Université Mohamed Khider de Biskra
Faculté des Sciences Exactes et des Sciences de la Nature et de la vie
Département des Sciences de la matière

MÉMOIRE DE MASTER

Sciences de la Matière
Filière : Chimie
Spécialité : Chimie Pharmaceutique

Réf. : Entrez la référence du document

Présenté et soutenu par :

KHALDI Menatallah Khadidja

Le :

Structure-based discovery of novel Cyclin Dependent Kinase 2 inhibitors for the treatment of cancer

---

Jury :

Mr	SELLAMI Mohamed	MCB	Université de Biskra	Président
Mme	KHAMOULI Saida	MCA	Université de Biskra	Rapporteur
Mme	AGGOUN Sihem	MAA	Université de Biskra	Examinateur

Année universitaire : 2023/2024

DEDICATION

This memoir is profoundly dedicated to all.

my beloved family (my mother, my father, and my all friends).

my thankful your endless love, Supports and Advice

Acknowledge

Praise to God, who has given me the strength, the courage and the will and the patience to complete my training and be able to accomplish thismodest research work.

I would like to express my full gratitude to my supervisor, Madame KHAMOULI Saida, Associate professor A at the University of Biskra

for having supervised me in the best possible way on my Master's memory

I thank her very much for the time she has devoted to this memory.

Thank you for your explanations, advice and patience.

My sincere thanks go deep gratitude to Mr. SELLAMi Mohamed Associate professor B at the University of Biskra for giving me the honor of beingthe president of the members of my jury,

I also extend my warm thanks to Mme. AGGOUN Sihem Maiter Assistant A at the University of Biskra for agreeing to read my manuscript and evaluate my work.

.

Table of Contents

List of Tables I
List of figures II
Liste of abbreviation III
Introduction 2
References 4

CHAPTER I: Overview of cancer disease and cyclin-dependent kinase 2

I.1 Introduction 6
I.2 Definition 6
I.3 Differences between Cancer Cells and Normal Cells 8
I.4 types of cancer 9
I.4.1 Epithelial and Blood Cell Cancers 9
I.4.2 Connective Tissue and Nervous System Cancers 9
I.5 Epidemiology 10
I.6 Cancer Causes 11
I.7 The stages of cancer 11
I.8 Symptoms 13
I.9 Early detection 13
I.10 Treatment 14
I.11 Cyclin dependent kinase 2 inhibitors (CDK2) 20
I.11.1 Definition of Cyclin-Dependent Kinases (CDK) 20
I.11.2 Cyclin-dependent kinases regulation Cell Cycle 21
I.11.3 Cyclin-dependent Kinases' Target Proteins 22
I.12.The Purines 23
I.12.1 Definition 23
I.12.2 Biological activities of purine analogue 23
References 25

Chapiter II: Computational Methods in Drug Discovery

II.1. Drug discovery 28
II.2. Steps in Modern Drug Discovery 29
II.3. Virtual Screening 34
II.3.1. Ligand-based Virtual Screening 34
II.3.1.1 Similarity Search 35
II.3.1.2 Ligand-based Pharmacophore Models 35
II.3.1.3 Quantitative Structure-Activity Relationship (QSAR) 36
II.3.2 Structure-Based Virtual Screening 36
II.3.2.1 Structure-Based Pharmacophore 37
II.3.2.2 .1 General Principle of Molecular Docking 37
II.3.2.2 .2 Types of Molecular Docking 38
II.3.2.2 .3 Search Algorithms 39
II.3.2.2 .4 Scoring Process 40
II.4. ADME-TOX 42
II.4. 1 Drug likeness 42
II.4.2 Pharmacokinetic and Toxicity Properties 42
II.4.2.1 Absorption 42
II.4.2.2 Distribution 42

II.4.2.3 Metabolism 43
II.4.2.4 Excretion 43
II.4.2.5 Toxicity 43
References 44

Chapiter III : Molecular docking and ADME study of 6-substituted 2- arylaminopurine derivative as anticancer agents

III.2. Materials and Methods 48
III.2.1. Computer System and Software 48
III.2.2 Methodology of calculations 50
III.3.2.1. Dataset 50
III.3.2.2. Ligand preparation 50
III.3.2.3. Protein preparation 56
III.3.2.4. Generation of Receptor Grid 57
III.3.2.5. Molecular docking 57
III.3.2.6. In silico ADME prediction 58
III.4. Results and discussions 60
III.4.1. Molecular Docking 60
III.4.2. Drug likeness and ADME prediction 67
III.4.2.1 Drug likeness 67
III.4.2.2 ADME prediction 69
References 72
Conclusion 75
Abstract

List of Tables

Table III.1: Structures of 6- substituted 2-arylaminopurine derivatives with their PIC50 **51**

Table III.2: Docking results of ligands and reference ligand with CDK2 proteins **60**

Table III.3: Docking results of ligands L_{32}, L_5,L_{24},L_{23},L_{29},L_{11} and L_{ref} **61**

TableIII.4: Drug-likeness parameters of top ligands and reference ligand **68**

TableIII.05: Absorption and distribution of top ligands and reference ligand **69**

TableIII.06: Metabolism and excretion of top ligands and reference ligand **71**

List of Figures

Figure I.1: Schema describing the formation of metastases **7**

Figure I.2: Schema represents the difference between normal cell and cancer cell **8**

Figure I.3: Distribution of cases and deaths for the top10 most common cancers in 2020 among females for Incidence and Mortality **10**

Figure I.4: Overview of the Photodynamic Therapy Function **19**

Figure I.5: Biological assembly of CDK2 **21**

Figure I.6: The mammalian CDKs involved in progression throughout the different phases of the cell cycle. **22**

Figure I.7: Structure of purine **23**

Figure 1 : Several major milestones in drug discovery during the 19th and 20th centuries **28**

Figure 2 : Drug Discovery Pipeline **34**

Figure 3: Classification of virtual screening methods **35**

Figure 4: Mechanisms of Docking **38**

Figure III.1: The graphical interface of the ChemDraw Pro12.0 software. **48**

Figure III.2: The graphical interface of the Meastro 13.9 software. **49**

Figure III.3: The graphical interface of the PDB database website **49**

Figure III.4: The graphical interface of the pkCSM pharmacokinetic properties. **50**

Figure III.5: Chemical structure of reference ligand (Lref) 1RO. **56**

Figure III.6: Structure of the Cyclin Dependent Kinases 2 (PDB ID 4EOS) **57**

FigureIII.7: 2D, 3D interactions of reference ligand with active site of CDK2 **62**

FigureIII.8: 2D, 3D interactions of ligand 32 with active site of CDK2 (ID:4EOS) **63**

FigureIII.9: 2D, 3D interactions of ligand 5 with active site of CDK2 (ID:4EOS) **64**

FigureIII.10: 2D, 3D interactions of ligand 24 with active site of CDK2 (ID:4EOS) **64**

FigureIII.11: 2D, 3D interactions of ligand 24 with active site of CDK2 (ID:4EOS) **65**

FigureIII.12: 2D, 3D interactions of ligand 11 with active site of CDK2 (ID:4EOS) **66**

FigureIII.13: 2D, 3D interactions of ligand 11 with active site of CDK2 (ID:4EOS) **66**

Abbreviation list

3D	Three dimensional
2D	Tow dimensional
ABCB	ATP-binding cassette sub-family B member 1
ABC	ATP-binding cassette transporters
ADMET	Absorption,Distribution,Metabolism,Excretion,Toxicity.
ADME	Adsorption, Distribution, Metabolism, and Excretion
Aflatoxin	Aflatoxins are a family of toxins produced by certain fungi that are found on agricultural crops
ALA	Alanine amino-acid
Alpha(2A)	Receptors decrease sympathetic outflow and blood pressure
Alpha(2B)	Subtype increases blood pressure
AMBER	Force field for molecular dynamics
Arsenic	Semimetallic element
ASP	Aspartic amino-acid
CADD	Computer -aided drug design
CDKs	Cyclin-dependent kinases
ChemScore	Empirical scoring function
CNS	Central nervous system
COX-2	Cyclooxygenase-2
DA	Daltons
DNA	Deoxyribonucleic acid ARN: Ribonucleic acid
EBV	Epstein-barr virus
4EOS	Cyclin dependent kinase 2 ID on protein data bank
FDA	The The Food and Drug Administration
FRB	Including the number of rotatable bonds
FU	One jump fraction or unbound fraction
GLU	Glutamic amino-acid
GI	Gastrointestinal absorption
HBA	Hydrogen bond acceptors
HBD	Number of hydrogen bond donors
HEB	Hemato-encephalic barrier
HIA	Humain intestinal absorption
HIS	Histidine amino-acid
HIV	Human immunodeficiency virus
HPV	Human papilloma virus
IC50	Half maximal inhibitory concentration
ILE	Isoleucine amino-acid
IV	Intravenous

Ki	The inhibitory constant
LEU	Leucine amino-acid
LogP	The partition coefficient
LYS	Lysine amino-acid
MC	Monte Carlo simulation
MDR1	Multi drug Resistance protein 1
MRI	Magnetic resonance imaging
MW	Molecular weight
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NBR	Number of rotatable bonds
NIH	The National Institutes of Health
NMR	Nuclear magnetic resonance
PDB	Protein Data Bank
PFM	The Pelvic Floor Muscles
PHE	Phenylalanine amino-acid
pkCSM	Predicting small-molecule pharmacokinetic properties using graph-based signatures
PLP	Piecewise Linear Potential scoring function
PSA	Polar surface area
QikProp	Tool of Schrodinger drug discovery software
QSAR	Quantitative Structure Activity Relationships.
RCSB	Research Collaboratory for Structural Bioinformatics Protein Data Bank
RD-QSAR	Receptor-Dependent Quantitative Structure Activity.
RMSD	Root Mean Square Deviation
1RO	Reference ligand ID on protein data bank
SP	Standard-Precision
TPSA	Topological polar surface area
VAL	Valine amino-acid
VDW	Van der waals
WHO	World Health Organization
XP	Extra-Precision

Introduction

Introduction

Cancer is one of the most serious diseases, and the second leading cause of death. It is described by overexcited and uncontrolled cell proliferation [1]. Around one third of the population worldwide is expected to be diagnosed with cancer during their lifetimes [2]. According to the International Agency for Research on Cancer, in 2020 about 19.3 million people were diagnosed with cancer and around 10 million people died of the disease. It is estimated that the morbidity and mortality of cancer will continue to rise, with cancer cases growing to 24.1 million new cases and 13.0 million deaths by 2030 [3]. When an appropriate treatment strategy is provided, around two thirds of human cancers can be treated effectively [4]. The development of treatment resistance is the major obstacle to successful cancer therapy. The most effective therapies frequently fail to result in a thorough tumor response, and eventually end up with treatment resistance and tumor deterioration [5].

Cyclin-dependent kinase (CDK) is a serine/threonine protein kinase family with a total of 20 members, including CDK1-CDK20[6]. The CDK family associates with cyclin and plays a vital role in controlling the cell cycle [7]. CDKs and cyclins are frequently observed to be upregulated in neoplastic cells. Hence, inhibitors of CDKs have been identified as potential therapeutic agents for the treatment of cancer [8]. CDK2 holds significant importance as a member of the CDK family, as evidenced by studies conducted by Whittaker et al [9].CDK2 plays a pivotal role in the regulation of the cell cycle within actively dividing cells, exhibiting significant functionality during the latter part of the G1 phase and throughout the entire S phase[10]. According to clinical studies, the elevation in CDK2 activity has been identified as a potential factor contributing to the onset of malignancies.

Computer aided drug design is a powerful tool in the search of promising drug candidates, particularly when used in tandem with current chemical biology screening techniques. Despite the fact that CADD makes use of several restrictions and approximations, this knowledge driven approach has become an essential part in the drug design process due to its ability to fast-track drug discovery by utilizing existing knowledge and theories on receptor-ligand.

Molecular docking is a key tool in structural molecular biology and computer-assisted drug design. The goal of ligand-protein docking is to predict the predominant binding mode(s) of a ligand with a protein of known three-dimensional structure.

Introduction

Successful docking methods search high-dimensional spaces effectively and use a scoring function that correctly ranks candidate dockings. Docking can be used to perform virtual screening on large libraries of compounds, rank the results, and propose structural hypotheses of how the ligands inhibit the target. [11].

The aim of this study is to utilize virtual screening methods such us molecular docking and ADMET to investigate 6-substituted 2-Arylaminopurines derivatives in order to select the promising compounds as anti-cancer activity.

Our work is divided into three chapters:

- The first chapter provides an overview of cancer, and the protein Cyclin-depandent kinase 2.
- The second chapter gave us an overview of the main virtual screening strategies, which are the predictive methods most commonly used to select new molecules of therapeutic interest.
- In the third chapter, we presented the different materials and methods used in this study. we present most of our results and a discussion.

In the end, we will provide a general conclusion that summarizes all the work.

Introduction

References

1. Alanazi, M. M., Mahdy, H. A., Alsaif, N. A., Obaidullah, A. J., Alkahtani, H. M., Al-Mehizia, A. A., ... & Eissa, I. H. (2021). New bis ([1, 2, 4] triazolo)[4, 3-a: 3′, 4′-c] quinoxaline derivatives as VEGFR-2 inhibitors and apoptosis inducers: Design, synthesis, in silico studies, and anticancer evaluation. Bioorganic Chemistry, 112, 104949.
2. Alanazi, M. M., Aldawas, S., & Alsaif, N. A. (2023). Design, synthesis, and biological evaluation of 2-mercaptobenzoxazole derivatives as potential multi-kinase inhibitors. Pharmaceuticals, 16(1), 97
3. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 71(3), 209-249.. https://doi.org/10.3322/caac.21660
4. Committee, C. P. R. S., Arteaga, C. L., Adamson, P. C., Engelman, J. A., Foti, M., Gaynor, R. B., ... & Fuller, N. R. (2014). AACR Cancer Progress Report 2014. Clinical cancer research: an official journal of the American Association for Cancer Research, 20(19 0), S1
5. El-Azab, A. S., Alaa, A. M., AlSaif, N. A., Alkahtani, H. M., Alanazi, M. M., Obaidullah, A. J., ... & Alharbi, A. (2020). Antitumor activity, multitarget mechanisms, and molecular docking studies of quinazoline derivatives based on a benzenesulfonamide scaffold: Cell cycle analysis. Bioorganic Chemistry, 104, 104345.
6. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The protein kinase complement of the human genome. Science, 298(5600), 1912-1934.
7. Roskoski Jr, R. (2016). Cyclin-dependent protein kinase inhibitors including palbociclib as anticancer drugs. Pharmacological research, 107, 249-275
8. Malumbres, M., Pevarello, P., Barbacid, M., & Bischoff, J. R. (2008). CDK inhibitors in cancer therapy: what is next?. Trends in pharmacological sciences, 29(1), 16-21
9. Whittaker, S. R., Mallinger, A., Workman, P., & Clarke, P. A. (2017). Inhibitors of cyclin-dependent kinases as cancer therapeutics. Pharmacology & therapeutics, 173, 83-105.
10. Ying, M., Shao, X., Jing, H., Liu, Y., Qi, X., Cao, J., ... & He, Q. (2018). Ubiquitin-dependent degradation of CDK2 drives the therapeutic differentiation of AML by targeting PRDX2. Blood, The Journal of the American Society of Hematology, 131(24), 2698-2711.
11. Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. Molecular modeling of proteins, 365-382.

Chapiter I

Overview of cancer disease and cyclin-depandent kinase 2

CHAPTER I: Overview of cancer disease and cyclin-dependent kinase 2

I.1 Introduction

Cancer is a complex and multifaceted group of diseases characterized by the uncontrolled growth and spread of abnormal cells. It can affect any part of the body and is often referred to by terms like neoplasms or malignant tumors. The hallmark feature of cancer is the rapid proliferation of these aberrant cells, which can infiltrate nearby tissues and organs and, in advanced stages, spread to distant sites through a process known as metastasis.

The deregulation of normal cellular processes, particularly those involved in cell division, lies at the heart of cancer development. Normally, the body tightly regulates cell growth and division to maintain tissue integrity and function. However, in cancer, this regulation breaks down, leading to uncontrolled cell proliferation. These abnormal cells continue to divide and accumulate, forming tumors that can interfere with the function of vital organs such as the liver, kidney, and lung.

Metastasis is a critical aspect of cancer progression and is responsible for much of the morbidity and mortality associated with the disease. During metastasis, cancer cells from the primary tumor invade nearby blood vessels or lymphatic channels, allowing them to travel to distant sites in the body. Once lodged in a new location, these cells can proliferate and form secondary tumors, further compromising organ function and often making treatment more challenging.

The impact of cancer on overall health can be devastating. As cancer cells proliferate and tumors grow, they compete with healthy tissues for vital nutrients and oxygen. Additionally, the presence of tumors can disrupt normal physiological processes, leading to a range of symptoms and complications depending on the affected organ system. In advanced stages, cancer can cause significant morbidity and ultimately lead to death, particularly if widespread metastases occur.

I.2 Definition

Cancer is a complex group of diseases characterized by abnormal cell growth and the ability of these cells to spread into surrounding tissues. It can originate in almost any part of the body. Normally, our cells undergo a tightly regulated process of growth, division, and replacement.

However, in cancer, this process becomes disrupted, leading to uncontrolled cell division and the formation of tumors.

Malignant tumors are characterized by the ability to invade surrounding tissues, disrupting their normal functions. Additionally, as malignant tumors grow, some cancer cells can detach and spread to distant parts of the body through the bloodstream or lymphatic system, a process known as metastasis. These cells can then form new tumors in distant organs or tissues, contributing to cancer spread.

Solid tumors are masses of tissue named after the type of cells they originate from, such as lung cancer arising from abnormal lung cells or breast cancer originating in breast tissue. In contrast, liquid tumors like leukemia develop in the blood and bone marrow.

Although benign tumors are not cancerous, they can still cause health issues depending on their size and location. Some benign tumors can grow significantly, causing symptoms or complications, and may require medical intervention due to their surrounding location. [1]

- **According to WHO**

Cancer is a generic term for a large group of diseases that can affect any part of the body. Other terms used are malignant tumours and neoplasms. One defining feature of cancer is the rapid creation of abnormal cells that grow beyond their usual boundaries, and which can then invade adjoining parts of the body and spread to other organs; the latter process is referred to as metastasis. Widespread metastases are the primary cause of death from cancer. [2]

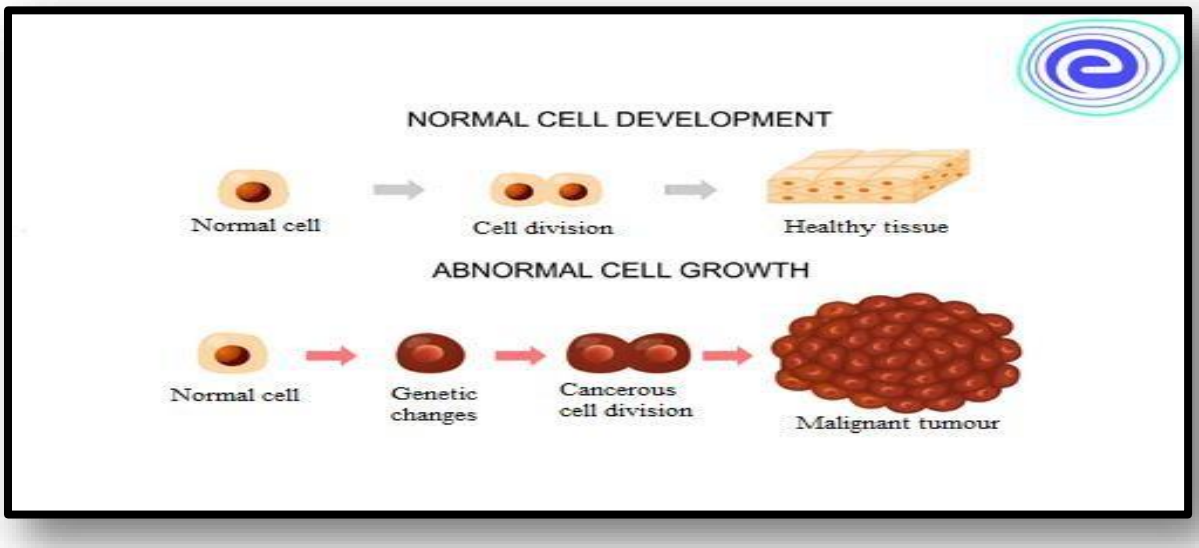


Figure I.1: Schema describing the formation of metastases

I.3 Differences between Cancer Cells and Normal Cells.

Cancer cells differ from normal cells in many ways that allow them to grow out of control and become invasive: [1]

• Unlike normal cells that mature into very distinct types with specific functions, cancer cells are less specialized, which helps them divide without stopping.

• Cancer cells ignore signals that would otherwise stop them dividing or that begin the process known as “programmed cell death,” or “apoptosis.”

• Cancer cells influence surrounding normal cells, molecules, and blood vessels — an area known as the microenvironment. For instance, cancer cells can induce nearby normal cells to form blood vessels that supply tumours with oxygen and nutrients and remove waste products.

• Cancer cells often evade or hide from the immune system that would normally remove them. They can even co-opt the immune system to help them grow and stay alive. [1]

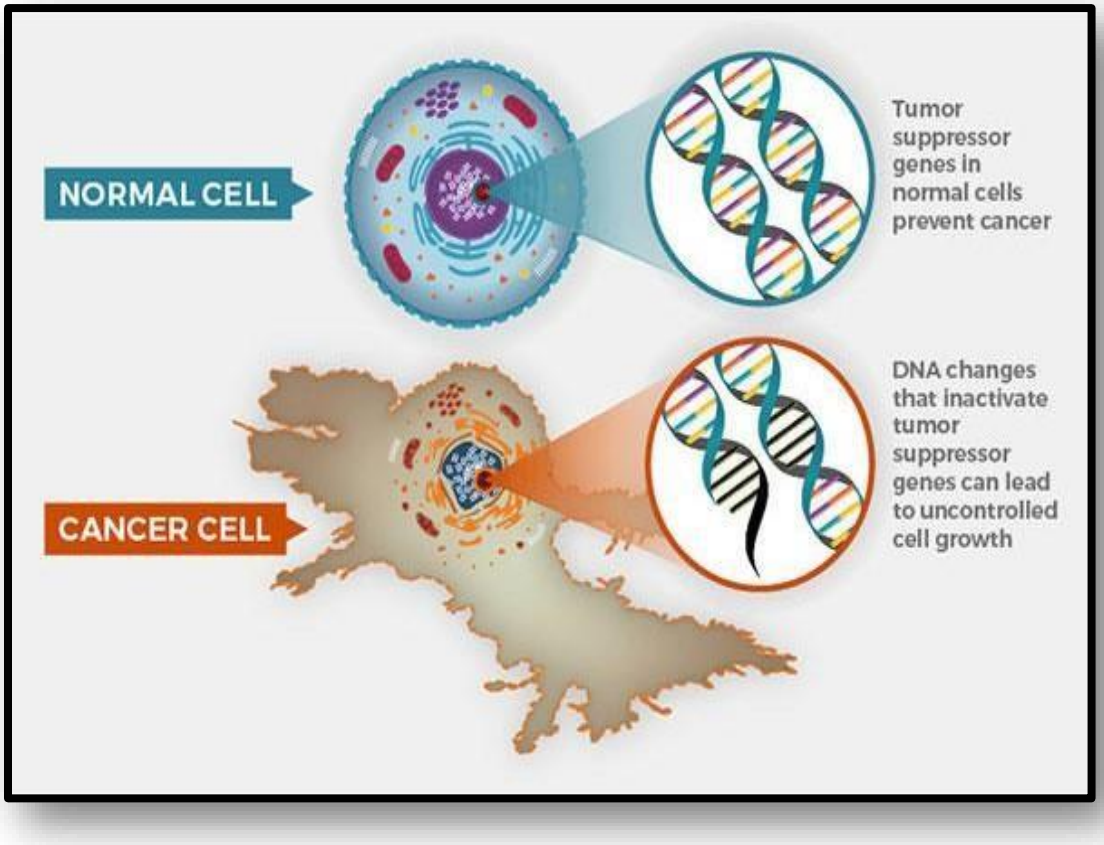


Figure I.2: Schema represents the difference between normal cell and cancer cell

I.4 types of cancer

Cancers are divided into various types that are:

I.4.1 Epithelial and Blood Cell Cancers:

a. Carcinomas: It starts in the tissue or the skin, which covers the glands and internal organ surface. It forms a solid tumor. Breast cancer, prostate cancer, colorectal cancer, lung cancer.

b. Leukemia's: Leukemia is a cancer of the blood. It begins when healthy blood cells grow uncontrollably and change. It is divided into 4 types, that are acute myeloid leukemia, acute lymphocytic leukemia, chronic myeloid leukemia, and chronic lymphocytic leukemia

c. Lymphomas: Lymphoma is cancer that begins in the lymphatic system and it is a network of glands and vessels that helps to fight with infection. Hodgkin lymphoma and non-Hodgkin lymphoma.

d. Multiple Myeloma: Multiple myelomas is cancer that begins in plasma cells, another type

e. Melanoma: It starts in cells that become melanocytes. These cells are specialized cells that make melanin, i.e., the pigment that gives the color to the skin. Mainly melanomas develop on the skin, but it can also develop in other pigmented tissue like an eye [3].

I.4.2 Connective Tissue and Nervous System Cancers:

a. Sarcomas: It starts in the tissues which connect and support the body. It can be formed in nerves, tendons, joints, fat, blood vessels, bone, lymph vessels, muscles, or cartilage.

b. Central Nervous System Cancers: Cancer that starts in brain tissues and spinal cord called "brain and spinal cord tumors", and others primary CNS lymphomas, vestibular schwannomas, gliomas, pituitary adenomas, primitive neuro-ectodermal tumors, meningiomas, and vestibular schwannomas.

c. Other Types of Tumors:

Germ Cell Tumors: It is the type of tumor that starts in the cells which give rise to eggs or sperms. This can be occurring anywhere in the body and either malignant or benign.

Neuroendocrine Tumors: Neuroendocrine tumors form from cells that release hormones into the blood in response to a signal from the nervous system. It forms from those cells which release hormones in blood in response to signal from the nervous system. These tumors, which can create higher-than-normal amounts of hormones, will cause many various symptoms. It may be either benign or malignant [3].

I.5 Epidemiology

According to the World Health Organization (WHO), one in six deaths worldwide is caused by cancer. The most common types of cancer identified by the WHO in 2020 were: [2]

- breast (2.26 million cases);
- Lung (2.21 million cases) ;
- Colon and rectum (1.93 million cases);
- Prostate (1.41 million cases) ;
- Skin (non-melanoma) (1.20 million cases)
- Stomach (1.09 million cases).

The Most Common causes of cancer Death in 2020 were: [2]

- lung (1.80 million deaths);
- colon and rectum (916 000 deaths);
- liver (830 000 deaths);
- stomach (769 000 deaths)
- breast (685 000 deaths). Each year, approximately 400 000 children develop cancer. The most common cancers vary between countries. Cervical cancer is the most common in 23 countries.

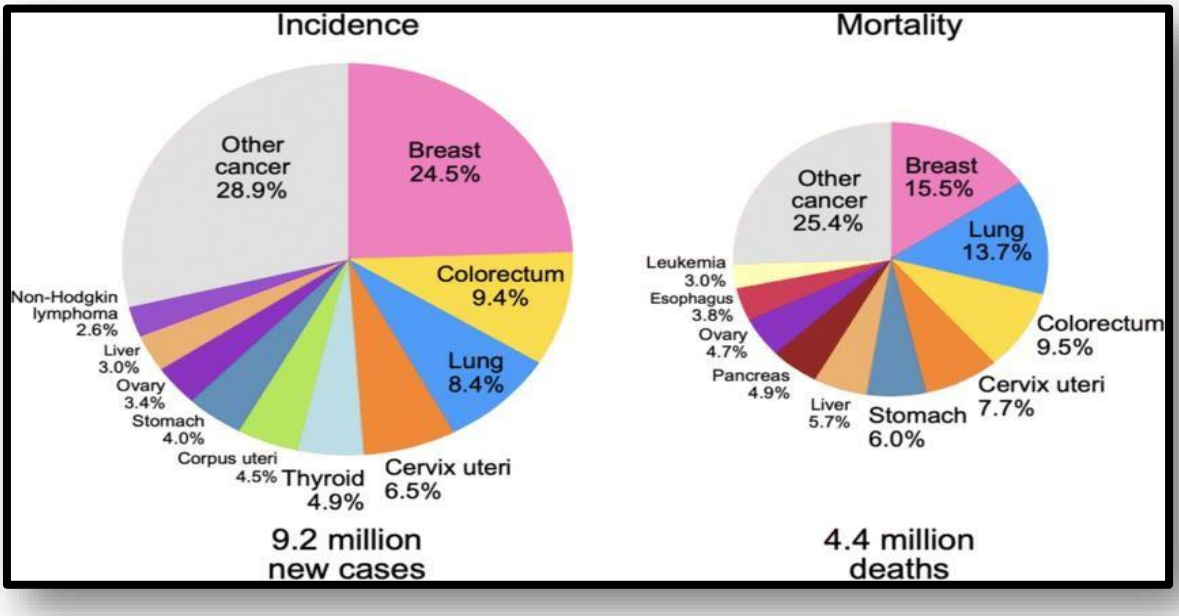


Figure I.3: Distribution of cases and deaths for the top10 most common cancers in 2020 among females for Incidence and Mortality

I.6 Cancer Causes

Cancer can be caused by a variety of factors, with significant contributors including tobacco consumption, which accounts for 22% of cancer deaths, and poor diet, obesity, lack of physical activity, and excessive alcohol consumption, collectively responsible for 10% of deaths. Additional causes include exposure to ionizing radiation, environmental pollutants, and infections. Approximately 15% of cancers globally are attributed to infections such as hepatitis B, hepatitis C, human papillomavirus (HPV), Helicobacter pylori, human immunodeficiency virus (HIV), and Epstein-Barr virus. These infections can lead to genetic changes that may result in cancer. Furthermore, inherited genetic defects from a patient's parents are responsible for 5-10% of cancer cases. Overall, cancer arises from the interaction between genetic factors and three categories of external agents: [3]

i. Physical Carcinogens: Ionizing radiation such as radon, ultraviolet rays from sunlight, uranium, radiation from alpha, gamma, beta, and X-ray-emitting sources.

ii. Chemical Carcinogens: Compounds like n-nitrosamines, asbestos, cadmium, benzene, vinyl chloride, nickel, and benzidine and contains about 60 known potent cancer-causing toxins or chemicals in cigarette smoking or tobacco consumption, a drinking water contaminant (arsenic), a food contaminant (aflatoxin).

iii. Biological Carcinogens: Infections from certain bacteria, viruses, or parasites and Pathogens like human papillomavirus (HPV), EBV or Epstein-Barr virus, hepatitis B and C,
Aging is also the cause of cancer. Age is the common incidence of cancer, which dramatically rises.

- **Genetics:** Genetic is the commonest cause for cancer or tumor-like Ovarian, breast, prostate, skin cancer, colorectal cancer. Individuals that eat heaps of cooked meat can also increase risk because of compounds fashioned at high temperatures. Proving that a substance doesn't cause or isn't associated with hyperbolic cancer risk is tough. [3]

I.7 The stages of cancer

Staging in cancer refers to the process of describing the size of a tumor and its extent of growth, along with assessing whether it has invaded surrounding tissues or metastasized to other parts of the body. Cancer staging may also involve evaluating the grade of the cancer, which indicates how closely the cancer cells resemble normal cells. [4]

- **Reasons for the importance of staging**

The reason staging matters is that it informs your treatment team about the treatments you require.

If your cancer is limited to one location, doctors could suggest a local therapy. Radiation treatment or surgery might be used for this. This may be sufficient to eradicate the malignancy entirely. Just one part of the body is treated by a local therapy. However, if your cancer has progressed, you may require treatment that goes throughout your entire body. [4]

- **The main stages of cancer are**

- Stage 0: abnormal cells are in their original place and have not invaded nearby tissues.
- Stage I: cancer is small and localized to one part of the body.
- Stage II: cancer is larger or has grown into nearby tissues or lymph nodes.
- Stage III: cancer is even larger or has spread to more lymph nodes or other organs near the original tumor.
- Stage IV: cancer has spread to distant organs or throughout the body. [4]

I.8 Symptoms

Early Symptoms: At the earliest stage cancer gives no sign or symptoms by which we cannot indicate the disease. Moreover, the symptoms or signs are shown in harm condition. Some common symptoms that may occur with cancer are as follows:

1. Persistent Cough or Blood-Tinged Saliva: If anyone is having cough from a month or blood in the mucus, then these are the sign of bronchitis or sinusitis, but they could be symptoms of neck, head or lung cancer.

2. A Change in Bowel Habits: It usually depends on the diet of a person and fluid intake. People with cancer felt that they need to have a bowel movement and also feel the same if they had if this symptom lasts more than a few days than it is a symptom of cancer. Mainly in cancer, there is continuous diarrhea.

3. Blood in the Stool: It is also the early symptom of cancer by which we can examine cancer. The evaluation includes colonoscopy *etc*.

4. Unexplained Anemia: People with low RBC in their blood from normal, then this condition is called anemia. Bowel cancer can cause iron-deficiency anemia. The evaluation includes X-ray studies or endoscopy of your lower and upper intestinal tracts.

5. Breast Lump or Breast Discharge: Most breast lumps are noncancerous tumors like cysts or adenomas, but all lumps are needed to check. The evaluation includes Ultrasound and x-ray study included MRI of the breast. Discharge from the breast is also the sign of cancer, and it is quite common, but not from only one nipple or bloody. [3]

6. Lumps in the Testicles: Men with cancer have an uncomfortable or painless lump on a testicle.

7. Change in Urination: The symptoms are slow urine flow, frequent urination, change in bladder function or small amounts of urine, caused by a urinary infection in women or by an enlarged prostate gland. Most men will suffer from enlargement of the prostate gland as they age, these may be the symptom of prostate
cancer. The evaluation includes PSA blood tests and the biopsy of the prostate.

8. Persistent back pain

9. Unexplained weight loss

10. Stomach pain and nausea

11. Bone pain

Late symptoms of cancer can vary depending on the type and location of the cancer, as well as how far it has spread. These symptoms may include changes in bowel or bladder habits, noticeable changes in the appearance of moles or warts, difficulty in swallowing or indigestion, and persistent sore throat or hoarseness. Other signs to watch out for include unexplained weight loss or loss of appetite, nausea, vomiting, fatigue, low-grade fevers, recurring infections, and pain in various parts of the body, such as bones. It's important to note that while many cancers may present with these general symptoms, each type of cancer often has additional specific symptoms. For instance, lung cancer commonly presents with chest pain, persistent cough with bleeding, and fatigue due to shortness of breath. [3]

I.9 Early detection

Cancer mortality is reduced when cases are detected and treated early.

I.9.1. Early diagnosis

When identified early, cancer is more likely to respond to treatment and can result in a greater probability of survival with less morbidity, as well as less expensive treatment. Significant improvements can be made in the lives of cancer patients by detecting cancer early and avoiding delays in care. [5]

Early diagnosis consists of three components:

Being aware of the symptoms of different forms of cancer and the importance of seeking medical advice when abnormal findings are observed, having access to clinical evaluation and diagnostic services, and ensuring timely referral to treatment services.

Early diagnosis of symptomatic cancers is relevant in all settings and the majority of cancers. Cancer programs should be designed to reduce delays in, and barriers to, diagnosis, treatment and supportive care.

I.9.2. The importance of early detection

Early detection of cancer, when it is found in its initial stages, can increase the effectiveness of treatment and lower the mortality rate. Cancer screenings can help detect early signs of cancer. Common cancer screenings include those for cervical and prostate cancer, which can be part of routine exams. Lung cancer screenings are regularly performed for individuals with certain risk factors. Dermatologists can perform skin cancer screenings if there are skin concerns or risks. The American Cancer Society recommends regular screenings for colorectal cancer starting at age 45, typically performed during a colonoscopy, although at-home testing kits may also detect some forms of colorectal cancer. For breast cancer, mammograms are recommended for women aged 45 and older, but screenings can start at age 40 if preferred. For those at high risk, earlier screenings may be advised.[5]

While recognizing cancer warning signs may help people with cancer seek diagnosis and treatment, some cancers may be harder to detect early and may not show symptoms until the later stages. [6]

I.10 Treatment

Since each kind of cancer has a unique treatment plan, a proper diagnosis is crucial for both appropriate and successful care. Typically, systemic therapy (chemotherapy, hormonal treatments, and targeted biological therapies) is used in conjunction with radiation and/or surgery. The disease and the patient receiving therapy are two important factors to consider when choosing a treatment plan. For the intended therapeutic outcome to be achieved, the treatment procedure must be completed within the allotted time. Establishing the treatment's objectives is a crucial first step. The main objective is usually to either cure the patient's cancer or significantly extend their life. Enhancing the patient's quality of life is also a major goal, which can be accomplished by providing palliative care for the patient when their cancer is

terminal. Certain common cancer types, like oral, breast, cervical, and colorectal cancer, have high cure rates when detected early and treated with best practices. [9]

Several alternatives for cancer therapy may be available, dependent on the kind and stage of the disease.

When receiving a localized treatment, a particular part of the body or tumor is targeted for procedures like surgery or local radiation therapy. methodical care. Chemotherapy, immunotherapy, and targeted therapy are examples of systemic pharmacological therapies that can have a whole-body effect. therapy for palliation. Cancer-related health problems, such as pain and difficulty breathing, are relieved as part of palliative care. Various cancer therapies are sometimes combined to eliminate the greatest number of malignant cells [3].

The most common types of treatment are:

I.10.1 Chemotherapy

Chemotherapy is a part of anti-cancer pharmaceutical therapy. It is a form of aggressive cancer treatment as well, employing toxic medications to kill cancer cells rapidly after they divide. The drugs may be given by mouth or into a blood vessel. Different types of drugs may be given together at the same time or one after the other. It can be used to decrease the size of a tumor or the number of cells in your body, in addition to lowering the likelihood that cancer will spread. It's a systemic therapy since it works on the whole body. Medical professionals define cytotoxic as the way chemotherapy acts in specific circumstances. Something that is "cytotoxic" to cells is said to be hazardous. [9]

It depends to:

- your type of cancer
- what the cancer cells look like under a microscope
- whether the cancer has spread
- your general health

I.10.2 Radiation Therapy

To destroy cancer cells, radiation treatment employs particles, radioactive seeds, and protons. Compared to normal bodily cells, cancer cells proliferate and divide more quickly. Radiation treatment harms cancer cells more than normal cells because radiation is most

destructive to rapidly developing cells. This results in cell death by stopping the cancer cells from proliferating and dividing. Radiation therapy can be used in conjunction with other medical interventions such hormone therapy, surgery, chemotherapy, or targeted therapy. [9] Radiation therapy can be administered internally (brachytherapy) or externally (from a machine outside your body). [10]

The two main types of radiation therapy are:

I.10.2.1 External beam: This is the most common form. It aims x-rays or particles at the tumor from outside the body.

I.10.2.2 Internal beam: This form delivers radiation inside your body. It may be given by radioactive seeds placed into or near the tumor; a liquid or pill that you swallow; or through a vein (intravenous, or IV).

I.10.3 Targeted Therapies

Targeted therapy has revolutionized cancer treatment by specifically targeting the mechanisms that drive cancer growth and spread, while sparing normal cells from damage. Unlike conventional chemotherapy, which can harm healthy cells along with cancerous ones, targeted therapy focuses on specific molecules or pathways within cancer cells that are essential for their growth and survival. By targeting these key molecules, targeted therapy disrupts the processes that allow cancer cells to multiply and metastasize. This approach can lead to various outcomes, such as inducing cancer cell death, inhibiting their proliferation, or preventing them from spreading further. Administered either intravenously or orally, targeted therapy medications can also help manage treatment side effects, enhance the immune system's response, and slow down the progression of cancer cells. [7]

Targeted therapy comes in several forms:

- monoclonal antibodies
- cancer growth inhibitors
- angiogenesis inhibitors
- vaccines

I.10.4 Hormonal therapy

The body naturally produces molecules known as hormones. They have an impact on cell development and activity and function as chemical messengers. The endocrine system is the collective term for the organs and glands that create hormones. Hormone treatment stops the growth of cancer cells by removing or blocking the hormones that cause some types of cancer. This medication is frequently used to treat tumors, such as some forms of breast cancer and prostate cancer, that may use hormones to develop and spread. Hormonal treatment comes in several forms. Usually, injections or pills are used to administer them. Depending on the specific medication, the adverse effects might differ. Fatigue, headaches, nausea, and soreness in the muscles or joints are examples of general adverse effects. Treating these hormone-dependent tumors can be accomplished in one of two ways by medications in this family: [11]

either by inhibiting the action of aromatase, an enzyme required to convert androgens to estrogens, or by acting as an antagonist to the action of hormones, such as fulvestrant and tamoxifen, which are anti-estrogen medications. They prevent estrogens from activating cancer cells by competing with them and taking up residence on the cell surface receptors**.**

I.10.5 Stem cell (bone marrow) transplant

Blood cells at their nascent stage of development are called stem cells. Stem cells are the source of all blood cells. Inside the bones lies a spongy substance called bone marrow. The bone marrow is the organ that produces stem cells. Your bone marrow is the material inside your bones that makes blood cells. A bone marrow transplant can use your own cells or cells from a donor. A bone marrow transplant allows your doctor to use higher doses of chemotherapy to treat your cancer. It may also be used to replace diseased bone marrow. [9]

There are 2 different types of stem cell transplants:

- high-dose treatment with stem cell support
- allogenic (donor) stem cell transplants

I.10.5.1 High-dose treatments

Standard chemotherapy is often followed by high-dose treatment with stem cell support. It can raise the likelihood of healing some cancers or leukemia and is used to eradicate any cancer cells that may still be present. Your stem cells are stored and then returned to you following high-dose therapy with stem cell support. You are now able to get chemotherapy at far larger dosages than before.

Autologous stem cell transplantation is another name for this procedure. Different malignancies, as well as some forms of leukemia and lymphomas, are treated with it. Certain uncommon non-cancerous illnesses can also be treated with it. [13]

I.10.6 Immunotherapy

The immune system of your body is used in immunotherapy, sometimes referred to as biological treatment, to combat cancer. Because your immune system fails to identify cancer as an invader, it might remain uncontrolled in your body. With the aid of these treatments, your antibodies are better able to identify cancer and utilize your body's own defenses to eliminate cancerous cells.

Immunomodulatory, which mostly consist of interferons and interleukin-2, are used to boost the immune system's ability to eliminate cancer cells. The three primary functions of interferons, which are cytokines generated by macrophages and non-T, non-B lymphocytes, are to modulate cellular immunity through immunomodulatory means, to inhibit proliferation, and to combat viruses. Certain leukemia, lymphomas, and other malignant disorders can be treated with interferons alpha-2a and alfa-2b. Interleukin-2: is a growth factor that promotes T lymphocyte activation and proliferation. It is used to treat malignant melanoma and some metastatic types of kidney cancer. [14]

I.10.7 Laser Therapy

A laser treatment targets and destroys cancer cells with an extremely narrow beam of light. Tumors and precancerous growths can be eliminated using laser treatment. Tumors obstructing the esophagus, colon, or stomach should shrink. Assist in treating the bleeding symptoms of cancer Seal nerve endings to lessen pain following surgery. Close lymphatic veins during surgery to minimize edema and prevent the spread of malignant cells. Often, a tiny, illuminated tube inserted within the body administers laser treatment. The light is directed towards the cancer cells by the thin fibers at the end of the tube. The skin is also treated using lasers. Most frequently, lasers are used in conjunction with chemotherapy and radiation therapy for the treatment of cancer. [15]

I.10.8 Hyperthermia

Hyperthermia is a therapeutic approach that harnesses heat to selectively target and eliminate cancer cells while preserving normal cells. This technique can be applied to various

scenarios, including targeting a localized area of cells like a tumor, treating specific parts of the body such as an organ or limb, or even addressing cancer throughout the entire body.

The heat is delivered from a machine outside the body or through a needle or probe placed in the tumor.

I.10.9 Photodynamic Therapy

In photodynamic therapy, a person gets a shot of a drug that is sensitive to a special type of light. The drug stays in cancer cells longer than it stays in healthy cells. Then, the doctor directs light from a laser or other source at the cancer cells. The light changes the drug to a substance that kills the cancer cells. [15]

A drug administered by injection initially spreads throughout the body, being absorbed by cells both healthy and cancerous. However, cancer cells tend to retain this medication longer compared to normal cells. After one to three days, the drug remains predominantly in the cancer cells while dissipating from healthy cells. Subsequently, a light source such as a laser is directed at the cancerous area. This light exposure triggers the drug to produce a type of oxygen that actively combats and destroys cancer cells. This process helps in killing the cancer cells, damaging blood cells within the tumor, and also aids the body's immune system in targeting and attacking the tumor.

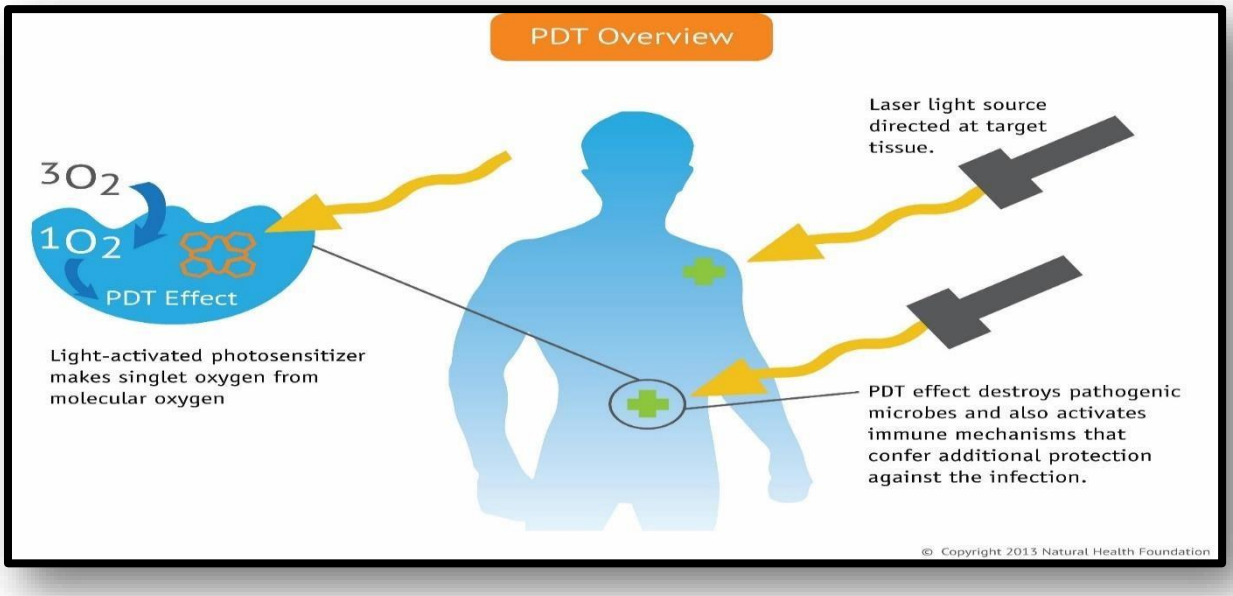


Figure I.4: Overview of the Photodynamic Therapy Function

I.10.10 Cryotherapy

This treatment, also known as cryosurgery, employs extremely cold gas to freeze and destroy cancer cells. It is occasionally used to treat precancerous cells, which are cells on the skin or cervix that have the potential to develop into cancer. Additionally, physicians can apply cryotherapy to internal cancers like those in the prostate or liver by using a specialized tool. Cryotherapy is done using a cotton swab that has been dipped into liquid nitrogen or a probe that has liquid nitrogen flowing through it. [15]

- The procedure is done in your health care provider's office. It usually takes less than a minute.
- The freezing may cause some discomfort. Your provider may apply a numbing medicine to the area first.

I.11 Cyclin dependent kinase 2 inhibitors (CDK2)

I.11.1 Definition of Cyclin-Dependent Kinases (CDK)

Serine/threonine kinases known as Cyclin-dependent kinases (or CDKs) bind to cyclin proteins to create a complex, which is necessary for the complete activation of their kinase activity. In response to both internal and extracellular cues, CDKs are essential for controlling cell division and modifying transcription. CDKs are a family of multifunctional enzymes that can modify various protein substrates involved in cell cycle progression.

The kinases are arranged in a route that makes sure each cell duplicates its DNA correctly during cell division and that the DNA is evenly segregated between the two daughter cells. Apoptosis is caused by deregulation of transcription or any of the cell cycle phases; however, if left untreated, these abnormalities can lead to a number of disorders, including cancer, stroke, and neurodegenerative illnesses like Parkinson's or Alzheimer's disease. The current state of knowledge about the properties of Cyclin-dependent kinases as putative pharmaceutical targets is reviewed in this study. CDKs must also be in a particular phosphorylation state with some sites phosphorylated and others dephosphorylated in order for activation to occur. Correct phosphorylation depends on

the action of other kinases and a second class of enzymes called phosphatases that are responsible for removing phosphate groups from proteins. [17]

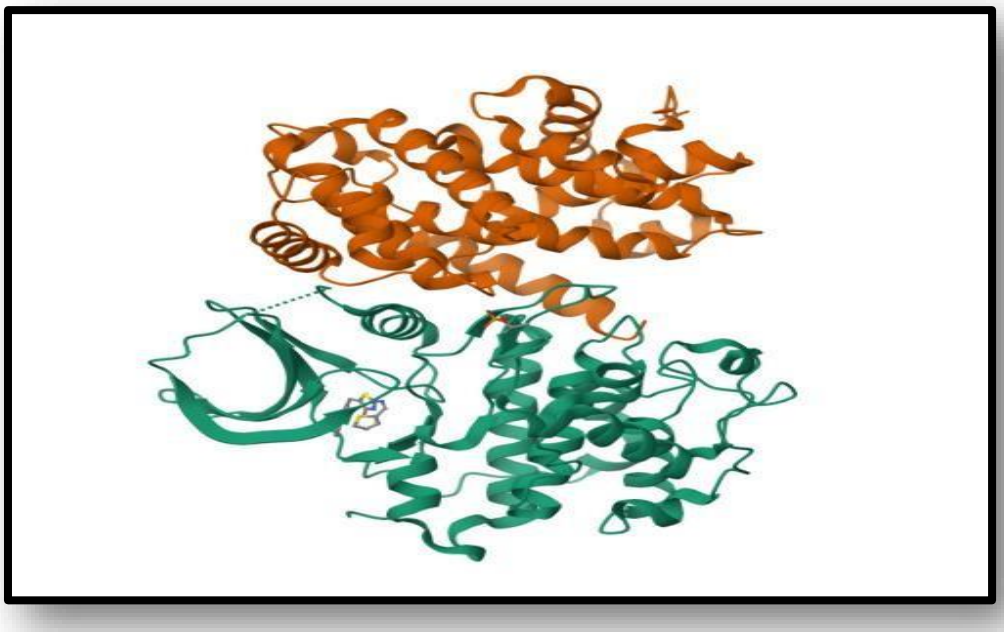


Figure I.5: Biological assembly of CDK2

I.11.2 Cyclin-dependent kinases regulation Cell Cycle

Every eukaryote has a different cyclin, and each one operates at a different phase of the cell cycle. The stage at which cyclins combine with Cyclin-dependent kinases determines the names assigned to each cyclin. There are several different types of cyclins, including G1/S, G1-, S-, and M-phase cyclins. Cyclins in the M-phase generate M-CDK complexes, which guide the entrance of cells into mitosis, while cyclins in the G1 phase form G1-CDK complexes, which propel cells through the G1 phase.

Throughout the whole cell cycle, the same quantities of all Cyclin-dependent kinases are present. The synthesis of new cyclin molecules is required for cell cycle progression, which affects how each step of cyclin production and malfunctions. The cells generate both G1- and at various points in the G1 phase, they generate G1/S-cyclins, and in the G2 phase, they generate M-cyclin molecules. To progress through the cell cycle, cyclin degradation is equally important. At particular points during the cell cycle, some enzymes cause cyclins to malfunction. The associated CDKs stop working when cyclin levels drop. Cyclins that do not breakdown can result in cell cycle arrest. [17]

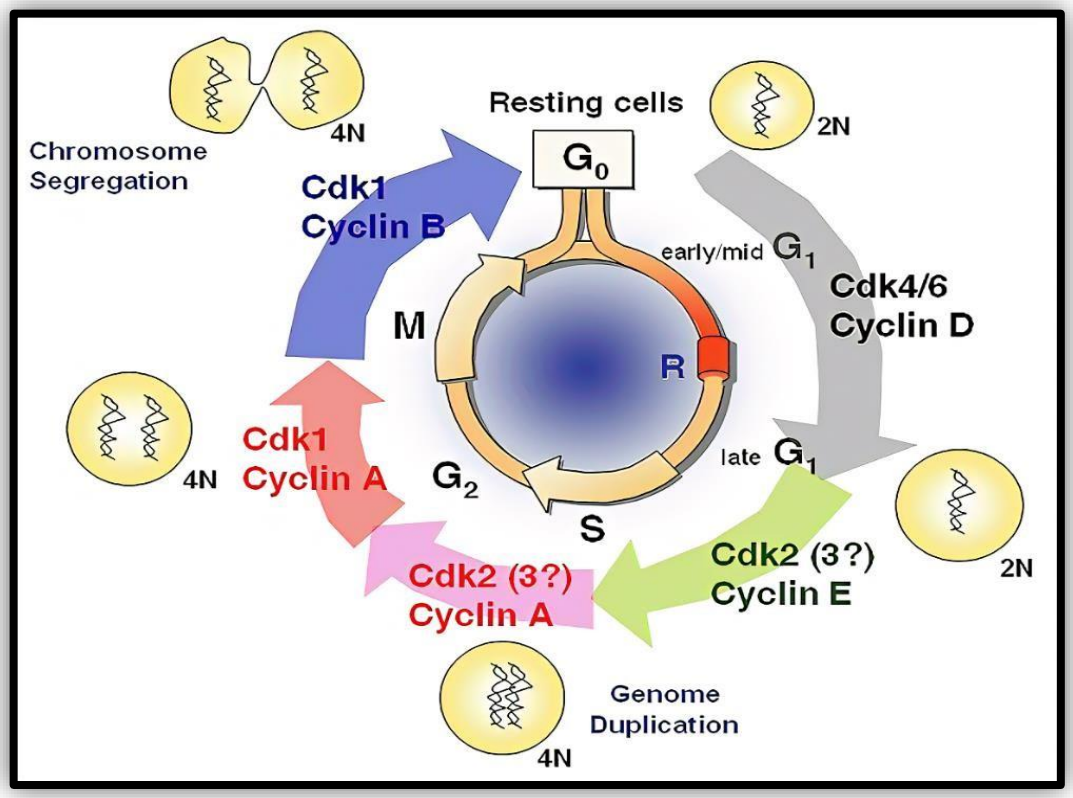


Figure I.6: The mammalian CDKs involved in progression throughout the different phases of the cell cycle.

I.11.3 Cyclin-dependent Kinases' Target Proteins

Each of the cyclin-CDK complexes in a cell alters a particular group of protein substrates. Proper phosphorylation of these substrates ought to happen at specific times in order for the cell cycle to continue. Because cyclin-CDK complexes recognize multiple substrates, they can correlate the various events that happen during each phase of the cell cycle. For example, at the start of S phase, S-CDK catalyzes the phosphorylation of the proteins that start DNA replication by admitting DNA replication complexes to make. During mitosis, M-CDKs phosphorylate a broad range of proteins. These involve condensing proteins that are vital for the substantial condensation of mitotic chromosomes, and laming proteins, which make a stabilizing network under the nuclear membrane, which dismantles during mitosis. M-CDKs also control the assembly of the mitotic spindle by phosphorylating proteins that regulate microtubule behavior. The effect of these coordinated phosphorylation reactions is the final division of chromosomes during mitosis. [18]

I.12. The Purines

I.12.1 Definition

Purine is an aromatic, heterocyclic, chemical molecule made up of an imidazole ring fused to a pyrimidine ring. Purines are the most prevalent kind of nitrogen-containing heterocyclic molecule found in nature, along with substituted purines and their tautomers1. Organic molecules with carbon that have a ring structure that includes atoms other than carbon, including sulfur, oxygen, or nitrogen, are known as heterocyclic compounds. A conjugated ring comprising unsaturated bonds, lone pairs, or empty orbitals that exhibits stabilization stronger than would be predicted by the stabilization of conjugation alone is said to be aromatic. The term "imidazole "designates the parent chemical C3H4N2, whereas imidazole's are a family of heterocycles that differ in their substituents but have similar ring structures. [19]

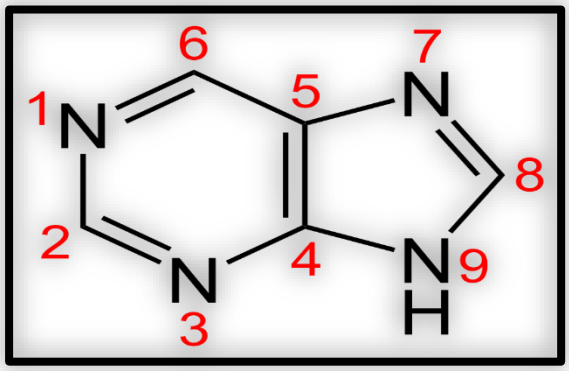


Figure I.7: Structure of purine

This aromatic molecule, with a double aromatic nucleus, is found in nucleotides of nucleic acids (DNA and RNA), ATP (which is an energy source molecule for the cell), coenzymes like NAD+, NADP+, and coenzyme A, molecules used to transmit a cellular signal such as cAMP and G-protein GTP, pigments like xanthine, alkaloids derived from purine such as caffeine and theobromine, and uric acid. In the body, the metabolism of purines allows the elimination of nitrogen in the form of uric acid excreted in the urine.

I.12.2 Biological activities of purine analogue

Purine derivatives, found in a wide range of naturally occurring chemicals and therapeutically effective molecules, are of great importance to both biologists and chemists due to their diverse biological activities. They exhibit antimicrobial properties, including antifungal

and antibacterial activities. Purine derivatives are known for their antiviral, cardiotonic, antineoplastic, anti-tubercular, antiulcer, and antibacterial qualities. Specifically, O6-alkylguanine derivatives play a significant role in inducing cancer, mutations, and cell death by causing GC to AT transition mutations, which inactivate the O6-alkylguanine-DNA alkyl transferase (AGT) protein, enhancing the efficacy of alkylating agents in chemotherapy).[19] Novel purine derivatives have shown high efficacy against bacterial, mycobacterial, autoimmune, cytostatic, and viral infections, with certain 6-chloropurines displaying notable antiviral and cytotoxic effects. Additionally, some purine derivatives exhibit significant antitumor activity, inhibiting the growth of specific human malignant tumor cell lines, such as cervical carcinoma (HeLa), laryngeal carcinoma (Hep2), and pancreatic carcinoma (MiaPaCa2). For example, nucleosides with a 5-substituted pyrimidine moiety can stop the growth of murine mammary cancer cells expressing the TK gene of herpes simplex virus type 1 (HSV-1). Compounds like 1-(6-Chloropurin-7-yl)-2-(2,3-O,O-dibenzyl-2-buten-4-olidylidene)ethane and 1-[6-(N-Pyrrole)purin-9-yl]-2-(2,3-O,O-dibenzyl-2-buten-4-olidylidene)ethane have been evaluated for their antitumor activity against various malignant cell lines, including murine leukemia (L1210/0), murine mammary carcinoma (FM3A), pancreatic carcinoma (MiaPaCa2), breast carcinoma (MCF7), cervical carcinoma (HeLa), laryngeal carcinoma (Hep2), and human T-lymphocytes (Molt4/C8 and CEM/0).[20]

References

[1] CCRA, PIP, UNDERSTANDING CANCER: What is Cancer and Types of Cancer, Canada, 2019.

[2] https://www.who.int/health-topics/cancer#tab=tab_1

[3] Saini, A., Kumar, M., Bhatt, S., Saini, V., & Malik, A. (2020). Cancer causes and treatments. Int. J. Pharm. Sci. Res, 11, 3121-3134.

[4] https://www.cancerresearchuk.org/about-cancer/what-is-cancer/stages-of-cancer#:~:text=Staging%20is%20a%20way%20of,has%20spread%20into%20surrounding%20tissues

[5] https://www.who.int/activities/promoting-cancer-early-diagnosis

[6] Castagnola, E., Mikulska, M., & Viscoli, C. (2015). Prophylaxis and empirical therapy of infection in cancer patients. Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, 3395.

[7] Vial, N., Vallard, A., Jmour, O., Rehailia-Blanchard, A., Mrad, M. B., Trone, J. C., ... & Magné, N. (2019). Innovations en radiothérapie: un regard sur 2018. Bulletin du Cancer, 106(1), 48-54.

[8] Talbert M, Willoquet G, Gervais R. Guide pharmaco(2011) : étudiants et professionnels en soins infirmiers. Rueil-Malmaison .

[9] https://www.cancer.org/cancer/managing-cancer/treatment-types.html

[10] Martini, M., et Ann Med(2014)., Pl3K1AKT signaling pathway and cancer: an updated review. .46(6): p. 372-83.

[11] https://linkinghub.elsevier.com/retrieve/pii/C20110062178

[12] DeVita, V. T., Lawrence, T. S., & Rosenberg, S. A. Cancer: principles & practice of oncology: primer of the molecular biology of cancer/editors, Vincent T. DeVita, Jr, Theodore S. Lawrence, Steven A. Rosenberg; with 82 contributing authors. (No Title).

[13] Jeffrey Schriber, Hillard Lazarus, Robert Peter Gale – (2000)High-Dose Cancer Therapy: Pharmacology, Hematopoietins, Stem Cells"

[14] Cramer, J. D., Burtness, B., & Ferris, R. L. (2019). Immunotherapy for head and neck cancer: Recent advances and future directions. Oral oncology, 99, 104460.

[15] Konstantinopoulos, P. A., Ceccaldi, R., Shapiro, G. I., & D'Andrea, A. D. (2015). Homologous recombination deficiency: exploiting the fundamental vulnerability of ovarian cancer. Cancer discovery, 5(11), 1137-1154.

[16] https://medlineplus.gov/ency/patientinstructions/000901.htm

[17] Shahab, M., Zheng, G., Khan, A., Wei, D., & Novikov, A. S. (2023). Machine Learning-Based Virtual Screening and Molecular Simulation Approaches Identified Novel Potential Inhibitors for Cancer Therapy. Biomedicines, 11(8), 2251.

[18] Tank, J. G., & Thaker, V. S. (2011). Cyclin dependent kinases and their role in regulation of plant cell cycle. Biologia Plantarum, 55, 201-212..

[19] Fields, R. D., Jessen, K. R., Zalc, B., Burnstock, G., Jacobson, K. A., Costanzi, S., ... & Inoue, K. Purinergic Signalling in Neuron–Glia Interactions: Novartis Foundation Symposium 276.

[20] Sachan Dinesh, Gangwar Shikha, Gangwar Bhavana, Sharma Nidhi, Sharma Dileep 2012 Journal of Pharmaceutical and Scientific Innovation BIOLOGICAL ACTIVITIES OF PURINE ANALOGUES: A REVIEW JPSI 1 (2), p29 34.

ChapiterII

Computational Methods in drug Discovry

Chapiter II: Computational Methods in Drug Discovery

II.1. Drug discovery

Drug discovery is a captivating and dynamic field that encompasses the exploration, identification, and development of new therapeutic compounds to treat various diseases and improve human health. It combines elements of chemistry, biology, pharmacology, and medicine, drawing on interdisciplinary approaches to unearth innovative solutions for medical challenges. Over the years, drug discovery has evolved significantly, transitioning from traditional methods rooted in natural remedies to sophisticated techniques like computational modeling, high-throughput screening, and molecular biology. This evolution has led to the discovery of groundbreaking medications that have revolutionized healthcare and extended life expectancy. Today, drug discovery continues to be at the forefront of scientific research, driving progress towards more effective and targeted treatments for complex medical conditions..]1[

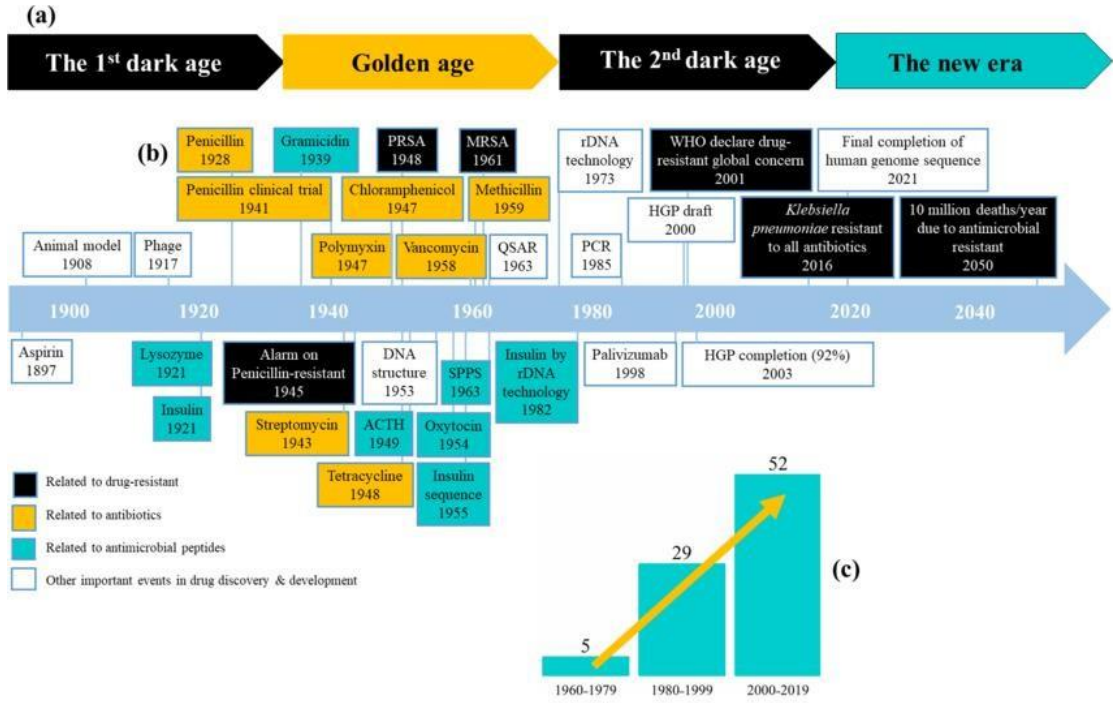


Figure 1 : Several major milestones in drug discovery during the 19th and 20th centuries

II.2. Steps in Modern Drug Discovery

Step 1: Target identification

Target identification is the first key stage in the drug discovery pipeline. Generally speaking, a drug target is the specific binding site of a drug in vivo through which the drug exerts its action [1]. A specific drug target might have the following characteristics:

1 .The drug target is a biomolecule(s), normally a protein that could exist in isolated or complex modality.

2 .The biomolecules have special sites that match other.

3. The biomolecular structure might change when the biomolecule binds to small molecules and the changes in structure normally are reversible.

4. Following the change in the biomolecule's structure various physiological responses occur and induce regulation of the cell, organ, tissue, or body status.
5. The physiological responses triggered by the changes in biomolecule structure play a major role in complex regulation and have a therapeutic effect on pathological conditions.
6. The expression, activity, and structure of the biomolecule might change over the duration of the pathological process.
7. Small molecules binding to the biomolecules are drugs. As is apparent from the above discussion, a drug target is a key molecule involved in a particular metabolic or signal transduction pathway that is specific to a disease condition or a specific disease. However, the term 'drug target' itself has several limitations and is debated within the pharmaceutical industry. In this respect, several points should be kept in mind.

First, a drug target is a relative concept. For starters, a drug target is, just like other biomolecules, also a biomolecule involved in a transduction pathway. The difference between the two is only in their location and role in the transduction pathway. Another aspect is that a drug target is disease-dependent, that is, every target is involved in a special spectrum of diseases.

Second, most human diseases are rather complicated and involve many risk factors, so there are clearly many different drug targets with respect to a specific disease. Targeting a specific target could not conceivably cure a kind of disease. However, the involvement of many targets in a disease does not mean that each target shares equally in the pathogenesis of the disease and thus drugs targeting these targets would not be equally effective in the therapy of the disease.

Third, drug targets can change, which means that with the development of insights into biomolecules and their role in the pathogenesis of a certain disease, drug targets might be not as important as or may be much more important than currently believed. In fact, the establishment of drug targets is based on understanding of the pathogenesis of the disease.

Fourth, there are many drugs targeting the same target and one drug may have more than one target. The relationship between a drug and its target is not one-to- one but one-to-many or many-toone.

Fifth, when a new drug target is discovered and validated, researchers usually hope to obtain more specific drugs targeting the target. However, a key understanding to keep in mind is that the body is a subtle organism and a more specific drug might disrupt the homeostasis of the body. Compared to aspirin, rofecoxib is a specific COX-2 inhibitor. However, studies had shown that rofecoxib increases cardiovascular risks, resulting in rofecoxib's withdrawal from the drug market.

Sixth, a drug target usually refers to a single biomolecule. According to whether there are drugs available, a drug target can be classified into two classes: established drug targets and potential drug targets. The former are those for which there is a good scientific understanding, supported by a lengthy publication history regarding both how the target functions in normal physiology and how it is involved in human pathology. Furthermore, there are many drugs targeting this target. The latter are those biomolecules whose functions are not fully understood and which lack drugs targeting them. Potential targets suggest directions for completely new drug research.

Step 2: Target validation

New target validation is the basis of completely new drug exploration and the initial step of drug discovery. New drug target validation might be of great help not only to new drug research and development but also provide more insight into the pathogenesis of target related diseases [2]. Basically, the target validation process might include six steps:

1. Discovering a biomolecule of interest.

2. Evaluating its potential as a target.

3. Designing a bioassay to measure biological activity.

4. Constructing a high-throughput screen.

5. Performing screening to find hits.

6. Evaluating the hits.

The drug discovery process starts with the identification or growing evidence of, biological targets that are believed to be connected to a particular condition or pathology. Information supporting the role of these targets in disease modulation can come from a variety of sources [3]. Traditionally, the targets have been researched and largely discovered in academic laboratories, and to a lesser extent in the laboratories of pharmaceutical and biotechnology companies. Basic research into understanding the fundamental, essential processes for signaling within and between cells and their perturbation in conditions has been the basic approach for establishing potential targets suitable for drug intervention.

Step 3: Lead discovery

Once a disease- associated molecular target has been identified and validated in disease models, in the lead generation phase, compounds are identified which interact intact animals or disease-related cell-based models that can provide information about the integrative response of an organism to a pharmacological intervention and hereby help to predict the possible profile of new drugs in patients. This is accomplished primarily with knock-out or knock-in animal models; small molecule molecular target *in vitro* usually precedes the validation of the therapeutic concept *in vivo*; together this defines its clinical potential. Validation involves studies in molecular target *in vitro* usually proceeds with the target protein and modulate its activity. Libraries of compounds that are either synthetic chemicals, peptides, natural or engineered proteins, or antibodies are exposed to the target in a manner that will detect and isolate those members of the library that interact with and, preferably, have an effect on the target [4-8]. The compounds selected are called “leads”. Initially screening can be performed by searching for compounds that bind to the target, but binding is not sufficient for therapeutic activity. More recent screening procedures include an activity-based readout as part of the initial screening assay. For example, if the goal is to inhibit a protein that is involved in activating the expression of a particular gene or set of genes, the assay can include readout to determine if the expression of the gene is reduced by

the compound. Such assays can be cell based, but more often they are enzymatic assays that can be performed in a high-throughput manner for compounds that bind to the target, but binding is not sufficient for therapeutic activity. More recent screening procedures include an activity-based readout as part of the initial screening assay. For example, if the goal is to inhibit a protein that is involved in activating the expression of a particular gene or set of genes, the assay can include readout to determine if the expression of the gene is reduced by the compound. Such assays can be cell-based, but more often they are enzymatic assays that can be performed in a high throughput manner.

Step 4: Lead optimization

Lead optimization is a process that begins with a compound that displays an interesting biological action and ends with the identification of the best analog. Molecules are chemically modified and subsequently characterized in order to obtain compounds with suitable properties to become a drug. Leads are characterized with respect to pharmacodynamic properties such as efficacy and potency *in vitro* and *in vivo*, Physiochemical properties, pharmacokinetic properties, and toxicological aspects.

Potency - refers to the amount of drug required for its specific effect to occur.

Efficacy - measures the maximum strength of the effect itself, at saturating drug concentrations.

Pharmacokinetics - determines the fate of xenobiotics. It explains about "What the body does to the drug". It often divided into areas examining the extent and rate of adsorption, distribution, metabolism, and excretion (ADME).

Pharmacodynamics– It determines the biochemical and physiological effects of drugs, the mechanism of drug action and the relationship between drug concentration and effect. It explains about "What the drug does to the body". This process ideally requires the simultaneous optimization of multiple parameters and is thus a time consuming and costly step. This is often the tightest bottleneck in drug discovery. However, by turning a biologically active chemical into an effective and safe drug, lead optimization contributes essentially towards added value in the drug discovery process.

Step 5: Pre-clinical and clinical development

Pre-clinical development: The pre-clinical development includes the following: develop large scale synthesis; animal safety studies; carcinogenicity tests; drug delivery; elimination and metabolism studies; drug formulation experiments; dose-ranging studies in animals. Wide ranging dosages of the compounds are introduced to

the cell line or animal in order to obtain preliminary efficacy and pharmacokinetic information.

Clinical development

The NIH organizes clinical trials into 5 different types:

1. Treatment trials: test experimental treatments or a new combination of drugs.

2. Prevention trials: look for ways to prevent a disease or prevent it from returning.

3. Diagnostic trials: find better test or procedures for diagnosing a disease.

4. Screening trials: test methods of detecting diseases.

5. Quality of life trials: explore ways to improve comfort & quality of life for individuals with a chronic illness.

Pharmaceutical clinical trials are commonly classified into 4 phases.

Phase 0 - A recent designation for exploratory, first in human trials designed to expedite the development of promising therapeutic agents by establishing early on whether the agent behaves in human

subjects as was anticipated from preclinical studies.

Phase 1 - A small group of healthy volunteers (20-80) are selected to assess the safety, tolerability, pharmacokinetics, & pharmacodynamics of a therapy. Normally include dose ranging studies so that doses for clinical use can be set/adjusted.

Phase 2- Performed on larger groups (20-300) & are designed to assess the activity of the therapy, & continue phase1 safety assessments.

Phase 3 -Randomized controlled trials on large patient groups (hundreds to thousands) aimed at being the definitive assessment of the efficacy of the new therapy, in comparison with standard therapy. Side effects are also monitored. It is typically expected that there be at least two successful phase 3 clinical trials to obtain approval from the FDA. Once a drug has proven acceptable, the trial results are manufacturing procedures, formulation details, shelf life, etc. This document is submitted to the FDA for review.

Phase 4 - Post-launch safety monitoring & ongoing technical support of a drug may be mandated or initiated by the pharmaceutical company designed to detect rare or long-term adverse effects over a large patient population & timescale than was possible during clinical trials.]3[

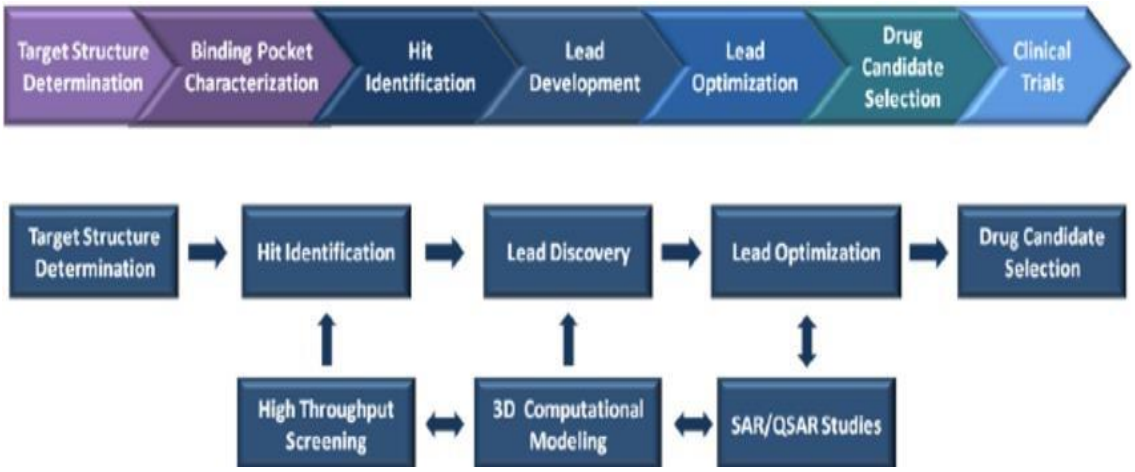


Figure 2 : Drug Discovery Pipeline

II.3. Virtual Screening

Screening methods aim to eliminate presumed inactive compounds or undesirable molecules while prioritizing compounds most likely to be active, which is easier to implement at a much lower cost than experimental screenings. Screening methods are divided into two main families: "ligand-based" and "structure-based," and the choice of their use is based on data availability.

II.3.1. Ligand-based Virtual Screening

Ligand-based screening methods rely on prior knowledge of ligands with activity against the therapeutic target. It will thus be possible to use these ligands as a starting point to identify other similar compounds, having activity characteristics common to the known ligands of the target. [9,10]

Different types of molecular descriptors can be calculated to quantify the similarity between compounds. Depending on the number of known ligands for the therapeutic target, several methods can be used: similarity search, pharmacophore screening, or QSAR approaches. [10]

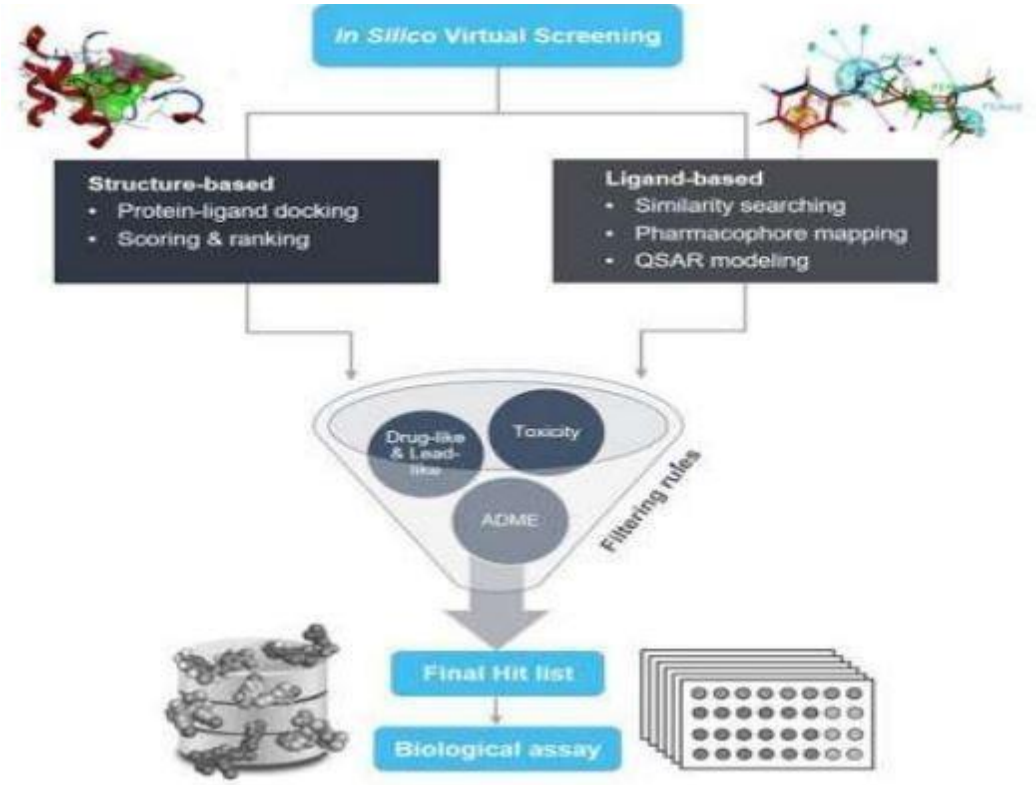


Figure 3: Classification of virtual screening methods

I1.3.1.1 Similarity Search

Similarity search is used when a known ligand for the therapeutic target is identified. Similarity metrics will then be used to compare reference ligands to screened molecules using appropriate molecular descriptors to predict their activity profiles.

II.3.1.2 Ligand-based Pharmacophore Models

The term "pharmacophore" was first used by Lemont.Kier in 1971, later published in 1978 [11]. The pharmacophore is defined as the three-dimensional arrangement of features enabling a molecule to have specific biological activity. It represents the set of steric and electrostatic molecular properties called pharmacophore points.

pharmacophoric (hydrogen bond donors and acceptors, electrostatic interactions, aromatic and hydrophobic groups) necessary to establish optimal supramolecular

interactions with a specific biological target to trigger or block its biological response. A pharmacophore can thus be used as a reference for screening chemical libraries, searching for molecules that fit into it.

A pharmacophore is called "ligand-based" when it is determined from active compounds used as reference without considering the structure of their receptor. Conversely, a pharmacophore is called "structure-based" when it is constructed from the binding site structure of the target under study.

The majority of pharmacophore construction methods operate on the same principle.

II.3.1.3 Quantitative Structure-Activity Relationship (QSAR)

Quantitative Structure-Activity Relationship (QSAR) properties are fundamental in the field of computational chemistry and molecular modeling. QSAR models are used to predict the activity of chemical compounds based on their chemical structure. Key properties in QSAR include physicochemical descriptors such as molecular weight, partition coefficient (logP), and solubility; electronic descriptors like molecular orbital energies and dipole moment; topological descriptors that describe the molecule's structure without considering its geometry; and geometrical descriptors that account for the three-dimensional shape of the molecule. The QSAR modeling process involves collecting data, selecting relevant descriptors, building the model using machine learning or statistical techniques, interpreting the model to understand the structure-activity relationship, and using the model to predict the activity of new compounds. QSAR models are used in drug discovery, toxicity assessment, environmental chemistry, and chemical engineering, aiding in the optimization of material properties based on their molecular structure.

II.3.2 Structure-Based Virtual Screening

As the name suggests, these methods use the structure of the target to discover new active compounds. When the 3D structure of the biological target of interest is available, so-called structure-based methods can be used for virtual screening. There are two types of experimental methods to obtain the 3D structure of a target: X-ray crystallography and nuclear magnetic resonance (NMR). When the experimental 3D structure has not yet been resolved, it is also possible to resort to homology modeling.

After obtaining a structure of the therapeutic target, several "structure-based" approaches can be employed for the identification of active compounds:

pharmacophore model construction, de novo ligand design, or methods such as RD-QSAR and docking, which are the most popular. Each of these methods requires the prior identification of a binding site on the therapeutic target.

II.3.2.1 Structure-Based Pharmacophore

As described earlier, "ligand-based" pharmacophore models allow the identification of new active compounds based on knowledge of ligands of the target under study. However, when the structure of the target has been resolved, the use of receptor shape, volume, or physicochemical information can make pharmacophore methods more powerful. Structure-based pharmacophore models can be built either from the receptor structure or from a receptor-ligand complex. [12]

II.3.2.2 Molecular Docking

In silico docking, also known as molecular docking, is a computational technique aimed at predicting the structure of a complex formed by two molecules (target and ligand) or

more molecules. The formation of these complexes relies on the recognition of the three-dimensional structure of a ligand by a receptor site and controls the activity of numerous molecules. Most often, the receptor is a protein that has one or more specific active sites, more or less accessible depending on the case. The ligand is generally a small, flexible foreign molecule. [13]

The use of docking methods for drug design began over 30 years ago. These methods have two main objectives: (i) the identification of ligands for the target under study from large chemical libraries, through the prediction of affinity scores or actual affinity values (Ki, IC50, etc.), and (ii) determining the binding modes adopted by the ligands in the receptor.

II.3.2.2 .1 General Principle of Molecular Docking

Molecular docking is divided into two distinct stages. The first section consists of search algorithms, which are capable of generating a large number of possible structures and determining the binding mode. The second section is dedicated to scoring, which are mathematical methods used to estimate the interaction strength and binding affinity between two molecules after the docking phase. The best result fo r docking is the protein-ligand complex with the lowest energy. [14]

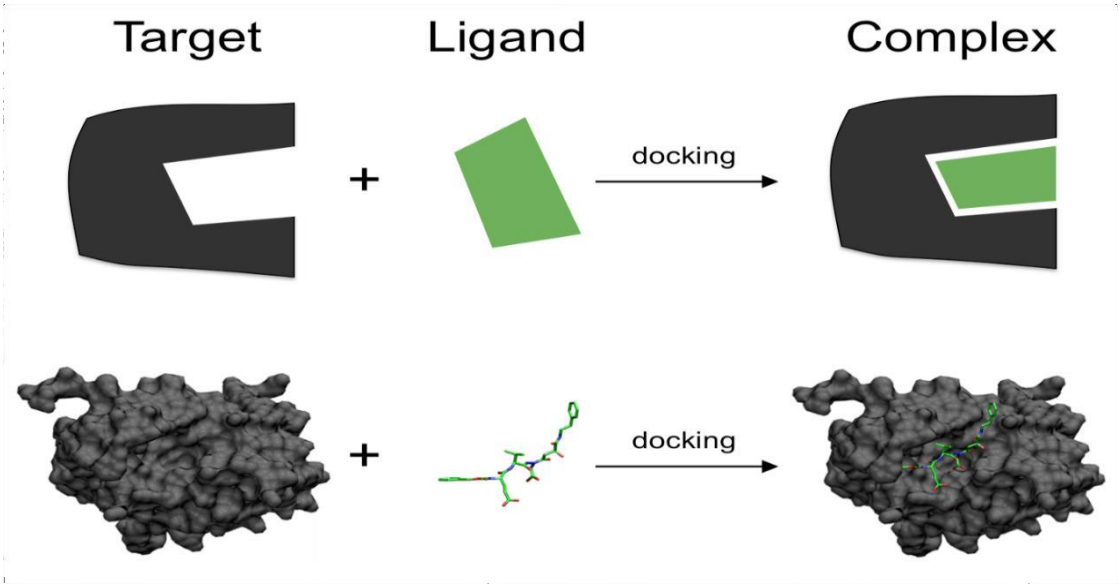


Figure 4: Mechanisms of Docking

II.3.2.2 .2 Types of Molecular Docking

There are three types of molecular docking:

A. Rigid Docking

The binding mechanism of a ligand to its receptor has been considered a static process in which the ligand acts as a key with a complementary shape to that of the lock it is capable of "opening," the receptor (the "lock-and-key" model). In the case of rigid docking methods, the search for the optimal pose is limited to the positioning by the enumeration of all possible rotations and translations for a ligand within the binding site. The interest of rigid docking methods lies in their execution speed, which allows for an initial screening of very large chemical libraries before proceeding, for example, to the implementation of flexible docking methods. [15]

B. Flexible Docking

The biological ligand-receptor system has been depicted as a dynamic system, where ligand binding to an interaction site occurs through a conformational selection process followed by an adaptation step of the ligand and the receptor to each other. It is therefore important for docking methods to consider this flexibility of the ligand and

the receptor. When docking methods account for ligand flexibility, two steps are performed successively throughout the docking process.

There are several types of algorithms for handling ligand flexibility: systematic methods (fragmentation/reconstruction), random methods, and simulation methods (molecular dynamics). [16]

C. Semi-Flexible Docking

When exploring the conformational space of ligands, the number of degrees of freedom in the search space can be significant for highly flexible molecules. Therefore, the use of exhaustive search methods often seems inappropriate as they require significant simplifications in sampling. Fragmentation algorithms are thus employed to incrementally construct the ligand within the protein's active site. The ligand's conformational space is then limited to the vicinity of an initial set of simplified states. This construction-based search strategy is adopted by programs such as DOCK, FLExX, and Hammerhead. [17]

II.3.2.2 .3 Search Algorithms

There are various algorithms that take into account ligand flexibility, including systematic algorithms and random or stochastic algorithms. [18]

A. Systematic Algorithm

Systematic search algorithms aim to explore all degrees of freedom of ligands by rotating from 0 to 360° of all rotatable bonds using a chosen incremental step. As a result, the number of generated conformations can be very high, leading to what is called combinatorial explosion. To address this issue, systematic search methods often use a fragmentation-reconstruction algorithm. The general principle is to cut the ligand into rigid and flexible fragments. One or more anchors can be defined between points where rotations are possible. First, one or more fragments that must be rigid are placed within the active site and interact with the target. Then, the ligand is reconstructed by placing the flexible fragments successively while exploiting torsional angles. This method has been integrated into several programs such as Dock, FlexX, and Surflex. [19].

B. Stochastic Algorithm

The stochastic approach involves making random changes in the three-dimensional structure of the ligand. The genetic algorithm is one of the most important stochastic algorithms. This method has been implemented in several programs including AutoDock and GOLD. Regarding stochastic search, the Monte Carlo (MC) method and genetic algorithm can be mentioned. [20]

In the Monte Carlo method, the ligand is randomly placed within the interaction site, where the interaction will be evaluated by an objective function (scoring function), and thus this conformation is scored. A new conformation is generated by making random changes in the rotatable bonds of the ligand and its spatial position (translation and rotation). After each change, the ligand is minimized, and its energy is calculated (scored). Then, if the new configuration is better than the previous one according to a probability function derived from Boltzmann's law, called the Metropolis criterion, it will be immediately accepted. [21]

II.3.2.2 .4 Scoring Process

A. Principle

The score is a numerical value useful for quantifying the degree to which a ligand binds to a receptor. It is typically an approximation of the free energy resulting from the transition from the free form of the protein and the ligand to the associated complex form. The thermodynamic principle is as follows:

$$\Delta G = \Delta G_{\text{complexe}} - \Delta G_{\text{ligand}} - \Delta G_{\text{protéine}}$$

Utilization of scoring functions to mathematically estimate the binding affinity between a receptor and each of the poses generated during docking is twofold. Firstly, they help determine the best conformation of the ligand in question. This conformation is called the "first pose." The other use of scores is to rank the first poses of each ligand to establish a final ranking of the most promising molecules. [22]

B. Families

Scoring functions can be classified into several categories: force field-based scoring functions, empirical scoring functions, and knowledge-based scoring functions. These classes of functions are based on a set of ligand-protein complexes. Hence, they have

the limitation of being able to predict only interactions that do not deviate too much from those recorded in the studied complexes. [23]

- Force Field-based Scoring Functions: In this type of functions, the score is the sum of the ligand-receptor interaction energy, the internal energy of the ligand, and the internal energy of the protein, with the internal energy of the protein being negligible when its conformation does not vary. The energies are calculated from a van der Waals term (represented by a Lennard-Jones potential) and an electrostatic term (Coulomb potential) coupled with a distance-dependent dielectric function that attenuates the contributions of charge-charge interactions. Many docking software use force field-based scoring functions. However, the values of force fields are not uniform, However, the values of the force fields are not always the same. Thus, the G- Score function (based on the Tripos force field) and the one implemented in AutoDock (based on the AMBER force field) are examples of this type of scoring function.
- Empirical Scoring Functions: This type of scoring function approximates the free energy of binding by summing weighted interaction terms derived from structural parameters. The different weights of the scoring function are adjusted to prioritize reproducing experimental data, such as binding constants from a training set of protein-ligand complexes. Most docking programs implement this type of scoring function, reflecting their effectiveness (in terms of precision/speed ratio). However, the main disadvantage of these empirical functions is their strong dependence on the data used for calibration, which, if poorly parameterized, can limit their transferability to different systems. Among the main empirical scoring functions are ChemScore, PLP, and LigScore.
- Knowledge-based Functions: These functions stem from the analysis of three-dimensional structures of experimentally determined ligand-protein complexes. Rules defining the preferred geometry of interactions are deduced from these structures using statistical methods.
- This alternative to empirical functions is more tolerant of interactions within the complex. Their expressions are less strict than in the case of empirical functions. The PMF function is part of this class of functions.

Chapiter II: Computational Methods in Drug Discovery

II.4. ADME-TOX

ADME is a four-letter acronym for Absorption, Distribution, Metabolism, and Excretion, which describes the movement of molecules in the body in space and time known as "pharmacokinetics". It is an important concept that describes the potential impact of a chemical or drug on a living system in the context of cell biology and biochemistry. Indeed, the movement and metabolism of molecules are determined by the physicochemical properties of the molecule as well as the host system [24]. Currently, the development of new drugs is no longer solely based on the perpetual increase in activity against a specific target. Indeed, ADMET parameters must be taken into account from the early stages of research to avoid drug failures in these more advanced stages of development and thus minimize the losses it induces [25].

II.4.1 Drug likeness

This rule describes a set of criteria for estimating the oral bioavailability of a compound based on its two-dimensional (2D) structure. The criteria include the molecular weight (MW), which should not exceed 500 daltons (Da), the partition coefficient (LogP), which should be less than or equal to 5, the number of hydrogen bond donors (nOH, NH), which should be less than or equal to 5, and the number of hydrogen bond acceptors (nO, N), which should be less than or equal to 10.

Additional criteria have been introduced by Veber, including the number of rotatable bonds (FRB) which should be less than 10, and the polar surface area of the molecule (PSA) which should be less than 140 $Å^2$. The polar surface area represents the sum of the surfaces of the polar atoms in the molecule and helps predict intestinal absorption and passage through the blood-brain barrier.

I1.4.2 Pharmacokinetic and Toxicity Properties

II.4.2.1 Absorption

Absorption refers to the penetration of the drug into the body

In other words, absorption encompasses all the phenomena involved in the transfer of the active drug substance from its site of administration into the bloodstream.

II.4.2.2 Distribution

Distribution depends on the drug's ability to cross barriers, which can occur through passive diffusion or transporter action

II.4.2.3 Metabolism

Drug metabolism is generally divided into two stages. During the first stage, molecules are primarily oxidized to form more polar and therefore more soluble compounds. This process is often followed by a second conjugation stage. The metabolites resulting from these different reactions may have different properties compared to the initial compound [24].

II.4.2.4 Excretion

Elimination is defined as the volume of plasma containing the amount of drug eliminated per unit of time. If, strictly speaking, the clearance of a drug can be renal, usually, the two main routes to consider in drug studies [26].

II.4.2.5 Toxicity

The toxicity of a drug candidate is one of the most feared parameters by developers, especially when it enters clinical phases. Indeed, a drug candidate showing serious side effects during clinical phases is immediately abandoned, usually permanently. It is therefore very important to try to predict the toxicity of a compound based on its structure during the early stages of R&D development [6].

References

1. Drews, J. (2000). Drug discovery: a historical perspective. science, 287(5460), 1960-1964.
2. Wang, S., Sim, T. B., Kim, Y. S., & Chang, Y. T. (2004). Tools for target identification and validation. Current opinion in chemical biology, 8(4), 371-377.
3. Du, G. H. (2004). Evaluation and validation of drug targets. Acta Pharmacol Sin, 25, 1566.
4. Leach, A. R., Shoichet, B. K., & Peishoff, C. E. (2006). Prediction of protein− ligand interactions. Docking and scoring: successes and gaps. Journal of medicinal chemistry, 49(20), 5851-5855.
5. http://dx.doi.org/10.1021/jm060999m
6. Enyedy, I. J., & Egan, W. J. (2008). Can we use docking and scoring for hit-to-lead optimization?. Journal of computer-aided molecular design, 22, 161-168.
7. https://doi.org/10.1007/s10822-007-9165-4.
8. Reddy, N. P., Reddy, T. C., Aparoy, P., Achari, C., Sridhar, P. R., & Reddanna, P. (2012). Structure based drug design, synthesis and evaluation of 4-(benzyloxy)-1-phenylbut-2-yn-1-ol derivatives as 5-lipoxygenase inhibitors. European journal of medicinal chemistry, 47, 351-359.
9. Ajay, Walters, W. P., & Murcko, M. A. (1998). Can we learn to distinguish between "drug-like" and "nondrug-like" molecules?. Journal of medicinal chemistry, 41(18), 3314-3324.
10. Toropov, A. A., Toropova, A. P., Raska Jr, I., Leszczynska, D., & Leszczynski, J. (2014). Comprehension of drug toxicity: software and databases. Computers in biology and medicine, 45, 20-25.
11. Toropov, A. A., Toropova, A. P., Raska Jr, I., Leszczynska, D., & Leszczynski, J. (2014). Comprehension of drug toxicity: software and databases. Computers in biology and medicine, 45, 20-25.
12. Valerio Jr, L. G. (2009). In silico toxicology for the pharmaceutical sciences. Toxicology and applied pharmacology, 241(3), 356-370..
13. Bawden, D. A. V. I. D. (1990). Applications of Two-dimensional Chemical Similarity Measures to Database Analysis and Querying'. Concepts and Applications of Molecular Similarity, 65-76.
14. Hawkins, P. C., Skillman, A. G., & Nicholls, A. (2007). Comparison of shape-matching and docking as virtual screening tools. Journal of medicinal chemistry, 50(1), 74-82.

15. Mason, J. S., Good, A. C., & Martin, E. J. (2001). 3-D pharmacophores in drug discovery. Current pharmaceutical design, 7(7), 567-597.
16. Gelbin, A., Schneider, B., Clowney, L., Hsieh, S. H., Olson, W. K., & Berman, H. M. (1996). Geometric parameters in nucleic acids: sugar and phosphate constituents. Journal of the American Chemical Society, 118(3), 519-529.
17. Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. Journal of molecular biology, 267(3), 727-748.
18. Shulman-Peleg, A., Mintz, S., Nussinov, R., & Wolfson, H. J. (2004). Protein-protein interfaces: Recognition of similar spatial and chemical organizations. In Algorithms in Bioinformatics: 4th International Workshop, WABI 2004, Bergen, Norway, September 17-21, 2004. Proceedings 4 (pp. 194-205). Springer Berlin Heidelberg.
19. Olivecrona, M., Blaschke, T., Engkvist, O., & Chen, H. (2017). Molecular de-novo design through deep reinforcement learning. Journal of cheminformatics, 9, 1-14.
20. van Gunsteren, Wilfred F. ; Berendsen, Herman J. C Angewandte Chemie,(September 1990), vol. 102, issue 9, pp. 1020-1055
21. Louacheni, F.Docking protéine-ligand avec AutoDock, Mémoire de Master, Département d'informatique, Francophone Institute for computer sciences, Hanoi, (2014).
22. Beautrait, A. (2008). Développement et validation de la plateforme de criblage virtuel VSM-G et étude du domaine FAT de la kinase d'adhérence focale FAK (Doctoral dissertation, Université Henri Poincaré-Nancy I).
23. Dar, A. M., & Mir, S. (2017). Molecular docking: approaches, types, applications and basic challenges. J Anal Bioanal Tech, 8(2), 1-3.
24. Brooijmans, N., & Kuntz, I. D. (2003). Molecular recognition and docking algorithms. Annual review of biophysics and biomolecular structure, 32(1), 335-373.
25. Krid, A., & Bencharif, M. (2008). Modélisation de nouvelles molécules biologiques actives liées au stress oxydant.
26. Guillemain, H. (2012). Evaluation et application de méthodes de criblage in silico (Doctoral dissertation, Paris, CNAM) .

Chapiter III

Molecular docking and ADME study of 6-substituted 2-aryaminopurine derivatives as anticancer agents

1. Introduction :

Heterocyclic compounds have a wide range of applications in medicinal chemistry, and about 80% of the drugs on the market contain heterocyclic characteristics. Among them [1], Purines compounds are the most studied basic components of active molecules, and have shown excellent activity characteristics in anti-tumor, anti-viral, and anti-infective drugs, the investigation of purines, holds immense promise in the field of drug discovery and development due to their diverse pharmacological activities and therapeutic potential.

In computer -aided drug design (CADD), molecular docking is widely used for identifying protein-ligand interaction modes. Molecular docking is one of the most frequently used methods in drug design [2], because of its ability to predict how small molecules interact with the appropriate target binding sites. Molecular docking was used to explore the mode of interaction of target protein (CDK2) to its inhibitors to find the most stable configuration that is similar to the bioactive one.

Furthermore, the evaluation of Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADME-TOX) properties is critical for assessing the pharmacokinetic behavior, metabolic fate, and safety profiles of purine derivatives in drug development. ADME-TOX studies provide essential information regarding the bioavailability, metabolic stability, and potential toxicity of these compounds, guiding decision-making processes throughout the drug discovery pipeline.

Therefore, the objective of this study is to conduct a comprehensive investigation into the molecular docking interactions and ADME-TOX profiles of purine derivatives as potential therapeutic agents.

2. Materials and Methods

III.2.1. Computer System and Software

- **Microcomputer**

We used a laptop (Asus) running Windows 10 Professional, which incorporates an Intel processor (Core TM I3-5005U CPU @ 2.00 GHz) with 8.00 GB of RAM. All software used are installed under the 64-bit operating system, version 22H2.

- **Chemdraw pro 12.0**

Chem draw is a comprehensive tool designed for chemists and biologists, incorporating a wide range of intelligent tools to facilitate researchers' work on a daily basis. In addition to the features of ChemDraw Prime, it includes numerous innovative tools such as NMR prediction or name-to-structure function.

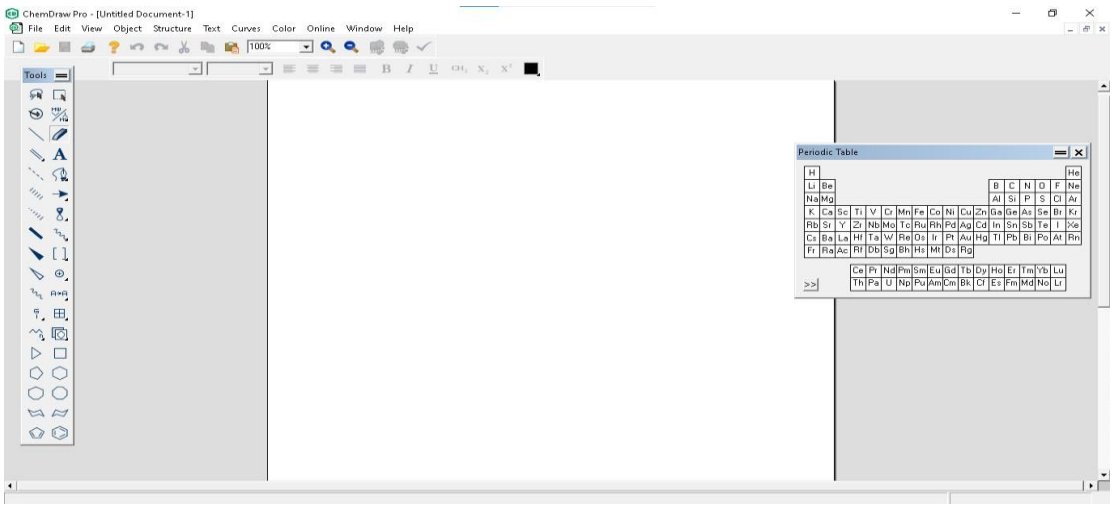


Figure III.1: The graphical interface of the ChemDraw Pro12.0 software.

- **Schrödinger**

The Schrödinger software suite contains a broad array of computational chemistry and molecular modeling tools that can be used to study the interaction of peptides with proteins. These include molecular docking using Glide and Piper, relative binding free energy predictions with FEP+, conformational searches using Macro Model and Desmond, and structural refinement using Prime and PrimeX. [3]

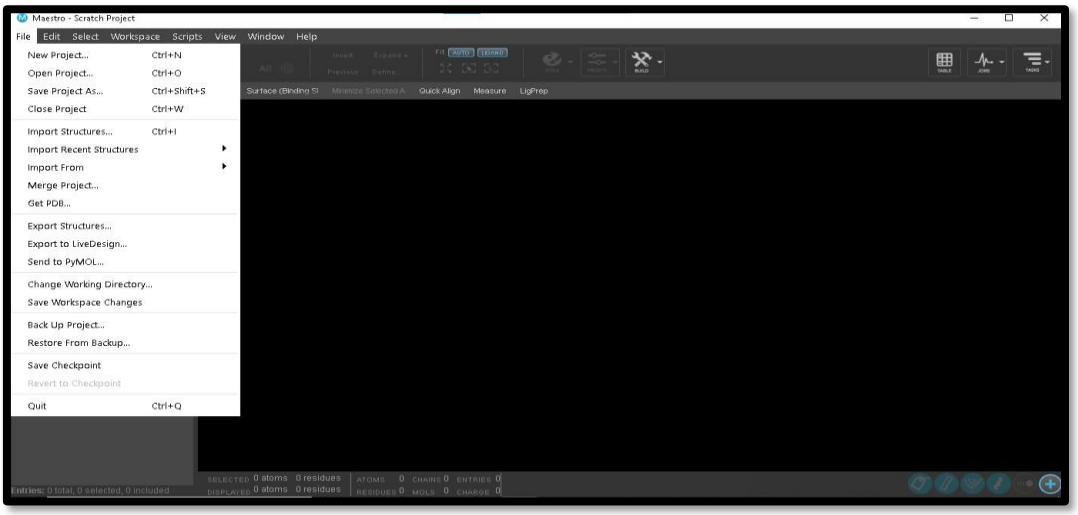


Figure III.2: The graphical interface of the Meastro 13.9 software.

- **Protein Data Bank (PDB)**

The Protein Data Bank (PDB) the single global repository of experimentally determined 3D structures of biological macromolecules and their complexes was established in 1971, becoming the first open-access digital resource in the biological sciences. The PDB archive currently houses ~130,000 entries (May 2017). It is managed by the Worldwide Protein Data Bank organization, [4] which includes the RCSB Protein Data Bank [5], the Protein Data Bank in Europe [6], and BioMagResBank , [7-8]and it's free online.

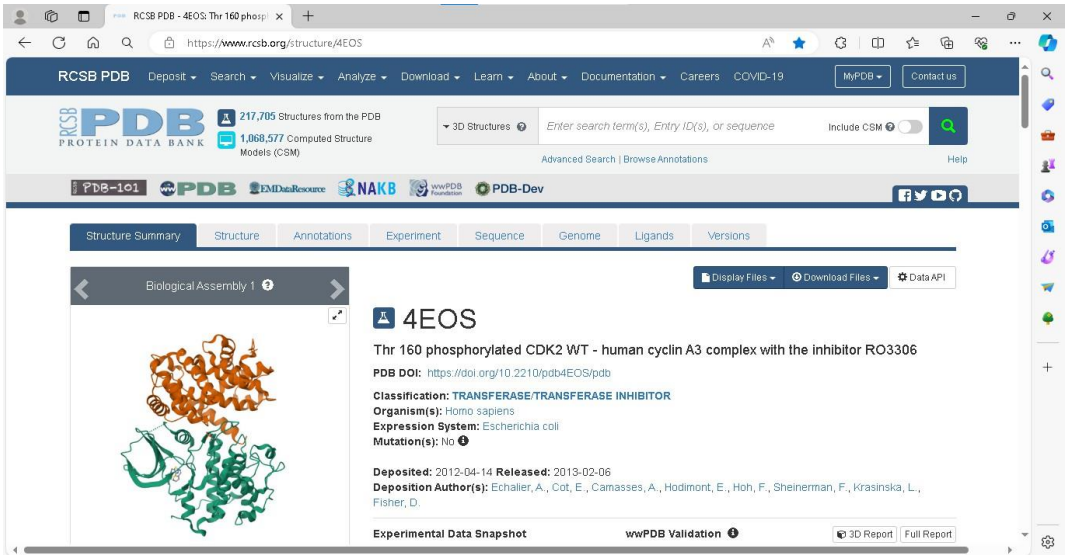


Figure III.3: The graphical interface of the PDB database website

- **pkCSM**

Drug development has a high attrition rate, with poor pharmacokinetic and safety properties a significant hurdle. Computational approaches may help minimize these risks. We have developed a novel approach (pkCSM) which uses graph-based signatures to develop predictive models of central ADMET properties (absorption, distribution, metabolism, and excretion, Toxicity) for drug development. pkCSM performs as well or better than current methods. A freely accessible web server [9], which retains no information submitted to it, provides an integrated platform to rapidly evaluate pharmacokinetic proprieties

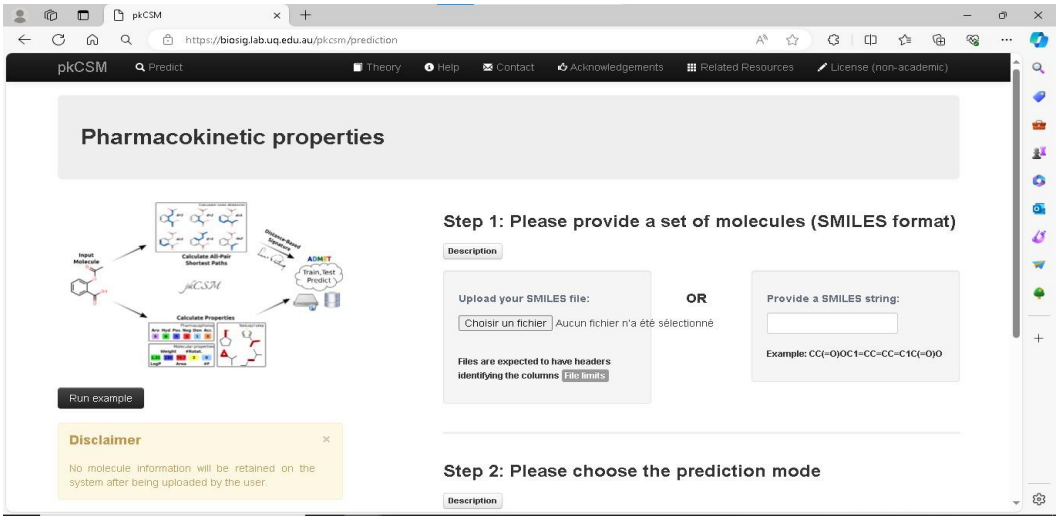


Figure III.4: The graphical interface of the pkCSM pharmacokinetic properties.

III.2.2 Methodology of calculations

III.3.2.1. Dataset

In this study, the dataset comprising 38 compounds of 6- substituted 2-arylaminopurine derivatives was selected from the literature along with their in vitro inhibitory activities (table III.1) [10] . Chem-Draw Professional 12.0 software was used to sketch all the dataset structures saved in ". mol" format. The maximal half inhibitory concentration (IC_{50}) in µM was converted to pIC_{50} (negative log of the IC_{50}).

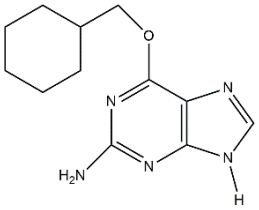
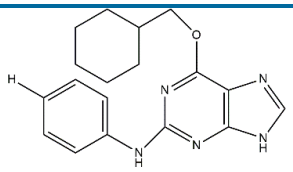
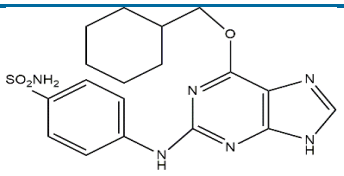
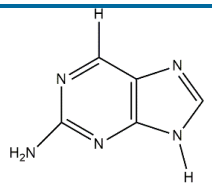
III.3.2.2. Ligand preparation

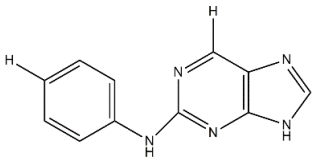
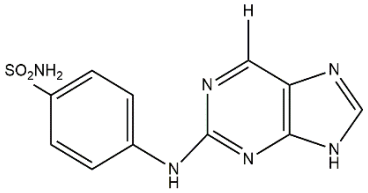
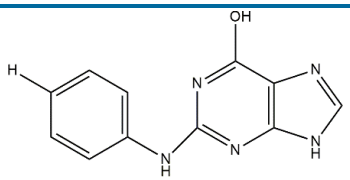
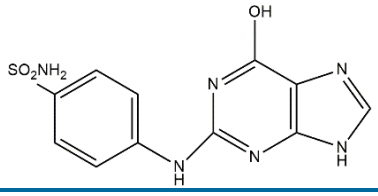
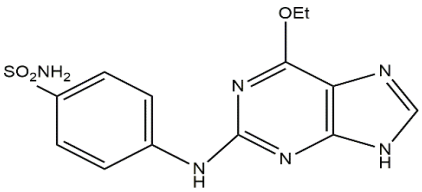
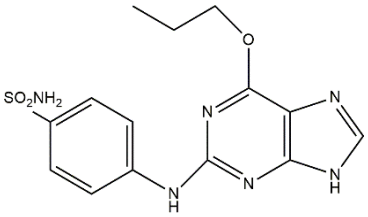
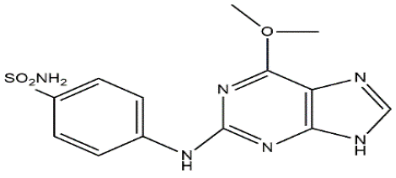
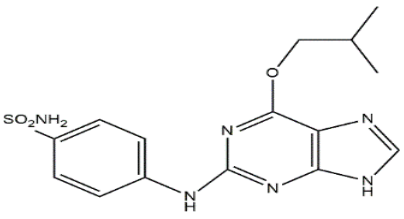
Ligand preparation was done using LigPrep (Schrödinger) from the application menu. The 2D structures of the ligands with bond directions were converted by LigPrep to the full 3-dimensional structure by assigning the force field as OPLS-2005. LigPrep can generate the expected ionized forms at significant concentrations corresponding to

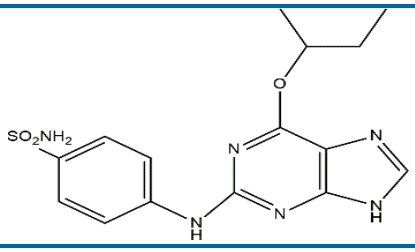
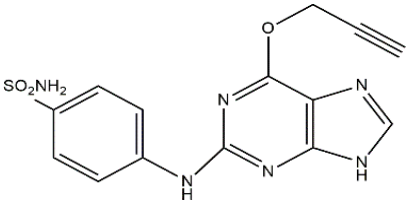
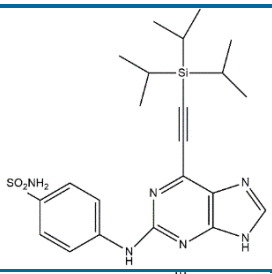
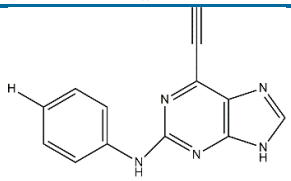
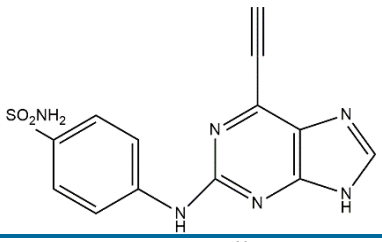
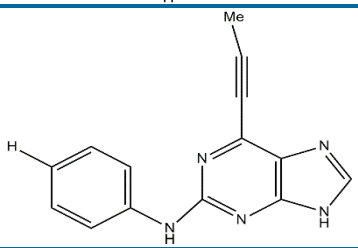
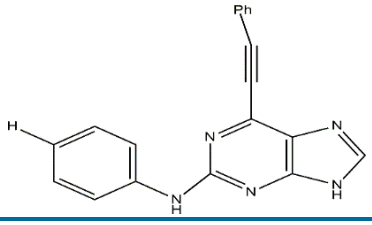
the pH 7.0±3.0 and generate the tautomers based on probability to recognize counter ions for removal and to add or remove hydrogens to achieve charge neutrality. It generates 32 stereochemical structures per ligand. Considering all the reasonable ionization states of ligands, the lowest energy conformer is important for docking studies.

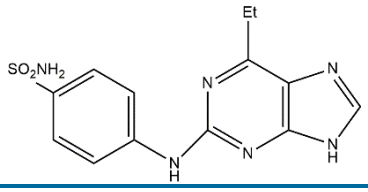
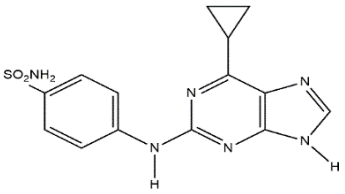
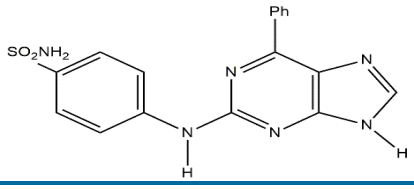
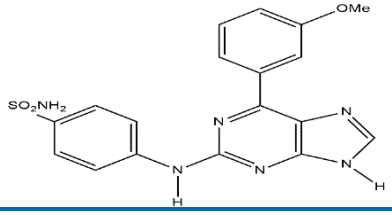
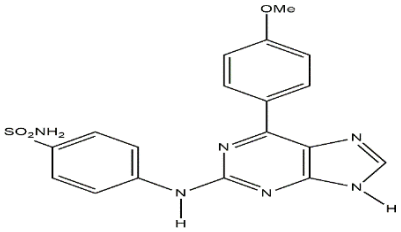
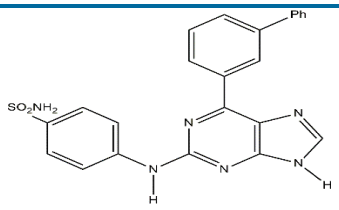
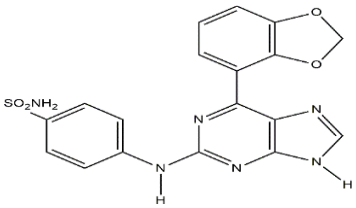
In our work we have used 38 ligands that are able to interact with the CDK2 receptor and subsequently form stable complexes. The chemical structure of the co-crystallized ligand (1RO) which is naturally complexed with the protein "4EOS" is given in the figure (figure III.5). [11]

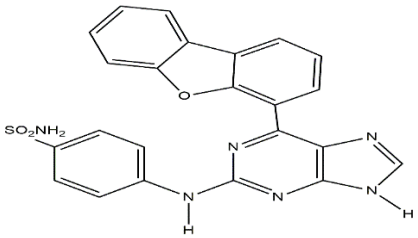
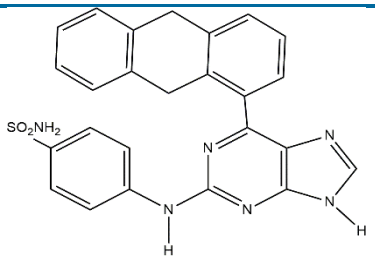
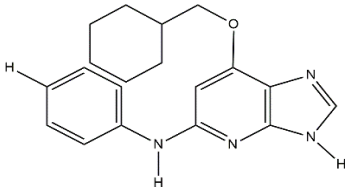
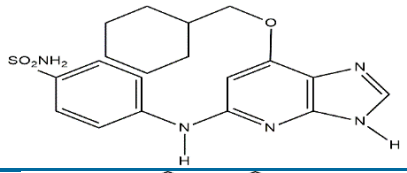
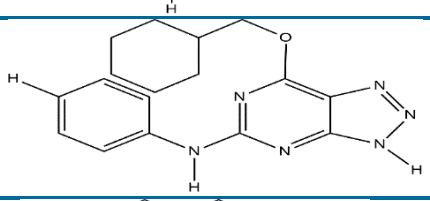
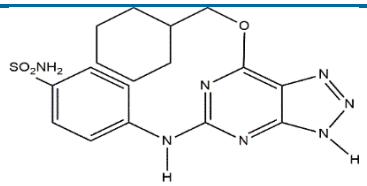
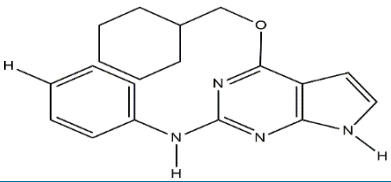
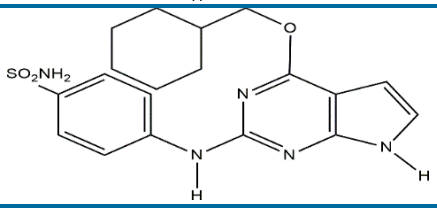
Table III.1: Structures of 6- substituted 2-arylaminopurine derivatives with their PIC50

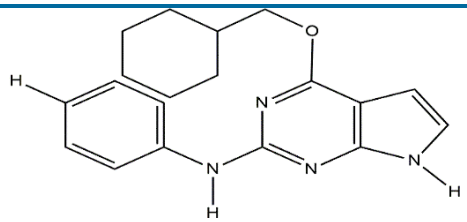
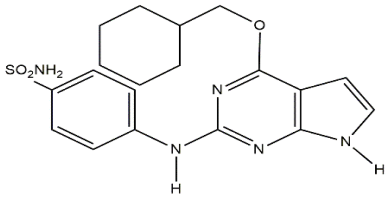
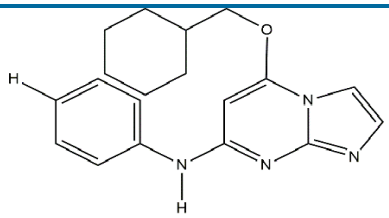
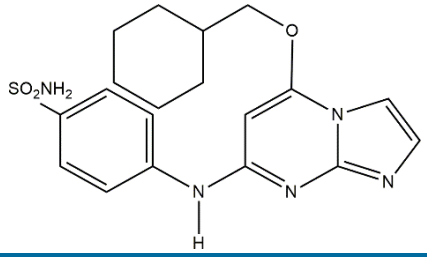
Compounds	Structure	pIC50
1		4.7696
2		6.0132
3		8.301
4		5.3979

5		1.65
6		5.8239
7		4.4685
8		4.284
9		7.585
10		8.0969
11		8
12		8.5229

13		9
14		7.7212
15		7.7696
16		4.6778
17		6.0809
18		5.1938
19		5.0506

20		6.6576
21		7.7212
22		7.6198
23		7.4089
24		7.284
25		7.3565
26		6.2366

27		5.8539
28		5.1739
29		4.5467
30		6.5376
31		4.7447
32		8.699
33		4.7959
34		7.7447

35		4.3872
36		7.5528
37		5
38		4.4437

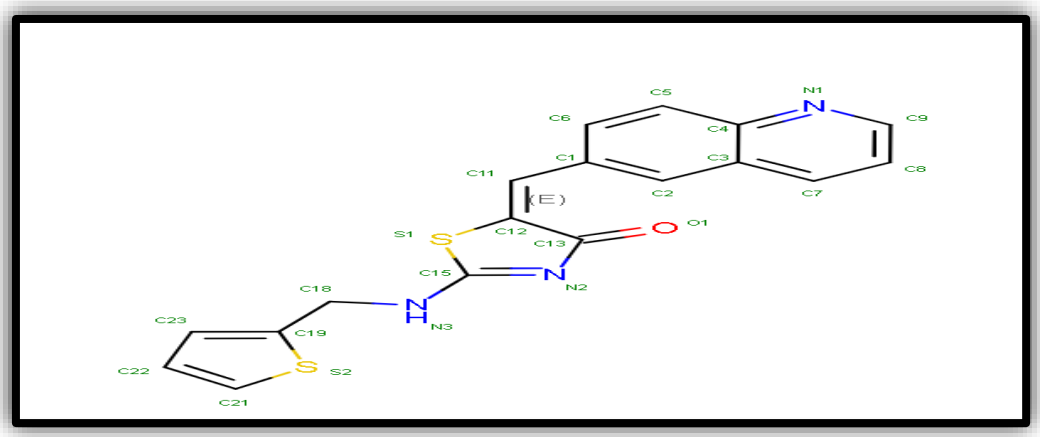


Figure III.5: Chemical structure of reference ligand (Lref) 1RO.

III.3.2.3. Protein preparation

The 3-dimensional (3D) structure of protein Cyclin dependent kinase was downloaded from the Protein Data Bank (PDB) [12] with the PDB ID(4EOS) , based on the parameters of resolution , and R value-free. Targeted three-dimensional structural coordinates are pre-processed using Protein Preparation Wizard module, available in

Schrödinger with a default set of parameters like assigning-bond orders, filling missing hydrogen's capping the termini, side chains & loops, and removing waters. Additionally, various states of tautomerization and protonation were projected in favor of ligand at a pH level of 7.0. Finally, the optimization of the protein H bonds to renovate the overlying hydrogen and minimization are performed in the presence of the force field OPLS-2005 with a RMSD value of 0.30 Å. [13-14]

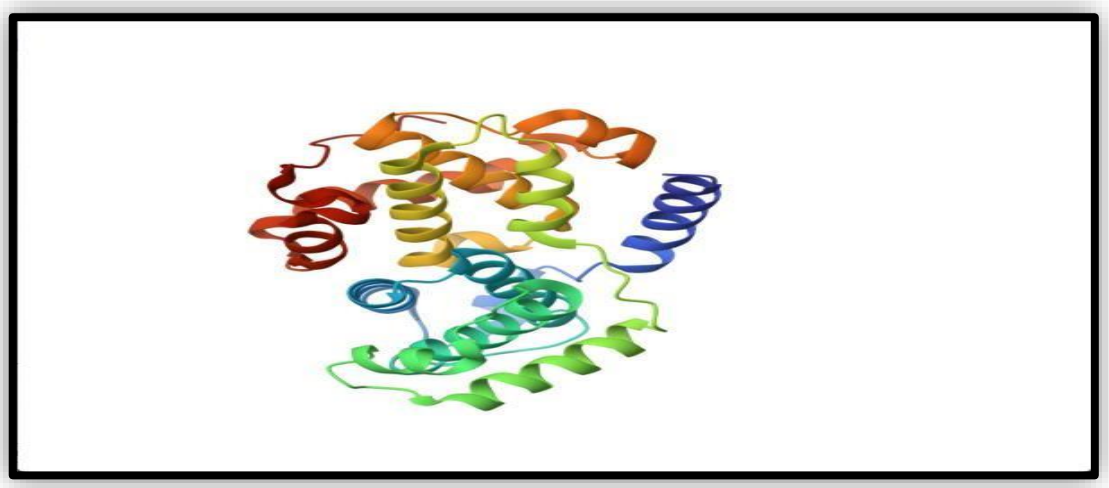


Figure III.6: **S**tructure of the Cyclin Dependent Kinases 2 (PDB ID 4EOS)

III.3.2.4. Generation of Receptor Grid

Cocrstallized structures downloaded from PDB had ligand was separated. The partial atomic charge cutoff was 0.25 and Van der Waals radii of receptor atoms were 1.0 Å by defaults. The centroid of the workspace ligand has an enclosing box to represent the activity of receptor. Center of the bound ligand in receptor was selected to generate a grid box by using the above protocol in the receptor grid generation module of Maestro. Furthermore, the receptor grid box was generated in each direction and the box was set at the center of the cognate ligands with allowance for the binding pocket to accommodate any ligand [15]. The dock main after the grid generation was 20Å for each dimension (x, y , z).

III.3.2.5. Molecular docking

Molecular docking studies were carried out using Glide docking protocol granting full flexibility to the ligands. The SP was performed on the output complexes

in order to reduce the initially collected 5 poses per compound to 3 A rescoring of the top-ranking SP pose of each compound was then performed with the XP scoring function of Glide and output poses were rescored by postprocess. The compounds which are docked most favorably were ranked based on XP Gscore.

III.3.2.6. In silico ADME prediction

The evaluation of drug likeness for all compounds was carried out using qikprop tool of Schrodinger drug discovery software [16]. Qik Prop provides ranges for comparing an exact molecule's properties with those of 95% of known drugs. QikProp also flags 30 types of reactive functional groups that may cause false positives in high throughput screening assays. It also evaluates the suitability of derivatives based on Lipinski's rule of five [16.17], which is essential to ensure drug-like pharmacokinetic profile while using rational drug design.

- **Gastrointestinal absorption(GI)**

It is the ability of a compound to cross the gastrointestinal barrier into the bloodstream. It is fundamentally characterized by one of three mechanisms, including facilitated dissemination, passive disseminations and active transportation.

- **HIA (humain intestinal absorption)**

The intestine is the main absorption site for medicines administered orally. So the forecast of this parameter reflects the prediction of the proportion of the drug absorbed by the small intestine. It is an element that determines the effectiveness of the drug by influencing the bioavailability.

- **P-gp inhibiteur**

Membrane protein Glycoprotein P belongs to the superfamily of ATP binding tape carriers. It is sometimes referred to as MDR1 or 2 ABCB1. ABC). Given that it can detect various drugs with distinct structural properties, many of which are also CYP3A4 substrates, this is likely the most promiscuous efflux carrier.

- **P-gp substrate**

Medium transport modulation by P glycoprotein has important pharmacokinetic implications for p-gp substrate that can be exploited for specific therapeutic benefits or lead to contraindications.

- **Penetration BBB (blood-brain barrier)**

It is one of the most selective barriers that separates circulating blood from the central nervous system. Medicines that act in the CNS must cross the hemato-encephalic barrier (HEB) to reach their biological target. The advantage of predicting BBB permeable drugs is to identify brain toxic substances and know drugs with a target at the central level.

- **FU (one jump fraction or unbound fraction)**

In plasma, drugs exist in balance between a plasma protein-bound and unbound form. The effectiveness of a drug can be affected by the degree to which it is bound to the plasma proteins, because the more it is tied, the less it passes through the cell membranes and less it reaches the biological target.

- **Cytochrome P450 inhibitors**

P450 is a detoxifying enzyme in the body, it can inhibit several drugs as it can activate others. Cytochrome P450 plays a role in the oxidation of drugs to facilitate their excretion. Knowledge of drugs inhibited by these isoenzymes is important. These are divided into two categories: Phase I isoenzymes (oxidative reactions) and Phase II (conjugation reactions). The human cytochrome P450 family (phase I enzymes) contains 57 isoenzymes, the main of which are: 1A2, 3A4, 2C9, 2C19 and 2D6. The main seat of cytochromes responsible for phase I reactions is the liver [17,18].

III.4. Results and discussions:

III.4.1. Molecular Docking

The docking studies of the ligands to protein active sites were performed by an advanced molecular docking program Schrodinger Maestro-13.9 version for determining the binding affinities of the compounds. It is based on a type of semi-flexible docking and is generally used for ligand protein bonding; the ligand being considered flexible and the main enzyme chain has been kept rigid. Once the ligand-receptor complex is formed, it will adapt to the most stable conformation, i.e. to the lowest energy level. Results in the following table:

Table III.2: Docking results of ligands and reference ligand with CDK2 proteins

Row	Docking score	Glide GScore	Glide Energy
01	-7.014	-7.242	-31.039
02	-7.925	-8.204	-47.052
03	-7.431	-7.891	-34.354
04	-7.967	-8.58.2	-36.758
05	-8.184	-8.689	-42.097
06	-7.387	-8.049	-35.804
07	-5.388	-5.571	-41.864
08	-7.615	-8.451	-43.381
09	-4.909	-5.179	-44.178
10	-5.234	-5.575	-48.960
11	-8.657	-8.930	-46.152
12	-5.376	-5.605	-45.239
13	-5.712	-5.943	-44.835
14	-5.062	-5.569	-39.156
15	-2.699	-5.132	-38.636
16	-5.039	-5.163	-40.119
17	-5.115	-5.956	-39.368
18	-3.416	-5.678	-43.460
19	-4.947	-6.038	-44.888
20	-7.507	-7.810	-41.892
21	-5.867	-6.260	-54.550
22	-7.177	-7.572	-43.720
23	-8.211	-8.607	-55.977
24	-8.195	-8.632	-52.113
25	-5.872	-6.428	-54.352

26	-7.444	-8.035	-61.797
27	-4.837	-5.274	-57.704
28	-7.589	-7.820	-59.596
29	-8.599	-8.767	-50.319
30	-6.176	-6.177	-49.173
31	-5.895	-5.89	-50.592
32	-7.686	-8.536	-43.653
33	-3.872	-4.796	-47.850
34	-5.814	-5.816	-50.362
35	-8.244	-8.260	-36.646
36	-8.137	-8.137	-42.275
37	-4.078	-6.225	-53.707
38	-4.133	-6.279	-50.177
L_{ref}	-7.511	-7.569	-45.061

The table III.2 shows the results of molecular ligand docked with CDK2 , according to the results obtained we've choose six ligands :L_5,L_{11},L_{32},L_{23},L_{24},L_{29} because of their lowest docking score energy, which shows these complexes are more stable, and also can be classified in the following order:

L_{ref} <L_{32}<L_5<L_{24}<L_{23}<L_{29}<L_{11}.

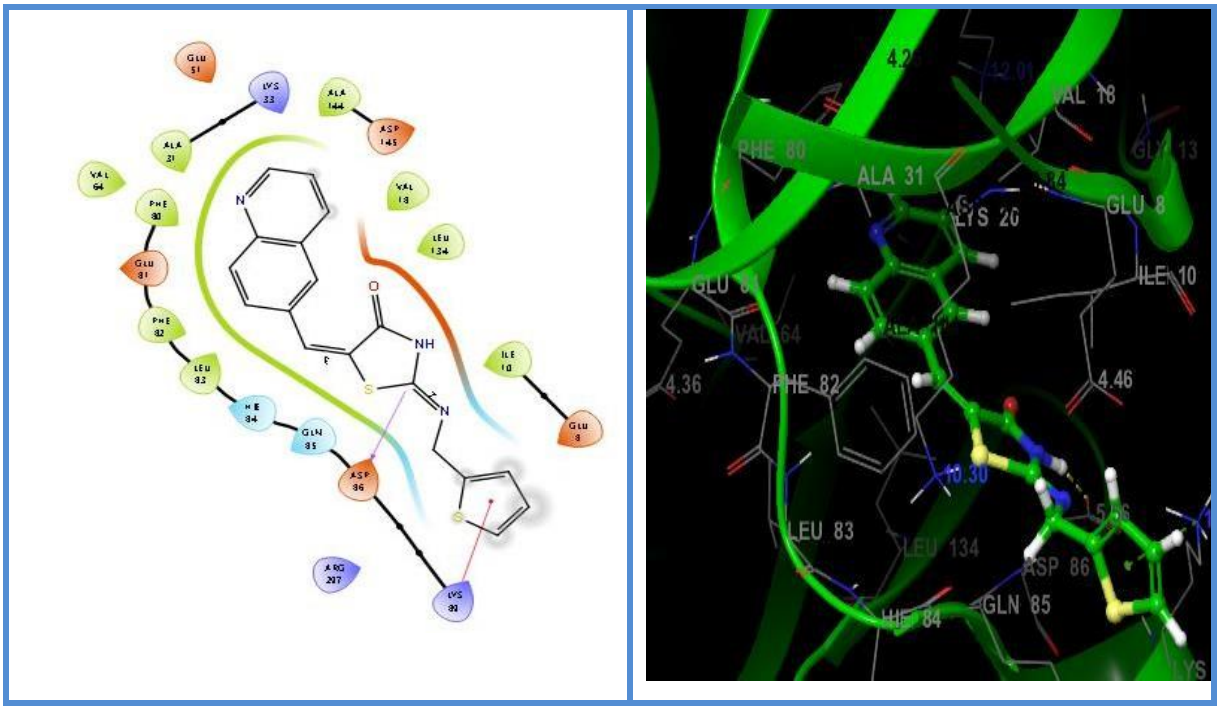
Table III.3: Docking results of ligands $L_{32,}$ $L_{5,}$$L_{24}$$L_{23,}$$L_{29}$,$L_{11}$ and L_{ref}

Compound	Glide GScore	Glide Energy	Hydrogen bond interaction	Hydrophobic interaction	Pi-Pi Stackin,
L_{11}	-8.930	-46.152	ASP86 ASP86	ALA31,ILE10,LEU134,VAL18,VAL64,LEU83,PHE82,PHE80ALA144,	PHE82
L_{29}	-8.767	-46.763		ALA31,ILE10,LEU134,VAL18,VAL64,LEU83,PHE82,PHE80ALA144	PHE82
L_5	-8.689	-42.097	GLU51,HIS84, ASP 145	Leu83,Leu134,PHE83 ,ALA31,PHE146,ALA144 ,VAL64, ALA10	PHE82
L_{24}	-8.632	-52.133	LYS33,ASP 145,GLU 12	LEU83,PHE82,PHE80,VAL64,LEU55,ALA144,ALA31 ,PHE144,VAL18	
L_{23}	-8.607	-55.977	GLU51, ILE10	:LEU134,LEU83,PHE82,PHE80, VAL64,ALA31,PHE146,LEU55, ALA144, VAL18	
L_{32}	-8.536	-43.653	Leu83	ALA31,ALA144,VAL64,VAL18,ILE10,LEU134	PHE80 PHE82
L_{REF}	-7.569	-45.061	ASP 86	LEU83,PHE82,PHE80,VAL64,ALA31,ILE10,LEU134 ,VAL18,ALA144	LYS 89

- **Reference ligand**

The objective of the study of reference protein-ligand interactions is to compare with other ligands, in order to choose the best ligand that has a good affinity with the CDK2 receptor. The results of molecular docking (Figure III.6, Table III.3) show that the reference ligand interacts with the active site CDK2. The glide score and glide energy values for reference ligand were -7.569 Kcal/mol and -45.061 kcal/mol, respectively.

The reference ligand (1RO) formed hydrogen bond with the residue ASP 86 of protein CDK2, one interaction pi-cation type with LYS 89 residues , and Hydrophobic interactions with the residues: LEU83,PHE82,PHE80,VAL64,ALA31,ILE10 ,LEU134 ,VAL18,ALA144 in the active site of CDK2

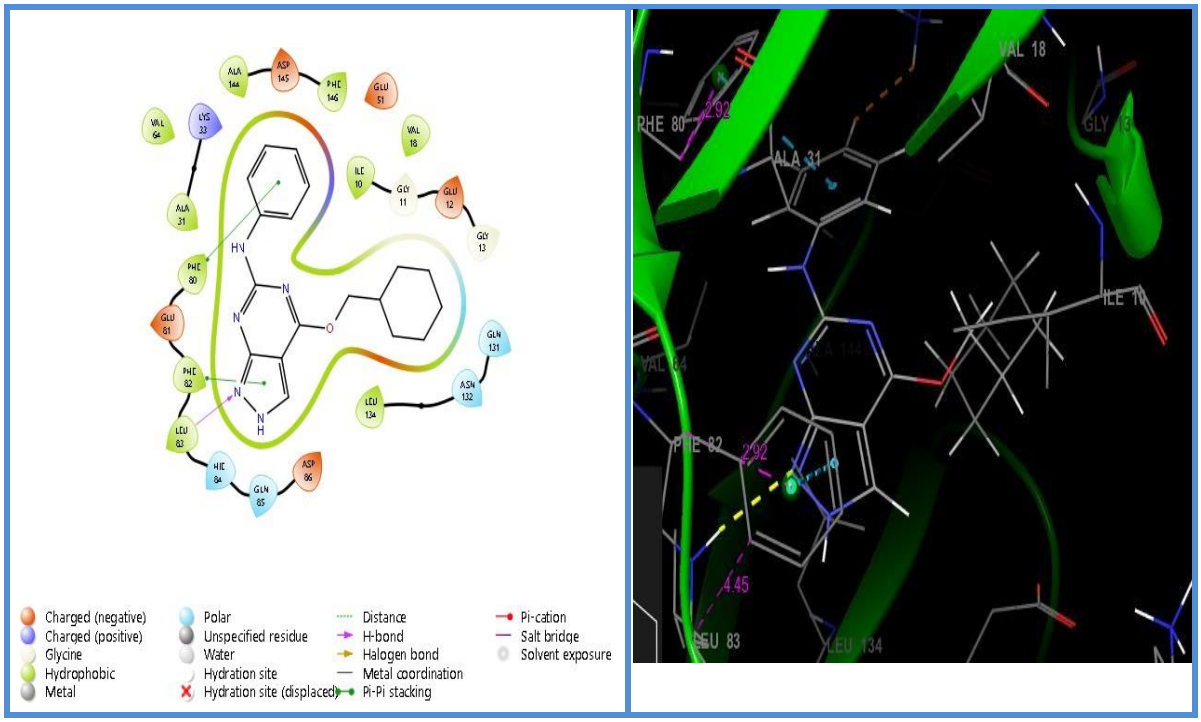


FigureIII.7: 2D, 3D interactions of reference ligand with active site of CDK2 (ID:4EOS)

- **Ligand 32**

The results of molecular docking (Figure III.8, Table III.3) show that ligand 32 interaction with the active site CDK2. The glide score and glide energy values for ligand were --8.536Kcal/mol and -43.653 kcal/mol, respectively.

Figure III.8 has revealed the presence of four interactions in the complex(ligand32-CDK2) active site. Two of them are Pi-Pi stacking type between the ligand 32 and both PHE80, PHE82 residues of the CDK2 active site. The existences of π-π stacking interactions play a crucial role in determining the structure and stability of the complex (Ligand 32-CDK2 residues). The third type of interaction is H-bond between Leu83 residue of the protein and the ligand 32.

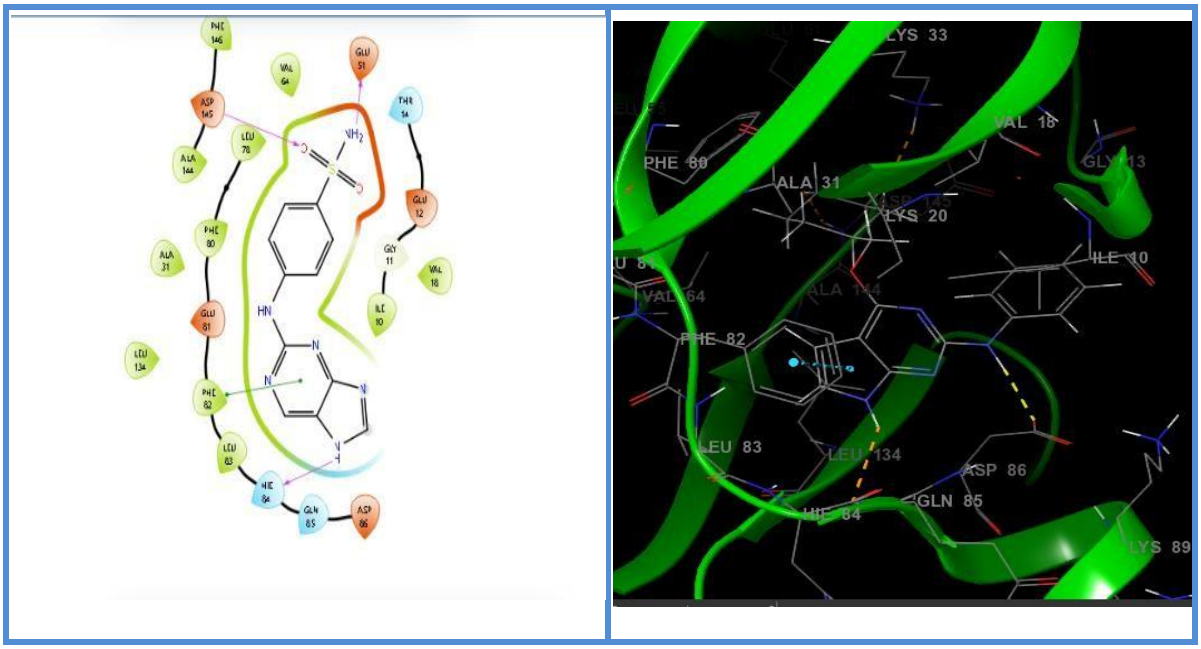


FigureIII.8: 2D, 3D interactions of ligand 32 with active site of CDK2 (ID:4EOS)

- **Ligand 5**

The CDK2-ligand 5 complex is characterized by a score energy of-8.689 kcal/mol with energy score -42.097 kcal/mol (Figure III.9, Table III.2).

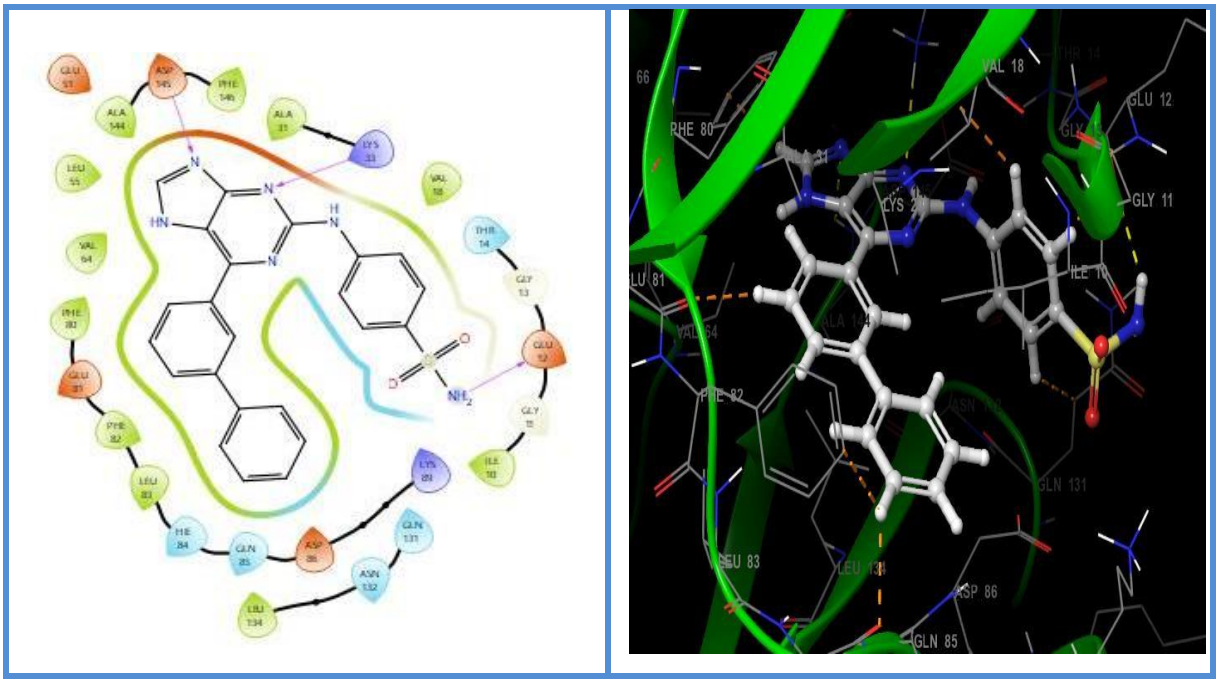
The visual analysis of ligand 5 and the active site residues of CDK2 indicates the presence of three hydrogen bonds with the residues GLU51, HIS 84 and ASP 145 , also the ligand 5 has Pi-Pi stacking bond with the residue PHE82, and hydrophobic interactions with the residues : Leu83,Leu134,PHE83,ALA31 ,PHE146,ALA144, VAL64, ALA10 in the active site of CDK2



FigureIII.9: 2D, 3D interactions of ligand 5 with active site of CDK2 (ID:4EOS)

- **Ligand 24**

The construction of the CDK2-Ligand 24 complex by molecular docking results a glide score of-8.632 kcal/mol with energy score -52.133 kcal/mol (Figure III.10, Table III.3).Visual analysis of ligand 24 and active site residues CDK2 indicates the presence three hydrogen bonds with residues LYS33, ASP 145, and GLU 12, and hydrophobic interactions with residues LEU83, PHE82, PHE80, VAL64 ,LEU55,ALA144,ALA31 ,PHE144,VAL18 of CDK2 active site.

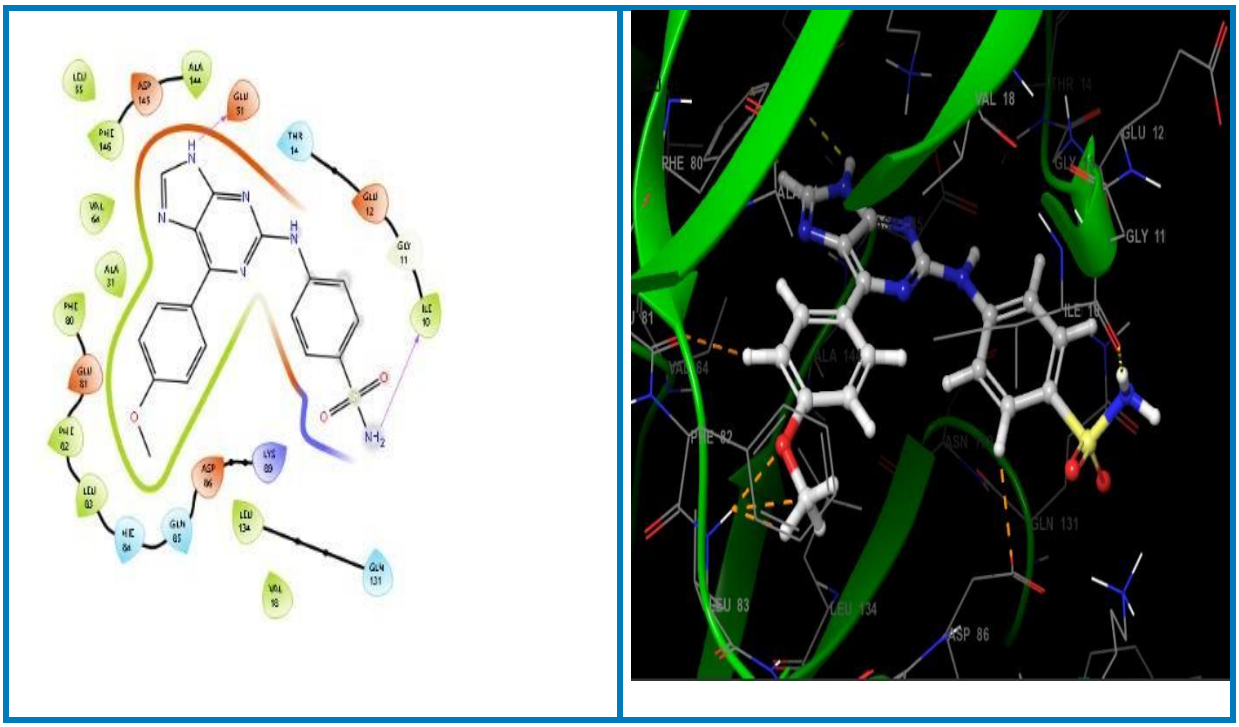


FigureIII.10: 2D, 3D interactions of ligand 24 with active site of CDK2 (ID:4EOS)

- **Ligand 23**

The CDK2-ligand 23 complex is characterized by a glide score -8.607 kcal/mol with energy score -55.977 kcal/mol (Figure III.11, Table III.3

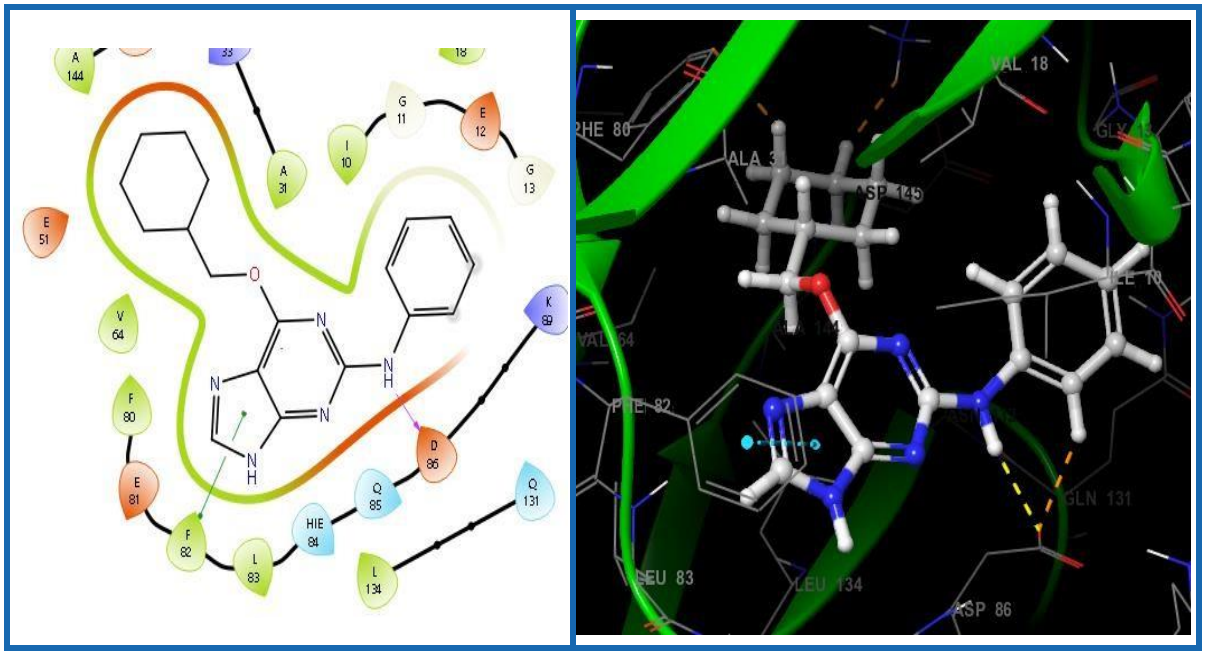
The visual analysis shows the presence of the following interactions: The ligand 23 formed two hydrogen bonds with residues GLU51, ILE10 ,and hydrophobic interactions with residues ; LEU134,LEU83,PHE82,PHE80, VAL64,ALA31,PHE146,LEU55,ALA144, VAL18 in the active site of CDK2



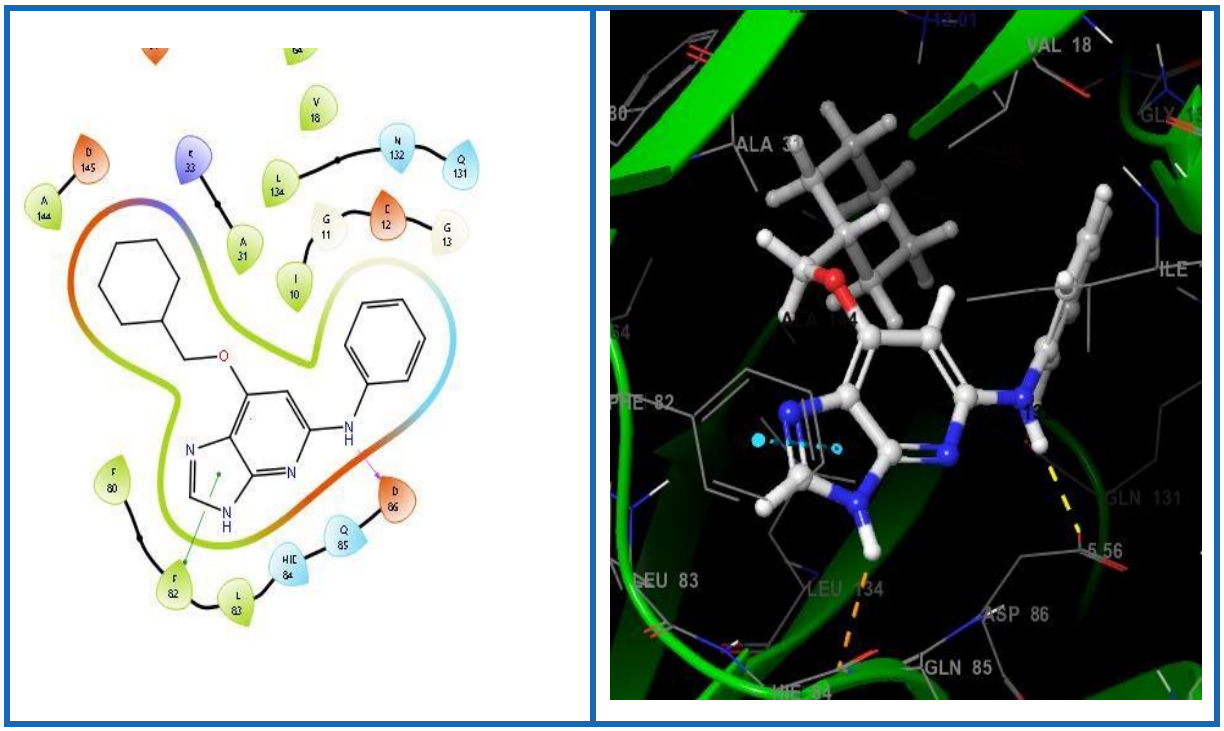
FigureIII.11: 2D, 3D interactions of ligand 24 with active site of CDK2 (ID:4EOS)

- **Ligand 29 and Ligand 11**

The molecular docking results shows the both Ligand 11 and 29 have the lowest docking score energy which refers to their good affinity. have more Glide scores due to more lipophilic character. The energy evidence and hydrogen bonding interaction of compound 29 and compound 11 are shown in Figures 12 , 13, and table III.3.



FigureIII.12: 2D, 3D interactions of ligand 11 with active site of CDK2 (ID:4EOS)



FigureIII.13: 2D, 3D interactions of ligand 11 with active site of CDK2 (ID:4EOS)

III.4.2. Drug likeness and ADME prediction

Chemicals' drug-likeness is a qualitative feature [18], beneficial for early-stage drug development. From the standpoint of this concept, it would be ideal to encode the equilibrium between a compound's molecular characteristics that effects its pharmacokinetics and eventually optimizes their absorption, distribution, metabolism and excretion (ADME) in the human body like a medicine., the quickest strategy for appreciating the drug-likeness of a set is to apply “rules”, they have been applied. As first, the most commonly Lipinski's rules are used [19,20].

III.4.2.1 Drug likeness

Continuous, our parameters determined that good absorption or permeation is more likely to occur when: the molecular weight (MW<500da), number of hydrogen bond donors (HBDs <5), to estimate hydrophobicity of molecules used the partition coefficient octanol/water (Log p< 5), and the number of hydrogen bond acceptors (HBA<10). In this rule of five-score, there are a total of four violations of Lipinski's rules. Veber identified the other two descriptors.[21]: number of rotatable bonds (NBR<10) and topological polar surface area (PSA <140 Å2). The TPSA is an important measure for predicting molecular transport properties, especially in the areas of blood-brain barrier (BBB) penetration and intestinal absorption [22,23]. It is well known that molecules with a TPSA of 140 $Å^2$ have a great ability to penetrate in an environment that is hydrophobic, like biological membranes. However, this could explain their quick penetration in hydrophilic settings, for instance, the core of transporter proteins [24].

TableIII.4: Drug-likeness parameters of top ligands and reference ligand

	Lipinski's rules					**Veber's rules**		
Ligands	MW< 500DA	Log p ≤ 5	HBD ≤ 5	HBA ≤10	Lipinski score	nRotB≤10	TPSA ≤ 140 Å2	Veber score
L_5	290.299	3.695	3	6	4	5	67.490	2
L_{11}	323.397	-0.467	2	4	4	3	125.713	2
L_{23}	396.423	-0.245	3	7	4	5	123.820	2
L_{24}	442.494	1	4	8	4	5	69.019	2
L_{29}	322.409	3.599	2	4	4	5	122.320	2
L_{32}	323.397	3.690	2	3	4	4	121.797	2
L_{ref}	351.456	4.0564	1	3	4	3	146.889	1

MW: molecular weigth. HBA: Num. H-bond acceptors.
HBD: Num. H-bond donors. NROT: Num. rotatable bonds.
LogP : Log Po/w (XLOGP3) . TPSA : Topological polar surface.

The Lipinski and Veber criteria are satisfied by all ligands, indicating that their theoretical results are optimal(table III.4).The association between appropriate aqueous solubility and intestinal permeability, as well as these physicochemical molecular properties that represent the first steps in oral bioavailability, indicating that all chemical ligands are able to exhibit biological activity without having problems with oral absorption. Unlike L_{ref} ,The results show that the L_{ref} doesn't fit perfectly within the margin of the criteria imposed by the veber rule, which means it may have problems with oral absorption, and all ligands with logP values less than 5 So they have a good

solubility in water, better gastric tolerance and effective elimination by the kidneys and good permeability through the cell membrane.

III.4.2.2 ADME prediction

The pharmacokinetic properties of the top ligands (L_5,L_{11},L_{23},L_{24},L_{29},L_{32})and the co-crystallized compound were estimated with the aid of pkCSM. Also,. The outcomes are shown in the tableIII.5 , which presents the ADME predictions.

TableIII.05: Absorption and distribution of top ligands and reference ligand

	Absorption and Distribution						
Ligands	Water solubility Log S (logmol/L)	Caco2 permeability (log Papp in 10-6 cm/s)	Intestinal absorption (human) (%)	P-glycoprotein substrate	VDss (human) (log L/kg)	BBB permeability (log BB)	CNS permeability (log PS)
L_5	-2.898	-0.359	74.38	YES	0.413	-1.807	-3.861
L_{11}	-2.86	1.343	100	YES	-0.04	1.359	-2.003
L_{23}	-2.892	0.02	77.78	YES	-0.287	-1.807	-3.861
L_{24}	-2.92	0.476	75.85	YES	-0.233	-1.664	-3.572
L29	-2.90	1.343	100	YES	-0.031	1.339	-2.506
L_{32}	-3.309	1.073	100	YES	0.846	1.193	-2.131
L_{ref}	-4.485	1.399	88.126	NO	0.043	-0.253	-1.79

The absorption and distribution were predicted from many properties. For the absorption, the Log S parameter was used to evaluate water solubility of ligands (L5,L11,L23,L24,L29,L32) and of the reference ligand (tableIII.5). We obtained values of(-3.309 to -2.86 log mol/L) these values indicate that all ligands are soluble in aqueous solutions due to their long R group's chain, unlike the reference ligand ,which it has log S value equals -4.485 that indicate reference ligand is Moderately soluble in aqueous solutions. chain. For the Caco2 permeability, a high Caco2 permeability result would translate in predicted values >0.9. (table III.13) shows that all ligands have a

high Caco2 permeability results starting by L_{11},L_{32},L_{29} with Caco2 permeability results 1.343,1.073,1.343 log Papp in 10-6cm/s respectively. We can note that these ligands establish a good Caco2 permeability result that are clise to than L_{ref} that has Caco2 permeability result equals 1.399 log Papp in 10-6cm/s.

For the intestinal absorption (%absorbed) (HIA) results were shown are between (74.38-100), we found that L_{11}, L_{29}, L_{32} (100) are slightly better absorbed than reference ligand (88.126).we should know that for a given compound it predicts the percentage that will be absorbed through the human intestine molecule with an absorbance off less than30% is considered to be poorly absorbed.

P-glycoprotein (P-gp) is a membrane carrier involved in the kinetics of many xenobiotics. This protein that allows the flow of xenobiotics is present within the different organs responsible for the absorption and excretion of drugs. all compounds except Lref are P-gp substrate

For the evaluation of the distribution of the compounds, we used the following factors: distribution's volume for human (log L/kg), factor of blood-brain barrier permeability (log BB), and CNS permeability (log PS). For a given compound log BB the >0.3 readily cross the blood-brain-barrier, log BB<-1 are poorly distributed. Can be seen that Ligands: (L_{11},L_{29},L_{32}) establish log BB values greater than 0.3,that means theses ligands are readily cross the blood-brain-barrier gives a positive response (a toxic effect on the CNS). Unlike the other ligands that shows results lower than -1 even better than reference ligand value.

According to L_{32} analyses, this compound has a log VDss value of 0.846, whereas the reference ligand has a smaller log VDss value of 0.043. Thus, the reference ligand is less advantageous than L_{32} in terms of what constitutes an acceptable VDss value according to Refs.Even L_5 has an acceptable result of VDss equals 0.413. we should know VDss considered low when logVDss<-0.15, and high when logVDss >0.45. Whereas the CNS permeability values of L_{11},L_{29},L_{32},L_{ref} close to each other (tableIII.5). The BBB is an intricate barrier that isolates the peripheral tissue from the central nervous system (CNS). The BBB regulates the movement of substances, cells, and nutrients from the blood to the brain and from the brain to the blood in order to

keep the CNS in a state of homeostasis. Additionally, it helps removing cellular metabolites and toxins from the brain and transports them to the blood.

TableIII.06: Metabolism and excretion of top ligands and reference ligand

Metabolism and excretion								
Ligands	CYP2D6 substrate	CYP3A4 substrate	CYP1A2 inhibitior	CYP2C19 inhibitior	CYP2C9 inhibitior	CYP2D6 inhibitior	CYP3A4 inhibitior	Total Clearance
L_5	No	No	Yes	No	No	No	No	0.709
L_{11}	No	No	Yes	Yes	Yes	No	No	0.937
L_{23}	No	No	No	No	Yes	No	No	0.655
L_{24}	No	No	No	Yes	Yes	No	Yes	0.306
L_{29}	No	Yes	Yes	Yes	Yes	No	No	0.963
L_{32}	No	Yes	Yes	Yes	Yes	No	No	-0.202
L_{ref}	No	Yes	Yes	Yes	Yes	No	No	-0.145

The tableIII.6 shows the results of an ADME study, specifically focused on metabolism and excretion of ligands L_5,L_{11},L_{23},L_{24},L_{29},L_{32},L_{ref}.

We noted that the ligands(L_{11},L_{23},L_{29},L_{32}) there is an inhibitory effect on CYP2C19, so we have a change in the activity of this isoform but on the other ligands(L_5,L_{24}andL_{ref}) ther's no modification noted.

We also noted that the CYP3A4 isoform is not inhibited by ligands(L_{11},L_{23},L_{32},L_{ref},),which is not with ligands(L_5,L_{24},L_{29}) which have an inhibitory potential on this isoform.

We have also shown positive CYP2C9 inhibition results for: (L_{24},L_{ref})but no inhibitions or modifications recorded for the rest of the ligands.

The last isoform of CYP2D6 can be inhibited by (L_5,L_{11},L_{23},L_{24},L_{29},L_{32},L_{ref}) Ligands. and its structure is modified by these ligands.

References

[1] Li, J. J. (Ed.). (2013). Heterocyclic chemistry in drug discovery. John Wiley & Sons.

[2] Heravi, Y. E., Sereshti, H., Saboury, A. A., Ghasemi, J., Amirmostofian, M., & Supuran, C. T. (2017). 3D QSAR studies, pharmacophore modeling, and virtual screening of diarylpyrazole-benzenesulfonamide derivatives as a template to obtain new inhibitors, using human carbonic anhydrase II as a model protein. Journal of Enzyme Inhibition and Medicinal Chemistry., 32(1), 688–700.

[3] www.PDB; wwpdb.org

[4] RCSB PDB; rcsb.org

[5] PDBe; pdbe.org

[6] BMRB; www.bmrb.wisc.edu

[7] http://www.pdb.org/

[8 http://structure.bioc.cam.ac.uk/pkcsm

[9] https://www.rcsb.org/

[10] Keretsu, S., Balasubramanian, P. K., Bhujbal, S. P., & Cho, S. J. (2017). Receptor-guided 3D-Quantitative Structure–Activity Relationship and Docking Studies of 6-Substituted 2-Arylaminopurines as CDK2 Kinase Inhibitors. Bulletin of the Korean Chemical Society, 38(11), 1275-1284.

[11] Selvaraj, C., Dinesh, D. C., Panwar, U., Abhirami, R., Boura, E., & Singh, S. K. (2021). Structure-based virtual screening and molecular dynamics simulation of SARS-CoV-2 Guanine-N7 methyltransferase (nsp14) for identifying antiviral inhibitors against COVID-19. Journal of Biomolecular Structure and Dynamics, 39(13), 4582-4593.

[12] Ganguly, S., & Debnath, B. (2014). Molecular docking studies and ADME prediction of novel isatin analogs with potent anti-EGFR activity. Med chem, 4(8), 558-68.

[13] Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced drug delivery reviews, 64, 4-17..

[14] Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2012). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced drug delivery reviews, 64, 4-17.

[15] Khamouli, S., Belaidi, S., Ouassaf, M., Lanez, T., Belaaouad, S., & Chtita, S. (2022). Multi-combined 3D-QSAR, docking molecular and ADMET prediction of 5-azaindazole derivatives as LRRK2 tyrosine kinase inhibitors. Journal of Biomolecular Structure and Dynamics, 40(3), 1285-1298.

[16] Pires, D. E., Blundell, T. L., & Ascher, D. B. (2015). pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. Journal of medicinal chemistry, 58(9), 4066-4072.

[17] Tripathi, S. K., Selvaraj, C., Singh, S. K., & Reddy, K. K. (2012). Molecular docking, QPLD, and ADME prediction studies on HIV-1 integrase leads. Medicinal Chemistry Research, 21, 4239-4251.

[18] Schaftenaar, G., & de Vlieg, J. (2012). Quantum mechanical polar surface area. Journal of computer-aided molecular design, 26, 311-318.

[19] Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. Journal of medicinal chemistry, 45(12), 2615-2623..

[20] Hopkins, A. L., Keserü, G. M., Leeson, P. D., Rees, D. C., & Reynolds, C. H. (2014). The role of ligand efficiency metrics in drug discovery. Nature reviews Drug discovery, 13(2), 105-121.. https://doi.org/10.1038/nrd4163

[21] Leeson PD, Springthorpe B (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. Nat Rev Drug Discov 6: 881-890.

[22] Ouassaf, M., Belaidi, S., Lotfy, K., Daoud, I., & Belaidi, H. (2018). Molecular docking studies and ADMET properties of new 1.2. 3 triazole derivatives for anti-breast cancer activity. Journal of Bionanoscience, 12(1), 26-36.

[23] Prasanna, S., & Doerksen, R. J. (2009). Topological polar surface area: a useful descriptor in 2D-QSAR. Current medicinal chemistry, 16(1), 21-41.

[24] Zhao, Y. H., Abraham, M. H., Le, J., Hersey, A., Luscombe, C. N., Beck, G., ... & Cooper, I. (2002). Rate-limited steps of human oral absorption and QSAR studies. Pharmaceutical research, 19, 1446-1457.
https://doi.org/10.1023/A:1020444330011

Conclusion

Conclusion

This manuscript's work has highlighted the various virtual screening approaches for evaluating biological activity and structural characteristic of 6-Substituted 2-Arylaminopurines derivatives with the aim of developing and discovering new bioactive molecules that are more effective against the cancer. In this memo, we discussed a number of important studies.

In the first part, we focused our attention to study and visualize the predominant interactions and affinity energies of 38 compounds of the purines derivatives to the active site of the CDK2 enzyme (PDB ID: 4EOS), by applying molecular docking investigations. These 38 selected molecules were docked into the active site of CDK2 (PDB ID: 4EOS) in order to assess and gain an in-depth insight into their ability to bind into the active site of the target enzyme. The docking analysis allowed us to evaluate their score and to highlight the ligands: (L5, L11, L23,L24,L29 and L32) associated with the active site as potential and powerful inhibitors of the 4EOS this justifies by the presence of different types of interactions, in addition that these ligands have lower energy levels compared to the reference ligand and also other complexes. Through the Maestro 13.9 program.

Drug likeness properties were analyzed for L5, L11, L23, L24, L29, L32 as well as the reference ligand in the second part. The computational results show that all these ligands are in agreement with the drug likeness rules (Lipinski and Veber’s rules), suggesting that these compounds theoretically have high ADME properties and will not have problems with oral absorption.

Finally, an in silico ADME study was performed for the selected more active ligands and the reference ligand to assess their pharmacokinetic properties. Based on the analysis of the results, it can be concluded that L5, L23, and L24 are the most effective ligands and have pharmacokinetic parameters that are within the acceptable range for human, better than the reference ligand.

In conclusion, the promising results for L5,L23,L24 ligands could be evaluated to generate better drug candidates for cancer treatment.

Abstract

Cyclin-dependent kinase inhibition is considered a promising target for cancer treatment for its crucial role in cell cycle regulation. In this work, we are focused on the development of new potential inhibitors Cyclin-dependent kinase 2 in order to enrich the therapeutic classes of anti-cancer drugs.

Molecular docking studies were applied to investigate the binding mode of the promising compounds and CDK2 receptor by calculating their docking energies and visualizing their interaction with active site of CDK-2 protein compared to the crystalized ligand. The ADME studies and drug-likeness showed proper pharmacokinetic properties which helped in structure requirements prediction for the observed antitumor activity. The in-silico results allowed us to conclude that ligands: L5, L23, and L24 could be a potential inhibitor of the Cyclin-dependent kinase 2.

Résumé

L'inhibition de la kinase cycline-dépendante est considérée comme une cible prometteuse pour le traitement du cancer en raison de son rôle crucial dans la régulation du cycle cellulaire. Dans ce travail nous nous sommes intéressés au développement de nouveaux inhibiteurs potentiels de la kinase cycline-dépendante2 afin d'enrichir les classes thérapeutiques des médicaments anticancéreux.

Les études de docking moléculaire ont été appliquées pour étudier le mode de liaison des composés prometteurs et le récepteur CDK2 en calculant leurs énergies de docking et en visualisant leur interaction avec le site actif de la protéine CDK2 par rapport au ligand cristallisé. Les études ADME et drug-likeness ont montré des propriétés pharmacocinétiques appropriées qui ont aidé à prédire les exigences structurelles pour l'activité anti-tumorale observée. Les résultats in-silico nous ont permis de conclure que les ligands L5, L23, et L24 pourraient être un inhibiteurs potentiel de la kinase cycline-dépendante 2.

République Algérienne Démocratique et Populaire
inistère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Med Khider Biskra
aculté des Sciences Exactes et des Sciences de la Nature et de la Vie
Département des Sciences de la Matière
Filière de Chimie

الجمهورية الجزائرية الديموقراطية الشعبية
وزارة التعليم العالي و البحث العلمي
جامعة محمد خيضر بسكرة
كلية العلوم الدقيقة و علوم الطبيعة و الحياة
قسم علوم المادة
شعبة الكيمياء

تصريـح شرفـــي

خاص بالالتزام بقواعد النزاهة العلمية لإنجاز بحث

(ملحق القرار 1082 المؤرخ في 2021/12/27)

أنا الممضي أسفله،

السيد(ة): منة الله خديجة خالدي

الصفة: طالب سنة ثانية ماستر كيمياء تخصص: كيمياء حيوية

الحامل(ة) لبطاقة التعريف الوطنية رقم: 1100[illegible] الصادرة بتاريخ: 2022/09/13

المسجل بكلية: علوم الطبيعة والحياة قسم: علوم المادة

والمكلف بانجاز أعمال بحث : مذكرة ماستر في الكيمياء

عنوانها: Structure Based discovery of novel cyclin-Dependant Kinase-2- (CDK2) inhibitors For the Treatment of Cancer

أصرح بشرفي أني ألتزم بمراعاة المعايير العلمية والمنهجية ومعايير الأخلاقيات المهنية والنزاهة الأكاديمية المطلوبة في انجاز البحث المذكور أعلاه وفق ما ينص عليه القرار رقم 1082 المؤرخ في 2021/12/27 المحدد للقواعد المتعلقة بالوقاية من السرقة العلمية ومكافحتها.

التاريخ: 2024/03/11

إمضاء المعني بالأمر

KHALDI