4.05	-1.2
P1	-5.63	/	/	/	/	/	/	/
P2	-6.81	0.3	O 23	OH	TYR 346	H-acceptor	3	-0.8
			O 37	OG	SER 281	H-acceptor	3.01	-1
P3	-7.38	/	O 33	SD	MET 344	H-donor	3.85	-1
			O 34	OG1	THR 406	H-acceptor	2.83	-1.8
			6-ring	CA	GLY 277	pi-H	4.22	-0.8
Hela (PDB ID: 3F81)							
Compounds	**S (Kcal/mol)**	**IC$_{50}$ (μM)**	**Compound atoms**	**Receptor atoms**	**Receptor residues**	**Type of bond interaction**	**Distance (Å)**	**Energies (Kcal/mol)**
Native ligand	-4.68	/	SAE 15	OD2	ASP 92	H-donor	3.48	-12
			OAD 1	NH2	ARG 130	H-acceptor	2.85	-7.1
			OAB 3	N	ARG 130	H-acceptor	3.32	-2
			OAB 3	NE	ARG 130	H-acceptor	3.03	-4.8
			OAC 4	N	TYR 128	H-acceptor	3.15	-1.4
			OAC 4	N	SER 129	H-acceptor	3.17	-2.3
			OAD 1	NE	ARG 130	Ionic	3.98	-0.5
			OAD 1	NH2	ARG 130	Ionic	2.85	-5.5
			OAB 3	NE	ARG 130	Ionic	3.03	-4.3
			OAB 3	NH2	ARG 130	Ionic	3.53	-1.8

P1	-4.71	/	O 1	OD2	ASP 92	H-donor	3.14	-2.5
			6-ring	N	TYR 23	pi-H	4.44	-0.7
P2	-5.15	1.12	6-ring	CG	GLU 126	pi-H	3.63	-0.7
P3	-6.16	/	O 23	N	ASP 164	H-acceptor	3.07	-1.6
HepG2 (PDB ID: **4ASD**)								
Compounds	**S (Kcal/mol)**	**IC$_{50}$ (µM)**	**Compound atoms**	**Receptor atoms**	**Receptor residues**	**Type of bond interaction**	**Distance (Å)**	**Energies (Kcal/mol)**
Native ligand	-9.94	/	N12 40	OE2	GLU 885	H-donor	2.75	-5.5
			N14 42	OE2	GLU 885	H-donor	3.19	-0.8
			O15 44	N	ASP 1046	H-acceptor	2.85	-2.3
			N26 46	N	CYS 919	H-acceptor	3.32	-2.2
P1	-7.22	/	O 1	OE2	GLU 885	H-donor	2.68	-5.9
			6-ring	CB	LEU 840	pi-H	4.12	-0.6
			6-ring	CD2	LEU 840	pi-H	4.20	-0.8
			6-ring	CE	LYS 868	pi-H	3.82	-0.7
P2	-7.23	0.37	C 41	6-ring	PHE 1047	H-pi	3.76	-0.6
P3	-5.81	/	O 22	NH1	ARG 1027	H-acceptor	2.59	-3.7

In the second case, we can see that P1 is the only one that has interactions with 1UOM with an energy score of -6.43 kcal/mol which is a very low value compared to the native ligand that gives -9.94 kcal/mol. Figure A3 (see Appendix A) revealed that P1 forms strong H-donor interaction with GLU 353 (2.94 Å). Likewise, strong H-donor interaction with GLU 353 (3.02 Å) is observed for the native ligand. In the third case, P3 gave the best S score when binding with 1CP3 of -7.35kcal/mol. This compound has a weak H-acceptor interaction with PHE 252 (3.30Å). The amino acid TYR 204 forms a pi-H interaction with P2, while P1 shows no interactions with this receptor Figure A4 (see Appendix A).

The binding mode of P3 with 2HY8 shows a weak H-donor interaction with MET 344 (3.85 Å), strong H-acceptor interaction with THR 406 (2.83 Å). This compound has the lowest energy score of -7.38 kcal/mol compared to other complexes. P2 establishes two strong hydrogen bonds, the first one with TYR 346 (3.00 Å) and the second with SER 281 (3.01 Å), whereas P1 has no interactions Figure A5 (see Appendix A). P3 forms a strong H-acceptor interaction with ASP 164 (3.07 Å) of the 3F81 receptor. This compound has the best score of -6.16 kcal/mol which is lower than the native ligand of -4.68kcal/mol. On the other hand, P1 is involved in a weak H-donor interaction with ASP 92 (3.14 Å). Similarly, weak H-donor interaction with ASP 92 (3.48 Å) is observed for the native ligand as illustrated in Figure A6 (see Appendix A). As shown in Figure A7 (see Appendix A), P1 establishes a strong H-donor interaction with GLU 885 (2.68 Å) with the active site residues of 4ASD.

Two H-donor interactions were noticed for the native ligand, while P2 has the best binding free energy of -7.23 kcal/mol. Moreover, P3 formed a strong H-acceptor interaction with ARG 1027 (2.59 Å).

Overall, these results match those observed in the MEP analysis that revealed the most reactive sites, indicating that the nitro group in P1 is susceptible to electrophilic attack, and thus the nitro group is constantly involved in H-donor interaction with the active site residues of four enzymes 2XP2, 1UOM, 3F81, and 4ASD. Furthermore, the carbonyl group in P2 established H-acceptor interaction with the active site residues of 2XP2 and 2HY8. Additionally, the carbonyl of the sulfonyl group showed hydrogen acceptor bonds when binding with 2XP2 and 2HY8 receptors.

III.4 Conclusion

In this study, a combined approach based on the K-means algorithm with statistical analysis has been applied successfully to classify 500 cytotoxic agents using 21 molecular descriptors into three distinct clusters. The percentage of molecules in each cluster is 50%, 24.88%, and 25.12% for cluster 1, cluster 2, and cluster 3, respectively. Each cluster groups a homogeneous class of molecules with respect to their molecular descriptors. Silhouette analysis, used as cluster validation approach proves that the molecules are very well clustered and there is no misallocation or overlapping clusters. The silhouette indices obtained for each cluster are 0.53 for cluster 1, 0.60 for cluster 2, and 0.57 for cluster 3. Therefore, the average silhouette index obtained is 0.56.   The value of the Hopkins statistic obtained is 0.922 confirming that the dataset is highly clusterable. *In silico* screening of pharmacological properties ADME and drug-likeness of molecules showed that Cluster 1 molecules exhibit the best ADME profile and oral bioavailability as they have shown high intestinal absorption, low P-gp activity, and good drug-likeness. The quantitative analysis of electrostatic potential reveals that the paragon molecules have large electron-deficient regions, particularly for paragon 2. These electrophilic regions are susceptible to be attacked by nucleophilic reagents. Interestingly, our docking results indicate that all compounds exhibit high binding affinity with the studied receptors that could explain the inhibition activity observed for these molecules against various cancer cell lines. Besides, this finding also confirms the results of cytotoxicity assay. The most striking result to emerge from our study is the mean in category of descriptors for cluster 1 as depicted in Table A1 (see Appendix A), i.e., the mean of mass, mean of log P, mean of heterocyclic rings, etc). This result provides a guide to assist the medicinal chemist in selecting the optimal molecules to synthesize in order to test and promote only the candidates that exhibit strong efficacy, low toxicity, and good pharmacokinetic profiles to improve clinical outcomes of drug therapies.

III.5 References

1.   Zhang X-B, Zou C-L, Duan Y-X, et al (2015) Activity guided isolation and modification of juglone from Juglansregia as potent cytotoxic agent against lung cancer cell lines. BMC Complement Altern Med 15:396. https://doi.org/10.1186/s12906-015-0920-0

2.   Shen Y-C, Chang Y-T, Lin C-L, et al (2011) Synthesis of 1-Substituted Carbazolyl-1,2,3,4-tetrahydro- and Carbazolyl-3,4-dihydro-β-carbolineAnalogs as Potential Antitumor Agents. Marine Drugs 9:256–277. https://doi.org/10.3390/md9020256

3.   Zhang Q-Q, Sun Z-Y, Feng X-Y, et al (2019) Thymol derivatives from the roots of Eupatorium chinense and their cytotoxic activities. Phytochemistry Letters 29:165–168. https://doi.org/10.1016/j.phytol.2018.12.008

4.   Zhou B, Hu Z-J, Zhang H-J, et al (2019) Bioactive staurosporine derivatives from the Streptomyces sp. NB-A13. Bioorganic Chemistry 82:33–40. https://doi.org/10.1016/j.bioorg.2018.09.016

5.   Zhu S-L, Wu Y, Liu C-J, et al (2013) Synthesis and in vitro cytotoxic activity evaluation of novel heterocycle bridged carbothioamide type isosteviol derivatives as antitumor agents. Bioorganic & Medicinal Chemistry Letters 23:1343–1346. https://doi.org/10.1016/j.bmcl.2012.12.091

6.   Yang C-R, Peng B, Cao S-L, et al (2018) Synthesis, cytotoxic evaluation and target identification of thieno[2,3- d]pyrimidine derivatives with a dithiocarbamate side chain at C2 position. European Journal of Medicinal Chemistry 154:324–340. https://doi.org/10.1016/j.ejmech.2018.05.028

7.   Wu P, He P, Zhao S, et al (2014) Effects of Ursolic Acid Derivatives on Caco-2 Cells and Their Alleviating Role in Streptozocin-Induced Type 2 Diabetic Rats. Molecules 19:12559–12576. https://doi.org/10.3390/molecules190812559

8.   Wang J, Chang Y, Dong X, et al (2018) Cytotoxic and glycosaminoglycan priming activities of novel 4-anilinequinazoline β-D-xylosides. Carbohydrate Research 463:6–13. https://doi.org/10.1016/j.carres.2018.04.007

9. Trinh BTD, Nguyen HD, Nguyen HT, et al (2019) Cytotoxic limonoids from the bark of Walsuracochinchinensis. Fitoterapia 133:75–79. https://doi.org/10.1016/j.fitote.2018.11.008

10. Gao H, Zhang X, Pu X-J, et al (2019) 2-Benzoylbenzofuran derivatives possessing piperazine linker as anticancer agents. Bioorganic & Medicinal Chemistry Letters 29:806–810. https://doi.org/10.1016/j.bmcl.2019.01.025

11. Adem FA, Kuete V, Mbaveng AT, et al (2018) Cytotoxicbenzylbenzofuran derivatives from Dorsteniakameruniana. Fitoterapia 128:26–30. https://doi.org/10.1016/j.fitote.2018.04.019

12. Bian J, Li T, Weng T, et al (2017) Synthesis, evaluation and quantitative structure–activity relationship (QSAR) analysis of Wogonin derivatives as cytotoxic agents. Bioorganic & Medicinal Chemistry Letters 27:1012–1016. https://doi.org/10.1016/j.bmcl.2016.12.076

13. Guo J, Yang Y, Wang N, Liu Z (2018) Synthesis and cytotoxicity screening of derivatives of the simplified ecteinascidinpentacyclic skeleton as anticancer agents. Tetrahedron Letters 59:3202–3205. https://doi.org/10.1016/j.tetlet.2018.07.027

14. Song Z-L, Wang M-J, Li L, et al (2016) Design, synthesis, cytotoxic activity and molecular docking studies of new 20(S)-sulfonylamidinecamptothecin derivatives. European Journal of Medicinal Chemistry 115:109–120. https://doi.org/10.1016/j.ejmech.2016.02.070

15. El-Zahabi HSA, Khalifa MMA, Gado YMH, et al (2019) New thiobarbituric acid scaffold-based small molecules: Synthesis, cytotoxicity, 2D-QSAR, pharmacophore modelling and in-silico ADME screening. European Journal of Pharmaceutical Sciences 130:124–136. https://doi.org/10.1016/j.ejps.2019.01.023

16. Chen Z, Duan H, Wang M, et al (2015) Synthesis, cytotoxicity and haemolytic activity of Pulsatillasaponin A, D derivatives. Bioorganic & Medicinal Chemistry Letters 25:2550–2554. https://doi.org/10.1016/j.bmcl.2015.04.049

17. Zeng F, Chen C, Al Chnani AA, et al (2019) Dibrefeldins A and B, A pair of epimers representing the first brefeldin A dimers with cytotoxic activities from

Penicilliumjanthinellum. Bioorganic Chemistry 86:176–182. https://doi.org/10.1016/j.bioorg.2019.01.042

18. Gu W, Hao Y, Zhang G, et al (2015) Synthesis, in vitro antimicrobial and cytotoxic activities of new carbazole derivatives of ursolic acid. Bioorganic & Medicinal Chemistry Letters 25:554–557. https://doi.org/10.1016/j.bmcl.2014.12.021

19. Hong T, Jiang W, Dong H-M, et al (2015) Synthesis and cytotoxic activities of E-resveratrol derivatives. Chinese Journal of Natural Medicines 13:375–382. https://doi.org/10.1016/S1875-5364(15)30029-7

20. Ibrahim SRM, Mohamed GA, Al Haidari RA, et al (2018) Fusarithioamide B, a new benzamide derivative from the endophytic fungus Fusariumchlamydosporium with potent cytotoxic and antimicrobial activities. Bioorganic & Medicinal Chemistry 26:786–790. https://doi.org/10.1016/j.bmc.2017.12.049

21. Li W-Q, Wang X-L, Qian K, et al (2013) Design, synthesis and potent cytotoxic activity of novel podophyllotoxin derivatives. Bioorganic & Medicinal Chemistry 21:2363–2369. https://doi.org/10.1016/j.bmc.2013.01.069

22. Li J, Ni G, Li L, et al (2019) New iridal-type triterpenoid derivatives with cytotoxic activities from Belamcandachinensis. Bioorganic Chemistry 83:20–28. https://doi.org/10.1016/j.bioorg.2018.08.039

23. Liao S-R, Qin X-C, Wang Z, et al (2016) Design, synthesis and cytotoxic activities of novel 2,5-diketopiperazine derivatives. European Journal of Medicinal Chemistry 121:500–509. https://doi.org/10.1016/j.ejmech.2016.06.002

24. Lin X, Tian D, Fu Y, et al (2019) Synthesis of novel guttiferone E and xanthochymol derivatives with cytotoxicities by inducing cell apoptosis and arresting the cell cycle phase. European Journal of Medicinal Chemistry 162:765–780. https://doi.org/10.1016/j.ejmech.2018.11.046

25. Luo Y, Wang K, Zhang M, et al (2015) Synthesis of new ent-labdanediterpene derivatives from andrographolide and evaluation on cytotoxic activities. Bioorganic & Medicinal Chemistry Letters 25:2421–2424. https://doi.org/10.1016/j.bmcl.2015.03.086

26. Dar AM, Rah B, Mir S, et al (2018) DNA binding, artificial nuclease activity and cytotoxic studies of newly synthesized steroidal pyrimidines. International Journal of Biological Macromolecules 111:52–61. https://doi.org/10.1016/j.ijbiomac.2017.12.128

27. Marques RA, Gomes AOCV, de Brito MV, et al (2018) Annonalide and derivatives: Semisynthesis, cytotoxic activities and studies on interaction of annonalide with DNA. Journal of Photochemistry and Photobiology B: Biology 179:156–166. https://doi.org/10.1016/j.jphotobiol.2018.01.016

28. Ni J, Guo M, Cao Y, et al (2019) Discovery, synthesis of novel fusidic acid derivatives possessed amino-terminal groups at the 3-hydroxyl position with anticancer activity. European Journal of Medicinal Chemistry 162:122–131. https://doi.org/10.1016/j.ejmech.2018.10.059

29. Phan NHT, Thuan NTD, Ngoc NT, et al (2014) Two tirucallane derivatives from Paramignyascandens and their cytotoxic activity. Phytochemistry Letters 9:78–81. https://doi.org/10.1016/j.phytol.2014.04.011

30. Sang X-N, Chen S-F, Tang M-X, et al (2017) α-Pyrone derivatives with cytotoxic activities, from the endophytic fungus Phoma sp. YN02-P-3. Bioorganic & Medicinal Chemistry Letters 27:3723–3725. https://doi.org/10.1016/j.bmcl.2017.06.079

31. Shan Y, Hong T, Wang Y-F, et al (2015) Synthesis and cytotoxicity of longistylin C derivatives. Chinese Journal of Natural Medicines 13:311–315. https://doi.org/10.1016/S1875-5364(15)30021-2

32. Shan W-G, Wang H-G, Chen Y, et al (2017) Synthesis of 3- and 29-substituted celastrol derivatives and structure-activity relationship studies of their cytotoxic activities. Bioorganic & Medicinal Chemistry Letters 27:3450–3453. https://doi.org/10.1016/j.bmcl.2017.05.083

33. Thieury C, Lebouvier N, Le Guével R, et al (2017) Mechanisms of action and structure-activity relationships of cytotoxic flavokawain derivatives. Bioorganic & Medicinal Chemistry 25:1817–1829. https://doi.org/10.1016/j.bmc.2017.01.049

34.   Wang K, Li Y, Zhang L-J, et al (2014) Synthesis and in vitro cytotoxic activities of sorafenib derivatives. Chinese Chemical Letters 25:702–704. https://doi.org/10.1016/j.cclet.2014.03.020

35.   Yang J, Su J-C, Lei X-P, et al (2018) Acylphloroglucinol derivatives from the leaves of Syzygiumsamarangense and their cytotoxic activities. Fitoterapia 129:1–6. https://doi.org/10.1016/j.fitote.2018.06.002

36.   Zou Z-X, Tan G-S, Zhang G-G, et al (2017) New cytotoxic apigenin derivatives from Selaginelladoederleinii. Chinese Chemical Letters 28:931–934. https://doi.org/10.1016/j.cclet.2017.01.011

37.   Zi C-T, Yang L, Xu F-Q, et al (2019) Synthesis and antitumor activity of biotinylatedcamptothecin derivatives as potent cytotoxic agents. Bioorganic & Medicinal Chemistry Letters 29:234–237. https://doi.org/10.1016/j.bmcl.2018.11.049

38.   Zhou M, Zhou K, Lou J, et al (2015) New biphenyl derivatives from the leaves of Nicotianatabacum and their cytotoxic activity. Phytochemistry Letters 14:226–229. https://doi.org/10.1016/j.phytol.2015.10.025

39.   Yang C-J, Song Z-L, Goto M, et al (2017) Design, semisynthesis and potent cytotoxic activity of novel 10-fluorocamptothecin derivatives. Bioorganic & Medicinal Chemistry Letters 27:4694–4697. https://doi.org/10.1016/j.bmcl.2017.09.012

40.   Zhang C-B, Liu Y, Liu Z-F, et al (2017) Synthesis and cytotoxic activity of novel tetrahydrobenzodifuran–imidazolium salt derivatives. Bioorganic & Medicinal Chemistry Letters 27:1808–1814. https://doi.org/10.1016/j.bmcl.2017.02.053

41.   Zhang Z-J, Tian J, Wang L-T, et al (2014) Design, synthesis and cytotoxic activity of novel sulfonylurea derivatives of podophyllotoxin. Bioorganic & Medicinal Chemistry 22:204–210. https://doi.org/10.1016/j.bmc.2013.11.035

42.   Wang W, Wang D, Wang Z, et al (2017) Synthesis of new sarsasapogenin derivatives with cytotoxicity and apoptosis-inducing activities in human breast cancer MCF-7 cells. European Journal of Medicinal Chemistry 127:62–71. https://doi.org/10.1016/j.ejmech.2016.12.011

43. Teng Y-N, Wang Y, Hsu P-L, et al (2018) Mechanism of action of cytotoxic compounds from the seeds of Euphorbia lathyris. Phytomedicine 41:62–66. https://doi.org/10.1016/j.phymed.2018.02.001

44. Taha M, Sultan S, Azlan M, et al (2017) Synthesis of a series of new 6-nitrobenzofuran-2-carbohydrazide derivatives with cytotoxic and antioxidant activity. New Horizons in Translational Medicine 4:23–30. https://doi.org/10.1016/j.nhtm.2017.08.002

45. Zou M, Yu S-S, Wang K, et al (2013) Glycosylation of ent-kaurene derivatives and an evaluation of their cytotoxic activities. Chinese Journal of Natural Medicines 11:289–295. https://doi.org/10.1016/S1875-5364(13)60031-X

46. Pandey SK, Pratap S, Marverti G, et al (2019) Synthesis, spectroscopic, crystal structure and in vitro cytotoxicity studies of N-thiophenoyl-N′-substituted phenyl thiocarbamide derivatives. Journal of Molecular Structure 1180:447–454. https://doi.org/10.1016/j.molstruc.2018.12.011

47. Supratman U, Naibaho W, Salam S, et al (2019) Cytotoxic Triterpenoids from the Bark of Chisocheton patens Blume (Meliaceae). Phytochemistry Letters 30:81–87. https://doi.org/10.1016/j.phytol.2019.01.034

48. Djurendić EA, Ajduković JJ, Sakač MN, et al (2012) Synthesis and cytotoxic activity of some 17-picolyl and 17-picolinylidene androstane derivatives. European Journal of Medicinal Chemistry 54:784–792. https://doi.org/10.1016/j.ejmech.2012.06.030

49. Song M, Wang S, Wang Z, et al (2019) Synthesis, antimicrobial and cytotoxic activities, and molecular docking studies of N-arylsulfonylindoles containing an aminoguanidine, a semicarbazide, and a thiosemicarbazide moiety. European Journal of Medicinal Chemistry 166:108–118. https://doi.org/10.1016/j.ejmech.2019.01.038

50. Boff L, Munkert J, Ottoni FM, et al (2019) Potential anti-herpes and cytotoxic action of novel semisynthetic digitoxigenin-derivatives. European Journal of Medicinal Chemistry 167:546–561. https://doi.org/10.1016/j.ejmech.2019.01.076

51. Promgool T, Kanokmedhakul K, Tontapha S, et al (2019) Bioactive homogentisic acid derivatives from fruits and flowers of Miliusavelutina. Fitoterapia 134:65–72. https://doi.org/10.1016/j.fitote.2019.02.007

52. Khazem M, Gaslonde T, Dumontet V, et al (2014) Cytotoxic turrianes from Kermadeciaelliptica: Hemisynthesis and biological activities of kermadecin A derivatives. Phytochemistry Letters 10:249–254. https://doi.org/10.1016/j.phytol.2014.10.006

53. Xu C, Wang Z-T (2011) Synthesis and Cytotoxic Activity of 12-Methyleneurea-14-deoxyandrographolide Derivatives. Chinese Journal of Natural Medicines 9:46–50. https://doi.org/10.1016/S1875-5364(11)60019-8

54. Tang L, Fu L, Lu C, et al (2017) New cytotoxic phloroglucinol derivatives from Agrimoniapilosa. Fitoterapia 118:69–72. https://doi.org/10.1016/j.fitote.2017.02.010

55. Yang YJ, Yang YN, Jiang JS, et al (2010) Synthesis and cytotoxic activity of heterocycle-substituted phthalimide derivatives. Chinese Chemical Letters 21:902–904. https://doi.org/10.1016/j.cclet.2010.03.039

56. Paraskevopoulos G, Krátký M, Mandíková J, et al (2015) Novel derivatives of nitro-substituted salicylic acids: Synthesis, antimicrobial activity and cytotoxicity. Bioorganic & Medicinal Chemistry 23:7292–7301. https://doi.org/10.1016/j.bmc.2015.10.029

57. Abd El-Karim SS, Syam YM, El Kerdawy AM, Abdelghany TM (2019) Newthiazol-hydrazono-coumarin hybrids targeting human cervical cancer cells: Synthesis, CDK2 inhibition, QSAR and molecular docking studies. Bioorganic Chemistry 86:80–96. https://doi.org/10.1016/j.bioorg.2019.01.026

58. Chen J, Dong X, Liu T, et al (2009) Design, synthesis, and quantitative structure–activity relationship of cytotoxic γ-carboline derivatives. Bioorganic & Medicinal Chemistry 17:3324–3331. https://doi.org/10.1016/j.bmc.2009.03.037

59. Chen J, Liu T, Wu R, et al (2011) Design, synthesis, and biological evaluation of novel γ-carboline ketones as anticancer agents. European Journal of Medicinal Chemistry 46:1343–1347. https://doi.org/10.1016/j.ejmech.2011.01.057

60. Kilic-Kurt Z, Bakar-Ates F, Aka Y, Kutuk O (2019) Design, synthesis and in vitro apoptotic mechanism of novel pyrrolopyrimidine derivatives. Bioorganic Chemistry 83:511–519. https://doi.org/10.1016/j.bioorg.2018.10.060

61. Nam NH, Van Kiem P, Ban NK, et al (2011) Chemical constituents of Mallotusmacrostachyus growing in Vietnam and cytotoxic activity of some cycloartane derivatives. Phytochemistry Letters S1874390011000954. https://doi.org/10.1016/j.phytol.2011.07.007

62. Shen C-C, Cheng J-J, Lay H-L, et al (2012) Cytotoxic Apigenin Derivatives from *Chrysopogonaciculatis*. J Nat Prod 75:198–201. https://doi.org/10.1021/np2007796

63. HyperChemv8(2009) Molecular Modeling System, Hypercube Inc, 1115 NW 4th Street, Gainesville, FL 32601, USA

64. Molecular Operating Environment (MOE) (2014) Chemical Computing Group Inc, 1010 Sherbrooke St. West, Suite # 910, Montreal QC, Canada H3A 2R7

65. MarvinSketch (2019) ChemAxon. http://www.chemaxon.com/products/marvin/marvinsketch/. Accessed 16 Jul 2019

66. SwissADME(2019). http://www.swissadme.ch/. Accessed 22 sept 2019

67. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian 09, Revision A.02, Gaussian, Inc, Wallingford CT

68. Lu T, Chen F (2012) Multiwfn: A multifunctional wavefunctionanalyzer. J ComputChem 33:580–592. https://doi.org/10.1002/jcc.22885

69. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. Journal of Molecular Graphics 14:33–38. https://doi.org/10.1016/0263-7855(96)00018-5

70. Cui JJ, Tran-Dubé M, Shen H, et al (2011) Structure Based Drug Design of Crizotinib (PF-02341066), a Potent and Selective Dual Inhibitor of Mesenchymal–Epithelial Transition Factor (c-MET) Kinase and Anaplastic Lymphoma Kinase (ALK). J Med Chem 54:6342–6363. https://doi.org/10.1021/jm2007613

71. Renaud J, Bischoff SF, Buhl T, et al (2003) Estrogen Receptor Modulators: Identification and Structure−Activity Relationships of Potent ERα-Selective Tetrahydroisoquinoline Ligands. J Med Chem 46:2945–2957. https://doi.org/10.1021/jm030086h

72. Mittl PRE, Di Marco S, Krebs JF, et al (1997) Structure of Recombinant Human CPP32 in Complex with the Tetrapeptide Acetyl-Asp-Val-Ala-Asp Fluoromethyl Ketone. J BiolChem 272:6539–6547. https://doi.org/10.1074/jbc.272.10.6539

73. Lu H, Lei M, Schulze-Gahmen U. Crystal Structure of the Complex between Human Pak1-kinase and 3-Hydroxystaurosporine. https://www.rcsb.org/structure/2HY8

74. Wu S, Vossius S, Rahmouni S, et al (2009) Multidentate Small-Molecule Inhibitors of *Vaccinia* H1-Related (VHR) Phosphatase Decrease Proliferation of Cervix Cancer Cells. J Med Chem 52:6716–6723. https://doi.org/10.1021/jm901016k

75. McTigue M, Murray BW, Chen JH, et al (2012) Molecular conformations, interactions, and properties associated with drug efficiency and clinical performance among VEGFR TK inhibitors. Proceedings of the National Academy of Sciences 109:18281–18289. https://doi.org/10.1073/pnas.1207759109

76. Daoud I, Melkemi N, Salah T, Ghalem S (2018) Combined QSAR, molecular docking and molecular dynamics study on new Acetylcholinesterase and Butyrylcholinesterase inhibitors. Computational Biology and Chemistry 74:304–326. https://doi.org/10.1016/j.compbiolchem.2018.03.021

77.　Aggarwal CC, Reddy CK (2016) Data Clustering: Algorithms and Applications. CRC Press, Boca Raton

78.　Hartigan JA, Wong MA (1979) Algorithm AS 136: A K-Means Clustering Algorithm. Applied Statistics 28:100-108. https://doi.org/10.2307/2346830

79.　MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, eds L. M. Le Cam & J. Neyman, 1:281–297. Berkeley, CA: University of California Press

80.　Lloyd S P (1982)　Least squares quantization in PCM. Technical Note, Bell Laboratories.*IEEE Transactions on Information Theory*, 28:128–137

81.　Forgy E W (1965) Cluster analysis of multivariate data: efficiency vs interpretability of classifications. *Biometrics* 21:768–769

82.　Hastie T, Tibshirani R, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd ed. Springer, New York, NY

83.　Kaufman L, Rousseeuw PJ (1990) Finding groups in data: an introduction to cluster analysis. Wiley series in probability and mathematical statistics, John Wiley& Sons, New York

84.　Husson F, Josse J, Le S, Mazet J (2020) FactoMineR: Factor Analysis and Data Mining with R. R package version 1.01. https://CRAN.Rproject.org/package=FactoMineR

85.　Rousseeuw P, Struyf A, Hubert M (2017) Cluster: Finding Groups in Data: Cluster Analysis, R package version 2.06. https://CRAN.R-project.org/ package=cluster

86.　Lawson RG, Jurs PC (1990) New index for clustering tendency and its application to chemical problems. J ChemInfComputSci 30:36–41. https://doi.org/10.1021/ci00065a010

87.　Han J, Kamber M (2012) Data mining: concepts and techniques, 3rd ed. Elsevier, Burlington, MA c.　Lloyd S P (1982) Least squares quantization in PCM. Technical Note, Bell Laboratories.*IEEE Transactions on Information Theory*, 28:128–137

88. Lebart L, Morineau A, Piron M (1995) Statistique exploratoire multidimensionnelle. Dunod, Paris

89. Husson F, Le S, Pages J (2010) Exploratory Multivariate Analysis by Example Using R. Taylor & Francis, Boca Raton

90. Yang NJ, Hinner MJ (2015) Getting Across the Cell Membrane: An Overview for Small Molecules, Peptides, and Proteins. In: Gautier A, Hinner MJ (eds) Site-Specific Protein Labeling. Springer New York, New York, NY, pp 29–53

91. Pajouhesh H, Lenz GR (2005) Medicinal chemical properties of successful central nervous system drugs. Neurotherapeutics 2:541–553. https://doi.org/10.1602/neurorx.2.4.541

92. Nobili S, Landini I, Mazzei T, Mini E (2012) Overcoming tumor multidrug resistance using drugs able to evade P-glycoprotein or to exploit its expression: OVERCOMING TUMOR MULTIDRUG RESISTANCE. Med Res Rev 32:1220–1262. https://doi.org/10.1002/med.20239

93. Binkhathlan Z, Lavasanifar A (2013) P-glycoprotein Inhibition as a Therapeutic Approach for Overcoming Multidrug Resistance in Cancer: Current Status and Future Perspectives. CCDT 13:326–346. https://doi.org/10.2174/15680096113139990076

94. Rodriguez-Antona C, Ingelman-Sundberg M (2006) Cytochrome P450 pharmacogenetics and cancer. Oncogene 25:1679–1691. https://doi.org/10.1038/sj.onc.1209377

95. Zhao L, Sun N, Tian L, et al (2019) Strategies for the development of highly selective cytochrome P450 inhibitors: Several CYP targets in current research. Bioorganic & Medicinal Chemistry Letters 29:2016–2024. https://doi.org/10.1016/j.bmcl.2019.06.040

96. World Health Organization (2018) https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed 13 Jul 2020

97. Clément G, Slenzka K (2006) Fundamentals of Space Biology: Research on Cells, Animals, and Plants in Space. Springer New York, New York, NY

98. Imberty A, Hardman KD, Carver JP, Perez S (1991) Molecular modelling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A. Glycobiology 1:631–642. https://doi.org/10.1093/glycob/1.6.631

CHAPTER Ⅳ: Molecular Docking/Dynamic simulations, MEP, ADME-TOX based analysis of xanthone derivatives as CHK1 inhibitors
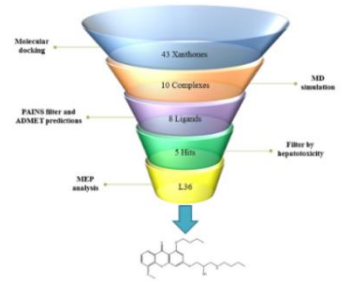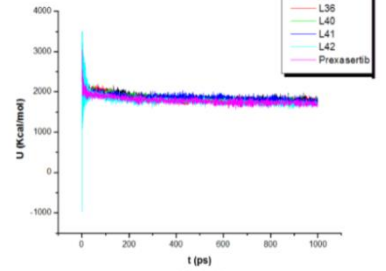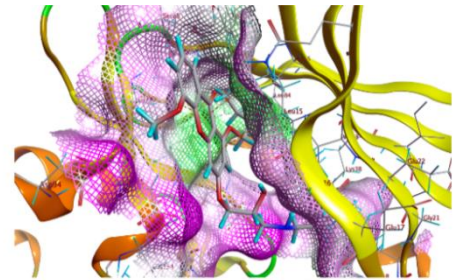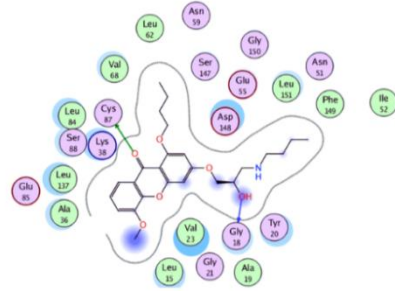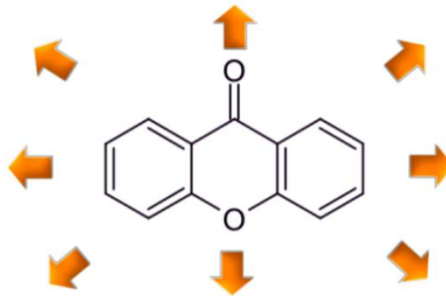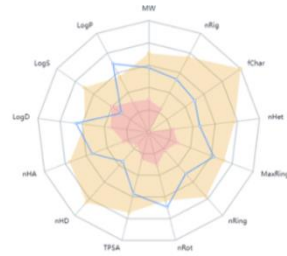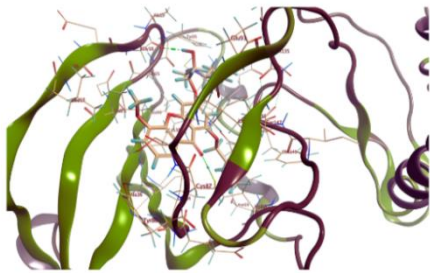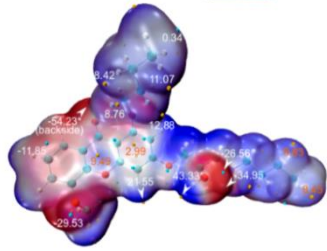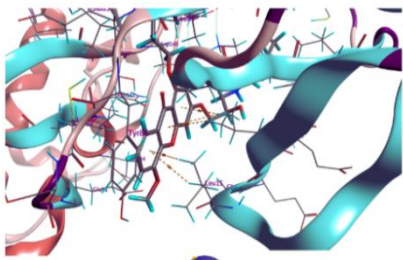
IV.1 Introduction

In silico pharmacological tools, viz., molecular docking, molecular dynamics, virtual screening, ADME-TOX predictions are regularly used to predict how drugs affect biological targets, optimize lead compounds, or discover novel hits. These techniques collect knowledge from the target crystal structure and the chemical structure of small molecules to compute the interaction energy score, the physicochemical and pharmacokinetic properties, and toxicological profile, providing new insight into the biochemical mechanism of drug and its side effects [1].

In this study, we investigated the inhibitory activity of cytotoxic xanthone compounds against the CHK1 pathway using a hierarchical *in silico* approach to identify and characterize new potential CHK1 inhibitors beneficial for cancer therapy. For this, a molecular docking study was conducted for the forty-three xanthone derivatives, along with standard Prexasertib toward the selected CHK1 protein structures 7AKM and 7AKO. In order to elucidate the dynamic behavior of target-ligand complex and validate the interactions of docking conformers, five top-scored complexes for each target subjected for molecular dynamics simulation along with Prexasertib against both targets. PAINS filter analysis was carried out for the top eight compounds to remove the problematic leads. Moreover, *in silico* ADME-TOX studies are used to predict the pharmacokinetics, pharmacodynamics, and toxicological properties of the selected eight hits and the reference Prexasertib. The quantitative analysis of electrostatic potential was performed for the lead compound L36 to identify the reactive sites and possible non-covalent interactions.

IV.2   Materials and methods

IV.2.1   Biological database

A series of forty-three cytotoxic xanthone derivatives were collected from the prior report published by Park *et al* [2] to create a database for analysis. All compounds have shown promising cytotoxicity *in vitro* against colon (HCT15), breast (T47D), cervical (HeLa), and stomach (NCI-N87) cancer cell lines. The details of chemical structure of xanthone compounds along with their experimental inhibitory activity expressed as IC_{50} (µM) are presented in (Table IV. 1).
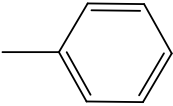
Table IV. 1 Chemical structure of xanthone compounds with their IC_{50} values
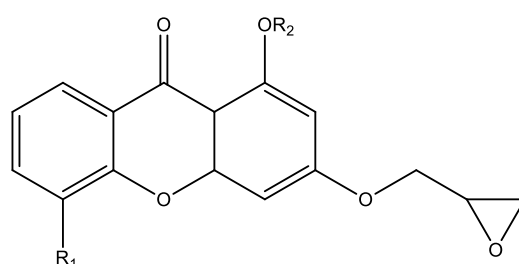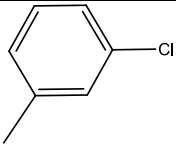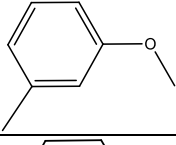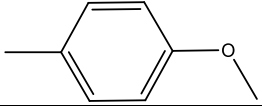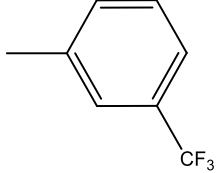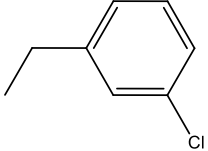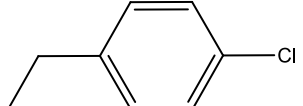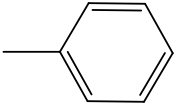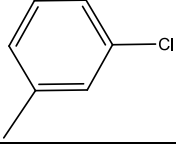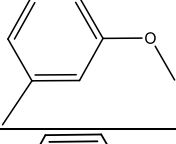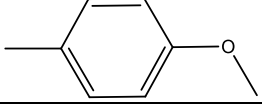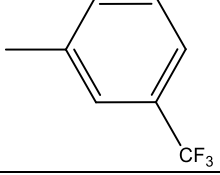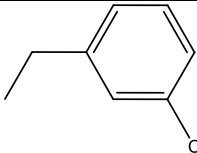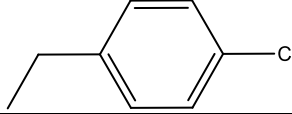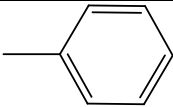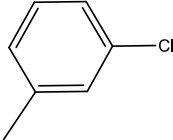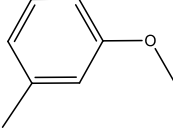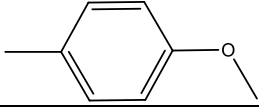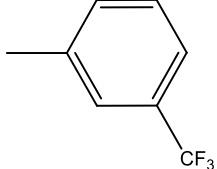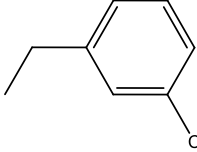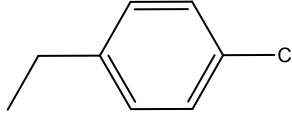
Group I (Compounds 1-22)
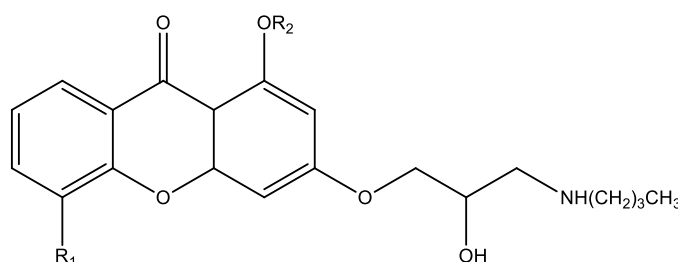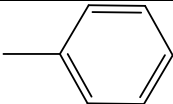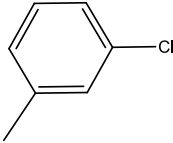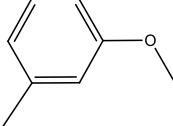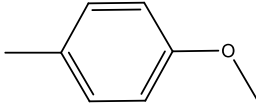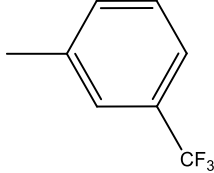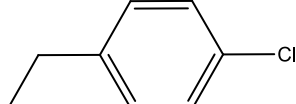
Group II (Compounds 23-43)

Comp.	R₁	R₂	IC_{50} (µM) HCT15	IC_{50} (µM) T47D	IC_{50} (µM) HeLa	IC_{50} (µM) NCI-N87
1	H	C₂H₅	3.96	3.60	0.80	3.91
2	H	i-C₃H₇	1.63	1.94	1.62	0.29
3	H	C₄H₉	2.45	3.75	1.81	4.55
4	H	C₅H₁₁	3.28	>50	1.79	10.95
5	H	(phenyl)	1.50	3.06	1.92	5.07

6	H	3-chlorophenyl	1.96	3.30	3.57	1.89
7	H	3-methoxyphenyl	1.49	3.73	1.57	3.26
8	H	4-methoxyphenyl	1.72	2.96	1.58	3.73
9	H	4-(CF$_3$)phenyl	5.08	2.30	4.28	1.36
10	H	3-ethyl-4-chlorophenyl	4.13	1.36	1.46	3.18
11	H	4-ethyl-... -chlorophenyl	5.27	0.97	4.91	2.02
12	OCH$_3$	C$_2$H$_5$	0.82	0.83	1.17	1.19
13	OCH$_3$	i-C$_3$H$_7$	0.31	1.75	1.87	3.12
14	OCH$_3$	C$_4$H$_9$	0.22	1.63	1.08	0.31
15	OCH$_3$	C$_5$H$_{11}$	0.37	4.57	4.06	1.69
16	OCH$_3$	phenyl	0.34	2.02	1.66	1.75
17	OCH$_3$	3-chlorophenyl	0.17	1.55	1.70	0.95
18	OCH$_3$	3-methoxyphenyl	0.35	4.05	2.43	1.32
19	OCH$_3$	4-methoxyphenyl	0.42	7.68	1.74	1.41
20	OCH$_3$	4-(CF$_3$)phenyl	0.36	2.87	4.27	2.24

21	OCH₃		1.08	2.74	0.47	1.11
22	OCH₃		45.22	>50	>50	1.49
23	H	C₂H₅	28.17	7.02	>50	1.77
24	H	i-C₃H₇	43.4	4.71	36.62	2.05
25	H	C₄H₉	8.61	1.54	2.07	1.63
26	H	C₅H₁₁	4.58	1.27	2.56	0.06
27	H		3.09	2.81	7.05	1.44
28	H		2.24	1.17	2.83	0.17
29	H		1.91	2.57	5.86	0.07
30	H		2.17	1.41	4.81	1.80
31	H		2.45	0.14	2.26	0.24
32	H		2.28	1.72	2.56	0.38
33	H		2.13	0.63	1.37	3.63
34	OCH₃	C₂H₅	10.6	4.28	23.48	2.24
35	OCH₃	i-C₃H₇	9.60	2.67	14.39	0.08
36	OCH₃	C₄H₉	6.02	1.13	7.65	1.64
37	OCH₃	C₅H₁₁	1.81	0.78	3.84	0.48

38	OCH$_3$		2.08	2.40	2.25	1.23
39	OCH$_3$		1.59	0.67	2.40	1.16
40	OCH$_3$		1.74	1.36	2.30	1.32
41	OCH$_3$		2.29	1.52	1.82	0.25
42	OCH$_3$		2.62	2.00	2.62	1.98
43	OCH$_3$		1.49	0.19	2.60	0.40

IV.2.2   Molecular docking study

IV.2.2.1   Targets preparation

Both X-ray crystal structures of the CHK1 protein (PDB ID: 7AKM), (PDB ID: 7AKO) [3] were downloaded from the Protein Data Bank (http://www.rcsb.org/pdb/). The properties of enzymes and energy score values are displayed in (Table IV. 2). Clément *et al* [4], suggest that the resolution of protein structures between 1.5 Å and 2.5 Å is considered as good quality for docking studies. Indeed, the resolution values of both structures of CHK1 kinase protein 7AKM and 7AKO belong to this interval. The R-value is lower than 0.2 for both target structures which emphasizes the reliability of the model, as Morris *et al* [5] demonstrated that the structure refined to $R \leq 0.2$ generally indicates that the protein has the correct topology. The structure preparation for docking starts by removing water molecules, ions, and cofactors, as well as adding explicit hydrogen and assigning charges. The identification of the active site cavity was carried out by the site finder wizard in MOE software [6] that revealed several cavities for each calculation. Out of the detected cavities, cavity 1 was selected because phosphothiophosphoric acid-adenylate ester (AGS) and staurosporine (STU) were fixed in it for both targets. Their active site cavity contains 43, and 34 amino acids corresponding to 7AKM and 7AKO, respectively.

Table Ⅳ. 2 Enzymes properties, energy score, and RMSD values

Enzymes ID	Amino acids number	Resolution (Å)	R-Value	Co-crystallized ligand	S score (kcal/mol)	RMSD (Å)
7AKM	298	1.93 Å	0.197	AGS	-7.2	1.81
7AKO	300	1.8 Å	0.19	STU	-8.52	0.51

To validate the accuracy of docking results, the native ligands were redocked into their respective binding pockets (Figure Ⅳ. 1). Thus, the root-mean-square deviation (RMSD) between the docked and co-crystallized conformation was calculated, as well as the binding free energy (S score) as indicated in (Table Ⅳ. 2). The predicted RMSD values of both complexes as follows 1.81Å and 0.51 Å are less than 2 Å confirm the accuracy of this method [7]. To further increase the confidence of our results, Prexasertib was used as standard drug and docked into targets allowing the comparison of binding affinity with the xanthone compounds. The docking protocol was applied as described in our previous studies [8, 9].
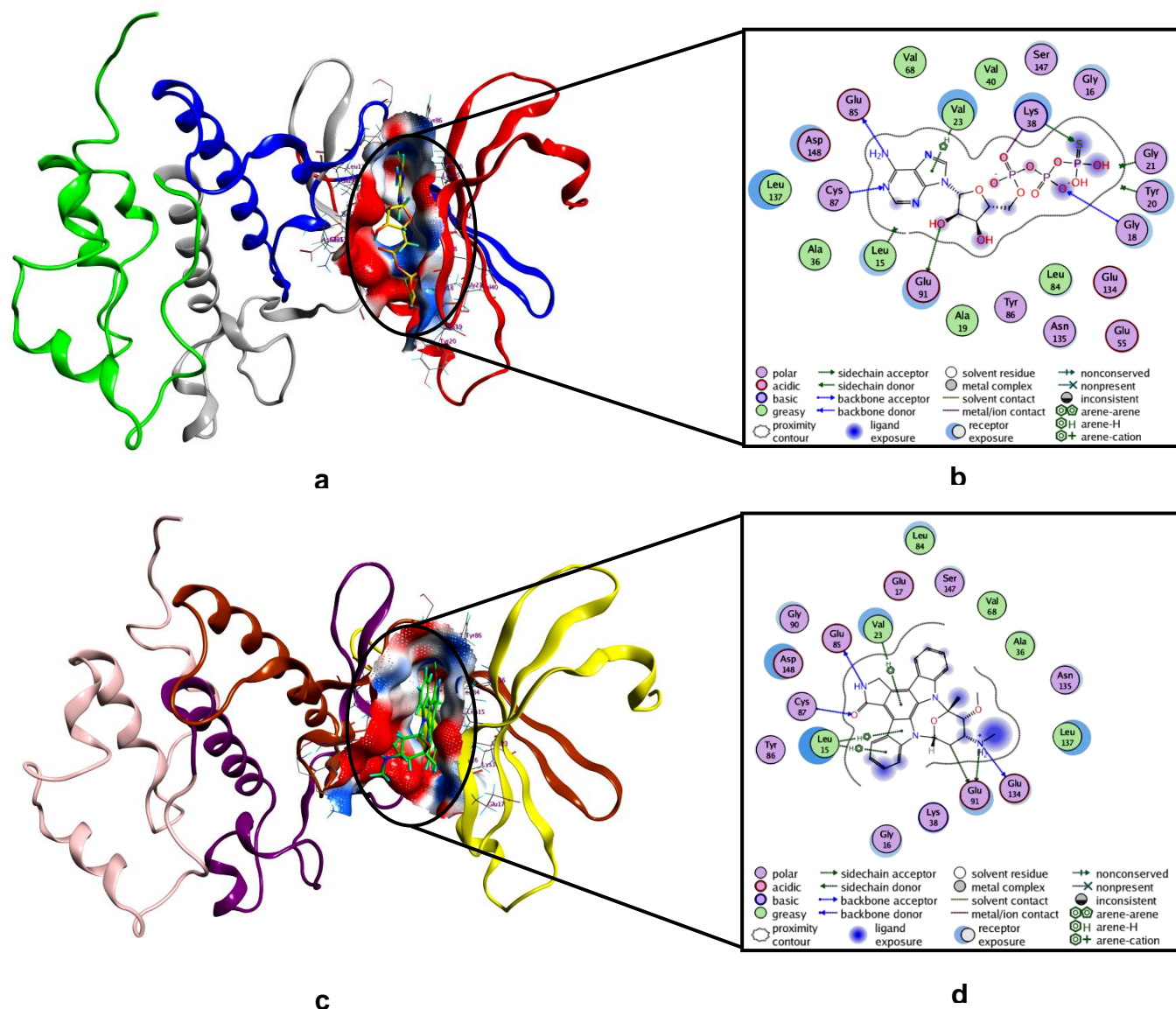
Figure IV. 1 (a) 3D crystal structure of 7AKM complexed with AGS (native ligand). (b) 2D representation of the interactions of AGS with the binding pocket of 7AKM. (c) 3D crystal structure of 7AKO complexed with STU (native ligand). (d) 2D representation of the interactions of STU with the binding pocket of 7AKO. In the electrostatic contour maps, blue regions represent the electropositive groups, red regions represent the electronegative groups, and white for the neutral regions

IV.2.2.2 Compounds preparation

The 3D structures of compounds were drawn by MarvinSketch software [10] and saved in "MDL mol" format. Thereafter, the forty-three molecules were optimized by applying the MM+ force field of molecular mechanics with the Polak-Ribiere conjugate gradient algorithm of 0.01 Kcal/mol in hyperchem8.08 software [11]. Further geometry

optimization was done by using the density functional theory (DFT) at B3LYP/6-31+G(d,p) level of theory in the Gaussian 09 program [12] to obtain the lowest energy conformation. For all stationary points, there is no imaginary frequency at the optimized molecular geometries indicating that the optimized structures are at the minimum on the potential energy surface. The optimized 3D structures of the whole set were imported to create a database input in MOE software for docking studies.

A molecular docking study was conducted to predict the binding mode of ligands within the target active sites and assess the binding free energy of protein-ligand complexes. MOE allows the performance of semi-rigid docking where the protein structure is kept rigid while the ligand is flexible, and thus the optimal geometry of the ligand will be determined during docking. We used a triangle matcher algorithm where the best score is ranked with the London dG scoring function, refined with FF, and re-scored using the force field-based scoring function GBVI/WSA dG that estimates the free energy of binding of the ligand from a given orientation. Furthermore, the maximum number of poses generated for each compound was set to 100 to obtain stable conformation [13, 14].

IV.2.3 Molecular dynamics simulation

To further validate the interactions of docking conformers, five top-scored complexes for each target were selected for molecular dynamics simulation along with Prexasertib. Moe software was used to perform an MD run and explore the stability of twelve complexes as follows 7AKM-L43, 7AKM-L36, 7AKM-L41, 7AKM-L30, 7AKM-L33, 7AKO-L36, 7AKO-L40, 7AKO- L42, 7AKO-L41, 7AKO-L31, 7AKM-Prexasertib, and 7AKO-Prexasertib. The chemical structures of eight selected ligands are depicted in (Figure Ⅳ. 2).
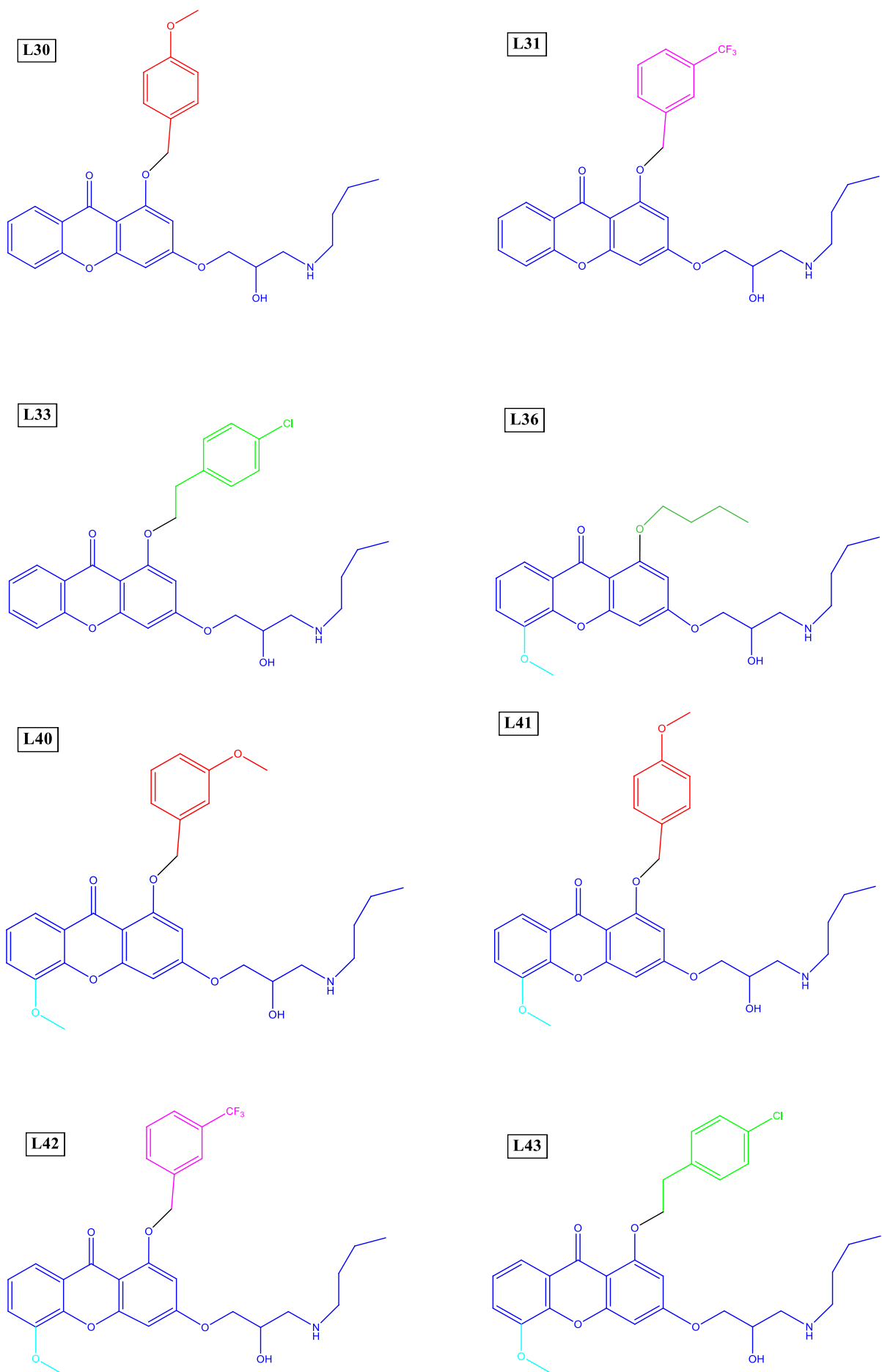
Figure Ⅳ. 2 Chemical structures of the selected xanthone hits

MD simulation was carried out by using the Nose–Poincare–Andersen (NPA) algorithm that reveals the interactions of amino acids residues in each system [15, 16]. The energy minimization step implements the Merck molecular force field MMFF94x force field [17]. Moreover, a cut-off radius of 8.0 Å was considered for short-range van der Waals interactions, and the Berendsen thermostat algorithm was used to monitor the simulation temperature [18]. Afterward, 100 ps was chosen for equilibrium time interval, and the production phase was set to 900 ps were performed for each system. All the other parameters were kept at default setting. In order to analyze the conformational stability of complexes, the fluctuation of potential energies U (Kcal/mol) is plotted as a function of simulation time (ps) by using Origin 6.0 software [19].

IV.2.4   ADME-TOX and drug-likeness prediction

In silico ADME-TOX studies are used to predict the pharmacokinetic, pharmacodynamics, and toxicological properties of compounds resulting in the identification of promising drug candidates. Prior to ADME-TOX prediction, PAINS filter analysis was conducted for the top eight compounds to remove the problematic leads. Pan assay interference compounds (PAINS) are chemical compounds that produce false positive results in high-throughput screenings [20]. In this study, PAINS-Remover (http://cbligand.org/PAINS) was used to check any alerts for the selected hits [21].

ADME-TOX predictions were performed by pkCSM [22] and ADMETlab 2.0 [23] web tools for the studied eight hits and the standard Prexasertib. pkCSM is a novel *in silico* interface that applies graph-based structural signatures to predict the pharmacokinetic and toxicological parameters of substances. Likewise, ADMETlab 2.0 is a freely available platform that uses multiple endpoints to assess the ADME-TOX properties of chemicals. The molecular structures of compounds were introduced in smile file format into these freely available tools and subsequently, we selected the most important pharmacological and toxicological parameters to provide ADME-TOX profile. pkCSM tool was used to predict Water solubility, Caco-2 permeability, human intestinal absorption (HIA), P-glycoprotein (P-gp) substrate, P-gp I inhibitor, P-gp II inhibitor, steady state volume of distribution (VDss), blood brain barrier (BBB) permeability, cytochrome P450 inhibitors including CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4 inhibitors as well as CYP2D6 substrate and CYP3A4 substrate. It was also used to predict the toxicological properties such as AMES toxicity, maximum tolerated dose, human ether a-go-go gene (hERG) I and II inhibitors, rat

lethal dosage (LD50), hepatotoxicity, and skin sensitisation. ADMETLAB 2.0 web tool was used to evaluate excretion parameters and medicinal chemistry properties including clearance, half-life time $T_{1/2}$, Lipinski rule, Pfizer rule, GSK rule, golden triangle, and synthetic accessibility score (SAscore). Prexasertib was used as a reference drug.

For qualitative analysis of the physicochemical quality of compound L36 along with control Prexasertib, a radar chart was retrieved from ADMETlab 2.0 web tool. This plot consists of 13 important physicochemical properties, including Molecular weight MW, number of hydrogen bond acceptors nHA, number of hydrogen bond donors nHD, number of rotatable bonds nRot, number of rings nRing, number of atoms in the biggest ring MaxRing, number of heteroatoms nHet, formal charge fChar, number of rigid bonds nRig, topological polar surface area TPSA, the logarithm of aqueous solubility value LogS, the logarithm of the n-octanol/water distribution coefficient LogP and the logarithm of the n-octanol/water distribution coefficients at physiological pH (7.4) Log D.

IV.2.5   Molecular electrostatic potential

Upon comparing and integrating the findings acquired from molecular docking, dynamics simulation, and ADME-TOX prediction, L36 was subjected to MEP analysis. The multifunctional wavefunction analyzer program, Multiwfn 3.7 [24], was used for the quantitative analysis of electrostatic potential (ESP) on van der Waals surface combined with the VMD 1.9.1 software [25]. Moreover, in order to extract quantitative information around molecular van der Waals surface, a set of statistically defined descriptors by Politzer *et al* [26] have been investigated. These descriptors concern the following statistical quantities: positive variance (PV), negative variance (NV), overall variance (OV), positive surface area (PS) and negative surface area (NS).

IV.3   Results and discussion

IV.3.1   Molecular docking

A molecular docking study was conducted for the forty-three xanthone derivatives along with standard Prexasertib into the selected CHK1 protein structures (7AKM and 7AKO).The results of total binding energy of docked complexes with their distances, types of interactions, the key residues, and the involved atoms of compounds and receptors are summarized in Table B1 and Table B2 (see Appendix B). Interestingly, all compounds showed various types of hydrogen bonds (H-donor and H-acceptor) and hydrophobic

interactions (pi-cation and pi-H) toward the active sites of enzymes. The forty-three xanthone derivatives revealed various interactions toward CHK1 protein structures, and the free energy scores S (Kcal/mol) were in the range of [-8.22,-6.33] and [-8.14,-6.19] for 7AKM and 7AKO, respectively, as reported in Table B3 and Table B4 (see Appendix B).

The co-crystallized ligand AGS displayed seven binding interactions with the binding pocket of 7AKM, two hydrogen donor bonds toward GLU 91 (2.82 Å) and GLU 85 (2.82 Å), three hydrogen acceptor bonds with LYS 38 (4.08 Å) GLY 18 (3.42 Å) CYS 87 (3.19 Å), ionic interaction with LYS 38 (2.94 Å) and hydrophobic interaction (pi-H) with VAL 23 Table B1 (see Appendix B). Moreover, STU was involved in nine binding interactions toward the active site of 7AKO, four hydrogen donor bonds with GLU 91 (3.44 Å, 3.29 Å), GLU 85 (2.93 Å), and GLU 134 (2.99 Å), hydrogen acceptor bond with CYS 87 (3 Å), ionic interaction with GLU 91 (3.29 Å), and three hydrophobic interactions (pi-H) toward LEU 15 and VAL 23 Table B2 (see Appendix B). In short, the native ligands AGS and STU bind into the active pockets of 7AKM and 7AKO, respectively, via different types of interactions toward several amino acids, among which the common residues between the two enzymes are as follows GLU 91, GLU 85, CYS 87, and VAL 23 as displayed in (Figure Ⅳ. 1). These amino acid residues might be responsible for the inhibitory activity of CHK1.

The top five derivatives showing the best affinity, namely L43, L36, L41, L30, and L33, give a binding score of -8.22, -8.16, -8.13, -8.11, and -8.08 kcal/mol, respectively. These xanthone compounds gave better binding affinity toward 7AKM compared to standard Prexasertib with S score of -7.2 kcal/mol Table B3 (see Appendix B) and also better than the native ligand AGS with S score of -7.19 kcal/mol (Table Ⅳ. 2). L43 gave the best energy score of -8.22 kcal/mol compared to other ligands, indicating that this complex is more stable. Additionally, this compound has shown promising cytotoxicity *in vitro* with IC_{50} of 1.49, 0.19, 2.6, and 0.4μM against colon, breast, cervical, and stomach cancer cell lines, respectively.

Imberty *et al* [27] suggest that interaction distances between 2.5-3.1 Å represent strong hydrogen bonds and 3.1-3.55Å for weak bonds. L43 forms two hydrogen interactions with the active site residues, a weak hydrogen donor bond with CYS 87 (3.47 Å) and a strong hydrogen acceptor bond with LYS 38 (2.98 Å) as illustrated in Figure Ⅳ. 3 and (Table B6, see Appendix B). The complex formed by L36 gives four hydrophobic interactions (pi-H) with LEU 15 and VAL 23 residues. Likewise, this compound exhibits potent inhibitory

activity with IC$_{50}$ values of 6.02, 1.13, 7.65, and 1.64 μM against colon, breast, cervical, and stomach cancer cell lines, respectively.
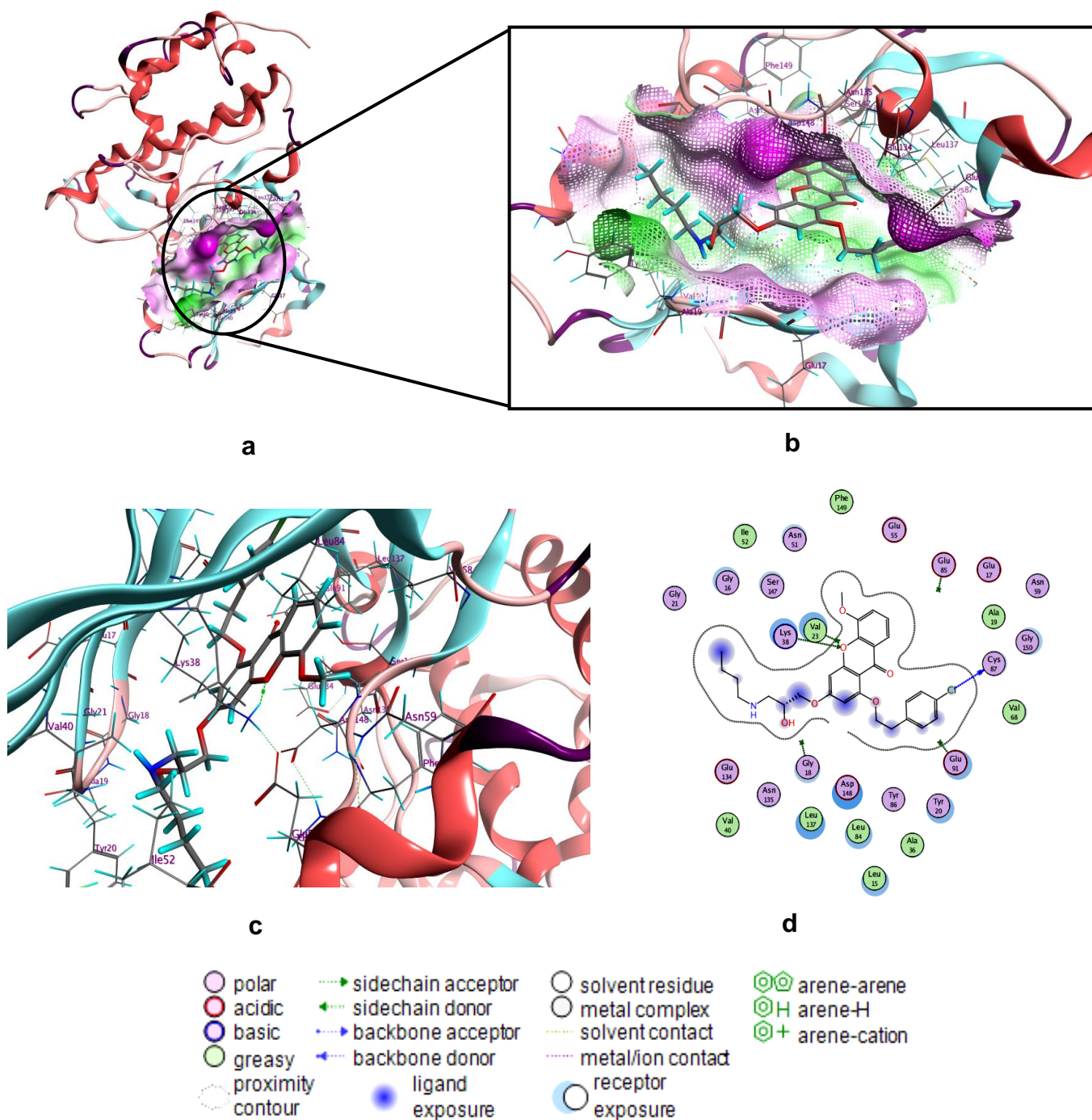
Figure Ⅳ. 3 Full view (a) and magnified view (b) of L43 docked within the active pocket of 7AKM. 3D (c) and 2D (d) representations showing the binding pattern of 7AKM-L43 docking complex. In the lipophilicity contour maps, purple regions represent the hydrophilic groups, green regions represent the lipophilic groups, and white for the neutral regions
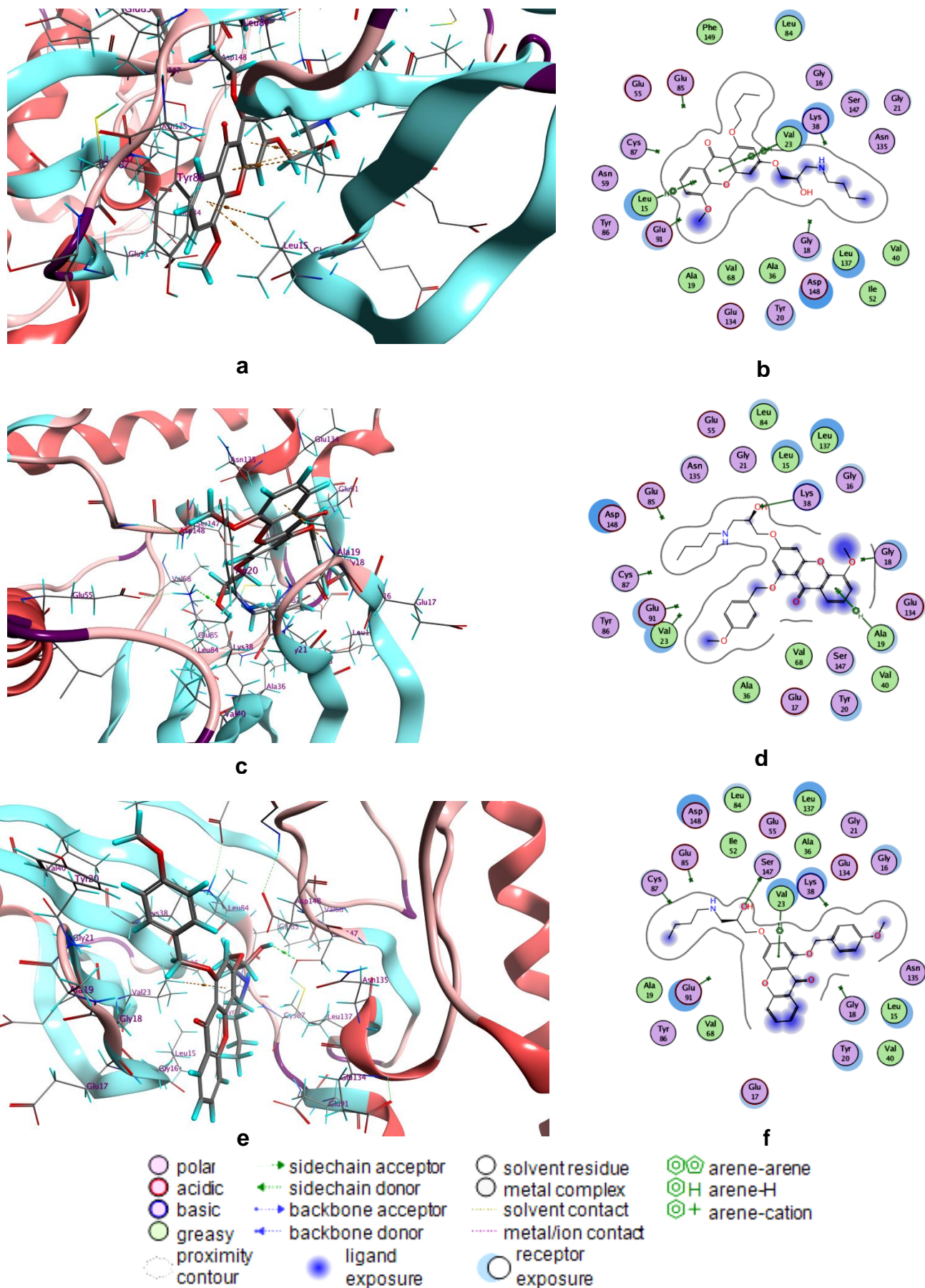
Figure Ⅳ. 4 Three dimensional representations of the best pose docked within 7AKM active site, namely L36 (a), L41(c), and L30 (e). Two dimensional representations describing the molecular interactions between the protein-ligand complex, namely L36 (b), L41 (d), and L30 (f)

L41 establishes two weak interactions with the receptor pocket. First, a hydrogen acceptor bond with LYS 38 (3.18 Å), where the second involve a pi-H bond with ALA 19. Strong hydrogen donor interaction with SER 147 (2.91 Å) and pi-H interaction with VAL 23 are observed for L30 are depicted in (Figure Ⅳ. 4).
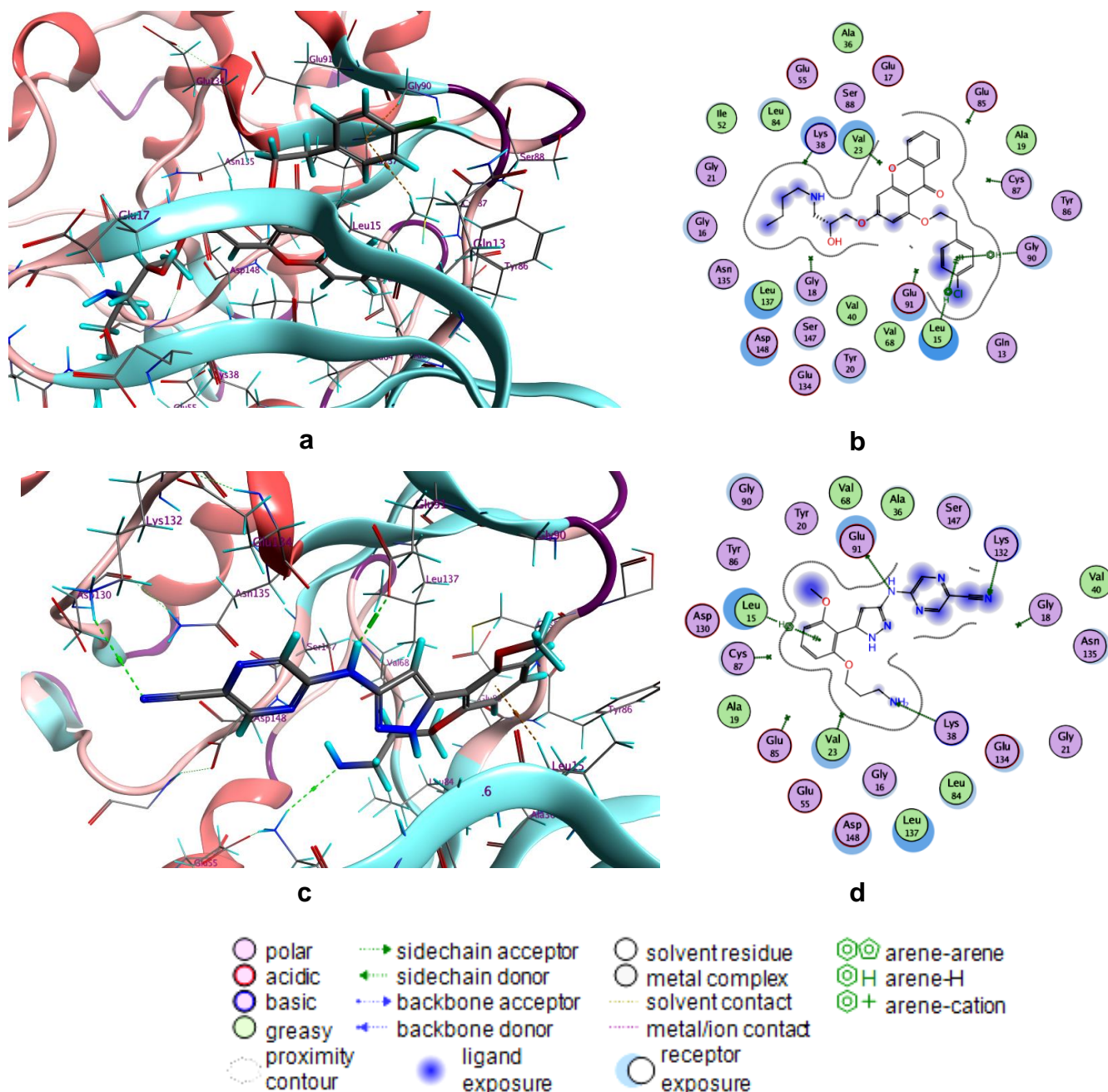
Figure Ⅳ. 5 Three dimensional representations of the best pose docked within 7AKM active site, namely L33 (a) and Prexasertib (c). Two dimensional representations describing the molecular interactions between the protein-ligand complex, namely L36 (b) and Prexasertib (d)

The interactions profile of L33 captured two pi-H interactions with LEU 15 and GLY 90. On the other hand, Prexasertib developed four binding interactions, a strong hydrogen donor bond with GLU 91 (2.99 Å), two weak hydrogen acceptor bonds toward LYS 38 (3.64 Å) and LYS 132 (3.77 Å), and hydrophobic interaction with LEU 15 (Figure Ⅳ. 5). In the second case, the prioritized compounds including, L36, L40, L42, L41, and L31, give a binding score of -8.14, -8.08 8.08, -7.98, -7.97, and -7.9 kcal/mol, respectively. These compounds were better fitted to the target pocket than the standard Prexasertib with S score of -7.41 kcal/mol Table B4 (see Appendix B). L36 gave the best binding energy of -8.14 kcal/mol compared to other complexes Table B4 (see Appendix B). It is noteworthy that ligand 36 also has the second-lowest energy with a very close score value of -8.16 kcal/mol when binding with 7AKM enzyme Table B3 (see Appendix B).
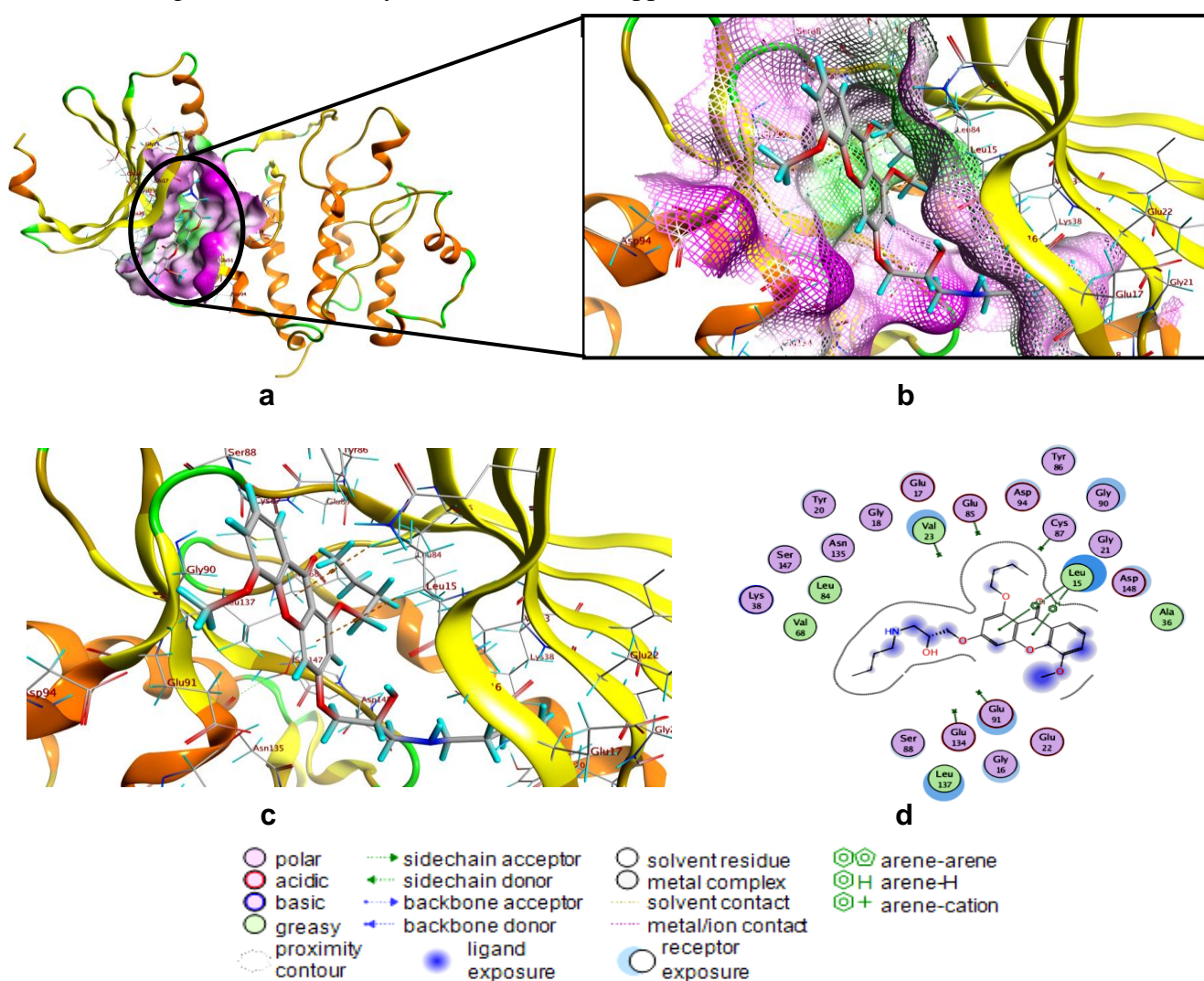
Figure Ⅳ. 6 Full view (a) and magnified view (b) of L36 docked within the active pocket of 7AKO. 3D (c) and 2D (d) representations show the binding pattern of L36-7AKO docking complex. In the lipophilicity contour maps, purple regions represent the hydrophilic groups, green regions represent the lipophilic groups, and white for the neutral regions
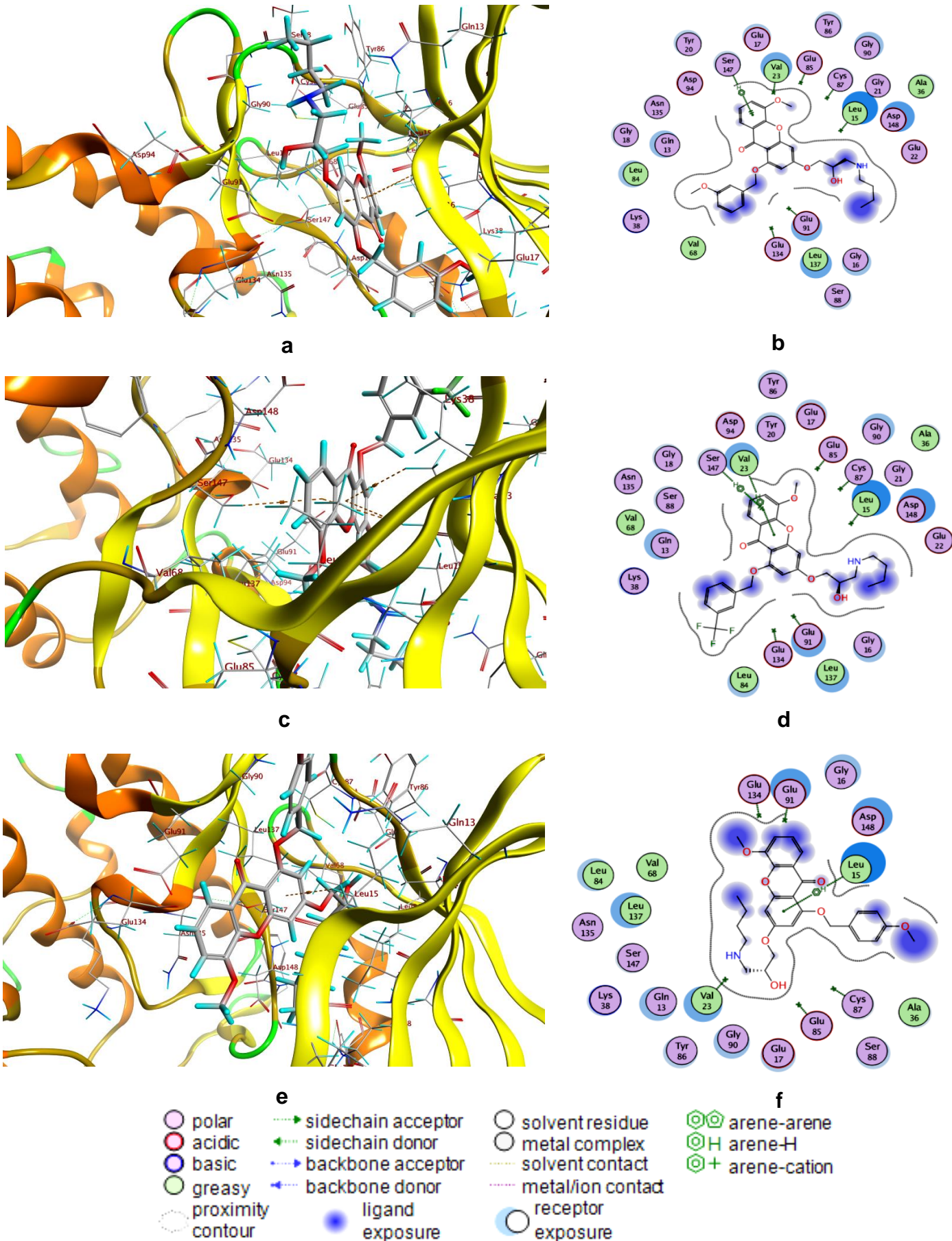
Figure Ⅳ. 7 Three dimensional representations of the best pose docked within 7AKO active site, namely L40 (a), L42(c), and L41 (e). Two dimensional representations describing the molecular interactions between the protein-ligand complex, namely L40 (b), L42 (d), and L41 (f)

As shown in (Figure Ⅳ. 6), L36 forms two pi-H interactions with LEU 15. Similarly, L40 forms pi-H interaction with SER 147, and L42 shows two hydrophobic interactions with VAL 23 and SER 147 displayed in (Figure Ⅳ. 7). Furthermore, L41 establishes a pi-H bond with LEU 15 while L31 maintains three hydrophobic interactions with the receptor cavity toward LEU 15 and VAL 23 residues Table B6 (see Appendix B). While Prexasertib displays a weak hydrogen acceptor bond with TYR 86 (3.43 Å) (Figure Ⅳ. 8).
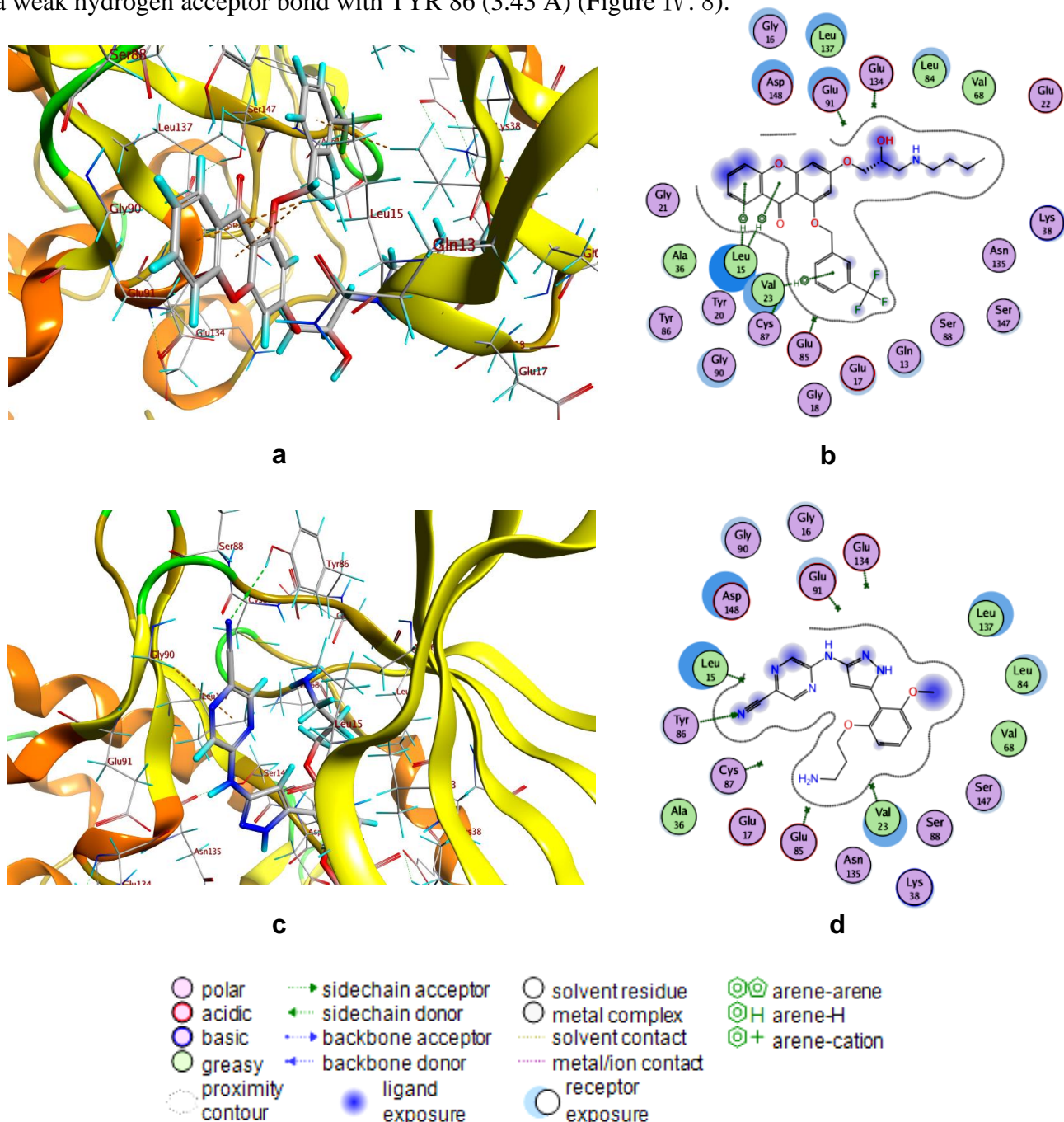
Figure Ⅳ. 8 Three dimensional representations of the best pose docked within 7AKO active site, namely L31 (a), and Prexasertib (c). Two dimensional representations describing the molecular interactions between the protein-ligand complex, namely L31 (b), and Prexasertib (d)

IV.3.2 MD simulation

MD studies are used to elucidate the dynamic behavior of target-ligand complex throughout simulation time. The flexibility of both receptor and ligand is incorporated in MD simulations allowing further optimization of ligand conformation retrieved from docking studies. MD simulation study was performed for the studied eight compounds along with the standard drug Prexasertib against both targets to examine the binding stability of the target-ligand complex in physiological conditions.

The docking complexes were used as input to the simulation process for 1000 ps time intervals each, and the value of potential energy was collected in every 0.5 ps. Figures Ⅳ. 9 and Ⅳ. 1 represent the graph of potential energy U (Kcal/mol) as a function of simulation time (ps) for both structures 7AKM and 7AKO, respectively.
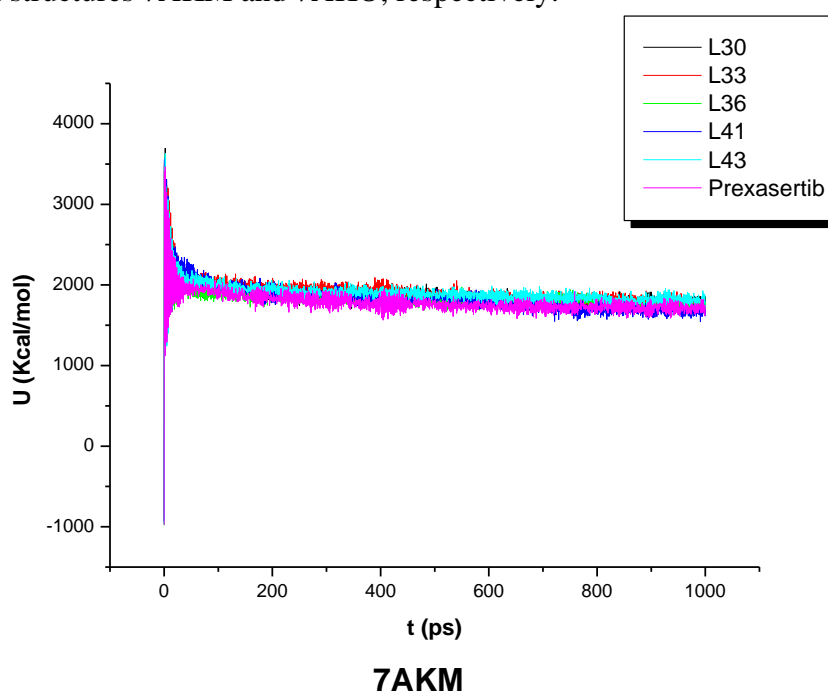
7AKM

Figure Ⅳ. 9 Potential energy plot of the top-scoring five hits and standard Prexasertib with 7AKM as a function of simulation time

The docked xanthone complexes superimposed very well with 7AKM-Prexasertib system during the whole run with some slight deviations between them as illustrated in (Figure Ⅳ. 9). 7AKM-Prexasertib complex displayed noticeable fluctuations during the first 40 ps as the potential energies changed from -978.49 to 1979.59 Kcal/mol, then stabilized until the end of simulation. On the other hand, 7AKO-L42 system exhibited energy variation from -968.93 to 2009.16 Kcal/mol over the first 40 ps and then stabilized for the rest of simulation time as depicted in (Figure Ⅳ. 10). Moreover, the typical potential energy value

for both targets 7AKM and 7AKO was observed around 2000 kcal/mol. In summary, the results indicate the comparative stability of receptor- ligand complexes during the whole run.
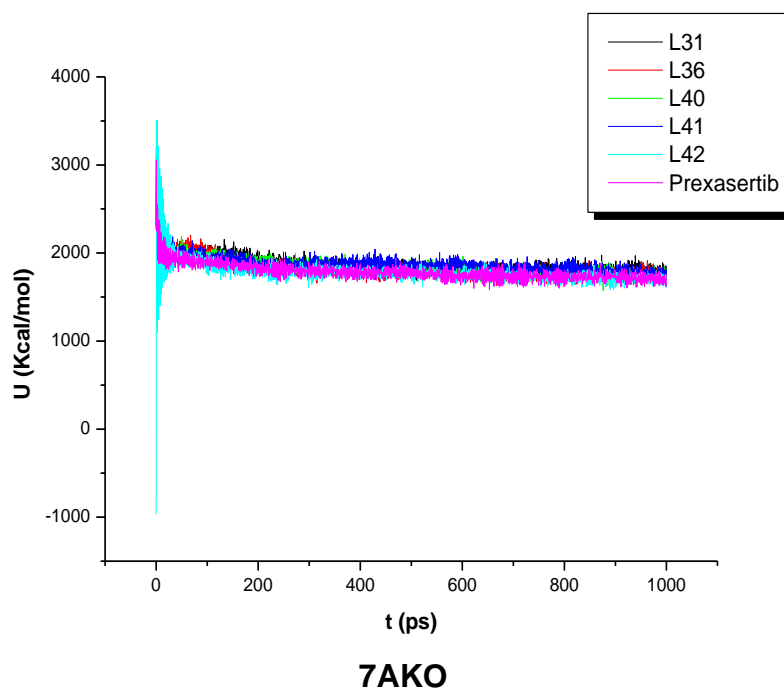
7AKO

Figure Ⅳ. 10 Potential energy plot of the top-scoring five hits and standard Prexasertib with 7AKO as a function of simulation time

The analysis of binding interactions of MD complexes revealed that L43 maintains three binding interactions with 7AKM amino acids: a strong H-acceptor bond with LYS 132 (2.44 Å) and two pi-H bonds toward GLU 134 and ASN 135 (Table Ⅳ. 4). Instead of the four pi-H bonds of L36 formed in molecular docking, a strong H-donor with GLY 18 (2.69 Å) and a weak H-donor bond with CYS 87 (3.81 Å) were observed in MD simulation results (Figure Ⅳ. 11).
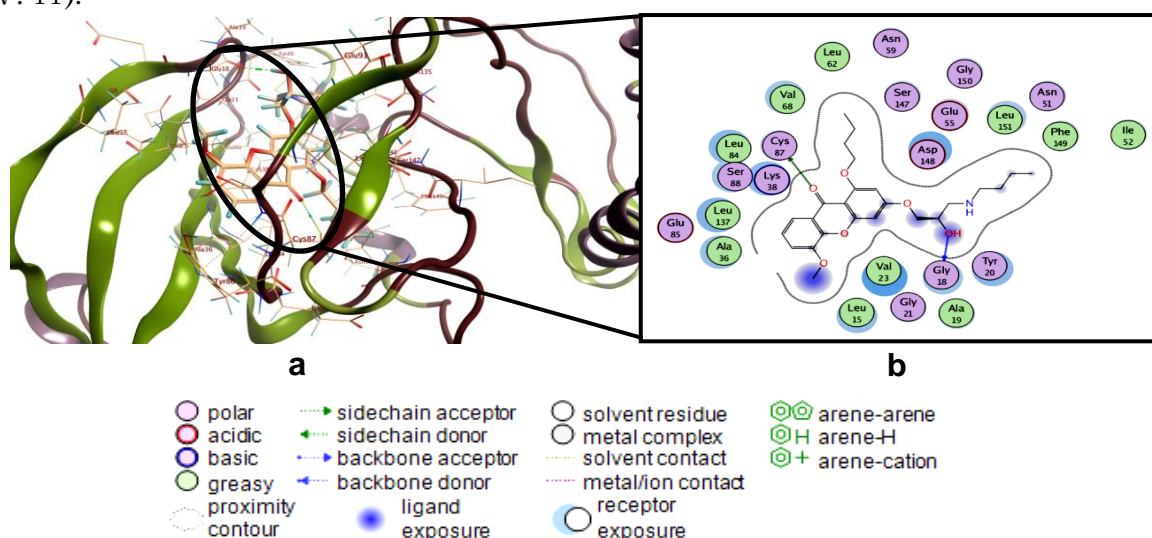
Figure Ⅳ. 11 3D interaction and 2D pose view of 7AKM-L36 derived from molecular dynamics simulations
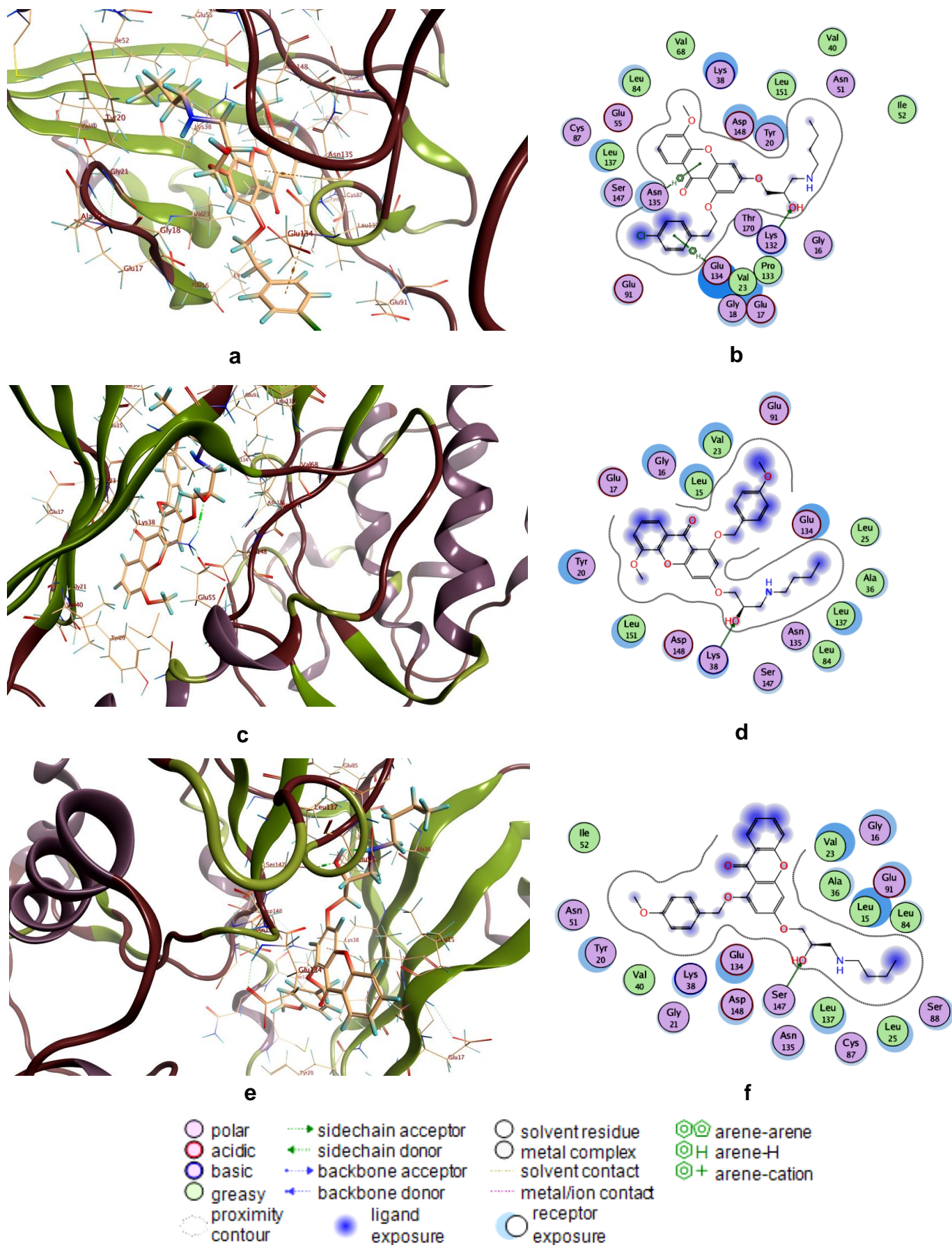
Figure Ⅳ. 12 3D and 2D interaction profiles of L43 (a, b), L41(c, d), and L30 (e, f) into

7AKM captured during molecular dynamics simulations

The complex 7AKM-L41 has a weak H-acceptor bond with LYS 38 (3.23 Å) similarly, this interaction was also shown in molecular docking. Moreover, L30 was involved in strong H-acceptor interaction with SER 147 (2.86 Å) as illustrated in (Figure Ⅳ. 12). However, L33 showed a weak H-acceptor bond with ASN 135 (3.19 Å). Furthermore, Prexasertib developed strong hydrogen donor interaction toward GLU 91 (2.62 Å) (Figure Ⅳ. 13).
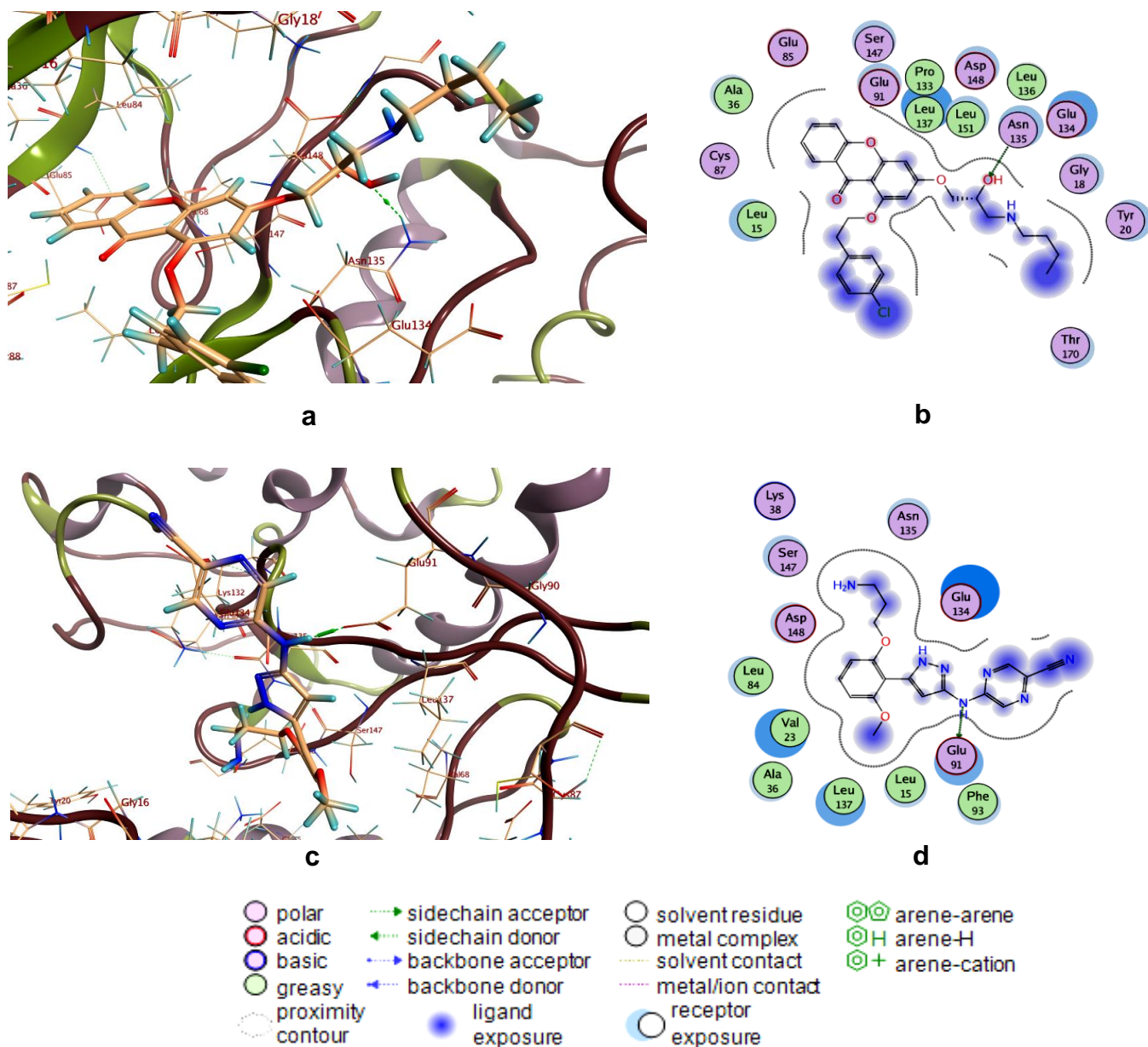
Figure Ⅳ. 13 3D and 2D interaction profiles of L33 (a, b) and Prexasertib (c, d) into 7AKM captured during molecular dynamics simulations

In contrast to the weak pi-H interactions of top ranked ligands with the active site of 7AKO in molecular docking, MD simulations revealed strong hydrogen donor and acceptor interactions where all of them belong to the strong bond range, as well as several weak pi-H

interactions Table B8 (see Appendix B). These findings suggest that the complexes of molecular dynamic simulation are more stable compared to molecular docking results. Therefore, L36 establishes two interactions, one strong H-donor bond with GLU 17 (2.91 Å) and pi-H bond into GLU 134 (Figure Ⅳ. 14).
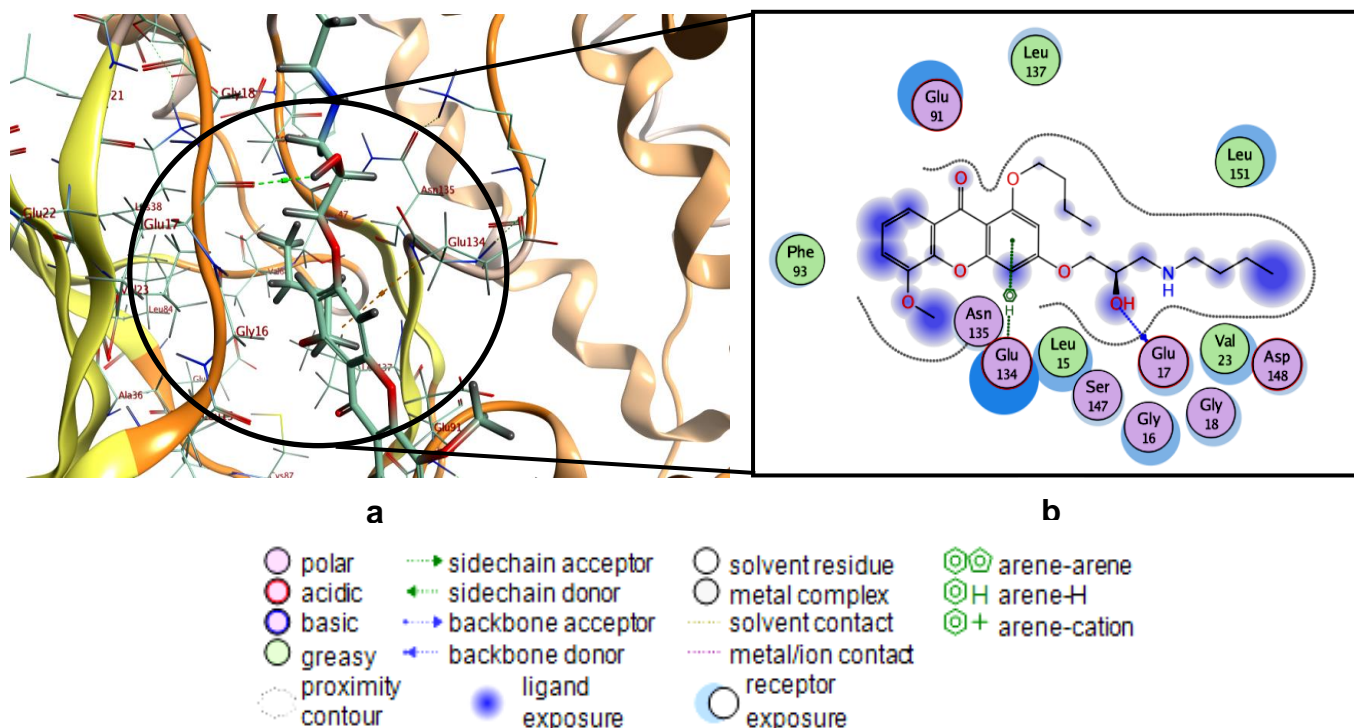
Figure Ⅳ. 14 3D interaction and 2D pose view of 7AKO-L36 derived from molecular dynamics simulations

L40 forms two strong H-donor interactions with the same amino acid GLU 91, one is formed with O30 with a distance value of (2.5 Å) and the second one with N31 (2.75 Å). Additionally, L40 shows two hydrophobic interactions with VAL 23. Moreover, O32 and N33 atoms of L42 established two strong H-donor bonds with GLU 91 with distance values of 2.61 and 2.75 Å, respectively. At the same time, two strong H-acceptor interactions were observed with TYR 20 (2.92 Å) and GLY 21 (2.67 Å).

L41 was the only ligand that maintains the same type of interaction with the same residue, as was observed in molecular dynamic calculation, pi-H bond with LEU 15 (Figure Ⅳ. 15). In the case of L31, a strong H-acceptor interaction was noted with LYS 132 (3.11 Å). Five binding interactions were observed for Prexasertib, two hydrogen donor bonds toward GLU 134 (2.74 Å) and GLU 91 (2.65 Å), hydrogen acceptor bond with TYR 86 (2.6 Å), and two hydrophobic interactions (pi-H) toward LEU 137 (Figure Ⅳ. 16).
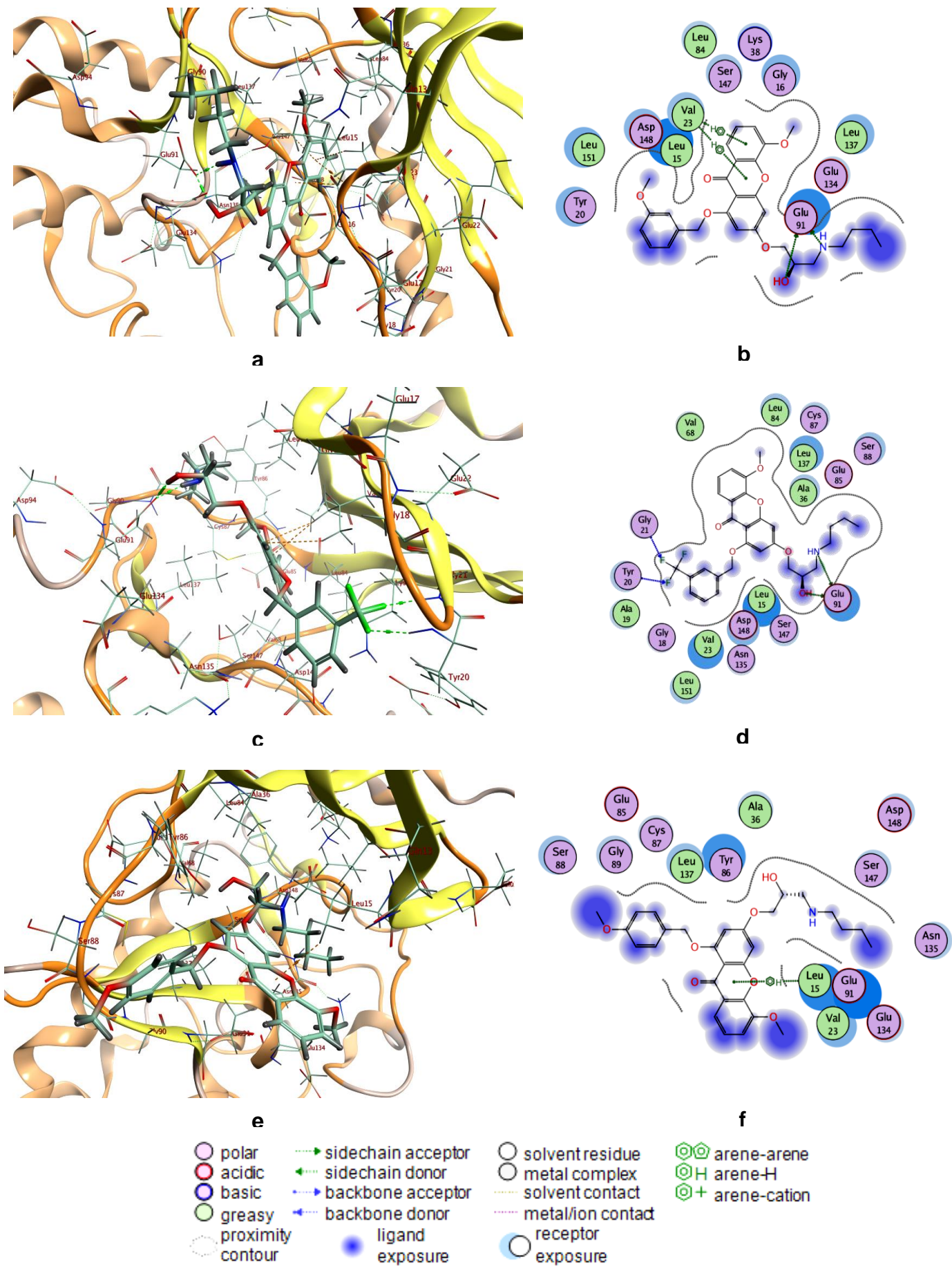
Figure Ⅳ. 15 3D and 2D interaction profiles of L40 (a, b), L42 (c, d), and L41 (e, f) into

7AKO captured during molecular dynamics simulations
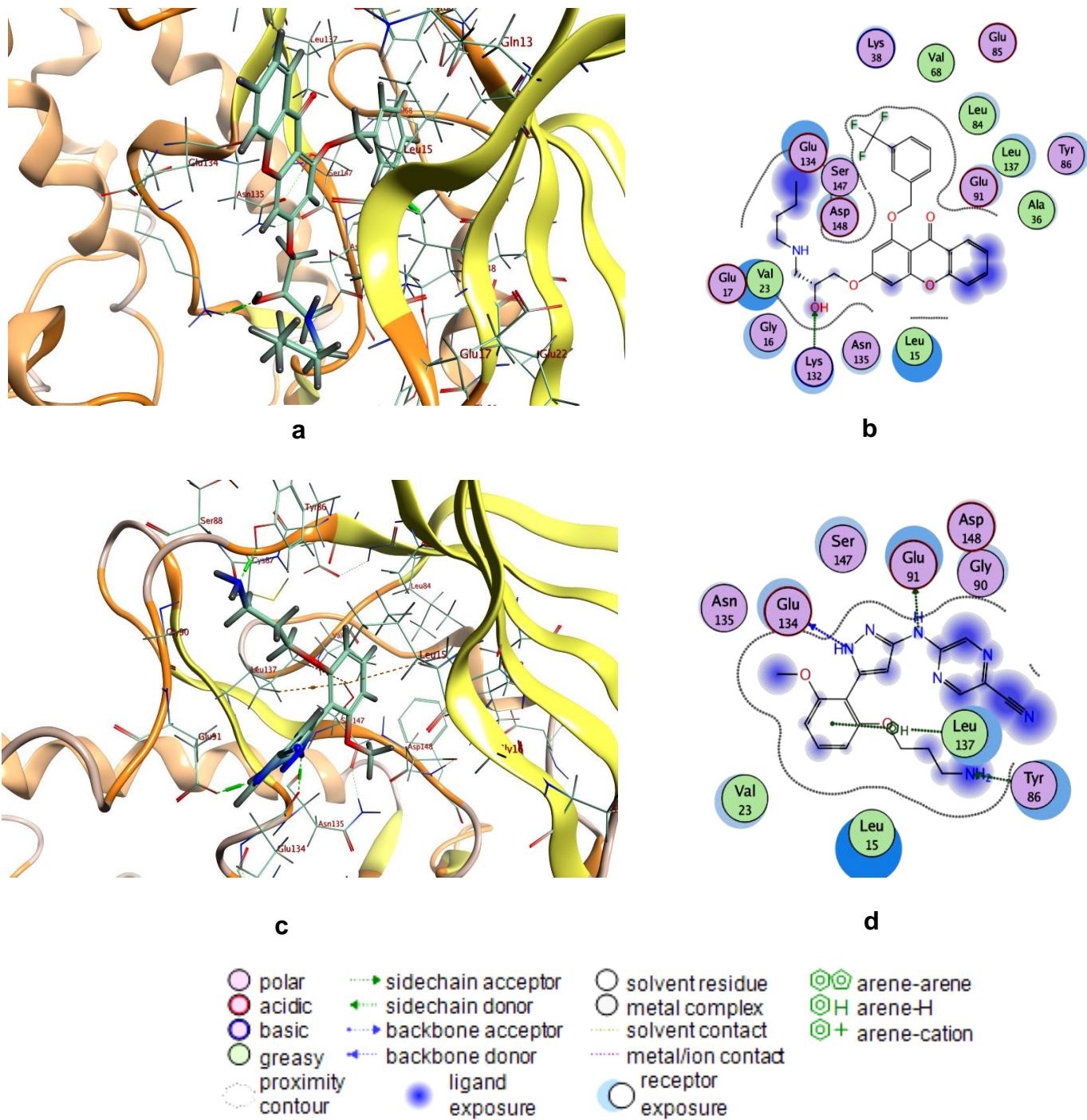
Figure Ⅳ. 16 3D and 2D interaction profiles of L31 (a, b) and Prexasertib (c, d) into 7AKO captured during molecular dynamics simulations

Taken together, MD results support molecular docking results and demonstrate the stability of studied complexes in physiological conditions. These findings confirm that the selected eight xanthone compounds are verifiable CHK1 inhibitors implying a good correlation between the *in silico* and *in vitro* studies.

IV.3.3 *In silico* analysis of ADME-TOX properties

The ADME-TOX properties of the top eight ranked compounds and the standard Prexasertib are presented in (Tables IV.3 and IV.4). The results obtained indicate that the xanthone analogs have poor water solubility (from -6.15 to -4.49) compared with the standard Prexasertib (-2.78). As shown in (Table IV.3), all compounds exhibited high intestinal absorption with percentages ranging from 91.34% to 95.58%. On the other hand, the reference molecule Prexasertib displayed a lower HIA percentage of 76.3%. The predicted apparent permeability of Caco-2 cell line pointed that the majority of compounds can transport through the intestinal mucosa membrane, thus only L40 and L41 as their values are below 0.9. Prexasertib showed lower caco-2 permeability with a value of 0.69. As can be seen, all candidate compounds are estimated to be P-gp substrates, as well as P-gp I and P-gp II inhibitors. By contrast, Prexasertib was found to be P-gp substrate but P-gpI and P-gpII non-inhibitor. Nearly all compounds have moderate distribution with values from -0.12 to 0.11 except for L36, which is estimated to be highly distributed with the highest value of 0.62. However, Prexasertib gave better distribution with a value of 1.18, suggesting that this drug is distributed more in tissue rather than in plasma. Furthermore, five derivatives L31, L40, L41, L42, and L43, are expected to have poor BBB permeability, while L30 L33 and L36 have moderate BBB permeability. Similarly, Prexasertib is predicted to have poor BBB permeability by displaying a lower value of -1.74.

Table IV.3 ADME-TOX prediction including absorption, distribution, metabolism, excretion, and toxicity properties of the selected eight compounds and the standard Prexasertib

Category	Model name	L30	L31	L33	L36	L40	L41	L42	L43	Prexasertib
Absorption	Water solubility	-5.06	-5.89	-6.15	-5.15	-4.49	-4.5	-5.34	-5.53	-2.78
	Caco-2 permeability	1.24	1.26	1.23	1.31	0.85	0.88	1.35	1.31	0.69
	HIA	94.98	91.34	92.77	94.77	95.53	95.58	92.61	94.05	76.3
	P-gp substrate	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	P-gp I inhibitor	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
	P-gp II inhibitor	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No

Category	Model name	L30	L31	L33	L36	L40	L41	L42	L43	Prexasertib
Distribution	VDss	-0.12	0.11	0.01	0.62	-0.09	-0.12	0.07	-0.04	1.18
	BBB permeability	-0.96	-1.17	-0.91	-0.98	-1.18	-1.19	-1.4	-1.14	-1.74
Metabolism	CYP2D6 substrate	No	No	No	No	No	No	No	No	No
	CYP3A4 substrate	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
	CYP1A2 inhibitor	No	No	No	No	No	No	No	No	No
	CYP2C19 inhibitor	No	Yes	Yes	Yes	No	No	Yes	Yes	No
	CYP2C9 inhibitor	No	Yes	Yes	No	No	No	No	Yes	No
	CYP2D6 inhibitor	No	No	No	No	No	No	No	No	No
	CYP3A4 inhibitor	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Excretion	Clearance	6.22	5.09	5.18	6.87	8.34	8.18	7.2	6.97	3.47
	Renal OCT2 substrate	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
	$T_{1/2}$	0.1	0.04	0.05	0.17	0.16	0.14	0.07	0.07	0.24

Continued Table Ⅳ. 3

Category	Model name	L30	L31	L33	L36	L40	L41	L42	L43	Prexasertib
Toxicity	AMES toxicity	No	No	No	No	No	No	No	No	No
	Max tolerated dose	0.42	0.46	0.42	0.43	0.36	0.35	0.4	0.36	0.57
	HERG I inhibitor	No	No	No	No	No	No	No	No	No
	HERG II inhibitor	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	rat toxicity (LD50)	2.36	2.32	2.29	2.39	2.41	2.4	2.36	2.33	2.24
	Hepatotoxicity	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
	Skin Sensitisation	No	No	No	No	No	No	No	No	No

ADME-TOX cut-offs: Caco-2 permeability high if >0.90, HIA poorly absorbed if <30%, VDss considered low if VDss<-0.15 and high if VDss>0.45, BBB permeability is poor when less than -1 and can cross when above 0.3, clearance is high when >15 ml/min/kg moderate if 5-15 ml/min/kg and low if CL<5

Regarding the CYP450 inhibition interactions, xanthone analogs were found to be both CYP3A4 substrate and inhibitor. While none of them inhibited CYP2D6 nor, act as CYP2D6 substrates. Moreover, our compounds were inert toward CYP1A2 and CYP2D6 isoenzymes. Results revealed that L31, L33, L36, L42, and L43 might inhibit CYP2C19, in contrast to the remaining compounds such as L30, L40, and L41 that were non-inhibitors. The prediction suggests no inhibition of the compounds toward CYP2C9 except for L31, L33, and L43. In contrast, the reference inhibitor Prexasertib was totally inert toward all CYP450 isoenzymes.

Results indicate that xanthone derivatives have moderate clearance as their values range from 5.09 to 8.34 ml/min/kg. While Prexasertib has a low clearance with a value of 3.47 compared to other compounds. It was also shown that all compounds are predicted to be

OCT2 substrates. Conversely, Prexasertib is expected to be OCT2 non-substrates. Half-life represents the time required for the plasma concentration of a drug to decrease by 50%. The predicted values range between 0.04 and 0.17. L36 gave the highest half time of 0.17 h. When compared to xanthone compounds, Prexasertib shows a higher half life time of 0.24.

In terms of toxicity, the Ames test indicated the absence of mutagenicity in all compounds. Likewise, no skin sensitisation was detected for them. They also showed relatively low max tolerated dose values between 0.35 and 0.46. Whereas Prexasertib exhibited higher max tolerated dose of 0.574. The studied xanthone compounds are not inhibitors of hERG I however, they act as HERG II inhibitors, which are comparable with the standard drug Prexasertib. The oral rat acute toxicity (LD50) values of our compounds were predicted to range from 2.29 to 2.41. Among the studied derivatives, L40 was the most toxic with a value of 2.41, compared to Prexasertib, which gives a score of 2.24. Another important safety concern in drug discovery was predicted by the hepatotoxicity descriptor, indicating that L36 is the only compound that showed no hepatocellular toxicity compared to other derivatives. Meanwhile, Prexasertib is predicted to be hepatotoxic. Despite the therapeutic potency of Prexasertib in clinical trials, the safety of its use is hampered by several adverse side effects, including hepatotoxicity, hematotoxicity (such as neutropenia, leukopenia, anemia, fatigue, nausea), and lung infection [28]. Therefore, the development of novel CHK1 inhibitors with superior efficacy and better tolerability in patients is of great importance in combating cancer.

Table Ⅳ. 4 Medicinal chemistry profiles and drug-likeness parameters predicted for the top eight hits and the standard Prexasertib

Category	Model name	L30	L31	L33	L36	L40	L41	L42	L43	Prexasertib
Medicinal chemistry	Lipinski Rule	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes
	Pfizer Rule	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	GSK Rule	No	No	No	No	No	No	No	No	Yes
	Golden Triangle	Yes	No	Yes	Yes	No	No	No	No	Yes
	SAscore	2.84	3.01	2.9	2.95	2.99	2.96	3.12	3.01	3.44
	Pains	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Almost all compounds accept Lipinski rule except for L31, L42, and L43 as described in (Table Ⅳ. 4). Furthermore, the selected analogs satisfy Pfizer rule however none of them accept GSK rule. Besides, the compounds did not fulfill golden triangle rule except for L30, L33, and L36. However, the standard drug Prexasertib accepts Lipinski, Pfizer, GSK, and golden triangle rules. The synthetic accessibility score estimates the ease of synthesis of drug-like molecules and is expected to be good for our xanthones with score values less than 6, which are considered easy to synthesize. Likewise, Prexasertib showed a similar SA score of 3.44.

These findings suggest that L36 has the best ADME-TOX profile as it was the only hit without hepatotoxicity among the studied compounds. Furthermore, it is a non-mutagenic compound, hERG I non-inhibitor, and does not induce skin sensitisation also exhibits the highest max tolerated dose, highest $T_{1/2}$, and relatively high Caco and HIA absorption. It also displayed superior binding affinity and satisfied Lipinski, Pfizer, and golden triangle rules.

The concept of bioavailability radar plot is very useful in the drug design process, therefore 13 important physicochemical properties were plotted to visually evaluate the drug-likeness of L36 and Prexasertib, as illustrated in (Figure Ⅳ. 17). The orange area between the upper and lower limits represents the ideal range of physicochemical properties. Hence, these limits are determined by ADMELAB 2.0 tool based on Drug-Like soft rule as follows MW (100,600), nHA (0,12) , nHD (0,7) , nRot (0,11), nRing (0,6), MaxRing (0,18), nHet (1,15), fChar (-4,4), nRig (0,30), TPSA (0,140), LogS (-4,0.5), LogP (0,3) and Log D (1,3). Figure Ⅳ. 17 revealed that L36 falls outside the optimal range in four properties, among which three are above the upper limit (LogP, LogD, and nRot), and one is below the lower limit (LogS). The remaining nine characteristics are within the ideal scope. These findings indicate that decreasing the number of rotatable bonds and increasing the solubility would improve the bioavailability profile of this compound. This could be achieved by introducing hydrophilic substituents into the xanthone scaffold to reduce the lipophilic character and thus enhance its solubility. On the other hand, Prexasertib showed a set of favorable physicochemical properties within the optimal range, denoting a potential drug candidate. Collectively, L36 remains the best hit among the studied xanthones in terms of binding affinity and toxicity profile. It also displays a good bioavailability profile
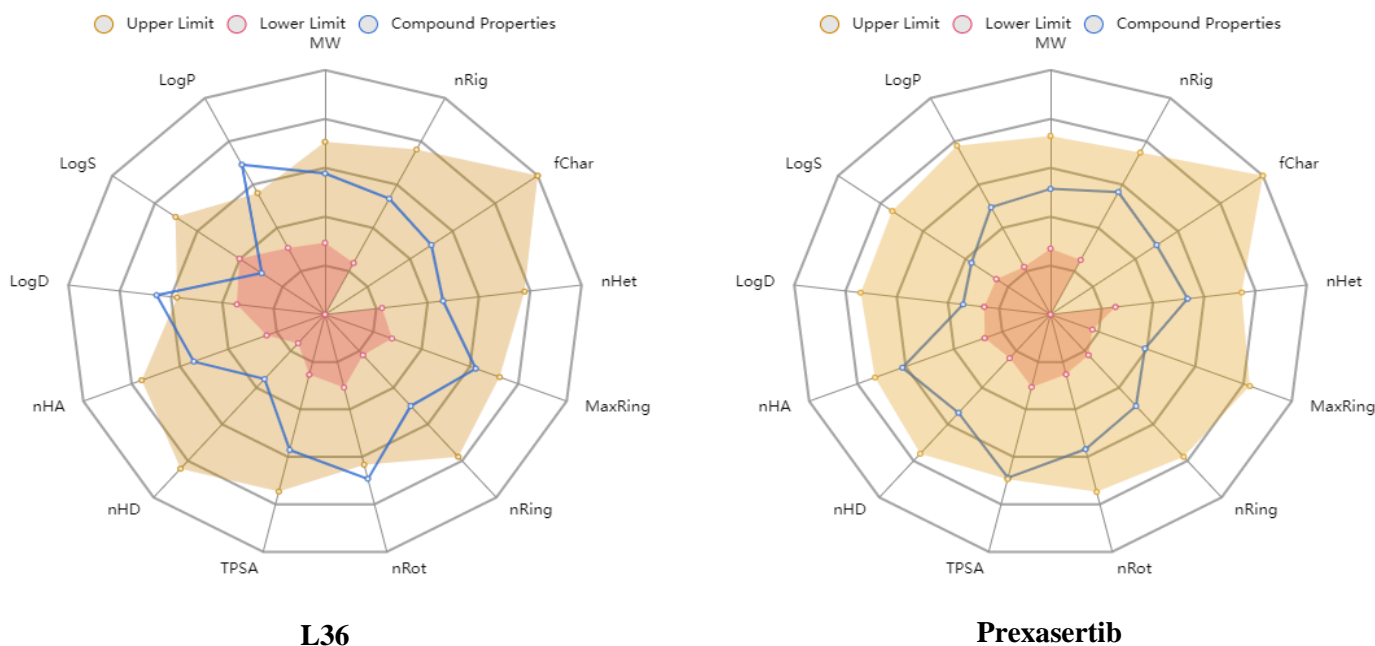
L36   **Prexasertib**

Figure Ⅳ. 17 Bioavailability radar chart of L36 and standard Prexasertib

IV.3.4   Quantitative MEP analysis of L36

MEP is a robust tool for visualizing and predicting non-covalent molecular interactions occurring without charge transfer [29]. The sign of electrostatic potential at point r is determined by one of the two dominant charge distributions (electrons and nuclei). Regions having a positive electrostatic potential are characterized by a low concentration of electrons, justifying their electrophilic character. A negative electrostatic potential indicates a region of high electron density capable of undergoing electrophilic attack. The map of ESP van der Waals surfaces along with surface extrema and some of statistically defined descriptors of the selected compound L36 is depicted in (Figure Ⅳ. 18). The blue regions are prone to nucleophilic attack, and the red regions are sites for electrophilic attack. Orange and cyan spheres correspond to the positions of maxima and minima, respectively. As can be seen in (Figure Ⅳ. 18), the large positive surface of L36 occupies 63 % of the total surface, implying electrophilic zones vulnerable to nucleophilic attack. The global surface minimum with a value of -54.23 Kcal/mol is observed around the carbonyl of xanthone scaffold, denoting the most electron-rich site, vulnerable for electrophilic attack. While the global surface maximum with a value of 43.33 Kcal/mol is located around the hydrogen of hydroxyl substituent, indicating potential hydrogen bond donor and revealing the low electronic density in this area caused by the withdrawing effect of oxygen linked to xanthone moiety.

However, the strong electron-withdrawing effect of amine group is also taken into account. Furthermore, the contribution of NV is about 79.40% compared to the overall variance, expressed as the sum of positive and negative contributions. This reveals that the degree of variability in the negative regions is much more pronounced compared to the positive regions. Collectively, MEP analysis indicates that the hydroxyl substituent is the most reactive site in L36. Interestingly, these findings positively correlate with MD results that showed that the hydroxyl function of L36 maintains H-donor interaction with the target pocket residues of 7AKM and 7AKO with GLY 18, and GLU 17, respectively.
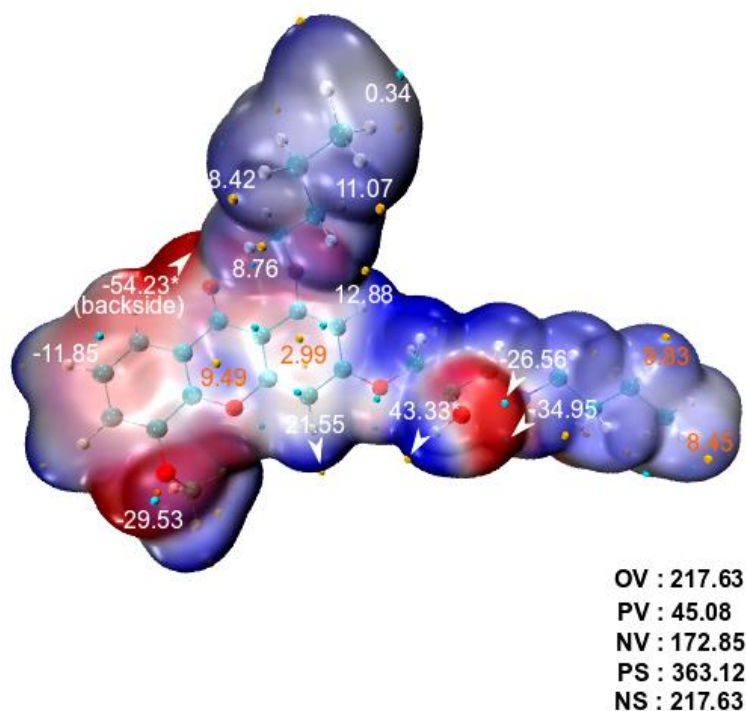
Figure IV. 18 ESP-mapped van der Waals surfaces (kcal/mol) of L36, using a color scale ranging from red (negative ESP) through white (neutral ESP) to blue (positive ESP). The blue regions indicate a vulnerable site for nucleophilic attack, and the red regions are sites for electrophilic attack. All the iso-surface maps were rendered by VMD software based on the surface analysis result of Multiwfn program. The grid spacings were set to 0.2 Bohr, and the van der Waals surface denotes the iso-surface of $\rho = 0.001 \ e/bohr^3 \ \equiv a.u.$ Value with a star indicates the global extrema. The bold numbers in the bottom right-hand corner are the overall ESP variance (OV), positive variance (PV), negative variance (NV), positive surface area (PS), and negative surface area (NS), whose units are $[Kcal/mol]^2, Å^2$, respectively
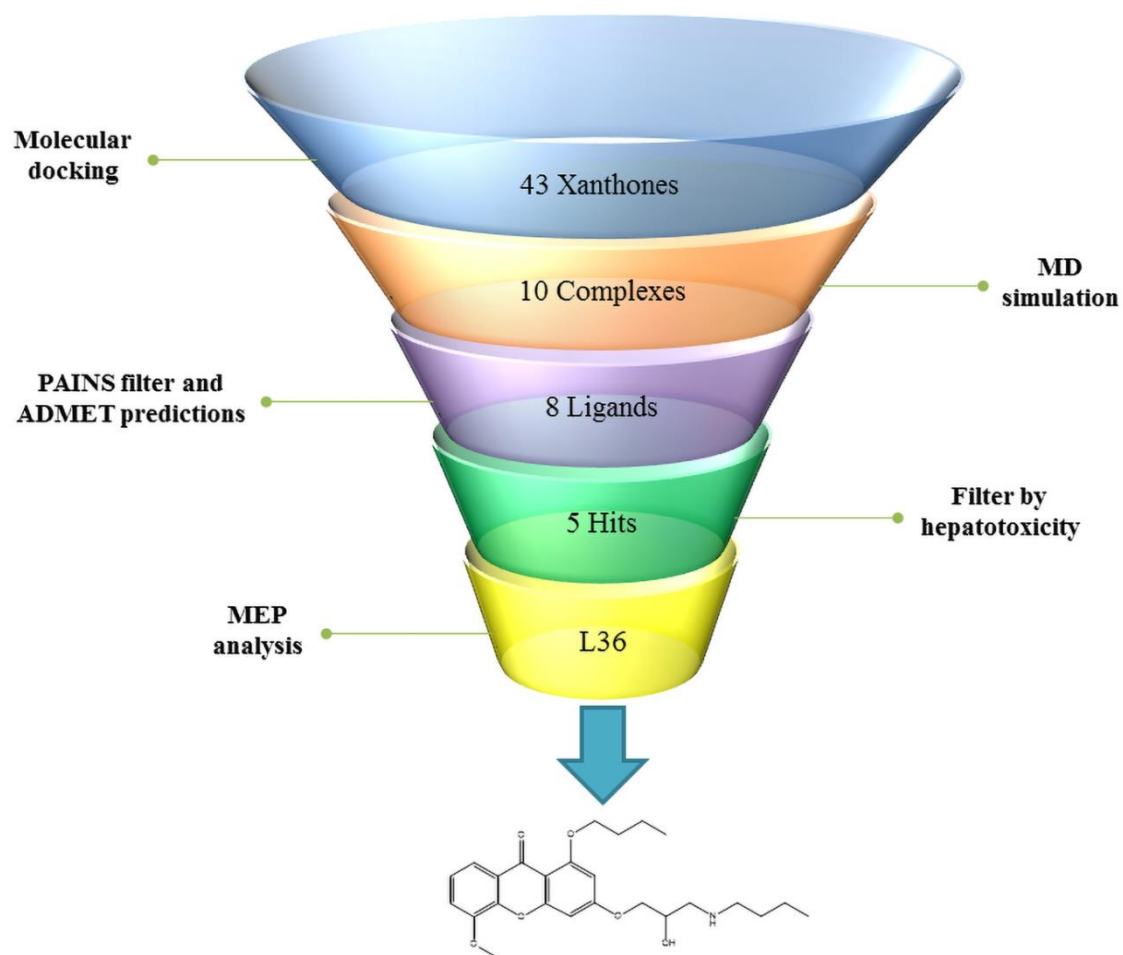
Figure Ⅳ. 19 Flowchart of the strategy used in the study

IV.4 Conclusion

In this paper, a simulation-based investigation was conducted to examine the CHK1 inhibitory activity of cytotoxic xanthone derivatives using a hierarchical workflow for molecular docking, MD simulation, ADME-TOX prediction, and MEP analysis. The forty-three xanthone derivatives revealed various interactions toward CHK1 protein structures, and the free energy scores S were in the range of [-8.22,-6.33Kcal/mol] and [-8.14,-6.19Kcal/mol] for 7AKM and 7AKO, respectively. Eight top hits were identified based on their binding affinity energies, namely L43, L42, L41, L40, L36, L33, L31, and L30. Interestingly, these compounds showed better binding affinity toward 7AKM and 7AKO active sites than the standard drug Prexasertib, which emphasizes the validity of our strategy. For further validation, an MD simulation study was performed for the studied eight compounds, together with the standard drug Prexasertib against both targets to examine the binding stability of the target-ligand- complex. Taken together, MD results support molecular docking results and validate the stability of studied complexes in physiological conditions. These findings confirm that the selected eight xanthone compounds are verifiable CHK1 inhibitors implying a good correlation between *in silico* and *in vitro* studies. Moreover, *in silico* analysis of ADME-TOX properties revealed that the selected eight hits exhibit good ADME-TOX profiles and are comparable to Prexasertib. Among the top eight hits, five compounds satisfy Lipinski rule vis., L30, L33, L36, L40, and L41. Furthermore, L36 shows the best ADME-TOX profile as it was the only hit without hepatotoxicity among the studied compounds. Besides, it is non-mutagenic compound, hERG I non-inhibitor, and does not induce skin sensitization also exhibits the highest max tolerated dose, highest $T_{1/2}$ and relatively high Caco and HIA absorption. Also, it displayed superior binding affinity and satisfied Lipinski, Pfizer, and golden triangle rules indicating a potent drug candidate. The quantitative analysis of electrostatic potential of the lead compound L36 indicates that the hydroxyl substituent is the most reactive site. Our study provides new unexplored insights into xanthones as CHK1 inhibitors and identified L36 as a potential drug candidate that could undergo further *in vivo* assays and optimization, laying a solid foundation for the development of CHK1 inhibitors and cancer drug discovery.

IV.5 References

1. Rasool N, Majeed A, Riaz F, Hussain W (2020) Identification of novel inhibitory candidates against two major Flavivirus pathogens via CADD protocols: in silico analysis of phytochemical binding, reactivity, and pharmacokinetics against NS5 from ZIKV and DENV. Struct Chem 31:2189–2204. https://doi.org/10.1007/s11224-020-01577-x

2. Park S, Hong E, Kwak SY, et al (2016) Synthesis and biological evaluation of C1- O -substituted-3-(3-butylamino-2-hydroxy-propoxy)-xanthen-9-one as topoisomerase IIα catalytic inhibitors. European Journal of Medicinal Chemistry 123:211–225. https://doi.org/10.1016/j.ejmech.2016.07.046

3. Day M, Parry-Morris S, Houghton-Gisby J, et al (2021) Structural basis for recruitment of the CHK1 DNA damage kinase by the CLASPIN scaffold protein. Structure 29:531-539.e3. https://doi.org/10.1016/j.str.2021.03.007

4. Clément G, Slenzka K (2006) Fundamentals of Space Biology: Research on Cells, Animals, and Plants in Space. Springer New York, New York, NY

5. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. Proteins 12:345–364. https://doi.org/10.1002/prot.340120407

6. Molecular Operating Environment (MOE) (2014) Chemical Computing Group Inc, 1010 Sherbrooke St. West, Suite # 910, Montreal QC, Canada H3A 2R7

7. Daoud I, Mesli F, Melkemi N, et al (2021) Discovery of potential SARS-CoV 3CL protease inhibitors from approved antiviral drugs using: virtual screening, molecular docking, pharmacophore mapping evaluation and dynamics simulation. Journal of Biomolecular Structure and Dynamics 1–18. https://doi.org/10.1080/07391102.2021.1973563

8. Belkadi A, Kenouche S, Melkemi N, et al (2021) K-means clustering analysis, ADME/pharmacokinetic prediction, MEP, and molecular docking studies of potential cytotoxic agents. Struct Chem 32:2235–2249. https://doi.org/10.1007/s11224-021-01796-w

9. Djebaili R, Melkemi N, Kenouche S, et al (2021) Combined Conceptual-DFT, Quantitative MEP Analysis, and Molecular Docking Study of Benzodiazepine Analogs. Orbital: Electron J Chem 13:301–315. https://doi.org/10.17807/orbital.v13i4.1607

10. MarvinSketch (2019) ChemAxon. http://www.chemaxon.com/products/marvin/marvinsketch/. Accessed 16 Jul 2019

11. HyperChem v8 (2009) Molecular Modeling System, Hypercube Inc, 1115 NW 4th Street, Gainesville, FL 32601, USA

12. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian 09, Revision A.02, Gaussian, Inc, Wallingford CT

13. Chenafa H, Mesli F, Daoud I, et al (2021) In silico design of enzyme α-amylase and α-glucosidase inhibitors using molecular docking, molecular dynamic, conceptual DFT investigation and pharmacophore modelling. Journal of Biomolecular Structure and Dynamics 1–22. https://doi.org/10.1080/07391102.2021.1882340

14. Daoud I, Melkemi N, Salah T, Ghalem S (2018) Combined QSAR, molecular docking and molecular dynamics study on new Acetylcholinesterase and Butyrylcholinesterase inhibitors. Computational Biology and Chemistry 74:304–326. https://doi.org/10.1016/j.compbiolchem.2018.03.021

15. Bond SD, Leimkuhler BJ, Laird BB (1999) The Nosé–Poincaré Method for Constant Temperature Molecular Dynamics. Journal of Computational Physics 151:114–134. https://doi.org/10.1006/jcph.1998.6171

16. Sturgeon JB, Laird BB (2000) Symplectic algorithm for constant-pressure molecular dynamics using a Nosé–Poincaré thermostat. The Journal of Chemical Physics 112:3474–3482. https://doi.org/10.1063/1.480502

17. Parikesit AA, zahroh H, Nugroho AS, et al (2015) The Computation of Cyclic Peptide with Prolin-Prolin Bond as Fusion Inhibitor of DENV Envelope Protein through Molecular Docking and Molecular Dynamics Simulation. arXiv:151101388 [q-bio]. https://doi.org/10.13140/2.1.4133.3760

18. Berendsen HJC, Postma JPM, van Gunsteren WF, et al (1984) Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics 81:3684–3690. https://doi.org/10.1063/1.448118

19. Origin (Pro) (6.0) (1999) OriginLab Corporation, Northampton, MA, USA

20. Bolz SN, Adasme MF, Schroeder M (2021) Toward an Understanding of Pan-Assay Interference Compounds and Promiscuity: A Structural Perspective on Binding Modes. J Chem Inf Model 61:2248–2262. https://doi.org/10.1021/acs.jcim.0c01227

21. Baell JB, Holloway GA (2010) New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. J Med Chem 53:2719–2740. https://doi.org/10.1021/jm901137j

22. Pires DEV, Blundell TL, Ascher DB (2015) pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. J Med Chem 58:4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104

23. Xiong G, Wu Z, Yi J, et al (2021) ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res 49:W5–W14. https://doi.org/10.1093/nar/gkab255

24. Lu T, Chen F (2012) Multiwfn: A multifunctional wavefunction analyzer. J Comput Chem 33:580–592. https://doi.org/10.1002/jcc.22885

25. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. Journal of Molecular Graphics 14:33–38. https://doi.org/10.1016/0263-7855(96)00018-5

26. Murray JS, Brinck T, Lane P, et al (1994) Statistically-based interaction indices derived from molecular surface electrostatic potentials: a general interaction properties function (GIPF). Journal of Molecular Structure: THEOCHEM 307:55–64. https://doi.org/10.1016/0166-1280(94)80117-7

27. Imberty A, Hardman KD, Carver JP, Perez S (1991) Molecular modelling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A. Glycobiology 1:631–642. https://doi.org/10.1093/glycob/1.6.631

28. Hong D, Infante J, Janku F, et al (2016) Phase I Study of LY2606368, a Checkpoint Kinase 1 Inhibitor, in Patients With Advanced Cancer. JCO 34:1764–1771. https://doi.org/10.1200/JCO.2015.64.5788

29. Murray JS, Politzer P (2011) The electrostatic potential: an overview. WIREs ComputMolSci 1:153–163. https://doi.org/10.1002/wcms.19

General Conclusion

Despite the advances in medical and pharmaceutical sciences, there is still a great need for more and better drugs with higher activity and fewer undesired side effects. The application of computational technology in drug design and discovery offers considerable potential for reducing development costs. These techniques are widely used to interpret the relationship between biological activity and chemical structure. The main objective of this work is to conduct two independent studies that contribute to the discovery and development of anticancer drugs.

In the first study, a combined approach based on K-means algorithm and statistical analysis has been applied successfully to classify 500 cytotoxic agents using 21 molecular descriptors. This study aims to define a set of rules for structural features, providing a guide to assist the medicinal chemist in selecting the optimal molecules to synthesize and thereby test and promote only the candidates that exhibit strong efficacy, low toxicity, and good pharmacokinetic profiles to improve clinical outcomes of drug therapies. The percentage of molecules in each cluster is 50%, 24.88%, and 25.12% for cluster 1, cluster 2, and cluster 3, respectively. Silhouette analysis is used as a cluster validation method, confirming that all the molecules are very well clustered. Moreover, in silico screening of pharmacological properties ADME and evaluation of drug-likeness of molecules showed that cluster 1 molecules have the best ADME profile and drug-likeness. Additionally, molecular docking was performed for the paragons molecules on six different targets. Docking studies support the results observed in the MEP analysis showing that the favorable reactive sites of molecules are involved in strong hydrogen interactions with the functional residues of receptors.

In the second study, a simulation-based investigation was conducted to examine the CHK1 inhibitory activity of cytotoxic xanthone derivatives using a hierarchical workflow for molecular docking, MD simulation, ADME-TOX prediction, and MEP analysis. A molecular docking study was conducted for the forty-three xanthone derivatives along with standard Prexasertib into the selected CHK1 protein structures 7AKM and 7AKO. Eight top hits were identified based on their free energy scores showed better binding affinity than the reference Prexasertib which emphasizes the validity of our strategy. Furthermore, MD studies support molecular docking results and validate the stability of studied complexes in physiological

conditions. These findings confirm that the selected eight xanthones are verifiable CHK1 inhibitors implying a good correlation between *in silico* and *in vitro* studies. Moreover, L36 showed the best ADME-TOX profile as it was the only hit without hepatotoxicity among the studied compounds. Besides, it displayed superior binding affinity and satisfied Lipinski, Pfizer, and golden triangle rules indicating a potent drug candidate. Finally, the quantitative analysis of electrostatic potential was performed for L36 to identify the reactive sites and possible non-covalent interactions.

Recommendations for future research that would be very beneficial include the development of a ligand-based pharmacophore model by 3D-QSAR pharmacophore generation for the forty-three xanthone derivatives, and thus a 3D query is used for the virtual screening of large databases. The pharmacophore model provides information about the most important structural features for activity which could help in designing new potential CHK1 inhibitors for cancer therapy.

Appendix A

Table A1 Results obtained using the test statistic for cluster 1

Molecular descriptors	z_k	Mean in category	Overall mean	sd in category	Overall sd	p.value
Hydrogen donor	-3.143361	1.3679245	1.5741627	1.242528	1.35918	1.67E-03
heterocyclic rings	-4.34888	1.4575472	1.7703349	1.112998	1.48996	1.37E-05
Log p	-6.551781	-0.2668868	0.8027751	1.876098	3.382124	5.69E-11
double bonds	-8.385037	2.4292453	3.3421053	1.651098	2.255286	5.07E-17
TPSA	-9.473377	73.5779718	93.4725357	32.108369	43.504276	2.71E-21
Hetero-atoms N+O+S	-9.783489	5.6745283	7.2727273	1.843579	3.384071	1.33E-22
Hydrogen donor + acceptor	-10.244272	4.8349057	6.291866	1.926995	2.946246	1.26E-24
Hydrogen acceptor	-11.266461	3.4669811	4.7177033	1.167098	2.299724	1.92E-29
Oxygen atoms	-11.920292	3.1745283	4.6507177	1.530419	2.565414	9.28E-33
Number of rings	-12.129848	3.4339623	4.5430622	1.593139	1.894164	7.34E-34
Rotatable bonds	-12.628486	4.8773585	7.2296651	2.30546	3.858735	1.47E-36
Chiral centers	-12.756367	0.5849057	2.7033493	1.670318	3.440268	2.87E-37
Flexibility	-15.714565	4.2122729	5.9009499	1.23722	2.22611	1.20E-55
Surface grid	-16.519684	562.532217	693.539115	80.058959	164.283809	2.65E-61
Mass	-16.565628	375.954057	496.359833	80.979087	150.571149	1.23E-61
Polarizability	-17.029683	38.9159434	52.5247608	8.229919	16.554504	4.95E-65
Volume	-17.358715	968.352453	1268.97895	166.227816	358.766609	1.69E-67

Table A2 Results obtained using the test statistic for cluster 2

Molecular descriptors	z_k	Mean in category	Overall mean	sd in category	Overall sd	p.value
Log p	16.731131	5.68E+00	0.8027751	1.6925158	3.3821239	7.77E-63
Chiral centers	13.927741	6.83E+00	2.7033493	2.8459925	3.4402679	4.30E-44
Volume	9.468625	1.56E+03	1268.97895	191.625217	358.766609	2.84E-21
Flexibility	8.812227	7.59E+00	5.9009499	1.8476637	2.2261098	1.23E-18
Polarizability	8.585525	6.48E+01	52.5247608	10.1487408	16.5545039	9.04E-18
Surface grid	8.516432	8.14E+02	693.539115	85.4389309	164.283809	1.65E-17
double bonds	6.259117	4.56E+00	3.3421053	2.3413317	2.2552856	3.87E-10
Mass	6.125284	5.76E+02	496.359833	87.701394	150.571149	9.05E-10
Number of rings	3.279142	5.08E+00	4.5430622	1.4732003	1.8941643	1.04E-03
Rotatable bonds	2.13925	7.94E+00	7.2296651	3.3892087	3.8587352	3.24E-02
Dipolar moment	-2.231093	4.63E+00	5.633832	1.7956474	5.2333448	2.57E-02
Hydrogen donor + acceptor	-2.925194	5.55E+00	6.291866	1.6955967	2.9462458	3.44E-03
Hydrogen acceptor	-3.225036	4.08E+00	4.7177033	1.3037373	2.2997238	1.26E-03
TPSA	-3.601449	8.00E+01	93.4725357	24.511897	43.5042761	3.16E-04
Hetero-atoms N+O+S	-5.674194	5.62E+00	7.2727273	1.6688855	3.3840709	1.39E-08
Sulfur atoms	-6.259531	5.88E-02	0.3779904	0.2352941	0.5915625	3.86E-10
heterocyclic rings	-7.219752	8.43E-01	1.7703349	0.8134301	1.4899601	5.21E-13
Nitrogen atoms	-9.449154	8.14E-01	2.2440191	1.1265981	1.7561308	3.42E-21
aromatic rings	-12.165403	8.92E-01	2.5143541	0.7784709	1.5470406	4.75E-34

Table A3 Results obtained using the test statistic for cluster 3

Molecular descriptors	z_k	Mean in category	Overall mean	sd in category	Overall sd	p.value
Hetero-atoms N+O+S	16.951175	12.1538462	7.2727273	2.1562487	3.3840709	1.89E-64
Hydrogen acceptor	16.232926	7.8942308	4.7177033	1.7700406	2.2997238	2.95E-59
Hydrogen donor + acceptor	14.752944	9.9903846	6.291866	2.4396374	2.9462458	2.95E-49
TPSA	14.533308	147.271826	93.4725357	32.1872075	43.5042761	7.45E-48
Oxygen atoms	13.360962	7.5673077	4.6507177	2.2393527	2.5654144	1.02E-40
Mass	13.071631	663.835481	496.359833	80.079922	150.571149	4.78E-39
Rotatable bonds	12.478667	11.3269231	7.2296651	3.0301127	3.8587352	9.76E-36
heterocyclic rings	12.201912	3.3173077	1.7703349	1.5207818	1.4899601	3.04E-34
Polarizability	11.164061	68.2507692	52.5247608	9.506319	16.5545039	6.11E-29
Nitrogen atoms	10.914702	3.875	2.2440191	1.2985754	1.7561308	9.81E-28
Number of rings	10.769558	6.2788462	4.5430622	1.1473777	1.8941643	4.79E-27
Volume	10.667217	1594.62308	1268.97895	225.015247	358.766609	1.45E-26
Surface grid	10.642925	842.316154	693.539115	129.873653	164.283809	1.88E-26
Aromatic rings	9.824948	3.8076923	2.5143541	1.1525634	1.5470406	8.79E-23
Flexibility	9.417991	7.6849095	5.9009499	1.2843569	2.2261098	4.60E-21
Sulfur atoms	6.626399	0.7115385	0.3779904	0.5312011	0.5915625	3.44E-11
Hydrogen donor	4.513423	2.0961538	1.5741627	1.6083869	1.3591802	6.38E-06
double bonds	3.478373	4.0096154	3.3421053	2.431742	2.2552856	5.04E-04
Dipolar moment	3.365505	7.1325148	5.633832	3.5561787	5.2333448	7.64E-04
Log p	-9.045471	-1.8003846	0.8027751	1.8841678	3.3821239	1.49E-19
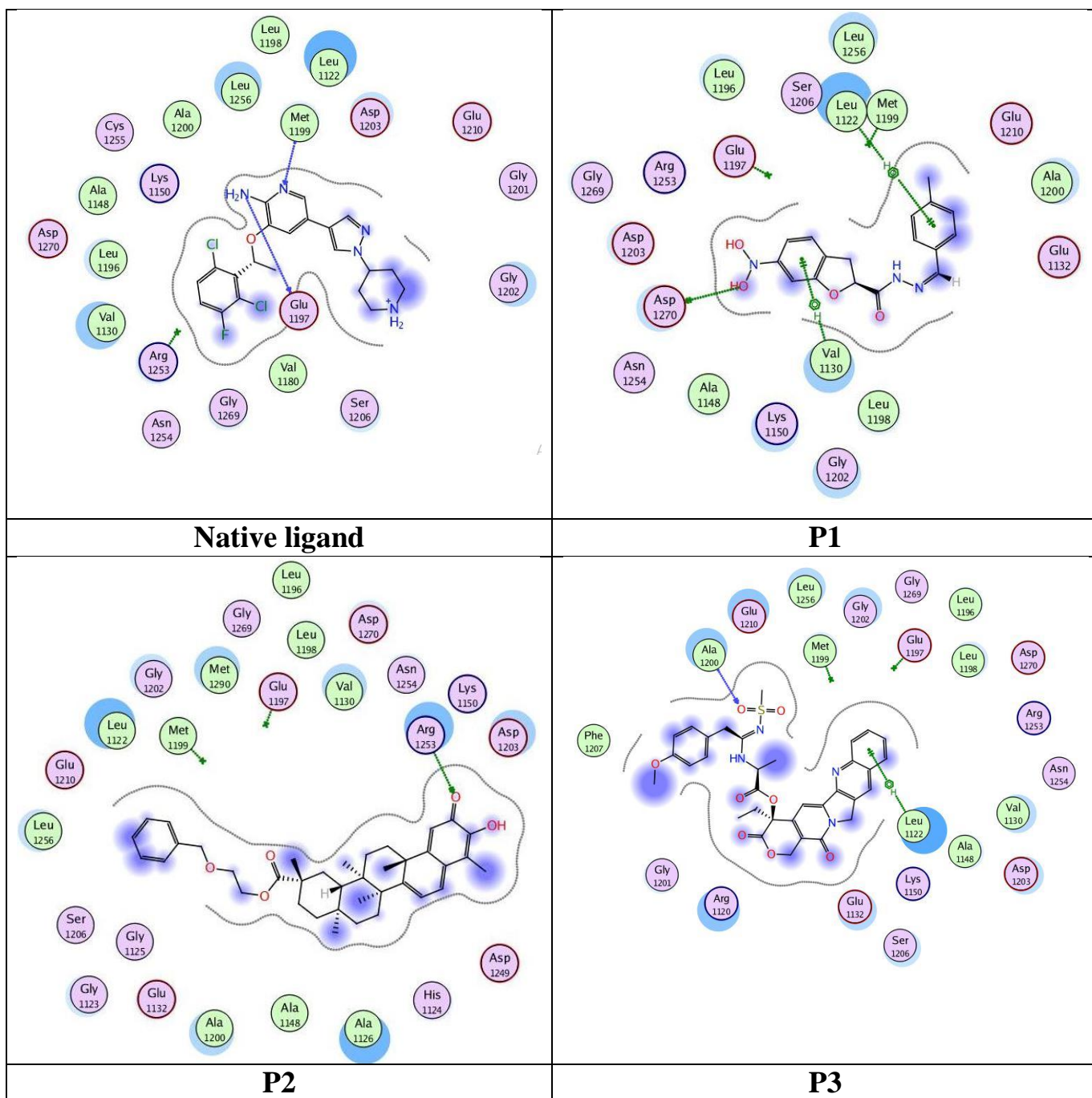
Figure A1 The binding modes of native ligand, P1, P2 and P3 into 2XP2
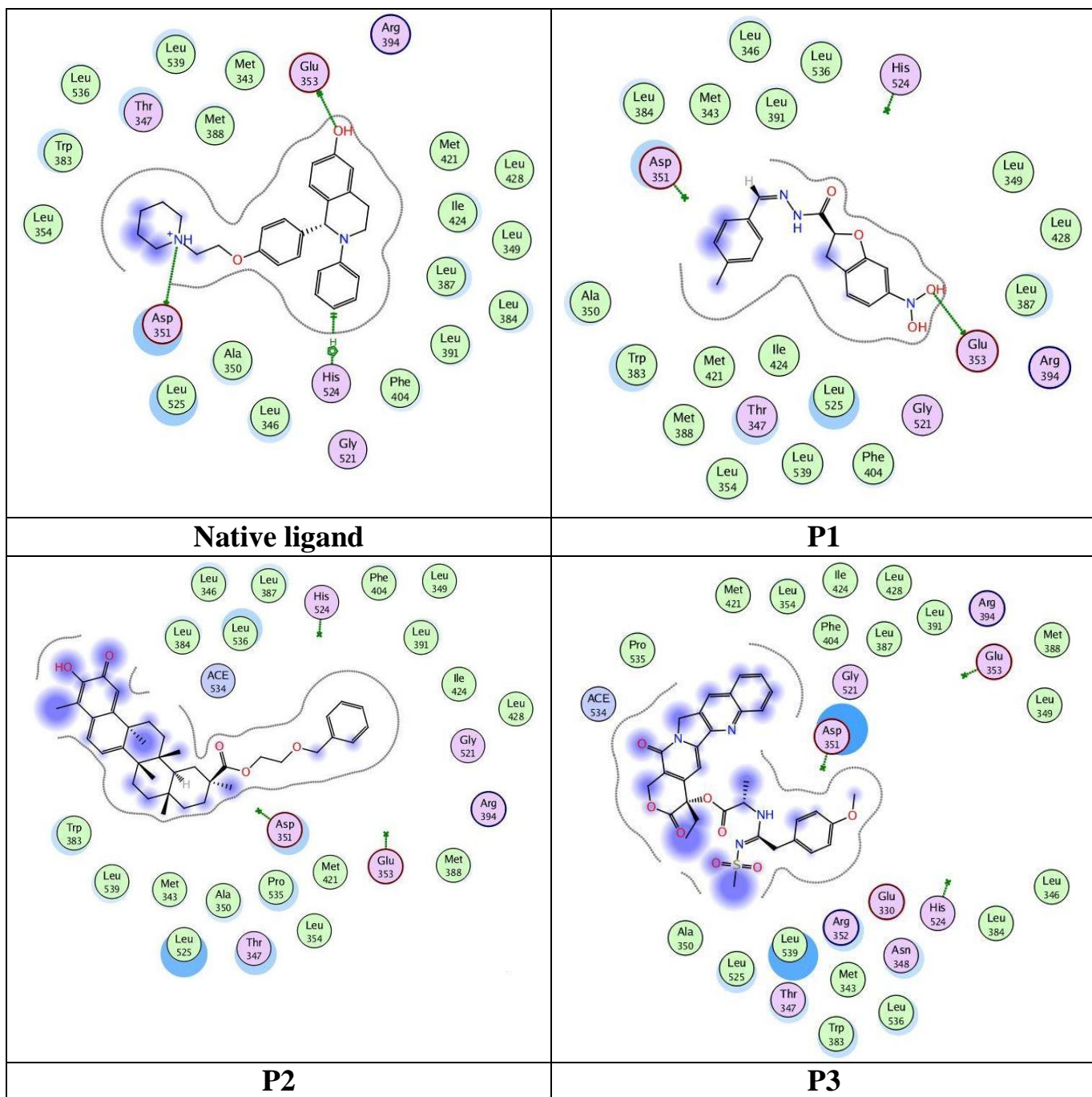
Figure A2 The binding modes of native ligand, P1, P2 and P3 into 1UOM
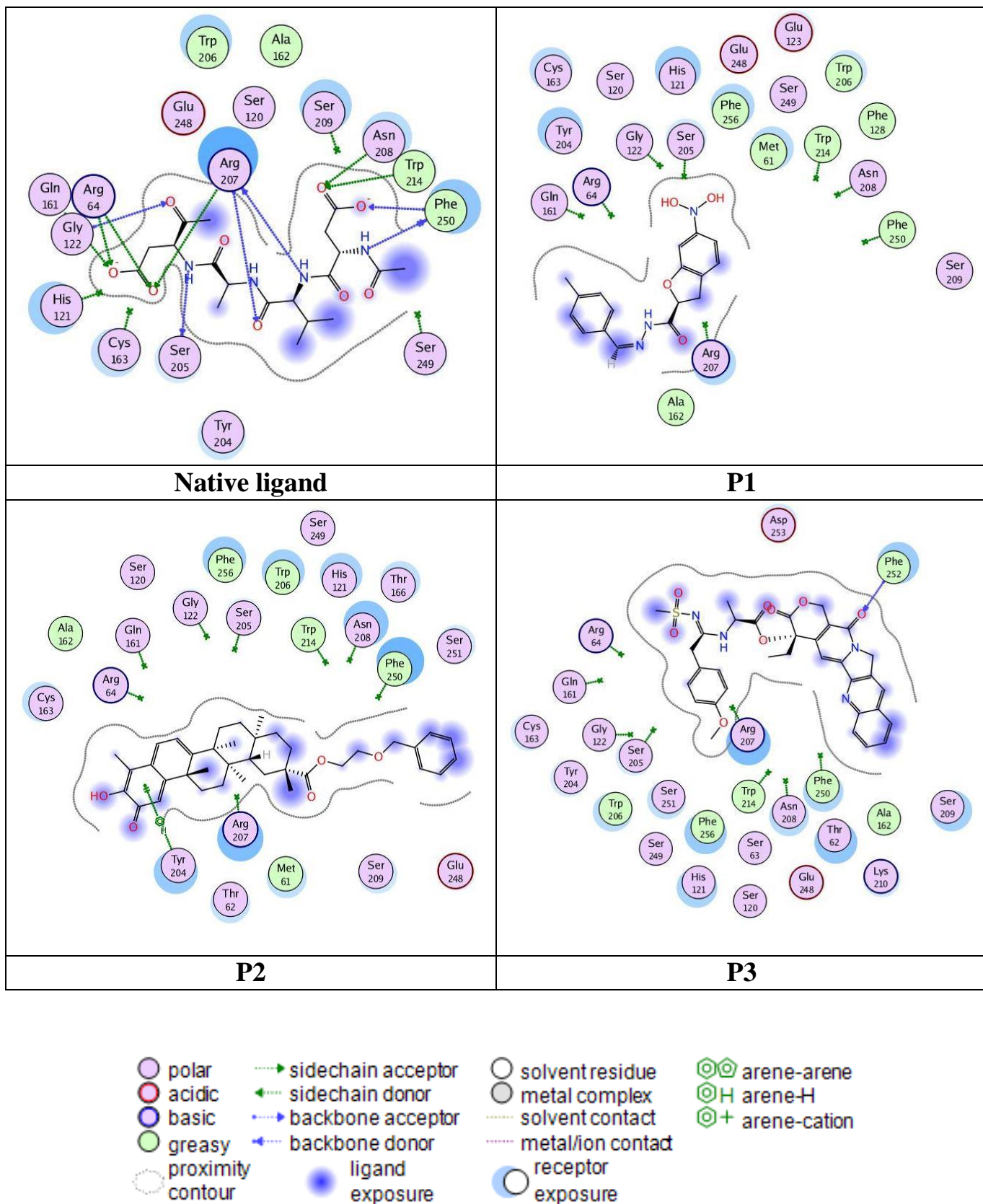
Figure A3 The binding modes of native ligand, P1, P2 and P3 into 1CP3
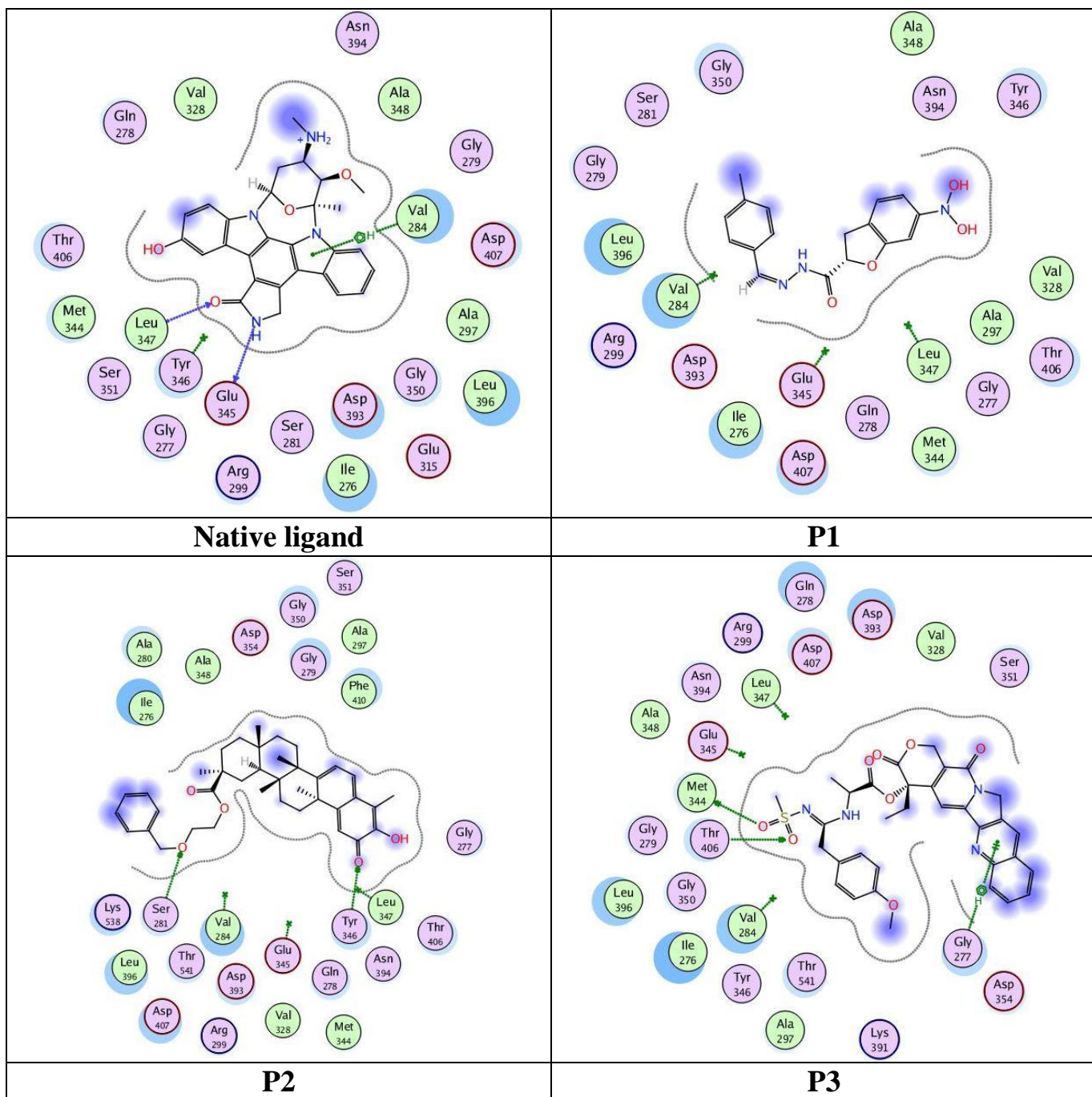
Figure A4 The binding modes of native ligand, P1, P2 and P3 into 2HY8
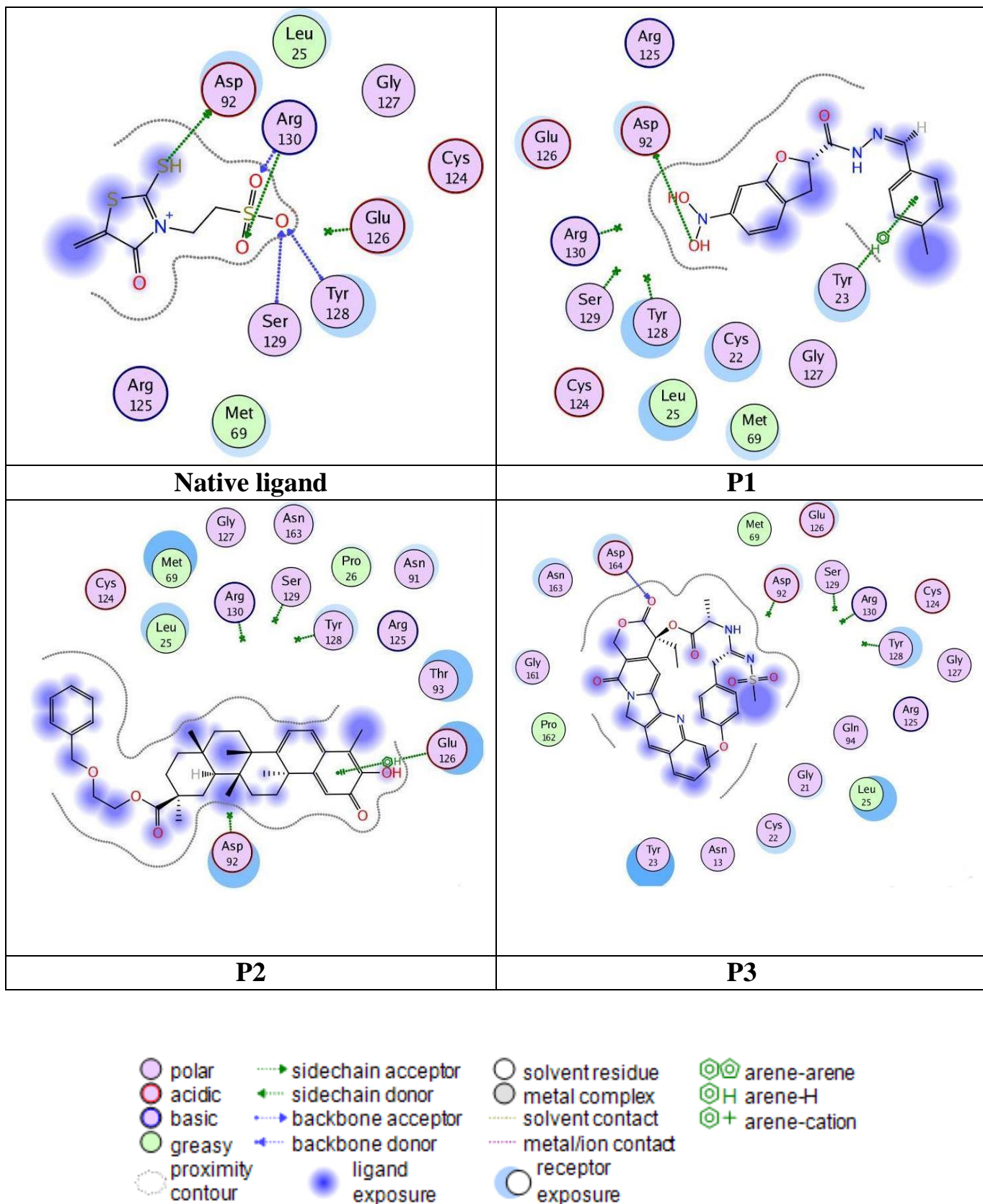
Figure A5 The binding modes of native ligand, P1, P2 and P3 into 3F81
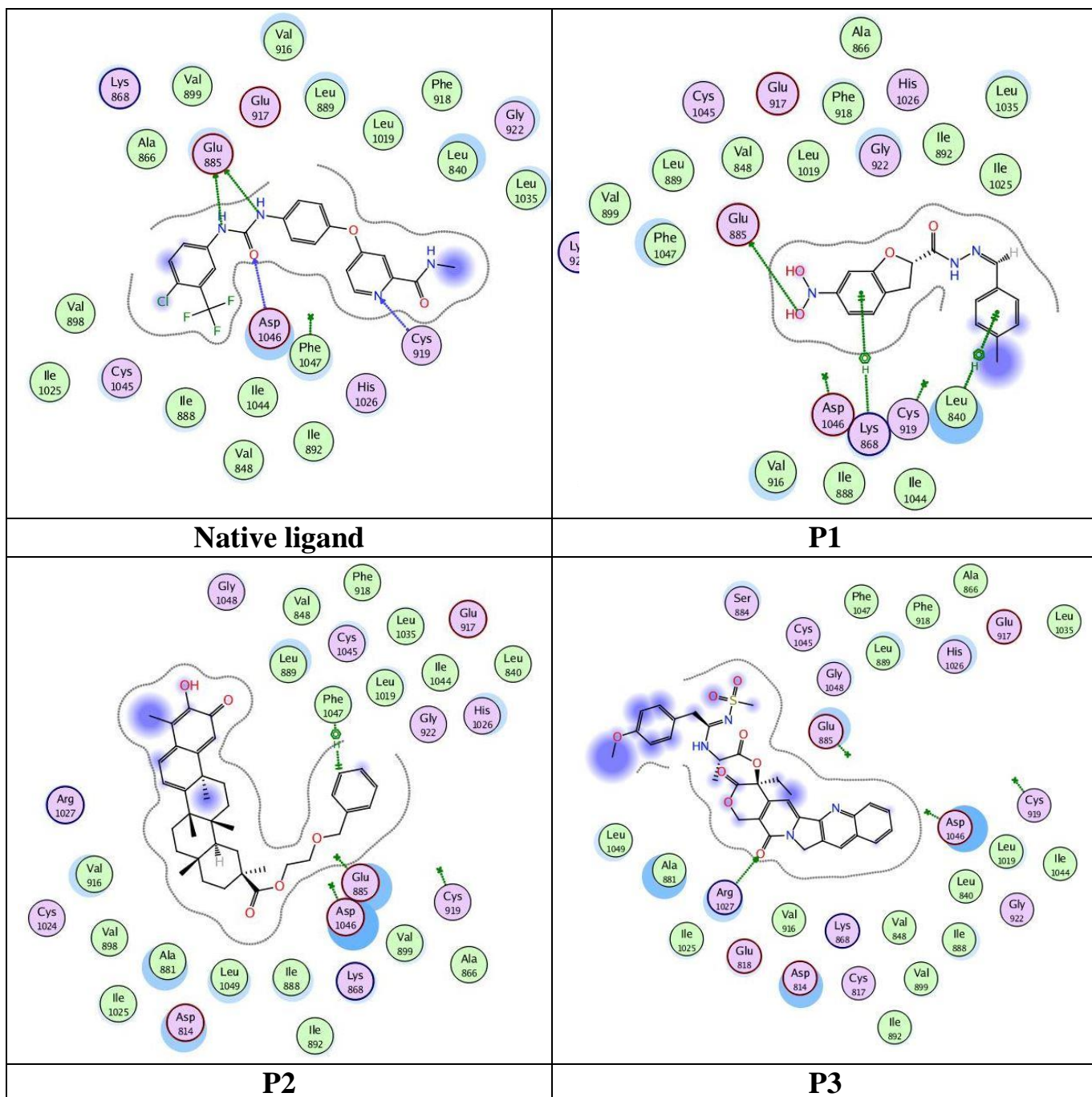
Figure A6 The binding modes of native ligand, P1, P2 and P3 into 4ASD

Appendix B

Table B1 Docking results of xanthone compounds and the native ligand with7AKM

Ligand	Compound atoms	Receptor atoms	Receptor residues	Type of bond interaction	Distance (Å)	Energies (Kcal/mol)
AGS	O2' 21	OE2	GLU 91	H-donor	2.82	-3.8
	N6 29	O	GLU 85	H-donor	2.82	-2.9
	S1G 2	NZ	LYS 38	H-acceptor	4.08	-4.3
	O2B 7	CA	GLY 18	H-acceptor	3.42	-0.7
	N1 30	N	CYS 87	H-acceptor	3.19	-3.1
	O1A 10	NZ	LYS 38	Ionic	2.94	-4.9
	5-ring	CG1	VAL 23	pi-H	3.71	-1.6
L1	O 22	N	ASP 148	H-acceptor	2.92	-1
	6-ring	CD1	LEU 15	pi-H	4.19	-0.8
L2	O 22	N	ASP 148	H-acceptor	2.91	-1
	6-ring	CB	LEU 15	pi-H	3.95	-0.6
	6-ring	CD1	LEU 15	pi-H	4.19	-0.8
	6-ring	CG1	VAL 23	pi-H	3.83	-0.6
L3	O 22	N	PHE 149	H-acceptor	3.19	-0.7
	6-ring	CB	LEU 15	pi-H	3.89	-1
	6-ring	CG1	VAL 23	pi-H	3.95	-0.6
	6-ring	CG2	VAL 23	pi-H	4.06	-0.6
L4	O 22	N	ASP 148	H-acceptor	2.9	-1
	6-ring	CB	LEU 15	pi-H	3.95	-0.6
	6-ring	CD1	LEU 15	pi-H	4.21	-0.8
	6-ring	CG1	VAL 23	pi-H	3.82	-0.7
L5	6-ring	CB	LEU 15	pi-H	4.01	-0.6
	6-ring	CD2	LEU 15	pi-H	4.19	-0.6
L6	6-ring	CB	LEU 15	pi-H	4	-0.6
	6-ring	CD2	LEU 15	pi-H	4.18	-0.6
L7	6-ring	CB	LEU 15	pi-H	3.92	-0.6
	6-ring	CB	LEU 15	pi-H	4.02	-0.6
	6-ring	CD2	LEU 15	pi-H	4.2	-0.6
L8	6-ring	CD1	LEU 15	pi-H	4.15	-0.7
L9	6-ring6-ring	CB	LEU 15	pi-H	4	-0.6
	ring	CD1	LEU 15	pi-H	4.11	-0.7
	6-ring	CD2	LEU 15	pi-H	4.17	-0.6
L10	O 18	N	ASP 148	H-acceptor	2.9	-1.1
	6-ring	CB	LEU 15	pi-H	3.95	-0.7
	6-ring	CD1	LEU 15	pi-H	4.26	-0.8
	6-ring	CA	GLY 18	pi-H	4.02	-0.7
	6-ring	CG1	VAL 23	pi-H	3.83	-0.6
L11	O 18	N	ASP 148	H-acceptor	2.89	-1.2
	6-ring	CD1	LEU 15	pi-H	4.19	-0.8
L12	6-ring	CD1	LEU 15	pi-H	3.98	-0.7
	6-ring	CG1	VAL 23	pi-H	3.88	-0.7

L13	O 22	N	ASP 148	H-acceptor	2.86	-1.3
	6-ring	CB	LEU 15	pi-H	4.12	-0.7
	6-ring	CG1	VAL 23	pi-H	4.10	-0.7
	6-ring	CG2	VAL 23	pi-H	4.01	-0.6
L14	6-ring	CD1	LEU 15	pi-H	3.98	-0.7
	6-ring	CG1	VAL 23	pi-H	3.88	-0.7
	6-ring	CG2	VAL 23	pi-H	4.44	-0.6
	6-ring	CG2	VAL 23	pi-H	3.75	-1.2
L15	O 22	N	ASP 148	H-acceptor	2.88	-1.2
	6-ring	CB	LEU 15	pi-H	4.15	-0.9
L16	6-ring	CG1	VAL 23	pi-H	3.88	-0.7
	6-ring	NZ	LYS 38	pi-cation	4.57	-0.7
	6-ring	CD	LYS 38	pi-H	3.74	-0.7
L17	O 21	N	ASP 148	H-acceptor	2.88	-1.1
	6-ring	CG2	VAL 23	pi-H	4.43	-0.6
	6-ring	CG2	VAL 23	pi-H	3.9	-0.9
	6-ring	CD	LYS 38	pi-H	3.78	-0.8
L18	6-ring	CB	LEU 15	pi-H	3.99	-0.9
	6-ring	CG2	VAL 23	pi-H	4.23	-0.7
	6-ring	CG2	VAL 23	pi-H	4.02	-0.8
L19	O 21	CA	ASP 148	H-acceptor	3.09	-0.7
	6-ring	CG1	VAL 23	pi-H	4.15	-0.8
	6-ring	CG2	VAL 23	pi-H	3.6	-0.6
L20	6-ring	CB	LEU 15	pi-H	4	-0.9
	6-ring	CG2	VAL 23	pi-H	4.22	-0.6
	6-ring	CG2	VAL 23	pi-H	4.02	-0.8
L21	6-ring	CB	LEU 15	pi-H	4.13	-0.8
	6-ring	CG1	VAL 23	pi-H	4.01	-0.6
L22	O 18	N	ASP 148	H-acceptor	2.86	-1.3
	6-ring	CB	LEU 15	pi-H	4.01	-1
L23	O 23	NZ	LYS 38	H-acceptor	3.21	-0.8
	6-ring	CB	LEU 15	pi-H	4.02	-0.9
	6-ring	CA	GLY 90	pi-H	3.78	-0.6
L24	6-ring	CB	LEU 15	pi-H	3.89	-0.8
	6-ring	CD2	LEU 15	pi-H	4.18	-0.7
L25	6-ring	CB	LEU 15	pi-H	3.98	-0.7
	6-ring	CD1	LEU 15	pi-H	4.04	-0.6
	6-ring	CG1	VAL 23	pi-H	3.97	-0.6
	6-ring	CG2	VAL 23	pi-H	3.62	-1
L26	O 26	OE2	GLU 91	H-donor	2.99	-1.8
L27	6-ring	CG2	VAL 23	pi-H	4.1	-0.6
	6-ring	NZ	LYS 38	pi-cation	4.04	-4.2
L28	6-ring	CB	LEU 15	pi-H	3.75	-0.6
	6-ring	CA	GLY 90	pi-H	3.94	-1.6
L29	O 30	OE2	GLU 91	H-donor	2.93	-1.7
	6-ring	CD2	LEU 15	pi-H	4.31	-0.6
	6-ring	CA	GLY 90	pi-H	3.83	-0.6
L30	O 30	OG	SER 147	H-donor	2.91	-0.8
	6-ring	CG2	VAL 23	pi-H	4.06	-0.7
L31	6-ring	CD2	LEU 15	pi-H	4.43	-0.6

L32	6-ring	CB	LEU 15	pi-H	4.04	-0.7
	6-ring	CD1	LEU 15	pi-H	4.15	-0.6
	6-ring	CD2	LEU 15	pi-H	3.96	-0.9
L33	6-ring	CD2	LEU 15	pi-H	4.48	-0.6
	6-ring	CA	GLY 90	pi-H	3.94	-1.2
L34	O 23	OD2	ASP 148	H-donor	3.05	-1.1
	O 23	NZ	LYS 38	H-acceptor	3.09	-1.1
	6-ring	CB	LEU 15	pi-H	4.01	-0.8
L35	6-ring	CG1	VAL 23	pi-H	3.81	-0.7
	6-ring	CG2	VAL 23	pi-H	3.74	-0.8
L36	6-ring	CB	LEU 15	pi-H	4.08	-0.8
	6-ring	CD1	LEU 15	pi-H	4.12	-0.6
	6-ring	CG1	VAL 23	pi-H	3.91	-0.7
	6-ring	CG2	VAL 23	pi-H	3.68	-0.9
L37	6-ring	CD1	LEU 15	pi-H	3.98	-0.7
	6-ring	CG1	VAL 23	pi-H	3.86	-0.7
	6-ring	CG2	VAL 23	pi-H	4.36	-0.6
	6-ring	CG2	VAL 23	pi-H	3.72	-1.2
L38	6-ring	CB	LEU 15	pi-H	4.04	-0.8
	6-ring	CD2	LEU 15	pi-H	4.43	-0.6
	6-ring	CA	GLY 90	pi-H	3.84	-0.9
L39	6-ring	CG2	VAL 23	pi-H	3.79	-0.8
L40	O 26	NZ	LYS 38	H-acceptor	3.23	-0.6
	O 30	NZ	LYS 38	H-acceptor	2.95	-4.3
	6-ring	CG1	VAL 23	pi-H	3.98	-0.7
L41	O 30	NZ	LYS 38	H-acceptor	3.18	-1.4
	6-ring	N	ALA 19	pi-H	3.79	-2.1
L42	6-ring	CA	GLY 18	pi-H	3.65	-0.7
	6-ring	CG1	VAL 23	pi-H	4.32	-0.6
	6-ring	CG2	VAL 23	pi-H	3.87	-0.8
L43	Cl 25	O	CYS 87	H-donor	3.47	-0.4
	O 8	NZ	LYS 38	H-acceptor	2.98	-1.1

Table B2 Docking results of xanthone compounds and the native ligand with 7AKO

Ligand	Compound atoms	Receptor atoms	Receptor residues	Type of bond interaction	Distance (Å)	Energies (Kcal/mol)
STU	C24 4	OE2	GLU 91	H-donor	3.44	-0.8
	N1 36	O	GLU 85	H-donor	2.93	-3.9
	N4 56	OE2	GLU 91	H-donor	3.29	-4.2
	N4 56	O	GLU 134	H-donor	2.99	-3.2
	O5 39	N	CYS 87	H-acceptor	3	-4.3
	N4 56	OE2	GLU 91	Ionic	3.29	-2.8
	5-ring	CB	LEU 15	pi-H	3.9	-0.8
	6-ring	CD2	LEU 15	pi-H	4.2	-0.7
	6-ring	CG1	VAL 23	pi-H	4.09	-0.7
L1	6-ring	CG1	VAL 23	pi-H	4.23	-0.8
L2	6-ring	CG1	VAL 23 VAL	pi-H	4.27	-0.6
	6-ring	CG1	23	pi-H	4.09	-0.6
	6-ring	CG2	VAL 23	pi-H	4.27	-0.8
L3	6-ring	CB	LEU 15	pi-H	3.92	-1.1
	6-ring	CG2	VAL 23	pi-H	4.07	-0.8
L4	O22	N	PHE 149	H-acceptor	2.94	-2.7
	6-ring	CB	SER 147	pi-H	4.33	-0.8
L5	6-ring	CA	GLY 16	pi-H	4.76	-0.9
	6-ring	N	PHE 149	pi-H	4.26	-1.4
L6	6-ring	CG1	VAL 23 VAL	pi-H	4.29	-0.7
	6-ring	CG1	23	pi-H	4.15	-0.7
	6-ring	CG2	VAL 23	pi-H	4.33	-0.7
L7	C 3	OE1	GLU 55	H-donor	3.4	-0.8
	6-ring	N	PHE 149	pi-H	4.51	-1
L8	C 3	OE1	GLU 55	H-donor	3.39	-0.9
	6-ring	CB	LEU 15	pi-H	4.23	-0.7
	6-ring	N	PHE 149	pi-H	4.51	-1
L9	C 6	OE1	GLU 55	H-donor	3.38	-0.9
	6-ring	CG2	VAL 23 PHE	pi-H	3.65	-0.6
	6-ring	N	149	pi-H	4.49	-1
L10	6-ring	CG1	VAL 23	pi-H	4.34	-0.6
L11	6-ring	CG1	VAL 23	pi-H	3.93	-0.6
	6-ring	CG2	VAL 23	pi-H	3.9	-0.6
L12	6-ring	CG1	VAL 23	pi-H	4.23	-0.6
L13	6-ring	CG1	VAL 23	pi-H	4.45	-0.6
L14	O 22	N	PHE 149	H-acceptor	3.05	-2.2
L15	6-ring	CB	LEU 15	pi-H	4.3	-0.7
L16	6-ring	CB	LEU 15	pi-H	3.97	-0.9
	6-ring	CG1	VAL 23	pi-H	4.38	-0.6
	6-ring	CG2	VAL 23	pi-H	4.02	-1
L17	6-ring	CB	LEU 15	pi-H	4.27	-0.6

L18	6-ring	CB	LEU 15	pi-H	4.38	-0.8
L19	6-ring	CB	LEU 15	pi-H	4.04	-0.6
L20	6-ring	CB	LEU 15	pi-H	3.97	-0.8
	6-ring	CG1	VAL 23	pi-H	4.35	-0.6
	6-ring	CG2	VAL 23	pi-H	4.02	-0.9
L21	6-ring	CB	LEU 15	pi-H	3.97	-0.8
	6-ring	CG1	VAL 23	pi-H	4.46	-0.6
L22	Cl30	OE1	GLU 55	H-donor	3.21	-0.9
	6-ring	CB	LEU 15	pi-H	4.03	-0.7
L23	O 23	OD2	ASP 148	H-donor	2.93	-3.1
	6-ring	CG1	VAL 23	pi-H	4.26	-0.7
L24	6-ring	CB	LEU 15	pi-H	4.17	-0.7
	6-ring	CD2	LEU 15	pi-H	4.19	-0.6
	6-ring	CD2	LEU 15	pi-H	4.24	-0.6
L25	O 25	OE2	GLU 91	H-donor	3.04	-1.9
	6-ring	CB	LEU 15	pi-H	4.19	-0.7
	6-ring	CD2	LEU 15	pi-H	4.28	-0.6
L26	6-ring	CB	LEU 15	pi-H	4.18	-0.8
	6-ring	CD2	LEU 15	pi-H	4.1	-0.8
L27	6-ring	CB	LEU 15	pi-H	4.5	-0.6
	6-ring	CD2	LEU 15	pi-H	4.32	-0.7
	6-ring	CB	SER 147	pi-H	4.43	-0.6
L28	6-ring	CD2	LEU 15	pi-H	3.99	-0.8
	6-ring	CA	GLY 90	pi-H	4.17	-0.6
L29	6-ring	CB	LEU 15	pi-H	4.28	-0.7
L30	6-ring	CD2	LEU 15	pi-H	4.35	-0.6
	6-ring	CB	SER 147	pi-H	4.51	-0.7
L31	6-ring	CD2	LEU 15	pi-H	4.32	-0.7
	6-ring	CD2	LEU 15	pi-H	4.41	-0.6
	6-ring	CG1	VAL 23	pi-H	4.22	-0.6
L32	Cl 25	OD1	ASN 59	H-donor	3.9	-0.5
	6-ring	CB	LEU 15	pi-H	4.12	-0.8
	6-ring	CD2	LEU 15	pi-H	4.01	-1
	6-ring	CA	GLY 90	pi-H	3.85	-0.6
L33	6-ring	CB	LEU 15	pi-H	3.95	-0.9
	6-ring	CD2	LEU 15	pi-H	4.01	-0.9
	6-ring	CA	GLY 90	pi-H	3.97	-1.6
L34	6-ring	CG1	VAL 23	pi-H	4.19	-0.8
L35	6-ring	CB	LEU 15	pi-H	4.1	-0.7
L36	6-ring	CB	LEU 15	pi-H	4.55	-0.7
	6-ring	CD2	LEU 15	pi-H	4.34	-0.6
L37	6-ring	CB	LEU 15	pi-H	4.26	-0.6
L38	6-ring	CB	LEU 15	pi-H	3.89	-0.8
	6-ring	CD2	LEU 15	pi-H	4.28	-0.6
	6-ring	CA	GLY 16	pi-H	4.37	-0.8
L39	6-ring	CG1	VAL 23	pi-H	4.31	-0.6
	6-ring	CG2	VAL 23	pi-H	4.34	-0.7
	6-ring	CB	SER 147	pi-H	4.34	-0.8

L40	6-ring	CB	SER 147	pi-H	4.29	-1
L41	6-ring	CB	LEU 15	pi-H	3.91	-0.7
L42	6-ring	CG2	VAL 23	pi-H	4.45	-0.6
	6-ring	CB	SER 147	pi-H	4.36	-0.8
L43	O 30	OE2	GLU 91	H-donor	2.9	-2.9
	6-ring	CG2	VAL 23	pi-H	4.24	-0.6

Table B3 Docking energy scores of xanthone compounds toward 7AKM

Ligand	S Score (Kcal/mol)
Prexasertib	-7.2
L43	-8.22
L36	-8.16
L41	-8.13
L30	-8.11
L33	-8.08
L29	-8.02
L42	-7.96
L38	-7.95
L27	-7.91
L21	-7.82
L26	-7.77
L22	-7.69
L39	-7.63
L40	-7.63
L37	-7.62
L32	-7.59
L35	-7.57
L31	-7.56
L25	-7.53
L16	-7.52
L9	-7.51

Ligand	S Score (Kcal/mol)
L6	-7.47
L18	-7.42
L10	-7.42
L20	-7.41
L11	-7.39
L34	-7.35
L5	-7.34
L28	-7.32
L24	-7.27
L17	-7.25
L14	-7.25
L19	-7.23
L7	-7.23
L23	-7.18
L15	-7.09
L4	-7.02
L8	-6.96
L3	-6.96
L13	-6.80
L12	-6.54
L2	-6.49
L1	-6.33

Table B4 Docking energy scores of xanthone compounds toward 7AKO

Ligand	S Score (Kcal/mol)
Prexasertib	-7.41
L36	-8.14
L40	-8.08
L42	-7.98
L41	-7.97
L31	-7.90
L43	-7.89
L29	-7.88
L26	-7.86
L20	-7.75
L39	-7.74
L27	-7.71
L30	-7.71
L32	-7.67
L23	-7.67
L38	-7.66
L21	-7.62
L33	-7.58
L37	-7.49
L16	-7.45
L24	-7.43
L9	-7.39

Ligand	S Score
L19	-7.39
L34	-7.39
L18	-7.39
L22	-7.25
L25	-7.21
L14	-7.20
L17	-7.18
L28	-7.18
L10	-7.16
L11	-7.14
L12	-7.09
L15	-7.06
L8	-7.02
L7	-7.00
L35	-7.00
L6	-6.97
L1	-6.73
L13	-6.66
L5	-6.65
L3	-6.58
L4	-6.49
L2	-6.19

Table B5 Docking results of top five scoring ligands and prexasertib with the active site of 7AKM

Ligand	Compound atoms	Receptor atoms	Receptor residues	Type of bond interaction	Distance (Å)	Energies (Kcal/mol)
L43	Cl 25	O	CYS 87	H-donor	3.47	-0.4
	O 8	NZ	LYS 38	H-acceptor	2.98	-1.1
L36	6-ring	CB	LEU 15	pi-H	4.08	-0.8
	6-ring	CD1	LEU 15	pi-H	4.12	-0.6
	6-ring	CG1	VAL 23	pi-H	3.91	-0.7
	6-ring	CG2	VAL 23	pi-H	3.68	-0.9
L41	O 30	NZ	LYS 38	H-acceptor	3.18	-1.4
	6-ring	N	ALA 19	pi-H	3.79	-2.1
L30	O 30	OG	SER 147	H-donor	2.91	-0.8
	6-ring	CG2	VAL 23	pi-H	4.06	-0.7
L33	6-ring	CD2	LEU 15	pi-H	4.48	-0.6
	6-ring	CA	GLY 90	pi-H	3.94	-1.2
Prexasertib	N 5	OE2	GLU 91	H-donor	2.99	-3.6
	N 6	NZ	LYS 38	H-acceptor	3.64	-0.6
	N 9	NZ	LYS 132	H-acceptor	3.77	-0.8
	6-ring	CB	LEU 15	pi-H	4.05	-0.7

Table B6 Docking results of top five scoring ligands and prexasertib with the active site of 7AKO

Ligand	Compound atoms	Receptor atoms	Receptor residues	Type of bond interaction	Distance (Å)	Energies (Kcal/mol)
L36	6-ring	CB	LEU 15	pi-H	4.55	-0.7
	6-ring	CD2	LEU 15	pi-H	4.34	-0.6
L40	6-ring	CB	SER 147	pi-H	4.29	-1
L42	6-ring	CG2	VAL 23	pi-H	4.45	-0.6
	6-ring	CB	SER 147	pi-H	4.36	-0.8
L41	6-ring	CB	LEU 15	pi-H	3.91	-0.7
L31	6-ring	CD2	LEU 15	pi-H	4.32	-0.7
	6-ring	CD2	LEU 15	pi-H	4.41	-0.6
	6-ring	CG1	VAL 23	pi-H	4.22	-0.6
Prexasertib	N 9	OH	TYR 86	H-acceptor	3.43	-0.8

Table B7 Binding interactions of the selected five ligands and prexasertib toward 7AKM generated through molecular dynamics simulation

Ligand	Compound atoms	Receptor atoms	Receptor residues	Type of bond interaction	Distance (Å)	Energies (Kcal/mol)
L43	O 30	NZ	LYS 132	H-acceptor	2.44	-0.4
	6-ring	CA	GLU 134	pi-H	4.01	-0.6
	6-ring	CA	ASN 135	pi-H	4.16	-0.6
L36	O 15	SG	CYS 87	H-donor	3.81	-0.8
	O 25	O	GLY 18	H-donor	2.69	-1.9
L41	O 30	NZ	LYS 38	H-acceptor	3.23	-3.2
L30	O 30	OG	SER 147	H-acceptor	2.86	-1.6
L33	O 30	ND2	ASN 135	H-acceptor	3.19	-0.6
Prexasertib	N 5	OE2	GLU 91	H-donor	2.62	-6.4

Table B8 Binding interactions of the selected five ligands and prexasertib toward 7AKO
generated through molecular dynamics simulation

Ligand	Compound atoms	Receptor atoms	Receptor residues	Type of bond interaction	Distance (Å)	Energies (Kcal/mol)
L36	O 25	O	GLU 17	H-donor	2.91	-1.4
	6-ring	CB	GLU 134	pi-H	4.41	-0.6
L40	O 30	OE2	GLU 91	H-donor	2.5	-1.4
	N 31	OE2	GLU 91	H-donor	2.75	-1.3
	6-ring	CG2	VAL 23	pi-H	4.31	-0.7
	6-ring	CG2	VAL 23	pi-H	3.98	-0.6
L42	O 32	OE2	GLU 91	H-donor	2.61	-3.6
	N33	OE2	GLU 91	H-donor	2.75	-1.5
	F 1	N	TYR 20	H-acceptor	2.92	-0.9
	F 3	N	GLY 21	H-acceptor	2.67	-1
L41	6-ring	CD2	LEU 15	pi-H	4.08	-0.8
L31	O 32	NZ	LYS 132	H-acceptor	3.11	-4
Prexasertib	N 3	O	GLU 134	H-donor	2.74	-2.7
	N 5	OE2	GLU 91	H-donor	2.65	-5.7
	N 6	OH	TYR 86	H-acceptor	2.6	-1.6
	6-ring	CD1	LEU 137	pi-H	4.09	-0.8
	6-ring	CD2	LEU 137	pi-H	4.02	-0.6

ORIGINAL RESEARCH

K-means clustering analysis, ADME/pharmacokinetic prediction, MEP, and molecular docking studies of potential cytotoxic agents

Ahlem Belkadi [1,2] · Samir Kenouche [3] · Nadjib Melkemi [1] · Ismail Daoud [4,5] · Rachida Djebaili [1]

Abstract

In this paper, a combined approach based on a k-means algorithm and statistical analysis has been applied successfully to classify 500 cytotoxic agents using 21 molecular descriptors. The k-means algorithm applied on the first two principal components split the observations into homogeneous clusters, in order to examine similarities and dissimilarities between molecules within and between clusters. Three clusters are clearly distinguished and the percentage of molecules in each cluster is 50%, 24.88%, and 25.12% for cluster 1, cluster 2, and cluster 3, respectively. Silhouette analysis is used as a cluster validation method, confirming that all the molecules are very well clustered. The silhouette indices obtained for each cluster are 0.53 for cluster 1, 0.60 for cluster 2, and 0.57 for cluster 3. Therefore, the average silhouette index obtained is 0.56. The value of the Hopkins statistic obtained is 0.922 confirming that the dataset is highly clusterable. In addition, the paragons of each cluster have been identified. The test statistic was performed to characterize each cluster by a subset of molecular descriptors. Moreover, in silico screening of pharmacological properties ADME and evaluation of drug-likeness of molecules showed that cluster 1 molecules have the best ADME profile and drug-likeness. The quantitative analysis of molecular electrostatic potential (MEP) on van der Waals surface was performed to identify the nucleophilic and electrophilic sites in the paragons molecules. Finally, molecular docking was performed for the paragon molecules on six different targets. Docking studies support the results observed in the MEP analysis showing that the favorable reactive sites of molecules are involved in strong hydrogen interactions with the functional residues of receptors.

Keywords k-means clustering · Cytotoxic activity · Statistical analysis · ADME · Drug-likeness · Molecular docking

Introduction

Cancer has been a great threat to humankind for thousands of years, and despite the significant efforts of chemists in developing cancer treatments, unfortunately, it continues to be a scientific challenge to researchers worldwide. Until now, there is no effective agent that can strongly control the unlimited cell growth and selectively target cancer without damaging the healthy cells, causing severe side effects such as bone marrow suppression, GI tract lesions, hair loss, nausea, and the rapid development of clinical resistance [1]. The side effects vary from one person to another. Some of them disappear after a few days, but it usually takes a few months to recover from chemotherapy. The first practical chemotherapeutic agent was discovered accidentally during World War II, nitrogen mustard which had antileukemic properties, a powerful alkylating agent that damage DNA, leading to cell death [2, 3].

Another important development in DNA-reactive drugs came with the discovery of cisplatin, which was also discovered by chance, provided a major advance in the treatment of testicular and ovarian carcinomas [4, 5]. The success of these initial agents led to the development and synthesis of a large

✉ Ahlem Belkadi
  ahlem.belkadi@univ-biskra.dz

1 Computational and Pharmaceutical Chemistry, LCME Laboratory, University of Mohamed Khider Biskra, 07000 Biskra, Algeria

2 Present address: University of Mohamed Khider Biskra, BP 145, 07000 Biskra, RP, Algeria

3 Modeling of Chemical Systems using Quantum Calculations, Applied Chemistry Laboratory, University of Mohamed Khider Biskra, 07000 Biskra, Algeria

4 Laboratory of Natural and Bioactive Substances, LASNABIO, University of Abou-Bakr Belkaid, Tlemcen, Algeria

5 Department of Matter Sciences, University of Mohamed Khider Biskra, BP 145 RP, 07000 Biskra, Algeria

ORIGINAL RESEARCH

Molecular docking/dynamic simulations, MEP, ADME-TOX-based analysis of xanthone derivatives as CHK1 inhibitors

Ahlem Belkadi[1] · Samir Kenouche[2] · Nadjib Melkemi[1] · Ismail Daoud[3,4] · Rachida Djebaili[1]

Abstract

CHK1 is a promising molecular target that gained immense attention recently for the development of cancer therapeutics. In this study, a simulation-based investigation was conducted to examine the CHK1 inhibitory activity of cytotoxic xanthone derivatives using a hierarchical workflow for molecular docking, MD simulation, ADME-TOX prediction, and MEP analysis. A molecular docking study was conducted for the forty-three xanthone derivatives along with standard prexasertib into the selected CHK1 protein structures 7AKM and 7AKO. Eight top hits were identified based on their free energy scores, namely L43, L42, L41, L40, L36, L33, L31, and L30, which showed better binding affinity (from − 8.22 to − 8.08 kcal/mol) (from − 8.14 to − 7.9 kcal/mol) toward 7AKM and 7AKO, respectively than the reference prexasertib which emphasizes the validity of our strategy. Furthermore, MD studies support molecular docking results and validate the stability of studied complexes in physiological conditions. These findings confirm that the selected eight xanthones are verifiable CHK1 inhibitors implying a good correlation between in silico and in vitro studies. Moreover, in silico ADME-TOX studies are used to predict the pharmacokinetic, pharmacodynamic, and toxicological properties of the studied eight hits and the standard prexasertib. Indeed, L36 showed the best ADME-TOX profile as it was the only hit without hepatotoxicity among the studied compounds. Besides, it displayed superior binding affinity and satisfied Lipinski, Pfizer, and golden triangle rules indicating a potent drug candidate. The quantitative analysis of electrostatic potential was performed for L36 to identify the reactive sites and possible non-covalent interactions.

Keywords CHK1 · Xanthones · Molecular docking · MD simulation · ADME-TOX · MEP analysis · Prexasertib

Introduction

The checkpoint kinase 1, commonly known as CHK1 or CHEK1, is a member of the serine/threonine kinase family of enzymes that mediate DNA damage response (DDR) and cell cycle progression [1]. It is well established that CHK1 overexpressed in numerous tumors, including breast, colon, liver, gastric, and nasopharyngeal carcinoma [2]. Moreover, many cancers rely on CHK1-mediated cell cycle arrest, especially if they are deficient in p53 [3]. Therefore, about 50% of all human cancers contain p53 mutations indicating that many tumors are dependent on CHK1 signaling pathway [4]. Numerous studies, both in vivo and in vitro, have shown that CHK1 inhibition is an effective targeting strategy that inhibits cancer cells proliferation, stops cell cycle arrest, and may also reverse tumor cell resistance to chemotherapeutics [5]. One of the most studied agents in this class is prexasertib, a clinical candidate inhibitor that selectively binds to CHK1, blocking DNA repair leading to the accumulation of damaged DNA that eventually triggers apoptosis in cancer cells. Indeed, prexasertib showed promising results for phase I and II trials as a monotherapy in patients with advanced solid tumors (squamous cell carcinomas) and in combination with cytotoxic DNA damaging agents, including Cisplatin

✉ Ahlem Belkadi
ahlem1.belkadi@gmail.com

1 Group of Computational and Pharmaceutical Chemistry, LCME Laboratory, University of Mohamed Khider Biskra, 07000 Biskra, Algeria

2 Group of Modeling of Chemical Systems Using Quantum Calculations, Applied Chemistry Laboratory, University of Mohamed Khider Biskra, 07000 Biskra, Algeria

3 Laboratory of Natural and Bioactive Substances, LASNABIO, University of Abou-Bakr Belkaid Tlemcen, Tlemcen, Algeria

4 Department of Matter Sciences, University of Mohamed Khider Biskra, BP 145 RP, 07000, Biskra, Algeria