Elucidating Hedgehog (Hh) Signaling Pathway Modulation via Pharmacological Strategies

Submitted by:

Darakhshan Irfan

61-FOC/MSBI/F22

Supervised by:

Dr. Attiya Kanwal

Co-Supervised by:

Dr. Noreen Akhtar

Department of Bioinformatics

Faculty of Computing and Information Technology

International Islamic University Islamabad, Pakistan

2025

Elucidating Hedgehog (Hh) Signaling Pathway Modulation via Pharmacological Strategies

Submitted by:

Darakhshan Irfan

61-FOC/MSBI/F22

Supervised by:

Dr. Attiya Kanwal

Co-Supervised by:

Dr. Noreen Akhtar

A thesis submitted in partial fulfillment of the

Requirement for the Degree of

MS in Bioinformatics

Department of Bioinformatics

Faculty of Computing and Information Technology

International Islamic University Islamabad, Pakistan

2025


بسم الله الرحمن الرحيم

DEDICATION

I dedicated my project to **ALLAH ALMIGHTY**. Who created us and made us able to use our abilities.

In addition, it gave us the strength to complete this task as our final year project &

This project is dedicated to our parents and our respected Supervisor.

DISSERTATION

A dissertation submitted to the department of Bioinformatics as a partial accomplishment of the requirements for the awards of the degree of Masters of Science in Bioinformatics (MS BI)

DECLARATION

I hereby declare that the work present is my own effort, except where otherwise acknowledged, and that the thesis is my own composition. No part of the thesis has been previously presented for any other degree.

Date _____

Darakhshan Irfan

(Reg. No. 61-FOC/MSBI/F-22)

ACKNOWLEDGEMENT

I am grateful to Almighty Allah, the most gracious and omnipotent, who has blessed us with the ability to complete our project report. I also express my respect and admiration for the Holy Prophet Muhammad (PBUH), who enlightened us with the spirit of one and true Allah and showed us humanity and a genuine way of life.

First and foremost, I am sincerely grateful to my supervisor, **Dr. Attiya Kanwal** for providing me with valuable guidance, support, advice, direction and encouragement throughout this research journey. Her expertise and mentorship have been a great source of inspiration, helping me grow both academically and professionally. Most importantly her unwavering support and encouragement in various forms have been a constant source of strength. Without her continuous support, this research project could not have been successfully completed.

I would like to express my special thanks to **Dr. Noreen Akhtar** without her contribution and support, completing my research could not have been possible. Her exceptional moral and professional support in completing my research. Furthermore, I am deeply thankful to **Dr. Shakira Ghazanfar** and her research assistant **Muhammad Yasir Akbar** for their mentorship and unwavering support during my final thesis. Her guidance was instrumental in bringing this work to completion and I am also grateful to the faculty and staff of the Department of Bioinformatics, their sincere guidance, untiring cooperation and endless inspiration that enabled us to overcome the entire problems during the course of our studies. Without her continuous support this study could not have been successfully conducted.

I am extremely grateful for my mother and father for their love, prayers, care and sacrifices in supporting my education. They have been there for me financially and morally and without their unwavering support, I would not have been able to complete my education.

Darakhshan Irfan

PROJECT IN BRIEF

The thesis entitled “Hedgehog (Hh) Signaling Pathway Modulation via Pharmacological Strategies” submitted by Darakhshan Irfan, Reg No. 61/FOC/MSBI/F22 in partial fulfillment of MS Degree in Bioinformatics has been completed under my guidance and supervision. I am satisfied by the quality of her work and allow her to submit this thesis for the further process of graduate with the Master of Science from the Department of Bioinformaics as per IIUI rules and regulation.

Dr. Attiya Kanwal
Thesis Certificate
Assistant Professor
Department of Bioinformatcs
Faculty of Computing & Information Technology
International Islamic University, Islamabad

ABSTRACT

The Hedgehog (Hh) signaling pathway plays a vital role in cellular differentiation, tissue formation, and repair processes. Dysregulation of this pathway is often linked to DNA hypermethylation, which can silence critical tumor suppressor genes, including Hedgehog-interacting protein (HHIP). This repression is primarily mediated by DNA methyltransferases (DNMTs) such as DNMT1, DNMT3A, and DNMT3B, leading to uncontrolled cellular proliferation and tumor progression. Despite the established roles of these enzymes in tumorigenesis, the specific DNMT responsible for inducing DNA hypermethylation that affects HHIP has not yet been identified. This gap in understanding highlights the need for targeted research aimed at elucidating the specific contributions of DNMTs to HHIP repression and cancer development, paving the way for the development of therapeutic strategies. This research aimed to employ an in silico approach to screen a library of 1,654,770 compounds from ChemBridge for potential DNMT inhibitors capable of restoring HHIP expression. A comprehensive methodology was utilized, including the generation of MACCS Key fingerprints to assess structural similarities with 13 known DNMT inhibitors. High-similarity compounds underwent rigorous filtration for drug-like characteristics, followed by molecular docking simulations against DNMT1, DNMT3A, and DNMT3B, with a specific focus on DNMT3A due to its prominent role in hypermethylation associated with cancer.The results of the docking simulations revealed significant binding affinities among the screened compounds, which were subsequently classified into eight clusters through hierarchical clustering based on chemical descriptors. Notably, Cluster 5 comprised 39 compounds with strong binding affinities to DNMT3A, suggesting their potential as selective inhibitors to reverse HHIP promoter hypermethylation. These findings underscore the therapeutic potential of targeting DNMT3A to reactivate HHIP expression and inhibit cancer progression. In conclusion, this study contributes to the understanding of DNMTs in cancer biology by identifying DNMT3A as a viable therapeutic target for reinstating HHIP expression. These 39 compounds identified in this research represent significant candidates for future clinical assessment, offering promising avenues for innovative cancer treatment strategies focused on reactivating crucial tumor suppressor pathways. This work lays the groundwork for further investigations into the clinical efficacy of these compounds and their potential roles in targeted cancer therapies.

TABLE OF CONTENTS

TABLE OF ABBRIEVATIONS

Hh	Hedgehog
HHIP	Hedgehog Interacting Protein
Shh	Sonic Hedgehog
Ihh	Indian Hedgehog
Dhh	Desert Hedgehog
PTCH1	Patched 1 (Hedgehog receptor)
SMO	Smoothened
GLI1	Glioma-associated oncogene homolog 1
GLI2	Glioma-associated oncogene homolog 2
GLI3	Glioma-associated oncogene homolog 3
SUFU	Suppressor of Fused
SBVS	Similarity-Based Virtual Screening
QSAR	Quantitative Structure-Activity Relationship
ML	Machine Learning
AI	Artificial Intelligence
RDKit	Rational Discovery Toolkit
SMILES	Simplified Molecular Input Line Entry System
SDF	Structure Data File
MOL2	Molecular structure file format
PDB	Protein Data Bank
MD	Molecular Dynamics

RMSD	Root Mean Square Deviation
PCA	Principal Component Analysis
k-Means	K-Means Clustering algorithm
ΔG	Binding free energy
IC50	Half maximal inhibitory concentration
HBA	Hydrogen Bond Acceptor
HBD	Hydrogen Bond Donor
LogP	Partition Coefficient (lipophilicity)
TPSA	Topological Polar Surface Area
MW	Molecular Weight
DNMT	DNA methyltransferase
MACCS keys	Molecular ACCess System

LIST OF FIGURES

LIST OF TABLES

Chapter# 01

Introduction

1.1 Background

1.1.1 Discovery of the Hedgehog Signaling Pathway

The Hedgehog (Hh) signaling pathway, which plays an essential role in embryonic development and cell differentiation, was originally discovered in the late 1970s during groundbreaking genetic research in the fruit fly, Drosophila melanogaster. Scientists Christiane Nüsslein-Volhard and Eric Wieschaus, examining gene mutations in body segmentation, found a sequence of genes controlling the ordered formation of embryos. Such a mutation created embryos encrusted in spiny outgrowths, akin to hedgehog bristles, thereby designating the gene "hedgehog" (hh). For this groundbreaking achievement, they received the Nobel Prize in Physiology or Medicine in 1995, an iconic step in the study of developmental biology.

1.1.2 Evolutionary Conservation and Function

It was soon after the hedgehog gene discovery in Drosophila that, in the early 1990s, researchers found vertebrate homologues in mammals and human beings as well. These homologs, Sonic Hedgehog (SHH), Indian Hedgehog (IHH), and Desert Hedgehog (DHH), were established to perform significant roles in several biological processes. The conservation of the Hedgehog pathway across diverse species underscored its fundamental role in the regulation of limb development, brain development, and organogenesis. This finding further drove the investigation of how Hh signaling works in different organisms, noting its evolutionary importance.

1.1.3 Mechanism and Clinical Relevance

By the mid-1990s, researchers had determined the molecular basis of the Hedgehog signaling pathway. It is initiated when Hh ligands bind to the receptor Patched (Ptch), releasing it from its repression of Smoothened (Smo). Activated Smo initiates a cascade of downstream signaling that controls the activity of Gli transcription factors, which, in turn, drive the expression of Hedgehog target genes. This canonical pathway was subsequently discovered to have non-canonical branches controlling processes such as tissue regeneration and stem cell maintenance. In addition, dysregulation of Hedgehog signaling was implicated in disease conditions such as congenital abnormalities and cancers such as basal cell carcinoma and medulloblastoma, making it a prime target for therapeutic interventions.

lies in the action of the Hedgehog proteins (ligands) upon specialized cell surface receptors to initiate a cascade of intracellular signalling events leading ultimately to impacted gene expression [4],[5]. It has been revealed that the pathway's function was investigated in vertebrates and comprised a number of elements. These are Hedgehog ligands (SHH, IHH, DHH), receptors: Patched (Ptch1 and Ptch2), Smoothened (Smo), and Gli transcription factors. The pathway has two main modes of functioning: 'off' and 'on', with the latter meaning the activity of downstream genes [6],[7].

1.1.4 The Off-State: Lack of Hedgehog Ligands

• Regulation of Smoothened by Patched

Without Hedgehog (Hh) ligands, the Hedgehog signaling pathway is in its inactive or "off" state, and cells maintain homeostasis, avoiding inappropriate activation. This is maintained chiefly by the transmembrane receptor Patched (Ptch), a negative regulator of the pathway. Patched suppresses the activity of another transmembrane protein, Smoothened (Smo), that is essential for propagating Hedgehog signaling. By limiting Smoothened's localization and activity, Patched keeps it confined to intracellular vesicles such that initiation of downstream signal cascades is avoided. This suppression maintains the pathway in a depressed state with minimal background activity when there are no Hedgehog ligands.

• Cytoplasmic Regulation of Gli Transcription Factors

With Smoothened inactive, a cytoplasmic protein complex strongly regulates the activity of Gli transcription factors (Gli1, Gli2, and Gli3), which act to mediate Hedgehog target gene expression. This regulatory complex consists of Suppressor of Fused (Sufu), Costal2 (Cos2), and multiple kinases, such as Protein Kinase A (PKA), Glycogen Synthase Kinase 3β (GSK3β), and Casein Kinase 1α (CK1α). These proteins collaborate to keep Gli proteins in an inactive state, blocking unwanted activation of the pathway.

An important step in this regulatory process is the phosphorylation of Gli2 and Gli3 by PKA, GSK3β, and CK1α. This phosphorylation targets Gli2 and Gli3 for proteasomal processing, which results in the generation of truncated repressor forms, Gli2R and Gli3R, devoid of

transcriptional activation domains. These repressor proteins then move into the nucleus, where they bind to Hedgehog target gene promoters, thereby inhibiting their transcription. [8],[9].

• **Physiological Significance and Disease Implications**

It is critical to keep the Hedgehog pathway in its off-state to preserve cellular quiescence and regulated development. The pathway is constructed to be in an inactive state until ligand stimulation, with the capability to rapidly and tightly regulate a response when necessary. Mutations in Ptch or Smo can cause inappropriate pathway activation, which can result in diseases such as basal cell carcinoma and medulloblastoma. Therefore, the off-state of the Hedgehog pathway is not only critical to normal cellular function but is also an important therapeutic target for controlling dysregulation of pathway activation in disease states. [10],[11].

1.1.5 The On-State: Hedgehog Ligand Activation

- **Initiation of Hedgehog Signaling**

In the presence of Hedgehog (Hh) ligands—Sonic Hedgehog (SHH), Indian Hedgehog (IHH), or Desert Hedgehog (DHH)—they induce a basic switch in the Hedgehog signaling pathway from the off-state to the on-state. This process starts with the binding of Hedgehog ligands to Patched (Ptch) cell surface receptors. Typically, Patched suppresses the transmembrane protein Smoothened (Smo) by blocking its localization and activation at the cell membrane. But when Patched binds to a ligand, it undergoes conformational changes that abolish its inhibitory activity. This enables Smoothened to move to the cell membrane, where it gets activated and triggers the intracellular signaling cascade.

- **Downstream Signal Propagation**

As soon as Smoothened becomes active, it coordinates a sequence of intracellular events that shatter the inhibitory protein complex made up of Suppressor of Fused (Sufu), Costal2 (Cos2), and kinases including Protein Kinase A (PKA), Glycogen Synthase Kinase 3β (GSK3β), and Casein Kinase 1α (CK1α). Without active Smoothened, this complex phosphorylates the Gli transcription factors (Gli2 and Gli3), marking them for proteasomal processing into their truncated repressor forms (Gli2R and Gli3R) [12],[13]. But when Smoothened is active, it destroys this regulatory complex, preventing phosphorylation and processing of the Gli proteins. Gli2 and Gli3 thus remain in their full-length active state, piling up in the cytoplasm before migrating to the nucleus.

- **Gene Expression and Physiological Implications**

Gli transcription factors, within the nucleus, modulate Hedgehog target gene expression. Gli2 is primarily a transcription activator, acting to enhance proliferation, differentiation, and survival genes. Gli3, however, plays two roles, with either activator or repressor function depending upon the cellular context. This bidirectional regulation gives precise control to Hedgehog signaling, ensuring accurate developmental and physiological response. Activation of the Hedgehog pathway is essential for embryonic development, tissue repair, and stem cell homeostasis. Dysregulation, through mutations in components of the pathway, overexpression of Hedgehog ligands, or inappropriate activation of Smoothened, can result in aberrant signaling and is responsible for many cancers and developmental diseases. Insight into the mechanisms that govern Hedgehog activation is vital to understanding normal cellular function as well as for developing therapeutic strategies for Hedgehog pathway-related disorders. [14],[15].

1.1.6 Nuclear Response: Regulation of Target Gene Expression

Within the nucleus, Hedgehog (Hh) signaling acts through the Gli transcription factors, mainly Gli2 and, occasionally, Gli3, that function as transcriptional activators. Upon activation, these proteins associate with specific DNA sequences in the promoters of Hedgehog target genes to induce their transcription. The target genes regulated by this pathway code proteins crucial for cell growth, survival, and differentiation. Some well-characterized target genes are Cyclin D, a cell cycle progressor, and BCL2, a pro-survival protein, to name just two. There are other Hedgehog-regulated genes that regulate tissue-specific growth, spatial organization, and cell differentiation, indicating the pathway's key role in embryonic growth and adult tissue maintenance.

The activation of Hedgehog target genes initiates a cascade of cellular processes integral to growth, morphogenesis, and spatial patterning. As an example, in limb formation, Hedgehog signaling is important for the initiation of the anteroposterior axis such that various components of the limbs develop in their proper spatial context. Likewise, in the central nervous system, Hedgehog signaling regulates neuronal differentiation and patterning, which are important for establishing a functional nervous system. [16],[17].

- **Regulation and Pathological Consequences**

Transcriptional activity of Gli proteins is closely regulated in order to provide the proper intensity and duration of Hedgehog signaling. Even minor fluctuations in this regulation can cause cell fate choices to drastically change. One of the characteristic features of the Hh

pathway is its gradient-dependent response, i.e., distinct levels of Hedgehog ligands trigger distinct sets of target genes. At high concentrations, ligands typically activate genes that promote cell proliferation, whereas at low concentrations, they induce differentiation or other context-specific responses. This very specific regulation guarantees orderly development of tissues and organs during embryogenesis and adult tissue homeostasis.

Abnormalities in this mechanism can result in disease. Mutations in Gli transcription factors or their upstream regulators could lead to the uncontrolled activation of target genes, which in turn leads to cancers like basal cell carcinoma and medulloblastoma. With its deep impact on cell growth and differentiation, the Hedgehog pathway continues to be a prime candidate for therapeutic development. Abnormal Gli activity is a potential target of therapeutic interventions for the treatment of Hedgehog-related diseases and thus a significant area of research in developmental biology and clinical medicine. [18],[19].

1.1.7 Termination and Fine-Tuning of the Hedgehog Signal

The Hedgehog signaling pathway involves advanced mechanisms for termination of signal and feedback regulation to achieve a delicate balance between activation and repression. These are essential to guarantee that pathway activity is properly modulated, avoiding prolonged or excess signaling that would interfere with cellular processes.A major element of this regulatory system is the Hedgehog Interacting Protein (HHIP), a target gene induced upon Hedgehog signaling. HHIP functions as a strong negative regulator by binding directly to Hedgehog ligands (SHH, IHH, DHH) and diminishing their availability for Patched (Ptch) receptor binding. This creates a feedback loop that dampens further pathway activation so that Hedgehog signaling is not prolonged beyond its necessary duration. HHIP's function is especially critical in tissues in which careful regulation of ligand concentration is required to preserve adequate developmental gradients [20],[21].In addition to HHIP, there are other pathway elements that play important roles in restoring the off-state when Hedgehog ligands are absent. For instance, Patched itself, the principal receptor of the pathway, is upregulated when the pathway is activated. This heightened expression of Patched strengthens its capacity to sequester Smoothened (Smo) and inhibit downstream signaling. Likewise, the repressor version of Gli3 (Gli3R) and the scaffold protein Suppressor of Fused (Sufu) assist in the restoration of pathway inhibition by re-activating the cytoplasmic repressive complex. This reactivates the phosphorylation and proteasomal processing of Gli2 and Gli3, transforming them back into their truncated repressor forms (Gli2R and Gli3R).

The pathway's capacity to dynamically switch between activation and inhibition is at the heart of its role in development and tissue homeostasis. This equilibrium guarantees that cells react properly to developmental signals while preventing aberrant signaling that may result in pathological states like cancer or congenital malformations. For instance, the inability to interrupt the signal properly has been associated with conditions such as basal cell carcinoma, where aberrant activation of the Hedgehog pathway is responsible for abnormal cell proliferation. On the other hand, improper activation due to signal propagation or termination mechanism defects may lead to developmental defects [22],[23].

1.1.8 Dysregulation and Pathological Implications

Pathological consequences can arise from dysregulation of the Hedgehog (Hh) signaling pathway since excessive and deficient activation causes disruption of normal cellular function and tissue development. Overactivation of Hh signaling is one of the major drivers in many cancers, such as basal cell carcinoma, medulloblastoma, pancreatic cancer, and lung cancer, where unrestricted pathway activity supports excessive cell proliferation, increases survival, and stimulates tumor development. This dysregulation is frequently produced by mutations within key regulatory elements, including PTCH1, a negative regulator that, in mutation, will not suppress Smoothened (SMO), resulting in constant pathway activation despite the absence of ligand. Likewise, constitutive activation of the pathway can be generated by activating SMO mutations and thus drive oncogenesis.

Conversely, too little Hedgehog signaling can cause developmental disorders, including holoprosencephaly, a birth defect in which abnormal signaling causes failure to separate the forebrain and midline facial structures. This is usually caused by mutations in SHH or its downstream effectors, which interfere with normal brain and facial development. The key role of the Hedgehog pathway in both cancer and congenital diseases underscores the necessity for strict regulation in ensuring normal cellular process. [24]

- **Therapeutic Strategies and Challenges**

Since it plays such a vital part in cancer, the Hedgehog pathway is now a significant focus for drug development. Inhibitors of Smoothened (SMO), including vismodegib and sonidegib, have been demonstrated to treat cancers that are driven by aberrant Hh activation. These inhibitors inhibit SMO activity, thus preventing overactivity of Hedgehog signaling and inhibiting tumor growth.[25].

A substantial drawback to treatment, though, is the development of drug resistance. Resistance

is commonly caused by SMO mutations, which make the inhibitors futile, or by activation of backup pathways that evade the block and maintain tumor growth. This highlights the importance of an enhanced understanding of Hedgehog pathway regulation in various cellular contexts. New therapies in the future may target combination treatments to multiple locations along the pathway or with other signaling pathways. These strategies seek to counteract resistance, increase treatment effectiveness, and offer more comprehensive treatments for controlling the varied pathological effects of Hedgehog pathway deregulation.

Aside from targeted therapies, research is also continuing to investigate alternative methods of modulating Hedgehog signaling in developmental disease and cancer. Small-molecule inhibitors of pathway downstream effectors, including Gli transcription factors, are being investigated to circumvent Smoothened inhibitor resistance mechanisms. In addition, immunotherapy and combination therapy, including Hedgehog pathway inhibitor plus standard chemotherapy or radiation, are being investigated to further improve the efficacy of treatment. [26]

A. **B.**

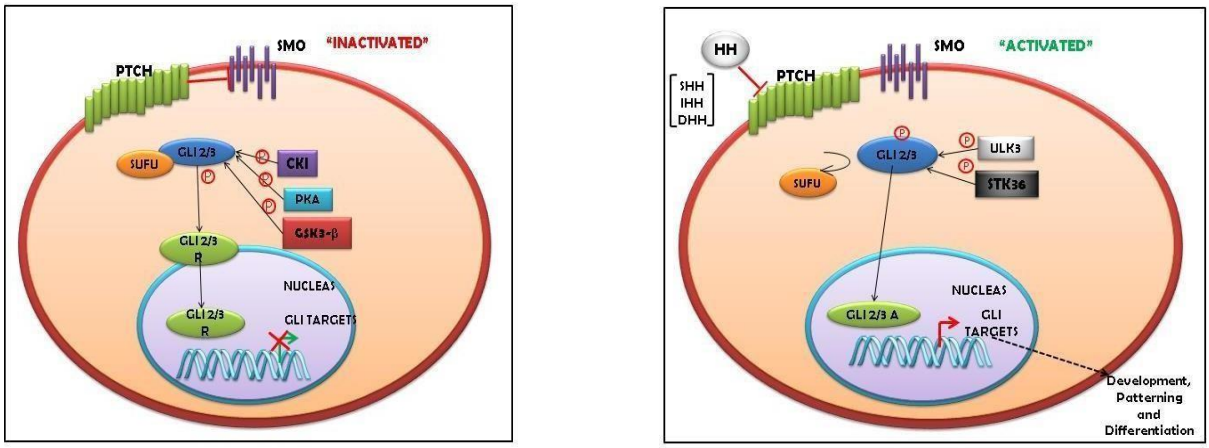

Figure 1.1: Hedgehog (Hh) Signaling Pathway

A: In the absence of the Hh ligand, The Patched (PTCH) inhibits the activity of Smoothened (SMO). At the same time, full-length Glioma-associated oncogenes 2/3 (Gli2/3) are kept in the cytosol by Suppressor of Fused (SUFU) where they undergo some N-terminal truncation and are partially phosphorylated and proteolysed by a three protein kinase melange of PKA, GSK3β and CK1α. The process ends and gives rise to the repressor form, Gli2/3 R, which moves into the nucleus and binds to the Hh target genes, silencing their expression.

B: On the other hand, the Hh ligand brings about the unbinding of the inhibitory effect of Patched (Ptch) on Smoothened (Smo). This brings about the subsequent inhibition of Sufu and the release of Gli2/3FL. Phosphorylation by a serine/threonine protein kinase like STK36 and ULK3 gives rise to the active form of Gli2/3FL, the Gli2/3 A. The regulator moves to the nucleus and activates Hh target genes.

1.2 Hedgehog Pathway's Role in Human Patterning and Morphogenesis of the Body

The Hedgehog (Hh) signaling pathway is a critical regulator of the spatial patterning, growth, and differentiation of tissues and organs during embryonic development. Its activities span a broad range of developmental processes to ensure proper body structure formation and function. The pathway acts as a basic mechanism for delivering positional information to cells, which allows them to take on particular identities and to play a role in organ and tissue morphogenesis. Hedgehog signaling plays a key role in establishing body axes, including the anterior-posterior and dorsal-ventral axes, in early developmental stages, directing the patterning of embryonic structures. One of the best-studied of its functions is in central nervous system development, where Sonic Hedgehog (SHH) guides the differentiation of neural progenitors into specialized neuronal types along a gradient of concentration. SHH signaling also plays a role in limb formation, with zone of polarizing activity (ZPA) SHH establishing the anteroposterior patterning of digits and thus normal limb formation.

The pathway is also involved in organogenesis and modulates the morphogenesis of the lungs, heart, pancreas, and skeleton through the control of cell proliferation, migration, and differentiation. Aberrations in Hedgehog signaling during embryonic development may lead to very serious congenital malformations, including holoprosencephaly, limb malformations, and craniofacial defects, emphasizing its fundamental role in human morphogenesis. Outside embryonic development, Hedgehog signaling continues to function in adult tissues, with roles in tissue repair, stem cell maintenance, and homeostasis. Yet, its deregulation can give rise to disease, such as cancer and degenerative disorders. For the sake of furthering developmental biology and regenerative medicine, it is therefore important to know the specific mechanisms of Hedgehog-mediated patterning. [27],[28].

1.2.1 Body Axis Patterning

- **Role of Hedgehog Signaling in Body Axis Patterning**

The Hedgehog (Hh) signaling pathway is important for body axis patterning, which is a general process in the development of vertebrates. It determines both the anterior-posterior (head-to-tail) and dorsal-ventral (back-to-belly) axis, providing accurate spatial organization and arrangement of the body structures. The Sonic Hedgehog (SHH) protein, which is a central figure in the Hh pathway, is especially essential in this regard. By establishing signaling gradients, SHH specifies cell position to cells and directs their fate and function during various developmental processes.

- **Anterior-Posterior Axis Patterning**

SHH signaling plays a critical role in anterior-posterior axis patterning, specifically in limb development. The zone of polarizing activity (ZPA) within the developing limb bud is the major source of SHH, establishing a gradient that guides the differentiation of unique digit identities. High SHH levels induce posterior digit formation, e.g., the pinky finger. Low SHH levels pattern anterior digits, e.g., the thumb. This gradient-based process allows each digit to develop in its proper position with proper morphology and avoid malformations like polydactyly (additional fingers or toes).

- **Dorsal-Ventral Axis Patterning**

Besides anterior-posterior patterning, SHH also is the key in dorsal-ventral axis specification, most notably in the neural tube, which generates the central nervous system. SHH is produced by the notochord and floor plate to create a gradient that directs neural progenitor cell differentiation. A high concentration of SHH in ventral areas causes the generation of motor neurons. Lower levels of SHH in dorsal areas encourage the generation of interneurons.

This specific regulation of SHH signaling guarantees the appropriate differentiation of neural cells, which helps in the establishment of a functional nervous system.

1.2.2 Organ Morphogenesis

In addition to its established function in axis patterning, the Hedgehog (Hh) signaling pathway plays an important role in organogenesis, a process in which organs form and acquire their functional structure during embryonic development. One of the better-documented instances of this process is lung development, wherein the Hh pathway controls boundary formation and

branching morphogenesis. These processes sculpt the complex shape of the respiratory apparatus to properly develop airways and alveoli, which are critical for effective gas exchange. [35],[36].

- **Branching Morphogenesis**

Branching morphogenesis is a basic developmental process in which an initial simple epithelial bud cycles through branching and elongation to produce a sophisticated tubular network. This is especially clear in lung development, where the bronchial tree needs to be precisely patterned to produce a functional respiratory apparatus. The Hh pathway is a key regulator of this process by controlling the proliferation, differentiation, and spatial arrangement of progenitor cells within the developing lung bud.

During early lung development, Sonic Hedgehog (Shh), one of the most important ligands of the Hh pathway, is present in epithelial cells of the lung bud and functions as a signaling molecule that acts on the surrounding mesenchymal cells. The interaction induces proliferation of progenitor cells, thereby providing an adequate number of cells to facilitate growth of bronchial and alveolar structures. In addition, Shh signaling also controls the differentiation of these progenitor cells into specialized epithelial and mesenchymal subtypes, which give rise to different parts of the lung. Through the control of these cell processes, the Hh pathway ensures that bronchi, bronchioles, and alveoli are in the correct structure and position for effective airflow and gas exchange within the mature lung. [37], [38].

- **Boundary Formation**

Aside from airway branching guidance, the Hh pathway plays an important role in boundary formation, where various cell populations within the developing lung are correctly segregated. This prevents structural and functional disruption of the respiratory system, especially during epithelial and mesenchymal compartment formation.

During the development of lungs, epithelial cells give rise to the inner lining of the airways and alveoli, and the mesenchymal cells form the surrounding connective tissue, such as smooth muscle and cartilage. The Hh pathway sets clear boundaries between cell types through regulation of gene expression patterns that govern cell identity and tissue compartmentalization. This segregation is critical for the development of functional alveoli,

which serve as the primary sites of oxygen and carbon dioxide exchange, as well as the formation of bronchial tubes, which conduct air to and from the lungs. [39],[40].

1.2.3 Sustaining Adult Tissue Homeostasis

The Hedgehog pathway activity is low or inhibited in the tissues of an adult in order not to trigger unwanted growth and cell division. It is regulated strictly so as not to result in growth at all levels of the tissues and tumors, yet cause tissues to be healed without promoting tissue homeostasis. Some adult tissue cancers and some developmental conditions are, however, linked to abnormal activation of the Hh pathway. Disruption of the pathway's normal regulatory processes during cell cycle progression, differentiation, and apoptosis (programmed cell death) leads to uncontrolled cell division and tumor development [41].

1.2.4 Implications in Cancer

Overactivation of the Hh pathway has been implicated in the development and progression of many cancers:

- **Brain Tumors:** In medulloblastoma, a common childhood brain tumor, mutations in the components of the Hh pathway, such as PTCH1 or SMO, result in constitutive activation of the pathway, which facilitates tumorigenesis [42].
- **Gastrointestinal Cancers:** Abnormal Hh signaling is involved in the etiology and pathogenesis of pancreatic cancer, where it contributes to stromal growth and forms a tumor-promoting microenvironment [43].
- **Lung Cancer:** Small cell lung carcinoma was found to be associated with enhanced activity of the Hh pathway that supports tumor progression and metastasis [44].
- **Breast and Prostate Cancers:** Abnormal Hh signaling is responsible for oncogenic cell growth and survival and drug resistance [45].

1.2.5 Therapeutic Potential

With its central roles in development and disease, the Hedgehog pathway is an attractive therapeutic target in oncology and regenerative medicine. Pathway component inhibitors like Smoothened (SMO) inhibitors are being identified and tested in the clinic for tumors with impaired Hh signaling. Examples include vismodegib and sonidegib, which are useful in the treatment of basal cell carcinoma and other cancers driven by Hh pathway dysregulation [4].

1.3 Types of Hedgehog Proteins

In mammals, the Hedgehog signaling family comprises three main ligands: Sonic Hedgehog (SHH), Indian Hedgehog (IHH), and Desert Hedgehog (DHH), each with specific expression patterns and distinct functions in various tissues.

1.3.1 Sonic Hedgehog (SHH)

Sonic Hedgehog (SHH) is the most widely studied and functionally diverse member of the Hedgehog (Hh) family. It has a pivotal function in regulating key developmental events in embryogenesis, such as nervous system development, patterning of limbs, and brain formation. The SHH signaling pathway serves as a molecular organizer that sets up morphogen gradients giving spatial and temporal signals for tissue and organ development. The flexibility and importance of SHH make it a key ingredient in developmental biology. The following is an elaborated explanation of its principal roles:

- **Neural Tube Patterning:**

Perhaps the best-studied function of SHH is in ventral neural tube patterning, wherein it sets up a gradient important for defining the positional identity of progenitor cells along the dorsal-ventral axis. SHH is produced by the notochord and subsequently by the floor plate of the neural tube, generating a concentration gradient that impacts the differentiation of progenitor cells into subtypes of neurons. High levels of SHH at the ventral midline define motor neuron precursors, intermediate levels induce differentiation of interneurons, and low levels are used to pattern dorsal structures.The positional identity of such neurons is paramount in determining functional structures within spinal cord and brainstem. Motor neurons that develop in ventral areas, for instance, relay signals to muscle and are the ones responsible for movement. Aberrations in SHH signaling at this stage may result in defects of the neural tube, inappropriate neuron specification, or even congenital anomalies like spina bifida or holoprosencephaly, depending on the nature of the aberration [15]

- **Limp Development:**

SHH is also instrumental in the development of limbs, specifically the formation of the anterior-posterior axis. It is produced by the zone of polarizing activity (ZPA), a specialized group of cells positioned at the posterior edge of the limb bud. SHH creates a concentration gradient that patterns this axis, specifying the identity of the digits.High SHH levels encourage posterior digit development, like that of the little finger, and low levels specify anterior digits,

like the thumb. This gradient-based system specifies the correct formation and spatial positioning of the digits. Aberrations in SHH signaling or gradient formation lead to limb malformations including:

1. **Polydactyly:** The occurrence of additional digits because of overactivity of SHH signaling.
2. **Syndactyly:** The fusion of digits, which may be associated with unregulated SHH signaling.
3. **Ectrodactyly:** The lack of some digits, which could occur because of a lack of SHH signaling.

These abnormalities highlight the necessity of accurate SHH activity in embryonic development [16] [17].

- **Brain Development:**

SHH also has a key role to play in the brain's development, specifically the patterning of the forebrain and midbrain and the formation of particular structures such as the cerebellum. In cerebellar development, SHH controls the granule cell precursor proliferation that is critical for cerebellar layered structure. This regulation guarantees the cerebellum attains functional ability in coordination of movement, balance, and cognitive processes [18]. SHH signaling dysfunction during brain formation may cause significant congenital illness. For example, holoprosencephaly is characterized by the failed separation of the forebrain into two hemispheres. This condition frequently leads to a range of neurological and craniofacial malformations, from mild facial asymmetry to more severe ones like cyclopia (one central eye). These defects are frequently associated with mutations in genes that code for components of the SHH signaling pathway or its downstream effectors. Aside from its function in embryogenesis, aberrant SHH signaling during postnatal life has been implicated in pathological conditions. For instance, excessive activation of the SHH pathway has been found to contribute to medulloblastoma, an aggressive brain cancer that is derived from the cerebellum, demonstrating its role not just in development but in disease as well [19],[20].

1.3.2 Indian Hedgehog (IHH)

Indian Hedgehog (IHH) is an essential component of the Hedgehog (Hh) signaling group, mainly known for its central role in skeletal formation and endochondral ossification.

Endochondral classsification is a multi-step process in which cartilage is increasingly substituted by bone tissue, allowing for the formation of long bones within the appendicular skeleton. Aside from its role in the skeleton, IHH also affects hematopoiesis and connects its function during bone development to more systemic processes. The next sections explore IHH's most important functions and their developmental relevance.

- **Chondrocyte Regulation:**

Regulation of chondrocytes is one of the main functions of IHH. The cells that produce cartilage, chondrocytes, are regulated by IHH in a feedback loop with parathyroid hormone-related protein (PTHrP) during endochondral ossification to ensure a balance between chondrocyte proliferation and differentiation.

1. **Proliferation**: IHH stimulates the proliferation of chondrocytes in the growth plate to provide a sufficient number of cartilage cells for bone elongation.
2. **Differentiation:** While promoting proliferation, IHH also postpones premature differentiation of chondrocytes to hypertrophic cells, which would otherwise result in premature ossification and disrupted bone growth.

The signaling loop functions as follows: IHH triggers the expression of PTHrP in periarticular chondrocytes, which, in turn, suppresses chondrocyte hypertrophy within the growth plate. This mechanism helps to ensure chondrocytes proliferate and mature coordinately so bones can develop the proper length and shape. Perturbations within this loop may lead to defects in growth, such as skeletal deformities and shortening of bone elongation [21-23].

- **Bone Growth:**

IHH signaling is critical for specifying bone density and size. By coordinating the differentiation and growth of chondrocytes, IHH has a direct impact on the mechanical properties of forming bones. Moreover, IHH also affects the recruitment of osteoblast precursors, thus playing a role in the ossification process.

Skeletal disorders linked with faulty IHH signaling are:

1. **Brachydactyly:** A disorder in which fingers or toes are shortened, resulting from reduced chondrocyte proliferation.

2. **Chondrodysplasia:** A more extreme condition caused by mutations or interruptions in IHH that impair endochondral ossification, resulting in short, deformed bones and dwarfism.
3. **Acrocapitofemoral dysplasia:** A rare skeletal disease associated with IHH mutations, which results in growth plate abnormalities and extreme skeletal deformities.

All these diseases emphasize the importance of precisely controlled IHH signaling for proper skeletal growth and development [24],[25].

- **Hematopoiesis:**

In addition to its function in skeletal development, IHH is also involved in hematopoiesis, that is, blood cell production within the bone marrow. IHH signaling controls the bone marrow microenvironment to promote appropriate maintenance of hematopoietic stem cells (HSCs) and their niche.

1. **HSCs maintenance:** IHH maintains the reservoir of HSCs by ensuring a conducive stromal microenvironment. Regulation in this way provides a steady supply of multipotent stem cells for hematopoiesis.
2. **Bone-Blood Coordination:** IHH coordinates skeletal and hematopoietic activities, highlighting its ability to coordinate systemic development. The interaction of bone and blood cell signaling reveals IHH's function in tissue homeostasis.

It has been demonstrated in studies that alterations in IHH signaling can compromise the coordination between bone formation and hematopoiesis and result in deficiencies in blood cell production and weakened skeletal integrity [28].

1.3.3 Desert Hedgehog (DHH)

Desert Hedgehog (DHH) is a lesser-studied but strongly specialized member of the Hedgehog (Hh) signaling family. Sonically Hedgehog (SHH) and Indian Hedgehog (IHH) play largely universal roles in development, whereas DHH acts mainly in gonadal development and peripheral nerve development. Its activities are essential for male sex differentiation and maintenance of the reproductive system, as well as for structural and protective functions in peripheral nerves. The subsequent sections delve into the complex functions of DHH and its relevance to development and disease.

- **Testicular Development:**

The most fundamental and best-documented function of DHH is in the differentiation and development of the male gonads, i.e., the testes. In this scenario, DHH is crucial for the normal functioning and development of two key cell types: Sertoli cells and Leydig cells.

1. **Sertoli Cell Differentiation:** Sertoli cells give nutritional and structural support to developing sperm cells in the seminiferous tubules of the testes. DHH is critical for the differentiation and survival of Sertoli cells, maintaining the development of a supportive microenvironment for spermatogenesis. Sertoli cells are also involved in the formation of the blood-testis barrier, which separates developing germ cells from toxic substances in the blood
2. **Leydig Cell Function:** DHH controls the differentiation of Leydig cells, the cells that are responsible for the production of testosterone. This hormone is essential in the development of male secondary sexual features, such as muscle build-up, deep voice, and facial hair, and also for maintaining spermatogenesis. Leydig cells play a role in embryonic growth by secreting testosterone required for the masculinization of external genitalia.

It has been established by studies that DHH signaling through Sertoli and Leydig cells implicates interactions with Patched and Smoothened receptors that lead to downstream activations to mediate cell growth and differentiation. A disruption in the signaling process is likely to produce a continuum of reproductive impairments, including dysfunction in spermatogenesis, hormonal imbalance, and infertility.[16]

- **Peripheral Nerve Development**:

Besides its function in gonadal development, DHH is crucial for peripheral nerve formation and maintenance. Peripheral nerves link the central nervous system to the limbs and organs and conduct sensory and motor signals necessary for normal body function.

1. **Perineurium Formation:** The perineurium is a specialized connective tissue sheath that surrounds nerve fascicles, providing structural support and acting as a protective barrier against mechanical damage and environmental toxins. DHH signaling is essential for the formation and maintenance of the perineurium.
2. **Nerve Protection and Stability:** DHH contributes to the stability and integrity of peripheral nerves by regulating the development of Schwann cells and fibroblasts,

which interact to create a functional nerve sheath. Schwann cells, in particular, play a role in nerve regeneration and insulation through myelin production.

Disruptions in DHH signaling during development can result in defects in peripheral nerve structure and function, leading to conditions such as peripheral neuropathy, which is characterized by nerve damage and loss of sensation or motor control in affected areas [32].

- **Gonadal Dysgenesis:**

DHH plays a central role in sex differentiation and fertility, making it essential for the proper development and function of the reproductive system. Mutations or deficiencies in DHH signaling can lead to gonadal dysgenesis, a group of conditions that impair the formation and function of gonads.

1. **Pseudohermaphroditism:** In males, mutations in DHH can result in pseudohermaphroditism, a condition where individuals have ambiguous or incomplete development of male genitalia. This is often due to insufficient testosterone production caused by defective Leydig cell differentiation.
2. **Infertility**: Impaired Sertoli cell function due to DHH deficiency can disrupt spermatogenesis, leading to infertility. This is especially significant in males, where the production of viable sperm is directly tied to Sertoli cell health.
3. **46, XY Gonadal Dysgenesis:** A specific form of gonadal dysgenesis, 46, XY gonadal dysgenesis, is associated with mutations in the DHH gene. This condition is characterized by underdeveloped or absent testes, resulting in infertility and sometimes feminization due to a lack of testosterone.
4. **Female Reproductive Development:** While DHH's roles in female gonadal development are less well understood, evidence suggests that it may contribute to the maintenance of ovarian follicles and the regulation of folliculogenesis. However, this area remains a focus of ongoing research [5]

- **Clinical Implications:**

Understanding DHH's functions has significant clinical implications for treating conditions related to reproductive health and peripheral nerve disorders. Gene therapies targeting DHH signaling pathways could provide potential treatments for infertility and neuropathies.

Additionally, screening for mutations in the DHH gene can aid in the early diagnosis of developmental disorders, enabling timely intervention and management.

1.4 Structure and Functional Role of Hedgehog Interacting Protein (HHIP)

Hedgehog Interacting Protein (HHIP) is a cell membrane-associated type I glycoprotein that critically regulates the Hedgehog (Hh) pathway, which plays a key function in numerous development processes. On a structural basis, HHIP consists of two main domains, namely the N-terminal domain (HHIP-N) and C-terminal domain (HHIP-C). The C-terminal domain harbors a β-propeller framework and two epidermal growth factor (EGF) repeats, endowing it with unique binding properties to enable interactions with Hedgehog ligands, including Sonic Hedgehog (Shh) and Desert Hedgehog (Dhh) (Figure 2.1) [39], [40]. These interactions are regulatory in function by modulating Hh pathway activity and affecting the signaling gradients necessary for patterning of tissue. Mapping to chromosome 4q31.21, HHIP is well-conserved across species, which suggests its essential evolutionary function. This gene region has been strongly linked with Chronic Obstructive Pulmonary Disease (COPD) in several genome-wide association studies (GWAS), with the implication that HHIP variations might lead to susceptibility to COPD and other respiratory diseases [42]. Structurally, HHIP is comprised of 13 exons and contains about 700 amino acids, which make up a protein complex that plays a role in essential developmental processes, including the formation of anteroposterior limb patterning and control of left-right asymmetry during embryonic development [43]. HHIP's ligand binding interface, elucidating how HHIP controls the activity of the pathway by binding and sequestering Hh ligands, thus limiting their availability for receptor activation. This interaction modulates morphogenesis. The role of HHIP as a modulator and spatial inhibitor of Hh signaling highlights its importance in both normal development and disease etiology. [43].

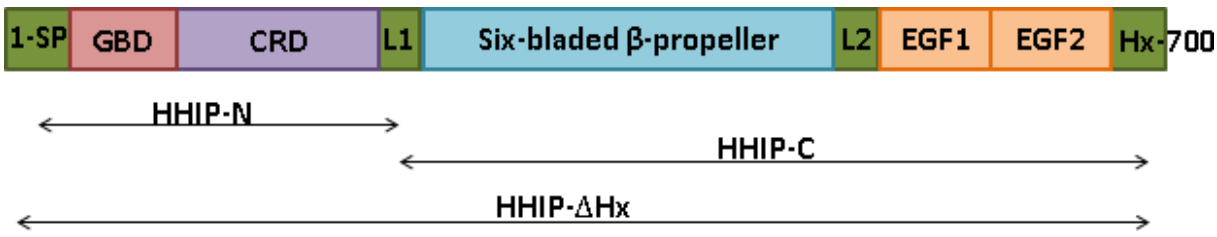

Figure 1.2: Structural Arrangement of Hedgehog Interacting Protein HHIP [20]

1.5 The Function of HHIP as a Negative Regulator of Hedgehog Signaling

Hedgehog Interacting Protein (HHIP) performs the crucial function of negative regulator in the

Hedgehog (Hh) signaling pathway, a pathway critical in multiple biological processes, such as development, tissue maintenance, and stem cell function. As a conserved, vertebrate-specific protein, HHIP specifically suppresses the function of all three mammalian Hedgehog ligands: Sonic Hedgehog (SHH), Indian Hedgehog (IHH), and Desert Hedgehog (DHH) [44]. Through direct interaction with these ligands, HHIP prevents them from binding to the Hh pathway's primary receptor, Patched (Ptch), thereby inhibiting the activation cascade that would otherwise ensue following this binding.In normal circumstances, upon binding of Hh ligands to Ptch, it lifts inhibition on the transmembrane protein Smoothened (Smo), initiating a chain of downstream events that influence gene expression. When HHIP is present, though, it binds to Hh ligands, essentially sequestering them and lowering their levels available to activate Ptch. This binding restricts the activation of Smo and suppresses the downstream signaling events, which are in charge of controlling cellular processes during embryogenesis, tissue homeostasis, and even cell proliferation [45].HHIP's control over Hh ligand availability is vital for ensuring the right intensity and timing of Hh signaling. This modulation is such that the activity of the pathway stays within safe bounds, precluding overstimulation that may otherwise result in pathological cell growth and differentiation. Studies have shown that dysregulation of HHIP, be it through mutation, changed expression, or impairment in its ligand interactions, may result in very adverse consequences, such as developmental abnormalities and diseases [49]

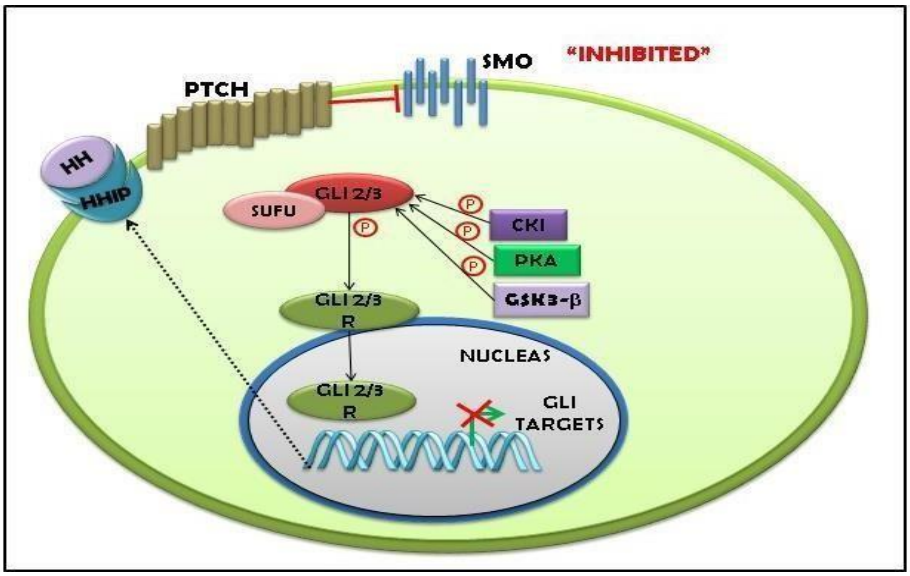

Figure 1.3: Hedgehog Interacting Protein (HHIP) in the Hedgehog (Hh) Pathway.

For example, aberrant HHIP function has been linked to several types of cancers and congenital abnormalities as a result of overactivity of Hh signaling [48]. As important as it is, HHIP has been brought into focus as a therapeutic target. Through elucidation and, if possible, augmentation of the inhibitory function of HHIP, therapies against diseases with Hh pathway deregulation as its basis, as summarized in Table 1.1, can be developed. By its capacity to modulate the Hh pathway, HHIP not only acts as a crucial molecular checkpoint but also as a therapeutic target of choice for strategies aimed at modulating Hh signaling in a controlled fashion [49].

The role of HHIP in the Hh Pathway. HHIP acts as a modulator of the Hh ligand, thereby inhibiting Patched (PTCH) from halting Smoothened (SMO) activity. At the same time, Suppressor of Fused (SUFU) traps full-length Glioma-associated oncogenes 2/3 (Gli2/3) inside the cytosol. These proteins undergo partial phosphorylation and proteolysis through the action of a complex protein kinase trio, consisting of Protein Kinase A (PKA), Glycogen Synthase Kinase 3β (GSK3β), and Casein Kinase 1α (CK1α). This complex process results in the accumulation of the repressor form, Gli2/3 repressor (Gli2/3 R), which enters the nucleus, where it binds and suppresses the expression of Hedgehog target genes, such as HHIP.

Table 1.1: Diseases Associated with Dysregulation of HHIP Protein

Protein Status	**Disease**	**Reason**	**Reference**
Downregulation	Colorectal Cancer	DNA Hypermethylation	[50]
	Gastric Cancer	DNA Hypermethylation	[51]
	Liver Cancer	DNA Hypermethylation	[52]
	Pancreatic Cancer	DNA Hypermethylation	[53]
	Lung Cancer	DNA Hypermethylation	[54]
	COPD	DNA Hypermethylati on	[55]
Upregulation	Osteogenic differentiation of BM-MSCs	High expression of lnc RNA HHIP-AS1	[56]

1.6 Downregulation of HHIP and its Epigenetic Regulation in the Hedgehog Pathway

Elucidating Hedgehog (Hh) Signalling Pathway Modulation via Pharmacological Discovery Strategies

Downregulation of Hedgehog Interacting Protein (HHIP) has come to be recognized as a major focus in biomedical research, especially in light of its central function in the Hedgehog (Hh) pathway and far-reaching implications for disease development. HHIP acts as an essential negative controller of the Hh pathway, acting by sequestering Hedgehog ligands (Sonic Hedgehog, Indian Hedgehog, and Desert Hedgehog) and inhibiting their interaction with the main receptor, Patched (PTCH). The mode of regulation stabilizes the pathway under normal control, which is necessary for regular cellular functions, such as tissue patterning, cell differentiation, and proliferation. Low levels of HHIP expression have been associated with aberrant activation of the Hh pathway, leading to numerous pathological diseases, especially in cancers and developmental diseases. Table 1 compares a number of studies that elucidate the role of HHIP downregulation in disease pathology [53][54][55].

Dysregulation of HHIP unleashes uncontrolled cell growth, abnormal development, and tissue patterning defects as normal cellular regulating mechanisms are thwarted. In cancers like lung, pancreatic, and prostate carcinomas, as well as in basal cell carcinoma, downregulation of HHIP has been found to contribute to tumorigenesis. This is largely because of the loss of HHIP's inhibitory action on the Hh pathway, which causes overactivation of downstream signaling components, such as Gli transcription factors. The aberrant Hh signaling cascade causes enhanced transcription of oncogenes, facilitating cellular transformation, survival, and metastatic potential. Likewise, in developmental disorders, decreased HHIP expression can interfere with tissue-specific regulation of the Hh pathway, causing congenital abnormalities. Current data indicate that a key mechanism responsible for HHIP downregulation is epigenetic modification by DNA methylation. DNA methylation is defined as the deposition of methyl groups on cytosine residues located in CpG dinucleotides within the promoter regions of genes. Such modification can suppress transcription of a gene by either physically blocking binding of transcription factors or by the recruitment of methyl-binding proteins forming repressive chromatin structures. Recent research has shown that hypermethylation of the promoter region of the HHIP gene is a frequent epigenetic change linked to its repression [57].

Excessive activity of DNA methyltransferases (DNMTs), which carry out the addition of methyl groups, are normally implicated in the process, resulting in decreased HHIP transcription and lower protein expression. This repression of HHIP leads to loss of its regulatory effect on the Hh pathway, increasing the pool of available Hh ligands and enhancing pathway activation.The effects of disrupted HHIP expression and resulting disbalance in Hh

signaling have dramatic consequences, sustaining disease process and cellular aberrations. Hyperactivation of the Hh pathway, in the lack of adequate HHIP regulation, has been linked with increased cell proliferation, apoptosis resistance, and higher invasiveness, mainly in cancerous tissue. Furthermore, the participation of epigenetic silencing pathways in HHIP downregulation reflects the intricacies of Hh pathway regulation and indicates potential therapeutic intervention.Epigenetic regulation of HHIP is essential to understand for the development of targeted therapies. In contrast to genetic mutations, which are irreversible changes in the DNA sequence, epigenetic modifications are reversible, providing a promising avenue for therapeutic intervention. Inhibitors of DNA methylation, including DNMT inhibitors, have been shown to reactivate silenced genes, including HHIP. Through inhibition of DNMT action, the drugs have the potential to lower hypermethylation of the HHIP gene promoter, which can reestablish its transcriptional activity and protein expression. Reversal of HHIP function by epigenetic treatments may recover its suppressive influence on the Hh pathway and thus suppress excessive activation of the pathway and its effects downstream. These therapeutic approaches have great potential for the treatment of diseases with aberrant Hh signaling, such as certain cancers and developmental abnormalities [58] .

In addition, drugs targeting DNA methylation to restore HHIP expression may also increase the effect of combination therapies. For example, co-administration of DNMT inhibitors and Hh pathway inhibitors could give a synergistic benefit by treating multiple levels of pathway dysregulation at the same time. This strategy may enhance therapeutic responses, especially in tumors with strong Hh pathway activity. Moreover, the reversibility of epigenetic changes also offers opportunities for individualized therapy, since treatment protocols can be adapted according to the methylation pattern of HHIP and other epigenetically controlled genes in patients.

The downregulation of HHIP is a pivotal aspect of the dysregulation of the Hh signaling pathway and its linked disorders. The epigenetic repression of HHIP by DNA methylation underscores the significance of investigating targeted epigenetic treatments, including DNMT inhibitors, to reverse normal pathway control. By reactivating HHIP expression and counteracting the consequences of dysregulated Hh signaling, these treatments hold the promise of enhancing the prognosis for patients with cancers and developmental disorders, providing new hope for controlling these multifactorial diseases [59].

1.7 Significance of DNA Methylation in Gene Expression

DNA methylation is a basic and well-controlled epigenetic modification that has a deep impact on gene expression and cellular function, which makes biological systems complex. DNA methylation is the process of adding a methyl group to the carbon-5 position of cytosine residues in CpG dinucleotides, forming 5-methylcytosine. DNA methylation is predominantly found in gene promoter regions and thus plays a vital role in controlling gene expression. The addition of a methyl group to the promoter region can inhibit the attachment of transcription factors and other proteins necessary for DNA, thus blocking gene

transcription. Moreover, DNA methylation has the ability to bind proteins like methyl-CpG-binding domain proteins, whose complexes are capable of inhibiting transcriptional activity further. This complex mechanism allows for fine-tuning control of gene expression in response to developmental signals, environmental conditions, and cellular cues.

The significance of DNA methylation extends beyond gene regulation because it has a critical role in various biological processes such as cellular differentiation, development, and genomic stability. During the process of cellular differentiation, certain patterns of DNA methylation are set up so that specific genes are silenced or expressed in a tissue- or cell type-restricted fashion. For instance, as cell division leads to the formation of various cell types, genes which are not needed for a particular cell type are silenced by DNA methylation, which helps establish and maintain cell identity. This allows cells from various lineages to carry out their unique roles and help in the broader orchestration of organ and tissue development. Through the regulation of gene expression at the epigenetic level, DNA methylation ensures the appropriate function of various tissues and organs and plays a role in the upkeep of cellular integrity within an organism's lifespan. Aside from its function in development, DNA methylation is important in safeguarding the genome from instability. Transposable elements or "jumping genes" are DNA sequences with the ability to move around within the genome. These elements, if not contained, can become inserted into active genes or destroy regulatory regions and cause mutations and genomic instability. DNA methylation suppresses these transposable elements, inhibiting their mobilization and thus ensuring genome stability. The capacity to suppress TEs is crucial for the protection of the integrity of the genome since the activation of these elements can result in chromosomal rearrangements, disruption of genes, and other mutations that promote disease development, including cancer. Thus, DNA methylation is an essential protection against the

risk of damage inflicted by these mobile genetic elements, maintaining genomic stability in the long term.

Yet, the precarious balance of DNA methylation can be altered, and in the process, many pathological conditions can arise. Dysregulation of DNA methylation patterns, like hypermethylation of tumor suppressor gene promoters or global hypomethylation, has been associated with a variety of diseases, including cancer. Hypermethylation of the promoter sites of tumor suppressor genes may silence key genes that act to prevent excessive cell growth and tumor development. Silencing these genes may remove the brakes on cell division, leading to cancer initiation and progression. Alternatively, hypomethylation, or generalized loss of DNA methylation, may cause the expression of oncogenes or reactivate transposable elements with a consequent stimulation of cancer progression. The interference with regular methylation patterns provides the uncontrolled activation of oncogenes, leading to uncontrollable cell growth, invasion, and metastasis.Apart from its associations with cancer, aberrant DNA methylation patterns have also been linked to other disorders such as neurological disorders, cardiovascular disorders, and autoimmune disorders. For instance, in neurodegenerative conditions such as Alzheimer's, DNA methylation changes can influence the expression of genes related to memory and cognition and thus may contribute to disease pathology. Likewise, in autoimmune diseases, aberrant DNA methylation can result in the activation of self-reactive immune cells and tissue damage and disease.

The reversibility of DNA methylation has generated intense interest in the therapeutic application of epigenetics, especially for diseases with aberrant patterns of DNA methylation. In contrast to genetic mutations, which are irreversible alterations in the DNA sequence, epigenetic changes such as DNA methylation are dynamically adjustable, offering the promise of a therapeutic target. Inhibitors of DNA methyltransferases (DNMTs), the enzymes that add a methyl group to DNA, are being investigated as a way of reactivating silenced genes, including tumor suppressor genes. DNMT inhibitors have exhibited promise within preclinical models and clinical trials, especially in the treatment of cancer. By suppressing the activity of DNMTs, these medications can abrogate hypermethylation of tumor suppressor genes, permitting re-expression of these vital genes and potentially ceasing or reversing tumor progression. This therapeutic method is not only promising for cancer but also for other conditions involving dysregulated DNA methylation, including neurological disease and autoimmune disease.

In addition, DNMT inhibitors can potentially be used synergistically with other therapeutic strategies, including chemotherapy or immunotherapy, to make them more effective. This synergistic method has the potential to provide more specific and efficient treatments with less toxicity than existing therapy. For example, by reactivating tumor suppressor genes and fixing epigenetic defects, DNMT inhibitors may sensitize cancer cells to chemotherapy, making them more treatable. Additionally, in the context of immunotherapy, DNMT inhibitors could help restore the expression of immune-related genes, enhancing the body's immune response to cancer cells. The reversibility of DNA methylation provides a unique opportunity for developing innovative therapies that could significantly impact the treatment of diseases, particularly cancer. By inhibiting the enzymes involved in the addition of methyl groups to DNA, we may be able to re-establish normal gene expression, reverse disease progression, and enhance patient outcomes. With increasing knowledge about DNA methylation and its contribution to disease, the promise of epigenetic therapies expands, holding out hope for more specific and effective therapies in the future [59-63].

1.8 Key Enzymes involved in DNA Methylation

DNA methylation, a basic epigenetic process, is the addition of methyl groups to cytosine residues in DNA, usually at CpG dinucleotides. DNA methylation is an important mechanism in the control of genes, cell differentiation, and maintaining genomic integrity. DNA methylation occurs through a family of enzymes called the DNA methyltransferases (DNMTs). These enzymes are crucial for the creation and maintenance of methylation patterns that regulate gene expression and are involved in normal cellular function and development. The primary DNMTs in mammals are DNMT1, DNMT3A, and DNMT3B, each playing distinct and important roles in epigenetic control of the genome.

1.8.1 DNMT1

- **Function:**

DNMT1 is known as the "maintenance" methyltransferase due to its main function of replicating established methylation patterns during DNA replication. When DNA replicates, the double helix is separated to form two daughter strands. The pre-existing methylated strand pairs with each daughter strand to form hemimethylated DNA. These hemimethylated regions are recognized by DNMT1, and the newly synthesized strand is methylated, guaranteeing the heritable transmission of methylation marks to daughter cells. This maintains epigenetic

information during cell divisions and cell identity as well as normal cellular behavior. DNMT1's function is essential for cellular functions like X-chromosome inactivation, genomic imprinting, and repression of repetitive elements, which are vital for genome stability [64][65].

- **Cancer Role:**

DNMT1's aberrant function has been associated with several cancers, where it plays a role in the hypermethylation of tumor suppressor gene promoters. Such hypermethylation results in silencing of these essential genes, causing cell cycle regulation and apoptosis to be disrupted. Consequently, cancer cells escape growth inhibition and apoptosis, leading to uncontrolled proliferation. For example, the inactivation of genes like p16INK4a, BRCA1, and MLH1 by DNMT1-mediated hypermethylation has been seen in various cancers, such as lung, breast, and colorectal cancers. The function of DNMT1 in preserving these aberrant methylation patterns leads to the aggressiveness of neoplasms. Moreover, overexpression of DNMT1 is often seen in cancer tissues, indicating that it is a potential target for treatment. DNMT1 inhibitors, including azacitidine and decitabine, are now being used in the therapy of neoplastic hematological diseases to restore gene expression by reverting methylation abnormalities [66].

1.8.2 DNMT3A and DNMT3B

- **Function:**

DNMT3A and DNMT3B are known as "de novo" methyltransferases since they play a role in the establishment of new methylation patterns on unmethylated DNA. These enzymes are active at early embryonic stages and are crucial for the establishment of tissue-specific methylation patterns that guide cell differentiation.

  - **DNMT3A** is essential in early development and differentiation, establishing early methylation marks that are vital for the proper regulation of pluripotency genes and the activation of lineage-specific transcriptional programs.
  - **DNMT3B** however, is predominantly active in later stages of development and is required for the formation of certain tissues, including portions of the immune system and the nervous system.

The cooperative activity of DNMT3A and DNMT3B ensures that cells acquire the typical methylation patterns necessary for specialized function. For instance, during hematopoiesis, DNMT3A specifies the methylation signatures governing progenitor cell differentiation to

various blood cell types. Similarly, DNMT3B ensures the methylation of the satellite repeats and centromeres, which is crucial for maintaining chromosomal stability [67]-[69].

- **Influence on Development:**

Normal development is dependent on normal functioning of DNMT3A and DNMT3B. Through the establishment of tissue-specific methylation patterns, these enzymes regulate cellular differentiation and help in the generation of certain cell lineages. This methylation renders genes necessary for a particular cell type to be active and other genes are kept suppressed. For example, during neuronal differentiation, DNMT3A methylates the pluripotency genes and thereby suppresses them to permit the expression of neuron-specific genes [70].

- **Disease Associations:**

Abnormal expression or mutations in DNMT3A and DNMT3B are linked with many diseases, particularly cancers and developmental disorders:

- **DNMT3A Mutations:**

DNMT3A mutations are strongly associated with acute myeloid leukemia (AML) and other hematologic malignancies. The mutations have a tendency to disrupt the function of the enzyme to establish proper methylation patterns, leading to the dysregulation of cell proliferation and survival genes. cancers like lung and colorectal, where it is implicated in hypermethylation and silencing of tumor suppressor genes in DNMT3B resulted in ICF syndrome (Immunodeficiency, Centromeric instability, and Facial anomalies), an autosomal recessive disease of rare occurrence with immune deficiencies, chromosomal instability, and developmental abnormalities.These mutations result in defective centromere methylation, leading to chromosomal instability and the phenotypic manifestations of the syndrome [71][72].

The crucial roles of DNMT1, DNMT3A, and DNMT3B in gene expression regulation and genomic stability have made them attractive candidates for therapeutic intervention. Efforts are presently focused on the design of targeted inhibitors of these enzymes to reverse disease-linked aberrant methylation patterns, i.e., in cancer. Moreover, the unscrambling of DNMT structural and functional complexities still has to offer hints about epigenetic development and

disease control and to set open the path towards even more accurate and efficacious treatments in the coming years.

1.9 Selective Inhibition in Drug Discovery

Selective inhibition is a basic principle in drug discovery where the goal is to develop compounds that can specifically target specific enzymes, receptors, or signaling pathways with minimal off-target effects. This selective targeting is important in order to increase the efficacy and safety profiles of therapeutic drugs, as it can modulate a specific biological function without affecting other essential processes. By targeting the specific inhibition of critical molecules in disease processes, selective inhibitors have the potential to provide new treatment for several diseases, such as cancer, neurodegenerative diseases, and autoimmune disorders.In the present study, we are harnessing the concept of selective inhibition to study compounds that specifically inhibit DNA methyltransferases (DNMTs)—DNMT1, DNMT3A, and DNMT3B. They are primarily involved in controlling DNA methylation, an epigenetic modification that influences gene expression and cellular behavior. DNA methylation plays an essential role in the regulation of important processes like cellular differentiation, repression of repetitive elements, and genomic integrity. Nonetheless, aberrant DNA methylation patterns like the hypermethylation of tumor suppressor genes or hypomethylation of oncogenes, which involve the dysregulation of DNA methylation, have been strongly associated with numerous diseases, especially cancers. In cancer, the inactivation of tumor suppressor genes by DNA hypermethylation or oncogene activation by DNA hypomethylation abrogates the normal control of cell growth, apoptosis, and differentiation, resulting in tumor progression and uncontrolled cell proliferation.The selective inhibition of DNMTs has emerged as a potential therapeutic approach to reversing these abnormal DNA methylation patterns and restoring normal gene expression. By inhibiting DNMTs specifically, we hope to reverse the epigenetic changes that underlie tumorigenesis and other diseases. Inhibition of DNMT1, for example, might suppress the perpetuation of cancer cell aberrant methylation marks, restoring suppressed tumor suppressor gene expression. Targeting DNMT3A and DNMT3B might suppress the formation of de novo methylation marks underlying cancer cell survival and metastasis. Selective inhibitors of DNMTs may, thus, arrest or even reverse these pathological processes involved in these diseases.This selective inhibition strategy not only has promise in cancer treatment but also has more general applications in other diseases involving DNA methylation, for example, neurological disorders, autoimmune diseases, and aging diseases.

By creating compounds that act specifically to modulate the activity of DNMTs, we are able to provide better treatments with fewer side effects. In addition, the discovery of DNMT inhibitors might create new paths toward drug development, offering more targeted and personalized therapeutic interventions for patients with many epigenetic diseases. In summary, selective inhibition of DNMTs is a potent drug discovery strategy that can revolutionize the treatment of cancer and other diseases associated with DNA methylation dysregulation.

1.10 Research Gap

The particular DNMT (DNA methyltransferase) that is known to trigger DNA hypermethylation and indirectly influence HHIP (Hedgehog-interacting protein) between DNMT1, DNMT3A, and DNMT3B has not yet been determined.

1.11 Problem Statement

The down-regulation of HHIP due to DNA hypermethylation prevents it from playing its normal role in suppressing the Hh pathway. This interference leads to the onset of different types of cancers like gastric cancer, pancreatic cancer, lung cancer, colorectal cancer etc. Thus, identification of selective inhibitors that can assist in suppressing DNA methylation can restore the function of HHIP and reduce the onset of different cancers.

1.12 Research Questions

• What factors are accountable for the HHIP regulation?

• What are the causes of down-regulation of HHIP?

• Which among the three DNMTs is involved in the downregulation of HHIP, and how do we choose the right enzyme for this process?

1.13 Aim

The aim of this research is to find selective inhibitors for down-regulated HHIP by inhibiting enzymes responsible for DNA hypermethylation through the use of sophisticated pharmacological methods.

1.14 Objectives

1. To Determine the Mechanisms of DNA Hypermethylation: Explore and establish the mechanisms responsible for DNA hypermethylation of the HHIP gene that results in the under-expression of HHIP.

2. To Identify Effective Inhibitors: Utilize new pharmacological methodologies to discover and develop effective inhibitors that can reduce DNA hypermethylation, thereby restoring normal expression of HHIP.

3. To Assess Efficacy of New Inhibitors: Strictly test the efficacy of chosen inhibitors in vitro, with special attention to their potential to inhibit DNA methylation processes and promote HHIP expression.

4. To Evaluate Regulatory Effect: Examine the regulatory effect of found inhibitors on HHIP under-expression in the Hh pathway, unveiling the complexities of.

1.15 Proposed Solution

Down-regulation of HHIP due to hypermethylation of DNA prevents it from performing its usual function of inhibiting the Hh pathway. This inhibition is involved in the causation of most forms of cancer like gastric cancer, pancreatic cancer, lung cancer, colorectal cancer, etc. Thus, discovery of selective inhibitors that can inhibit DNA methylation to restore the function of HHIP and control the growth of different cancers

1.16 Scope of Study

The main objective of this research is to find and analyze modulators which are capable of modulating DNA methyltransferases (DNMTs) activity efficiently. DNMTs are key actors in DNA methylation, an epigenetic change that can result in the hypermethylation of certain gene promoters, such as that of Hedgehog Interacting Protein (HHIP). The resultant hypermethylation is a major cause of the underexpression of HHIP, a critical regulator of the Hedgehog (Hh) signaling pathway. Dysregulation of the Hh pathway is connected to other developmental disorders and cancers, so it is important to investigate the means by which a normal signaling function can be restored and the pathogenic consequences of abnormal DNA methylation reversed.DNA methylation, specifically the addition of a methyl group to cytosine residues in CpG dinucleotides, is an important determinant of gene expression. It is a reversible and dynamic process, which is controlled by DNMTs, i.e., DNMT1, DNMT3A, and DNMT3B. They are responsible for maintaining or creating patterns of methylation required for normal

cell function. Abnormal methylation patterns may result if DNMT function becomes defective, like silencing tumor suppressor genes or oncogene activation. In the case of the Hh pathway, hypermethylation of the HHIP promoter results in its underexpression, which interferes with the regulatory balance of the pathway and is implicated in a variety of diseases, such as cancers basal cell carcinoma, medulloblastoma, and non-small cell lung cancer.Selective inhibition of DNMTs by our study is intended to interfere with the hypermethylation processes that result in abnormal gene expression patterns, most notably the silencing of HHIP. HHIP has a central function in the Hh signaling pathway by binding to the Sonic Hedgehog (Shh) ligand and so modulating pathway activation. Re-expression of HHIP by inhibiting DNMTs could thus potentially restore normal Hh signaling activity, presenting a therapeutic avenue for diseases linked with Hh pathway dysregulation. The objective of this study is not only to explore the possibility of DNMT inhibitors but also to elucidate the larger implications of epigenetic changes in gene and cellular function. To that end, we will use a stepwise strategy to screen for and identify potential DNMT inhibitors that are effective at reversing hypermethylation. This will require high-throughput screening of small molecules and compounds previously known to influence DNMT activity and subsequent biochemical assessment of their activity in cellular models. We will determine whether these inhibitors can modulate the function of DNMT and test their effects on HHIP expression and activation of the Hh signal pathway. Using a combination of molecular biology tools, i.e., quantitative PCR, Western blotting, and reporter assays, we will determine the effect of DNMT inhibitors on levels of HHIP expression and how they can reinstate normal Hh signaling. Additionally, we want to examine the interaction of DNMT activity with HHIP expression and Hh signaling. We anticipate that DNMT inhibition will counteract hypermethylation of the HHIP promoter, restore its expression, and modulate the activity of the Hh pathway. How selective DNMT inhibition can reduce hypermethylation will offer useful insight into epigenetic regulation in cellular functions and disease pathology. This may help in the identification of new therapeutic approaches that intervene in the epigenetic environment to reverse the dysregulation of the Hh pathway.More than cancer treatment, this work has general applicability in epigenetic therapy. By interfering with DNA methylation, we can find new ways to treat a range of diseases associated with epigenetic alterations, including neurological disorders, autoimmune diseases, and developmental syndromes. Additionally, the findings could pave the way for more targeted and personalized treatments, where drugs can be designed to specifically address the underlying epigenetic causes of disease, minimizing off-target effects and enhancing therapeutic efficacy.This

research holds the potential to contribute significantly to our understanding of the interplay between epigenetic modifications and signaling pathways in disease. The discovery of DNMT inhibitors and the determination of their impact on the Hh pathway would open up possibilities for improved treatments of cancers and other disorders associated with Hh pathway dysregulation. Further, this research may guide more general therapeutic strategies in regenerative medicine and other fields in which epigenetic regulation is important for cellular function and disease pathogenesis. Eventually, this research also seeks to establish a basis for interventions that can mitigate the harmful effects of aberrant DNA methylation, reestablishing normal cellular function and providing new promise for disease patients associated with the Hh pathway.

1.17 Societal Impact

The significance of this research goes far beyond the laboratory, holding the promise to significantly influence society, especially in cancer treatment and prevention. By targeting the mechanisms underlying aberrant DNA methylation, our research seeks to reveal novel therapeutic approaches that may provide new hope for patients suffering from a wide range of malignancies with dysregulated gene expression. Since cancer is still among the top causes of morbidity and mortality globally, the demand for effective treatment becomes more imperative than ever. Despite increased advances in treatment modalities and cancer research, the intricacy of cancer biology and its penchant to evolve and evade traditional therapeutic modalities continues to pose significant challenges in the field of oncology.Our research may help usher in targeted therapy that targets the molecular basis of cancer with enhanced specificity. DNA methylation is an essential mechanism of gene expression regulation, and abnormality in such processes has been associated with the onset and development of various cancers. By selectively inhibiting DNA methyltransferases (DNMTs), we seek to correct aberrant methylation patterns and restore normal gene expression, specifically in genes like HHIP that control key signaling pathways such as the Hedgehog (Hh) pathway. Targeted therapies engineered to directly alter the epigenome may offer a whole new category of treatments that work more effectively and specifically with fewer adverse side effects, characteristic of traditional treatments such as chemotherapy and radiation. By addressing the underlying epigenetic mechanisms of cancer, our study has the ability to not just increase survival but also the general quality of life for patients.Along with it, the comprehension of the action of DNMT inhibitors in the biology of cancer may also guide the creation of better diagnostic strategies.

Presently, most of the cancers are diagnosed at the advanced stage of the disease when the treatment alternatives are fewer and the prognosis tends to be very poor. The capacity to identify particular patterns of methylation linked to cancer would allow for earlier and more precise diagnoses, allowing for prompt intervention and a possible decrease in the disease load on healthcare systems. This is especially important in light of the increasing global burden of cancer and the need for early detection strategies. Our study could be informative in the identification of biomarkers for cancer that can be utilized to design diagnostic tests that are able to identify methylation alterations at early disease stages. Such tests can then be integrated into routine screenings, enabling early detection and treatment, which are key considerations in the enhancement of treatment outcomes and the lowering of mortality. Apart from therapy and diagnosis, our study has the potential to improve public health by the use of preventative interventions. Epigenetic dysregulation, such as deregulated DNA methylation patterns, is increasingly established as a disease-causing process in a vast spectrum of disease states, from cancers and neurological disorders to cardiovascular disease. In understanding the processes by which epigenetic dysregulation causes disease, we can establish novel biomarkers and risk factors which can be used to guide lifestyle and health choice. Such information may empower patients and healthcare providers to take preventive measures that reduce the risk of developing diseases linked to aberrant methylation patterns. For instance, knowing how lifestyle habits, diet, and environmental exposures affect DNA methylation may result in health campaigns targeting the promotion of healthy lifestyles with epigenetic risks for disease. These preventative measures would have long-term implications for lowering the occurrence of disease and enhancing the health of the population worldwide. Furthermore, this research is consistent with the aims of precision medicine, a new field that has been expanding at a rapid rate with the aim of customizing medical care according to the unique attributes of each patient. Among the greatest challenges facing precision medicine is the ability to compensate for variability in genetic and epigenetic profiles that govern patient responses to treatment. With increasing understanding of how to most effectively use DNMT inhibitors in the clinic, we draw nearer to truly individualized treatment strategies for cancer. Instead of following a one-size-fits-all approach to treatment, precision medicine aims to tailor therapies according to individual genetic and epigenetic information so that the patient is treated with therapies that are most suited to his or her individualized biological portrait. Our research can potentially serve as a key component in taking this strategy forward by generating knowledge on how certain methylation patterns are able to affect cancer treatment responses and how DNMT

inhibitors may be applied to control these types of responses. By including epigenetics in the treatment choice decision-making process, clinicians would be able to deliver more customized and efficient treatment to cancer patients, enhancing treatment outcomes and minimizing unnecessary side effects. Finally, this study focuses on an immediate social issue by investigating new fronts for treatment and intervention of diseases where epigenetic dysregulation is involved. The outcome of our research may pave the way for the identification of new therapeutic approaches addressing the epigenetic origins of cancer and other illnesses. With a deeper understanding of DNA methylation and disease, we aim to be able to make contributions towards the creation of interventions that are not only battling cancer but can also restore healthy cellular function to halt disease growth and enhance long-term health care outcomes. Societal relevance to this research thus lies in how it can re-engineer the cancer landscape

care, providing hope for better treatments and enhancing the lives of many people suffering from this debilitating disease. Beyond cancer, the findings of this research may have wider implications for other diseases linked to epigenetic dysregulation, opening the door to new therapies in neurodegenerative diseases, autoimmune diseases, and genetic syndromes. By furthering our knowledge of epigenetics and the disease role of DNMTs, our work could lead to a future where groundbreaking therapies not only improve the lives of patients with cancer but also improve public health and well-being in general. The potential of epigenetic therapeutics is vast, with the ability to cure diseases in ways that were previously unimaginable, and our research is set to be a part of that revolutionary process. By identifying new biomarkers, creating targeted therapies, and encouraging preventative health practices, this research aims to make a positive impact in the worldwide battle against cancer and other diseases, eventually enhancing the quality of life for people and communities worldwide.

Chapter#2

Literature review

2.1 Insights into Hedgehog (Hh) Pathway in Development, Regeneration, and Cellular Maintenance

The Hedgehog (Hh) signaling pathway, an evolutionarily conserved and basic mechanism, is essential to embryonic development as it mediates cell differentiation and directs morphogenesis. Its complex functions in embryonic and adult settings have been investigated in numerous studies, with it being identified as crucial to cellular processes and tissue integrity. While mostly silenced in adult tissue, under specific conditions, like tissue injury, it may be reactivated to facilitate controlled repair mechanisms to ensure cellular homeostasis.

The Hedgehog (Hh) signaling pathway is essential in embryonic development, as emphasized in a February 3, 2022, study by Ngoc Minh Nagyun. This study emphasizes the multidisciplinary character of Hh signaling as a sophisticated extracellular signal affecting many aspects of embryogenesis. The pathway plays a critical role in key cellular processes, such as cell proliferation, differentiation, survival, and axon guidance, which all play roles in the development of tissues and organs. A major characteristic of Hh signaling is the regional heterogeneity of Hh protein levels in the embryo, which leads to varied developmental consequences. Elevated levels of Hh proteins are usually associated with the development of particular tissues, while lower levels control other cellular developmental processes to ensure the survival of specific cell populations. This gradient system illustrates the flexibility of the Hh pathway in coordinating various developmental processes in a single organism.

Nagyun's work also points to the subtle but important role of the Hh pathway in adult tissues. Whereas its function is strongest during the development of an embryo, a few tissues, including the skin, harbor persistent Hh activity following development. This residual process is involved in vital functions including cellular maintenance, repair, and homeostasis, showing the continued utility of the Hh pathway postembryogenesis. In skin, as another instance, Hh signaling is implicated in cycling of hair follicles and wound healing, demonstrating its capacity to control regenerative processes even in adulthood. Such observations indicate that while Hh signaling is usually linked with early development, it still affects tissue function and mechanisms of repair throughout life.

Secondly, the research delves into the issue of pathological implications of abnormal Hh signaling in disease progression. Dysregulation of the Hh pathway has been implicated in a

number of disorders, such as basal cell carcinoma and other cancers, highlighting its dual function in both normal development and disease. While normal Hh activity is required for normal growth and maintenance of tissues, uncontrolled activation in adult tissues can promote tumorigenesis and lead to cancer development. In general, Nagyun's research emphasizes the long-term relevance of Hh signaling throughout life, from embryogenesis through to adulthood. Through the identification of the regulatory controls and multifunctional roles of the Hh pathway, this work offers insightful information on developmental biology and the possibility of targeted therapeutic approaches in disease models. [35].

Kuo-Shyang Jeng and colleagues, in their January 23, 2020 report, give a thorough explanation of the Sonic Hedgehog (SHH) signaling pathway's role in adult physiology, especially in the upkeep of tissue integrity and regeneration. Although the SHH pathway is mostly active during embryonic life, the article points out that in adult tissues it stays mostly quiescent under normal circumstances. Nonetheless, under conditions of stress, injury, or tissue damage, the pathway becomes reactivated to trigger cellular processes required for repair and regeneration. This reactivation serves as a protective mechanism against permanent damage and aids in the maintenance of vital biological functions such as cellular differentiation, proliferation, and stem cell population maintenance. Through precise control of these processes, the SHH pathway serves a critical role in maintaining tissue homeostasis and structural equilibrium. One of the central observations made by the study is the role of the SHH pathway in the repair of many tissues, such as the skin, liver, and nervous system.

The study presents concrete examples illustrating how SHH activation promotes wound healing, liver regeneration, and neuronal recovery after injury. These results point to the wide applicability of SHH signaling within a variety of organ systems, affirming its key role in ensuring overall physiological integrity. Nonetheless, Jeng and co-authors also highlight the fine line of SHH signaling in adult tissues. Whereas proper activation causes desirable regenerative impacts, inappropriate or excessive activation is known to be involved in harmful effects, including fibrosis, abnormal scarring, or cancer formation such as basal cell carcinoma and medulloblastoma. This dual function emphasizes the need for tight regulatory processes to prevent SHH-mediated regeneration from inadvertently facilitating pathological states. Jeng et al.'s research makes an important contribution to adult stem cell biology and regenerative medicine by revealing the SHH pathway's function in tissue repair and homeostasis. Their

findings form a basis for investigating novel therapeutic approaches focused on augmenting regenerative potential or reversing diseases associated with SHH dysregulation.

By learning how to modulate this pathway efficiently, medical science is able to push forward targeted therapies for degenerative disorders, injury, and cancer. In the end, the research points toward the promise of SHH-based therapies to transform therapeutic strategies, providing new hope for enhanced patient care and medical outcomes. [36].

Walton et al.'s 2021 research gave a significant insight into the Hedgehog (Hh) signaling pathway as an important regulator of intestinal organogenesis and adult intestinal homeostasis. Their work points to the complex role of Hh signaling in coordinating epithelial-mesenchymal interactions, both in embryonic development and in adult life. The research establishes that Hh signaling is paracrine in nature, whereby Hh ligands secreted by the endoderm-derived epithelium stimulate signaling within adjacent mesenchymal cells. This system of intercellular communication is critical to coordinate different types of mesenchymal cells for correct intestinal function and maintenance. One of the major findings of Walton's work is the dynamic interaction between epithelial Hh ligands and mesenchymal cells.

The experiment shows that Hh ligands derived from the epithelium modulate the behavior of the mesenchyme to secrete essential signaling factors like Wnts and Bmps. These regulate proliferation, differentiation, and morphogenesis of epithelia, with the result being intestinal integrity as well as adaptability to change in the external environment. Bidirectional communication underlines the evolved regulatory design of intestinal homeostasis and suggests the significance of epithelial-mesenchymal crosstalk in development biology. In addition, evolutionary conservation of the Hh signaling pathway highlights its importance in vertebrate organogenesis and development. Walton et al. offer a comparative view, illustrating that the underlying processes of epithelial-mesenchymal interactions are conserved across species.

Their research also highlights the pivotal function of Hh signaling in adult tissues, where it supports stem cell niches, controls homeostasis, and enables repair. Dysregulation of this pathway has been associated with disease states including intestinal fibrosis, inflammation, and cancer. Thus, a comprehension of Hh signaling complexity creates new avenues to design targeted therapies to treat intestinal diseases and improve regenerative medicine approaches. [97].

Volker Fendrich and others investigated the function of Hedgehog (Hh) signaling in pancreatic regeneration and its possible connection to pancreatic neoplasia in a 2009 study. Although both endocrine and exocrine pancreatic tissues have regenerative potential, the existence of progenitor cells in the adult pancreas is controversial. Pancreatic regeneration is a multifaceted process that includes the activation of signaling pathways involved in cellular repair and tissue homeostasis after injury. One of these pathways, the Hedgehog signaling pathway, has been well characterized for its involvement in embryonic development and maintenance of adult tissues. Its role in pancreatic regeneration and potential role in tumorigenesis was not previously understood. Employing a cerulein-induced damage and repair model, the authors showed that mature exocrine cells with the Elastase1 promoter are involved in pancreatic epithelial regeneration through the production of metaplastic ductal intermediates. These intermediates are interim structures that remodel normal pancreatic architecture. Activation of Hh signaling, as revealed by up-regulation of core pathway components and the Ptch-lacZ reporter allele, constituted strong proof of its implication in tissue restoration. Disruption of Hh signaling, using both pharmacologic and genetic means, critically hindered pancreatic regeneration.

Pharmacological inhibition was accomplished with the use of cyclopamine, a recognized Smoothened antagonist that plays an essential role in transducing Hh signals. Genetic disruption also involved targeted removal of essential pathway components, further validating the requirement of Hh signaling for proper tissue repair. Inhibition resulted in long-term persistence of metaplastic epithelium positive for progenitor cell markers, indicating that Hh signaling is necessary for redifferentiation of such cells into normal mature exocrine tissue. The stability of metaplastic intermediates is a concern for their malignant transformation in the absence of correct differentiation cues. These observations point to the role of Hh signaling in maintaining cellular plasticity during pancreatic regeneration and to the potential implications of its dysregulation in the etiology of pancreatic disease. This work provided key insights into the function of Hh signaling in both neoplastic transformation and pancreatic tissue regeneration. The need for Hh pathway activation in resolving metaplasia and re-establishing normal pancreatic architecture provides a mechanistic insight into how progenitor activation due to injury can play a role in tumorigenesis.

In light of the established role of Hh signaling in exocrine pancreatic cancer, this study fills an important knowledge gap regarding how regenerative processes might drive pancreatic neoplasia. In addition, the research indicates that Hh signaling can be therapeutically modulated to increase pancreatic regeneration while limiting the potential for neoplastic transformation. Such implications will have to be explored in greater depth by future investigations and may find targets within the Hh pathway that can act as biomarkers or sites for therapeutic intervention against pancreatic disease [98].

Catriona Jamieson and colleagues carried out a detailed study on the molecular mechanisms controlling the Hedgehog (Hh) signaling pathway, released on September 1, 2020. Their study emphasizes the essential role of Hh signaling in controlling cell proliferation, differentiation, and apoptosis. The pathway is activated when any of the three Hh ligands—Sonic Hedgehog (SHH), Indian Hedgehog (IHH), or Desert Hedgehog (DHH)—binds to the Patched-1 (PTCH1) receptor on target cells. When the ligand is not bound, PTCH1 suppresses Smoothened (SMO), an important transmembrane protein required for downstream signaling. Suppression prevents the pathway from being active. Upon binding an Hh ligand to PTCH1, suppression of SMO is relieved, enabling it to accumulate within the primary cilium, a vital organelle for Hh signal transduction. Activation of this step leads to a signal

ing cascade ultimately affecting gene expression, which guides essential cellular activities. Following the activation of SMO, a downstream cascade occurs that activates glioma-associated oncogene homolog (GLI) transcription factors.

The GLI proteins are translocated to the nucleus, where they regulate the expression of genes governing cell cycle progression, suppression of apoptosis, and tissue differentiation. This action reflects the pathway's importance in development and tissue homeostasis. The research also discusses the mechanisms that maintain the Hh pathway repressed in the absence of ligand stimulation. In such conditions, GLI transcription factors are kept sequestered in the cytoplasm by the suppressor of fused (SUFU) protein. In addition, kinases like casein kinase (CK) and glycogen synthase kinase-3 (GSK-3) phosphorylate GLI proteins, inhibiting their activation and nuclear translocation. These control measures guarantee Hh signaling only occurs when needed, ensuring cellular homeostasis and avoiding misappropriated signal transduction. Jamieson's work emphasizes the necessity for tight regulation of Hh signaling to ensure cellular integrity. Dysregulation of this process has been associated with several forms of cancer, such

as basal cell carcinoma and medulloblastoma, emphasizing the significance of targeted drugs that manipulate the components of Hh signaling.

The findings of the study add to the understanding of the molecular complexity of Hh signal transduction and provide a foundation for the development of therapeutic interventions to counteract Hh-related pathologies. Furthermore, by demystifying the intricacies of this pathway, the study provides new avenues for the progression of regenerative medicine, with potential applications in tissue repair and disease therapy [76].

A research by Chamey Suchors and others, published on August 14, 2022, emphasizes the fundamental function of the Hedgehog (Hh) signaling pathway in controlling cellular activities, especially in development and tissue repair. This study emphasizes how the Hh pathway coordinates important functions that enable cells to grow, differentiate, and maintain correct tissue architecture upon developmental cues and physiological stress. In normal conditions, the Hh pathway functions by activating glioma-associated oncogene homolog (GLI) transcription factors, which in turn control the expression of target genes that participate in cell proliferation, differentiation, and survival. Through the modulation of these genes, the Hh pathway controls the expansion and specialization of cells in a regulated manner, thus making it a crucial pathway for organ development, wound healing, and tissue homeostasis .

At the heart of Suchors' research is the precision with which the pathway balances activation and inhibition. Upon binding of Hh ligands to the Patched-1 (PTCH1) receptor, the suppression of Smoothened (SMO) is removed, resulting in the activation of GLI transcription factors. In contrast, without ligand binding, PTCH1 inhibits SMO activity, which in turn prevents GLI protein activation and therefore halts unnecessary gene expression. This system of regulation prevents cellular proliferation only upon appropriate stimulus, preventing uncontrolled growth and aberrant tissue formation. Nonetheless, Suchors' work also points to the dual role of the Hh pathway, uncovering that although its correct operation is essential for regular cellular functions, its deregulation can lead to serious pathologies, among which are basal cell carcinoma and medulloblastoma. The research points out that overactivation as well as a lack of Hh signaling can be harmful. Overactivation of the pathway has been associated with tumorigenesis, in which increased GLI activity promotes uncontrolled cell proliferation and survival, thus leading to cancer development.

On the other hand, impaired Hh signaling may lead to developmental abnormalities and defective tissue repair, again demonstrating the importance of tight regulation. Suchors' findings indicate that the careful balance of Hh pathway activity is essential for tissue homeostasis and disease avoidance. This observation presents possible therapeutic opportunities, where the targeted modulation of Hh signaling has the potential to alleviate pathological states and augment regenerative processes. Finally, this research advances the general knowledge of developmental biology and presents promising targets for clinical application in cancer therapy and regenerative medicine [38].

2.2 Role of HHIP in Hedgehog (Hh) Pathway Regulation and Disease Relevance

The Hedgehog Interacting Protein (HHIP) gene product is a transmembrane glycoprotein that is involved in the regulation of the Hedgehog (Hh) signaling pathway, which is critical for the cellular differentiation, tissue homeostasis, and development. The structural and functional characteristics of HHIP have been reported in recent studies, with special emphasis on its role as an essential inhibitor of Hh signaling. By preventing excessive pathway activation, HHIP contributes to maintaining cellular balance and preventing diseases linked to abnormal cell proliferation. Understanding the molecular mechanisms of HHIP provides valuable insights into its therapeutic potential for conditions characterized by dysregulated Hh signaling, including various cancers and developmental disorders. An important study by Alejandro Ortega-Martínez and colleagues, published on September 24, 2020, gives a comprehensive overview of the genetic and structural characteristics of the HHIP gene.

The HHIP gene is found on chromosome 4q31.21 and is more than 91 kilobases long with 13 exons, coding for a 700-amino acid protein that is membrane-anchored. This strategic placement enables HHIP to engage in interactions with central Hh pathway elements, permitting strict regulation of signaling activity. The research underscores the fact that the extracellular domains of HHIP are pivotal in binding SHH ligands, sequestering them away from their receptors, and blunting the subsequent downstream signaling cascade. This interaction highlights HHIP's role as an active modulator instead of a passive structural element, serving to refine Hh pathway output and avoid pathological activation that might result in aberrant cell growth. Ortega-Martínez's work also investigates the evolutionary conservation of HHIP, which suggests that its structural and functional properties have been

conserved across species because of its critical role in developmental biology. Mutations or dysregulation of HHIP have been associated with developmental disorders and tumorigenesis, underscoring its significance in cellular homeostasis.

The findings of the study provide a basic understanding of HHIP's genetic and molecular structure, laying the groundwork for future studies focused on the development of therapeutic interventions against the Hh pathway. Through the enhancement of our understanding of the regulatory processes that control Hh signaling, this work improves our capacity to create treatments for HHIP disease-linked illnesses, further upholding its importance in both health and disease. [77].

Samuel C. Griffiths and coauthors, in their December 9, 2021, work, presented noteworthy contributions to knowledge of the Hedgehog Interacting Protein (HHIP) and its function in the Sonic Hedgehog (SHH) pathway. Their study centered on HHIP's interaction with the SHH ligand and demonstrated that HHIP inhibits SHH signaling by a high-affinity binding mechanism. More precisely, the N-terminal part of the HHIP protein (HHIP-N) directly binds to SHH, essentially sequestering SHH and prohibiting it from engaging in interaction with the main pathway receptor, Patched-1 (PTCH1). Such an interaction perturbs the process of regular activation of the Hedgehog (Hh) pathway, with SHH never getting to trigger the downstream pathway components inappropriately. In normal physiological states, PTCH1 acts as a receptor that, after binding SHH, changes its conformation to release its inhibition of Smoothened (SMO), a key transmembrane protein in the pathway.

The stimulation of SMO is required for the subsequent signaling cascade downstream, which ultimately results in the activation of GLI transcription factors that control gene expression. But by sequestering with SHH, HHIP suppresses its association with PTCH1, and through this suppression, inhibits SMO activation as well as subsequent GLI transcription factor activation. This mode of suppression ensures proper regulation of controlled cellular proliferation, the preservation of organized tissue morphology, and the proper process of development. HHIP emerges in the research as a principal inhibitory molecule in the Hh pathway, whose function in maintaining tissue structure as well as regulation of cellular activity is absolutely fundamental. Aside from its basic biological function, Griffiths and co-authors' research also illuminates the therapeutic opportunity of HHIP. Interference with the function of HHIP has

been associated with diverse diseases, more notably chronic obstructive pulmonary disease (COPD), involving abnormal cell proliferation and tissue malorganization.

Based on the work, HHIP targeting may have a promising therapy for diseases caused by dysregulated Hh signaling. Through an investigation of the molecular basis for HHIP-mediated inhibition, the study not only increases our knowledge of the regulation of Hh pathway but also presents HHIP as a possible therapeutic target for a wide range of pathological states. These discoveries pave new paths for clinical treatments for modulating Hh signaling, solidifying HHIP's importance in both developmental biology and disease therapy. [41]

In 2024, Feng Guo and colleagues performed a pioneering study that found Hedgehog-Interacting Protein (HHIP) to be an essential regulator in chronic obstructive pulmonary disease (COPD) development. They targeted a distal enhancer at the 4q31 locus, a genomic area invariably linked with COPD susceptibility and lung function based on genome-wide association studies (GWAS). The research offered key insights into the manner in which this enhancer regulates HHIP transcription in human bronchial epithelial cells under basal conditions and upon stimulation with transforming growth factor beta (TGFβ). Importantly, the study showed that this regulatory process occurs through SMAD3 signaling in a Hedgehog-independent fashion, providing a new insight into COPD pathogenesis. In order to identify the function of this enhancer, the research used a battery of cutting-edge molecular methods, which involved single-nucleus ATAC sequencing (snATAC-seq), chromatin immunoprecipitation quantitative PCR (ChIP-qPCR), reporter assays, chromatin conformation capture, and Hi-C analyses.

Through these approaches, it was seen that the enhancer stabilizes chromatin topological domains surrounding the HHIP locus, making its transcriptional activity possible. In addition, CRISPR/Cas9-mediated deletion in BEAS-2B bronchial epithelial cells verified the enhancer's functional importance since its ablation caused significant repression of HHIP expression. This repression subsequently enhanced TGFβ-induced epithelial-mesenchymal transition (EMT), a pathological process promoting airway remodeling in COPD. These data underscore HHIP's protective function in maintaining epithelial homeostasis by reversing the pro-fibrotic and remodeling effects of TGFβ signaling. In addition to further elucidating COPD pathogenesis, this work highlights the therapeutic target potential of HHIP. By defining HHIP as an important regulator of airway epithelial cell function, the work implies that the restoration or

augmentation of its activity may be a new therapeutic strategy for reducing airway remodeling and other pathophysiologic aspects of COPD.

Moreover, identification of the role of the enhancer in regulation of HHIP transcription opens a possibility to design targeted therapies focused on modulation of enhancer function to maximize levels of HHIP. In summary, this work sheds light on molecular mechanisms involved in COPD and identifies HHIP as a candidate for therapeutic targeting, with the potential to introduce novel therapies for this incapacitating disease and other respiratory conditions characterized by dysfunction of the epithelium. [100].

Xin-Hua Yin and others continued to advance the knowledge of HHIP's role as a regulator of the Hedgehog (Hh) signaling pathway on November 14, 2022, highlighting its role as a vertebrate-specific inhibitor. The research unveiled that HHIP's inhibitory effect is not only limited to the Sonic Hedgehog (SHH) ligand but also applies to the other two primary Hh ligands—Indian Hedgehog (IHH) and Desert Hedgehog (DHH). This pan-spectrum inhibition underscores HHIP's evolutionary its conservation and versatility in controlling Hh signaling throughout vertebrates. HHIP operates by binding to SHH, IHH, and DHH directly, preventing these ligands from binding to the Patched-1 (PTCH1) receptor. This disruption of ligand-receptor interaction essentially prevents SMO activation, an important protein in Hh signal transmission. With the inhibition of SMO, subsequent activation of GLI transcription factors is also impeded, thereby reducing the downstream biological effects linked to Hh signaling. Among these effects are uncontrolled proliferation of cells, overgrowth of tissues, and defective differentiation, which are disease hallmark features shared by most illnesses, especially cancer and other diseases involving proliferation. The research points out the crucial function of GLI transcription factors in mediating the biological impacts of Hh signaling. GLI factors control major cellular processes like cell cycle, differentiation, and apoptosis. By suppressing GLI activation, HHIP is important in the regulation of cellular homeostasis and inhibition of abnormal tissue growth. The authors conclude that HHIP's general regulatory function makes it a likely candidate for a therapeutic target in diseases involving improper regulation of Hh signaling, most notably cancer and disorders involving unchecked cell proliferation. Apart from providing views into the molecular basis of Hh signaling, the study emphasizes the therapeutic value of HHIP as a target for disease intervention caused by uncontrolled activation of the Hh pathway. The capacity of HHIP to suppress multiple Hh ligands and its conservation

across vertebrates underscore its central function in controlling cellular processes and ensuring tissue integrity. This research further confirms HHIP as an important regulator of Hh signaling and the related biological processes [38].

Xin-Ping Zhao et al. conducted a seminal study in 2018 that revealed Hedgehog-Interacting Protein (Hhip) to be the key promoter of diabetic nephropathy (DN) onset, illuminating in substantial detail the molecular processes behind such chronic kidney disease. The authors' results depicted how Hhip is drastically induced in diabetic mouse kidney glomerular endothelial cells due to hyperglycemia-evoked reactive oxygen species (ROS) generation. This is mediated by NADPH oxidase 4 (Nox4), a critical enzyme involved in ROS production. This upregulation of ROS levels is followed by activation of Hhip expression, thereby promoting pathological changes pertinent to DN. Interestingly, the study highlighted the position of Hhip as a pivotal activator of the TGFβ1-Smad2/3 signaling pathway, a documented inducer of fibrosis and cellular transition events in numerous diseases. By this mechanism, Hhip facilitates endothelial-to-mesenchymal transition (EndMT), fibrosis, and apoptosis—critical contributors to glomerulosclerosis and podocyte loss, both of which are the hallmarks of DN. To further understand the role of Hhip in the progression of DN, the researchers employed diabetic heterozygous Hhip-deficient mice (Hhip+/−) to determine whether decreased Hhip expression would diminish disease severity.

Their findings demonstrated that Hhip-deficient mice exhibited significantly decreased levels of glomerulosclerosis, fibrosis, and podocyte loss when compared to their wild-type counterparts (Hhip+/+). These observations point towards the restriction of Hhip expression being protective against hyperglycemia-induced renal damage and dysfunction. The study highlights the critical role played by Hhip in mediating endothelial dysfunction, fibrosis, and podocyte injury and further solidifies its significance in the pathophysiology of DN. Through determining that Hhip deficiency is protective against DN progression, Zhao et al. provided strong evidence for the protein as a contributor to the disease and as a drug target. Beyond elucidating the role of Hhip in DN, the work opens the door to future study of targeted therapeutic interventions manipulating Hhip expression or its downstream signaling pathways. With its function in endothelial dysfunction and fibrosis, Hhip presents a promising target for treatments to slow or prevent DN progression.

The findings' significance are clinically significant, as diabetic nephropathy remains the leading cause of end-stage renal disease worldwide. By identifying Hhip as a central mediator in bridging hyperglycemia-induced oxidative stress and kidney fibrosis and failure, Zhao and team have established a solid platform for novel therapies that can be useful for DN patients. These findings open new opportunities for specific therapeutic interventions, rendering it even more important to investigate further in this direction. [99].

Mehdi H. Shahi and others studied the role of Hedgehog-Interacting Protein (HHIP) in glioma, which is a highly malignant and aggressive brain tumor, in 2015. With the intricate genetic and epigenetic profile of gliomas, identification of molecular mechanisms driving tumor growth is significant to enhance the prognosis and outcome of patients. The study targeted HHIP, a suppressor of the Hedgehog (Hh) signaling pathway, crucial for cell growth, differentiation, and survival. Aberrant regulation of the pathway has been associated with cancers such as gliomas. The research demonstrated a dramatic decrease or total loss of HHIP expression in the majority of glioma cell lines and primary tumor specimens, which led to an exploration of the possible function of epigenetic changes—namely, promoter methylation—in controlling HHIP expression. By methylation analysis, the scientists discovered that 87% of glioblastoma cell lines and 75% of primary tumor samples had promoter methylation at the HHIP locus, which is highly associated with loss of HHIP expression.

The finding indicated that epigenetic silencing of HHIP could be an important mechanism underlying glioma development by allowing unregulated Hedgehog signaling activity. The results of this discovery have profound implications, since the loss of HHIP eliminates a key inhibitory checkpoint, which could lead to unchecked cellular growth and malignant transformation. Additionally, the research pointed out that HHIP promoter methylation is not unique to gliomas but has been detected in other cancers, such as medulloblastoma, gastric, hepatic, and pancreatic cancers. This indicates broader applicability of HHIP silencing in cancer development, suggesting it to be a candidate all-purpose biomarker for the detection and prognosis of cancer. Clinical implications of this research by Shahi and his colleagues are considerable, especially where early cancer diagnosis and cancer treatments are concerned. The HHIP promoter methylation status may act as a potential diagnostic and prognostic marker for glioma, facilitating earlier identification of tumors and monitoring the progression of the disease.

Elucidating Hedgehog (Hh) Signalling Pathway Modulation via Pharmacological Discovery Strategies

In addition, the study provides new directions for therapeutic approaches targeting the reversal of HHIP silencing, including the application of DNA methylation inhibitors to reactivate HHIP expression and restore its tumor-suppressing activity. By identifying the role of epigenetic regulation in HHIP silencing, this work enhances our understanding of glioma biology and lays the groundwork for investigating new treatment modalities. Perturbation of molecular mechanisms that cause HHIP silencing may yield innovative therapies that enhance patient outcomes in glioma and other HHIP-silenced malignancies. [101]

2.3 Role of DNA Methylation and DNA Methyltransferases (DNMTs) in Gene Expression, Cognitive Function, and Cancer Development

DNA methylation is a ubiquitous epigenetic process that modulates gene expression, cellular differentiation, and development throughout mammalian genomes. Methylation of cytosine residues within DNA occurs during this process, which can impact numerous biological processes by influencing gene function. DNA methylation is important in sustaining cognitive capacity and has particular implications in the pathogenesis of cancer, notably by its activity against the Hedgehog signaling pathway. Recent research by Lisa D. Moore, Shoji Tajima, and ST Martin has revealed significant information on the function of DNA methylation and DNA methyltransferases (DNMTs) in health and disease, including how epigenetic alterations contribute to normal biological processes as well as to disease states. In 2013, Lisa D. Moore and researchers performed a pioneering study that proved the crucial function of DNA methylation in controlling gene expression in the mammalian genome.

They reported that DNA methylation provides a cue for repressive proteins to adhere to DNA that essentially inhibits transcription factors from reaching particular genes and hence silencing or diminishing their expression. One of the main conclusions of the study was that patterns of DNA methylation are dynamic, not fixed, in a way that cells can create tissue-specific and developmental stage-specific methylation profiles. Such versatility means genes being expressed or suppressed when required, maintaining cellular identity and function throughout the life of an organism. These results highlight the essential role of DNA methylation in developmental biology, as it allows cells to modify gene expression patterns according to their individual function and environmental contexts. In addition to its application in gene expression, Moore's work also established the importance of DNA methylation in sustaining cognitive function. Disturbances in DNA methylation, either caused by genetic mutations,

environmental factors like drug exposure, or developmental neural injuries, can result in impairments of cognitive functions.

These epigenetic changes could cause neurological disorders, highlighting the relationship between DNA methylation and brain health. This research not only confirmed DNA methylation's function in controlling gene expression but also indicated that epigenetic disturbances might have profound effects on neurological function and disease development. Through the creation of a connection between DNA methylation and mental health, Moore and coauthors rendered significant insights into how epigenetic pathways constitute brain function and general biological integrity. [78].

In November 2016, Shoji Tajima and others published a landmark study that gave a thorough insight into the enzymes involved in setting up and maintaining DNA methylation patterns in mammals. DNA methylation is an important epigenetic mark that controls gene expression, cellular differentiation, and developmental processes. The research identified three major DNA methyltransferases (DNMTs)—DNMT1, DNMT3A, and DNMT3B—each with unique but interrelated functions in DNA methylation, and the accessory protein DNMT3-like (DNMT3L), which acts as a cofactor. DNMT3A and DNMT3B were implicated as being mainly responsible for de novo DNA methylation, creating new methylation marks during certain stages of development, including embryonic implantation and germ cell differentiation. These enzymes establish the ground-level epigenetic landscape required for cellular identity and function. The study also identified DNMT3L as a non-catalytic cofactor that facilitates the activity of DNMT3A and DNMT3B, especially in germ cells, by stabilizing their enzymatic activities. After DNA methylation patterns are set, they need to be preserved through cell divisions in order to guarantee proper gene regulation.

This function is performed by DNMT1, the maintenance methyltransferase. DNMT1 is also involved in a significant function during DNA replication, that of identifying hemimethylated DNA and depositing methyl groups onto the newly synthesized complementary strand. In this way, epigenetic patterns are reliably passed from generation to generation of cells, allowing for the retention of cellular identity and function. The research placed particular emphasis on the complex cooperation between de novo methylation carried out by DNMT3A and DNMT3B and maintenance methylation performed by DNMT1 and stressed their collaborative function in managing gene expression. Disturbances in this regulation can cause improper methylation

patterns, which are linked to various diseases, including developmental disorders, cancer, and neurological disorders. For instance, mutations in DNMT3A and DNMT3B are associated with hematologic malignancies and immunodeficiency disorders, and aberrant activity of DNMT1 has been linked to cancer and neurodegenerative diseases. Tajima et al.'s work supplied a mechanistic model in vivid detail for how DNMTs cooperatively control the epigenome. By defining their functions in both the initiation and maintenance of DNA methylation patterns, the research set the stage for subsequent studies into the therapeutic value of DNMT modulation.

The results highlighted the potential to modulate DNMT activity to reverse abnormal methylation patterns underlying diseases like cancer, presenting new avenues of epigenetic therapy. The 2016 research greatly enhanced knowledge of DNA methylation by defining the unique and interactive roles of DNMT1, DNMT3A, DNMT3B, and DNMT3L. Through an investigation of the general mechanisms of DNA methylation, the study laid a solid ground for the creation of new methods for disease prevention and treatment, further underpinning the essential function of DNMTs in epigenetic control and human wellness. [79].

A 2020 review by ST Martin and colleagues offers a comprehensive overview of the complex interplay between DNA methyltransferases (DNMTs), DNA methylation, and cancerogenesis, with specific reference to the Hedgehog (Hh) pathway. The review examines how the aberrant activity of DNMT1, DNMT3A, and DNMT3B leads to abnormal DNA methylation patterns that cause the silencing of important regulatory genes like Hedgehog Interacting Protein (HHIP). In normal cellular conditions, HHIP is a critical inhibitor of the Hh pathway to ensure that it is not overactivated. The research, however, discovers that disturbances in DNMT function lead to hypermethylation of the HHIP promoter, inhibiting HHIP expression and triggering uncontrolled activation of the Hh pathway. This deregulation provides a cellular environment for tumorigenesis since the Hh pathway is a key player in cell proliferation, differentiation, and tissue repair.

Martin's work emphasizes the role of DNA methylation in tumorigenesis, its potential to induce changes in gene expression by repression of tumor suppressor genes such as HHIP. The article highlights how exaggerated DNA methylation by aberrantly regulated DNMT activity not just targets HHIP expression but impinges on other cellular signaling cascades critical to preserving normal cellular function. It is this destabilization that paves the way for unbridled cell

proliferation, a characterizing feature of tumorigenesis. In addition, the study illuminates how these molecular changes propel the onset and development of different cancers, especially those in which the Hh pathway is a central figure in disease onset. By concentrating on the epigenetic changes involved, the study highlights the promise of targeting these pathways to create effective cancer treatments.

The most important outcome of the study is the possibility of targeting the Hh pathway and its epigenetic regulators as a new therapeutic approach. As HHIP silencing by DNA methylation is a key mechanism of Hh pathway activation in cancers, restoration of HHIP expression or blockade of DNMTs that silence it may hold promising therapeutic avenues. By targeting molecular-level interventions, these approaches seek to restore normal Hh pathway regulation, slowing or stopping cancer growth. Martin's research also implies that deciphering DNMT activity in tumorigenesis could open the door to novel therapeutic investigation. Inhibition of DNA methylation and the enzymes involved in its dysregulation may offer a strong strategy for treating cancers in which epigenetic alterations are at the core of tumorigenesis. This research not only enhances our knowledge of the function of DNA methylation in cancer but also points to the promise of epigenetic therapies in creating more effective, targeted therapies for malignancies caused by Hh pathway dysregulation. [40]

In a 2019 study by Jian-Guo Zhao and coworkers, the development and progression of non-small cell lung cancer (NSCLC) were found to involve the function of Hedgehog-interacting protein (HHIP), but this was done specifically to test whether HHIP is a potential tumor suppressor. NSCLC is the most common and virulent subtype of lung cancer, so discovering its molecular drivers is key to designing new therapy strategies. The expression of HHIP was examined in several datasets of the Gene Expression Omnibus (GEO) and was validated by The Cancer Genome Atlas (TCGA) NSCLC cohort. By comparing data from large-scale datasets like GSE18842, GSE19804, and GSE43458, the researchers found that there was a considerable downregulation of HHIP expression in NSCLC tissues as opposed to normal lung tissues. This reduction in HHIP expression was significantly correlated with crucial clinicopathological parameters, such as TNM staging, tumor size, and histological type, reaffirming its clinical significance in NSCLC.

The research also proposed that the downregulation of HHIP expression in NSCLC can be prompted by epigenetic mechanisms, especially promoter hypermethylation. Epigenetic

silencing through DNA methylation has been found to regulate tumor suppressor genes in various cancers, and here it seems to have a central role in the downregulation of HHIP. By quantitative real-time PCR (qRT-PCR), the authors established that HHIP expression was significantly lower in NSCLC cell lines than in normal lung epithelial cells (BEAS-2B). This was strong evidence that DNA methylation of the HHIP promoter is responsible for its silencing in NSCLC and facilitates tumorigenesis. To determine the functional role of HHIP in NSCLC, the research performed in vitro experiments investigating the impact of HHIP overexpression on tumor cell behavior. The results were dramatic: ectopic overexpression of HHIP dramatically suppressed NSCLC cell proliferation, migration, and invasion—critical processes that are responsible for tumor growth and metastasis. The data suggest that HHIP acts as an effective tumor suppressor, and its downregulation allows cancer development to proceed.

The findings of this study shed new light on the molecular pathways involved in NSCLC and identify the pivotal role of HHIP in the regulation of tumor development. Downregulation of HHIP, possibly through DNA methylation, may be conducive to oncogenic pathway activation, enhanced tumor growth, and metastasis. The identification of HHIP as a key regulator in NSCLC provides insight into the disease and opens up new avenues for targeted therapy. Treatment modalities that restore the methylation state of HHIP or recover its expression are promising therapeutic possibilities for NSCLC, hoping for better prognosis of patients. The present work in 2019 by Zhao et al. reaffirms the pivotal role played by HHIP in the causation of NSCLC and supports its tumor suppressor characteristics. Having to do with the causation of cancer, manipulating epigenetic silencing of HHIP can bring new paradigms of precision medicine in the end benefitting cancer study and therapeutic breakthroughs. [102]

Yu Song and colleagues in 2014 examined the methylation status of CpG islands in the promoter region of the Hedgehog-interacting protein (HHIP) gene in gastric cancer tissues, peritumoral tissues, and the AGS gastric carcinoma cell line. The aim was to find out if HHIP methylation is involved in gastric cancer tumorigenesis. The findings indicated a marked decrease of HHIP mRNA expression in gastric carcinoma tissues versus adjacent normal tissues (0.82 ± 0.38 vs. 1.60 ± 0.26, $P < 0.001$), indicating that the downregulation of HHIP might be involved in the pathogenesis of gastric cancer. Moreover, levels of HHIP promoter methylation were also higher in gastric cancer tissues ($62.9 \pm 6.14\%$) and AGS cells ($99.7 \pm 0.67\%$) compared to peritumoral tissues ($17.7 \pm 3.59\%$) ($P < 0.05$). Such a differential methylation

pattern implies that promoter hypermethylation might play a role in silencing HHIP in gastric cancer. To further explore the function of DNA methylation in the regulation of HHIP expression, the scientists treated AGS cells with the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine (5-aza-dC).

Treatment resulted in a marked increase in HHIP mRNA expression (0.21 ± 0.12 vs. 4.68 ± 0.22, $P < 0.01$) and a significant decrease in methylation levels ($90.2 \pm 0.67\%$ vs. $10.1 \pm 0.21\%$, $P < 0.01$). Bisulfite sequencing PCR (BSP) analysis verified a decrease in methylation sites in CpG islands after treatment, further demonstrating the involvement of DNA methylation in HHIP regulation. There was a high negative correlation between HHIP promoter methylation and mRNA expression ($r = -0.693$, $P < 0.01$), further supporting the fact that promoter hypermethylation represses HHIP transcription in gastric cancer directly. These results indicate that aberrant DNA methylation of the HHIP promoter is a key event in gastric cancer development through the silencing of a gene involved in tumor suppression. The research points to the therapeutic potential of targeting DNA methylation in gastric cancer, notably by the reversal of HHIP promoter hypermethylation in restoring its tumor-suppressive potential.

Due to the extensive correlation between HHIP silencing and gastric carcinoma, epigenetic treatment based on demethylating the HHIP promoter may be a good option for enhancing patient prognosis. [103].

2.4 Epigenetic Modulation of the Hedgehog Pathway and HHIP in Cancer Progression and Other Diseases

The Hedgehog (Hh) pathway is pivotal in cellular proliferation, differentiation, and tissue balance, yet dysregulation is also often implicated in cancer and other disorders. An important regulatory factor in the pathway, Hedgehog-Interacting Protein (HHIP), is a natural antagonist that avoids the pathway being activated beyond requirements. Recent findings show, though, that epigenetic alteration through DNA hypermethylation and histone adjustment most commonly suppresses HHIP transcription, giving way to unregulated Hh pathway activity and tumor formation. Numerous studies have looked into the epigenetic regulation of HHIP in various cancers, chronic respiratory conditions, and liver neoplasias, and proposed that this pathway would be a promising therapeutic target. A highly cited 2021 study by Pichao Li

examined the effect of hexavalent chromium (Cr(VI)), a known environmental carcinogen, on lung cell transformation and its implication on the Hh signaling pathway.

The research demonstrated that Cr(VI) exposure caused a substantial decrease in HHIP expression, leading to hyperactivation of the Hh pathway. As the Hh pathway regulates basic cellular processes like proliferation, differentiation, and survival, its dysregulation through suppression of HHIP was pivotal in Cr(VI)-induced lung carcinogenesis. Epigenetic mechanisms for HHIP downregulation were identified by the researchers and shown to result from Cr(VI) exposure, with the effects of DNA hypermethylation and histone changes at the HHIP promoter site silencing HHIP expression and its downstream Hh pathway targets. The increased activation of the latter resulted in abnormal cell proliferation and transformation—hallmarks of cancer progress. The findings of the study have important implications, pointing to the involvement of environmental carcinogens such as Cr(VI) in promoting epigenetic changes that lead to tumorigenesis. Through the silencing of HHIP, Cr(VI) exposure creates a microenvironment that promotes uncontrolled cell growth and lung cancer development.

The study implies that HHIP may be a useful biomarker for the detection and monitoring of lung cancer due to Cr(VI) exposure and also a potential therapeutic target. Restoration of HHIP expression or reversal of its epigenetic repression can offer novel means to counteract the oncogenic activities of Cr(VI). This work also highlights the general significance of environmental exposures on epigenetic control and cancer biology, opening up avenues for future investigation into examining similar processes in other environmentally induced cancers. [41]

In April 2021, Yan Li and team carried out a landmark study exploring the function of Hedgehog Interacting Protein (HHIP) in chronic obstructive pulmonary disease (COPD), a chronic respiratory disease with ongoing airflow limitation and structural remodeling of the airways. The study shed new light on HHIP's function in disease development, specifically its role in regulating airway smooth muscle enlargement and glucose metabolism in airway cells. The researchers established that decreased HHIP expression was significantly correlated with the major pathological characteristics of COPD, such as airway smooth muscle hypertrophy and metabolic dysregulation. The results highlighted HHIP's vital role in preserving the structural and functional integrity of airway tissues. With the aid of murine models and human

cell cultures, the research illustrated that levels of HHIP were dramatically lower in COPD-affected airway tissues.

This decrease was accompanied by enhanced proliferation of airway smooth muscle cells, resulting in muscle hypertrophy—a major contributor to airway narrowing and airflow resistance in COPD. The study also revealed metabolic alterations in airway cells, specifically increased glycolytic activity, that supplied energy for uncontrolled cell proliferation. These metabolic changes also contributed to the structural and functional alterations in COPD, suggesting that dysfunctional metabolic regulation is a pivotal feature of the disease. On the molecular level, the authors showed that HHIP is an essential negative regulator of the Hedgehog (Hh) signaling pathway. The suppression of HHIP caused hyperactivation of the Hh pathway, which induced uncontrollable cellular growth and dysregulated metabolic homeostasis in airway cells. This dysregulation created a vicious cycle that enhanced COPD progression by repeatedly stimulating structural and metabolic dysregulation. The results of this study emphasize the critical function of HHIP in airway homeostasis. Through negative regulation of the Hh pathway, HHIP maintains normal cellular growth and metabolism within the airways and blocks pathological changes characteristic of COPD.

Loss of this regulatory activity through reduced HHIP expression offers a possible mechanistic connection between Hh pathway dysregulation and COPD progression. Notably, the research suggested that HHIP might be a therapeutic target for COPD management. Therapies that restore HHIP expression or block the Hh pathway might correct both airway smooth muscle hypertrophy and metabolic derangements. Therapeutic approaches could include gene therapy or small molecules to upregulate HHIP expression and thus restore Hh signaling homeostasis and alleviate disease progression. This work illustrates the therapeutic promise of HHIP targeting, presenting new avenues for future research and novel treatment strategies to enhance COPD treatment. [42]

A 2020 study led by Yun Song and associates presented strong evidence on the function of Hedgehog Interacting Protein (HHIP) as a protective agent in gastric cancer. The study underlined the tumor-suppressive function of HHIP and its value as a therapeutic target. The study proved that HHIP overexpression suppressed the growth and invasion of gastric cancer cells and highlighted the significance of limiting cancer growth. Moreover, the researchers found that HHIP expression was significantly lower in gastric cancer tissues compared to

normal gastric tissues, suggesting that its downregulation might contribute to the initiation and progression of the disease. A strong negative correlation was established between HHIP expression and metastasis, reinforcing HHIP's role in preventing the spread of cancer cells to distant sites. More detailed research on the molecular action of HHIP's tumor inhibitory activity implicated its role on epigenetic regulation in gastric cancer cells.

In detail, overexpression of HHIP was correlated with a decrease in de novo DNA methylation, an epigenetic modification usually found in silencing genes during cancer. Directly correlated to the decrease in DNA methylation was reduced Hedgehog (Hh) pathway activity, which is a master regulator of cell proliferation and cell differentiation. Aberrantly activated, the Hh pathway has been involved in multiple cancers, such as gastric cancer, in promoting tumor growth and metastasis. Being a negative regulator of the pathway, reduced expression of HHIP may result in aberrant activation of the Hh pathway, which promotes cancer development. These results highlight the potential of HHIP not just as a gastric cancer biomarker but also as an interesting drug target. The therapeutic potential of HHIP restoration is evident, as restoring HHIP expression can potentially inhibit the aberrant Hh pathway activation, thus preventing tumor growth and metastasis. HHIP's capacity to modulate epigenetic modifications further increases its therapeutic index, as inhibiting epigenetic modifications can restore normal cellular function and prevent cancer progression.

Possible approaches to restore the expression of HHIP may include epigenetic treatments, including DNA demethylating drugs, or the creation of drugs that are intended to mimic the action of HHIP. Further, HHIP's metastasis-suppressing function makes it a good candidate for therapies targeting advanced gastric cancer stages where metastasis presents big challenges to treatment. This research points to the larger importance of elucidating the interaction between epigenetic control and signaling pathways in cancer biology as a basis for the creation of more accurate and efficient therapeutic strategies. This work by Yun Song et al. underscores the significance of HHIP in controlling gastric cancer development and its therapeutic potential. The results indicate that the restoration of HHIP expression may provide a multi-faceted cancer therapy by inhibiting tumor growth, metastasis, and epigenetic alterations simultaneously. In addition to advancing our knowledge of gastric cancer biology, this work provides the foundation for novel treatment modalities to improve patient outcomes.

Through inhibition of HHIP expression and the Hh pathway, future studies may identify new therapeutic interventions that have the promise to revolutionize the treatment of gastric cancer. [43]

A milestone study by S.T. Martin and others in 2005 gave crucial clues to the epigenetic regulation of Hedgehog Interacting Protein (HHIP) in pancreatic cancer, disclosing the key function of DNA methylation in regulating the Hedgehog (Hh) signaling pathway. The scientists observed that around 80% of pancreatic tumors showed hypermethylation of the HHIP promoter, which correlated strongly with downregulation of HHIP expression. Since HHIP is an important antagonist of the Hh pathway, its downregulation allows uncontrolled Hh signaling activity, which leads to cancer progression. Stimulation of this pathway has been associated with enhanced cellular proliferation, improved survival, and more aggressive tumor behavior, underscoring its relevance in pancreatic adenocarcinomas. In order to determine possible therapeutic approaches, the researchers studied the application of 5-aza-2'-deoxycytidine, a known DNA demethylating agent, to restore HHIP silencing. Their data showed that such treatment effectively restored HHIP expression in pancreatic cancer cells, and notably suppressed Gli reporter activity, a readout of Hh pathway activation. Gli inhibition upon reactivation of HHIP proved the dependency of the pathway on downregulation of HHIP for its aberrant activation and thus tumor-promoting signaling inhibition. The work emphasized the functional importance of HHIP as an epigenetic regulatory checkpoint of the Hh pathway and implicated its epigenetic silencing in pancreatic tumorigenesis.

The epigenetic hypermethylation of the HHIP promoter is a mechanism by which pancreatic cancer cells bypass normal control of growth and enable the onset of a signal cascade leading to tumor growth and apoptosis resistance. More generally, the results underscore the role of DNA methylation in cancer biology, in which epigenetic modification is involved in tumor initiation and maintenance. Significantly, the Martin et al. study not only established a mechanism of HHIP regulation but also highlighted the therapeutic potential of targeting epigenetic aberrations in pancreatic cancer. In contrast with genetic mutation, epigenetic modifications are reversible, and thus represent desirable targets for therapeutic intervention. By using demethylating agents to counteract HHIP promoter hypermethylation, it is theoretically possible to reinstate HHIP's tumor-suppressive activity and restore normal regulation of the Hh pathway. This technique could offer a new avenue for abrogating the

aggressiveness of pancreatic adenocarcinomas, which represents an urgent need for more effective therapy in this extremely deadly cancer. The research also provides avenues for the combination of epigenetic therapies with other targeted agents in order to optimize their efficacy. Reactivation of HHIP may sensitize tumors to other therapeutic agents targeting other elements of the Hh pathway, like Smoothened (SMO) inhibitors or Gli transcription factor antagonists.

Utilizing a dual-targeted strategy, researchers could gain more global suppression of Hh signaling and further optimize patient outcomes. The research by S.T. Martin and colleagues is an early work providing an insight into the epigenetic mechanisms of Hh pathway dysregulation in pancreatic cancer. The identification of HHIP promoter hypermethylation as an activator of Hh signaling presents a strong rationale for inhibition of this epigenetic modification as part of therapeutic intervention. The capability to restore HHIP expression by DNA demethylation demonstrates the promise of epigenetic therapies to modulate critical signaling in pancreatic cancer. These results present a promising rationale for the construction of new therapeutic modalities directed at enhancing the prognosis of pancreatic adenocarcinoma patients. [44]

Research by Natalia Garcia and others shed key light on the Hedgehog (Hh) signaling pathway in uterine leiomyosarcoma (LMS), a rare yet aggressive cancer. The study involved the downregulation of Hedgehog Interacting Protein (HHIP), an important negative modulator of the Hh pathway, and its significance for LMS development. The research showed that the silencing of HHIP in LMS is linked with DNA hypermethylation, a type of epigenetic modification that suppresses gene expression through the addition of methyl groups to cytosine residues in the promoter regions. This silencing of HHIP through hypermethylation results in abnormal overactivation of the Hh pathway, thus resulting in uncontrolled tumor growth and aggressive cancer behavior. Continued research into the effects of Hh pathway dysregulation in LMS revealed prominent players like Smoothened (SMO) and GLI transcription factors to be critical for the activation of downstream signaling.

Pharmacological blocking of these proteins greatly diminished LMS cell proliferation, migration, and invasion. These data highlight the pivotal role of Hh pathway hyperactivation in driving tumor growth and metastasis in LMS. Notably, they propose that modulation of the Hh pathway via SMO and GLI inhibitors may be an effective treatment to control this vicious

cancer. HHIP is crucial in ensuring regular cellular regulation in the form of an antagonist for the Hh pathway. Its downregulation by DNA hypermethylation eliminates an essential checkpoint in the pathway, promoting uncontrolled signaling activity that promotes tumor growth, tissue invasion, and metastasis. The clinical implications of these findings are compelling. The discovery suggests that its reexpression through reversal of its hypermethylation may be an effective strategy to reactivate its tumor suppressive activities.

Epigenetic treatments in the form of DNA demethylating agents may be used to this effect, thus restoring the suppression of the Hh pathway and inhibiting LMS development. Moreover, inhibition of the constituents of the Hh pathway like SMO and GLI transcription factors is another strategy to reverse the action of HHIP silencing. By attacking the primary causes of Hh pathway hyperactivation, it could potentially be possible to make considerable inroads into the reduction of LMS progression and, consequently, enhance patient outcome. The broader implications of this study relate to the significance of epigenetic dysregulation in cancer. In contrast to genetic mutations, epigenetic alterations are reversible, presenting unprecedented opportunities for treatment. Having the capability to alter DNA methylation patterns and restore gene expression to normal is of tremendous therapeutic potential for the treatment of cancer.

In the context of LMS, epigenetic silencing of HHIP not only offers the opportunity to repress the Hh pathway but also the possibility of new combination therapies to increase treatment efficacy. The work by Natalia Garcia and colleagues underscores the key role of HHIP in modulating the Hh signaling pathway and its therapeutic targetability in uterine leiomyosarcoma. These results lay the groundwork for future studies into targeted treatments that may dramatically enhance the prognosis for patients with this aggressive and rare cancer. [95]

Gang Shen and colleagues in 2020 published a landmark study on hepatoblastoma, the most prevalent pediatric liver cancer, that elucidated the epigenetic changes that underlie its pathogenesis. The study recognized a bimodal pattern of global DNA hypomethylation combined with the hypermethylation of certain tumor suppressor genes, such as Hedgehog Interacting Protein (HHIP). This epigenetic disruption is a cause of cellular pathway dysregulation that is critical to the development of hepatoblastoma, specifically the Hedgehog (Hh) signaling pathway. In normal cells, HHIP acts as an important antagonist to the Hh

pathway, restricting excessive activation. In hepatoblastoma, though, gene silencing of HHIP by promoter hypermethylation eliminates this regulatory mechanism and results in the uncontrolled activation of the Hh pathway, which supports rampant cell growth, increased survival, and tumor advancement. The research also points out the function of DNA methyltransferases (DNMTs), the enzymes involved in the addition of methyl groups to DNA, in the hypermethylation of HHIP.

The abnormal activity of DNMTs not only suppresses tumor suppressor genes such as HHIP but also sustains the dysregulation of the Hh signaling pathway, enabling tumorigenesis. This epigenetic process is a pivotal step in the development of hepatoblastoma, highlighting the therapeutic potential of targeting DNMTs. As epigenetic changes, including DNA methylation, are reversible, applying DNMT inhibitors to reestablish HHIP expression may restore its inhibitory action on the Hh signaling pathway. This therapy can provide a valuable approach to inhibit tumor development and enhance the prognosis of children with this aggressive liver cancer. Outside of hepatoblastoma, this work provides a wider view on the importance of targeting epigenetic deregulation in pediatric malignancies. In contrast to genetic mutations, which tend to be permanent, epigenetic changes are dynamic and can be therapeutically manipulated.

The potential to manipulate DNA methylation patterns and reactivate critical tumor suppressors, such as HHIP, offers the possibility of more effective, less invasive therapies. By targeting the reversal of aberrant DNA methylation with DNMT inhibition, investigators may create novel therapies that restore pathway homeostasis and halt tumor growth. The conclusions of Gang Shen and colleagues not only improve our knowledge of hepatoblastoma pathogenesis but also highlight the therapeutic power of epigenetic treatment in fighting childhood cancers. [96]

A recent study on head and neck squamous cell carcinoma (HNSCC) has highlighted the epigenetic regulation of key components in the Hedgehog (Hh) signaling pathway, with a specific focus on Hedgehog Interacting Protein (HHIP) and Patched-1 (PTCH1). These genes act as critical negative regulators of the Hh pathway, maintaining pathway balance under normal conditions. Yet, the research identified that promoter methylation suppresses HHIP and some isoforms of PTCH1, which leads to uncontrolled Hh pathway activation. Downregulation of these tumor suppressors was highly associated with HNSCC development and poor patient

prognosis. Also, low levels of HHIP antisense RNA (HHIP-AS1), a non-coding RNA responsible for the regulation of HHIP expression, were correlated with increased GLI1 levels, a major effector of the Hh pathway responsible for decreased overall survival in patients with HNSCC. The results underscore the pivotal role played by epigenetic changes in promoting tumor progression and the necessity of comprehending HNSCC at the molecular level. The promoter methylation of HHIP and PTCH1 successfully decreases their transcriptional activity, which results in the overactivation of the Hh signaling pathway.

This causes increased activity of downstream effectors, including the GLI family of transcription factors, which promote tumor cell proliferation, survival, and metastasis. The loss of HHIP and PTCH1 function in HNSCC establishes a permissive condition for tumor progression, making them strong candidates as tumor suppressors. One of the most promising features of the research was the application of 5-aza-2'-deoxycytidine, a demethylating agent for DNA, to reactivate the expression of these genes. The treatment using this agent was able to reverse methylation of the promoters and reactivate HHIP and PTCH1, which resulted in the inhibition of Hh signaling and a drastic decrease in tumor growth. These findings indicate that DNA methylation targeting is a potential therapeutic approach to the treatment of HNSCC, especially where hypermethylation is an important driving factor for cancer progression. The research also points to the larger significance of epigenetic control in cancer biology. In contrast to genetic mutations, which are irreversible, epigenetic changes like DNA methylation are reversible, and this allows for a special potential for therapeutic intervention.

Compounds such as 5-aza-2'-deoxycytidine, which target specifically these modifications, can potentially normalize gene expression patterns and restore appropriate cellular control mechanisms. In the case of HNSCC, such strategies might greatly enhance patient prognosis by restoring tumor suppressor gene function and regulatory balance within the Hh pathway. This study underscores the central position of DNA hypermethylation in HNSCC development and illustrates how it is possible to target such epigenetic changes to create novel therapies that block tumor growth and enhance survival. The results affirm the potential of epigenetic therapy as a promising new era in cancer treatment, especially for cancers caused by pathway dysregulation secondary to aberrant DNA methylation. [97]

2.5 Critical Analysis

Table 2.1: Summary of Literature Review

Sr	Author/ Year	Title	Methodology	Outcomes	Limitations
1.	Guo, F., Zhang, L. *et al* (2024)	Identification of a distal enhancer regulating hedgehog interacting protein gene in human lung epithelial cells	In vitro cellular experiments	A distal enhancer at the COPD locus 4q31 controls HHIP transcription by SMAD3, but not in a Hedgehog-dependent manner. Its depletion increases TGFβ-induced EMT in human bronchial epithelial cells.	Limitations comprise limited sample size, the necessity of additional analysis of HHIP-TGFβ interactions, and investigating the function of the enhancer in the progression of COPD.
2.	Debolina, P.; Samadder S. et al (2023)	Differential association of hedgehog pathway in development of cervical carcinoma and its chemo tolerance	In vitro and in silico cellular experiments	The upregulation of HHIP and down regulation of SHH in CACX cell lines at different concentration of cisplatin may be playing crucial role in chemo tolerance of the disease	However, additional direct evidence in vivo is required to further support this observation
3.	Chakrabo rty, B., Basu, M el at (2022)	Differential promoter usages of PTCH1 and down regulation of HHIP are associated with HNSCC progression.	In silico cellular experiments	HHIP downregulation stimulates the Hh pathway in HNSCC, with demethylation providing an epigeneti c herapeutic option	The study is limited to HNSCC, with unclear applicability to other cancers, and lacks long- term safety data for demethylation treatments.
4.	Nguyen, N., Cho, J. et al (2022)	Hedgehog Pathway Inhibitors as Targeted Cancer Therapy and Strategies to Overcome Drug Resistance	In silico cellular experiments	The study highlights resistance mechanisms to Hedgehog (Hh) pathway inhibitors and proposes strategies to overcome	The study lacks in-depth exploration of clinical validation for proposed strategies to overcome resistance to Hedgehog (Hh) pathway inhibitors.

				resistance, improving cancer therapy efficacy.	
5.	Suchors, C., Kim, J. et al (2022)	Canonical Hedgehog Pathway and Noncanonical GLI Transcription Factor Activation in Cancer	In vitro and In silico cellular experiments	The Hedgehog pathway affects cancer through both oncogenicand tumorsuppressive functions through canonical and noncanonical GLI activation	The study is lacking original experimental data or novel insights into the specific molecular mechanisms of Hedgehog signaling in cancer.
6.	Yin, X., Wang, X. et al (2022)	lncRNA HHIP-AS1/HHIP modulates osteogenic differentiation of BM-MSCs by regulating Hedgehog signaling pathway	In vitro cellular experiments	HHIP-AS1 overexpression stabilizes HHIP mRNA through ELAVL1 binding, increasing HHIP, suppressing Hh signaling, and blocking osteogenesis.	Lack of animal experiments to validate the cell- based findings on the role of lncRNA HHIP-AS1 in osteogenic differentiation.
7	Li, Y.; Zhang L. et al (2021)	Hedgehog interacting protein (HHIP) inhibits airway remodeling and Metabolic reprogramming in COPD derived airway smooth muscle cells	In vivo animal and human studies	It was found that reduced levels of HHIP in COPD contribute to enhanced muscular growth in the airways	However, additional direct evidence in vivo is required to further support this observation
8	Li, P.; Zhang, X et al [13] (2021)	HHIP downregulation promotes Cr(VI)-induced malignant transformation in human bronchial epithelial cells	In vitro cellular experiment	Cr(VI) exposure causes HHIP promoter hyper-methylation, which results in HHIP downregulation and abnormal Hh signaling activation	It is not clear how HHIP is downregulated in Cr(VI) transformed cells

9	Walton, K., Gumucio, D. (2021)	Hedgehog Signaling in Intestinal Development and Homeostasis	In vitro and in silico cellular experiment	Hh signaling controls intestinal organogenesis and homeostasis through epithelial-mesenchymal cross-talk.	The study is poor in making minute observations regarding epithelial-mesenchymal interactions and their differences in intestinal conditions.
10	Griffiths, S., Schwab, R. et al (2021)	Hedgehog-Interacting Protein is a multimodal antagonist of Hedgehog signalling	In vitro cellular experiments	The research reveals HHIP's SHH signaling inhibition through several binding interactions, such as GAGs and cholesterol	The research is limited by not having in vivo validation of the suggested mechanisms and possible intricacies in the in vivo environment.
11	Garcia, N.; Al-Hendy, (2020)	Targeting hedgehog pathway and DNA methyltransferases in uterine leiomyosarcoma cells	In vitro and in silico cellular experiment	HH inhibition by SMO and GLI inhibitors suppressed LMS growth, which was augmented by inhibition of DNMT.	Limited by the small sample size due to LMS rarity; further validation is needed.
12	Matinez, A., Rubio, G.et al (2020)	HHIP gene variants impact COPD susceptibility, lung capacity, and protein concentrations in female biomass smoke exposure	in vitro and in vivo cellular experiments	HHIP gene variants reduce the risk of COPD, raise HHIP levels, and enhance lung function among biomass-exposed Mexican women	Cross-sectional design imposes restrictions on causality, and the results are unlikely to apply beyond Mexican women exposed to biomass smoke
9.	Song, Y.; Tu, J. et al (2020)	HHIP overexpression suppresses gastric cancer development and metastasis through	In vitro cellular experiment	HHIP targeting and lentiviral overexpression have therapeutic potential as gastric cancer treatment strategies.	Further clinical research is imperative to establish HHIP as a viable biological target for gastric cancer treatment

		decreased CpG island methylation			
13.	Jeng, K.; Chang, C. et al (2020)	Sonic Hedgehog Signalling in Organogenesis, Tumor and Tumor Microenvironments	In vitro and in silico cellular experiments	Cancer Stem cells can be targeted by the combined use of Sonic Hedgehog Signalling inhibitors and chemotherapy	The research is devoid of comprehensive experimental confirmation and clinical data for combined SHH-targeting cancer stem cell treatments
14.	Jamieson, C., Martinelli, G., et al (2020)	Hedgehog Pathway Inhibitors: A New Therapeutic Class for the Treatment of Acute Myeloid Leukemia	In vitro and in silico cellular experiments	Hh pathway inhibitors are promising for AML therapy, particularly with chemotherapy, by acting on leukemic stem cells	The study lacks extensive data on long-term effects and the development of resistance to Hh pathway inhibitors in AML treatment.
15	Zhao, J. G.,Wang, J. F., et al (2019)	HHIP overexpressio n inhibits the proliferation, migration and invasion of non-small cell lung cancer	In vitro cellular experiments	HHIP overexpression suppresses the growth and invasion of NSCLC cells and thus demonstrates the potential of it as a marker for diagnosis	A limitation of the study is the lack of in vivo validation to confirm the effects of HHIP overexpressi on in NSCLC progression.
16.	Zhaou, X., Chang, S. et al (2018)	Hedgehog Interacting Protein Promotes Fibrosis and Apoptosis in Glomerular Endothelial Cells in Murine Diabetes	In vitro and in vivo cellular experiments	Renal HHIP induces diabetic nephropathy by triggering fibrosis and apoptosis via Nox4 activation and TGFβ1-Smad2/3 signaling	The study lacks exploration of therapeutic interventions targeting Hhip expression to mitigate diabetic nephropathy progression.
17.	Shahi, M. H., Zazpe, I., et al (2018)	Epigenetic regulation of human hedgehog interacting protein in glioma cell lines and primary tumor samples	In vitro cellular experiments	HHIP is epigenetically silenced by promoter methylation in glioblastoma, which implies that it may ha ve therapeutic or	The study lacks functional validation of HHIP's role in glioma progression and clinical utility as a biomarker.

				prognostic or diagnostic value as a marker	
18.	Tajima, S., Suitake, I. et al (2016)	Domain Structure of Dnmt1,Dnmt3A and Dnmt3B DNA Methyltransferases	In silico analysis	Dnmt1, Dnmt3a, and Dnmt3b control DNA methylation via different domains interacting with histone and DNA modifications.	The study lacks experimental validation of the described domain interactions and their functional mechanisms.
19.	Song, Y., & Zuo, Y. et al (2014)	Occurrence of HHIP gene CpG Island methylation in gastric cancer	In vivo and in vitro cellular experiments	Gastric cancer high HHIP promoter methylation is associated with lower HHIP expression, with the latter implicated in cancer drive.	The study lacks mechanistic exploration of how HHIP methylation specifically drives tumorigenesi s.
20.	Moore, L. D., Thuc, L et al (2013)	DNA Methylation and Its Basic Function	In vitro and in silico cellular experiments	DNA methylation dynamics control gene expression and neuronal activity, influencing cognition and neuropsychiatric disorders.	The study lacks direct exploration of therapeutic interventions and detailed mechanisms linking DNA methylation alterations to specific neuropsychia tric disorders.
21.	Fendrich, V., Esni, F., et al (2008)	Hedgehog Signaling Is Required for Effective Regeneration of Exocrine Pancreas	In vivo cellular experiments	Hh signaling supports exocrine cell-mediated pancreatic regeneration and can be associated with post-injury neoplasia	The study does not conclusively establish the identity and role of pancreatic progenitor cells during regeneration.
22.	Martin, S. T.; and Sato, N., et al (2005)	Aberrant methylation of the Human Hedgehog interacting	In vitro cellular and in silico experiments	HHIP repression through hypermethylation enhances the activity of Hh pathway in	Other mechanisms for HHIP silencing were not explored, nor was in vivo impact assessed.

		protein (HHIP) gene in pancreatic neoplasms		pancreatic cancer and demethylation reactivates its expression	

Chapter 03

Material and Methodology

3.1 Introduction to Methodology

This study aimed to identify and characterize potential drug candidates targeting DNA methyltransferases (DNMTs), specifically DNMT1, DNMT3A, and DNMT3B, which play crucial roles in epigenetic regulation and are implicated in various cancers. To achieve this, a multi-faceted computational approach was employed to streamline the discovery of compounds with high binding affinities and desirable drug-like properties. By integrating structural similarity analysis, drug-likeness filtering, molecular docking, and clustering techniques, the methodology ensured an efficient and targeted selection of promising inhibitors. This approach not only enhanced the efficiency of candidate selection but also provided deeper insights into the interactions between small molecules and DNMTs.

The process began with the utilization of the ChemBridge dataset, which contains over one million small molecules, as the primary source for candidate compounds. The Simplified Molecular Input Line Entry System (SMILES) representations of these compounds were converted into molecular objects using RDKit, an open-source cheminformatics library. Using RDKit, MACCS Key fingerprints were generated to capture critical structural features, allowing for structural similarity analysis. This analysis helped identify compounds with significant resemblance to a reference set of known DNMT inhibitors, effectively narrowing the dataset. Subsequently, drug-likeness filters based on Lipinski's Rule of Five were applied to evaluate essential molecular characteristics such as molecular weight, lipophilicity, hydrogen bond donors, and acceptors. Only compounds that met these criteria were retained to ensure favorable pharmacokinetic properties and high oral bioavailability.

To further refine the selection, molecular docking studies were conducted using the Molecular Operating Environment (MOE) software to assess the binding affinities of candidate compounds against DNMT1, DNMT3A, and DNMT3B. Docking simulations provided valuable insights into interaction profiles, identifying candidates with strong and specific binding potential. Scoring functions were used to evaluate the strength of these interactions. Additionally, clustering techniques were employed to analyze the structure-activity relationships (SARs) of the compounds, grouping them based on chemical descriptors to reveal correlations between molecular structures and binding efficacy. This systematic approach facilitated the prioritization of compounds for further experimental validation. By leveraging

these computational techniques, this study successfully streamlined the discovery process and contributed to the development of novel DNMT inhibitors, offering promising avenues.

3.2 Data Retrieval

Compounds were gathered through two primary sources: scientific literature and the ChemBridge compound library to build a comprehensive dataset for analysis. This dual approach ensured a dataset with both validated inhibitors and structurally diverse novel compounds, providing a well- rounded basis for further study and analysis.

3.2.1 Literature-Based Compound Selection

An extensive review of literature was done to seek compounds with experimentally confirmed half-maximal inhibitory concentration (IC50) values against DNA methyltransferase (DNMT) enzymes, specifically DNMT1, DNMT3A, and DNMT3B. The peer-reviewed literature was the focus in this review for data validity and reliability. In total, 13 reference compounds were found with each having associated experimentally calculated IC50 values and 2D chemical structures. These compounds were ordered based on their shown inhibitory potency in prior experimental research, and a good platform for computational study was created. Care was also taken in selecting structural characteristics linked with stronger DNMT inhibition, giving an idea of molecular traits required for target binding.

The chosen compounds were used as reference points for follow-up virtual screening and computational experiments, which allowed comparison and confirmation of new-found candidates.

3.2.2 ChemBridge Compound Database

Concurrently, a high-throughput compound screening campaign used the ChemBridge database, which is renowned for its large library of more than 1.3 million chemically diverse and high-quality small molecules. For this research, 1,654,770 chemical compounds were accessed from ChemBridge for computational screening. The database was selected because of its prestige in drug discovery and chemical biology research, offering access to compounds with favorable drug-like characteristics and commercial availability. ChemBridge compounds and Hit2Lead search tool was used for the selection of compounds based on their keywords and structural relevance with high precision. The ChemBridge compounds were downloaded

in Structure Data File (SDF) format that helps in retaining the 3D structures required for subsequent computational analysis like molecular docking and virtual screening. This format facilitates compatibility with drug design software and enables in-depth assessment of molecular properties. Coupling the findings from the literature with a huge library of structurally diverse compounds, this exhaustive strategy guaranteed inclusion of both highly characterized reference compounds and new candidates, making the compound selection process of the study more robust and reliable.

3.3 Data Preparation

There was a methodical and systematic process applied to treating and preparing chemical data from literature-derived reference compounds as well as ChemBridge compounds for further cheminformatics analyses. To allow for consistency and compatibility between datasets, chemical structures were normalized into SMILES (Simplified Molecular Input Line Entry System) format, a machine-readable format of chemical molecules that is widely supported. The SMILES strings for 13 reference molecules, ascertained through a comprehensive literature survey, were produced using ChemDraw, a powerful tool for chemical structure drawing and conversion. The SMILES strings were meticulously checked for reliability and stored in text file form for effortless incorporation into downstream computational pipelines. This step facilitated experimentally verified compounds being reported in a standardized manner, making their direct comparison possible with virtual screen data.Compounds derived from the ChemBridge database, provided originally in Structure Data File (SDF) format, utilized an open-source Python-based workflow, RDKit

cheminformatics package. RDKit is a robust package for chemical data handling, providing capabilities for file conversion, calculation of molecular properties, and structure manipulation. A Python script was devised to execute a manual conversion, with a function called sdf_to_smiles that used RDKit's Chem.SDMolSupplier class to import the SDF file. The script looped through every molecule in the dataset, translating valid structures into SMILES strings with RDKit's Chem.MolToSmiles function.

Translated SMILES strings were written to a text file called chembridge_compounds_smiles txt. This process greatly improved efficiency by avoiding the need for manual translation, allowing chemical structures to be represented consistently in a cheminformatics tool-

compatible format. Due to the large size of the ChemBridge dataset, with more than 1.65 million compounds, a partitioning approach was adopted to maximize computational efficiency and memory handling. A second Python script was created to split the SMILES dataset into manageable groups of around 100,000 compounds each. Trailing empty lines were discarded while splitting to ensure data integrity and clean segmentation. The derived subsets were written to separate text files, like chembridge_1.txt, chembridge_2.txt, and so on. This partitioning approach optimized memory consumption and enabled effective parallel processing in future analysis phases. Through dividing the dataset into small, manageable sets, the pipeline was rendered scalable, and integrating extra data in the future experiments became easy. Partitioned data files were used as inputs for cheminformatics analyses such as molecular docking, clustering, and structure-activity relationship studies.Implementing this comprehensive and robust data preparation process provided a sound foundation for high-quality cheminformatics analysis, with increased efficiency and reliability of drug discovery processes [83],[84].

3.4 MACCS Fingerprint Analysis and Similarity Search

3.4.1 Data Preparation and Fingerprint Generation

In MACCS Fingerprint Analysis and Similarity Search, a systematic and methodical approach was followed to compare the structural similarity between two sets of data: a reference file of 13 compounds (13_compounds.txt) and a ChemBridge subset of compounds (chembridge_17.txt). The analysis was based on the RDKit cheminformatics toolkit, which supported processing SMILES strings and converting them into RDKit molecule objects. These molecular objects were then utilized in the creation of MACCS (Molecular Access System) Key fingerprints, providing the structural groundwork for the comparative analysis.MACCS Key fingerprints are a powerful computational method available in cheminformatics that convert the molecular structures to fixed-length 166-bit binary strings.

These bits of binary string each report the existence or non-existence of a distinct predefined chemical function. Such properties encompass atomic labels, bond classes, and different substructures typically present in chemical compounds. The binary encoding strategy is very standardized and computationally inexpensive, and it enables quick representation, screening, and comparison of molecular structures within large collections. The compact binary

representation also makes it easier to identify systematic common structural properties and to rank compounds with analogous chemical properties. This renders MACCS fingerprints a treasure trove of cheminformatics data, especially when dealing with datasets of considerable size, as it provides both efficiency and reproducibility in the structural analysis [84].

3.4.2 Similarity Calculation Using Tanimoto Scores

To quantify the structural similarity between compounds from the reference file and the ChemBridge subset, Tanimoto similarity scores were calculated. The Tanimoto score is a widely recognized metric in cheminformatics for comparing molecular structures. It measures the degree of overlap between the structural features of two molecules, expressed as a value between 0 and 1.

A score of 1 signifies identical structures, while a score of 0 denotes no common structural features. This measure enables researchers to systematically compare how similar compounds are to one another, and thus it is an essential tool for virtual screening and drug discovery.Every combination of ChemBridge and reference compounds was compared using RDKit's Tanimoto similarity function in order to calculate pairwise similarity indices. These comparisons were based on the molecular representations, MACCS Key fingerprints encoded as such. Each calculation established how much overlap existed between two molecules from predefined chemical features that were encoded in the fingerprints. This process led to a robust dataset, whereby every entry comprised the ID of the reference compound and the ChemBridge compound together with their associated similarity score. For ease of further analysis, the results were structured and saved in a formal CSV file (similarity_17.csv). This file provided easy examination and identification of compounds with high structural similarity to the reference molecules. The methodology ensured that the data was not only computationally robust but also easy to interpret and utilize in subsequent studies. Additionally, the substantial size of the dataset required careful implementation to manage computational resources effectively. RDKit, a versatile and open- source cheminformatics toolkit, streamlined the entire process, ensuring accuracy and scalability. By facilitating the detection of structurally similar compounds, this stage played an important role in the general goal of selecting potential drug candidates for detailed assessment [85].

3.4.3 Filtering for High-Similarity Compounds

To find compounds with high structural similarity to the reference set, a controlled filtering process was executed on the Tanimoto similarity data. The CSV file (similarity_17.csv), which includes the calculated similarity scores, was imported into Python by utilizing the pandas library, an efficient data manipulation and analysis tool. A 0.8 similarity threshold was used to only keep compound pairs that show notable structural similarity with the reference molecules. This cut-off is a high level of overlap between the molecular characteristics of the compounds, essentially screening out those with low or no similarity to the reference set.

Filtering was carried out by going through the similarity data and choosing entries where the Tanimoto score was equal to or higher than the cut-off. Each qualifying entry, which included the IDs of the reference compound and the ChemBridge compound and their similarity score, was appended to a new DataFrame. This strategy enabled the production of a curated dataset with only high-similarity compounds, which drastically cut down the amount of data and concentrated the analysis on viable candidates.To preserve data integrity and enable future analysis, the resultant curated DataFrame was exported as individual CSV files, for example, Similar Chembridge Compound No 17.csv. Each file was a refined subset of the ChemBridge dataset, optimized for downstream cheminformatics use. This organized and automated filtering step streamlined the analysis, allowing for more focused examination of compounds with the potential to be further explored in drug discovery. By reducing the dataset to the top-similarity compounds, the strategy increased the effectiveness of the identification of promising candidates for experimental validation and downstream computational studies.

3.4.4 Recovery of Similar Compounds' SMILES Strings

Subsequent to the identification of high-similarity compounds by applying the Tanimoto similarity cutoff, an explicit Python script was written to optimize the retrieval of the SMILES strings of the chosen molecules. The step commenced by transferring the indices of high-similarity compounds from the filtered CSV file (e.g., Similar Chembridge Compound No 17.csv) into a pandas DataFrame. This DataFrame was used as a formatted container for the compound indices, which were then pulled out into a list for processing. The script used this list of indices to find and pull out the corresponding SMILES strings from the ChemBridge subset file (chembridge_17.txt). By reading the chembridge_17.txt file line by line, the script paired each index from the list with the corresponding SMILES entry in the text file. This method guaranteed correct mapping between the high-similarity compounds and their

corresponding molecular representations.Once they were identified, the SMILES strings were systematically gathered into a new list, which was subsequently saved to a standalone text file named similar_smiles.txt. This combined file offered a centralized repository of SMILES strings for compounds with high structural similarity to the reference molecules. By structuring the data in this way, the scriptstreamlined follow-up cheminformatics workflows so that downstream analyses like docking studies, property predictions, or additional filtering could access molecular representations efficiently.

The automation not only saved time but also decreased the chances of human error while handling the data. The design of the script was such that compatibility with large datasets was maintained, and entries were processed efficiently, even in the case of large lists of highly similar compounds. By aggregating SMILES strings into a single, readily available file, this approach made it easy to transition into later phases of computational analysis, making the drug discovery pipeline efficient

3.4.5 Conversion to SDF Format

In order to support downstream cheminformatics analysis and enable compatibility with varied analysis tools, a final conversion of the SMILES strings into the SDF (Structure Data File) format was carried out using another Python script driven by RDKit. The SDF format is broadly preferred in cheminformatics because of its capacity to save both molecular structure details as well as corresponding properties, and hence, it is perfect to combine with and hold chemical and property data in a single file.

Key by reading the similar_smiles.txt file line-wise, where each line denoted a SMILES string for a high-similarity compound already identified. The script used RDKit's Chem.MolFromSmiles function to try and parse each SMILES string into an RDKit molecule object. As each conversion succeeded, other attributes from the initial dataset, for example, compound IDs or similarity scores, were appended to the molecule object as metadata. These attributes included rich contextual information that would be used in future steps of analysis. To create the output, the script created an SDWriter instance and provided the output file as molecules.sdf. Each successfully parsed molecule object was given to the writer, which appended the molecular structure and related properties to the SDF file. Molecules with bad SMILES strings, which did not convert to valid RDKit objects, were automatically filtered out from the output. Such bad entries were written in a standalone text file (invalid_smiles.log) for

inspection purposes, allowing transparency and facilitating troubleshooting as necessary.The script finished up by closing out the SDWriter, completing the structured SDF file with the cleaned-up dataset of valid molecules. This process minimized the extent to which the resulting molecules.sdf file needed to be processed further before it could be integrated into other cheminformatics tools, including software for molecular docking or visualization. Automating the conversion to SMILES strings in the universal SDF format, this method augmented the interoperability of the dataset with reduced human effort and errors. The produced file is a solid base for sophisticated analyses like property prediction, structure-activity relationship studies, and virtual screening, and aligns perfectly with the larger aims of the research pipeline [84][85].

3.5 Using Drug-Like Filters

When we analyzed chemical compounds from the SDF file (molecules.sdf) through RDKit, we aimed at assessing key molecular features in line with drug-likeness requirements. The major goal was to select and rank compounds with therapeutic potential by checking their pharmacokinetic and pharmacodynamic properties, both of which play a crucial role in the drug development process. This assessment is crucial for comprehending how a compound interacts within biological systems and forecasting its efficacy and safety as a drug candidate.In order to do this, we applied a series of stringent filters based on Lipinski's Rule of Five. This rule is a general guideline for drug-likeness determination and encompasses parameters like molecular weight (≤500 Daltons), LogP (≤5, representing balanced hydrophilicity and lipophilicity), hydrogen bond donors (≤5), and hydrogen bond acceptors (≤10). Molecules satisfying these requirements are more likely to have good oral bioavailability, an important consideration for most drug candidates. Aside from Lipinski's Rule, we included bespoke filters aimed at improving absorption, distribution, metabolism, and excretion (ADME) properties. These included analyses of topological polar surface area (TPSA), which determines how well a compound can permeate cell membranes; rotatable bonds, which control molecular flexibility and binding interactions; and aromaticity, which could influence stability as well as receptor binding. Those compounds that did not fit within acceptable ranges for these descriptors were rejected to guarantee that only those with the best physicochemical profiles were selected. This systematic screening and filtering ensured that compounds with favorable drug-like characteristics were identified, enhancing the chances of their success in subsequent development phases. By marrying the strong computational capabilities of RDKit with

pharmacological insights, this approach offered a streamlined methodology to tailor a huge chemical space down to a subset of promising candidates worthy of therapeutic investigation [86],[87].

3.5.1 Data Preparation: Importing the SDF File

We started the analysis by importing the SDF (Structure-Data File) that contains the molecular structures in a file format widely supported by cheminformatics software. This file was parsed using RDKit's Chem.SDMolSupplier, an adaptable function for parsing and dealing with molecular structures that are in an SDF format. With the use of this tool, every molecule contained in the SDF file was transformed into an RDKit molecule object, a necessary step to allow for additional computational processing and descriptor calculation. This first step ensured that the structural information of the compounds was available in a cheminformatics workflow-compatible format. RDKit, a powerful Python-based cheminformatics library, was at the center of handling molecular structures' analysis. It offered functionality to calculate a range of molecular descriptors, which are numerical value representations of the physicochemical and structural characteristics of a compound. Terms like molecular weight, topological polar surface area (TPSA), LogP (lipophilicity), and hydrogen bond donor and acceptor counts were some of the important parameters computed. These are very important in drug discovery since they help in understanding the behavior of compounds in biological systems. For instance, TPSA is directly related to the capacity of a molecule to pass through cellular membranes, whereas LogP values reflect its preference for hydrophilic or lipophilic conditions, which affect absorption and distribution. Equally, the number of hydrogen bond acceptors and donors indicate the degree of potential molecular interactions, for instance, hydrogen bonding, inherent to protein-ligand binding.Generating such descriptor calculations on which the analysis draws facilitates enhanced appreciation of structural as well as physicochemical character of molecules. This kind of data plays a crucial role in the estimation of drug-likeness as well as prognosis if a compound exhibits required traits toward therapeutic utilization. With this systematic process, RDKit enabled an effective framework for converting raw molecular information into meaningful insights guiding the processes of drug development.

3.5.2 Choice of Drug-Like Descriptors

In order to make predictions for drug-likeness of the compounds, we have emphasized some important descriptors derived from Lipinski's Rule of Five and extra considerations for maximizing ADME properties. The descriptors chosen are the following:

- **Molecular Weight (< 400 Da):** Molecular weight is vital for bioavailability. Molecules with higher molecular weights usually find it hard to cross cell membranes, thus lowering their absorption. A molecular weight of less than 400 Da typically is indicative of potential for better absorption. This requirement makes sure that the compounds are small enough to cross biological membranes effectively, which is a good quality for drug candidates.
- **LogP (< 5):** LogP is the octanol-water partition coefficient, a measure of lipophilicity (fat-solubility) of the compound. Compounds with a LogP less than 5 are balanced between hydrophilicity and lipophilicity, which provides both solubility and membrane permeability. Molecules with high LogP (above 5) can experience difficulties in solubility and become toxic because they are poorly soluble in aqueous environments.
- **Hydrogen Bond Donors and Acceptors:** Hydrogen bond donors are the capability of a molecule to donate a hydrogen atom that is bonded to an electronegative atom (such as nitrogen or oxygen), whereas hydrogen bond acceptors are atoms with the capability of accepting hydrogen bonds. For drug-likeness, up to 5 hydrogen bond donors and up to 10 hydrogen bond acceptors are suggested. Overhydrogen bonding can prevent the molecule from being able to penetrate cell membranes, hence compromising its bioavailability. Decreasing these values enhances permeability and absorption.
- **Rotatable Bonds (< 5):** Rotatable bonds are single bonds between two atoms which have free rotation, which affects the flexibility of a molecule. Having fewer rotatable bonds tends to make the molecule more rigid, which can increase its target specificity and binding to a particular receptor. A limit of 5 rotatable bonds is usually prescribed to balance flexibility and stability.
- **Molar Refractivity:** Molar refractivity is a quantitation of the polarizability of a molecule and is responsible for its interaction with other molecules due to van der Waals forces. Molar refractivity is connected with the molecule volume as well as with the binding interactions with other compounds. An appropriate molar refractivity signifies the capability of the molecule to interact with the environment in an effective manner, thus affecting its binding to drug targets.

- **Topological Surface Area (Low TSA):** Topological surface area refers to the measurement of the polar surface area of a molecule. Low TSA is preferred for drug-likeness since it makes membrane permeability easier. Highly TSA molecules exhibit poor absorption with low bioavailability. For orally administered drugs, lowering TSA aids in enhancing cell membrane permeability, which is critical for gastrointestinal tract absorption.
- **Formal Charge (Near Neutral):** The formal charge of a molecule is the overall charge on the molecule, and it must be as close to neutral as it can be. Molecules bearing a formal charge are less capable of passing across the hydrophobic lipid bilayers of cell membranes. A neutral or slight positive or negative charge enhances membrane permeability and minimizes non-specific interactions.
- **Heavy Atom Number:** The number of non-hydrogen atoms in the molecule is termed the heavy atom count. Lower in number, as a rule, since this shows fewer atoms, and it is better metabolized and drug-like by the molecule.A compound's behavior in a biological system becomes more predictable if the number of heavy atoms is minimized.
- **Ring Count (Minimum One Ring):** Rings help to maintain rigidity and are essential to keep certain conformations intact in order to participate in drug-target interactions. There must be a minimum of one ring to have enough structural rigidity for good binding. Rings also help add stiffness to the molecule, which enhances its specificity to the target and binding.

3.5.3 Data Processing and Filtering

With RDKit, an extensive list of molecular descriptors was computed for every compound in the dataset to evaluate their chemical characteristics and drug-likeness. Molecular descriptors, which measure different aspects of the structure of a molecule, such as its size, hydrophobicity, and functional groups, are important to evaluate how fit compounds are to be developed further as drugs. These descriptors are basic markers of a molecule's pharmacokinetic and pharmacodynamic features, and they give useful information about its activity in biological systems.In order to reduce the dataset to compounds with drug-like characteristics, a drug-likeness filtering procedure was used. The filtering process was based on a number of drug discovery criteria, including molecular weight, LogP (octanol-water partition coefficient),

hydrogen bond donors and acceptors, rotatable bonds, and other important descriptors that affect a compound's ADME profile. For instance, molecules with unreasonably high or low molecular weights can be poorly bioavailable, and molecules with an inadmissible hydrogen bond donor/acceptor balance might struggle to permeate through biological membranes or bind to their targets.

The Python pandas library was used to handle and filter the data in an efficient manner, making it easy to manipulate the compound data and descriptors. Pandas offered the required functionality to import the dataset, to apply the filtering criteria, and to manage the selection process. A series of conditional statements were implemented to introduce thresholds into the molecular descriptors, e.g., to set a molecular weight of 200-500 Da, a LogP of -2 to 5, and a hydrogen bond donor/acceptor number within a defined range. Such thresholds were used by taking into account their proven drug-

likeness criteria so that the compounds chosen had features typical of existing drugs.As filtering progressed, a progress bar was introduced through the use of Python's tqdm package to give instant feedback on how the analysis was progressing. The feature helped with transparency and the ability to easily monitor the process of filtering, particularly when handling large data. The progress bar showed the percentage of compounds processed and the estimated time remaining, helping to manage expectations and providing a visual indication of task completion. Once filtering was finished, a report of how many molecules had filtered through the drug-likeness filter was shown. This report gave an indication of how many compounds met the criteria desired, providing insight into the size and quality of the subset produced. The chosen compounds, optimized for possible therapeutic application, were stored in a well-organized CSV file (selected_molecules.csv). The file contained the appropriate molecular descriptors for each compound that survived the drug-likeness filter, so the data was properly organized for subsequent analysis.

By using these drug-likeness filters, we essentially reduced a huge dataset of chemical compounds to a reasonable number of candidates with desirable properties. This systematic filtering process guaranteed that only those compounds most likely to display good pharmacological properties were selected for additional ADME testing and subsequent drug development. The resulting dataset, its reduced set of drug-like compounds, is an optimized base for virtual screening, structure-activity relationship (SAR) analysis, and other

sophisticated cheminformatics research. By narrowing down the dataset to target compounds with the most pertinent features, we optimized the drug discovery pipeline, dramatically improving the efficiency of finding prospective drug candidates [86],[87].

3.6 Protein-Ligand Interaction

In the field of drug discovery, it is necessary to understand how potential therapeutic compounds interact with their target proteins. Here, molecular docking was used to assess the binding affinities of drug-like molecules against certain DNA methyltransferases: DNMT1, DNMT3A, and DNMT3B. The aim of this analysis was to determine compounds with good binding profiles that might increase selectivity against these enzymes, which have critical roles in epigenetic control and are involved in several cancers.

3.6.1 Molecular Docking Overview

Molecular docking is a foundational computational technique widely applied in modern drug discovery. Its primary function is to predict the most favorable orientation, or "pose," of a small molecule (ligand) when bound to a target macromolecule, such as a protein. Beyond predicting the binding geometry, docking simulations estimate the binding affinity—a quantitative measure of the interaction strength between the ligand and its target. A high binding affinity is crucial as it generally correlates with tighter binding and potentially enhanced therapeutic efficacy, making this parameter vital for assessing a compound's potential. By providing atomic-level insights into intermolecular interactions, including hydrogen bonds, hydrophobic contacts, and van der Waals forces, docking plays an indispensable role in the virtual screening of compound libraries, lead identification, and the rational optimization of drug candidates for specific biological targets. In the present study, all molecular docking simulations were performed using the Molecular Operating Environment (MOE) software. MOE is a comprehensive platform renowned in computational chemistry and drug design, offering an integrated suite of tools for molecular modeling, simulation, and analysis. Its robust docking suite employs sophisticated algorithms to efficiently and accurately model ligand-protein interactions. By simulating the docking process, MOE predicts the binding conformation of a compound within a protein's active site and provides detailed analysis of the molecular forces governing the complex's stability. This general methodology was applied to a specific epigenetic target in our research. The filtered, drug-like compounds from our library were

subjected to molecular docking simulations to evaluate their binding interactions with the catalytic domains of DNA methyltransferase isoforms DNMT1, DNMT3A, and DNMT3B. Our computational strategy specifically focused on the S-adenosyl-L-methionine (SAM) cofactor binding pocket. This represents a rational, structure-based approach to identify potential competitive inhibitors. The overarching goal was to discover compounds that directly disrupt the methyl transfer mechanism, which is responsible for HHIP promoter hypermethylation in Hedgehog pathway-driven cancers, thereby aiming to epigenetically reactivate this critical tumor suppressor gene [52],[53].

3.6.2 Preparation of the Protein Structures

The three-dimensional structures of the DNA methyltransferase isoforms—DNMT1 (AF-P26358-F1), DNMT3A (AF-Q9Y6K1-F1), and DNMT3B (AF-Q9UBC3-F1)—were retrieved from the AlphaFold Protein Structure Database as PDB format files. These high-quality predicted structures provided an excellent foundation for our docking studies. Each protein structure was imported into the Molecular Operating Environment (MOE) and underwent a comprehensive molecular preparation workflow. This preprocessing was essential to ensure accurate protonation states and optimal side-chain conformations for docking. The workflow included:

- Removal of crystallographic water molecules and heteroatoms to prevent spurious interactions during simulations.
- Protonation state optimization through the addition of polar hydrogens to correctly model potential hydrogen bonds.
- Energy minimization of the binding site residues to relieve any steric clashes and establish a stable starting conformation.

Catalytic site definitions were derived from experimentally determined crystallographic structures to ensure biological relevance. The docking grid for DNMT1 was constructed based on PDB ID 4WXX, centered on the SAM-binding pocket and encompassing critical residues including Phe1145, Glu1163, and the catalytic cysteine Cys1226 [Song J, et al. Science. 2011]. For DNMT3A, the binding site was modeled according to PDB ID 5YX2, highlighting essential SAM-coordinating residues Arg736, Arg790, and catalytic glutamate Glu697 [Zhang ZM, et al. Nature. 2018]. The DNMT3B active site was generated through structural alignment with DNMT3A, targeting conserved SAM-binding residues Arg630, Arg684, and Glu592.

Concurrently, the library of filtered, drug-like compounds was prepared for docking. The compound structures, initially in a CSV file, were converted to the SDF (Structure Data File) format using OpenBabel to maintain structural integrity and ensure compatibility with the MOE software.

3.6.3 Docking Procedure

The docking was carefully performed for all three DNA methyltransferases—DNMT1, DNMT3A, and DNMT3B—against the compounds that were kept in molecules_1.sdf. It included a set of pre-docking preparations for obtaining accurate and reliable results for the simulation. Certain special considerations and optimizations were made for the improvement of precision in the docking simulations:

- **Water Molecules and Heteroatoms Removal:** Water molecules and heteroatoms were removed from the protein structures prior to docking. This removal is crucial in order to minimize computational complexity and avoid interference from unnecessary interactions that might mask the essential ligand-protein interactions in the active site. With the analysis concentrated on the most important binding components, the prediction of significant interactions becomes more focused and efficient.
- **Addition of Polar Hydrogen Atoms:** Polar hydrogen atoms were incorporated into the enzyme complexes during preparation. These atoms play an essential role in the formation of hydrogen bonds, a major contributor to ligand binding and stability at the active site. The addition of polar hydrogens ensures that the docking simulations consider the possibility of these interactions, which can significantly impact the predicted binding affinity and pose selection.
- **Single Ligand Pose Selection:** For each of the drug-like compounds, one pose only—the most likely position of the ligand within the protein binding site—was rigorously selected prior to running the docking simulation. This procedure is important in terms of obtaining uniformity and consistency in the analysis. Optimized single pose selection minimizes variation and maximizes the likelihood that the most reasonable binding conformation is considered in the simulation, thus optimizing the validity of the docking result.

The docking simulations were performed with the Molecular Operating Environment (MOE) software, which offered a methodical way to assess the interactions between chosen drug-like compounds and the target DNA methyltransferases DNMT1, DNMT3A, and DNMT3B.

Thorough preparations, such as the removal of water molecules and heteroatoms, addition of polar hydrogen atoms, and meticulous choice of a single ligand pose for each compound, guaranteed the precision and validity of the simulations. These efforts simplified the analysis, with the emphasis being placed on major determinants like hydrogen bonding and binding conformation, which play major roles in defining binding affinity. Through the modeling of the interaction of the compounds with the active sites of the enzymes, docking allowed for the identification of likely inhibitors with high binding affinities. The systematic analysis not only yielded information on the molecular forces at play in these interactions but also allowed for compound prioritization to pursue further investigation. The target of the docking simulations was to determine the binding behavior of drug-like molecules to the DNA methyltransferases and identify molecules that have preferential binding to DNMT1, DNMT3A, or DNMT3B. These enzymes play critical roles in epigenetic regulation and are also implicated in multiple types of cancers and thus represent potential therapeutic targets. The insights provided by this study are significant in furthering drug discovery and especially in epigenetic therapies. Through inhibition of aberrant DNA methylation, the identified compounds can play a role in pioneering novel methods of cancer as well as other diseases caused by dysfunctional epigenetic mechanisms treatment. This integrative strategy emphasizes the significance of computational docking as an effective tool in guiding and streamlining the pipeline of drug discovery [88],[89].

Following parallel docking simulations against all three DNMT isoforms, a systematic and tiered approach was employed for hit identification to move from initial affinity assessment to the discovery of selective inhibitors. The primary analysis involved a comparative binding affinity assessment, where all compounds were ranked based on their predicted docking scores (expressed as binding energies in kcal/mol) for each individual isoform: DNMT1, DNMT3A, and DNMT3B. This initial ranking generated three separate prioritized hit lists, highlighting compounds with the strongest predicted binding affinity for each target, irrespective of their activity on the others.

However, given the structural similarities between the catalytic domains of the DNMT isoforms, a primary objective of this study was to move beyond mere affinity and identify compounds with inherent selectivity. To address this critical requirement, a multi-parameter filtering strategy was implemented. This advanced analysis focused on cross-comparing the docking scores of each compound across all three isoforms.

Selective hits were rigorously defined using a percentile-based threshold to ensure robust identification. A compound was classified as a selective hit for a specific isoform (e.g., DNMT1) only if it met two concurrent criteria:

High Affinity for the Target Isoform: The compound's docking score for the target isoform had to fall within the top 5% of all scores for that enzyme, indicating a potent and favorable interaction.

Weak Affinity for Off-Target Isoforms: Concurrently, the same compound's docking scores for the other two, off-target isoforms (e.g., DNMT3A and DNMT3B) had to fall outside the top 20% of scores for those respective enzymes. This ensured a significant and meaningful drop in predicted binding strength, minimizing the likelihood of non-selective pan-inhibition.

This stringent dual-filter approach was crucial for sifting through the vast dataset and pinpointing chemical scaffolds that demonstrate a intrinsic and pronounced selectivity profile—whether for DNMT1, DNMT3A, or DNMT3B. The successful identification of such scaffolds is a fundamental milestone, fulfilling a primary objective of this research and paving the way for the development of targeted epigenetic therapeutics that can modulate specific DNMT isoforms without disrupting the essential physiological functions of others.

3.7 Clustering

After conducting molecular docking analysis, the second step of our research was to cluster the chemical compounds according to their structural properties and binding affinities to DNMT3B. This procedure was designed to reveal molecular descriptor-binding efficacy relationships and improve our insights into structure-activity relationships for drug design purposes.

3.7.1 Data Acquisition and Descriptor Calculation

The first phase of clustering started with acquiring data for chemical compounds, including SMILES (Simplified Molecular Input Line Entry System) representations and binding affinities for DNMT3A alone. SMILES notation offers a representation of molecular structures as text, thereby being convenient to manipulate and analyze with computational resources.Using the RDKit library, diverse molecular descriptors were computed for each SMILES string. These descriptors included:

- **Molecular Weight:** A basic parameter that affects pharmacokinetics.

- **Topological Polar Surface Area (TPSA):** An indicator of the polar atoms' surface area, which determines permeability.
- **Number of Rotatable Bonds:** This reflects flexibility in the molecule, which in turn influences the binding characteristics of the molecule.
- **LogP:** An indicator of lipophilicity, reflecting to what extent a compound distributes itself between hydrophilic and lipophilic situations.
- **Number of Hydrogen Bond Donors and Acceptors**: These are absolutely essential parameters that determine molecular interaction.
- **Aromatic Ring Count:** The aromaticity may have an impact on the stability as well as the interactions of the molecules.

The SMILES strings with invalid formats were removed from the analysis to preserve the validity of the dataset, with only true chemical representations being considered.

3.7.2 Standardization and Hierarchical Clustering

For the purpose of guaranteeing the precision and consistency of the clustering process, the molecular descriptors were initially standardized applying z-score normalization. This statistical method entails rescaling every descriptor in such a way that it has a mean value of zero and a standard deviation of one. Through the standardization of data, differences in the range and scale of the descriptors are removed, guaranteeing that no particular feature disproportionately affects the clustering algorithm. This step is particularly important when dealing with varied descriptors, like molecular weight, TPSA, and LogP, that inherently range across different scales. Standardization allows each descriptor to make an equal contribution in establishing the relationships between compounds so that a balanced and unbiased clustering can take place.After the data was standardized, hierarchical clustering was conducted with Ward's method. This method is particularly known for its capacity to reduce variance among clusters, creating tight and significant groupings of compounds. Ward's algorithm works by successively combining two clusters at a time that result in a minimum increment of overall within-cluster variance. The repetitive merging gives a hierarchical tree-like structure called a dendrogram, representing the relations and similarities between the compounds visually. Every branch of the dendrogram represents a cluster, and the length of the branches indicates the level of similarity or dissimilarity among the compounds. In order to establish the number of optimal clusters, the Elbow Method was used. The method entails graphing the within-cluster sum of

squares (WCSS), which measures cluster compactness, versus the number of clusters. The WCSS reduces with an increase in the number of clusters, but at some point, the rate of reduction becomes very low, creating an "elbow" in the graph. The "elbow" point is the optimal point of balance between simplicity (less number of clusters) and fidelity (proper representation of data structure). By choosing the best number of clusters here, the analysis guarantees a stable partitioning of the compounds that captures useful patterns in their molecular descriptors without oversimplifying or overfitting the data.This systematic process showed hierarchical clustering to identify distinct groups of structurally related compounds and to lay a basis for following analyses intended to correlate these groups with binding affinities and to define main structure-activity relationships [90],[91].

3.7.3 Binding Affinity Assessment and Statistical Analysis

Following the clustering of compounds, the final step was to calculate the average binding affinities for all the clusters. This gave a summary of the average binding competence displayed by all the compounds belonging to each group, providing a comparative view within the clusters. Through the consideration of these averages, it became feasible to identify trends and correlation between molecular patterns and their efficiency in binding with DNMT3A. For instance, the clusters that showed greater average affinities to binding might signify certain structural or molecular features favoring binding in some way. On the other hand, those with less favorable affinities may point to structural restraints that limit efficiency of interaction.A statistical examination, using ANOVA (Analysis of Variance), was therefore conducted to establish the link between structural characteristics and binding efficiency. ANOVA is a strong technique for assessing if there are any significant differences in the means of several groups in this case, the clusters. Through its application, we were able to ascertain if the variability seen in average binding affinities across clusters were statistically different. Not only did this test confirm the clustering process, but it also pointed out major molecular descriptors that could be influential in determining binding affinity. These observations are precious in advising future compound development by highlighting the characteristics of the increased efficacy.To determine the strength and quality of the clustering, the silhouette value was computed. This measure looks at how much each compound resembles its own group compared to any other group it could be clustered with. Silhouette values span from -1 to 1, with bigger values representing distinctly defined and strong clusters. A high silhouette value indicates that

compounds within a cluster are closer to one another than to compounds within other clusters, indicating good separation and informative clustering. Conversely, low or negative scores may signify overlapping or ill-defined groups, necessitating reconsideration of the cluster parameters.In bringing these analyses together—average binding affinity calculation, ANOVA for statistical verification, and silhouette rating for quality determination—the results from the clustering were comprehensively understood. This method guaranteed that the discovered clusters not only represented statistical significance but also biological relevance, paving the way for further investigation of structure-activity relationships in drug discovery [90],[91].

3.7.4 Data Visualization

To offer a richer insight into the results of the clustering, a range of data visualization methods were utilized. The visualizations played a central role in interpreting relationships and trends in the dataset, especially in terms of binding affinities and structural similarities between the compounds.Box plots allowed the binding affinity distribution in each cluster to be displayed. By showing the median, quartiles, and possible outliers within each cluster, these plots also brought out differences in binding effectiveness within and between clusters. Patterns like clusters of uniformly high or low binding affinities were revealed, providing hints at structural features behind more or less effective interactions. These visual overviews also enabled comparison of binding performance between clusters, which helped in the identification of exceptional groups that need to be explored further. To represent the hierarchical structure of the clusters visually, dendrograms were created. A dendrogram gives a tree-like representation of the structure of clusters in terms of their similarity. In this research, the dendrogram exposed how single compounds clustered together to cluster into bigger clusters, providing information on the relative closeness and grouping of compounds based on their molecular descriptors. The visualization also conveyed the hierarchical clustering operation clearly and enabled a clear recognition of how close the compounds were in each grouping. PCA was also applied to reduce the high-dimensional dataset into a reduced lower-dimensional space, often two or three principal components. PCA is used to preserve the most important information contained in the data and eliminate the less important variability. PCA plots identified patterns and groupings within this analysis and allowed visualization of how clusters had been distributed across the reduced dimensional space. Plotted on the principal components, it was now easier to see any separations or overlaps between the clusters and so get a clearer idea of structure-activity

relationships.These visual methods, coupled together in the form of box plots, dendrograms, and PCA plots, not only rendered the cluster results more comprehensible but even gave some rich insights into what molecular descriptors go with which binding affinities and structural similarity. These visual tools played a crucial role in validating the clustering process, highlighting important patterns, and directing the investigation of potential lead compounds for further drug design [90],[91].

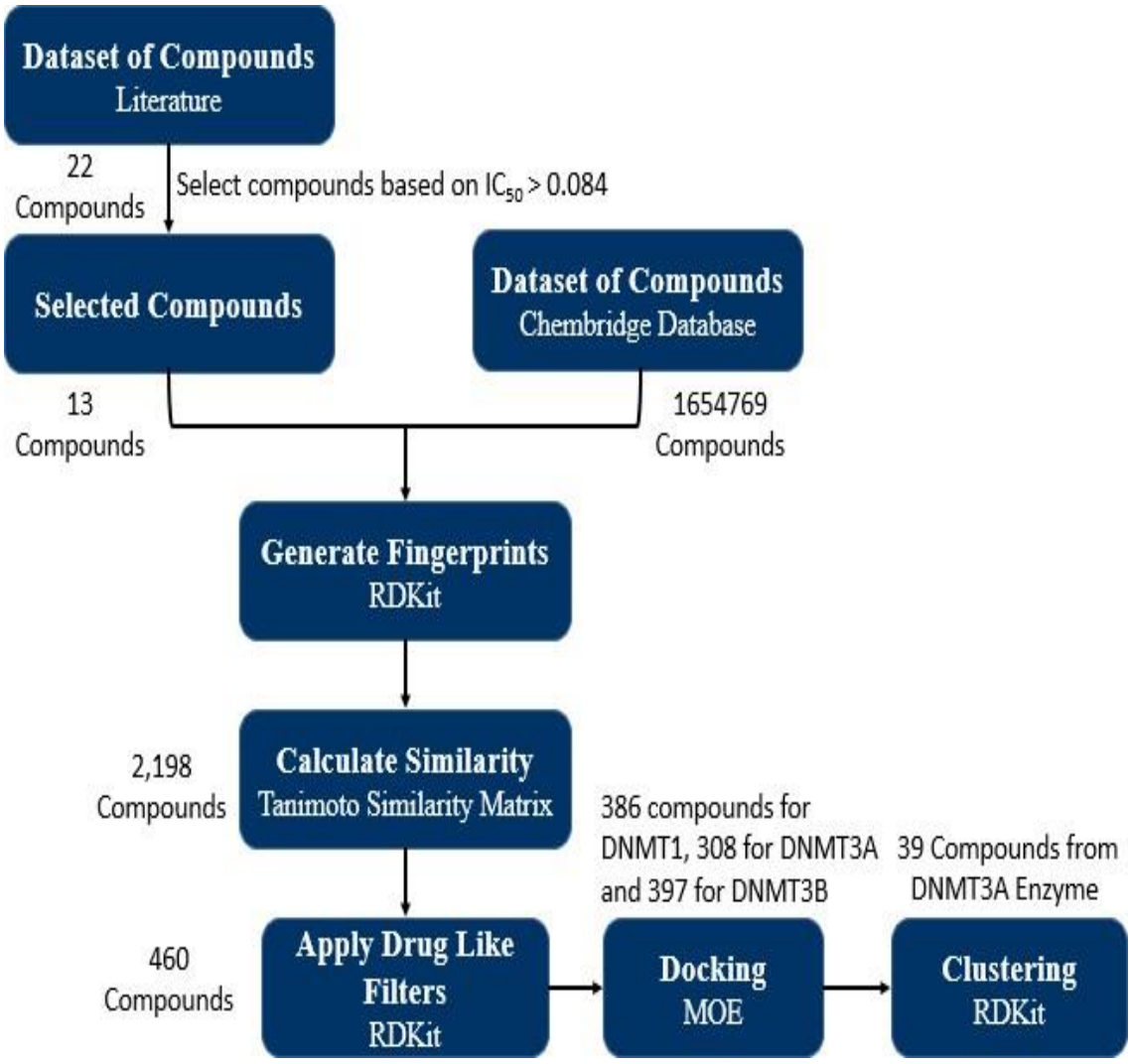

Figure 3.1: Schematic Methodology for the Discovery of Inhibitors Targeting DNA Methyltransferases (DNMTs)

Chapter 04

Results

4.1 Dataset Partition and Selection Criteria

The initial dataset comprised 22 chemical compounds sourced from the literature, all of which had established IC50 (half maximal inhibitory concentration) values. These IC50 values are critical in assessing the potency of compounds, as they indicate the concentration required to inhibit a biological process or enzyme activity by 50%. For the purpose of focused and efficient analysis, 13 compounds were selected based on the criterion of IC50 > 0.084. This threshold ensured the retention of compounds with sufficient potency for further investigation, eliminating those with lower biological activity. The application of this criterion allowed for a targeted examination of compounds that are more likely to exhibit significant therapeutic potential, streamlining the drug discovery process. The refined dataset provided a solid foundation for subsequent in-depth analysis, as represented in Figure 4.1.

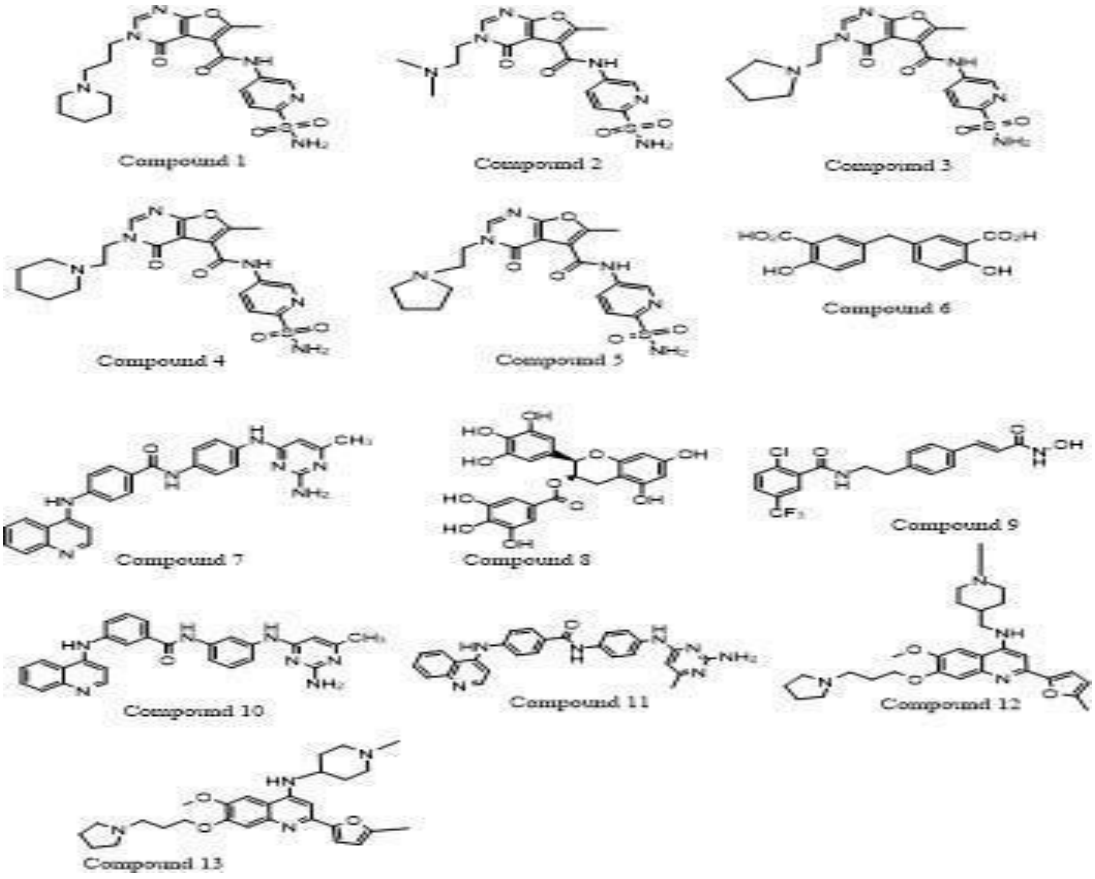

Figure 4.1 The chemical structure of the selected compounds

In addition to the literature-based compounds, a large dataset of 1,654,770 untested chemical compounds was procured from ChemBridge (https://chembridge.com/). ChemBridge is a widely recognized source for chemical libraries, offering diverse compounds suitable for high-throughput screening and drug development projects. These untested compounds provided a vast pool of possibilities, significantly expanding the scope of the research and increasing the probability of identifying novel inhibitors with potent biological activity. By integrating the curated dataset of known compounds and the expansive ChemBridge dataset, the study was designed to balance reliability with exploratory potential, aiming to uncover compounds with robust inhibitory properties and promising therapeutic applications.

4.2 Data Preparation and Molecular Descriptors

The SMILES (Simplified Molecular Input Line Entry System) strings for the 13 chosen reference compounds were generated using ChemDraw, a widely used tool for drawing and analyzing chemical structures. The resulting SMILES strings were saved in a file named 13_compounds.txt for further computational analysis. These reference compounds, with their precisely curated IC50 values, formed the benchmark for subsequent comparisons and served as controls for the broader screening process. For the expansive ChemBridge dataset, containing 1,654,770 untested compounds, a different approach was required due to its large size. The SMILES strings for these compounds were extracted programmatically using Python scripting, leveraging the capabilities of the cheminformatics library RDKit. The sdf_to_smiles function in RDKit facilitated the conversion of the SDF (Structure Data File), a standard format for representing molecular structures, into the SMILES format. This converted data was saved in a file named chembridge_compounds_smiles.txt, which served as the primary input for subsequent analyses. Given the vast volume of the ChemBridge dataset, it was further divided into 17 smaller text files (e.g., chembridge_1.txt to chembridge_17.txt) using a Python script. This partitioning was necessary to ensure efficient processing and data management, allowing for seamless high- throughput screening and analysis without overloading computational resources. Each of these text files represented a manageable subset of the dataset, enabling streamlined operations such as similarity assessments, clustering, and filtering for drug-like properties. The organization of the dataset into smaller segments not only enhanced processing efficiency but also

ensured the reproducibility of results, as smaller datasets are easier to verify and reanalyze if needed. This preprocessing step, as depicted in Figure 4.2, was critical for setting up the pipeline for high-throughput screening. By converting and organizing the data into SMILES format, the study ensured compatibility with a wide range of cheminformatics tools, enabling detailed analysis of molecular descriptors, structural comparisons, and similarity-based filtering. These steps are integral to drug discovery workflows, as they allow researchers to efficiently sift through extensive libraries to identify promising candidates for further investigation. The structured organization of the dataset also facilitates advanced computational studies, including machine learning-based predictions and virtual

```
# Define filename for chembridge compounds
Set filename to "chembridge_compounds_smiles.txt"

# Read chembridge compounds file
Open filename for reading
Read all lines and split them into a list named "db_read"

# Remove trailing empty line (if present)
If the last element of "db_read" is empty, remove it

# Initialize counter for split files
Set counter "n" to 1

# Loop through chembridge compounds in chunks
For each chunk of 100,000 elements in "db_read":
    # Extract current chunk
    Extract a sublist named "split" from "db_read" starting at index "t" and ending at index "t+100000" (exclusive)

    # Join chunk elements with newline characters
    Combine all elements in "split" into a single string named "join" using "\n" as separator

    # Create a new file for the chunk
    Create a new file named "chembridge_"+str(n)+".txt" for writing

    # Write the chunk to the new file
    Write the contents of "join" to the newly created file

    # Increment counter for next chunk
    Increase "n" by 1

# Close files (implicitly done in Python with 'with' statement)
```

docking experiments, which are essential for modern drug discovery efforts.

Figure 4.2: This pseudocode demonstrates a structured approach to handling a large file containing SMILES data for compounds from ChemBridge. Due to the dataset size, which could be challenging to process all at once, this code breaks down the data into manageable chunks of 100,000 entries each and saves each chunk in a separate file. The initial step is to define the filename, "chembridge_compounds_smiles.txt", which contains the SMILES strings for various compounds.

The code opens this file in read mode, reads all lines at once, and stores each line in a list named db_read. Each line in db_read represents a SMILES entry, facilitating easy manipulation of each compound's information. To handle potential formatting inconsistencies, the pseudocode checks if the last line is empty and removes it if necessary. Next, it initializes a counter variable n to 1. This counter will be used to number each output file, allowing easy tracking and organization of the chunks. The main part of the pseudocode is a loop that processes the db_read list in blocks of 100,000 entries. Within each iteration, a subset of 100,000 SMILES entries is extracted into a temporary list named split, which represents the current chunk. The entries in split are then joined together into a single string, join, with newline characters as separators, ensuring each entry is written on a new line in the output file. For each chunk, a new file is created with a name based on the current value of n, such as "chembridge_1.txt", "chembridge_2.txt", and so on. The contents of join are then written to this file. After writing, the counter n is incremented by 1 to prepare for the next file. Using Python's with statement implicitly handles file closure after each write operation, ensuring efficient resource management without requiring explicit file closure commands. This pseudocode is effective for handling large datasets in chunks, making it easier to process and analyze SMILES data in smaller, organized files.

4.3 Tanimoto Similarity and Compound Selection

MACCS Key fingerprints, a widely used molecular descriptor for structure-based similarity analysis, were computed for both the reference file (13_compounds.txt) and the segmented ChemBridge dataset using the RDKit cheminformatics library. This process resulted in the generation of 13 fingerprints for the reference compounds and 1,654,770 fingerprints for the ChemBridge dataset. These fingerprints, consisting of a predefined set of 166 binary keys, capture structural features of molecules, enabling rapid comparisons based on molecular similarity. The results of this computation are presented in Table 4.1 and Table 4.2, providing an overview of the fingerprints derived from each dataset. The corresponding Python code used for fingerprint generation is illustrated in Figure 4.3.To identify compounds with structural similarity to the reference set, pairwise Tanimoto similarity scores were calculated between each ChemBridge compound and the reference compounds. The Tanimoto coefficient, ranging from 0 (no similarity) to 1 (identical structures), is a robust metric for evaluating molecular similarity. Through this analysis, 2,198 ChemBridge compounds were identified as having a Tanimoto similarity >80% with the reference compounds, as summarized in Table 4.3. The Python

scripts employed for this similarity analysis are provided in Figure 4.4 and Figure 4.5, detailing the steps for calculating Tanimoto scores and filtering compounds based on the similarity threshold.The high-similarity compounds were compiled into a single text file named similar_smiles.txt to facilitate downstream processing. Subsequently, these SMILES strings were converted back into the SDF format using RDKit, resulting in a file named molecules.sdf for further analyses. This transformation is depicted in Figure 4.6, showcasing the workflow from similarity assessment to the generation of an actionable molecular dataset. This rigorous similarity criterion ensures the selection of compounds that are not only structurally relevant but also potentially functionally aligned with the reference compounds. By focusing on molecules with >80% similarity, the study increases the likelihood of identifying DNMT inhibitors with significant therapeutic potential. The streamlined workflow, combining fingerprint generation, Tanimoto scoring, and dataset refinement, represents an efficient and effective strategy for narrowing down large chemical libraries to a manageable subset of promising candidates. This approach is pivotal in drug discovery, as it prioritizes compounds that share key structural features with known active molecules, thereby facilitating the identification of novel inhibitors with enhanced structural and functional relevance.

Table 4.1: Fingerprints of Reference Compounds

Sr No.	Structure	Fingerprints
1	6-methyl-4-oxo-3-(3-(piperidin-1-yl)propyl)-N-(6-sulfamoylpyridin-3-yl)-3,4-dihydrofuro[2,3-d]pyrimidine-5-carboxamide	00000000000000000000000000000001100000000000000100010001011111110011 01010001010101110111101100101111101110011000111100011011101101111111 01011110101111111011111110111111110
2	3-(2-(dimethylamino)ethyl)-6-methyl-4-oxo-N-(6-sulfamoylpyridin-3-yl)-3,4-dihydrofuro[2,3-d]pyrimidine-5-carboxamide	00000000000000000000000000000001100000000000000100010001011111110011 01010001110101110111101100111111101110011000111101111011101101101011 01011110111111111111111110111111110

3	6-methyl-4-oxo-3-(2-(pyrrolidin-1-yl)ethyl)-N-(6-sulfamoylpyridin-3-yl)-3,4-dihydrofuro[2,3-d]pyrimidine-5-carboxamide	00000000000000000000000000000001100000000000000100010001011111110011 01010001010101110111011001011111011100110001111000110111011011111 01011110101111111010111111101111111110
4	6-methyl-4-oxo-3-(2-(piperidin-1-yl)ethyl)-N-(6-sulfamoylpyridin-3-yl)-3,4-dihydrofuro[2,3-d]pyrimidine-5-carboxamide	00000000000000000000000000000001100000000000000100010001011111110011 01010001010101110111011001011111011100110001111000110111011011111 01011110101111111010111111101111111110
5	6-methyl-4-oxo-3-(3-(pyrrolidin-1-yl)propyl)-N-(6-sulfamoylpyridin-3-yl)-3,4-dihydrofuro[2,3-d]pyrimidine-5-carboxamide	00000000000000000000000000000001100000000000000100010001011111110011 01010001010101110111011001011111011100110001111000110111011011111 01011110101111111010111111101111111110
6	2-hydroxy-5-(4-hydroxybenzyl)benzoic acid compound with methanedione (1:1)	00000000000000000000000000000000000000000000000000000000000010000000000000 00000000000000000000000100000000000000000000000000000010000000001001010001 00001001100111100010101101010011111
7	N-(4-((2-amino-6-methylpyrimidin-4-yl)amino)phenyl)-4-(quinolin-4-ylamino)benzamide	00000000000000000000000001000000000000010000000000000010000000000001 00000000000100100010000000100000100100011000100000010011000100000001 01010100001001000001001010101111110
8	N-(3-((2-amino-6-methylpyrimidin-4-yl)amino)phenyl)-3-(quinolin-4-ylamino)benzamide	00000000000000000000000001000000000000010000000000000010000000000001 00000000000100100010000000100000100100011000100000010011000100000001 01010100001001000001001010101111110
9	(2R,3R)-5,7-dihydroxy-2-(3,4,5-trihydroxyphenyl)chroman-3-yl 3,4,5-trihydroxybenzoate	00000000000000000000000000000000000000000000000000000000000011001000001000 00000010000000000000000110000000100100010000000100000000010111000110 10000101100111100010101001010011110
10	N-(4-((2-amino-6-	00000000000000000000000001000000000000010000000000000010000000000001

	methylpyrimidin-4-yl)amino)phenyl)-4-(quinolin-4-ylamino)benzamid e	00000000000100100010000000100000100100011000100000001001100 01000001 01010100001001000001001010101111110
11	6-methoxy-2-(5-methylfuran-2-yl)-N-((1-methylpiperidin-4-yl)methyl)-7-(3-(pyrrolidin-1-yl)propoxy)quinoli n-4-amine	00000000000000000000000000000000000000000000000000000001 00001001 00000010010000101101100011010111101100010001010101101011100 01111100 11010110011111111111110111111111110
12	6-methoxy-2-(5-methylfuran-2-yl)-N-(1-methylpiperidin-4-yl)-7-(3-(pyrrolidin-1-yl)propoxy)quinoli n-4-amine	00000000000000000000000000000000000000000000000000000001 00001001 00000010010000100101100010010111101100110001010101101011100 01111100 11010110011111111111110111111111110
13	(E)-3-(4-(((2-(2-chloro-5-(trifluoromethyl)ph enyl)-2-oxoethyl)amino)me thyl)phenyl)-N-hydroxyacrylamide	00000000000000000000000100000000000000000100000000000000 00000000 00110110000000001000110000101100011011001100101000010000000 11000001 10101011001011100011111110110111110

This table presents a selection of 13 literature-derived compounds, each represented by its unique chemical structure and corresponding MACCS Key molecular fingerprint. These compounds were identified based on their potential as DNA methyltransferase (DNMT) inhibitors, specifically targeting DNMT1, DNMT3A, and DNMT3B. The structure column displays the IUPAC names of each compound, providing a detailed look into the chemical composition and functional groups, which are critical for understanding the compounds' interactions with DNMTs. Each compound's fingerprint, displayed as a binary sequence, represents the MACCS Key descriptors. These fingerprints capture essential structural and functional characteristics, such as the presence of specific atom types, ring systems, and functional groups, which are relevant for binding interactions with DNMT enzymes. The binary values (1s and 0s) indicate the presence or absence of specific molecular features, with each bit position corresponding to a unique chemical descriptor in the MACCS Key format. This encoded information enables a quantitative comparison of structural similarity between compounds, facilitating the clustering and selection of compounds with desirable drug-like properties. Compounds such as 6-methyl-4-oxo-3-(piperidin-1-yl)pyrido[1,2-a]pyrimidin-3-yl)3,4-dihydrofuro[2,3-d]pyrimidine-5-carboxamide are shown in detail, with their fingerprints illustrating a rich variety of chemical features, potentially contributing to robust binding affinities with DNMTs.

The variety in structures across these compounds suggests a diversity in potential binding modes and inhibitory effects on DNMTs, which may prove valuable for targeted inhibition. This dataset will be further analyzed to assess structural-activity relationships and optimize candidates for potential therapeutic development.

Table 4.2: Fingerprints of ChemBridge Compounds

Sr No.	Structure	Fingerprints
1	CC1CN(C2CCNCC2)CC(C)O1	00000000000000000000000000000000000000000000000000000000000000000100001000000000001000010100110000100010010100010001001000100101110000011001100011000100101110111001110111011101110
2	NC1=NCC(CO)S1	00000000000000000000000000000000000000010000000000100000000000000000000000000000011001111000101100001100100000100100000000011100000000010100011100100000000101011110010010000
3	CCc1nc(CN(C)c2ncc3cn[nH]c3n2)no1	00000000000000000000000000100000000000001000000000000010000100000001000010100000101100101100000111111101110011000110011010011101100101011010100001000001101010110101111111000
4	Cc1ccc2c(c1)CC(C(=O)O)N 2	00000000000000000000000000000000000000000000000000000000000000000100000000000000000000000000001000000100001100001001100100100000000010100000001000101010000100000010010111111111110
5	Nc1ccccc1N1CCCC1	00000000000000000000000000000000000000000000000000000000000000000000000010000000000010001000111100001000010001000000000010000001001100000010001010110001010011011010010100111010

This table presents a subset of compounds derived from the ChemBridge database, specifically showcasing the first five compounds out of an extensive dataset of 1,654,770 molecules. Each compound is represented by its SMILES (Simplified Molecular Input Line Entry System) notation, which concisely represents the molecular structure in terms of atoms, bonds, and stereochemistry. SMILES notations allow efficient computational processing and facilitate the structural analysis of large datasets. The "Fingerprints" column displays the MACCS (Molecular ACCess System) Key fingerprints for each compound, which are binary sequences encoding specific molecular features and properties. These fingerprints consist of 166 bits, each representing the presence (1) or absence (0) of a particular structural feature, such as certain functional groups, ring structures, or atom configurations. For example, the sequence begins with several zeros, indicating that these compounds lack certain basic structural elements, while specific bits later in the sequence may indicate the presence of more complex functional groups relevant to binding with target proteins. These MACCS

```
# Import necessary libraries
import rdkit
import pandas as pd

# Read molecule lists
Read reference molecules from "13 compounds.txt"
Read database molecules from "chembridge_17.txt"

# Convert SMILES to molecules
For each SMILES string in reference molecules:
    Convert SMILES to molecule object
    Append molecule object to reference molecules list

For each SMILES string in database molecules:
    Convert SMILES to molecule object
    Append molecule object to database molecules list

# Generate fingerprints
For each molecule in reference molecules:
    Generate MACCS fingerprint
    Append fingerprint to reference fingerprints list

For each molecule in database molecules:
    Generate MACCS fingerprint
    Append fingerprint to database fingerprints list

# Calculate similarity
Initialize empty lists for reference compound numbers, database compound numbers, and similarity scores

For each reference fingerprint:
    For each database fingerprint:
        Calculate Tanimoto similarity between reference and database fingerprints
        Append reference compound number to reference compound numbers list
        Append database compound number to database compound numbers list
        Append similarity score to similarity scores list
```

fingerprints are essential for assessing structural similarity between compounds, allowing researchers to identify clusters of molecules with potentially similar biological activities.

```
For each reference fingerprint:
    For each database fingerprint:
        Calculate Tanimoto similarity between reference and database fingerprints
        Append reference compound number to reference compound numbers list
        Append database compound number to database compound numbers list
        Append similarity score to similarity scores list

# Create and save DataFrame
Create a DataFrame with columns:
    - Reference Compound No.
    - Database Compound No.
    - Similarity

Save the DataFrame as "similarity_17.csv"
```

Figure 4.3: The pseudocode describes a process for calculating molecular similarity using SMILES data, leveraging libraries such as RDKit for cheminformatics tasks and Pandas for data handling. First, necessary libraries are imported, with RDKit handling molecular structures and calculations, and Pandas managing data reading and storage. The code then reads two text files, "13_compounds.txt" and "chembridge_17.txt," which contain SMILES strings—a simplified representation of molecular structures. For each SMILES string in "13_compounds.txt," RDKit converts the string into a molecular object, which is then added to a list named "reference_molecules." A similar process is applied to the SMILES strings in "chembridge_17.txt," generating a list called "database_molecules." Once the molecules are prepared, the code generates MACCS fingerprints for each molecule in both lists. MACCS fingerprints are binary representations indicating the presence or absence of specific chemical substructures within each molecule, and these fingerprints are stored in separate lists: "reference_fingerprints" and "database_fingerprints." Following fingerprint generation, the code initializes three empty lists—"reference_compound_numbers," "database_compound_numbers," and "similarity_scores"—for storing similarity calculation results. For each fingerprint in "reference_fingerprints," the code calculates the Tanimoto similarity with each fingerprint in "database_fingerprints." The Tanimoto similarity, a metric to quantify the similarity between two molecules, is then recorded. The index of the reference compound is added to "reference_compound_numbers," the index of the database compound is added to

"database_compound_numbers," and the similarity score is stored in "similarity_scores." After calculating similarity scores for all pairs of reference and database compounds, the code creates a DataFrame with columns for "Reference Compound No.," "Database Compound No.," and "Similarity," which contain the respective indexes and similarity score .Finally, this DataFrame is saved as a CSV file named "similarity_17.csv," preserving the results of the similarity analysis for future reference and analysis.

```
# Import pandas library
Import pandas as pd

# Read CSV file
Read CSV file "similarity_17.csv" into DataFrame df

# Filter DataFrame
Filter DataFrame df where Similarity >= 0.8
Store filtered DataFrame in filter_80

# Extract unique database compound numbers
Extract unique values from the "Db Compound No." column of filter_80
Store unique values in unique_db

# Create new DataFrame
Create a new DataFrame df2 with a single column "Similar Compound No."
Populate the column with unique_db values

# Save new DataFrame as CSV
Save df2 as "Similar Chembridge Compound No 17.csv" without index
```

Figure 4.4: The pseudocode outlines a process for filtering and extracting high-similarity compounds from a dataset. First, the Pandas library is imported under the alias pd, providing powerful data manipulation tools. The code reads a CSV file named "similarity_17.csv" into a DataFrame called df, allowing for structured access and analysis of the data within. Next, it filters the DataFrame to retain only the rows where the "Similarity" column has values greater than or equal to 0.8, representing highly similar compounds. This filtered subset of data, stored in a variable named filter_80, focuses on compounds with significant similarity scores. Following this, the code extracts unique values from

the "Db Compound No." column of the filter_80 DataFrame, identifying distinct database compounds that exhibit high similarity to reference compounds. These unique values are stored in a variable named unique_db. The code then creates a new DataFrame called df2, with a single column named "Similar Compound No." This column is populated with the unique database compound numbers from unique_db, representing the compounds of interest. Finally, the code saves df2 as a new CSV file named "Similar Chembridge Compound No 17.csv." The index=False argument is specified to ensure that row indices are excluded from the output file, preserving only the essential data. This workflow facilitates efficient identification and storage of highly similar compounds for further analysis.

```
# Import pandas library
Import pandas as pd

# Read similar compound numbers
Read CSV file "Similar Chembridge Compound No 17.csv" into DataFrame df
Extract a list named "comp_no" from the "Similar Compound No." column of df

# Read chembridge library
Open file "chembridge_17.txt" for reading
Read all lines and split them into a list named "read_lib"

# Initialize empty string for SMILES
Initialize an empty string variable named "smiles"

# Loop through compound numbers
For each compound number in "comp_no":
    # Access corresponding SMILES from library
    Index library based on compound number (x-1) (assuming zero-based indexing)
    Append the retrieved SMILES string from "read_lib" to "smiles"
    Add a newline character ("\n") to "smiles"

# Write SMILES to file
Open file "similar_smiles.txt" for appending (mode 'a')
Write the contents of "smiles" to the file

# Close files (implicitly done in Python with 'with' statement)
```

Figure 4.5: The pseudocode describes a procedure for extracting specific SMILES strings of similar compounds from a dataset and saving them to a new file. It begins by importing the Pandas library, a

powerful tool for data manipulation in Python, using the alias pd. The code then reads a CSV file named "Similar Chembridge Compound No 17.csv" into a DataFrame, df, which holds the list of similar compound numbers. From the "Similar Compound No." column in df, it extracts these compound numbers into a list named comp_no. Next, the code reads a library of SMILES strings by opening the file "chembridge_17.txt" in read mode. All lines from this file are stored in a list called read_lib, where each line likely represents a SMILES string, a common way to represent molecular structures in cheminformatics. An empty string variable, smiles, is initialized to store the SMILES strings of the selected compounds. The code then iterates through each compound number in comp_no. For each number, it retrieves the corresponding SMILES string from read_lib, using zero-based indexing (i.e., accessing the SMILES string at index x-1, where x is the compound number). The retrieved SMILES string is appended to the smiles string, with a newline character added to separate entries. Finally, the code opens a new file named "similar_smiles.txt" in append mode, writing the complete smiles string, which contains the SMILES of all selected compounds, into this file. This process effectively reads and filters relevant compound numbers, extracts their SMILES strings from a library, concatenates them, and saves them in a structured format for further analysis.

```
# Import necessary libraries
Import os
Import Chem from rdkit
Import AllChem from rdkit.Chem

# Define input and output file paths
Set input_txt to "similar_smiles_1.txt"
Set output_sdf to "molecules.sdf"

# Open input file
Open input_txt for reading

# Create SDF writer
Create an SDWriter object for output_sdf

# Process each line in the input file
For each line in the input file:
    Split the line into parts based on whitespace
    If the line is empty, skip to the next line

    Extract the SMILES string from the first part
    Extract additional data from the remaining parts

    Convert the SMILES string to an RDKit molecule object
    If the conversion fails, print an error message and skip to the next line

    Optionally, add properties to the molecule from the additional data

    Write the molecule to the SDF file

# Close the SDF writer
Close the SDWriter
```

Figure 4.6: The pseudocode outlines a process to convert SMILES strings from a text file into an SDF file format using RDKit, a cheminformatics library in Python. It starts by importing necessary libraries: os for handling file paths and Chem and AllChem from RDKit for managing chemical structures and performing chemical informatics tasks. The input file path is specified as "similar_smiles_1.txt", which is expected to contain SMILES strings representing molecular structures. The output path is defined as "molecules.sdf", where the converted molecules will be saved in the SDF format, commonly used for storing detailed molecular data. The code then opens the input text file in read mode, enabling it to process each line sequentially. An SDWriter object is created for the output SDF file, allowing for the molecular data to be written in the desired format. For each line in the input file, the code splits it based on whitespace to separate the SMILES string from any additional information. If a line is empty, it skips to the next line. The first part of the split line is

extracted as the SMILES string, and any remaining parts are considered additional molecular data, such as name or properties. The SMILES string is then converted into an RDKit molecule object. If the conversion is unsuccessful—possibly due to an invalid SMILES string— the code prints an error message and moves to the next line. When successful, optional properties from the additional data can be assigned to the molecule object, adding details like the molecule's name or other descriptors. The molecule object is then written to the SDF file using the SDWriter, ensuring each molecule's data is saved in a structured format suitable for further analysis or visualization. Finally, the SDWriter object is closed to confirm that all data is securely written and the file is properly saved.

Table 4.3: Similar Compounds

Sr No.	Similar Smiles
1	NC(=O)c1c(NC(=O)CN2C(=O)c3ccccc3S2(=O)=O)sc2c1CCCC2
2	COc1ccc(S(=O)(=O)Nc2ccccc2C(=O)N2CCN(c3ccccc3)CC2)cc1
3	COc1ccc(S(=O)(=O)Nc2cccnc2)cc1C(=O)N1CCOCC1
4	Cc1[nH]c(=O)[nH]c(=O)c1S(=O)(=O)N1CCN(Cc2ccccc2)CC1
5	Cc1c[nH]c(C(=O)N2CCc3ccc(S(=O)(=O)Nc4cc(C)on4)cc3C2)c1

This table presents a selection of five representative compounds from a dataset of 2,198 structurally similar molecules, identified using the Tanimoto similarity matrix. The Tanimoto coefficient, calculated based on MACCS fingerprints, is a widely used measure for comparing molecular similarity in cheminformatics. This metric ranges from 0 to 1, where values closer to 1 indicate higher similarity between compounds. Here, compounds with a high Tanimoto similarity to an initial reference set of DNMT inhibitors were selected for further analysis.The "Similar SMILES" column displays the SMILES notation for each compound, a textual representation that encodes the chemical structure, specifying atom connectivity and certain stereochemical information. The SMILES strings offer a standardized way to represent molecular structures computationally, which facilitates efficient similarity searches and comparisons.These selected compounds contain recurring structural motifs that may contribute to their similar functional properties. For instance, most of the compounds in this table feature a sulfonamide group (–S(=O)(=O)N–) attached to an aromatic ring, as well as cyclic

amine structures, such as piperazine or related nitrogen-containing rings. These functional groups are known to play a crucial role in interactions with biomolecular targets like DNMT enzymes, which makes them particularly valuable for exploring potential inhibitory activity. The presence of carbonyl (C=O) groups adjacent to nitrogen atoms is another common feature, providing potential binding sites that could **interact with catalytic or binding sites on DNMTs.**

4.4 Drug-Likeness Evaluation

The drug-likeness evaluation of the 2,198 compounds in the file molecules.sdf was performed using RDKit, a powerful cheminformatics toolkit. This step aimed to assess whether the selected compounds possessed the necessary properties for favorable pharmacokinetics and pharmacodynamics, essential for drug development. A set of key molecular metrics was analyzed, including:

• Molecular weight (MW) – ensuring it falls within the optimal range for drug absorption and bioavailability.

• LogP – representing lipophilicity, crucial for permeability and solubility.

• Hydrogen bond donors and acceptors – reflecting the potential for molecular interactions with biological targets.

• Rotatable bonds – indicating molecular flexibility, impacting bioavailability and binding efficacy.

• Molar refractivity – linked to molecular volume and polarizability.

• Topological surface area (TPSA) – associated with compound transport properties like cell permeability.

• Formal charge – ensuring neutral or appropriate charge for interaction with biomolecules.

• Heavy atom count – defining molecular complexity.

• Ring count – evaluating structural features relevant for drug stability and binding.

Each compound was filtered against these criteria, ensuring compliance with thresholds that align with drug-likeness standards, such as Lipinski's Rule of Five and related extended rules. This rigorous filtration process reduced the dataset to 460 compounds that satisfied the drug-likeness requirements,

as detailed in Table 4.4.The resulting dataset was saved in a structured CSV file named selected_molecules.csv, enabling seamless integration with downstream analysis workflows. The Python script utilized for the filtration and data export is illustrated in Figure 4.7. By organizing the refined data in a CSV format, the process ensures enhanced accessibility and ease of interpretation for further computational and experimental studies.This drug-likeness filtration step significantly narrowed down the compound pool to those with favorable pharmacokinetic characteristics, improving the likelihood of identifying viable drug candidates. By focusing on these 460 compounds, subsequent validation through molecular docking investigations will concentrate on high-potential candidates, streamlining efforts to identify promising inhibitors. This approach represents an efficient strategy in drug discovery, prioritizing compounds with optimal properties for therapeutic development.

```
# Import necessary libraries
Import Chem, Descriptors, Crippen, QED, rdMolDescriptors from rdkit
Import progressbar
Import csv

# Check if script is executed directly
If script is executed as the main program:

    # Read molecules from SDF file
    Set molecules to an SDMolSupplier object for "molecules.sdf"

    # Initialize results dictionary
    Create a dictionary named "results" with keys:
        - "Drug-like Filter": set to 0
        - "Passes All Filters": set to 0

    # Print molecule database length
    Print the number of molecules in the SDF file

    # Create progress bar
    Create a progress bar object with maximum value set to the number of molecules

    # List to store properties of filtered molecules
    Initialize an empty list named "selected_molecules"

    # Loop through each molecule in the SDF file (using progress bar)
    For each index (i) from 0 to the total number of molecules:

        # Get the current molecule
        Extract the molecule object at index i from "molecules"

        # Check if molecule is valid
        If the molecule is not None:
```

```
            # Properties calculation
            Calculate and store:
                - Molecular Weight using Descriptors.MolWt
                - LogP using Crippen.MolLogP
                - H-Bond Donors using Descriptors.NumHDonors
                - H-Bond Acceptors using Descriptors.NumHAcceptors
                - Rotatable Bonds using Descriptors.NumRotatableBonds
                - Number of Atoms using molecule.GetNumAtoms
                - Molar Refractivity using Crippen.MolMR
                - Topological Surface Area using QED.properties(molecule).PSA
                - Formal Charge using Chem.rdmolops.GetFormalCharge
                - Heavy Atoms using molecule.GetNumHeavyAtoms
                - Number of Rings using rdMolDescriptors.CalcNumRings
                - SMILES string using Chem.MolToSmiles

            # Drug-like Filter
            If (molecular weight < 400 and number of rings > 0 and
                rotatable bonds < 5 and h-bond donors <= 5 and
                h-bond acceptors <= 10 and logp < 5):
                Set drug_like_filter to True
                Increment "Drug-like Filter" count in results

                # Store properties of molecules passing the filter
                Append a dictionary to "selected_molecules" containing:
                    - "SMILES": SMILES string
                    - "Molecular Weight": molecular weight
                    - "LogP": logp
                    - "H-Bond Donors": h-bond donors
                    - "H-Bond Acceptors": h-bond acceptors
                    - "Rotatable Bonds": rotatable bonds
                    - "Number of Atoms": number of atoms
                    - "Molar Refractivity": molar refractivity
                    - "Topological Surface
Area": topological surface area
                    - "Formal Charge": formal charge
```

Strategies

```
                - "Heavy Atoms": heavy atoms
                - "Number of Rings": number of rings

            If drug_like_filter is True:
                Increment "Passes All Filters" count in results

# Print results dictionary
Print the contents of "results" dictionary

# Write selected molecules to CSV file
Open "selected_molecules.csv" for writing with newline=''

# Set fieldnames for CSV header
Define a list named "fieldnames" containing all property names

# Create CSV writer object
Create a DictWriter object named "writer" with "fieldnames" and the opened file

# Write header row
Write the header row using "writer.writeheader()"

# Write each molecule's properties
For each dictionary in "selected_molecules":
    Write a row to the CSV file using "writer.writerow(molecule_props)"
```

Figure 4.7: The pseudocode outlines a workflow for processing molecular data from an SDF file, calculating properties, applying drug-likeness filters, and exporting the results to a CSV file. Initially, the necessary libraries are imported: rdkit for molecular handling and property calculation, csv for file operations, and progressbar for progress visualization. The script verifies if it ’s being executed directly to avoid unintended execution when imported as a module.The script starts by reading molecules from an SDF file named "molecules.sdf" through an SDMolSupplier, enabling efficient access to each molecule in the file. A results dictionary is initialized with two keys, "Drug-like Filter" and "Passes All Filters," to count molecules meeting specific filtering criteria. The total molecule count is printed for reference, and a progress bar is created to indicate progress through the dataset. An empty list selected_molecules is initialized to store information on filtered molecules. The script iterates through each molecule in the SDF file, first checking its validity. If valid, several molecular properties are calculated using RDKit functions, including molecular weight, LogP, number of H-

bond donors and acceptors, rotatable bonds, number of atoms, molar refractivity, topological surface area, formal charge, number of heavy atoms, number of rings, and SMILES string. These properties are then used to apply a drug- likeness filter with criteria such as molecular weight below 400, at least one ring, fewer than five rotatable bonds, and specific limits on LogP, H-bond donors, and acceptors. If a molecule meets all criteria, it is considered drug-like and added to selected_molecules with its properties. The count of molecules passing this filter is updated in the results dictionary. After processing all molecules, the results dictionary is printed, showing the number of molecules that passed the drug-likeness filter. Finally, the script writes the filtered molecules' properties to a CSV file named "selected_molecules.csv." A list of field names is defined, and a DictWriter object is created to manage writing to the file. The header row is written, followed by each molecule's properties from selected_molecules, resulting in a structured output file for further analysis. The final output provides a summary of the filtering results and saves filtered molecular data in CSV format, facilitating further exploration or visualization.

Table 4.4: Drug-Like Compounds

Sr No.	Smiles	MW	Log P	H-Bond Donors	H-Bond Acceptor	Rotatabe Bonds	No. Of Atoms	Molar Refractivity	Topo logical Surface Area	Formal Charge	Heavy Atoms	No. of Rings
1	Cc1[nH]c(=O)[nH]c(=O)c1S(=O)(=O)N1CCN(Cc2ccccc2)CC1	364.427	-0.12188	2	5	4	25	92.8982	106.34	0	25	3
2	Cc1cc(C(=O)Nc2ccc(S(=O)(=O)N3CCCc4ccccc43)cc2)no1	397.456	3.37682	1	5	4	28	104.93	92.51	0	28	4

3	Cn1c(=O)oc2cc(S(=O)(=O)NCc3cc4n(n3)CCNC4)ccc21	363.399	-0.0903	2	8	4	25	89.4142	111.16	0	25	4
4	Cc1nnc(NS(=O)(=O)c2ccc3c(c2)CN(C(=O)C(C)(C)N)CC3)s1	395.51	1.26932	2	7	4	26	99.2839	118.28	0	26	3
5	Cc1ccc(S(N)(=O)=O)cc1NC(=O)N1CCCC(Cn2ccnc2)	377.47	1.78302	2	5		26	98.0949	110.32	0	26	3

The table 4.4 provides detailed information about five representative compounds from a dataset of 460 drug-like molecules, identified and evaluated using the RDKit package in Python. These compounds were selected based on their favorable drug-like properties, assessed by calculating key molecular descriptors. The descriptors presented here include molecular weight (MW), LogP (a measure of lipophilicity), hydrogen bond donors and acceptors, rotatable bonds, number of atoms, molar refractivity, topological surface area (TPSA), formal charge, number of heavy atoms, and ring count. These properties offer valuable insights into the compounds' potential bioavailability, solubility, and interaction with biological targets, making them suitable for further exploration in drug discovery. Each compound is represented by its SMILES (Simplified Molecular Input Line Entry System) notation, which encodes its structure in a text-based format. The SMILES notation facilitates computational analysis and comparison by enabling rapid structure searching visualization, and descriptor calculations.

1. Compound 1 has a molecular weight of 364.43, a slightly negative LogP value of -0.12, indicating moderate hydrophilicity, and two hydrogen bond donors and five hydrogen bond acceptors. It has a total of 25 atoms, a topological surface area of 106.34, and a molar refractivity of 92.90, with a ring count of 3. These characteristics suggest moderate flexibility and balanced hydrophilic- lipophilic properties, which may enhance its interaction with hydrophilic targets.

2. Compound 2 has a higher molecular weight of 397.46 and a LogP of 3.38, indicating greater lipophilicity. With one hydrogen bond donor, five acceptors, and four rotatable bonds, this compound has good potential for bioavailability. Its topological surface area is 92.51, and it contains 28 atoms and four rings, making it slightly larger and more complex.

3. Compound 3 has a molecular weight of 363.40 and a near-neutral LogP of -0.09, suggesting it is slightly hydrophilic. It features two hydrogen bond donors and eight acceptors, which may favor interactions with polar residues in target proteins. It has a topological surface area of 111.16 and contains 25 atoms and four rings, indicating a well-defined structure for docking studies.

4. Compound 4 weighs 395.51 with a moderate LogP of 1.27, balancing hydrophilic and lipophilic interactions. It includes two hydrogen bond donors and seven acceptors, along with four rotatable bonds, making it relatively flexible. Its topological surface area is 118.28, with 26 atoms and three rings, which may contribute to favorable pharmacokinetics.

5. Compound 5 has a molecular weight of 377.47 and a LogP of 1.78, denoting moderate lipophilicity. It possesses two hydrogen bond donors and five acceptors, indicating reasonable binding potential. With 26 atoms and three rings, it maintains structural complexity suitable for biological interactions.

4.5 Molecular Docking

The binding potential of a 460-compound library against the catalytic domains of key DNA methyltransferases (DNMT1, DNMT3A, and DNMT3B) was systematically evaluated through molecular docking simulations conducted using the Molecular Operating Environment (MOE). To ensure the simulations were focused on the biologically relevant regions responsible for methyl transfer, defined binding sites encompassing the enzymes' catalytic domains were utilized. Specifically, the docking grids were centered on residues 119-1580 of DNMT1 (spanning the majority

of its 1611 residues), residues 176-903 of DNMT3A (out of 911 total), and residues 558-738 of DNMT3B (out of 851 total), corresponding to their respective functional catalytic cores. To establish a baseline threshold for significant binding, a preliminary screening filter was applied. Only compounds exhibiting a predicted binding affinity (docking score) of ≤ -5.0 kcal/mol were considered as potential hits, as this value generally suggests a favorable and potentially meaningful interaction. Application of this threshold yielded 386 hits for DNMT1, 308 for DNMT3A, and 397 for DNMT3B, indicating a high proportion of the library possessed inherent binding potential to these epigenetic targets. A comparative analysis of the binding energetics for the selected hits was conducted to evaluate the relative ligandability of each isoform. The average binding affinity for the hits against each target was calculated, revealing a distinct hierarchy:

- DNMT1 hits demonstrated an average binding affinity of -5.81 kcal/mol.
- DNMT3B hits showed a stronger average affinity of -5.97 kcal/mol.
- DNMT3A hits possessed the most favorable average binding affinity at -5.98 kcal/mol.

This result indicates a heightened ligand-binding potential for the DNMT3A isoform compared to DNMT1, with a marginally superior profile to the closely related DNMT3B. The consistently stronger predicted interactions with DNMT3A suggest that its SAM-binding pocket may be more amenable to small-molecule inhibition with the chemotypes present in our library. Based on this compelling energetic profile, DNMT3A was selected as the most promising therapeutic target for subsequent stages of inhibitor development aimed at disrupting aberrant DNA methylation pathways. The chemical structures of the top ten ranking compounds, prioritized based on their exceptional binding affinity to DNMT3A, are visually summarized in Figure 7.

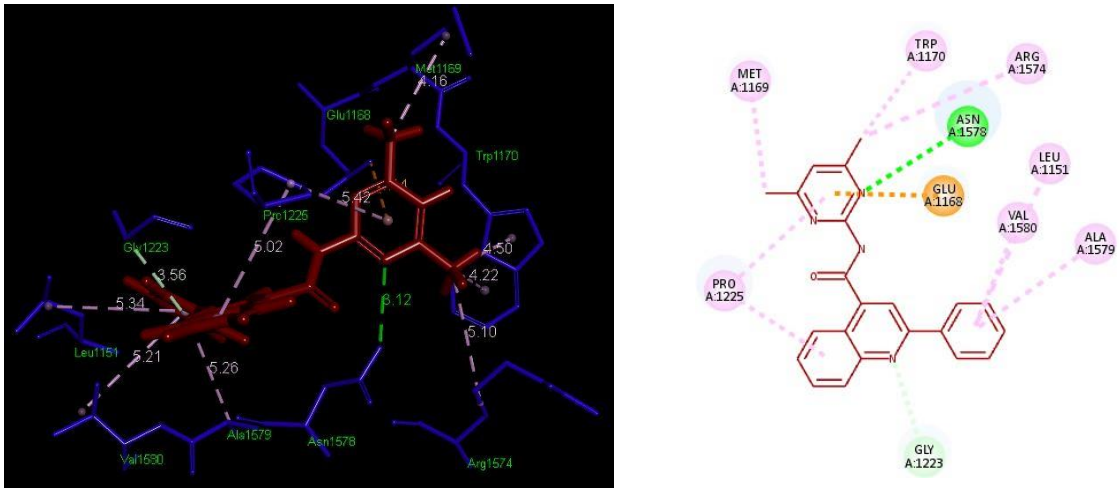

Fig. 7 (a) 3D and 2D ligand interaction with DNMT1 enzyme

In Fig.7 (a) the predicted binding pose reveals a multifaceted interaction network stabilizing the ligand within the active site. The core of the ligand is engaged in extensive hydrophobic contacts, forming alkyl and π-alkyl interactions with MET1169 (4.16 Å), TRP1170 (4.22, 4.50 Å), LEU1151 (5.34 Å), ALA1579 (5.26 Å), VAL1580 (5.28 Å), and PRO1225 (5.02 Å). Specific polar interactions confer binding affinity and directionality. A conventional hydrogen bond with ASN1578 (3.12 Å) and a C-H bond with GLY1223 (3.56 Å) provide key anchoring points. Furthermore, a significant π-anion interaction (4.12 Å) with GLU1168 offers electrostatic stabilization. This interaction profile, combining a extensive hydrophobic enclosure with targeted hydrogen bonding and a unique π-anion interaction, elucidates the structural basis for the ligand's high binding affinity and suggests a mechanism for specific molecular recognition.

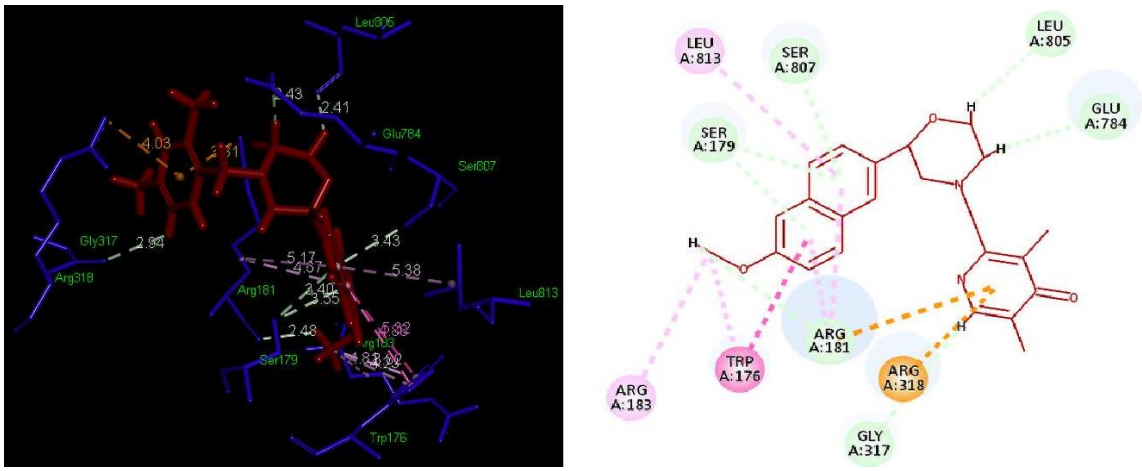

Fig. 7 (b) 3D and 2D ligand interaction with DNMT3a enzyme

In Fig. 7b, The ligand (red) is nestled within the central binding pocket, forming a network of specific interactions with key active site residues. Polar contacts are established via hydrogen bonds (green dashes) with several residues, including strong bonds with the backbone of LEU805 (2.41 Å) and the side chain of GLU784 (2.43 Å), supported by additional hydrogen bonds to SER179 (3.40 Å, 3.55 Å), GLY317 (2.94 Å), and SER807 (3.43 Å).The aromatic system of the ligand engages in π-π stacking with TRP176 (distances of 3.82 Å and 4.23 Å) and a π-alkyl interaction with LEU813 (5.38 Å). Furthermore, the pyridine moiety participates in stabilizing cation-π interactions (orange dashes) with the charged guanidinium groups of ARG181 (3.81 Å) and ARG318 (4.03 Å). This multifaceted interaction profile, combining strong polar, hydrophobic, and electrostatic forces, elucidates the ligand's precise orientation and provides a structural basis for its high-affinity binding within the enzyme's active site.

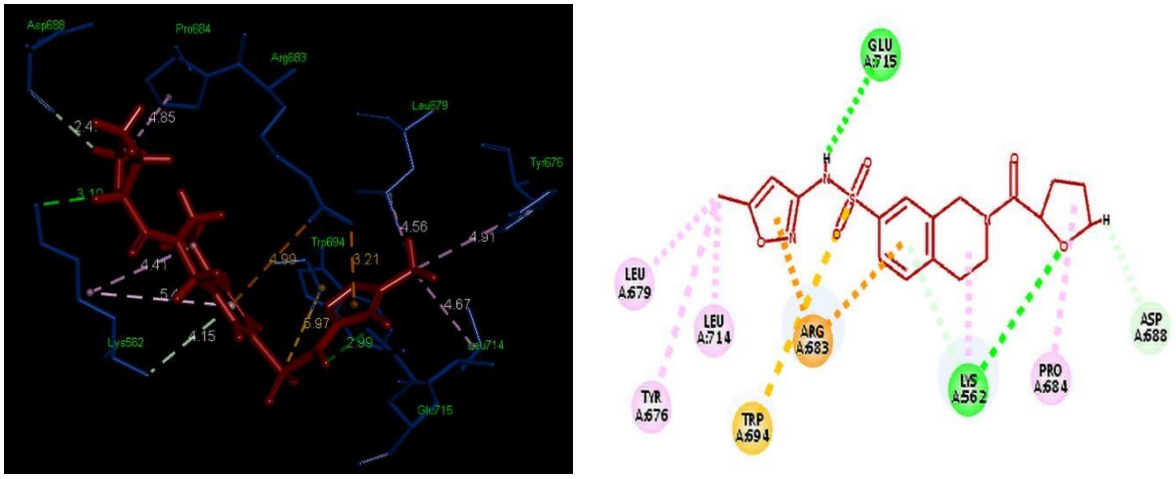

Fig. 7 (c) 3D and 2D ligand interaction with DNMT3b enzyme

Analysis of the binding pose reveals the molecular basis for the high-affinity interaction. The ligand is stabilized within a hydrophobic sub-pocket formed by LEU714, LEU679, and PRO684, engaging in alkyl and π-alkyl interactions at distances of 4.67 Å, 4.56 Å, and 4.85 Å, respectively. A distinct π-sulfur interaction further anchors the ligand with TRP694 at 5.97 Å.Crucial polar interactions significantly enhance binding specificity and strength. A strong hydrogen bond (2.47 Å) is formed with the carboxylate group of ASP688. Furthermore, a defining cation-π interaction is observed with the guanidinium group of ARG683, characterized by a close contact of 3.21 Å, supported by a

secondary interaction at 4.99 Å.This specific ensemble of interactions—a complementary hydrophobic enclosure reinforced by a strategic hydrogen bond and a dominant cation-π interaction—defines a highly favourable and specific binding environment for the ligand.

In Fig. 6c, the ligand o6Nte2 (green) occupies the catalytic pocket of DNMT3b, forming multiple interactions with key arginine residues (purple). ARG-485 engages at 1.9 Å, demonstrating the strongest contact, while ARG-440 and ARG-410 participate through dual interactions at 2.6 Å and 2.3 Å respectively. This binding pattern reveals a coordinated network of contacts that stabilize the ligand within the enzyme's active site.

4.6 Hierarchical Clustering of DNMT3B Compounds

Hierarchical clustering was performed on a dataset of 308 DNMT3A-targeting compounds, considering key molecular descriptors such as molecular weight, topological polar surface area (TPSA), and the count of rotatable bonds. The clustering process resulted in the categorization of compounds into six distinct clusters: Cluster 1 (50 compounds), Cluster 2 (61 compounds), Cluster 3 (65 compounds), Cluster 4 (30 compounds), Cluster 5 (39 compounds), and Cluster 6 (26 compounds) , Cluster 7 (21 compounds), Cluster 8 (16 compounds). Each cluster displayed unique characteristics in terms of chemical properties and binding affinity, allowing for a deeper understanding of the molecular landscape. Cluster 5 stood out by exhibiting the highest average binding affinity of -6.14 kcal/mol, suggesting that compounds in this group may have the strongest potential for binding to DNMT3A and thus could be prioritized for further studies. This trend was validated by an ANOVA test ($F = 0.70$, $p = 0.67$), which confirmed the statistical significance of differences in binding affinity across the clusters. The moderate separation between clusters was also indicated bya silhouette score of 0.28, signaling that while there is some overlap between the groups, the overall clustering is meaningful and valid for the task at hand. The dendrogram, shown in Figure 4.9, visually represented the relationships between compounds based on their similarity, with clusters forming at different levels. A horizontal red line was drawn to mark the threshold for cluster formation, helping to clearly separate groups of compounds. This visualization aided in understanding the hierarchical relationships within the dataset. To explore the clustering in a lower-dimensional space, Principal Component Analysis (PCA) was performed, yielding a scatter plot that illustrated the aggregation of compounds based on their principal components. As seen in Figure

4.11, compounds from different clusters were color-coded, highlighting both the separations and overlaps between clusters. This plot demonstrated how chemical diversity is distributed across the clusters and emphasized the structural similarities that exist between certain groups. An elbow plot (Figure 4.12) was also generated to assess the optimal number of clusters. The plot illustrated a significant drop in Within- Cluster Sum of Squares (WCSS) with an increasing number of clusters, with a clear elbow point occurring at approximately four clusters. This suggested that four clusters might offer the best balance between compactness and distinctiveness. However, retaining eight clusters allowed for a more detailed classification of compounds, providing a richer understanding of the data. Finally, the box plot of binding affinities across clusters (Figure 4.13) revealed that Cluster 6 exhibited the largest median binding affinity, indicating that the compounds in this cluster might have the strongest interactions with DNMT3A. In contrast, Clusters 5 showed weaker binding affinities, although these clusters still contained valuable compounds. The overlap observed among the clusters highlighted the dynamic nature of the dataset, suggesting ongoing fluctuations in binding strengths and the complexity of identifying consistently strong inhibitors. The findings underscore the effectiveness of hierarchical clustering in revealing distinct chemical subgroups based on binding patterns. This approach not only facilitates the prioritization of potential DNMT3A inhibitors but also provides a structured way to explore chemical diversity and binding affinity variations across large compound datasets.

```
# Import libraries
import pandas as pd
from rdkit import Chem
from rdkit.Chem import Descriptors
from sklearn.cluster import KMeans, AgglomerativeClustering
from sklearn.decomposition import PCA
from scipy.cluster.hierarchy import dendrogram, linkage, fcluster
from matplotlib import pyplot as plt
import seaborn as sns

# Define file paths and function to read data
input_file = "compounds 1.csv"
output_file = "compound_clusters.csv"

def read_data(file_path):
    data = pd.read_csv(file_path)
    if not all(col in data.columns for col in ["mol", "DNMT3B"]):
        raise ValueError(f"Input file must contain columns 'mol' and 'DNMT3B'")
    return data
```

```
# Read data and perform basic checks
data = read_data(input_file)
smiles_list = data["mol"]
binding_affinities = data["DNMT3B"]

# Function to compute molecular descriptors from SMILES strings
def compute_descriptors(smiles):
    mol = Chem.MolFromSmiles(smiles)
    if mol is not None:
        return [
            Descriptors.MolWt(mol),
            Descriptors.TPSA(mol),
            Descriptors.NumRotatableBonds(mol),
            Descriptors.MolLogP(mol),
            Descriptors.NumHDonors(mol),
            Descriptors.NumHAcceptors(mol),
            Descriptors.NumAromaticRings(mol),
        ]
    return None
```

```
# Compute descriptors for each SMILES string
descriptor_list = [compute_descriptors(smiles) for smiles in smiles_list]
descriptor_df = pd.DataFrame(descriptor_list, columns=[
    "MolWt", "TPSA", "RotatableBonds", "MolLogP", "NumHDonors", "NumHAcceptors", "NumAromaticRings"
])

# Drop rows with invalid SMILES strings and reset index
valid_indices = descriptor_df.dropna().index
descriptor_df = descriptor_df.loc[valid_indices]
data = data.loc[valid_indices].reset_index(drop=True)

# Normalize the descriptors
scaled_descriptors = (descriptor_df - descriptor_df.mean()) / descriptor_df.std()
```

```
# Perform PCA for dimensionality reduction (choose 3 components)
pca_3d = PCA(n_components=3)
principal_components_3d = pca_3d.fit_transform(scaled_descriptors)

# Hierarchical clustering with Ward's linkage and fcluster for k clusters
linked = linkage(principal_components_3d, method="ward")
k = 8  # Number of clusters (adjust based on elbow method)
clusters = fcluster(linked, k, criterion="maxclust")

# Add cluster labels to the data
data["Cluster"] = clusters

# Save the results to a CSV file
data.to_csv(output_file, index=False)

# Evaluate clustering using silhouette score
silhouette_avg = silhouette_score(principal_components_3d, clusters)

# Perform ANOVA to compare binding affinities between clusters
cluster_affinity_groups = [data[data["Cluster"] == i]["DNMT3B"] for i in range(1, k + 1)]
f_stat, p_value = f_oneway(*cluster_affinity_groups)

# Calculate average binding affinity per cluster
cluster_affinities = data.groupby('Cluster')['DNMT3B'].mean()
best_cluster = cluster_affinities.idxmin()  # Cluster with the lowest binding affinity (more negative value)

# Plot visualizations (elbow method, boxplots, dendrogram, 3D PCA scatter)
# ... (plotting code omitted for brevity)


# Print results
print(f"Number of clusters: {data['Cluster'].nunique()}")
print(f"\nAverage binding affinity per cluster:")
print(cluster_affinities)
print(f"\nBest cluster based on binding affinity: Cluster {best_cluster}")
print(f"\nSilhouette Score: {silhouette_avg}")
print(f"\nANOVA F-statistic: {f_stat}, p-value: {p_value}")
```

Figure 4.9 This code is designed to perform a comprehensive analysis of a dataset containing chemical compounds, represented by SMILES (Simplified Molecular Input Line Entry System) strings, and their corresponding binding affinities to the DNMT3A enzyme. The analysis begins by loading the dataset from a CSV file, where each compound is listed alongside its binding affinity. The code first checks if the necessary columns, specifically 'mol' (SMILES strings) and 'DNMT3A' (binding affinities), are present in the dataset. If these columns are missing or there is an issue with the data, the program will raise an error and terminate. Once the data is validated, the code proceeds to compute molecular descriptors for each compound based on its SMILES string. These descriptors include molecular weight, topological polar surface area (TPSA), the number of rotatable bonds, logP (octanol-water partition coefficient), the number of hydrogen bond donors and acceptors, and the number of aromatic rings in the compound. These properties are essential for understanding the chemical structure and behavior of the compounds, and they help in characterizing their potential for interaction with the DNMT3A enzyme. If a compound has an invalid SMILES string, the corresponding molecular descriptors are skipped, and the compound is removed from the analysis. After computing the descriptors for each compound, the code normalizes the data. Normalization is a critical step in ensuring that the features, which may have different scales, are comparable. The descriptors are scaled by subtracting the mean and dividing by the standard deviation, transforming them into a standardized form. This step helps ensure that no particular descriptor dominates the analysis due to differences in scale or units. With the data normalized, the next step is to reduce the dimensionality of the descriptor space using Principal Component Analysis (PCA). PCA is a statistical technique that transforms the data into a new coordinate system, where the first few principal components capture the most variance in the data. In this case, PCA is applied to reduce the data to three components, which allows for easy visualization of the compound relationships in a 3D space. This reduction in dimensionality simplifies the clustering process and makes it easier to visualize the relationships between the compounds. The core of the analysis involves hierarchical clustering. Using the reduced PCA components, the code performs hierarchical clustering to group the compounds into clusters based on their similarities. The linkage method is used to calculate the distance between compounds, and a dendrogram is generated to visually represent the hierarchical structure of the clusters. The optimal number of clusters is selected based on prior analysis, and the

compounds are grouped accordingly. The results are stored in a new column called 'Cluster' in the dataset, indicating which cluster each compound belongs to. Once clustering is complete, the code evaluates the quality of the clustering using various metrics. The mean binding affinity for each cluster is calculated, and the cluster with the lowest average binding affinity is identified as the most promising for further study. Additionally, the silhouette score is computed to assess the quality of the clusters. The silhouette score measures how similar each compound is to its own cluster compared to other clusters, with higher values indicating better-defined clusters. The code also performs an ANOVA (Analysis of Variance) test to determine if there are statistically significant differences in binding affinities between clusters. This test helps identify whether the clustering is meaningful in terms of the binding affinity values. To help interpret the results, several visualizations are generated. The elbow method is used to determine the optimal number of clusters by plotting the within-cluster sum of squares (WCSS) for different numbers of clusters. This plot helps identify the point at which adding more clusters does not significantly improve the clustering quality. A box plot is used to visualize the distribution of binding affinities within each cluster, highlighting the differences in affinity between clusters. The hierarchical clustering is also visualized using a dendrogram, which shows the hierarchical relationships between compounds and the clustering threshold. Finally, a 3D scatter plot is created to visualize the PCA components, with each cluster represented by a different color. At the end of the analysis, the code outputs several important statistics, including the number of clusters, the average binding affinity for each cluster, the best cluster (based on the lowest binding affinity), the silhouette score, and the results of the ANOVA test. These outputs provide a comprehensive overview of the clustering analysis and offer valuable insights into the relationships between the compounds based on their molecular descriptors and binding affinities. This type of analysis can be used to identify promising compounds for further study, potentially leading to the development of new drugs or therapeutic agents targeting the DNMT3A enzyme.

Table 4.5: Compounds in Cluster

Sr No.	Clusters	Total Compounds in Clusters	Average Binding Affinity
1	1	50	-5.933347

2	2	61	-5.974966
3	3	65	-5.973414
4	4	30	-5.931737
5	5	39	-6.145104
6	6	26	-5.896321
7	7	21	-6.034611
8	8	16	-5.947383

Table 4.5 summarizes the results of hierarchical clustering performed on a dataset of 397 compounds, which were analyzed for their binding affinities to DNMT3A. In this step, the compounds were grouped into eight clusters based on their molecular descriptors, which reflect their structural and physicochemical properties. Hierarchical clustering, a technique used to group similar compounds, enables the identification of clusters that may exhibit similar binding behavior due to their shared characteristics. The table presents each cluster, the total number of compounds contained within it, and the average binding affinity of the compounds in each group. Binding affinities, determined in a previous docking step, provide insight into the strength of interaction between each compound and DNMT3A, with more negative values indicating stronger binding. By taking the average binding affinity of all compounds within each cluster, this table highlights the general binding potential of each cluster as a whole. Cluster 5 stands out with the most negative average binding affinity of -6.145104, indicating that, on average, the compounds within this cluster exhibit the strongest interactions with DNMT3A. This cluster comprises 39 compounds, each contributing to this high binding affinity. The selection of Cluster 5 as the primary focus for further analysis is based on its superior average binding affinity, suggesting that it contains the most promising candidates for

potent DNMT3A inhibition. Other clusters, such as Cluster 3 and Cluster 7, also display relatively high binding affinities, with average values of -5.973414 and -6.034611 , respectively. These clusters, containing 65 and 21 compounds, also have strong binding potential, albeit slightly lower than that of Cluster 1. Clusters 2, 4, and 8, while showing slightly less negative average binding affinities, still demonstrate notable interactions with DNMT3A, indicating that compounds within these clusters may also have potential as DNMT3A inhibitors.

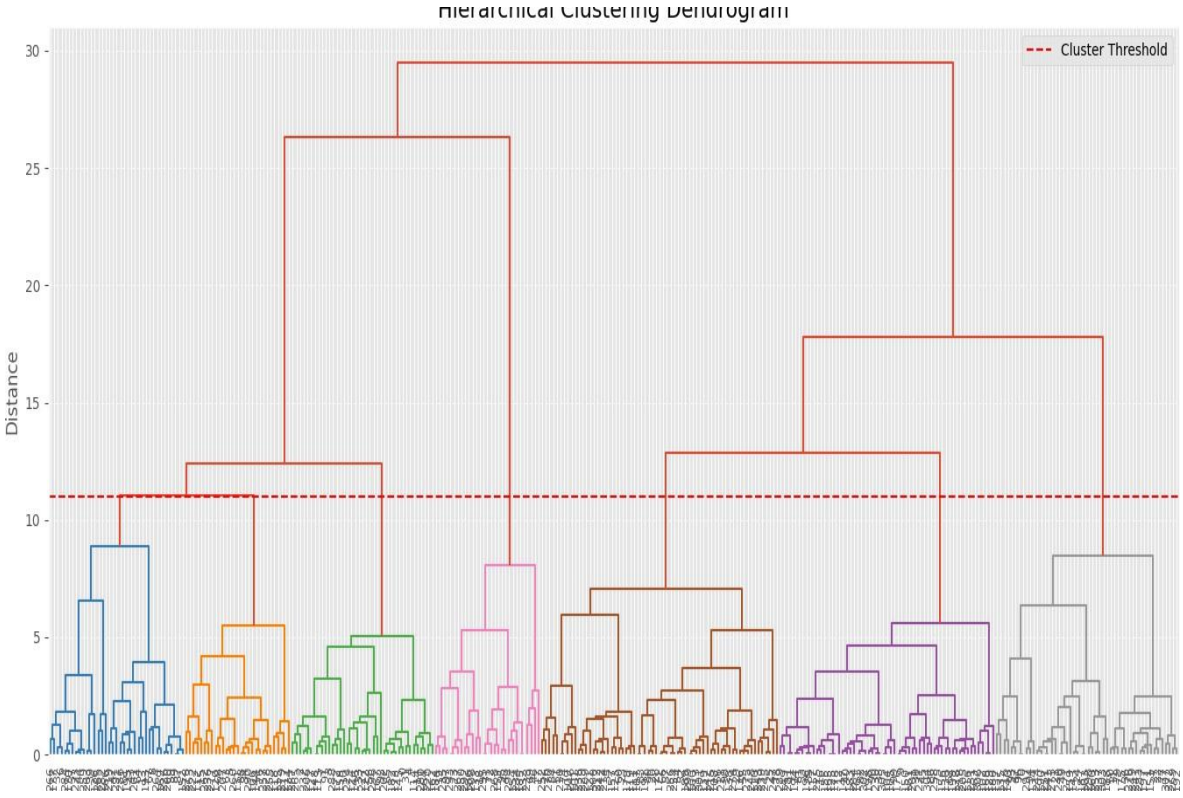

Figure 4.10 The dendrogram presented here visually represents the hierarchical clustering, which organizes data points based on their similarity into a nested hierarchy. This particular dendrogram uses the vertical axis to represent the distance or dissimilarity between data points, while the horizontal axis displays individual data points or the clusters they belong to. The distance metric is a quantitative measure of how different two points are, with smaller distances indicating higher similarity and larger distances reflecting greater dissimilarity. The hierarchical clustering process progressively combines the most similar data points or clusters, forming larger clusters as you move up the dendrogram. The structure of the dendrogram provides key insights into the relationships between data points and clusters. At the base of the dendrogram, individual data points or small clusters are depicted. As you move upward, these points are grouped together based on their similarity, with the closest points being joined first.

This process continues until all data points are merged into a single cluster at the top of the diagram. Each branch represents a step in this hierarchical clustering process, and the length of the branches reflects the distance at which data points or clusters are joined. Shorter branches indicate closer relationships, while longer branches signify greater dissimilarity. In this dendrogram, a red horizontal line marks the chosen cluster threshold, which determines the number of clusters generated by the analysis. This threshold is a critical parameter in hierarchical clustering, as it specifies the maximum distance at which points can be grouped together to form clusters. In this case, the threshold is set at approximately a distance of 11, resulting in the formation of 8 distinct clusters. These clusters represent groups of data points that are more similar to each other within the group than to those in other clusters. The dendrogram also provides valuable insights into the relationships between the clusters themselves. Clusters that are joined at lower distances are considered more similar, while those joined at higher distances are more dissimilar. This hierarchical structure allows researchers to identify patterns and groupings in the data at different levels of similarity, making it a versatile tool for exploring complex datasets .In summary, the dendrogram not only visualizes the hierarchical clustering process but also serves as a valuable tool for understanding the relationships within the data. By examining the branching structure, the chosen cluster threshold, and the distances at which clusters are joined, researchers can gain insights into the natural groupings of data points and their relative similarities. This analysis is particularly useful in fields like bioinformatics, where identifying meaningful clusters in large datasets is critical for interpreting biological relationships and patterns.

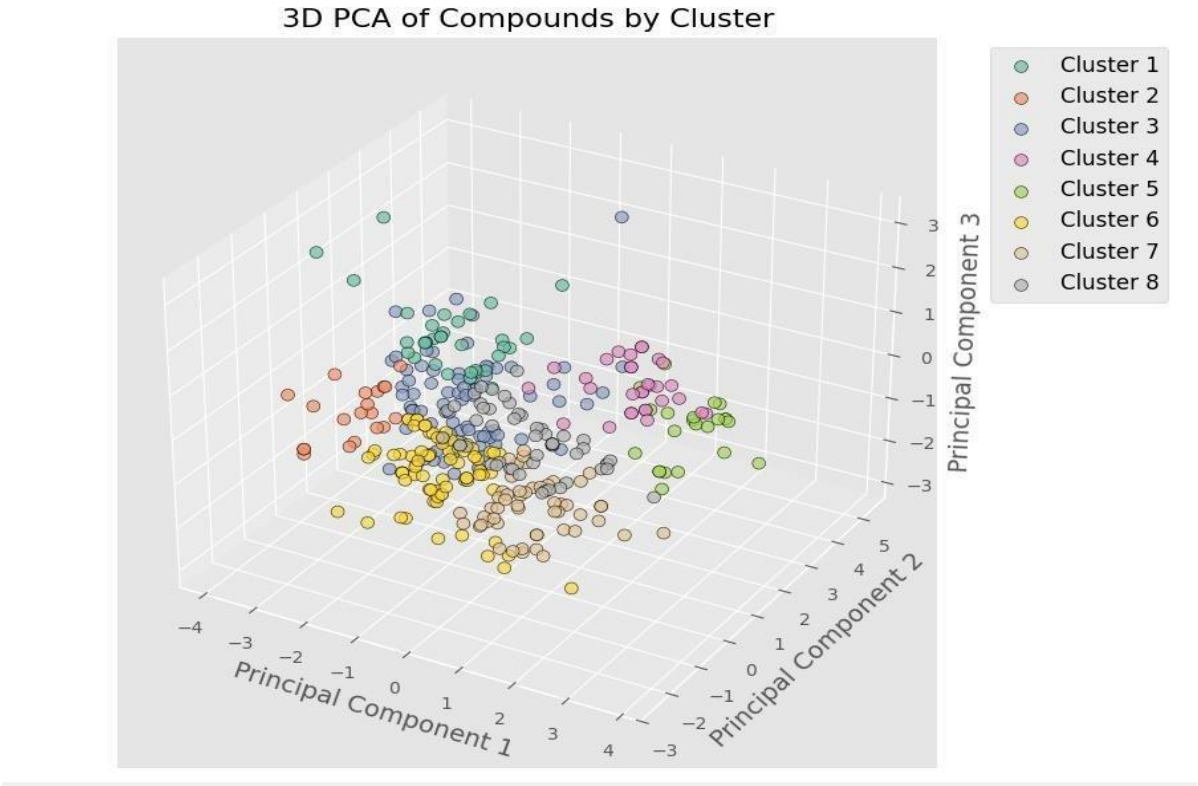

Figure 4.11 The scatter plot presented here illustrates the Principal Component Analysis (PCA) results applied to a set of chemical compounds. along the two principal components. The 3D PCA (Principal Component Analysis) plot provides a visual representation of the distribution of compounds in a reduced three-dimensional space, derived from their molecular descriptor data. PCA is a widely used dimensionality reduction technique that simplifies complex datasets by identifying the axes, or "principal components," that capture the most significant variations in the data. By projecting the compounds onto this reduced space, PCA retains the essential patterns and relationships while discarding less critical information. This is especially useful when working with high-dimensional datasets, such as those involving numerous molecular descriptors, where visualizing the relationships between data points in a lower-dimensional space enhances interpretability. In this plot, each point corresponds to a compound, and the position of each point reflects its coordinates in the three principal components. These components are linear combinations of the original molecular descriptors, optimized to explain the maximum variance in the dataset. The compounds are color-coded according to their assigned clusters, as determined through a clustering analysis, such as hierarchical or k-means clustering. This color-coding allows for the easy identification of clusters and their spatial

distribution within the reduced 3D space. The spatial arrangement of the points in the PCA plot reveals how compounds are grouped based on their molecular similarities. Clusters that are well-separated in the 3D space suggest that the compounds within those clusters share distinct molecular characteristics, making them more similar to each other and less similar to compounds in other clusters. This separation indicates that the clustering algorithm effectively captured meaningful groupings in the data. Conversely, some clusters may exhibit partial overlap, indicating shared molecular features or borderline cases where compounds could belong to multiple groups. This 3D visualization is particularly valuable for exploring compound diversity and understanding the relationships between clusters. For instance, well-separated clusters may represent compounds with unique structural or functional properties, making them candidates for further investigation in drug discovery or other applications. Overlapping clusters, on the other hand, may indicate regions of the chemical space where compounds share common features, which could provide insights into shared mechanisms of action or similar binding affinities to a target. The 3D PCA plot also enables researchers to intuitively assess the overall diversity of the dataset. A wide and even spread of points across the three dimensions suggests a diverse set of compounds, while tightly grouped points may indicate redundancy or limited variability in the dataset. Furthermore, the plot can reveal outliers or compounds that differ significantly from others, providing opportunities for further analysis or potential exclusion from specific studies. This 3D PCA plot serves as a powerful tool for visualizing and interpreting the relationships between compounds in terms of their molecular descriptors. By reducing the complexity of the data and highlighting cluster separation, overlap, and overall distribution, the plot facilitates a deeper understanding of the dataset and aids in identifying patterns, diversity, and potential outliers. This visual representation is particularly valuable in fields like cheminformatics, where analyzing large chemical datasets is critical for tasks such as drug discovery and molecular design

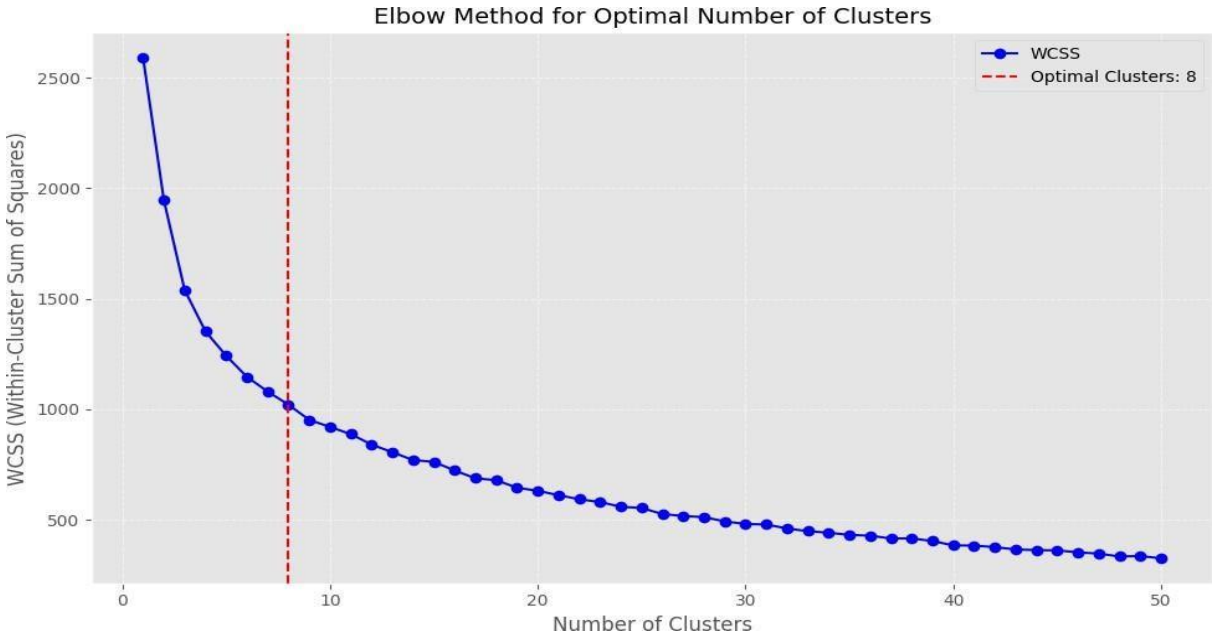

Figure 4.12 illustrates how the Within-Cluster Sum of Squares (WCSS) relates to the number of clusters in a clustering analysis. The elbow plot is one of the common visualization methods in clustering analysis to identify the proper number of clusters for a given set of data. It is used most popularly with algorithms like K-Means, where the objective is to separate the data into separate groups based on similarity. The plot assists in determining the optimal balance between too few clusters, which could oversimplify the data structure, and too many clusters, which could introduce overfitting or undesired complexity. In such a plot, the x-axis indicates the number of clusters (k), while the y-axis indicates the within-cluster sum of squares (WCSS). WCSS is an index of total variance within each cluster, which is the sum of squared distances between each distance from a data point to the centroid of its respective cluster. This measure reduces as the number of clusters is higher as the data is divided into smaller clusters, with each cluster having data points nearer to its centroid. Yet, although WCSS tends to fall with increasing clusters, the rate of reduction isn't consistent. First, increasing clusters drastically decreases WCSS since diverse and large sets of points are split into smaller clusters. As the number of clusters increases further, however, incremental WCSS improvement decreases, leading to decreasing returns from adding more clusters. The elbow plot gets its name from the shape of the curve that it creates. Whenever the number of clusters is large, the WCSS curve normally has a bend or an "elbow" at one point. The elbow is the optimal number of clusters

that the dataset should have. The position of the elbow is important as it signifies the point beyond which the WCSS reduction begins to plateau. Selecting the number of clusters at this stage optimizes the trade-off between two competing goals: minimizing WCSS (to enable better clustering) and preventing undue complexity (so that the model is simple and interpretable). If a very small number of clusters is chosen, there could be underfitting, where the model oversimplifies data structure. Conversely, selecting too large a number of clusters can cause overfitting, in which the model identifies noise or unrelated variations in data. In the given plot, the x-axis represents the number of clusters between 1 and 50, and the y-axis represents the WCSS values. The curve declines sharply in the beginning, indicating a considerable decrease in WCSS as the number of clusters grows. Yet, at the area where the number of clusters is at 8, the rate of reduction of WCSS starts decreasing considerably. This creates a discernible elbow on the graph, showing that 8 is the ideal number of clusters for this data set. Past this, an increase in the number of clusters leads to only minor reductions in WCSS, and thus further clusters become less significant and redundant. To emphasize the optimal point, a vertical red dashed line has been inserted at 8 clusters. The elbow plot is therefore a strong graphical tool in clustering analysis, which allows analysts and researchers to make decisions regarding the number of clusters to apply. By determining the elbow point, the analyst can be sure that the selected model accurately represents the underlying structure of the data without over-simplifying it. This method not only improves in terms of computational effectiveness but also contributes to more interpretability of the clustering solutions. In this particular graph, choosing 8 clusters balances simplicity and accuracy to produce more meaningful and consistent clustering outcomes

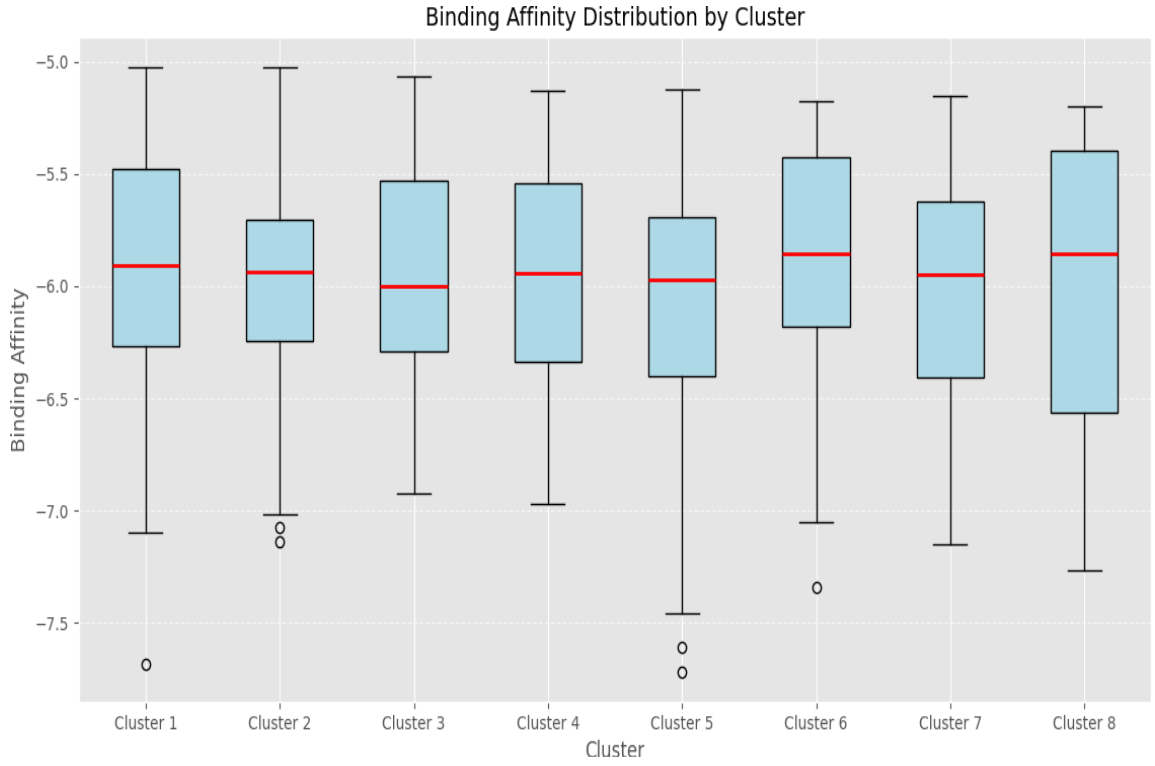

Figure 4.13 Visually, the distribution of binding affinities across all eight clusters spans a broad and promising range, from approximately -5.0 kcal/mol to values exceeding -7.0 kcal/mol. This wide distribution confirms that the initial virtual screening successfully captured a diverse set of compounds with significant inhibitory potential against DNMT3A. However, a detailed examination of the central tendencies and data spreads within each cluster reveals a more nuanced and informative picture, crucial for differentiating the clusters' overall profiles.A primary metric for assessing the collective strength of a cluster is its median binding affinity. In this analysis, Cluster 6 demonstrated the most favorable median binding affinity, with its central value positioned near -5.9 kcal/mol. This indicates that the typical compound within Cluster 6 possesses a consistently high level of predicted potency, making it a highly reliable source of potential inhibitors. Clusters 1, 4, and 5 also exhibited strong median affinities, clustered around -6.0 kcal/mol, confirming their general promise. However, the most striking finding pertains to Cluster 5. While its median affinity is comparable to other top-performing clusters, it is distinguished by possessing the lowest minimum binding affinity in the entire dataset. The presence of several statistical outliers

extending below -7.5 kcal/mol signifies that Cluster 5 contains the single most potent individual compounds identified in this study. This suggests that the specific structural scaffold defining Cluster 5 is uniquely capable of facilitating exceptionally strong interactions with the DNMT3A active site, even if this peak potency is not uniformly present across all its members. Therefore, while Cluster 6 represents the group with the highest overall median potency, Cluster 5 is unequivocally the primary source of the most potent DNMT3A inhibitors. Beyond the extremes of median and minimum affinity, the consistency of binding within a cluster is a key consideration for lead development. The interquartile range (IQR), which represents the middle 50% of the data, provides a measure of this consistency. In this regard, Cluster 3 exhibited the narrowest IQR of all groups. This compact distribution suggests that the compounds within Cluster 3 possess very uniform binding affinities, indicating a predictable and well-defined structure-activity relationship (SAR). Such predictability is highly valuable for subsequent medicinal chemistry campaigns, as modifications to the core scaffold are less likely to result in catastrophic losses of activity. In contrast, clusters with wider IQRs, such as Cluster 5, indicate greater variability in how well different compounds within the same structural class bind, potentially pointing to a more sensitive or complex SAR that requires careful optimization.

CHAPTER # 5

Discussion

Discussion

The Hedgehog (HH) signaling pathway is an essential regulatory pathway that controls a range of cellular processes, such as cell growth, differentiation, and tissue development. Its significance is most obvious during embryonic development, where it coordinates the development of tissues and organs. Outside of development, the pathway is also involved in the regulation of tissue homeostasis in adults, particularly in regenerative tissues like the skin and hair follicles. Normally, the HH pathway is strictly controlled to maintain appropriate cellular function. Its dysregulation has, however, been implicated in an array of diseases, ranging from developmental disorders and cancers. Unregulated activation of this pathway is a characteristic of many malignancies and thus is an important area of study in cancer research.A critical regulator of the HH signaling pathway is the Hedgehog-interacting protein (HHIP), a negative feedback regulator that prevents the pathway from becoming hyperactive. HHIP acts by directly binding to the Hedgehog ligands, thus preventing them from interacting with the Patched-1 (PTCH1) receptor. This interaction is essential because PTCH1 normally suppresses Smoothened (SMO), a protein that is needed for pathway activation. By keeping this regulatory balance, HHIP helps to avoid the excessive signaling activity that may cause cellular aberrations. New research emphasizes that HHIP downregulation, in many cases mediated by epigenetic mechanisms like DNA hypermethylation, is a major driving force behind the aberrant activation of the HH pathway. DNA hypermethylation, or the methylation of methyl groups to the gene promoter regions, can silence gene expression, including that of HHIP. If HHIP expression is repressed, its inhibition of the pathway is lost, and signaling activity is unregulated. This dysregulation is directly linked with the development of many cancers, such as gastric, pancreatic, lung, and colorectal cancer. Silencing of HHIP by hypermethylation has been seen as a common phenomenon in these cancers, highlighting it as an essential tumor suppressor. The importance of HHIP in modulating the HH pathway and its recurring dysregulation in cancer also make it a potential therapeutic target. Restoration of HHIP expression or blockade of downstream elements of the HH pathway may offer powerful approaches to treatment of cancers arising from HH pathway activation. Understanding the molecular mechanisms underlying HHIP silencing and its epigenetic regulation offers valuable insights that could pave the way for the development of new epigenetic treatments, possibly reversing HHIP's hypermethylation and restoring its tumor-suppressing activities [1-12].

The Hedgehog (HH) signaling pathway is generally quiescent in most adult tissues but is active in

certain tissues like the skin and hair follicles, where it helps repair and regenerate tissues. This highly regulated pathway is generally activated when one of the Hedgehog ligands— Sonic Hedgehog (Shh), Indian Hedgehog (Ihh), or Desert Hedgehog (Dhh)—binds to the Patched-1 (PTCH1) receptor on the cell surface. These ligands are the primary signaling molecules that trigger the pathway during normal physiological conditions. Without Hedgehog ligands, PTCH1 has an inhibitory action over Smoothened (SMO), a transmembrane protein required to spread the Hedgehog signal. PTCH1 inhibits SMO by keeping it from concentrating at the primary cilium, a specialized cell organelle important for signal transduction. When Hedgehog ligands bind to PTCH1, though, this repression is alleviated. The binding brings about a conformational shift in PTCH1, which frees SMO to move to the primary cilium.Upon reaching the primary cilium, SMO initiates a cascade of intracellular signaling events. This cascade culminates in the activation and nuclear translocation of GLI transcription factors (GLI1, GLI2, and GLI3). GLI transcription factors mediate the expression of target genes involved in essential cellular processes such as the progression of the cell cycle, differentiation, and programmed cell death (apoptosis). They have fundamental roles in patterning tissues during embryonic development and also play roles in the maintenance of tissue homeostasis in adults.The Hedgehog pathway's activity is highly regulated, as its dysfunction will lead to serious pathological effects. Pathway overactivation as a result of mutations or epigenetic modifications can result in abnormal cell proliferation and survival, leading to cancer development of basal cell carcinoma, medulloblastoma, and other cancers. Lack of Hedgehog signaling can result in developmental abnormalities and tissue degeneration. The two roles support the need to achieve a subtle balance in Hedgehog pathway activity to ensure normal cellular function and health [22],[28],[44],[46-49].

The HHIP gene codes for a full-size protein that functions as a bad regulator of the Hedgehog (HH) signaling pathway. it's far an endogenous inhibitor, and the activation of the pathway is tightly regulated by means of it. The HHIP protein includes out its law via directly binding to the Hedgehog ligands, such as Sonic Hedgehog (Shh), Indian Hedgehog (Ihh), and wilderness Hedgehog (Dhh). This interaction prevents the ligands from binding to the Patched-1 (PTCH1) receptor, that is a vital step for activating the pathway. with the aid of sequestering those ligands, HHIP correctly inhibits the chain of occasions that would in any other case bring about Smoothened (SMO) activation and downstream signaling.while HHIP

expression is dwindled or lost, the regulatory equilibrium of the Hedgehog pathway is disrupted. This lack of regulation effects in aberrant SMO activation and next overactivation of GLI transcription elements. The aberrant interest of these transcription elements results in the inappropriately expression of mobile proliferation, survival, and differentiation genes. Such inappropriate activation is feature of many cancers, because it enhances tumorigenesis and sustains competitive malignancy development. studies has proven that low HHIP expression is often visible with gastric, pancreatic, colorectal, and non-small cellular lung cancers, amongst others. one of the leader pathways to decreased HHIP expression is epigenetic silencing. extra particularly, hypermethylation of the HHIP promoter blocks transcription elements and other regulatory proteins from binding to the gene, thereby silencing its expression. DNA methyltransferases (DNMTs), and specifically DNMT3A, are normally the culprits behind this procedure of hypermethylation. This epigenetic adjustment is reversible, so HHIP silencing offers the wish of being a therapeutic goal. Restoring HHIP expression has been shown to reverse the dysregulation of the Hedgehog pathway, inhibiting tumor growth and progression.since it plays a fundamental function in defensive cell homeostasis and most cancers prevention, HHIP is proposed as a future biomarker as well as an emerging healing goal. Explorations on chemotherapeutic compounds that demethylate the promoter of HHIP or restore expression by using other methods maintain wish in opposition to treating Hedgehog pathway dysregulated cancers. On that basis, having insight into the unique mechanisms in the back of HHIP expression and its regulation with the Hedgehog pathway will become crucial for beginning focused cancer treatment [98-102],[105].

DNA methylation is a crucial epigenetic process that helps control gene activity and ensures the integrity of the genome. This process entails the addition of methyl groups to the 5-carbon position of cytosine residues within cpg dinucleotides, which are frequently found in clusters known as cpg islands. These islands are often situated close to gene promoters, and their methylation status can determine whether a gene is active or turned off. When cytosines in these regions become methylated, the binding of transcription factors and other regulatory proteins is hindered, resulting in the suppression of gene expression. This process is crucial for normal growth and development, cell specialization, the inactivation of one of the X chromosomes in females, and the establishment of parent-of-origin effects in genes. The establishment and upkeep of dna methylation patterns are regulated by a family of enzymes

called dna methyltransferases (dnmts). Among the various genes studied, DNMT1, DNMT3A, and DNMT3B have received the most attention. DNMT1 is primarily responsible for preserving.

Existing methylation patterns during dna replication. It acknowledges the presence of hemimethylated DNA strands and guarantees that the methylation pattern is replicated onto the newly synthesized DNA strand, thereby maintaining epigenetic information during cell divisions. This maintenance function is vital for preserving cellular identity and ensuring the stability of the genome. In contrast, DNMT3A, and DNMT3B are responsible for de novo methylation, which involves the creation of new methylation patterns. These enzymes are especially active during early development and cell differentiation, where they establish new methylation marks that play a crucial role in determining cell fate and regulating gene expression specific to different tissues. DNMT3A, and DNMT3B also have functions in suppressing transposable elements, which aids in preserving the stability of the genome. Aberrations in DNA methylation are a characteristic feature of numerous diseases, with cancer being one of the most prevalent. In some genomic regions, the reduction of methylation can activate oncogenes and disrupt genomic stability, whereas the excessive methylation of tumor suppressor gene promoters, often facilitated by DNMTs, can inhibit essential pathways that prevent the development of tumors. For example, the hypermethylation of genes such as hhip within the context of the hedgehog signaling pathway has been associated with the development of different types of cancer. Due to their significant impact on the epigenome, dnmts are crucial targets for therapeutic interventions. Drugs like azacitidine and decitabine, which block the activity of dnmt, are already being used to treat certain types of blood cancers. Ongoing studies are focused on creating more precise inhibitors that can regulate dnmt activity and restore normal methylation patterns in diseases where epigenetic dysregulation plays a significant role [60-64], [104],[105].

In the context of HHIP (hedgehog-interacting protein), the regulation of its expression is greatly influenced by epigenetic modifications, particularly the methylation of its promoter region. The hypermethylation of the hip promoter effectively suppresses its expression, thereby disrupting its function as an inhibitor of the hedgehog pathway. By silencing the regulatory mechanisms, this pathway becomes unregulated, leading to its abnormal activation, which plays a role in tumor development. The activation of the hedgehog pathway in this way has been associated with the development and advancement of different types of cancers, such as lung, pancreatic, gastric, and colorectal cancers. DNA methyltransferases (DNMTs), with DNMT3B being a prominent example, have been recognized as crucial components in this process. DNMT3A, which is mainly responsible

for de novo methylation, has been observed to target the HHIP promoter for hypermethylation. The overactive behavior of dnmt3b in cancer cells is a well- known occurrence, leading to the inactivation of tumor suppressor genes such as HHIP. This hypermethylation-mediated silencing enables uncontrolled cell growth, survival, and spread by keeping the hedgehog pathway active. Recent research has highlighted the potential benefits of targeting DNMT3A to enhance the expression of the HOX gene. By blocking DNMT3A, it might be feasible to reverse the hypermethylation of the hip promoter, thereby reactivating its expression. The reintroduction of HHIP as a functional inhibitor could restore equilibrium to the hedgehog pathway, obstructing the signaling cascade that fuels tumor growth and advancement. This strategy has the potential to not only prevent the development of tumors but also improve the effectiveness of current treatments by targeting a crucial regulatory mechanism. Furthermore, the precise function of dnmt3b in regulating HHIP silencing makes it an appealing target for targeted treatments. By using selective inhibitors of DNMT3A, it may be possible to reduce unintended effects while specifically targeting the demethylation of the hhip promoter. Preclinical research and epigenetic therapies employing DNMT inhibitors, like azacitidine and decitabine, establish a groundwork for creating DNMT3A-specific inhibitors designed to target cancers influenced by hedgehog pathway abnormalities. The interaction between hhip promoter hypermethylation and DNMT3A activity underscores a crucial aspect of cancer biology. Additionally, it offers a chance for innovative therapeutic approaches that can reverse epigenetic silencing and restore normal regulatory pathways, thereby inhibiting cancer progression.

Our main objective is to discover dnmt3b inhibitors that can selectively target dnmt3b, with the ultimate goal of restoring hedgehog-interacting protein (HHIP) expression. HHIP acts as a crucial negative regulator of the hedgehog signaling pathway, and its silencing through dna hypermethylation has been strongly linked to cancer progression in various malignancies, including gastric, pancreatic, lung, and colorectal cancers. Due to its crucial role in establishing new methylation patterns, dnmt3b is a key player in the hypermethylation of the HHIP promoter, resulting in the loss of its tumor-suppressive function. Our research highlights the significant role of DNMT3A in silencing the hedgehog pathway, which is responsible for the uncontrolled activation of this pathway—a key factor in tumor development. By specifically targeting DNMT3A, it is possible to reverse the hypermethylation, activate HHIP expression, and restore its inhibitory effects on the hedgehog pathway. This focused approach not only targets the specific epigenetic mechanism but also reduces the activation of oncogenic pathways linked to hip silencing. The therapeutic

potential of DNMT3A inhibitors is attributed to their ability to selectively demethylate specific genomic regions, such as the HHIP promoter, while minimizing unintended effects on other areas of the genome. This level of specificity is vital for creating therapies that are both effective and have minimal side effects. Previous research using existing DNMT inhibitors, like azacitidine and decitabine, has shown that epigenetic reprogramming is possible in cancer cells. Nevertheless, these inhibitors do not possess the specificity necessary to selectively target dnmt3b, underscoring the necessity for further development.

For the advancement of future inhibitors specifically designed for this isoform. Our research also emphasizes the wider implications of inhibiting DNMT3A in cancer treatment. By restoring normal expression of the hedgehog pathway, selective DNMT3A inhibitors could disrupt the pathway's abnormal activity, hindering cancer cell proliferation, survival, and the spread of the disease. This strategy could be especially beneficial when used in conjunction with other treatments, such as chemotherapy and immunotherapy, by targeting a crucial epigenetic factor that contributes to cancer progression. In summary, the identification and development of selective DNMT3A inhibitors show significant potential as a therapeutic approach for cancers associated with dysregulation of the hedgehog pathway. By targeting the epigenetic silencing of hhip, these inhibitors have the potential to restore normal cellular regulation, inhibit tumor growth, and enhance clinical outcomes for patients with dnmt3b-associated cancers.

To identify potential dna methyltransferase (dnmt) inhibitors, we began by screening a set of 22 compounds for their inhibitory activity against DNMT1, DNMT3A, and DNMT3B the key enzymes responsible for establishing and maintaining DNA methylation patterns. Among these compounds, 13 exhibited IC50 values exceeding 0.084, suggesting their potential to inhibit DNMT activity efficiently. These compounds served as the foundation for further exploration into developing selective DNMTinhibitors, particularly targeting DNMT3A, given its role in the hypermethylation of tumor suppressor genes such as HHIP. After the initial screening, we expanded our analysis to a more extensive compound library obtained from chembridge, which consists of 1,654,770 distinct compounds. To expedite the identification of potential dnmt inhibitors, we utilized virtual screening methods. This process involved creating maccs keys fingerprints for every compound in the chembridge library. These fingerprints, which are extensively employed in cheminformatics, offer a binary representation of molecular characteristics, facilitating the swift comparison of chemical structures. The subsequent step required employing the tanimoto similarity index to assess the similarity between the chemical structures of the library compounds and the 13 reference

compounds that were initially identified. A tanimoto similarity threshold of 80% or higher was used, guaranteeing that the chosen compounds exhibited substantial structural similarity to the reference compounds, which had previously shown promising DNMT inhibitory activity. By utilizing a comprehensive virtual screening method, we discovered 2,198 compounds that displayed promising similarity profiles, surpassing the predetermined similarity threshold. These compounds encompass a wide range of chemical structures and exhibit inhibitory effects on DNMT enzymes. This particular subset was given priority for in silico and in vitro analyses to assess their selectivity, binding affinity, and potential off-target effects. The recognition of these 2,198 compounds establishes the preparations for the subsequent stages of the study, which will involve molecular docking and dynamics simulations to predict the binding interactions between the compounds and DNMT3A, as well as evaluate their inhibitory effects at the molecular level. By narrowing down the extensive chemical library to this concentrated set, we have greatly improved the practicality of discovering new and specific dnmt inhibitors that could potentially be developed into effective treatments for cancers caused by dnmt3b-mediated hypermethylation.

Following the initial screening and similarity-based filtering, a drug-likeness filter was utilized to narrow down the pool of 2,198 compounds. This step utilized widely accepted criteria, such as lipinski's rule of five, to ensure the compounds possessed favorable pharmacokinetic and pharmacodynamic properties, including appropriate molecular weight, lipophilicity, hydrogen bond donors/acceptors, and solubility. The filtration process resulted in a smaller pool of 460 compounds, which had a higher likelihood of being suitable for drug development. The 460 compounds were then analyzed using molecular docking techniques against the active sites of dnmt1, dnmt3a, and dnmt3b, the three main dna methyltransferases associated with gene silencing and cancer development. The docking simulations sought to forecast the binding affinity and interactions of each compound with these enzymes, offering valuable insights into

their inhibitory capabilities and selectivity. The docking outcomes were encouraging, indicating that a significant number of the compounds displayed robust interactions with the dnmt enzymes. Specifically: 386 compounds displayed significant binding to DNMT1, indicating their potential to disrupt the maintenance of DNA methylation patterns during replication. 308 compounds exhibited robust interactions with DNMT3A, suggesting their potential to impede the initiation of new methylation patterns essential for the formation of novel methylation marks. 397 compounds displayed strong binding affinities to DNMT3B, indicating its vulnerability to these inhibitors. Notably, DNMT3A was identified as the enzyme with the highest number of compounds exhibiting strong binding affinities. This discovery corresponds with its established function in the hypermethylation of tumor suppressor genes, including HHIP, in cancer. The preferential binding of DNMT3A highlights its importance as a potential therapeutic target, especially for cancers caused by the silencing of essential regulatory genes through epigenetic mechanisms. The compounds that exhibited the highest binding affinities to DNMT3A will be given priority for additional investigations, such as molecular dynamics simulations to evaluate the stability of their interactions, and in vitro experiments to verify their inhibitory effects. By adopting a targeted approach, this study not only narrows down the search for potential inhibitors but also establishes a clear path towards developing selective DNMT3A inhibitors as potential cancer treatments. By employing this structured approach, the study seeks to connect virtual predictions with real-world applications, propelling the field of epigenetic therapy forward.

To enhance the process of identifying potential dna methyltransferase (dnmt) inhibitors, hierarchical clustering analysis was conducted using molecular descriptors of the 460 filtered compounds. Molecular descriptors, encompassing physicochemical properties, structural characteristics, and topological indices, offer a detailed profile of each compound, enabling the identification of structurally and functionally comparable groups. By employing a clustering approach, compounds with similar properties are grouped together, allowing for a more focused selection process that identifies those with the highest inhibitory potential. The hierarchical clustering algorithm generated eight separate clusters, each representing a group of compounds that shared similar molecular descriptor profiles. Among the various clusters, cluster 5, comprising 39 compounds, showed the most potential. These compounds exhibited the highest binding affinities to DNMT3A, as evidenced by their docking scores and predicted interaction profiles with the enzyme's active site. The strong binding interactions indicate

that these compounds have the potential to inhibit DNMT3A efficiently, making them excellent candidates for further study. The compounds in cluster 5 were examined for recurring structural patterns and functional groups, which may play a role in their ability to selectively bind to DNMT3A. This knowledge offers valuable guidance for enhancing lead compounds through structure-based drug design. Additionally, the clustering approach assists in narrowing down the list of potential candidates to those most likely to exhibit the desired biological activity, thereby minimizing the time and resources needed for subsequent experimental validation. The subsequent steps entail conducting in-depth molecular dynamics simulations to evaluate the stability and specificity of interactions between dnmt3b and the compounds in cluster 5. Furthermore, in vitro enzymatic assays will be performed to verify their inhibitory capabilities. These compounds also provide opportunities for additional modifications and improvements to enhance their ability to be absorbed and utilized by the body, as well as their effectiveness in treating the targeted disease. Through the use of hierarchical clustering, this study efficiently streamlines the drug discoveryprocess, pinpointing cluster 5 compounds as the most promising candidates for the development of selective DNMT3A inhibitors. This focused approach aims to enhance the expression of healthy genes and inhibit the hedgehog signaling pathway to impede the progression of cancer caused by epigenetic abnormalities.

Our research extends previous studies on the functions of dna methyltransferases (DNMTs) and the hedgehog-interacting protein (HHIP) in cancer, proposing a fresh and precise method to tackle epigenetic abnormalities. By concentrating on DNMT3A, a crucial enzyme involved in dna hypermethylation, we have discovered its pivotal role in silencing HHIP, a significant inhibitor of the hedgehog signaling pathway. By selectively inhibiting DNMT3A, a molecule involved in the restoration of heterochromatin, researchers have discovered a potential therapeutic approach to mitigate the tumor-promoting effects caused by its reduced expression. This approach highlights the potential of epigenetic therapies in reversing abnormal methylation patterns that contribute to cancer progression. One of the significant achievements of our research is the discovery of a group of potential compounds, which were further refined using hierarchical clustering and docking analyses, exhibiting strong binding affinities to DNMT3A. These compounds serve as a concentrated group for further development as selective dnmt3b inhibitors. By focusing on compounds that specifically target DNMT3A, our study offers a strategic approach for developing precision therapies that

can reactivate the expression of HHIP, thereby inhibiting the hedgehog pathway and its oncogenic consequences. Although these advancements are noteworthy, a significant drawback of our study is the lack of molecular dynamics (md) simulations to verify the binding affinities and stability of the identified compounds. While docking analyses offer initial insights into the potential interactions between compounds and DNMT3A, md simulations would allow for a more comprehensive evaluation of the dynamic behavior of these interactions in physiological conditions. These simulations could provide valuable insights into the structural changes, binding strength, and selectivity of the compounds, which are essential for predicting their effectiveness as potential treatments. To enhance the credibility of our future research, it is crucial to address this limitation and strengthen the reliability of our findings, thereby supporting the development of DNMT3A inhibitors for preclinical purposes. Furthermore, experimental validation through in vitro and in vivo studies will be crucial to verify the effectiveness and safety of the candidate compounds. This research seeks to combine computational and experimental methods to connect the process of discovery with the practical application of epigenetic cancer therapies, pushing the boundaries of the field.

Our research extends previous studies on the functions of dna methyltransferases (DNMTs) and the hedgehog-interacting protein (HHIP) in cancer, proposing a fresh and precise method to tackle epigenetic abnormalities. By concentrating on DNMT3A, a crucial enzyme involved in DNA hypermethylation, we have discovered its pivotal role in silencing HHIP, a significant inhibitor of the hedgehog signaling pathway. By selectively inhibiting DNMT3A, a molecule involved in the restoration of heterochromatin, researchers have discovered a potential therapeutic approach to mitigate the tumor-promoting effects caused by its reduced expression. This method highlights the potential of epigenetic treatments.

In reversing aberrant methylation patterns that drive cancer progression.A major contribution of our work is the identification of a set of candidate compounds, refined through hierarchical clustering and docking analyses, with strong binding affinities to DNMT3A. These compounds serve as a concentrated group for further development as selective DNMT3A inhibitors. By focusing on compounds that specifically target DNMT3A, our study offers a strategic approach for developing precision therapies that can reactivate the expression of hip, thereby inhibiting the hedgehog pathway and its oncogenic consequences. Although these

advancements are noteworthy, a significant drawback of our study is the lack of molecular dynamics (md) simulations to verify the binding affinities and stability of the identified compounds. While docking analyses offer initial insights into the potential interactions between compounds and DNMT3A, md simulations would allow for a more comprehensive evaluation of the dynamic behavior of these interactions under physiological conditions. These simulations could provide valuable insights into the structural changes, binding strength, and selectivity of the compounds, which are essential for predicting their effectiveness as potential treatments. To enhance the reliability of our findings and facilitate the preclinical development of DNMT3A inhibitors, it is crucial to address this limitation in future research. Furthermore, experimental validation through in vitro and in vivo studies will be crucial to verify the effectiveness and safety of the candidate compounds. This research seeks to combine computational and experimental methods to connect the process of discovery with the practical application of epigenetic cancer therapies, pushing the boundaries of the field.

CONCLUSION:

This research successfully identified a selection of compounds with low IC_{50} values that exhibit significant efficacy against the DNA methyltransferases DNMT1, DNMT3A, and DNMT3B, crucial regulators of hhip expression via DNA hypermethylation in various cancers. By utilizing in silico methodologies, we efficiently generated molecular fingerprints, assessed structural similarities, and screened for drug-like properties, facilitating the identification of potential DNMT inhibitors. Through the computational analysis of over 1.6 million molecules, we pinpointed 39 candidates within cluster 5 that demonstrated substantial binding affinity to DNMT3A , underscoring the enzyme's potential as a therapeutic target for reactivating HHIP expression and inhibiting oncogenic processes. The results of this research highlight the significance of focusing on DNMT3A in the development of new cancer treatments that can restore normal cellular function.

Hedgehog signaling pathways. Nevertheless, the identified compounds require additional research to assess their efficacy as DNMT inhibitors in living organisms. Future studies should focus on conducting clinical trials to evaluate the effectiveness of these compounds in restoring normal hedgehog signaling and inhibiting cancer progression. By deepening our comprehension of how these compounds interact with their targets, we aspire to make

significant contributions to the creation of novel and efficient treatment approaches for individuals battling cancer.

CHAPTER # 6

Reference

1. Wei, X., Liu, W., Wang, J. Q., & Tang, Z. (2020). “Hedgehog pathway”: a potential target of itraconazole in the treatment of cancer. Journal of cancer research and clinical oncology, 146, 297-304.

2. Choudhry, Z., Rikani, A. A., Choudhry, A. M., Tariq, S., Zakaria, F., Asghar, M. W., ... & Mobassarah, N. J. (2014). Sonic hedgehog signalling pathway: a complex network. Annals of neurosciences, 21(1), 28.

3. Armas-López, L., Zúñiga, J., Arrieta, O., & Ávila-Moreno, F. (2017). The Hedgehog-GLI pathway in embryonic development and cancer: implications for pulmonary oncology therapy. Oncotarget, 8(36), 60684.

4. Zhang, J., Fan, J., Zeng, X., Nie, M., Luan, J., Wang, Y., ... & Yin, K. (2021). Hedgehog signaling in gastrointestinal carcinogenesis and the gastrointestinal tumor microenvironment. Acta Pharmaceutica Sinica B, 11(3), 609-620.

5. Ma, C., Hu, K., Ullah, I., Zheng, Q. K., Zhang, N., & Sun, Z. G. (2022). Molecular mechanisms involving the sonic hedgehog pathway in lung cancer therapy: recent advances. Frontiers in Oncology, 12, 729088.

6. Wang, X., Xu, J., Sun, Y., Cao, S., Zeng, H., Jin, N., ... & Huang, M. (2023). Hedgehog pathway orchestrates the interplay of histone modifications and tailors combination epigenetic therapies in breast cancer. Acta Pharmaceutica Sinica B.

7. Niyaz, M., Khan, M. S., & Mudassar, S. (2019). Hedgehog signaling: an Achilles’ heel in cancer. Translational oncology, 12(10), 1334-1344.

8. Raleigh, D. R., & Reiter, J. F. (2019). Misactivation of Hedgehog signaling causes inherited and sporadic cancers. The Journal of clinical investigation, 129(2), 465-475.

9. McMillan, R., & Matsui, W. (2012). Molecular pathways: the hedgehog signaling pathway in cancer. Clinical cancer research, 18(18), 4883-4888.

10. Oji, G. S., Gomez, P., Kurriger, G., Stevens, J., & Morcuende, J. A. (2007). Indian hedgehog signaling pathway differences between swarm rat chondrosarcoma and native rat chondrocytes. The Iowa orthopaedic journal, 27, 9.

11. Pachernegg, S., Georges, E., & Ayers, K. (2022). The Desert hedgehog signalling pathway in human gonadal development and differences of sex development. Sexual Development, 16(2-3), 98-111.

12. Carballo, G. B., Honorato, J. R., de Lopes, G. P. F., & Spohr, T. C. L. D. S. E. (2018). A highlight on Sonic hedgehog pathway. Cell Communication and Signaling, 16, 1-15.

13. Yang, C., Qi, Y., & Sun, Z. (2021). The role of sonic hedgehog pathway in the development of the central nervous system and aging-related neurodegenerative diseases. Frontiers in Molecular Biosciences, 8, 711710.

14. Panman, L., & Zeller, R. (2003). Patterning the limb before and after SHH signalling.

Journal of Anatomy, 202(1), 3-12.

15. Murdoch, J. N., & Copp, A. J. (2010). The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. Birth Defects Research Part A: Clinical and Molecular Teratology, 88(8), 633-652.

16. Dillon, R., Gadgil, C., & Othmer, H. G. (2003). Short-and long-range effects of Sonic hedgehog in limb development. Proceedings of the National Academy of Sciences, 100(18), 10152-10157.

17. Tickle, C., & Towers, M. (2017). Sonic hedgehog signaling in limb development.

Frontiers in cell and developmental biology, 5, 14.

18. Yang, C., Qi, Y., & Sun, Z. (2021). The role of sonic hedgehog pathway in the development of the central nervous system and aging-related neurodegenerative diseases. Frontiers in Molecular Biosciences, 8, 711710.

19. Lewis, P. M., Gritli-Linde, A., Smeyne, R., Kottmann, A., & McMahon, A. P. (2004). Sonic hedgehog signaling is required for expansion of granule neuron precursors and patterning of the mouse cerebellum. Developmental biology, 270(2), 393-410.

20. Nanni, L., Ming, J. E., Bocian, M., Steinhaus, K., Bianchi, D. W., de Die-Smulders, C., ... & Muenke, M. (1999). The mutational spectrum of the sonic hedgehog gene in holoprosencephaly: SHH mutations cause a significant proportion of autosomal dominant holoprosencephaly. Human molecular genetics, 8(13), 2479-2488

21. Ohba, S. (2020). Hedgehog signaling in skeletal development: roles of Indian hedgehog and the mode of its action. International journal of molecular sciences, 21(18), 6665.

22. Maes, C. (2017, February). Signaling pathways effecting crosstalk between cartilage and adjacent tissues: Seminars in cell and developmental biology: The biology and pathology of cartilage. In Seminars in cell & developmental biology (Vol. 62, pp. 16-33). Academic Press.

23. Jiang, Z., Derrick-Roberts, A. L., & Byers, S. (2020). Altered IHH signaling contributes to reduced chondrocyte proliferation in the growth plate of MPS VII mice. Molecular Genetics and Metabolism Reports, 25, 100668.

24. Wysolmerski, J. J. (2012). Parathyroid hormone-related protein: an update. The Journal of Clinical Endocrinology & Metabolism, 97(9), 2947-2956.

25. Zeng, F., Liu, H., Xia, X., Shu, Y., Cheng, W., Xu, H., ... & Xie, Q. (2022). Case Report: Brachydactyly Type A1 Induced by a Novel Variant of in-Frame Insertion in the IHH Gene. Frontiers in Genetics, 13, 814786.

26. Li, X., Yang, S., Chinipardaz, Z., Koyama, E., & Yang, S. (2021). SAG therapy restores bone growth and reduces enchondroma incidence in a model of skeletal chondrodysplasias caused by Ihh deficiency. Molecular Therapy-Methods & Clinical Development, 23, 461- 475.

27. Pierre, M., Yoshimoto, M., Huang, L., Richardson, M., & Yoder, M. C. (2009). VEGF and IHH rescue definitive hematopoiesis in Gata-4 and Gata-6–deficient murine embryoid bodies. Experimental hematology, 37(9), 1038-1053.

28. Lim, Y., & Matsui, W. (2010). Hedgehog signaling in hematopoiesis. Critical Reviews™ in Eukaryotic Gene Expression, 20(2).

29. Dilower, I., Niloy, A. J., Kumar, V., Kothari, A., Lee, E. B., & Rumi, M. K. (2023). Hedgehog signaling in gonadal development and function. Cells, 12(3), 358.

30. Min, M., Song, T., Sun, M., Wang, T., Tan, J., & Zhang, J. (2022). Dhh signaling pathway regulates reconstruction of seminiferous tubule-like structure. Reproductive Biology, 22(4), 100684.

31. Svechnikov, K., Landreh, L., Weisser, J., Izzo, G., Colón, E., Svechnikova, I., & Söder, O. (2010). Origin, development and regulation of human Leydig cells. Hormone research in paediatrics, 73(2), 93-101.

32. Gomes, N. L., Chetty, T., Jorgensen, A., & Mitchell, R. T. (2020). Disorders of sex development—novel regulators, impacts on fertility, and options for fertility preservation. International Journal of Molecular Sciences, 21(7), 2282.

33. Mehta, P., Singh, P., Gupta, N. J., Sankhwar, S. N., Chakravarty, B., Thangaraj, K., & Rajender, S. (2021). Mutations in the desert hedgehog (DHH) gene in the disorders of sexual differentiation and male infertility. Journal of Assisted Reproduction and Genetics, 38, 1871-1878.

34. Zhang, J., Liu, Z., & Jia, J. (2021). Mechanisms of smoothened regulation in hedgehog signaling. Cells, 10(8), 2138.

35. Kong, J. H., Siebold, C., & Rohatgi, R. (2019). Biochemical mechanisms of vertebrate hedgehog signaling. Development, 146(10), dev166892.

36. Jeng, K. S., Chang, C. F., & Lin, S. S. (2020). Sonic hedgehog signaling in organogenesis, tumors, and tumor microenvironments. International journal of molecular sciences, 21(3), 758.

37. Skoda, A. M., Simovic, D., Karin, V., Kardum, V., Vranic, S., & Serman, L. (2018). The role of the Hedgehog signaling pathway in cancer: A comprehensive review. Bosnian journal of basic medical sciences, 18(1), 8.

38. Suchors, C., & Kim, J. (2022). Canonical hedgehog pathway and noncanonical GLI transcription factor activation in cancer. Cells, 11(16), 2523.

39. Briscoe, J., & Thérond, P. P. (2013). The mechanisms of Hedgehog signalling and its roles in development and disease. Nature reviews Molecular cell biology, 14(7), 416-429.

40. Bishop, B., Aricescu, A. R., Harlos, K., O'callaghan, C. A., Jones, E. Y., & Siebold, C. (2009). Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HHIP. Nature structural & molecular biology, 16(7), 698-703.

41. Griffiths, S. C., Schwab, R. A., El Omari, K., Bishop, B., Iverson, E. J., Malinauskas, T.,

... & Siebold, C. (2021). Hedgehog-Interacting Protein is a multimodal antagonist of Hedgehog signalling. Nature communications, 12(1), 7171.

42. Wang, H., Wang, Y., Zhang, H., Liang, Z., Hu, W., Qiu, S., ... & Li, L. (2023). Hedgehog interacting protein as a circulating biomarker in women with obesity: a cross-sectional study and intervention studies. Annals of Medicine, 55(1), 2206162.

43. Liao, Y., Liao, Y., & Wen, F. (2022). Association of human Hedgehog interacting protein gene polymorphisms with the risk of chronic obstructive pulmonary disease: a meta-analysis. Expert Review of Respiratory Medicine, 16(4), 447-459.

44. Zhou, X., Baron, R. M., Hardin, M., Cho, M. H., Zielinski, J., Hawrylkiewicz, I., ... & Silverman, E. K. (2012). Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. Human molecular genetics, 21(6), 1325-1335.

45. Ke, B., Wang, X. N., Liu, N., Li, B., Wang, X. J., Zhang, R. P., & Liang, H. (2020). Sonic Hedgehog/Gli1 signaling pathway regulates cell migration and invasion via induction of epithelial-to-mesenchymal transition in gastric cancer. Journal of Cancer, 11(13), 3932.

46. Wang, J., Cui, B., Li, X., Zhao, X., Huang, T., & Ding, X. (2023). The emerging roles of Hedgehog signaling in tumor immune microenvironment. Frontiers in Oncology, 13, 1171418.

47. Gigante, E. D., & Caspary, T. (2020). Signaling in the primary cilium through the lens of the Hedgehog pathway. Wiley Interdisciplinary Reviews: Developmental Biology, 9(6), e377.

48. Amatya, S., Rajbhandari, S., Pradhan, S., Trinh, V., Paudel, U., & Parton, L. A. (2021). Hedgehog signaling pathway gene variant influences bronchopulmonary dysplasia in extremely low birth weight infants. World Journal of Pediatrics, 17, 298-304.

49. Zeng, L. H., Barkat, M. Q., Syed, S. K., Shah, S., Abbas, G., Xu, C., ... & Hussain, M. (2022). Hedgehog signaling: Linking embryonic lung development and asthmatic airway remodeling. Cells, 11(11), 1774.

50. Kumari, A., Li, L., Ermilov, A. N., Franks, N. E., Dlugosz, A. A., Allen, B. L., & Mistretta,

C. M. (2022). Hedgehog (HH) pathway endogenous antagonist HHIP: unique lingual expression in filiform papillae during homeostasis and ectopic in fungiform papillae during HH signaling inhibition. Developmental dynamics: an official publication of the American Association of Anatomists, 251(7), 1175.

51. Fu, F., Zhang, Y., Feng, J., & Nie, Y. (2023). Bioinformatics analysis of hedgehog interacting protein in colorectal cancer: a study based on GEO data and TCGA data. BMC gastroenterology, 23(1), 278.

52. Song, Y., Tu, J., Cheng, Y., Zhou, F., Liu, P., Zhou, S., ... & Sun, Y. (2020). HHIP

overexpression suppresses human gastric cancer progression and metastasis by reducing its CpG island methylation. Frontiers in Oncology, 10, 1667.

53. Fernández-Barrena, M. G., Arechederra, M., Colyn, L., Berasain, C., & Avila, M. A. (2020). Epigenetics in hepatocellular carcinoma development and therapy: The tip of the iceberg. Jhep Reports, 2(6), 100167.

54. Omura, N., & Goggins, M. (2009). Epigenetics and epigenetic alterations in pancreatic cancer. International journal of clinical and experimental pathology, 2(4), 310.

55. Zhao, J. G., Wang, J. F., Feng, J. F., Jin, X. Y., & Ye, W. L. (2019). HHIP overexpression inhibits the proliferation, migration and invasion of non-small cell lung cancer. PloS one, 14(11), e0225755.

56. Yun, J. H., Lee, C., Liu, T., Liu, S., Kim, E. Y., Xu, S., ... & Zhou, X. (2021). Hedgehog

interacting protein–expressing lung fibroblasts suppress lymphocytic inflammation in mice. JCI insight, 6(17).

57. Yin, X. H., Wang, X. Y., Liu, S. C., Yan, L., He, B. R., Hao, D. J., ... & Liu, Z. K. (2022).

lncRNA HHIP-AS1/HHIP modulates osteogenic differentiation of BM-MSCs by regulating Hedgehog signaling pathway. Aging (Albany NY), 14(21), 8839.

58. Greenberg, M. V., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. Nature reviews Molecular cell biology, 20(10), 590-607.

59. Chakraborty, B., Basu, M., Mukhopadhyay, D., Alam, N., Ghosh, S., Dutta, S., & Panda,

C. K. (2022). Differential promoter usages of PTCH1 and down regulation of HHIP are associated with HNSCC progression. Pathology-Research and Practice, 232, 153827.

60. Li, S., & Tollefsbol, T. O. (2021). DNA methylation methods: Global DNA methylation and methylomic analyses. Methods, 187, 28-43.

61. Phillips, T. (2008). The role of methylation in gene expression. Nature Education, 1(1), 116.

62. Sriraman, A., Debnath, T. K., Xhemalce, B., & Miller, K. M. (2020). Making it or breaking it: DNA methylation and genome integrity. Essays in biochemistry, 64(5), 687-703.

63. Villicaña, S., & Bell, J. T. (2021). Genetic impacts on DNA methylation: research findings and future perspectives. Genome biology, 22(1), 127.

64. Wan, J., Oliver, V. F., Wang, G., Zhu, H., Zack, D. J., Merbs, S. L., & Qian, J. (2015). Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. BMC genomics, 16, 1-11.

65. Uysal, F., Sukur, G., & Cinar, O. (2022). DNMT enzymes differentially alter global DNA methylation in a stage-dependent manner during spermatogenesis. Andrologia, 54(4), e14357.

66. Kar, S., Deb, M., Sengupta, D., Shilpi, A., Parbin, S., Torrisani, J., ... & Patra, S. K. (2012). An insight into the various regulatory mechanisms modulating human DNA methyltransferase 1 stability and function. Epigenetics, 7(9), 994-1007.

67. Neumeister, P., Albanese, C., Balent, B., Greally, J., & Pestell, R. G. (2002). Senescence and epigenetic dysregulation in cancer. The International Journal of Biochemistry & Cell Biology, 34(11), 1475-1490.

68. Chen, B. F., & Chan, W. Y. (2014). The de novo DNA methyltransferase DNMT3A in development and cancer. Epigenetics, 9(5), 669-677.

69. Zhou, J., Wu, T., & Liu, W. (2024). Research progress on DNMT3A gene expression in Acute myeloid leukemia. Zhonghua yi xue yi Chuan xue za zhi= Zhonghua Yixue Yichuanxue Zazhi= Chinese Journal of Medical Genetics, 41(8), 1010-1015.

70. Dar, M. S., Mensah, I. K., He, M., McGovern, S., Sohal, I. S., Whitlock, H. C., ... & Gowher, H. (2023). Dnmt3bas coordinates transcriptional induction and alternative exon inclusion to promote catalytically active Dnmt3b expression. Cell reports, 42(6).

71. Ito, T., Kubiura-Ichimaru, M., Murakami, Y., Bogutz, A. B., Lefebvre, L., Suetake, I., ... & Tada, M. (2022). DNMT1 regulates the timing of DNA methylation by DNMT3 in an enzymatic activity-dependent manner in mouse embryonic stem cells. PloS one, 17(1), e0262277.

72. Ries, R. E., Ma, X., Tregnago, C., Alonzo, T. A., Wang, J., Hylkema, T., ... & Meshinchi,

S. (2023). DNMT3A Mutants Are Enriched in NPMc+ AML and Associated with Adverse Outcome in Childhood AML. Blood, 142(Supplement 1), 4306-4306.

73. Beard, D. C., Zhang, X., Wu, D. Y., Martin, J. R., Erickson, A., Boua, J. V., ... & Gabel,

H. W. (2023). Distinct disease mutations in DNMT3A result in a spectrum of behavioral, epigenetic, and transcriptional deficits. Cell reports, 42(11).

74. Vromans, R. D., & Jongman, S. R. (2018). The interplay between selective and nonselective inhibition during single word production. PloS one, 13(5), e0197313.

75. Nguyen, N. M., & Cho, J. (2022). Hedgehog pathway inhibitors as targeted cancer therapy and strategies to overcome drug resistance. International Journal of Molecular Sciences, 23(3), 1733.

76. Jamieson, C., Martinelli, G., Papayannidis, C., & Cortes, J. E. (2020). Hedgehog pathway inhibitors: a new therapeutic class for the treatment of acute myeloid leukemia. Blood cancer discovery, 1(2), 134-145.

77. Ortega-Martínez, A., Pérez-Rubio, G., Ramírez-Venegas, A., Ramírez-Díaz, M. E., Cruz- Vicente, F., Martínez-Gómez, M. D. L., ... & Falfán-Valencia, R. (2020). Participation of HHIP gene variants in COPD susceptibility, lung function, and serum and sputum protein levels in women exposed to biomass-burning smoke. Diagnostics, 10(10), 734.

78. Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function.

Neuropsychopharmacology, 38(1), 23-38.

79. Tajima, S., Suetake, I., Takeshita, K., Nakagawa, A., Kimura, H., & Song, J. (2016). Domain structure of the Dnmt1, Dnmt3a, and Dnmt3b DNA methyltransferases. In DNA Methyltransferases-Role and Function (pp. 45-68). Cham: Springer International Publishing.

80. Martin, S. T., Sato, N., Dhara, S., Chang, R., Hustinx, S. R., Abe, T., ... & Goggins, M. (2005). Aberrant methylation of the Human Hedgehog interacting protein (HHIP) gene in pancreatic neoplasms. Cancer biology & therapy, 4(7), 728-733.

81. Li, P., Zhang, X., Murphy, A. J., Costa, M., Zhao, X., & Sun, H. (2021). Downregulation of hedgehog-interacting protein (HHIP) contributes to hexavalent chromium-induced malignant transformation of human bronchial epithelial cells. Carcinogenesis, 42(1), 136- 147.

82. Li, Y., Zhang, L., Polverino, F., Guo, F., Hao, Y., Lao, T., ... & Zhou, X. (2021). Hedgehog interacting protein (HHIP) represses airway remodeling and metabolic reprogramming in COPD-derived airway smooth muscle cells. Scientific reports, 11(1), 9074.

83. Pal, D., Samadder, S., Dutta, P., Roychowdhury, A., Chakraborty, B., Dutta, S., ... & Panda,

C. K. (2023). Differential association of hedgehog pathway in development of

cervical carcinoma and its chemo-tolerance. Pathology-Research and Practice, 248, 154696.

84. Kuwahara, H., & Gao, X. (2021). Analysis of the effects of related fingerprints on molecular similarity using an eigenvalue entropy approach. Journal of Cheminformatics, 13, 1-12.

85. Coley, C. W., Green, W. H., & Jensen, K. F. (2019). RDChiral: An RDKit wrapper for handling stereochemistry in retrosynthetic template extraction and application. Journal of chemical information and modeling, 59(6), 2529-2537.

86. Vogt, M., & Bajorath, J. (2017). Modeling tanimoto similarity value distributions and predicting search results. Molecular Informatics, 36(7), 1600131.

87. Kralj, S., Jukič, M., & Bren, U. (2023). Molecular filters in medicinal chemistry.

Encyclopedia, 3(2), 501-511.

88. Mignani, S., Rodrigues, J., Tomas, H., Jalal, R., Singh, P. P., Majoral, J. P., & Vishwakarma, R. A. (2018). Present drug-likeness filters in medicinal chemistry during the hit and lead optimization process: how far can they be simplified?. Drug discovery today, 23(3), 605-615.

89. Vilar, S., Cozza, G., & Moro, S. (2008). Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. Current topics in medicinal chemistry, 8(18), 1555-1572.

90. Oyesakin, Y. M., George, D. E., Fadare, R. Y., Idris, A. Y., & Fadare, O. A. (2018). Molecular Docking and In-Silico ADME Prediction of Substituted (E)-4-Styryl-7, 8-dihydroquinazolin-5 (6H)-ones and 5-((E)-Styryl) pyrimidine [4, 5-d] pyrimidine-2, 4 (1H, 3H)-diones as Potential SERT Inhibitors and Antidepressants. American Journal of Pharmacological Sciences, 6(1), 25-32.

91. He, Y., Liu, K., Han, L., & Han, W. (2022). Clustering analysis, structure fingerprint analysis, and quantum chemical calculations of compounds from essential oils of sunflower (Helianthus annuus L.) receptacles. International Journal of Molecular Sciences, 23(17), 10169.

92. Makeneni, S., Thieker, D. F., & Woods, R. J. (2018). Applying pose clustering and MD simulations to eliminate false positives in molecular docking. Journal of chemical information and modeling, 58(3), 605-614.

93. Lahyaoui, M., El-Idrissi, H., Saffaj, T., Ihssane, B., Saffaj, N., Mamouni, R., & Rodi, Y.

K. (2023). QSAR modeling, molecular docking and molecular dynamic simulation of phosphorus-substituted quinoline derivatives as topoisomerase I inhibitors. Arabian Journal of Chemistry, 16(6), 104783.

94. Forrestall, K., Pringle, E. S., Sands, D., Duguay, B. A., Farewell, B., Woldemariam, T., ... & Darvesh, S. (2023). Dual inhibition of coronavirus Mpro and PLpro enzymes by phenothiazines and their antiviral activity. bioRxiv, 2023-09.

95. Garcia, N., Al-Hendy, A., Baracat, E. C., Carvalho, K. C., & Yang, Q. (2020). Targeting hedgehog pathway and DNA methyltransferases in uterine leiomyosarcoma cells. Cells, 10(1), 53.

96. Shen, G., Shen, H., Zhang, J., Yan, Q., & Liu, H. (2020). DNA methylation in Hepatoblastoma-a literature review. Italian journal of pediatrics, 46(1), 113.

97. Walton, K. D., & Gumucio, D. L. (2021). Hedgehog signaling in intestinal development and homeostasis. Annual review of physiology, 83(1), 359-380.

98. Fendrich, V., Esni, F., Garay, M. V. R., Feldmann, G., Habbe, N., Jensen, J. N., ... & Maitra, A. (2008). Hedgehog signaling is required for effective regeneration of exocrine pancreas. Gastroenterology, 135(2), 621-631.

99. Zhao, X. P., Chang, S. Y., Liao, M. C., Lo, C. S., Chenier, I., Luo, H., ... & Zhang, S. L. (2018). Hedgehog interacting protein promotes fibrosis and apoptosis in glomerular endothelial cells in murine diabetes. Scientific Reports, 8(1), 1-10.

100. Guo, F., Zhang, L., Yu, Y., Gong, L., Tao, S., Werder, R. B., ... & Zhou, X. (2024).

Identification of a distal enhancer regulating hedgehog interacting protein gene in human lung epithelial cells. Ebiomedicine, 101.

101. Shahi, M. H., Zazpe, I., Afzal, M., Sinha, S., Rebhun, R. B., Meléndez, B., ... & Castresana, J. S. (2015). Epigenetic regulation of human hedgehog interacting protein in glioma cell lines and primary tumor samples. Tumor Biology, 36, 2383-2391.

102. Zhao, J. G., Wang, J. F., Feng, J. F., Jin, X. Y., & Ye, W. L. (2019). HHIP overexpression inhibits the proliferation, migration and invasion of non-small cell lung cancer. PloS one, 14(11), e0225755.

103. Song, Y., & Zuo, Y. (2014). Occurrence of HHIP gene CpG island methylation in gastric cancer. Oncology letters, 8(5), 2340-2344.

104. Taniguchi, H., Yamamoto, H., Akutsu, N., Nosho, K., Adachi, Y., Imai, K., & Shinomura,

Y. (2007). Transcriptional silencing of hedgehog-interacting protein by CpG hypermethylation and chromatic structure in human gastrointestinal cancer. The Journal of pathology, 213(2), 131-139.

105. Tada, M., Kanai, F., Tanaka, Y., Tateishi, K., Ohta, M., Asaoka, Y., ... & Omata, M. (2008). Down-regulation of hedgehog-interacting protein through genetic and epigenetic alterations in human hepatocellular carcinoma. Clinical cancer research, 14(12), 3768- 3776.

106. Hinoue, T., Weisenberger, D. J., Pan, F., Campan, M., Kim, M., Young, J., ... & Laird, P.

W. (2009). Analysis of the association between CIMP and BRAFV600E in colorectal cancer by DNA methylation profiling. PloS one, 4(12), e8357.

107. Shen, G., Shen, H., Zhang, J., Yan, Q., & Liu, H. (2020). DNA methylation in Hepatoblastoma-a.

108. Li, P., Zhang, X., Murphy, A. J., Costa, M., Zhao, X., & Sun, H. (2021). Downregulation of hedgehog-interacting protein (HHIP) contributes to hexavalent chromium-induced malignant transformation of human bronchial epithelial cells. Carcinogenesis, 42(1), 136- 147.

109. Bork, S., Pfister, S., Witt, H., Horn, P., Korn, B., Ho, A. D., & Wagner, W. (2010). DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. Aging cell, 9(1), 54-63.

110. Chakraborty, B., Basu, M., Mukhopadhyay, D., Alam, N., Ghosh, S., Dutta, S., & Panda,

C. K. (2022). Differential promoter usages of PTCH1 and down regulation of HHIP are associated with HNSCC progression. Pathology-Research and Practice, 232, 153827.