T.R.

GEBZE TECHNICAL UNIVERSITY

GRADUATE SCHOOL

THE IDENTIFICATION OF POTENTIAL TARGETS OF CYCLIN DEPENDENT KINASE INHIBITORS (CDK-i) USING *IN SILICO* METHODS AND CONFIRMATION WITH *C. ELEGANS* AND CELL LINE DISEASE MODELS

RIDVAN AZİZ AYAZ

A THESIS OF MASTER OF SCIENCE

DEPARTMENT OF BIOTECHNOLOGY

ADVISOR: PROF. DR. ELİF DAMLA ARISAN

JANUARY 2024

T.R.

GEBZE TECHNICAL UNIVERSITY

GRADUATE SCHOOL

THE IDENTIFICATION OF POTENTIAL TARGETS OF CYCLIN DEPENDENT KINASE INHIBITORS (CDK-I) USING *IN SILICO* METHODS AND CONFIRMATION WITH *C. ELEGANS* AND CELL LINE DISEASE MODELS

RIDVAN AZİZ AYAZ

A THESIS OF MASTER OF SCIENCE

DEPARTMENT OF BIOTECHNOLOGY

ADVISOR: PROF. DR. ELİF DAMLA ARISAN

2. ADVISOR: ASSOC. PROF. MEHMET ÖZBİL

JANUARY 2024

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

SİKLİN BAĞIMLI KİNAZ İNHİBİTÖRLERİ (CDK-I) HEDEFLERİNİN *İN SİLİKO* METODLAR KULLANILARAK TANIMLANMASI VE *C. ELEGANS* VE HÜCRE HATTI HASTALIK MODELLERİNDE DOĞRULANMASI

RIDVAN AZİZ AYAZ

YÜKSEK LİSANS TEZİ
BİYOTEKNOLOJİ ANABİLİM DALI

DANIŞMAN: PROF. DR. ELİF DAMLA ARISAN
2. DANIŞMAN: DOÇ. DR. MEHMET ÖZBİL

OCAK 2024

MASTER of SCIENCE JURY APPROVAL FORM

A thesis submitted by Rıdvan Aziz AYAZ defended on 31/01/2024 before the jury formed with the 30/01/2024 date and 2024-09 numbered decision of the GTU Graduate Administration Board, has been accepted as a MASTER of SCIENCE thesis in the Department of Biotechnology, Master of Science Program.

JURY

MEMBER
(THESIS ADVISOR) : Prof. Dr. Elif Damla ARISAN

MEMBER : Prof. Dr. Pınar OBAKAN YERLİKAYA

MEMBER : Asst. Prof. Hüseyin ÇİMEN

APPROVAL

Gebze Technical University Graduate Administration Board
..... /..... /..... date and ...../..... numbered decision.

SIGNATURE/SEAL

Dedicated to my beloved family...

ABSTRACT

Small molecules, ligands, and drugs can exhibit promiscuous binding within an organism, and this varied selectivity can hinder the efficacies of therapeutics in diseases. While such variations and lack of predictability can result in a wide range of side effects in patients, detailed profiling of drug-target interactions and the utilization of simplified model organisms can highlight the hidden potentials of drugs and ease side effect management, especially in cancer.

In this study, we employed computational approaches such as molecular docking, molecular dynamics and expression profiling to identify secondary or non-canonical targets of cyclin-dependent kinase inhibitor (CDK-i) molecules. We also evaluated the effects of few CDK-i molecules on MIA PaCa-2 and PANC-1 cells using 2D gel/MS, and on the model organism *Caenorhabditis elegans* via physiological assessments.

Our findings indicate that CDK-i molecules may promiscuously interact with other CDKs and many other potential targets, especially ATP-binding domain containing proteins, resulting in a few key pathway alterations. Most notably, metabolic and biosynthesis of cofactors pathways were highlighted in humans, while MAPK, FoxO pathways, and developmental phenotypes were identified in *C. elegans*. Upon evaluating the most potent targets, PRKCA, NOS2, HADHA in humans, and DAF-2 and DAF-4 in *C. elegans*, were found as targets of cyclin-dependent kinase inhibitors. Moreover, palbociclib and roscovitine delayed larval development, while abemaciclib increased lipid content in *C. elegans*. In MIA PaCa-2 and PANC-1 cells, palbociclib increased Peroxisome 1 and 6 levels. Additionally, CDK levels are differentially higher in embryonic stages, and any drug and toxicology studies in *C. elegans* can benefit from target profiling in varied larval stages. Lastly, we highlighted a few key effects of CDK-i drugs in various cancers and as well as in *C. elegans*, and suggested that abemaciclib, roscovitine, and ribociclib might have potential effects on other cancers.

Keywords: CDK inhibitor, Receptor Screening, *in silico*, Molecular Modeling, *Caenorhabditis elegans*

ÖZET

Küçük moleküller, ligandlar ve ilaçlar bir organizma içinde seçicisiz bağlanma sergileyebilir ve bu değişen seçicilik, hastalıklardaki terapilerin etkinliğini engelleyebilir. Bu tür değişkenlikler ve öngörülebilirlik eksikliği, hastalarda geniş bir yan etki yelpazesine neden olabilir. Ancak ilaç-hedef etkileşimlerinin detaylı profillemesi ve basitleştirilmiş model organizmaların kullanımı, ilaçların gizli potansiyellerini ortaya çıkarabilir ve özellikle kanserde yan etki yönetimini kolaylaştırabilir.

Bu çalışmada, sikline-bağımlı kinaz inhibitör (CDK-i) moleküllerinin ikincil veya amaç dışı hedeflerini belirlemek için moleküler kenetlenme, moleküler dinamik simulasyonları ve ifade profillemesi gibi hesaplamalı yaklaşımlar kullanılmıştır. Ayrıca, bu moleküllerden bazılarının etkileri 2D jel/MS yaklaşımı ile MIA PaCa-2 ve PANC-1 hücrelerinde, fizyolojik değerlendirmeler aracılığı ile de *Caenorhabditis elegans* model organizmasında değerlendirilmiştir.

Bulgular, sikline-bağımlı kinaz inhibitörlerin diğer sikline-bağımlı kinazları ve birçok diğer potansiyel hedefle seçicisiz etkileşimde bulunabileceğini, özellikle ATP bağlanma bölgesi içeren diğer proteinleri de hedefleyeceğini, bu durumun birkaç önemli yolakta değişikler ile sonuçlanabileceğini göstermektedir. Özellikle, insanlarda metabolik ve koenzimlerin biosentezi yolakları vurgulanırken, *C. elegans*'da MAPK, FoxO yolakları ve gelişimsel fenotipler ön plana çıkmıştır. Potansiyel hedefler değerlendirdikten sonra, insan için PRKCA, NOS2, HADHA ve *C. elegans*'da da DAF-2 ve DAF-4, sikline-bağımlı kinaz inhibitörlerinin hedefleri olarak belirlenmiştir.

Ayrıca, palbosiklib ve roskovitin *C. elegans*'ta larva gelişimini geciktirirken, abemasiklib lipit miktarını arttırmıştır. Palbociclib, MIA PaCa-2 ve PANC-1 hücrelerinde ise, Peroksizom 1 ve 6 seviyelerini artırmıştır. Bunlara ek olarak, CDK seviyeleri embriyonik aşamalarda daha yüksektir ve *C. elegans*'daki herhangi bir ilaç ve toksikoloji çalışması, çeşitli larva aşamalarında hedef profilinin çıkarılması ile zenginleştirilebilir. Son olarak, çeşitli kanser türlerine ek olarak *C. elegans*'ta CDK-i ilaçlarının birkaç önemli etkisi vurgulanmış ve abemasiklib, roskovitin ve ribosiklib'in diğer kanserler üzerinde potansiyel etkileri olabileceği önerilmiştir.

Anahtar Kelimeler: CDK inhibitörü, Reseptör Taraması, *in-siliko*, Moleküler Modelleme, *Caenorhabditis elegans*

ACKNOWLEDGEMENTS

I would like to express my thanks to the Scientific and Technical Research Council of Turkey (TÜBİTAK) BIDEB 2210-C National Master's Scholarship Program for considering this study as a part of the Priority Fields in Science and Technology. Additionally, I would like to thank to TÜBİTAK ARDEB for partially supporting this study, with project number 118Z100. Their support played a crucial role in the completion of this thesis.

Undoubtedly, success and progress are a team effort. I am deeply indebted to my supervisor, Prof. Dr. Elif Damla ARISAN, for her unwavering guidance, continuous support, and valuable insights. I am particularly grateful for her trust and encouragement, allowing me the freedom to pursue and explore my ideas, which has been pivotal in shaping the trajectory of my academic journey.

I extend my heartfelt appreciation to my co-advisor, Assoc. Prof. Mehmet ÖZBİL, for his valuable contributions, constructive feedback, and encouragement that greatly enriched the content of this thesis.

To my friends, thank you for the motivation, companionship, and support, making this academic endeavor more fulfilling.

Last but not least, a special thanks goes to my family for their unconditional love, understanding, and encouragement. Your belief in me has been a driving force behind my academic and personal achievements. Importantly, I also want to express my love to our cat, whose silent companionship provided comfort during long hours of writing.

Thank you all for being part of my academic journey. Your support has been invaluable, and I truly appreciate each contribution.

TABLE OF CONTENTS

	Page
ABSTRACT	vi
ÖZET	vii
ACKNOWLEDGEMENTS	viii
TABLE of CONTENTS	ix
LIST of ABBREVIATIONS and ACRONYMS	xi
LIST of FIGURES	xii
LIST of TABLES	xiii
1. INTRODUCTION	1
1.1. Drug Promiscuity on Multiple Targets	1
1.2. Cancers and Chemotherapeutics	3
1.3. Cyclin-Dependent Kinases	6
1.4. Cyclin-Dependent Kinase Inhibitors In Cancer	9
1.5. Computational Approaches In Drug Target Determination	13
1.5.1. Molecular Dynamics Simulations	14
1.5.2. Molecular Docking Simulations	15
1.5.3. Receptor Screening	16
1.6. Model Organism *Caenorhabditis elegans*	17
1.7. Aim of This Study	19
2. MATERIALS AND METHODS	20
2.1. Computational Methods	21
2.1.1. Sequence and Structure Similarity Assessments for CDKs	21
2.1.2. Protein and Ligand Structure Preparation	22
2.1.3. Receptor Screening	24
2.1.4. Network and Gene Set Enrichment	25
2.1.5. Network Hierarchy Analysis	25
2.1.6. Expression Profiling for Human and *C. elegans* Targets	26
2.1.7. Molecular Dynamics Simulations and Binding Free Energies	27
2.2. Mass Spectrometry Analysis of Human Cell Lines	29
2.3. *C. elegans* Methods	31
2.3.1. Culturing	31
2.3.2. Cryopreservation	32
2.3.3. L1 Larvae Synchronization	32
2.3.4. Dauer Larvae Synchronization	33
2.3.5. Development Assays	33
2.3.6. Swimming Assays	34
2.3.7. Lipid Accumulation Assays	35

2.4. Laboratory Materials 36
2.4.1. Chemicals, Buffers and Solutions 36
2.4.1.1. Nematode Growth Medium 37
2.4.1.2. M9 Buffer 37
2.4.1.3. KPO_4 Mix Buffer 37
2.4.1.4. Bleach Solution 37
2.4.1.5. Oil Red O Staining Solution 38
2.4.2. Laboratory Equipment 38
2.5. Statistical Analysis, Graphics, and Data Processing 39

3. RESULTS 40
3.1. Similarities Between CDKs in Human and *C. elegans* 40
3.2. Receptor Screening Results 45
3.3. Functional Enrichments and Hierarchy Analysis 48
3.4. Expression Profiles of Selected Targets in Human and *C. elegans* 54
3.5. MD Analysis of Identified Targets 57
3.6. Proteomic Effects of Drugs in Human Cell Lines 62
3.7. Physiological Effects of Drugs on *C. elegans* 64

4. DISCUSSIONS AND CONCLUSIONS 68
4.1. Discussions 68
4.1.1. Structural Similarity of CDKs and Drug Molecular Weight Might Indicate Varied Selectivity 68
4.1.2. CDK-is Can Alter Metabolic Processes 69
4.1.3. Drug Effects Can Variate Between Different Biological Conditions 71
4.1.4. CDK-is Show Binding to Possible Targets In Molecular Dynamics Simulations 72
4.1.5. Abemaciclib, Palbociclib and Roscovitine Alter MAPK, FoxO and TGF-Beta Pathways in *C. elegans* 73
4.1.6. *C. elegans* CDKs Required in Embryonic Stages 74
4.1.7. CDK-is Might Be Repurposed in Other Cancers 76
4.2. Conclusions 79

REFERENCES 80

BIOGRAPHY 92

PUBLICATIONS WITHIN the SCOPE of THESIS STUDY 93

APPENDICES 94

LIST of ABBREVIATIONS and ACRONYMS

CDK	:	Cyclin-dependent Kinase
CDK-i	:	Cyclin-dependent Kinase Inhibitor
FUDR	:	Floxuridine
kDa	:	kilo-Dalton
kcal/mol	:	Kilocalorie per mol
MD	:	Molecular Dynamics
PME	:	Particle mesh Ewald
NVT	:	Constant-volume ensemble
NPT	:	Constant-pressure ensemble
3D	:	3-Dimensional
ID	:	Identifier
TM-Score	:	Template modeling score
CPU	:	Central Processing Unit
GPU	:	Graphical Processing Unit
fc	:	fold-change
E2F	:	Transcription factor E2F
RB	:	Retinoblastoma-associated protein
MAPK	:	Mitogen-activated protein kinase
FoxO	:	Forkhead box O
KEGG	:	Kyoto Encyclopedia of Genes and Genomes
PRKCA	:	Protein kinase C alpha
HADHA	:	Trifunctional enzyme subunit alpha, mitochondrial
PEX5	:	Peroxisomal targeting signal 1 receptor
NTRK1	:	High affinity nerve growth factor receptor
ADSS2	:	Adenylosuccinate synthetase isozyme 2
NOS2	:	Nitric oxide synthase, inducible
GAPDH	:	Glyceraldehyde-3-phosphate dehydrogenase
SHMT2	:	Serine hydroxymethyltransferase, mitochondrial
EPRS1	:	Bifunctional glutamate/proline–tRNA ligase
MAPK14	:	Mitogen-activated protein kinase 14
LET-756	:	Lethal protein 756
PMK-1	:	Mitogen-activated protein kinase pmk-1
MEK-2	:	Dual specificity mitogen-activated protein kinase kinase mek-2
DAF-2	:	Abnormal dauer formation protein 2
RPM	:	Round per Minute
MM/GBSA	:	Molecular Mechanics with Generalised Born and Surface Area Solvation
NGM	:	Nematode Growth Medium
MS	:	Mass Spectrometry
ANOVA	:	Analysis of Variance

LIST OF FIGURES

Page

Figure 1.1 : Promiscuous drugs can bind multiple target proteins. 2
Figure 1.2 : Dose response relationship of chemotherapeutic drugs. 5
Figure 1.3 : CDK4/6 is a key regulator of the cell cycle. 8
Figure 1.4 : CDK inhibitors and corresponding target CDKs. 12
Figure 1.5 : Abemaciclib and CDK-2 as an example to molecular docking. 15
Figure 1.6 : Schematic representation of screening methodologies. 16
Figure 1.7 : *Caenorhabditis elegans* life cycle at 20 °C. 18

Figure 2.1 : Graphical methodology workflow of this study 20
Figure 2.2 : Graphical abstract for utilized methodology in *C. elegans* experiments. 31
Figure 2.3 : Data acquisiton in *C. elegans* development assays 34
Figure 2.4 : Image of Oil Red O stained *C. elegans* worms. 35

Figure 3.1 : Multiple sequence alignments of Cyclin-Dependent Kinases. 41
Figure 3.2 : Sequence identity percentage matrix and phylogenetic tree for Human and *C. elegans* CDKs. 42
Figure 3.3 : Combined structural similarity matrix for human and *C. elegans* CDKs. 43
Figure 3.4 : Superimposition of Human and *C. elegans* CDKs. 44
Figure 3.5 : Superimposition of Human and *C. elegans* CDK 5. 44
Figure 3.6 : Human and *C. elegans* dataset sizes and distributions. 47
Figure 3.7 : Drug efficiency scores of drugs for Human and *C. elegans* CDKs. 48
Figure 3.8 : Functional enrichment results of human and *C. elegans* datasets. 50
Figure 3.9 : Protein interaction network for human datasets. 52
Figure 3.10 : Protein interaction network for *C. elegans* datasets. 53
Figure 3.11 : Gene expression heatmap for human cancer cell lines and tumor datasets. 56
Figure 3.12 : Gene expression profile of selected targets in *C. elegans* larval stages. 57
Figure 3.13 : RMSD graphs of selected targets and drug molecules. 58
Figure 3.14 : Binding free energy results of selected targets of human and *C. elegans*. 59
Figure 3.15 : Human target-drug interaction diagrams. 60
Figure 3.16 : *C. elegans* drug-target interaction diagrams. 61
Figure 3.17 : Average body areas of *C. elegans* individuals grown in drug-free and drug-containing environments. 65
Figure 3.18 : Physiological Effects of Drugs on *C. elegans* swimming and oil accumulation. 67

Figure 4.1 : CDK levels in *C. elegans* larval stages. 75
Figure 4.2 : Arbitrary effect potential scores of CDK-I drugs against cancers. 77

LIST OF TABLES

		Page
Table 2.1:	Uniprot identification codes for human and *C. elegans* CDKs.	21
Table 2.2:	Selected drug molecules and properties.	23
Table 2.3:	Selected primary targets of drugs for minimum affinity cutoff in human and *C. elegans*.	24
Table 2.4:	Protein structure identifiers for human and *C. elegans*.	27
Table 2.5:	Chemicals used in this study.	36
Table 2.6:	Chemicals used in *C. elegans* developmental assay.	36
Table 2.7:	Instruments used in this study.	38
Table 3.1:	Cutoff values for human and *C. elegans* datasets.	45
Table 3.2:	Names of the identified targets.	54
Table 3.3:	List of the differentially regulated proteins in abemaciclib-treated PANC-1 cells.	62
Table 3.4:	List of the differentially regulated proteins in palbociclib-treated PANC-1 cells.	63
Table 3.5:	List of the differentially regulated proteins in abemaciclib-treated MIA PaCa-2 cells.	64
Table 3.6:	List of the differentially regulated proteins in palbociclib-treated MIA PaCa-2 cells.	64

1. INTRODUCTION

1.1. Drug Promiscuity on Multiple Targets

Throughout history, humans have sought remedies from various natural sources to address a spectrum of illnesses, spanning infectious diseases to psychiatric disorders. There are more than 22,700 known diseases have been documented, including 14,790 rare conditions and 6,404 drug-associated conditions [1]. In the realm of modern medicine, there are 16,576 recorded drugs available [2]. Therapeutic agents, often derived from plants, are procured for use either as drugs or to produce bioactive compounds [3]. In pharmacology, the binding of a drug is a nuanced process that involves the interaction between the drug molecule and specific proteins within the body [4]. This interaction is analogous to a molecular handshake, where the drug recognizes and engages with its target in a highly selective manner. The key to this interaction lies in the molecular structure of the drug and the complementary features of the target protein [5]. Generally, it was exemplified like that drugs are three-dimensional puzzle pieces designed to fit precisely into a corresponding cavity on the protein surfaces. This interaction is governed by a myriad of forces, including electrostatic attractions, hydrogen bonding, and hydrophobic interactions [6]. For drugs targeting receptors, this binding event induces a conformational change in the receptor. This change, related to flipping a switch, triggers a cascade of cellular events. In the case of enzymes, the drug may dock onto the enzyme's active site, disrupting its normal catalytic function. Addition to receptors and enzymes, drugs may also modulate signaling pathways by interacting with key components of these pathways, and can influence the transmission of signals, ultimately regulating cellular responses [7].

When drugs and their interactions with the enzyme or other protein targets are considered, concepts of affinity and specificity are important in this dance of molecules. Affinity describes the strength of the interaction, while specificity ensures that the drug selectively engages with its intended target within a sea of other molecules [8]. In the intricate relationship between a drug and its target, the process of binding and stability involves a delicate interaction of molecular forces and dynamic adaptations.When a

drug unexpectedly binds to a non-intended target within the biological system, it can lead to a cascade of repercussions. These may encompass disruptions in normal cellular signaling pathways, eliciting unintended physiological responses. The consequences can extend to adverse reactions, diminishing the drug's intended efficacy, and even instigating alterations in pharmacokinetics. Such off-target binding events can also trigger immunogenic responses, causing an immune reaction against the drug itself or inducing unwanted immune responses in the body [9]. Additionally, there is the potential for drug-drug interactions, where the presence of the unintended binding might interfere with the action of other medications.

Some drugs exhibit efficacies in treating conditions beyond their primary indications. Small molecules, owing to their inherent characteristics, can display affinity and activity on multiple targets, a phenomenon known as promiscuity [10]. This property can prove both beneficial and detrimental [11].

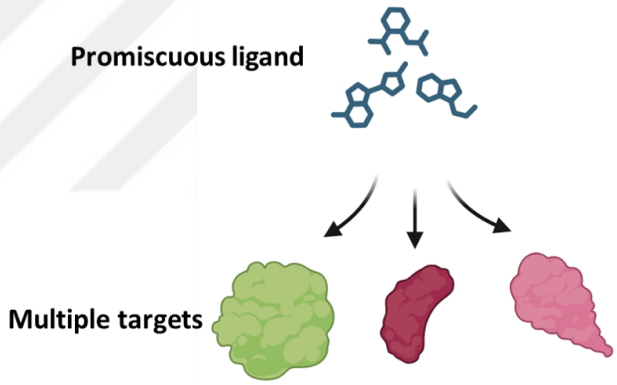


Figure 1.1: Promiscuous drugs can bind multiple target proteins.

The compromised selectivity resulting from non-specific binding may not only affect the drug's therapeutic benefits but could also pave the way for unforeseen long-term consequences. It underscores the importance of rigorous testing and thorough understanding of a drug's interactions within the biological environment to mitigate these risks and enhance the overall safety profile of pharmaceutical interventions. Numerous drugs demonstrate the ability to act on various targets, encompassing both canonical and non-canonical (unexpected) interactions [12]. The latter type of interactions is hypothesized to underlie the occurrence of side effects associated with therapeutic agents [13]. Managing these side effects becomes a critical consideration in

clinical settings, particularly in the treatment of diseases such as cancer [14]. Understanding the intricate landscape of drug interactions and their potential consequences is pivotal for effective disease management in the clinical realm.

1.2. Cancers and Chemotherapeutics

Cancer is a wide range of complications of uncontrolled division of cells within the body and while it is generally associated with tumor progression as well as many other physiological symptoms; it also consists of many genetic, chemical and external factors prior to progression and to treatment stages [15]. All cancers share uncontrolled cell growth with potential to spread, and there are also many differences in the way that cancer manifests itself in different individuals [16]. Although cancer general mechanisms are understandable, various cases originating from various organs require different sets of approaches [17]. Additionally, it is known that each cancer case might show different expression, interaction, and metabolism profiles [18].

Cancers are typically named based on the tissue or organ of origin, further classified into subtypes according to the type of cell from which the tumor arises. Common solid tumor types include carcinomas, sarcomas (such as osteosarcoma, chondrosarcoma, and leiomyosarcoma), neuroendocrine tumors (including carcinoid tumors, small cell lung cancer, and pancreatic neuroendocrine tumors), and gastrointestinal stromal tumors. A comprehensive understanding of many cancer types has been achieved through extensive research [19]. While variations in different cancers can be easily understandable, there can also be significant variations in how same cancer type presents itself in different patients [20]. Two patients with lung cancer may have different genetic mutations that affect the course of the disease, or they may have been exposed to different environmental factors that influence the severity of the cancer. Additionally, patients with the same type of cancer may respond differently to treatment based on their age, lifestyle factors, and overall health [21]. To better understand these variations and develop personalized cancer treatments, new techniques for analyzing individual tumors at the molecular level are concurrently being developed and utilized [22]. These techniques, such as genomic sequencing and proteomics, can help identify the specific genetic and protein-level changes that are driving cancer growth in each patient, and develop targeted therapies that are tailored to their unique needs.

Clinically, main cancer treatment options are radiotherapy to target and destroy cancerous cells or tumors, immunotherapy to target cancerous cells, tissues, or tumors via immunological biomolecules, and also surgically removing such formations in the body. In addition, and maybe as a well-known approach, chemotherapy drugs are widely used to hinder cell division in a controlled manner or altogether, while utilizing other approaches and enhancing immunity in aim to overcome cancers [23].

The treatment options for cancer depend on several factors, including the type and stage of the cancer, the patient's overall health, and the patient's preferences [24]. Some of the most common cancer therapy options include surgery, radiotherapy, immunotherapy,chemotherapy, and targeted therapy. In surgery, surgical invasive methods are often used to remove solid tumors, and this is usually the first line of treatment for early-stage cancers. The goal of surgery is to remove as much of the cancer as possible while preserving the surrounding healthy tissue. Another known approach, radiation therapy uses high-energy radiation to kill cancer cells. It can be used alone or in combination with other treatments, such as surgery and chemotherapy. Radiation therapy is often used to shrink tumors before surgery or to kill any remaining cancer cells after surgery [25]. Another emerging option, immunotherapy is based on the idea of using the body's own immune system to fight cancer. It works by targeting and activating the immune cells that can recognize and attack cancer cells. Immunotherapy has shown promising results in treating several types of cancer, including melanoma, lung cancer, and bladder cancer [26]. Another option, and maybe the most notoriously known among the public because of its many side effects [27], the chemotherapy approach can be simply explained as using drugs to kill cancer cells. These drugs can be taken orally or given intravenously. Chemotherapy is often used to treat cancers that have spread to other body parts or shrink tumors before surgery. Chemotherapy drugs affect cells at specific stages of a cell's cycle; however they do not differentiate between cancerous and healthy cells (Figure: 1.2) [28]. Many healthy cells also have a rapid cell cycle, such as those in the gastrointestinal tract and hair follicles and thus are more susceptible to the effects of chemotherapy resulting in nausea and hair loss during treatment [29]. Chemotherapy is a systemic cancer treatment involving various chemotherapeutic agents, such as antimetabolites, alkylating agents, anthracyclines, topoisomerase inhibitors, and platinum compounds, that target rapidly

dividing cancer cells. These agents can be used alone or in combination with each other and with other cancer treatments like surgery, radiation therapy, immunotherapy, and targeted therapy. There are many different kinds of chemotherapy agents available, and their combinations are based on factors such as the type and stage of cancer, the patient's overall health status, and potential side effects [30].

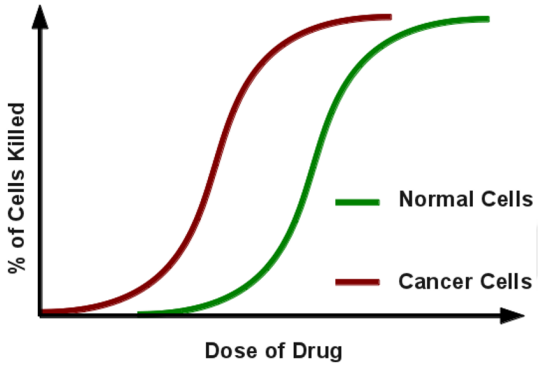


Figure 1.2: Dose response relationship of cell killing by chemotherapeutic drugs on normal and cancer cells. At high doses the percentage of normal and cancer cells killed is very similar **[28]**.

In some cases, the drug side effects may be a result of the drug's interaction with a specific organ or tissue. Example to this, cisplatin is a chemotherapeutic agent used to treat a variety of cancers, but it is known to cause nephrotoxicity, or damage to the kidneys [31] while similarly doxorubicin is known to cause cardiotoxicity [32]. Other factors that contribute to drug toxicity include the patient's genetic makeup and lifestyle factors. Patients with certain genetic variations may metabolize drugs differently, leading to differences in toxicity and efficacy [33]. Lifestyle factors such as smoking and alcohol consumption can also affect the metabolism of drugs and increase the risk of toxicity. To minimize drug toxicity, researchers and clinicians aim to develop drugs that are highly selective for cancer cells and have minimal impact on healthy cells. This approach, known as targeted therapy, has led to the development of several successful cancer drugs, including imatinib (Glivec®) which targets multiple tyrosine kinases such as CSF1R, ABL, c-KIT, FLT3, and PDGFR-β in chronic myelogenous leukemia cells [34], and trastuzumab (Herceptin), which targets HER2 in breast cancer cells [35].

Lastly, a promising and emerging approach to cancer is using targeted therapy. Targeted therapy approach uses drugs that target specific molecules or proteins involved in cancer cell growth and spread. These drugs can be more effective and have fewer side effects than traditional chemotherapy drugs [36]. Targeted therapy is often used to treat cancers that have specific genetic mutations, such as breast cancer and lung cancer [37].
Uncontrolled cell growth and many variations in cancers, as well as necessity of efficient targeting of specific macromolecules and enzymes in cancers highlighted the potential target properties of cyclin-dependent kinases (CDKs) over the years.

1.3. Cyclin-Dependent Kinases

When cells undergo division, many important enzymes take part in various signaling cascades. One of the many, but a well-known and essential set of enzymes is Cyclin Dependent Kinases (CDK's), are protein kinases characterized by needing a separate subunit - a cyclin - that provides domains essential for enzymatic activity [38]. CDKs also play important roles in the modulate transcription in response to several extra- and intracellular cues. While CDK's are evolutionary conserved in many eukaryotic cell division and DNA replication machinery [39], the evolutionary expansion of the CDK family in mammals led to the division of CDKs into three cell-cycle-related subfamilies (Cdk1, Cdk4 and Cdk5) and five transcriptional subfamilies (Cdk7, Cdk8, Cdk9, Cdk11 and Cdk20). CDKs respond to the extracellular and intracellular signals to regulate cell division, acting as the catalytic subunits by forming a heterodimer complex with the cyclins, which function as the regulatory subunits [38].
In human (*Homo sapiens*) cells, 20 CDKs and 29 cyclins exist [40]. The expression of CDKs follows a cyclical pattern throughout the cell cycle [41]. CDKs that directly involved in regulating cell-cycle transitions and division are CDK1, CDK2, CDK3, CDK4, CDK6, and CDK7, while CDK7–11 mediate gene transcription [42]. These factors converge on the cell cycle, activating intracellular signaling networks that impact CDK4 and CDK6, propelling cell-cycle progression from G0 or G1 phase into the S phase.
CDK4/6 activity is positively regulated by the association with D-type cyclins (cyclin D1, cyclin D2, and cyclin D3). Active cyclin D/CDK4/6 complexes initiate the phosphorylation of the tumor suppressor protein RB (encoded by RB1). RB recruits

co-repressors, repressing the transcription of target genes regulated by E2F transcription factors (E2Fs), inhibiting G1/S transition. Sequential phosphorylation inactivates RB, leading to cell-cycle progression from G1 into S phase. Phosphorylated RB de-represses E2F transcription factors, inducing the transcription of G1/S target genes like cyclin E (CCNE), CCNA, CCNB, dihydrofolate reductase (DHFR), ribonucleotide reductase M1 (RRM1), RRM2, polo-like kinase 1 (PLK1), spindle checkpoint protein MAD2, and BUB1 mitotic checkpoint serine/threonine kinase (BUB1), involved in G1–S entry and cell-cycle progression [42].

During late G1 phase, target genes (cyclins E1 and E2) of E2Fs are activated. The cyclin E/CDK2 complex, specifically, modulates RB to override the restriction point of the G1/S phase at the boundary, initiating S phase and forming a positive feedback loop [43].

The activities of CDK4/6 and CDK2 coordinate cell-cycle progression into S phase, known as the ”restriction point” [44]. Towards the end of S phase, cyclin A removes cyclin E, forming a new complex, cyclin A/CDK2. The cyclin A/CDK2 complex concludes S phase by phosphorylating CDC6 and E2F1, driving the cell-cycle transition from S phase to G2 phase. Subsequently, it activates CDK1 through cyclin A, leading cells into the transition to the M phase [42].

During mitosis, CDK1 activity is maintained by the cyclin B/CDK1 complex. Deregulated CDK1 expression enables chromosome separation and completes mitosis and cytokinesis. CDK1 is the only CDK essential for cell-cycle progression, initiating mitosis and ensuring critical events to occur [45]. Additionally, cyclin C/CDK3 initiates RB phosphorylation, driving cells from G0 to S phase. The cell cycle exit, transitioning into the reversible or permanent G0 phase, is also governed by cyclin C/CDK3 [46].

CDKs are relatively small proteins, typically weighing between 34 to 40 kDa, and are predominantly composed of the kinase domain. When cyclin is absent, CDK displays minimal kinase activity; only the cyclin-CDK complex serves as an active kinase. This activity is subject to further regulation through phosphorylation and interactions with other binding proteins. CDKs phosphorylate serines and threonines, categorizing them as serine-threonine kinases [38]. Vital roles and functions of kinases in cellular metabolism have been thoroughly explored and documented in the existing literature.

In normal cells, the activity of CDKs is tightly regulated to ensure that cells divide only

when there are no abnormal conditions within the control checkpoints. However, in cancer cells, the activity of CDKs is often dysregulated, leading to uncontrolled cell division and tumor growth [47]. There are several ways in which CDK activity can become dysregulated in cancer. One commonly observed way CDK activity becomes dysregulated is through the overexpression of cyclins [48]. Another way that CDK activity becomes dysregulated in cancer is through mutations or amplifications of CDK genes. Mutations in the CDK4 gene have been linked to an increased risk of developing melanoma, while amplification of the CDK6 gene has been found in several types of cancer [49]. In addition, cancer cells often have an overactive CDK-cyclin complex, which drives the cells to divide more rapidly than normal cells. This overactivity can also make cancer cells resistant to chemotherapy and radiation therapy, as these treatments are designed to target rapidly dividing cells. Because of their crucial role in cell cycle regulation and their dysregulation in cancer, CDK4/6 have emerged as promising targets for cancer therapy and drugs that inhibit CDK4/6 activity have shown remarkable clinical activity in various cancers, particularly hormone receptor-positive breast cancer, where they have become a standard treatment option [50].

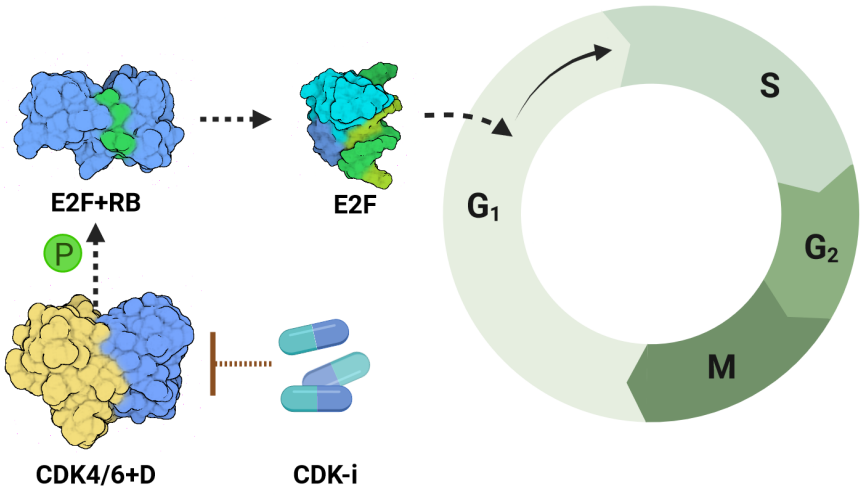


Figure 1.3: CDK4/6 is a key regulator of the cell cycle, acting by forming a complex with cyclin D. This complex can directly phosphorylate retinoblastoma (RB), then release transcription factor E2F and promote the transcription of cell cycle-related genes, promoting the cell cycle from the G1 mitosis. CDK inhibitor molecules can hinder the activity of CDKs (CDK4/6 in this case) to stop the relevant cascades and stop the cell cycle progression.

CDK4/6 inhibitors, such as palbociclib, ribociclib, and abemaciclib, work by inhibiting CDK4/6 activity, leading to cell cycle arrest and inhibition of cell proliferation. They have demonstrated impressive efficacy in preclinical and clinical studies, particularly in combination with endocrine therapy, and have significantly improved outcomes for patients with advanced breast cancer [51].

However, complex intracellular interactions of these kinase family proteins and other kinases and their involvement in multiple pathways highlights the effects of inhibition of these proteins outside of cancers or any disease condition.

1.4. Cyclin-Dependent Kinase Inhibitors In Cancer

Because of the critical role that CDKs play in cancer development and progression, Cyclin-Dependent Kinase inhibitors(CDK-i) have emerged as a potential cancer therapeutic targets and have shown remarkable clinical activity in various cancers, particularly hormone receptor-positive breast cancer, where they have become a standard treatment option [50].

CDK inhibitors target CDKs directly, either by binding to the CDK itself or to the cyclin that activates it. By inhibiting CDK activity, these drugs can slow or stop the growth of cancer cells. There are several types of CDK inhibitors, including selective CDK4/6 inhibitors, pan-CDK inhibitors, and CDK9 inhibitors. Selective CDK4/6 inhibitors target CDK4 and CDK6, which are involved in the G1 phase of the cell cycle and inhibiting these CDKs can slow or stop the progression of cancer cells through the cell cycle, which can slow tumor growth. Selective CDK4/6 inhibitors have been approved for treating hormone receptor-positive and HER2-negative breast cancer and are being tested in clinical trials for other types of cancer [52]. Pan-CDK inhibitors target multiple CDKs, including CDK1, CDK2, CDK4, CDK6, and CDK9, therefore these inhibitors have a broader spectrum of activity than selective CDK inhibitors and may be more effective against cancers that are resistant to selective inhibitors [53]. However, pan-CDK inhibitors may also have more side effects than selective inhibitors. The majority of CDK inhibitors are designed to target the ATP binding pocket. However, the highly similar catalytic domain and three-dimensional structure of CDK kinases pose a challenge in achieving the necessary selectivity [54].

Novel CDK inhibitors have shown promise in preclinical studies and clinical trials, and

several have been approved for the treatment of certain types of cancer. Additionally, CDK inhibitors may be more effective when used in combination with other cancer therapies, such as chemotherapy or radiation therapy, and may help to overcome resistance to treatment and improve outcomes for patients with cancer.

Currently, several CDK inhibitor drugs have been approved for the treatment of certain types of cancer, and many others are currently being tested in clinical trials [55]. The development of CDK inhibitor drugs might represent an important advance in cancer therapy, as these drugs have the potential to offer more targeted and less toxic treatments than traditional chemotherapy drugs.

The first CDK inhibitor to enter clinical trials was flavopiridol, which was initially developed as an anti-inflammatory agent but was found to have potent CDK inhibitory activity. Flavopiridol was tested in clinical trials for the treatment of various types of cancer, including leukemia, lymphoma, and solid tumors, but it was ultimately found to have limited efficacy and significant toxicity [56].

More recently, the focus of CDK inhibitor drug development has shifted towards the development of selective CDK inhibitors that target specific CDKs involved in the dysregulated cell cycle in cancer cells. The first selective CDK inhibitor to be approved by the FDA was palbociclib, which targets CDK4 and CDK6 and is approved for the treatment of hormone receptor-positive, HER2-negative breast cancer. Since the approval of palbociclib, several other selective CDK inhibitors have been developed and are in various stages of clinical development. These include abemaciclib, which also targets CDK4 and CDK6 and is approved for the treatment of hormone receptor-positive, HER2-negative breast cancer, and ribociclib, which also targets CDK4 and CDK6 and is approved for the treatment of advanced breast cancer. To date, the CDK inhibitors (CDK-is), specifically those that block the enzyme activity of CDK4 and CDK6 (CDK4/6), have been approved by the FDA for treating metastatic hormone receptor positive breast cancer [55]. Several drugs are in the spotlight that target the cell cycle and have shown promise in the treatment of cancer.

Among these drugs, Abemaciclib, Palbociclib, Roscovitine, Ribociclib, Trilaciclib, Lerociclib, Narazaciclib, Atuveciclib and PHA-767491 are currently either approved or in clinical trials (Table: 2.2).

Abemaciclib (Verzenio®) is a CDK4/6 inhibitor that is approved for the treatment of

hormone receptor-positive (HR+), HER2-negative advanced breast cancer. It is an orally administered drug that works by inhibiting the activity of CDK4 and CDK6, which are key regulators of the cell cycle. Abemaciclib binds to the ATP-binding site of CDK4 and CDK6, preventing their interaction with cyclin D and subsequent phosphorylation of the Rb protein. This leads to cell cycle arrest and inhibition of cell proliferation. Abemaciclib is effective as a monotherapy and in combination with other drugs such as endocrine therapy and chemotherapy [57].

Palbociclib (Ibrance®) is another CDK4/6 inhibitor that is approved for the treatment of HR+, HER2-negative advanced breast cancer. Like Abemaciclib, it works by inhibiting the activity of CDK4 and CDK6, leading to cell cycle arrest and inhibition of cell proliferation. Palbociclib is given orally and is often used in combination with endocrine therapy to improve outcomes for patients with advanced breast cancer [58].

Roscovitine(Seliciclib) is a small molecule inhibitor of CDKs that has been shown to have broad activity against several CDKs including CDK1, CDK2, CDK5, and CDK7. It has been studied in several types of cancer, including melanoma, breast cancer, and leukemia, with promising results. Given that,it is a first generation CDK inhibitor which has a broad-range purine inhibitor, which inhibits CDK1, CDK2, CDK5 and CDK7, but is a poor inhibitor for CDK4 and CDK6 [59].

Ribociclib (Kisqali®) is a CDK4/6 inhibitor that is approved for the treatment of HR+, HER2-negative advanced breast cancer. Like Abemaciclib and Palbociclib, it works by inhibiting the activity of CDK4 and CDK6, leading to cell cycle arrest and inhibition of cell proliferation. Ribociclib is given orally and is often used in combination with endocrine therapy to improve outcomes for patients with advanced breast cancer [60].

Trilaciclib (G1T28), a selective CDK4/6 inhibitor, is approved to reduce chemotherapy-induced myelosuppression in patients undergoing topotecan-containing or platinum and etoposide-containing chemotherapy for extensive stage small cell lung cancer [61]. Lerociclib (G1T38), another selective CDK4/6 inhibitor, is in clinical trials for treating solid tumors. It hinders CDK4 and CDK6 activity, resulting in cell cycle arrest and reduced cell proliferation. Studied in combination with endocrine therapy and other drugs, lerociclib shows promise in various cancers, including breast and ovarian cancer [62].

Narazaciclib (ON123300), a multi-kinase inhibitor, targets CDKs, Aurora kinases, and

PLK1. It induces cell cycle arrest and apoptosis in cancer cells. Under clinical trials, ON123300 is being evaluated on various cancers [63].

Atuveciclib (BAY-1143572), a selective CDK9 inhibitor, disrupts transcriptional regulation, leading to cell death in cancer cells. Clinical trials are ongoing, exploring its efficacy in treating acute myeloid leukemia and other cancers [64].

PHA-767491, a Cdc7/CDK9 inhibitor, suppresses glioblastoma growth and invasiveness in preclinical trials[65].

As a non-CDK inhibitor, FUDR (5-fluorodeoxyuridine) interferes with DNA synthesis in cancer cells. Converted to its active form, 5-fluorouracil (5-FU), it inhibits thymidylate synthase, leading to cell death. FUDR is commonly used in colorectal cancer treatment and has applications in pancreatic and liver cancer therapy.

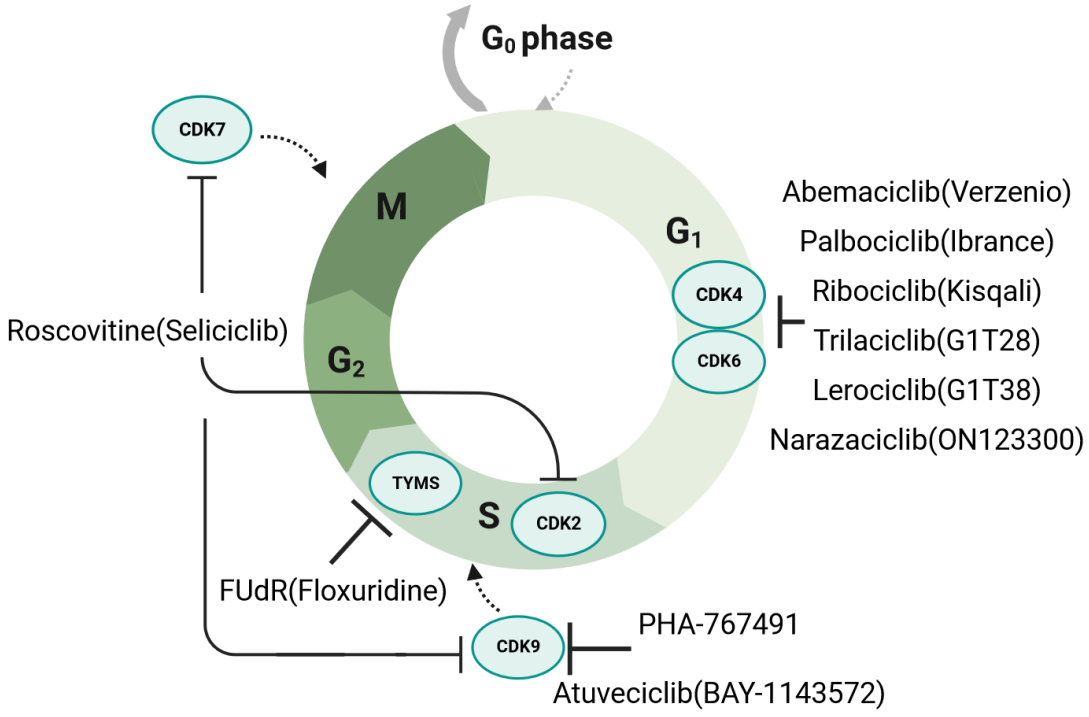


Figure 1.4: Cyclin-dependent kinase inhibitors in this study, and corresponding target CDKs in mitosis cycle. Cell cycle associated proteins are placed on the representing ring, CDK7 and CDK9 are transcriptional regulators. Roscovitine inhibits CDK2, CDK7 and CDK9, while FUdR, the non-targeted therapy drug, inhibits TYMS. Other inhibitor molecules target CDK4 and CDK6.

1.5. Computational Approaches In Drug Target Determination

Computer-aided drug discovery/design methods have played a major role in the development of therapeutically important small molecules for over decades. Algorithms, softwares, and methodologies are commonly utilized in the study of protein functions, drug-protein interactions, and the physiological behavior of biomolecules [66]. This multidisciplinary approach contributes to advancing our understanding of biochemical systems and facilitates more informed decision-making in healthcare.

Traditional drug development from scratch to market costs approximately $2.558 billion over 10 to 15 years, and this large investment results in a drug market success rate of approximately 13% [67]. Rejection of potential drugs, especially during phase II and phase III clinical development, is associated with unexpected clinical side effects and cross-reactivity. These unexpected results may include proteins or genes that are potential candidates for disease, biological signaling pathways, disease-associated microRNAs, biomarkers, critical junctions in biological networks, or molecular functions [68]. This may be potentially related to insufficient knowledge about the drug target, adverse pharmacokinetic manifestations due to interactions with the target, or off-target side effects. Acknowledging this, there is a compelling need to modulate drug targets systematically in order to mitigate observed pathologies and achieve the desired biological responses. This modulation involves elucidating off-targets, particularly those associated with promiscuous kinase inhibitors, to refine drug specificity and enhance therapeutic efficacy [69]. In this regard, protein structural information is essential to study drug and protein interactions in atomic level.

Structural data in biochemistry elucidates the three-dimensional arrangement of biological molecules, offering crucial insights into their functions and interactions. Therefore, high quality structural data is important.

AlphaFold, a powerful deep learning system [70], has revolutionized structural biology by accurately predicting protein structures, advancing the understanding of complex molecular architectures and accelerating drug discovery efforts, especially when other sources of structure data is scarce. Structure-based methods are, in principle, require

essential target and ligand structure information. These approaches encompass ligand docking, pharmacophore, and ligand design methods.

Importantly, molecular dynamics simulations (MD), molecular docking simulations, inverse or drug screening, as well as data or system-oriented bioinformatic methodologies are potent approaches in drug target identification.

1.5.1. Molecular Dynamics Simulations

Molecular dynamics simulation (MD) computationally mimics biomolecular motions over time, offering insights into proteins, nucleic acids, and lipids' behaviors at an atomic or molecular level [71].

Molecular dynamics simulation (MD) is a computational technique used to simulate the motions and interactions of biomolecules, such as proteins, nucleic acids, and lipids, over time. MD simulations are widely used in biochemistry and biophysics research, as they provide a way to study the behavior of molecules at an atomic or molecular level that is difficult or impossible to achieve experimentally [72]. The behavior of a molecule is modeled as a collection of atoms, each of which is assigned a set of properties such as mass, charge, and position. The behavior of these atoms is then calculated using classical mechanics, based on the principles of Newtonian physics [73]. The interactions between atoms are typically modeled using force fields, which describe the potential energy of the system as a function of the positions of the atoms.

MD simulations can be used to study a wide range of molecular phenomena, including protein folding, protein-protein and protein-ligand interactions, membrane dynamics, and DNA replication. By simulating these processes, researchers can gain insights into the underlying mechanisms and develop new hypotheses about the behavior of biomolecules. One of the major advantages of MD simulations is their ability to capture the dynamic behavior of biomolecules. Unlike static experimental techniques such as X-ray crystallography or NMR spectroscopy, which provide a snapshot of a molecule in a particular conformation, MD simulations can track the motions of the molecule over time. This allows researchers to study the effects of thermal fluctuations, conformational changes, and other dynamic processes that are critical for understanding the behavior of biomolecules [72]. Another advantage of MD simulations is their ability to provide detailed structural information about molecular interactions. By tracking the motions of

atoms in a molecule, MD simulations can reveal the precise geometries of molecular interactions, such as hydrogen bonds, salt bridges, and hydrophobic interactions. This information can be used to design new drugs or other therapeutic agents that target specific molecular interactions with high specificity and efficacy [72]. Also, with the advancements in the computer science, especially in graphical processing unit (GPU) technologies, mathematical functions like fourier transforms and integral equations that drive the MD can be further accelerated [74].

1.5.2. Molecular Docking Simulations

Molecular docking is a computational method crucial in drug design, predicting how small molecule drugs bind to target proteins or biomolecules. It involves modeling the 3D structures of the target and the drug candidate, aiming to predict their binding affinity and mode of interaction based on chemical and physical properties [75]. In cancer research, molecular docking plays a crucial role in the design and development of new drugs, as it provides a way to predict the potential effectiveness of a drug before it is tested *in vitro* or *in vivo*. The process of molecular docking involves modeling the three-dimensional structure of the target protein or biomolecule, as well as the small molecule drug candidate. The goal is to predict the binding affinity and binding mode of the drug candidate to the target protein, based on the chemical and physical properties of both the drug and the protein.

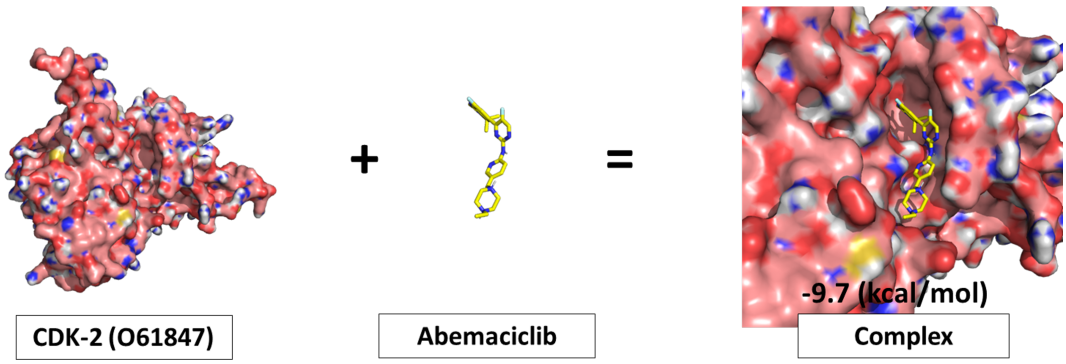


Figure 1.5: Abemaciclib and CDK-2 as an example to molecular docking. Molecular docking predicts the binding conformation of the ligand in the target protein and calculates an affinity.

Molecular docking has several advantages in cancer research. First, it can be used to identify new drug candidates that have the potential to target specific proteins or pathways involved in cancer growth and progression. By screening large databases of small molecule compounds against specific target proteins, researchers can identify potential drug candidates with high binding affinity and specificity.

1.5.3. Receptor Screening

In receptor screening (target screening), these methods play a crucial role in identifying and evaluating potential drug targets by analyzing the interactions between receptors and ligands at the molecular level.

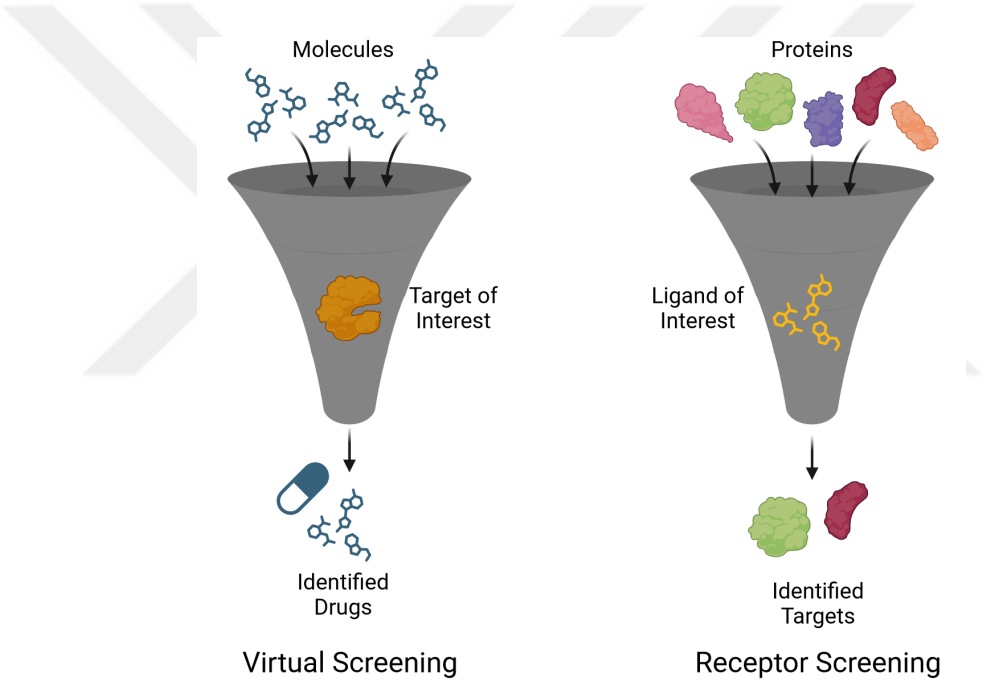


Figure 1.6: Schematic representation of virtual and receptor screening methodology. In receptor screening, a large library of proteins is individually screened with binding (docking) affinities for the drug of interest, and identified target candidates are further processed.

In cancer drug development, both in vitro and *in vivo* experiments are essential steps in evaluating the efficacy and safety of potential drug candidates. While *in vitro* and *in silico* experiments provide valuable insights into the mechanisms of action and toxicity of potential drug candidates, they cannot fully replicate the complex environment of

living organisms. Therefore, *in vivo* experiments are also necessary to evaluate the safety and efficacy of potential drug candidates. *In vivo* experiments are conducted using whole organisms, such as mice or rats, and allow researchers to study the effects of potential drug candidates on the entire organism, including interactions with various organs and tissues. *In vivo* experiments are often conducted in the later stages of drug development, after promising candidates have been identified through *in vitro* experiments.

1.6. Model Organism *Caenorhabditis elegans*

Model organisms play a pivotal role in advancing our understanding of biological processes and diseases, serving as invaluable platforms for research. These organisms, which encompass a diverse range including fruit flies, zebrafish, rats, and mice, have been extensively studied and established as crucial tools in unraveling the fundamental mechanisms of various biological conditions, notably cancer [76]. Gregor Mendel's groundbreaking studies on peas marked the inception of utilizing model organisms for biological investigations [77]. The term "model organisms" collectively refers to these species that offer unique advantages in terms of cost-effectiveness and logistical simplicity compared to human subjects. Notably, model organisms' shorter lifespans and accelerated development cycles enable real-time observation of cancer progression and responses to therapeutic interventions. Genetic manipulations in these organisms facilitate the development of specific cancer types or the study of particular cancer-related genes and pathways [78]. One crucial aspect of employing model organisms in cancer research is their role in drug screening. This step allows for the assessment of the efficacy and toxicity of potential cancer drugs before advancing to human clinical trials. The use of model organisms not only expedites the research process but also contributes to reducing the need for extensive animal testing and alleviating ethical concerns associated with human clinical trials [79].

Model organisms refers to species like fruit flies, zebrafish, rats, and mice, which have become fundamental in understanding various biological processes, particularly those related to cancer. While findings from model organisms may not always seamlessly translate to *Homo sapiens* (human), their integration with other research tools, including *in vitro* and *in vivo* experiments, patient-derived tumor samples, and computational approaches, ensures the relevance of discoveries to human cancer biology.

Caenorhabditis elegans, stands out as a significant model organism in biological and biomedical research. This roundworm species, residing in the soil and subsisting on bacteria, has proven invaluable for diverse research applications. With a concise lifespan of 2-3 weeks, *C. elegans* is ideal for studying aging and lifespan in animals. Its simple body structure, coupled with a manageable number of cells, makes it an easily tractable and manipulable model organism [80].

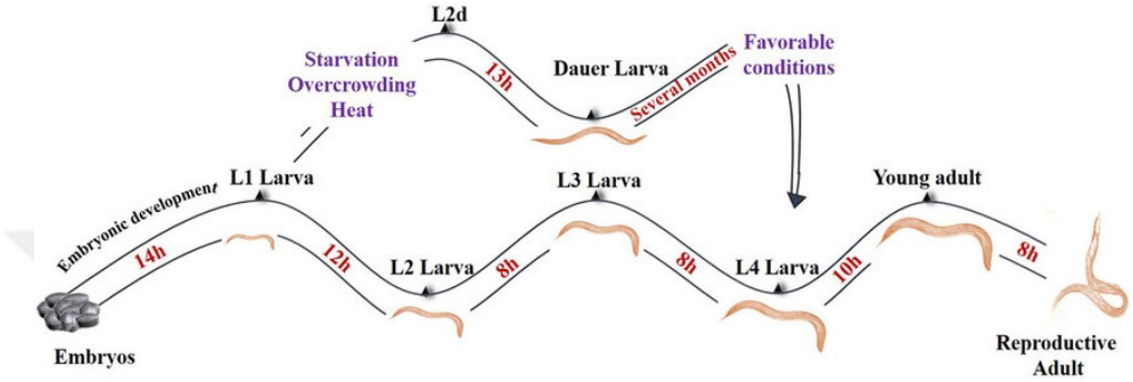


Figure 1.7: *Caenorhabditis elegans* life cycle at 20 °C. *C. elegans* undergoes distinct larval stages during its development, providing unique opportunities to study various aspects of biological processes. This feature enhances our understanding of how treatments may influence not only mature organisms but also those in earlier developmental phases (original figure from Bouyanfif et al. (2019) **[81]**).

One important aspect of *C. elegans* is that it has a defined number of cells that remains the same throughout adulthood [82]. This characteristic makes it a powerful model for studying drug effects and toxicology assessments. *C. elegans* also was extensively used to study the molecular mechanisms of cancer such as the Wnt pathway, which is involved in the regulation of cell proliferation and differentiation [83]. Additionally, *C.elegans* is used to identify compounds that enhance the efficacy of chemotherapy drugs in treating breast cancer, identify genes involved in tumor growth and metastasis in melanoma study the effects of a specific mutation in the *BRCA1* gene on breast cancer development [84].

1.7. Aim of This Study

In this thesis, it is hypothesized that receptor screening using molecular docking, molecular dynamics simulations, and systems biology methods can be used to understand and identify “off” or “secondary” targets of CDK-i drugs in humans and *C.elegans*. Utilizing this combined methodology, clear and replicable drug side-effect prediction approaches can be created, drug-response indicative *C.elegans* assays and/or models can be established, and potential outcomes in the integration process of this model organism in further studies can be highlighted.

2. MATERIALS AND METHODS

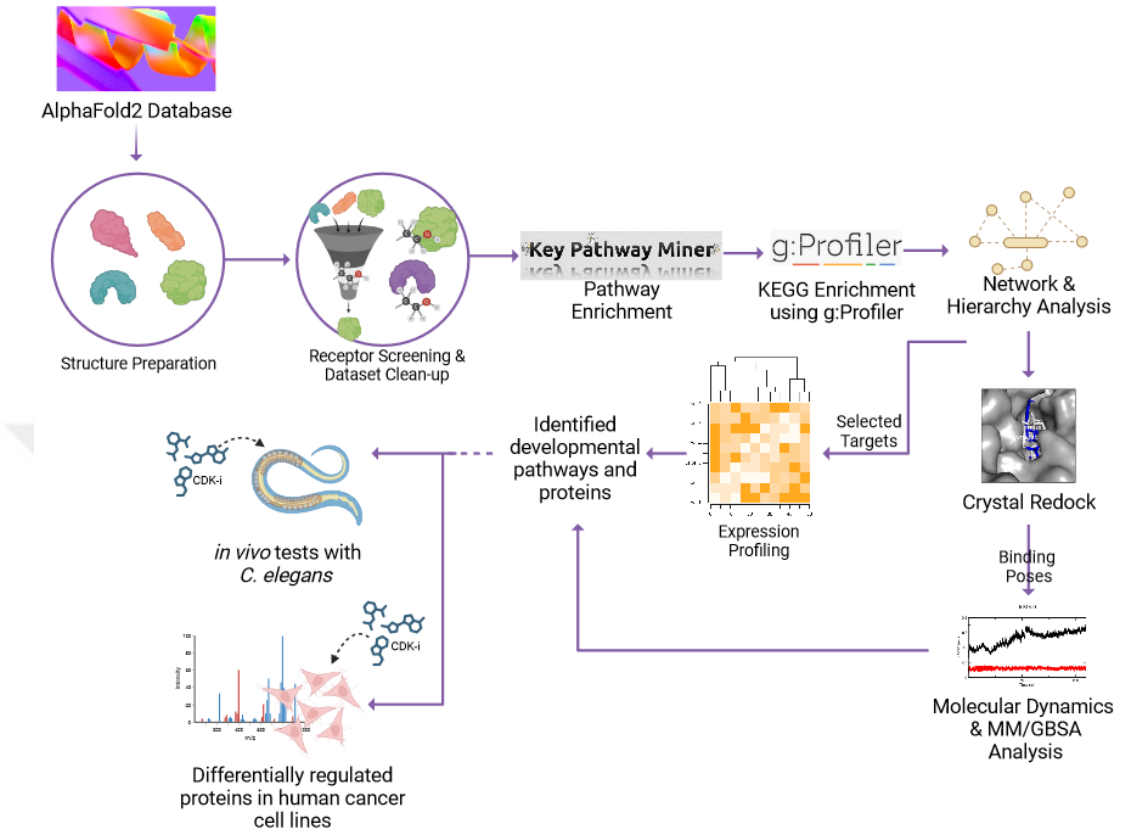


Figure 2.1: Graphical methodology workflow of this study. Human and *C. elegans* protein structure data were obtained from AlphaFold and parameterized. CDK-i drugs and FUDR were downloaded from their respective databases and parametrized. Subsequently, drug structures were blind-docked to every protein in both human and *C. elegans* proteomes, and affinities were determined sorted, and a cutoff reduction was applied. Following this, each drug-organism group (10 X 2 = 20 groups) was separately enriched using g:Profiler. The results from g:Profiler were then employed for further enrichments, either in cancer conditions (OncoDB data and cell lines) or in *C. elegans* stages (Expression Atlas), to evaluate non-canonical targets and pathways of CDK-i's. Additionally, *C. elegans* nematodes were treated with CDK-i drugs *in vivo* to highlight potential physiological effects.

2.1. Computational Methods

2.1.1. Sequence and Structure Similarity Assessments for CDKs

In order to compare and understand the selectivity of CDK inhibitors, human CDKs and *C. elegans* CDKs were assessed for similarities by comparing sequential and structural properties. Firstly, human and *C. elegans* CDK proteins were identified in the Uniprot [85] database using their respective identification codes. Human and *C. elegans* CDKs obtained from the database using the corresponding identifiers (Table: 2.1).

Table 2.1: Uniprot Identification Codes (ID) for human and *C. elegans* CDKs.

Human	**Uniprot ID**	***C. elegans***	**Uniprot ID**
CDK1	P06493	CDK-1	P34556
CDK2	P24941	CDK-2	O61847
CDK3	Q00526	CDK-4	Q9XTR1
CDK4	P11802	CDK-5	G5ECH7
CDK5	Q00535	CDK-7	G5EFV5
CDK6	Q00534	CDK-8	P90866
CDK7	P50613	CDK-9	Q9TVL3
CDK8	P49336	CDK-12	P46551
CDK9	P50750		

Obtained CDK IDs were used to perform multiple sequence alignment using the Clustal Omega [86] software version 1.2.4, with no iteration. The obtained alignments were visualized along with a percentage similarity matrix to uncover homological similarities. Then, the same CDKs for human and *C. elegans* were obtained from the Alphafold database using the corresponding IDs, and then structurally compared via the LovoAlign [87] tool. Structural similarities were represented with TM score [88].

2.1.2. Protein and Ligand Structure Preparation

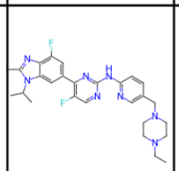
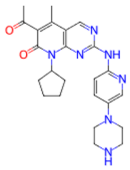
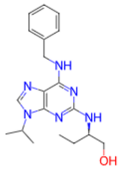
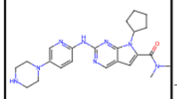
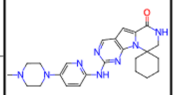
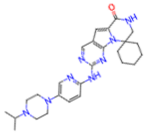
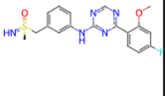
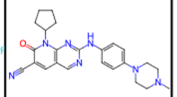
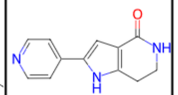
In order to screen proteome-wide proteins to identify targets of cyclin-dependent kinase inhibitor molecules, protein structure data was prepared.

Firstly, structural data for proteins were downloaded in bulk as ”Model Organism Proteomes” from the Alphafold Protein Structure Database [89] for both Human and *C. elegans*. Alphafold structure database represents structural prediction quality via pLDDT [70] score, and recommends not to interpret aminoacids with low pLDDT scores [90]. Therefore, after obtaining human and *C. elegans* proteomes, each downloaded protein structure data file was processed by means of text editing to remove amino acids with pLDDT scores below 50, and correspondingly saved for using in molecular docking calculations.

In the molecular dynamics phase of identified targets for Human, relevant proteins were downloaded from the RCSB Database [91] and used after removing unwanted entries such as water and salt. Missing parts were corrected using the PDBFIXER [92] tool when necessary.

Structural data for drug molecules of Abemaciclib, Palbociclib, Roscovitine, Ribociclib, Trilaciclib, Lerociclib, Atuveciclib, Narazaciclib, PHA-767491, and FUDR, were downloaded from ChEMBL [93], Drugbank [94], or PubChem [95] databases.

Table 2.2: Selected drug molecules and properties. FUDR is also included.

	Abemaciclib	Palbociclib	Roscovitine	Ribociclib	Trilaciclib	Lerociclib	Atuveciclib	Narazaciclib	PHA-767491	FUDR
Structure										
Primary Targets	• CDK4 • CDK6	• CDK4 • CDK6	• CDK2 • CDK7 • CDK9	• CDK4 • CDK6	• CDK4 • CDK6	• CDK4 • CDK6	• CDK9	• CDK4 • CDK6 • ARK5	• CDC7 • CDK9	• TYMS
Molecular Weight	506.61	447.54	354.46	434.55	446.56	474.61	387.44	429.53	213.24	246.19
Heavy Atoms	37	33	26	32	33	35	27	32	16	17
Rule of Five	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trade/Other Names	Verzenio	Ibrance	Seliciclib	-	G1T28	G1T38	BAY-1143572	ON123300	-	Floxuridine
Current Phase	Approved	Approved	Phase 2	Approved	Approved	Phase 3	Phase 1	Phase 1	Preclinical	Approved

2.1.3. Receptor Screening

Receptor screening or inverse screening methodology were employed in order to identify high affinity targets of the selected drugs(CDK-is and FUDR) within the human and *C. elegans* proteomes. All protein structural data files for Human and *C. elegans* were downloaded from the AlphaFold database and were processed to remove low-score amino acids, as described. After that, all protein structures were further processed using the OpenBabel [96] program (Linux, version 3.0) to merge nonpolar hydrogens and calculate partial charges using Kolmann [97] charges. The drug molecules to be used were also prepared for docking studies using the same approach except gasteiger [98] charges were applied. Each drug was docked to the protein structures in both the human and *C. elegans* proteomes using a blind docking approach [99] via Smina [100] program (a branch of Autodock VINA, Linux, version 2019) with an exhaustiveness value of at least 30. Upon screening, binding values of the drugs to all proteins were recorded, and these values were then ranked to prioritize the best-binding potential targets for each drug group. Primary targets of each drug were selected as cutoff values, based on the literature for human proteins. In *C. elegans*, orthologs or homologs of the human counterparts were selected as primary(canonical) targets for cutoff application.

Table 2.3: Selected primary targets of drugs for minimum affinity cutoff in human and *C. elegans*.

	Human	***C. elegans***
Abemaciclib	CDK6	CDK-4
Palbociclib	CDK6	CDK-4
Ribociclib	CDK6	CDK-4
Trilaciclib	CDK6	CDK-4
Lerociclib	CDK6	CDK-4
Roscovitine	CDK2/CDK7/CDK9	CDK-2/CDK-7/CDK-9
Narazaciclib	CDK4/ARK5	CDK-4
Atuveciclib	CDK9	CDK-9
PHA-767491	Cdc7/CDK9	CDK-9
FUDR	TYMS	TYMS-1

Each drug's primary target's binding affinities were assigned as the threshold for the ranked values. Proteins associated with bindings less successful than this threshold were removed from the list.

Additionally, when drug binding energies for CDKs were represented, obtained binding energy values were normalized by dividing by the total calculated heavy atoms in the drugs (Table 2.2, Heavy Atoms) then multiplied by 10 and represented as "*efficiency score*" to facilitate a comparability of binding energies among drugs.

2.1.4. Network and Gene Set Enrichment

After obtaining affinity lists of drugs both in human and *C. elegans* proteome, lists are called as datasets. Then, each dataset underwent protein-protein interaction enrichment using the KeyPathwayMiner [101] (KPM) tool to correct for missing or redundant proteins. Network enrichment methodology is an effective approach for identifying biological circuits and pathways [102], [103]. BioGrid [104] protein-protein interaction (PPI) data source was used in KPM. While enriching datasets based on their interactions, each dataset was also turned into a protein-protein interaction network by KPM. Subsequently, the enriched interaction network datasets were individually transferred to gProfiler [105] and analyzed by selecting data source as only "KEGG". Upon analysis, output was analyzed to illuminate the biological pathways represented by proteins in these datasets, focusing on KEGG Pathways (Kyoto Encyclopedia of Genes and Genomes) [106] terms. These findings were then separately integrated to each dataset's interaction network using Cytoscape 3.10 [107] (Microsoft Windows™) and the StringApp [108] plugin for visualizing highlighted pathways and contributing proteins.

2.1.5. Network Hierarchy Analysis

Highly connected proteins in the cell are the most important for its survival [109]. Therefore, in order to obtain further detail and reduction within the datasets, a hierarchical priority approach was utilized and the most potent targets were identified by ordering highly connected potential targets within each dataset.

Protein-protein interactions were obtained and visualized by clustering relevant nodes

within the KEGG terms. Clustered nodes are then hierarchically ordered to identify proteins (nodes) with the highest interactions (degree) within these clusters. After ordering and clustering all datasets individually, all the graphs were combined in a single diagram for human and *C. elegans* seperately. While selecting the most potent targets after hierarchical clustering, redundancy was avoided and relevance to corresponding biological enrichment was preferred. Additionally, through the same plugin, STRINGDB [110] enrichment was performed for *C. elegans* datasets, obtaining Worm Phenotype Ontology (WPO) [111] terms and revealing phenotypic significance indications in these datasets.

2.1.6. Expression Profiling for Human and *C. elegans* Targets

Gene expression levels of potentially identified targets for human and *C. elegans* from previous steps were profiled to understand possible druggable target levels and enrich the interpretation of corresponding pathways or physiological differences. For human datasets, original data from the Cancer Cell Line Encyclopedia [112] were downloaded through the Expression Atlas web server [113] with accession code E-MTAB-2770 [114]. Gene expression levels of identified potential target proteins were extracted from this data by selecting specific commonly used cell lines. For expression levels of selected targets in cancer sample sites, OncoDB data were downloaded, curated and visualized. OncoDB data comprised of original data from The Cancer Genome Atlas Pan-Cancer Analysis Project (TCGA) [115], and represented as bulk for cancer tumor samples [116]. For *C. elegans*, data with the accession E-MTAB-2812 [117] were downloaded from the Expression Atlas web server, and the gene expression levels of potentially identified target proteins were visualized throughout all larval stages.

During visualizations, heatmaps were created to highlight the quantities between conditions by representing gene expression levels of selected targets with the highest degree as ”+ or - fold-change(fc)” compared to the average of all biological states and scaled during visualization.

2.1.7. Molecular Dynamics Simulations and Binding Free Energies

Some of the most potently identified targets from the datasets of human and *C. elegans* were simulated with their corresponding drug molecules for drug-protein binding conformations, aminoacid interactions and binding free energy evaluations. Identified target structures specific to *C. elegans* were stabilized through preliminary MD simulations using the structures obtained from the AlphaFold database, with amino acids unaltered this time. Identified target structures specific to human were downloaded from the RCSB database [91], fixed for missing residues, cleaned and stabilized through preliminary MD simulations. Additionally, human CDK6 protein with abemaciclib complex (PDB: 5L2I) was also included to the analysis workflow for comparability.

Selected protein names and corresponding PDB codes from human and *C. elegans* datasets are listed in Table 2.4.

Table 2.4: Protein structure identifiers for human and *C. elegans*.

Human	**PDB ID**	***C. elegans***	**Alphafold ID**
CDK6	5L2I	CDK-4	Q9XTR1
PRKCA	4RA4	LET-756	Q11184
PEX5	2C0L	PMK-1	Q17446
ADSS2	2V40	MEK-2	Q10664
HADHA	5ZQZ	DAF-2	Q968Y9
NOS2	2NSI		
GAPDH	1U8F		
SHMT2	6M5O		
EPRS1	7F99		
MAPK14	1KWP		

In molecular dynamics (MD) simulations, open-source GROMACS 2022.4 [118] software was used. Protein structure files from either Alphafold server or RCSB were downloaded and prepared by removing the water, ion and any other atoms other than in the file, as explained. After this step, topologies were created with AMBER03 force

field and protein structures were solvated in TIP3P water model while aiming to achieve 0.15 mM of calcium and sodium ion concentration in the system. After that, the system was energy minimized via the steepest-descent algorithm to achieve energy convergence of <1000 kJ mol-1 nm-1. The Energy-minimized system was then equilibrated in NVT ensemble and NPT ensembles , respectively. After this step, production MDs were run for varying durations, and simulated with 2 fs time steps. Neighbor searching was performed every 10 steps and the PME algorithm was used for electrostatic interactions.A single cut-off of 1 nm was used for Van der Waals interactions. Temperature coupling was done with the V-rescale algorithm. When simulating solute proteins or complexes, 310.15 K and 293.15 K thermostat temperature was set for human and *C. elegans*, respectively. Pressure coupling was done with the Parrinello-Rahman algorithm, and graphical processing unit (GPU) utilization via CUDA Runtime libraries [119] were preferred when possible. Simulations were run for at least 30 ns in all cases, and stability assessments were done using root-mean square deviation (RMSD) analysis.

When drug-protein (ligand MD) complexes were simulated for molecular dynamics, drug molecules were parameterized for the AMBER force field using the ACPYPE [120] server and then placed into relevant target structures prior to simulation workflows. At least 3 replicate trajectories for protein-drug complex simulations were produced from random Maxwell–Boltzmann distribution samples for each binding free energy calculations.

After producing the drug-protein complex trajectories, binding free energies were calculated. For this, firstly, each system was fixed for periodic-boundary conditions and transformed. Then MM/GBSA (molecular mechanics generalized Born surface area) calculations were accomplished using the gmx_MMPBSA tool [121] with GB (generalized-Born) approach and Generalized Born ESURF were calculated using LCPO surface areas. After calculating the free energies of binding from the complex trajectories, the delta was obtained as Complex - Receptor - Ligand , and the average Delta of frames was represented.

Additionally, after obtaining drug-protein complex trajectories, representative structures were obtained via gmx cluster tool within the GROMACS suite, and these obtained clusters were then visualized using MOE 2022 (Chemical Computing Group) for drug

binding diagrams. For diagram visualizations, complexes with least root-mean square deviations (RMSD) were selected.

2.2. Mass Spectrometry Analysis of Human Cell Lines

In order to distinguish protein-level differences and regulations resulting from CDK-i treatments, PANC-1 and MIA PaCa-2 cells were treated with abemaciclib and palbociclib, 2D gel electrophoresis coupled with MALDI-TOF protein mass spectrometry and peptide mass fingerprinting (PMF) methodology were utilized previously.

Simply, PANC-1 (CRL-1469), and MIA PaCa-2 (CRL-1420), cell lines were purchased from American Type Culture Collection (ATCC, Rockville, MD, USA). The cells were cultured in Dulbecco's modified Eagle's medium (DMEM; GIBCO- Life Technologies, Carlsbad, CA) supplemented with 10% fetal bovine serum (Pan Biotech GmbH, Aidenbach, Germany) and 1% penicillin/streptomycin (GIBCO, Invitrogen Co) and incubated in 37 °C with 5% CO2 (HeraCell 150i ThermoLab systems, Beverly, MA, USA). Abemaciclib and palbociclib were dissolved in dimethyl sulfoxide (DMSO) at an initial stock concentration of 10 mM and stored as aliquots at -20°C.

Cells were seeded at a density of 1×10^6 cells in 100-mm dishes and treated with a selected dose of abemaciclib and palbociclib for 24h. Proteins were extracted using RIPA buffer (Thermo Scientific, USA), and protein concentrations were determined with a modified Bradford assay (Bio-Rad, USA). Before the first-dimension separation, protein samples were cleaned up using the ReadyPrep 2D Cleanup Kit (Bio-Rad), to eliminate any contaminating non-protein agents. The precipitated proteins were resuspended in 2DE lysis buffer containing 7 M urea, 2 M thiourea, 4% CHAPS, 30 mMTris pH 8.5, and 1x protease inhibitor cocktail. Protein extracts (500 µg each) were loaded onto 11 cm non-linear pH 3-10 immobilized pH gradient (IPG) strips (Bio-Rad, USA) by passive rehydration. First dimension separation based on isoelectric points was performed with a Protean isoelectric focusing cell (Bio-Rad) using the following conditions: 20 min at 250 V with rapid ramp, 2 h at 4,000 V with slow ramp, and 8,000 V with rapid ramp until a total of 25,000 V/h was reached, at 20˚C. Following isoelectric focusing, strips were washed with buffer I (6 M Urea, 375 mM Tris-HCl pH 8.8, 2% SDS, 20% glycerol, 2% w/v DTT) for 30 min and then with buffer

II (6 M Urea, 375 mM Tris-HCl pH 8.8, 2% SDS, 20% glycerol, 2.5% w/v iodoacetamide) for 30 min at room temperature. The second-dimension separation based on molecular weights was achieved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE; 12%) using the Criterion Dodeca gel running system (Bio-Rad, USA). Gels were stained with colloidal Coomassie Blue and visualized by VersaDoc4000 MP (Bio-Rad) using QuantityOne software (Version 4.6.7, Bio-Rad).

For DIGE experiments, the proteins were resuspended in DIGE buffer containing 7 M urea, 2 M thiourea, 30 mM Tris, and 4% CHAPS and then labelled with CyDye minimal dyes using the CyDye™ DIGE Fluor Minimal Labeling Kit (Lumiprobe, USA); the experiments were carried out in the dark. In brief, 50 µg of protein samples were used in each experiment. The pH of the protein extracts was adjusted to 8.5, and then extracts were incubated with the dyes Cy2, Cy3, and Cy5 at 4˚C for 30 min. Labelling reactions were stopped by adding 10 mM lysine. Labeled protein samples were then combined, and a standard 2-DE experiment was performed.

The Coomassie-stained 2-DE gels were analyzed with PDQuest Advance 2-DE Analysis Software (Version 8.0.1 build 055, Bio-Rad), which provided the tools for spot intensity calibration, spot detection, and background subtraction. Stain speckles were filtered, and the standardized areas of interest from all gels were warped and matched, and the quantity of each spot was normalized by the total valid spot intensity using the linear regression model. The statistical significance of image analysis was determined by the Student's t-test (statistical level of $p<0.05$ was considered significant). Protein spots with differences in expression more than 2-fold were selected and excised from gels using ExQuest Spot Cutter (Bio-Rad). DIGE gels were also visualized with VersaDoc MP4000 (Bio-Rad) using three different light sources.

In-gel digestion was performed using an in-gel tryptic digestion kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Digested peptides were then desalted and concentrated with C18 ZipTips (Millipore, Burlington, MA, USA). Peptides were eluted with a matrix solution containing 50% (v/v) acetonitrile (Sigma Aldrich, St. Louis, MO, USA) 0.1% (v/v) trifluoroacetic acid (Sigma Aldrich), and 10 mg/ml alpha-cyano-4- hydroxycinnamic acid (α-CHCA) (Sigma Aldrich), and were spotted onto MALDI target plate. An AB SCIEX TOF/TOF 5800 instrument (AB Sciex, Framingham, MA, USA) was used to generate peptide mass fingerprints (PMFs)

for protein identification. PMFs were searched in the MASCOT database (Matrix Science, Boston, MA, USA) using ProteinPilot software 4.0.8085 revision 148085 (AB Sciex). Search parameters included were; human as taxonomy, trypsin as the enzyme, ±50 ppm as the precursor tolerance, ±0.4 Da as the MS/MS fragment tolerance, and carbamidomethyl and MetOxidation as the variable modifications. Only hits with $p<0.05$ value were accepted for protein identification.

2.3. *C. elegans* Methods

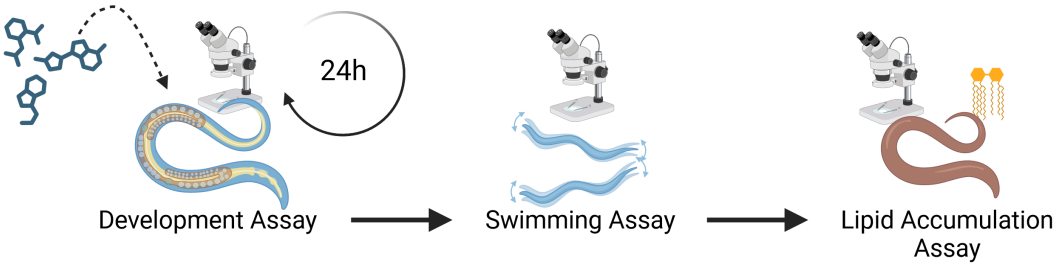


Figure 2.2: Graphical abstract for utilized methodology in *C. elegans* experiments. Nematodes were treated with drugs and imaged every 24 hours, after that, swimming assays and lipid accumulation assays were subsequently conducted.

2.3.1. Culturing

For maintenance and breeding of the *C. elegans* organisms, *E. coli* OP50 bacteria were spread on the Nematode Growth Medium (NGM) and allowed to grow overnight at 37°C to provide nutrition, then nematodes were transferred and grown on NGM and as previously generalized by Sydney Brenner [122]. Cholesterol was also added to the NGM in order to trigger the egg hatching and cuticle motility. NGM plates were prepared in advance on a large scale and stored at +4°C until ready for use. One day before transferring worms to a new growth environment, *E. coli* OP50 was added to NGM plates and incubated at 37°C overnight to allow bacterial growth. The following day, the worms were transferred to the nutrient medium. Except when treated with bleach solution, being transferred, or being imaged in assays, all *C. elegans* individuals were maintained in an incubator at a constant temperature of 20°C, within a laboratory

with relatively stable 16-18°C. Within this thesis, all *C. elegans* based methodologies were conducted using the N2 Bristol strain of the *C. elegans*. N2 Bristol strain were provided by the Caenorhabditis Genetics Center(CGC), which is funded by the NIH Office of Research Infrastructure Programs (P40 OD010440).

2.3.2. Cryopreservation

C. elegans nematodes were also cryopreserved additionally to continuously bred live stocks. For cryopreservation, firstly, freshly starved L1-L2 populated *C. elegans* plates were washed with the S-buffer and transferred in a sterile test tube. After that, equal volume of S Buffer + 30% glycerin were add and mixed. Next, animals within this mix were transferred to cryovials and labeled, and immediately stored in -80°C using Mr. Frosty cryopack.

When required, cryo-stored nematodes were revived by thawing the vials at room temperature until all ice had turned to liquid and pouring the contents onto an bacteria seeded NGM plate.

2.3.3. L1 Larvae Synchronization

When the worms reached the reproductive stage and began laying eggs, they were collected in a 15 mL centrifuge tube after being washed with dH_2O. To remove bacteria, a washing step was performed by centrifuging the tube at 1500 *g* for 2 minutes, discarding the upper liquid. Approximately 7-8 mL of dH_2O was added to the worms, followed by another centrifugation step. After discarding the upper liquid, 5 mL of the bleaching solution was added to the worms, and they were continuously vortexed for a maximum of 6 minutes. During this period, worms were observed under a stereo microscope, and the process was terminated when they were sufficiently disrupted.

To minimize the effect of the bleaching solution, approximately 7-8 mL of M9 buffer was quickly added. The tube was then centrifuged at 1500 *g* for 2 minutes. After discarding the upper liquid, 7-8 mL of M9 buffer was added, followed by another centrifugation step. This washing procedure, aimed at removing remnants of the bleaching solution, was repeated at least 3 to 5 times. Meanwhile, the now-collected worms were disrupted, and eggs were released into the buffer. To facilitate the emergence of worms from the

eggs, they were continuously vortexed at 800 RPM overnight at 20°C in an incubator. The following day, L1 larval stage worms had hatched from the eggs and were ready to use in the assays.

2.3.4. Dauer Larvae Synchronization

Dauer larvae synchronization method was used to obtain dauer larvae. In order to observe the effects on worms in the dauer phase, which is known to exhibit significant differences from their normal metabolic and physiological states [123] individuals of *C. elegans* were placed in a nutrient-rich environment 10-15 days prior to the intended use, following the generalized method by Cassada and Russell [124]. On the day of usage, worms were collected from this environment and treated with 1% SDS to selectively obtain Dauer larvae.

Firstly, worms were washed off from the plates, pelleted using a centrifuge, for 2-3 times and resuspended in 1% SDS. After this step, worms were incubated in 1% SDS for 30 minutes with gentle agitation. After, the remaining worms were washed 1-5 times with sterile water to remove all SDS in the media.

At this point, the tube contains living dauer larvae and the carcasses of non-dauer larvae.

2.3.5. Development Assays

The drugs abemaciclib, palbociclib, roscovitine, and FUDR were individually tested at a single dose (1μM in NGM) using the methodology adapted from the Wang and colleagues [125]. The developmental effects, swimming speed, and lipid accumulations potentially induced by the drugs were observed. To achieve this, *C. elegans* larvae were synchronized and grown until the L1 stage, and then transferred to bacteria seeded treatment and control plates. Worms were then observed under a dissection microscope coupled with a camera at 24-hour intervals in the environment containing the respective drugs. The observations continued either until the nutrient was completely depleted or until they exhibited egg-laying behavior.

To observe the effect of drugs on the development of larvae in the dauer phase, *C. elegans* individuals were placed in a nutrient-rich environment 10-15 days prior and were collected from this environment and treated with 1 % SDS to obtain only the

dauer larvae before experiments. Obtained dauers were then transferred to bacteria seeded control and treatment plates and then observed in the environment containing the respective drugs at 24-hour intervals until the nutrient was completely depleted or until they exhibited egg-laying behavior.

Captured images were used to quantify body sizes using ImageJ software [126] (NIH, USA) and recorded in pixels squared (px^2).

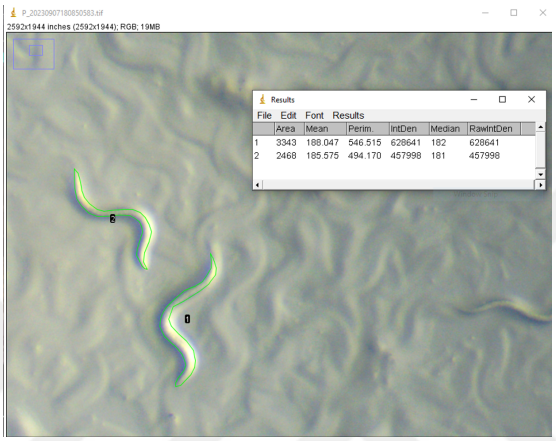


Figure 2.3: Outlines of the individual *C. elegans* worms were selected for data acquisition.

2.3.6. Swimming Assays

After monitoring their growth in developmental assays, nematodes were directly collected using the M9 buffer. Subsequently, they were transferred onto a microscope slide and recorded under a dissection microscope equipped with a Swiftcam SC500 camera and the relevant software for 1 minute. Following these recordings, visual inspection involved counting the periodic waves of the worms' heads, and the movement of the body waves per minute was noted. Results were represented in comparison to control groups.

2.3.7. Lipid Accumulation Assays

After monitoring the growth and swimming capacities of worms, they were collected, fixed, and stained with Oil Red O to observe neutral lipid contents.

Firstly, following development and swimming assays, worms were recollected with 1xPBS in a microfuge tube. Centrifuged at 400 *g* for 2 minutes, the supernatant was aspirated to cleanse excess bacteria with 1x PBS until it appeared clean. Subsequently, 120 µL of ice-cold methanol was added to the nematodes, and they were incubated on ice for 2 minutes.

After fixation, the nematodes were centrifuged for 1 minute at 1000 *g*, and the supernatant was removed. Following this, 600 µL of 60% isopropanol was added and incubated at room temperature for 15 minutes, allowing the worms to dehydrate. Next, 500 µL of Oil Red O working solution was added to the dehydrated worms. The worms in Oil Red O working solution were incubated overnight at room temperature in an orbital shaker with mild agitation.

On the next day, 1x PBS with 0.1% Triton-X100 solution was added to fill the microfuge tubes and allowed to settle for 3-5 minutes. This step was repeated three times to clear excess stain. Subsequently, stained worms were transferred to clean NGM plates, and their images were captured using a camera-mounted stereo microscope.

Later, utilizing the ImageJ program, Integrated Density values were normalized with the Optical Density (OD) equation and recorded.

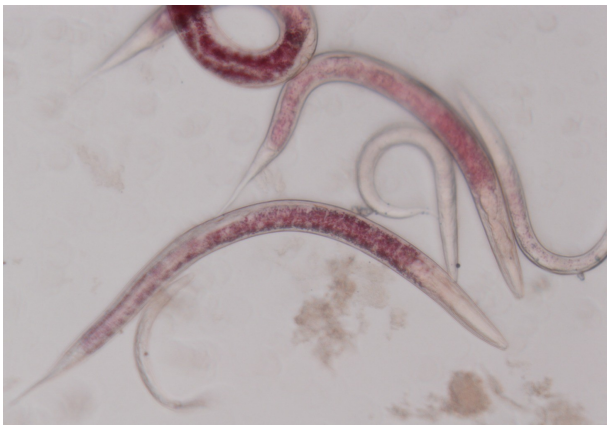


Figure 2.4: Image of Oil Red O stained *C. elegans* worms.

2.4. Laboratory Materials

2.4.1. Chemicals, Buffers and Solutions

Table 2.5: Chemicals used in this study.

Chemical	Provider	Cat/Lot
Agar Technical	Biolife	52X009
Peptone	Biolife	27P039
LB Agar	BioShop	0L67747/ LBL406.1
LB Broth	BioShop	0K67307/ LBL405.1
CaCl	Neofroxx	10035-04-08
$MgSO_4$	Merck	1.05886
K_2HPO_4	TEKKIM	TK.930147
KH_2PO_4	TEKKIM	TK.200990
Na_2HPO_4	Merck	106586
NaCl	Kimyalab	V54903.901H
Cholesterol	Biobasic	CD0122
NaOH	Kimyalab	V06498.901
Sodium Azide	Biobasic	DB0613
Glycerol	Isolab	LR0690709AJW
Oil Red O	Isolab	955.D01
SDS	AFG Bioscience	368131
Household Bleach	Migros	-

Table 2.6: Chemicals used in *C. elegans* developmental assay.

Chemical	Provider	Cat/Lot
Floxuridine	Cayman	14154
Abemaciclib	Selleckchem	S5716
Palbociclib	Selleckchem	S1579
Roscovitine	Selleckchem	S1153

2.4.1.1. Nematode Growth Medium

In order to prepare Nematode Growth Medium (NGM), 2.5 g of peptone, 17 g of agar, 3 g of sodium chloride (NaCl) and 1 L of distilled H_2O (dH_2O) were mixed in a flask and autoclaved at 121°C for 15 minutes. Subsequently, the mixture's temperature was allowed to drop to approximately 65°C with magnetic stirring. After autoclaving, 1 mL of 1M $CaCl_2$, 1 mL of 1M $MgSO_4$, 25 mL of 1M KPO_4 buffer mixture (pH:6) and 1 mL of cholesterol (5 mg/mL in ethanol) were added, with approximately 5 minutes between each addition to prevent crystallization.

To prepare the KPO_4 buffer mixture, 108.3 g KH_2PO_4 and 35.6 g K_2HPO_4 were mixed in 1 L of dH_2O and autoclaved.

2.4.1.2. M9 Buffer

For M9 buffer preparation, 3 g of KH_2PO_4, 6 g of Na_2HPO_4, and 5 g of NaCl were dissolved in 1 L of H_2O and autoclaved for 15 minutes. Then, 1 mL of 1 M $MgSO_4$ was added and prepared M9 Buffer stored in room temperature.

After 2 week, check for visible contamination before use.

2.4.1.3. KPO_4 Mix Buffer

1M KPO_4 Mix Buffer was prepared by mixing 108.3 g KH_2PO_4 and 35.6 g K_2HPO_4 and completing the mix to the 1 L with dH_2O. After that, the buffer was autoclaved.

2.4.1.4. Bleach Solution

10 mL of Bleaching solution can be prepared by adding 2 mL of household bleach, 0.5 mL of 10M NaOH and completing this mix to 10 mL with dH_2O in a vessel.

Bleach solution should be kept away from the sunlight and should be prepared fresh every time.

2.4.1.5. Oil Red O Staining Solution

Oil Red O dye was used for staining neutral lipids in the *C. elegans*. Oil Red O solutions were prepared in Stock and Working solutions.

Oil Red O Stock Solution

Firstly, 0.5 g ORO powder was dissolved in 100 mL of 2-propanol and incubated at 85°C for 2 hours in a water bath. After incubating, this stock solution were filtrated using 0.22 µm filter and stored away from direct light.

Oil Red O Working Solution

30 mL of Oil Red O stock solution was diluted by 20 mL of dH_2O.

2.4.2. Laboratory Equipment

Various instruments were used for *C. elegans* based methodologies.

Table 2.7: Instruments used in this study.

Instrument	Make	Model
Stereo Microscope	SOIF	SZM45-T2/L
Digital Microscope	Dino-Lite	Dino-Lite Plus
Microscope Camera	Swiftcam	SC500
Cooling Incubator	Elektro-Mag	M7050R
Biosafety Cabinet	Grafen Scientific	GMS90
Magnetic Stirrer	DLAB	MS-H380-Pro
Balance	Weightlab Instruments	WSA-224
Vortex	BIOSAN	Bio Vortex V1
Multi Speed Vortex	BIOSAN	MSV-3500
Water Bath	BIOSAN	WB-4MS
Incubator	Grafen Scientific	GS-OVN-120
Centrifuge	Nüve	NF1200R
Autoclave	Nüve	NC-150
Freezing Container	Thermo Scientific	Mr. Frosty

2.5. Statistical Analysis, Graphics, and Data Processing

Statistical representation graphs were created with Graphpad Prism 9 with pairwise comparisons using ANOVA and Tukey with always at least 3 replicates. Data graphics were created with RStudio 2022.7 (R lang 4.0.3) using “pheatmap”, “dplyr”, “ggplot2” and “ggridges” packages. Color scheme was obtained from “wesanderson” color package. Structural images were created using PyMOL Molecular Graphics System 2.1.0 [127]. Infographics were created using BioRender [128] unless stated or cited. Text-based data processing was done using BASH shell and Microsoft Excel.

3. RESULTS

In order to understand variation, selectivity, possible other targets, and consequential side effect and reusability properties of cyclin-dependent kinase inhibitor drugs, CDK similarity assessments, receptor screening, functional and network enrichments, molecular dynamics simulations, as well as *in vivo C. elegans* physiological evaluation studies were conducted.

3.1. Similarities Between CDKs in Human and *C. elegans*

Pharmacological drugs or small molecules, as well as macromolecules can effect multiple molecular targets, and this promiscuity can be beneficial or detrimental [11]. Firstly, to understand the basis behind selectivity of the CDK inhibitors, human CDKs and *C. elegans* CDKs were assessed for similarities. Human and *C. elegans* CDKs were aligned, and percentage similarities were obtained.

Multiple sequence alignments revealed that human and *C. elegans* CDKs have important sequence differences, and are mainly unaligned on the N and C termini (Figure: 3.1). In human, highest sequence similarity was between CDK2 and CDK3 (76.4%), and CDK4 and CDK6 (70.3%) (Figure:3.2 -B). In *C. elegans*, the highest sequence similarities were between CDK-1 and CDK-5 (47.1%) (Figure:3.2 -C). When similarities between organisms were considered, the highest similarities were between human CDK5 and *C. elegans* CDK-5 (73.6%), CDK7 and CDK-7 (58.6%), CDK9 and CDK-9(53.2%)(Figure:3.2 -A).

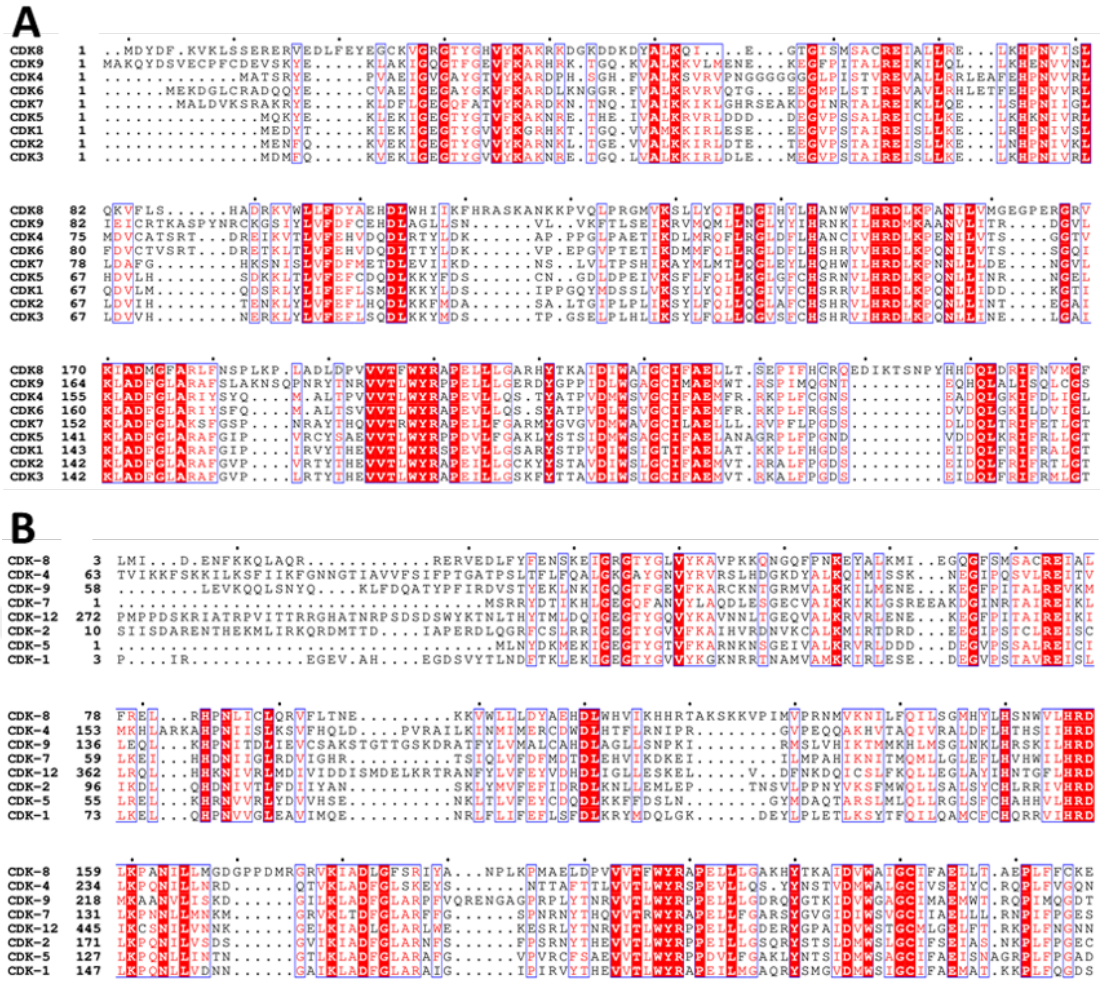


Figure 3.1: Multiple sequence alignments of human Cyclin-Dependent Kinases (**A**) and *C. elegans* Cyclin-dependent Kinases (**B**). Red highlighted white texts show identical residues, boxed sections indicate conserved regions. Leading and trailing unaligned sequences were trimmed.

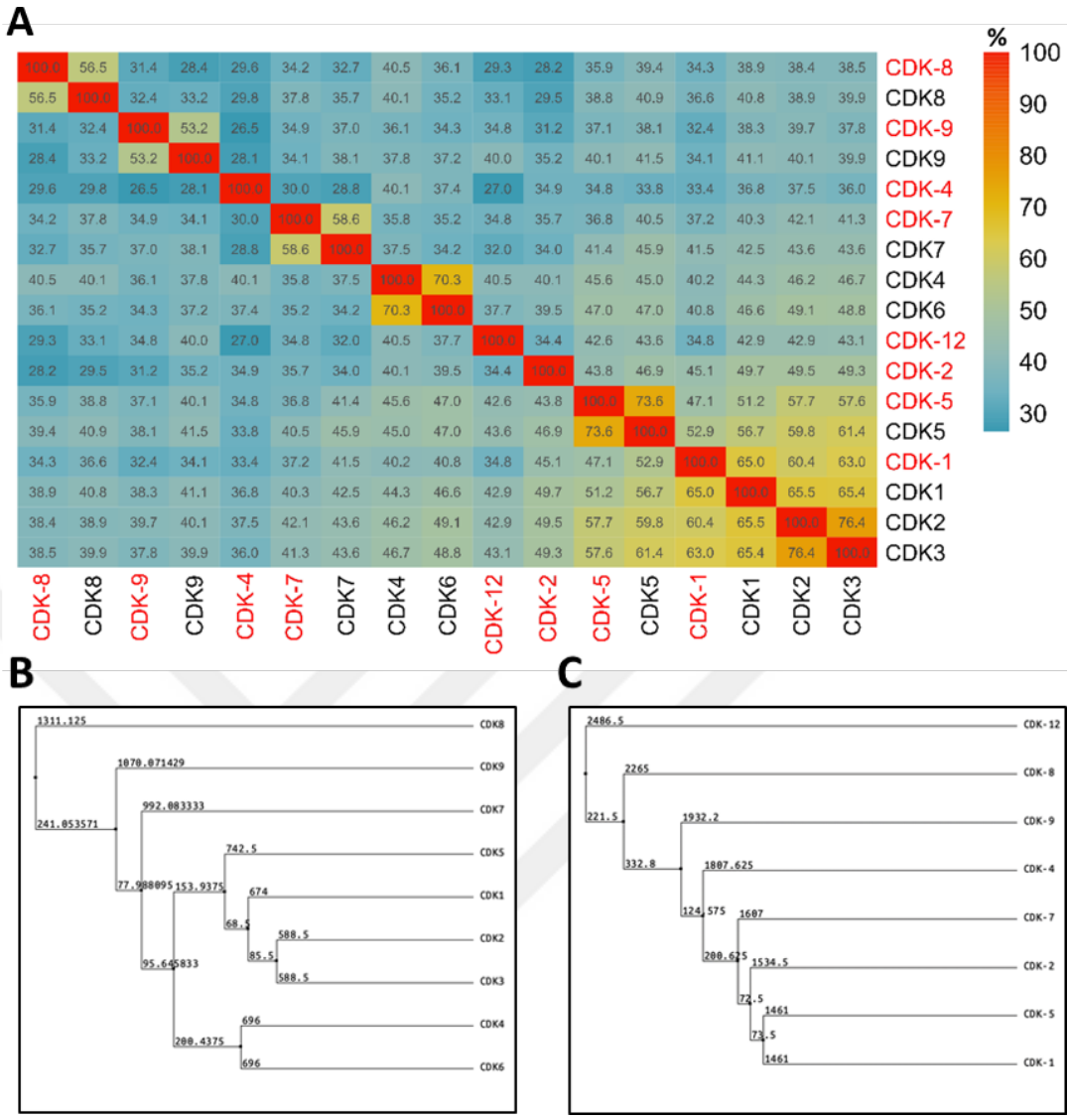


Figure 3.2: Sequence identity percentage matrix (**A**) and phylogenetic tree for human (**B**) and *C. elegans* CDKs (**C**). Protein names in black and red indicate human and C. elegans CDKs, respectively.A higher percentage (towards red) indicate a higher identity. **A**: CDK-5,CDK-1,CDK1, CDK2 and CDK3 shown varied similarities between, while CDK4-CDK6, CDK5-CDK-5 and CDK1-CDK-1 shown relatively higher similarities. **B**: Phylogenetic tree of Human CDKs. CDK4 and CDK6, CDK2 and CDK3 are highly similar. **C**: Phylogenetic tree of *C. elegans* CDKs. CDK-1 and CDK-5 are highly similar.

The structure of proteins is determined by their sequences [129], and similar sequences will result in similar structures. However, proteins with dissimilar sequences can also have similar structures [130]. In order to assess the presence of this condition and highlight the possible drug binding similarities of human and CDKs, CDKs are

structurally aligned using the LovoAlign tool [87]. In human, CDKs pairs with highest structurally similarity was CDK1-CDK3 (0.94), CDK1-CDK2(0.93), CDK2-CDK3(0.93) and CDK4-CDK6(0.90). In *C. elegans* the highest structurally similar CDK pairs were CDK-7-CDK-1 (0.81), CDK-5-CDK-7(0.80) and CDK-5-CDK-1 (0.78). When structural similarities of CDKs between organisms were considered, the highest similarities were between Human CDK5 and *C. elegans* CDK-5 again (0.99), CDK2-CDK-5(0.89) and CDK1-CDK-1(0.88) (Figure 3.3).

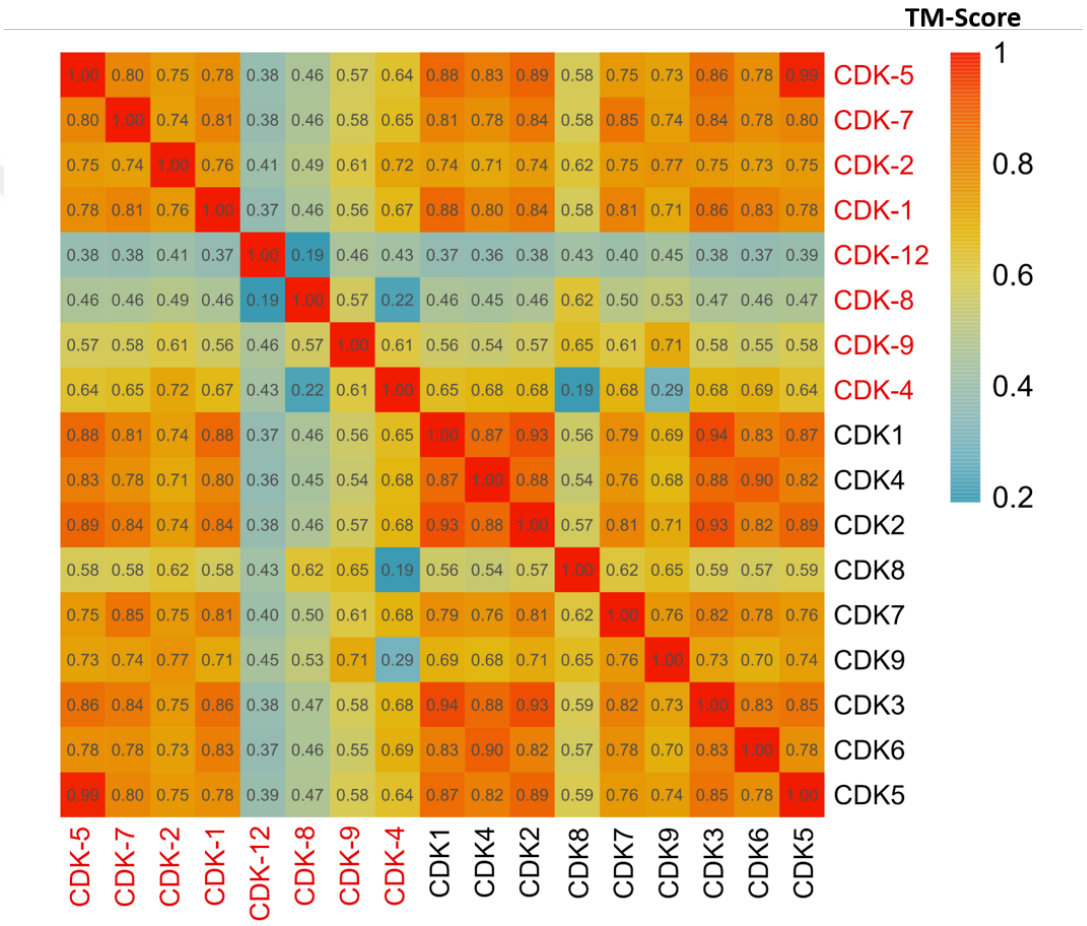


Figure 3.3: Combined structural similarity matrix for human and *C. elegans* CDKs. Red colored names indicate C. elegans CDK proteins. Human and *C. elegans* CDKs show relatively low structural similarities between, except CDK5 and CDK-5.

Additionally, when human and *C. elegans* CDK groups were superimposed, it was observed that proteins are primarily structurally aligned on conserved kinase domains (Figure: 3.4). Most structurally similar CDK pairs of human and *C. elegans*, CDK5 and CDK-5(Figure: 3.3) aligned perfectly while showing relatively low sequental similarity with 73.6% (Figure: 3.5).

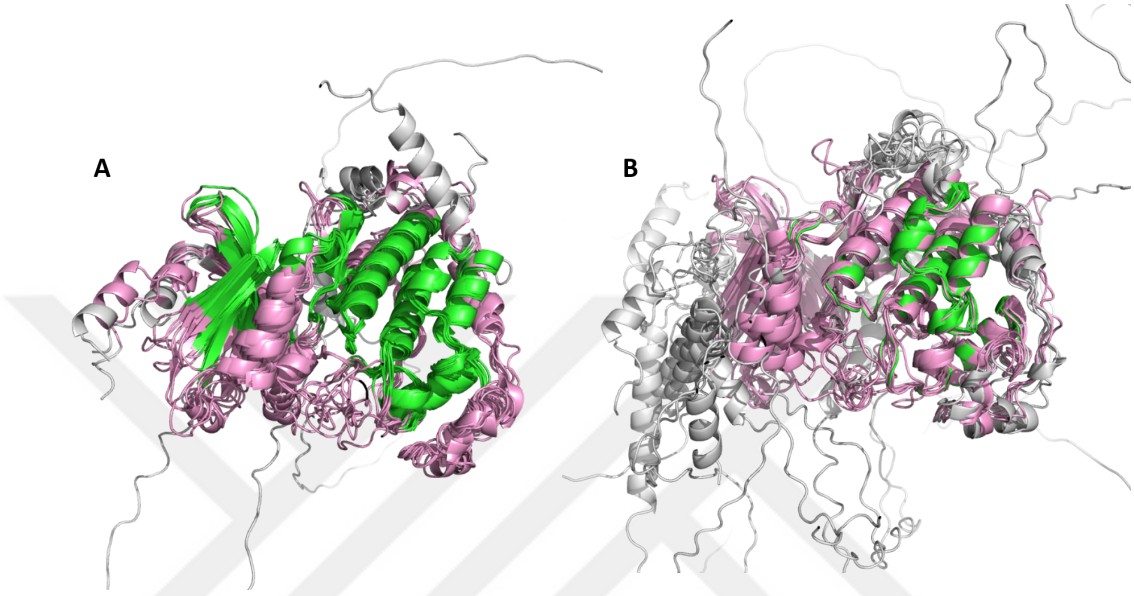


Figure 3.4: Superimposition of Human (**A**) and *C. elegans* (**B**) CDKs. Regions in pink indicate conserved regions in alignment while green regions indicate identical residues in alignment.

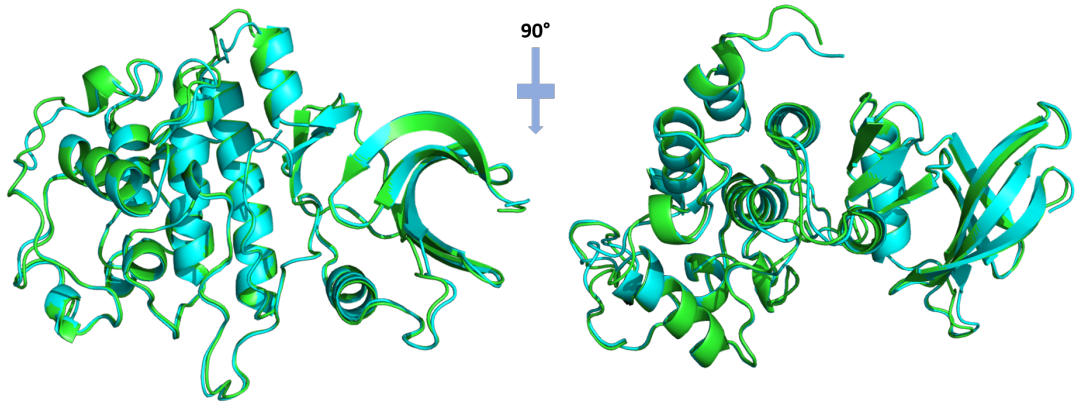


Figure 3.5: Superimposition of Human and *C. elegans* Cyclin-dependent Kinase 5. Green colored cartoon representation shows *C. elegans* CDK-5, while blue color shows human CDK5. The structure rotated 90° at the vertical axis in right-hand side of the figure.

3.2. Receptor Screening Results

In aim to identify targets of CDK-i drugs in human and *C. elegans* proteome, receptor screening (inverse docking) methodology was utilized. Approximately 430,850 individual docking simulations were performed in total. Upon screening, canonical target affinities were identified for each dataset to determine the cutoffs, utilizing drug-target pairs as listed below.

Table 3.1: Cutoff values (kcal/mol) for human and *C. elegans* dataset reductions were obtained upon receptor screening.

	Human	***C. elegans***
Abemaciclib	-9.4 (CDK6)	-9.3 (CDK-4)
Palbociclib	-7.5 (CDK6)	-9.3 (CDK-4)
Ribociclib	-7.9 (CDK6)	-7.2 (CDK-4)
Trilaciclib	-10.3 (CDK6)	-10.1 (CDK-4)
Lerociclib	-8.7 (CDK6)	-8.5 (CDK-4)
Roscovitine	-8.3 (CDK7)	-8.4 (CDK-7)
Narazaciclib	-8.9 (CDK4)	-8.1 (CDK-4)
Atuveciclib	-9.3 (CDK9)	-7.6 (CDK-9)
PHA-767491	-7.6 (CDK9)	-5.6 (CDK-9)
FUDR	-6.5 (TYMS)	-6.8 (TYMS-1)

After obtaining cutoff values, data reduction and cleaning were performed by removing proteins with less successful bindings than their cutoff values from the datasets.

After reducing the dataset sizes by removing low affinity binder proteins, in human, the smallest dataset was found as atuveciclib with 404 elements, and the biggest dataset was found as narazaciclib with 7206 elements. The total number of screened proteins in human proteome for each drug was 23391.

In *C. elegans*, the smallest dataset was palbociclib with a number of 399 and the biggest dataset was PHA-767491 with 12845.A total number of screened proteins in *C. elegans* proteome for each drug was 19694 (Figure: 3.6-A).

In order to statistically analyze further, and for also highlight any relevant information, affinity distribution histograms were plotted for each human and *C. elegans* datasets.

Upon analyzing distributions, firstly, it was observed that, in all datasets distributions, affinities were distributed binominally. Additionally, roscovitine showed more distinct affinities around -9 kcal/mol in both human and *C. elegans* datasets. In human datasets, palbociclib, narazaciclib and lerociclib, in *C. elegans* datasets PHA-767491, lerociclib and atuveciclib showed wider range in terms of affinity (Figure:3.6-B and C).

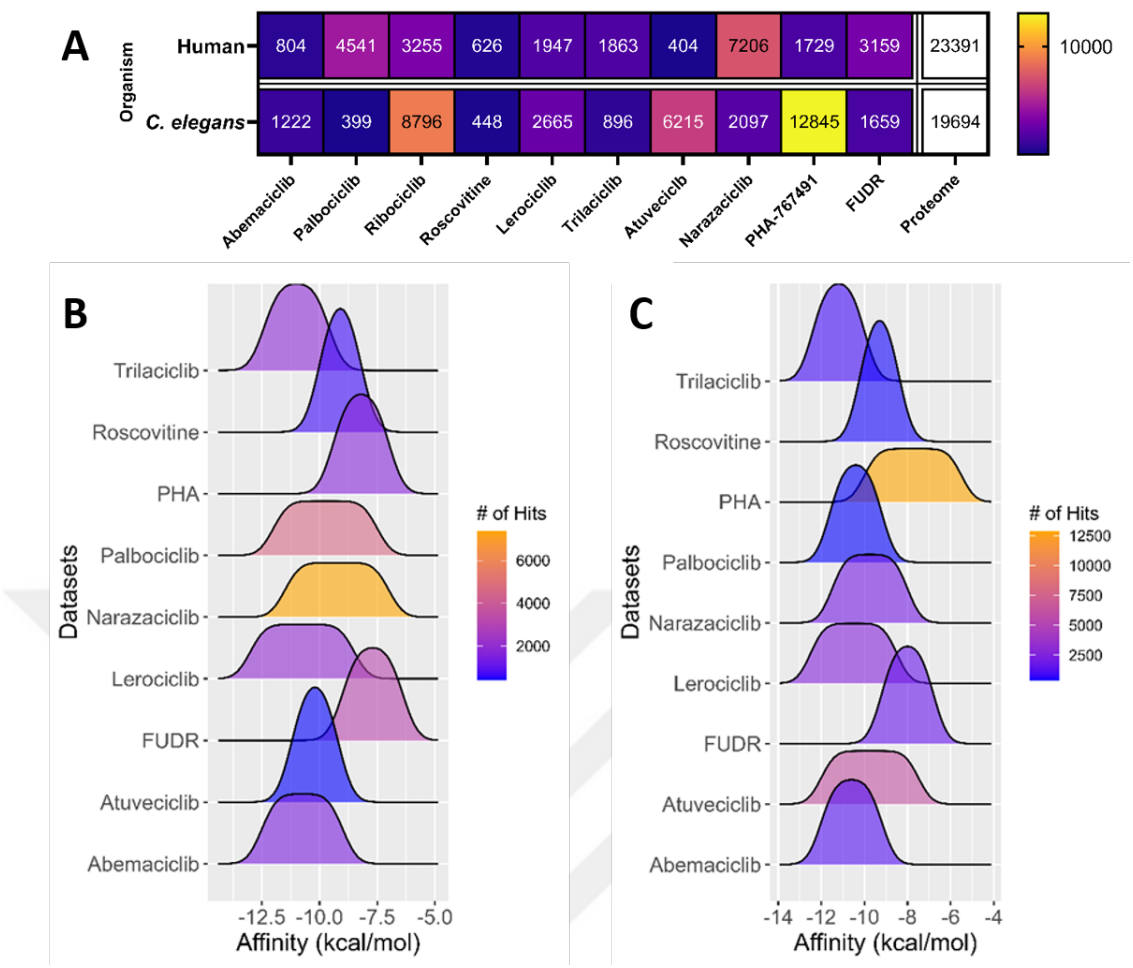


Figure 3.6: Human and *C. elegans* dataset sizes and distributions. **A:** Dataset sizes after affinity cut-off. Number of elements in each dataset represented with the value in corresponding cells. White cells represent proteome size for organisms. In human, smallest dataset was atuveciclib, and biggest dataset was found as narazaciclib. In *C. elegans*, smallest dataset was palbociclib and biggest dataset was PHA-767491. **B:** Human affinity-size distribution graph. **C:** *C. elegans* affinity-size distribution graph. Roscovitine docking affinities were around -9 kcal/mol in both human and *C. elegans* datasets. Colors indicate number of elements in datasets, peak height indicates stacking of the same affinities. Coloring towards yellow indicates higher dataset size. In both distribution graphs, median affinity was around -9 kcal/mol for all datasets.

In addition to these findings, with the screening of the proteomes, binding affinities of drugs for the previously noted canonical CDK targets were revealed. Obtained affinity values were normalized as explained in Section 2.1.3. Both in human and in *C. elegans*, PHA-767491, FUDR and Roscovitine had the highest efficiencies for the CDKs while Lerociclib, ribociclib and palbociclib showed low efficiencies. (Figure: 3.7)

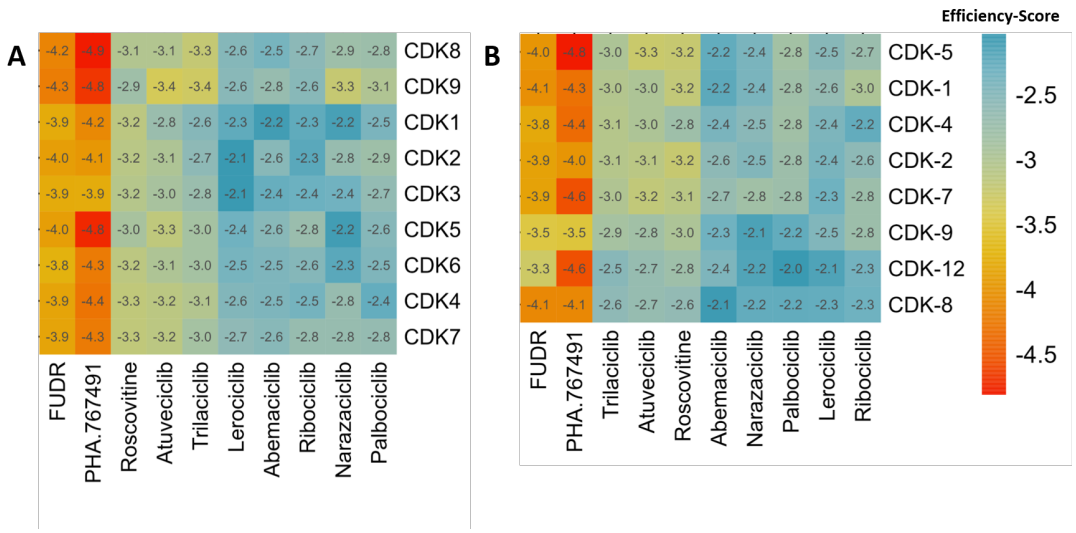


Figure 3.7: Drug efficiency scores of drugs for Human (**A**) and *C. elegans* (**B**) Cyclin-dependent Kinases. Similar scores were neighbored, lower values represent better binding towards the red color. FUDR and PHA-767491 molecules have shown higher affinities for Cyclin-dependent Kinases in both human and *C. elegans*.

3.3. Functional Enrichments and Hierarchy Analysis

In order to determine the most potential targets by their relevant cumulative biological meaning in each receptor screening, all of the cleaned-up datasets were analyzed one-by-one via functional enrichment using gProfiler. To perform functional enrichments, firstly reduced datasets were provided to KPM to create enriched, corrected protein-protein interaction networks, which convey their relevant biological pathways or conditions. Upon network enrichment, dataset sizes were drastically reduced further, around 35% and up to 49% reduction in human datasets and 80 to 91% reductions in *C. elegans* datasets were observed. Overall, in terms of size, *C. elegans* datasets were reduced much more drastically than human datasets (Figure: 3.8-C).

Consequently, network-enriched and reduced datasets were functionally enriched using gProfiler. After analysis, top-ranking KEGG terms were highlighted.

In human datasets, for each dataset, ”Metabolic Pathways” term was initially identified as the top-ranking KEGG term. For abemaciclib, ”Fc gamma R-mediated phagocytosis”; for palbociclib, ”Cofactor Biosynthesis”; for roscovitine, ”Progesterone-mediated oocyte maturation”; for ribociclib, ”Neurotrophin signaling pathway”; for FUDR, ”Cofactor Biosynthesis”; for lerociclib, ”Peroxisome”; for trilaciclib, ”Fatty acid degradation”;

for narazaciclib, ”Cofactor Biosynthesis”; for atuveciclib, ”Peroxisome”, and as well as for PHA-767491, ”Cofactor Biosynthesis” terms were highlighted (Figure 3.8 -A).

In *C. elegans* datasets, the ”MAPK Signaling Pathway” was predominantly highlighted for abemaciclib, palbociclib, FUDR, trilaciclib, and narazaciclib. Meanwhile, the ”FoxO Signaling Pathway” was highlighted for ribociclib, lerociclib, and atuveciclib. Additionally, the ”TGF-beta Signaling Pathway” was identified as a highlighted KEGG term for roscovitine, and the ”Wnt Signaling Pathway” for PHA-767491 upon functional enrichments (Figure 3.8 -B).

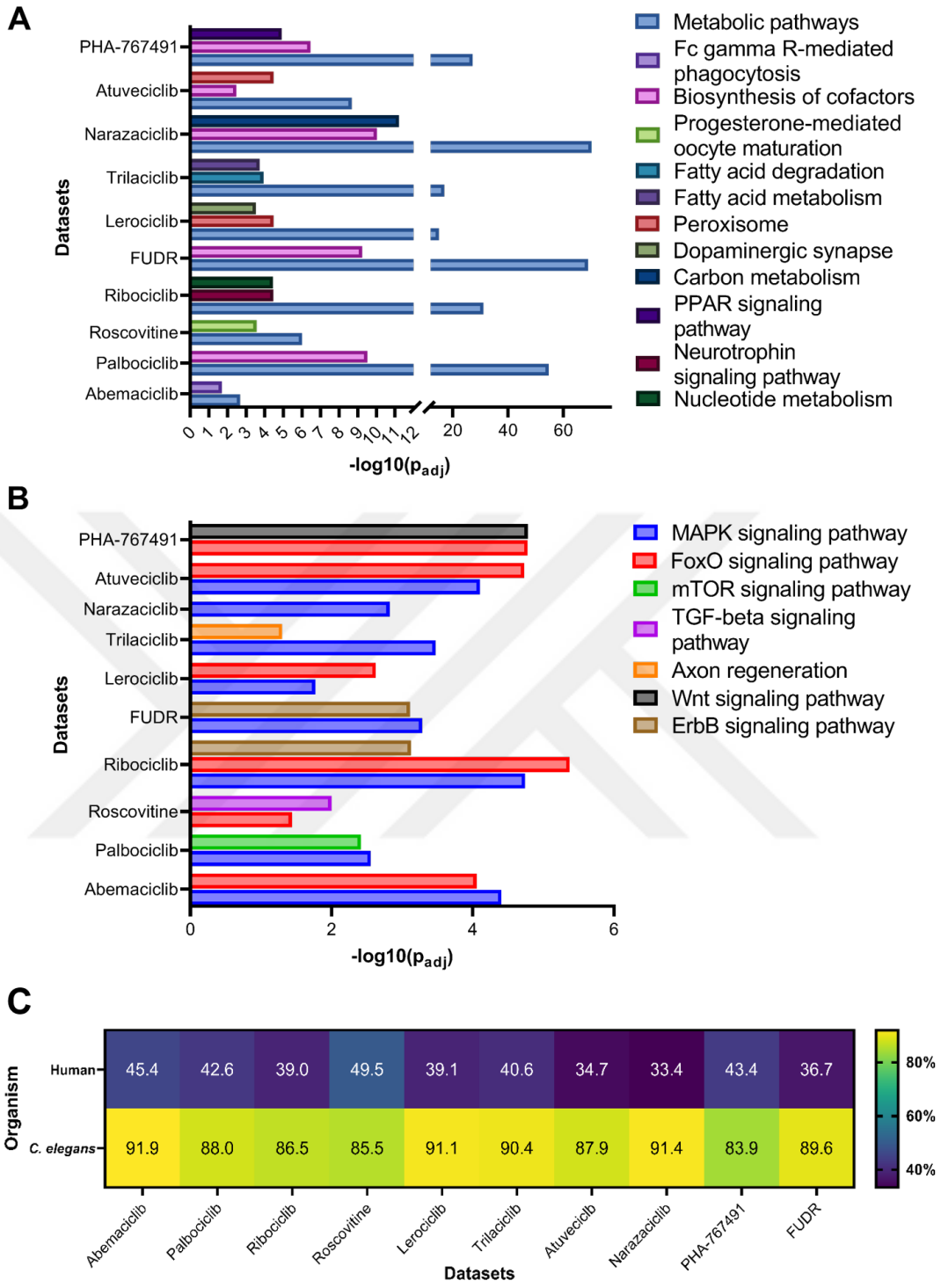


Figure 3.8: Functional enrichment results of human and *C. elegans* datasets. **A:** Human datasets KEGG terms after Gene set enrichments. **B:** *C. elegans* datasets KEGG terms after Gene set enrichments. **C:** Reduction rates after KPM network enrichments.

Proteins with the highest degree of connectivity in the cell are crucial for its survival [109]. Therefore, in order to obtain most potent targets, highest connected potential targets were obtained from each dataset. Upon hierarchically ordering the proteins in the protein interaction network; biologically relevant highest degree nodes were selectively identified as most potential targets.

Upon clustering the nodes according to their functional clusters and also hierarchically ordering the nodes within clusters in terms of interactions, PRKCA in ”FC gamma R-mediated Phagocytosis”, HADHA in ”Fatty Acid Metabolism”, PEX5 and NOS2 in ”Peroxisome”, GAPDH in ”Carbon Metabolism”, NTRK1 in ”Neurotrophin Signaling Pathway”, SHMT2 in ”Biosynthesis of Cofactors”,and MAPK14 within ”Progesteron-mediated Oocyte Maturation” proteins were found as highest degree nodes in human interactome as seen in Figure 3.9.

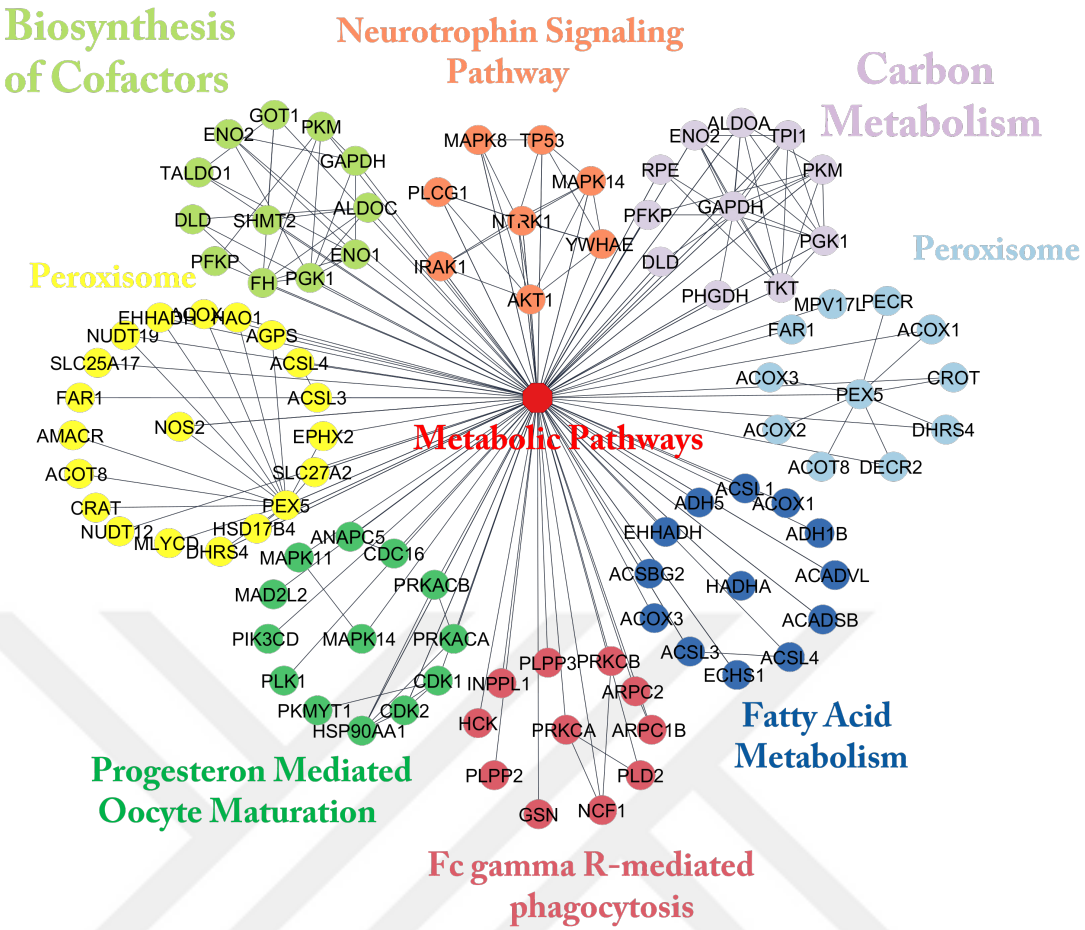


Figure 3.9: Clustered protein interaction network for human datasets through functional enrichments. Each dataset is indicated by a different color. Within clusters, proteins with the highest overall interactions in the network are positioned at the center of the clusters hierarchically.

Continuing with the hierarchical analysis of *C. elegans* functional interactions, after clustering the proteins with functional roles in KEGG and phenotypic roles in WPO, LET-756, LET-60, MPK-1, PMK-1 and SEM-5 in MAPK Signaling pathway, DAF-2 in mTOR and MAPK Signaling Pathways, LET-60 also in FoxO signaling pathway, and DAF-4 and LIN-35 in TGF-beta signaling pathway were identified as highest degree nodes. Within these pathways, all pathways were associated with developmental phenotypes with either ”Developmental Phenotype” or ”Organism Development Variant”. In MAPK signaling pathway, WPO term ”Fertility Phenotype” was also discovered (Figure: 3.10).

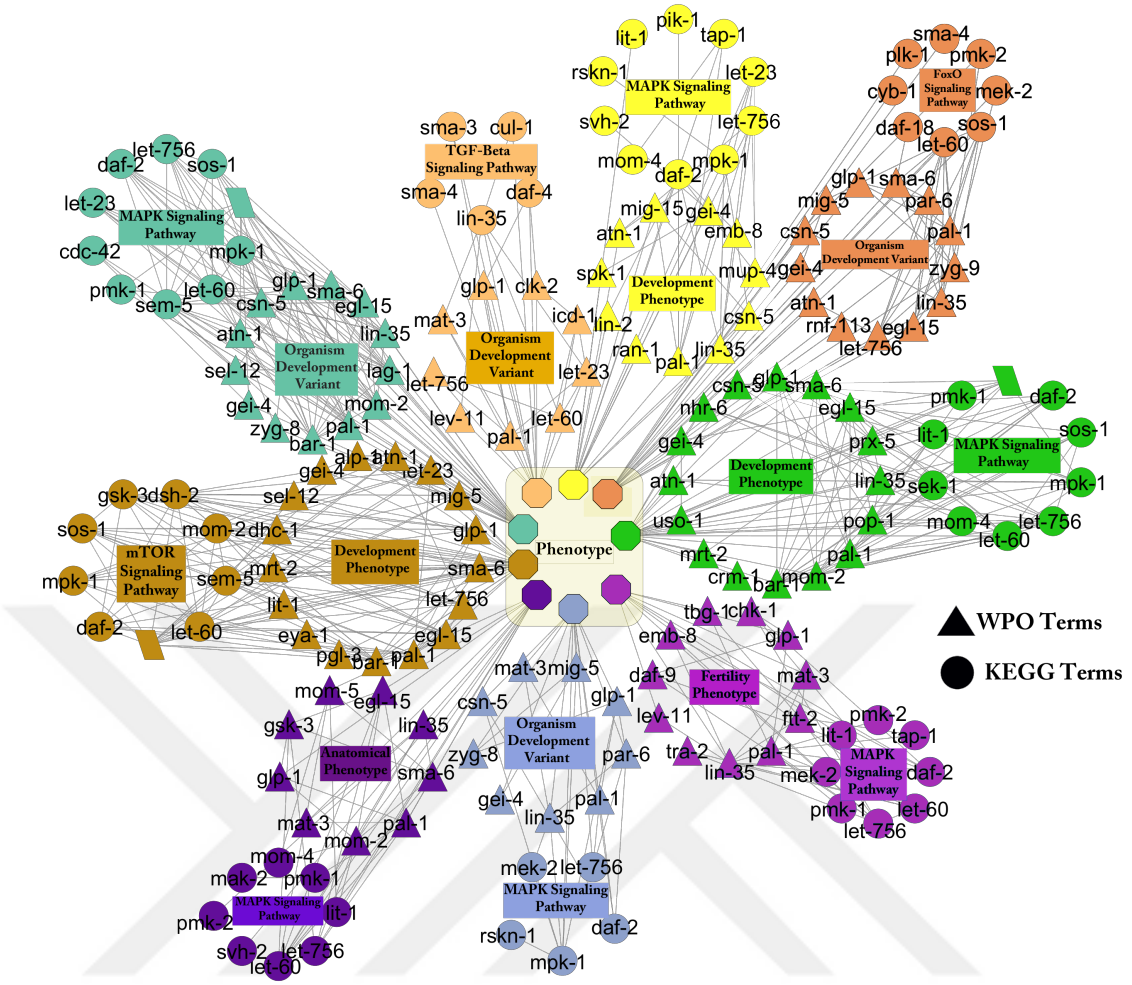


Figure 3.10: Functional enrichment-based clustered protein interaction network for *C. elegans* datasets. Each dataset is represented with a different color. Protein names are represented as gene names. Within clusters, proteins with the most interactions throughout the network are positioned hierarchically towards the bottom of the clusters. Triangles (close to the center) represent clusters corresponding to Worm Phenotype Ontology (WPO) terms, while circles represent clusters corresponding to KEGG terms. Parallelograms indicate merged low-degree nodes.

In order to further distinguish drug-specific potential targets, hierarchically ordered proteins for each drug were listed (Appendix-A). Within this listing, biologically relevant highest-degree nodes were selectively identified as the most potential targets, and assigned to drugs for further drug-target evaluations as listed in Table 3.2.

Table 3.2: Protein names of the identified targets in human and *C. elegans* for each dataset.

	Human	***C. elegans***
Abemaciclib	PRKCA	LET-756
Palbociclib	SHMT2	MEK-2
Roscovitine	MAPK14	DAF-4
Ribociclib	NTRK1	DAF-2
Trilaciclib	HADHA	PMK-1
Lerociclib	NOS2	SOS-1
Narazaciclib	GAPDH	LIT-1
Atuveciclib	PEX5	MPK-1
PHA-767491	EPRS1	BAR-1
FUDR	ADSS2	LET-60

3.4. Expression Profiles of Selected Targets in Human and *C. elegans*

It is known that mRNA levels are parallel with protein levels [131], and a high level of a protein might indicate a higher chance of druggability of that protein. Therefore, in order to understand the targetability of the potentially identified proteins and potential co-expressions, mRNA expression levels were profiled for human and *C. elegans* targets, in parallel with the relevant biological enrichments. Relevant cumulative data were obtained from respective databases, and potent targets with the highest degrees for each drug dataset, and additionally, CDK expression levels were profiled for human cancer tumor samples, various human cell lines, and *C. elegans* larval stages.

After expression profiling, in human cancer cell lines, the *SHMT2* gene and *CDK4* levels are relatively higher in the SK-MM-2 cell line compared to other genes, while the *HADHA* gene is lower. Similarly, in the HOS cell line, the *NOS2* gene is prominently elevated. The same can be said for the other identified targets; *PRKCA* in A498 and A172 cell lines, *NTRK1* in HMCB, *GAPDH* and *PEX5* in MIA Paca-2, *EPRS1* in SK-MM-2, *ADSS2* in A172. Additionally, when compared to the average between cell lines, *CDK4* and *CDK6* expression levels were observed lowest in MIA Paca-2 (Figure: 3.11-A).

When primary cancer tumor samples from TCGA were evaluated, selected targets were expressed relatively higher in bile duct-originated Cholangiocarcinoma (CHOL) than in other samples. In GBM (glioblastoma) samples, *CDK2*, *CDK4*, *CDK6*, *HADHA* and *SHMT2* were relatively overexpressed than the samples average. *NTRK1* and *NOS2* had higher levels in PCPG (pheochromocytoma and paraganglioma) when compared to the average of all samples. Additionally, it was observed that, gene expressions between samples with the same origin were also vary, like in KIRC and KIRP, *SHMT2* and *PRKCA* expression levels varied between samples (Figure: 3.11-B).

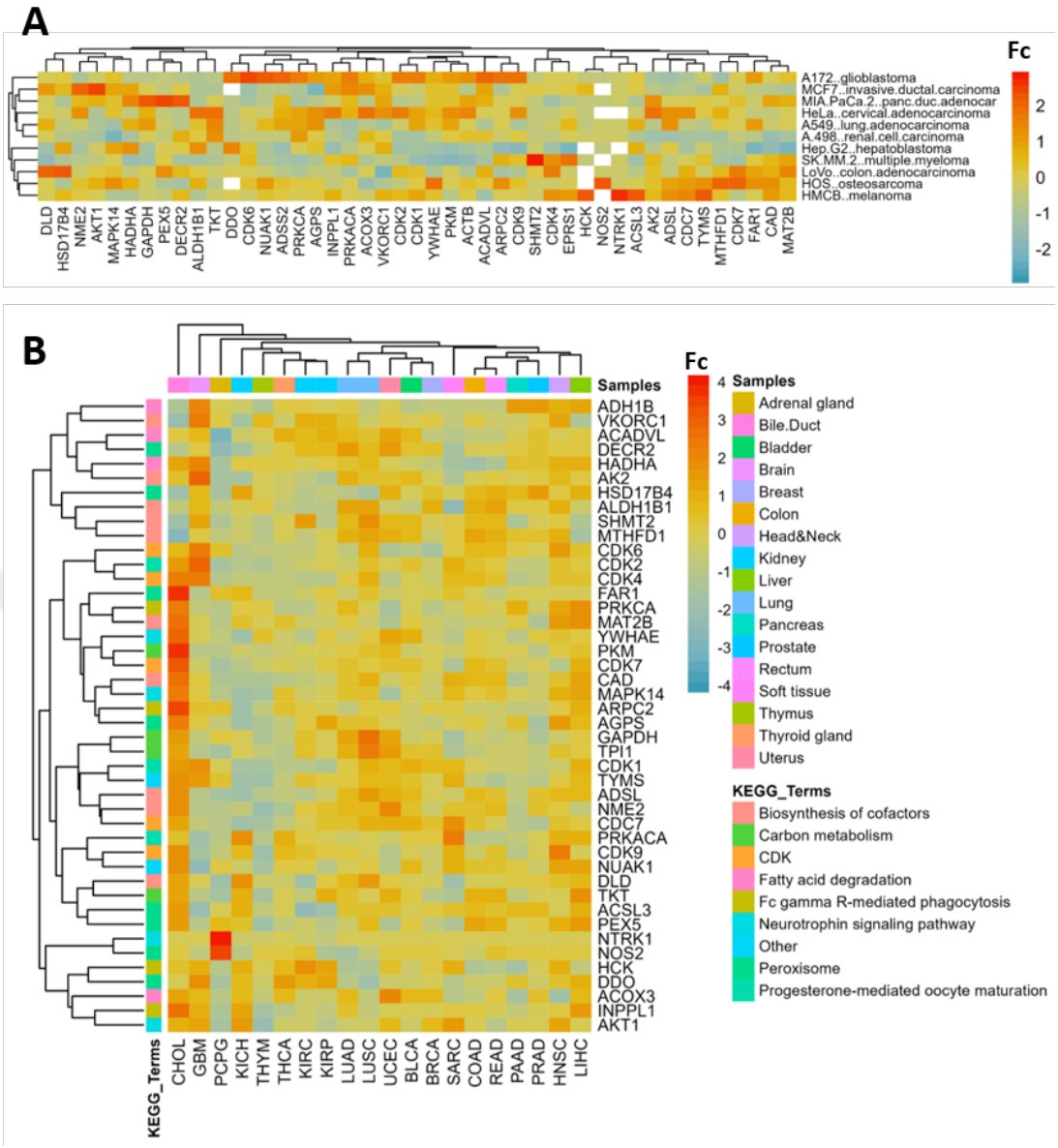


Figure 3.11: Gene expression profile of identified potential targets in human cancer cell lines (**A**, vertical axis) and human primary cancer sites (**B**, horizontal axis). **A**: Data obtained and curated from Human Cancer Cell Line Atlas. The *ACTB* gene data was also selected from the original data and added to the heatmap for visualization and comparability. **B**: Data obtained from OncoDB. The topmost colored bar represents the overall set of cancer samples, while the leftmost colored bar indicates the KEGG terms relevant to the identified targets. CDKs are grouped and shown under "KEGG Terms" for simplification. Similar expression levels in the samples are neighbored. From blue to red color indicates Fc. White cells indicate no data. **Fc**: Fold-chance

When the gene expression levels of potential targets for *C. elegans* were assessed in larval developmental stages, it was observed that the *let-756* is primarily expressed highly at the L1 stage, while there is a high expression levels of many genes during the embryonic stage. Another notable aspect is the expression of the *daf-2* gene in the neurons of the L1 stage. Additionally, the general clustering of the expression levels in L1 and embryo stages was also evident. Also, all Cyclin-dependent Kinases of interest in *C. elegans* were highly upregulated in embryonic stages compared to other larval stages (Figure: 3.12).

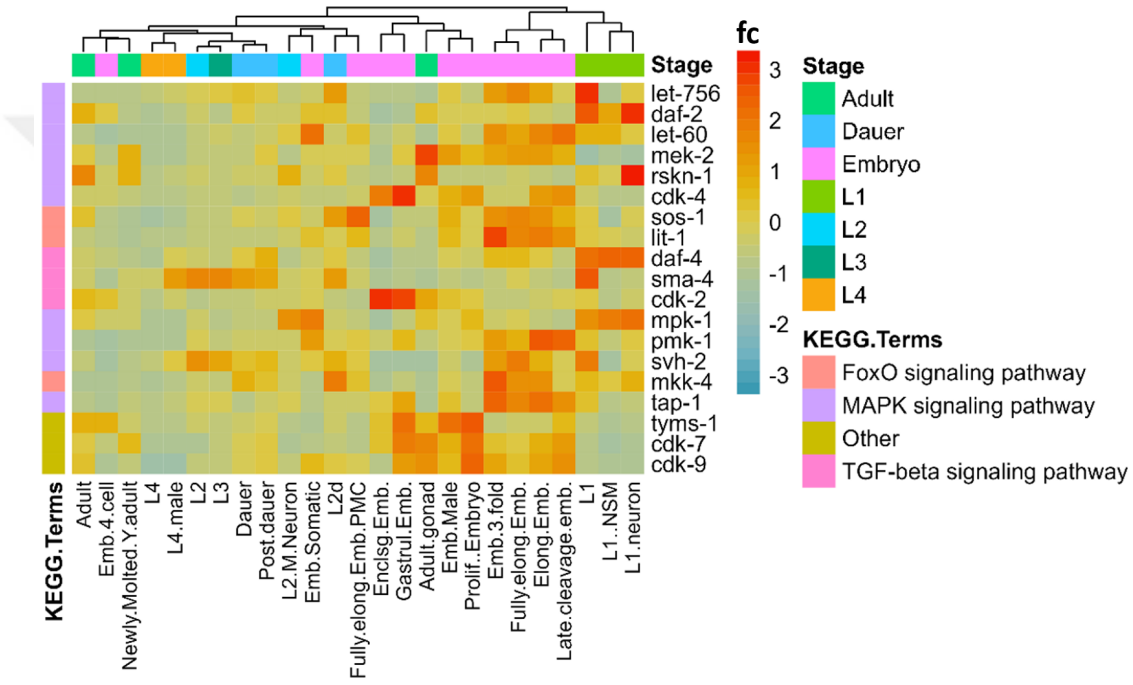


Figure 3.12: Gene expression heatmap for *C. elegans* larval developmental stages (horizontal axis) and identified potential targets (vertical axis). Data obtained and curated from The Expression Atlas. The topmost colored bar represents the overall set of larval stages, while the leftmost colored bar indicates the KEGG terms relevant to the identified targets. From blue to red color indicates Fc. Fc: Fold-chance

3.5. MD Analysis of Identified Targets

With the network hierarchy analysis and expression profiling, most potently identified targets were subjected to MD simulations to evaluate drug-target dynamics.

In the initial ligand-free solute (protein in water 0.15 mM salt, protein only) MD simulations of the selected targets, while MAPK14 showed deviations at the beginning,

all human structures reached to stability in RMSD within 30 ns. In *C. elegans* solute MD simulations, most structures reached to stability in varying timesteps except DAF-4 and SOS-1 structures (Figure: 3.13-A and B). The abemaciclib molecule, simulated by binding to the human target PRKCA, exhibited a higher initial deviation compared to all other protein-ligand complexes The drug molecules in all complexes show stability across the simulation durations. Palbociclib molecule in *C. elegans* target PMK-1, did not show stability in given timeframe, and ribociclib in DAF-2 showed late instabilities in the MD (Figure: 3.13-B and C).

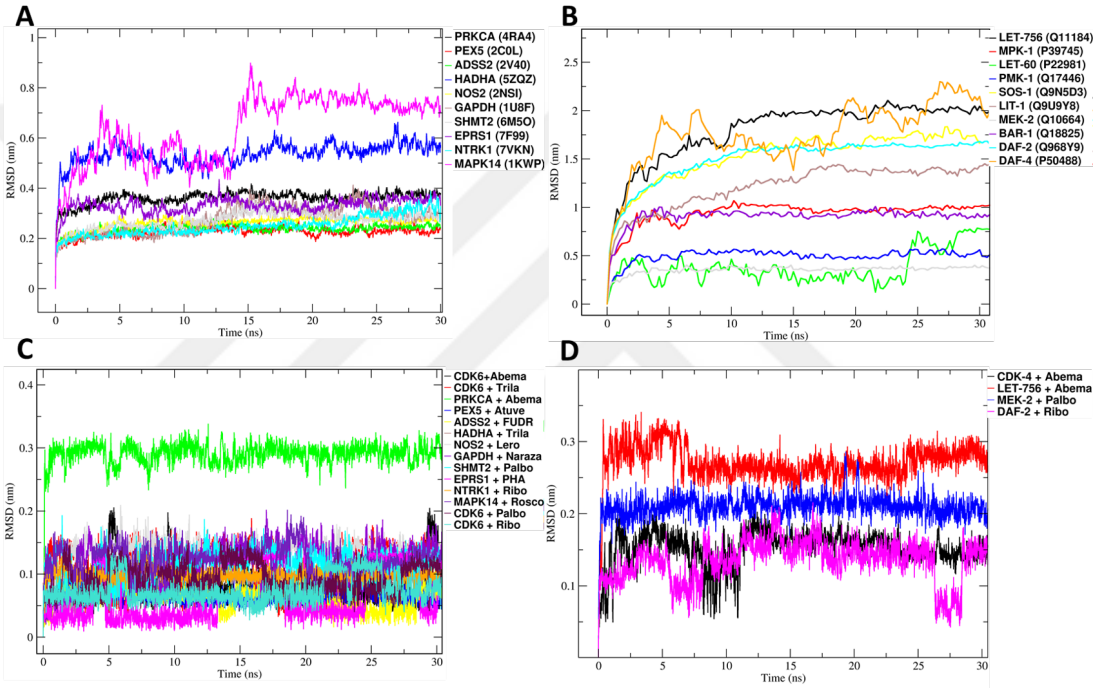


Figure 3.13: RMSD graphs of ligand-free, solute MD simulations of the selected targets and drug molecules in selected potential target protein-ligand complexes. **A:** RMSD results for identified human target protein structures. **B:** RMSD results for identified *C. elegans* target protein structures. Each color in a graph represents different target protein RMSD, while monikers in brackets in legend also indicates corresponding RSCB PDB (**A**) or Alphafold (**B**) identifier. **C:** Human protein-ligand complexes. **D:** *C. elegans* protein-ligand structures. Each color represents the analysis results of a different drug molecule in the respective complexes.

In addition to the MD results, binding free energies of the drugs to their identified targets were calculated using MM/GBSA method. In MM/GBSA analysis, trilaciclib in human target HADHA, showed the lowest binding free energy in all complexes

(*ΔG*:-56.710 kcal/mol), even lower than the primary target CDK6 of the trilaciclib (*ΔG*:-52.410 kcal/mol). Abemaciclib in human target PRKCA again showed similar free energy (*ΔG*:-48.113 kcal/mol) when compared to binding to primary target CDK6 (*ΔG*:-53.160 kcal/mol) Lerociclib·NOS2 as *ΔG*:-44.673 kcal/mol, ribociclib·NTRK1 as *ΔG*:-41.717 kcal/mol, atuveciclib·PEX5 as *ΔG*:-33.810 kcal/mol, palbociclib·SHMT2 as *ΔG*:-32.543 kcal/mol and roscovitine·MAPK14 were calculated as *ΔG*:-22.883 kcal/mol. In *C. elegans* target-drug complexes, ribociclib·DAF-2 as *ΔG*:-37.260 kcal/mol, palbociclib·MEK-2 as *ΔG*:-34.463 kcal/mol, trilaciclib·PMK-1 *ΔG*:-33.043 kcal/mol, abemaciclib·LET-756 as *ΔG*:-20.493 kcal/mol and abemaciclib·CDK-4 binding were calculated as *ΔG*:-17.667 kcal/mol (Figure : 3.14)

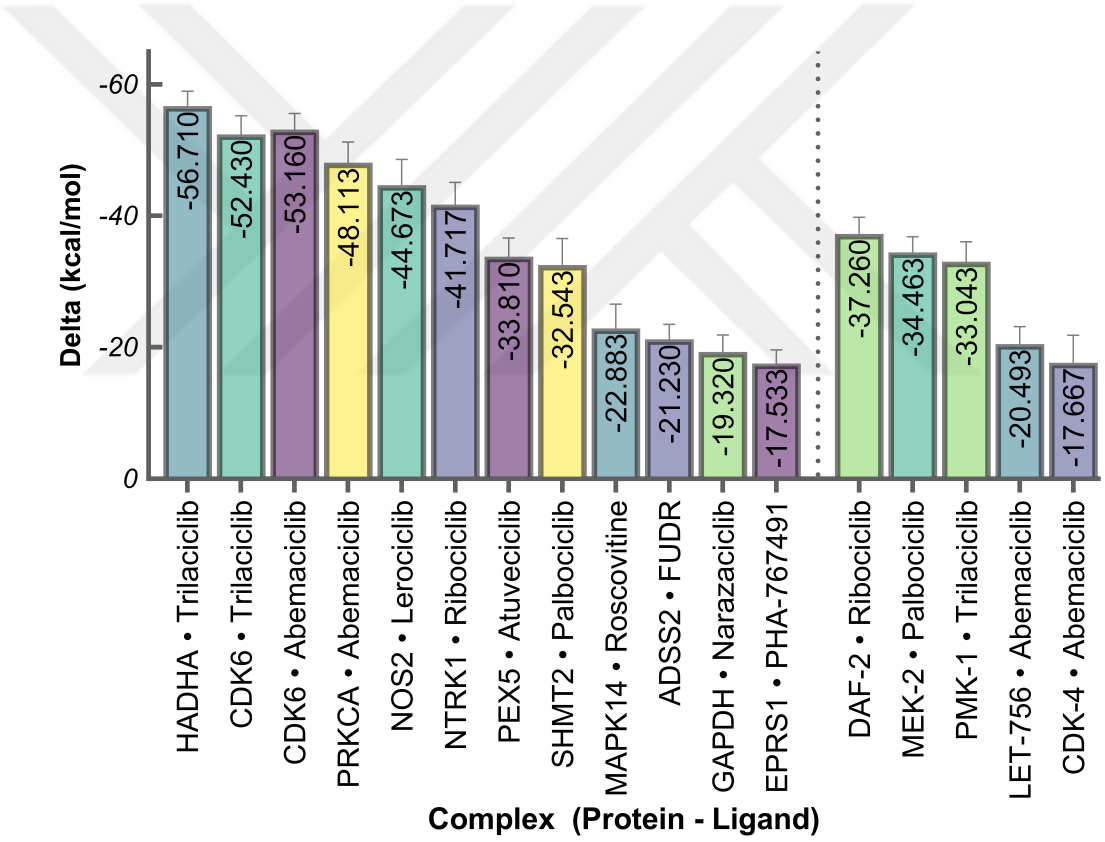


Figure 3.14: Binding free energy results of selected targets of human and *C. elegans*. The result obtained by MM/GBSA analysis. Human target-drug binding energy calculations were places left side while *C. elegans* target-drug results right-hand side of the bar graph. HADHA-Trilaciclib showed higher binding free energy than CDK6-Trilaciclib and Abemaciclib. (Error Bars: SEM, $n \geq 3$)

Afterwards, each drug-target trajectories were clustered to obtain drug-target residue interaction conformation diagrams. When residue interaction diagrams were evaluated, solvent exposure of ligands was observed in all drug-target conformations (Figure: 3.15 and Figure: 3.16). Abemaciclib, in the canonical target CDK6 and in the identified target PRKCA shown similar arene interactions, and again showed similar solvent exposures (Figure: 3.15).

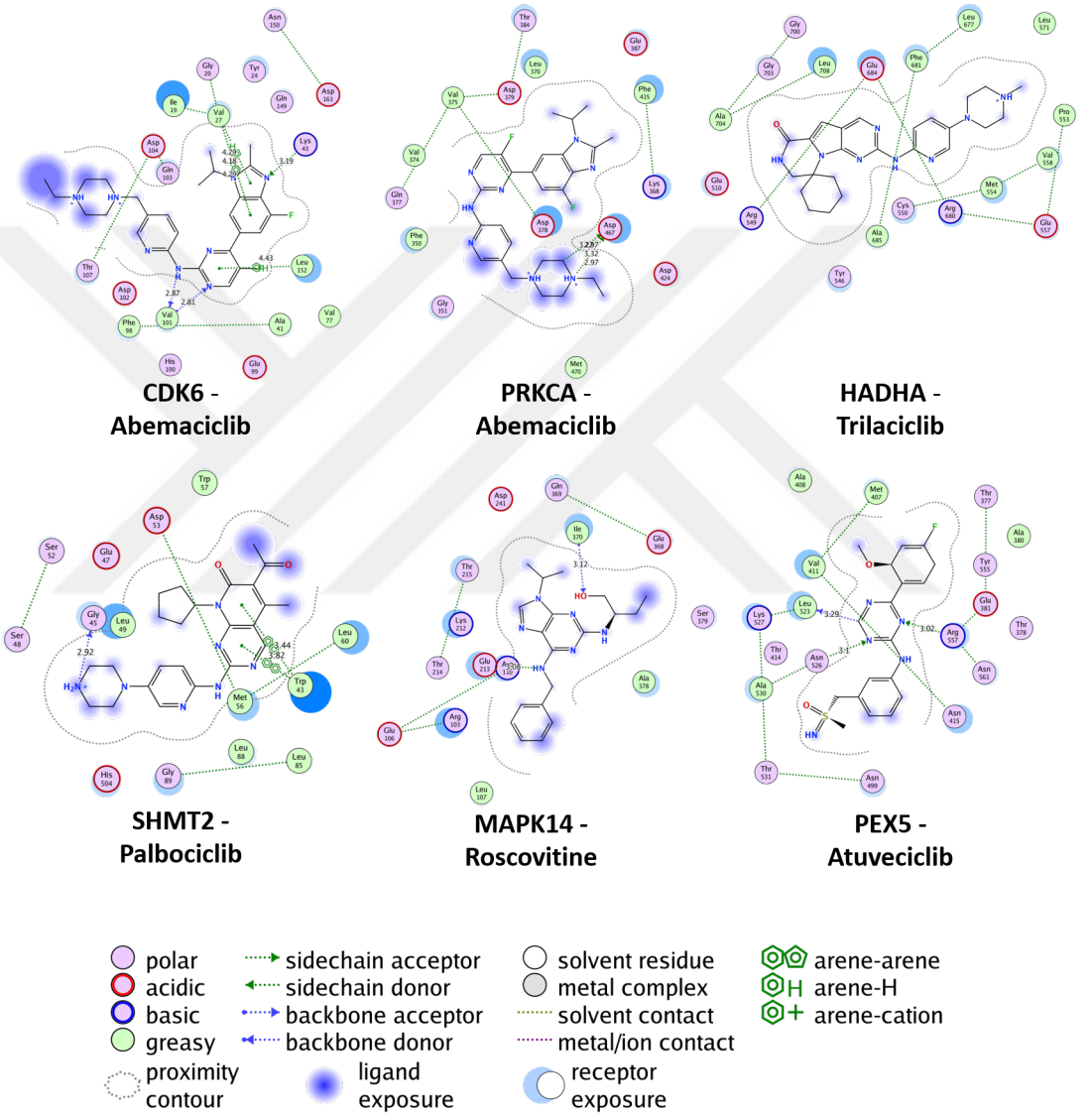


Figure 3.15: Human target-drug interaction diagrams. From top-left to bottom-right; CDK6-Abemaciclib, PRKCA-Abemaciclib, HADHA-Trilaciclib, SHMT2-Palbociclib, MAPK14-roscovitine, PEX5-Atuveciclib and complexes, respectively. The legend in the bottom represents the type of interactions.

When abemaciclib were inspected for conformation within the possible target *C. elegans* CDK-4, it shown misfitting within the binding pocket of the target and was more exposed to solvent when compared to its canonical binding (CDK6, Figure: 3.15). In addition, when Abemaciclib is inspected for LET-756 binding, more interacting residues were observed and buried more within the protein compared to CDK-4 (Figure: 3.16).

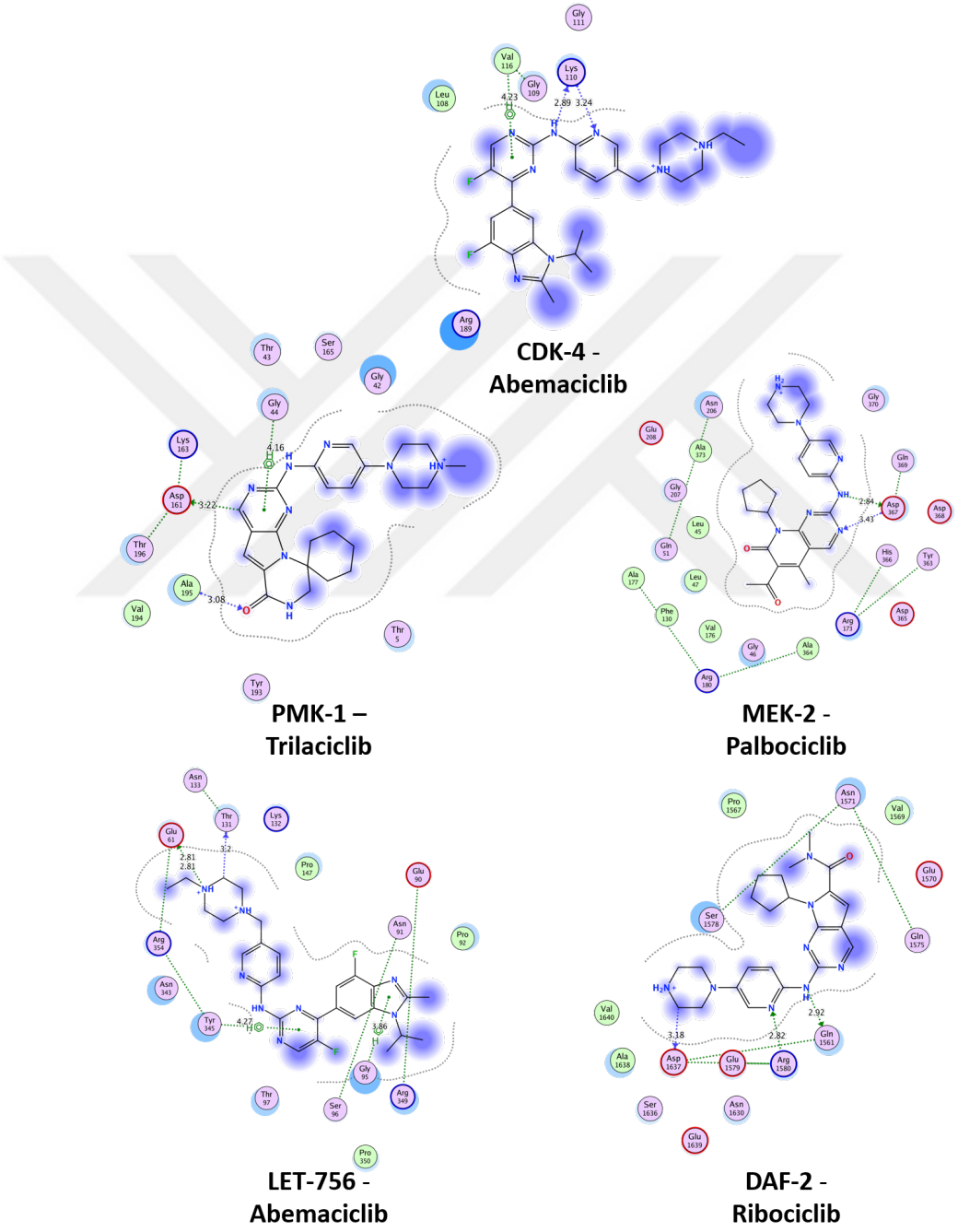


Figure 3.16: *C. elegans* drug-target interaction diagrams. CDK-4-Abemaciclib, PMK-1-Trilaciclib, MEK-2-Palbociclib, LET-756-Abemaciclib and DAF-2-Ribociclib complexes, respectively.

3.6. Proteomic Effects of Drugs in Human Cell Lines

In order to identify drug induced differentiation in protein levels, 2D gel/Mass spectrometry analysis methodology was utilized previously, as explained. Upon evaluating the differential protein levels in PANC-1 and MIA Paca-2 cells using 2D gel/MS methodology, Gelsolin protein was highly downregulated (16.67-fold) in PANC-1 cells after abemaciclib treatment. When same cell lines were treated with palbociclib, Ezrin and Radixin drastically downregulated (100-fold). Continuing with the MIA Paca-2 cell lines, abemaciclib upregulated Tropomyosin alpha-3 chain (24.34), while palbociclib-treated cells exhibiting upregulated Peroxiredoxin-6 (7.99) and Phosphoglycerate mutase 1 (3.33). In MIA Paca-2 cells, abemaciclib and palbociclib altered the regulation of the tropomyosin alpha-3 chain (Table 3.3 to 3.6).

Table 3.3: List of the differentially regulated proteins identified by 2D gel/MALDI-TOF mass spectrometry in PANC-1 cell lines after abemaciclib treatment.

Accession Number	Protein Description	Protein Score	E-value	Ratio	
P43487	Ran-specific GTPase-activating protein	102	1.3E-06	2.86	↓
P06396	Gelsolin	60	2E-02	16.67	↓
P61978	Heterogeneous nuclear ribonucleoprotein K	82	1E-04	2.00	↓
Q13838	Spliceosome RNA helicase BAT1	84	8E-05	2.22	↓
O00148	ATP-dependent RNA helicase DDX39	59	2E-02	2.22	↓
P04264	Keratin, type II cytoskeletal 1	94	8.1E-06	2.70	↓
P35908	Keratin, type II cytoskeletal 2 epidermal	71	2E-03	3.33	↓
P00568	Adenylate kinase isoenzyme 1	173	1E-013	1.81	↑

Table 3.4: List of the differentially regulated proteins identified by 2D gel/MALDI-TOF mass spectrometry in PANC-1 cell lines after palbociclib treatment.

Accession Number	Protein Description	Protein Score	E-value	Ratio	
P15311	Ezrin	262	1.3E-22	100.00	↓
P35241	Radixin	59	2E-02	100.00	↓
P13645	Keratin, type I cytoskeletal 10	124	8.1E-09	7.14	↓
P30086	Phosphatidylethanolamine-binding protein 1	73	1E-03	2.80	↑
Q12931	Heat shock protein 75 kDa, mitochondrial	372	1E-33	2.22	↓
P10768	S-formylglutathione hydrolase	536	5.1E-50	1.55	↑
P04083	Annexin A1	245	6E-21	2.12	↑
O00151	PDZ and LIM domain protein 1	192	1.3E-15	2.12	↑
P37837	Transaldolase	84	7.9E-05	2.12	↑
Q15365	Poly(rC)-binding protein 1	144	8.1E-11	2.15	↑
P57721	Poly(rC)-binding protein 3	68	3E-03	2.15	↑
Q06830	Peroxiredoxin-1	682	1E-64	2.53	↑

Table 3.5: List of the differentially regulated proteins identified by 2D gel/MALDI-TOF mass spectrometry in MIA PaCa-2 cell lines after abemaciclib treatment.

Accession Number	Protein Description	Protein Score	E-value	Ratio	
P06753	Tropomyosin alpha-3 chain	53	1E-01	24.34	↑
P30101	Protein disulfide isomerase A3	271	2E-23	1.98	↑
P02768	Serum albumin	78	2.9E-04	17.61	↑
P07437	Tubulin beta chain	39	3E+00	2.63	↑

Table 3.6: List of the differentially regulated proteins identified by 2D gel/MALDI-TOF mass spectrometry in MIA PaCa-2 cell lines after palbociclib treatment.

Accession Number	Protein Description	Protein Score	E-value	Ratio	
P06753	Tropomyosin alpha-3 chain	89	2.5E-05	1.35	↓
P15311	Ezrin	289	2.6E-25	2.45	↓
P30041	Peroxiredoxin-6	468	3.2E-43	7.99	↑
P18669	Phosphoglycerate mutase 1	212	1.3E-17	3.33	↑
P06733	Alpha-enolase	530	2E-49	2.02	↑

3.7. Physiological Effects of Drugs on *C. elegans*

In addition to computational methodologies, abemaciclib, palbociclib, roscovitine and FUDR drugs were tested on *C. elegans* to understand and establish the connections to the possible effects of inhibition of the identified targets.

In developmental assays, worms grown and observed from the L1 stage in a drug-free environment were significantly larger in body size (1766.73 px^2) compared to worms grown in environments containing palbociclib (1566.45 px^2) and roscovitine (1487.56 px^2), while no significant difference was observed in body size between worms treated with abemaciclib (1738.89 px^2) and FUDR (1680.81 px^2) at the end of the initial 24-hour treatment. At 48 hours, no significant differences were observed in any of the

drug-treated groups compared to the control group in worms grown from L1 larvae (Figure: 3.17-A).

When individuals grown from dauer larvae were considered, no significant differences were observed at 24th hour and 48th hours compared to the control group (Figure: 3.17-B).

In both L1 and dauer larvae developmental assays, the control group for each day showed a significant difference compared to the control group of the previous day until the 48th hour. L1 worms in control groups started to lose size difference after 76th hour (Figure: 3.17-A and B).

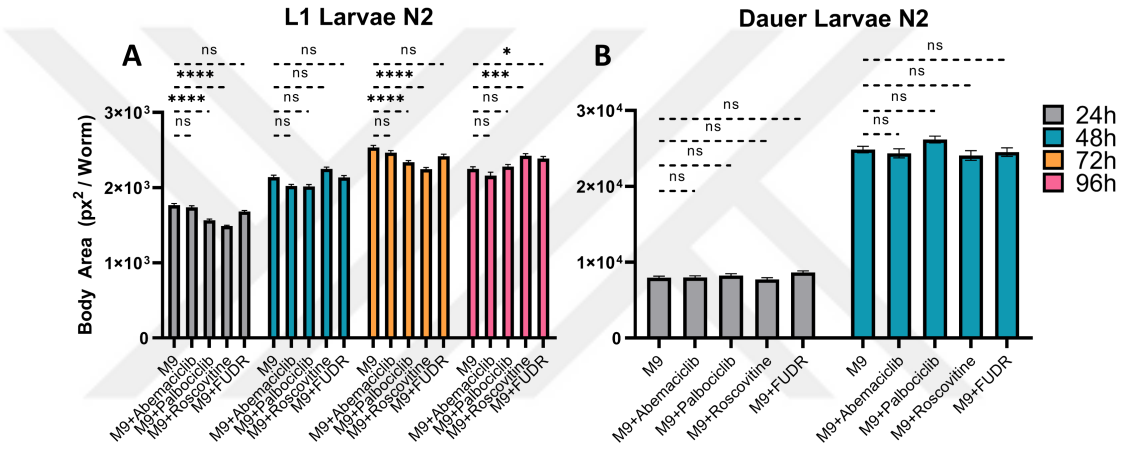


Figure 3.17: Average body areas of *C. elegans* individuals grown in drug-free (M9) and 1 µM drug-containing environments. Starting from L1 larvae (**A**) and from dauer larvae (**B**) to onwards, at 24-hour intervals. After 24 hours palbociclib and roscovitine treated nematodes were significantly smaller compared to control. At the 48th hour, treatments were same with control, and in the 72nd hour, same pattern with the 24h was observed. At the 96th hour, Roscovitine and FUDR shown differences with the control. In dauer grown larvae, no significant differences were observed between treatment and control groups. Each color group represents measurements taken at 24-hour intervals (ns: P>0.05; *: P≤0.05, **: P≤0.01, ***: P≤0.001, ****: P≤0.0001, Error Bars: SEM, n>53).

C. elegans swimming capacity can be representative for overall fitness, muscle health and energy metabolism [132]. In swimming assays, worms grown on drug-containing and control NGMs from both L1 and dauer larvae, shown no significant differences compared to the control worms.

After development and swimming assays, lipid accumulations were quantified in L1 grown and dauer grown *C. elegans* by staining neural lipids inside of the worms and transforming the pixel densities to OD. In nematodes grown from L1, with abemaciclib treated group exhibited significantly higher lipid accumulation (OD 0.2876) compared to the control group (OD 0.2450). The other drug groups did not show a significant difference compared to the control group. In *C. elegans* grown from dauer, no significant changes were observed when any drug group was compared to the control group.

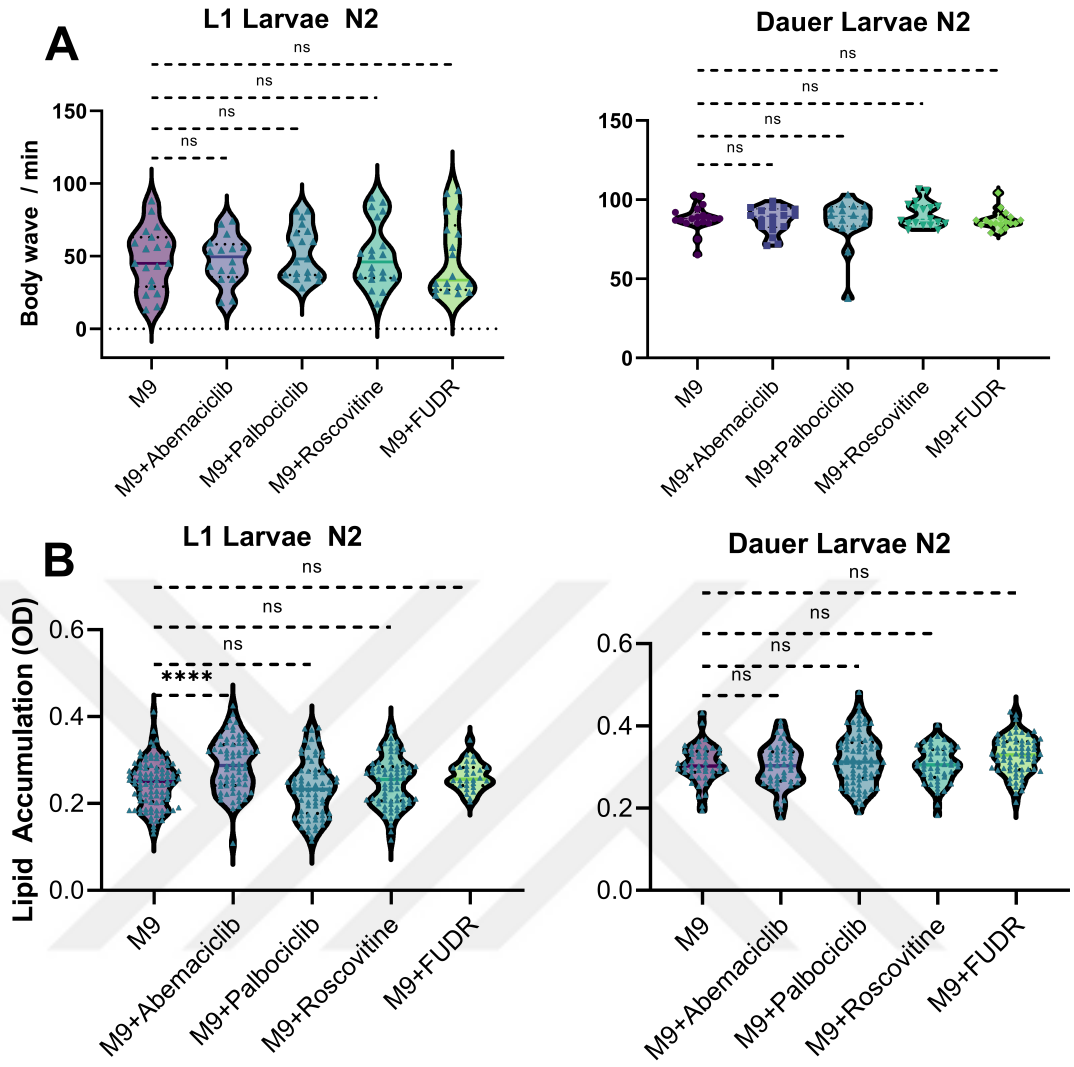


Figure 3.18: **A:** Swimming worms were observed for 1 minute. In both L1 grown and dauer grown larvae, no significant differences were observed between control and treatment groups. (ns:P>0.05,n>16) **B:** Body lipid accumulations of *C. elegans* individuals after growing drug-free and drug-containing environments were evaluated. In L1 grown larvae, abemaciclib treated *C. elegans* shown significantly higher lipid accumulation than control (****: P≤0.0001). In dauer grown larvae, no significance was observed after treatments (ns: P>0.05, ****: P≤0.0001, 28<n<101).

4. DISCUSSIONS AND CONCLUSIONS

This study explores potential targets of Cyclin-dependent Kinase inhibitors (CDK-i) like abemaciclib, palbociclib, and roscovitine beyond their known canonical targets CDK4/6 and CDK2. Additional CDK-i molecules (ribociclib, trilaciclib, lerociclib, narazaciclib, atuveciclib, and PHA-767491) and the non-CDK inhibitor FUDR are included for comprehensive analysis.

4.1. Discussions

4.1.1. Structural Similarity of CDKs and Drug Molecular Weight Might Indicate Varied Selectivity

Sequence comparisons of CDKs within and between human and *C. elegans* reveal low similarity (40.1% and 37.4%) between *C. elegans* CDK-4 and human CDK4/6, suggesting evolutionary divergence (Figure 3.2). When structural similarities were investigated, human CDK5 and CDK1 counterparts in *C. elegans* were highly similar, suggesting CDK5 and CDK1 targeting drugs might show similar biochemical properties in *C. elegans*. (Figure 3.3, 3.4 and 3.5). Additionally, when human CDKs and *C. elegans* CDKs are structurally compared, conserved residues were mostly aligned within the kinase domain, highlighting the possibility of further potential in promiscuity of kinase inhibitor drugs (Figure 3.4). Additionally, structural and functional similarities between *C. elegans* and human CDK proteins support the adaptability of drug pharmacokinetics between humans and *C. elegans*.

Upon receptor screening, most of the affinities in both organism datasets were exhibiting binominal distribution. While this finding does not inherently indicate any reliability measurement of the methodology, it represents that after the cutoff application, there were still variations in the curated datasets (Figure 3.6), and apart from the applied workflow, other forms of sophisticated curation, reduction and/or grouping approaches can still be utilized. Additionally, when drug affinities for human and *C. elegans* CDKs were normalized, FUDR and PHA-767491 molecules has the

highest binding efficiencies (Figure 3.7). This might be due to that both molecules having the smallest molecular weights among the selected drugs (246.19 and 213.24, respectively, Table 2.2), and possibly being able to fit inside many pockets in protein structures. Continuing with that, this outcome also clearly correlates for PHA-767491 in *C. elegans* affinity distribution graph, and trilaciclib in both human and *C. elegans* graphs by smaller molecules showing greater variation of affinities in the screened dataset (Figure 3.6).

Additionally, overall aminoacid composition percentages were obtained from all datasets. In all datasets, as well as proteome-wide percentages, leucine was the most occurring amino acid (Appendix-A). Apart from this finding, the same phenomenon was also seen in Uniprot database statistics [133], and it is possible to say that leucine is the most frequent amino acid in the world. This might be related to the importance of leucine in cell biology and evolutionary conserved functions of leucine, such as but not limited to mTOR activation [134], protein synthesis and integration to other pathways. As such, because mTOR and other biological terms like these were also observed in enrichment phases of the study, there might be a connection between the overall metabolic facilities of the cell with the frequency, or itself of leucine. Additionally, upon calculating the varying amino acidic percentages for datasets with this methodology, it can be speculated that, drug-favored amino acids for each drug or molecule might be identified. Yet, further evaluations are needed.

4.1.2. CDK-is Can Alter Metabolic Processes

The receptor screening approach gave the potential receptor target proteins for each drug, each dataset was still quite large, and while each of these proteins can be targets of interest, further selective reduction and identification were applied by gene enrichment and hierarchical ordering.

In human datasets, for each dataset, ”Metabolic Pathways” term was initially identified as the top-ranking KEGG term (Figure 3.8-A). This finding, at first, elicited a rethinking process for the methodology, and might be due to various reasons. Firstly, while cut-off applied datasets were subjected to network enrichment (Figure: 3.8-C) (experimental reasoning to correct protein target dataset lists), most of the interacting

proteins were associated with various enzymatic and intra-cellular processes, therefore highlighted as metabolic. Secondly, human gene and protein functions are extensively studied, enriched with knowledge, and supported by high-quality publications (Gene Ontology statistics [135]: "Number of annotations by evidence", and UniProt KB statistics [133]: "Taxonomic Origin"). Therefore, each functional enrichment might have gained priority in the enrichment scoring algorithm and associated with each other in a deeper level, resulting a consensus KEGG term. This is more likely to be the reason, since non-network enriched datasets were also highlighted as "Metabolic Pathways" in the methodology validation steps, and human enrichments scores were drastically higher (15 to 70 $-\log_{10}P_{adj}$) than *C. elegans* enrichments scores (2 to 5 $-\log_{10}P_{adj}$) (Figure: 3.8).

In human functional enrichments, "biosynthesis of cofactors" term was highlighted repeatedly in datasets. As previously stated, CDK inhibitor molecules primarily target ATP binding sites, and this property can be a source of promiscuity [54]. Based on this information, that finding might indicate that varied selectivity of the CDK-i molecules towards kinase and ATP dependent enzyme families in the relevant cofactor pathways, and it is reasonable to think that hindrance of such CDKs can alter relevant cellular functions. Additionally, other terms such as "Fc gamma R-mediated phagocytosis", "Peroxisome" can indicate relevant alterations in such pathways.

Supporting this ATP-binding domain occupancy claim, in MS results of abemaciclib and palbociclib treated MIA PaCa-2 cells and PANC-1 cells exhibited down-regulations in enzymes with ATP binding domains (Tables 3.3 to 3.6). Addition to this, palbociclib treated PANC-1 cells had up-regulated Peroxiredoxin-1 protein (Table: 3.4)). This finding correlates with the findings of "Peroxisome" pathways in functional enrichments (Figure: 3.8-B) to indicate CDK-i molecules can hinder peroxisomal activities within the cell.

In *C. elegans* datasets, upon enrichments, "MAPK signaling pathway" was mainly highlighted in functional enrichment results (Figure: 3.8-B). MAPK signaling in *C. elegans* mostly associated with stress response and resistance to environmental cues [136], [137], [138]. Additionally, when WPO enrichments are investigated, Datasets with "MAPK signaling pathway" terms are also mainly associated with "developmental" phenotypes (Figure: 3.16). This can be interpolated as to expect

developmental variations when nematodes were altered in MAPK signaling pathways. Again in *C. elegans* datasets, in roscovitine group, "TGF-beta signaling pathway" was primarily highlighting the possible relevance between this pathway and identified potential targets. "TGF-beta signaling pathway" associated with growth and specialized tissue development previously [139], [140], [141]. Such as the MAPK pathway, and these findings with relevant literature again indicates possible phenotypical and physiological outcomes with alteration of these pathways via inhibiting the identified targets within (Figure 3.9 and 3.10).

4.1.3. Drug Effects Can Variate Between Different Biological Conditions

It is known that mRNA levels correlate with protein levels, and higher levels of proteins might indicate a higher chance of interactions by the drugs in the cellular level [131]. With that in mind, expression levels of identified targets were analyzed for cancer site samples and cancer cell lines in human, as well as for larval stages in *C. elegans*. Between human cancer tumor samples (TCGA-OncoDB) and corresponding types of cancer cell lines, expression levels of primary targets showed some differences (up-down regulation) (Figure 3.11). The lack of exact correlation between cell lines and cancers might be due to individual divergence (differentiation) of the cell lines from the pooled cancer sample. This can again indicate and remind the variability of the cancers among the population. When target expressions in cancer samples are inspected, it is observed and concluded that CDK4/6, canonical target of CDK inhibitors, also varies highly in different cancer samples. This observation suggests that drug interaction dynamics between the conditions can also variate.

Potentials of a few CDK-i molecules on noncanonical targets and cancer types have been highlighted previously [142], [143], [144], [145]. Relevantly, potential of roscovitine on GSK3B using *C. elegans* Alzheimer's disease models when combined with epibrassinolide were also previously evaluated [146]. In addition, in a recent study, KT-474 (SAR444656), an IRAK4 degrader shown as the first heterobifunctional degrader to provide initial proof of concept for activity outside of cancer [147]. Additionally, structural evaluations on promiscuity and noncanonical bindings of CDK

inhibitors were also discovered in the literature. For example, it was shown that abemaciclib potentially binds to DYRK1A and HIP kinase [148], while palbociclib pharmacological inhibits the PI5P4K alpha / beta and disrupts cell energy metabolism [149]. Also, interactions between PDXK and roscovitine and analogs also discovered previously [150]. Such findings, and with the results within this study again highlights the potentials of noncanonical effects of the CDK-i molecules in-and-outside of cancers.

4.1.4. CDK-is Show Binding to Possible Targets In Molecular Dynamics Simulations

When identified targets of CDK-is were analyzed via molecular dynamics simulations, it was observed that human target-drug interaction dynamics were relatively more stable when compared to the *C. elegans* protein-drug complexes (Figure: 3.13). This may be attributed to the origin of the structure data. Human protein structures were obtained from experimental results while *C. elegans* structures obtained from Alphafold. While Alphafold structure prediction methodology is highly reliable, it should be noted that local energy minimum and aminoacid sidechain placements can be wrong and high-energetic, resulting in unstable complexes at intra-regions in the overall protein [151] *in silico*.

Receptor screening identifies potential targets, and hierarchical interaction analysis prioritizes highly interacting targets for further investigation (Table Appendix-A.1). With receptor screening, enrichments, and MD methodologies, PRKCA, HADHA, PEX5, NOS2 and NTRK1 proteins have identified and computationally evaluated as highly potential CDK inhibitor targets in human. In *C. elegans*, primarily via receptor screening and enrichments, LET-756, DAF-4, DAF-2, PMK-1 and MEK-2 proteins were identified as possible targets for CDK-i drugs (Figure: 3.14). *C. elegans* CDK-4, homolog of CDK4 in human, is also evaluated for inhibition by abemaciclib, and found unlikely given the high binding free energy.

4.1.5. Abemaciclib, Palbociclib and Roscovitine Alter MAPK, FoxO and TGF-Beta Pathways in *C. elegans*

After computational approaches, it was logical to validate the used methodology *in vivo* using *C. elegans* model organism, as well as bridge the further relevant theoretical findings with real world observations. Therefore, abemaciclib, palbociclib, roscovitine and FUDR drugs were applied on *C. elegans* nematodes during both L1 and dauer phases at 1 μM concentrations.

Drug testing in *C. elegans* shows abemaciclib increases fat accumulation (Figure 2.4) without affecting larval development (Figure 3.17), while palbociclib and roscovitine delays larval development without impacting fat accumulation when worms develop throughout L1 to adulthood, and not during dauer to adulthood. This finding supports the idea that physiological changes, especially retardation of the body growth, is potentially due to alterations in the MAPK and/or TGF-beta signaling pathways by inhibition or combined hindrances of MEK-2 and/or DAF-4, by either roscovitine or palbociclib. DAF-4 is a key element in TGF-beta pathway, and responsible for entry into and exit from the dauer larval stage, normal adult lifespan and body size determination [152]. Additionally, it was previously stated that, *C. elegans* worms with defects in TGF-Beta signaling pathway can display *lon* (increased length) or *sma* (small body size) phenotypes [153], [154]. Therefore, it is likely that roscovitine inhibits DAF-4 at 1 μM concentration and hinders TGF-beta pathway. In turn, N2 strain *C. elegans* exhibits smaller body sizes during larval development compared to untreated worms.

When increased lipid content of abemaciclib treated worms were evaluated, it was previously discovered and explained that LET-756 mutations on *C. elegans* results in slow growth, and transparent looking animals. Additionally, partial loss-of-function of LET-756 exhibited as slow-growing, partially-fertile and uncoordinated worms [155]. This information contradicts with the findings within this study, and it is possible that various other reasons might result with increase in lipids rather than treatment of abemaciclib. However, if the enrichments of abemaciclib in *C. elegans* are re-evaluated, it can be seen that abemaciclib dataset also closely enriched for ”FoxO signaling pathway” (Figure: 3.8 -B). Additionally, when FoxO Signaling Pathway in network diagram (Figure: 3.10) and abemaciclib potential targets in hierarchy list of

C. elegans (Table Appendix-A.1) were inspected, DAF-18 and DAF-2 proteins can be seen as other possible targets of abemaciclib. In parallel to this, it was previously stated that, DAF-2 and DAF-18 proteins, as well as DAF-16, are important elements aging, metabolism, longevity and dauer formation [156], and DAF-2 mutants accumulate lipids prior to dauer formation [157]. In terms of possible effects of CDK-i's on *C. elegans*, inhibition of DAF-2 by abemaciclib can potentially trigger pre-dauer metabolic remodelling, therefore increase in lipid content while no significant variations in body size were observed.

Eventually, it is shown and discussed that, MEK-2, DAF-2 and DAF-4 are identified as possible targets of primarily palbociclib, abemaciclib, and roscovitine, respectively and as shown and explained, promiscuity of the CDK inhibitors highlight and remind the varied targets in both human and *C. elegans*.

Additionally, while drug treated *C. elegans* larvae shown developmental variations compared to control groups, additional research can highlight effects of these drugs on fertility, survival rates and stress resilience. However, such methods might require further and complicated understanding about bioavailability of such drugs.

4.1.6. *C. elegans* CDKs Required in Embryonic Stages

Human CDK4/6 homolog in *C. elegans* is CDK5 or CDK1 in terms of structural and sequence similarity. Therefore, in C. elegans drug treatment assays, it was hypothesized that, CDK-4, CDK-1 or CDK-5 protein inhibition should be considered canonical effect of the drugs, therefore upon profiling the levels of CDKs in the larval stages, CDK levels were lower compared to embryonic stages (Figure: 4.1), and treatments of the nematodes started from either L1 larvae or dauer Larvae.

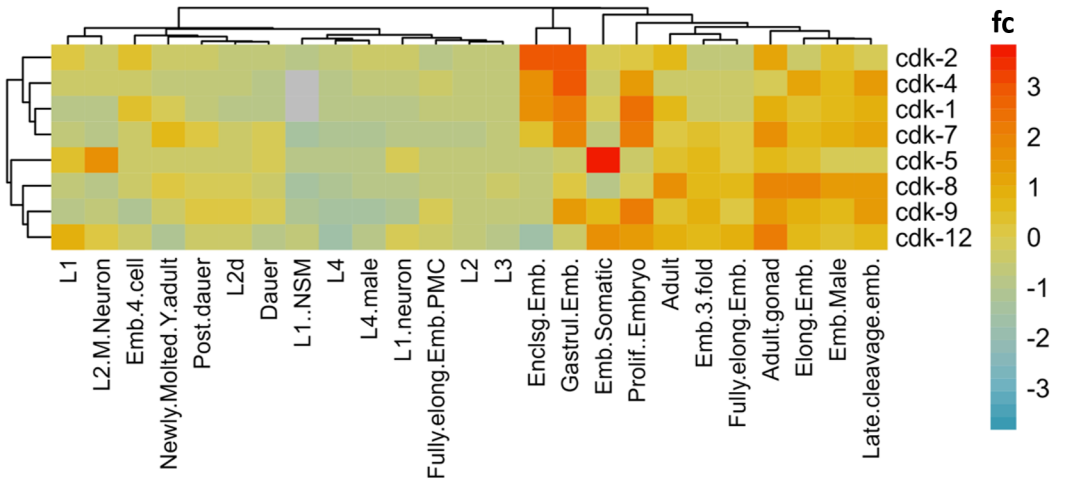


Figure 4.1: CDK levels in *C. elegans* larval stages. CDKs in general, expressed highy in Embriyonic stages.

Lastly, with the findings from computational approaches and observed physiological variations in *C. elegans*, several noteworthy conclusions about CDK-is can be discussed.

Firstly, *C. elegans* and human CDKs are evolutionary divergent, but show structural similarities and can be targeted by human CDK-is. Secondly, in *C. elegans*, cyclin-dependent kinases are not predominantly expressed after embryonic development is completed. CDK-I's might potentially show higher efficacies by interacting with noncanonical targets in other cancer or disease conditions than which their primary efficacies are evaluated. Also, receptor screening methodology, supported with various computational biology and biochemistry approaches can be used to highlight the biological interactions, side-effects and repurposing potentials of the CDK inhibitors and other drugs. Moreover, CDK inhibitors might be exhibiting the effects of non-canonical or off-target inhibitions when primary targets are either compensated in a condition that of redundancy; and/or these primary targets are minimally expressed in a condition that a secondary target is highly expressed.

Additionally, with the cumulated data, it is highlighted that, in order to ease the complexity in the studies regarding cancer or other diseases, as well as drug toxicology or drug response studies in *C. elegans*, evaluations and assays can be planned in a manner that target macromolecule of a drug should be available in the larval stage, and target macromolecule or protein should not be compensated upon inhibition. Also,

possible effects of the drug should be indicated by the *C. elegans*, or can be evaluated by molecular approaches.

4.1.7. CDK-is Might Be Repurposed in Other Cancers

Repurposing potential of the drugs on various cancers are also evaluated. In order to create an arbitrary scoring for potential effect of CDK inhibitors on cancers, simple calculation steps were accomplished. For this, Z-scores of the expressions of the targets were obtained and averaged, then ordered. Upon scoring the drugs against TCGA-OncoDB datasets, it is plausible that abemaciclib, roscovitine and ribociclib, have potential effects on cholangiocarcinoma (bile duct cancer) (Figure 4.2) and treatments with this drugs might result positive outcomes.

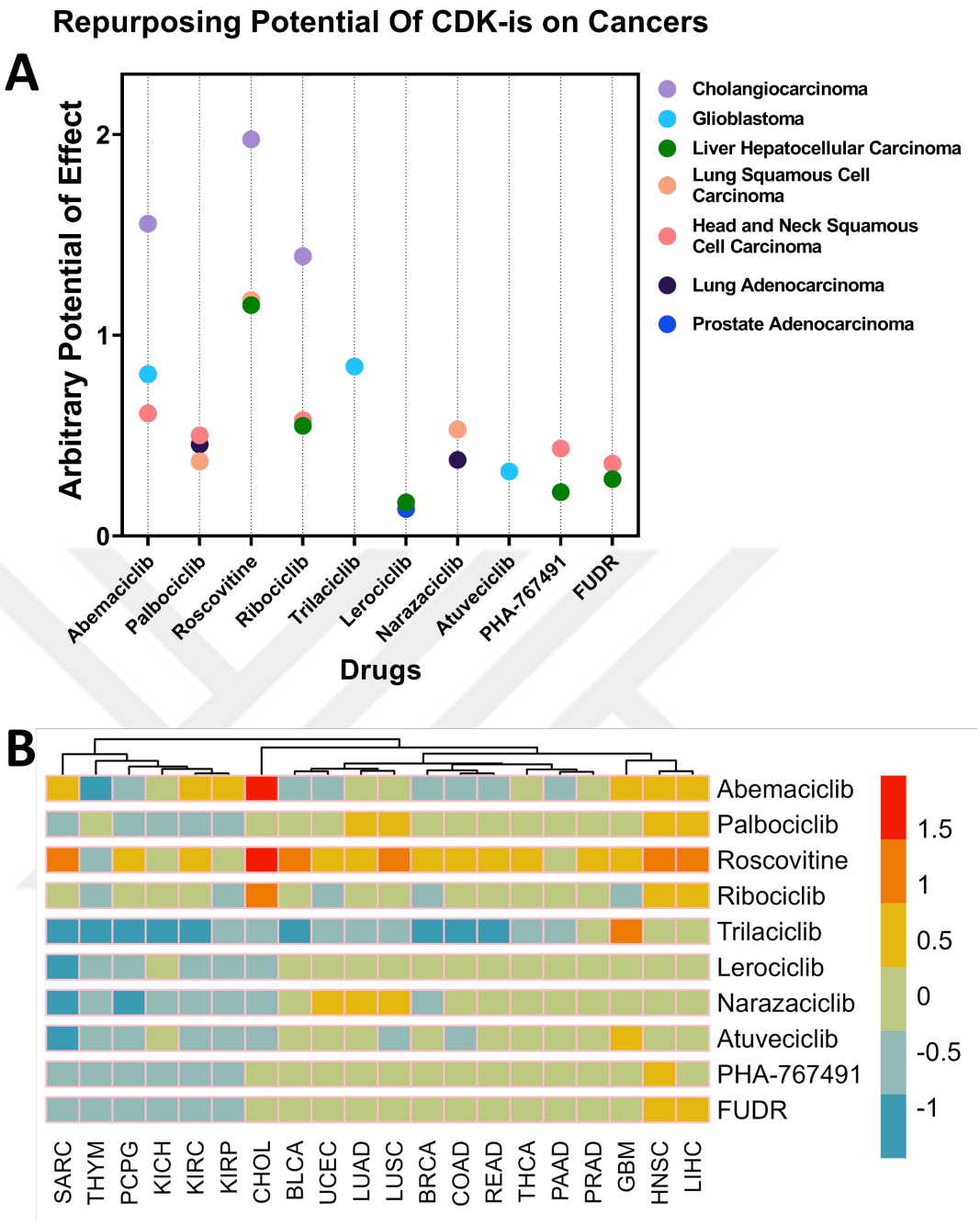


Figure 4.2: Arbitrary effect potential scores of CDK-I drugs against cancers. Potential targets, and hierarchically ordered other targets within the human datasets were evaluated using their expression levels in cancer samples and interpolated for possibility score by sorting best to worst, then discarding lowest scores. **A:** Each dot with various color represent cancer types. Higher score (vertically higher) represents theoretical potential of effect of CDK-I on the cancer type. **B:** Raw data of the arbitrary effect potential scores of CDK-i drugs against cancers. Lower values (blue) to higher values (red) indicate power of arbitrary potential of effect.

Apart from results and methodologies previously discussed, it is important to highlight that this study were consisted of many computational approaches and steps. While, computational (dry-lab) methodologies considered much cheaper to conduct in research compared to laboratory work in general, efficiency and time management is still paramount, and indicates that possible cost and energy savings still can be achieved. Within this study roughly 2654.87 kWh electrical energy was consumed for computation, resulting 1.1 – 1.3 t of Carbon Dioxide (CO_2) equivalent (calculated via epa.com [158]). Knowing that global pharmaceutical industry is 55% more carbon emission intense than the automotive industry [159], this study can represent the potential of *in silico* methods and efficient models in cancer and drug research.

4.2. Conclusions

In conclusion, our results suggest that inhibitors targeting cyclin-dependent kinases may engage in diverse interactions not only with other cyclin-dependent kinases but also with several other potential targets, leading to notable alterations in key pathways. Specifically, pathways related to ”metabolic” and ”biosynthesis of cofactors” were emphasized in human datasets, while ”MAPK”, ”FoxO” pathways, and developmental phenotypes were observed in *C. elegans*. Through an assessment of the most potent targets, we identified PRKCA, NOS2, HADHA in humans, and DAF-2 and DAF-4 in *C. elegans* as targets of CDK inhibitors. Upon treating *C. elegans* with few CDK-is, it was observed that palbociclib and roscovitine caused a delay in larval development, whereas abemaciclib led to an increase in lipid content, supporting the computational findings and indicating that CDK-is in fact can exhibit varied efficacies in other types of cancers.

REFERENCES

[1] Espe S., (2018), “MalaCards: The human disease database”, J. Med. Libr. Assoc.

[2] “Statistics | DrugBank Online — go.drugbank.com” (n.d.), `https://go.drugbank.com/stats`, [Accessed 08-01-2024].

[3] Fabricant D. S., Farnsworth N. R., (2001), “The value of plants used in traditional medicine for drug discovery.”, Environmental Health Perspectives, 109(suppl 1), 69–75.

[4] McElnay J. C., D’Arcy P. F., (1983), “Protein binding displacement interactions and their clinical importance”, Drugs, 25(5), 495–513.

[5] A T., Va B., (2018), “Molecular docking: From lock and key to combination lock”, J. Mol. Med. Clin. Appl.

[6] Patil R., Das S., Stanley A., Yadav L., Sudhakar A., Varma A. K., (2010), “Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing”, PLoS One, 5(8), e12029.

[7] Ramsay R., Tipton K., (2017), “Assessment of enzyme inhibition: A review with examples from the development of monoamine oxidase and cholinesterase inhibitory drugs”, Molecules, 22(7), 1192.

[8] Radhakrishnan M. L., Tidor B., (2007), “Specificity in molecular design: a physical framework for probing the determinants of binding specificity and promiscuity in a biological environment”, J. Phys. Chem. B, 111(47), 13419–13435.

[9] Rudmann D. G., (2013), “On-target and off-target-based toxicologic effects”, Toxicol. Pathol., 41(2), 310–314.

[10] Huggins D. J., Sherman W., Tidor B., (2012), “Rational approaches to improving selectivity in drug design”, J. Med. Chem., 55(4), 1424–1444.

[11] Cerisier N., Petitjean M., Regad L., Bayard Q., Réau M., Badel A., Camproux A.-C., (2019), “High impact: The role of promiscuous binding sites in polypharmacology”, Molecules, 24(14), 2529.

[12] Youdim M. B. H., Buccafusco J. J., (2005), “Multi-functional drugs for various CNS targets in the treatment of neurodegenerative disorders”, Trends Pharmacol. Sci., 26(1), 27–35.

[13] Wiseman A., (2003), “Do unpredictable toxic side-effects of drugs reflect their promiscuous binding energies?”, Trends Pharmacol. Sci., 24(2), 61–62.

[14] Lustberg M. B., Kuderer N. M., Desai A., Bergerot C., Lyman G. H., (2023), “Mitigating long-term and delayed adverse events associated with cancer treatment: implications for survivorship”, Nat. Rev. Clin. Oncol., 20(8), 527–542.

[15] Institute of Medicine (US) Roundtable on Environmental Health Sciences, Research, and Medicine (2002), “Cancer and the Environment: Gene-Enviroment Interaction”, National Academies Press (US), Washington (DC).

[16] “What is Cancer? — cancer.net” (n.d.), https://www.cancer.net/navigating-cancer-care/cancer-basics/what-cancer, [Accessed 09-01-2024].

[17] Debela D. T., Muzazu S. G., Heraro K. D., Ndalama M. T., Mesele B. W., Haile D. C., Kitui S. K., Manyazewal T., (2021), “New approaches and procedures for cancer treatment: Current perspectives”, SAGE Open Med., 9, 20503121211034366.

[18] Budhu A., Terunuma A., g G., Hussain S. P., Ambs S., Wang X. W., (2014), “Metabolic profiles are principally different between cancers of the liver, pancreas and breast”, Int. J. Biol. Sci., 10(9), 966–972.

[19] Hassanpour S. H., Dehghani M., (2017), “Review of cancer from perspective of molecular”, J. Canc. Res. Pr., 4(4), 127–129.

[20] Fisher R., Pusztai L., Swanton C., (2013), “Cancer heterogeneity: implications for targeted therapeutics”, Br. J. Cancer, 108(3), 479–485.

[21] Krzyszczyk P., Acevedo A., Davidoff E. J., Timmins L. M., Marrero-Berrios I., Patel M., White C., Lowe C., Sherba J. J., Hartmanshenn C., O’Neill K. M., Balter M. L., Fritz Z. R., Androulakis I. P., Schloss R. S., Yarmush M. L., (2018), “The growing role of precision and personalized medicine for cancer treatment”, Technology (Singap World. Sci.), 6(3-4), 79–100.

[22] Gambardella V., Tarazona N., Cejalvo J. M., Lombardi P., Huerta M., Roselló S., Fleitas T., Roda D., Cervantes A., (2020), “Personalized medicine: Recent progress in cancer therapy”, Cancers (Basel), 12(4), 1009.

[23] Bracci L., Schiavoni G., Sistigu A., Belardelli F., (2014), “Immune-based mechanisms of cytotoxic chemotherapy: implications for the design of novel and rationale-based combined treatments against cancer”, Cell Death Differ., 21(1), 15–25.

[24] “Cancer treatment - Mayo Clinic — mayoclinic.org” (n.d.), https://www.mayoclinic.org/tests-procedures/cancer-treatment/about/pac-20393344, [Accessed 09-01-2024].

[25] “How Radiation Therapy Is Used to Treat Cancer — cancer.org” (n.d.), https://www.cancer.org/cancer/managing-cancer/treatment-types/radiation/basics.html, [Accessed 09-01-2024].

[26] Koury J., Lucero M., Cato C., Chang L., Geiger J., Henry D., Hernandez J., Hung F., Kaur P., Teskey G., Tran A., (2018), “Immunotherapies: Exploiting the immune system for cancer treatment”, J. Immunol. Res., 2018, 1–16.

[27] Lindley C., McCune J. S., Thomason T. E., Lauder D., Sauls A., Adkins S., Sawyer W. T., (1999), “Perception of chemotherapy side effects cancer versus noncancer patients”, Cancer Pract., 7(2), 59–65.

[28] Corrie P. G., (2008), “Cytotoxic chemotherapy: clinical aspects”, Medicine (Abingdon), 36(1), 24–28.

[29] Tobias J. S., Hochhauser D., (2014), “Cancer and its Management”, 7 Edition, John Wiley & Sons, Nashville, TN.

[30] “Cancer Drugs — cancer.gov” (n.d.), https://www.cancer.gov/about-cancer/treatment/drugs, [Accessed 09-01-2024].

[31] Miller R. P., Tadagavadi R. K., Ramesh G., Reeves W. B., (2010), “Mechanisms of cisplatin nephrotoxicity”, Toxins (Basel), 2(11), 2490–2518.

[32] Volkova M., Russell 3rd R., (2011), “Anthracycline cardiotoxicity: prevalence, pathogenesis and treatment”, Curr. Cardiol. Rev., 7(4), 214–220.

[33] Lauschke V. M., Ingelman-Sundberg M., (2016), “The importance of patient-specific factors for hepatic drug response and toxicity”, Int. J. Mol. Sci., 17(10), 1714.

[34] “Imatinib (Glivec)” (n.d.), https://www.cancerresearchuk.org/about-cancer/treatment/drugs/imatinib, Accessed: 2024-1-15.

[35] “Trastuzumab” (2006), https://www.cancer.gov/about-cancer/treatment/drugs/trastuzumab, Accessed: 2024-1-15.

[36] “Targeted Therapy for Cancer — cancer.gov” (n.d.), https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies, [Accessed 09-01-2024].

[37] Wilkes G. M., (2018), “Targeted therapy: Attacking cancer with molecular and immunological targeted agents”, Asia Pac. J. Oncol. Nurs., 5(2), 137–155.

[38] Malumbres M., (2014), “Cyclin-dependent kinases”, Genome Biol., 15(6), 122.

[39] Liu J., Kipreos E. T., (2000), “Evolution of cyclin-dependent kinases (CDKs) and CDK-activating kinases (CAKs): Differential conservation of CAKs in yeast and metazoa”, Mol. Biol. Evol., 17(7), 1061–1074.

[40] Cao L., Chen F., Yang X., Xu W., Xie J., Yu L., (2014), “Phylogenetic analysis of cdk and cyclin proteins in premetazoan lineages”, BMC evolutionary biology, 14(1), 1–16.

[41] Cicenas J., Valius M., (2011), “The CDK inhibitors in cancer research and therapy”, J. Cancer Res. Clin. Oncol., 137(10), 1409–1418.

[42] Ding L., Cao J., Lin W., Chen H., Xiong X., Ao H., Yu M., Lin J., Cui Q., (2020), “The roles of cyclin-dependent kinases in cell-cycle progression and therapeutic strategies in human breast cancer”, Int. J. Mol. Sci., 21(6), 1960.

[43] Bertoli C., Skotheim J. M., de Bruin R. A. M., (2013), “Control of cell cycle transcription during G1 and S phases”, Nat. Rev. Mol. Cell Biol., 14(8), 518–528.

[44] Zetterberg A., Larsson O., Wiman K. G., (1995), “What is the restriction point?”, Current opinion in cell biology, 7(6), 835–842.

[45] Santamaría D., Barrière C., Cerqueira A., Hunt S., Tardy C., Newton K., Cáceres J. F., Dubus P., Malumbres M., Barbacid M., (2007), “Cdk1 is sufficient to drive the mammalian cell cycle”, Nature, 448(7155), 811–815.

[46] Ren S., Rollins B. J., (2004), “Cyclin c/cdk3 promotes rb-dependent g0 exit”, Cell, 117(2), 239–251.

[47] Shapiro G. I., (2006), “Cyclin-dependent kinase pathways as targets for cancer treatment”, J. Clin. Oncol., 24(11), 1770–1783.

[48] Montalto F. I., De Amicis F., (2020), “Cyclin D1 in cancer: A molecular connection for cell cycle control, adhesion and invasion in tumor and stroma”, Cells, 9(12), 2648.

[49] Ghafouri-Fard S., Khoshbakht T., Hussen B. M., Dong P., Gassler N., Taheri M., Baniahmad A., Dilmaghani N. A., (2022), “A review on the role of cyclin dependent kinases in cancers”, Cancer Cell Int., 22(1), 325.

[50] Liang X.-B., Dai Z.-C., Zou R., Tang J.-X., Yao C.-W., (2023), “The therapeutic potential of CDK4/6 inhibitors, novel cancer drugs, in kidney diseases”, Int. J. Mol. Sci.

[51] O’Leary B., Finn R. S., Turner N. C., (2016), “Treating cancer with selective CDK4/6 inhibitors”, Nat. Rev. Clin. Oncol., 13(7), 417–430.

[52] Shah M., Nunes M. R., Stearns V., (2018), “CDK4/6 inhibitors: Game changers in the management of hormone receptor–positive advanced breast cancer?”, Oncology (Williston Park), 32(5), 216–222.

[53] Yuan K., Wang X., Dong H., Min W., Hao H., Yang P., (2021), “Selective inhibition of CDK4/6: A safe and effective strategy for developing anticancer drugs”, Acta Pharm. Sin. B., 11(1), 30–54.

[54] Łukasik P., Baranowska-Bosiacka I., Kulczycka K., Gutowska I., (2021), “Inhibitors of cyclin-dependent kinases: Types and their mechanism of action”, Int. J. Mol. Sci., 22(6), 2806.

[55] Zhang M., Zhang L., Hei R., Li X., Cai H., Wu X., Zheng Q., Cai C., (2021), “CDK inhibitors in cancer therapy, an overview of recent development”, Am. J. Cancer Res., 11(5), 1913–1935.

[56] Ettl T., Schulz D., Bauer R. J., (2022), “The renaissance of cyclin dependent kinase inhibitors”, Cancers (Basel), 14(2), 293.

[57] “Abemaciclib” (n.d.), `https://medlineplus.gov/druginfo/meds/a617049.html`, Accessed: 2024-1-15.

[58] “Palbociclib” (n.d.), `https://medlineplus.gov/druginfo/meds/a615013.html`, Accessed: 2024-1-15.

[59] Cicenas J., Kalyan K., Sorokinas A., Stankunas E., Levy J., Meskinyte I., Stankevicius V., Kaupinis A., Valius M., (2015), “Roscovitine in cancer and other diseases”, Ann. Transl. Med., 3(10), 135.

[60] “Ribociclib” (n.d.), `https://medlineplus.gov/druginfo/meds/a617008.html`, Accessed: 2024-1-15.

[61] “Trilaciclib injection” (n.d.), `https://medlineplus.gov/druginfo/meds/a621020.html`, Accessed: 2024-1-15.

[62] PubChem (n.d.*a*), “Lerociclib”, `https://pubchem.ncbi.nlm.nih.gov/compound/Lerociclib`, Accessed: 2024-1-15.

[63] PubChem (n.d.*b*), "8-cyclopentyl-2-((4-(4-methylpiperazin-1-yl)phenyl)amino)-7-oxo-7,8-dihydropyrido[2,3-d]pyrimidine-6-carbonitrile", https://pubchem.ncbi.nlm.nih.gov/compound/56649281, Accessed: 2024-1-15.

[64] "ClinicalTrials.Gov" (n.d.), https://www.clinicaltrials.gov/study/NCT02345382, Accessed: 2024-1-15.

[65] Erbayraktar Z., Alural B., Erbayraktar R. S., Erkan E. P., (2016), "Cell division cycle 7-kinase inhibitor PHA-767491 hydrochloride suppresses glioblastoma growth and invasiveness", Cancer Cell Int., 16(1), 88.

[66] Sliwoski G., Kothiwale S., Meiler J., Lowe E. W., (2014), "Computational methods in drug discovery", Pharmacological reviews, 66(1), 334–395.

[67] Zhong F., Xing J., Li X., Liu X., Fu Z., Xiong Z., Lu D., Wu X., Zhao J., Tan X., et al. (2018), "Artificial intelligence in drug design", Science China Life Sciences, 61, 1191–1204.

[68] Yang Y., Adelstein S. J., Kassis A. I., (2012), "Target discovery from data mining approaches", Drug discovery today, 17, S16–S23.

[69] Morphy R., (2010), "Selectively nonselective kinase inhibition: striking the right balance", J. Med. Chem., 53(4), 1413–1437.

[70] Jumper J., Evans R., Pritzel A., Green T., Figurnov M., Ronneberger O., Tunyasuvunakool K., Bates R., Žídek A., Potapenko A., Bridgland A., Meyer C., Kohl S. A. A., Ballard A. J., Cowie A., Romera-Paredes B., Nikolov S., Jain R., Adler J., Back T., Petersen S., Reiman D., Clancy E., Zielinski M., Steinegger M., Pacholska M., Berghammer T., Bodenstein S., Silver D., Vinyals O., Senior A. W., Kavukcuoglu K., Kohli P., Hassabis D., (2021), "Highly accurate protein structure prediction with AlphaFold", Nature, 596(7873), 583–589.

[71] Nair P. C., Miners J. O., (2014), "Molecular dynamics simulations: from structure function relationships to drug discovery", In Silico Pharmacol., 2(1), 4.

[72] Hollingsworth S. A., Dror R. O., (2018), "Molecular dynamics simulation for all", Neuron, 99(6), 1129–1143.

[73] Zhou K., Liu B., (2022), "Molecular dynamics simulation", Elsevier Science Publishing, Philadelphia, PA.

[74] Jones D., Allen J. E., Yang Y., Drew Bennett W. F., Gokhale M., Moshiri N., Rosing T. S., (2022), "Accelerators for classical molecular dynamics simulations of biomolecules", J. Chem. Theory Comput., 18(7), 4047–4069.

[75] Meng X.-Y., Zhang H.-X., Mezei M., Cui M., (2011), "Molecular docking: a powerful approach for structure-based drug discovery", Curr. Comput. Aided Drug Des., 7(2), 146–157.

[76] Müller B., Grossniklaus U., (2010), "Model organisms–a historical perspective", J. Proteomics, 73(11), 2054–2063.

[77] Mendel G., Punnett R. C., Burndy Library, (1866), "Versuche über Pflanzen-Hybriden /", Im Verlage des Vereines,, Brünn :.

[78] Taormina G., Ferrante F., Vieni S., Grassi N., Russo A., Mirisola M. G., (2019), "Longevity: Lesson from model organisms", Genes (Basel), 10(7), 518.

[79] Martinez-Pacheco S., O'Driscoll L., (2021), "Pre-clinical in vitro models used in cancer research: Results of a worldwide survey", Cancers (Basel), 13(23), 6033.

[80] Leung M. C. K., Williams P. L., Benedetto A., Au C., Helmcke K. J., Aschner M., Meyer J. N., (2008), "Caenorhabditis elegans: an emerging model in biomedical and environmental toxicology", Toxicol. Sci., 106(1), 5–28.

[81] Bouyanfif A., Jayarathne S., Koboziev I., Moustaid-Moussa N., (2019), "The nematode caenorhabditis elegans as a model organism to study metabolic effects of ω-3 polyunsaturated fatty acids in obesity", Adv. Nutr., 10(1), 165–178.

[82] Maduro M. F., (2015), "Developmental robustness in the caenorhabditis elegans embryo", Molecular reproduction and development, 82(12), 918–931.

[83] Kobet R. A., Pan X., Zhang B., Pak S. C., Asch A. S., Lee M.-H., (2014), "Caenorhabditis elegans: A model system for anti-cancer drug discovery and therapeutic target identification", Biomol. Ther. (Seoul), 22(5), 371–383.

[84] Kyriakakis E., Markaki M., Tavernarakis N., (2015), "Caenorhabditis elegans as a model for cancer research", Mol. Cell. Oncol., 2(2), e975027.

[85] UniProt Consortium (2023), "UniProt: The universal protein knowledgebase in 2023", Nucleic Acids Res., 51(D1), D523–D531.

[86] Sievers F., Higgins D. G., (2018), "Clustal omega for making accurate alignments of many protein sequences", Protein Sci., 27(1), 135–145.

[87] Martínez L., Andreani R., Martínez J. M., (2007), "Convergent algorithms for protein structural alignment", BMC Bioinformatics, 8(1), 306.

[88] Zhang Y., Skolnick J., (2004), "Scoring function for automated assessment of protein structure template quality", Proteins, 57(4), 702–710.

[89] Varadi M., Anyango S., Deshpande M., Nair S., Natassia C., Yordanova G., Yuan D., Stroe O., Wood G., Laydon A., Žídek A., Green T., Tunyasuvunakool K., Petersen S., Jumper J., Clancy E., Green R., Vora A., Lutfi M., Figurnov M., Cowie A., Hobbs N., Kohli P., Kleywegt G., Birney E., Hassabis D., Velankar S., (2022), "AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models", Nucleic Acids Res., 50(D1), D439–D444.

[90] Database A. P. S., (n.d.), "AlphaFold Protein Structure Database — alphafold.ebi.ac.uk", `https://alphafold.ebi.ac.uk/faq#faq-12`, [Accessed 09-01-2024].

[91] Burley S. K., Bhikadiya C., Bi C., Bittrich S., Chao H., Chen L., Craig P. A., Crichlow G. V., Dalenberg K., Duarte J. M., Dutta S., Fayazi M., Feng Z., Flatt J. W., Ganesan S., Ghosh S., Goodsell D. S., Green R. K., Guranovic V., Henry J., Hudson B. P., Khokhriakov I., Lawson C. L., Liang Y., Lowe R., Peisach E., Persikova I., Piehl D. W., Rose Y., Sali A., Segura J., Sekharan M., Shao C., Vallat B., Voigt M., Webb B., Westbrook J. D., Whetstone S., Young J. Y., Zalevsky A., Zardecki C., (2023), "RCSB protein data bank (RCSB.org): delivery

of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning", Nucleic Acids Res., 51(D1), D488–D508.

[92] Eastman P., Swails J., Chodera J. D., McGibbon R. T., Zhao Y., Beauchamp K. A., Wang L.-P., Simmonett A. C., Harrigan M. P., Stern C. D., Wiewiora R. P., Brooks B. R., Pande V. S., (2017), "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics", PLoS Comput. Biol., 13(7), e1005659.

[93] Davies M., Nowotka M., Papadatos G., Dedman N., Gaulton A., Atkinson F., Bellis L., Overington J. P., (2015), "ChEMBL web services: streamlining access to drug discovery data and utilities", Nucleic Acids Res., 43(W1), W612–W620.

[94] Wishart D. S., Feunang Y. D., Guo A. C., Lo E. J., Marcu A., Grant J. R., Sajed T., Johnson D., Li C., Sayeeda Z., Assempour N., Iynkkaran I., Liu Y., Maciejewski A., Gale N., Wilson A., Chin L., Cummings R., Le D., Pon A., Knox C., Wilson M., (2018), "DrugBank 5.0: a major update to the DrugBank database for 2018", Nucleic Acids Res., 46(D1), D1074–D1082.

[95] Kim S., Chen J., Cheng T., Gindulyte A., He J., He S., Li Q., Shoemaker B. A., Thiessen P. A., Yu B., Zaslavsky L., Zhang J., Bolton E. E., (2023), "PubChem 2023 update", Nucleic Acids Res., 51(D1), D1373–D1380.

[96] O'Boyle N. M., Banck M., James C. A., Morley C., Vandermeersch T., Hutchison G. R., (2011), "Open babel: An open chemical toolbox", J. Cheminform., 3(1), 33.

[97] Bayly C. I., Cieplak P., Cornell W., Kollman P. A., (1993), "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model", J. Phys. Chem., 97(40), 10269–10280.

[98] Geidl S., Bouchal T., Raček T., Svobodová Vařeková R., Hejret V., Křenek A., Abagyan R., Koča J., (2015), "High-quality and universal empirical atomic charges for chemoinformatics applications", J. Cheminform., 7(1), 59.

[99] Hetényi C., van der Spoel D., (2002), "Efficient docking of peptides to proteins without prior knowledge of the binding site", Protein Sci., 11(7), 1729–1737.

[100] Koes D. R., Baumgartner M. P., Camacho C. J., (2013), "Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise", J. Chem. Inf. Model., 53(8), 1893–1904.

[101] List M., Alcaraz N., Dissing-Hansen M., Ditzel H. J., Mollenhauer J., Baumbach J., (2016), "KeyPathwayMinerWeb: online multi-omics network enrichment", Nucleic Acids Res., 44(W1), W98–W104.

[102] Ulitsky I., Shamir R., (2007), "Identification of functional modules using network topology and high-throughput data", BMC Syst. Biol., 1(1), 8.

[103] Batra R., Alcaraz N., Gitzhofer K., Pauling J., Ditzel H. J., Hellmuth M., Baumbach J., List M., (2017), "On the performance of de novo pathway enrichment", NPJ Syst. Biol. Appl., 3(1), 6.

[104] Oughtred R., Rust J., Chang C., Breitkreutz B.-J., Stark C., Willems A., Boucher L., Leung G., Kolas N., Zhang F., Dolma S., Coulombe-Huntington J., Chatr-Aryamontri A., Dolinski K., Tyers M., (2021), "The BioGRID database: A

comprehensive biomedical resource of curated protein, genetic, and chemical interactions", Protein Sci., 30(1), 187–200.

[105] Kolberg L., Raudvere U., Kuzmin I., Adler P., Vilo J., Peterson H., (2023), "g:profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update)", Nucleic Acids Res., 51(W1), W207–W212.

[106] Kanehisa M., Furumichi M., Tanabe M., Sato Y., Morishima K., (2017), "KEGG: new perspectives on genomes, pathways, diseases and drugs", Nucleic Acids Res., 45(D1), D353–D361.

[107] Shannon P., Markiel A., Ozier O., Baliga N. S., Wang J. T., Ramage D., Amin N., Schwikowski B., Ideker T., (2003), "Cytoscape: a software environment for integrated models of biomolecular interaction networks", Genome Res., 13(11), 2498–2504.

[108] Doncheva N. T., Morris J. H., Gorodkin J., Jensen L. J., (2019), "Cytoscape StringApp: Network analysis and visualization of proteomics data", J. Proteome Res., 18(2), 623–632.

[109] Jeong H., Mason S. P., Barabási A. L., Oltvai Z. N., (2001), "Lethality and centrality in protein networks", Nature, 411(6833), 41–42.

[110] Szklarczyk D., Gable A. L., Lyon D., Junge A., Wyder S., Huerta-Cepas J., Simonovic M., Doncheva N. T., Morris J. H., Bork P., Jensen L. J., Mering C. v., (2019), "STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets", Nucleic Acids Res., 47(D1), D607–D613.

[111] Schindelman G., Fernandes J. S., Bastiani C. A., Yook K., Sternberg P. W., (2011), "Worm phenotype ontology: integrating phenotype data within and beyond the c. elegans community", BMC Bioinformatics, 12(1), 32.

[112] Barretina J., Caponigro G., Stransky N., Venkatesan K., Margolin A. A., Kim S., Wilson C. J., Lehár J., Kryukov G. V., Sonkin D., Reddy A., Liu M., Murray L., Berger M. F., Monahan J. E., Morais P., Meltzer J., Korejwa A., Jané-Valbuena J., Mapa F. A., Thibault J., Bric-Furlong E., Raman P., Shipway A., Engels I. H., Cheng J., Yu G. K., Yu J., Aspesi Jr P., de Silva M., Jagtap K., Jones M. D., Wang L., Hatton C., Palescandolo E., Gupta S., Mahan S., Sougnez C., Onofrio R. C., Liefeld T., MacConaill L., Winckler W., Reich M., Li N., Mesirov J. P., Gabriel S. B., Getz G., Ardlie K., Chan V., Myer V. E., Weber B. L., Porter J., Warmuth M., Finan P., Harris J. L., Meyerson M., Golub T. R., Morrissey M. P., Sellers W. R., Schlegel R., Garraway L. A., (2012), "The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity", Nature, 483(7391), 603–607.

[113] Team E. G. E., (n.d.), "Experiment; Expression Atlas EMBL-EBI — ebi.ac.uk", `https://www.ebi.ac.uk/gxa/experiments/E-MTAB-2770/Downloads`, [Accessed 09-01-2024].

[114] Encyclopedia C. C. L., (2013), "RNA-seq of 1019 human cancer cell lines from the Cancer Cell Line Encyclopedia", `https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-2770`.

[115] Weinstein J. N., Collisson E. A., Mills G. B., Shaw K. R., Ozenberger B. A., Ellrott K., Shmulevich I., Sander C., Stuart J. M., (2013), "The cancer genome atlas pan-cancer analysis project", Nature genetics, 45(10), 1113–1120.

[116] Tang G., Cho M., Wang X., (2022), "Oncodb: an interactive online database for analysis of gene expression and viral infection in cancer", Nucleic Acids Research, 50(D1), D1334–D1339.

[117] arrayexpress@ebi.ac.uk Role: submitter Affiliation: European Bioinformatics Institute A. A. A. A. E., ArrayExpress A., (2010), "Deep sequencing of the Caenorhabditis elegans transcriptome using RNA isolated from various developmental stages under various experimental conditions RW0001 - uninfected worms", https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-2812.

[118] Abraham M. J., Murtola T., Schulz R., Páll S., Smith J. C., Hess B., Lindahl E., (2015), "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers", SoftwareX, 1-2, 19–25.

[119] Corp. N., (n.d.), "CUDA Toolkit - Free Tools and Training — developer.nvidia.com", https://developer.nvidia.com/cuda-toolkit, [Accessed 11-01-2024].

[120] Kagami L., Wilter A., Diaz A., Vranken W., (2023), "The ACPYPE web server for small-molecule MD topology generation", Bioinformatics.

[121] Valdés-Tresanco M. S., Valdés-Tresanco M. E., Valiente P. A., Moreno E., (2021), "Gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS", J. Chem. Theory Comput., 17(10), 6281–6291.

[122] Brenner S., (1974), "The genetics of caenorhabditis elegans", Genetics, 77(1), 71–94.

[123] Fielenbach N., Antebi A., (2008), "C. elegans dauer formation and the molecular basis of plasticity", Genes Dev., 22(16), 2149–2165.

[124] Cassada R. C., Russell R. L., (1975), "The dauerlarva, a post-embryonic developmental variant of the nematode caenorhabditis elegans", Dev. Biol., 46(2), 326–342.

[125] Wang L., Graziano B., Bianchi L., (2023), "Protocols for treating c. elegans with pharmacological agents, osmoles, and salts for imaging and behavioral assays", STAR Protoc., 4(2), 102241.

[126] Schneider C. A., Rasband W. S., Eliceiri K. W., (2012), "NIH image to ImageJ: 25 years of image analysis", Nat. Methods, 9(7), 671–675.

[127] Schrödinger, LLC (2015), The PyMOL molecular graphics system, version 1.8.

[128] "Scientific Image and Illustration Software | BioRender — biorender.com" (n.d.), https://biorender.com/, [Accessed 13-01-2024].

[129] Anfinsen C. B., (1973), "Principles that govern the folding of protein chains", Science, 181(4096), 223–230.

[130] Krissinel E., (2007), "On the relationship between sequence and structure similarities in proteomics", Bioinformatics, 23(6), 717–723.

[131] Liu Y., Beyer A., Aebersold R., (2016), "On the dependency of cellular protein levels on mRNA abundance", Cell, 165(3), 535–550.

[132] Laranjeiro R., Harinath G., Burke D., Braeckman B. P., Driscoll M., (2017), "Single swim sessions in c. elegans induce key features of mammalian exercise", BMC Biol., 15(1), 30.

[133] "UniProt-Statistics-Amino acid composition, uniprot.org" (n.d.), `https://www.uniprot.org/uniprotkb/statistics#amino-acid-composition`, [Accessed 14-01-2024].

[134] Wilkinson D. J., Hossain T., Hill D. S., Phillips B. E., Crossland H., Williams J., Loughna P., Churchward-Venne T. A., Breen L., Phillips S. M., Etheridge T., Rathmacher J. A., Smith K., Szewczyk N. J., Atherton P. J., (2013), "Effects of leucine and its metabolite β-hydroxy-β-methylbutyrate on human skeletal muscle protein metabolism", J. Physiol., 591(11), 2911–2923.

[135] "Gene Ontology resource" (n.d.), `https://geneontology.org/stats.html`, Accessed: 2024-1-14.

[136] Kim D. H., Liberati N. T., Mizuno T., Inoue H., Hisamoto N., Matsumoto K., Ausubel F. M., (2004), "Integration of caenorhabditis elegans MAPK pathways mediating immunity and stress resistance by MEK-1 MAPK kinase and VHP-1 MAPK phosphatase", Proc. Natl. Acad. Sci. U. S. A., 101(30), 10990–10994.

[137] Inoue H., Hisamoto N., An J. H., Oliveira R. P., Nishida E., Blackwell T. K., Matsumoto K., (2005), "The c. elegans p38 MAPK pathway regulates nuclear localization of the transcription factor SKN-1 in oxidative stress response", Genes Dev., 19(19), 2278–2283.

[138] Li W., Wang D., Wang D., (2018), "Regulation of the response of caenorhabditis elegans to simulated microgravity by p38 mitogen-activated protein kinase signaling", Sci. Rep.

[139] Gumienny T. L., Savage-Dunn C., (2013), "TGF-β signaling in c. elegans", WormBook, 1–34.

[140] Savage-Dunn C., Padgett R. W., (2017), "The TGF-β family in*Caenorhabditis elegans*", Cold Spring Harb. Perspect. Biol., 9(6), a022178.

[141] Goodman M. B., Savage-Dunn C., (2022), "Reciprocal interactions between transforming growth factor beta signaling and collagens: Insights from caenorhabditis elegans", Dev. Dyn., 251(1), 47–60.

[142] Chou A., Froio D., Nagrial A. M., Parkin A., Murphy K. J., Chin V. T., Wohl D., Steinmann A., Stark R., Drury A., Walters S. N., Vennin C., Burgess A., Pinese M., Chantrill L. A., Cowley M. J., Molloy T. J., Waddell N., Johns A., Grimmond S. M., Chang D. K., Biankin A. V., Sansom O. J., Morton J. P., Grey S. T., Cox T. R., Turchini J., Samra J., Clarke S. J., Timpson P., Gill A. J., Pajic M., (2018), "Tailored first-line and second-line CDK4-targeting treatment combinations in mouse models of pancreatic cancer", Gut, 67(12), 2142–2155.

[143] Franco J., Witkiewicz A. K., Knudsen E. S., (2014), "CDK4/6 inhibitors have potent activity in combination with pathway selective therapeutic agents in models of pancreatic cancer", Oncotarget, 5(15), 6512–6525.

[144] "ClinicalTrials.gov — clinicaltrials.gov" (n.d.), https://clinicaltrials.gov/study/NCT00555906#study-overview, [Accessed 14-01-2024].

[145] Altenburg J. D., Farag S. S., (2015), "The potential role of PD0332991 (palbociclib) in the treatment of multiple myeloma", Expert Opin. Investig. Drugs, 24(2), 261–271.

[146] Obakan Yerlikaya P., Arısan E. D., Coker Gurkan A., Okumus O. O., Yenigun T., Ozbey U., Kara M., Palavan Unsal N., (2021), "Epibrassinolide prevents tau hyperphosphorylation via GSK3β inhibition in vitro and improves caenorhabditis elegans lifespan and motor deficits in combination with roscovitine", Amino Acids, 53(9), 1373–1389.

[147] Ackerman L., Acloque G., Bacchelli S., Schwartz H., Feinstein B. J., La Stella P., Alavi A., Gollerkeri A., Davis J., Campbell V., McDonald A., Agarwal S., Karnik R., Shi K., Mishkin A., Culbertson J., Klaus C., Enerson B., Massa V., Kuhn E., Sharma K., Keaney E., Barnes R., Chen D., Zheng X., Rong H., Sabesan V., Ho C., Mainolfi N., Slavin A., Gollob J. A., (2023), "IRAK4 degrader in hidradenitis suppurativa and atopic dermatitis: a phase 1 trial", Nat. Med.

[148] Kaltheuner I. H., Anand K., Moecking J., Düster R., Wang J., Gray N. S., Geyer M., (2021), "Abemaciclib is a potent inhibitor of DYRK1A and HIP kinases involved in transcriptional regulation", Nat. Commun., 12(1), 6607.

[149] Chen S., Chandra Tjin C., Gao X., Xue Y., Jiao H., Zhang R., Wu M., He Z., Ellman J., Ha Y., (2021), "Pharmacological inhibition of PI5P4Kα/β disrupts cell energy metabolism and selectively kills p53-null tumor cells", Proc. Natl. Acad. Sci. U. S. A., 118(21), e2002486118.

[150] Tang L., Li M.-H., Cao P., Wang F., Chang W.-R., Bach S., Reinhardt J., Ferandin Y., Galons H., Wan Y., Gray N., Meijer L., Jiang T., Liang D.-C., (2005), "Crystal structure of pyridoxal kinase in complex with roscovitine and derivatives", J. Biol. Chem., 280(35), 31220–31229.

[151] Binbay F. A., Rathod D. C., George A. A. P., Imhof D., (2023), "Quality assessment of selected protein structures derived from homology modeling and AlphaFold", Pharmaceuticals (Basel), 16(12), 1662.

[152] Gunther C. V., Georgi L. L., Riddle D. L., (2000), "A caenorhabditis elegans type I TGF beta receptor can function in the absence of type II kinase to promote larval development", Development, 127(15), 3337–3347.

[153] Suzuki Y., Yandell M. D., Roy P. J., Krishna S., Savage-Dunn C., Ross R. M., Padgett R. W., Wood W. B., (1999), "A BMP homolog acts as a dose-dependent regulator of body size and male tail patterning in caenorhabditis elegans", Development, 126(2), 241–250.

[154] Watanabe N., Ishihara T., Ohshima Y., (2007), "Mutants carrying two sma mutations are super small in the nematode c. elegans", Genes Cells, 12(5), 603–609.

[155] Roubin R., Naert K., Popovici C., Vatcher G., Coulier F., Thierry-Mieg J., Pontarotti P., Birnbaum D., Baillie D., Thierry-Mieg D., (1999), “let-756, a c. elegans fgf essential for worm development”, Oncogene, 18(48), 6741–6747.

[156] Murphy C. T., Hu P. J., (2013), “Insulin/insulin-like growth factor signaling in c. elegans”, WormBook, 1–43.

[157] Xie M., Roy R., (2015), “AMP-activated kinase regulates lipid droplet localization and stability of adipose triglyceride lipase in c. elegans dauer larvae”, PLoS One, 10(6), e0130480.

[158] Us Epa O., (2015), “Greenhouse gas equivalencies calculator”.

[159] Greever C., Ramirez-Aguilar K., Connelly J., (2020), “Connections between laboratory research and climate change: what scientists and policy makers can do to reduce environmental impacts”, FEBS Lett., 594(19), 3079–3085.

BIOGRAPHY

Rıdvan Aziz AYAZ enrolled in the Molecular Biology and Genetics department in 2015. During his undergraduate studies, he conducted research on the computational and bioinformatic characterization, as well as the intracellular localization of a novel protein in the same department at Gebze Technical University. Additionally, he interned at the X-Ray Crystallography Laboratory, focusing on small molecule characterization and other analytical chemistry approaches in the Chemistry department.

After graduating in 2021, he joined the Biotechnology MSc program at the Institute of Biotechnology, Gebze Technical University, where he conducted relevant studies throughout his thesis and other projects. In his Master's, he received the TÜBİTAK BİDEB 2210-C National Priority Areas Master's Scholarship Program in recognition of his academic achievements.

Apart from his academic pursuits, he is a data science enthusiast, a *C. elegans* worm breeder, and has a passion for music.

PUBLICATIONS WITHIN the SCOPE of THESIS STUDY

Ayaz R.A., Arısan E.D., Özbil M., (2022), ”Structural Similarity of Cyclin Dependent Kinases In *Caenorhabditis elegans* and Human And Their Targeting By Selective Inhibitors”, 6th FBE Lisansüstü Araştırmalar Sempozyumu - Gebze Technical University, 1-2 June 2022

Ayaz R.A., Özbil M., Arısan E.D., (2023), ”The Identification of Potential Targets of Cyclin Dependent Kinase Inhibitors (CDK-i) Using *in silico* Methods And Confirmation With *C. elegans* and Human Cell Line Disease Models”, 7th FBE Lisansüstü Araştırmalar Sempozyumu - Gebze Technical University, 31 May-1 June 2023

APPENDICES

APPENDIX-A. HIERARCHICAL ORDER OF THE POTENTIAL TARGETS

Table A.1: Hierarchical order of the potential targets. Higher degree (upper in the target lists) targets are priority candidates for CDK inhibitors. **Dataset H**: Human datasets that correspond to *in silico* receptor screening. **Dataset C**: *C. elegans* datasets. **Degree:** Order of hierarchy that corresponds to number of protein interactions.

Dataset ^{H}	Targets	Degree	*Dataset* ^{H}	Targets	Degree	*Dataset* ^{C}	Targets	Degree	*Dataset* ^{C}	Targets	Degree
Abemaciclib	PRKCA	13	Lerociclib	NOS2	22	Abemaciclib	let-756	14	Lerociclib	let-60	24
	INPPL1	10		PEX5	20		daf-2	11		sos-1	12
	PRKCB	6		HSD17B4	7		let-60	10		sma-4	4
	HCK	6		ACSL3	5		lit-1	5		mek-2	3
	ARPC2	5		ACSL4	5		pmk-1	2		pmk-2	2
	NCF1	5		AGPS	5		tap-1	2		cyb-1	2
	ARPC1B	4		DHRS4	4		mek-2	1			
Palbociclib	SHMT2	87	Narazaciclib	GAPDH	90	Palbociclib	let-756	8	Narazaciclib	daf-2	19
	ENO1	77		PKM	70		mpk-1	8		let-756	13
	GAPDH	51		TKT	57		daf-2	5		mpk-1	12
	PKM	42		TPI1	45		mek-2	2		let-23	10
	FH	31		ALDOA	39		rskn-1	1		lit-1	7
	ENO2	26		PFKP	36					tap-1	3
	ALDOC	26		PGK1	36					mom-4	2
	DLD	26		DLD	36					svh-2	1
	PGK1	25		PHGDH	35						
Roscovitine	HSP90AA1	58	PHA-767491	EPRS1	15	Roscovitine	lin-35	20	PHA-767491	bar-1	125
	CDK2	16		CAD	11		sma-3	3		lit-1	39
	CDK1	12		VKORC1	10		daf-4	3		sel-12	30
	MAPK14	11		ADSL	9		sma-4	2		mom-2	30
	PRKACA	10		UROD	6		cul-1	2		pop-1	23
	PRKACB	7		SPR	6					dsh-2	21
	MAD2L2	5		AK2	6					sma-4	19
	PLK1	5		MAT2B	6					hmp-2	18
	MAPK11	4		GSS	5					pry-1	17
	CDC16	4		RDH11	5						
Ribociclib	NTRK1	402	Atuvecicilib	PEX5	14	Ribociclib	daf-2	79	Atuvecicilib	daf-2	52
	TP53	130		DECR2	3		let-60	59		let-60	48
	AKT1	49		FAR1	3		sem-5	53		sos-1	22
	YWHAE	47		DHRS4	2		let-23	48		mpk-1	22
	MAPK14	35		CROT	2		mpk-1	35		lit-1	16
	IRAK1	29		ACOX1	2		sos-1	34		sma-4	13
	ABL1	27		MPV17L	2		lit-1	28		pmk-1	12
	MAPK8	23		ACOX2	1		pmk-1	16		mek-2	4
Trilaciclib	ACSL3	7	FUDR	ADSL	16	Trilaciclib	let-756	15	FUDR	let-756	25
	HADHA	7		ADSS2	16		let-60	12		daf-2	16
	ECHS1	6		MTHFD1	15		lit-1	6		let-60	16
	ADH5	4		ALDH1B1	15		pmk-1	5		lit-1	8
	ACSL4	4		MAT2B	12		mom-4	2		mpk-1	6
	ACADVL	3		UGDH	12		mak-2	1		pmk-1	4
	ACADSB	3		SGK1	12		svh-2	1		mek-2	2
	EHHADH	2		UGP2	11		mkk-4	1		mkk-4	2

FULL NAMES OF THE IDENTIFIED TARGETS

Table A.2: Full names and functions of the identified targets in human and *C. elegans* for each dataset.

		Abemaciclib	Palbociclib	Roscovitine	Ribociclib	Trilaciclib	Lerociclib	Narazaciclib	Atuveciclib	PHA-767491	FUDR
Human	Target	PRKCA	SHMT2	MAPK14	NTRK1	HADHA	NOS2	GAPDH	PEX5	EPRS1	ADSS2
	Name	Protein kinase C alpha type	Serine hydroxymethyltransferase, mitochondrial	Mitogen-activated protein kinase 14	High affinity nerve growth factor receptor	Trifunctional enzyme subunit alpha, mitochondrial	Nitric oxide synthase, inducible	Glyceraldehyde-3-phosphate dehydrogenase	Peroxisomal targeting signal 1 receptor	Bifunctional glutamate/proline–tRNA ligase	Adenylosuccinate synthetase isozyme 2
	Function	ATP binding , calcium-dependent protein kinase C activity, enzyme binding , histone H3T6 kinase activity, integrin binding ,	chromatin binding , glycine hydroxymethyltransferase activity , mRNA 5'-UTR binding , mRNA regulatory element binding translation repressor activity, protein homodimerization activity , pyridoxal phosphate binding ,	ATP binding , enzyme binding , MAP kinase activity , MAP kinase kinase activity , mitogen-activated protein kinase p38 binding	ATP binding, GPI-linked ephrin receptor activity, identical protein binding, kinase binding, nerve growth factor binding	3-hydroxyacyl-CoA dehydratase activity, 3-hydroxyacyl-CoA dehydrogenase activity , acetyl-CoA C-acetyltransferase activity , enoyl-CoA hydratase activity , fatty-acyl-CoA binding	arginine binding, calmodulin binding, flavin adenine dinucleotide binding, nitric-oxide synthase activity	aspartic-type endopeptidase inhibitor activity, disordered domain specific binding, glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity, identical protein binding, microtubule binding, NAD binding, NADP binding, peptidyl-cysteine S-nitrosylase activity	enzyme binding, peroxisome matrix targeting signal-1 binding, peroxisome membrane targeting sequence binding , peroxisome targeting sequence binding, protein carrier chaperone, small GTPase binding	ATP binding, glutamate-tRNA ligase activity, GTPase binding, identical protein binding, proline-tRNA ligase activity, protein homodimerization activity, RNA stem-loop binding, zinc ion binding	adenylosuccinate synthase activity, GTP binding , magnesium ion binding, phosphate ion binding
C. elegans	Target	LET-756	MEK-2	DAF-4	DAF-2	PMK-1	SOS-1	LIT-1	MPK-1	BAR-1	LET-60
	Name	Protein let-756 (Lethal protein 756)	Dual specificity mitogen-activated protein kinase kinase mek-2 (MAP kinase kinase mek-2)	Cell surface receptor daf-4 (Abnormal dauer formation protein 4)	Insulin-like receptor (IR)(Abnormal dauer formation protein 2)	Mitogen-activated protein kinase pmk-1 (Stress-activated protein kinase pmk-1)	Son of sevenless homolog (Guanine nucleotide exchange factor for RAS)	Serine/threonine kinase NLK (Loss of intestine protein 1) (Nemo-like kinase)	Mitogen-activated protein kinase mpk-1 (MAP kinase sur-1)	Beta-catenin/armadillo-related protein 1 (Protruding vulva protein 1)	Ras protein let-60 (Abnormal cell lineage protein 34) (Lethal protein 60)
	Function	fibroblast growth factor receptor binding, growth factor activity, protein kinase binding, RNA polymerase II-specific DNA-binding	ATP binding, MAP kinase kinase activity, protein serine kinase activity,	ATP binding, BMP binding, transforming growth factor beta receptor activity, transmembrane receptor protein	ATP binding, insulin receptor activity, peptide hormone binding, protein kinase binding	ATP binding, MAP kinase activity, metal ion binding, protein kinase activity	guanyl-nucleotide exchange factor activity, protein heterodimerization activity	ATP binding, beta-catenin binding, MAP kinase activity, metal ion binding, protein kinase activity	ATP binding, MAP kinase activity, metal ion binding, protein serine kinase activity	alpha-catenin binding, cadherin binding, DNA-binding transcription factor binding, protein phosphatase binding	GDP binding , GTP binding, GTPase activity, phospholipase binding, protein serine/threonine kinase binding

APPENDIX-B. AMINOACID COMPOSITIONS OF DATASETS

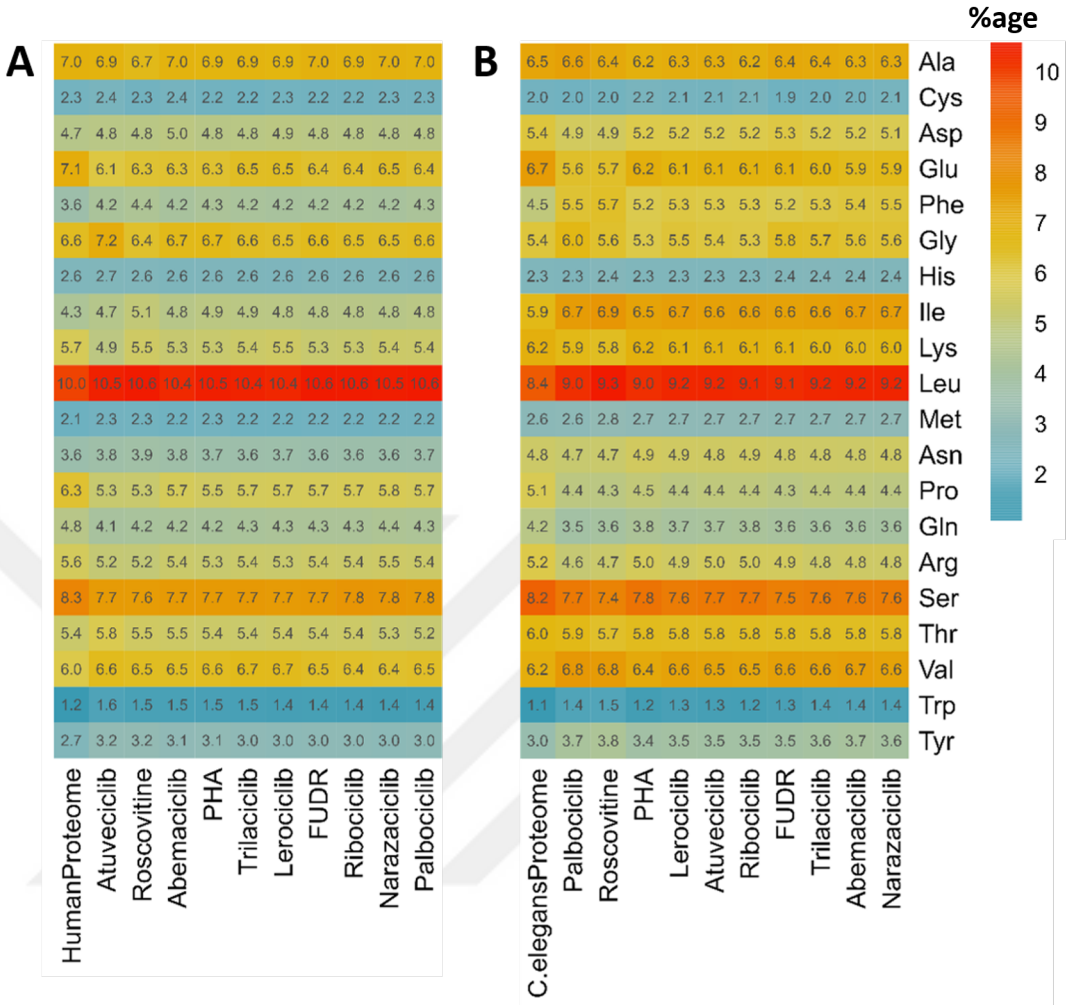


Figure B.1: Amino acid composition percentages of datasets. **A:** Human datasets, amino acid composition percentage heatmap. **B:** *C. elegans* datasets, amino acid composition percentage heatmap. Higher percentages were indicated towards red color. Proteome-wide amino acid percentages were placed at first columns.