Deciphering the Molecular Mechanistic of the H3K27M Mutation and PRC2 complex binding in Diffuse Intrinsic Pontine Glioma (DIPG)

By

ALINA RAHMAN

Master of Science in Bioinformatics

Fall 21-MSBI-NUST00000360994

Supervised by:

Prof. Dr. Ishrat Jabeen

School of Interdisciplinary Engineering & Sciences (SINES)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

September, 2023

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by Miss. **Alina Rahman** Registration No. **00000360994** of **SINES** has been vetted by undersigned, found complete in all aspects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature with stamp: _____

DR. ISHRAT JABEEN
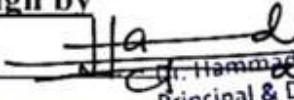Professor
SINES - NUST Sector H-12
Islamabad

Name of Supervisor: **Dr. Ishrat Jabeen**

Date: _21 | 09 | 2023_____

Signature of HoD with stamp: _____

Dr. Mian Ilyas Ahmad
HoD Engineering
Professor
SINES - NUST, Sector H-12
Islamabad

Date: _____ 2.1 SEP 2023

Countersign by

Signature (Dean/Principal): _____ M. Cheema

Principal & Dean
SINES - NUST, Sector H-1
Islamabad

Date: _____ 2 2 SEP 2023

CERTIFICATE OF ORGANILITY

I hereby declare that the research work presented in this thesis has been generated by me as a result of my own research work. Moreover, none of its contents are plagiarized or submitted for any kind of assessment or higher degree. I have acknowledged and referenced all the main sources of help in this work.

Alina Rahman

Fall 2021-MS BI 00000360994

_____

Dedication

This thesis is dedicated to my beloved sister, *Nawal Rahman* (late) and to all the children who are affected by DIPG. I have put all of my efforts to understand the disease that took her away from us too soon. Through this work, I hope to contribute to the body of knowledge surrounding this disease, with the aim that one day it may lead to improved treatments and ultimately a cure.

In loving memory of **Nawal**, this thesis is dedicated with profound gratitude and a deep sense of love for the sister I hold in my heart forever.

ACKNOLEDGEMENTS

I begin by expressing my heartfelt gratitude to Allah, I am deeply thankful for His support in helping me to accomplish my goal of obtaining my MS degree. A special note of appreciation goes out to my dedicated teachers who have been instrumental in shaping my academic path. I am profoundly indebted to my supervisor, *Dr. Ishrat Jabeen*. Her unwavering support, expert guidance, and endless patience have been my guiding lights during moments of research challenges. Her mentorship has not only enriched my academic pursuits but also nurtured my growth as a researcher. I extend my sincere gratitude to the esteemed members of my guidance committee: *Dr. Muhammad Tariq Saeed*, *Dr. Uzma Habib* from SINES, NUST, and *Dr. Najam Us Sahar Sadaf Zaidi* from ASAB NUST. Your thoughtful insights and encouraging words have played a crucial role in refining my research and shaping my academic perspective.

It is with deep gratitude that I turn to my family, who have been my unwavering pillars of support for me. My mother, *Bushra Rahman*, stands as the epitome of selflessness and determination. All credit goes to her. Her sacrifices and boundless encouragement have paved the way for all of my accomplishments throughout my career. To my father, the late *Fazal-ur-Rahman Khalid*, whose love and guidance continue to inspire me till now, I pay tribute to him with a heavy heart. Though he is no longer with us, his influence remains etched in my journey forever. My beloved and beautiful sisters, *Areeba Rahman, Arshia Rahman*, and my late *Nawal Rahman*, have been my steadfast companions through every phase of my life. Their unwavering belief in me has been a source of strength and inspiration for me.

Finally, I extend my gratitude to all those who, in ways both seen and unseen, have contributed to my academic and personal growth. May the knowledge I have gained and the accomplishments I have achieved stand as a testament to the collective efforts that have shaped my path.

Table of Contents

Table of Figures

Table of Equations

Table of Tables

ABSTRACT

DIPG, a fatal pediatric brainstem tumor with a 0% survival rate, is hindered by its inoperable location, limiting research. It is driven by the H3K27M mutation in histone 3, causing structural changes in PRC2 and aberrant gene activation that leads to the childhood brainstem tumors. This study explores the H3K27M mutation's impact on PRC2 and gene regulation. The methodology employed includes extensive molecular dynamics (MD) simulations of the Mutant peptide (H3K27M) – PRC2 complex and Wild Type peptide (H3K27me3) – PRC2 complex. Subsequent to MD simulations, various analyses including RMSD and RMSF plots, counts and stability of Peptide-Protein interactions, and binding free energy analysis were conducted. Results unveiled that, in the absence of EED and SUZ12, the protein displayed heightened RMSD values spanning from 0-20 Å, indicative of instability. However, upon binding with the peptides (mutant or wild-type) and EED, SUZ12, the RMSD plot displayed stability in the PRC2 complex. Notably, the H3K27M-PRC2 complex exhibited higher RMSD values ranging from 1.95-5 Å compared to the H3K27me3-PRC2 complex, which maintained a range of 0.9 – 1.8 Å. Furthermore, the simulation revealed the straightening of the SBD subdomain of PRC2 in complex with the mutant peptide, a typically bent configuration in the active state of PRC2. This alteration may have contributed to PRC2 inactivity. Importantly, this analysis highlighted the pivotal role of ASN-668 in EZH2 for robust binding of H3K27M to PRC2, resulting in stable bonding and PRC2 inactivation. Moreover, interactions between SRM and SET-I regions of EZH2, involving Lys_660 and Valine_657 of SET-I and Phe_145, ASN_142, and Leu_149 of SRM, were thought to activate the lysine-binding substrate. However, the mutant peptide disrupted these interactions, particularly by interacting with Lys_660 of Set-I, pulling K-660 away from SRM's residue and preventing SET-I from becoming active which consequently inhibit lysine binding. This is potentially crucial in deactivating the PRC2 complex. The presented binding free energy calculations validated the hypothesis present in the literature that the H3K27M peptide exhibits stronger binding affinity to the PRC2 complex, as reflected by a more negative ΔG_bind value i.e. -126.3079 kcal/mol as compared to the wild-type H3K27me3 i.e. -42.3238 kcal/mol. This shift in binding energy

underscores the mutation's disruptive effect on the normal functioning of the PRC2 complex, implying a potential role in aberrant gene regulation. In summary, the presented computational exploration yields valuable insights into the structural and functional consequences of the H3K27M mutation within the PRC2 complex. These findings advance our understanding of epigenetic deregulation in cancer and may offer promising avenues for future therapeutic interventions.

Chapter 1
INTRODUCTION

1   Introduction

Pediatric brain tumors are the second most common type of cancer in children, after leukemia. These tumors are categorized as either supratentorial or infratentorial lesions. This can also be classified as congenital brain tumors (CBT) (diagnosed in the first 60 days of life), tumors of infancy (younger than 1 year of age), and tumors of older children. The prognosis of pediatric brain tumors depends on factors such as the age at diagnosis, histological classification, and the degree to which the tumor can be resected [1].

1.1   Understanding Brainstem Tumors:

Brainstem high grade tumors, accounting for approximately 10-20% of all central nervous system cancers in children aged 0–14 years in the United States, pose a significant challenge in pediatric oncology. Most of these tumors, known as gliomas, originate from glial cells—the supportive cells of the brain [2]. Within the realm of brainstem gliomas, two distinct categories exist that are low-grade gliomas and high-grade gliomas. Around 20% of cases belong to the low-grade group, typically developing in the midbrain, medulla, or dorsal pons. These tumors often exhibit focal characteristics and can be surgically resected with relatively positive prognoses. However, the remaining 80% of brainstem tumors arise from the pons and present a grave concern. Termed Diffuse Intrinsic Pontine Glioma (DIPG), these tumors spread extensively throughout the pons, leading to an extremely poor prognosis. Figure 1.1 is an MRI image of a DIPG patient showing the presence of tumor in the stem region of brain which is specifically located in PONS region [3].

Figure 1.1 MRI image of brainstem tumor location [3]

1.2   Diffuse Intrinsic Pontine Glioma?

Diffuse Intrinsic Pontine Glioma (DIPG) is a devastating pediatric brain tumor that arises in the pons, a critical brainstem region. This aggressive and infiltrative tumor primarily affects children between the ages of 5 and 10 and is characterized by its diffuse growth pattern and resistance to current treatment modalities. Around 300 children are diagnosed with DIPG each year [4]. Among the various genetic alterations observed in DIPG, the H3K27M mutation has emerged as a significant molecular driver, contributing to the pathogenesis of the disease.

1.2.1   DIPG in Pakistan:

The incidence of DIPG in Pakistan is not well-known because there is not enough research that is being carried out in Pakistan. Only 3 Pakistani studies are being retrieved from PubMed that are related to DIPG [5] [6] [7]. However, these studies did not report the incidence of DIPG in Pakistan.

1.2.2   Etiology of DIPG:

The formation of this specific tumor may be linked to how the brain grows. According to studies, particular cells present in high concentrations while the brain tissue is developing cause the tumor. The fact that the tumor is less common in adults but more prevalent in children aged 5 to 10 lends credence to this theory. During this time, the brain tissue is actively maturing. Previously, different studies have reported a type of cell similar to the cells that produce neurons (called neural precursor-like cells) in the distinct ventral pons region of the human brain. The ventral pons is a part of the brainstem where DIPG tumors typically form. These cells are connected, in terms of their location and when they appear, to the occurrence of DIPG [1].

Interestingly, the number of these neural precursor-like cells during mid-childhood follows a bell-shaped pattern that matches the occurrence of DIPG tumors. In simpler terms, as these cells increase during mid-childhood, the likelihood of DIPG tumors also tends to rise. On the other hand, these cells are not present in the midbrain region, which is why DIPGs are rarely found there [1].

These findings suggest that the development of DIPG tumors may be linked to the presence and activity of specific cells in the ventral pons during the mid-childhood, an important time for brain development. [1]

1.2.3   Clinical Representation of DIPG:

The signs and symptoms of Diffuse Intrinsic Pontine Glioma (DIPG) can vary depending on the location and size of the tumor. The following are typical indications and symptoms linked to DIPG:

1. **Cranial nerve dysfunction:** DIPG often affects the brainstem area, which controls many cranial nerves. This can result in various cranial nerve dysfunctions, such as double vision, drooping of the eyelids, and difficulty with eye movements, facial weakness or asymmetry, difficulty swallowing, and changes in speech.

2. **Difficulties with coordination and motor function:** DIPG can interfere with the normal functioning of the cerebellum, which is responsible for coordinating movements. Children with DIPG may experience problems with balance, coordination, walking, and fine motor skills. They may exhibit clumsiness, stumbling, and difficulty with tasks that require precise movements.

3. **Weakness and paralysis:** The tumor's location in the brainstem can cause weakness or paralysis in the muscles of the face, arms, or legs. This may manifest as difficulty moving certain body parts, muscle weakness, or a loss of muscle tone.

4. **Headaches:** Many children with DIPG experience persistent headaches, which may be severe and worsen over time. The headaches may be associated with nausea and vomiting and may be more pronounced in the morning or with changes in position.

5. **Vision changes:** DIPG can affect the optic pathways and cause vision problems. Children may experience blurred vision, double vision, changes in the visual field, or other visual disturbances.

6. **Changes in behavior and personality:** Some children with DIPG may exhibit changes in behavior, mood, and personality. These changes can include irritability, mood swings, depression, withdrawal, and alterations in cognitive function, such as difficulties with memory and attention.

1.2.4 How is DIPG Diagnosed?

DIPG is usually diagnosed with a combination of imaging tests, including:

- **Magnetic resonance imaging (MRI):** An MRI is a test that uses a strong magnetic field and radio waves to create detailed images of the brain and other organs. MRI is the most common way to diagnose DIPG. [8]

- **Computed tomography (CT scan):** A CT scan uses X-rays to create detailed images of the brain and other organs. CT scans are often used to confirm the diagnosis of DIPG and to look for any other tumors or abnormalities in the brain. [8]

- **Biopsy:** In some cases, a biopsy may be necessary to confirm the diagnosis of DIPG. A biopsy is a procedure in which a small piece of tissue is removed from the tumor and examined under a microscope. [8]

1.3   Treatments of DIPG:

The treatment options for Diffuse Intrinsic Pontine Glioma (DIPG) are limited due to the tumor's location in the brainstem and its diffuse nature. Currently, there is no known cure for DIPG. However, the following treatment modalities are commonly used to manage the symptoms and prolong survival:

1. **Radiation therapy:** Radiation is the primary treatment for DIPG. It aims to shrink the tumor and alleviate symptoms. Radiation therapy is typically delivered over several weeks, with daily sessions targeting the tumor. This treatment can provide temporary relief by reducing the size of the tumor and relieving pressure on surrounding structures. [8]

2. **Clinical trials:** Participation in clinical trials is an important option for children with DIPG. Researchers are continuously investigating new approaches, including novel therapies and targeted treatments, to improve outcomes for DIPG patients. Clinical trials may offer access to experimental drugs or strategies that are not yet widely available. [8]

3. **Surgery:** Surgery can be done in rare cases of DIPG, but it is not usually recommended because of the critical location of the tumor. [8]

4. **Supportive care:** Supportive care focuses on managing symptoms and enhancing the quality of life for children with DIPG. A comprehensive approach to treatment may be necessary, incorporating various disciplines such as physical therapy, occupational therapy, speech therapy, and palliative care. Supportive care addresses pain, improves mobility and function, provides emotional support, and enhances overall well-being. [8]

1.4   Pathophysiology:

One of the most widely accepted explanations for the development of diffuse intrinsic pontine glioma (DIPG) is a mutation in a gene called H3. This mutation explicitly affects the N-terminal tail of a protein called histone H3. Recent studies have discovered that these mutations are commonly found in high-grade gliomas located in the midline of the brainstem and occur at a high frequency in DIPG cases.

The driving force behind the development of DIPG is the substitution of a specific amino acid called lysine at position 27 in the H3 gene. This mutation, known as the H3K27 mutation, plays a significant role in the oncogenesis of DIPG. [9] It is essential to note that the mutation can occur in any variant of the histone H3 gene, whether it is H3.1, H3.2, or H3.3. Regardless of the variant affected, the mutation could lead to the development of DIPG.

The H3K27 mutation and the subsequent loss of a particular epigenetic mark called H3K27me3 are crucial in understanding DIPG. These molecular changes are characteristic of DIPG and likely represent the first genomic event that triggers the transformation of normal cells into cancerous cells in these tumors. As a result, DIPG can be viewed as a uniform group of tumors categorized as glial neoplasms, characterized by a distinct epigenetic alteration resulting from the histone H3 mutation.

Interestingly, all the H3K27 mutations identified in DIPG have similar epigenomic consequences on a polycomb repressive complex 2 (PRC2) complex, despite the distinct functions and genomic distribution of the different histone H3 variants. This suggests that the mutation leads to a shared epigenetic disruption, further highlighting the importance of these alterations in DIPG. [10]

1.5   Problem Statement:

While it is well-established that the H3K27M mutation perturbs the function of the Polycomb Repressive Complex 2 (PRC2), a critical epigenetic regulator, the precise molecular mechanisms governing this interaction remain incompletely understood. Specifically, the intricate interplay between the PRC2 complex and the H3K27M mutation, and how this interaction contributes to the growth and dissemination of DIPG tumors, has not been fully elucidated.

1.6   Problem Solution:

The proposed solution for the above problem is to understand the molecular mechanistic of H3K27M and PRC2 complex interaction to unveil how this mutation induces conformational changes in PRC2 complex.

1.7   Objectives:

The goals of the research are:

- ✓ To evaluate the impact of H3K27M mutation on conformational changes of the PRC2 complex,
- ✓ How the binding free energies of the H3K27M mutant-induced conformations compare to those of the wild-type H3K27me3.

By achieving these objectives, we aim to comprehensively understand the molecular alterations caused by the H3K27M mutation and their functional consequences in DIPG.

CHAPTER 2

LITTERATURE REVIEW

2   Literature Review

2.1   Role of Epigenetics in DIPG:

Epigenetics refers to the study of heritable changes in gene expression or cellular phenotype that occurs without alterations to the underlying DNA sequence [11]. It involves modifications to DNA and its associated proteins, which can influence gene activity and gene regulation. Epigenetics play a major role in the development and progression of DIPG. It involves modifications to DNA and its associated proteins that can influence gene activity [12].
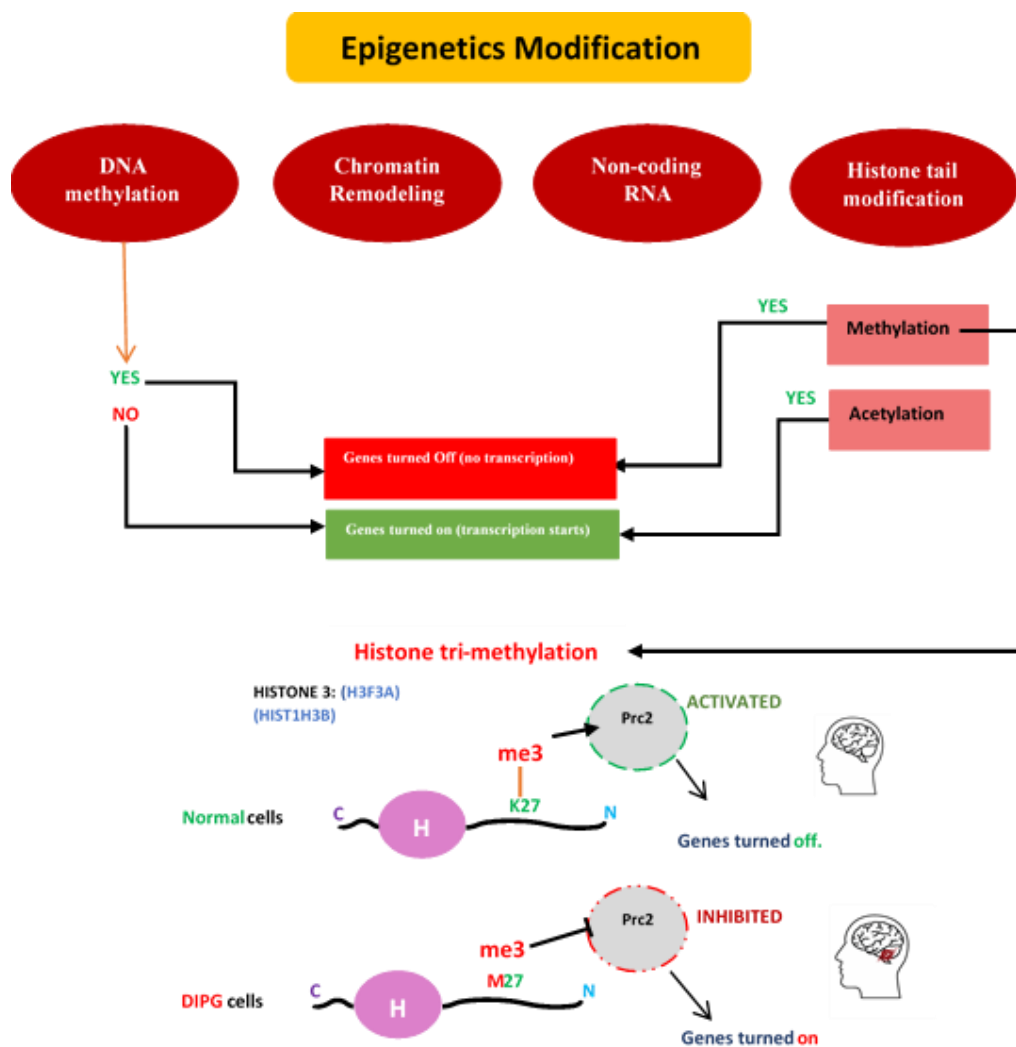
Figure 2.1 Role of Epigenetics in DIPG [13]

The figure 2.1 is depicting the role of epigenetic modifications in DIPG and how Histone methylation is linked with the occurrence of DIPG [13].

2.2   Structure and Function of PRC2 Complex:

Epigenetic regulation, including DNA methylation and de-methylation, histone modification, chromatin remodeling, incorporation of histone variants and chromatin modulation by non-coding RNAs, play a key role in modulating the chromatin state and gene expression without altering the DNA sequence. Polycomb group proteins (PcGs) are important epigenetic regulators that form large multimeric protein complexes with diverse catalytic activities and cellular roles. Among these, polycomb repressive complexes 1 and 2 (PRC1 and PRC2) have been extensively studied. PRC1 utilizes its catalytic subunit, either RING1A or RING1B ubiquitin ligase, to mono-ubiquitylate lysine 119 on histone H2A (H2AK119ub1). On the other hand, PRC2 catalyzes mono-, di-, and tri-methylation on lysine 27 of histone H3 (H3K27me1, H3K27me2, and H3K27me3). Both PRC1 and PRC2 exhibit various forms with distinct subunit compositions, potentially influencing the complexes' structure and mechanisms. [13]

2.3   PRC2 Composition and Activation:

 PRC2 complex is a distinctive molecular structure profoundly involved in epigenetic regulation across various eukaryotic cells. With its impact on over 2000 genes distributed among 10 chromosomes, the PRC2 complex plays a pivotal role in cellular development and differentiation, as underscored by experiments on mice where the absence of core subunits resulted in deleterious morphological defects and premature deaths due to transcription of damaging proteins. [14] PRC2 is a protein complex found in various organisms that plays a role in controlling gene expression. It adds a chemical mark called trimethylated K27 to a protein called histone, which is involved in packaging DNA. This mark is associated with gene silencing, meaning it can turn off the expression of specific genes. [14]

2.3.1   Structure of PRC2 complex:

PRC2 is comprised of three core subunits: enhancer of zeste homolog 1 or 2 (EZH1/2), embryonic ectoderm development (EED) protein, and suppressor of zeste 12(SUZ12). The PRC2

core complex is responsible for the trimethylation of Lysine 27 on histone 3 (H3K27), leading to the H3K27me3 modification on the nucleosome. By preserving histone modifications throughout the cell cycle, including cytokinesis, the PRC2 prevents aberrant transcriptional activity, which could be severely detrimental. These subunits work together to carry out the function of PRC2. EZH2 is initially inactive but becomes active after interacting with EED and SUZ12. EED can also interact with the product of PRC2, known as H3K27me3, to activate PRC2 allosterically. Through various experimental studies and structural analysis, it was elaborated that how PRC2 works [15]. The interaction between Ezh2, EED, and Suz12 is very close, with Ezh2 enveloping EED and Suz12, making connections with both subunits. The catalytic part of PRC2 involves specific regions of Ezh2, including the SET activation loop and SET. Notably, Ezh2 possesses a flexible region called the stimulation-responsive motif (SRM) that responds to the presence of H3K27me3. The interaction of SRM with H3K27me3 and EED induces conformational changes, which triggers a communication pathway that ultimately enhances the catalytic activity of PRC2 complex. [16]
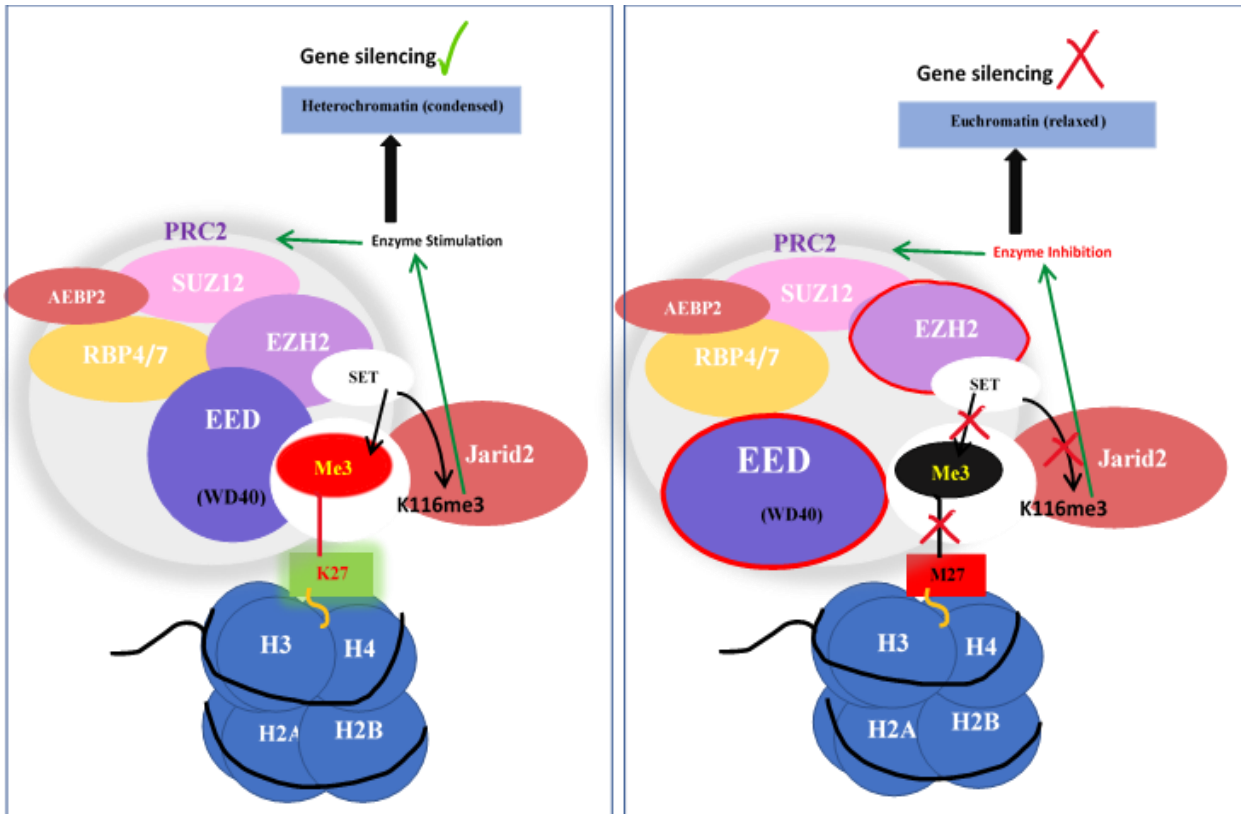
Figure 2.2 Mode of action of w.t and mutant peptide on PRC2 complex [13]

2.3.1.1   Subdomains of Core Components:

Each core component of PRC2 complex has subdomains that play different crucial roles in PRC2's activity. These subdomains play different roles in stabilizing the PRC2 complex by forming extensive intra and inter subdomain interactions.

EZH2	
N-terminal	
SBD(SANT1L binding domain)	Forms intermolecular interactions with SANT1 domain and tightens the EZH2 belt-like structure around EED like a buckle.
EBD(EED binding domain)	Occupies surface groove across the bottom face of EED WD40 domain.
BAM (β-addition motif)	It is the continuation of EBD on the side face of EED and consists of 3 β strands that are added to β-propeller fold of WD40 repeats.
SAL (SET activation loop)	It is the Ezh2 loop region that extends away from EED surface to the back of SET domain and is directly connected to SRM.

SRM(stimulation responsive motif)	It sits on the stimulating peptide at the center of top face of EED and forms a sandwich like assembly with EED and H3K27me3 peptide.
SANT1	It is enriched with bulky aromatic and hydrophobic residues
C-terminal	
MCSS (motif connecting SANT1 and SANT2)	It connects SANT1L with SANT2L. It harbors two helices. One is Zinc binding motif and the second one is β -hairpin.
SANT2	It coordinates a zinc ion as well.
CXC(cysteine rich domain)	It makes strong interactions with the Vefs helical and MCSS domains on one side and the SET domain on the other.
SET	It is positioned above EED and Vefs domain adjacent to SRM. It is the catalytic domain of EZH2. It is in the autoinhibited conformation which is relieved whenEzh2 associates with EED and Suz12.

Table 1 Details of each Subdomain
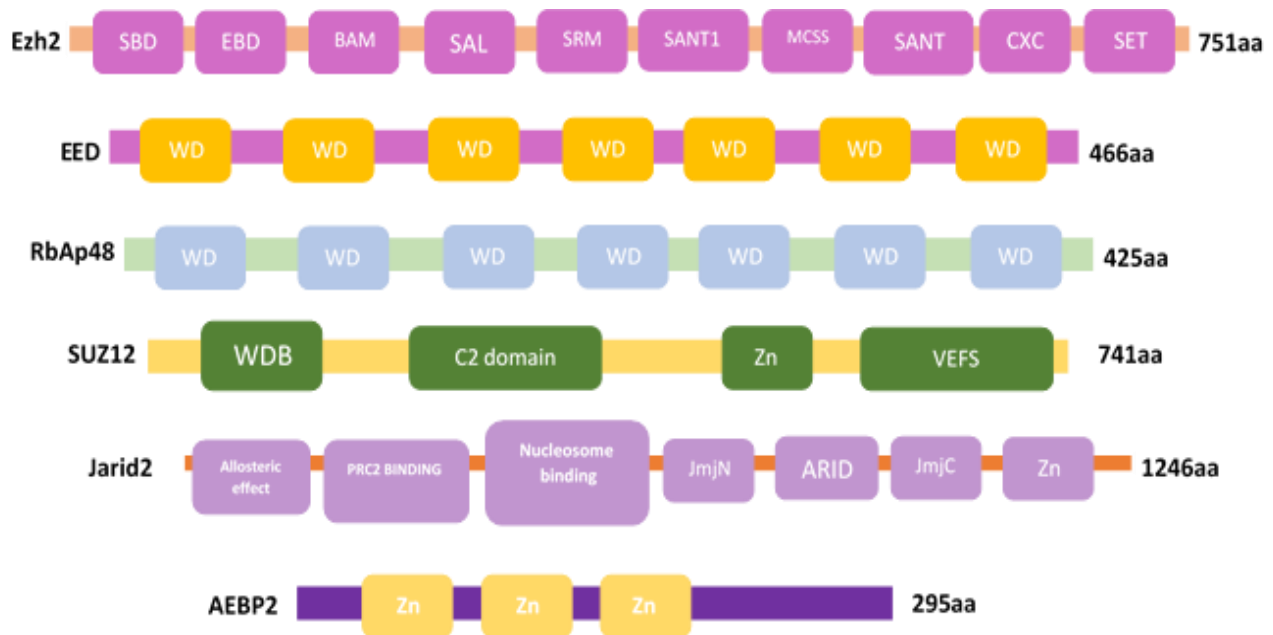
Figure 2.3 Subdomains of Core Components [16]

2.3.1.2   Active and Basal States:

H3K27me3 peptide interaction with SRM brings SRM in structured format and allows it to bind to the catalytic SET-1 domain of EZH2. These coupled interactions are associated with a 20⁰ counterclockwise rotation of SET-1 domain and opening of substrate-binding cleft, resulting in PRC2 activity. H3K27me3 peptide interaction with SRM brings SRM in structured format and

allows it to bind to the catalytic SET-1 domain of EZH2. These coupled interactions are associated with a 20º counterclockwise rotation of SET-1 domain and opening of substrate-binding cleft, resulting in PRC2 activity.

Compact Active State	JARID2 K116me3 sits in the middle of an aromatic cage (F97, Y148, and Y365) with hydrogen-bond interactions (R414:F117 and W364:R115) that stabilize the JARID2 peptide backbone.	Additional interaction between EZH2 D136 and the backbone amide of JARID2 K116 helps position the EZH2 SRM helix next to the EZH2 SET domain. The helical SRM forms contacts with the SET domain.
		SBD is bent and SANT1 is packed against the SBD. SRM is ordered and adopts an alpha helical structure, and it is stably bound.
Basal State	No binding of any cofactor	EBD/SANT1 straight and SRM motif is mobile and disordered. SBD-SANT1 module stands tall over the EED. SBD is straight and SBD-SANT1 module stands tall over the EED

Table 2 Different states of PRC2 Complex

2.4   Function of PRC2 complex:

The primary function of PRC2 is to add a methyl group as a chemical tag to histone H3. This chemical tag is added to a particular amino acid called lysine 27 (K27) in histone H3. The enzyme EZH2, part of the PRC2 complex, adds this methyl group to K27 [17]. However, EZH2 alone has limited activity and only works at a partial capacity by itself. It must combine two other proteins, EED and Suz12, to form the complete PRC2 complex. When EED and Suz12 join EZH2, they activate their full potential. In the absence of EED and Suz12, EZH2 is auto-inhibited, meaning its activity is reduced. Once EZH2 adds three methyl groups to K27 to make H3K27me3, EED binds to the new histone through a part of its structure called the WD40 domain. This interaction changes the shape of EZH2 and enhances its catalytic activity. This change is called allosteric activation. PRC2 can then move to other nearby histones and add the H3K27me3 mark. EED plays a role in this process by allowing PRC2 to bind to the H3K27me3   domain, helps stabilize the PRC2

complex. PRC2 also relies on other proteins, RBBP4 and RBBP7, to bind to unmodified histones and carry out its full methyltransferases activity. [16] [18]

In short, PRC2 is a group of proteins that adds methyl groups to histone H3 at K27. For its full activity, it needs EZH2, EED, and Suz12. EED helps PRC2 bind to its product, H3K27me3, and Suz12 stabilizes the complex. RBBP4 and RBBP7 are essential for PRC2's binding to unmodified histones. When a specific mutation happens in the amino acid K27, which stops methyl groups from being added, it breaks the link between EED and the changed histone, which causes the H3K27me3 mark to be lost. This loss might result in the activation of genes and contribute to the development of cancer.

2.5   Mutations in DIPG:

Mutation in H3 proteins were identified as most prominent driver mutations in DIPG. Two studies initially reported an unexpected finding of genetic mutations in histone genes in high-grade gliomas. In one study, the genomes of seven diffuse intrinsic pontine glioma (DIPG) tumors were sequenced, along with targeted sequencing of additional 43 DIPG tumors and 36 pediatric glioblastomas outside the brainstem. It was discovered that 78% of DIPG tumors and 22% of non-brainstem pediatric glioblastomas had mutations in either the H3F3A or HIST1H3B gene, which are responsible for encoding specific types of histone proteins [19]. HIST1H3B encodes canonical histone H3 (H3.1 and H3.2), while H3F3A encodes the variant histone H3.3 [20] [21]. In all cases of DIPG, mutations in H3F3A and HIST1H3B resulted in a substitution of lysine 27 with methionine (K27M) in histone H3. Another study involved sequencing exosomes from 48 pediatric glioblastomas, revealing somatic mutations in histone genes, including H3.3K27M and H3.3G34R/V, in 31% of the cases [22]. The modification of lysine 27 on histone proteins plays a crucial role in regulating both active gene transcription and gene silencing. This residue can be mono-, di-, or tri-methylated (H3K27me1-3) by the Polycomb Repressive Complex 2 (PRC2) through its catalytic subunit Ezh1/2 [14] [23]
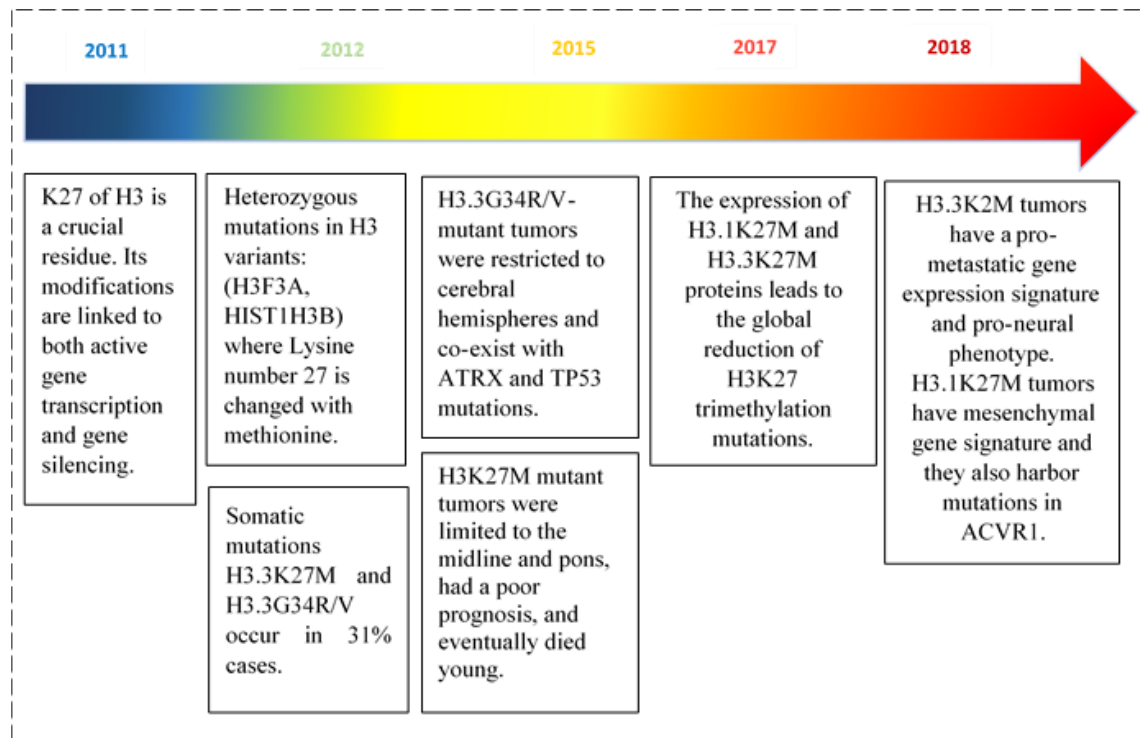
Figure 2.4 Mutations in DIPG Reported in successive studies

2.6 H3K27 methylation and PRC2 complex induced gene silencing:

The H3K27M mutation results in a conformational change in the PRC2 complex which is responsible for gene regulation and epigenetic control of gene expression. The mutation causes a structural change in the PRC2 complex which affects its ability to bind to target genes and regulate their expression. This results in the activation of genes which are normally repressed by PRC2 and thus leads to the formation of tumor in the brainstem of children. Various models have been studied to understand how the H3K27M mutation contributes to tumor development. It has been found that H3K27M alone is not sufficient to promote tumorigenesis [15]. Instead, it must cooperate with other modifications, such as the up-regulation of PDGFRA and deletion of TP53, to efficiently generate tumors. When H3K27M mutant genes were expressed in human pluripotent stem cells, they showed a moderate increase in proliferation. However, the mutant cells exhibited increased proliferation only when combined with p53 depletion and PDGFRA overexpression. Similarly, in mouse neural stem cells with H3.3K27M, Trp53 mutations, and PDGFRA

overexpression, tumors were formed with gene expression patterns similar to human DIPG cells [24].

Further investigation into the distribution of H3K27M mutant proteins and Ezh2 suggests that multiple mechanisms are responsible for the decrease in H3K27M methylation. H3K27M mutant proteins are found abundantly at promoters and actively transcribed genes where the PRC2 complex is absent or reduced, similar to wild-type H3.3. Paradoxically, levels of H3K27me3 are still present at hundreds of genomic locations. These opposite changes need further investigation to understand how H3K27M selectively retains H3K27me3 at specific genes in vivo. It is proposed that the high amount of H3K27M sequesters PRC2 at poised enhancers but not at active promoters, leading to the global loss of H3K27me3.
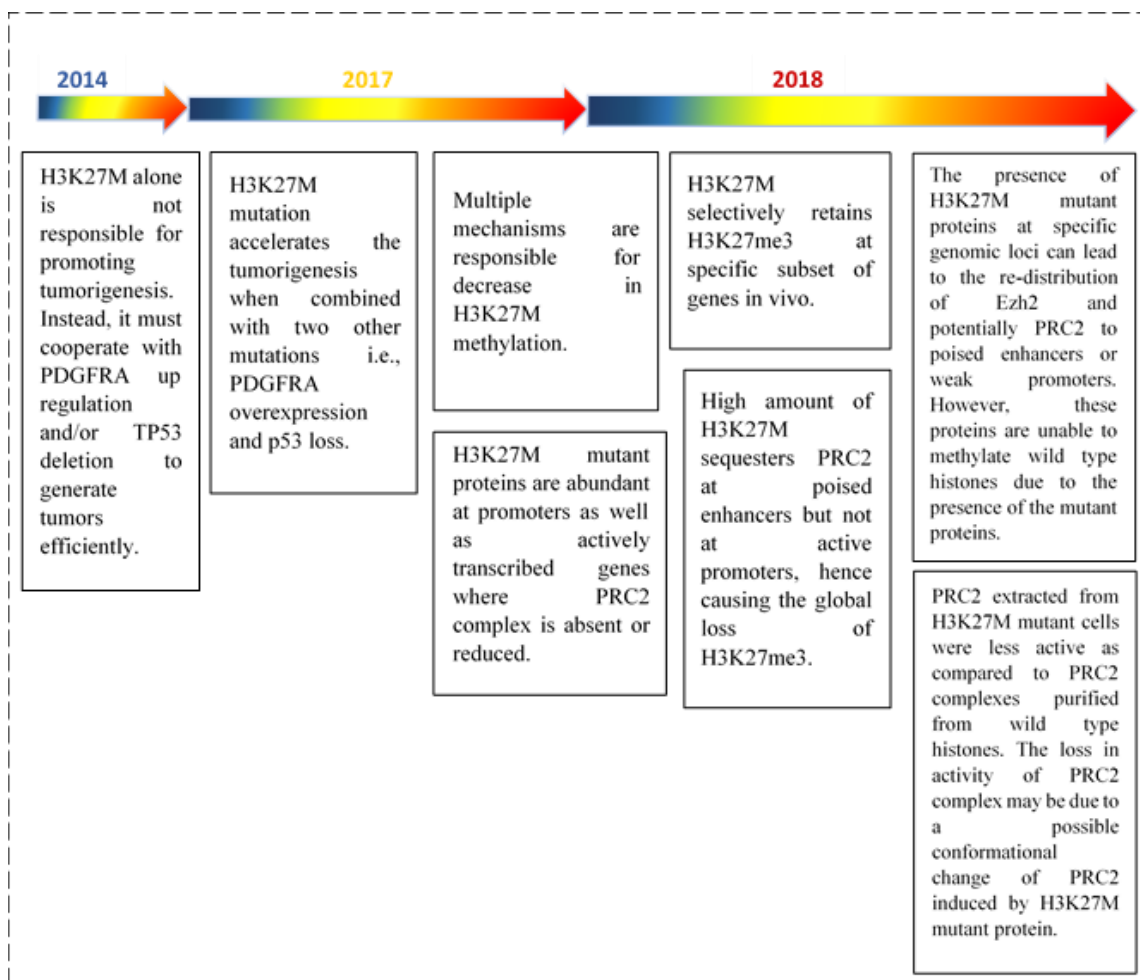

Figure 2.5 H3K27M and PRC2 complex in DIPG [23] [24] [25] [26]

Moreover, H3K27M mutant proteins are few or absent at strong PRC2 sites with H3K27 methylation. This raises doubts that the global loss of H3K27 methylation is solely due to PRC2 inhibition. Studies using mouse embryonic stem cells have shown that Ezh2 is present at poised enhancers and weak promoters in H3K27M mutant cells compared to wild-type cells. However, the levels of H3K27 methylation in H3K27 mutant cells are low despite the presence of Ezh2 at these sites. These findings suggest that the presence of H3K27M mutant proteins causes the redistribution of Ezh2 and possibly PRC2 to poised enhancers and weak promoters. Still, they are unable to methylate wild-type histones. Another study hypothesized that PRC2 extracted from H3K27M mutant cells is less active than PRC2 complexes purified from wild-type histones. The decrease in the activity of PRC2 may be due to a possible change in the shape of PRC2 induced by the H3K27M mutant protein [24].

In conclusion, the existing literature provides a strong foundation for further research into the epigenetic mechanisms driving DIPG, with a particular focus on the H3K27M mutation and its interaction with the PRC2 complex. Continued investigation into the molecular intricacies of this interaction, coupled with translational research and collaborative initiatives, holds promise for improving the prognosis and treatment options for DIPG patients in the future.

CHAPTER 3

METHODOLOGY

3   Methodology

3.1   Protein Structure Collection:

To run MD simulations, RCSB Protein Data Bank (PDB) database is being used to search for the three dimensional crystal structure of protein of interest which in our case is PRC2 complex. PRC2 structure with PDB Id "5HYN" [15] was being chosen for simulations which is determined by X-ray diffraction. 5HYN contains 3 core subunits of PRC2 i.e., EZH2, EED and SUZ12. It also contains oncohistone, mutant peptide i.e. H3K27M where Lysine number 27 of wild-type peptide is being replaced by Methionine number 27 and JARID2 K116me3 which mimics the functioning of H3K27me3 peptide [15]. 5HYN also contains zinc ions and co-factor SAH. 5HYN is present in oligomer state, so there are total 4 complexes, each containing 3 core subunits of PRC2 complex and H3K27M, JARID2K116me3 peptides [15].
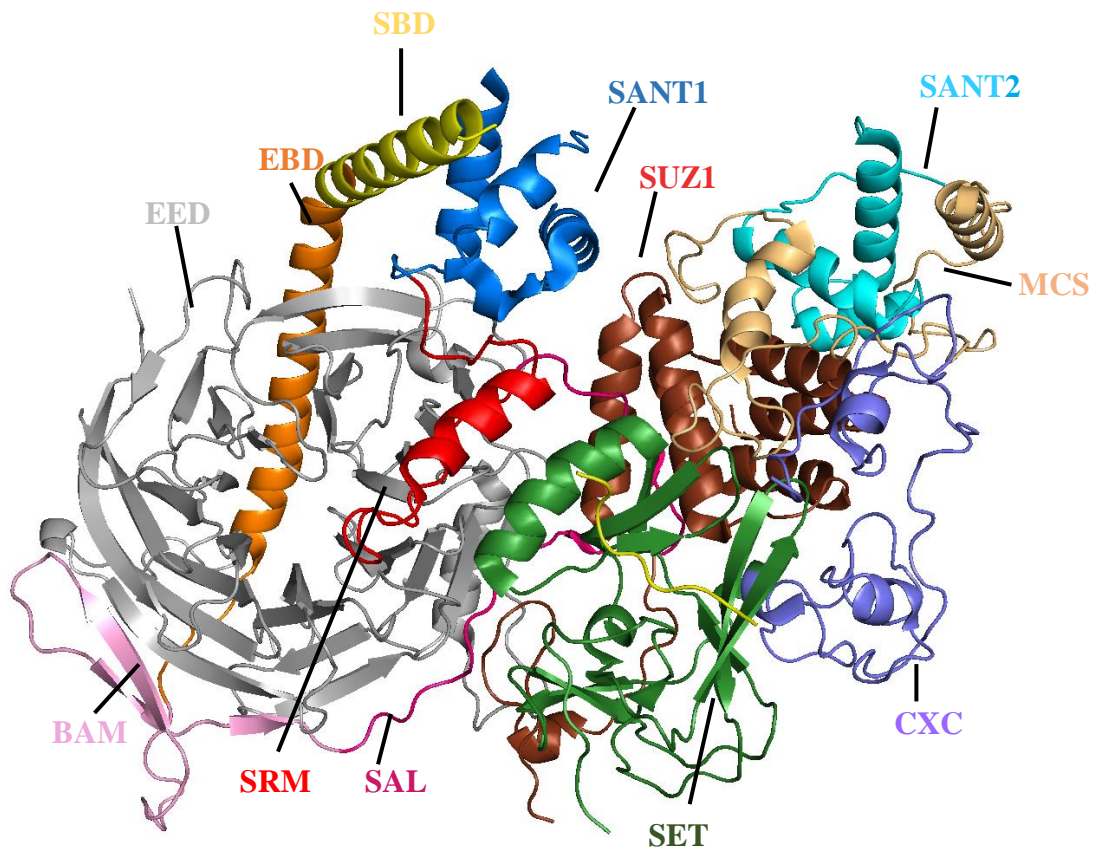
Figure 3.1 Positions of subdomains of PRC2 Complex [15]

Figure 3.1 is the structure of PRC2 complex with three important subunits of PRC2 complex that are EZH2, EED and SUZ12Vefs and 10 distinct subdomains of EZH2 represented with different colors according to the color scheme in the bottom part of figure 3.2. In the top part of figure 3.2, PRC2 is divided into three different lobes according to their contribution towards the function of PRC2 complex [16].
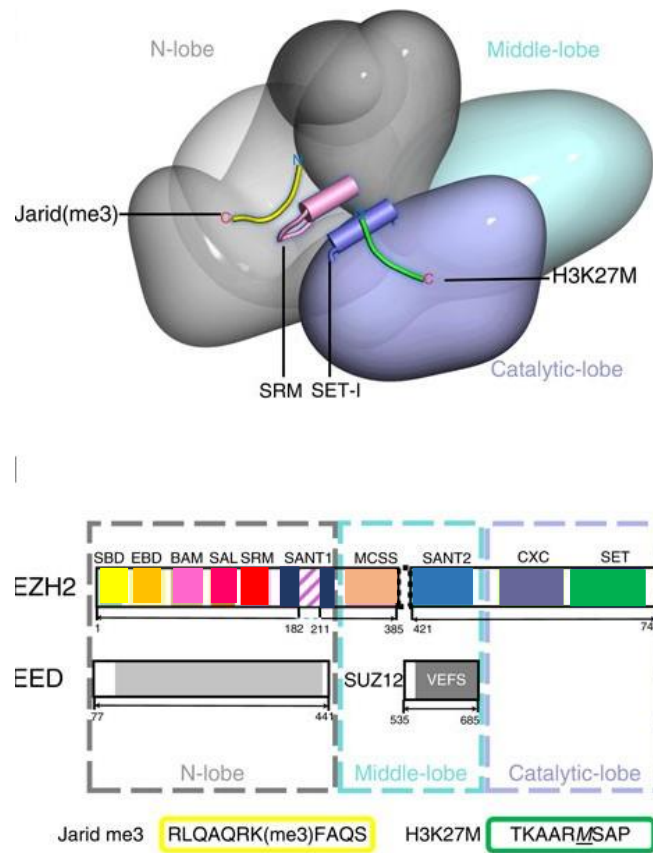
Figure 3.2 Schematic representation of 5HYN subdomains

Figure 3.3 shows the residues of EED that are present in aromatic cage that initially recognizes the already tri-methylated repressive histone mark [15].
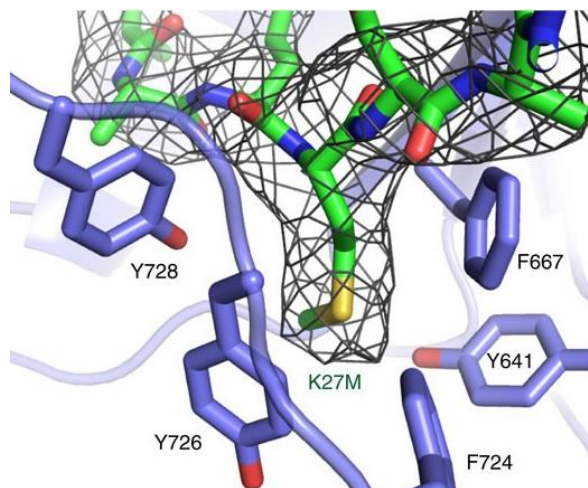
Figure 3.3 EED aromatic cage surrounding H3K27M

3.2   Structure Preprocessing:

- **Mutant peptide-PRC2 complex simulation**

The three dimensional crystal structure of PRC2 complex (5HYN), downloaded from RSCB PDB was then preprocessed because raw structure can't be used for MD simulations as it has so many missing hydrogens, residues, loops and sidechains. Firstly, only the 4 chains i.e., chain A (Ezh2), chain B (EED), chain C (SUZ12), chain D (H3K27M peptide) were retained along with six zinc ions and one cofactor SAH. Then missing atoms, residues and side chains are fixed using 'Protein Preparation Wizard'.

Water molecules and other heteroatoms are removed that are not part of the protein. Correct bond orders and hydrogen positions are assigned. pk_a states to ionizable residues are assigned [25] [26]. After preprocessing and reviewing and modifying the structure, it was further refined by optimizing its hydroxyl, Asn, Gln and His states using ProtAssign.

3.3   Defining Periodic boundary conditions:

The concept of Periodic boundary conditions is fundamental in Molecular Dynamic (MD) simulations. Periodic boundary conditions are applied to simulate a system (e.g. a protein in a water box) as if it were part of an infinitely repeating system. This allows to study local interactions

while avoiding edge effects and accurately capturing bulk-like behavior. PBC helps to simulate larger systems without needing an excessively large computational domain. This is especially advantageous when studying systems that involve a large number of molecules, such as biological macromolecules in solvent. Applying PBC prior to running md simulations usually involves adding a simulation box. In that simulation box a solvent, usually water is added. In our case, Desmond's System builder package is being used to define periodic boundary condition. Firstly, the solvent/water model TIP3P is selected. The TIP3P model specifies a 3-site rigid water molecule with charges and Lennard –Jones parameters assigned to each of the 3 atoms. After selecting water model, the shape and dimensions of simulation box are defined. In our case, orthorhombic box shape has been chosen. To ensure that there is enough buffer space around the protein to prevent interactions with periodic images, the distance is kept at 15 Å. In the end, the box volume is minimized to complement with the volume of protein. The overall charge of the system was 5. To neutralize the system, 5 counter ions Cl- are added to the system. In the end, 0.15M salt has been added to the system. OPLS4 force field is chosen to run the job. The equation of OPLS4 force field is consisted of different energy terms where

E_{bonds} = the bond energy, E_{angle} = angle energy, $E_{torsion}$= torsion energy, $E_{nonbonded}$ = non bonded energy.

$$E_{bonds} = \sum_{i} k_{b,i}(r_i - r_{o,i})^2$$

$$E_{angles} = \sum_{i} k_{b,i}(\theta_i - \theta_{o,i})^2$$

$$E_{torsion} = \sum_{i} \left[\frac{1}{2}V_{1,i}(1 + \cos \varphi_i) + \frac{1}{2}V_{2,i}(1 + \cos 2\varphi_i) + \frac{1}{2}V_{3,i}(1 + \cos 3\varphi_i) + \frac{1}{2}V_{4,i}(1 + \cos 4\varphi_i) \right]$$

$$E_{nonbond} = \sum_{i} \sum_{j>i} \left\{ \frac{q_i q_j e^2}{r_{ij}} + 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \right\}$$

Equation 1

3.4   Energy Minimization:

Prior to starting the MD simulations, energy minimization step is being performed. This step is important to identify the proper molecular arrangement in space since the drawn protein structures are not energetically favorable [27]. The potential energy of a molecule contains different energy components like stretching, bending, and torsion; hence, when an energy minimization job is run, it will immediately reach a minimum local energy value, and it might stop if the employed program is not exhaustive.
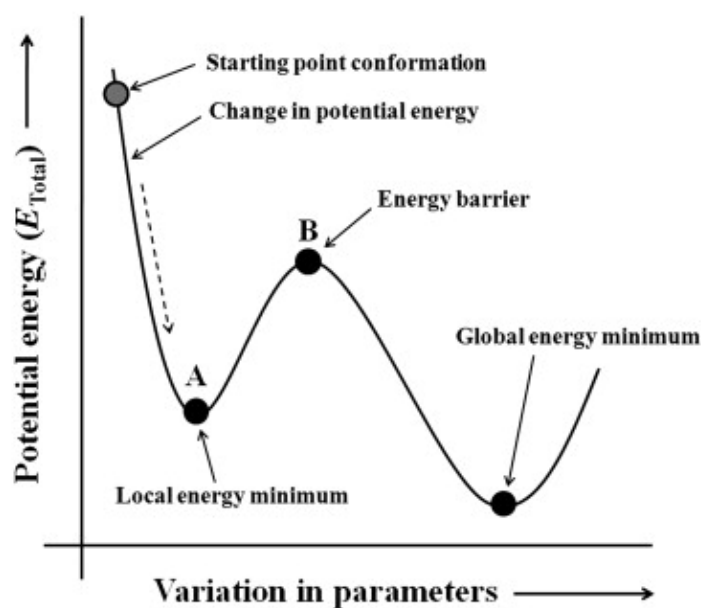
Figure 3.4 Energy Minimization Concept Diagram [27]

Moreover, most MD simulations start with an initial configuration of atoms that might not be physically realistic or might have steric clashes (atoms overlapping). Energy minimization helps to alleviate these clashes and create a more reasonable starting structure for the simulation. Energy minimization also ensures that system starts from a stable configuration, avoiding any immediate high-energy interactions that could lead to unrealistic dynamics. So, to energy minimize our protein of interest, desmond's energy minimization tool is being used in order to relax the protein and remove steric clashes. 'System Preparation' panel is being used to setup the energy minimization job.

3.5   System Equilibrium:

Equilibration involves allowing the system to adjust and stabilize before collecting actual simulation data. This step is critical for producing accurate and meaningful results in an MD simulation. There are two main types of equilibration:

- **NVT Equilibration:** In this step, the system's volume, number of particles, and temperature (NVT ensemble) are controlled. It allows the system to reach a state where the temperature is constant and atoms have appropriate kinetic energies. This is important to mimic the conditions of a real experiment where temperature is controlled.

- **NPT Equilibration**: In addition to controlling the temperature, NPT equilibration, also controls pressure along with volume, creating a constant pressure and temperature (NPT ensemble) environment. This step ensures that the system's density and pressure are at realistic values.

Reasons for Equilibration:

- **Thermalization:** Equilibration allows the system to adjust its energy distribution and achieve the desired temperature. Without equilibration, the initial high-energy state might lead to unrealistic dynamics.

- **Statistical Mechanics:** Equilibration is important for the system to sample various energy states and reach thermodynamic equilibrium. This ensures that the simulation provides meaningful statistical averages and properties.

- **Avoiding Artifacts:** Without equilibration, the simulation might contain artifacts due to the initial conditions, which could affect the interpretation of the simulation results.

- **Accurate Data Collection:** Equilibrated systems provide a more stable starting point for data collection. Starting simulations from an equilibrated state leads to more accurate and reproducible results.

In our case, to bring the system to the desired temperature and pressure, NPT ensemble is being chosen and temperature and pressure setting are specified. In the end, integration settings (time

step, constraints) and relaxation protocols are specified. 300 K temperature and 1.01325 bar pressure is specified for our particular system.

3.6   MD simulations

Lastly, the production of MD simulation is configured. In this step, the simulation time, time step, constraints, and other parameters are defined. Output frequencies for trajectory snapshots and other data are specified. For simulations, 100 nanoseconds simulation time is defined. Recording Interval for trajectory is defined to be 9.6 ps. For this particular recording interval, the approximate number of frames are 10416. To start the simulation run, the proposed job is submitted to Desmond using supercomputer.

3.7   Simulation Results Analysis:

After the completion of simulations, the trajectory is analyzed using Desmond's analysis tool called 'Simulation Interaction Diagram'. The simulation Interaction Diagram tool showed the RMSD, RMSF and other graphs for the analysis.

- **Wild-type peptide – PRC2 complex Simulations:**

For the simulations of wild-type peptide i.e. H3K27me3 and PRC2 complex, the same structure with PDB id 5HYN is used because the WT-PRC2 complex structures are not available on PDB. So, 5HYN is used by replacing the mutated residue methionine with the wild type tri-methylated residue Lysine. This specific job is done using Pymol software [28]. This structure is then opened in Schrodinger maestro and is simulated using the steps same as used for mutant peptide above [25].

3.8   Binding Free Energy Analysis:

Binding free energy refers to the energy change associated with the formation of a complex between two or more molecules, such as a ligand (e.g., a drug or small molecule) and a receptor (e.g., a protein or enzyme). It quantifies the strength of the interaction between the molecules and indicates whether the binding process is energetically favorable or unfavorable. Binding free

energy, often denoted as ΔG_bind (Delta G bind), represents the difference in the Gibbs free energy (ΔG) between the unbound state (isolated molecules) and the bound state (molecules in complex). It is a measure of the stability or affinity of the complex formation.

Thermodynamics: Binding free energy is a thermodynamic quantity, and it takes into account various energy contributions, including:

Table 3 Different Types of Energies

Different Types of Energies	
Columbic Interactions:	Electrostatic interactions between charged molecules.
Van der Waals Interactions:	Attractive forces between molecules due to fluctuations in electron density.
Hydrogen Bonding:	Specific, directional interactions between hydrogen and electronegative atoms (e.g., oxygen or nitrogen).
Hydrophobic Interactions:	Interactions between nonpolar molecules or regions that tend to cluster together in aqueous environments.
Covalent Interactions:	Formation or breaking of chemical bonds, if applicable.
Solvation Effects:	Changes in the energy due to interactions with solvent molecules.

Negative Values: A negative binding free energy (ΔG_bind < 0) indicates that the binding process is energetically favorable. In other words, the complex formation releases energy, and the molecules tend to stay bound together.

Positive Values: A positive binding free energy (ΔG_bind > 0) suggests that the binding process is energetically unfavorable. This means that the molecules are less stable in the complex than in their isolated states, and they tend to dissociate.

Zero Value: A ∆G_bind of zero indicates that the binding is energetically neutral. The complex formation neither releases nor consumes energy.

$$\Delta Gbind = \Delta EMM + \Delta Gsolv + \Delta GSA$$

Equation 2

Where ∆Gbind represents the change in free energy associated with the binding of two molecules. It quantifies the energy difference between the bound state (ligand and receptor together) and the unbound state (ligand and receptor separate). ∆EMM is the change in potential energy calculated using a molecular mechanics force field. It accounts for the energy changes associated with the internal molecular geometry and bonding. ∆Gsolv represents the change in energy due to the solvation or desolvation of the ligand and receptor in a solvent (typically water). ∆GSA is the change in energy associated with the change in the accessible surface area of the ligand and receptor upon binding.

MMGBSA Binding Energy Calculations:

Molecular mechanics-generalized Born surface area (MM-GBSA) method in Prime was used was used for the calculation of binding free energy of Protein-Peptide complexes [29]. The binding free energy ∆G_bind was determined according to the following equation:

∆Gbind = E_complex (minimized) – E_ligand (minimized) – E_receptor minimized

Equation 3

To calculate ∆G_bind using this equation, MMGBSA method uses following steps:

- Minimize the energy of the ligand in isolation (E_ligand (minimized)).
- Minimize the energy of the receptor in isolation (E_receptor (minimized)).
- Create the ligand-receptor complex and minimize its energy (E_complex (minimized)).
- Calculate ∆G_bind using the equation 3.

The result ΔG_bind represents the binding free energy, and its sign will tell whether the binding is favorable (negative) or unfavorable (positive). A more negative ΔG_bind value indicates a stronger binding interaction

CHAPTER: 4

RESULTS

4   Results

4.1   RMSD of catalytic domain of PRC2:

As stated in literature [30], the catalytic domain of PRC2 complex is not stable and active alone in the absence of other subunits of PRC2 complex. Simulation of the catalytic domain (EZH2-SET) without any other component i.e. EED and SUZ12 were done for 200 ns. The RMSD of catalytic domain showed high values ranging from 0 – 20 Å confirming that EZH2 is highly unstable throughout the course of simulations. The RMSF plot shows that the residues between 130-170 showed high fluctuations. These residues lie in the SRM domain of EZH2 which depicts that SRM is very flexible region in the absence of EED and SUZ12. Since most of the stability/activity of EZH2 comes from the interactions between SRM and EED, it can be inferred that due to lack of these interactions, EZH2 is unstable and is inactive and cannot govern its methyltransferase activity.
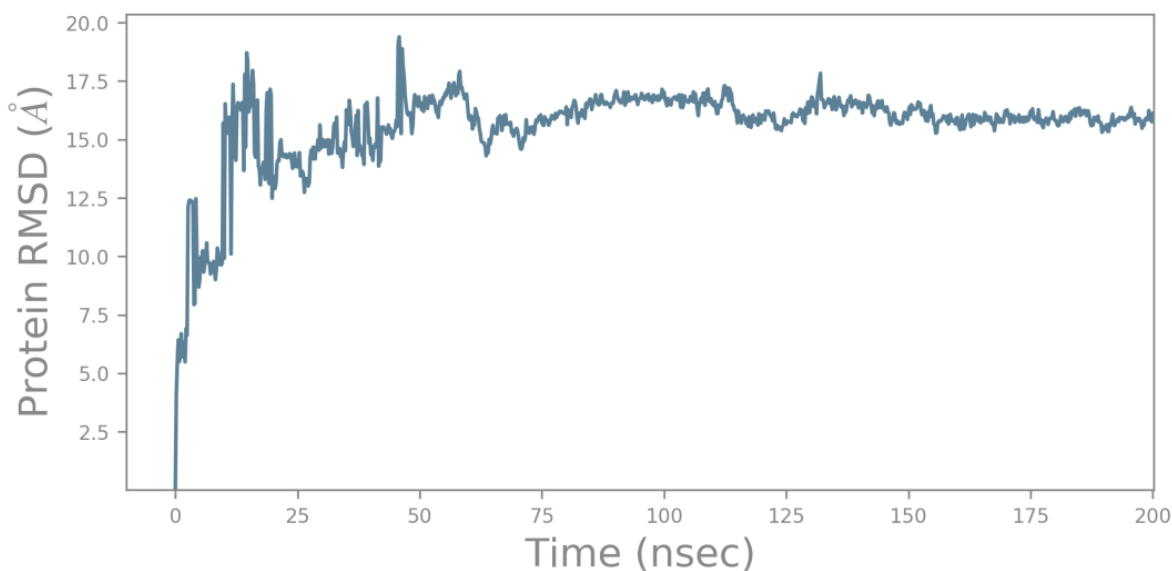
Figure 4.1 RMSD plot of catalytic Domain of PRC2
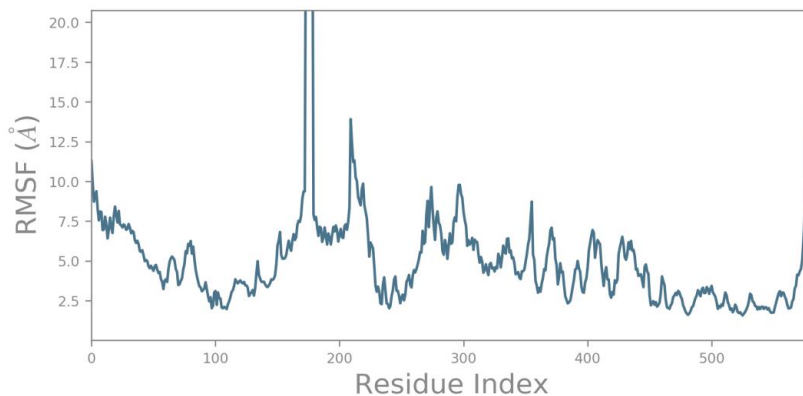
Figure 4.2 RMSF plot showing the most fluctuating residues of catalytic domain

4.2   MD Simulations of Mutant Peptide:

After running MD simulations of mutant peptide, the results are visualized by using simulation interaction diagram wizard.

4.2.1   Structural Stability Analysis of H3K27M-PRC2 complex:
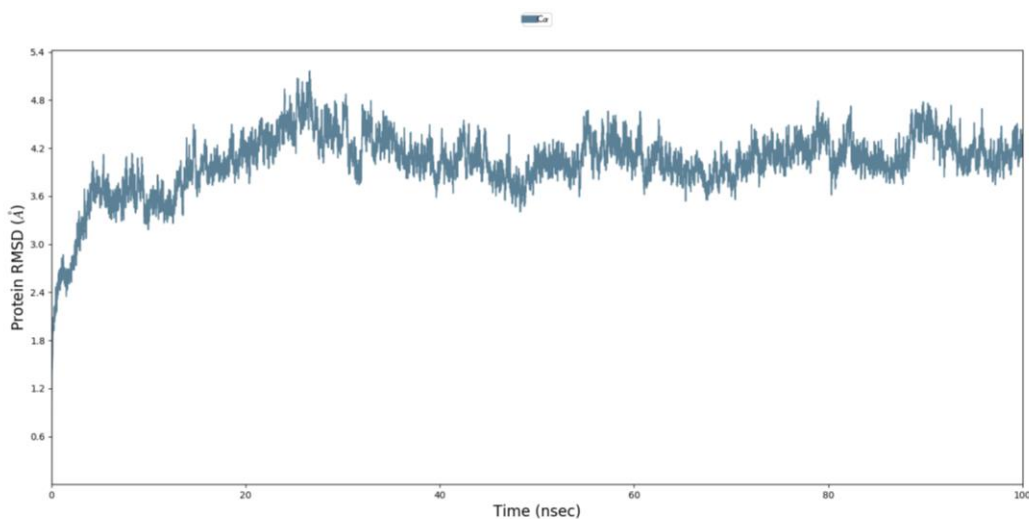
Figure 4.3 RMSD Plot of H3K27M-PRC2 complex

The RMSD plot depicts the changes in conformations of the protein-mutant peptide complex with time. The RMSD values are ranging from 1.95 Å (lowest) to 5 Å (highest). These values showed structural instability due to the presence of mutant peptide in complex with PRC2 complex.

Moreover, by the end of simulations at around 95-100 nseconds, SBD subdomain of EZH2 got straightened which is usually bent in active state and a fold of SRM domain disappeared. To get more clarity, we extracted 1st and last frame after simulations and superimposed them using chimera. The difference between these frames is shown in figure 4.4 and 4.5. These structural changes imply that the fold of SRM consisting of residues 135-141, might be important for the catalytic activity of PRC2 complex. Furthermore, the bent SBD domain is also important for maintaining the catalytic activity of EZH2.
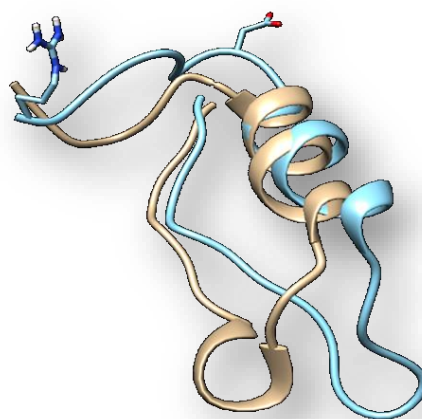
Figure 4.4 Superimposition of SRM subdomain from frame 1 and frame 10418
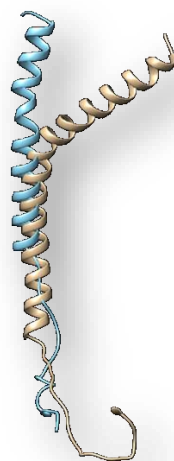
Figure 4.5 Superimposition of SRM subdomain from frame 1 and frame 10418

4.2.2   RMSF Plot of PRC2 complex:

The RMSF plot shows the local changes along the protein chain. Green lines in the plot show the residues of the PRC2 complex that are involved in the binding with the H3K27M peptide.
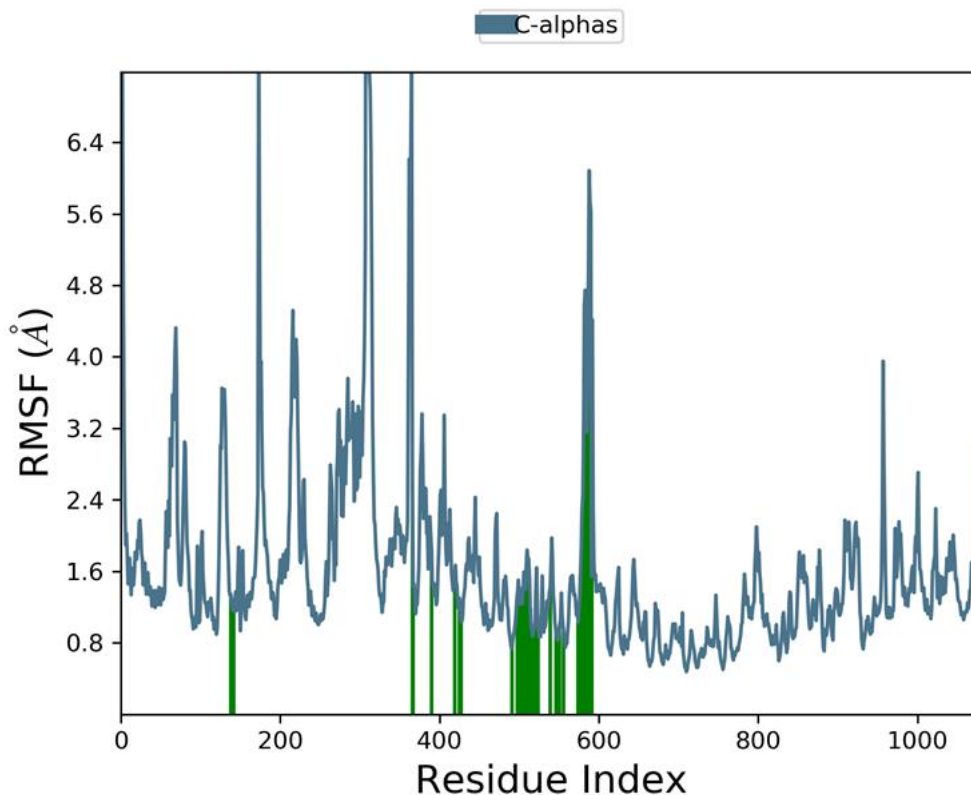
Figure 4.6 RMSF plot showing the most fluctuating residues of PRC2

The fluctuations in RMSD of protein is due to unstable residues within the protein. These unstable residues are given in the table below.

Table 4 Residues with high fluctuations

Chain	ResName	LigandContact	CA
A	Lys_10	No	9.823
A	GLY_11	No	8.191
A	PRO_12	No	7.045

A	GLU_211	No	8.060
A	SER_212	No	5.581
A	PRO_422	No	12.437
A	ASN_423	No	10.577
A	ILE_424	No	10.074
A	GLU_425	No	8.717
A	PRO_426	No	7.177
A	PRO_427	No	6.985
A	GLU_428	No	6.252
A	ILE_476	No	6.211
A	ILE_477	No	5.734
A	ALA_478	No	6.325
A	PRO_479	No	7.583

The above residues with higher RMSF values indicate greater atom fluctuations and, therefore, greater local flexibility in the protein.

4.2.3   Protein Ligand Contacts:

Protein-Ligand Contacts
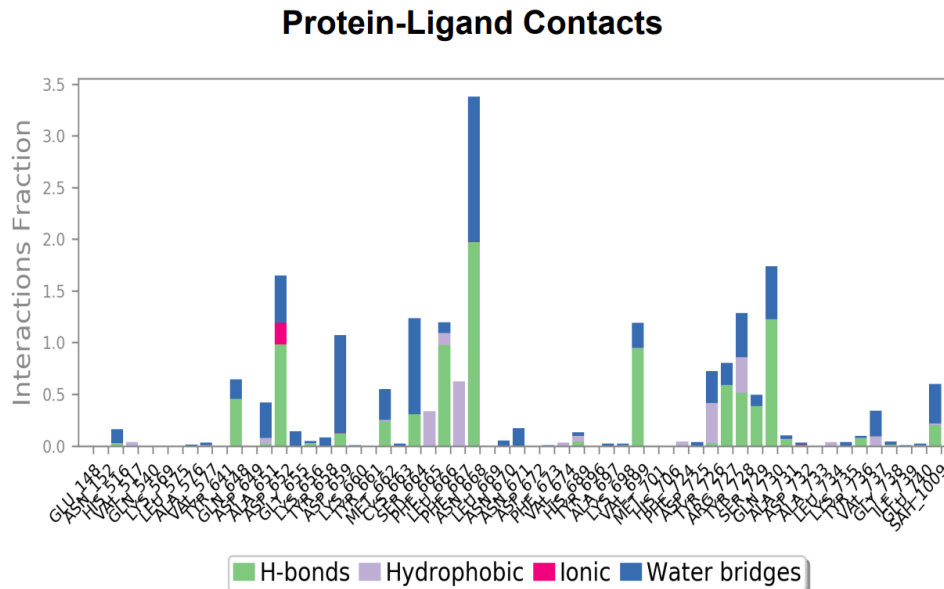
Figure 4.7 Histogram of Protein-peptide contacts

The above plot shows the PRC2 interactions with the H3K27M peptide throughout the simulation. These interactions can be categorized by type and summarized, as shown in the plot above. Protein-ligand interactions (or 'contacts') are categorized into four types: Hydrogen Bonds, Hydrophobic, Ionic and Water Bridges. In the above plot of PRC2 and H3K27M interactions, ASN 668 of chain A i.e. EZH2 is making highest interactions with the peptide. The majority of the interactions are hydrogen bonds and water mediated interactions and very few are hydrophobic. Hydrogen bonds are typically strong and highly specific, indicating that ASN is playing a crucial role in the binding affinity between PRC2 and mutant peptide. ASN's involvement in multiple types of interactions (hydrogen bonds, hydrophobic, and water-mediated) suggests that it is a key residue within the binding site of PRC2. Its interactions likely contribute to the overall binding energy and specificity of the complex. Other key residues include ASP-652, LEU-666, GLN-648, SER-664, VAL-699. Furthermore, two residues of SRM (stimulatory responsive motif) that are GLU_148 and ASN_152 are also making weak contacts with H3K27M peptide. The plot below is depicting the timeline representation of interactions and contacts (H-bonds, Hydrophobic, Ionic, water bridges) summarized in the figure 4.10. The top panel shows the total number of specific contacts the

protein PRC2 makes with the H3K27M peptide over the course of the trajectory. The bottom panel shows which residues interact with the peptide in each trajectory frame. As explained earlier, ASN 668 is making strong and more than one interactions with the peptide, which is represented by a darker shade of orange, according to the scale to the right of the plot.

4.2.4   Interaction of SRM with SET-I region:

The interaction of SRM region of EZH2 makes contacts with SET-I region of EZH2 so that SET-I region rotates in a way that it exposes the lysine-binding substrate making it active. The residues involved in these interactions are Lys_660 and Valine_657 of SET-I and Phe_145, ASN_142 and Leu_149 of SRM [31]. From interactions plot, it can be seen mutant peptide is making hydrogen bond interactions with Lys_660 of Set-I region which may leads to the distortion in the interactions between SET-I and SRM. We computed distances of K-660 with SRM's Leu-149, Phe-145, ASN-142 in native PRC2 structure and also in H3K27M-PRC2 complex and then took their average. The average distance in native PRC2 structure was 3.5 and in H3K27M-PRC2 complex was 10.73. In conclusion, it is possible that the interaction of Lys_660 with H3K27M peptide is pulling it away from SRM's residues disrupting its interaction with these residues, due to which SET-I is not rotated to its active form and the lysine binding substrate will not be opened for the binding of lysine to turn the genes off.
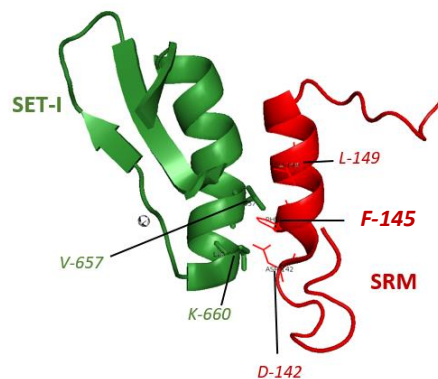
Figure 4.8 Binding Interface between SRM and SET-I regions [31]
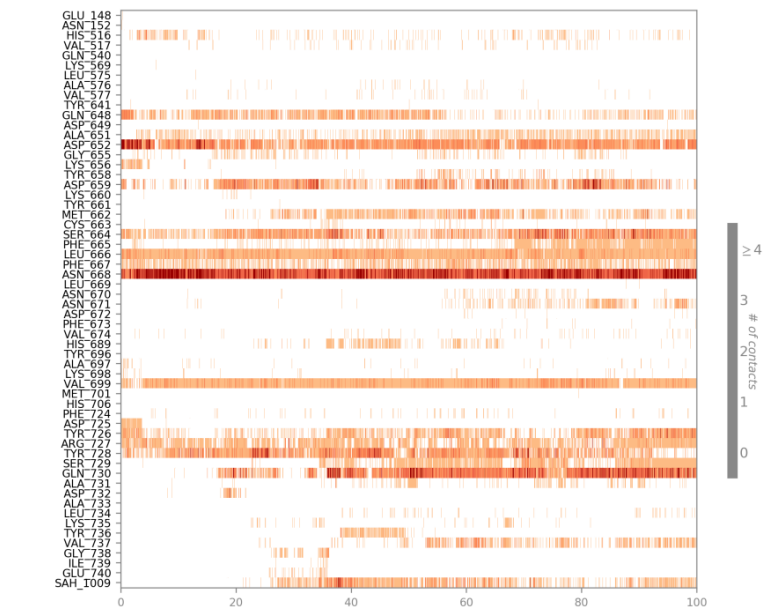
Figure 4.9 Timeline of Interactions

Below is the schematic of detailed peptide atom interaction with the PRC2 residues. Interaction that occur more than 30% of the simulation time in the selected trajectory (0.00 through 100.00 nsec), are shown. The diagram shows that ASN-668 of chain A is making more than one interactions. ASN-668 is donating a hydrogen to the carboxyl side chain of Methionine number 27. The amide side chain of Methionine number 27 is donating its hydrogen to the LEU-666 of chain A. Hence, these residues are involved in the stabilization of mutant peptide in the active site of PRC2 complex and ultimately stopping the PRC2 complex to function properly.
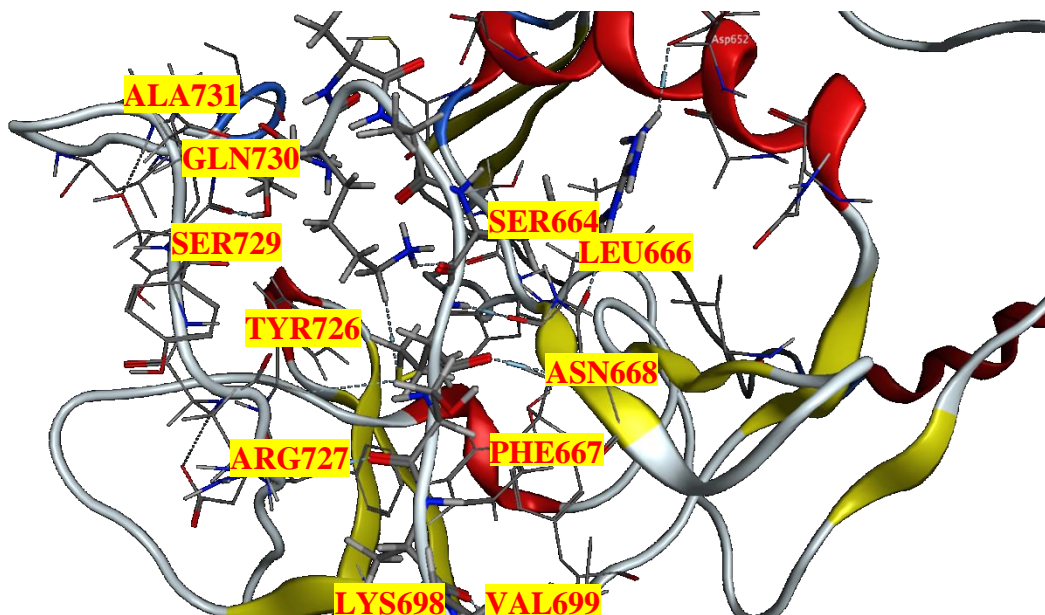
Figure 4.10 3D representation of Interactions between Peptide and PRC2

4.3   MD Simulations of Wild-Type Peptide:

4.3.1   Structural Stability Analysis of H3K27me3-PRC2 complex:

The RMSD plot of Wild-type peptide and PRC2 complex is showing a stable RMSD of 0.9 – 1.8Å. This implies that the protein's overall structure has remained relatively stable, with only minor fluctuations or changes in its conformation during the simulation. The RMSD plot of H3K27me3-PRC2 complex shows that the binding of wild-type peptide stabilizes the PRC2 complex as compared to mutant peptide H3K27M which caused relatively high RMSD of PRC2 complex and ultimately made it less stable.
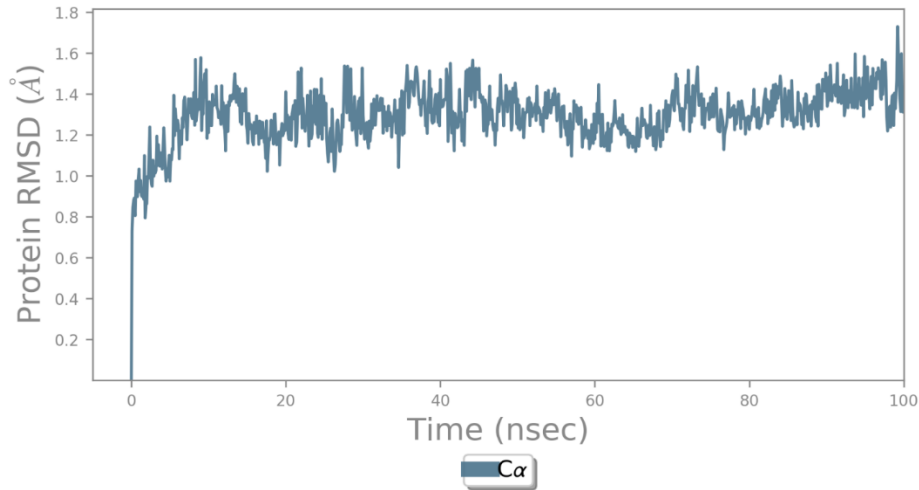
Figure 4.11 RMSD plot of H3K27me3-PRC2 complex

4.3.2   RMSF Plot of PRC2 complex:

The RMSF plot shows the local changes along the protein chain. Green lines in the plot show the residues of the PRC2 complex that are involved in the binding with the H3K27me3 peptide.
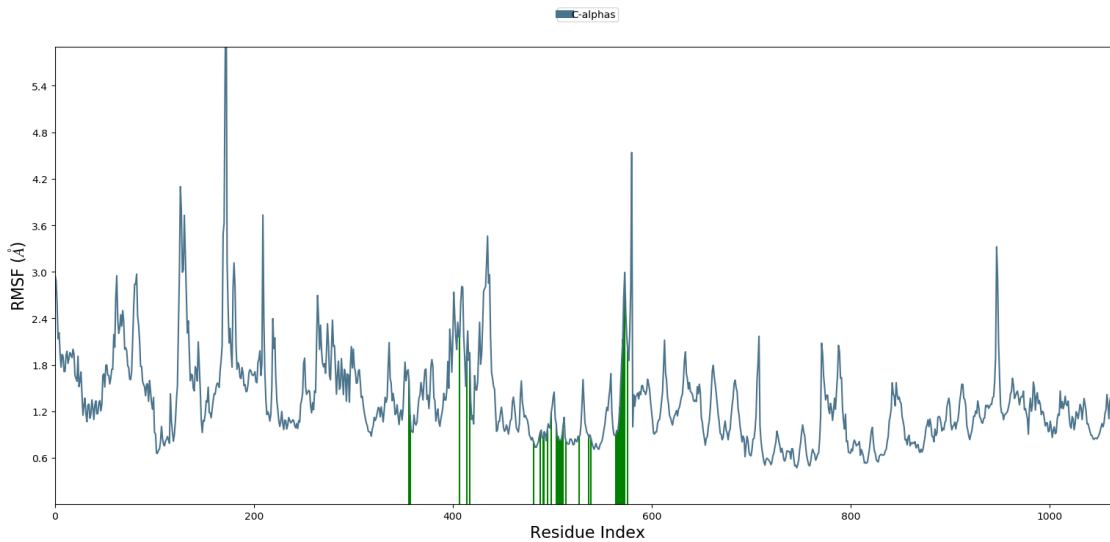
Figure 4.12 RMSF Plot of PRC2

In PRC2-H3k27me3 peptide only few residues showed high fluctuations unlike PRC2-H3K27M peptide. Those highly fluctuating residues are TYR_181 AND asn_182 with RMSF values 5.585 and 7.068 respectively.

4.3.3   PRC2-H3K27me3 Contacts:
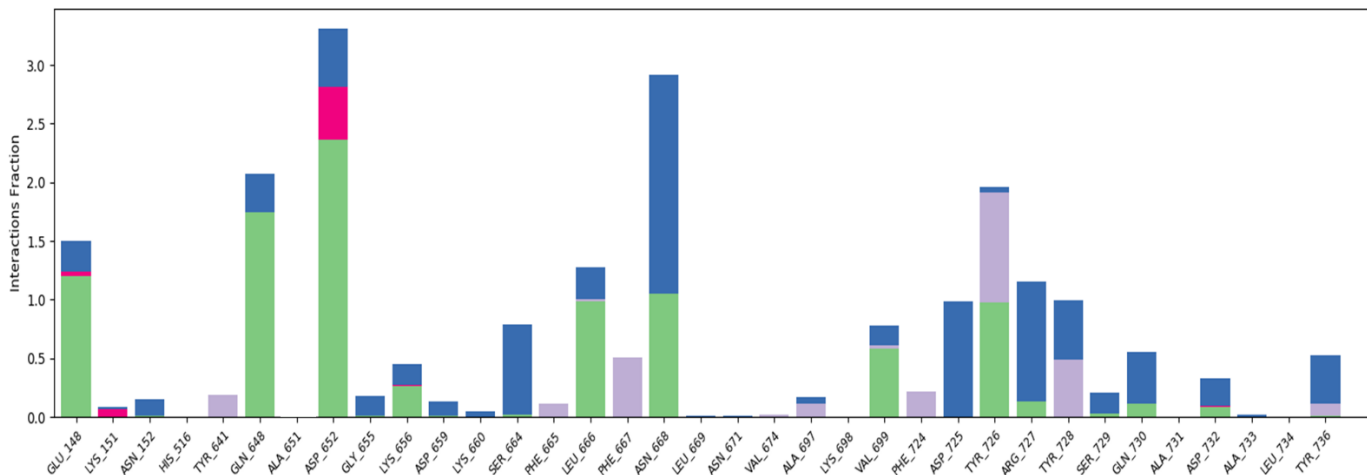
Figure 4.13 Histogram Plot of H3K27me3-PRC2 interactions

In H3K27me3-PRC2 ASN_668 is again involved in extensive interactions between peptide and PRC2 but here it is not showing any hydrophobic interactions as compared to H3K27M-PRC2 complex. In wild type-PRC2 complex, ASP_652 is showing extensive interactions with 1.739 Hydrogen bond contacts, 0.351 ionic interactions, and 0.720 water mediated interactions. The residues of lysine access channel i.e. TYR_726 and TYR_728 of chain A are making more extensive interactions with H3K27me3 as compared to H3K27M. The plot below shows the stability of hydrogen bonds during the time of simulations.
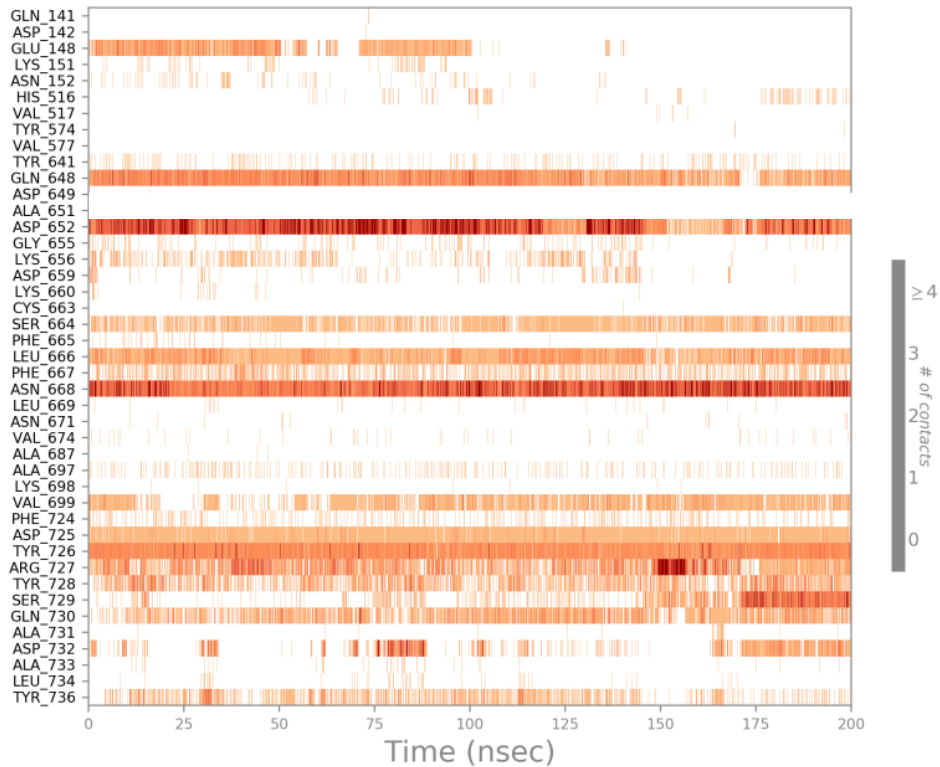
Figure 4.14 Timeline of H3K27me3-PRC2 interactions

The bottom plot is showing the 2D representation of H3K27me3-PRC2 interactions. ASN-668 is involved in many strong interactions with the atoms of H3K27me3. The C--terminal of Lysine number 27 is making 99 % interactions with the residue ASN_668. The trimethyl group at Lysine number 27 is making Pi-cation interactions with TYR-726 of lysine access channel. So it can be inferred from the interaction plots of both mutant and wild-type peptide – prc2 complexes that H3K27M fits with PRC2 complex in a similar way to that of H3K27me3.
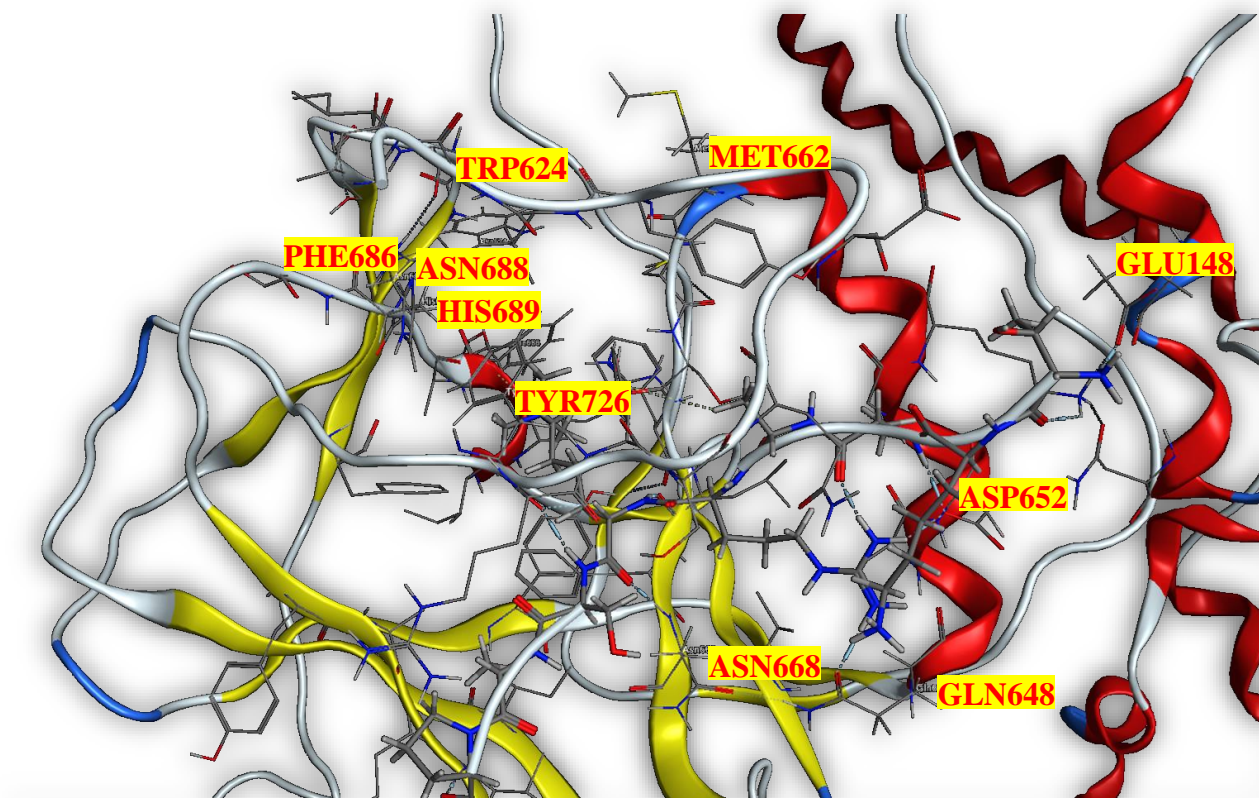
Figure 4.15 3D representation of H3K27me3-PRC2 interactions

4.4   Binding Free Energy Analysis of Mutant Vs Wild-Type Peptide:

The tables below contain the binding free energy values of mutant peptide-PRC2 complex and wild type-PRC2 complex.

Table 5 Energy values of Mutant Peptide/Wild type peptide-PRC2 complex

Energy Terms	H3K27M-PRC2 complex	H3K27me3-PRC2 complex
ΔG_bind	-126.3079	-42.3238
Columbic contribution	-167.6024	-104.9789
Covalent contribution	9.3856	8.6656
Hydrogen bonding contribution	-7.1331	-2.0659
Lipophilic (hydrophobic) contribution	-23.6102	-16.3144
Van der Waals (dispersion) contribution	-87.3737	-40.8406
Packing interactions	-0.7778	-0.4866
Self-contact contribution	0.0046	3.1807
Generalized Born solvent contribution	150.7991	110.5164
Total energy of the molecular complex	-16970.8917	-15774.7915

Interpretation:

4.4.1   ΔG_bind (Binding Free Energy):

- H3K27M-PRC2 Complex: ΔG_bind = -126.31 kcal/mol
- H3K27me3-PRC2 Complex: ΔG_bind = -42.32 kcal/mol

The binding free energy (ΔG_bind) of the H3K27M-PRC2 complex is significantly more negative than that of the H3K27me3-PRC2 complex. This indicates that the H3K27M peptide has a stronger binding affinity for the PRC2 complex compared to the H3K27me3 peptide.

- **Contributions to Binding Free Energy:**

The Coulombic, covalent, hydrogen bonding, lipophilic, van der Waals, and solvent contributions collectively make up the binding free energy.

In the H3K27M-PRC2 complex, these contributions are more pronounced, indicating stronger interactions between the mutant peptide and the PRC2 complex. In the H3K27me3-PRC2 complex, the contributions are comparatively weaker.

- **Self-Contact Contribution:**

The self-contact contribution in the H3K27M-PRC2 complex is negligible (0.00 kcal/mol), while it is positive (3.18 kcal/mol) in the H3K27me3-PRC2 complex. This suggests that the mutant peptide in the H3K27M-PRC2 complex does not have significant self-interactions, which may contribute to its higher binding affinity.

- **Generalized Born Solvent Contribution:**

The solvent contribution in both complexes is substantial, but it is higher in the H3K27M-PRC2 complex. This indicates that solvent effects play a significant role in stabilizing the complex.

In conclusion, the results suggest that the H3K27M mutation enhances the binding affinity of the peptide to the PRC2 complex, leading to a more negative binding free energy. This finding proves the experimental findings that H3K27M binds to PRC2 complex more strongly/tightly as compared to H3K27me3 [31]. This may have implications for the role of the mutation in the development of DIPG and its potential impact on epigenetic regulation.

CHAPTER: 5

DISCUSSION

5   Discussion

Diffuse Intrinsic Pontine Glioma (DIPG) is a devastating and highly aggressive pediatric brainstem tumor primarily affecting children between the ages of 0 and 14 years. It is characterized by its diffuse infiltrative growth pattern, rendering it inoperable, and its location within the pons, a critical region of the brainstem responsible for numerous vital functions [32] [33]. DIPG represents a grave clinical challenge due to its dismal prognosis, with a median survival of less than a year following diagnosis, and a 0% long-term survival rate despite decades of research and therapeutic interventions. This grim prognosis is compounded by the lack of effective treatment options, underscoring the urgent need for a deeper understanding of the disease and the development of innovative therapies [34]. DIPG accounts for approximately 10-20% of all pediatric central nervous system tumors, making it a relatively rare but devastating disease [35]. Its characteristic symptoms often include cranial nerve deficits, ataxia, and motor weakness, which result from the tumor's infiltration of critical brainstem structures. Due to the anatomical location of DIPG, biopsy or surgical resection is rarely performed, as it carries significant risks and provides limited therapeutic benefit. Consequently, the diagnosis of DIPG is primarily based on clinical and radiological findings [36]. Recent advancements in molecular biology have shed light on the underlying genetic alterations that drive DIPG. One of the most critical discoveries is the frequent occurrence of mutations in the gene encoding histone H3, specifically the substitution of lysine 27 with methionine (H3K27M), occurring in up to 80% of cases. These mutations lead to widespread epigenetic dysregulation and are considered a hallmark of DIPG. Histone H3 mutations are not exclusive to DIPG but are also found in other high-grade gliomas, particularly those located in the midline of the brain, emphasizing their importance in tumorigenesis. The H3K27M mutation is thought to be a pivotal driver of DIPG pathogenesis [37]. It disrupts the function of the Polycomb Repressive Complex 2 (PRC2), a key epigenetic regulator responsible for maintaining gene silencing through histone methylation. In DIPG, this mutation results in the loss of trimethylation at H3K27 (H3K27me3), a crucial epigenetic mark associated with facultative heterochromatin [38]. The loss of H3K27me3 leads to the derepression of genes that promote tumorigenesis and inhibit normal cellular differentiation, contributing to the aggressive behavior of DIPG tumors [29].

The Polycomb Repressive Complex 2 (PRC2) is a multi-subunit epigenetic complex that plays a central role in gene regulation through the modification of chromatin structure [30]. It is intricately involved in establishing and maintaining transcriptional repression during various cellular processes, including development, differentiation, and stem cell maintenance. The structure and activity of the PRC2 complex are crucial to its function and are closely related to the pathogenesis of Diffuse Intrinsic Pontine Glioma (DIPG) [36] [22] [39].

The PRC2 complex is composed of several core subunits, each with specific functions [30]:

- **Ezh2 (Enhancer of Zeste Homolog 2):** Ezh2 is the catalytic subunit of PRC2 and possesses histone methyltransferases activity. It is responsible for adding methyl groups to histone H3 on lysine 27 (H3K27), leading to the formation of H3K27me1/2/3 marks [40].
- **Eed (Embryonic Ectoderm Development):** Eed is essential for the recognition and binding of the H3K27me3 mark. It stabilizes the binding of PRC2 to chromatin and stimulates its methyltransferases activity.
- **Suz12 (Suppressor of Zeste 12 Homolog):** Suz12 is crucial for maintaining the integrity of the PRC2 complex and enhancing its histone methyltransferases activity. It interacts with both Ezh2 and Eed to form a stable core complex.
- **RbAp46/48 (Retinoblastoma Binding Protein 46/48):** RbAp46/48 proteins are involved in histone recognition and substrate binding. They help in positioning histone substrates for methylation by Ezh2.
- **Other Accessory Subunits:** PRC2 can also associate with additional accessory subunits, which can modulate its activity and target gene specificity. These subunits include AEBP2, JARID2, and PHF1/19.

The primary function of PRC2 is to catalyze the methylation of H3K27, leading to the formation of H3K27me3 marks on chromatin. These marks serve as epigenetic signals that indicate gene repression. The PRC2 complex achieves this through a series of molecular events:

Recognition of Target Genes: PRC2 identifies specific genomic loci through various mechanisms, including interaction with non-coding RNAs, DNA-binding proteins, and other chromatin-associated factors [41].

Histone Methylation: Once localized to target genes, PRC2 catalyzes the methylation of H3K27. Ezh2, as the catalytic subunit, transfers methyl groups from S-adenosylmethionine (SAM) to H3K27, leading to the progressive accumulation of H3K27me1, H3K27me2, and H3K27me3 marks.

Chromatin Compaction: The addition of H3K27me3 marks by PRC2 promotes the recruitment of other repressive chromatin modifiers and chromatin compaction, resulting in the repression of gene transcription.

PRC2 and DIPG Tumor: The relationship between the PRC2 complex and DIPG tumor is primarily centered on the frequent occurrence of mutations in histone H3, specifically the H3K27M mutation, which is found in a substantial proportion of DIPG cases. This mutation involves the substitution of lysine 27 with methionine in histone H3 (H3.3 or H3.1 variant), disrupting the normal epigenetic regulation mediated by PRC2. Several key points highlight this relationship [17]:

H3K27M Mutation Disrupts PRC2 Function: The H3K27M mutation inhibits the methylation of H3K27 by PRC2. This leads to a global loss of the repressive H3K27me3 marks in affected cells and results in aberrant gene expression.

Epigenetic Dysregulation in DIPG: The disruption of PRC2-mediated epigenetic regulation is a hallmark of DIPG. It leads to the derepression of genes that promote tumorigenesis and inhibit normal cellular differentiation, contributing to the aggressive behavior of DIPG tumors.

H3K27M and PRC2 Interaction: The precise molecular mechanisms governing the interaction between the H3K27M mutation and PRC2, and how this interaction contributes to DIPG tumorigenesis, are subjects of ongoing research. It is believed that the mutation disrupts the binding of PRC2 to chromatin and alters its specificity, leading to the activation of oncogenic pathways [38].

Therapeutic Implications: Understanding the interplay between H3K27M and PRC2 is critical for the development of targeted therapies for DIPG. Efforts are being made to identify strategies that can restore PRC2 function or counteract the effects of the mutation [30].

The methodology employed in this study aimed to investigate the structural and energetic consequences of the H3K27M mutation on the Polycomb Repressive Complex 2 (PRC2) and its implications for gene regulation. This discussion will provide an overview of the steps involved in the computational simulations and analysis of the PRC2 complex and its interactions with both mutant and wild-type histone peptides. The initial step involved retrieving the three-dimensional crystal structure of the PRC2 complex from the RCSB Protein Data Bank (PDB) database. The selected structure, identified by PDB ID "5HYN," was chosen due to its relevance to the study as it contains the core subunits of PRC2 (EZH2, EED, and SUZ12), the H3K27M mutant peptide, and JARID2 K116me3, which mimics the functioning of the H3K27me3 peptide. Additionally, the structure contains zinc ions and the co-factor SAH. Before conducting molecular dynamics (MD) simulations, the retrieved PRC2 structure required preprocessing to prepare it for analysis. This included retaining only the essential chains (chain A for EZH2, chain B for EED, chain C for SUZ12, and chain D for H3K27M peptide), resolving missing atoms, residues, loops, and sidechains using the Protein Preparation Wizard. Water molecules and other heteroatoms that were not part of the protein were removed, and correct bond orders and hydrogen positions were assigned. Ionizable residues were assigned pKa states, and further refinement was conducted to optimize hydroxyl, Asn, Gln, and His states using ProtAssign. Desmond's System Builder package was employed to define PBC. This involved selecting the TIP3P water model for the solvent, specifying the shape and dimensions of the simulation box (orthorhombic in this case), and ensuring an adequate buffer space around the protein to prevent interactions with periodic images. Counter ions (Cl-) were added to neutralize the system, and salt (0.15M) was introduced. The OPLS4 force field was selected for the simulations. Energy minimization was performed to optimize the molecular arrangement and alleviate steric clashes in the protein structure before initiating MD simulations. This step is essential for starting from a stable configuration and avoiding unrealistic dynamics. Desmond's energy minimization tool was used to relax the protein and remove steric clashes, with the System Preparation panel guiding the setup of the energy minimization job. Equilibration of the system was conducted to allow it to adjust and stabilize before data collection in MD simulations. Two types of equilibration were performed: NVT equilibration, which controls volume, number of particles, and temperature, and NPT equilibration, which additionally controls pressure. These steps are crucial to achieving accurate

and meaningful results in MD simulations. Proper temperature and pressure settings were specified (300 K and 1.01325 bar), and integration settings and relaxation protocols were defined. The production phase of MD simulations was configured, specifying parameters such as simulation time (100 nanoseconds), time step, constraints, and output frequencies for trajectory snapshots. Desmond's MD Simulations panel facilitated the setup of the simulation run, which was then submitted to Desmond using a supercomputer. Following the completion of MD simulations, the trajectory data were analyzed using Desmond's Simulation Interaction Diagram tool. This analysis included examining root mean square deviation (RMSD), root mean square fluctuation (RMSF), and other relevant graphs to assess the dynamics and stability of the protein complexes. To simulate the wild-type peptide (H3K27me3) in complex with PRC2, the 5HYN structure, which originally contained the mutant H3K27M peptide, was used. Using PyMol software, the mutant Methionine residue at position 27 was replaced with the wild-type Lysine residue. Additionally, a post-translational modification (PTM) was applied to convert the Lysine at position 27 to a trimethylated Lysine to mimic H3K27me3. This refined structure was then used for simulations following the same steps as for the mutant peptide. In this study, the molecular mechanics-generalized Born surface area (MM-GBSA) method implemented in Prime was employed to calculate the binding free energy (ΔG_bind) of the protein-peptide complexes. This involved considering various energy components, including Coulombic interactions, van der Waals interactions, hydrogen bonding, hydrophobic interactions, covalent interactions, and solvation effects. The negative or positive values of ΔG_bind indicated whether the binding process was energetically favorable or unfavorable, respectively.

5.1   Conclusion:

The results of our computational study provide a comprehensive understanding of the structural and functional implications of the H3K27M mutation within the context of the Polycomb Repressive Complex 2 (PRC2) and shed light on its potential role in aberrant gene regulation in Diffuse Intrinsic Pontine Glioma (DIPG). Firstly, we observed that in the absence of EED and SUZ12, PRC2 exhibited increased root-mean-square deviation (RMSD) values ranging from 0 to 20 Å, indicating a state of instability. However, when bound with either the mutant (H3K27M) or wild-type (H3K27me3) peptides, along with EED and SUZ12, the RMSD plot demonstrated

stability within the PRC2 complex. This suggests that the presence of these peptides and associated proteins contributes to the stabilization of PRC2. Notably, when comparing the H3K27M-PRC2 complex to the H3K27me3-PRC2 complex, we observed higher RMSD values in the former, ranging from 1.95 to 5 Å, while the latter maintained a range of 0.9 to 1.8 Å. This discrepancy in RMSD values suggests that the H3K27M mutation induces a distinct structural behavior within PRC2 compared to the wild-type configuration, possibly contributing to the inactivity of PRC2. A significant finding from our simulations was the straightening of the SBD (SANT2) subdomain of PRC2 when it was bound to the mutant peptide. This subdomain typically exhibits a bent configuration in the active state of PRC2. The observed straightening of the SBD subdomain in the presence of the mutant peptide may be associated with PRC2's reduced activity. Our analysis further highlighted the crucial role of ASN-668 in EZH2 for the strong binding of H3K27M to PRC2, leading to stable bonding and subsequent PRC2 inactivation. Additionally, we observed interactions between the SRM (SET-Responsive Motif) and SET-I regions of EZH2, involving Lys_660 and Valine_657 of SET-I and Phe_145, ASN_142, and Leu_149 of SRM. These interactions are thought to activate the lysine-binding substrate. However, the mutant peptide disrupted these interactions, especially by interacting with Lys_660 of Set-I. This interaction led to the displacement of K-660 from SRM's residue, thereby preventing SET-I from becoming active, which, in turn, inhibited lysine binding. This disruption of the SET-I region's activity is a critical factor in deactivating the PRC2 complex. Furthermore, our binding free energy calculations corroborated the hypothesis presented in the literature that the H3K27M peptide exhibits stronger binding affinity to the PRC2 complex, as indicated by a more negative ΔG_bind value compared to the wild-type H3K27me3. This shift in binding energy underscores the mutation's disruptive effect on the normal functioning of the PRC2 complex, suggesting its potential role in the aberrant regulation of gene expression in DIPG.

In summary, our computational exploration has provided valuable insights into the structural and functional consequences of the H3K27M mutation within the PRC2 complex. These findings contribute to our understanding of epigenetic dysregulation in cancer, particularly in the context of DIPG, and offer promising directions for future therapeutic interventions aimed at targeting this critical molecular pathway.

CHAPTER: 6

REFERENCES

6   References

1.   Subramanian S, A.T., *Childhood Brain Tumors.* StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, 2023.

2.   Patil, N., et al., *Epidemiology of brainstem high-grade gliomas in children and adolescents in the United States, 2000-2017.* Neuro Oncol, 2021. **23**(6): p. 990-998.

3.   Spine, M.B. *Tricia's Story (Hemangioblastoma)*. 2020; Available from: https://mayfieldclinic.com/mc_hope/story_tricia2020.htm.

4.   Childrens, N. *Diffuse Intrinsic Pontine Glioma*. 2020; Available from: https://www.nationwidechildrens.org/conditions/diffuse-intrinsic-pontine-glioma#:~:text=Diffuse%20intrinsic%20pontine%20glioma%2C%20also,and%20muscles%20of%20the%20face.

5.   Farrukh, S., et al. *Emerging Therapeutic Strategies for Diffuse Intrinsic Pontine Glioma: A Systematic Review*. in *Healthcare*. 2023. MDPI.

6.   Khan, D.A., et al., *Treatment options for paediatric brainstem gliomas.* JPMA. The Journal of the Pakistan Medical Association, 2019. **69**(9): p. 1400-1402.

7.   Laghari, A., et al., *Pediatric brainstem gliomas: An institutional experience.* Asian Journal of Neurosurgery, 2019. **14**(04): p. 1144-1150.

8.   Vanan, M.I. and D.D. Eisenstat, *DIPG in Children – What Can We Learn from the Past?* Frontiers in Oncology, 2015. **5**.

9.   Castel, D., et al., *Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes.* Acta neuropathologica, 2015. **130**: p. 815-827.

10.   Chan, K.-M., et al., *The histone H3. 3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression.* Genes & development, 2013. **27**(9): p. 985-990.

11.   Greally, J.M., *A user's guide to the ambiguous word'epigenetics'.* Nature Reviews Molecular Cell Biology, 2018. **19**(4): p. 207-208.

12.   Park, J. and C. Chung, *Epigenetic and Metabolic Changes in Diffuse Intrinsic Pontine Glioma.* Brain Tumor Research and Treatment, 2023. **11**(2): p. 86.

13.   Chittock, E.C., et al., *Molecular architecture of polycomb repressive complexes.* Biochemical Society Transactions, 2017. **45**(1): p. 193-205.

14.   Margueron, R. and D. Reinberg, *The Polycomb complex PRC2 and its mark in life.* Nature, 2011. **469**(7330): p. 343-349.

15.   Zhang, X. and Z. Zhang, *Oncohistone mutations in diffuse intrinsic pontine glioma.* Trends in cancer, 2019. **5**(12): p. 799-808.

16.   Chammas, P., I. Mocavini, and L. Di Croce, *Engaging chromatin: PRC2 structure meets function.* British Journal of Cancer, 2020. **122**(3): p. 315-328.

17.   Laugesen, A., J.W. Højfeldt, and K. Helin, *Role of the polycomb repressive complex 2 (PRC2) in transcriptional regulation and cancer.* Cold Spring Harbor perspectives in medicine, 2016. **6**(9): p. a026575.

18.   Shi, Y., et al., *Structure of the PRC2 complex and application to drug discovery.* Acta Pharmacologica Sinica, 2017. **38**(7): p. 963-976.

19.   *Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas.* Nature genetics, 2012. **44**(3): p. 251-253.

20.   Burgess, R.J. and Z. Zhang, *Histone chaperones in nucleosome assembly and human disease.* Nature structural & molecular biology, 2013. **20**(1): p. 14-22.

21.   Talbert, P.B. and S. Henikoff, *Histone variants—ancient wrap artists of the epigenome.* Nature reviews Molecular cell biology, 2010. **11**(4): p. 264-275.

22.   Vizán, P., et al., *Role of PRC 2-associated factors in stem cells and disease.* The FEBS journal, 2015. **282**(9): p. 1723-1735.

23.   Schuettengruber, B., et al., *Genome regulation by polycomb and trithorax: 70 years and counting.* Cell, 2017. **171**(1): p. 34-57.

24.   Stafford, J.M., et al., *Multiple modes of PRC2 inhibition elicit global chromatin alterations in H3K27M pediatric glioma.* Science advances, 2018. **4**(10): p. eaau5935.

25.   2023-3:, S.R., *Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2023.*

26.   *Proceedings of the 2006 ACM/IEEE conference on Supercomputing.* 2006. Tampa, Florida: Association for Computing Machinery.

27.   Roy, K., S. Kar, and R.N. Das, *Chapter 5 - Computational Chemistry*, in *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, K. Roy, S. Kar, and R.N. Das, Editors. 2015, Academic Press: Boston. p. 151-189.

28.   Schrödinger, L.a.W.D., *PyMol*. 2020.

29.   Choudhary, M.I., et al., *In silico identification of potential inhibitors of key SARS-CoV-2 3CL hydrolase (Mpro) via molecular docking, MMGBSA predictive binding energy calculations, and molecular dynamics simulation.* Plos one, 2020. **15**(7): p. e0235030.

30.   Kouznetsova, V.L., et al., *Polycomb repressive 2 complex—Molecular mechanisms of function.* Protein Science, 2019. **28**(8): p. 1387-1399.

31.   Guo, Y., S. Zhao, and G.G. Wang, *Polycomb gene silencing mechanisms: PRC2 chromatin targeting, H3K27me3'readout', and phase separation-based compaction.* Trends in Genetics, 2021. **37**(6): p. 547-565.

32.   Warren, K., *Diffuse intrinsic pontine glioma: poised for progress.* Frontiers in oncology, 2012. **2**: p. 38080.

33.   Duffner, P.K., M.E. Cohen, and A.I. Freeman, *Pediatric brain tumors: an overview.* CA: a cancer journal for clinicians, 1985. **35**(5): p. 287-301.

34.   Walker, D.A., et al., *A multi-disciplinary consensus statement concerning surgical approaches to low-grade, high-grade astrocytomas and diffuse intrinsic pontine gliomas in childhood (CPN Paris 2011) using the Delphi method.* Neuro-oncology, 2013. **15**(4): p. 462-468.

35.   Hargrave, D., U. Bartels, and E. Bouffet, *Diffuse brainstem glioma in children: critical review of clinical trials.* The lancet oncology, 2006. **7**(3): p. 241-248.

36.   Bartels, U., et al., *Proceedings of the diffuse intrinsic pontine glioma (DIPG) Toronto Think Tank: advancing basic and translational research and cooperation in DIPG.* Journal of neuro-oncology, 2011. **105**: p. 119-125.

37.   Cholewa-Waclaw, J., et al., *The role of epigenetic mechanisms in the regulation of gene expression in the nervous system.* Journal of Neuroscience, 2016. **36**(45): p. 11427-11434.

38.   Diehl, K.L., et al., *PRC2 engages a bivalent H3K27M-H3K27me3 dinucleosome inhibitor.* Proceedings of the National Academy of Sciences, 2019. **116**(44): p. 22152-22157.

39.   Riising, E.M., et al., *Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide.* Molecular cell, 2014. **55**(3): p. 347-360.

40.   von Schimmelmann, M., et al., *Polycomb repressive complex 2 (PRC2) silences genes responsible for neurodegeneration.* Nature neuroscience, 2016. **19**(10): p. 1321-1330.

41.   Thornton, S.R., et al., *Polycomb repressive complex 2 regulates lineage fidelity during embryonic stem cell differentiation.* PloS one, 2014. **9**(10): p. e110498.