

# **Molecular Dynamics Simulations of the Substrate- and Product Specificity and Mechanism of DNA- and Protein Lysine Methyltransferases**

A thesis accepted at the Faculty 4: Energy-, Process- and Bio-Engineering of the University of Stuttgart and the Stuttgart Center for Simulation Science in partial fulfillment of the requirements for the degree of

**Doktor der Naturwissenschaften/PhD in Natural Science (Dr. rer. nat.)**

by

**Philipp Schnee**

born 27.10.1995 in Stuttgart, Germany

Main Referee: Prof. Dr. Albert Jeltsch (University of Stuttgart)  
Co-Referee: Prof. Dr. Martin Zacharias (Technical University Munich)  
Committee Chair: Prof. Dr. Markus Morrison (formerly Rehm) (University of Stuttgart)

Date of Defense: 17.04.2024

**University of Stuttgart**

Institute of Biochemistry and Technical Biochemistry

2024

### **Eidesstattliche Erklärung**

Hiermit versichere ich, dass ich die vorliegende Arbeit mit dem Titel

„Molecular Dynamics Simulations of the Substrate- and Product Specificity and Mechanism of DNA- and Protein Lysine Methyltransferases“

selbstständig verfasst habe und dabei keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

### **Declaration of Authorship**

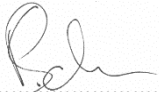
I hereby certify that the dissertation entitled

„Molecular Dynamics Simulations of the Substrate- and Product Specificity and Mechanism of DNA- and Protein Lysine Methyltransferases“

is entirely my own work except where otherwise indicated. Passages and ideas from other sources have been clearly indicated.

Name: Philipp Schnee

Unterschrift/Signature:

.....

Datum/Date:

08.01.2024

# Table of contents

Abstract .....	VI
Zusammenfassung.....	IX
Acknowledgements .....	XII
List of publications.....	XIV
Author contributions .....	XV
List of figures .....	XVI
List of abbreviations .....	XVIII
1. Introduction.....	1
1.1. Epigenetics .....	1
1.2. Chromatin structure and its regulation .....	1
1.3. DNA methylation.....	2
1.3.1. DNA methyltransferases .....	3
1.4. Lysine methylation.....	4
1.4.1. Histone lysine methylation.....	5
1.4.2. Non-histone protein lysine methylation .....	5
1.5. Protein lysine methyltransferases .....	6
1.5.1. Structure of SET domain PKMTs.....	7
1.5.2. Different structural arrangements of SET domain PKMTs .....	9
1.5.3. Autoinhibition of SET domain PKMTs.....	11
1.5.4. Placeholder residues .....	11
1.5.5. Target lysine deprotonation .....	12
1.5.6. Reaction mechanism of SET domain PKMTs .....	14
1.5.7. Substrate specificity of SET domain PKMTs .....	15
1.5.8. Discovery of PKMT super-substrates.....	16
1.5.9. Product Specificity of SET domain PKMTs .....	18
1.6. Histone lysine 36 methylation .....	19
1.6.1. SETD2.....	20
1.6.2. NSD2 .....	22
1.7. Lysine demethylation.....	23

1.8.	Molecular Dynamics Simulation .....	25
1.8.1.	Modelling.....	25
1.8.2.	Structure of MD Simulations .....	27
1.8.3.	Steered molecular dynamics simulations.....	29
2.	Aims of this work.....	31
2.1.	MD Simulation of the somatic cancer mutation R882H of DNMT3A .....	31
2.2.	MD Simulation of the somatic cancer mutation T1150A of NSD2 .....	32
2.3.	Mechanistic basis of super-substrate peptides .....	32
3.	Material and Methods.....	33
3.1.	MD simulations and representation .....	33
3.1.1.	MD simulations of the DNMT3A/L hetero tetramer complexed with DNA .....	34
3.1.2.	sMD simulations of the peptide association process into the NSD2 active site .....	34
3.1.3.	MD simulations of NSD2.....	35
3.1.4.	MD simulations of peptides in solution .....	36
3.1.5.	sMD simulations of the peptide association process into the SETD2 active site .....	36
3.1.6.	MD simulations of SETD2 .....	37
3.2.	Trajectory analysis .....	38
3.2.1.	Contact maps analysis .....	38
3.2.2.	NSD2 active site volume calculation .....	38
3.2.3.	Clustering of peptide conformations in solution .....	38
3.3.	SETD2 purification and peptide hairpin validation using FRET .....	39
3.4.	<i>In vitro</i> methylation assay to test peptide inhibitors.....	39
4.	Results .....	41
4.1.	MD Simulation of the somatic cancer mutation R882H of DNMT3A .....	42
4.1.1.	DNMT3A R882H establishes more inter subunit contacts than the DNMT3A WT .....	42
4.2.	MD simulation of the T1150A cancer mutant of NSD2 .....	45
4.2.1.	NSD2 T1150A can accommodate a H3K36me2 peptide and SAM in sMD simulations .....	48
4.2.2.	NSD2 T1150A loses contacts responsible for restricting the active site volume .....	50
4.3.	Mechanistic basis of the enhanced methylation activity towards super-substrates .....	52
4.3.1.	The SETD2 super-substrate peptide prefers a hairpin conformation in solution .....	53
4.3.2.	Experimental investigation of conformational preferences .....	55

4.3.3. A hairpin conformation has easier access into the SETD2 active site.....	57
4.3.4. Hairpin structures unfold into an extended conformation upon binding to SETD2 .....	59
4.3.5. Hairpin conformation in peptides lead to a faster methylation by SETD2.....	62
4.3.6. ssK36 establishes more and different TS-like conformation than H3K36.....	63
4.4. Peptides can function as competitive inhibitors for PKMTs .....	68
4.5. MD Simulation of NSD2 in complex with a new NSD2-specific super-substrate.....	70
5. Discussion.....	75
5.1. MD Simulation of the somatic cancer mutation R882H of DNMT3A .....	75
5.2. MD Simulation of the T1150A cancer mutant of NSD2 .....	78
5.3. Mechanistic basis of the SETD2 super-substrate peptide .....	81
5.3.1. Hairpins in histone tails might facilitate the binding towards PKMTs .....	82
5.3.2. Peptide hairpin conformations unfold upon binding to PKMTs and establish distinct contacts .....	85
5.4. Super-substrate peptides function as PKMT inhibitors .....	87
6. References.....	90
7. Appendix.....	109
7.1. Appendix I (not included in the published thesis) .....	109
7.2. Appendix II .....	110
7.3. Appendix III .....	180

## Abstract

While the genetic information within each cell is encoded as the base pair sequence in the DNA, cellular differentiation and adaption to environmental signals are dictated by variations in gene expression. Epigenetics describes these often stable, yet reversible, changes in gene expression patterns, which do not involve alterations in the DNA sequence. Major epigenetic signals are DNA and histone lysine methylation. These modifications are deposited by DNA methyltransferases (DNMTs) and protein lysine methyltransferases (PKMTs) by transferring a methyl group from S-adenosyl-L-methionine (SAM) to the respective target. Subsequently, set modifications are read by chromatin remodeling enzymes, altering the accessibility of genes depending on the actual modifications. Hence, DNMTs and PKMTs function as key players in the regulation of genome stability, gene expression, DNA repair and cellular differentiation. Their activity is controlled by factors like substrate- and product specificity, autoinhibition, and conformational changes upon interaction with substrates. Cancer mutations in DNMTs and PKMTs disturb these regulatory mechanisms, which makes their understanding a main target in modern epigenetic research.

In this work, molecular dynamics (MD) simulation in combination with biochemical experiments were used to investigate the catalytic mechanism of these enzymes in detail. In pursuit of this goal, two approaches were applied. By simulating cancer mutants of DNMTs and PKMTs and comparing the obtained simulation results to simulation results of the wild type enzyme (WT), distinctions between mutant and WT can be found. This was achieved for the somatic cancer mutant DNMT3A R882H, which frequently occurs in blood cancers. In earlier methylation experiments a change in flanking sequence preference was determined for R882H. The mechanism behind this observation was revealed by MD simulations of the DNMT3A/L-heterotetramer (3L-3A-3A-3L) in this work. Conducted simulations showed that the mutated R882H residue had a decreased contact to a guanine, three base pairs away from the methylation site, and instead interacted with the adjacent 3A subunit. The lost contact is directly connected to a different affinity for certain DNA substrates, explaining the change in flanking sequence preference. Moreover, R882H was found to have a dominant effect even in a heterozygous state. Extended contacts analysis of the MD simulation data showed that R882H not only interacted with the adjacent subunit but led to a rearrangement of a small contact network, increasing the overall binding affinity of DNMT3A R882H dimers compared to WT. Since the flanking sequence preference of DNMT3A tetrameric complexes is determined in the 3A-3A interface, which is preferentially formed by R882H, the dominant effect of R882H in R882H/WT mixed complexes was rationalized by the MD simulation results.

Biochemical characterization of the PKMT NSD2 and its somatic cancer mutation T1150A revealed an altered product specificity and a change from a dimethyltransferase to a trimethyltransferase. Changes

in the methylation state of histone 3 lysine 36 (H3K36), a methylation target of NSD2, are known to be associated with diverse biological outcomes, as dimethylated H3K36 (H3K36me<sub>2</sub>) and trimethylated H3K36 (H3K36me<sub>3</sub>) exhibit distinct downstream effects on gene transcription and chromatin structure. Therefore, a combination of MD and steered MD (sMD) simulation techniques was used in this work to investigate the reason for the altered product specificity of NSD2 T1150A. The analysis showed that in contrast to NSD2 WT, NSD2 T1150A was able to accommodate the H3K36me<sub>2</sub> peptide and SAM simultaneously in a productive conformation in the active site. Volume calculations of the active site revealed that larger volumes occurred more often for T1150A compared to WT, enabling the productive accommodation of the higher methylation state. The reason for this was found in a subsequent contact analysis. In NSD2 WT, T1150 was engaged in contacts with Y1092 and L1120, which oriented these residues effectively reducing the volume of the active site. The T1150 side chain hydroxyl group interacted with the Y1092 backbone nitrogen. The side chain methyl group hydrophobically interacted with the L1120 side chain. These two contacts were lost in NSD2 T1150A. As a consequence, the orientation of Y1092 and L1120 was more flexible and the active site volume increased. The presented results precisely explain the molecular mechanism behind the altered product specificity observed in biochemical experiment with NSD2 T1150A.

The substrate specificity of the PKMT SETD2 regarding its natural H3K36 target sequence has been previously mapped using Celluspot peptide array methylation. This revealed that the canonical H3 amino acids were not ideal at many positions. Based on this, an artificial peptide substrate was designed that contained the most favorable amino acid at each position. Methylation experiments showed that the 15 amino-acid long super-substrate peptide (ssK36), which differed at four positions from the original H3K36, was methylated more than 100-fold faster than the canonical H3K36 peptide. The crystal structure of SETD2 with bound ssK36 peptide was resolved, but did not entirely explain the highly increased methylation activity of SETD2. The second approach in this work focuses on the mechanistic reasons behind this massive increase in reaction rate, using a combination of *in vitro* methylation and FRET experiments, MD and sMD simulation techniques to cover multiple steps of the catalytic process. MD simulations of the free peptides in solution showed a preference for ssK36 to form hairpin conformation, whereas H3K36 preferred an extended conformation. This preference was based on the four introduced mutations. Moreover, it was demonstrated in sMD simulations that hairpin-shaped peptides had easier access into the active site of SETD2, compared to extended conformations. In fact, methylation experiments confirmed that chemically induced hairpins increased the methylation activity of peptides by SETD2. Additionally, in MD simulation of the ssK36-SETD2 complex it was observed that the four mutations established a unique contact profile with SETD2, leading to more and different transition state-like conformation compared to the H3K36-SETD2 complex. The transferability of this approach was demonstrated as a new super-substrate peptide was specifically designed for NSD2. The molecular mechanism behind the increased methylation rate was

again investigated by MD simulation presented in this work. Remarkably, SETD2 and NSD2 were shown to be specific for their respective super-substrate and did not show increased activity for the other.

The optimized enzyme interactions of the super-substrate peptides were then used as a starting point to establish a PKMT-specific inhibition assay in which ssK36 was demonstrated to function as a substrate-competitive SETD2-specific inhibitor.

In conclusion, the MD simulations conducted in this work revealed yet unknown reasons for the dominant effect of DNMT3A R882H in heterozygous states and explained the altered product specificity for NSD2 T1150A. Moreover, the features of the artificially designed super-substrate peptides, which caused a ~100-fold activity increase of the PKMTs SETD2 and NSD2, were precisely described in various MD simulation approaches and validated by wet-lab experiments. The found molecular mechanisms in this work explain biochemical results of DNMTs and PKMTs at an atomistic resolution and suggest novel strategies for the design of a new class of substrate-competitive PKMT inhibitors.



## Zusammenfassung

Die genetische Information einer jeden Zelle ist in Form der Basenpaarsequenz der DNA codiert. Zelldifferenzierung und Anpassung an Umweltsignale werden jedoch durch Änderungen der Genexpression bewerkstelligt. Epigenetik beschreibt diese meist stabilen, aber reversiblen Veränderungen der Genexpression, denen keine Änderungen der DNA-Sequenz zugrunde liegt. Zu den wichtigsten epigenetischen Signalen gehören die DNA- und Histonlysinmethylierung. Diese Modifikationen werden durch DNA-Methyltransferasen (DNMTs) und Proteinlysinmethyltransferasen (PKMTs), durch Übertragung einer Methylgruppe von S-Adenosyl-L-Methionin (SAM) auf das jeweilige Zielmolekül, gesetzt. Die methylierte DNA oder Proteine werden anschließend von Enzymen erkannt, welche die Zugänglichkeit von Genen, je nach Modifikation, verändern. DNMTs und PKMTs haben deshalb eine Schlüsselrolle bei der Regulation von Genomstabilität, Genexpression, DNA-Reparatur und Zelldifferenzierung. Ihre Aktivität wird von Faktoren wie der Substrat- und Produktspezifität, Autoinhibition sowie Konformationsänderungen während der Interaktion mit Substraten gesteuert. Krebsmutationen in DNMTs und PKMTs stören diese Regulationsmechanismen, wodurch ihre Erforschung zu einem Hauptziel in der modernen epigenetischen Forschung geworden ist.

In dieser Arbeit wurden Molekulardynamik (MD) Simulationen in Kombination mit biochemischen Experimenten durchgeführt, um die katalytischen Mechanismen dieser Enzyme im Detail zu untersuchen. Um dieses Ziel zu erreichen, wurden zwei Ansätze angewendet. Durch Simulationen von Krebsmutationen von DNMTs und PKMTs und Vergleich mit Simulationen des jeweiligen Wildtyp Enzyms (WT) können Unterschiede zwischen Mutanten und WT gefunden werden. Dieser Ansatz wurde auf die somatische Krebsmutation DNMT3A R882H angewandt, die häufig in verschiedenen Leukämieformen auftritt. In früheren Methylierungsexperimenten wurde eine veränderte Präferenz von R882H für die Flankierungssequenz von CpG Methylierungsstellen festgestellt. Der Mechanismus hinter dieser Beobachtung wurde durch MD Simulationen des DNMT3A/L-Heterotetramers (3L-3A-3A-3L) in dieser Arbeit aufgeklärt. Die durchgeführten Simulationen zeigten, dass die mutierte R882H-Aminosäure einen verringerten Kontakt zu einem Guanin, drei Basenpaare vom zu methylierenden CpG Dinukleotid entfernt, aufwies. Stattdessen interagierte sie mit der benachbarten 3A-Untereinheit. Diese verringerte Kontaktintensität steht in direktem Zusammenhang mit einer veränderten Affinität für bestimmte DNA-Substrate, wodurch die Änderung der Flankierungssequenzpräferenz erklärt werden konnte. Darüber hinaus wurde für R882H beschrieben, dass selbst im heterozygoten Zustand ein dominanter Effekt zu beobachten ist. Eine erweiterte Kontaktanalyse der MD Simulationsdaten zeigte, dass R882H nicht nur mit der benachbarten Untereinheit interagierte, sondern dies auch zu einer Restrukturierung eines kleinen Kontaktnetzwerks führte, welches die Bindungsaffinität von DNMT3A R882H-Dimeren im Vergleich zu WT-Dimeren erhöhte. Da die Flankierungssequenzpräferenz von DNMT3A Tetrameren in der 3A-3A-Schnittstelle bestimmt wird, welche vorzugsweise von R882H

Untereinheiten gebildet wird, wurde der dominante Effekt von R882H in gemischten WT/R882H Komplexen durch die Ergebnisse der MD Simulationen erklärt.

Die biochemische Charakterisierung der PKMT NSD2 und ihrer somatischen Krebsmutation T1150A wies eine veränderte Produktspezifität und einen Wechsel von einer Dimethyltransferase zu einer Trimethyltransferase auf. Veränderungen im Methylierungsgrad von Histon 3 Lysin 36 (H3K36), ein Methylierungsziel von NSD2, sind mit unterschiedlichen biologischen Effekten verbunden. Dimethyliertes H3K36 (H3K36me<sub>2</sub>) und trimethyliertes H3K36 (H3K36me<sub>3</sub>) weisen unterschiedliche Auswirkungen auf die Genexpression und Chromatinstruktur auf. Um die Gründe für die veränderte Produktspezifität von NSD2 T1150A zu untersuchen, wurde in dieser Arbeit eine Kombination von MD und *steered* MD (sMD) Simulationsmethoden angewendet. Die Analyse zeigte, dass NSD2 T1150A im Gegensatz zu NSD2 WT ein H3K36me<sub>2</sub>-Peptid und SAM gleichzeitig in einer produktiven Konformation im aktiven Zentrum binden konnte. Volumenberechnungen des aktiven Zentrums zeigten, dass größere Volumina bei T1150A im Vergleich zu WT häufiger vorkamen und die gemeinsame Bindung ermöglicht wurde. Der Grund für die größeren Volumina konnte in einer nachfolgenden Kontaktanalyse gefunden werden. In NSD2 WT interagierte T1150 mit Y1092 und L1120, wodurch diese Reste ausgerichtet, und das Volumen des aktiven Zentrums reduziert wurde. Die Hydroxylgruppe der Seitenkette von T1150 interagierte dabei mit dem Stickstoff der Backbone-Atome von Y1092. Zusätzlich war die Methylgruppe der Seitenkette in einem hydrophoben Kontakt mit der Seitenkette von L1120. Diese beiden Kontakte gingen in NSD2 T1150A aufgrund der Mutation verloren. Die Orientierung von Y1092 und L1120 war dadurch weniger strukturiert, und das Volumen des aktiven Zentrums vergrößerte sich. Die vorgestellten Ergebnisse erklären präzise den molekularen Mechanismus hinter der experimentell beobachteten veränderten Produktspezifität für NSD2 T1150A.

Jüngst wurde die Substratspezifität von der PKMT SETD2 mit Hilfe von Cellusspots-Peptid-Array-Methylierung hinsichtlich der natürlichen H3K36-Zielsequenz kartiert. Dabei wurde festgestellt, dass die H3-Aminosäuresequenz an einigen Positionen nicht ideal war. Basierend darauf, wurde ein artifizielles Peptidsubstrat entworfen, welches an jeder Stelle die optimale Aminosäure enthielt. Methylierungsexperimente zeigten, dass das 15 Aminosäuren lange Super-Substratpeptid (ssK36), welches an vier Positionen von der natürlichen H3-Sequenz abweicht, mehr als 100-mal schneller methyliert wurde als das H3K36-Peptid. Die Kristallstruktur von SETD2 mit gebundenem ssK36-Peptid erklärte die stark erhöhte Methylierungsaktivität jedoch nicht vollständig. Der zweite Ansatz dieser Arbeit fokussiert sich auf die mechanistischen Gründe hinter diesem massiven Anstieg der Reaktionsrate. Um dies zu erreichen, wurde eine Kombination von *in vitro* Methylierungs- und FRET-Experimenten sowie MD- und sMD Simulationsmethoden angewendet, um mehrere Schritte des katalytischen Prozesses abzubilden. MD Simulationen der freien Peptide in Lösung zeigten, dass ssK36 präferentiell eine Haarnadelkonformation ausbildete, während H3K36 eher gestreckt vorlag. Diese

Präferenz basierte auf den vier eingeführten Mutationen. Darüber hinaus wurde in sMD Simulationen gezeigt, dass Peptide in Haarnadelkonformationen einen besseren Zugang in das aktive Zentrum von SETD2, im Vergleich zu gestreckten Strukturen hatten. Methylierungsexperimente belegten, dass durch chemisch induzierte Haarnadelstrukturen in Peptiden, die Methylierungsaktivität von SETD2 erhöht werden konnte. Zusätzlich wurde in MD Simulationen des ssK36-SETD2-Komplexes beobachtet, dass die vier Mutationen ein einzigartiges Kontaktprofil mit SETD2 ausbildeten. Dadurch wurden einerseits unterschiedliche aber auch eine höhere Anzahl an Übergangszustands-ähnlichen Konformation im Vergleich zum H3K36-SETD2-Komplexen ausgebildet.

Die Übertragbarkeit dieses Ansatzes wurde durch ein neues Super-Substrat-Peptid, welches speziell für NSD2 entworfen wurde, demonstriert. Der molekulare Mechanismus hinter der erhöhten Mehtylierungsrate wurde erneut durch MD Simulation, untersucht. Bemerkenswerterweise zeigten SETD2 und NSD2 spezifisch nur für ihr eigenes Super-Substrat eine erhöhte Aktivität und nicht für das jeweils andere Super-Substrat.

Die optimierten Interaktionen der Super-Substratpeptide wurden anschließend als Ausgangspunkt verwendet, um einen PKMT-spezifischen Inhibitionstest zu etablieren, bei welchem ssK36 als substratkompetitiver SETD2-spezifischer Inhibitor fungierte.

Zusammenfassend zeigten die in dieser Arbeit durchgeführte MD Simulationen bisher unbekannte Gründe für den dominanten Effekt von DNMT3A R882H im heterozygoten Zustand und erklärten die veränderte Produktspezifität für NSD2 T1150A. Darüber hinaus wurden die Merkmale des artifiziellen entworfenen Super-Substratpeptides, durch verschiedene MD Simulationsansätze beschrieben und durch biochemische Experimente validiert. Die, in dieser Arbeit gefundenen, molekularen Mechanismen für DNMTs und PKMTs erklären biochemische Ergebnisse in atomarer Auflösung und zeigen neue Strategien für die Gestaltung einer neuen Klasse von substratkompetitiven PKMT-Inhibitoren auf.

## Acknowledgements

My deepest appreciation goes to my supervisor Prof. Dr. Albert Jeltsch for more reasons that one could list here. I am immensely grateful for the guidance and mentorship throughout my academic journey. From the very first lecture in my bachelor's study - to the opportunity to conduct my PhD in his institute to collaborative efforts on publishing papers, patents, reviews, Professor Jeltsch's support, professionalism, and advice have been instrumental in shaping my understanding of science and research skills. His commitment to fostering academic growth and excellence has left an indelible mark on my academic pursuits, and I am fortunate to have had the privilege of working alongside such a dedicated mentor.

I extend my genuine gratitude to Prof. Dr. Jürgen Pleiss for his persistent support, innovative ideas, and key role in building a yet unseen GPU infrastructure tailored for Molecular Dynamics simulation. His support greatly facilitated my academic journey. From the beginning of our collaboration, Prof. Pleiss has consistently demonstrated a commitment to raising an environment of academic quality.

I am thankful for Prof. Dr. Martin Zacharias from the Technical University of Munich to be my second examiner. Special thanks go to Prof. Dr. Markus Morrison for his kind acceptance to participate in my examination committee.

Special thanks to Dr. Sara Weirich for her guidance in the lab, which made me more confident and independent. Additionally, I want to highlight Dr. Mina Saad Khella, and Michael Choudalakis who are not only co-authors in publications, but showed incredible support in developing techniques and approaches presented in this work. Moreover, thanks to the rest of the PKMT-group for the teamwork atmosphere, valuable discussions and continuous help. Additionally, I am thankful for Dr. Philipp Rathert for valuable comments in seminars and meetings. I would like to thank the people behind the scenes for making the work smooth and perfectly organized: Priv. Doz. Dr. Hans. Rudolph, Elisabeth Tosta and Lea Irsigler. Very special thanks to Regina, Dragica and Branka for all their helping hand, nice talks and laughs.

I would like to express my sincere gratitude to SimTech EXC 2075 390740016 for providing the financial support that made me pursue my PhD. Especially my project network PN2-5 for the insightful discussions.

My appreciation goes to Dr. Sven Benson for his pivotal role in sharpening my understanding for challenges outside of the university. His mentorship not only strengthened my scientific courage but also exposed me to the significance of navigating challenges within “yesterday” timelines.

I am grateful for Dr. Peter Stockinger, my steadfast companion, from the moment we first met until the present. He has been more than a study friend – he has been my partner in crime. His enduring

support and camaraderie have made the academic journey to something more, turning challenges into shared triumphs.

I am deeply thankful for the committed support of my parents throughout my academic journey. Their encouragement and sacrifices have been the bedrock of my success, with my father's profound wisdom and guidance serving as a beacon, steering me through challenges and enriching my personal and scholarly growth.

## List of publications

**Schnee P**, Choudalakis M, Weirich S, Khella M. S, Carvalho H, Pleiss J, Jeltsch A\*, (2022) Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide. *Commun. Chem.* 5, 139, doi.org/10.1038/s42004-022-00753-w

**Schnee P**, Weirich S, Jeltsch A\*, (2023) Charakterisierung der Substratspezifität von Protein Methyltransferasen – Methoden und Anwendungen. *BIOspektrum* 29, 249-251, doi.org/10.1007/s12268-023-1930-y

**Schnee P**, Pleiss J, Jeltsch A\*, (2024) Approaching the catalytic mechanism of protein lysine methyltransferases by biochemical and simulations techniques. *Critical Reviews In Biochemistry & Molecular Biology* 7, 1-49, doi.org/10.1080/10409238.2024.2318547

**Schnee P**, Jeltsch A, Weirich S (2023) Artifizielles Peptid mit PKMT-inhibitorischer Wirkung - Universität Stuttgart PCT-Patentanmeldung

Khella M. S., **Schnee P**, Weirich S, Bui T, Bröhm A, Bashtrykov P, Pleiss J, Jeltsch A\*, (2023) The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation. *J Biol Chem* 104796, doi.org/10.1016/j.jbc.2023.104796

Weirich S, Kusevic D, **Schnee P**, Reiter J, Pleiss J, Jeltsch A\*, (2024) Discovery of new NSD2 non-histone substrates and design of a super-substrate – *Communications Biology*, Manuscript submitted for publication.

Mack A, Emperle M, **Schnee P**, Adam S, Pleiss J, Bashtrykov P, Jeltsch A\*, (2022) Preferential Self-interaction of DNA Methyltransferase DNMT3A Subunits Containing the R882H Cancer Mutation Leads to Dominant Changes of Flanking Sequence Preferences. *J Mol Biol* 15;434(7):167482, doi.org/10.1016/j.jmb.2022.167482.

## Author contributions

Schnee et al. 2022, Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide: **P.S.** conducted the MD simulation and fluorescence spectroscopy experiments. **P.S.** and A.J. did the data analysis and interpretation of the data. **P.S.** and A.J. prepared the manuscript and figures.

Schnee et al. 2023, Charakterisierung der Substratspezifität von Protein Methyltransferasen Methoden und Anwendungen: **P.S.** created the first draft of the manuscript.

Schnee et al. 2024, Approaching the catalytic mechanism of protein lysine methyltransferases by biochemical and simulations techniques: **P.S.** and A.J. prepared the draft manuscript. **P.S.** prepared the figures.

Khella et al. 2023, The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation: **P. S.** designed and conducted the MD simulation experiments and data analysis thereof. **P.S.** prepared the draft and figures for the MD simulation part of the manuscript.

Weirich et al. 2024, Discovery of new NSD2 non-histone substrates and design of a super-substrate: S.W. and A.J. devised the study. D.K. and S.W. conducted the biochemical experiments. **P.S.** conducted the MD simulations and the data analysis thereof. **P.S.** prepared the draft and figures for the MD simulation part of the manuscript.

Mack et al. 2022, Preferential Self-interaction of DNA Methyltransferase DNMT3A Subunits Containing the R882H Cancer Mutation Leads to Dominant Changes of Flanking Sequence Preferences: **P.S.** performed and analyzed the MD simulations. **P.S.** prepared the draft and figures for the MD simulation part of the manuscript.

## List of figures

<b>Figure 1:</b> DNA and protein methylation influence gene transcription by recruiting chromatin remodeling enzymes .....	2
<b>Figure 2:</b> DNA Methyltransferase (DNMT) 3A and its DNMT3L cofactor form a heterotetramer.....	4
<b>Figure 3:</b> Protein Lysine Methyltransferases (PKMTs) transfer up to three methyl groups to specific lysine residues in proteins.....	7
<b>Figure 4:</b> Binding mode of cofactor SAM and protein substrate for SET and non-SET domain-containing PKMTs.....	8
<b>Figure 5:</b> Cartoon representation of multiple SET domain PKMT architectures.....	10
<b>Figure 6:</b> The autoinhibitory loop (AL) and placeholder residue need conformational changes to overcome autoinhibition.....	12
<b>Figure 7:</b> PKMTs deprotonate the target lysine prior to the methyl group transfer .....	14
<b>Figure 8:</b> Geometric criteria for a bimolecular nucleophilic substitution ( $S_N2$ ) mechanism.....	15
<b>Figure 9:</b> Substrate specificity profile of SETD2 led to the super-substrate peptide (ssK36) .....	17
<b>Figure 10:</b> Proposed control mechanism for the product specificity of PKMTs .....	19
<b>Figure 11:</b> Reaction mechanisms of lysine demethylation. ....	24
<b>Figure 12:</b> Binding mode of PKMTs to nucleosomal substrates.....	26
<b>Figure 13:</b> Molecular dynamics simulation use bonded and non-bonded forces to model the interactions between atoms .....	28
<b>Figure 14:</b> DNMT3A R882H establishes more and different contacts in the RD interface and different interactions with the DNA.....	44
<b>Figure 15:</b> Product specificity change of NSD2 T1150A and NSD1 T2029A compared to WT enzymes on H3.1 protein and nucleosomes <i>in vitro</i> . ....	47
<b>Figure 16:</b> sMD simulation of SAM association into the complex of either NSD2 WT or T1150A with a H3K36me1 or me2 peptide substrate .....	49
<b>Figure 17: Measurement of the active site pocket volume in NSD2 WT and T1150A complexes .....</b>	<b>51</b>
<b>Figure 18:</b> The natural H3K36 peptide differs at four positions from the artificially designed super-substrate peptide (ssK36).....	53
<b>Figure 19:</b> Clustering of the H3K36 and ssK36 peptide conformations in solution observed in MD simulations reveals a hairpin conformation preference for ssK36 .....	55
<b>Figure 20:</b> H3K36 and ssK36 peptides show different conformational preferences in solution .....	57
<b>Figure 21:</b> Hairpin conformations facilitate the access of peptides into the binding cleft of SETD2 ...	59
<b>Figure 22:</b> H3K36 and ssK36 peptides unfold upon binding to SETD2.....	61
<b>Figure 23:</b> Hairpin formation and resolution upon binding increase SETD2 methylation activity. ....	62



<b>Figure 24:</b> The complex of SETD2-ssK36 established significantly more TS-like conformations than SETD2-H3K36.....	64
<b>Figure 25:</b> Contact profiles of the H3K36 and ssK36 peptides bound to SETD2 observed in the MD simulations .....	66
<b>Figure 26:</b> The enhanced methylation activity of SETD2 towards ssK36 can be summarized as four steps. ....	67
<b>Figure 27:</b> The ssK36 peptide functions as a substrate-competitive inhibitor for SETD2.....	69
<b>Figure 28:</b> The complex of NSD2 and ssK36(NSD2) establishes more TS-like conformations than the NSD2-H3K36 complex.....	71
<b>Figure 29:</b> Contact profile analysis reveals different contact maps for H3K36 and ssK36 in complex with NSD2 .....	72
<b>Figure 30:</b> The H3K36 and ssK36(NSD2) peptide establish different contacts with NSD2 .....	74
<b>Figure 31:</b> Possible mechanism for the binding of a PKMT towards a nucleosome and the recognition of the target lysine .....	85

## List of abbreviations

Ade	Adenine
AL	Autoinhibitory Loop
AML	Acute Myeloid Leukemia
AOL	Amine Oxidase Like
ASH1L	Absent, Small, or Homeotic Discs 1-Like
ASH2L	Absent, Small, or Homeotic Discs 2-Like
AWS	Associated With SET
CpG	Cytosine-Guanine
CTD	C-terminal Domain
Cyt	Cytosine
DNA	Deoxyribose Nucleic Acid
DNMT1	DNA Methyltransferase 1
DNMT3A	DNA Methyltransferase 3A
DNMT3B	DNA Methyltransferase 3B
DNMT3L	DNMT3-like
DNMTs	DNA Methyltransferases
DTT	Dithiothreitol
DOT1L	Disruptor of Telomeric silencing 1-like
EDTA	Ethylenediaminetetraacetic Acid
EM	Electron Microscopy
EZH2	Enhancer of Zeste Homolog 2
FAD	Flavine-Adenine Dinucleotide
FRET	Förster Resonance Energy Transfer
GLP	G9a Like Protein
GPU	Graphical Processing Unit
GST	Glutathione-S-Transferase
Gua	Guanine
G9a	aka EHMT2 (Euchromatic histone-lysine N-methyltransferase 2)
H1	Histone H1
H2A	Histone H2A
H2B	Histone H2B
H3	Histone H3

H3K27me	Histone H3 lysine 27 methylation
H3K36me	Histone 3 lysine 36 methylation
H3K4me	Histone 3 lysine 4 methylation
H3K9me	Histone H3 lysine 9 methylation
H4	Histone H4
HDACs	Histone Deacetylase Complexes
HP1	Heterochromatin Protein-1
KDM	Lysine Demethylase
Kme1	Monomethyl lysine
Kme2	Dimethyl lysine
Kme3	Trimethyl lysine
LSD	Lysine Specific Demethylase
MALDI	Matrix-assisted Laser Desorption/Ionization
MBD	Methylated DNA Binding Domain
MD	Molecular Dynamics
MLL	Mixed Lineage Leukemia
MYND	Myeloid Translocation Protein 8, Nery, and DEAF-1
NSD1/2/3	The Nuclear Receptor-Binding SET domain 1/2/3
PDB	Protein Data Bank
PMT	Protein Methyltransferase
PKMT	Protein Lysine Methyltransferase
PRDM9	PR Domain Containing 9
PTM	Post Translational Modification
PWWP	Proline-Tryptophan-Tryptophan-Proline domain
QM/MM	Quantum Mechanics/Molecular Mechanics
RNA	Ribose Nucleic Acid
RNAP II	RNA Polymerase 2
SAH	S-Adenosyl-L-Homocysteine
SAM	S-Adenosyl-L-Methionine
SET	Suppressor of Variegation 3-9, Enhancer of Zeste, Trithorax
SETD2	SET Domain-containing protein 2
SET-I	SET Insertion domain
sMD	Steered Molecular Dynamics
SMYD	SET and MYND Domain-containing protein

S <sub>N</sub> 2	Nucleophilic Substitution
SRI	Set2–Rpb1 Interaction Domain
ssK36	Super-substrate around lysine 36
STAT1/3	Signal Transducer and Activator of Transcription 1/3
SUV39H1/2	Suppressor of Variegation 3-9 Homolog 1/2
TETs	Ten Eleven Translocation enzyme
Thy	Thymine
TS	Transition State
WHS	Wolf-Hirschhorn Syndrome
WRAD	WDR5, RbBP5, Ash2L and Dpy30 complex
5caC	5-carboxy cytosine
5fC	5-formyl cytosine
5hmC	5-hydroxymethyl cytosine
5mC	5-methyl cytosine

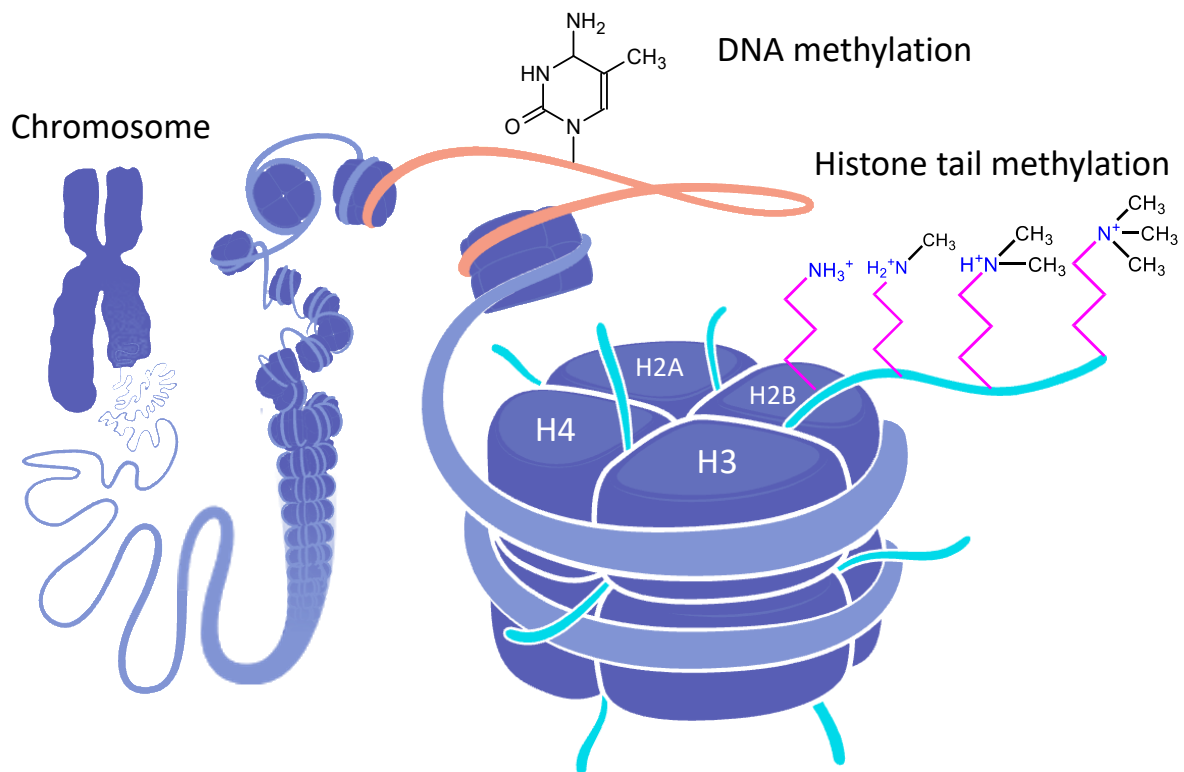
# 1. Introduction

## 1.1. Epigenetics

Every cell has its genetic information encoded as the base pair sequence in the DNA. However, cellular differentiation is driven by differences in the expression of genes. Epigenetics describes the mechanisms of these often stable, but still reversible, changes in gene expression patterns that do not involve alterations in the DNA sequence (Allis & Jenuwein, 2016). How the expression of genes is regulated under certain circumstances, is one of the key questions in epigenetics. The highly regulated and reversible changes add a dynamic layer of complexity beyond the static genetic code. A unilateral flow of DNA to RNA to protein is therefore no longer feasible, since proteins themselves regulate gene expression and react to environmental changes.

## 1.2. Chromatin structure and its regulation

Chromatin is the complex of proteins and DNA within the nucleus of eukaryotic cells. The smallest structural units of chromatin are nucleosomes, which consist of a stretch of about 146 base pairs of DNA wrapped around an octamer of histone proteins. The octamer contains two copies of the core histones, namely H2A, H2B, H3, and H4. Nucleosomes are connected by linker DNA segments, varying in length, forming a “beads-on-a-string” structure. This linear arrangement of nucleosomes can be further compacted to form higher-order chromatin structures, such as 30 nm chromatin fibers and highly condensed chromosomes. This leads to a remarkable size reduction, as the genetic information for a human cell lined up on a string would stretch over 2 meters. The compaction enables the DNA to fit into the eukaryotic nucleus with a diameter of 5-16  $\mu\text{M}$ . Besides space optimization, chromatin compaction plays a pivotal role in the regulation of gene expression. In condensed chromatin regions, nucleosomes are packed tightly, restricting the access of other proteins responsible for e.g. transcription, DNA replication and repair. This repressed state of chromatin is referred to as heterochromatin. Chromatin regions that are less condensed and accessible for gene transcription are referred to as euchromatin. The conversion of heterochromatin to euchromatin, or *vice versa*, is regulated by modifications on DNA- and protein level (**Fig. 1**).



**Figure 1: DNA and protein methylation influence gene transcription by recruiting chromatin remodeling enzymes.** Chromosomes consist of smaller subunits called nucleosomes, which themselves consist of a protein octamer with DNA wrapped around it. DNA methylation at cytosines, and lysine methylation of histone proteins are epigenetic modifications, which are read by reader enzymes, recruiting chromatin remodeling enzymes. Depending on the actual modifications, the chromatin structure is being tightened or loosened, directly influencing gene transcription.

### 1.3. DNA methylation

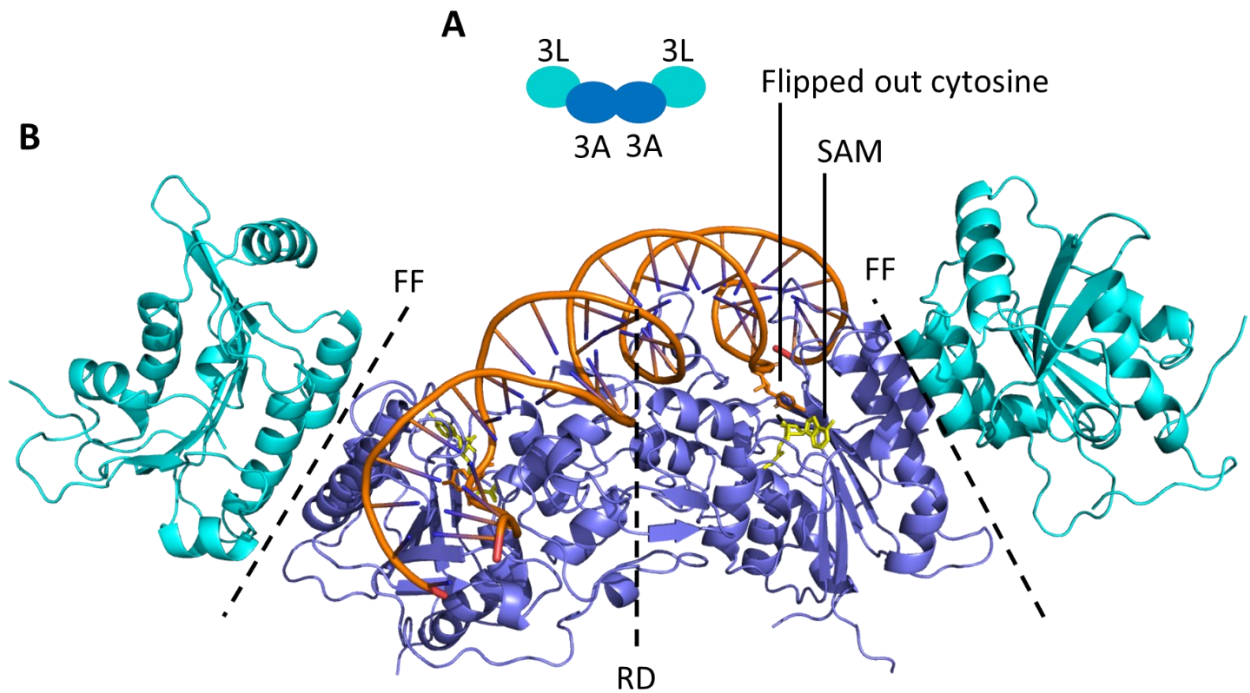
DNA methylation occurs at the fifth carbon (C5) in the pyrimidine base cytosine. The methylation reaction is catalyzed by DNA methyltransferases (DNMTs) using the cofactor S-adenosyl-L-methionine (SAM) as a methyl group donor, which is converted to S-adenosyl-L-homocysteine (SAH). DNA methylation in gene promoters is correlated with silenced gene expression. In general, it contributes to formation of heterochromatin and is a more stable modification compared to protein methylation. Moreover, DNA methylation is a crucial signal in many biological processes including development and gametogenesis, parental imprinting, X chromosome inactivation, as well as maintenance of genome integrity (Jurkowska, Jurkowski, et al., 2011). Cytosine-guanine (CpG) islands are genomic regions with a high frequency of cytosine-guanine dinucleotides, often associated with gene promoters and commonly unmethylated. At promoters, DNA methylation serves as a repressive signal, hindering the interaction of transcriptional activators and facilitating the recruitment of transcriptional repressors that incorporate methylated DNA binding domains (MBDs) (Razin & Riggs, 1980; Tate & Bird, 1993; Yin et al., 2017). Repressed promoters with methylated CpG islands are found in regions, where silencing

is desired, like centromeric heterochromatin, imprinted genes or transposons (Howard et al., 2008; Jurkowska, Jurkowski, et al., 2011).

### 1.3.1. DNA methyltransferases

In humans, three DNMTs catalyze the methylation of cytosine (Gowher & Jeltsch, 2018). Whereas DNA methyltransferase 1 (DNMT1) maintains the DNA methylation after DNA replication and works preferentially on hemimethylated CpG dinucleotide sites (Bestor et al., 1988; Fatemi et al., 2001; Goyal et al., 2006), DNA methyltransferase 3A and 3B (DNMT3A and 3B) are responsible for *de novo* DNA methylation. *De novo* methylation is important during early development, germ cell differentiation as well as imprinting, a process in which specific genes are marked with methyl groups based on their parental origin (Gowher & Jeltsch, 2001; Okano et al., 1998). DNMT3A and 3B methylate CpG and non CpG sites and have no preference for hemimethylated sequences, distinguishing them from DNMT1 (Gowher & Jeltsch, 2001). The DNMT3-like (DNMT3L) protein lacks catalytic activity, but serves as a scaffold protein for DNMT3A and DNMT3B, enhancing their *de novo* DNA methylation activity (Bourc'his et al., 2001). The catalytically active C-terminal domain (CTD) of DNMT3A in complex with the CTD of DNMT3L forms a linear heterotetrameric complex with the two DNMT3A subunits in the center and the DNMT3L at the edges (3L-3A-3A-3L, **Fig. 2A**) (Jia et al., 2007). The different subunits are connected by two interfaces. The DNMT3A/3L interface, called FF interface, and the central interface between the DNMT3A subunits, denoted as RD interface. The binding of DNMT3L at the FF interface helps to organize the active site and SAM binding pocket of DNMT3A, which explains the stimulation of DNMT3A activity (Jia et al., 2007). Crystal structures of the DNMT3A/3L complex bound to DNA showed that the DNMT3L subunits are not in contact with DNA, whereas the two DNMT3A subunits of the tetramer interact with two CpG sites of the substrate DNA, which involves the flipping of the target bases (**Fig. 2B**) (Zhang et al., 2018). The FF interface also supports DNMT3A/3A interactions, allowing the replacement of DNMT3L subunits in the DNMT3A/3L heterotetramer by two DNMT3A subunits yielding a DNMT3A homotetramer (Jurkowska et al., 2008; Jurkowska, Rajavelu, et al., 2011).

DNA methylation is a dynamic process and can be reversed either passively during DNA replication or actively by DNA demethylases enzymes called ten eleven translocation enzymes (TETs). TET enzymes oxidize 5-methyl cytosine (5mC) in progressive oxidation reactions resulting in 5-hydroxymethyl cytosine (5hmC), 5-formyl cytosine (5fC) and 5-carboxy cytosine (5caC) (Ito et al., 2011; Tahiliani et al., 2009).



**Figure 2: DNA Methyltransferase (DNMT) 3A and its DNMT3L cofactor form a heterotetramer. A|** Schematic representation of the tetrameric DNMT3A (blue) – DNMT3L (cyan) (DNMT3A/L) complex. **B|** Cartoon representation of the DNMT3A/3L complex with SAM (yellow), bound DNA and flipped out cytosine (orange) (Protein Data Bank (PDB) 6W8B). The positions of the two FF interfaces between DNMT3A and DNMT3L subunits and the RD interface between the central DNMT3A subunits are indicated as black, dashed lines. Figure taken and modified from (Mack et al., 2022).

#### 1.4. Lysine methylation

SAM-dependent methylation occurs not only in DNA but also in proteins or peptides and can be found at side chains of lysine (K), arginine (R), aspartate (D), glutamate (E), histidine (H), asparagine (N), glutamine (Q), and cysteine (C) (Clarke, 2013). Due to the lone-pair electrons present in the  $\epsilon$ -amine of lysine, as well as its preference for localization on the protein surface, lysine residues are a favorable target for posttranslational modifications (PTM) (Luo, 2018). Protein lysine methylation stands apart from other types of modifications like acetylation, ubiquitination or SUMOylation for three reasons. Firstly, the addition of methyl groups to lysine does not affect the overall charge of the residue at physiological pH, unlike acylation modifications that convert the positively charged  $\epsilon$ -amine into a neutral amide. Secondly, lysine methylation represents the smallest PTM, resulting in only minor changes in the size of the side chain compared to other types of lysine modifications (Luo, 2018). Thirdly, up to three methyl groups can be transferred to a target lysine creating monomethyl lysine (Kme1), dimethyl lysine (Kme2) and trimethyl lysine (Kme3).

Methyl lysine recognition is challenging, since lysine methylation only subtly alters the physiological properties. Still, the methylated lysines' ability to engage in cation- $\pi$  interactions (increased dispersion



of the positive charge around neighboring hydrocarbons) and the ability to form hydrogen bonds as a donor and acceptor is reduced with progressive methylation providing some options for discrimination and readout (Luo, 2018). A regular characteristic for methyl lysine-specific reader proteins is to recognize Kme2 and Kme3 methyl lysine groups through a hydrophobic pocket containing aromatic residues (e.g., F, Y, and W). Given that K, Kme1, Kme2, and Kme3 all carry an overall +1 formal charge at physiological pH, the aromatic pocket serves as a binding site for cation- $\pi$  interactions (Luo, 2018).

#### **1.4.1. Histone lysine methylation**

Histone tails are the flexible ends of the histone proteins extending from the histone octamer and are a key target for lysine methylation (**Fig. 1**). The methylated lysine residues serve as a platform for the recruitment of proteins and protein complexes that interpret and regulate this modification. Such effector proteins contain specific domains which recognize the position of the lysine residue in the histone tail sequence and its methylation state (Cornett et al., 2019; Hyun et al., 2017; Yun et al., 2011). Eventually, a signal cascade of downstream effects is triggered, influencing the activity of chromatin remodelers, which alter the accessibility of DNA and thus gene transcription (**Fig. 1**). In the case of histone 3 lysine 9 (H3K9) methylation, found in constitutive heterochromatin, a cooperative mechanism involving other histone modifications and DNA methylation is found to trigger gene silencing. Protein lysine methyltransferases like SUV39H1 and H2 deposit H3K9 methylation, subsequently recognized by heterochromatin protein 1 (HP1) via an aromatic cage in its chromo domain (Kumar & Kono, 2020). HP1 binds H3K9me2/3 and acts as a transcriptional repressor by preventing the association of transcription factors and RNA polymerase (Schoelz & Riddle, 2022). Beyond the steric effect, HP1 further recruits DNMTs, which methylate CpG sites adjacent to the methylated lysine. The methylated DNA then acts as a foundation for MBD binding, which in turn recruit Histone Deacetylases (HDACs) (Jones et al., 1998). HDACs remove histone acetylation, thereby increasing the histone's positive charge (Bannister & Kouzarides, 2011). This strengthens the interaction with the negatively charged DNA sugar-phosphate backbone, leading to chromatin compaction (Bannister & Kouzarides, 2011).

#### **1.4.2. Non-histone protein lysine methylation**

In addition to lysine methylation at histone tails, this modification was found at non-histone proteins like p53 (West & Gozani, 2011), E2F1 (Couture et al., 2006), STAT3 (Jinbo Yang et al., 2010) and the androgen receptor (Gaughan et al., 2011; Ko et al., 2011). Lysine methylation of non-histone proteins influences their functionality in multiple ways (Hamamoto et al., 2015). The methylation can serve as a signal to deploy other PTMs including ubiquitination, thereby affecting e.g. protein stability. Proteins

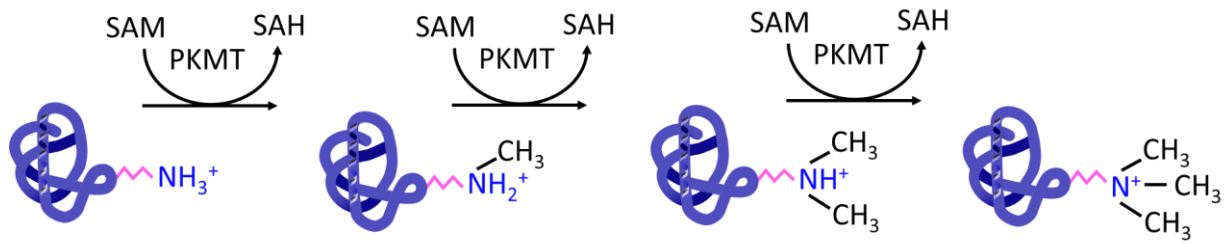
binding to methylated residues or other deployed PTMs: (i) stimulate or inhibit the target protein, (ii) regulate protein-protein interactions, (iii) affect the subcellular localization.

### 1.5. Protein lysine methyltransferases

The transfer of methyl groups from SAM to proteins is catalyzed by enzymes called protein methyltransferases (PMTs). If the methylated amino acid is a lysine residue, they are denoted as protein lysine methyltransferases (PKMTs) (**Fig. 3**). Over 60 characterized PKMTs are encoded in the human genome and can be categorized into two classes: SET domain-containing PKMTs (class V methyltransferase, called SET due to its discovery in the *Drosophila* enzymes named Suppressor of variegation 3-9, Enhancer of zeste, and Trithorax) and non-SET domain PKMTs (class I methyltransferases, also called 7-beta strand MTases) (Copeland et al., 2009; Falnes et al., 2016; Luo, 2012; Richon et al., 2011). Notably, more than 90% of PKMTs belong to the SET domain family (Luo, 2018). Still, non-SET domain-containing PKMTs and especially disruptor of telomeric silencing 1-like (DOT1L) as the most-studied representative member of this family, were shown to have major roles in cellular process and disease development (McLean et al., 2014; Nguyen & Zhang, 2011; Sarno et al., 2020).

PKMTs generally display a high specificity, targeting only defined lysine residues in one or few substrate proteins. Remarkably, histone lysine methylation is redundant, meaning that one lysine could be methylated by more than one PKMT. This redundancy has advantages: (i) different enzymes can be differentially regulated leading to a dramatic increase in the complexity of the regulatory network; (ii) PKMTs with a redundant substrate specificity can be recruited to different genomic loci like enhancers, promoters or gene bodies; (iii) PKMTs with redundant substrate specificities can transfer varying numbers of methyl groups. For example, PKMTs NSD1 (aka KMT3B), NSD2 (aka MMSET, WHSC1), NSD3 (aka WHSC1L1) and ASH1L transfer up to two methyl groups to histone H3 lysine 36 (H3K36) while SETD2 (aka KMT3A, HYPB, SET2) catalyzes trimethylation at the same lysine residue (Edmunds et al., 2008; Gregory et al., 2007; Li et al., 2009).

PKMTs transfer up to three methyl groups to a target lysine. They can do so in two different ways. In a distributive mechanism, each round of catalysis results in product dissociation and rebinding of a fresh substrate is needed for a second turnover. Hence, each methylation event is independent leading to the stochastic generation of Kme1, Kme2 and Kme3, depending on the product specificity of the PKMT. In contrast, in a processive reaction mechanism, multiple rounds of catalysis proceed on the same substrate before dissociation of the product (Gowher & Jeltsch, 2001; van Dongen et al., 2014).

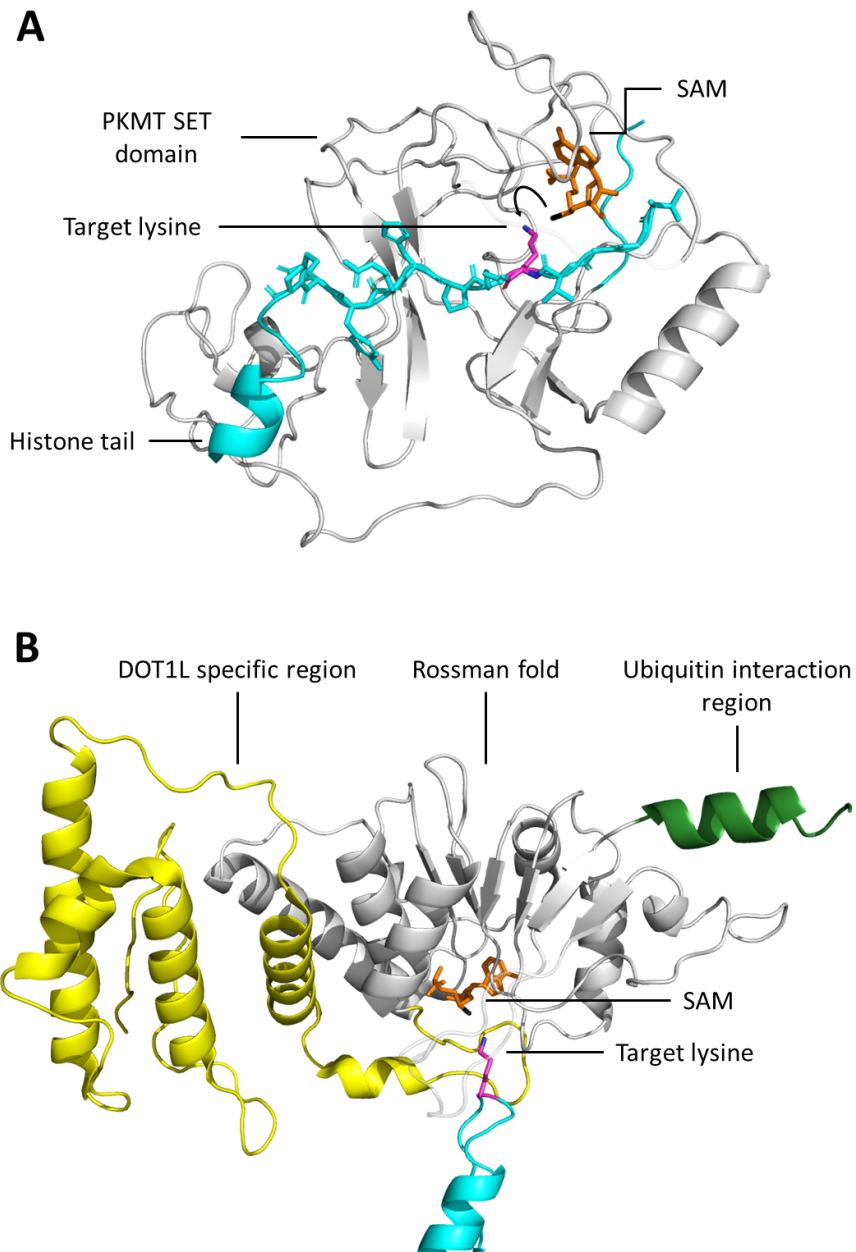


**Figure 3: Protein Lysine Methyltransferases (PKMTs) transfer up to three methyl groups to specific lysine residues in proteins.** The cofactor S-adenosyl-L-methionine (SAM) provides the methyl group. It is released after the transfer as S-adenosyl-L-homocysteine (SAH). Figure taken from (Schnee et al., 2023).

### 1.5.1. Structure of SET domain PKMTs

The SET domain of PKMTs is responsible for the methylation activity of this class of enzymes. It consists of approximately 130 amino acids, is often flanked by a pre-SET and post-SET domain (Qian & Zhou, 2006) and sometimes contains the domain insertion SET-I. SET domain-containing PKMTs bind the protein substrate and the methyl group providing cofactor SAM at opposing binding faces (Cheng et al., 2005). This is contrary to non-SET domain-containing PKMTs, like DOT1L, where the protein substrate and SAM are accommodated within a single, extended binding cleft (Min et al., 2003).

In the SET domain of PKMTs, the target lysine is brought in close proximity to the SAM methyl group through a hydrophobic tunnel. Here, the lysine hydrocarbon side chain interacts with tyrosine, phenylalanine and tryptophan residues via hydrophobic interactions (Qian & Zhou, 2006; Trievel et al., 2003). The positively charged  $\epsilon$ -amine group interacts with these residues through cation- $\pi$  interactions (Luo, 2018). After insertion, the  $\epsilon$ -amine group is oriented by multiple tyrosine residues and primed for the methyl group transfer. Meanwhile, SAM binds at the opposing site via contacts with its nucleobase and sugar moiety. The methyl group is then inserted into the active site and transferred to the deprotonated lysine  $\epsilon$ -amine group (**Fig. 4**). The detailed mechanistic features of different SET domain architectures, autoinhibition, lysine deprotonation and methyl group transfer are described in the following chapters.

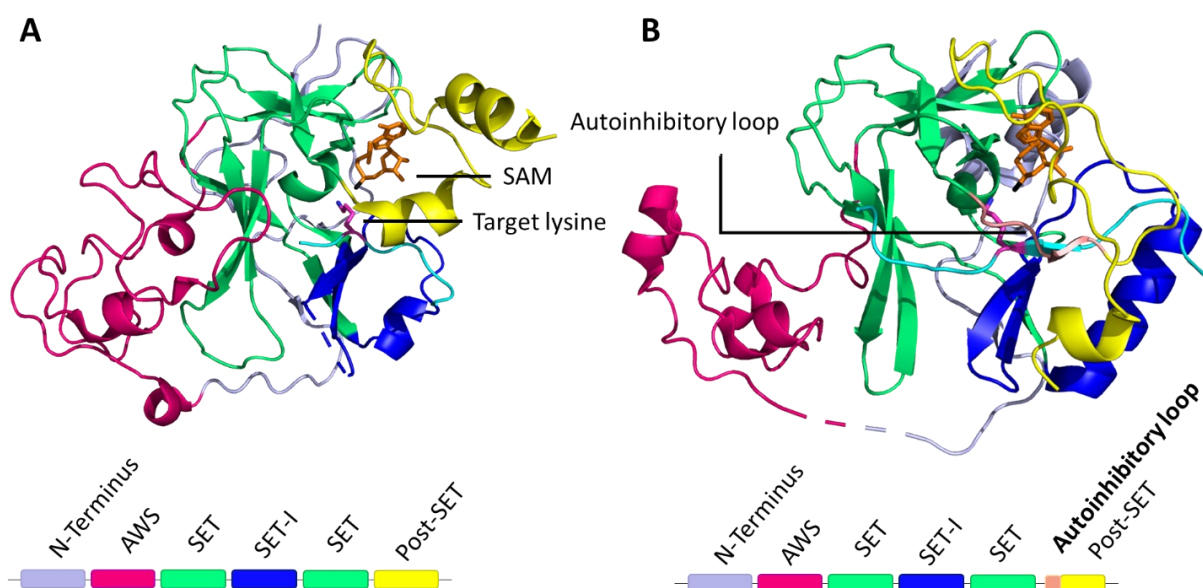


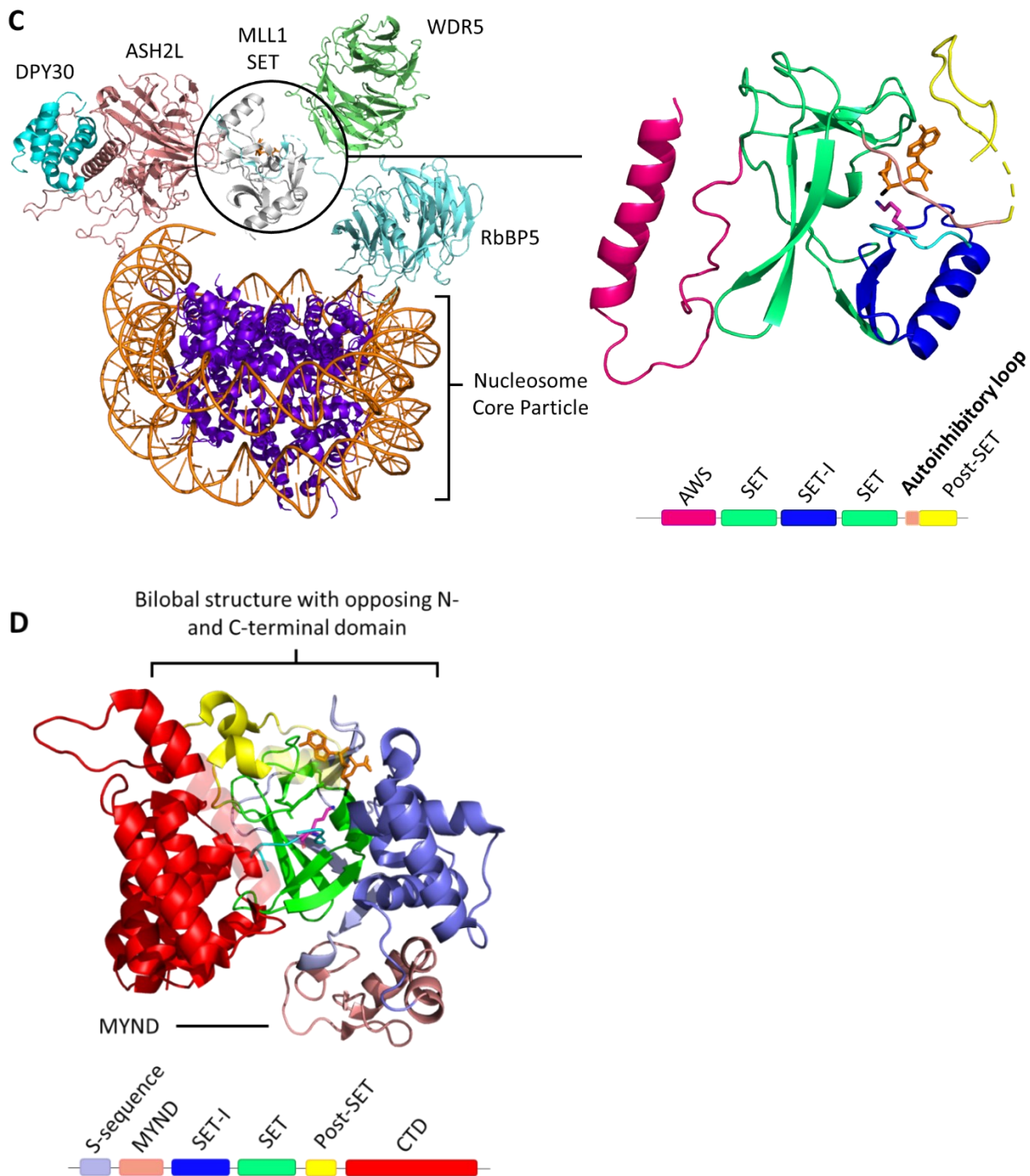
**Figure 4: Binding mode of cofactor SAM and protein substrate for SET and non-SET domain-containing PKMTs.** **A|** In SET domain PKMTs, the protein substrate (cyan) and SAM (orange, methyl group is colored black) bind at opposing sites. The target lysine (pink) is inserted into a narrow tunnel, where it undergoes deprotonation and is oriented for the methyl group transfer (image created using simulation results of PDB 6VDB). **B|** The non-SET domain-containing PKMT DOT1L binds the target lysine (pink) and cofactor SAM (orange, methyl group is colored black) in the same pocket (PDB 1NW3). The architecture consists of a DOT1L specific region (yellow), a 7-beta sheet Rossman fold (white) and a ubiquitin interaction region (forest green).

### 1.5.2. Different structural arrangements of SET domain PKMTs

Phylogenetic analysis of SET domain sequences revealed that human SET domain PKMTs can be classified into subfamilies, each characterized by unique architectures (Wu et al., 2010). G9a (aka EHMT2, KMT1C), SUV39H1 (aka KMT1A) and SUV39H2 (aka KMT1B) belong to the classical PKMT subfamily, where their SET domains catalyze the methyl group transfer without prior conformational changes (**Fig. 5A**) (Schnee et al., 2023; Tachibana et al., 2001). In contrast, NSD1, NSD2, NSD3 and SETD2 are part of the PKMT subfamily with an autoinhibitory loop (AL) (**Fig. 5B**) (An et al., 2011; Bennett et al., 2017; Yang et al., 2016). In this subfamily, the apo form of the SET domain is expected to have a highly reduced activity and needs to undergo conformational changes for substrate binding and enzyme activity. Other PKMTs act in complexes with additional proteins or contain specific domains to: (i) bind to certain structures like nucleosomes; (ii) recognize specific modifications on the substrate; and/or (iii) regulate their own activity. An example for this are Mixed lineage leukemia (MLL) SET domains, which are inactive on their own but become catalytically active in the presence of binding partners such as WDR5, RbBP5, ASH2L, and DPY30, collectively referred to as WRAD (**Fig. 5C**) (Borkin et al., 2015; Cao et al., 2014; Grebien et al., 2015).

The SET domain of PKMTs can feature insertions like the MYND domain (Myeloid translocation protein 8, Nery and DEAF-1). PKMTs with a MYND domain insertion represent the “SET and MYND Domain-containing protein” (SMYD) subfamily. The MYND domain is responsible for protein-protein interactions possibly recruiting the enzymes to specific substrate proteins. Additionally, SMYD enzymes are characterized by a bilobal architecture with the protein substrate in the middle (**Fig. 5D**) (Ferguson et al., 2011; Mazur et al., 2014; Mzoughi et al., 2016; Saddic et al., 2010; Sirinupong et al., 2011; Sirinupong et al., 2010).





**Figure 5: Cartoon representation of multiple SET domain PKMT architectures. A** | SET domain-containing PKMT G9a complexed with the 9 amino acid long H3K36 peptide (cyan) with the target lysine (pink), and cofactor SAM (orange, PDB 5JIY). SET domain-containing PKMTs incorporate zinc ions for structural stability in their “associate with SET” (AWS) domain (magenta), post-SET (yellow) or MYND (rose) domain depending on the enzyme (Dillon et al., 2005; Wu et al., 2011). However, they are not involved in catalysis or conformational changes. For simplicity, zinc ions are therefore not shown in protein structures presented in this work. **B** | SETD2 complexed with the 14 amino acid long H3K36 peptide, and cofactor SAM (PDB 5JLB). The autoinhibitory loop (rose) is in an open position to accommodate the protein substrate. **C** | MLL1 SET domain (white) associated with WDR5 (green), RbBP5 (light blue), ASH2L (rose) and DPY30 (cyan) bound to a nucleosome core particle (PDB 6PWV). MLL1 SET domain complexed with the 8 amino acid long H3K4 peptide (PDB 6UH5) **D** | SMYD2 complexed with the 10 amino acid long peptide  $E\alpha$  (PDB 4O6F). Distinct features are the bilobal or clamshell-like structure and the MYND domain. Figure taken from (Schnee et al., 2023)

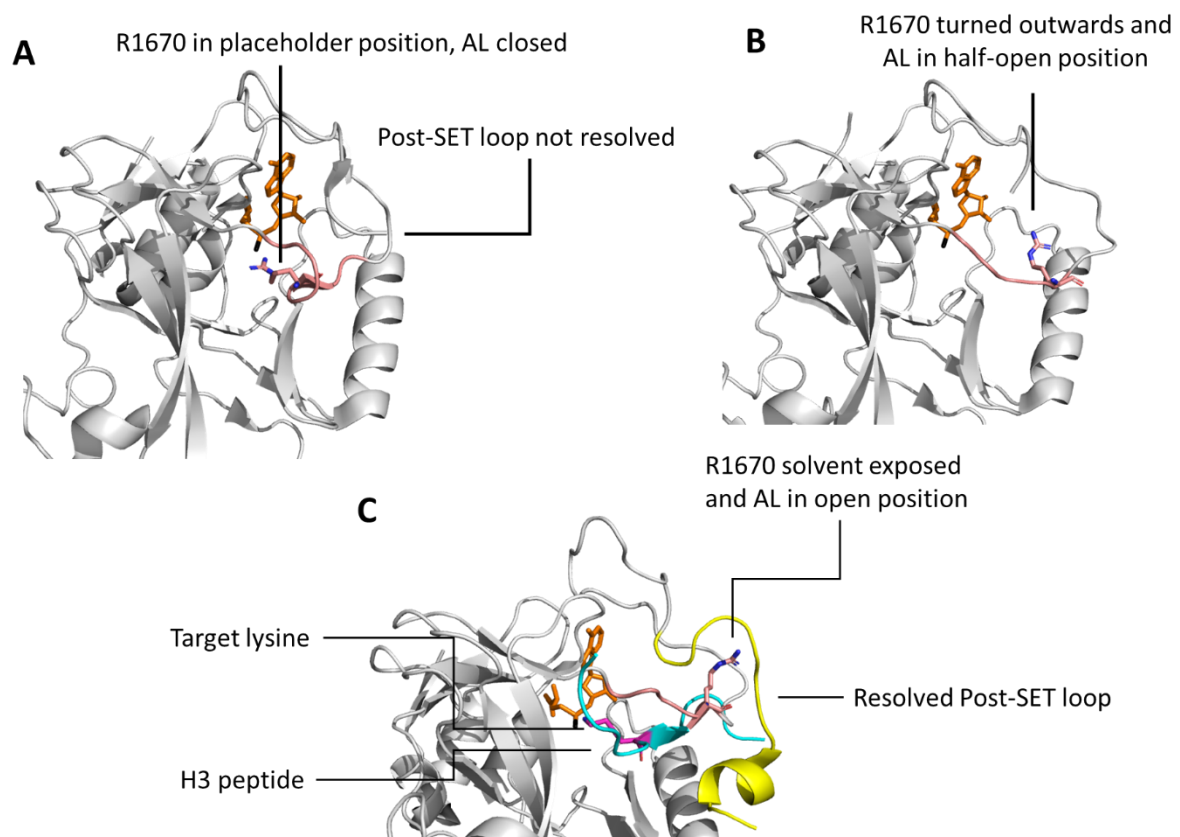
### 1.5.3. Autoinhibition of SET domain PKMTs

NSD1 was one of the first PKMTs for which an AL was described. Crystal structure analysis and Molecular Dynamics (MD) simulations of NSD1 with bound cofactor SAM, but without bound substrate, showed that a loop of approx. 14 amino acids is placed on top of the substrate binding cleft, effectively blocking the entrance of a target peptide (**Fig. 5B**) (Triebel et al., 2003; Xiao et al., 2003; Zhang et al., 2003). The AL is positioned between the SET and Post-SET domain. Multiple PKMTs were shown to have an AL in their structure, but their sequences are not conserved (Couture et al., 2005; Xiao et al., 2005). Studies on ASH1L demonstrated that stabilizing the closed position of the AL, achieved by enforcing hydrophobic interactions between AL and enzyme through mutations, decreased the ASH1L methylation activity (Rogawski et al., 2015). Adding to this, the AL was speculated to regulate the product specificity of PKMTs. Mutational studies of the ASH1L AL turned the enzyme from a dimethyltransferase to a trimethyltransferase (An et al., 2011; Rogawski et al., 2015). This result may provide an explanation for different product specificities among PKMTs with high sequence similarity in the active site, but not in the AL (Schnee et al., 2023). This was postulated for PKMTs NSD1, DIM-5 and SETD2, which share a high active site sequence similarity but possess different AL residues and exhibit differing product specificities (SETD2 and DIM-5 are trimethyltransferases, NSD1 is a dimethyltransferase) (Qiao et al., 2011). Together, these findings suggested a regulatory role of the AL. However, the MD simulation experiments and crystal structure analysis, which led to this conclusion were conducted with peptides as substrates. The mechanistic principles for the interaction with larger substrates like nucleosomes remain to be described.

### 1.5.4. Placeholder residues

The AL plays a pivotal role in regulating the substrates binding of SET domain PKMTs by sterically blocking the binding cleft. In addition to its steric hindrance, the AL was also found to position residues directly in the active site, at the position of the target lysine. Notably, specific residues function as “placeholder” residues in this context, stabilizing the AL in its closed conformation. For instance, NSD1 employs C2062 as a placeholder residue (Morishita & di Luccio, 2011), NSD2 uses C1183 (Jaffe et al., 2013), SETD2 relies on R1670 (Yang et al., 2016), and ASH1L uses S2259 (Yang et al., 2016). Mutational studies on ASH1L demonstrated the impact of such interactions, where the placeholder residue serine was exchanged for methionine, which might establish stronger interactions with the hydrophobic lysine binding tunnel. This strengthened the closed conformation of the AL and led to a heavily decreased methylation activity (Rogawski et al., 2015). Remarkably, methionine was not found to be a placeholder residue in any SET domain PKMT, indicating that its binding strength into the active site might be too strong (Schnee et al., 2023). In the case of SETD2, the placeholder residue R1670 can adopt multiple conformations (Yang et al., 2016). In the AL closed state, R1670 was observed to occupy

the lysine binding tunnel (**Fig. 6A**). Crystal structures depicting the AL in a half-open position showed R1670 slightly flipped outwards, away from the active center (**Fig. 6B**). Furthermore, structures with the AL in a fully open conformation, and substrate bound, showed R1670 completely flipped outwards and exposed to the solvent. Moreover, the usually unresolved Post-SET loop (Q1676-K1703 for SETD2) was captured in front of the bound peptide substrate, engaging in hydrophobic interactions with the core enzyme (**Fig. 6C**). This loop is not resolved in crystal structures without bound peptide, indicating its high flexibility in this state.



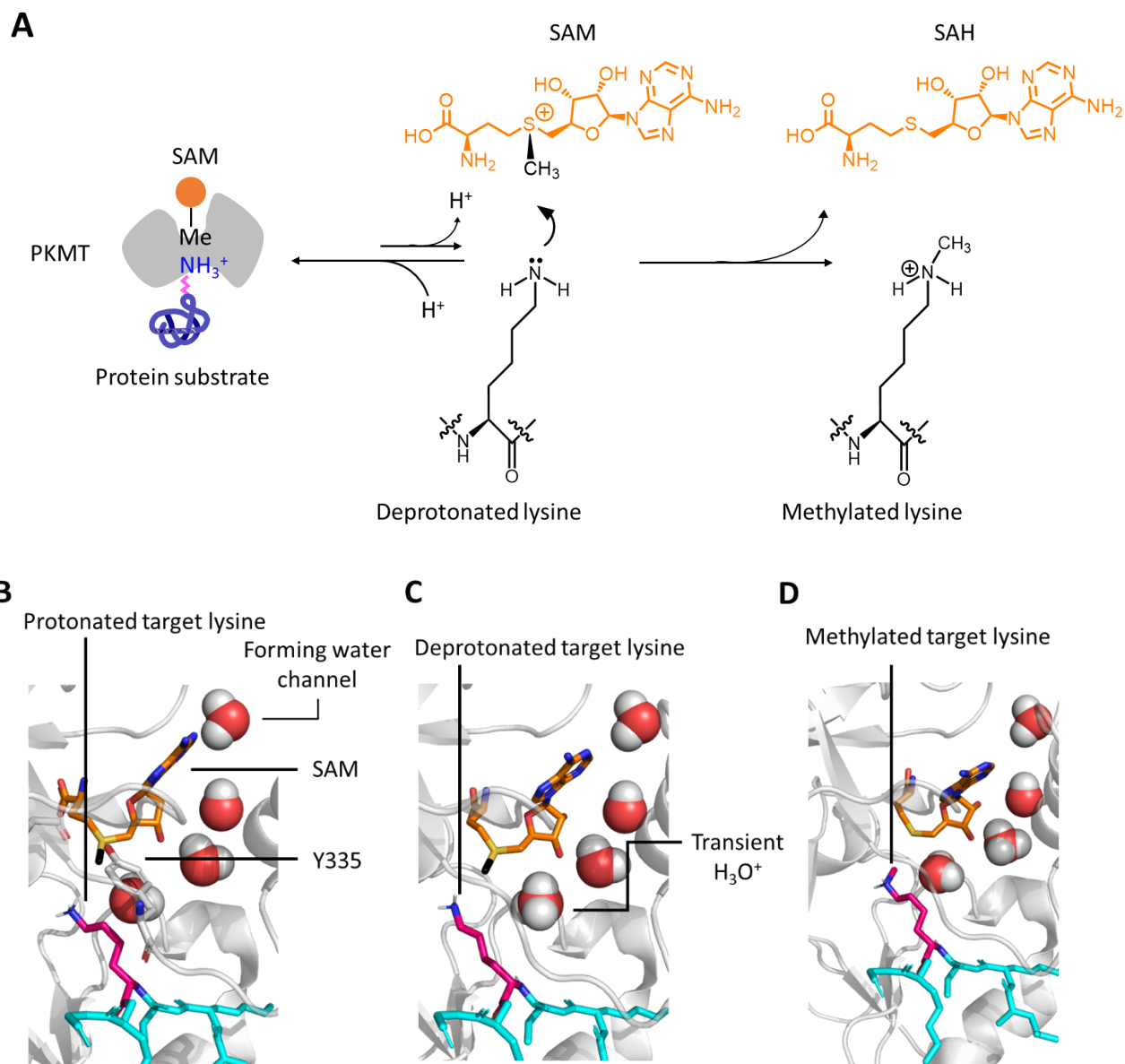
**Figure 6: The autoinhibitory loop (AL) and placeholder residue need conformational changes to overcome autoinhibition. A** | In the binary PKMT-SAM state, the placeholder residue occupies the target lysine channel. In SETD2 the placeholder residue R1670 (rose) can adopt multiple conformations. If no peptide is bound, the AL is in a closed position and R1670 occupies the target lysine channel (PDB 4H12). **B** | In a half-opened position, the AL starts to lift, and R1670 turns outwards (PDB 5JLE). **C** | When a peptide substrate (cyan, target lysine in pink) is bound, the AL is in an open position and R1670 becomes solvent exposed. The Post-SET loop (yellow) is closed on top of the bound peptide (PDB 5JLB). Figure taken and modified from (Schnee et al., 2023).

### 1.5.5. Target lysine deprotonation

The side chains of K, Kme1 and Kme2 possess lone-pair electrons on their  $\epsilon$ -amine groups, making them targets for methylation. However, due to their high  $pK_a$  values (10.2–10.7), K, Kme1, and Kme2 predominantly exist in a protonated state under physiological conditions (pH 7.4), where they are unreactive as a nucleophile. Therefore, one critical requirement for lysine methylation catalyzed by



PKMTs is the deprotonation of the target lysine (**Fig. 7A**) (Trievel et al., 2002). To explain the deprotonation mechanism, MD and Hybrid Quantum Mechanics/Molecular Mechanics Quantum (QM/MM) simulations have been employed, showcasing a conserved mechanism for the deprotonation of the target lysine in multiple SET domain-containing PKMTs (Zhang & Bruice, 2007b). In this mechanism, the deprotonation of the side chain nitrogen ( $N\epsilon$ ) occurs through the transient formation of dynamic water channels in the enzyme's active site (Hu & Zhang, 2006; X. Zhang & T. Bruice, 2008a, 2008b). A water molecule, which was frequently observed in the crystal structures of SET domain-containing PKMTs, was suggested to transfer the proton through a chain of water molecules into the aqueous solvent and finally to a buffer molecule (**Fig. 7B-D**). Additionally, electrostatic interactions between the positive charges of the SAM sulfonium moiety and the protonated  $N\epsilon$  atom decrease the  $pK_a$  of the latter from 10.9 to 8.2 (Zhang & Bruice, 2007b). This could explain the necessity of a basic reaction buffer for PKMTs and the weak *in vitro* methylation activity in acidic and even neutral buffers, as the deprotonation of the target lysine is impeded (Wilson et al., 2002; Zhang et al., 2002). Based on this, Bruice and Zhang suggested a stepwise process, in which: (i) the water channel appears; (ii) the target lysine is deprotonated; (iii) the target lysine is methylated using the cofactor SAM; (iv) the proton is transferred into the solvent (Zhang & Bruice, 2007a, 2007b, 2007c; X. Zhang & T. Bruice, 2008a, 2008b, 2008c). This model is applicable to multiple SET domain-containing PKMTs but not to non-SET containing PKMTs like DOT1L (**Fig. 4B**). In the class of non-SET domain PKMTs, a water channel was not observed, and the amino acids located at the target lysine channel appear incapable of facilitating a direct deprotonation. It was speculated that their more hydrophobic active site could reduce the  $pK_a$  of the target lysine and that the carboxylate of SAM could help in the subsequent deprotonation process (Cheng et al., 2005; Cortopassi et al., 2016; Min et al., 2003).

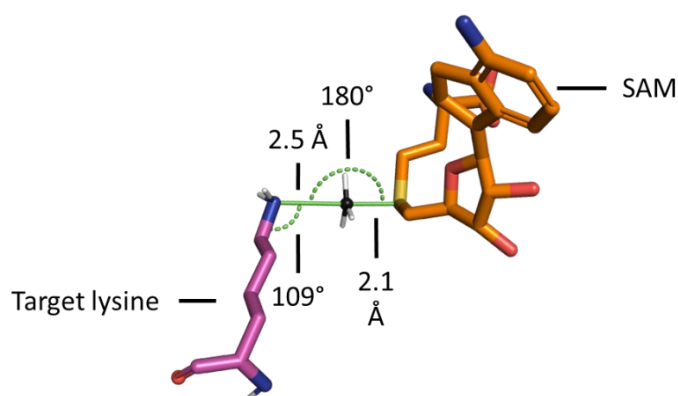


**Figure 7: PKMTs deprotonate the target lysine prior to the methyl group transfer. A|** Schematic depiction illustrating the obligatory target lysine deprotonation prior to the PKMT-catalyzed methyl group transfer. **B|** The protonated target lysine (pink) is oriented by e.g., PKMT SET7/9 Y335 (white, sticks), while the water channel (red spheres) is already present (prepared using PDB 1XQH). **C|** The lysine proton is transferred to a nearby water molecule. **D|** After lysine deprotonation, the SAM methyl group is rapidly transferred to the deprotonated target lysine thereby preventing reprotonation. The excess proton is transferred into the bulk solvent. **B-D|** Figure taken and modified from (Schnee et al., 2023).

### 1.5.6. Reaction mechanism of SET domain PKMTs

After a successful deprotonation, target lysine and SAM must be oriented in a conformation that facilitates the subsequent bimolecular nucleophilic substitution reaction ( $S_N2$ ), leading to the methyl group transfer. QM/MM simulations of SET7/9 (aka SETD7, SET7, SET9, KMT7) were first to describe the details of the  $S_N2$  mechanism (Hu & Zhang, 2006). In this mechanism, the deprotonated lysine  $N\epsilon$  acts as the nucleophile, whereas the SAM sulphonium cation ( $S^+$ ) as the leaving group. The free

electron pair of N $\epsilon$  is present in a sp<sup>3</sup> orbital at an 109° angle. The S<sub>N</sub>2 reaction occurs at an aliphatic sp<sup>3</sup> carbon center (the C-atom of the transferred methyl group), with the electronegative sulphonium leaving group attached to it. The nucleophile attacks the carbon at a minimal distance of approximately 4.4–4.6 Å (Chen et al., 2019). A combination of computational modeling, QM/MM and kinetic isotope effect studies have demonstrated that PKMTs can stabilize two distinct transitions states (TS) when methylating substrates (Chen et al., 2019; Linscott et al., 2016; Poulin et al., 2016). The SET8 enzyme (aka Pr-SET7, SETD8, KMT5A) exhibits an early S<sub>N</sub>2 TS (with a C–S distance of 2.0 Å and a C–N $\epsilon$  distance of 2.4 Å), while a late S<sub>N</sub>2 TS was observed for NSD2 (with a C–S distance of 2.5 Å and a C–N $\epsilon$  distance of 2.1 Å) (**Fig. 8**). Breaking of the C–S bond and the formation of the new bond between C and the nucleophile occurs instantaneously through a trigonal bipyramidal TS in which the carbon atom is sp<sup>2</sup> hybridized. The nucleophile attacks the carbon at a 180° angle to the leaving group, optimizing the overlap between the nucleophile's lone pair and the C–S antibonding orbital. Subsequently, the leaving group is pushed off at the opposite side, the TS structure collapses, and the methyl group covalently binds to the nitrogen atom, while SAH is released as a product (Copeland et al., 2009). Important to note is that the methyl group is transferred rapidly to the target lysine once a geometry favorable for the reaction has been achieved, due to the high group transfer potential of SAM, where it then prevents reprotonation.



**Figure 8: Geometric criteria for a bimolecular nucleophilic substitution (S<sub>N</sub>2) mechanism.** A transition state (TS)-like conformation can be approximated by using the depicted metrics.

### 1.5.7. Substrate specificity of SET domain PKMTs

PKMTs are highly regulated enzymes and aberrant methylation of proteins could result in misregulation of chromatin states or protein activity. A specific recognition of the protein substrate by PKMTs is therefore indispensable. A suitable technique to decipher the substrate specificity of PKMTs are Celluspot peptide arrays (Bock et al., 2011; Sara Weirich & Albert Jeltsch, 2022). In this method, peptides are synthesized on a cellulose membrane using solid-phase peptide synthesis. A large variety of different peptides can be synthesized on a single membrane, increasing the screening capacity. Each

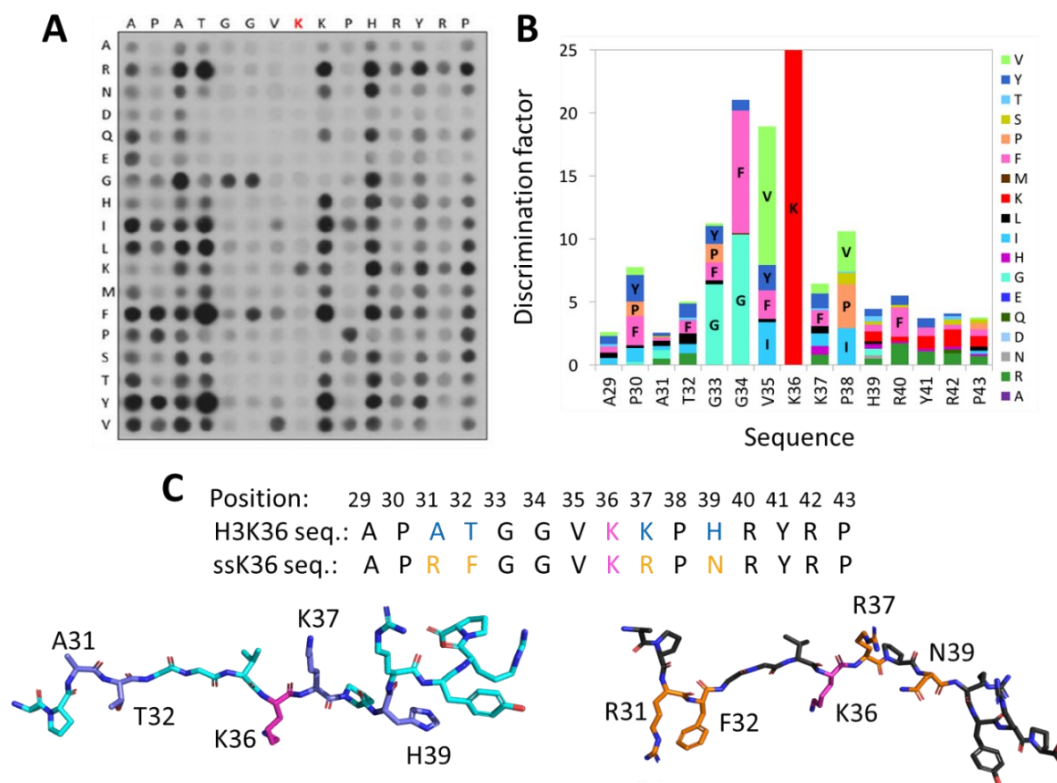
spot on the membrane represents an individual peptide sequence (**Fig. 9A**). The membrane is then incubated with the PKMT of interest and radioactively labeled SAM in buffer, allowing the detection of methylation through autoradiography. The signal intensity of the different peptide spots directly indicates which peptide sequences are preferred by the PKMT. By creating peptide arrays containing all possible single amino acid substitutions of an original substrate sequence, a PKMT-specific substrate specificity profile can be generated (**Fig. 9B**) (Dhayalan et al., 2011; Kudithipudi, Kusevic, et al., 2014; Kudithipudi, Lungu, et al., 2014; Kusevic et al., 2017; Rathert, Dhayalan, Ma, et al., 2008; Schuhmacher et al., 2015; Weirich et al., 2016). With the obtained specificity profile as a consensus sequence, novel protein substrate candidates have been identified (Dhayalan et al., 2011; Rathert, Dhayalan, Murakami, et al., 2008; Schuhmacher et al., 2020; Weirich et al., 2020). Certain PKMTs displayed a strict substrate specificity, limiting the range of substrates available for methylation (Kudithipudi et al., 2012; Schuhmacher et al., 2015). In contrast, other PKMTs showed a relaxed substrate specificity, allowing them to methylate a broad spectrum of substrates (Rathert, Dhayalan, Murakami, et al., 2008).

A strict substrate specificity of PKMTs could be explained by precise interaction between enzyme and substrate, ensuring an accurate readout of the substrate sequence. On the other hand, explaining the promiscuity of certain PKMTs is more challenging. Recognizing multiple lysine residues necessitates a complex network of specific interactions, while still avoiding off-target effects to prevent aberrant methylation profiles. A hollow active site with loose contacts therefore appears as a too simplistic explanation. Various models have been suggested to clarify the promiscuity of PKMTs. One hypothesis suggests that the structural flexibility of both, the enzyme's active site and the substrate allows for the adoption of numerous dynamic conformations, enabling PKMTs to recognize multiple substrates (Luo, 2018). An alternative model proposes that certain PKMTs identify their substrates based on their backbone atoms rather than their side chains (Al Temimi et al., 2019; Luo, 2018). The SET-I domain splits the SET domain and is speculated to be one of the key factors in determining substrate specificity, since it is the least conserved region among SET domain-containing PKMTs and heavily interacts with the substrate (**Fig. 5**) (Ronen Marmorstein, 2003). However, the SET-I hypothesis is challenged by the observation that PKMTs with similar substrate specificity, such as SETDB1 and SUV39H1, both methylating H3K9, exhibit large differences in their SET-I sequence (Ronen Marmorstein, 2003; Qian & Zhou, 2006).

#### **1.5.8. Discovery of PKMT super-substrates**

One striking example of the complex mechanism behind the substrate specificity of PKMT has recently been discovered for SETD2. Hereby, the canonical substrate, H3 residues A29-P43, was deemed suboptimal for SETD2 (Schuhmacher et al., 2020). Surprisingly, multiple single amino acid exchanges

caused a higher methylation of the corresponding peptide substrates. By combination of the preferred amino acids, a novel, non-natural peptide sequence, referred to as “super-substrate K36” (ssK36), was created. The ssK36 peptide, differed at four positions from the canonical H3K36 sequence and was methylated about 100-fold more efficiently (**Fig. 9C**). Methylation differences between H3K36 and ssK36 were even larger in a protein context (Schuhmacher et al., 2020). The crystal structure of the ssK36 peptide complexed to the SET domain of SETD2 was resolved, and subtle differences were observed compared to the H3K36-SETD2 structure (Schuhmacher et al., 2020). Three of the four amino acids altered in ssK36 established distinct contacts: ssK36-R31 forms an H-bond/salt bridge with SETD2-E1674, ssK36-F32 is bound into a pocket formed by SETD2-E1674 and SETD2-Q1676, and ssK36-R37 interacts with the backbone of SETD2-A1700. Despite these alterations, the overall structures of the ssK36-SETD2 and H3K36-SETD2 complexes remained very similar. Consequently, the crystal structures could not fully explain the substantial enhancement in the methylation rate of ssK36.



**Figure 9: Substrate specificity profile of SETD2 led to the super-substrate peptide (ssK36).** **A** | Cellspot peptide array with the 15-residue long H3K36 peptide sequence as the starting sequence, incubated with SETD2 and radioactively labeled SAM. Positions were individually mutated to any other amino acid except tryptophan and cysteine. At several positions amino acids are preferred which differ from the original H3 sequence. **B** | Quantification of the peptide array methylation data generates a PKMT specific specificity profile, highlighting the preference for each position. **C** | Combination of preferred residues led to the super-substrate peptide (ssK36) (black) sequence, differing at 4 positions (orange) from the canonical H3K36 peptide (cyan) sequence. SETD2 was demonstrated to have a strongly enhanced methylation efficiency towards ssK36. Figure taken from (Philipp Schnee et al., 2022; Schuhmacher et al., 2020).

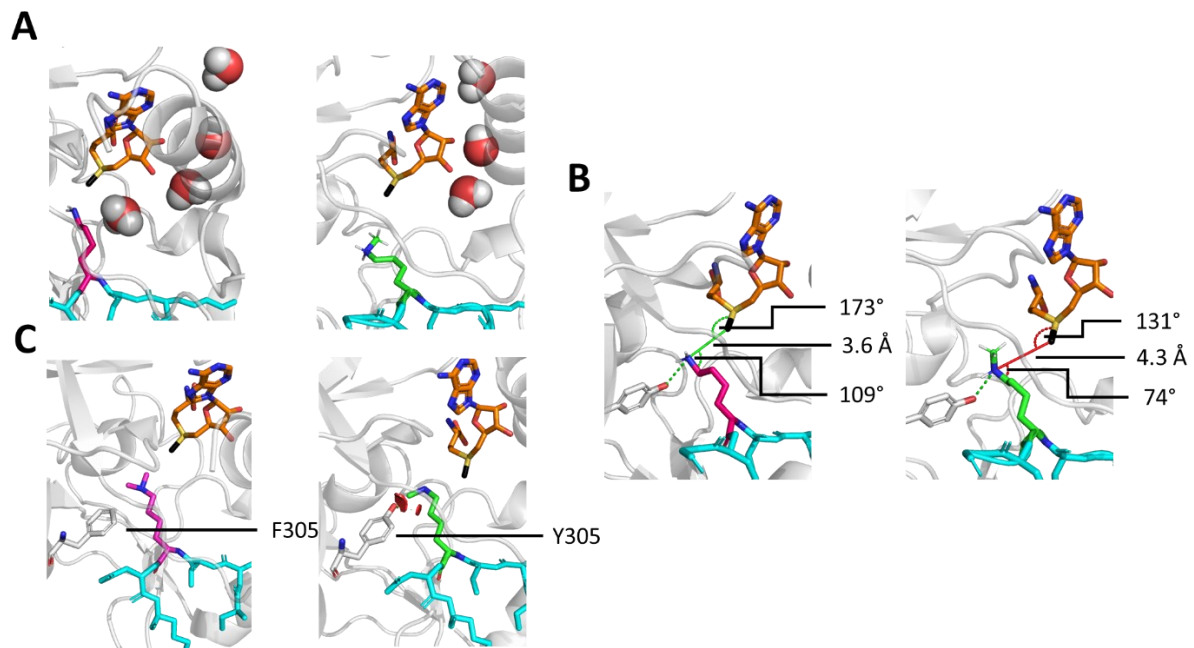
### 1.5.9. Product Specificity of SET domain PKMTs

Product specificity in PKMTs refers to their capability to transfer a precise number of methyl groups to their lysine residue target: one, two or three methyl groups, creating Kme1, Kme2 or Km3, respectively. Despite structural similarities in the SET domain, PKMTs exhibit distinct substrate but also product specificities, requiring unique mechanisms to control the number of methylation steps.

One potential mechanism to control the number of transferred methyl groups is speculated to be the target lysine deprotonation. As described earlier, the deprotonation of the lysine N $\epsilon$  is facilitated via a chain of water molecules (Zhang & Bruice, 2007b). In the context of product specificity, the presence or absence of the water channel could define the outcome. MD simulations of the monomethyltransferase SET7/9 revealed a water channel's presence only in the SAM-bound state complexed with an unmethylated K4 peptide, suggesting a role in lysine deprotonation and regulating mono-methylation. In the presence of SAH, or K4me1, the water channel was absent, preventing further methylation after monomethylation. (**Fig. 10A**). Mechanistically, the methyl group of the monomethylated peptide takes the position of the proton that would be removed through the water channel. Deprotonation and further methylation of Kme1 is therefore impossible (X. Zhang & T. Bruice, 2008b).

Another possible regulation mechanism refers to the S<sub>N</sub>2 reaction mechanism used by PKMTs. If methyl group and lysine N $\epsilon$  are too distant, the transfer is unlikely. In MD and QM/MM simulations of SET7/9 complexed with the K4 or K4me1 peptide, the distance between the SAM sulfur group and lysine N $\epsilon$  was greater for K4me1 (6.1 Å) than for K4 (5.7 Å) (Zhang & Bruice, 2007b). This difference in distance may be attributed to the active site potentially being too narrow. After the first methylation, a reorientation of Kme1 is not possible due to steric constraints. Consequently, a productive state, in which monomethylated lysine N $\epsilon$  and SAM methyl group come in close proximity, cannot form (**Fig. 10B**). Adding to the described mechanisms, in multiple sequence alignments it had been identified, that PKMTs possessing a tyrosine at the so-called "F/Y-switch" position are limited to catalyzing mono- or dimethylation. In contrast, enzymes with a phenylalanine or another hydrophobic residue at this position display di- or trimethyltransferase activity (Collins et al., 2005). This phenomenon was observed for several SET domain-containing PKMTs and it could even be used to manipulate the product specificity. For instance, the trimethyltransferase DIM-5 could be converted into a mono/dimethylase by the F281Y mutation (Zhang et al., 2003) and the monomethyltransferase SET7/9 could be changed to a dimethylase through the Y305F mutation (Del Rizzo et al., 2010; Zhang et al., 2003). The mechanistic basis of the F/Y-switch solely relies on the presence of a single hydroxyl group. The missing hydroxyl group in the Y to F mutants creates additional space in the active site, facilitating the accommodation of water molecules and proper reorientation of already transferred methyl groups. In contrast, F-to-Y mutations, which turn trimethyltransferases into mono- or

dimethyltransferases, could be based on steric effects caused by the additional hydroxyl group making the active site too narrow to accommodate multiple methyl groups at the lysine N $\epsilon$  (**Fig. 10C**) (Chu et al., 2012; Hu & Zhang, 2006). The concept of the active site volume as a regulator for product specificity may also provide insights into somatic cancer mutations altering product specificity as later shown in this work.



**Figure 10: Proposed control mechanism for the product specificity of PKMTs.** PKMTs catalyze the transfer of a distinct number of methyl groups to their lysine target (pink). Multiple mechanisms have been proposed, regarding the control of this process. **A** | Restricted second methylation caused by a disrupted water channel (red, spheres) and blocked lysine deprotonation of monomethylated target lysine (green, PDB 1XQH). **B** | The S<sub>N</sub>2 geometry cannot be adopted in the presence of a monomethyl substrate. **C** | The F/Y-switch position controls the product specificity of certain PKMTs. Phenylalanine (white) at this position creates additional space in the active site, allowing accommodation of a dimethyl product. In contrast, a tyrosine with its additional hydroxyl group causes clashes, preventing the formation of the dimethylated product. Figure taken and modified from (Schnee et al., 2023).

## 1.6. Histone lysine 36 methylation

Methylation of lysine 36 of histone H3 (H3K36) and especially the di- and trimethylation (H3K36me<sub>2</sub> and me<sub>3</sub>) are important histone modifications affecting many cellular processes (Eric J. Wagner & Phillip B. Carpenter, 2012). NSD1, NSD2, NSD3, ASH1L and SETD2, SMYD5 and PRDM9 are the PKMTs responsible for H3K36 methylation in human cells. While NSD1, NSD2, NSD3 and ASH1L can only introduce mono- and dimethylation of H3K36 *in vitro* and *in vivo* (Eric J. Wagner & Phillip B. Carpenter, 2012). SETD2 and SETD5 are responsible to introduce up to trimethylation at H3K36 in gene bodies, the SET and MYND domain-containing 5 (SMYD5) and PR/SET 9 (PRDM9) do so at promoter regions (Edmunds et al., 2008; Gregory et al., 2007; Li et al., 2009; Powers et al., 2016; Sessa et al., 2019; Zhang

et al., 2022). H3K36me2 is enriched at intergenic regions and promoters while H3K36me3 is enriched at gene bodies of active genes (Lam et al., 2022). H3K36me3 levels can be controlled in several ways, such as demethylation by the eraser protein KDM4A (Klose et al., 2006) or through the stability of SETD2. Homeostatic SETD2 protein levels in mammalian cells are low, as it is readily degraded by the ubiquitin–proteasome system (Zhu et al., 2017). SETD2 is also negatively regulated at the transcriptional level by the microRNA miR-106b-5p. Overexpression of miR-106b-5p was found to reduce SETD2 expression (Xiang et al., 2015).

The biological functions of H3K36 methylation encompass the regulation of gene expression, DNA repair, recombination and gene splicing (Eric J. Wagner & Phillip B. Carpenter, 2012). The diverse effects arise through the physical interaction of H3K36 PKMTs, predominantly SETD2, with RNA polymerase II (RNAP II), RNA-binding proteins, and transcriptional elongation factors (Li et al., 2019). H3K36 methylation is associated with both active gene transcription marks and gene repression. Its impact on gene transcription is controlled by adjacent histone modifications and their respective reader proteins (Eric J. Wagner & Phillip B. Carpenter, 2012). As a repressive modification, H3K36 methylation functions to suppress the aberrant initiation of transcription within coding regions of gene bodies in particular during active gene expression. This repression is facilitated by recruiting deacetylase complexes and the DNA methylation machinery (Lam et al., 2022; Li et al., 2019; Eric J. Wagner & Phillip B. Carpenter, 2012). The connection between H3K36 methylation and DNA methylation is established through the PWWP domains of DNMT3A and DNMT3B, which preferentially bind to H3K36me2 and H3K36me3, respectively (Dukatz et al., 2019).

### **1.6.1. SETD2**

The SET domain-containing protein 2 (SETD2), has a size of 230 kDa, which corresponds to 2564 amino acids. This enzyme is a major writer of H3K36me3 in mammals, depositing the modification primarily at gene bodies of actively transcribed genes. SETD2 was thought to be the sole protein responsible for H3K36me3, but a recent study has indicated that SETD5 can also deposit H3K36me3 at active gene bodies *in vivo* (Sessa et al., 2019). Furthermore, PKMTs SMYD5 and PRDM9 have been shown to deposit H3K36me3 at promoter regions and during meiosis, respectively (Powers et al., 2016; Zhang et al., 2022).

Human SETD2 contains several functional domains. These include a SET domain flanked by an “associated with SET” (AWS) domain and a post-SET domain, which together are responsible for the methyltransferase activity (Sun et al., 2005). SETD2 also contains a Set2–Rpb1 interaction (SRI) domain, which interacts with RNA polymerase II (RNAPII) (Li et al., 2005). The largest subunit of RNAPII contains a CTD that is hyperphosphorylated during active transcription. SETD2 interacts specifically with this



phosphorylated form of the RNAPII CTD through its SRI domain (Li et al., 2005; Sun et al., 2005). This allows SETD2 to selectively associate with actively transcribed regions of the genome. As a result, H3K36me3 is generally deposited at the 3' end of actively transcribed gene bodies and it is associated with euchromatin (Bannister et al., 2005; Mitchell et al., 2023). The SETD2 deposited H3K36me3 modification interacts with a large group of H3K36me3-binding proteins to participate in numerous cellular processes. One such process is *de novo* DNA methylation at genomic sites enriched with H3K36me3. This is performed by the reader protein DNMT3B (Baubec et al., 2015). DNA methylation aids in repressing cryptic transcription, which is the false initiation of transcription from intragenic sites of protein-coding genes (Neri et al., 2017).

SETD2 is additionally involved in the regulation of cell size. MicroRNA-mediated SETD2 knockdown in human cells caused an increase in cell size and total protein content accompanied by an increased protein synthesis rate *in vitro* (Molenaar et al., 2022). It is speculated that SETD2 might indirectly regulate cell size by influencing cell cycle dynamics or by directly controlling protein synthesis rates (Molenaar et al., 2022). Overexpression of the oncohistone H3.3K36M, which diminishes H3K36me3, also causes an increase in cell volume (Molenaar et al., 2022). Crystal structures revealed that the H3K36M mutation inhibits the catalytic activity of SETD2, as the introduced methionine mutation binds into the lysine binding channel of the active site and functions as a competitive enzyme inhibitor (Yang et al., 2016). Despite H3K36me3 loss through H3K36M overexpression, the cell size increase was smaller than for SETD2 knockdown (Molenaar & van Leeuwen, 2022). This indicates that SETD2-mediated regulation of cell size is not entirely dependent on H3K36me3 and highlights that the biological role of SETD2 is not limited to H3K36me3 deposition.

Besides histone protein targets, recent studies found that SETD2 methylation occurs at non-histone substrates. Among these, SETD2 has been observed to monomethylate K525 of the signal transducer and activator of transcription 1-alpha/beta (STAT1) (Chen et al., 2017). This methylation promotes STAT1 phosphorylation and activation, connecting SETD2 with the amplification of IFN $\alpha$ -dependent antiviral immunity signaling pathways (Chen et al., 2017). Other SETD2 targets are K735 of EZH2, K40 of  $\alpha$ -tubulin and K68 of actin (Park et al., 2016; Seervai et al.; Yuan et al., 2021).

Numerous somatic mutations of the SETD2 gene were found in cancer tissues, especially in pediatric high-grade gliomas (Fontebasso et al., 2013). Additionally, frameshift, non-sense and missense mutations in SETD2 are driver mutations in renal cell carcinoma (cRCC), indicating a loss-of-function mechanism. This is supported by the reduced amount of H3K36me3 in cRCC and an unaffected amount of H3K36me2 (Kudithipudi, 2014).

### 1.6.2. NSD2

The nuclear receptor SET domain-containing 2 (NSD2) catalyzes up to dimethylation of H3K36 and non-histone proteins. Down-regulation of NSD2 significantly decreases the methylation of H4K20, leading to the increased accumulation of 53BP1 (Pei et al., 2011). NSD2 interacts with phosphatase and tensin homolog deleted on chromosome 10 (PTEN) via its CTD and stimulates the dimethylation of PTEN in cells (Zhang et al., 2019). The latter is recognized by the specific domain of 53BP1 to recruit PTEN into sites of DNA damage. This is suspected to represent one pathway to regulate the sensitivity of cells to DNA damage (Chen et al., 2020).

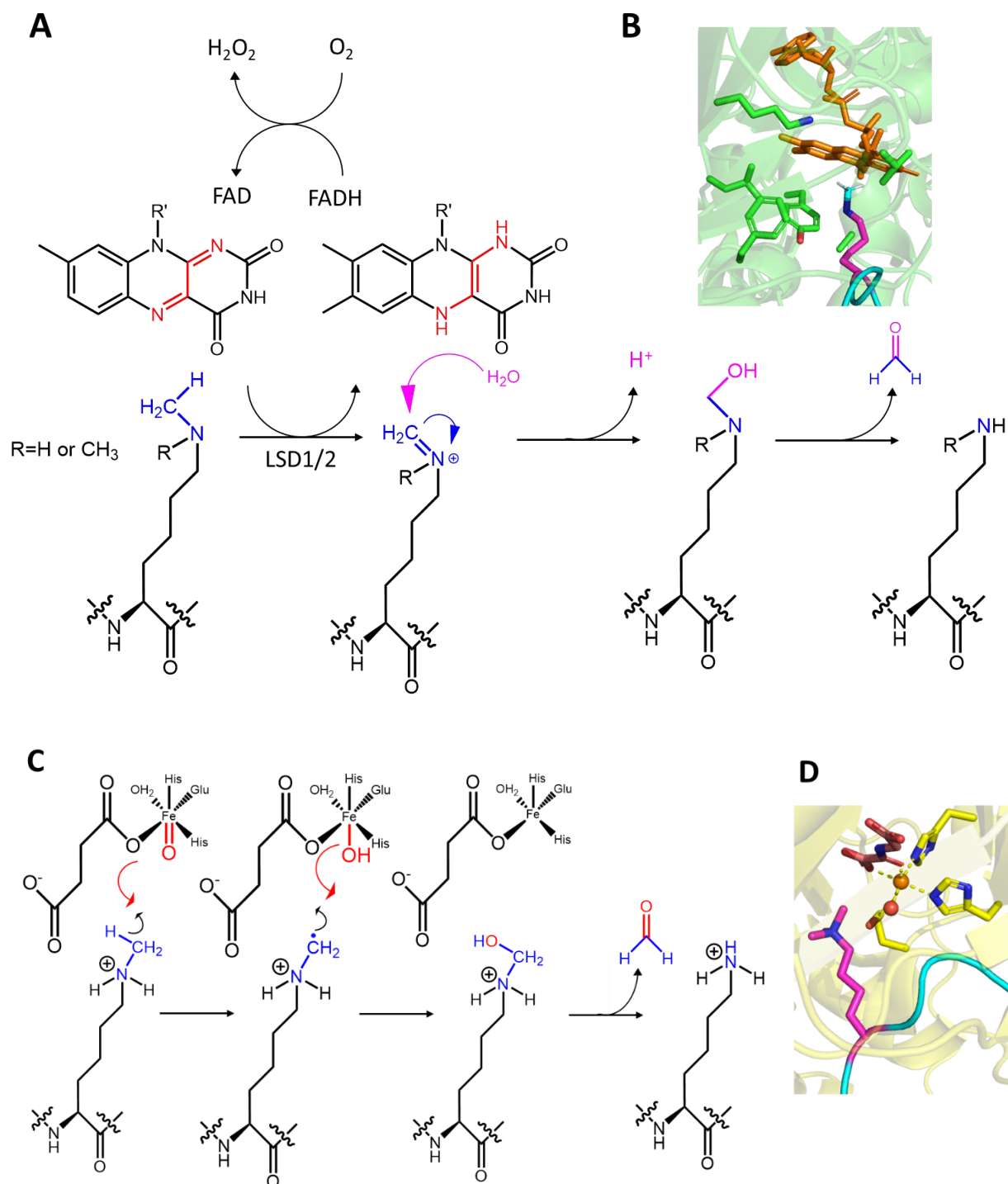
NSD2 dysfunction is linked to many diseases ranging from developmental disorders to cancers (Lam et al., 2022; Li et al., 2019; Eric J. Wagner & Phillip B. Carpenter, 2012). Heterozygous loss of NSD2 is responsible for the developmental disease called Wolf-Hirschhorn syndrome (WHS) (Bergemann et al., 2005). Moreover, missense mutations in NSD2 were observed in various types of cancers like lung cancers (Sengupta et al., 2021a; Yuan et al., 2021), hematological cancers (Jaffe et al., 2013) and head and neck squamous cell carcinomas (Cancer Genome Atlas, 2015). Epithelial–mesenchymal transformation (EMT) is a crucial process in cancer development, in which epithelial cells acquire characteristics of mesenchymal cells during tumorigenesis, development and progression (Brabletz et al., 2005; Yang et al., 2004). Recently, it was found that the overexpression of NSD2 occurs in 15% of patients with t(4;14)-positive multiple myeloma and that Twist-1 participates in driving the expression of EMT-related genes and contributes to tumor migration (Cheong et al., 2020). NSD2 interacts with Twist-1, which leads to an increase in H3K36me2 and promotion of EMT (Ezponda et al., 2013).

Contrary to the straightforward impact of gene deletions causing loss-of-function changes, understanding the biological effects of single point mutations is a more complex task. Many somatic missense mutations in PKMTs have been detected in diverse cancer types, and have been shown to alter the enzyme's activity, substrate specificity, product specificity, or other enzymatic properties (Brohm et al., 2019; Oyer et al., 2014; Weirich et al., 2017; Weirich et al., 2015). A frequent NSD2 missense single point mutation is E1099K, which was detected in leukemic patients. This mutant was comprehensively characterized and shown to be hyperactive (Jaffe et al., 2013; Oyer et al., 2014; Pierro et al., 2020; Swaroop et al., 2019). E1099K was demonstrated to firstly enhance the binding towards nucleosomal substrates by interacting with the negatively charged DNA, and secondly destabilize the AL of NSD2 through breaking the salt bridge mediated by E1099. Eventually, this led to an enhanced nucleosome associations and higher activity (Li et al., 2019; Sato et al., 2021). However, the effects of other frequent missense cancer mutants like T1150A in NSD2 are still unknown.

## 1.7. Lysine demethylation

Histone lysine methylation is a dynamic modification and can be removed by a group of enzymes called lysine demethylase (KDMs) (Hyun et al., 2017). The chemically inert nature of methyl lysine restricts the potential mechanisms of enzymes to remove a methyl group from the  $\epsilon$ -amine of a protein lysine residue. Two mechanisms of enzymatic lysine demethylation were characterized and involve amino oxidation and hydroxylation (Luo, 2018). Lysine-specific demethylases (LSDs) are flavine-adenine-dinucleotide (FAD) dependent and use the amine oxidase like (AOL) domain for the amino oxidation to remove lysine methylation (**Fig. 11A-B**). The hydroxylation reaction to demethylate lysine residues is carried out by KDMs bearing characteristic JmjC domains (**Fig. 11C-D**). JmjC-domain-containing KDMs can remove methyl groups from Kme1/2/3 whereas LSD enzymes can only act on Kme1/2 as substrates (Cole, 2008; Nowak et al., 2016).

Similar to PKMTs, KDMs have been shown to demethylate methyl lysine in non-histone proteins such as ER $\alpha$  (Zhang et al., 2013), E2F1 (Kontaki & Talianidis, 2010), DNMT1 (Nicholson & Chen, 2009) and STAT3 (J. Yang et al., 2010). Many non-histone targets are substrates of LSD1, even though a well-defined sequence motive is missing (Luo, 2018). This raises questions as LSD1 was shown to bind its H3K4me2 substrate in a highly sequence-specific manner (Luo, 2018). The promiscuous sequences and the specific recognition of LSD1 substrates at the same time are contrary and further work will be needed to clarify this issue. Two factors could potentially be responsible for this interesting effect: (i) the substrate-binding pocket of LSD1 is flexible and adopts multiple conformations to accommodate different substrates; (ii) the substrate specificity of LSD1 is altered by recruitment of regulatory partners like the androgen receptor, which interacts with the LSD1 SWIRM domain, changing the preference from H3K4me2 to H3K9me2 (Luo, 2018; Wu et al., 2012).



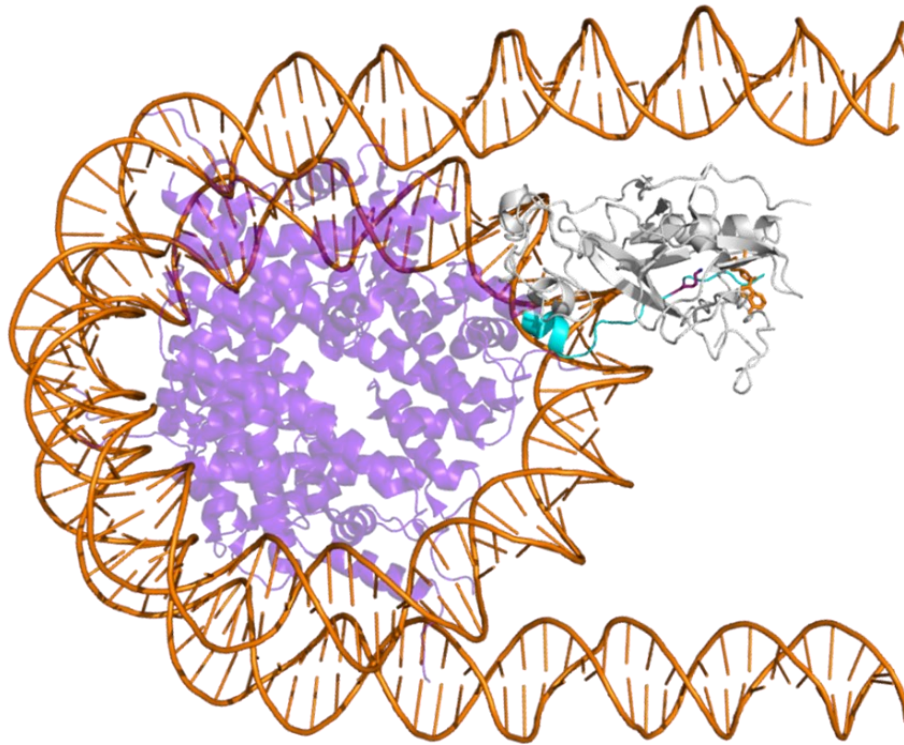
**Figure 11: Reaction mechanisms of lysine demethylation. A** | Chemical mechanism of the demethylation reaction catalyzed by LSDs. **B** | Hydrophobic interactions of monomethylated lysine (pink) with its surrounding residues (green) in the catalytic chamber with FAD (orange, PDB 6VYP). **C** | Chemical mechanism of demethylation reaction catalyzed by JmjC-domain-containing KDMs (Luo, 2018). **D** | Representative structure and catalytic site of JmjC-domain-containing KDMs. KDM4A is shown as an example (PDB 2OQ6). Residues H188, E180, H276 (yellow), a water molecule (red) and an  $\alpha$ -ketoglutarate analogue (rose) coordinate the iron (orange). Figure taken and modified from (Kong et al., 2011; Luo, 2018).

## 1.8. Molecular Dynamics Simulation

Molecular Dynamics Simulations (MD) are computational methods used to study the dynamic behavior of molecules and atoms in space over time. Empirically derived physics principles are applied to model the interactions and movements of atoms in the simulated system. By numerically solving the equation of motion for each atom, MD simulations provide insights into the dynamic behavior of biomolecular systems at an atomistic resolution. Researchers use MD simulations to investigate a wide range of questions including protein folding, ligand association to a protein, and cancer mutants. In the following chapters, the different MD simulation techniques used in this work are briefly described.

### 1.8.1. Modelling

At present, the Protein Data Bank (PDB) holds more than 206 000 experimentally solved structures, including more than 200 000 proteins (<https://www.rcsb.org/> retrieved on November 11<sup>th</sup>, 2023). Due to the increasing resolution in crystal structures and cryo-Electron Microscopy (EM), resolved structures help to identify protein architectures, interactions with substrates or cofactors and unravel active site compositions in atomistic details. Over time, larger complexes have been resolved and complexes of multiple proteins together with nucleosomes have become state-of-art. The information from these structural data is highly valuable as it was shown for PKMTs SETD2 and NSD2, that the binding to a nucleosome causes the DNA around the histone octamer to partially unwrap. This facilitates the methylation of histone tails (**Fig. 12A**). SETD2 and NSD2 establish important contacts with the DNA, enhancing the affinity towards nucleosome substrates compared to protein substrates.



**Figure 12: Binding mode of PKMTs to nucleosomal substrates. A|** NSD2 (grey) was found to bind nucleosome substrates by partially unwrapping the DNA (orange). By this conformational rearrangement, the H3 histone tails (cyan) is made accessible and the target lysine (pink) binds in the NSD2 binding cleft (PDB 7E8D). Figure taken and modified from (Schnee et al., 2023).

Whereas structural data provide important information about the substrate recognition of enzymes by complex formation or arrangement of key amino acids, it has limitations, which need to be considered. Proteins are dynamic entities with flexible structures. Said structures are difficult to localize in x-ray crystallography. The outcome is that flexible regions or residues are not resolved. This is critical since flexible residues can be involved in highly relevant processes like conformational changes, association reactions or sampling of protein-protein interactions.

Another limitation is the necessity for stable complex formation to conduct structural studies. This is specifically relevant for enzymes since they catalyze reactions by binding substrates and releasing products. Inhibitors can be used to circumvent this problem, but critical information about catalysis is therefore lost. For example, SET domain-containing PKMTs use water channels to deprotonate the target lysine residue in the active site. The water channel is only visible if the cofactor SAM is bound and not for the cofactor product SAH (X. Zhang & T. Bruice, 2008a). However, in resolved structures mostly SAH has been used as cofactor, since the methyl group had already been transferred to a substrate, or SAM derivatives like sinefungin were used in the first place for stable complex formation. Alternatively, substrate inhibitors, like the H3K36M peptide can be used, constantly binding to SET domain-containing PKMTs, since methionine is not methylated (Fang et al., 2016; Schnee et al., 2023;

Schuhmacher et al., 2020; Yang et al., 2016; Zhang et al., 2017). Hence, interactions between the missing target lysine and the PKMTs active site are lost. Most importantly, the TS of the catalyzed reaction is never visible in static crystal or cryo-EM structures using inhibitors. The TS is however the most critical conformation between enzyme and substrate.

The highlighted limitation can be overcome by using modelling approaches. Here, resolved structures function as a template from which key factors can be modified. Based on enzymes with similar sequences or architectures, missing residues can be modelled using tools like PyMOD 3.0 (Janson & Paiardini, 2021). Homology models can be generated from enzymes for which no resolved structure has yet been published by tools like AlphaFold2 (Jumper et al., 2021), or SWISS-MODEL (Waterhouse et al., 2018). A K36M inhibitor can be reverted to lysine or the cofactor SAH can be exchanged for SAM using PyMOL (L. L. C. Schrödinger, 2015). These modifications create a more accurate model of the dynamic interaction between PKMT, cofactor and protein substrate.

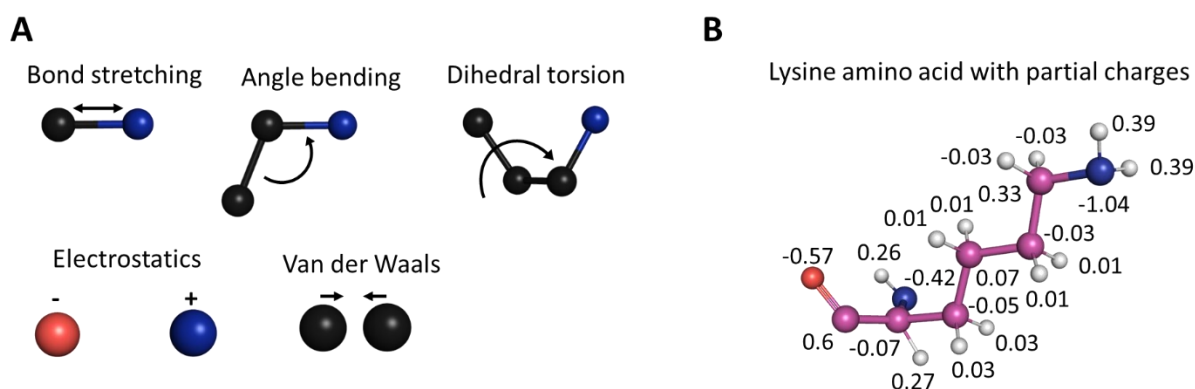
### 1.8.2. Structure of MD Simulations

Structural models lack dynamic behavior and critical information about conformational changes due to substrate binding, product release or the orientation during the TS is not accessible. This was shown for the Post-SET loop of SETD2, which blocks the active site, but opens up upon binding to protein substrates (Yang et al., 2016). While at least snapshots of this process were captured in crystal structures (PDB 5LSU for closed, 5V21 for open), the complete process involving the regulatory contacts between enzyme and substrate can only be investigated using MD simulation approaches.

At the beginning of an MD simulation, a starting structure, based on a crystal structure, cryo-EM structure or modelling results (or a combination of all these approaches), is used. Afterwards, each atom is given a random velocity, based on a temperature dependent Maxwell Boltzmann distribution (Meller, 2001; Yu & Dalby, 2020). Subsequently, the atoms start to move, following the distributed velocities for a short time interval. Next, forces are applied on each atom modeling bonded and non-bonded interactions. Bonded interactions describe the chemical bonds between atoms by bond potentials, which encompasses stretching, bending, and torsional rotation (**Fig. 13A**). Non-bonded interactions involve Van-der-Waals interactions described by the Lennard-Jones potential accounting for both attractive and repulsive forces. Electrostatic forces between charged atoms, carrying full or partial charges, are modeled by a Coulomb potential and an appropriate dielectric constant (**Fig. 13B**). Electrons are not explicitly considered in classical MD simulations. To calculate the forces applied on each atom, a molecular mechanics force field is employed. Force fields are derived and calibrated to fit to the results of precise quantum mechanical calculations and often incorporate specific experimental measurements to validate the obtained parameters (Monticelli & Tieleman, 2013). A

force field consists of atoms masses, descriptions of bond energies between atoms depending on length, angles and dihedrals. Comparison of simulation and experimental data indicates that force fields have improved substantially in recent years (Chmiela et al., 2018; Dasetty et al., 2019; Lindorff-Larsen et al., 2012; Lopes et al., 2015). Still, force fields use empiric data and the uncertainty introduced by these approximations should be considered when interpreting simulation results.

Moreover, in a classical MD simulation, no covalent bonds are formed or broken meaning that chemical reactions cannot be directly represented. Hybrid quantum mechanics/molecular mechanics (QM/MM) simulations model a small part of a system using quantum mechanical calculations, taking electron densities into account. Here, bond breakage and formation can be modeled. The remainder of the system is modeled by classical MD simulation. QM/MM simulations are frequently employed to study chemical reactions and enzyme catalysis that involve changes in covalent bonds. However, QM/MM simulations pay for the accurate calculations with heavily reduced simulation times and higher computational demand.



**Figure 13: Molecular dynamics simulation use bonded and non-bonded forces to model the interactions between atoms. A|** In MD simulation, atoms are connected via spring-like bonds. The bonded forces comprise bond stretching, angle bending and dihedral torsion. Non-bonded interactions consist of Van-der-Waals and electrostatic forces. **B|** Electrostatic forces are applied between fully or partially charged atoms. Van-der-Waals forces are described using the Lennard-Jones Potential. Figure taken and modified from (Schnee et al., 2023).

By solving Newton’s laws of motion, the calculated forces applied on each atom result in an acceleration vector, which is then applied to update the velocities. This results in a sequential progression, iteratively applying forces to atoms and updating their positions and velocities. Ultimately, the outcome is a dynamic, three-dimensional, atomic-level resolution representation of a dynamic protein, e.g. the behavior of a PKMT in the presence of its substrate and SAM (Hollingsworth & Dror, 2018; Hospital et al., 2015).

To ensure numerical stability of the molecular structures, the time steps in an MD simulation must be short, typically only a few femtoseconds ( $10^{-15}$  s) each. In contrast, conformational changes of enzymes



take place on timescales of nanoseconds, microseconds, up to seconds. Hence, billions of time steps are needed to capture the complete dynamics. In each step, all interatomic interactions in the system are calculated, which is computationally expensive. Luckily, the continuous and almost exponential advancements in high performance computing power, especially the availability of Graphics Processing Units (GPU), have allowed for more efficient and longer simulations. An example benchmark simulation with a 159-residue protein (DHFR) in explicit solvent (<24000 atoms) using OpenMM (Eastman & Pande, 2010; Eastman et al., 2017) yielded >1ms per day on a Nvidia RT2080Ti (<https://openmm.org/benchmarks>). To simulate even longer and larger systems, coarse-grained MD simulations offer a loophole. Here, one artificial particle represents a group of atoms rather than a single atom. Thereby, the resolution decreases but accessible timescales increase by orders of magnitude (Marrink & Tieleman, 2013). However, these approaches are no longer first-principle methods as they are based on purely empirical force fields which need to be established beforehand and calibrated with experimental data very carefully.

Due to limited computational resources, it is necessary to restrict the system size. Therefore, periodic boundary conditions are used. The system is confined in a box, assuming that the properties of the actual system can be approximated by a virtual infinite system of repeating side-by-side boxes. If a molecule or atom passes through the initial box boundary, it will re-enter the neighboring box from the opposite boundary, forming a periodic space (Meller, 2001; Yu & Dalby, 2020).

### **1.8.3. Steered molecular dynamics simulations**

Steered molecular dynamics (sMD) simulation is a technique in which controlled external forces are used to guide an under-sampled process. Rare events like the association of a ligand to a protein are too slow to be observed in realistic timeframes or only under immense computational cost. During the simulation, the majority of resources is wasted on unimportant events like the ligand floating in the solvent. In the sMD approach two groups of molecules are connected with a very weak spring, slowly guiding the ligand to the protein. Thereby, reactions that otherwise would be too slow to be modelled in MD simulations are accelerated and at the same time the conformational sampling is concentrated along a specific, predefined reaction coordinate (Yang et al., 2019).

The sMD method has been successfully utilized in a variety of investigations. Especially, binding affinities between proteins and ligands are of special interest, as the simulation data can be validated with experiments (Chang, 2004). For example, the chaperone FK506-binding protein dimerizes upon contact with Tacrolimus (FK506), which made it a valuable tool for chemically induced dimerization applications where it can be used to manipulate protein localization, signalling pathways and protein

activation (Fegan, 2010). In this case, the sMD obtained dissociation constant ( $k_D$ ) was calculated and agreed nicely with the experimentally determined  $k_D$  (Lee & Olson, 2006).

## 2. Aims of this work

The methylation reactions catalyzed by DNMTs and PKMTs are highly regulated, and aberrant methylation of protein and DNA is associated with neurologic disorders, cardiovascular diseases, and various types of cancer. The overarching goal of this work was to use molecular dynamics simulation in combination with biochemical experiments to investigate the catalytic machinery of these enzymes. In pursuit of this goal, two approaches were used:

1. By simulating cancer mutants of DNMTs and PKMTs and comparing the obtained simulation results to simulations of the wild type (WT), distinctions between mutant and WT can be found. These differences help to reveal the molecular mechanisms behind the cancer mutation and lead to a better understanding of the pathophysiological effects.
2. Investigating the substrate specificity of PKMTs by comparing canonical peptide substrates to artificially designed super-substrate peptides. A super-substrate peptide is methylated with a high efficiency by the PKMT for which it was designed for. By uncovering the properties which cause the increased methylation, a detailed knowledge about the full catalytic process of PKMTs is gained. This extends our understanding of how PKMT interact with their substrates but could also serve as a basis to develop PKMT specific inhibitors.

### 2.1. MD Simulation of the somatic cancer mutation R882H of DNMT3A

DNMT3A mutations were observed in 20–40% of patients with particular subtypes of Acute Myeloid Leukemia (AML), a cancer of blood forming cells in the bone marrow (Ley et al., 2010; Yan et al., 2011). DNMT3A R882H is the most abundant missense mutations (66%) among others (Tate et al., 2019). Despite the high clinical relevance, the pathogenic mechanism of this mutation is unclear. R882 is located in the RD interface of DNMT3A tetramers, which forms the DNA binding site. Biochemical studies of DNMT3A R882H have shown that the mutation leads to a massive change in the flanking sequence preferences compared to the WT. This leads to alterations of DNA methylation patterns in cancer cells (Mack et al., 2022). The genetic data clearly indicated a dominant behavior of the mutation. To understand this behavior, the DNMT3A/L heterotetramer was investigated by MD simulations. DNMT3A WT and DNMT3A R882H mutant were each complexed with a DNA substrate and simulated to identify potential differences in complex conformations and interface stability. Found discrepancies could shed light on the pathogenic mechanism behind this mutation.

## **2.2. MD Simulation of the somatic cancer mutation T1150A of NSD2**

Biochemical characterization of the PKMT NSD2 and its somatic cancer mutation T1150A revealed a hyperactivity of this enzyme. Besides the found hyperactivity, the product specificity was changed from a dimethyltransferase to a trimethyltransferase. Changes in the H3K36 methylation state are known to be associated with diverse biological outcomes, because H3K36me2 and H3K36me3 exhibit distinct downstream effects on gene transcription and chromatin structure (DiFiore et al., 2020). Therefore, a combination of MD and sMD simulation techniques was used to investigate the reason for the altered product specificity of NSD2 T1150A. This allowed for an atomistic and dynamic analysis of the NSD2 active site and its interaction with the SAM cofactor and the protein substrate.

## **2.3. Mechanistic basis of super-substrate peptides**

SETD2 transfers up to three methyl groups to H3K36. The substrate specificity of SETD2 regarding H3K36 has been previously mapped using Celluspot peptide array methylation (Schuhmacher et al., 2020). This revealed that the canonical H3 amino acids were not ideal at many positions. Based on this, an artificial peptide substrate was designed that contained the most favorable amino acid at each position. Methylation experiments showed that the 15 amino-acid long super-substrate peptide (ssK36), which differed at four positions from the original H3K36 was methylated more than 100-fold faster than the canonical H3K36 peptide (Schuhmacher et al., 2020). The crystal structure of SETD2 with bound ssK36 peptide was resolved but did not entirely explain the highly increased methylation activity of SETD2. To elucidate the mechanistic reasons behind this massive increase in reaction rate, a combination of biochemical methylation and FRET experiments, MD and sMD simulation techniques were used to cover multiple steps of the catalytic process, considering (i) the peptide in solution; (ii) the association of the peptide into the enzyme's active site; (iii) the established contacts in the enzyme-peptide complex. Comparing the H3K36 peptide and the highly active ssK36 peptide at all catalytic stages offers the opportunity to understand the factors influencing the methylation activity of SETD2 and potentially other PKMTs. Likewise, a similar workflow was applied to the PKMT NSD2, for which a new super-substrate peptide was designed by Dr. Sara Weirich (Weirich et al., 2023). Again, the PKMT-peptide complex properties were investigated by MD simulation, showing the transferability of this approach. The catalytical advantages of the super-substrate peptides were then used as a starting point to establish a PKMT-specific inhibition assay in which ssK36 was tested for its SETD2-specific inhibitory strength.

### 3. Material and Methods

In the upcoming chapters the main methods used in the presented studies are briefly summarized. The detailed methods can be found in the corresponding attached publications. The exact code for the MD simulation protocols, input files, trajectory analysis scripts, software and library versions can be found in the respective DaRUS repository of the University of Stuttgart:

- [doi.org/10.18419/darus-2463](https://doi.org/10.18419/darus-2463) for the MD simulation protocols and trajectory analysis scripts of DNMT3A (Mack et al., 2022)
- [doi.org/10.18419/darus-2508](https://doi.org/10.18419/darus-2508) for peptides in solution, SETD2 association and SETD2-peptide complex simulations and analysis scripts (Philipp Schnee et al., 2022)
- [doi.org/10.18419/darus-3263](https://doi.org/10.18419/darus-3263) for NSD2 T1150A simulation and analysis scripts (Khella et al., 2023a)
- [doi.org/10.18419/darus-3815](https://doi.org/10.18419/darus-3815) for NSD2 super-substrate simulation and analysis scripts (Weirich et al., 2023)

The manuscript of NSD2 and its super-substrate (manuscript #4 (Weirich et al., 2023), appendix II) was submitted for review, which could influence the accessibility of the DaRUS data repository. If the repository is not accessible, refer to reviewer access at:

<https://darus.uni-stuttgart.de/privateurl.xhtml?token=3560453f-56c3-45a7-a365-140c6a5973f3>

#### 3.1. MD simulations and representation

All MD simulations were performed using OpenMM 7.4.2 (for DNMT3A simulations) or 7.5.1 (for peptide, SETD2 and NSD2 simulations) (Eastman & Pande, 2010; Eastman et al., 2017) utilizing the NVIDIA CUDA (Nvidia, 2010) GPU platform. The systems were parameterized using the General Amber force field (GAFF) and AMBER 14 all-atom force field (Case, 2014). SAM was modelled based on the coordinates of SAH and parameterized using ANTECHAMBER from AmberTools (18.0) (Junmei Wang, 2001). The non-bonded interactions were treated with a cut-off at 10 Å. Additionally, the Particle Mesh Ewald method (Tom Darden, 1993) was used to compute long-range coulomb interactions with a 10 Å non-bonded cut-off for the direct space interactions. Energy minimization of the systems was performed until a 10 kJ/mole tolerance energy was reached. Simulations were run using a 2 fs integration time step. The Langevin integrator (Bussi & Parrinello, 2007) was used to maintain the system temperature at 300 K with a friction coefficient of 1 ps<sup>-1</sup>. The initial velocities were assigned randomly to each atom using a Maxwell–Boltzmann distribution at 300 K. A cubic water box with a 10 Å padding to the nearest solute atom was filled with water molecules using the tip4p-Ew model (Horn et al., 2004) at pH 7. The ionic strength of 0.1 M NaCl was applied, by adding the corresponding number

of Na<sup>+</sup> and Cl<sup>-</sup> ions. The pK<sub>a</sub> values of proteins and peptides were assigned according to the AMBER ff14SB force field (Maier et al., 2015). Production runs were performed under periodic boundary conditions, and trajectories were written every 10,000 steps (20 ps). Deviating protonation states, equilibration protocols and other specifications for the individual system setups are described in the following paragraphs below. All structures and MD simulation snapshots were visualized using PyMOL (2.4.1) (L. L. C. Schrödinger, 2015).

### **3.1.1. MD simulations of the DNMT3A/L hetero tetramer complexed with DNA**

MD simulations were conducted using structures PDB 6W8B for DNMT3A WT and PDB 6W89 for DNMT3A R882H. SAM was introduced instead of SAH and zebularine was changed to cytosine not covalently bound to cytosine 710. Simulations were run using a 1 fs integration time step. A rectangular box shape was used to minimize the amount of additional water molecules. The system charge was neutralized by adding 36 Na<sup>+</sup> ions for the DNMT3A WT simulations and 38 Na<sup>+</sup> ions for the DNMT3A R882H simulations. To equilibrate the solvent, a 3 ns pressure coupled equilibration with Monte Carlo barostat (Faller & de Pablo, 2002) was performed at a pressure of 1 atm. This was followed by a 3 ns equilibration with restrained protein and DNA backbone with a force of 2 kJ/mole × Å<sup>2</sup>. The restraints were removed and followed by a 3 ns equilibration. With the resulting system, 6 independent 25 ns production runs were performed for each DNMT3A WT and DNMT3A R882H. Further details can be found in manuscript #1 (appendix I).

### **3.1.2. sMD simulations of the peptide association process into the NSD2 active site**

The structures of the SET domain of human NSD2 WT and NSD2 T1150A (Y991–K1220) were modeled based on the cryo-EM structure of NSD2 the E1099K/T1150A double mutant in complex with a nucleosome (PDB 7CRO). Missing amino acids and reverting of the mutations K1099E (to obtain NSD2 T1150A) and K1099E/ A1150T double mutant (to obtain NSD2 WT), were modeled using PyMOD 3.0 (Janson & Paiardini, 2021). The missing part of the post-SET loop (amino acids P1206–K1220) was modeled based on the SET domain of SETD2 (PDB 5V21) using PyMOD 3.0, since no structure of NSD2 complexed with the H3K36 peptide (29-APATGGVKKPHRYRP-43) and post-SET loop has been resolved. Subsequently, the histone tail of PDB 7CRO was replaced by the H3K36 peptide from PDB 5V21. In this peptide methionine 36 was mutated to lysine and manually deprotonated as required for the S<sub>N</sub>2 mechanism. Methyl groups were introduced at the lysine side chain nitrogen using PyMOL. Parametrization of methylated lysine in the different methylation states was accomplished using AMBER 14 GAFF and ff14SB. The Zn<sup>2+</sup> ions were modeled using the cationic dummy atom method (Pang, 1999, 2001). Cysteines 1016, 1018, 1026, 1032, 1041, 1046, 1052, 1145, 1192, 1194 and 1199

were treated as unprotonated to ensure proper  $Zn^{2+}$  binding. The system charge was neutralized by adding 3  $Na^+$ . SAM was modelled as described before and placed  $\sim 27$  Å away from the SAM-binding pocket, representing the starting point for the association studies (distance between SAM methyl group and side chain nitrogen of the peptide lysine 36). To equilibrate the solvent, a 5 ns pressure coupled equilibration with Monte Carlo barostat was performed at a pressure of 1 atm. The C-alpha ( $C\alpha$ ) atoms of NSD2, the peptide, and the SAM atoms were restrained with a force of 100, 100 and 5  $\text{kJ/mol} \times \text{Å}^2$ , respectively. The restraints were taken off successively, starting with the NSD2  $C\alpha$  restraints, followed by a 5 ns equilibration with the peptide and SAM still being restrained. Subsequently, the SAM and peptide restraints were removed as well, followed by 0.1 ns equilibration with no restraints. A distance-dependent force of  $0.2 \times \text{distance between centroid 1 (lysine 36 side chain nitrogen and its attached two hydrogen atoms) and centroid 2 (SAM methyl group and its attached three hydrogen atoms) (kJ/mol)/\text{Å}^2$ ) was used. This force was not static but rather represents a spring pulling the center of mass (COM) of centroid 1, towards the COM of centroid 2, while still giving room for conformational changes. The lysine hydrogen atoms were replaced with carbon atoms as appropriate for the Kme1, Kme2 and Kme3 simulations. Two additional weaker forces were used to guide SAM into a proper binding position in the SAM binding pocket (force2:  $0.1 \times \text{distance between centroid 3 (SAM atoms N1, C2, N3) and centroid 4 (NSD2 L1202 atoms N, C}\alpha, C) (\text{kJ/mol})/\text{Å}^2$ ); force3:  $0.05 \times \text{distance between centroid 5 (SAM atoms N0, C}\alpha, C\beta) and centroid 6 (NSD2 F1149 atoms C}\alpha, C, O) (\text{kJ/mol})/\text{Å}^2$ ). For production, sMD simulations were conducted for 100 replicates à 35 ns for each protein variant NSD2 WT and NSD2 T1150A. Further details can be found in manuscript #2 (appendix I).

### 3.1.3. MD simulations of NSD2

General simulation parameters and starting structures of the SET domain of NSD2 WT and T1150A complexed with the H3K36 peptide were modeled as described above for the sMD experiments (chapter 3.1.2). SAM was positioned in the SAM binding pocket based on the coordinates of SAH in PDB 7CRO. A 5 ns pressure coupled equilibration with Monte Carlo barostat was performed at a pressure of 1 atm. NSD2 and peptide  $C\alpha$  atoms as well as SAM atoms were restrained with a force of 100, 100 and 5  $\text{kJ/mol} \times \text{Å}^2$ , respectively. The restraints were taken off successively, starting with the  $C\alpha$  restraints, followed by a 5 ns equilibration with only SAM restrained. Subsequently, the SAM restraints were removed as well followed by 5 ns equilibration with no restraints. For production, 30 replicates à 100 ns were performed for each NSD2 variant WT and T1150A. Further details can be found in manuscript #2 (appendix I).

For the investigation of the NSD2 super-substrate ssK36(NSD2), NSD2 complexed with the 15 amino acid long H3K36 peptide was modelled as described above (chapter 3.1.2). The ssK36(NSD2) peptide

in complex with NSD2 was obtained by manually mutating H3K36 in PyMOL to generate the ssK36(NSD2) sequence (29-APKTGGVKRPNNYRP-43). The protein charge was neutralized and an ionic strength of 0.1 M NaCl was applied, by adding 30 Na<sup>+</sup> and 27 Cl<sup>-</sup> ions. A 5 ns pressure coupled equilibration with Monte Carlo barostat was performed at a pressure of 1 atm. NSD2 and peptide C $\alpha$  atoms as well as SAM atoms were restrained with a force of 100, 100 and 5 kJ/mol  $\times$  Å<sup>2</sup>, respectively. The restraints were taken off successively, starting with the C $\alpha$  restraints, followed by a 5 ns equilibration with only SAM restrained. Subsequently, the SAM restraints were removed as well followed by 5 ns equilibration with no restraints. For production, MD simulations were conducted in 50 replicates  $\times$  100 ns (total simulation time 5  $\mu$ s). Further details can be found in the attached manuscript #4 (appendix II).

#### **3.1.4. MD simulations of peptides in solution**

The peptide structures were retrieved from crystal structures of each peptide complexed with SETD2, PDB 5V21 for the H3K36 peptide (29-APATGGVKKPHRYRP-43) and PDB 6VDB for the ssK36 peptide (29-APRFGGVKRPNNRYRP-43). Missing amino acids were modelled using the PyMOL builder (L. L. C. Schrödinger, 2015). Since the crystal structures of SETD2 were complexed with K36M inhibitor peptides, M36 was mutated to K36 for both peptides using the PyMOL mutagenesis tool. The system charge was neutralized by adding 4 Cl<sup>-</sup> ions for the H3K36 simulations and 5 Cl<sup>-</sup> ions for the ssK36 simulations. To equilibrate the solvent, a 5 ns pressure coupled equilibration with Monte Carlo barostat was performed at a pressure of 1 atm, which was followed by a 10 ns free equilibration without constraints. The equilibration steps were repeated for each simulation replicate and followed by a 50 independent 70 ns production runs for each peptide. Further details can be found in manuscript #3 (appendix I).

#### **3.1.5. sMD simulations of the peptide association process into the SETD2 active site**

The structure of the SET domain of human SETD2 was retrieved from PDB 6VDB, missing amino acids were modelled using PDBFixer (P. Eastman, 2013) and the complexed ssK36 peptide was removed. The centroid structures from the peptide conformations clustering 1 and 2 (chapter 3.2.3) were placed 30 Å away from the binding cleft (distance between SAM methyl group and side chain nitrogen of the peptide lysine 36). The target lysine was manually deprotonated as required for the S<sub>N</sub>2 mechanism (X. Zhang & T. Bruice, 2008a). The Zn<sup>2+</sup> ions were modelled using the cationic dummy atom method (Pang, 1999, 2001). Cysteines 1499, 1501, 1516, 1520, 1529, 1533, 1539, 1631, 1678, 1680, 1685 were treated as unprotonated to ensure proper Zn<sup>2+</sup> binding. The system charge was neutralized by adding 7 Cl<sup>-</sup> ions. To model SETD2 in an open position, the post-SET loop (Q1691-K1703) was manually lifted



upwards using PyMOL based on the results described previously (Yang et al., 2016). In this conformation, the peptide binding cleft is no longer blocked by the post-SET loop and the autoinhibitory residue R1670 points outwards. The post-SET loop was not constrained in the open conformation. The open position of SETD2 was then used for the sMD simulations. To equilibrate the solvent, a 5 ns pressure coupled equilibration with Monte Carlo barostat was performed at a pressure of 1 atm. The C $\alpha$  atoms of SETD2, the peptide and the SAM atoms were restrained with a force of 100, 100 and 5 kJ/mole  $\times \text{\AA}^2$ , respectively. The restraints were taken off successively, starting with the SETD2 C $\alpha$  restraints, followed by a 5 ns equilibration with the peptide and SAM still being restrained. Subsequently, the SAM and peptide restraints were removed as well, followed by 0.1 ns equilibration with no restraints. A distance-dependent force of  $0.5 \times \text{distance of centroid 1 (K37 atoms NZ, HZ1, HZ2) and centroid 2 (SAM atoms S, C\epsilon, H10) (kJ/mol)/\text{\AA}^2}$  was used to pull the COM of the lysine 36 side chain nitrogen and its attached two hydrogen atoms, towards the COM of the SAM methyl group and its attached three hydrogen atoms. This force was not static but rather represents a spring pulling the COM of centroid 1, towards the COM of centroid 2, while still giving room for conformational changes. The sMD simulation was run for 100 replicates  $\times$  50 ns for each peptide.

To discriminate between the different conformational preferences of the peptides in the sMD experiment, an additional, distance-dependent repulsive force of  $-0.3 \times \text{distance between centroid 1 (A29 atoms N, C\alpha, C) and centroid 2 (P43 atoms C\gamma, C\delta, OXT) (kJ/mol)/\text{\AA}^2}$  was added to the system. This repulsive force decreased with the distance of A29 and P43, and it pushed the peptide ends apart if they got in close proximity, thus preventing the formation of a hairpin conformation. For this set up, additional 100 replicates  $\times$  50 ns were conducted for each peptide. Further details can be found in manuscript #3 (appendix I).

### 3.1.6. MD simulations of SETD2

The starting structures were retrieved from the PDB 5V21 for the SET domain of SETD2 complexed with the H3K36 peptide and PDB 6VDB for SETD2 complexed with the ssK36 peptide. Missing amino acids, M36K mutation, preparation of SAM and the Zn<sup>2+</sup> ions was carried out as described in the previous chapter (3.1.5). A 5 ns pressure coupled equilibration with Monte Carlo barostat was performed at a pressure of 1 atm. SETD2 and peptide C $\alpha$  atoms, as well as cofactor SAM atoms, were restrained with a force of 100, 100 and 5 kJ/mole  $\times \text{\AA}^2$ , respectively. The restraints were taken off successively, starting with the C $\alpha$  restraints, followed by a 5 ns equilibration with only SAM restrained. Subsequently, the SAM restraints were removed as well, followed by 5 ns equilibration with no restraints. For production, 15 replicates  $\times$  100 ns were performed for each complex, SETD2 with H3K36 and SETD2 with ssK36. Further details can be found in manuscript #3 (appendix I).

### **3.2. Trajectory analysis**

Basic trajectory preparations like atom or frame slicing as well as RMSD calculation, analysis of distances or angles e.g. for the S<sub>N</sub>2 Ts-like conformations were carried out by MDTraj (1.9.4) (McGibbon et al., 2015). Specific tools for contact analysis, clustering or volume calculation are specified below. The words frame and snapshot can be used interchangeably.

#### **3.2.1. Contact maps analysis**

The contact map analysis was performed utilizing Contact Map Explorer (0.7.1) (David W.H. Swenson, 2017). A contact was counted if at least one heteroatom of a residue was in a 4.5 Å<sup>3</sup> sphere surrounding one heteroatom from another residue excluding neighboring residues. Only for the analysis of the contacts of R882 and H882 in DNMT3A, a cut-off of 3.0 Å was used to get a higher resolution at this specific location. Further details can be found in the manuscripts #1, #2, #3 (appendix I) and the attached manuscript #4 (appendix II).

#### **3.2.2. NSD2 active site volume calculation**

Calculations of the volume around lysine 36 in the NSD2 WT and NSD2 T1150A simulations were performed using POVME3 (Wagner et al., 2017). Parameters for the calculation were a grid spacing of 0.4 Å, a distance cut of 0.4 Å, a contiguous points criterion of three, and a convex hull exclusion. The coordinates of the inclusion spheres were: 35.30, 40.30, 34.55 and 30.00, 39.77, 34.69 each with a radius of 3.0 Å. Out of the total simulation time of 3 μs for each NSD2 variant, 10% of the simulation frames were randomly chosen for analysis to decrease computational costs. This process was done in triplicates for each methylation state (me0, me1, and me2) for NSD2 WT and NSD2 T1150A. Further details can be found in manuscript #2 (appendix I).

#### **3.2.3. Clustering of peptide conformations in solution**

The conformations of the peptide in solution simulations were clustered by Enspara (0.1.1) (Porter et al., 2019). In total, 350,000 conformations from the MD simulations of H3K36 and ssK36 were processed. The clustering of the peptide conformations was carried out using the k-hybrid algorithm. The used metric was the conformational similarity based on backbone atom (Cα, C and N) root mean square deviation (RMSD). The conformations were distributed in 2, 3 or 5 clusters based on this metric. Each cluster was represented by a centroid structure, visualizing the peptide conformation, which most accurately characterizes the corresponding cluster. To ensure a hypothesis-free clustering and to

maximize the sampling space, the trajectories of both peptides were pooled before the clustering. Thereby, each conformation from either peptide had equal chances to be clustered with every other conformation. Further details can be found in manuscript #3 (appendix I).

### **3.3. SETD2 purification and peptide hairpin validation using FRET**

For the FRET experiments, peptides (Dabcyl-H3K36-Glutamine(EDANS), Dabcyl-ssK36-Glutamine(EDANS), H3K36-Glutamine(EDANS), ssK36-Glutamine(EDANS)) were obtained from JPT Peptide Technologies GmbH and dissolved in DMSO stock solutions at 5 mM. For the experiments, the peptides were diluted in 20mM HEPES pH 7.2, 200 mM KCl, 1 mM EDTA and 10% glycerol to a final concentration of 10  $\mu$ M. For control experiments, the peptides were diluted in 20 mM HEPES pH 7.2, 200 mM KCl and 10% glycerol to a concentration of 11  $\mu$ M and 1/10 of the volume of a Proteinase K solution (100  $\mu$ g/ml, dissolved in 20 mM HEPES pH 7, 100 mM NaCl and 10 mM  $MgCl_2$ ) was added and the mixture incubated for 1 h at 37 °C. The measurements were conducted using a Jasco FP-8300 spectrofluorometer. The excitation was set to 340 nm, and the emission was measured at 490 nm. The excitation and emission bandwidths were 5 and 10 nm, respectively. Samples were stirred at 600 rpm. The SETD2 binding kinetics were measured using 1  $\mu$ M peptide dissolved in 20 mM HEPES pH 7.2, 200 mM KCl, 1 mM EDTA and 10% glycerol supplemented with 2  $\mu$ M of SETD2 catalytic SET domain. Measurements were performed at 26 °C, except for the analysis of peptide conformations, a temperature gradient from 5 to 95 °C in 10 °C steps was applied.

The His<sub>6</sub>-tagged expression construct of SETD2 catalytic SET domain (S1347–K1711, UniProt: Q9BYW2) was expressed and purified. In short, the plasmid encoding the His<sub>6</sub>-tagged SETD2 SET domain was transformed into BL21 (DE3) Codon Plus cells and protein expression was induced by adding 0.2 mM IPTG at 17 °C for 16 h. Protein purification was conducted by affinity chromatography using Ni-NTA affinity agarose resin (Genaxxon bioscience). After washing the beads, the purified protein was eluted in elution buffer (220 mM imidazole, 30 mM KPI, 500 mM KCl, 0.2 mM DTT, 10% glycerol). Fractions containing SETD2 were pooled and dialyzed in dialysis buffer (20 mM HEPES pH 7.2, 200 mM KCl, 0.2 mM DTT, 1 mM EDTA, 10% glycerol) for 3 h at 8 °C. Subsequently, the protein was aliquoted, flash frozen in liquid nitrogen and stored at –80 °C. The purity of the protein preparation was analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) using 16% gel stained with colloidal Coomassie brilliant blue. Further details can be found in manuscript #3 (appendix I).

### **3.4. *In vitro* methylation assay to test peptide inhibitors**

The His<sub>6</sub>-tagged SETD2 catalytic SET domain (S1347–K1711) (9  $\mu$ M) was mixed with the unmodified GST-H3 protein (A1-L60, 4.5  $\mu$ M), and competitor peptides (increasing concentration, stock: 2 mg dissolved in 200  $\mu$ l DMSO, obtained from Shanghai RoyoBiotech Co., Ltd) in methylation buffer (20 mM

Tris/HCl HCl, pH 9, 1.5 mM MgCl<sub>2</sub> and 10 mM DTT) supplemented with 0.76 μM radioactively labeled SAM ([methyl-<sup>3</sup>H]-SAM (Perkin Elmer Inc., dissolved at 25 μM in 10 mM sulfuric acid)) over night at 37 °C. The reactions were stopped by the addition of SDS-PAGE loading buffer and heating for 5 min at 95 °C. Afterward, the samples were separated by Tricine-SDS-PAGE followed by the incubation of the gel in amplify NAMP100V (GE Healthcare) for 1 h on a shaker and drying of the gel for 2 h at 70 °C under vacuum. The signals of the transferred radioactively labeled methyl groups were detected by autoradiography using a Hyperfilm high performance autoradiography film (GE Healthcare) at -80 °C in the dark. The film was developed with an Optimax Typ TR machine after different exposure times. Quantification of scanned images was conducted with ImageJ.

## 4. Results

PKMTs and DNMTs influence a broad range of biological processes including gene expression, DNA repair, gene splicing and development. The activity of PKMTs and DNMTs is therefore highly regulated by distinct mechanisms to protect cellular functions and genome stability. One of these mechanisms is the sequence specificity of DNMTs and PKMTs, which ensures the methylation of the correct target residues. An altered sequence specificity leads to aberrant methylation patterns and could disturb the genome stability and cellular homeostasis. Another regulatory mechanism is the product specificity of PKMTs, since up to three methyl groups can be transferred to a single lysine residue. Remarkably, each methylation state can function as its own signal, causing different downstream effects. Moreover, investigating how PKMTs associate and interact with their substrate explains why certain substrates are preferred compared to others. The found mechanisms could be used to rationalize somatic cancer mutations on an atomistic resolution or serve as a basis for a coherent enzyme-specific inhibitor.

In this work, characterization of DNMT3A and SETD2 and NSD2 PKMTs regarding substrate- and product specificity was conducted by using a combination of MD simulation techniques and biochemical experiments. Precisely, the yet unclear molecular mechanism of the altered substrate specificity and multimerization of somatic cancer mutant DNMT3A R882H was revealed. Likewise, the NSD2 somatic mutation T1150A was found to alter the product specificity of this PKMT from a dimethyltransferase to trimethyltransferase, which could be pinned down to altered contacts in the active site. Lastly, based on the substrate specificities of the PKMTs SETD2 and NSD2, super-substrate peptides were designed, which were methylated ~100-fold faster compared to the canonical H3 peptide sequence. MD simulations with validating biochemical experiments identified the critical factors, which influenced the activity at multiple points during catalysis and lead to the very fast methylation of the super-substrates. The SETD2 super-substrate peptide was further validated as a competitive SETD2 inhibitor.

#### 4.1. MD Simulation of the somatic cancer mutation R882H of DNMT3A

Manuscript #1, appendix I

Mack A, Emperle M, **Schnee P**, Adam S, Pleiss J, Bashtrykov P, & Jeltsch A\*, (2022) Preferential Self-interaction of DNA Methyltransferase DNMT3A Subunits Containing the R882H Cancer Mutation Leads to Dominant Changes of Flanking Sequence Preferences. *J Mol Biol*, 434(7), 167482. doi.org/10.1016/j.jmb.2022.167482.

The somatic DNMT3A R882H mutation occurs frequently in AML, but the pathogenic mechanism of this mutation has remained elusive, despite many years of research. The mutation occurs mostly heterozygously and WT and R882H subunits co-exist in affected cells. R882 is located in the RD interface which is formed by the DNMT3A subunits in DNMT3A/L hetero tetramers (**Fig. 14A**). The RD interface also forms the DNA binding site of the complex. In this research project, my colleague Alexandra Mack conducted methylation experiments using DNMT3A, DNMT3A R882H and different DNA substrates. The results indicated that R882H causes strong changes in the flanking sequence preferences of DNMT3A. Specifically, DNMT3A R882H heavily disfavored substrates with thymine (Thy) and cytosine (Cyt) at the +1 position (the methylated CpG site marks the 0-position). These substrates were methylated 4 to 5-fold slower by R882H compared to the WT. If the +1 position is guanine (Gua) or adenine (Ade), it was methylated 2-fold faster by R882H compared to the WT. The change in flanking sequence preference extends up to the +4 site. Overall, R882H was more active than the WT at 13% and less active at 52% of all CGNNNN sites. The R882H characteristic flanking sequence preferences were observed in mixed tetrameric complexes containing WT and R882H subunits, indicating a dominant behavior of the R882H subunits. In mixed complexes, an R882H/R882H RD interface might thus be preferred.

The structure of DNMT3A R882H in complex with DNA (PDB 6W89) showed only subtle conformational change at the RD interface when compared with WT DNMT3A (PDB 6W8B) (**Fig. 14A**). Since the mutation is the only difference between the two enzymes, dynamic changes could be the reason for the altered sequence preference and for the preferred formation of DNMT3A R882H dimers at the RD interface.

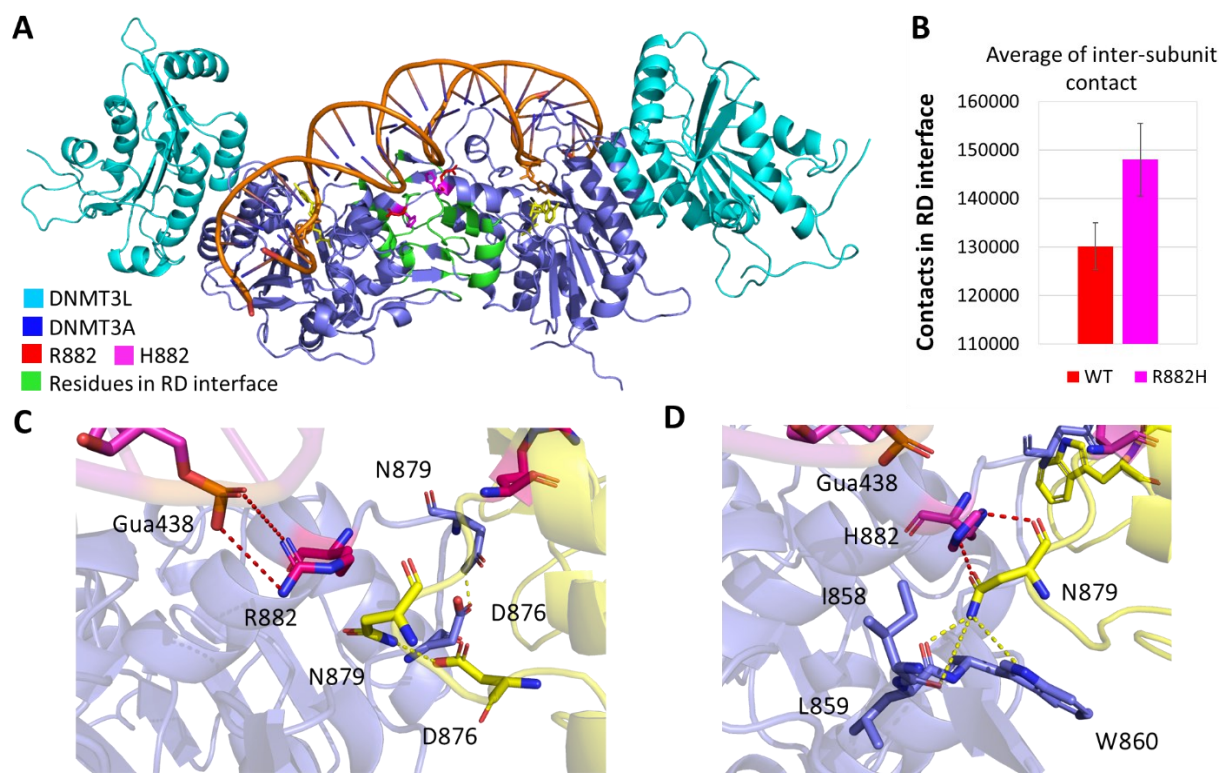
##### 4.1.1. DNMT3A R882H establishes more inter subunit contacts than the DNMT3A WT

The dynamic properties of the complexes of either DNMT3A WT or DNMT3A R882H together with DNA were inspected by MD simulations. After replacement of SAH in the structure by SAM, zebularine by cytosine, and removal of the covalent linkage between enzyme and DNA, 6 independent MD

simulation runs were conducted for both systems, each consisting of 3 ns restrained NPT equilibration, 3 ns restrained backbone equilibration and 3 ns unconstrained equilibration followed by 25 ns production runs. In a first analysis, it was investigated why the change in flanking sequence preference in R882H is a dominant effect and R882H has a preference to form at the RD interface, displacing the WT. For this, the MD simulation trajectories were analyzed, and a contact map generated for the residues in RD interface. Two residues were assumed to be in contact if the maximum distance between two heteroatoms was 4.5 Å (excluding neighboring residues). Contacts were counted in 2500 frames for each simulation replicate. The contact map shows how the amino acids at the RD interface dynamically behave over time. Examination of the contact maps revealed that DNMT3A R882H on average forms significantly more inter-subunit contacts at the RD interface than the WT (**Fig. 14B**). This result indicates that the R882H/R882H RD interface has a tighter packing and is more stable than the WT/WT RD interface, which agrees with its preferred formation.

To understand the molecular mechanism behind this finding, a more detailed analysis of the established contacts of R882 in WT and H882 in R882H was performed. The contact distance cut-off was reduced to 3.0 Å and new contact maps generated. Remarkably, it was observed that the R882 side chain points towards the DNA, contacting the phosphodiester group of Gua438 (**Fig. 14C**). This contact was present 96% of the simulation time over all replicates. Gua438 is of special interest since it connects the +3 and +4 site of the DNA. For the H882 mutant, this DNA contact was reduced 3-fold (31% of the simulation time), suggesting a different binding of DNA substrates. This is in agreement with the biochemical data showing altered flanking sequence preferences of WT and R882H up to position +4.

While R882 preferably interacts with the DNA, H882 adopts a different conformation pointing into the RD interface (**Fig. 14D**). Precisely, H882 was found to engage in contacts with N879 of the opposing DNMT3A R882H subunit. The H882-N879 contact was formed in 73% of the simulation time. In contrast, R882 was only in 20% of the simulation time engaged in the analogous contact (3.5-fold reduction). Moreover, the interaction with H882 caused a movement of N879 leading to the establishment of more inter-subunit contacts in the RD interface, namely N879 with I858, L859, and W860. This newly established contact network can explain the more stable binding of R882H subunits at the RD interface. In contrast, N879 in the WT complex is involved in contacts with residues from its own subunit, namely D876. Thus, less inter-subunit contacts are established for DNMT3A WT. This mechanism explains the preference to form DNMT3A R882H dimers at the RD interface.



**Figure 14: DNMT3A R882H establishes more and different contacts in the RD interface and different interactions with the DNA. A |** Overlay of representative structures of DNMT3A WT and R882H from a 25 ns MD simulation. The two DNMT3A subunits are shown in blue, DNMT3L in cyan and the bound DNA in orange. The RD interface is colored green. R882 and H882 are shown in red and pink, respectively. **B |** The R882H complex establishes more inter-subunit contacts at the RD interface than WT. In 6 independent 25 ns MS simulations of each complex, 2500 frames were analyzed per simulation replicate showing that the R882H complex forms more average inter-subunit RD contacts than WT ( $p$ -value =  $1.82 \times 10^{-3}$ ). The error bars correspond to the standard deviation of the average contact counts in the frames of individual MD runs. **C |** Detailed contacts of R882 in WT show that R882 (pink) points towards the DNA and interacts with Gua438 (contacts indicated by a red dotted lines). N879 (yellow) from the opposing DNMT3A subunit is involved in contacts with D876 in the same subunit (yellow, contacts indicated by yellow dotted lines). **D |** Detailed contact analysis of H882 showed that H882 points towards the RD interface and interacts with N879 from the opposing subunit (yellow, contact indicated by a pink dotted line). The movement of N879 causes this residue to establish additional inter-subunit contacts with I858, L859, and W860 (blue, contacts indicated by yellow dotted lines). Panels C and D show representative structures of the WT and R882H MD simulations runs. Figure taken and modified from (Mack et al., 2022).



## 4.2. MD simulation of the T1150A cancer mutant of NSD2

Manuscript #2, appendix I

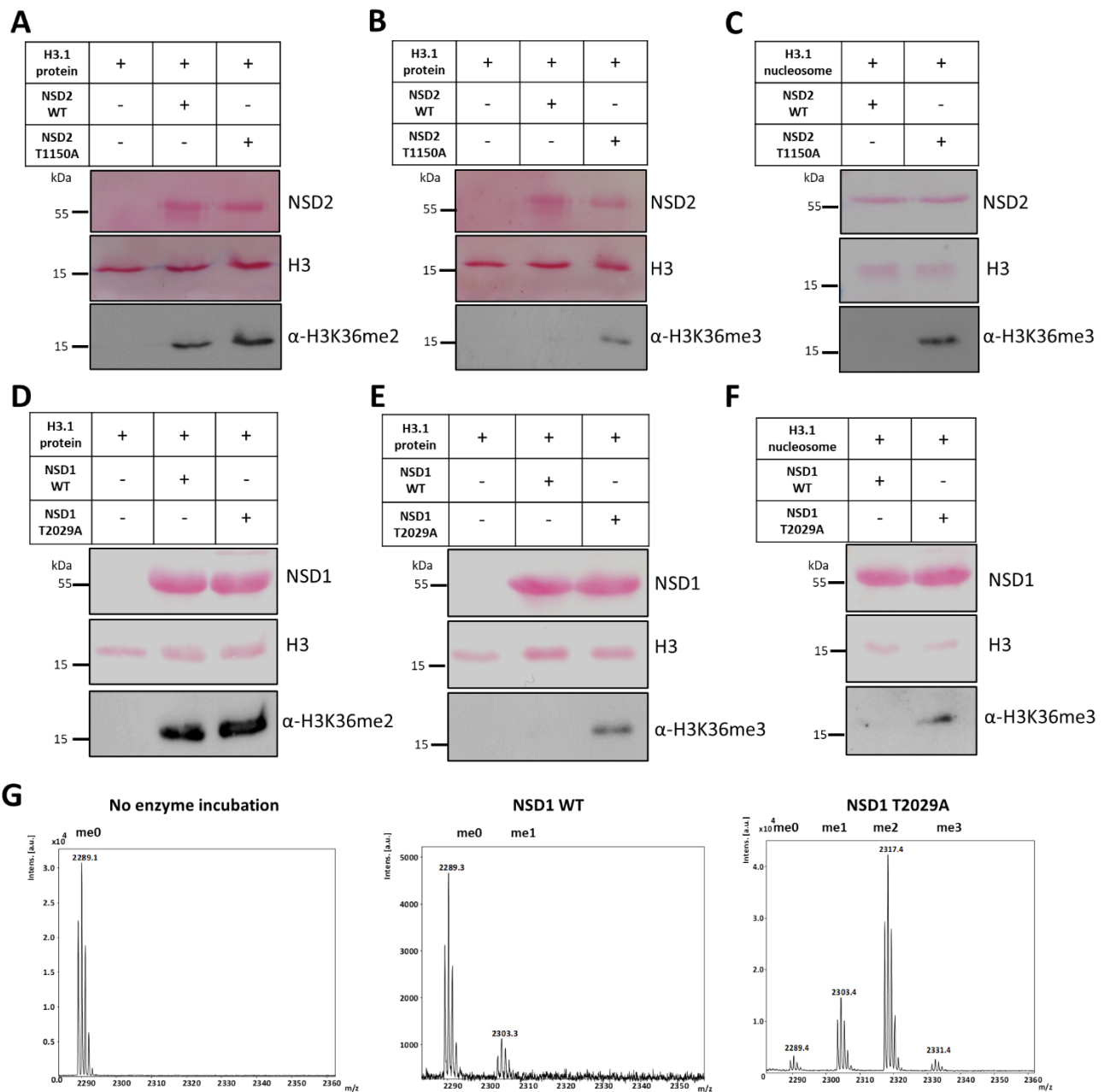
Khella M. S, **Schnee P**, Weirich S, Bui T, Brohm A, Bashtrykov P, Pleiss J, & Jeltsch A\*, (2023) The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation. *J Biol Chem*, 104796. doi.org/10.1016/j.jbc.2023.104796.

NSD2 has been shown to introduce up to dimethylation to H3K36 *in vitro* and *in vivo* (Dylan Husmann & Or Gozani, 2019; Li et al., 2019). The most frequently observed NSD2 SET domain missense mutation is E1099K, followed by T1150A (Bennett et al., 2017). Both mutations were specifically observed in patients with hematological cancers. While NSD2 E1099K has already been well characterized (Jaffe et al., 2013; Li et al., 2021; Oyer et al., 2014; Pierro et al., 2020; Sato et al., 2021), little was known about T1150A. *In vitro* methyltransferase assays, performed by my colleague Dr. Mina Saad Khella, uncovered that NSD2 T1150A and its paralogue NSD1 T2029A are hyperactive. Quantification of the methylation signals of NSD2 T1150A showed a ~9-fold higher activity on H3 peptide substrates, 4-fold on recombinant H3.1 protein, and 7-fold on recombinant mononucleosomes compared to NSD2 WT. Surprisingly, methylation experiments with subsequent Western Blot analysis using H3.1 recombinant protein and mononucleosomes as substrates discovered that NSD2 T1150A was able to generate H3K36me3 (**Fig. 15A-C**). This result was confirmed by methylation experiments of NSD1 T2029A with following MALDI mass spectrometry analysis using peptides as substrates (**Fig. 15D-F**). Additional controls showed that the H3K36me3 generation was not a consequence of hyperactivity of but due to a real change in product specificity of the mutant. This was demonstrated by MALDI mass spectrometry experiments, in which H3K36me3 was generated by T2029A under conditions of similar H3K36me2 activity of WT and T2029A.

The change in product specificity could rewrite the H3K36 methylation landscape leading to massive biological responses that may explain the carcinogenic effect of these mutations (DiFiore et al., 2020; Li et al., 2019; Eric J. Wagner & Phillip B. Carpenter, 2012). Since the observed change in product specificity is not a result of the hyperactivity, a distinct mechanism, which usually restricts trimethylation in NSD2 WT, might be damaged by the mutation. This was investigated using different MD simulation techniques of NSD2 WT and the NSD2 T1150A mutant.

For the correct setup of MD experiments, information about the reaction mechanism was essential, in particular, whether the enzyme acts in a processive or distributive manner. Several SET domain PKMTs were shown to exhibit a processive methylation of their target lysine residues (Dirk et al., 2007; Kwon et al., 2003; Patnaik et al., 2004; Zhang et al., 2003). In a processive reaction mechanism, the substrate is methylated multiple times without dissociation. The cofactor product SAH is still replaced after

catalysis by a new SAM molecule. In a distributive mechanism, the methylated substrate dissociates, and a new substrate needs to bind again for the next round of methylation. To investigate this for NSD1 and NSD2, Dr. Mina Saad Khella conducted methylation experiments using H3K36me1 peptide substrates and NSD1 T2029A (the only enzyme that was active enough for this kind of analysis) to detect the product methylation states by MALDI mass spectrometry. Noteworthy, starting with H3K36me0, a clear peak of dimethylated H3K36 product and some trimethylation was observed as mentioned before (**Fig. 15G**). In contrast, much lower methylation was detectable when the reaction was started under the same conditions with the H3K36me1 substrate (manuscript #2, appendix I). This result is not compatible with a distributive reaction mechanism and indicates that the enzyme catalyzes a processive methylation reaction.



**Figure 15: Product specificity change of NSD2 T1150A and NSD1 T2029A compared to WT enzymes on H3.1 protein and nucleosomes *in vitro*.** Western blot analysis using **A** | H3K36me2 specific antibody or **B** | H3K36me3 specific antibody after methylation of H3.1 recombinant protein with NSD2 WT or T1150A. **C** | Western blot analysis using the H3K36me3 specific antibody after methylation of H3.1 nucleosomes with NSD2 WT or T1150A. Western blot analysis signals using **D** | H3K36me2 specific antibody or **E** | H3K36me3 specific antibody after methylation of H3.1 recombinant protein with NSD1 WT versus T2029A mutant. **F** | Western blot analysis using the H3K36me3 specific antibody after methylation of H3.1 nucleosomes with NSD1 WT or T2029A. In all panels, the equal loading of substrates and enzymes is shown by Ponceau S staining. **G** | Methylation of the H3K36me0 peptide by NSD1 WT and T2029A analyzed by MALDI mass spectrometry. The figures show from left to right the mass spectra of the H3K36 peptide without NSD1 enzyme incubation, the H3K36 peptide after incubation with NSD1 WT or NSD1 T2029A. The samples were incubated in methylation buffer containing SAM for 4 h at 37 °C. The masses of the corresponding peptides are 2289 Da (H3K36me0), 2303 Da (H3K36me1), 2317 Da (H3K36me2), and 2331 Da (H3K36me3). Figure taken from (Khella et al., 2023a).

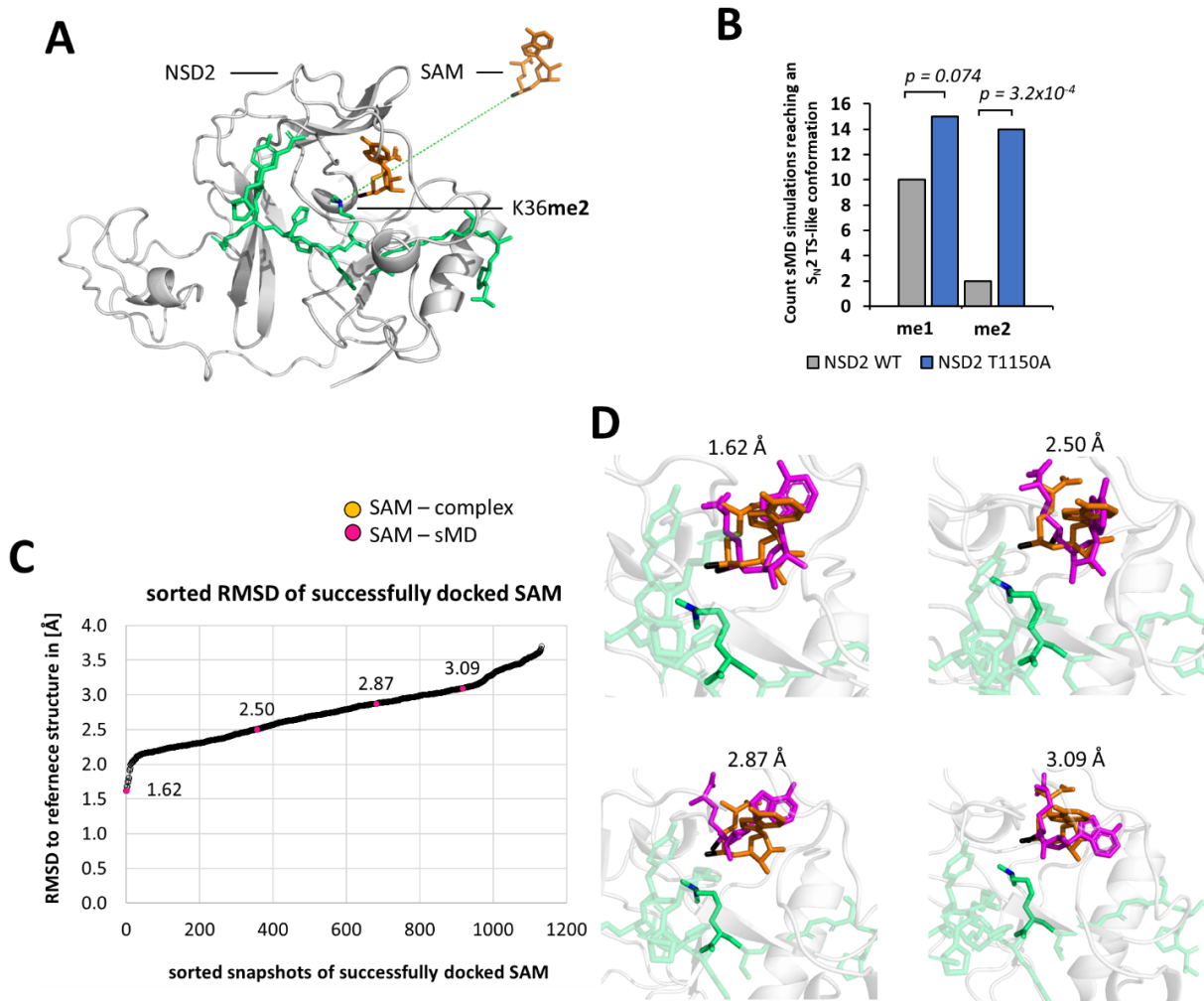
#### 4.2.1. NSD2 T1150A can accommodate a H3K36me2 peptide and SAM in sMD simulations

To study the change in product specificity of NSD2 T1150A, steered Molecular Dynamics (sMD) simulations were applied. In the sMD approach, external forces are used to guide bimolecular association processes (Yang et al., 2019). Thereby, reactions that otherwise would be too slow to be modeled in MD simulations are accelerated. At the same time, the conformational sampling is concentrated along a specific, predefined reaction coordinate. The association of a SAM molecule to NSD2 WT or T1150A containing a H3K36me1 or me2 peptide was simulated to study the potential generation of H3K36me3. If SAM successfully associates and is able to establish a productive conformation with the peptide, a higher methylation state could be facilitated. The system was built using the cryo-EM structure of NSD2 E1099K, T1150A bound to a nucleosome and SETD2 complexed with the H3K36M peptide as templates (PDB 7CRO and 5V21, see also Material and Methods Chapter 3.1.2). To simulate the SAM association process into the cofactor binding site, SAM was placed 27 Å above the NSD2 binding pocket and a weak attractive force of 0.2 kJ/(mol × Å<sup>2</sup>) was applied between the Nε-atom of K36 and the methyl group C-atom of SAM (**Fig. 16A**). In order to define criteria describing a successful docking, the geometric requirements for a S<sub>N</sub>2 transition state (TS)-like conformation were applied (chapter 1.5.6, Fig. 8). Since MD simulation cannot account for bond breaking or creating new bonds, descriptive criteria needed to be implemented to accurately describe a transition state (TS)-like conformation, wherein a methyl group transfer is most likely (Khella et al., 2023a; Philipp Schnee et al., 2022; Schnee et al., 2023). These geometric criteria are derived from the S<sub>N</sub>2 reaction mechanism and comprise:

- The distance between the lysine Nε and SAM methyl group C-atom is <4 Å.
- The angle between the lysine Nε - lysine Cδ bond and the virtual bond between lysine Nε and the SAM methyl group C-atom is in a range of 109° ± 30°.
- The angle between the lysine Nε- SAM methyl group C-atom and SAM sulfur S-atom is in a range of 180° ± 30°.

One hundred sMD simulations à 35 ns were performed for each NSD2 WT and T1150A and methylation state of the peptide. The number of simulation replicates were monitored in which at least once the S<sub>N</sub>2 criteria for a TS-like conformation were fulfilled. The analysis revealed 10 successful docking simulation replicates for NSD2 WT and 15 for NSD2 T1150A, both complexed with the H3K36me1 peptide. This corresponds to a non-significant difference (**Fig. 16B**). In contrast, complexed with the H3K36me2 peptide, NSD2 WT accommodated SAM successfully into the binding pocket in only 2 out of 100 simulations, whereas NSD2 T1150A reached a productive conformation in 14 out of 100 simulations (p-value 3.2 × 10<sup>-4</sup>) (**Fig. 16B**). To validate correct binding of SAM, the RMSD of successful SAM associations was compared to SAM already bound in NSD2, derived from crystal structures (PDB 7CRO), documenting good positional agreement (**Fig. 16C-D**).

The sMD experiment demonstrated an enhanced capability to accommodate SAM and the H3K36me2 substrate simultaneously for T1150A compared to WT. This fully recapitulated the result of the altered product specificity found in the methylation experiments and the ability to generate H3K36me3. No significant difference was found when H3K36me1 was the peptide substrate indicating that differences between NSD2 T1150A and WT only appear at higher methylation states.



**Figure 16: sMD simulation of SAM association into the complex of either NSD2 WT or T1150A with a H3K36me1 or me2 peptide substrate.** **A** | Starting position of the sMD simulations in which H3K36me1 or me2 is bound to NSD2 and SAM was positioned 27 Å away from the SAM binding pocket. **B** | Number of simulation replicates, in which at least once a  $S_N2$  TS-like conformation was reached. 100 sMD simulations à 35 ns of WT an T1150A complexed with peptides H3K36me1 or H3K36me2 were conducted. **C** | Comparison of the positions and conformations of SAM after successful sMD association with the positions and conformations of SAM modelled on the basis of available NSD2-Peptide-SAH structures. The figure shows all atom RMSD values of SAM, taken from successful SAM associations observed in 16 sMD simulations à 100 ns. The RMSD values in the small Å range document similar positions and orientations of SAM in both settings. **D** | Visualization of example structures. The RMSD value in Å is indicated at the top. SAM from sMD simulations is shown in purple. SAM modelled based on the SAH coordinates is shown in orange. Figure taken from (Khella et al., 2023a).

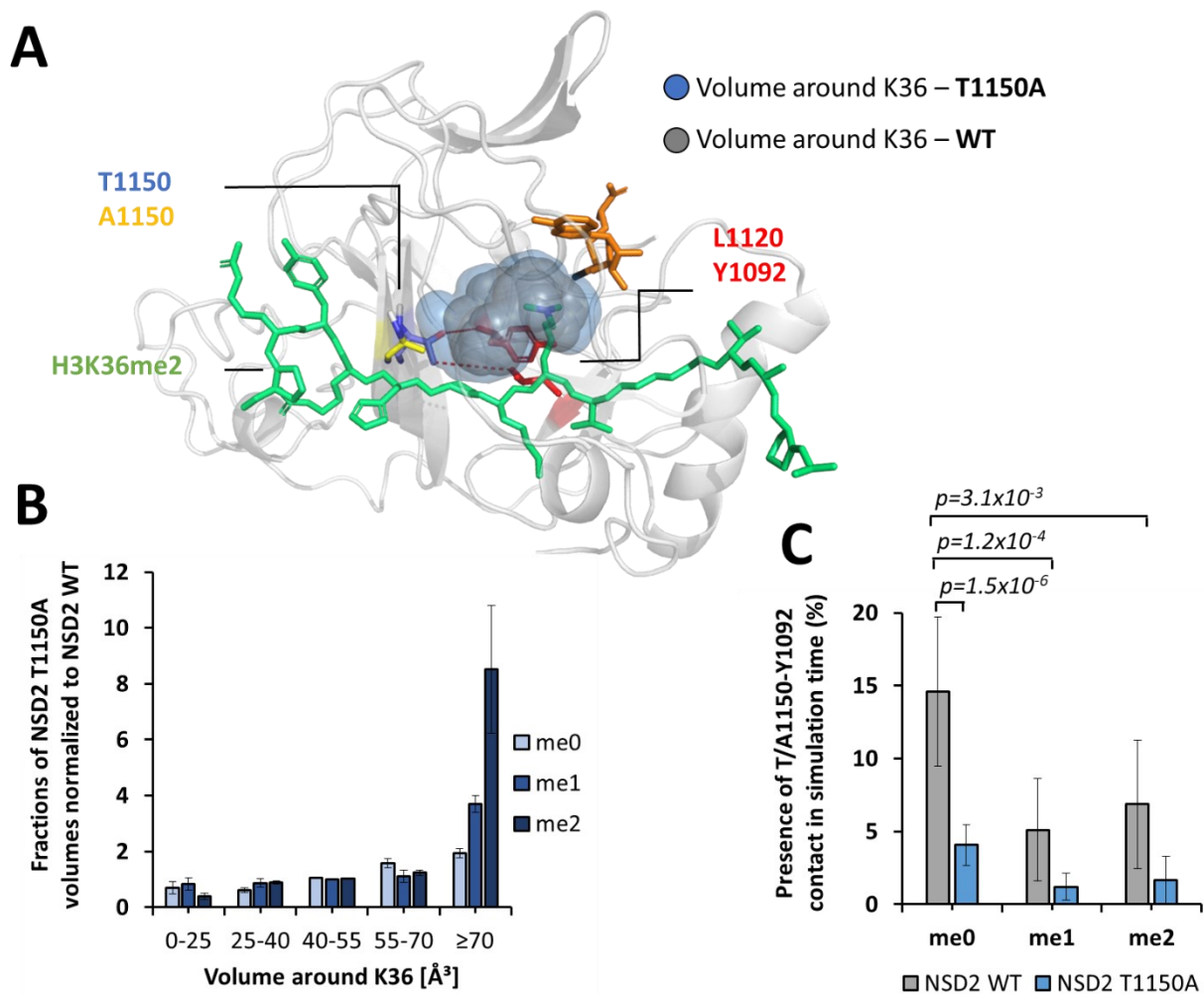
#### 4.2.2. NSD2 T1150A loses contacts responsible for restricting the active site volume

The sMD experiments showed that NSD2 T1150A had an increased ability to bind H3K36me2 and SAM simultaneously compared to NSD2 WT. To analyze the molecular foundations of this effect, the volume of the active site pocket was measured during the MD simulations. Furthermore, the contacts of the active site amino acids were examined. The analysis of the active site volume was carried out by simulating NSD2 WT and T1150A complexed with SAM and the H3K36 peptides in different methylation states. By having SAM already bound, a standardized comparison between NSD2 WT and NSD2 T1150A could be made and undersampling of NSD2 WT snapshots with productively bound SAM was avoided. For this, 30 MD simulations à 100 ns were conducted for each WT and T1150A. Out of this pool, 5000 randomly selected snapshots were used to calculate the active site volume around K36 (**Fig. 17A**). The analysis of the calculated volumes shows that large volumes ( $\geq 70 \text{ \AA}^3$ ) occurred more frequently with NSD2 T1150A, while small volumes ( $0\text{--}25 \text{ \AA}^3$ ) occurred more frequently with NSD2 WT. This effect increases with higher methylation levels of K36. For K36me2, large volumes occurred 8.5-fold more often for T1150A compared to WT (p-value 0.015, calculated by two-tailed t test assuming equal variance based on three replicates of the analysis) (**Fig. 17B**). The strong elevation of this effect with higher H3K36 methylation levels suggests that the active site tends to collapse with higher methylation levels. This effect is more pronounced with WT than with NSD2 T1150A. Overall, these findings clearly explain the increased capability of T1150A to accommodate SAM and H3K36me2 simultaneously.

Since the T1150A mutation is the only difference between the two proteins, the increased active site volume must be a direct consequence of this amino acid exchange. Contact analysis of the MD simulations of NSD2 complexed with the H3K36me0, me1, and me2 peptides revealed that T1150 is in contact with Y1092 and L1120 (**Figs. 17A**). Contacts were considered as established if the distance of a pair of heavy atoms from both amino acids was below  $4.5 \text{ \AA}$ . A polar interaction between the hydroxyl group of T1150 and the backbone amide of Y1092 was established in 15% of the simulation time with H3K36me0 but only 5% and 7% in the case of me1 and me2 (**Fig. 17C**). This contact orients Y1092 and restricts the volume of the active site pocket consequently disfavoring the methylation of me1 and me2 substrates. In the case of A1150, the contact with Y1092 is much less frequent, which supports further methylation of me1 and me2 substrates (4%, 1%, 2% for me0, me1, me2, respectively).

Moreover, a hydrophobic contact between the T1150 side chain methyl group and the C $\delta$ -atoms of L1120 was observed in 53% of the simulation time (average of me0, me1, and me2), but only in 24% of the time in the case of A1150 (p-value  $1.2 \times 10^{-14}$ , calculated by two-tailed t test assuming equal variance based on 30 replicates, data shown in Supplementary Information of manuscript #2 accessible through [doi.org/10.18419/darus-3263](https://doi.org/10.18419/darus-3263)). This interaction orients L1120, further restricting the volume of the active site pocket. Hence, the disruption of T1150 contacts to Y1092 and L1120 in T1150A is the

reason for the enlargement of the active site volume, leading to its change in product specificity and activity as a trimethyltransferase.



**Figure 17: Measurement of the active site pocket volume in NSD2 WT and T1150A complexes. A** | Overlay of structures NSD2 WT and T1150A with bound H3K36me2 and SAM as well as the corresponding volumes around K36. Red lines indicate the contacts of T1150 with Y1092 and L1120. **B** | Distribution of the volumes around K36 in Å<sup>3</sup> observed in MD simulations of NSD2 T1150A – peptide – SAM complexes, normalized to the corresponding values for NSD2 WT. **C** | Presence of the contact between T1150 and Y1092 in % of the simulation time in 30 simulations à 100 ns. Corresponding p-values were determined by two-tailed t test assuming equal variance based on the 30 sMD replicates. Figure taken from (Khella et al., 2023a).

### 4.3. Mechanistic basis of the enhanced methylation activity towards super-substrates

Manuscript #3, appendix I

**Schnee P**, Choudalakis M, Weirich S, Khella M. S, Carvalho H, Pleiss J, & Jeltsch A\*, (2022) Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide. *Communications Chemistry*, 5(1), 139. doi.org/10.1038/s42004-022-00753-w.

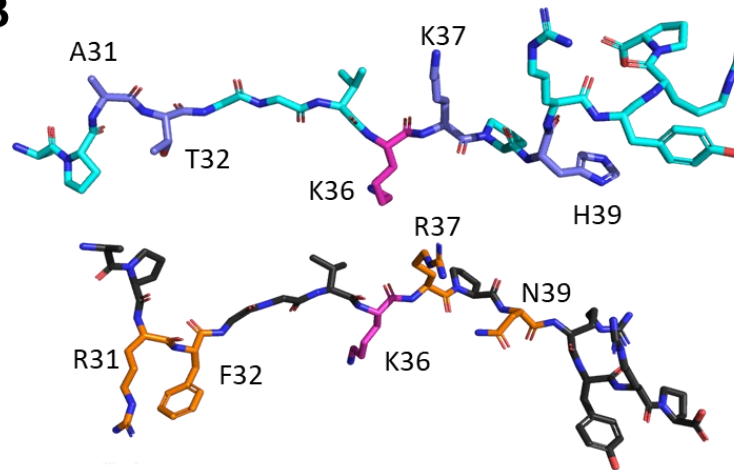
PKMTs have important regulatory functions in cells, but mechanisms determining their activity and specificity are incompletely understood. Naturally, SETD2 methylates the H3 histone tails at K36 and various non-histone protein substrates (Schuhmacher et al., 2020). Remarkably, the recently mapped substrate specificity of SETD2 revealed that H3 residues were not the most preferred ones at many positions of the target sequence (Schuhmacher et al., 2020). Based on this, an artificial peptide was designed that contained the most favorable residues at each position (**Fig. 18**). It was shown that this 15 amino-acid long super-substrate peptide (ssK36), which differed at four positions from the original H3K36 (A31R, T32F, K37R and H39N) was methylated more than 100-fold faster than the corresponding natural H3K36 peptide. The resolved crystal structure of the ssK36 peptide complexed to the SET domain of SETD2 showed only subtle differences to the H3K36-SETD2 structure and could not fully explain the heavily increased methylation activity.

To unravel the mechanistic basis for the enhanced methylation activity of SETD2 towards ssK36, a combination of MD simulations and biochemical experiments were conducted. The experiments considered multiple stages of the catalytic process: (i) the peptide conformations in solution; (ii) the association of the peptide into the enzyme's active site; (iii) the formation of transition-state (TS) like conformations in the enzyme-peptide complex and the established contacts.



**A**

Position:	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
H3K36 seq.:	A	P	A	T	G	G	V	K	K	P	H	R	Y	R	P
ssK36 seq.:	A	P	R	F	G	G	V	K	R	P	N	R	Y	R	P

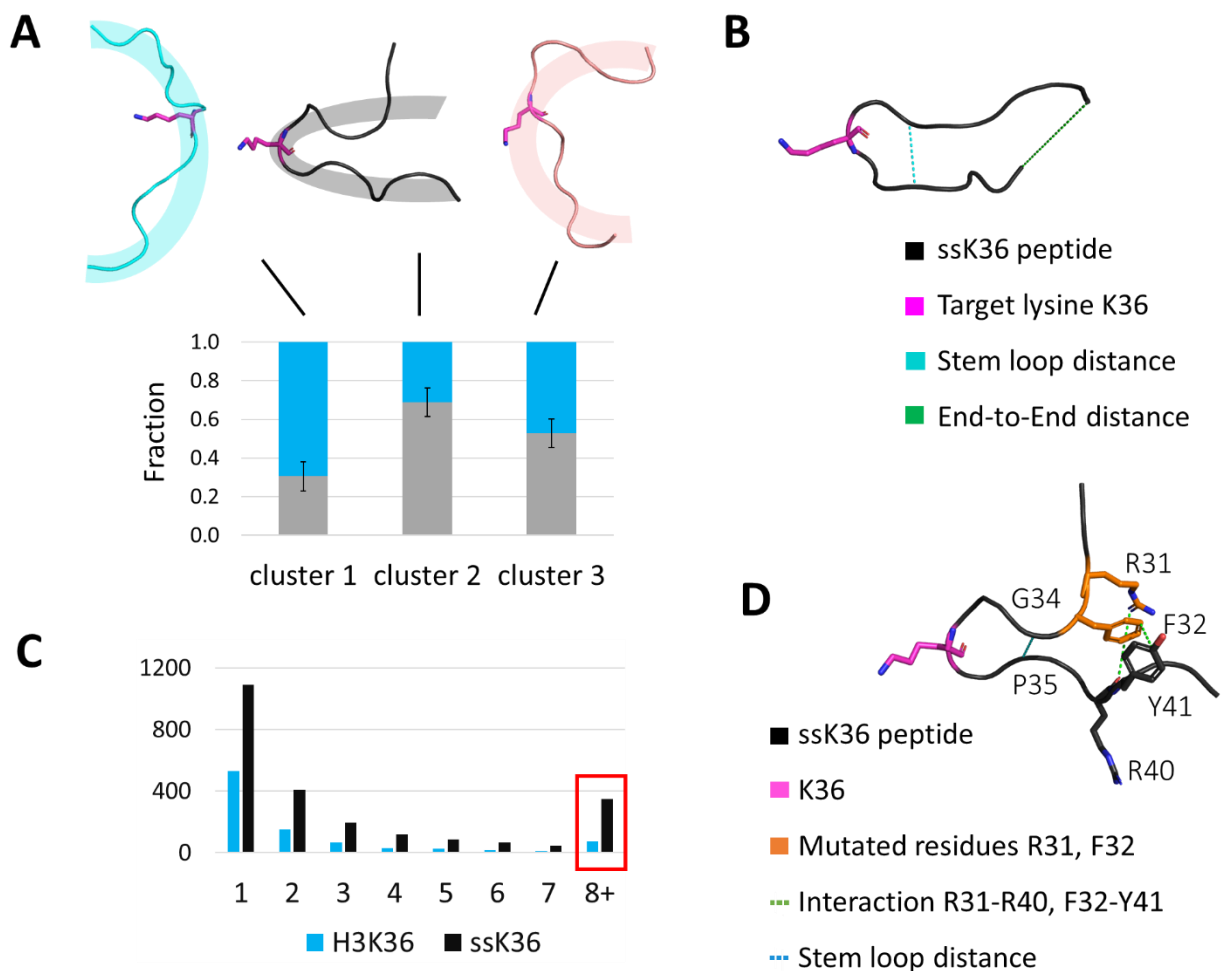
**B**

**Figure 18: The natural H3K36 peptide differs at four positions from the artificially designed super-substrate peptide (ssK36).** **A|** Sequences of the 15 amino acid long peptides H3K36 and ssK36. Mutated residues are colored blue in H3K36 and orange in ssK36. The target lysine 36 is colored pink. **B|** Visualization of peptide structures H3K36 and ssK36 in solution. The H3K36 peptide is depicted in cyan, the mutated residues are colored blue. The artificially designed ssK36 is shown in black and the ssK36 specific residues in orange. For both peptides, the target lysine 36 in colored pink. Figure taken and modified from (Philipp Schnee et al., 2022).

#### 4.3.1. The SETD2 super-substrate peptide prefers a hairpin conformation in solution

The first state at which the H3K36 and ssK36 peptide could show differences, which might influence their ability to be methylated by SETD2 is their conformation in solution. To observe potential conformational differences of both peptides in solution, each peptide was simulated in a water box for a total simulation time of 3.5  $\mu$ s (50 replicates  $\times$  70 ns for each peptide). To generate a variety of random starting positions, each replicate had an independent equilibration phase of 10 ns preceding the subsequent simulation. A total of 350,000 conformations from the MD simulations of H3K36 and ssK36 were processed by the clustering algorithm Enspara (Porter et al., 2019) and distributed into clusters. Clustering was based on conformational similarities measured as backbone atom root mean square deviation (RMSD). Each cluster was represented by a centroid structure, visualizing the peptide conformation, which most accurately characterizes the corresponding cluster. To ensure a hypothesis-free clustering and to maximize the sampling space, the trajectories of both peptides were pooled before the clustering. Thereby, each conformation from either peptide had equal chances to be clustered with every other conformation. Clustering into 3 groups revealed one cluster with an extended centroid structure, one sharply bent structure with a strong curvature centered at the target

lysine residue and one mixture between the extended and bent conformation (**Fig. 19A**). Next, it was determined how many conformations of H3K36 or ssK36 simulations were assigned to the respective clusters. This analysis revealed that cluster 1 mainly consisted of conformations observed in simulations of the H3K36 peptide (70%), while the majority of conformations in cluster 2 were produced in simulations with ssK36 (69%). Cluster 3 showed an almost equal distribution of H3K36 and ssK36 conformations (53% / 47%). This suggests that the H3K36 peptide preferably adopts an extended conformation in solution, whereas the ssK36 peptide prefers a hairpin conformation with a loop at the position of the target lysine. Decreasing the number of clusters led to an extended and a hairpin conformation, whereas the mixed conformation was not present anymore. Increasing the number of clusters led to a splitting of the mixed clusters into sub-clusters, while the extended and bent centroid structures for clusters 1 and 2 were maintained (Supplementary Fig. 1, manuscript #3, appendix I). To separate between transient and more stable conformations, the frequency of long-lasting hairpin structures of both peptides, defined as multiple consecutive MD simulation frames which displayed a hairpin conformation was measured (**Fig. 19B-C**). This analysis revealed that ssK36 is 4.8-fold more often in hairpin conformations with long lifetimes compared to H3K36, again indicating that hairpin stabilization is stronger for ssK36. A detailed analysis of the hairpin structures showed that the mutated amino acids in ssK36 are engaged in different intramolecular interactions, which stabilize the structure (Supplementary Fig. 2, manuscript #3, appendix I). These interactions, among others, include hydrogen bonds of R31 with the backbone atoms of R40 and stacking of F32 and Y41 (**Fig. 19D**). Such interactions were less pronounced for H3K36. This indicates that the different conformational preferences of H3K36 and ssK36 directly result from the four amino-acid exchanges in the ssK36 peptide.

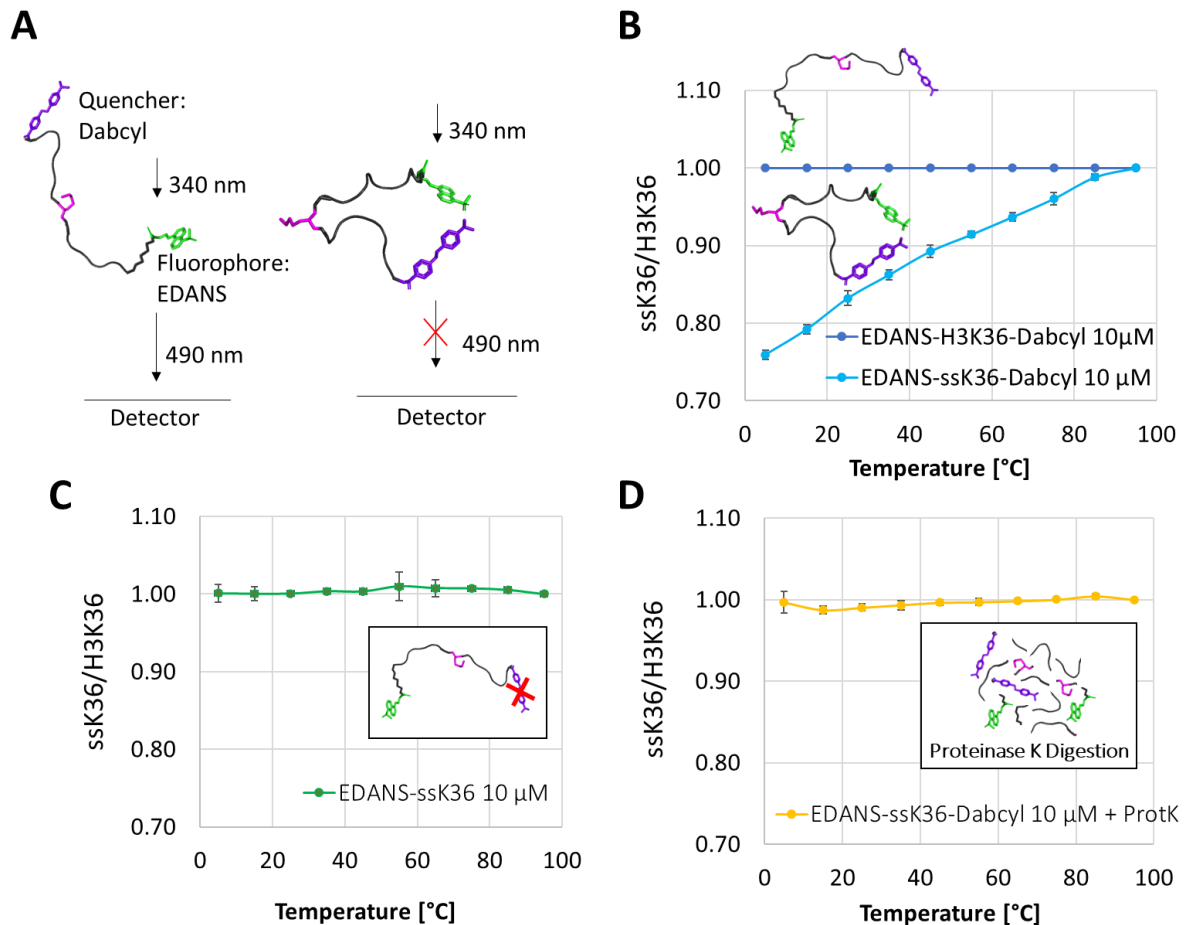


**Figure 19: Clustering of the H3K36 and ssK36 peptide conformations in solution observed in MD simulations reveals a hairpin conformation preference for ssK36.** **A** | A total of 350,000 conformations from MD simulations of H3K36 and ssK36 in solution were clustered based on backbone atom RMSD into 3 clusters. Cluster 1 is represented by an almost extended peptide conformation, which consists to 70% of conformations observed in simulations with H3K36. Cluster 2 is represented by a bent hairpin structure, with 69% of the conformations coming from ssK36 simulations. Cluster 3 is represented by a mixture of extended and bent structures and shows an almost equal contribution of H3K36 and ssK36 conformations (H3K36: 47%, ssK36: 53%). The bars represent the mean of three independent clustering replicates. Error bars represent the standard error of the mean. **B** | For further analysis, structures were assigned as hairpin conformations based on two parameters, a stem loop distance ( $G33 \text{ C}\alpha\text{-P38 C}\alpha$ )  $<7 \text{ \AA}$  and an end-to-end distance ( $A29 \text{ C}\alpha\text{-P43 C}\alpha$ )  $<15 \text{ \AA}$ . **C** | Distribution of the lifetimes of hairpin conformations, determined by the number of consecutive frames with the peptide in hairpin conformation in all peptide in solution simulations ( $50 \times 70 \text{ ns}$  for each peptide). Note the strong overrepresentation of long-living hairpin conformations of ssK36 (highlighted by the red box). **D** | The ssK36 specific residues F32 and R31 form contacts with Y41 and R40 supporting the formation of a hairpin conformation. Shown is the ssK36 peptide in hairpin conformation from a MD simulation replicate. Figure taken from (Philipp Schnee et al., 2022).

#### 4.3.2. Experimental investigation of conformational preferences

The peptide in solution simulations revealed different conformational preferences for H3K36 and ssK36. To challenge these simulation results experimentally, a FRET system was applied to determine the dynamic end-to-end distance distributions of both peptides in solution. For this, an EDANS

fluorophore was attached to the C-terminus of the peptides and a Dabcyl quencher to the N-terminus (Dabcyl-peptide-EDANS). EDANS is excited at 340 nm and emits fluorescence at 490 nm. Due to FRET, the emission is partially quenched by Dabcyl depending on the average end-to-end distance and the Förster radius of the FRET pair. Hence, if fluorophore and quencher are in close proximity, more FRET occurs, and less fluorescent light is detected (**Fig. 20A**). To assess the dynamic structural properties of both peptides in the FRET experiments, the fluorescence intensity of both peptides was measured at multiple temperatures. The intensity of Dabcyl-H3K36-EDANS was 33% higher than the intensity of Dabcyl-ssK36-EDANS at 5 °C. This difference suggests that the ssK36 peptide has a shorter average end-to-end distance (**Fig. 20B**). This result is in accordance with a preferential hairpin formation, which brings the quencher and fluorophore closer together, resulting in more FRET and less detectable fluorescence. As expected, the difference between H3K36 and ssK36 decreased with increasing temperature and completely disappeared at 95 °C. At this temperature, the hairpin conformations have unfolded, and no conformational differences of both peptides exist due to the enhanced thermal movement. After normalization to this endpoint, the differences between the two peptides were highly reproducible in two independent replicates of this experiment. To test these observations and tackle the question, of whether the experimental data do indeed show the suggested different conformational preferences, two control systems were used. In one system, only the fluorophore but no quencher was attached to each peptide (**Fig. 20C**). The observed data showed no temperature-dependent differences in their fluorescence. This indicates that the effect, which was seen before, indeed is due to temperature-dependent conformational preferences influencing the FRET efficiency. In a second control experiment, the FRET peptides with quencher and fluorophore were used, but digested with Proteinase K before the measurement (**Fig. 20D**). In this system, no conformational preferences are possible anymore. Again, no temperature-dependent differences between H3K36 and ssK36 were detected. This result confirms that the intact conformation of the peptides is necessary to observe the temperature-dependent FRET effects.



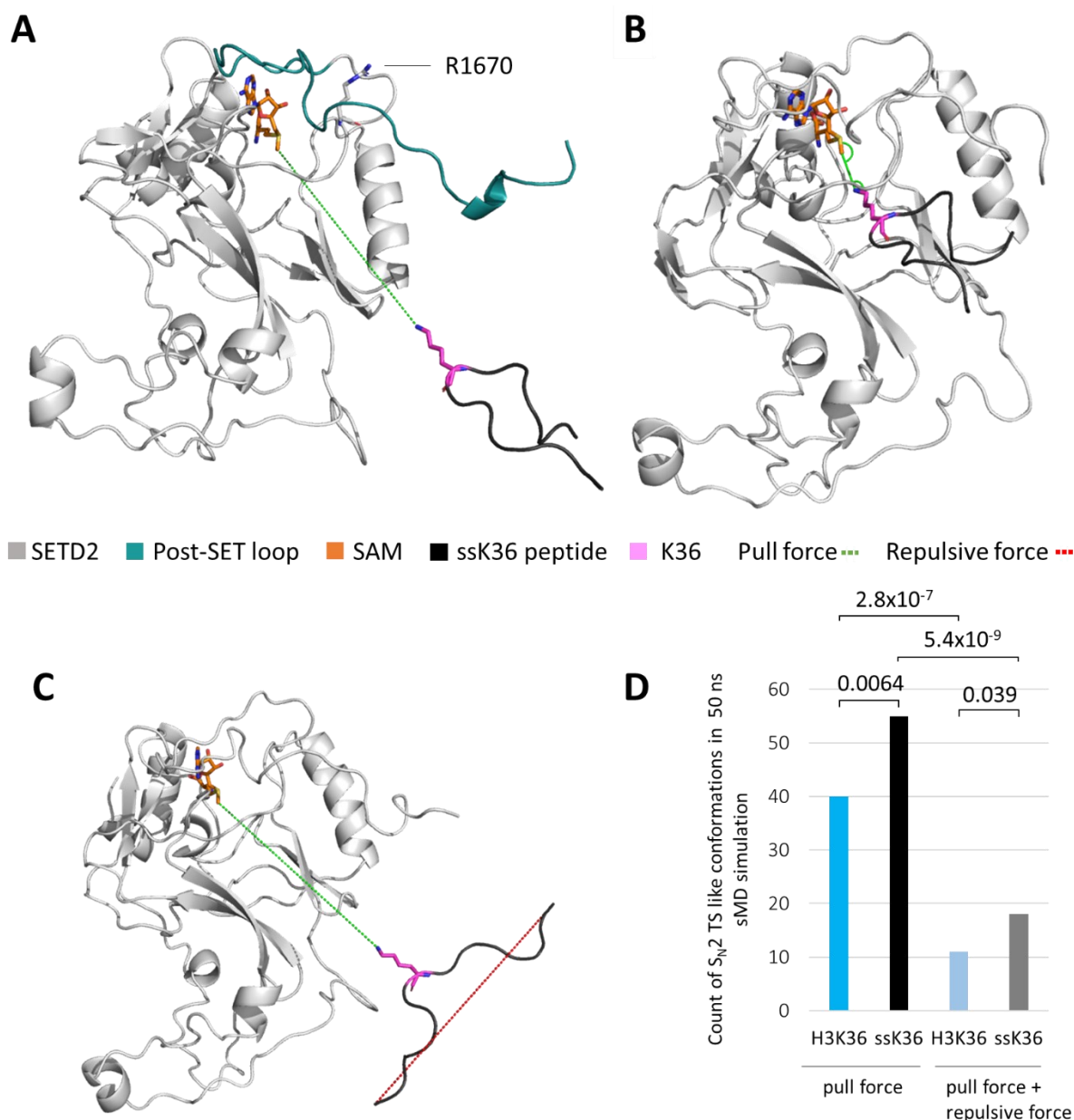
**Figure 20: H3K36 and ssK36 peptides show different conformational preferences in solution.** **A** | Scheme of the FRET system. H3K36 and ssK36 peptides were synthesized with an EDANS fluorophore attached to the C-terminus and a Dabcyl quencher at the N-terminus. EDANS was excited at 340 nm and fluorescence emission was measured at 490 nm. Because of FRET, the fluorescence emission is partially quenched by Dabcyl, and only the remaining fluorescence was measured. The FRET experiments were conducted at multiple temperatures starting at 5 °C up to 95 °C using peptide concentrations of 10 μM. **B** | At 5 °C, EDANS-H3K36-Dabcyl showed a 33% higher fluorescence intensity than EDANS-ssK36-Dabcyl. At 95 °C, both peptides displayed the same intensity. **C** | Control experiment without quencher showed no temperature-dependent difference between EDANS-H3K36 and EDANS-ssK36 fluorescence. **D** | Control experiment showed no difference in fluorescence intensity of EDANS-H3K36-Dabcyl and EDANS-ssK36-Dabcyl after their digestion with Proteinase K. In **B-D**, the fluorescence intensity of the ssK36 sample was normalized to the H3K36 sample. Data show average values of two independent experiments, error bars represent the deviation of the data points from the average. Figure taken from (Philipp Schnee et al., 2022).

#### 4.3.3. A hairpin conformation has easier access into the SETD2 active site

The results of the MD simulations and the FRET experiments suggested different conformational preferences for H3K36 and ssK36. To test whether these different conformations influence the binding rate to SETD2, sMD simulations were applied. For this, SETD2 was modelled in an open conformation, in which no peptide is bound, the post-SET loop (Q1676-K1703) is detached, and the AL in an open position. In this conformation, the placeholder residue R1670 points outwards, and does not block the active site (**Fig. 21A**). For sMD experiments, the H3K36 and ssK36 peptides were placed right above the open binding cleft, 30 Å away from the cofactor SAM with the K36 side chain facing towards the

SETD2 binding cleft. To accelerate the association process, a weak attractive force of  $0.5 \text{ kJ}/(\text{mol} \times \text{\AA}^2)$  was applied between the  $\text{N}\epsilon$ -atom of K36 and the methyl group C-atom of the SAM (**Fig. 21A**). In order to define criteria describing a successful docking, the geometric requirements for a transition state (TS)-like conformation were derived from the  $\text{S}_{\text{N}}2$  geometry (chapter 1.5.6, Fig. 8).

One hundred sMD simulations of 50 ns were performed for each peptide, and the number of docking events monitored which fulfilled the geometric  $\text{S}_{\text{N}}2$  criteria for TS-like conformations in at least one snapshot. The analysis of the sMD simulations revealed that both peptides were able to establish TS-like conformations following this definition. The H3K36 peptide successfully docked into the active site in 40 of 100 simulations, whereas the ssK36 peptide docked successfully in 55 out of 100 simulations (**Fig. 21B**). However, a caveat in this approach was that K36 is positioned in the middle of the peptide and the pulling on this residue with the sMD force automatically induced a hairpin conformation of the peptide due to the friction of the movement. Therefore, both peptides eventually formed hairpin conformations when approaching the active site of SETD2 (90% of H3K36 simulation replicates and 100% of ssK36 simulation replicates at least transiently showed a peptide end-to-end distance  $<15 \text{ \AA}$ ). To preserve the different conformational preferences of the peptides during the sMD experiment, an additional repulsive force of  $0.3 \text{ kJ}/(\text{mol} \times \text{\AA}^2)$  between the ends of the peptides was added to the system. This repulsive force increased with decreasing distance between N- and C-termini (N-C distance) and thereby counteracted the formation of a hairpin conformation during the association process (**Fig. 21C**). With this approach, a direct comparison between the docking efficiency of hairpin and extended conformations was possible. 100 replicates of this sMD experiment showed that the number of successful dockings strongly decreased if the peptides were not able to approach SETD2 in a hairpin state (**Fig. 21D**). Overall, with the repulsive force, H3K36 docked only 11 times successfully and ssK36 18 times, corresponding to a 3.6-fold (H3K36) and 2.8-fold (ssK36) decrease in docking efficiency when compared to the unconstrained settings. This indicates that the hairpin formation heavily increased the chances of both peptides to successfully dock into the peptide binding cleft of SETD2. Moreover, in both settings, a higher number of productive docking events was observed with ssK36 than with H3K36. All differences between ssK36 and H3K36 with and without repulsive force were highly significant (**Fig. 21D**). This finding underscores the higher intrinsic ability of ssK36 to adopt structures that are more likely to achieve TS-like conformations.



**Figure 21: Hairpin conformations facilitate the access of peptides into the binding cleft of SETD2.** **A** | sMD simulation of H3K36 and ssK36 driven by a distance-dependent external force between the N $\epsilon$ -atom of K36 and the methyl group C-atom of SAM. Shown is the starting position of sMD simulation replicates, in which the peptide was positioned 30 Å away from SAM, right above the peptide binding cleft of SETD2 in an open conformation. **B** | Example of a successful docking of ssK36 into the active site of SETD2 after 50 ns sMD. **C** | sMD simulation with an additional, distance-dependent repulsive force between the peptide ends to counteract hairpin formation. **D** | Number of successful docking events in 100 sMD simulations with H3K36 or ssK36 based on the S<sub>N</sub>2 TS-like criteria. Figure taken from (Philipp Schnee et al., 2022).

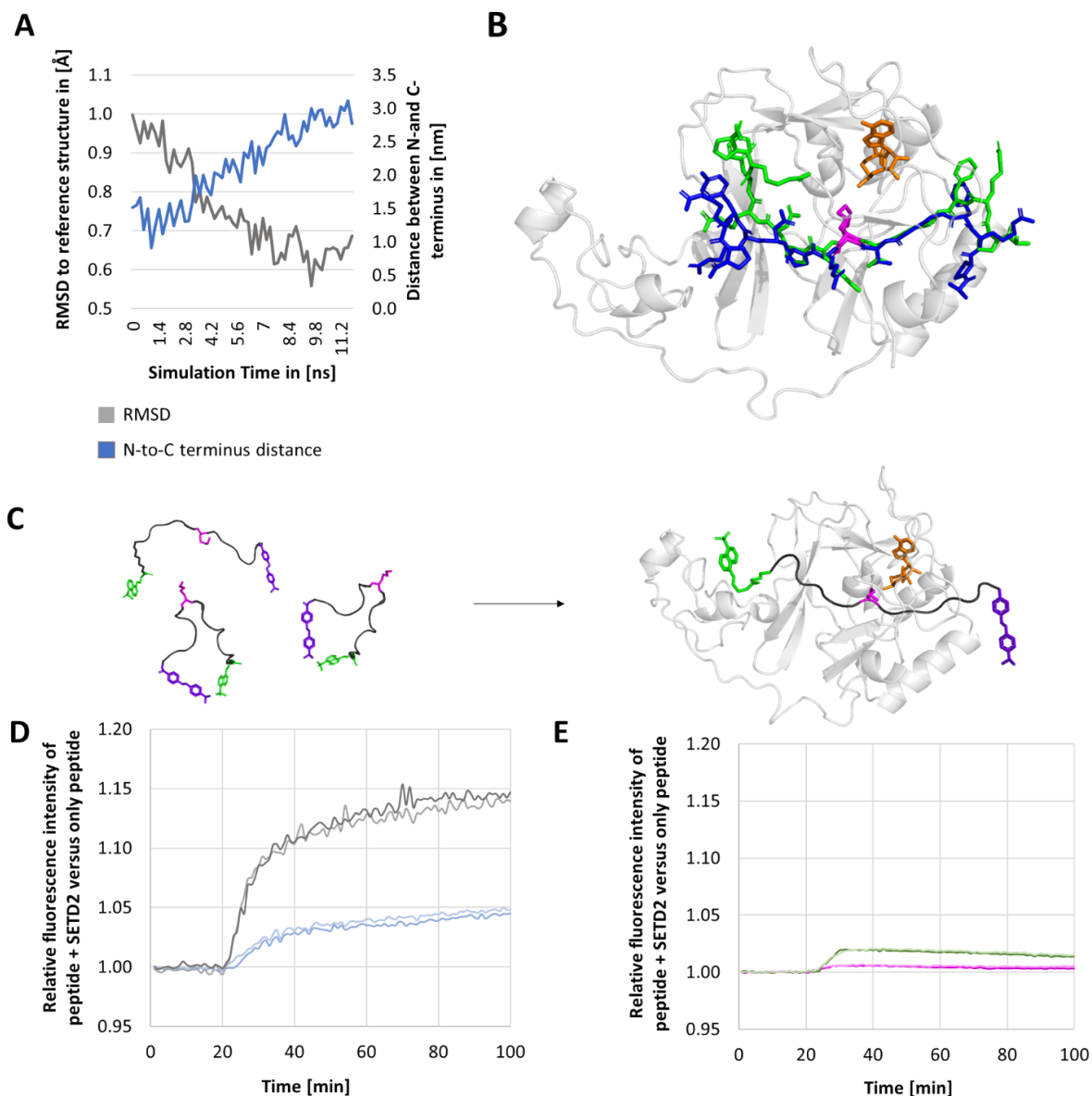
#### 4.3.4. Hairpin structures unfold into an extended conformation upon binding to SETD2

The sMD experiments showed easier access of hairpin conformations into the active site of SETD2 compared to extended peptide conformations. However, crystal structures of SETD2 complexed with either H3K36 or ssK36 show both peptides bound in an extended conformation in the binding cleft of SETD2 (PDB 5V21 for H3K36, PDB 6VDB for ssK36). Therefore, the conformational changes during the

sMD simulations were analyzed in more detail to find out if transitions occurred between the hairpin and extended conformations, after the initial docking of the hairpin peptide into SETD2. The mentioned crystal structures were used as references. The conformations of peptides observed in sMD simulations without repulsive force that reached a TS-like structure were compared to the reference conformations based on backbone RMSD. In fact, 84% of ssK36 peptides reduced their RMSD to the extended reference conformation during successful docking events, as well as 68% of the H3K36 peptides, indicating that in most cases, the structures of the docked peptides approached a conformation similar to crystal structures. The reduction in RMSD entails an unfolding process, in which the hairpin structure opens up and the N-C-terminal distance increases. To monitor this process, the N-C distance of the peptides was measured during sMD simulations with successful docking. For ssK36, 82% of the cases showed an increasing N-C distance during the simulation, as well as 85% of the H3K36 replicates. This indicates that in many simulations during the docking process, both peptides unfold from a hairpin conformation into an extended conformation similar to the crystal structures (**Fig. 22A**). Final structures of sMD simulations often showed a very good superposition with the extended peptide conformations observed in the crystal structure analysis, in particular at residues surrounding K36 (**Fig. 22B**).

The analysis of the sMD simulations suggested that an unfolding of hairpin conformations occurs upon binding to the active site of SETD2. To validate this observation experimentally, the H3K36 and ssK36 peptides with attached fluorophore and quencher were used to monitor the peptide conformation by FRET during association to SETD2 (**Fig. 22C**). Strikingly, the addition of SETD2 led to an increase in the fluorescence intensity of both peptides, but the extent of the intensity gain differed (**Fig. 22D**). The addition of SETD2 to H3K36 caused a 5% increase in intensity, while the intensity increased by 14.8% in the case of ssK36. This result suggests that dynamic hairpin conformations of both peptides are resolved during SETD2 binding since fluorophore and quencher are being separated. The finding that the addition of SETD2 caused a 3-fold stronger fluorescence increase for ssK36 than H3K36 can be connected to more hairpin conformations present in solution in the case of ssK36 and/or a more efficient binding of SETD2 to ssK36. To confirm that the intensity increase was caused by the interaction between quencher and fluorophore, peptides without quencher were used (**Fig. 22E**). Upon addition of SETD2, the intensity increased only slightly for both peptides, much less than for the peptides with the quencher attached (0.18% for EDANS-H3K36 and 1.2% for EDANS-ssK36).

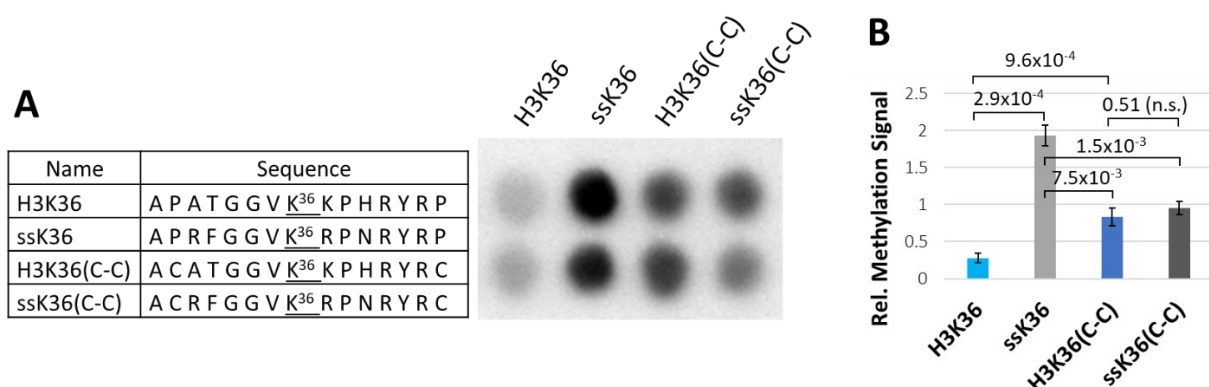




**Figure 22: H3K36 and ssK36 peptides unfold upon binding to SETD2.** **A** | RMSD between the peptide conformation in the sMD simulation and the extended peptide conformation in the crystal structure (grey) and peptide N-C terminal distance (blue) in the sMD simulations without repulsive force which resulted in a TS-like structure. The N-C distance was calculated as the distance between the backbone nitrogen atoms of A29 and P43. The RMSD and N-C distance were tracked until the minimum RMSD position was reached. Shown is a representative replicate from ssK36 sMD simulations. Additional examples are shown in the respective publication (Philipp Schnee et al., 2022). **B** | Representative sMD simulation structure with an RMSD of  $\sim 0.5$  Å overlaid with crystal structure PDB 6VDB. **C** | Schematic representation of EDANS-H3K36-Dabcyl and EDANS-ssK36-Dabcyl hairpin conformation unfolding upon binding to SETD2 leading to an increase in the N-C distance and reduced FRET. **D** | EDANS-H3K36-Dabcyl ( $1 \mu\text{M}$ ) and EDANS-ssK36-Dabcyl ( $1 \mu\text{M}$ ) were dissolved in buffer and the fluorescence was monitored. After 20 min, SETD2 ( $2 \mu\text{M}$ ) was added to each peptide resulting in an intensity increase of the fluorescence. To monitor fluorescence changes due to the changes in buffer volume and concentrations, control experiments were conducted with addition of buffer to the peptides instead of SETD2 (data shown in Supplementary Information of manuscript #3, accessible through [doi.org/10.18419/darus-2508](https://doi.org/10.18419/darus-2508)). The fluorescence signal of peptides after addition of SETD2 was expressed relative to the signal after addition of equal volumes of buffer. **E** | Peptides without quencher were used as a negative control to ensure that FRET was the reason for the fluorescence intensity increase observed in panel D. In each panel, the results of two independent experiments are shown. Figure taken from (Philipp Schnee et al., 2022).

#### 4.3.5. Hairpin conformation in peptides lead to a faster methylation by SETD2

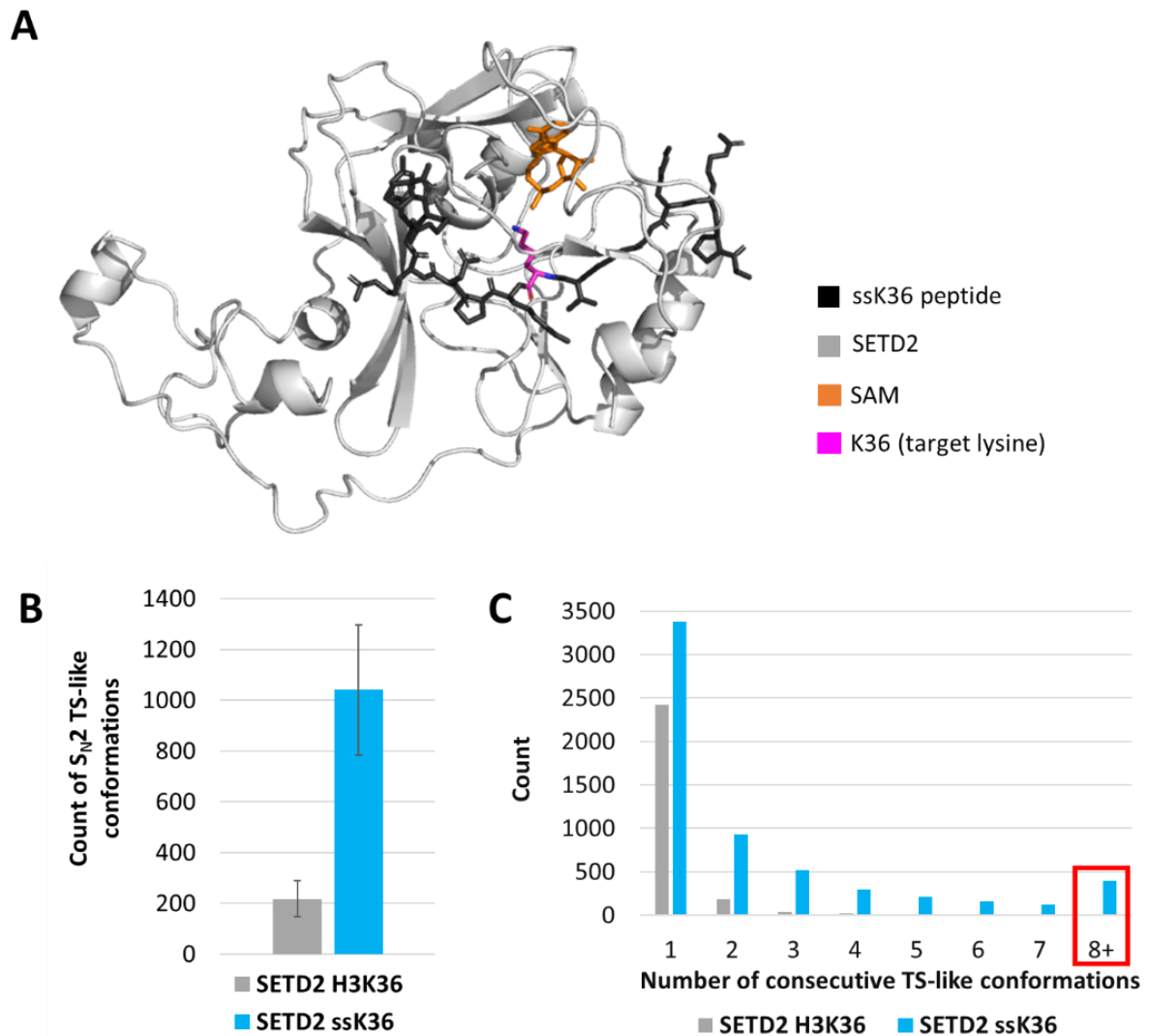
The MD simulation results predicted an association process, in which the binding of a peptide is preferred in a hairpin conformation. During the binding process, the hairpin unfolds, and the peptide adopts the extended conformations seen in crystal structures. To verify these predictions, Dr. Mina Saad Khella conducted Celluspot peptide array methylation experiments with H3K36 and ssK36 peptides and H3K36(C-C) and ssK36(C-C) variants thereof, which contain two cysteine residues at position 30 (P30C) and 43 (P43C) allowing them to form disulfide bonds under oxidative conditions. The disulfide bond arrests the peptide in a hairpin conformation with K36 presented in the loop of the hairpin (**Fig. 23**). The exact methods for the methylation assay can be found in manuscript#3, appendix II. The observed methylation rates show that in the chemically fixed hairpin conformation, H3K36(C-C) is methylated faster than H3K36. However, the chemically fixed hairpin conformation of ssK36, ssK36(C-C), is methylated worse than ssK36. H3K36(C-C) and ssK36(C-C) are methylated almost equally. These findings can be summarized as:  $H3K36 \ll H3K36(C-C) \sim ssK36(C-C) < ssK36$ . The fact that the hairpin stabilized H3K36(C-C) is methylated faster than the original H3K36 peptide clearly indicates that hairpin formation accelerates the reaction. However, the hairpin peptides H3K36(C-C) and ssK36(C-C) are methylated at a reduced rate when compared with ssK36, indicating that the unfolding of the peptide during complex formation is beneficial for methylation, because it allows for the formation of additional contacts between SETD2 and the peptide. Finally, the observation that the hairpin stabilized forms of H3K36 and ssK36 were methylated at similar rates supports the notion that the hairpin formation potential is one of the critical differences between both peptides explaining large parts of their different methylation rates by SETD2.



**Figure 23: Hairpin formation and resolution upon binding increase SETD2 methylation activity.** **A** | Sequences of the peptide substrates used for *in vitro* SPOT peptide methylation assay using radioactively labeled SAM (Kudithipudi, Kusevic, et al., 2014). Autoradiography image after 2 weeks of film exposure. The two lanes represent technical duplicates and contain identical peptides. **B** | Average of three independent experiments each containing a technical duplicate set of all spots. The error bars represent the standard error of mean. Signals were normalized to the average signal of all spots in each experiment. P-values were determined by t-test for a two-tailed distribution of paired values. Figure taken from (Philipp Schnee et al., 2022).

#### 4.3.6. ssK36 establishes more and different TS-like conformation than H3K36

The crystal structures of SETD2 with bound H3K36 or ssK36 peptide only resolve the ground state conformation of both complexes. Therefore, they were used as a starting point for MD simulations to investigate the interplay between protein and peptide over time in atomistic resolution to obtain more information about the TS-like conformations that are most relevant for enzymatic catalysis (**Fig. 24A**). In a total of 3  $\mu$ s simulation time (15 simulations à 100 ns for each peptide), ssK36 established significantly more  $S_N2$  TS-like structures than H3K36. The complex of SETD2 with ssK36 was, on average 1117 times per simulation in a state, in which all TS criteria were fulfilled, and the  $S_N2$  reaction could take place. In contrast, the complex of SETD2 with H3K36 reached the TS-like conformation only 305 times on average per simulation (p-value  $1.2 \times 10^{-8}$ , based on two-tailed t-test with equal variance) (**Fig. 24B**). The increased number of TS-like conformations of ssK36 was accompanied by an elevated lifetime of these conformations (**Fig. 24C**). TS-like conformations established by H3K36 were mostly transient. Only 13 events were identified in which  $S_N2$  criteria were fulfilled in 8 or more consecutive frames of the simulation. In contrast, the ssK36 peptide achieved long-lasting TS-like states frequently (397 events fulfilled  $S_N2$  criteria in 8 or more consecutive frames).



**Figure 24: The complex of SETD2-ssK36 established significantly more TS-like conformations than SETD2-H3K36. A|** Model of SETD2 SET domain complexed with ssK36 derived from crystal structure PDB 6VDB. **B|** Number of TS-like conformations reached in MD simulations of the H3K36 and ssK36-SETD2 complexes. The graph shows the average of 15 simulations à 100 ns. Error bars represent the standard error of mean (p-value  $1.2 \times 10^{-8}$ , based on two-tailed t-Test with equal variance). **C|** Stability of TS-like conformations formed by SETD2-H3K36 and SETD2-ssK36 complexes as determined by the distribution of events with TS-like conformations persisting over a given number of consecutive MD frames. Note the strong overrepresentation of long-living TS-like conformations of ssK36 (highlighted by the red box). Figure taken from (Philipp Schnee et al., 2022).

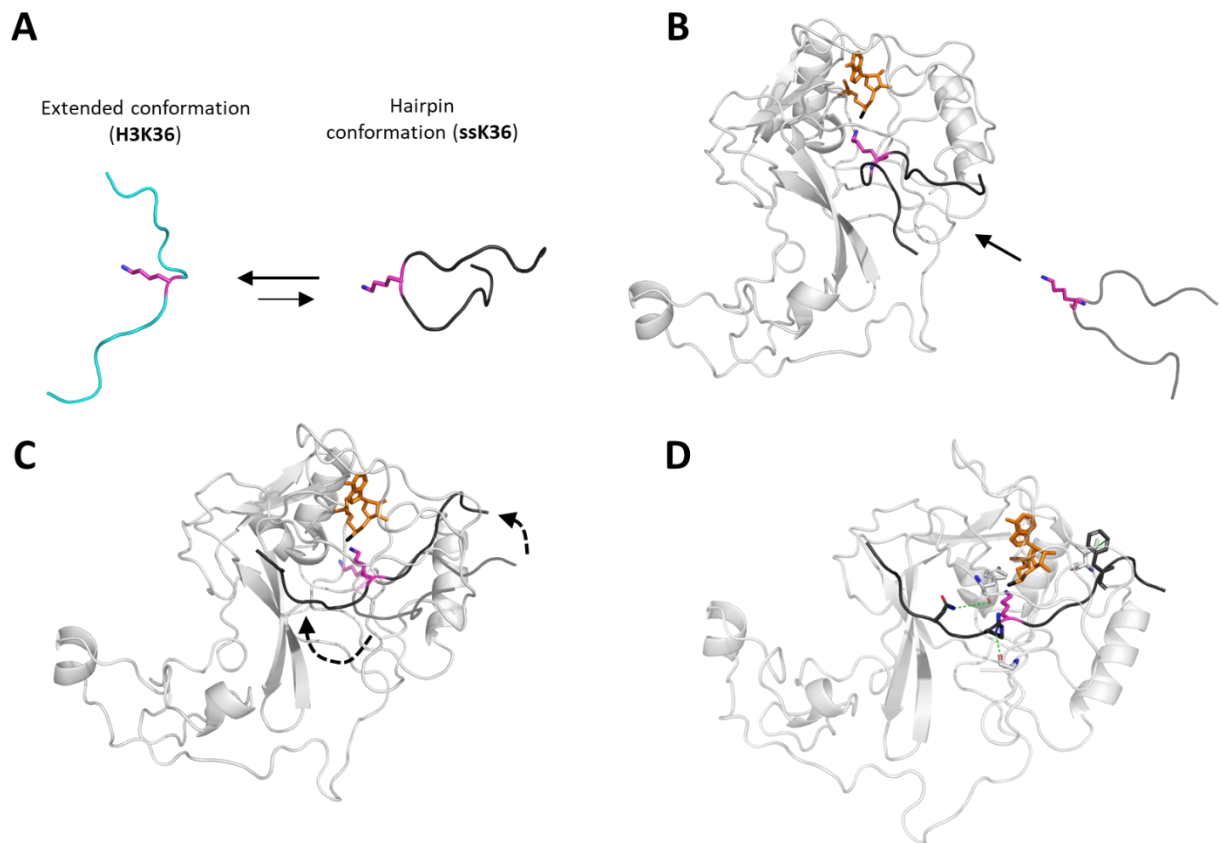
To identify the molecular mechanisms leading to the differing abilities of the H3K36- and ssK36-SETD2 complexes to reach TS-like structures, the contacts established between SETD2 and each peptide during the simulations were analyzed as described earlier for NSD2 (chapter 4.2.2). The resulting contact profile displayed contacts that were already observed in the crystal structure comparison before (Schuhmacher et al., 2020). However, a large difference in the contact map was found at residue 39, where H3 carries a histidine and ssK36 an asparagine, that was not detected in the static

crystal structure analysis (**Fig. 25A**). For detailed description of the contacts refer to original publication (manuscript #3, appendix I).

Special attention needs to be paid to the contacts during the TS-like conformation, since these contacts are most relevant in stabilizing a catalytically component state. Surprisingly, this state is not reached by H3K36 and ssK36 in the same way (**Fig. 25B**). The contact profiles, which only considered TS-like conformation snapshots, differ heavily. The result of the different contacts showed that H3K36 and ssK36 have different average conformation in their TS-like conformation. The N- and C-terminal part of the ssK36 peptide were shifted towards SETD2, whereas the middle part of the peptide was slightly detached from SETD2, potentially giving more flexibility for K36 to orient itself (Fig. 8C in manuscript #3, appendix I). A counter movement was observed for H3K36, which showed a closer binding to SETD2 around K36 and more detached at the N- and C-terminal part. This behavior might explain the overall higher number of TS-like conformation for ssK36 and the better stabilization leading to longer lifetimes.



of the peptide in a zipper-like process. This leads to an unfolding of the hairpin and adoption of an extended peptide conformation as observed in the crystal structures of SETD2 complexed with either H3K36 or ssK36 (**Fig. 26C**). Peptide-specific contacts with the enzyme are subsequently established stabilizing  $S_N2$  TS-like conformations between target lysine and Sam methyl group (**Fig. 26D**).



**Figure 26: The enhanced methylation activity of SETD2 towards ssK36 can be summarized as four steps. A|** Preferred hairpin formation of ssK36 in solution due to favorable intrapeptide contacts. **B|** Accelerated binding of the hairpin into the SETD2 peptide binding cleft with the K36 facing outside. **C|** Unfolding of the hairpin conformation in the SETD2-peptide complex by specific enzyme-peptide contacts in a zipper-like process leading to the adoption of an extended peptide conformation. **D|** Distinct contact networks lead to different TS-like conformations and better stabilization of TS-like conformation of ssK36 than H3K36. Figure taken from (Philipp Schnee et al., 2022).

#### 4.4. Peptides can function as competitive inhibitors for PKMTs

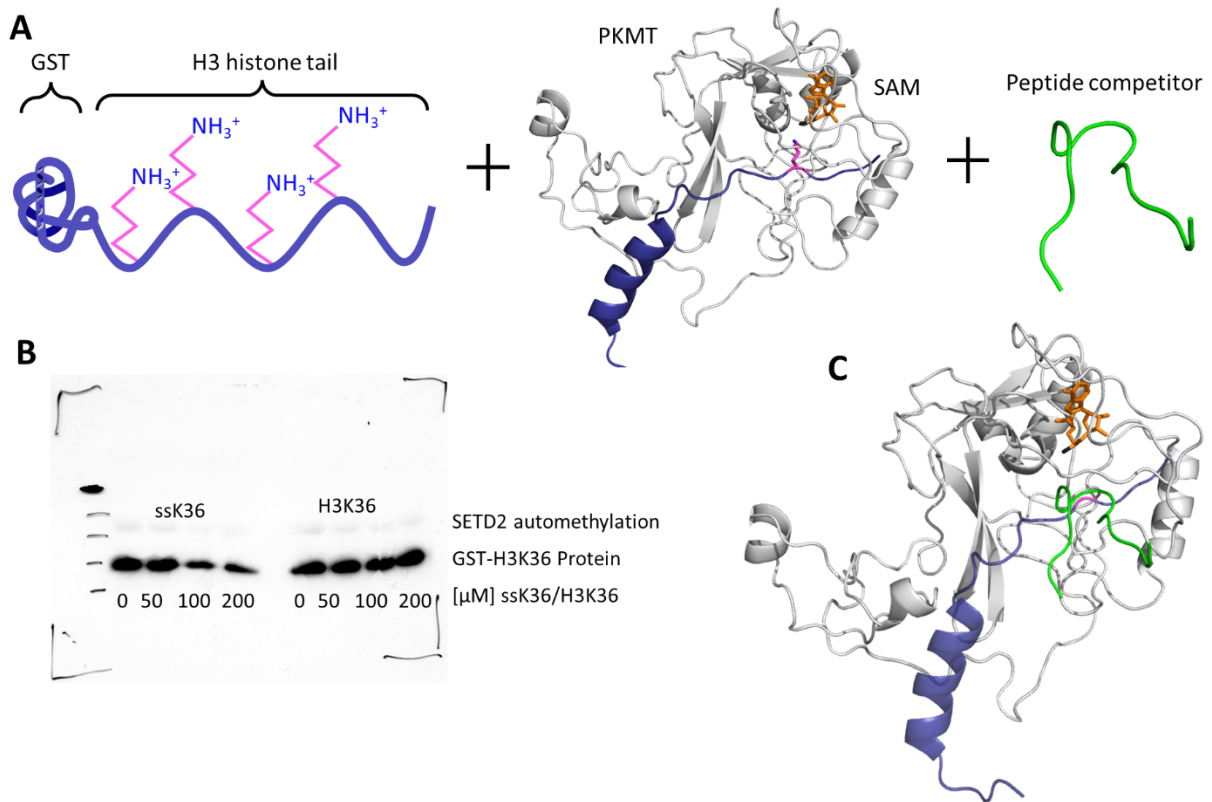
Appendix III

**Schnee P, Jeltsch A, Weirich S (2023).** Artifizielles Peptid mit PKMT-inhibitorischer Wirkung - Universität Stuttgart PCT-Patentanmeldung.

PKMTs play a key role in the dynamic regulation of cellular physiology and gene expression in all our cells. Given their key biological functions and implications in human diseases, there has been a growing interest in approaching these enzymes as potential therapeutic targets. Consequently, discovery and development of PKMT inhibitors has become a very active and fast-growing research area over the past decade. PKMT hyperactivity, overexpression, or aberrant methylation events are observed in different kinds of cancer (Hanley et al., 2023; Luo, 2015; Sato et al., 2021; Weirich et al., 2017). Developing specific inhibitors for mutated PKMTs could decrease the dysregulated methylation, potentially offering new avenues for disease treatment. A common strategy for developing PKMT inhibitors is to engage small molecules to selectively fit in SAM-binding pockets of specific PKMTs and compete with the cofactor SAM. This strategy however faces the obstacle, that SAM is used as a methyl group donor for all PKMTs but also other methyl group transferring enzymes like DNMTs. Specific, SAM-derivative inhibitors targeting only desired PKMTs have been developed (Luo, 2015), the screening effort to find such molecules however naturally increases with the number of potential off-target enzymes due to fewer specific properties in the SAM binding pocket. An alternative approach to inhibit PKMTs is to target the substrate binding pocket. Here, a strong readout was evolutionary developed to identify the correct target lysine in a large variety of histone and non-histone substrates. Naturally, specific contact networks are present and could serve as a guidance for inhibitor development. Peptides originating from PKMTs substrates could be the starting point to find substrate-competitive inhibitors.

To test this hypothesis, SETD2 was incubated with radioactively labelled SAM, Glutathione-S-Transferase (GST)-fused to a H3 peptide chain, mimicking a protein substrate, and either ssK36 or H3K36 peptides (**Fig. 27A**). The strongest methylation signal for the protein substrate was observed if no additional peptides were present (**Fig. 27B**). If the peptide concentration was increased, the substrate methylation signal dropped, indicating a competitive inhibition effect of the added peptide. Important to note is that this effect only appeared for the ssK36 peptide. The substrate methylation signal stayed constant with increasing concentration of H3K36 peptide. This indicates that the previously described properties of ssK36 are beneficial for the design of substrate competitive PKMT inhibitors (chapter 4.3) (**Fig. 27C**).





**Figure 27: The ssK36 peptide functions as a substrate-competitive inhibitor for SETD2. A|** Schematic representation of the components in the *in vitro* methylation assay containing: (i) the substrate GST-H3, where a GST protein (M1-K218) was fused to a H3 peptide chain (A1-60L) containing the target lysine 36 for SETD2; (ii) the investigated PKMT, in this case SETD2; (iii) radioactively labeled SAM; (vi) the peptide competitor. **B|** Autoradiography result of *in vitro* methylation of GST-H3 (4.5 μM) by SETD2 (9 μM) using radioactively labeled SAM (0.76 μM) and increasing concentration of peptide competitor ssK36 or H3K36. **C|** Representation of the inhibitory mechanism of the peptide competitor (green), which binds the targeted PKMT with a high affinity due to its hairpin association and favorable contacts, thereby displacing the natural substrate (faded purple).

#### 4.5. MD Simulation of NSD2 in complex with a new NSD2-specific super-substrate

Manuscript #4, appendix II

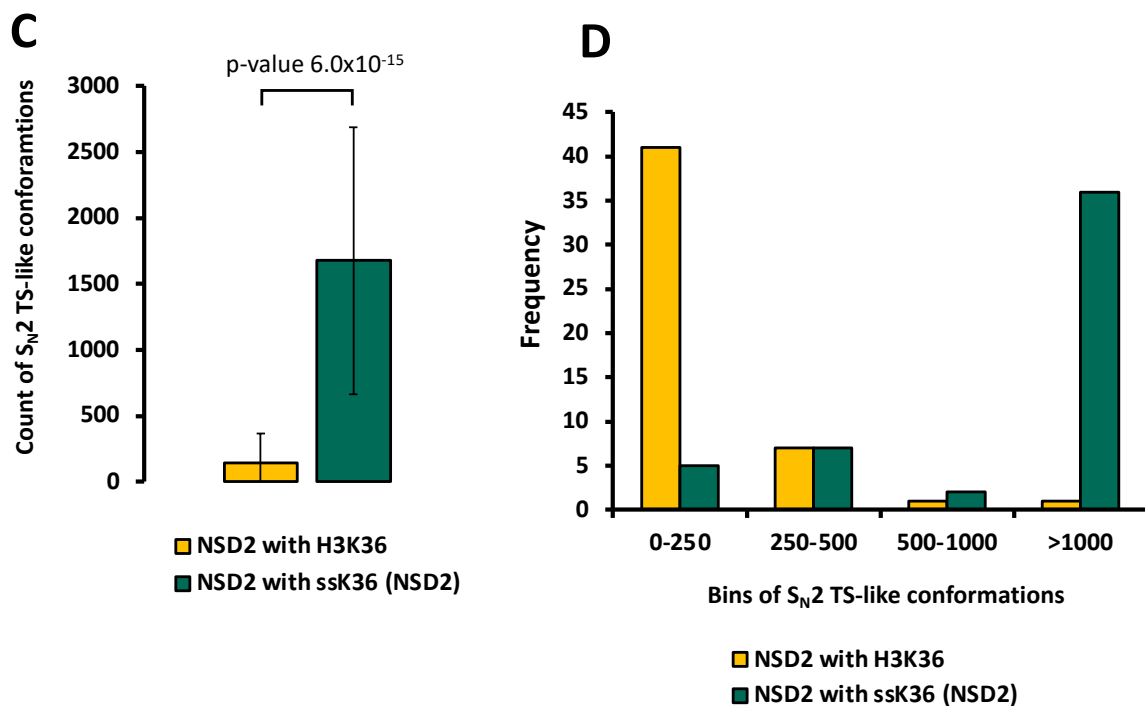
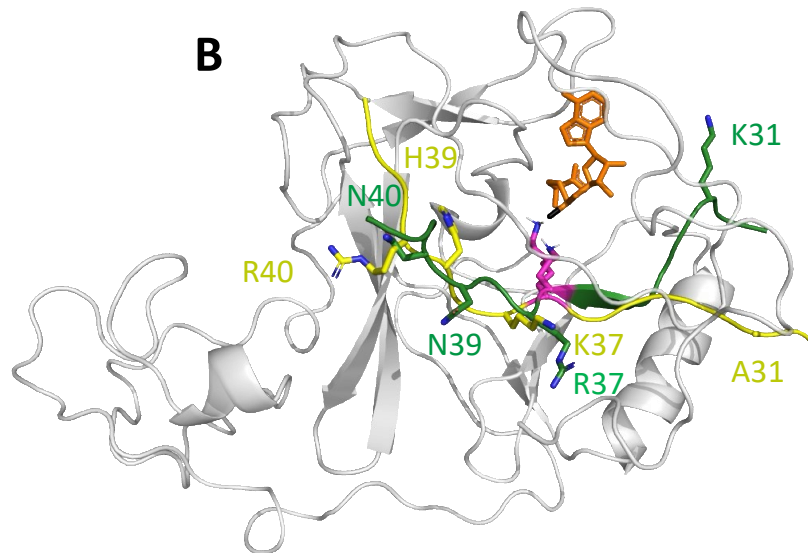
Weirich S, Kusevic D, **Schnee P**, Reiter J, Pleiss J, Jeltsch A (2023) Discovery of new NSD2 non-histone substrates and design of a super-substrate, *Communications Biology* – manuscript submitted for publication.

Peptides can function as competitive inhibitors for PKMTs. However, potential candidates need to display a higher affinity towards the targeted PKMT compared to their natural substrate. A suitable method to find peptides, which display such high affinities are methylation experiments with Celluspot peptide arrays. Using this method, a super-substrate for SETD2 was designed, which displayed a ~100-fold activity increase compared to the natural H3K36 peptide substrate (Schuhmacher et al., 2020). The molecular mechanism behind this activity increase was found using MD simulation techniques. To validate the transferability of this method, a novel super-substrate was designed for the PKMT NSD2 by Dr. Sara Weirich using the same workflow as applied for the development of the SETD2 super-substrate (Weirich et al., 2023). Subsequently, the properties of the new NSD2-specific super-substrate were again investigated by MD simulation.

Methylation experiments showed that NSD2 had a >100-fold higher activity towards the newly designed, NSD2-tailored super-substrate peptide ssK36(NSD2) compared to the canonical H3K36 peptide. ssK36(NSD2) differs at four positions from the H3 sequence: A31K, K37R, H39N, R40N (**Fig. 28A**). To investigate the molecular mechanism behind this massive increase in NSD2 activity, the H3K36 and ssK36 peptide were modelled in complex with NSD2 as described before (chapter 3.1.3). Each complex was subject to MD simulation for 50 replicates à 100 ns (**Fig. 28B**). To define criteria describing the likelihood of a methyl group transfer during the MD simulation and approximate activity, the geometric requirements for a TS-like conformation were applied as described before (chapter 1.5.6, Fig. 8). Strikingly, the complex of NSD2 and ssK36(NSD2) was able to establish significantly more TS-like conformation than the complex of NSD2 with H3K36 (**Fig. 28C**). The difference between the two peptides was even more pronounced after sorting the simulation replicates into bins. In 36 out of 50 replicates the ssK36 peptides established > 1000 TS-like conformation frames (5000 frames in total, per replicate) indicating that it adopted a very active conformation in which the TS-like state was reached frequently. This state was only achieved once in the case of H3K36 (**Fig. 28D**). Adding to this, the majority (41 replicates) of H3K36 simulations had 0-250 TS-like conformation frames, whereas only 4 of the ssK36(NSD2) simulation replicates had such a low number of TS-like conformation frames. These observations indicate that the four amino acid exchanged in ssK36(NSD2) cause a much better stabilization of TS-like conformations between target lysine and SAM.

**A**

Position:	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
H3K36:	A	P	A	T	G	G	V	K	K	P	H	R	Y	R	P
ssK36(NSD2):	A	P	K	T	G	G	V	K	R	P	N	N	Y	R	P



**Figure 28: The complex of NSD2 and ssK36(NSD2) establishes more TS-like conformations than the NSD2-H3K36 complex. A|** Sequences of the 15 amino acid long peptides H3K36 and ssK36(NSD2). Mutated residues are colored yellow in H3K36 and dark green in ssK36(NSD2). The target lysine 36 is colored pink. **B|** Overlay of representative structures from MD simulation runs of the 15 amino acid long H3K36 (yellow) and ssK36 (green) peptide, each complexed in the NSD2 SET domain (grey) binding cleft. The target lysine (pink) is deprotonated and inserted in the hydrophobic tunnel. SAM (orange) binds from the opposing site. **C|** Average number of TS-like conformations of 50 MD simulation replicates (100 ns) for each peptide complexed to NSD2. **D|** Histogram of simulation replicates, where every replicate was sorted in the indicated bins. Figure taken from (Weirich et al., 2023).

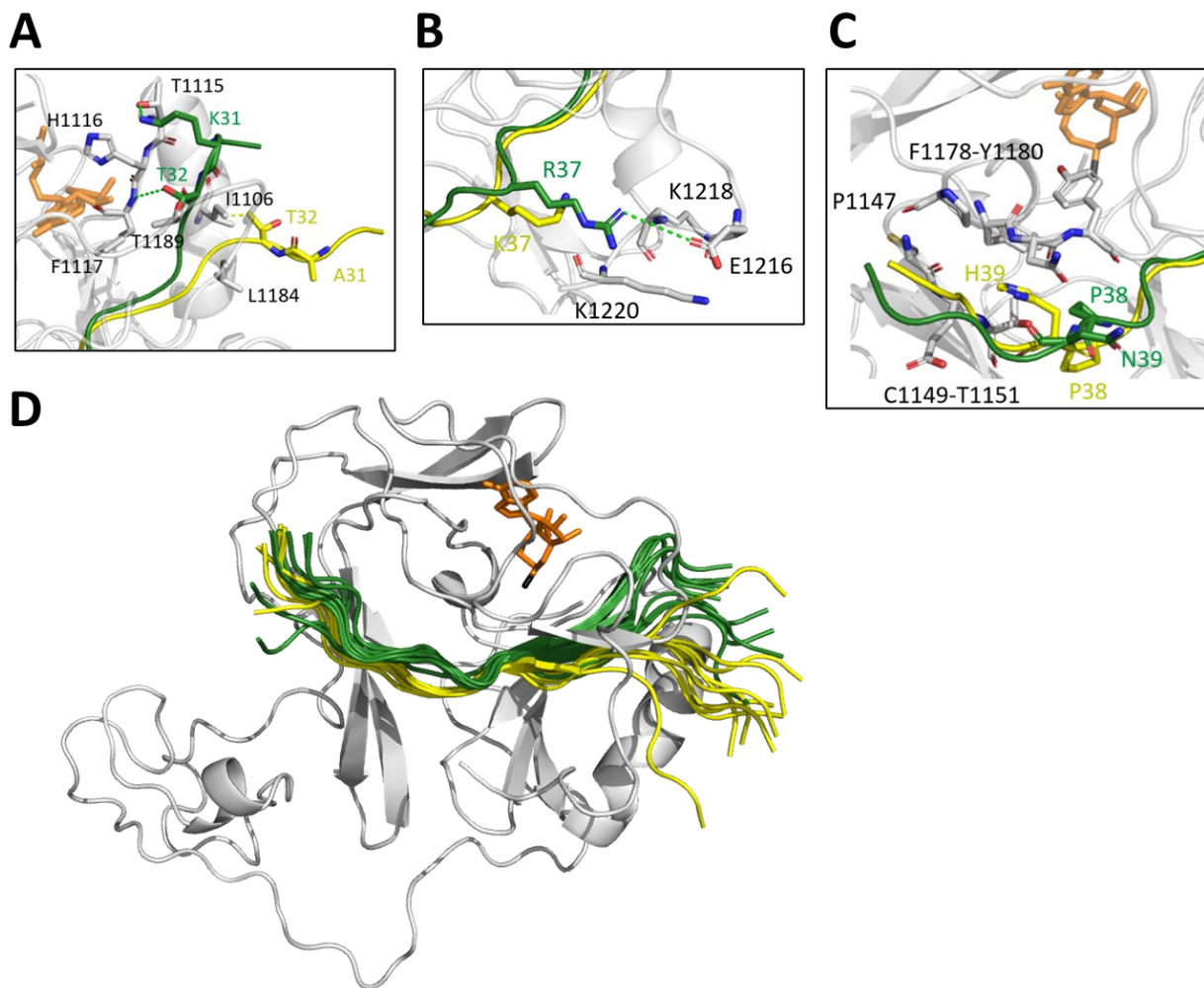
To investigate the molecular mechanism behind the better stabilization of the TS-like conformations, a contact map of the established contacts between peptide and NSD2 protein during the simulation was prepared using Contact Map Explorer as described before (chapter 4.3.6). The resulting contact maps for H3K36 and ssK36(NSD2) were then contrasted and differences extracted (**Fig. 29**).



**Figure 29: Contact profile analysis reveals different contact maps for H3K36 and ssK36 in complex with NSD2.** Contact map difference obtained after the subtraction of the H3K36 and ssK36 contact map. Shown are all NSD2 residues which exhibited noticeable changes (>5%). Yellow indicates that a specific contact was more often observed in simulations with H3K36. Green symbolizes a higher contact frequency for ssK36. Framed contacts displayed the largest difference.

The largest deviations between the two peptides were found at 3 out of the 4 mutated residues. At position 31, alanine was mutated to lysine in ssK36(NSD2). This caused peptide residues P30-T32 to strongly interact with NSD2 residues H1110-F1117 and E1187-T1189. The K31 side chain amino group was solvent exposed and interacted with the T1116 hydroxy group (**Fig. 30A**). This interaction led to an upward bending of the N-terminal part of the ssK36 peptide (A29-G33). In contrast, this part of the H3K36 peptide made contacts with I1106 and L1184-N1186, which guided the peptide to bend in the opposite direction. The second large difference between the contact maps was observed at peptide positions 37-39. The larger side chain of R37 enabled contacts primarily with the E1216 side chain and K1220, K1221 backbone atoms (**Fig. 30B**). Moreover, a higher contact frequency of ssK36(NSD2) P38 with NSD2 residues F1177-Y1179, which are next to the H39 pocket was observed. P38 in H3K36 was oriented differently, potentially influencing the structure of the backbone atoms in the peptide. Hence, K31, R37 and P38 established other contacts in ssK36, potentially stabilizing the TS-like conformation. In contrast, H39 in H3K36 was positioned in a pocket surrounded by NSD2 residues P1146-T1150, while N39 pointed outwards into to solvent leading to the loss of this interaction (**Fig. 30C**). This finding suggests that the H39 contact to NSD2 residues P1146-T1150 is unfavorable for TS-like conformations.

As a consequence of the described interactions, the mutated residues changed the conformation of the ssK36(NSD2) peptide in the NSD2 binding cleft. Especially at the N-terminus, the ssK36(NSD2) bended towards SAM, whereas H3K36 stayed straight (**Fig. 30D**). Due to this conformational change, K31 could also interact with the hydroxyl group of the SAM sugar moiety. Adding to this, the SET-I loop (D1114–Y1119), which was proposed to be contribute to the substrate specificity of PKMTs (Ronen Marmorstein, 2003; Qian & Zhou, 2006), changed its conformation towards SAM (**Fig. 30A**). Combined with K31, these interactions could stabilize SAM in the cofactor binding pocket and increase the probability of a methyl group transfer.



**Figure 30: The H3K36 and ssK36(NSD2) peptide establish different contacts with NSD2. A|** K31 in ssK36 interacts with NSD2 T1116, whereas residues T32-V35 in H3K36 interact with I1106. **B|** The longer side chain of R37 reaches out to E1216. **C|** H39 of H3K36 is positioned in a pocket made by P1146, C1148, E1149, and T1150. In contrast, N39 points into solvent which positions P38 closer towards residues F1177, N1178 and Y1179. **D|** Overlay of representative snapshots from MD simulation of H3K36 and ssK36 peptides complexed with NSD2. Throughout the simulation, both peptides change their conformation, especially ssK36 bends at the N-terminus, eventually interacting with SAM. Figure taken from (Weirich et al., 2023).

Since a super-substrate has already been designed for SETD2, it was investigated by Dr. Sara Weirich if NSD2 is specific for its own super-substrate or if NSD2 methylation is also increased for the SETD2-tailored super-substrate. Therefore, a peptide array methylation experiment (Supplementary Figure 4, manuscript#4, appendix II) was performed where the methylation of the NSD2 specific ssK36(NSD2) and ssK36(SETD2) were directly compared on one array. The methylation data clearly demonstrated that NSD2 only methylates the NSD2 super-substrate with high activity, while no methylation was observed for the SETD2 specific super-substrate.

## 5. Discussion

PKMTs and DNMTs function as key players in the regulation of genome stability, gene expression, DNA repair and cellular differentiation. Their activity is controlled by factors like substrate- and product specificity. Cancer mutations in these enzymes disturb the regulatory mechanisms, which makes their understanding a main target in modern epigenetic research. The goal of this work was to use MD simulation techniques in combination with methylation and FRET experiments to characterize somatic cancer mutations found in DNMT3A and the PKMT NSD2. Conducted MD simulations revealed yet unknown reason for the dominant effect of DNMT3A R882H in heterozygous states and explained the altered product specificity for NSD2 T1150A. Moreover, the features of the artificially designed super-substrate peptides, which caused a ~100-fold activity increase of the PKMTs SETD2 and NSD2, were precisely described by MD simulation and validated by wet-lab experiments. The found molecular mechanisms in this work explain biochemical results at an atomistic resolution and suggest novel strategies for the design of a new class of substrate-competitive PKMT inhibitors.

### 5.1. MD Simulation of the somatic cancer mutation R882H of DNMT3A

DNMT3A mutations in blood cancers show a strong enrichment of missense mutations (73%), which most times occur in heterozygous states combined with an intact WT allele (Yang et al., 2015). Among missense mutations, R882 mutations are most abundant (66%), two thirds of them are R882H (Mack et al., 2022). *In vitro* methylation of the DNMT3A R882H mutant displayed a change in the flanking sequence preferences for DNMT3A, mainly at the +1 site of the CpG site, extending up to +4. Consequently, DNMT3A R882H preferred different DNA substrates than DNMT3A WT. In addition, it was shown that the RD interface in mixed DNMT3A tetramers preferentially forms in a symmetric manner by two R882H subunits. The molecular mechanism for these observations of DNMT3A R882H were however unclear. Therefore, MD simulations were conducted comparing WT and R882H to investigate the detected pathogenic effects.

As a first step, the altered flanking sequence preference was investigated by developing contact maps of DNMT3A WT and R882H. The contact map analysis revealed strong differences for R882H in the interaction with the DNA substrate. Charged interactions between R882 and the sugar-phosphate DNA backbone at the +3/+4 flanking sites with Gua438 were found for DNMT3A WT. This contact was heavily reduced (~3-fold) in the DNMT3A R882H mutant. In contrast, H882 established more contacts in the RD interface, which connects the DNMT3A-subunits in the DNMT3A/L hetero tetramer. The reduced contact for H882 with Gua438 indicated a lower affinity towards DNA substrates with Gua at this position and explains the change in flanking sequence preference.

DNMTs were demonstrated to base their substrate specificity on the established contacts with their DNA substrate (Dukatz et al., 2020; Dukatz et al., 2022; Emperle et al., 2021). Especially arginine residues play a critical role in determining the substrate specificity. For instance, DNMT3A and DNMT3B display different flanking sequence preferences. Using different mutant cell lineages, it was shown that flanking sequence preferences do not stem from differential localization of DNMT3A and DNMT3B, but from minor variations in the structure of their catalytic domains (Mallona et al., 2021). Comparisons of the protein sequences and structures of DNMT3A (Zhang et al., 2018) or DNMT3B (Lin et al., 2020) complexed with DNA, found different residues contacting the +2 position of the methylated CpG in the major groove of the DNA. In DNMT3A, R836 was found to contact purines on the opposite strand from the methylated CpG, while at the same position K777 in DNMT3B contacts purines on the methylated strand. The proposed recognition mechanism was investigated further by MD simulations of DNMT3A complexed with multiple decameric DNA substrates containing favored and disfavored flanking sequences. In favored DNA sequences R836 plays a key role by interacting with the guanine that base-pairs with the cytosine at position +2. By introducing a lysine residue, as in DNMT3B, the flanking sequence preference of DNMT3A R836K was almost completely converted to resemble that of DNMT3B (Mallona et al., 2021). This result was later extended as R836 was shown to suppress activity at CNT sites but support methylation of CAC substrates, the preferred target for non-CpG methylation of DNMT3A in cells (Dukatz et al., 2022).

Besides the altered flanking sequence preference of DNMT3A R882H, biochemical data suggested a preference for R882H to form dimers at the RD interface in the inner position of the 3L-3A-3A-3L heterotetramers (Mack et al., 2022). Therefore, in a mixture of DNMT3A WT and R882H, DNMT3A WT is being displaced into the outer positions. This behavior was explained by MD simulation results performed in this work. Simulating the DNMT3A/L heterotetramer with and without R882H mutation in the DNMT3A subunits, displayed a higher number of contacts in the RD interface for R882H. More contacts stabilize the subunits in the RD interface leading to a higher affinity and a more tightly packing. H882 was observed to form more inter-subunit contacts with N879, but also rearranged a small network of inter-subunit contacts involving N879, I858, L859 and W860. The interaction of H882 with N879 could be a general mechanism for R882 cancer mutants, since somatic mutants R882C and R882S, which also occur in AML, have the ability to interact with the N879 side chain via polar interactions in a similar way as R882H (Jeltsch et al., 2021). Using the same mechanism, R882C and R882S could also orient N879 and cause a rearrangement of contacts involving I858, L859 and W860 in the RD interface. Moreover, R882C and R882S might therefore loose contact to Gua438, as seen for R882H. This hypothesis could be tested with additional MD simulations of R882S and R882C using the same workflow as described in this work. More contacts with residues from the opposing subunit would hint towards a preference to form dimers at the RD interface. A reduced contact to Gua438 would suggest a similar mechanism for the change in flanking sequence preference. This proposition



is supported by deep enzymology data revealing that R882C and R882S have similar flanking sequence preferences as R882H (Emperle et al., 2019).

Previously it was found that R882H stabilizes long oligomers consisting of DNMT3A subunits (Nguyen et al., 2019). This agrees with the increased affinity of DNMT3A R882H dimers in the RD interface caused by more inter-subunit contacts. Moreover, the flanking sequence preference is determined in the RD interface. This explains the observed dominant effect of this mutation on DNA methylation even in a heterozygous genetic background, since the DNMT3A WT is displaced from the RD interface by DNMT3A R882H. The dominant effect of R882H leads to aberrant methylation patterns in cells and explains the lack of methylation at major satellite repeats in affected patients (Kim et al., 2013).

A different mechanism was presented by Yang et al., who used a combination of MD simulation as well as Markov State Models to understand the catalytic process of DNMT3A mediated CpG methylation (W. Yang et al., 2023). Markov State Models can be used to discretize the MD simulation trajectory into a set of states and describe the transitions between these states as a Markov process. They found that the DNMT3A/3L heterotetramers are able to bend at 4 defined hinge positions. The bending movements at the hinge position are believed to be responsible for inserting the target cytosine into the DNMT3A active site. Two hinges are located at the RD interface, next to R882. It was speculated that R882H hinders the hinge swinging motion, thereby impeding the insertion of the target cytosine, effectively decreasing the methylation activity. This is in line with earlier studies showing that R882H had a reduced methylation activity (Emperle et al., 2018; Ley et al., 2010; O'Brien et al., 2014). A concrete mechanism explaining how R882H hampers the hinge has not been suggested. However, by connecting the results of Yang et al. with the results of this work, it is plausible, that the increase of contacts in the RD interface, observed in MD simulations of R882H in this work, lead to a stiffening of the hinge located next to R882. This could explain the decreased swinging motion observed in R882H.

A different study proposed a mechanism for the reduced activity, in which R882H mutation induced a “folded” conformation in SAM through different hydrogen bond formations at the SAM ribose moiety (Liu et al., 2019). In this mechanism, R891 functions as a mediator. In DNMT3A R882H, R891 changed its interaction network through formation of more hydrogen bonds with the methionine portion of SAM and reducing its interactions with the ribose ring. The conformational changes of SAM induced an increase in the reaction barrier of the rate determining step through forming less reactive geometries. The distal transmission from R882H to the active site was rationalized through internal cascade, in which H882 slightly compresses residues L883, R887, R891. This compression was not observed for R882. The suggested mechanism fits to the observed contact maps in this work, suggesting that R882 was preferably in contact with the DNA backbone reaching upwards. In contrast H882 caused a rearrangement of contacts heavily involving I858 and L859. I858 and L859 are located on the helix next to L883, R887 and R891, which could lead to the compression described by Liu et al.

This could be validated by measuring the conformation of SAM or the altered contacts as described by Liu et al. throughout the MD simulation conducted in this work.

In this context it is important to note that, the two proposed mechanisms might only explain a generally reduced methylation activity of DNMT3A R882H. However, the biochemical and MD simulation data in this work demonstrate and explain a change in flanking sequence preference, which is a more precise description of the pathogenic mechanism. This is supported by methylation experiments in which R882H was even more active on 13% of the used substrates than the WT and only less active on 52%. A simply reduced methylation activity falls short against this observation.

## **5.2. MD Simulation of the T1150A cancer mutant of NSD2**

H3K36 methylation is an important histone modification affecting many cellular processes (Wagner & Carpenter, 2012). H3K36me2 and H3K36me3 function as two distinct signals that have different downstream effects. The H3K36me2 mark occurs at intergenic regions and promoters, while the H3K36me3 mark is enriched at gene bodies of active genes (Cornett et al., 2019; Yun et al., 2011). H3K36me2 directly interacts with the DNA methyltransferase DNMT3A via its PWWP domain (Dukatz et al., 2019; Weinberg et al., 2019), while H3K36me3 is bound by the PWWP domain of DNMT3B (Baubec et al., 2015). NSD2 deposits up to dimethylation at H3K36 and was shown to be crucial for genome stability (Topchu et al., 2022). An example for this is the developmental disease Wolf-Hirschhorn syndrome, which is caused by a heterozygous loss of NSD2 (Bergemann et al., 2005). Unlike the clear-cut impact of gene deletions that always result in a loss of the gene function, unraveling the biological consequences of single point mutations is more challenging. Many somatic missense mutations in PKMTs have been detected in diverse cancer types, altering the enzyme's activity, substrate specificity, product specificity, or other enzymatic properties (Brohm et al., 2019; Oyer et al., 2014; Weirich et al., 2017; Weirich et al., 2015). The NSD2 missense single point mutation T1150A, was detected in leukemic patients and characterized in this work using a combination of biochemical and MD simulation approaches.

In different *in vitro* methylation experiments, Dr. Mina Saad Khella showed that NSD2 T1150A and its paralogue NSD1 T2029A are hyperactive and introduce trimethylation to H3K36. The hyperactivity of NSD2 T1150A was observed before (Li et al., 2021; Sato et al., 2021). However, the change in product specificity from di-to trimethylation was unknown and displays a completely new pathogenic mechanism of this mutation. The NSD2 T1150A mediated trimethylation was further confirmed by Dr. Mina Saad Khella in human cells highlighting its biological relevance. Despite the intensive biochemical work, the molecular mechanism of how T1150A overcomes the dimethylation barrier remained

elusive. In this work, the mechanism for the altered product specificity was revealed using MD and sMD simulation techniques combined with cutting edge analysis tools.

In the sMD experiments, the association of a SAM molecule into the SAM binding pocket of NSD2 WT or NSD2 T1150A complexed to a H3K36me2 was simulated. In the analysis it was found that NSD2 WT can barely accommodate SAM and H3K36me2 in a productive conformation at the same time. In contrast, NSD2 T1150A was able to do so, and TS-like conformation were observed frequently. This confirmed the Western blot and MALDI mass spectrometry results in which trimethylation catalyzed by NSD2 T1150A was detected. The drastic difference was not observed in sMD simulation with H3K36me1 peptides complexed to NSD2, showing that a distinct mechanism occurs only with a dimethylated peptide on the verge to trimethylation.

To investigate this, MD simulation of NSD2 WT and T1150A complexed with H3K36me2 and SAM were conducted. Using POVME3 (Wagner et al., 2017), the volume of the active site was calculated. The analysis showed that large volumes occurred more often in T1150A compared to the WT. The reason for this was found in a subsequent contact analysis. In NSD2 WT, T1150 engaged in contacts with Y1092 and L1120, which oriented these residues effectively reducing the volume of the active site. The T1150 side chain hydroxyl group interacted with the Y1092 backbone nitrogen. The side chain methyl group hydrophobically interacted with the L1120 side chain. These two contacts were lost in NSD2 T1150A because of the absence of the side chain hydroxyl and methyl group. The orientation of Y1092 and L1120 was more flexible and the active site volume increased. The importance of Y1092 for the structural integrity of the active site was also highlighted by the finding that the Y1971C mutation in NSD1 (NSD1 Y1971 corresponds to NSD2 Y1092) was demonstrated to abolish the enzyme activity (Khella et al., 2023a).

The active site volume has been identified before as a key factor in regulating the product specificity of PKMTs and restricting further methylation. In multiple sequence alignments it was observed, that PKMTs possessing a tyrosine at the so-called F/Y-switch position are limited to catalyze mono- or dimethylation. PKMTs possessing a phenylalanine or another hydrophobic residue in this position display di- or trimethyltransferase activity (Collins et al., 2005). Mutating the F/Y-switch residue to the respective other amino acid led to predictable changes in product specificity (Collins et al., 2005; Cortopassi et al., 2016; DiFiore et al., 2020; Qian & Zhou, 2006; Triebel et al., 2003; Wu et al., 2010; Zhang et al., 2003). The trimethyltransferase DIM-5 was converted into a dimethylase by mutating F281 into tyrosine (Zhang et al., 2003). The monomethyltransferases SET7/9 and SET8 were changed to dimethyltransferases through the corresponding Y-to-F mutation (Y305F for SET7/9 and Y334F for SET8) (Del Rizzo et al., 2010; Zhang et al., 2003). The F/Y-switch is not fixed to a defined position and more than one residue can take this function. The dimethyltransferase G9a was turned into a trimethyltransferase by Y1067F (Wu et al., 2010), but into a monomethyltransferase by F1152Y

indicating that two susceptible positions exist (Collins et al., 2005). The difference between tyrosine and phenylalanine is the side chain hydroxyl group. The hydroxyl group influences the volume in the active site in two ways and thereby the product specificity. Firstly, through the simple addition of the hydroxyl group. Secondly, because of the interactions of the hydroxyl group with other tyrosines (e.g., Y305, Y335 and Y245 in SET7/9) and water molecules in the active site (Hu & Zhang, 2006). These polar interactions build a network pulling the active site residues closer to each other (Schnee et al., 2023). In contrast, phenylalanine is not able to interact via polar contacts. The described mechanism for the F/Y-switch is in very good agreement to the results for NSD2 T1150A obtained in this study.

Moreover, NSD2 L1120 is the only amino acid residue near the target lysine side chain which is not conserved between the dimethyltransferase NSD2 and the trimethyltransferase SETD2. SETD2 contains a less bulky methionine at this position highlighting that L1120 might have an important role in controlling the product specificity of NSD2. Likewise, L1120M was shown to enhance the trimethylation activity of NSD2 (Sato et al., 2021), strengthening the important role of L1120 and its interaction with T1150 in the control of the product specificity of NSD2.

The consequence of the reduced active site volume is a failure to accommodate SAM and the dimethylated target lysine in the active site. This prevents the formation of catalytically competent TS-like conformations (Schnee et al., 2023). Product specificity was demonstrated to depend on the  $S_N2$  TS-like criteria, as in MD and QM/MM simulations of the monomethyltransferase SET7/9 complexed with the K4 or K4me1 peptide, the distance between the SAM sulfur atom and the lysine N $\epsilon$  was higher for K4me1 (6.1 Å) than for K4 (5.7 Å) (Zhang & Bruice, 2007b). After the first methylation, a reorientation of Kme1 was not possible due to steric constraints. A productive state in which monomethylated lysine N $\epsilon$  and SAM methyl group are close together did not occur.

The presented results precisely explain the molecular mechanism behind the altered product specificity observed for NSD2 T1150A. The hyperactivity of NSD2 T1150A was investigated before by Sato and coworkers, who simulated the binary NSD2 T1150A-SAM complex. Their MD simulation results showed that the autoinhibitory loop (AL) adopts more open conformations in NSD2 T1150A compared to WT. Hence, T1150A was suggested to destabilize the interactions that keep the AL closed. With an open AL, the probability for a substrate to successfully bind in the active site increases. It needs to be noted that the starting conformation of the MD simulations conducted by Sato et al. had the AL in a conformation in which the placeholder residue C1183 was already turned outwards (Schnee et al., 2023). As described earlier, prior steps are necessary to turn the placeholder residue in this position. These type of conformational changes of the AL were not observed in the MD simulation conducted in my work, because the ternary complex of NSD2-peptide and SAM was simulated. The increased probability of an open AL can however not explain the trimethylation activity of T1150A.

It is important to emphasize that T1150A does not occur in cancer alone, but in combination with the missense mutation E1099K (Bea et al., 2013). This mutation displayed hyperactivity in the same range as T1150A in *in vitro* methylation experiments (Li et al., 2021; Sato et al., 2021). Li and colleagues solved the cryo-EM structure of the NSD2 E1099K/T1150A mutant bound to a nucleosome (PDB 7CRO). In this structure, the linker DNA was partially unwrapped to accommodate NSD2 (**Fig. 12**) (Li et al., 2021). Furthermore, the positively charged lysine residue in E1099K was noticed to interact with the negatively charged sugar-phosphate backbone of the nucleosomal DNA. The authors proposed this to be a reason for the increased activity. This model was supported by the finding that NSD2 exhibits weak and non-specific lysine methylation activity on histone protein octamers as substrates, but specific and strong methylation was measured when nucleosomal substrates were used (Li et al., 2009). However, kinetic analysis revealed that NSD2 E1099K and T1150A have no increased nucleosome affinity compared to the WT (Li et al., 2021; Sato et al., 2021). Through MD simulations, Sato and coworkers found, that E1099K destabilizes the AL as observed for T1150A. Thus, a combination of the proposed mechanisms by Sato and Li for the hyperactivity of NSD2 E1099K/T1150A could potentially fully explain the pathogenic behavior. The destabilization of the AL, caused by E1099K and T1150A, need prior structural rearrangements of the AL by e.g. conformational changes of the Post-SET loop or the placeholder residue. These could be induced upon interactions with nucleosomes. The necessary conformational changes upon nucleosome binding are enhanced in NSD2 E1099K due to the improved E1099K-DNA contact. This is supported by the absence of the Post-SET loop in structures of PKMTs bound nucleosomes, indicating a high flexibility and an open state (Schnee et al., 2023). Using this explanation, both approaches can be reconciled. However, the destabilized AL and the increased interaction with the nucleosomal DNA can only account for the hyperactivity of NSD2 T1150A but cannot explain the change in product specificity described in this work.

The proposed mechanisms by Sato, Li and this work could operate simultaneously, as they are not mutually exclusive. Together, they might fully explain the pathogenic mechanism of the NSD2 T1150A cancer mutation and its biological behavior. The observed hyperactivity of T1150A can be explained by the destabilizing effect of T1150A on the AL, which needs prior interaction with a substrate, enhanced by NSD2 E1099K. The change in product specificity can only be explained by the increase in active site volume as observed in this work. The identification of the mechanisms using biochemical experiments and MD simulation showcases the strengths of a combined approach especially in the context of complex cancerogenic missense mutations.

### **5.3. Mechanistic basis of the SETD2 super-substrate peptide**

Many PKMTs methylate more than one substrate, which raises the question how PKMTs recognize different substrates. To distinguish different potential methylation substrates, PKMTs form contacts

to the amino acid residues surrounding the target lysine. This results in a specific readout of the substrate's amino acid sequence (Del Rizzo & Trievel, 2014). To systematically investigate the substrate specificity of PKMTs Celluspot peptide arrays can be used as described earlier (chapter 1.5.7). The resulting specificity profile of SETD2 revealed that the natural H3K36 substrate sequence is not ideal at many positions and other amino acids were preferred (Schuhmacher et al., 2020). Mutating the disfavored residues in the H3 sequence to preferred ones led to the design of the 15 amino acid long super-substrate peptide (ssK36), which was methylated ~100-fold faster than the canonical H3K36 peptide. The solved crystal structures of SETD2 complexed with the H3K36M or ssK36M peptides represented important steps towards the understanding of SETD2's substrate specificity by illustrating how H3K36M and ssK36M peptides bind into the active site of SETD2 and which contacts are established between SETD2 and the peptides in each complex (Schuhmacher et al., 2020; Yang et al., 2016; Zhang et al., 2017). However, despite the discovery of peptide-specific contacts, these structures were very similar overall, and they could not fully explain the 100-fold increase in the methylation rate of the ssK36. One reason for this is that static crystal structures only capture the conformation of the ground state, and it is unknown to which extent this represents the catalytically competent conformations able to reach the TS of the enzymatic reaction. Moreover, complex structures do not describe the pathway of the peptide entry into the active site, nor the different structures of the peptides in the solvent, which may affect the association reaction. Hence, critical questions regarding the substrate-specific activity of SETD2 remained unsolved. This created the urge to compare H3K36 and ssK36 using MD simulation techniques and reveal, which factors influence the methylation activity of SETD2. Determining these factors helps to understand how PKMTs interact and discriminate between substrates avoiding off-target methylation in cells. Moreover, the factors determining the interaction of PKMTs with super-substrates may function as a starting point for a rational inhibitor design.

In this work, the mechanisms of the sequence specific methylation of SETD2 were explored by a combination of various MD simulation techniques and biochemical experiments. The combined approach made use of the predictive power of simulations and the validation by biochemical FRET and methylation experiments, highlighting a powerful interplay. Together, multiple steps in catalysis were investigated at which differences between the H3K36 and ssK36 peptide could arise: (i) the conformations of the peptide in solution; (ii) the pathway of the peptide into the SETD2 binding cleft and the conformational changes accompanying it; (iii) the enzyme-peptide complex and the establishment of TS-like conformations priming the methyl group transfer.

### **5.3.1. Hairpins in histone tails might facilitate the binding towards PKMTs**

MD simulations of the H3K36 and ssK36 peptides in solution with subsequent backbone-based clustering showed that the ssK36 peptide preferably adopted a hairpin-like conformation. The H3K36

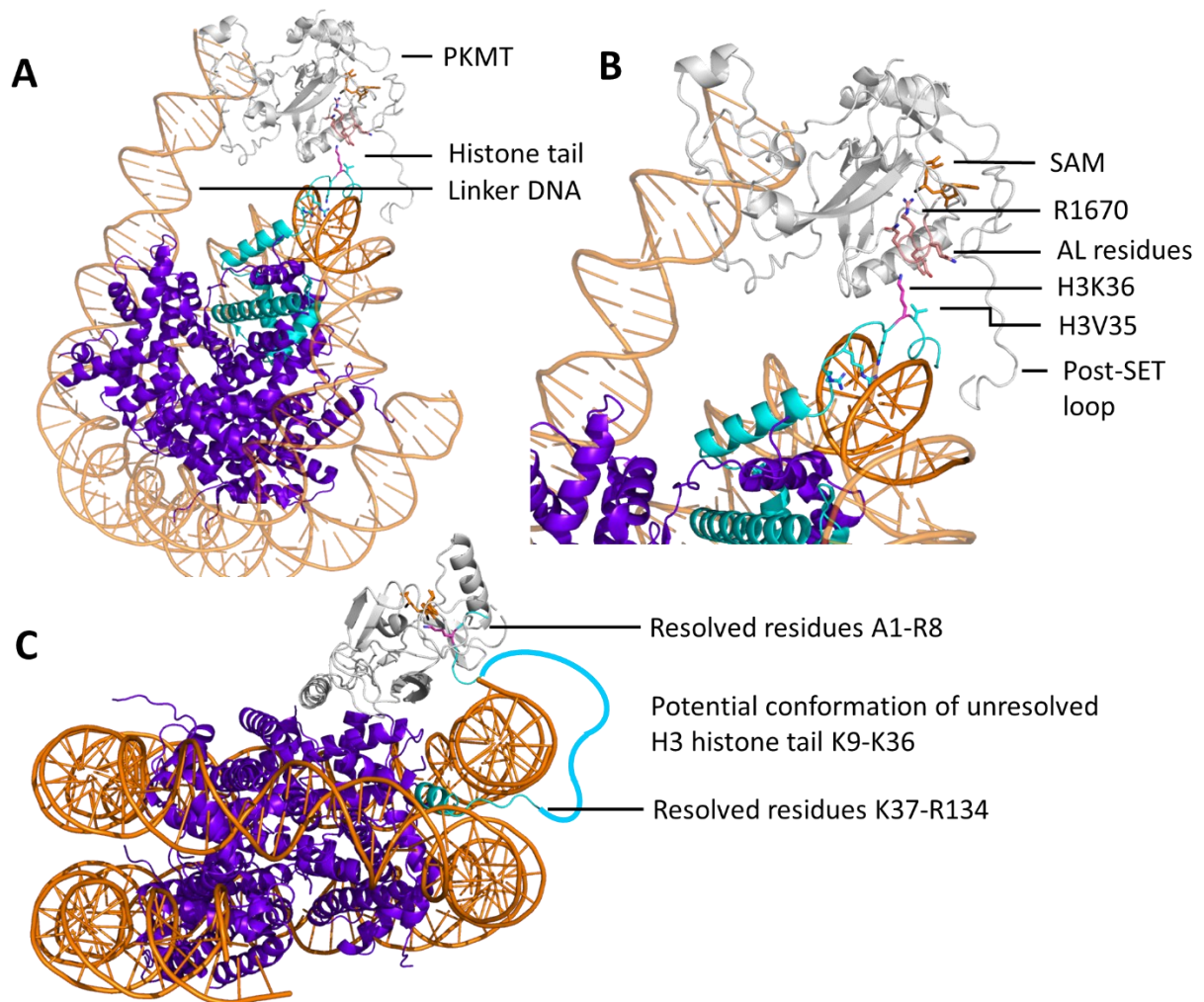
peptide preferred an extended conformation. It is important to note that despite these preferences, both peptides adopted extended and hairpin conformation throughout the simulation time and did not stay exclusively in their preferred shape. The differing conformational preferences were a direct result of the four introduced mutations in ssK36. Multiple intra-peptide contacts were established by the mutated residues stabilizing a hairpin shape. The hairpin conformation of the ssK36 peptide in solution was confirmed by FRET experiments. Remarkably, the hairpin conformation presents K36 and the adjacent V35 in the loop region facing outside, suggesting that these residues could act as first contact points during the docking of the peptide into the SETD2 binding cleft.

Thereby, contacts between the target lysine, surrounding amino acids and the PKMT are established first. This could resemble a probing process, since PKMTs need to identify distinct lysine residues along histone tails or other proteins. Exposed side chains of the target lysine and one or more characteristic surrounding residue(s) could support this probing process. This presumption is in agreement with multiple specificity profiles of PKMTs, showing that the most specific sequence readout occurs at the positions close to the target lysine (Schuhmacher et al., 2020; Schuhmacher et al., 2015; Weirich et al., 2016; Weirich et al., 2023; Weirich et al., 2020). An example for this is the hydrophobic V35 position, which was shown to be obligatory for substrate recognition by SETD2 (Schuhmacher et al., 2020). The key role of an outwards facing lysine was further supported by inhibition experiments, in which the inhibition constant of ssK36M towards SETD2 was only moderately (3.5-fold) improved compared to H3K36M, while the stimulation of ssK36 methylation was much stronger (Schuhmacher et al., 2020).

Besides the identification of the correct methylation target, the positively charged, outwards facing lysine side chain could further interact with solvent exposed residues of the AL. As mentioned earlier, the AL needs to undergo conformational changes to bind a substrate (chapter 1.5.3). The target lysine in the hairpin structure might therefore not only function as a probe, but may also be the start of the conformational changes of the AL, which are needed to overcome autoinhibition (Schnee et al., 2023). Structural data (PDB 3HNA) showed that residues R1214 and D1217 of the PKMT GLP (G9a-like protein) are located at the AL, solvent exposed and play a role in substrate recognition (Wu et al., 2010; Zheng et al., 2012). In SETD2 these AL residues could resemble Q1667, Q1669 and Y1671, which might interact with the target lysine and surrounding residues, lifting the AL up in an open position. Alanine mutants Q1669A and Y1671A lead to a disruption of the SETD2 methylation activity, emphasizing the importance of the side chains, supporting a possible interaction with the substrate (Zheng et al., 2012). R1670 is the placeholder residue in SETD2 and is also part of the AL. The complex role of R1670 was highlighted by a strong reduction of SETD2 methylation activity by the R1670G mutation (Zheng et al., 2012). A simple destabilization of the AL by R1670G should lead to a higher catalytic activity, which was not observed, highlighting potential interactions needed to facilitate catalysis.

Important to note is that MD simulation results suggesting this model are based on peptides as substrates. For larger substrate structures like proteins or nucleosomes, additional factors might come into play. Considering nucleosomes as substrates, a working model for the association between PKMT and histone tails could be envisioned. Cryo-EM structure of PKMTs, including SETD2 and NSD2, demonstrated that PKMTs bind nucleosome substrates by squeezing in between the nucleosome histone octamer particle and the linker DNA, thereby partially unwrapping it (**Fig. 12**) (Li et al., 2021; Liu et al., 2021; Sato et al., 2021). Thereby, the PKMT is positioned close to the location, where the H3 histone tail exits the octamer structure (**Fig. 31A**). The histone tail could interact with the nucleosomal DNA, dynamically wrapping around it. This bends the histone tail, creating transient hairpin conformations, exposing the target lysine and nearby residues, which are then recognized by the PKMTs (**Fig. 31B**) (Schnee et al., 2023). This model is in line with cryo-EM structures of MLL1, MLL3 or SetD2 in complex with nucleosomal substrates (PDB 6KIX, 6KIU, 6UH5) (Hsu et al., 2019; Xue et al., 2019). In these structures, the target lysine and amino acids around the target lysine are resolved (A1-R8 for PDB 6UH5) and positioned correctly in the SET domain of the respective PKMT. However, the H3 tail is not completely resolved, but misses the connecting residues K9-K36. Connecting the resolved parts, creates a flexible histone tail, wrapped around the nucleosomal DNA (**Fig. 31C**). This model is supported by a large number of positively charged residues in the H3 sequence, which could interact with the negatively charged DNA phosphate sugar backbone (Schnee et al., 2023).





**Figure 31: Possible mechanism for the binding of a PKMT towards a nucleosome and the recognition of the target lysine.** **A|** The PKMT (white) binds the nucleosome by partially unwrapping the linker DNA (orange), positioning the binding cleft towards the exit point of the histone tail (cyan, modified from PDB 7EA5). The histone tail interacts with the DNA thereby forming a hairpin-like conformation with the target lysine facing outside. **B|** First contacts between H3K36 (pink) and H3V35 and the SETD2 AL (modified from PDB 6VDB) lead to conformational changes in the AL and R1670 (rose) opening the SETD2 binding cleft. Figure taken and modified from (Schnee et al., 2023). **C|** Binding of the yeast MLL1 homologue Set1 (white) to a nucleosome (PDB 6UH5). Histone tail residues H3A1-H3R8 (cyan) are resolved and the target lysine K4 (pink) is correctly bound in the SET domain. The H3 tail reaches out from the nucleosome core body and is not resolved. Potentially, it wraps around the nucleosomal DNA to establish a stable binding in the Set1 SET domain.

### 5.3.2. Peptide hairpin conformations unfold upon binding to PKMTs and establish distinct contacts

The sMD data in this work indicated that hairpin conformations facilitate the entry in the SETD2 active site when compared with extended peptide conformations. This suggests that a transient hairpin conformation is needed during the initial docking of the substrate peptide. However, crystal structures of SETD2 complexed with either H3K36M or ssK36M show both peptides bound in an extended conformation in the binding cleft. In the sMD simulations, it was observed that the N- to C-terminal distance of the peptides increased during the peptide association. This implies a gradual unfolding of

the hairpin conformation in a zipper-like process during the binding. Residues in the middle of the peptide could position themselves at the start of the association with high flexibility and rotational freedom. Ultimately, outer residues bind last and have thus limited freedom to position themselves, decreasing their impact on substrate specificity. This was supported by specificity profiles of PKMTs showing a loose readout for residues towards the N- and C-terminus. The peptide unfolding was accompanied with a decrease in RMSD between the sMD conformations and the crystal structures. Eventually, the peptide conformations in the sMD simulations displayed a good overlay with the crystal structures. The simulation findings were confirmed by FRET experiments, in which the addition of SETD2 to each peptide led to an increase in fluorescence, indicating a separation of quencher and fluorophore.

The unfolding process of the hairpin peptide was found to play a role in catalysis. Preventing the unfolding process by chemically fixing the hairpin conformation for ssK36 using disulfide bonds, reduced the methylation by SETD2. This observation could imply that the contacts between peptide and enzyme at the N- and C-terminus of the peptide lead to a tighter binding and a more stable peptide conformation. This could be critical to reduce fluctuations and stabilize the target lysine in the middle of the peptide for the methyl group transfer. Without the stabilization, the fluctuations could destroy long lasting TS like conformation, which were shown in this work to be a critical for catalysis.

During the sMD simulation, it was discovered that the unfolding process was accompanied by a closing movement of the Post-SET loop. The Post-SET loop is highly flexible and not resolved in PKMT structures without bound substrate, exposing the binding cleft and AL. Upon substrate binding, the Post-SET loop closed behind the substrate further strengthening the binding (Yang et al., 2016; Zheng et al., 2012). These results agree with proposed open-close models of the Post-SET loop. These models were however purely based on static crystal structures. Not covered in this work is the release of the methylated product, which is suggested to involve an uplifting of the Post-SET loop (Yang et al., 2016).

In the last step of the enzymatic turnover cycle, the peptide is complexed in the PKMT binding cleft in an extended conformation. Here, contacts with the PKMT led to a stronger or weaker stabilization of S<sub>N</sub>2 TS-like conformations, depending on the peptide sequence. Results from MD simulations in this work of SETD2-peptide complexes revealed specific TS-like conformation for ssK36 that were stabilized more efficiently in SETD2-peptide complexes than the TS-like conformation of H3K36. The four introduced mutations formed unique contacts with the enzyme exclusively accessible for ssK36. Because of these contacts, the N- and C-terminal parts of ssK36 moved closer to SETD2 when compared with H3K36. However, the central part of the peptide showed the opposite trend and slightly detaches from SETD2, potentially giving more flexibility for K36 to reorient.

Interestingly, the peptides reached the TS-like conformation through different contact networks. This indicates that methylation of different peptide sequences by one PKMT may involve distinct, substrate-specific contact networks, adding another layer of substrate discrimination. This was further supported by the previously designed, NSD2-specific super-substrate ssK36(NSD2). The contacts between enzyme and peptide again differed heavily for H3K36 and ssK36(NSD2).

Based on the presented data in this work, the roughly 100-fold increase in SETD2 methylation rate for ssK36 can be attributed to two distinct effects, the accelerated association due to preferred hairpin formation and the better stabilization of TS-like conformations. The finding that kinetic barriers during association and formation of TS-like conformations are lowered with ssK36 is in agreement with previous data showing that the  $K_M$ -value of ssK36 was only marginally improved (1.4-fold), leaving most of the rate enhancement for  $k_{cat}$  (about 70-fold) (Schuhmacher et al., 2020). The overall rate enhancement of ssK36 is in a very good agreement with the results of the MD simulations presented in this work. The 4.8-fold increase in the stable hairpin formation of ssK36 combined with the 30-fold increase in TS-like conformations with long lifetimes could explain an up to 140-fold increase in reaction rate. Currently, it is not possible to determine the minimal lifetimes of hairpin conformations and TS-like structures that are needed for productive association and catalysis. Hence, the 140-fold change represents an upper limit of the potential effect, because the differences in both properties were smaller if lifetimes were not considered.

For NSD2, a >100-fold higher methylation activity of the NSD2-tailored super-substrate was observed. The MD simulation data showed that hyperactive conformations of NSD2 were observed 36-fold more often for ssK36(NSD2) than for H3K36. This is in rough agreement with the methylation data. It is important to note that whereas 36 simulation replicates with >1000 TS were observed for ssK36, only one was observed for H3K36. Increasing the number of replicates might increase the discrepancy between the two peptides, which could align biochemical with simulation results.

#### **5.4. Super-substrate peptides function as PKMT inhibitors**

PKMTs were shown to be hyperactive and deliver methylation at aberrant loci in different kinds of cancer (Hanley et al., 2023; Luo, 2015; Sato et al., 2021; Weirich et al., 2017). Developing specific inhibitors for mutated PKMTs could decrease the dysregulated methylation. Common strategies for the development of PKMT inhibitors focus on small molecules blocking the SAM-binding pockets to compete with the cofactor SAM (Feoli et al., 2022). Since SAM is used by variety of PKMTs and DNMTs, the risk of potential off-targets is high, meaning that other unrelated methyltransferases may also be inhibited. In contrast, substrate-competitive inhibitors could be tailored towards a binding cleft, which has been evolutionary optimized to heavily discriminate between potential substrates. This creates an

advantage for the design of substrate-competitive PKMT-specific inhibitors. In this work, it was tested if PKMT-specific super-substrate peptides could function as starting points for the development of PKMT inhibitors. In *in vitro* methylation assays it was found that the addition of ssK36 to a mixture of SETD2 and a protein substrate caused a decrease in substrate methylation. This effect was not observed after the addition of equal concentrations of the H3K36 peptide. This implies that the inhibitory strength is related to the peptide sequence and its properties. The transferability of this strategy was shown by the design of new NSD2-specific super-substrate peptide (ssK36(NSD2)). The sequence of ssK36(NSD2) differs from the SETD2 super-substrate (ssK36(SETD2)). Remarkably, the super-substrates were highly specific for their respective PKMT and did not display a high activity for the other (Weirich et al., 2023). This underlines the high potential of substrate-competitive inhibitors to act in an enzyme specific manner.

The results presented in this work can however only account for a qualitative proof-of concept since the reduction in PKMT activity was not comparable to industry relevant PKMT inhibitors. The inhibitor concentration of ssK36(SETD2) to achieve a 50% activity reduction ( $IC_{50}$ ) for SETD2 was roughly estimated to be 100  $\mu$ M using this assay. SAM competitive inhibitors displayed an  $IC_{50}$  value of 0.8  $\mu$ M (Feoli et al., 2022). Important to note is that the assay used in this work was just a starting point. In the presented set up, the peptide inhibitor was methylated by the investigated PKMT (Supplementary Fig. 1, appendix III). Due to the methylation, the peptide inhibitor dissociates from the PKMT and is unable to bind again, leading to a constant decrease of effective inhibitor concentration over time. Replacing the lysine with a methionine, which is not methylated and a known substrate competitive inhibitor (Fang et al., 2016), causes a drop in binding affinity (Schuhmacher et al., 2020). This is due to the different biochemical properties of lysine and methionine. As described earlier, the positive charge of lysine stabilizes the hairpin formation and interacts with SETD2 to overcome autoinhibition. One possibility to circumvent this problem could be to use non-natural amino acids like ornithine. Ornithine is one carbon shorter than lysine and could therefore still account for the biochemical properties found for the ssK36 peptide. Mass spectrometry experiments with multiple PKMTs showed that peptides containing ornithine at the target position were not methylated, since the ornithine  $N\epsilon$  and SAM methyl group were too far apart (Al Temimi et al., 2017). It is possible that a super-substrate peptide with an ornithine at the target lysine position leads to heavily increased inhibitory strength since a permanent binding is expected as seen for K36M peptides.

Although incorporation of ornithine at the target lysine position might decrease the  $IC_{50}$ , peptides as inhibitors in a clinical context do still have a variety of drawbacks: bioavailability, degradation through proteases, triggering immune responses and the costs of production to name a few. Some of these problems can be tackled by using exosome delivery systems (Hade et al., 2023; Kim et al., 2021; Xu et al., 2021). Still, small molecules display many advantages compared to peptides in terms of cell

permeability, oral administration, stability, established screening pipelines, synthesis and large-scale production. A promising approach could be to transfer the chemical features of ssK36 to small molecules. A first step towards this approach was shown in this work as the hairpin formation feature was transferred to a different peptide sequence, namely H3K36. Methylation experiments with a modified H3K36 sequence, having two cysteine residues and the N-and C-terminus and forcing the peptide in a hairpin conformation increased its capability to be methylation by SETD2 by ~2-fold. In earlier studies, disulfide bridges were used to enforce hairpin conformation in other peptides as well (Dhayalan et al., 2011). The SET domain-containing PKMT SET7/9 displayed an increased methylation activity towards an artificially designed peptide with a disulfide-stabilized hairpin structure compared to its canonical substrate.

If small molecules are considered, it could be beneficial to not chemically fix a hairpin during the design, but establish a charged interaction which only transiently support hairpin formation and allow for unfolding of the small molecule upon PKMT binding, as was shown in this work to be critical. Moreover, the generated contact profiles of super-substrates peptides in this work could be the basis for the design of functional groups in the small molecule as the critical interactions required for PKMT binding were identified. Despite the rational design approach using the found properties of super-substrates, screening of numerous small molecules and combinations of the properties is indispensable. The super-substrate properties can however heavily reduce the screening effort and function as a starting template providing key chemical properties from which combinatorial chemistry strategies could start.

## 6. References

- Al Temimi, A., Reddy, Y. V., White, P. B., Guo, H., Qian, P., & Mecinovic, J. (2017). Lysine Possesses the Optimal Chain Length for Histone Lysine Methyltransferase Catalysis. *Sci Rep*, *7*(1), 16148. <https://doi.org/10.1038/s41598-017-16128-4>
- Al Temimi, A. H. K., Teeuwen, R. S., Tran, V., Altunc, A. J., Lenstra, D. C., Ren, W., Qian, P., Guo, H., & Mecinovic, J. (2019). Importance of the main chain of lysine for histone lysine methyltransferase catalysis. *Org Biomol Chem*, *17*(23), 5693-5697. <https://doi.org/10.1039/c9ob01038f>
- Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat Rev Genet*, *17*(8), 487-500. <https://doi.org/10.1038/nrg.2016.59>
- An, S., Yeo, K. J., Jeon, Y. H., & Song, J. J. (2011). Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism. *J Biol Chem*, *286*(10), 8369-8374. <https://doi.org/10.1074/jbc.M110.203380>
- Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res*, *21*(3), 381-395. <https://doi.org/10.1038/cr.2011.22>
- Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C., & Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem*, *280*(18), 17732-17736. <https://doi.org/10.1074/jbc.M500796200>
- Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., Akalin, A., & Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*, *520*(7546), 243-247. <https://doi.org/10.1038/nature14176>
- Bea, S., Valdes-Mas, R., Navarro, A., Salaverria, I., Martin-Garcia, D., Jares, P., Gine, E., Pinyol, M., Royo, C., Nadeu, F., Conde, L., Juan, M., Clot, G., Vizan, P., Di Croce, L., Puente, D. A., Lopez-Guerra, M., Moros, A., Roue, G., . . . Campo, E. (2013). Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. *Proc Natl Acad Sci U S A*, *110*(45), 18250-18255. <https://doi.org/10.1073/pnas.1314608110>
- Bennett, R. L., Swaroop, A., Troche, C., & Licht, J. D. (2017). The Role of Nuclear Receptor-Binding SET Domain Family Histone Lysine Methyltransferases in Cancer. *Cold Spring Harb Perspect Med*, *7*(6). <https://doi.org/10.1101/cshperspect.a026708>
- Bergemann, A. D., Cole, F., & Hirschhorn, K. (2005). The etiology of Wolf-Hirschhorn syndrome. *Trends Genet*, *21*(3), 188-195. <https://doi.org/10.1016/j.tig.2005.01.008>
- Bestor, T., Laudano, A., Mattaliano, R., & Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of Molecular Biology*, *203*(4), 971-983. [https://doi.org/https://doi.org/10.1016/0022-2836\(88\)90122-2](https://doi.org/https://doi.org/10.1016/0022-2836(88)90122-2)
- Bock, I., Kudithipudi, S., Tamas, R., Kungulovski, G., Dhayalan, A., & Jeltsch, A. (2011). Application of Cellusspots peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails. *BMC Biochemistry*, *12*(48). <https://doi.org/10.1186/1471-2091-12-48>
- Borkin, D., He, S., Miao, H., Kempinska, K., Pollock, J., Chase, J., Purohit, T., Malik, B., Zhao, T., Wang, J., Wen, B., Zong, H., Jones, M., Danet-Desnoyers, G., Guzman, M. L., Talpaz, M., Bixby, D. L.,

- Sun, D., Hess, J. L., . . . Grembecka, J. (2015). Pharmacologic inhibition of the Menin-MLL interaction blocks progression of MLL leukemia in vivo. *Cancer Cell*, 27(4), 589-602. <https://doi.org/10.1016/j.ccell.2015.02.016>
- Bourc'his, D., Xu, G.-L., Lin, C.-S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the Establishment of Maternal Genomic Imprints. *Science*, 294(5551), 2536-2539. <https://doi.org/10.1126/science.1065848>
- Brabletz, T., Jung, A., Spaderna, S., Hlubek, F., & Kirchner, T. (2005). Migrating cancer stem cells — an integrated concept of malignant tumour progression. *Nature Reviews Cancer*, 5(9), 744-749. <https://doi.org/10.1038/nrc1694>
- Brohm, A., Elsayy, H., Rathert, P., Kudithipudi, S., Schoch, T., Schuhmacher, M. K., Weirich, S., & Jeltsch, A. (2019). Somatic Cancer Mutations in the SUV420H1 Protein Lysine Methyltransferase Modulate Its Catalytic Activity. *J Mol Biol*, 431(17), 3068-3080. <https://doi.org/10.1016/j.jmb.2019.06.021>
- Bussi, G., & Parrinello, M. (2007). Accurate sampling using Langevin dynamics. *Phys Rev E Stat Nonlin Soft Matter Phys*, 75(5 Pt 2), 056707. <https://doi.org/10.1103/PhysRevE.75.056707>
- Cancer Genome Atlas, N. (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, 517(7536), 576-582. <https://doi.org/10.1038/nature14129>
- Cao, F., Townsend, E. C., Karatas, H., Xu, J., Li, L., Lee, S., Liu, L., Chen, Y., Ouillette, P., Zhu, J., Hess, J. L., Atadja, P., Lei, M., Qin, Z. S., Malek, S., Wang, S., & Dou, Y. (2014). Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol Cell*, 53(2), 247-261. <https://doi.org/10.1016/j.molcel.2013.12.001>
- Case, D. A., V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossvary, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollman. (2014). AMBER 14. *University of California, San Francisco*.
- Chang, C.-E. G., Michael K. (2004). Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. *Journal of the American Chemical Society*, 40, 9. <https://doi.org/13156-13164>
- Chen, K., Liu, J., Liu, S., Xia, M., Zhang, X., Han, D., Jiang, Y., Wang, C., & Cao, X. (2017). Methyltransferase SETD2-Mediated Methylation of STAT1 Is Critical for Interferon Antiviral Activity. *Cell*, 170(3), 492-506 e414. <https://doi.org/10.1016/j.cell.2017.06.042>
- Chen, R., Chen, Y., Zhao, W., Fang, C., Zhou, W., Yang, X., & Ji, M. (2020). The Role of Methyltransferase NSD2 as a Potential Oncogene in Human Solid Tumors. *Onco Targets Ther*, 13, 6837-6846. <https://doi.org/10.2147/OTT.S259873>
- Chen, S., Kapilashrami, K., Senevirathne, C., Wang, Z., Wang, J., Linscott, J. A., & Luo, M. (2019). Substrate-Differentiated Transition States of SET7/9-Catalyzed Lysine Methylation. *J Am Chem Soc*, 141(20), 8064-8067. <https://doi.org/10.1021/jacs.9b02553>
- Cheng, X., Collins, R. E., & Zhang, X. (2005). Structural and sequence motifs of protein (histone) methylation enzymes. *Annu Rev Biophys Biomol Struct*, 34, 267-294. <https://doi.org/10.1146/annurev.biophys.34.040204.144452>

- Cheong, C. M., Mrozik, K. M., Hewett, D. R., Bell, E., Panagopoulos, V., Noll, J. E., Licht, J. D., Gronthos, S., Zannettino, A. C. W., & Vandyke, K. (2020). Twist-1 is upregulated by NSD2 and contributes to tumour dissemination and an epithelial-mesenchymal transition-like gene expression signature in t(4;14)-positive multiple myeloma. *Cancer Lett*, *475*, 99-108. <https://doi.org/10.1016/j.canlet.2020.01.040>
- Chmiela, S., Saucedo, H. E., Muller, K. R., & Tkatchenko, A. (2018). Towards exact molecular dynamics simulations with machine-learned force fields. *Nat Commun*, *9*(1), 3887. <https://doi.org/10.1038/s41467-018-06169-2>
- Chu, Y., Yao, J., & Guo, H. (2012). QM/MM MD and free energy simulations of G9a-like protein (GLP) and its mutants: understanding the factors that determine the product specificity. *PLoS One*, *7*(5), e37674. <https://doi.org/10.1371/journal.pone.0037674>
- Clarke, S. G. (2013). Protein methylation at the surface and buried deep: thinking outside the histone box. *Trends Biochem Sci*, *38*(5), 243-252. <https://doi.org/10.1016/j.tibs.2013.02.004>
- Cole, P. A. (2008). Chemical probes for histone-modifying enzymes. *Nature Chemical Biology*, *4*(10), 590-597. <https://doi.org/10.1038/nchembio.111>
- Collins, R. E., Tachibana, M., Tamaru, H., Smith, K. M., Jia, D., Zhang, X., Selker, E. U., Shinkai, Y., & Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *J Biol Chem*, *280*(7), 5563-5570. <https://doi.org/10.1074/jbc.M410483200>
- Copeland, R. A., Solomon, M. E., & Richon, V. M. (2009). Protein methyltransferases as a target class for drug discovery. *Nat Rev Drug Discov*, *8*(9), 724-732. <https://doi.org/10.1038/nrd2974>
- Cornett, E. M., Ferry, L., Defossez, P. A., & Rothbart, S. B. (2019). Lysine Methylation Regulators Moonlighting outside the Epigenome. *Mol Cell*, *75*(6), 1092-1101. <https://doi.org/10.1016/j.molcel.2019.08.026>
- Cortopassi, W. A., Kumar, K., Duarte, F., Pimentel, A. S., & Paton, R. S. (2016). Mechanisms of histone lysine-modifying enzymes: A computational perspective on the role of the protein environment. *J Mol Graph Model*, *67*, 69-84. <https://doi.org/10.1016/j.jmgm.2016.04.011>
- Couture, J. F., Collazo, E., Brunzelle, J. S., & Trievel, R. C. (2005). Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev*, *19*(12), 1455-1465. <https://doi.org/10.1101/gad.1318405>
- Couture, J. F., Collazo, E., Hauk, G., & Trievel, R. C. (2006). Structural basis for the methylation site specificity of SET7/9. *Nat Struct Mol Biol*, *13*(2), 140-146. <https://doi.org/10.1038/nsmb1045>
- Dasetty, S., Meza-Morales, P. J., Getman, R. B., & Sarupria, S. (2019). Simulations of interfacial processes: recent advances in force field development. *Current Opinion in Chemical Engineering*, *23*, 138-145. <https://doi.org/10.1016/j.coche.2019.04.003>
- David W.H. Swenson, S. R. (2017). *Contact Map Explorer*. In [https://github.com/dwhswenson/contact\\_map](https://github.com/dwhswenson/contact_map). <https://contact-map.readthedocs.io/en/latest/index.html>
- Del Rizzo, P. A., Couture, J. F., Dirk, L. M., Strunk, B. S., Roiko, M. S., Brunzelle, J. S., Houtz, R. L., & Trievel, R. C. (2010). SET7/9 catalytic mutants reveal the role of active site water molecules in lysine multiple methylation. *J Biol Chem*, *285*(41), 31849-31858. <https://doi.org/10.1074/jbc.M110.114587>



- Del Rizzo, P. A., & Trievel, R. C. (2014). Molecular basis for substrate recognition by lysine methyltransferases and demethylases. *Biochim Biophys Acta*, 1839(12), 1404-1415. <https://doi.org/10.1016/j.bbagr.2014.06.008>
- Dhayalan, A., Kudithipudi, S., Rathert, P., & Jeltsch, A. (2011). Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. *Chem Biol*, 18(1), 111-120. <https://doi.org/10.1016/j.chembiol.2010.11.014>
- DiFiore, J. V., Ptacek, T. S., Wang, Y., Li, B., Simon, J. M., & Strahl, B. D. (2020). Unique and Shared Roles for Histone H3K36 Methylation States in Transcription Regulation Functions. *Cell Rep*, 31(10), 107751. <https://doi.org/10.1016/j.celrep.2020.107751>
- Dillon, S. C., Zhang, X., Trievel, R. C., & Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol*, 6(8), 227. <https://doi.org/10.1186/gb-2005-6-8-227>
- Dirk, L. M., Flynn, E. M., Dietzel, K., Couture, J. F., Trievel, R. C., & Houtz, R. L. (2007). Kinetic Manifestation of Processivity during Multiple Methylations Catalyzed by SET Domain Protein Methyltransferases. *Biochemistry*, 46, 3905-3915. <https://doi.org/10.1021/bi6023644>
- Dukatz, M., Adam, S., Biswal, M., Song, J., Bashtrykov, P., & Jeltsch, A. (2020). Complex DNA sequence readout mechanisms of the DNMT3B DNA methyltransferase. *Nucleic Acids Res*, 48(20), 11495-11509. <https://doi.org/10.1093/nar/gkaa938>
- Dukatz, M., Dittrich, M., Stahl, E., Adam, S., de Mendoza, A., Bashtrykov, P., & Jeltsch, A. (2022). DNA methyltransferase DNMT3A forms interaction networks with the CpG site and flanking sequence elements for efficient methylation. *J Biol Chem*, 298(10), 102462. <https://doi.org/10.1016/j.jbc.2022.102462>
- Dukatz, M., Holzer, K., Choudalakis, M., Emperle, M., Lungu, C., Bashtrykov, P., & Jeltsch, A. (2019). H3K36me<sub>2/3</sub> Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to its Chromatin Interaction. *J Mol Biol*, 431(24), 5063-5074. <https://doi.org/10.1016/j.jmb.2019.09.006>
- Eastman, P., & Pande, V. S. (2010). OpenMM: A Hardware Independent Framework for Molecular Simulations. *Computing in science & engineering*, 12(4), 34-39. <https://doi.org/10.1109/MCSE.2010.27>
- Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L. P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R., & Pande, V. S. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol*, 13(7), e1005659. <https://doi.org/10.1371/journal.pcbi.1005659>
- Edmunds, J. W., Mahadevan, L. C., & Clayton, A. L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J*, 27(2), 406-420. <https://doi.org/10.1038/sj.emboj.7601967>
- Emperle, M., Adam, S., Kunert, S., Dukatz, M., Baude, A., Plass, C., Rathert, P., Bashtrykov, P., & Jeltsch, A. (2019). Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. *Nucleic Acids Res*, 47(21), 11355-11367. <https://doi.org/10.1093/nar/gkz911>
- Emperle, M., Bangalore, D. M., Adam, S., Kunert, S., Heil, H. S., Heinze, K. G., Bashtrykov, P., Tessmer, I., & Jeltsch, A. (2021). Structural and biochemical insight into the mechanism of dual CpG site binding and methylation by the DNMT3A DNA methyltransferase. *Nucleic Acids Res*, 49(14), 8294-8308. <https://doi.org/10.1093/nar/gkab600>

- Emperle, M., Dukatz, M., Kunert, S., Holzer, K., Rajavelu, A., Jurkowska, R. Z., & Jeltsch, A. (2018). The DNMT3A R882H mutation does not cause dominant negative effects in purified mixed DNMT3A/R882H complexes. *Sci Rep*, *8*(1), 13242. <https://doi.org/10.1038/s41598-018-31635-8>
- Ezponda, T., Popovic, R., Shah, M. Y., Martinez-Garcia, E., Zheng, Y., Min, D. J., Will, C., Neri, A., Kelleher, N. L., Yu, J., & Licht, J. D. (2013). The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial-mesenchymal transition and invasive properties of prostate cancer. *Oncogene*, *32*(23), 2882-2890. <https://doi.org/10.1038/onc.2012.297>
- Faller, R., & de Pablo, J. J. (2002). Constant pressure hybrid Molecular Dynamics–Monte Carlo simulations. *The Journal of Chemical Physics*, *116*(1). <https://doi.org/10.1063/1.1420460>
- Falnes, Pål Ø., Jakobsson, Magnus E., Davydova, E., Ho, A., & Małeckı, J. (2016). Protein lysine methylation by seven- $\beta$ -strand methyltransferases. *Biochemical Journal*, *473*(14), 1995-2009. <https://doi.org/10.1042/bcj20160117>
- Fang, D., Gan, H., Lee, J., Han, J., Wang, Z., Riester, S. M., Jin, L., Chen, J., Zhou, H., Wang, J., Zhang, H., Yang, N., Bradley, E. W., Ho, T. H., Rubin, B. P., Bridge, J. A., Thibodeau, S. N., Ordog, T., Chen, Y., . . . Zhang, Z. (2016). The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. *Science*, *352*, 1344-1348. <https://doi.org/10.1126/science.aae0065>
- Fatemi, M., Hermann, A., Pradhan, S., & Jeltsch, A. (2001). The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. Edited by J. Karn. *Journal of Molecular Biology*, *309*(5), 1189-1199. <https://doi.org/10.1006/jmbi.2001.4709>
- Fegan, A. W., Brian; Carlson, Jonathan C. T.; Wagner, Carston R. (2010). Chemically controlled protein assembly: techniques and applications. *Chemical reviews*, *6*, 22. <https://doi.org/10.1021/cr8002888>
- Feoli, A., Viviano, M., Cipriano, A., Milite, C., Castellano, S., & Sbardella, G. (2022). Lysine methyltransferase inhibitors: where we are now. *RSC Chem Biol*, *3*(4), 359-406. <https://doi.org/10.1039/d1cb00196e>
- Ferguson, A. D., Larsen, N. A., Howard, T., Pollard, H., Green, I., Grande, C., Cheung, T., Garcia-Arenas, R., Cowen, S., Wu, J., Godin, R., Chen, H., & Keen, N. (2011). Structural basis of substrate methylation and inhibition of SMYD2. *Structure*, *19*(9), 1262-1273. <https://doi.org/10.1016/j.str.2011.06.011>
- Fontebasso, A. M., Schwartzenruber, J., Khuong-Quang, D. A., Liu, X. Y., Sturm, D., Korshunov, A., Jones, D. T., Witt, H., Kool, M., Albrecht, S., Fleming, A., Hadjadj, D., Busche, S., Lepage, P., Montpetit, A., Staffa, A., Gerges, N., Zakrzewska, M., Zakrzewski, K., . . . Majewski, J. (2013). Mutations in SETD2 and genes affecting histone H3K36 methylation target hemispheric high-grade gliomas. *Acta Neuropathol*, *125*(5), 659-669. <https://doi.org/10.1007/s00401-013-1095-8>
- Gaughan, L., Stockley, J., Wang, N., McCracken, S. R., Treumann, A., Armstrong, K., Shaheen, F., Watt, K., McEwan, I. J., Wang, C., Pestell, R. G., & Robson, C. N. (2011). Regulation of the androgen receptor by SET9-mediated methylation. *Nucleic Acids Res*, *39*(4), 1266-1279. <https://doi.org/10.1093/nar/gkq861>
- Gowher, H., & Jeltsch, A. (2001). Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-

- CpG [correction of non-CpA] sites. *J Mol Biol*, 309(5), 1201-1208. <https://doi.org/10.1006/jmbi.2001.4710>
- Gowher, H., & Jeltsch, A. (2018). Mammalian DNA methyltransferases: new discoveries and open questions. *Biochemical Society Transactions*, 46(5), 1191-1202. <https://doi.org/10.1042/BST20170574>
- Goyal, R., Reinhardt, R., & Jeltsch, A. (2006). Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. *Nucleic Acids Research*, 34(4), 1182-1188. <https://doi.org/10.1093/nar/gkl002>
- Grebien, F., Vedadi, M., Getlik, M., Giambruno, R., Grover, A., Avellino, R., Skucha, A., Vittori, S., Kuznetsova, E., Smil, D., Barsyte-Lovejoy, D., Li, F., Poda, G., Schapira, M., Wu, H., Dong, A., Senisterra, G., Stukalov, A., Huber, K. V. M., . . . Superti-Furga, G. (2015). Pharmacological targeting of the Wdr5-MLL interaction in C/EBPalpha N-terminal leukemia. *Nat Chem Biol*, 11(8), 571-578. <https://doi.org/10.1038/nchembio.1859>
- Gregory, G. D., Vakoc, C. R., Rozovskaia, T., Zheng, X., Patel, S., Nakamura, T., Canaani, E., & Blobel, G. A. (2007). Mammalian ASH1L Is a Histone Methyltransferase That Occupies the Transcribed Region of Active Genes. *Molecular and Cellular Biology*, 27(24), 8466-8479. <https://doi.org/10.1128/MCB.00993-07>
- Hade, M. D., Suire, C. N., & Suo, Z. (2023). An Effective Peptide-Based Platform for Efficient Exosomal Loading and Cellular Delivery of a microRNA. *ACS Appl Mater Interfaces*, 15(3), 3851-3866. <https://doi.org/10.1021/acsami.2c20728>
- Hamamoto, R., Saloura, V., & Nakamura, Y. (2015). Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nature Reviews Cancer*, 15(2), 110-124. <https://doi.org/10.1038/nrc3884>
- Hanley, R. P., Nie, D. Y., Tabor, J. R., Li, F., Sobh, A., Xu, C., Barker, N. K., Dilworth, D., Hajian, T., Gibson, E., Szewczyk, M. M., Brown, P. J., Barsyte-Lovejoy, D., Herring, L. E., Wang, G. G., Licht, J. D., Vedadi, M., Arrowsmith, C. H., & James, L. I. (2023). Discovery of a Potent and Selective Targeted NSD2 Degrader for the Reduction of H3K36me2. *Journal of the American Chemical Society*, 145(14), 8176-8188. <https://doi.org/10.1021/jacs.3c01421>
- Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. *Neuron*, 99(6), 1129-1143. <https://doi.org/10.1016/j.neuron.2018.08.011>
- Horn, H. W., Swope, W. C., Pitner, J. W., Madura, J. D., Dick, T. J., Hura, G. L., & Head-Gordon, T. (2004). Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys*, 120(20), 9665-9678. <https://doi.org/10.1063/1.1683075>
- Hospital, A., Goni, J. R., Orozco, M., & Gelpi, J. L. (2015). Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem*, 8, 37-47. <https://doi.org/10.2147/AABC.S70333>
- Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., & Eden, A. (2008). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene*, 27(3), 404-408. <https://doi.org/10.1038/sj.onc.1210631>
- Hsu, P. L., Shi, H., Leonen, C., Kang, J., Chatterjee, C., & Zheng, N. (2019). Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS. *Mol Cell*, 76(5), 712-723 e714. <https://doi.org/10.1016/j.molcel.2019.10.013>

- Hu, P., & Zhang, Y. (2006). Catalytic Mechanism and Product Specificity of the Histone Lysine Methyltransferase SET7/9: An ab Initio QM/MM-FE Study with Multiple Initial Structures. *128*, 1272-1278. <https://doi.org/10.1021/ja056153>
- Husmann, D., & Gozani, O. (2019). Histone lysine methyltransferases in biology and disease. *Nature Structural & Molecular Biology*, *26*(10), 880-889. <https://doi.org/10.1038/s41594-019-0298-7>
- Hyun, K., Jeon, J., Park, K., & Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Experimental & Molecular Medicine*, *49*(4), e324-e324. <https://doi.org/10.1038/emm.2017.11>
- Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science*, *333*(6047), 1300-1303. <https://doi.org/10.1126/science.1210597>
- Jaffe, J. D., Wang, Y., Chan, H. M., Zhang, J., Huether, R., Kryukov, G. V., Bhang, H. E., Taylor, J. E., Hu, M., Englund, N. P., Yan, F., Wang, Z., Robert McDonald, E., 3rd, Wei, L., Ma, J., Easton, J., Yu, Z., deBeaumont, R., Gibaja, V., . . . Stegmeier, F. (2013). Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat Genet*, *45*(11), 1386-1391. <https://doi.org/10.1038/ng.2777>
- Janson, G., & Paiardini, A. (2021). PyMod 3: a complete suite for structural bioinformatics in PyMOL. *Bioinformatics*, *37*(10), 1471-1472. <https://doi.org/10.1093/bioinformatics/btaa849>
- Jeltsch, A., Adam, S., Dukatz, M., Emperle, M., & Bashtrykov, P. (2021). Deep Enzymology Studies on DNA Methyltransferases Reveal Novel Connections between Flanking Sequences and Enzyme Activity. *J Mol Biol*, *433*(19), 167186. <https://doi.org/10.1016/j.jmb.2021.167186>
- Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., & Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, *449*(7159), 248-251. <https://doi.org/10.1038/nature06146>
- Jones, P. L., Veenstra, G. J. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., & Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics*, *19*(2), 187-191. <https://doi.org/10.1038/561>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Junmei Wang, W. W., Peter A. Kollman and David A. Case. (2001). Antechamber, An Accessory Software Package For Molecular Mechanical Calculations. *Journal of Chemical Information and Computer Sciences*, *222*, 41.
- Jurkowska, R. Z., Anspach, N., Urbanke, C., Jia, D., Reinhardt, R., Nellen, W., Cheng, X., & Jeltsch, A. (2008). Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic Acids Res*, *36*(21), 6656-6663. <https://doi.org/10.1093/nar/gkn747>
- Jurkowska, R. Z., Jurkowski, T. P., & Jeltsch, A. (2011). Structure and function of mammalian DNA methyltransferases. *ChemBiochem*, *12*(2), 206-222. <https://doi.org/10.1002/cbic.201000195>
- Jurkowska, R. Z., Rajavelu, A., Anspach, N., Urbanke, C., Jankevicius, G., Ragozin, S., Nellen, W., & Jeltsch, A. (2011). Oligomerization and binding of the Dnmt3a DNA methyltransferase to

- parallel DNA molecules: heterochromatic localization and role of Dnmt3L. *J Biol Chem*, 286(27), 24200-24207. <https://doi.org/10.1074/jbc.M111.254987>
- Khella, M. S., Schnee, P., Weirich, S., Bui, T., Brohm, A., Bashtrykov, P., Pleiss, J., & Jeltsch, A. (2023). The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation. *J Biol Chem*, 104796. <https://doi.org/10.1016/j.jbc.2023.104796>
- Kim, G., Lee, Y., Ha, J., Han, S., & Lee, M. (2021). Engineering exosomes for pulmonary delivery of peptides and drugs to inflammatory lung cells by inhalation. *J Control Release*, 330, 684-695. <https://doi.org/10.1016/j.jconrel.2020.12.053>
- Kim, S. J., Zhao, H., Hardikar, S., Singh, A. K., Goodell, M. A., & Chen, T. (2013). A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. *Blood*, 122(25), 4086-4089. <https://doi.org/10.1182/blood-2013-02-483487>
- Klose, R. J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., & Zhang, Y. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature*, 442(7100), 312-316. <https://doi.org/10.1038/nature04853>
- Ko, S., Ahn, J., Song, C. S., Kim, S., Knapczyk-Stwora, K., & Chatterjee, B. (2011). Lysine methylation and functional modulation of androgen receptor by Set9 methyltransferase. *Mol Endocrinol*, 25(3), 433-444. <https://doi.org/10.1210/me.2010-0482>
- Kong, X., Ouyang, S., Liang, Z., Lu, J., Chen, L., Shen, B., Li, D., Zheng, M., Li, K. K., Luo, C., & Jiang, H. (2011). Catalytic mechanism investigation of lysine-specific demethylase 1 (LSD1): a computational study. *PLoS One*, 6(9), e25444. <https://doi.org/10.1371/journal.pone.0025444>
- Kontaki, H., & Talianidis, I. (2010). Lysine Methylation Regulates E2F1-Induced Cell Death. *Molecular Cell*, 39(1), 152-160. <https://doi.org/https://doi.org/10.1016/j.molcel.2010.06.006>
- Kudithipudi, S., Dhayalan, A., Kebede, A. F., & Jeltsch, A. (2012). The SET8 H4K20 protein lysine methyltransferase has a long recognition sequence covering seven amino acid residues. *Biochimie*, 94(11), 2212-2218. <https://doi.org/10.1016/j.biochi.2012.04.024>
- Kudithipudi, S., Jeltsch, A. (2014). Role of somatic cancer mutations in human protein lysine methyltransferases. *Biochim Biophys Acta*, 1846(2), 366-379. <https://doi.org/10.1016/j.bbcan.2014.08.002>
- Kudithipudi, S., Kusevic, D., Weirich, S., & Jeltsch, A. (2014). Specificity analysis of protein lysine methyltransferases using SPOT peptide arrays. *J Vis Exp*(93), e52203. <https://doi.org/10.3791/52203>
- Kudithipudi, S., Lungu, C., Rathert, P., Happel, N., & Jeltsch, A. (2014). Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1. *Chem Biol*, 21(2), 226-237. <https://doi.org/10.1016/j.chembiol.2013.10.016>
- Kumar, A., & Kono, H. (2020). Heterochromatin protein 1 (HP1): interactions with itself and chromatin components. *Biophys Rev*, 12(2), 387-400. <https://doi.org/10.1007/s12551-020-00663-y>
- Kusevic, D., Kudithipudi, S., Iglesias, N., Moazed, D., & Jeltsch, A. (2017). Clr4 specificity and catalytic activity beyond H3K9 methylation. *Biochimie*, 135, 83-88. <https://doi.org/10.1016/j.biochi.2017.01.013>

- Kwon, T., Chang, J. H., Kwak, E., Lee, C. W., Joachimiak, A., Kim, Y. C., Lee, J., & Cho, Y. (2003). Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. *EMBO J*, 22(2), 292-303. <https://doi.org/10.1093/emboj/cdg025>
- Lam, U. T. F., Tan, B. K. Y., Poh, J. J. X., & Chen, E. S. (2022). Structural and functional specificity of H3K36 methylation. *Epigenetics Chromatin*, 15(1), 17. <https://doi.org/10.1186/s13072-022-00446-7>
- Lee, M. S., & Olson, M. A. (2006). Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys J*, 90(3), 864-877. <https://doi.org/10.1529/biophysj.105.071589>
- Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. E., Kandoth, C., Payton, J. E., Baty, J., Welch, J., Harris, C. C., Lichti, C. F., Townsend, R. R., Fulton, R. S., Dooling, D. J., Koboldt, D. C., Schmidt, H., Zhang, Q., Osborne, J. R., . . . Wilson, R. K. (2010). DNMT3A Mutations in Acute Myeloid Leukemia. *New England Journal of Medicine*, 363(25), 2424-2433. <https://doi.org/10.1056/NEJMoa1005143>
- Li, J., Ahn, J. H., & Wang, G. G. (2019). Understanding histone H3 lysine 36 methylation and its deregulation in disease. *Cell Mol Life Sci*, 76(15), 2899-2916. <https://doi.org/10.1007/s00018-019-03144-y>
- Li, M., Phatnani, H. P., Guan, Z., Sage, H., Greenleaf, A. L., & Zhou, P. (2005). Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proceedings of the National Academy of Sciences*, 102(49), 17636-17641. <https://doi.org/10.1073/pnas.0506350102>
- Li, W., Tian, W., Yuan, G., Deng, P., Sengupta, D., Cheng, Z., Cao, Y., Ren, J., Qin, Y., Zhou, Y., Jia, Y., Gozani, O., Patel, D. J., & Wang, Z. (2021). Molecular basis of nucleosomal H3K36 methylation by NSD methyltransferases. *Nature*, 590(7846), 498-503. <https://doi.org/10.1038/s41586-020-03069-8>
- Li, Y., Trojer, P., Xu, C. F., Cheung, P., Kuo, A., Drury, W. J., 3rd, Qiao, Q., Neubert, T. A., Xu, R. M., Gozani, O., & Reinberg, D. (2009). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J Biol Chem*, 284(49), 34283-34295. <https://doi.org/10.1074/jbc.M109.034462>
- Lin, C. C., Chen, Y. P., Yang, W. Z., Shen, J. C. K., & Yuan, H. S. (2020). Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. *Nucleic Acids Res*, 48(7), 3949-3961. <https://doi.org/10.1093/nar/gkaa111>
- Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O., & Shaw, D. E. (2012). Systematic validation of protein force fields against experimental data. *PLoS One*, 7(2), e32131. <https://doi.org/10.1371/journal.pone.0032131>
- Linscott, J. A., Kapilashrami, K., Wang, Z., Senevirathne, C., Bothwell, I. R., Blum, G., & Luo, M. (2016). Kinetic isotope effects reveal early transition state of protein lysine methyltransferase SET8. *Proc Natl Acad Sci U S A*, 113(52), E8369-E8378. <https://doi.org/10.1073/pnas.1609032114>
- Liu, L., Shi, T., Houk, K. N., & Zhao, Y. L. (2019). Understanding the R882H mutation effects of DNA methyltransferase DNMT3A: a combination of molecular dynamics simulations and QM/MM calculations. *RSC Adv*, 9(54), 31425-31434. <https://doi.org/10.1039/c9ra06791d>

- Liu, Y., Zhang, Y., Xue, H., Cao, M., Bai, G., Mu, Z., Yao, Y., Sun, S., Fang, D., & Huang, J. (2021). Cryo-EM structure of SETD2/Set2 methyltransferase bound to a nucleosome containing oncohistone mutations. *Cell Discov*, 7(1), 32. <https://doi.org/10.1038/s41421-021-00261-6>
- Lopes, P. E. M., Guvench, O., & MacKerell, A. D. (2015). Current Status of Protein Force Fields for Molecular Dynamics Simulations. In A. Kukol (Ed.), *Molecular Modeling of Proteins* (pp. 47-71). Springer New York. [https://doi.org/10.1007/978-1-4939-1465-4\\_3](https://doi.org/10.1007/978-1-4939-1465-4_3)
- Luo, M. (2012). Current chemical biology approaches to interrogate protein methyltransferases. *ACS Chem Biol*, 7(3), 443-463. <https://doi.org/10.1021/cb200519y>
- Luo, M. (2015). Inhibitors of protein methyltransferases as chemical tools. *Epigenomics*, 7(8), 1327-1338. <https://doi.org/10.2217/epi.15.87>
- Luo, M. (2018). Chemical and Biochemical Perspectives of Protein Lysine Methylation. *Chem Rev*, 118(14), 6656-6705. <https://doi.org/10.1021/acs.chemrev.8b00008>
- Mack, A., Emperle, M., Schnee, P., Adam, S., Pleiss, J., Bashtrykov, P., & Jeltsch, A. (2022). Preferential Self-interaction of DNA Methyltransferase DNMT3A Subunits Containing the R882H Cancer Mutation Leads to Dominant Changes of Flanking Sequence Preferences. *J Mol Biol*, 434(7), 167482. <https://doi.org/10.1016/j.jmb.2022.167482>
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8), 3696-3713. <https://doi.org/10.1021/acs.jctc.5b00255>
- Mallona, I., Ilie, I. M., Karemaker, I. D., Butz, S., Manzo, M., Caflich, A., & Baubec, T. (2021). Flanking sequence preference modulates de novo DNA methylation in the mouse genome. *Nucleic Acids Res*, 49(1), 145-157. <https://doi.org/10.1093/nar/gkaa1168>
- Marmorstein, R. (2003). Structure of SET domain proteins: a new twist on histone methylation. *Trends in Biochemical Sciences*, 28(2), 59-62. [https://doi.org/https://doi.org/10.1016/S0968-0004\(03\)00007-0](https://doi.org/https://doi.org/10.1016/S0968-0004(03)00007-0)
- Marrink, S. J., & Tieleman, D. P. (2013). Perspective on the Martini model. *Chem Soc Rev*, 42(16), 6801-6822. <https://doi.org/10.1039/c3cs60093a>
- Mazur, P. K., Reynoird, N., Khatri, P., Jansen, P. W., Wilkinson, A. W., Liu, S., Barbash, O., Van Aller, G. S., Huddleston, M., Dhanak, D., Tummino, P. J., Kruger, R. G., Garcia, B. A., Butte, A. J., Vermeulen, M., Sage, J., & Gozani, O. (2014). SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. *Nature*, 510(7504), 283-287. <https://doi.org/10.1038/nature13320>
- McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernandez, C. X., Schwantes, C. R., Wang, L. P., Lane, T. J., & Pande, V. S. (2015). MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J*, 109(8), 1528-1532. <https://doi.org/10.1016/j.bpj.2015.08.015>
- McLean, C. M., Karemaker, I. D., & van Leeuwen, F. (2014). The emerging roles of DOT1L in leukemia and normal development. *Leukemia*, 28(11), 2131-2138. <https://doi.org/10.1038/leu.2014.169>
- Meller, J. a. (2001). Molecular Dynamics. In *Encyclopedia of Life Sciences*. Nature Publishing Group. <https://doi.org/10.1038/npg.els.0003048>

- Min, J., Feng, Q., Li, Z., Zhang, Y., & Xu, R. M. (2003). Structure of the Catalytic Domain of Human DOT1L, a Non-SET Domain Nucleosomal Histone Methyltransferase. *Cell*, *112*(5), 711-723. [https://doi.org/10.1016/S0092-8674\(03\)00114-4](https://doi.org/10.1016/S0092-8674(03)00114-4)
- Mitchell, B., Thor, S., & Piper, M. (2023). Cellular and molecular functions of SETD2 in the central nervous system. *J Cell Sci*, *136*(21). <https://doi.org/10.1242/jcs.261406>
- Molenaar, T. M., Malik, M., Silva, J., Liu, N. Q., Haarhuis, J. H. I., Ambrosi, C., Kwesi-Maliepaard, E. M., van Welsem, T., Baubec, T., Faller, W. J., & van Leeuwen, F. (2022). The histone methyltransferase SETD2 negatively regulates cell size. *J Cell Sci*, *135*(19). <https://doi.org/10.1242/jcs.259856>
- Molenaar, T. M., & van Leeuwen, F. (2022). SETD2: from chromatin modifier to multipronged regulator of the genome and beyond. *Cell Mol Life Sci*, *79*(6), 346. <https://doi.org/10.1007/s00018-022-04352-9>
- Monticelli, L., & Tieleman, D. P. (2013). Force Fields for Classical Molecular Dynamics. In L. Monticelli & E. Salonen (Eds.), *Biomolecular Simulations: Methods and Protocols* (pp. 197-213). Humana Press. [https://doi.org/10.1007/978-1-62703-017-5\\_8](https://doi.org/10.1007/978-1-62703-017-5_8)
- Morishita, M., & di Luccio, E. (2011). Structural insights into the regulation and the recognition of histone marks by the SET domain of NSD1. *Biochem Biophys Res Commun*, *412*(2), 214-219. <https://doi.org/10.1016/j.bbrc.2011.07.061>
- Mzoughi, S., Tan, Y. X., Low, D., & Guccione, E. (2016). The role of PRDMs in cancer: one family, two sides. *Curr Opin Genet Dev*, *36*, 83-91. <https://doi.org/10.1016/j.gde.2016.03.009>
- Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., & Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, *543*(7643), 72-77. <https://doi.org/10.1038/nature21373>
- Nguyen, A. T., & Zhang, Y. (2011). The diverse functions of Dot1 and H3K79 methylation. *Genes Dev*, *25*(13), 1345-1358. <https://doi.org/10.1101/gad.2057811>
- Nguyen, T. V., Yao, S., Wang, Y., Rolfe, A., Selvaraj, A., Darman, R., Ke, J., Warmuth, M., Smith, P. G., Larsen, N. A., Yu, L., Zhu, P., Fekkes, P., Vaillancourt, F. H., & Bolduc, D. M. (2019). The R882H DNMT3A hot spot mutation stabilizes the formation of large DNMT3A oligomers with low DNA methyltransferase activity. *J Biol Chem*, *294*(45), 16966-16977. <https://doi.org/10.1074/jbc.RA119.010126>
- Nicholson, T. B., & Chen, T. (2009). LSD1 demethylates histone and non-histone proteins. *Epigenetics*, *4*(3), 129-132. <https://doi.org/10.4161/epi.4.3.8443>
- Nowak, R. P., Tumber, A., Johansson, C., Che, K. H., Brennan, P., Owen, D., & Oppermann, U. (2016). Advances and challenges in understanding histone demethylase biology. *Current Opinion in Chemical Biology*, *33*, 151-159. <https://doi.org/https://doi.org/10.1016/j.cbpa.2016.06.021>
- Nvidia. (2010). Cuda Toolkit Documentation. *Comp. A J. Comp. Education*. <https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html>
- O'Brien, E. C., Brewin, J., & Chevassut, T. (2014). DNMT3A: the DioNysian MonsTer of acute myeloid leukaemia. *Ther Adv Hematol*, *5*(6), 187-196. <https://doi.org/10.1177/2040620714554538>



- Okano, M., Xie, S., & Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics*, *19*(3), 219-220. <https://doi.org/10.1038/890>
- Oyer, J. A., Huang, X., Zheng, Y., Shim, J., Ezponda, T., Carpenter, Z., Allegretta, M., Okot-Kotber, C. I., Patel, J. P., Melnick, A., Levine, R. L., Ferrando, A., Mackerell, A. D., Jr., Kelleher, N. L., Licht, J. D., & Popovic, R. (2014). Point mutation E1099K in MMSET/NSD2 enhances its methyltransferase activity and leads to altered global chromatin methylation in lymphoid malignancies. *Leukemia*, *28*(1), 198-201. <https://doi.org/10.1038/leu.2013.204>
- P. Eastman, M. S. F., J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M. R. Shirts, and V. S. Pande. (2013). OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *Journal of Chemical Theory and Computation*, *9*, 9. <https://doi.org/10.1021/ct300857j>
- Pang, Y.-P. (1999). Novel Zinc Protein Molecular Dynamics Simulations: Steps Toward Antiangiogenesis for Cancer Treatment. *J. Mol. Model*, *5*, 7. <https://doi.org/10.1007/s008940050119>
- Pang, Y.-P. (2001). Successful Molecular Dynamics Simulation of Two Zinc Complexes Bridged by a Hydroxide in Phosphotriesterase Using the Cationic Dummy Atom Method. *PROTEINS: Structure, Function, and Genetics*, *45*, 7. <https://doi.org/10.1002/prot.1138>
- Park, I. Y., Powell, R. T., Tripathi, D. N., Dere, R., Ho, T. H., Blasius, T. L., Chiang, Y. C., Davis, I. J., Fahey, C. C., Hacker, K. E., Verhey, K. J., Bedford, M. T., Jonasch, E., Rathmell, W. K., & Walker, C. L. (2016). Dual Chromatin and Cytoskeletal Remodeling by SETD2. *Cell*, *166*(4), 950-962. <https://doi.org/10.1016/j.cell.2016.07.005>
- Patnaik, D., Chin, H. G., Esteve, P. O., Benner, J., Jacobsen, S. E., & Pradhan, S. (2004). Substrate specificity and kinetic mechanism of mammalian G9a histone H3 methyltransferase. *J Biol Chem*, *279*(51), 53248-53258. <https://doi.org/10.1074/jbc.M409604200>
- Pei, H., Zhang, L., Luo, K., Qin, Y., Chesi, M., Fei, F., Bergsagel, P. L., Wang, L., You, Z., & Lou, Z. (2011). MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. *Nature*, *470*(7332), 124-128. <https://doi.org/10.1038/nature09658>
- Pierro, J., Saliba, J., Narang, S., Sethia, G., Saint Fleur-Lominy, S., Chowdhury, A., Qualls, A., Fay, H., Kilberg, H. L., Moriyama, T., Fuller, T. J., Teachey, D. T., Schmiegelow, K., Yang, J. J., Loh, M. L., Brown, P. A., Zhang, J., Ma, X., Tsirigos, A., . . . Carroll, W. L. (2020). The NSD2 p.E1099K Mutation Is Enriched at Relapse and Confers Drug Resistance in a Cell Context-Dependent Manner in Pediatric Acute Lymphoblastic Leukemia. *Mol Cancer Res*, *18*(8), 1153-1165. <https://doi.org/10.1158/1541-7786.MCR-20-0092>
- Porter, J. R., Zimmerman, M. I., & Bowman, G. R. (2019). Enspara: Modeling molecular ensembles with scalable data structures and parallel computing. *J Chem Phys*, *150*(4), 044108. <https://doi.org/10.1063/1.5063794>
- Poulin, M. B., Schneck, J. L., Matico, R. E., McDevitt, P. J., Huddleston, M. J., Hou, W., Johnson, N. W., Thrall, S. H., Meek, T. D., & Schramm, V. L. (2016). Transition state for the NSD2-catalyzed methylation of histone H3 lysine 36. *Proc Natl Acad Sci U S A*, *113*(5), 1197-1201. <https://doi.org/10.1073/pnas.1521036113>
- Powers, N. R., Parvanov, E. D., Baker, C. L., Walker, M., Petkov, P. M., & Paigen, K. (2016). The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. *PLoS Genet*, *12*(6), e1006146. <https://doi.org/10.1371/journal.pgen.1006146>

- Qian, C., & Zhou, M. M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell Mol Life Sci*, *63*(23), 2755-2763. <https://doi.org/10.1007/s00018-006-6274-5>
- Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D., & Xu, R. M. (2011). The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J Biol Chem*, *286*(10), 8361-8368. <https://doi.org/10.1074/jbc.M110.204115>
- Rathert, P., Dhayalan, A., Ma, H., & Jeltsch, A. (2008). Specificity of protein lysine methyltransferases and methods for detection of lysine methylation of non-histone proteins. *Mol Biosyst*, *4*(12), 1186-1190. <https://doi.org/10.1039/b811673c>
- Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai, Y., Cheng, X., & Jeltsch, A. (2008). Protein lysine methyltransferase G9a acts on non-histone targets. *Nat Chem Biol*, *4*(6), 344-346. <https://doi.org/10.1038/nchembio.88>
- Razin, A., & Riggs, A. D. (1980). DNA Methylation and Gene Function. *Science*, *210*(4470), 604-610. <https://doi.org/10.1126/science.6254144>
- Richon, V. M., Johnston, D., Sneeringer, C. J., Jin, L., Majer, C. R., Elliston, K., Jerva, L. F., Scott, M. P., & Copeland, R. A. (2011). Chemogenetic analysis of human protein methyltransferases. *Chem Biol Drug Des*, *78*(2), 199-210. <https://doi.org/10.1111/j.1747-0285.2011.01135.x>
- Rogawski, D. S., Ndoj, J., Cho, H. J., Maillard, I., Grembecka, J., & Cierpicki, T. (2015). Two Loops Undergoing Concerted Dynamics Regulate the Activity of the ASH1L Histone Methyltransferase. *Biochemistry*, *54*(35), 5401-5413. <https://doi.org/10.1021/acs.biochem.5b00697>
- Saddic, L. A., West, L. E., Aslanian, A., Yates, J. R., 3rd, Rubin, S. M., Gozani, O., & Sage, J. (2010). Methylation of the retinoblastoma tumor suppressor by SMYD2. *J Biol Chem*, *285*(48), 37733-37740. <https://doi.org/10.1074/jbc.M110.137612>
- Sarno, F., Nebbioso, A., & Altucci, L. (2020). DOT1L: a key target in normal chromatin remodelling and in mixed-lineage leukaemia treatment. *Epigenetics*, *15*(5), 439-453. <https://doi.org/10.1080/15592294.2019.1699991>
- Sato, K., Kumar, A., Hamada, K., Okada, C., Oguni, A., Machiyama, A., Sakuraba, S., Nishizawa, T., Nureki, O., Kono, H., Ogata, K., & Sengoku, T. (2021). Structural basis of the regulation of the normal and oncogenic methylation of nucleosomal histone H3 Lys36 by NSD2. *Nat Commun*, *12*(1), 6605. <https://doi.org/10.1038/s41467-021-26913-5>
- Schnee, P., Choudalakis, M., Weirich, S., Khella, M. S., Carvalho, H., Pleiss, J., & Jeltsch, A. (2022). Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide. *Communications Chemistry*, *5*(1), 139. <https://doi.org/10.1038/s42004-022-00753-w>
- Schnee, P., Pleiss, J., & Jeltsch, A. (2023). Approaching the catalytic mechanism of protein lysine methyltransferases by biochemical and simulations techniques.
- Schoelz, J. M., & Riddle, N. C. (2022). Functions of HP1 proteins in transcriptional regulation. *Epigenetics Chromatin*, *15*(1), 14. <https://doi.org/10.1186/s13072-022-00453-8>
- Schrödinger, L. L. C. (2015). The {PyMOL} Molecular Graphics System. *Version: 2.4.1*.

- Schuhmacher, M. K., Beldar, S., Khella, M. S., Brohm, A., Ludwig, J., Tempel, W., Weirich, S., Min, J., & Jeltsch, A. (2020). Sequence specificity analysis of the SETD2 protein lysine methyltransferase and discovery of a SETD2 super-substrate. *Commun Biol*, 3(1), 511. <https://doi.org/10.1038/s42003-020-01223-6>
- Schuhmacher, M. K., Kudithipudi, S., Kusevic, D., Weirich, S., & Jeltsch, A. (2015). Activity and specificity of the human SUV39H2 protein lysine methyltransferase. *Biochim Biophys Acta*, 1849(1), 55-63. <https://doi.org/10.1016/j.bbagr.2014.11.005>
- Seervai, R. N. H., Jangid, R. K., Karki, M., Tripathi, D. N., Jung, S. Y., Kearns, S. E., Verhey, K. J., Cianfrocco, M. A., Millis, B. A., Tyska, M. J., Mason, F. M., Rathmell, W. K., Park, I. Y., Dere, R., & Walker, C. L. (2020). The Huntingtin-interacting protein SETD2/HYPB is an actin lysine methyltransferase. *Science Advances*, 6(40), eabb7854. <https://doi.org/10.1126/sciadv.abb7854>
- Sengupta, D., Zeng, L., Li, Y., Hausmann, S., Ghosh, D., Yuan, G., Nguyen, T. N., Lyu, R., Caporicci, M., Morales Benitez, A., Coles, G. L., Kharchenko, V., Czaban, I., Azhibek, D., Fischle, W., Jaremko, M., Wistuba, II, Sage, J., Jaremko, L., . . . Gozani, O. (2021). NSD2 dimethylation at H3K36 promotes lung adenocarcinoma pathogenesis. *Mol Cell*, 81(21), 4481-4492 e4489. <https://doi.org/10.1016/j.molcel.2021.08.034>
- Sessa, A., Fagnocchi, L., Mastrototaro, G., Massimino, L., Zaghi, M., Indrigo, M., Cattaneo, S., Martini, D., Gabellini, C., Pucci, C., Fasciani, A., Belli, R., Taverna, S., Andreazzoli, M., Zippo, A., & Broccoli, V. (2019). SETD5 Regulates Chromatin Methylation State and Preserves Global Transcriptional Fidelity during Brain Development and Neuronal Wiring. *Neuron*, 104(2), 271-289 e213. <https://doi.org/10.1016/j.neuron.2019.07.013>
- Sirinupong, N., Brunzelle, J., Doko, E., & Yang, Z. (2011). Structural insights into the autoinhibition and posttranslational activation of histone methyltransferase SmyD3. *J Mol Biol*, 406(1), 149-159. <https://doi.org/10.1016/j.jmb.2010.12.014>
- Sirinupong, N., Brunzelle, J., Ye, J., Pirzada, A., Nico, L., & Yang, Z. (2010). Crystal structure of cardiac-specific histone methyltransferase SmyD1 reveals unusual active site architecture. *J Biol Chem*, 285(52), 40635-40644. <https://doi.org/10.1074/jbc.M110.168187>
- Sun, X. J., Wei, J., Wu, X. Y., Hu, M., Wang, L., Wang, H. H., Zhang, Q. H., Chen, S. J., Huang, Q. H., & Chen, Z. (2005). Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *J Biol Chem*, 280(42), 35261-35271. <https://doi.org/10.1074/jbc.M504012200>
- Swaroop, A., Oyer, J. A., Will, C. M., Huang, X., Yu, W., Troche, C., Bulic, M., Durham, B. H., Wen, Q. J., Crispino, J. D., MacKerell, A. D., Jr., Bennett, R. L., Kelleher, N. L., & Licht, J. D. (2019). An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. *Oncogene*, 38(5), 671-686. <https://doi.org/10.1038/s41388-018-0474-y>
- Tachibana, M., Sugimoto, K., Fukushima, T., & Shinkai, Y. (2001). Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem*, 276(27), 25309-25317. <https://doi.org/10.1074/jbc.M101914200>
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, 324(5929), 930-935. <https://doi.org/10.1126/science.1170116>

- Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., . . . Forbes, S. A. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, *47*(D1), D941-D947. <https://doi.org/10.1093/nar/gky1015>
- Tate, P. H., & Bird, A. P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. *Current Opinion in Genetics & Development*, *3*(2), 226-231. [https://doi.org/https://doi.org/10.1016/0959-437X\(93\)90027-M](https://doi.org/https://doi.org/10.1016/0959-437X(93)90027-M)
- Tom Darden, D. Y., and Lee Pedersen. (1993). An  $N \cdot \log(N)$  method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*, 4. <https://doi.org/10.1063/1.464397>
- Topchu, I., Pangen, R. P., Bychkov, I., Miller, S. A., Izumchenko, E., Yu, J., Golemis, E., Karanicolas, J., & Boumber, Y. (2022). The role of NSD1, NSD2, and NSD3 histone methyltransferases in solid tumors. *Cell Mol Life Sci*, *79*(6), 285. <https://doi.org/10.1007/s00018-022-04321-2>
- Triebel, R. C., Beach, B. M., Dirk, L. M. A., Houtz, R. L., & Hurley, J. H. (2002). Structure and Catalytic Mechanism of a SET Domain Protein Methyltransferase. *Cell*, *111*, 91-103. [https://doi.org/10.1016/s0092-8674\(02\)01000-0](https://doi.org/10.1016/s0092-8674(02)01000-0)
- Triebel, R. C., Flynn, E. M., Houtz, R. L., & Hurley, J. H. (2003). Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSM1. *Nat Struct Mol Biol*, *10*, 545-552. <https://doi.org/10.1038/nsb946>
- van Dongen, S. F., Elemans, J. A., Rowan, A. E., & Nolte, R. J. (2014). Processive catalysis. *Angew Chem Int Ed Engl*, *53*(43), 11420-11428. <https://doi.org/10.1002/anie.201404848>
- Wagner, E. J., & Carpenter, P. B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nature Reviews Molecular Cell Biology*, *13*(2), 115-126. <https://doi.org/10.1038/nrm3274>
- Wagner, J. R., Sørensen, J., Hensley, N., Wong, C., Zhu, C., Perison, T., & Amaro, R. E. (2017). POVME 3.0: Software for Mapping Binding Pocket Flexibility. *Journal of Chemical Theory and Computation*, *13*(9), 4584-4592. <https://doi.org/10.1021/acs.jctc.7b00500>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*, *46*(W1), W296-W303. <https://doi.org/10.1093/nar/gky427>
- Weinberg, D. N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K. N., Horth, C., McGuire, J. T., Xu, X., Nikbakht, H., Lemiesz, A. E., Marchione, D. M., Marunde, M. R., Meiners, M. J., Cheek, M. A., Keogh, M. C., Bareke, E., Djedid, A., Harutyunyan, A. S., . . . Lu, C. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature*, *573*(7773), 281-286. <https://doi.org/10.1038/s41586-019-1534-3>
- Weirich, S., & Jeltsch, A. (2022). Specificity Analysis of Protein Methyltransferases and Discovery of Novel Substrates Using SPOT Peptide Arrays. In R. Margueron & D. Holoch (Eds.), *Histone Methyltransferases: Methods and Protocols* (pp. 313-325). Springer US. [https://doi.org/10.1007/978-1-0716-2481-4\\_15](https://doi.org/10.1007/978-1-0716-2481-4_15)
- Weirich, S., Kudithipudi, S., & Jeltsch, A. (2016). Specificity of the SUV4-20H1 and SUV4-20H2 protein lysine methyltransferases and methylation of novel substrates. *J Mol Biol*, *428*(11), 2344-2358. <https://doi.org/10.1016/j.jmb.2016.04.015>

- Weirich, S., Kudithipudi, S., & Jeltsch, A. (2017). Somatic cancer mutations in the MLL1 histone methyltransferase modulate its enzymatic activity and dependence on the WDR5/RBBP5/ASH2L complex. *Mol Oncol*, *11*(4), 373-387. <https://doi.org/10.1002/1878-0261.12041>
- Weirich, S., Kudithipudi, S., Kycia, I., & Jeltsch, A. (2015). Somatic cancer mutations in the MLL3-SET domain alter the catalytic properties of the enzyme. *Clin Epigenetics*, *7*(1), 36. <https://doi.org/10.1186/s13148-015-0075-3>
- Weirich, S., Kusevic, D., Schnee, P., Reiter, J., Pleiss, J., & Jeltsch, A. (2023). Discovery of new NSD2 non-histone substrates and design of a super-substrate. - *Publication submitted for review.*
- Weirich, S., Schuhmacher, M. K., Kudithipudi, S., Lungu, C., Ferguson, A. D., & Jeltsch, A. (2020). Analysis of the Substrate Specificity of the SMYD2 Protein Lysine Methyltransferase and Discovery of Novel Non-Histone Substrates. *ChemBiochem*, *21*(1-2), 256-264. <https://doi.org/10.1002/cbic.201900582>
- West, L. E., & Gozani, O. (2011). Regulation of p53 function by lysine methylation. *Epigenomics*, *3*(3), 361-369. <https://doi.org/10.2217/epi.11.21>
- Wilson, J. R., Jing, C., Walker, P. A., Martin, S. R., Howell, S. A., M., B. G., Gamblin, S. J., & Xiao, B. (2002). Crystal Structure and Functional Analysis of the Histone Methyltransferase SET7/9. *Cell*, *111*(1), 105-115. [https://doi.org/10.1016/S0092-8674\(02\)00964-9](https://doi.org/10.1016/S0092-8674(02)00964-9)
- Wu, C.-Y., Hsieh, C.-Y., Huang, K.-E., Chang, C., & Kang, H.-Y. (2012). Cryptotanshinone down-regulates androgen receptor signaling by modulating lysine-specific demethylase 1 function. *International Journal of Cancer*, *131*(6), 1423-1434. <https://doi.org/https://doi.org/10.1002/ijc.27343>
- Wu, H., Min, J., Lunin, V. V., Antoshenko, T., Dombrowski, L., Zeng, H., Allali-Hassani, A., Campagna-Slater, V., Vedadi, M., Arrowsmith, C. H., Plotnikov, A. N., & Schapira, M. (2010). Structural biology of human H3K9 methyltransferases. *PLoS One*, *5*(1), e8570. <https://doi.org/10.1371/journal.pone.0008570>
- Wu, J., Cheung, T., Grande, C., Ferguson, A. D., Zhu, X., Theriault, K., Code, E., Birr, C., Keen, N., & Chen, H. (2011). Biochemical characterization of human SET and MYND domain-containing protein 2 methyltransferase. *Biochemistry*, *50*(29), 6488-6497. <https://doi.org/10.1021/bi200725p>
- Xiang, W., He, J., Huang, C., Chen, L., Tao, D., Wu, X., Wang, M., Luo, G., Xiao, X., Zeng, F., & Jiang, G. (2015). miR-106b-5p targets tumor suppressor gene SETD2 to inactivate its function in clear cell renal cell carcinoma. *Oncotarget; Vol 6, No 6*. <https://www.oncotarget.com/article/2926/text/>
- Xiao, B., Jing, C., Kelly, G., Walker, P. A., Muskett, F. W., Frenkiel, T. A., Martin, S. R., Sarma, K., Reinberg, D., Gamblin, S. J., & Wilson, J. R. (2005). Specificity and mechanism of the histone methyltransferase Pr-Set7. *Genes Dev*, *19*(12), 1444-1454. <https://doi.org/10.1101/gad.1315905>
- Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, I. A., Blackburn, G. M., & Gamblin, S. J. (2003). Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature*, *421*, 652-656. <https://doi.org/10.1038/nature01378>
- Xu, H., Liao, C., Liang, S., & Ye, B. C. (2021). A Novel Peptide-Equipped Exosomes Platform for Delivery of Antisense Oligonucleotides. *ACS Appl Mater Interfaces*, *13*(9), 10760-10767. <https://doi.org/10.1021/acsami.1c00016>

- Xue, H., Yao, T., Cao, M., Zhu, G., Li, Y., Yuan, G., Chen, Y., Lei, M., & Huang, J. (2019). Structural basis of nucleosome recognition and modification by MLL methyltransferases. *Nature*, *573*(7774), 445-449. <https://doi.org/10.1038/s41586-019-1528-1>
- Yan, X.-J., Xu, J., Gu, Z.-H., Pan, C.-M., Lu, G., Shen, Y., Shi, J.-Y., Zhu, Y.-M., Tang, L., Zhang, X.-W., Liang, W.-X., Mi, J.-Q., Song, H.-D., Li, K.-Q., Chen, Z., & Chen, S.-J. (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nature Genetics*, *43*(4), 309-315. <https://doi.org/10.1038/ng.788>
- Yang, J., Huang, J., Dasgupta, M., Sears, N., Miyagi, M., Wang, B., Chance, M. R., Chen, X., Du, Y., Wang, Y., An, L., Wang, Q., Lu, T., Zhang, X., Wang, Z., & Stark, G. R. (2010). Reversible methylation of promoter-bound STAT3 by histone-modifying enzymes. *Proceedings of the National Academy of Sciences*, *107*(50), 21499-21504. <https://doi.org/10.1073/pnas.1016147107>
- Yang, J., Huang, J., Dasgupta, M., Sears, N., Miyagi, M., Wang, B., Chance, M. R., Chen, X., Du, Y., Wang, Y., An, L., Wang, Q., Lu, T., Zhang, X., Wang, Z., & Stark, G. R. (2010). Reversible methylation of promoter-bound STAT3 by histone-modifying enzymes. *Proc Natl Acad Sci U S A*, *107*(50), 21499-21504. <https://doi.org/10.1073/pnas.1016147107>
- Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A., & Weinberg, R. A. (2004). Twist, a Master Regulator of Morphogenesis, Plays an Essential Role in Tumor Metastasis. *Cell*, *117*(7), 927-939. <https://doi.org/https://doi.org/10.1016/j.cell.2004.06.006>
- Yang, L., Rau, R., & Goodell, M. A. (2015). DNMT3A in haematological malignancies. *Nature Reviews Cancer*, *15*(3), 152-165. <https://doi.org/10.1038/nrc3895>
- Yang, S., Zheng, X., Lu, C., Li, G. M., Allis, C. D., & Li, H. (2016). Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase. *Genes Dev*, *30*(14), 1611-1616. <https://doi.org/10.1101/gad.284323.116>
- Yang, T., Zhang, W., Cheng, J., Nie, Y., Xin, Q., Yuan, S., & Dou, Y. (2019). Formation Mechanism of Ion Channel in Channelrhodopsin-2: Molecular Dynamics Simulation and Steering Molecular Dynamics Simulations. *Int J Mol Sci*, *20*(15). <https://doi.org/10.3390/ijms20153780>
- Yang, W., Zhuang, J., Li, C., & Cheng, G. J. (2023). Unveiling the methyl transfer mechanisms in the epigenetic machinery DNMT3A-3 L: A comprehensive study integrating assembly dynamics with catalytic reactions. *Comput Struct Biotechnol J*, *21*, 2086-2099. <https://doi.org/10.1016/j.csbj.2023.03.002>
- Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., & Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, *356*(6337), eaaj2239. <https://doi.org/10.1126/science.aaj2239>
- Yu, H., & Dalby, P. A. (2020). A beginner's guide to molecular dynamics simulations and the identification of cross-correlation networks for enzyme engineering. *Methods Enzymol*, *643*, 15-49. <https://doi.org/10.1016/bs.mie.2020.04.020>
- Yuan, G., Flores, N. M., Hausmann, S., Lofgren, S. M., Kharchenko, V., Angulo-Ibanez, M., Sengupta, D., Lu, X., Czaban, I., Azhibek, D., Vicent, S., Fischle, W., Jaremko, M., Fang, B., Wistuba, II, Chua, K. F., Roth, J. A., Minna, J. D., Shao, N. Y., . . . Gozani, O. (2021). Elevated NSD3 histone methylation activity drives squamous cell lung cancer. *Nature*, *590*(7846), 504-508. <https://doi.org/10.1038/s41586-020-03170-y>

- Yun, M., Wu, J., Workman, J. L., & Li, B. (2011). Readers of histone modifications. *Cell Research*, 21(4), 564-578. <https://doi.org/10.1038/cr.2011.42>
- Zhang, J., Lee, Y. R., Dang, F., Gan, W., Menon, A. V., Katon, J. M., Hsu, C. H., Asara, J. M., Tibarewal, P., Leslie, N. R., Shi, Y., Pandolfi, P. P., & Wei, W. (2019). PTEN Methylation by NSD2 Controls Cellular Sensitivity to DNA Damage. *Cancer Discov*, 9(9), 1306-1323. <https://doi.org/10.1158/2159-8290.CD-18-0083>
- Zhang, X., & Bruice, T. (2007a). Catalytic Mechanism and Product Specificity of Rubisco Large Subunit Methyltransferase: QM/MM and MD Investigations. *Biochemistry*, 2007, 5505-5514. <https://doi.org/10.1021/bi700119p>
- Zhang, X., & Bruice, T. (2007b). Histone Lysine Methyltransferase SET7/9: Formation of a Water Channel Precedes Each Methyl Transfer. *Biochemistry*, 46, 14838-14844. <https://doi.org/10.1021/bi7014579>
- Zhang, X., & Bruice, T. (2007c). A Quantum Mechanics/Molecular Mechanics Study of the Catalytic Mechanism and Product Specificity of Viral Histone Lysine Methyltransferase. *Biochemistry*, 46, 9743-9751. <https://doi.org/10.1021/bi700515q>
- Zhang, X., & Bruice, T. (2008a). Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases. *PNAS*, 105. <https://doi.org/10.1073/pnas.0801788105>
- Zhang, X., & Bruice, T. (2008b). Mechanism of Product Specificity of AdoMet Methylation Catalyzed by Lysine Methyltransferases: Transcriptional Factor p53 Methylation by Histone Lysine Methyltransferase SET7/9. *Biochemistry*, 47, 2743-2748. <https://doi.org/10.1021/bi702370p>
- Zhang, X., & Bruice, T. (2008c). Product Specificity and Mechanism of Protein Lysine Methyltransferases: Insights from the Histone Lysine Methyltransferase SET8. *Biochemistry*, 47, 6671-6677. <https://doi.org/10.1021/bi800244s>
- Zhang, X., Tamaru, H., Khan, S. I., Horton, J. R., Keefe, L. J., Selker, E. U., & Chen, X. (2002). Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell*, 111(1), 117-127. [https://doi.org/10.1016/s0092-8674\(02\)00999-6](https://doi.org/10.1016/s0092-8674(02)00999-6)
- Zhang, X., Tanaka, K., Yan, J., Li, J., Peng, D., Jiang, Y., Yang, Z., Barton, M. C., Wen, H., & Shi, X. (2013). Regulation of estrogen receptor alpha by histone methyltransferase SMYD2-mediated protein methylation. *Proc Natl Acad Sci U S A*, 110(43), 17284-17289. <https://doi.org/10.1073/pnas.1307959110>
- Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., & Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. *Mol Cell*, 12(1), 177-185. [https://doi.org/10.1016/s1097-2765\(03\)00224-7](https://doi.org/10.1016/s1097-2765(03)00224-7)
- Zhang, Y., Fang, Y., Tang, Y., Han, S., Jia, J., Wan, X., Chen, J., Yuan, Y., Zhao, B., & Fang, D. (2022). SMYD5 catalyzes histone H3 lysine 36 trimethylation at promoters. *Nat Commun*, 13(1), 3190. <https://doi.org/10.1038/s41467-022-30940-1>
- Zhang, Y., Shan, C. M., Wang, J., Bao, K., Tong, L., & Jia, S. (2017). Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. *Sci Rep*, 7, 43906. <https://doi.org/10.1038/srep43906>
- Zhang, Z.-M., Lu, R., Wang, P., Yu, Y., Chen, D., Gao, L., Liu, S., Ji, D., Rothbart, S. B., Wang, Y., Wang, G. G., & Song, J. (2018). Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature*, 554(7692), 387-391. <https://doi.org/10.1038/nature25477>

- Zheng, W., Ibanez, G., Wu, H., Blum, G., Zeng, H., Dong, A., Li, F., Hajian, T., Allali-Hassani, A., Amaya, M. F., Siarheyeva, A., Yu, W., Brown, P. J., Schapira, M., Vedadi, M., Min, J., & Luo, M. (2012). Sinefungin derivatives as inhibitors and structure probes of protein lysine methyltransferase SETD2. *J Am Chem Soc*, *134*(43), 18004-18014. <https://doi.org/10.1021/ja307060p>
- Zhu, K., Lei, P. J., Ju, L. G., Wang, X., Huang, K., Yang, B., Shao, C., Zhu, Y., Wei, G., Fu, X. D., Li, L., & Wu, M. (2017). SPOP-containing complex regulates SETD2 stability and H3K36me3-coupled alternative splicing. *Nucleic Acids Res*, *45*(1), 92-105. <https://doi.org/10.1093/nar/gkw814>



## 7. Appendix

### 7.1. Appendix I (not included in the published thesis)

Manuscript #1:

Mack A, Emperle M, **Schnee P**, Adam S, Pleiss J, Bashtrykov P, Jeltsch A\*, (2022) Preferential Self-interaction of DNA Methyltransferase DNMT3A Subunits Containing the R882H Cancer Mutation Leads to Dominant Changes of Flanking Sequence Preferences. *J Mol Biol* 15;434(7):167482, doi.org/10.1016/j.jmb.2022.167482.

Manuscript #2:

Khella M. S., **Schnee P**, Weirich S, Bui T, Bröhm A, Bashtrykov P, Pleiss J, Jeltsch A\*, (2023) The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation. *J Biol Chem* 104796, doi.org/10.1016/j.jbc.2023.104796

Manuscript #3:

**Schnee P**, Choudalakis M, Weirich S, Khella M. S, Carvalho H, Pleiss J, Jeltsch A\*, (2022) Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide. *Commun. Chem.* 5, 139, doi.org/10.1038/s42004-022-00753-w

## 7.2. Appendix II

Manuscript #4:

Weirich S, Kusevic D, **Schnee P**, Reiter J, Pleiss J, Jeltsch A\*, (2023) Discovery of new NSD2 non-histone substrates and design of a super-substrate – *Communications Biology*, Publication submitted for review.

# Discovery of new NSD2 non-histone substrates and design of a super-substrate

Sara Weirich, Denis Kusevic, Philipp Schnee, Jessica Reiter, Jürgen Pleiss & Albert Jeltsch\*

Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany

\*Corresponding author

Prof. Dr. Albert Jeltsch, Institute of Biochemistry, Faculty of Chemistry, University of Stuttgart

Allmandring 31, 70569 Stuttgart, Germany

## Running title

NSD2 new protein substrates and super-substrates

## Key words

Protein lysine methyltransferase; Histone methylation; H3K36; DNA damage repair; Enzyme specificity; Molecular dynamics simulation

## Abstract

NSD2 catalyzes dimethylation at H3K36 and it has very important roles in development and disease. We investigated the substrate sequence specificity of NSD2 and discovered strong sequence readout between the target site position -3 and +2. Strikingly, amino acid residues differing from the H3K36 target were preferred at some positions in the specificity profile. Combination of preferred residues allowed the design of a super-substrate which was methylated >100-fold faster by NSD2. Molecular dynamics simulations demonstrated that this activity increase is due to distinct hyperactive conformations of the enzyme-peptide complex. A search for human nuclear proteins matching the NSD2 specificity profile led to the discovery of 22 novel peptide substrates. In protein methylation studies, we identified ATRX and FANCM as novel NSD2 protein substrates in vitro and in human cells. Both proteins have important roles in DNA repair strengthening the connection of NSD2 and H3K36 methylation to DNA repair.

## Introduction

Together with DNA methylation and the presence of non-coding RNAs, histone posttranslational modifications (PTM) regulate many chromatin templated processes (Allis & Jenuwein, 2016), control cellular phenotypes and development of diseases (Jambhekar et al., 2019; Zhao et al., 2021). The most prominent histone PTMs are acetylation, phosphorylation, ubiquitination and methylation (Huang et al., 2015). Histone lysine methylation is an extensively studied modification and up to three methyl groups can be transferred on lysine residues by specific Protein lysine methyltransferases (PKMTs) (Boriack-Sjodin & Swinger, 2016; D. Husmann & O. Gozani, 2019). Dependent on the site and degree of methylation, effector proteins with specific binding properties are recruited and regulate further downstream biological processes (Patel & Wang, 2013). Besides the methylation of lysine residues of histone tails, high-throughput proteomic studies led to the identification of lysine methylation events at numerous non-histone proteins where it controls important processes, like protein degradation, protein localization and protein-protein interactions, which in turn regulate many biological processes (Biggar & Li, 2015; Cornett et al., 2019; Zhang et al., 2015). Up to date, the discovery of non-histone targets is incomplete for many PKMTs and it is often difficult to associate the matching substrates and PKMTs. One opportunity to search for possible non-histone substrates of PKMTs is to analyze their substrate specificity (Kudithipudi & Jeltsch, 2016).

The Nuclear Receptor Binding SET Domain Protein 2 (NSD2, also known as MMSET, or WHSC1), together with NSD1 and NSD3 form the nuclear receptor SET domain-containing (NSD) enzyme family. The NSD enzymes are key epigenetic enzymes that catalyze H3K36 mono- and dimethylation (Bennett et al., 2017; Lam et al., 2022; Li et al., 2019). H3K36 methylation in gene bodies is associated with active transcription and splicing and this modification is also involved in the regulation of heterochromatin formation, DNA replication, recombination and DNA repair (Lam et al., 2022; Li et al., 2019; E. J. Wagner & P. B. Carpenter, 2012). The NSD enzymes differ in protein length, where NSD2 with 1365 amino acids (aa) is the smallest among all NSD family members. All three NSD enzyme share the SET domain as catalytically functional part, flanked by AWS (Associated with SET) and Post-SET domains. In addition, they contain PWWP (proline-tryptophan-tryptophan-proline motif) domains and PHD (plant homeodomain) domains, which are important for mediating interaction with chromatin and other proteins (Sankaran et al., 2016). In contrast to all other NSD family members, NSD2 also contains a high mobility group (HMG) domain that interacts with the DNA binding domain of the androgen-receptor (AR) and results in enhanced nuclear translocation of both proteins (Kang et al., 2009). The gene encoding NSD2 was identified in 1998 when it was observed that patients suffering from the Wolf-Hirschhorn syndrome have either a partial or complete deletion of the NSD2 gene, which leads to a haploinsufficiency of NSD2, resulting in the name Wolf-Hirschhorn syndrome candidate 1 (WHSC1) (Stec et al., 1998). NSD2-deficient mice exhibit phenotypes similar to the human WHS, such as growth defects, deficiencies in midline fusion and congenital heart defects and die 10 days after birth (Nimura et al., 2009). Moreover, the Wolf-Hirschhorn syndrome is linked to the DNA damage response (DDR) (Hajdu et al., 2011).

Later it was discovered that NSD2 has PKMT activity and several studies reported that NSD2 catalyze dimethylation of H3K36 (Nimura et al., 2009), H3K4 and H3K9 (Kang et al., 2009), trimethylation of H3K27 (Kim et al., 2008), di- and trimethylation of H4K20 (Marango et al., 2008; Pei et al., 2011), and monomethylation of H4K44, but this reaction was not observed in the chromatin context (Li et al., 2009). However, there is disagreement among published reports regarding the methylation activities of NSD2 in particular at H3K4, H3K27 and H4K20 (Li et al., 2009). In addition, several non-histone substrates of NSD2 with important roles in tumorigenesis and tumor progression have been described. Song et al. (2020) detected Signal transducer and activator of transcription 3 (STAT3) as novel target

of NSD2. With the methylation of STAT3 at lysine 163 by NSD2, the STAT3 signaling pathway was activated, which promotes tumor angiogenesis (Song et al., 2021). Methylation of the Aurora kinase A (AURKA) by NSD2 was discovered, which reduces p53 stability resulting in increased cell proliferation and oncogenic activity (Park et al., 2018). NSD2 has been shown to be crucial for the recruitment of p53 binding protein (53BP1) to DNA double-strand breaks (Pei et al., 2011). Later, it was also shown to methylate K349 of Phosphatase and tensin homolog (PTEN) K349, which is bound by the Tudor domain of 53BP1 and facilitates its recruitment to DNA double strand breaks (Zhang et al., 2019).

In addition to its key role in WHS, many reports demonstrated that NSD2 is mutated in several tumors (Kudithipudi & Jeltsch, 2014; Oyer et al., 2014). Especially, the E1099K mutation, located in the active pocket in a loop next to the bound substrate is recurrently occurring in lung adenocarcinoma where it increases NSD2 activity promoting KRAS signaling and other oncogenic gene expression programs (Sengupta et al., 2021b). The effect of the hyperactivating mutations is comparable to the chromatin alterations often leading to NSD2 overexpressing in tumor cells (Oyer et al., 2014). The activating T1150A mutation in NSD2 is often found in leukemia patients (Sato et al., 2021). This mutation also changes the product specificity from H3K36me2 to H3K36me3 (Khella et al., 2023b).

In this study, we investigated the peptide sequence specificity of NSD2. We observed that amino acid residues different from the natural ones in the H3 tail were preferred at some positions in the substrate peptide. Combining four of these preferred residues led to the development of a super-substrate which was methylated at least 100-fold faster by NSD2. Molecular dynamics simulations demonstrated that this activity increase is caused by distinct hyperactive conformations of the enzyme-peptide complex. In addition, based on the specificity profile of NSD2, we identified K1033 of ATP-dependent helicase (ATRX) and K819 of Fanconi anemia group M (FANCM) protein as novel NSD2 protein substrates *in vitro* and demonstrated their methylation in cells. Both these proteins have important roles in DNA repair and they strengthen the connection of NSD2 and H3K36 methylation to DNA repair.

## Results

### *Substrate specificity analysis of NSD2*

Several previous publications reported that PKMTs methylate different substrates, but for many enzymes of this class, the full substrate spectrum is not known (Kudithipudi & Jeltsch, 2016). Knowledge about PKMT substrates is essential to understand their biological roles, because each individual protein methylation event can regulate a specific biological pathway (Biggar & Li, 2015; Cornett et al., 2019; D. Husmann & O. Gozani, 2019; Zhang et al., 2015). One approach to determine the full substrate spectrum of PKMTs and identify novel substrates is to characterize the sequence specificity of the enzyme. For this task, peptide SPOT arrays can be employed, as they allow to study the methylation of several hundred peptides in one experiment at moderate costs (Kudithipudi, Kusevic, et al., 2014; S. Weirich & A. Jeltsch, 2022). NSD2 is an essential PKMT which generates H3K36 mono- and dimethylation and has essential roles in chromatin regulation, cell physiology and cancer (Bennett et al., 2017). To investigate the substrate specificity of this enzyme, we have cloned the catalytic SET domain (aa 991-1240) of human NSD2 (Uniprot O96028) in a GST-tagged form, overexpressed it in *E. coli* cells, and purified the enzyme by affinity chromatography with good yield and quality (Supplementary Figure 1A). To confirm its methyltransferase activity, peptide arrays containing 15 aa long peptides of potential histone substrates and the corresponding K-to-A mutations as negative control were synthesized on a cellulose membrane using the SPOT technology. The peptide arrays were incubated with NSD2 in presence of radioactively labeled AdoMet as cofactor. As

expected, the autoradiographic image confirmed the methylation of H3K36 and loss of methylation for the negative controls (Supplementary Figure 1B). In addition, we detected strong methylation of H4K44 as previously shown (Li et al., 2009) and observed strong methylation of H1.5K168, which was previously identified as preferred substrate of NSD1 (Kudithipudi, Lungu, et al., 2014). In all these peptides, methylation occurs in a G V **K** (KR)  $\psi$  sequence context ( $\psi$  represents a hydrophobic amino acid residue). To test the PKMT activity of NSD2 with protein substrates, recombinant H3.1 and H4 were incubated with NSD2 in presence of radioactively labeled AdoMet. Since many PKMTs show an automethylation activity (Iglesias et al., 2018; Khella et al., 2020; Lee et al., 2019; Piao et al., 2016; Rathert, Dhayalan, Murakami, et al., 2008; Weil et al., 2018; Weirich et al., 2017), an additional sample without external protein substrate was included. After methylation, the samples were separated by SDS-PAGE and methyl transfer was detected by autoradiography. In this experiment, methylation of recombinant H3.1 and H4 was confirmed and in addition, a strong automethylation signal of NSD2 in absence and presence of recombinant H3.1 and H4 was detected (Supplementary Figure 1C).

Since NSD2 methylates H3K36 at peptide level and recombinant H3.1 at protein level, we next intended to investigate the detailed substrate sequence specificity of NSD2 using H3K36 as template sequence. Therefore, peptide SPOT array was synthesized in which each single position of the template sequence H3K36 (29-43) was exchanged against 18 natural amino acids (excluding cysteine and tryptophan) to create all possible single amino acid exchanges of the template sequence. Three independent methylation reactions were performed and the results of each methylation assay were quantitatively analyzed, normalized and the data averaged (Figure 1A and B). Calculating the standard deviation (SD) of the methylation activity of each single spot indicated a high reproducibility and a good quality of the data, because 90 % of the peptides had an SD less than  $\pm 20$  % and 60 % of the peptides even showed an SD smaller than  $\pm 10$  % (Supplementary Figure 2). Next, the discrimination factors were calculated for better visualization of the substrate specificity at each position as previously described (Figure 1C) (Kudithipudi, Kusevic, et al., 2014). The substrate specificity profile of NSD2 shows that it strongly recognizes the residues from the G33 to P38 of the H3 tail. The results revealed a preference for G at the -3 side, as well as preference for aromatic amino acids and glycine at the -2 position (F>G>Y). At the -1 position, only large aliphatic amino acids (I>L>V) are allowed, whereas at the +1 site few residues are strongly disfavored (A, D, E, G, P). At the +2 position, hydrophobic residues are preferred (V>I>L>P>T).

#### *Design of an NSD2 specific super-substrate*

Surprisingly, we observed at several positions of the NSD2 specificity analysis, that residues different from the natural ones occurring in H3 were preferred, for example K at the -5 site, F at -2, or N at the +3 and +4 sites (Figure 1A and B). We, therefore, speculated that it might be possible to develop an NSD2 specific H3K36 super-substrate (ssK36), which would provide additional information about the mechanism of NSD2 and its potential novel substrates. To this end, we first selected all strongly methylated single point mutations based on the substrate specificity profile from position -5 to +4 and synthesized the corresponding peptides again on an additional peptide array to validate the observations. As positive control H3K36 (Figure 2A; spot A1 and C9) and as negative control H3K36A (Figure 2A; spot A2 and C10) was added (Supplementary Table 1). After methylation with NSD2, direct comparison of the methylation signals was performed which reproduced the previous findings in most places. The red circles indicate mutations, which led to a strong increase in methylation activity and which were used for the next step of super-substrate development. With an additional array all possible double, triple, quadruple and quintuple combinations of these amino acid mutations were

further analyzed (Figure 2B and Supplementary Table 2), again using H3K36 (Figure 2B; spot A1 and A3) and H3K36A (Figure 2B; spot A2 and A4) as positive and negative controls. From this array the sequences of the most strongly methylated spots (marked with red circles) were synthesized on an extra array in the first line together with the corresponding K-to-A mutant sequences in the second line (Figure 2C and Supplementary Table Nr.3). The methylation data clearly showed that all sequences were much more preferred than the H3K36 sequence in position A1. For all sequences loss of methylation was observed in the K-to-A mutants indicating that methylation occurred at K36. The best and strongest signal we detected for spot A8 (marked with blue circle), which from now on will be designed as super-substrate ssK36 for NSD2. It contains 4 amino acids exchanges when compared with H3K36: A31K, K37R, H38N and R39N. As the H3 peptide contains P at position 2 and the end, we were concerned that this may cause artefacts. Hence, in some design steps, P30 and P43 were replaced by G and S, but this change does not affect the methylation levels (Supplementary Figure 3A). Since our lab already designed an H3K36 super-substrate for SETD2 (P. Schnee et al., 2022; Schuhmacher et al., 2020), we were interested to see if NSD2 is specific for the ssK36 designed here or if it can also methylate the ssK36 developed for SETD2. Therefore, a peptide array methylation experiment (Supplementary Figure 4) was performed where the methylation of the NSD2 specific ssK36 (spot B1) and SETD2 ssK36 (spot B3), together with its K-to-A mutants (spot B2 and B4, respectively) were directly compared on one array. As observed in Supplementary Figure 4, methylation of ssK36 (NSD2) was much stronger compared to the wildtype H3K36 sequence spotted at position A1 and A3. Moreover, the methylation data clearly demonstrated that NSD2 only methylates the ssK36 specially designed for NSD2, while no methylation was observed for the SETD2 specific ssK36.

Based on the successful design of an NSD2 specific ssK36 at peptide level, we next intended to investigate its methylation at protein level. For this, GST-tagged H3K36 (29-43) and ssK36 (29-43) were overexpressed and purified. For the protein methylation assay, equal protein amounts were incubated with NSD2 in the presence of radioactively labelled AdoMet separated by SDS-PAGE and the methylation was analyzed by autoradiography. As shown in Figure 2D, a strong methylation signal of the GST-tagged ssK36 was detected, but no signal for the wildtype H3K36 protein, even not after massive increase of image sensitivity (Supplementary Figure 5). By comparison of the intensities of the strong K36ss methylation signal and the weak NSD2 automethylation signal in the presence of H3K36, and considering the dynamic range of the analysis, we estimate that H3K36 is methylated at least 100-fold weaker than ssK36. This confirms that NSD2 strongly prefers the NSD2 super-substrate for methylation also at protein level. Moreover, we noticed, that NSD2 showed a much stronger automethylation signal in the presence of ssK36 than in the presence of H3K36, which provides interesting mechanistic clues that are described in the discussion section.

#### *MD Simulation of NSD2-peptide complexes*

Methylation experiments showed that NSD2 has a >100-fold higher activity towards the artificially designed super-substrate peptide ssK36 compared to the canonical H3K36 peptide. We next applied molecular dynamics (MD) simulations to reveal the molecular mechanism behind this massive increase in NSD2 activity. For this, we used NSD2 complexed with the H3 histone tail from PDB 7CRO as the starting point. The peptide in this complex was elongated using PDB 5V21 as a structural template. By mutating the H3K36 peptide sequence, the NSD2-ssK36 complex was modelled. The target K36 was manually deprotonated as required for the  $S_N2$  mechanism (Poulin et al., 2016; X. Zhang & T. C. Bruice, 2008). NSD2 in complex with ssK36 and H3K36 was subjected to MD simulations in 50 replicates à 100 ns and frames were recorded every 20 ps (Figure 3A). In order to define criteria describing the

likelihood of a methyl group transfer during the MD simulation and by this approximate the enzymatic NSD2 activity, the geometric requirements for a transition state (TS)-like conformation were applied, which were derived from the known  $S_N2$  geometry of methyl group transfer reactions (Supplementary Figure 6) (Khella et al., 2023b; P. Schnee et al., 2022). Strikingly, the complex of NSD2-ssK36 established significantly more TS-like conformations than the complex of NSD2 with H3K36 (Figure 3B). The difference between the two peptides was even more pronounced after sorting the simulation replicates into bins. In 36 out of 50 replicates the ssK36 peptides established >1000 TS-like conformation frames (5000 frames in total per replicate) indicating that it had adopted a very active conformation in which the TS-like state was reached frequently. This state was observed only once in the case of H3K36 (Figure 3C). Conversely, the majority (41 replicates) of H3K36 simulations had 0-250 TS-like conformation frames, whereas only 4 of the ssK36 simulation replicates had such a low number of TS-like conformation frames. These observations indicate that the 4 amino acid exchanges in ssK36 cause a much better stabilization of TS-like conformations of the target lysine together with AdoMet, because it can adopt a hyperactive conformation. Comparison of exemplary snapshots of complex structures revealed that the ssK36 peptide tends to bend towards the active site of NSD2 and the bound AdoMet, a structural change that was not observed with the H3K36 peptide (Figure 3A). MD simulation of peptides with sequences as were used for the peptide methylation experiments (P30G, P43S double mutant) were conducted. Analysis of the simulations agreed with simulation data with canonical peptide sequences. All PDB files and MD simulation protocols used in this study and MD simulation results using the peptide sequence as used for the peptide array are deposited at DaRUS (<https://doi.org/10.18419/darus-3263>)

#### *Contact analysis of H3K36 and ssK36-NSD2 complexes*

To investigate the molecular mechanisms behind the better stabilization of the TS-like conformations, a contact map of the established contacts between peptide and protein during the simulation was prepared using contact map explorer (Swenson & Roet, 2017). The analysis was based on distance criteria, and contacts were considered if the distance of a pair of heavy atoms from the peptide and an NSD2 residue was below 4.5 Å. The fraction of time in which a contact was established was measured and a contact profile created. The resulting contact maps for H3K36 and ssK36 were contrasted and differences extracted (Figure 4). The largest deviations between the two peptides were found at three of the four mutated residues. The A31K mutation in ssK36 causes the peptide residues P30-T32 to strongly interact with NSD2 residues H1110-F1117 and E1187-T1189. The K31 side chain amino group is solvent exposed and interacts with the T1116 hydroxy group (Figure 5). This interaction likely triggers a bending of the N-terminal part of the ssK36 peptide (A29-G33) towards the active site and AdoMet. In contrast, this part of the H3K36 peptide makes contacts with I1106 and L1184-N1186, which keeps the peptide in a straight orientation. The second large difference between the contact maps was observed at peptide positions 37-39. The larger side chain of R37 enables contacts primarily with the E1216 side chain and K1220, K1221 backbone atoms (Figure 5). In contrast, H39 in H3K36 is positioned in a pocket surrounded by NSD2 residues P1146-T1150, while N39 points outwards into to solvent leading to the loss of this interaction (Figure 5). Moreover, a higher contact frequency of ssK36 P38 with NSD2 residues F1177-Y1179 next to the H39 pocket is observed. P38 in H3K36 is oriented differently than in ssK36, potentially influencing the structure of the backbone atoms in the peptide. Hence, for K31, R37 and P38 new contacts are established in ssK36, potentially stabilizing the TS-like conformation. Conversely, our data suggest that the H39 contact to NSD2 residues P1146-T1150 is unfavorable for TS-like conformations and loss of this interaction stimulates catalytic activity.



As a consequence of these altered interactions, the mutated residues change the conformation of the ssK36 peptide in the NSD2 binding cleft. Especially at the N-terminus, ssK36 bends towards the AdoMet, whereas H3K36 stays straight (Figure 3A and 5C). Due to this conformational change, K31 can also interact with the hydroxyl group of the AdoMet sugar moiety. Moreover, the SET-I loop (D1114–Y1119), which is suspected to contribute to the substrate specificity of PKMTs (R. Marmorstein, 2003; Qian & Zhou, 2006), changes its conformation towards AdoMet and H1116 interacts with the AdoMet sugar moiety (Figure 5C). Combined with K31, these interactions could stabilize AdoMet in the cofactor binding pocket, bridge between AdoMet and the bound ssK36 peptide and hence increase the probability of a methyl group transfer. Again, simulations with peptide sequences as were used for the peptide methylation experiments (P30G, P43S double mutant) agreed with generated contact maps. All MD simulation analysis scripts used in this study and MD simulation results using the peptide sequence as used for the peptide array are deposited at DaRUS (<https://doi.org/10.18419/darus-3263>)

#### *Identification and methylation of non-histone peptide substrates by NSD2*

Next, we were interested to use the newly derived specificity profile of NSD2 to identify novel substrates of NSD2. To this end, a ScanSite search (Obenauer et al., 2003) with the specificity sequence motif was performed to identify human proteins that contain a sequence matching the sequence specificity profile. In order to cover most possible substrates, the search profile was expanded at this stage, also including less preferred but still allowed residues at several positions (Table 1):

Site	-1	K	+1	+2	+3	+4
<i>Motif</i>	ILV	K	KRVQNI	VILP	NGLSFTM IHQAEK	LNQGHK MF

Table 1: Motif used for the search for novel NSD2 non-histone substrates.

Since NSD2 is mainly localized in the nucleus ([www.proteinatlas.org/ENSG00000109685-NSD2](http://www.proteinatlas.org/ENSG00000109685-NSD2)), the search was restricted to nuclear proteins, which led to the identification of 226 substrate candidates. A peptide array with 15 aa long peptides of each candidate with the target lysine centered in the middle was synthesized using the SPOT synthesis method (Figure 6A, Supplementary Table Nr.4). As positive controls H3K36 (aa 29-43) (spot A1) and H4K44 (spot A3) were included and the corresponding K-to-A mutations served as negative controls (spot A2 and A3, respectively). The peptide array was incubated with NSD2 using radioactively labeled AdoMet as cofactor and the methyl transfer was detected by autoradiography revealing numerous methylation signals. Depending on the strength of the methylation signal and the biological relevance of the target protein, 25 peptides were selected and further analyzed by additional peptide arrays to determine if the predicted target lysine is the site of modification. For this, pairs of the 15 aa long WT sequences and the respective K-to-A mutants were prepared and incubated with NSD2 in presence of radioactively labelled AdoMet (Figure 6B, Supplementary Table 5). Methylation was validated in all of the peptides, and activity at the target lysine was confirmed in all but 3 of the candidate substrates which showed no loss of methylation in the K-to-A mutant indicating that the methylation did not occur at the predicted target lysine in these peptides. All three peptides have additional lysine residues, which may be methylated by NSD2. To confirm this, additional experiments would be required. In summary, 22 new peptide substrates of NSD2 with methylation at the predicted target lysine were discovered.

### *Investigation of the methylation of non-histone protein substrates by NSD2*

It is possible that some of the methylated peptide substrates are not methylated at protein level, because the NSD2 target site is rich in hydrophobic residues which might be buried in folded proteins making them inaccessible. For this reason, we cloned the 22 newly identified non-histone substrate candidates with GST-tag and expressed them in *E. coli*, and finally 17 of them were successfully purified. Roughly comparable amounts of the purified proteins were verified by SDS-PAGE and Western Blot using an anti-GST antibody (Supplementary Figure 7) and used for subsequent protein methylation assays. After incubation of the substrate proteins with NSD2 and radioactively labelled AdoMet, the samples were separated by SDS-PAGE and methyl transfer was detected by autoradiography. For two of the targets, ATRX and FANCM, a methylation signal was detected and loss of methylation of the corresponding K-to-R mutants (Figure 6C), which validated the methylation at the target lysine residues K1033 of ATRX and K819 FANCM which were identified in the peptide methylation experiments.

To analyze methylation of ATRX and FANCM by NSD2 at cellular level, we aimed to apply an H3K36me1 antibody, because the methylation sites in ATRX and FANCM have a similar sequence context as H3K36. To investigate if the anti-H3K36me1 antibody was also able to detect the methylation of the novel NSD2 substrates ATRX and FANCM, *in vitro* protein methylation reactions were performed by incubating either ATRX or FANCM with or without NSD2 enzyme in methylation buffer containing unlabeled AdoMet. The samples were separated by SDS-PAGE, transferred to nitrocellulose membranes and the membrane was probed with the anti-H3K36me1 antibody (Supplementary Figure 8). A clear signal was observed for the methylated FANCM protein, only when NSD2 was present. For ATRX, the H3K36me1 antibody was not able to discriminate perfectly between methylated and unmethylated proteins, but the signal for the methylated ATRX protein was much stronger than for the unmethylated protein. Additionally, the automethylated NSD2 was stained by the antibody, demonstrating that it has a rather broad Kme1 reactivity.

In the next step, ATRX and FANCM were cloned into the mammalian expression vector pEYFP-C1 and the full-length NSD2 enzyme into the mammalian expression vector pECFP-C1 and expression of all proteins was validated after transfection of the plasmids into HEK293 cells by Western blot using antibodies against the fluorophore tags (Supplementary Figure 9). For methylation analysis, HEK293 cells were cotransfected with two plasmids encoding CFP-tagged NSD2 enzyme and YFP-tagged ATRX or FANCM. As a control, HEK293 cells were only transfected with one of the substrate protein expression plasmids without the NSD2 plasmid. After isolation of the YFP-tagged substrate proteins by GFP trap, equal loading of methylated and unmethylated sample was analyzed by Western Blot using an anti-GFP antibody (Figure 6D). The cellular methylation of ATRX and FANCM was investigated using the previously verified H3K36me1 antibody. The results showed strong methylation of ATRX and FANCM, which were isolated from cells cotransfected with NSD2, whereas no methylation signal was observed without NSD2 cotransfection. This result indicates that NSD2 is the responsible methyltransferase for ATRX and FANCM methylation in HEK293 cells establishing both proteins as novel non-histone substrates of NSD2.

### **Discussion**

Investigation of the substrate specificity of PKMTs has important implications by providing details about the mechanism of peptide interaction and catalysis and allowing for unbiased searchers for

novel PKMT substrates. In this work, we have investigated the substrate specificity of the NSD2 PKMT, which is a well-established H3K36 mono- and dimethyltransferase with important roles in chromatin regulation, cell physiology and cancer. Our data revealed a highly specific interaction of NSD2 with the amino acid residues between the -3 and +2 sites surrounding the K36 target lysine. Interestingly, we observed at several positions of the H3K36 sequence that amino acid residues, differing from the natural ones in the H3 protein, were preferred by NSD2. A combination of four of these preferred residues led to the design of a super-substrate (ssK36) that was methylated >100 time more strongly both at peptide level and as GST-tagged protein. MD simulations revealed that the super-substrate peptide adopts a bend conformation in complex with NSD2 that may help K36 to approach the AdoMet more efficiently. This conformational change was driven by novel contacts established between the ssK36 and NSD2 which support catalysis and loss of unfavorable contacts engaged by H3K36. Interestingly, automethylation of NSD2 was reduced with H3K36 when compared to ssK36. This observation suggests that H3K36 binds to the active site of NSD2, but it stays in a catalytically incompetent conformation, which blocks the active site and reduces automethylation.

It is interesting to consider that the natural H3K36 methylation by NSD2 takes place in the nucleosomal context, where the histone tail binding occurs together with interactions of the enzyme with the other parts of the nucleosome and the linker DNA. Moreover, the H3 tail must be partially lifted up in order to make K36 accessible for methylation. This may require special conformational adaptations that are not ideal for the interaction with the isolated peptide, explaining why alternative peptide sequences can be better substrates. This mechanism could explain, why a super-substrate could also be designed for SETD2, another H3K36 methyltransferase (Khella et al., 2023b; Schuhmacher et al., 2020). More systematic studies will be required to find out, if super-substrate can also be designed for other PKMTs that do not normally methylate H3K36.

The role of NSD2 in the Wolf-Hirschhorn syndrome and H3K36 methylation implicated a direct connection to DNA repair. This conjecture was further validated when it was discovered that the interactome of NSD2 comprises many factors involved in the DNA repair including Poly(ADP-ribose) polymerase 1 (PARP1) (Huang et al., 2019). Subsequently, it was found that PARylation reduces NSD2 histone methyltransferase activity and impedes its chromatin binding (Huang et al., 2019). Different non-histone substrates of NSD2 have important role in DNA damage repair such as PTEN (Zhang et al., 2019) and Aurora kinase A (AURKA) (Park et al., 2018) and other studies showed that NSD2 enhances DNA damage repair leading to an increase in resistance to chemotherapeutic agents (Shah et al., 2016). This connection between NSD2 and DNA damage repair is further enhanced by our finding that ATRX and FANCM are direct targets of NSD2. These two proteins are helicases with important roles in DNA repair and R-loop metabolism (S. Yang et al., 2023). Future work will be needed to unravel the effects of NSD2 methylation on these factors and its biological implications.

## **Methods**

### *Cloning, expression and purification of proteins*

The DNA sequence encoding for the human NSD2 enzyme (aa 991-1240; Swiss Prot No. O96028) and the putative human substrate protein domains were amplified by PCR using cDNA isolated from HEK293 cells. Protein domains of the non-histone substrates were designed with the Scooby domain prediction tool (<http://www.ibi.vu.nl/programs/scoobywww/>) (George et al., 2005). All constructs were cloned into the pGex-6p-2 expression vector as GST-fusion proteins. The K-to-R mutations of the non-histone substrates were introduced using a megaprimer PCR mutagenesis method (Jeltsch &

Lanio, 2002). The ssK36 (29-43)-GST construct was cloned by site-directed mutagenesis using H3K36 (29-43)-GST plasmid as template taken from Schuhmacher et. al (2020) (Schuhmacher et al., 2020). For mammalian expression the coding sequence of the full-length NSD2 (Swiss Prot No. O96028) was cloned into the pECFP-C1 (Clontech, USA). The protein domains encoding for ATRX (aa 893-1188; Swiss Prot No. P46100) and FANCM (aa 723-933; Swiss Prot No. Q8IYD8) were cloned into the pEYFP-C1 vector (Clontech, USA). All cloning steps were confirmed by Sanger sequencing.

For protein overexpression, the plasmids were transformed into *E. coli* BL21-CodonPlus (DE3) cells (Novagen, USA) which were grown in LB medium at 37 °C until an OD<sub>600</sub> of 0.6 to 0.8 was reached. The culture was then shifted to 20 °C overnight (14 to 16 h) and protein expression was induced with 1 mM isopropyl-D-thiogalactopyranoside (IPTG). Afterward, the cells were harvested by centrifugation at 4,500 rpm, washed once with STE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA and 100 mM NaCl) and the cell pellet was stored at -20 °C. For purification, the cell pellet was thawed on ice, resuspended in sonication buffer (50 mM Tris/HCl pH 7, 150 mM NaCl, 1 mM DTT, 5 % (w/v) glycerol) supplemented with protease inhibitor cocktail and lysed by sonication (14 rounds, 30 % power, 4°C). Thereafter, the sample was centrifuged at 18,000 rpm for 90 min and the supernatant was loaded onto a Glutathione Sepharose 4B resin (GE Healthcare) column, which was pre-equilibrated with sonication buffer. Afterward, the beads were washed once with sonication buffer and twice with washing buffer (50 mM Tris/HCl pH 8, 500 mM NaCl, 1 mM DTT, 5 % (w/v) glycerol). Subsequently, the bound proteins were eluted with elution buffer (40 mM reduced glutathione, 50 mM Tris/HCl pH 8, 500 mM NaCl, 1 mM DTT, 5 % (w/v) Glycerol) and then dialyzed against low glycerol dialysis buffer 1 (20 mM Tris/HCl pH 7.4, 100 mM KCl, 0.5 mM DTT, 10 % (w/v) glycerol) for 3 h and afterwards over night against high glycerol dialysis buffer 2 (20 mM Tris/HCl pH 7.4, 100 mM KCl, 0,5 mM DTT, 60 % (w/v) glycerol). The purified proteins were analyzed by sodium-dodecyl-sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

#### *Peptide array synthesis*

Peptide arrays, containing 15 aa long peptides, were synthesized on cellulose membrane using the SPOT synthesis method (Frank, 2002) with an Autospot Multiprep synthesizer (Intavis AG). Each spot contained approximately 9 nmol peptide (Autospot Reference Handbook, Intavis AG) and the successful synthesis of the peptides on the cellulose membrane was qualitatively confirmed by bromophenol blue staining (Kudithipudi, Kusevic, et al., 2014; S. Weirich & A. Jeltsch, 2022). Data analysis and derivation of sequence motif and discrimination factor calculation was performed as described previously (Kudithipudi, Kusevic, et al., 2014; S. Weirich & A. Jeltsch, 2022).

#### *Peptide array methylation*

All peptide arrays were preincubated in methylation buffer containing 50 mM Tris pH 8.5, 50 mM NaCl and 0.5 mM DTT for 5 min at room temperature. Then, the membranes were incubated in methylation buffer supplemented with 260 nM NSD2 and 0.76 µM labelled [methyl-<sup>3</sup>H]-AdoMet (Perkin Elmer Inc., dissolved at 25 µM in 10 mM sulfuric acid) for 60 min at 25 °C on a shaker. Afterwards, the arrays were washed five times with 100 mM NH<sub>4</sub>HCO<sub>3</sub> and 1 % SDS, followed by incubation for 5 min in Amplify NAMP100V solution (GE Healthcare). The membranes were exposed to Hyperfilm™ high performance autoradiography films (GE Healthcare) in the dark at -80°C. Film development was performed with an Optimus TR developing machine.

### *Protein methylation assay*

The methylation of non-histone substrate proteins and recombinant H3.1 (NEB) was performed in methylation buffer (50 mM Tris pH 8.5, 50 mM NaCl and 0.5 mM DTT) supplemented with 3  $\mu$ M NSD2 and 0.76  $\mu$ M labelled [methyl- $^3$ H]-AdoMet (Perkin Elmer Inc., dissolved at 25  $\mu$ M in 10 mM sulfuric acid) for 4 h at 25°C. For methylation of the H3K36 (29-43)-GST and ssK36 (29-43)-GST proteins, 420 nM NSD2 was added and the mixture incubated over night at 25°C. The methylation reaction was stopped by the addition of SDS loading buffer and boiling for 5 min at 95°C. Equal amounts of target protein was confirmed by Coomassie Brilliant Blue staining and Western Blot using as primary antibody anti-GST (GE Healthcare, 27457701V). The methylated samples were separated by 12 % SDS-PAGE. Then, the SDS gel was incubated for 1 h in Amplify NAMP100V (Ge Healthcare) and dried for 90 min at 65°C under vacuum. The dried SDS gel was exposed to Hyperfilm<sup>TM</sup> high performance autoradiography films (GE Healthcare) in the dark at -80°C. Film development was performed with an Optimus TR developing machine.

### *Cell culture, transfection and immunoprecipitation*

HEK293 cells were grown in Dulbecco's Modified Eagle's Medium (Sigma) supplemented with 5 % fetal bovine serum, penicillin/streptomycin, and L-glutamine (Sigma) in an incubator providing 37°C and 5 % CO<sub>2</sub>. The pECFP-C1 tagged full-length NSD2 was co-transfected with pEYFP-C1 fused ATRX or FANCM using polyethylenimine (Polyscience, USA; according to manufacturer's instructions). 72 h after transfection, the cells were washed with PBS buffer and harvested by centrifugation at 500 g for 5 min. For methylation analysis the YFP-fused ATRX and FANCM substrate proteins were immunoprecipitated from cell extract using GFP-Trap<sup>®</sup> A (Chromotek) following the manufacturer's instructions. The samples were heated to 95°C for 5 min in SDS-gel loading buffer and resolved by 16 % SDS-PAGE. Analysis was performed by Western Blot using as primary antibody H3K36me1 (Abcam, UK; Cat. No: ab9048) or GFP antibody (Clontech, lot. 1404005).

### *MD Simulation of the NSD2-peptide complexes and trajectory analysis*

All-atom MD simulations were conducted similarly as described recently (Khella et al., 2023b; P. Schnee et al., 2022). In brief, the structure of human NSD2 (positions Y991–K1220) was modeled based on the cryo-EM structure of NSD2 E1099K, T1150A in complex with its nucleosome substrate (PDB: 7CRO) (Li et al., 2021). Reverting mutations of K1099E and A1150T were modeled using PyMOL (2.4.1) (L. Schrödinger, 2015). The missing part of the post-SET loop (positions P1206–K1220) was modeled based on the SET domain of SETD2 (PDB: 5V21) (Zhang et al., 2017) using PyMOD 3.0 (Janson & Paiardini, 2021), since no structure of NSD2 complexed with the H3K36 peptide and post-SET loop has been resolved. The histone tail of PDB 7CRO was replaced by the 15 aa long H3K36M peptide from PDB 5V21, and methionine 36 mutated to lysine. The H3K36 peptide (29-APATGGVKKPHRYRP-43) was manually mutated in PyMOL to generate the ssK36 peptide (29-APKTGGVKKRPNNYRP-43). The K36 was manually deprotonated as required for the S<sub>N</sub>2 mechanism (Poulin et al., 2016; X. Zhang & T. C. Bruice, 2008). AdoMet was modeled based on the coordinates in PDB 7CRO and parametrized using ANTECHAMBER from AmberTools (18.0) (Wang et al., 2001). The Zn<sup>2+</sup> ions were modeled using the cationic dummy atom method (Oelschlaeger, 2003; Pang, 1999). Cysteines 1016, 1018, 1026, 1032, 1041, 1046, 1052, 1145, 1192, 1194 and 1199 were treated as unprotonated to ensure proper Zn<sup>2+</sup> binding (Cheng & Zhang, 2007). The protein charge was neutralized and an ionic strength of 0.1 M NaCl was applied, by adding 30 Na<sup>+</sup> and 27 Cl<sup>-</sup> ions. Total number of atoms and water molecules for the

NSD2-H3K36-system: 61,128 and 10,000; for NSD2-ssK36-system: 62,503 and 10,000 . To equilibrate the solvent, a 5 ns pressure coupled equilibration with Monte Carlo barostat (Faller & de Pablo, 2002) was performed at a pressure of 1 atm. Initially, the C-alpha (C $\alpha$ ) atoms of NSD2, the peptide, and the AdoMet atoms were restrained with a force constant of 100 and 5 kJ mol<sup>-1</sup> Å<sup>-2</sup>, respectively. The restraints were removed successively, starting with the NSD2 C $\alpha$  restraints, followed by a 5 ns equilibration with the peptide and AdoMet still being restrained. Subsequently, the AdoMet and peptide restraints were removed as well, followed by 5 ns equilibration with no restraints. For production, MD simulations were conducted in 50 replicates à 100 ns (total simulation time 5  $\mu$ s).

In order to define criteria describing a catalytically competent conformation, the following geometric requirements for a transition state (TS)-like conformation were derived from the known S<sub>N</sub>2 geometry of methyl group transfer reaction (P. Schnee et al., 2022) (Supplementary Figure 6).

- (1) The distance between the lysine N $\epsilon$  and AdoMet methyl group C-atom is <4 Å.
- (2) The angle between the lysine N $\epsilon$ , the lysine C $\delta$  bond and the virtual bond between lysine N $\epsilon$  and the AdoMet methyl group C-atom is in a range of 109°  $\pm$  30°.
- (3) The angle between the lysine N $\epsilon$ , the AdoMet methyl group C-atom and AdoMet S-atom bonds is in a range of 180°  $\pm$  30°.

Data analysis was performed utilizing MDTraj (1.9.4) (McGibbon et al., 2015) to calculate the distances and angles necessary for the geometric criteria of an S<sub>N</sub>2 TS-like conformation. All structures were visualized using PyMOL (2.4.1). The contact map analysis was performed utilizing contact-map explorer (0.7.1) (Swenson & Roet, 2017). For the contact maps a cut-off of 4.5 Å was used for the analysis. A contact was counted if at least one heteroatom of a residue was in a 4.5 Å<sup>3</sup> sphere surrounding one heteroatom from another residue excluding neighboring residues.

#### *Data availability*

All biochemical data generated or analyzed during this study are included in the published article and its supplementary files. Modelled structures of NSD2 bound to different peptides and cofactors, source data of the results of the MD analysis, MD simulations codes and analysis scripts are provided on DaRUS (<https://doi.org/10.18419/darus-3263>).

#### *Supporting information*

This article contains supporting information.

#### *Author contributions*

S.W. and A.J. devised the study. D.K. and S.W. conducted the biochemical experiments. P.S. conducted the MD simulations with the help of J.R. A.J. supervised the work. J.P. contributed to the supervision of the MD simulations. S.W., D.K., P.S., and A.J. prepared the figures. S.W., P.S., and A.J. wrote the manuscript draft. All authors were involved in data analysis and interpretation. The final manuscript was approved by all authors.

### *Funding*

This work was supported by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy EXC 2075 390740016 in PN2-5.

### *Conflict of interest*

The authors declare that they have no conflicts of interest with the contents of this article.

### **References**

- 1 Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17**, 487-500 (2016).
- 2 Jambhekar, A., Dhall, A. & Shi, Y. Roles and regulation of histone methylation in animal development. *Nat Rev Mol Cell Biol* **20**, 625-641 (2019).
- 3 Zhao, S., Allis, C. D. & Wang, G. G. The language of chromatin modification in human cancers. *Nat Rev Cancer* **21**, 413-430 (2021).
- 4 Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chemical reviews* **115**, 2376-2418 (2015).
- 5 Boriack-Sjodin, P. A. & Swinger, K. K. Protein Methyltransferases: A Distinct, Diverse, and Dynamic Family of Enzymes. *Biochemistry* **55**, 1557-1569 (2016).
- 6 Husmann, D. & Gozani, O. Histone lysine methyltransferases in biology and disease. *Nature structural & molecular biology* **26**, 880-889 (2019).
- 7 Patel, D. J. & Wang, Z. Readout of epigenetic modifications. *Annu Rev Biochem* **82**, 81-118 (2013).
- 8 Zhang, X., Huang, Y. & Shi, X. Emerging roles of lysine methylation on non-histone proteins. *Cellular and molecular life sciences : CMLS* **72**, 4257-4272 (2015).
- 9 Biggar, K. K. & Li, S. S. Non-histone protein methylation as a regulator of cellular signalling and function. *Nat Rev Mol Cell Biol* **16**, 5-17 (2015).
- 10 Cornett, E. M., Ferry, L., Defossez, P. A. & Rothbart, S. B. Lysine Methylation Regulators Moonlighting outside the Epigenome. *Molecular cell* **75**, 1092-1101 (2019).
- 11 Kudithipudi, S. & Jeltsch, A. Approaches and Guidelines for the Identification of Novel Substrates of Protein Lysine Methyltransferases. *Cell Chem Biol* **23**, 1049-1055 (2016).
- 12 Bennett, R. L., Swaroop, A., Troche, C. & Licht, J. D. The Role of Nuclear Receptor-Binding SET Domain Family Histone Lysine Methyltransferases in Cancer. *Cold Spring Harb Perspect Med* **7** (2017).
- 13 Li, J., Ahn, J. H. & Wang, G. G. Understanding histone H3 lysine 36 methylation and its deregulation in disease. *Cellular and molecular life sciences : CMLS* **76**, 2899-2916 (2019).
- 14 Lam, U. T. F., Tan, B. K. Y., Poh, J. J. X. & Chen, E. S. Structural and functional specificity of H3K36 methylation. *Epigenetics & chromatin* **15**, 17 (2022).
- 15 Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol* **13**, 115-126 (2012).
- 16 Sankaran, S. M., Wilkinson, A. W., Elias, J. E. & Gozani, O. A PWWP Domain of Histone-Lysine N-Methyltransferase NSD2 Binds to Dimethylated Lys-36 of Histone H3 and Regulates NSD2 Function at Chromatin. *J Biol Chem* **291**, 8465-8474 (2016).

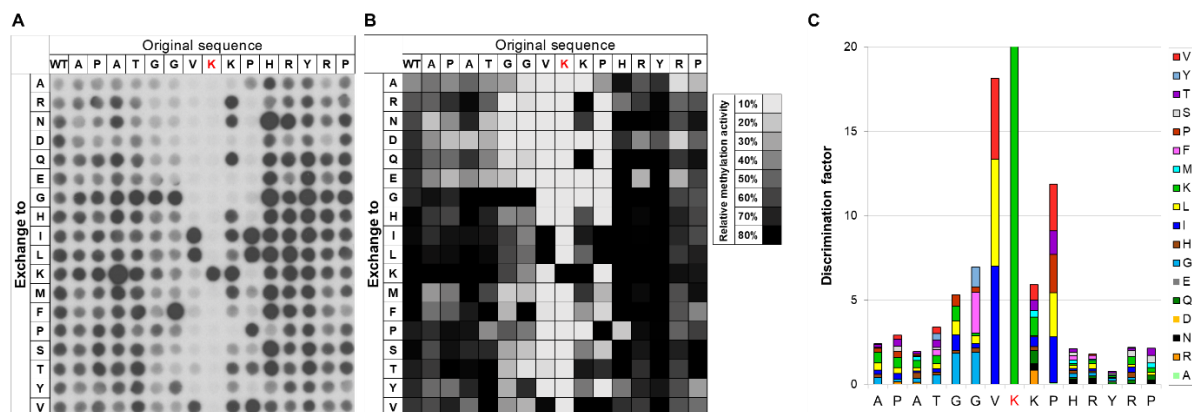
- 17 Kang, H. B. *et al.* The histone methyltransferase, NSD2, enhances androgen receptor-mediated transcription. *FEBS Lett* **583**, 1880-1886 (2009).
- 18 Stec, I. *et al.* WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphia gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum Mol Genet* **7**, 1071-1082 (1998).
- 19 Nimura, K. *et al.* A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome. *Nature* **460**, 287-291 (2009).
- 20 Hajdu, I., Ciccia, A., Lewis, S. M. & Elledge, S. J. Wolf-Hirschhorn syndrome candidate 1 is involved in the cellular response to DNA damage. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 13130-13134 (2011).
- 21 Kim, J. Y. *et al.* Multiple-myeloma-related WHSC1/MMSET isoform RE-IIBP is a histone methyltransferase with transcriptional repression activity. *Mol Cell Biol* **28**, 2023-2034 (2008).
- 22 Marango, J. *et al.* The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor. *Blood* **111**, 3145-3154 (2008).
- 23 Pei, H. *et al.* MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. *Nature* **470**, 124-128 (2011).
- 24 Li, Y. *et al.* The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J Biol Chem* **284**, 34283-34295 (2009).
- 25 Song, D. *et al.* NSD2 promotes tumor angiogenesis through methylating and activating STAT3 protein. *Oncogene* **40**, 2952-2967 (2021).
- 26 Park, J. W., Chae, Y. C., Kim, J. Y., Oh, H. & Seo, S. B. Methylation of Aurora kinase A by MMSET reduces p53 stability and regulates cell proliferation and apoptosis. *Oncogene* **37**, 6212-6224 (2018).
- 27 Zhang, J. *et al.* PTEN Methylation by NSD2 Controls Cellular Sensitivity to DNA Damage. *Cancer Discov* **9**, 1306-1323 (2019).
- 28 Kudithipudi, S. & Jeltsch, A. Role of somatic cancer mutations in human protein lysine methyltransferases. *Biochimica et biophysica acta* **1846**, 366-379 (2014).
- 29 Oyer, J. A. *et al.* Point mutation E1099K in MMSET/NSD2 enhances its methyltransferase activity and leads to altered global chromatin methylation in lymphoid malignancies. *Leukemia* **28**, 198-201 (2014).
- 30 Sengupta, D. *et al.* NSD2 dimethylation at H3K36 promotes lung adenocarcinoma pathogenesis. *Molecular cell* **81**, 4481-4492.e4489 (2021).
- 31 Sato, K. *et al.* Structural basis of the regulation of the normal and oncogenic methylation of nucleosomal histone H3 Lys36 by NSD2. *Nat Commun* **12**, 6605 (2021).
- 32 Khella, M. S. *et al.* The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation. *J Biol Chem* **299**, 104796 (2023).
- 33 Weirich, S. & Jeltsch, A. Specificity Analysis of Protein Methyltransferases and Discovery of Novel Substrates Using SPOT Peptide Arrays. *Methods in molecular biology (Clifton, N.J.)* **2529**, 313-325 (2022).
- 34 Kudithipudi, S., Kusevic, D., Weirich, S. & Jeltsch, A. Specificity analysis of protein lysine methyltransferases using SPOT peptide arrays. *Journal of visualized experiments : JoVE*, e52203 (2014).
- 35 Kudithipudi, S., Lungu, C., Rathert, P., Happel, N. & Jeltsch, A. Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1. *Chemistry & biology* **21**, 226-237 (2014).
- 36 Weil, L. E. *et al.* Oligomerization and Auto-methylation of the Human Lysine Methyltransferase SETD6. *J Mol Biol* **430**, 4359-4368 (2018).



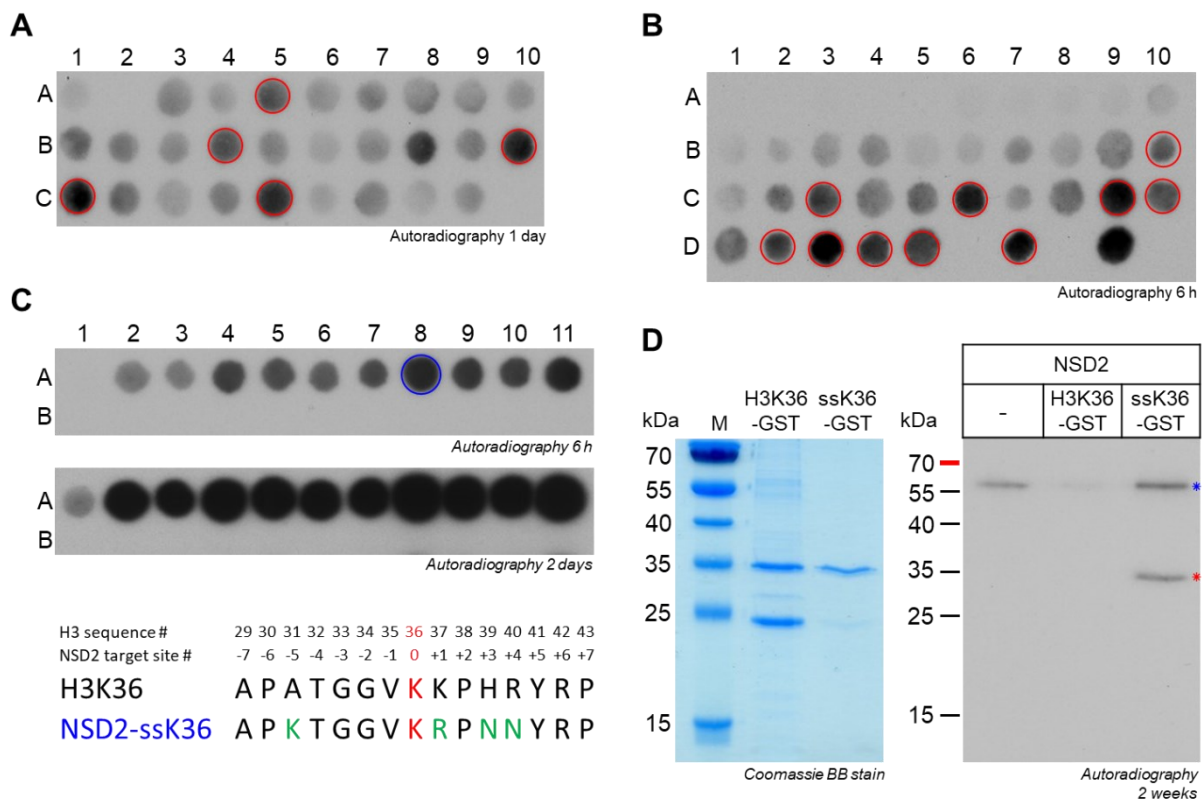
- 37 Piao, L. *et al.* Automethylation of SUV39H2, an oncogenic histone lysine methyltransferase, regulates its binding affinity to substrate proteins. *Oncotarget* **7**, 22846-22856 (2016).
- 38 Iglesias, N. *et al.* Automethylation-induced conformational switch in Clr4 (Suv39h) maintains epigenetic stability. *Nature* **560**, 504-508 (2018).
- 39 Lee, C. H. *et al.* Automethylation of PRC2 promotes H3K27 methylation and is impaired in H3K27M pediatric glioma. *Genes Dev* **33**, 1428-1440 (2019).
- 40 Weirich, S., Kudithipudi, S. & Jeltsch, A. Somatic cancer mutations in the MLL1 histone methyltransferase modulate its enzymatic activity and dependence on the WDR5/RBBP5/ASH2L complex. *Mol Oncol* **11**, 373-387 (2017).
- 41 Khella, M. S., Brohm, A., Weirich, S. & Jeltsch, A. Mechanistic Insights into the Allosteric Regulation of the Clr4 Protein Lysine Methyltransferase by Autoinhibition and Automethylation. *Int J Mol Sci* **21** (2020).
- 42 Rathert, P. *et al.* Protein lysine methyltransferase G9a acts on non-histone targets. *Nat Chem Biol* **4**, 344-346 (2008).
- 43 Schuhmacher, M. K. *et al.* Sequence specificity analysis of the SETD2 protein lysine methyltransferase and discovery of a SETD2 super-substrate. *Communications biology* **3**, 511 (2020).
- 44 Schnee, P. *et al.* Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide. *Communications Chemistry* **5**, 139 (2022).
- 45 Poulin, M. B. *et al.* Transition state for the NSD2-catalyzed methylation of histone H3 lysine 36. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 1197-1201 (2016).
- 46 Zhang, X. & Bruice, T. C. Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 5728-5732 (2008).
- 47 Swenson, D. E. H. & Roet, S. Contact Map Explorer. [https://github.com/dwhswenson/contact\\_map](https://github.com/dwhswenson/contact_map). <https://contact-map.readthedocs.io/en/latest/index.html> (2017).
- 48 Marmorstein, R. Structure of SET domain proteins: a new twist on histone methylation. *Trends Biochem Sci* **28**, 59-62 (2003).
- 49 Qian, C. & Zhou, M. M. SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cellular and molecular life sciences : CMLS* **63**, 2755-2763 (2006).
- 50 Obenauer, J. C., Cantley, L. C. & Yaffe, M. B. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic acids research* **31**, 3635-3641 (2003).
- 51 Huang, X. X. *et al.* Defining the NSD2 interactome: PARP1 PARylation reduces NSD2 histone methyltransferase activity and impedes chromatin binding. *Journal of Biological Chemistry* **294**, 12459-12471 (2019).
- 52 Shah, M. Y. *et al.* MMSET/WHSC1 enhances DNA damage repair leading to an increase in resistance to chemotherapeutic agents. *Oncogene* **35**, 5905-5915 (2016).
- 53 Yang, S., Winstone, L., Mondal, S. & Wu, Y. Helicases in R-loop Formation and Resolution. *J Biol Chem* **299**, 105307 (2023).
- 54 George, R. A., Lin, K. & Heringa, J. Scooby-domain: prediction of globular domains in protein sequence. *Nucleic acids research* **33**, W160-163 (2005).
- 55 Jeltsch, A. & Lanio, T. Site-directed mutagenesis by polymerase chain reaction. *Methods in molecular biology (Clifton, N.J.)* **182**, 85-94 (2002).

- 56 Frank, R. The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports-- principles and applications. *Journal of immunological methods* **267**, 13-26 (2002).
- 57 Li, W. *et al.* Molecular basis of nucleosomal H3K36 methylation by NSD methyltransferases. *Nature* **590**, 498-503 (2021).
- 58 Schrödinger, L. The PyMOL molecular graphics system, version 1.7. 6.6. *Schrödinger LLC* (2015).
- 59 Zhang, Y. *et al.* Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. *Sci Rep* **7**, 43906 (2017).
- 60 Janson, G. & Paiardini, A. PyMod 3: a complete suite for structural bioinformatics in PyMOL. *Bioinformatics (Oxford, England)* **37**, 1471-1472 (2021).
- 61 Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc* **222** (2001).
- 62 Oelschlaeger, P. S., Rolf D.; Pleiss, Juergen. Modeling domino effects in enzymes: molecular basis of the substrate specificity of the bacterial metallo-beta-lactamases IMP-1 and IMP-6. *Biochemistry* **30**, 12 (2003).
- 63 Pang, Y.-P. Novel Zinc Protein Molecular Dynamics Simulations: Steps Toward Antiangiogenesis for Cancer Treatment. *J. Mol. Model* **5**, 7 (1999).
- 64 Pang, Y.-P. Successful Molecular Dynamics Simulation of Two Zinc Complexes Bridged by a Hydroxide in Phosphotriesterase Using the Cationic Dummy Atom Method. *PROTEINS: Structure, Function, and Genetics* **45**, 7 (2001).
- 65 Cheng, X. & Zhang, X. Structural dynamics of protein lysine methylation and demethylation. *Mutat Res* **618**, 102-115 (2007).
- 66 Faller, R. & de Pablo, J. J. Constant pressure hybrid Molecular Dynamics–Monte Carlo simulations. *The Journal of Chemical Physics* **116** (2002).
- 67 McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical journal* **109**, 1528-1532 (2015).

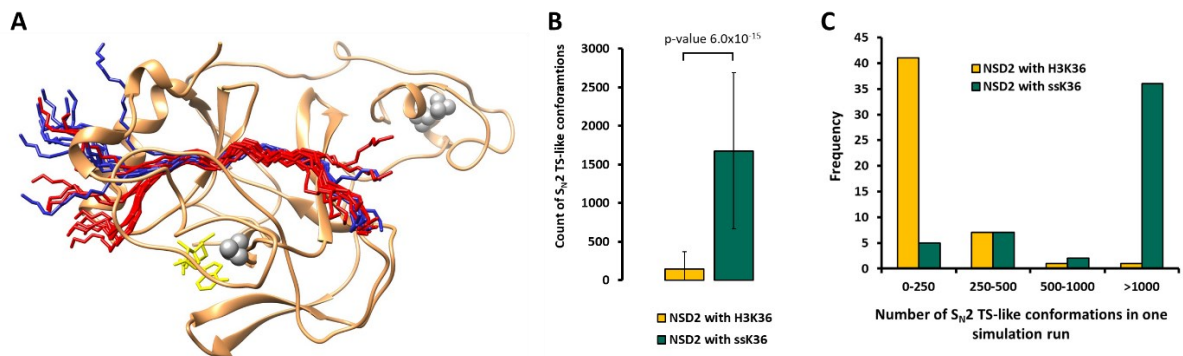
## Figures and figure legends



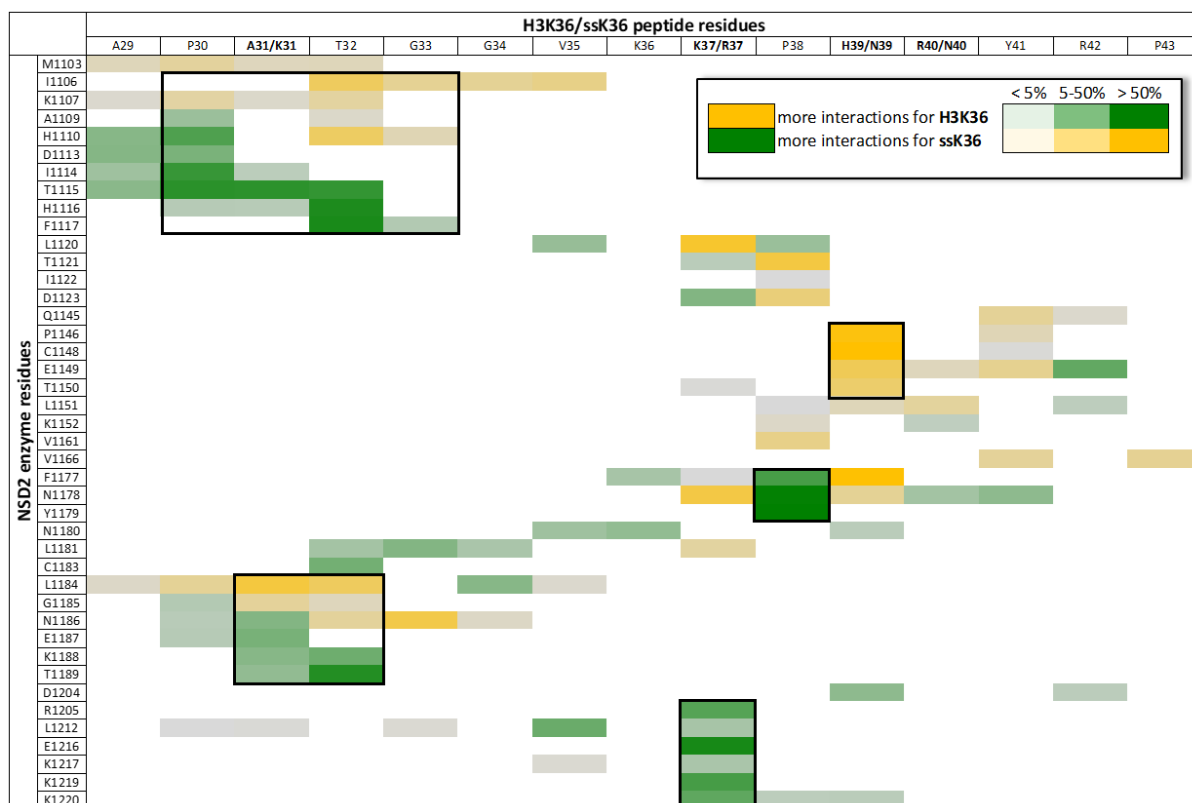
**Figure 1. Substrate specificity analysis of NSD2.** **A)** Autoradiographic image (3 days of exposure) of a peptide SPOT array methylated by NSD2 in presence of radioactively labeled AdoMet. The horizontal axis represents the H3K36 (29-43) template sequence with the target lysine highlighted in red. The vertical axis indicates the residues that were sequentially exchanged at the position corresponding to the row. **B)** Data from three independent experiments were averaged after normalizing the full activity to 1. The activity is displayed in grey-scale as indicated. **C)** Discrimination factors for the recognition of each amino acid at the corresponding position of the H3 substrate by NSD2 represented in a bar diagram.



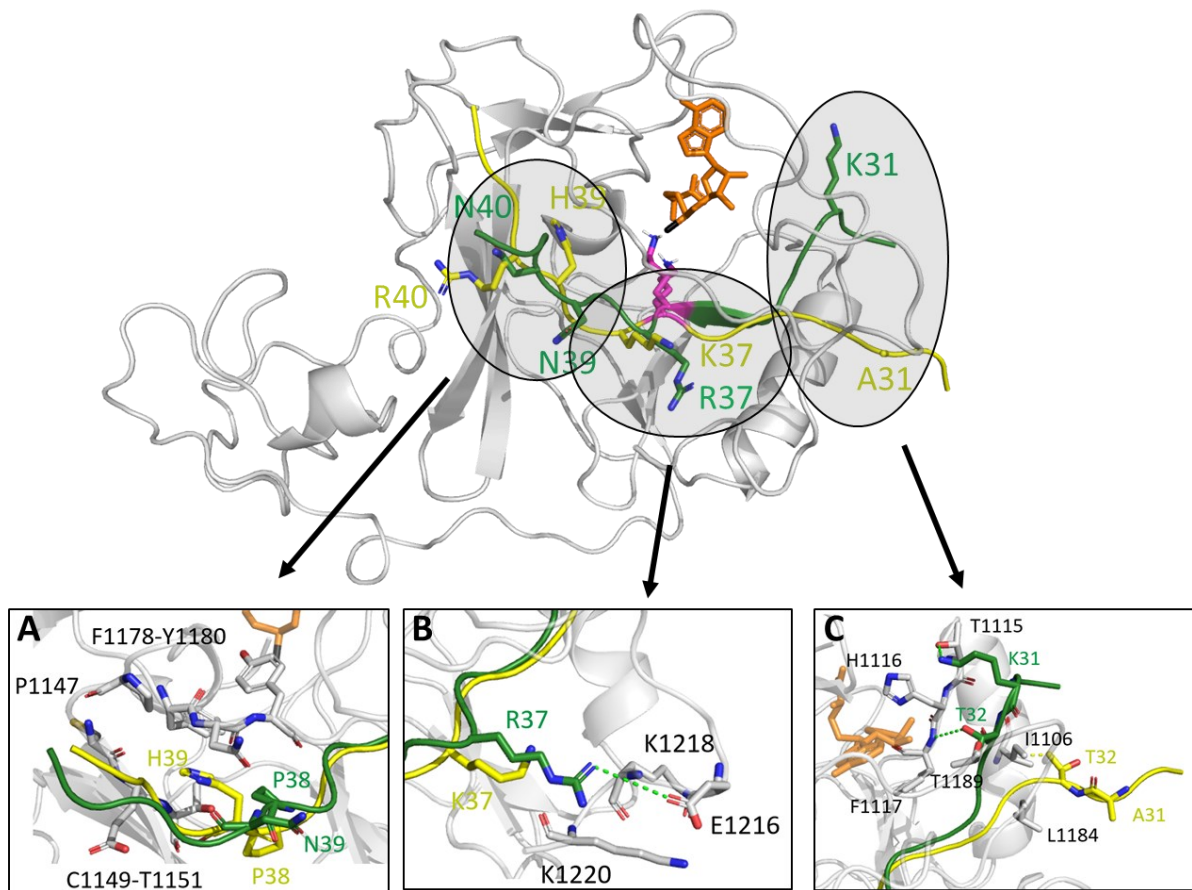
**Figure 2. Development of an H3K36 super-substrate specific for NSD2.** **A)** Based on the substrate specificity profile in Figure 1A all strongly methylated sequences with single point mutations from position -5 to +4 were synthesized on an extra array (Supplementary Table 1). As positive control H3K36 (spot A1 and C9) and as negative control H3K36A (spot A2 and C10) was added. Spots labelled with red circle indicate single point mutations used for the next array in panel B. **B)** Peptide array containing all possible double, triple, quadruple and quintuple amino acid mutation combinations from panel A (Supplementary Table 2). As positive and negative controls H3K36 (spot A1 and A3) and H3K36A (spot A2 and A4) were used. Strongest spots are marked with red circles and further used for the next peptide array in panel C. **C)** In first line all peptide sequences selected from panel B were synthesized next to each other together with its K-to-A mutants in the second line (Supplementary Table 3). Two autoradiographic images with exposure after 6 h and 2 days are presented. The best and strongest signal was detected for spot A8 labelled with a blue circle. Direct comparison of the natural H3 sequence and the NSD2 specific H3K36 super-substrate (ssK36) reveals 4 amino acids exchanges A31K, K37R, H38N and R39N. **D)** GST-tagged H3K36 (29-43) and ssK36 (29-43) were overexpressed and purified. For the protein methylation assay, equal protein amounts, verified by Coomassie staining, were incubated with NSD2 in the presence of radioactively labelled AdoMet separated by SDS-PAGE. Methylation was analyzed by autoradiography after 2 weeks of exposure. The corresponding band of the expected size is marked with a red asterisk. Automethylation of NSD2 is labelled with a blue asterisk.



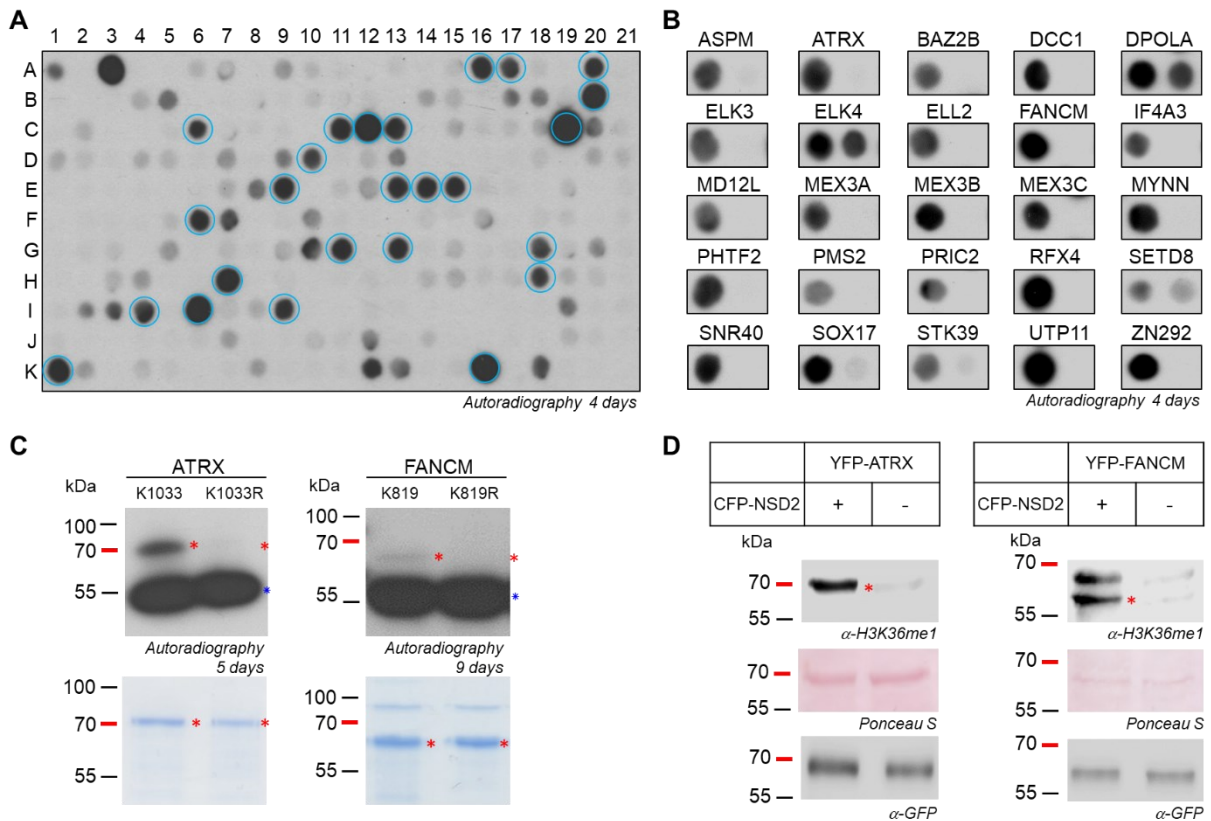
**Figure 3. The NSD2-ssK36 peptide complex establishes more  $S_N2$  TS-like conformations than the NSD2-H3K36 complex. A)** Superposition of 10 randomly selected frames taken from the ssK36 and H3K36 MD simulations after superposition of the NSD2 proteins. The peptides are shown in red (ssK36) and blue (H3K36) ribbon. NSD2 is shown in tan ribbon, with Zinc ions in grey and AdoMet in yellow. **B)** Average number of  $S_N2$  TS-like conformations of 50 MD simulation replicates (100 ns) for each peptide complexed to NSD2. **C)** Histogram of simulation replicates, where every replicate was sorted in the indicated bins depending on how many TS-like conformations were observed during the simulation run.



**Figure 4. The H3K36 and ssK36 peptides establish different contacts with NSD2.** The figure shows the contact map difference as the results of the subtraction of the H3K36 and ssK36 contact map including all NSD2 residues which exhibited noticeable changes. Yellow indicates that a specific contact was more often observed in simulations with H3K36. Green symbolizes a higher contact frequency for ssK36. Framed regions displayed the largest difference.



**Figure 5. Overlay and starting structure of the 15 aa long H3K36 (yellow) and ssK36 (green) peptide complexed in the NSD2 (grey) binding cleft.** The target lysine (pink) is inserted in the hydrophobic tunnel of the NSD2 SET domain. AdoMet (orange) binds from the opposing site. **A)** H39 of H3K36 is positioned in a pocket made by P1146, C1148, E1149, and T1150. In contrast, N39 points into solvent which positions P38 closer towards residues F1177, N1178 and Y1179. **B)** The longer side chain of R37 in ssK36 contacts E1216, which is not possible for K37 in H3K36. **C)** K31 in ssK36 interacts with NSD2 T1116, whereas residues T32-V35 in H3K36 interact with I1106.



**Figure 6. Identification of NSD2 non-histone substrates. A)** 15 aa long peptides (with the target K in the middle) of potential non-histone substrates identified in ScanSite searches (Obenauer et al., 2003) were synthesized on a SPOT peptide array and methylated by NSD2. As positive controls H3K36 (aa 29-43) and H4K44 (aa 161-175) were included in spot A1 and A3. As negative controls, the corresponding K-to-A mutants were used: H3K36A (aa 29-43) in spot A2 and H4K44A (aa 161-175) in spot A4 (Supplementary Table 4). 25 substrate peptides were selected for further analysis based on their methylation strength and the biological relevance marked with blue circles. **B)** To determine the correct target site position additional peptide array methylation experiments were performed, including the K-to-A mutants of each selected candidate substrate (Supplementary Table 5). Methylation of the predicted target lysine was validated in all but 3 of the candidate substrates (DPOLA, ELK4 and SETD8) which showed no loss of methylation in the K-to-A mutant. **C)** Purified wildtype and K-to-A mutant proteins of ATRX and FANCM were methylated by the NSD2 enzyme. Equal protein amounts were verified by Coomassie staining and the bands of expected size are marked with red asterisk. The autoradiographic image after 5 days for ATRX and 9 days for FANCM shows methylation in the wildtype protein and loss of methylation in the corresponding mutants. Automethylation of NSD2 is labelled with a blue asterisk. **D)** HEK293 cells were transfected with YFP-tagged ATRX or FANCM with or without CFP-tagged NSD2. After cell lysis the substrate proteins were purified by GFP trap and equal loading of the sample was verified by Ponceau S staining and Western Blot using anti-GFP antibody. To determine ATRX and FANCM methylation at cellular level Western Blot with the previously verified anti-H3K36me1 antibody (Supplementary Figure 8) was performed.



# Discovery of new NSD2 non-histone substrates and design of a super-substrate

Sara Weirich, Denis Kusevic, Philipp Schnee, Jessica Reiter, Jürgen Pleiss & Albert Jeltsch\*

## Supplementary Information

### Supplementary Figures

Supplementary Figure 1. Investigation of the protein quality and methylation activity of NSD2.

Supplementary Figure 2. Additional information related to Figure 1.

Supplementary Figure 3. Comparison of the methylation level at peptides when P30 and P43 were replaced by G and S.

Supplementary Figure 4. NSD2 specifically methylates the NSD2 super-substrate designed here, but not the previously designed super-substrate for SETD2.

Supplementary Figure 5. Additional information related to Figure 2D.

Supplementary Figure 6. Additional information related to Figure 2.

Supplementary Figure 7. Purification and Western Blot analysis of NSD2 non-histone substrate candidate proteins.

Supplementary Figure 8. Validation of the anti-H3K36me1 antibody for detection of methylated ATRX and FANCM proteins.

Supplementary Figure 9. Immunoblot detection of the expression of NSD2 full-length, ATRX and FANCM in HEK293 cells.

### Supplementary Tables

Supplementary Table 1: Sequences of the peptide SPOT array shown in Figure 2A.

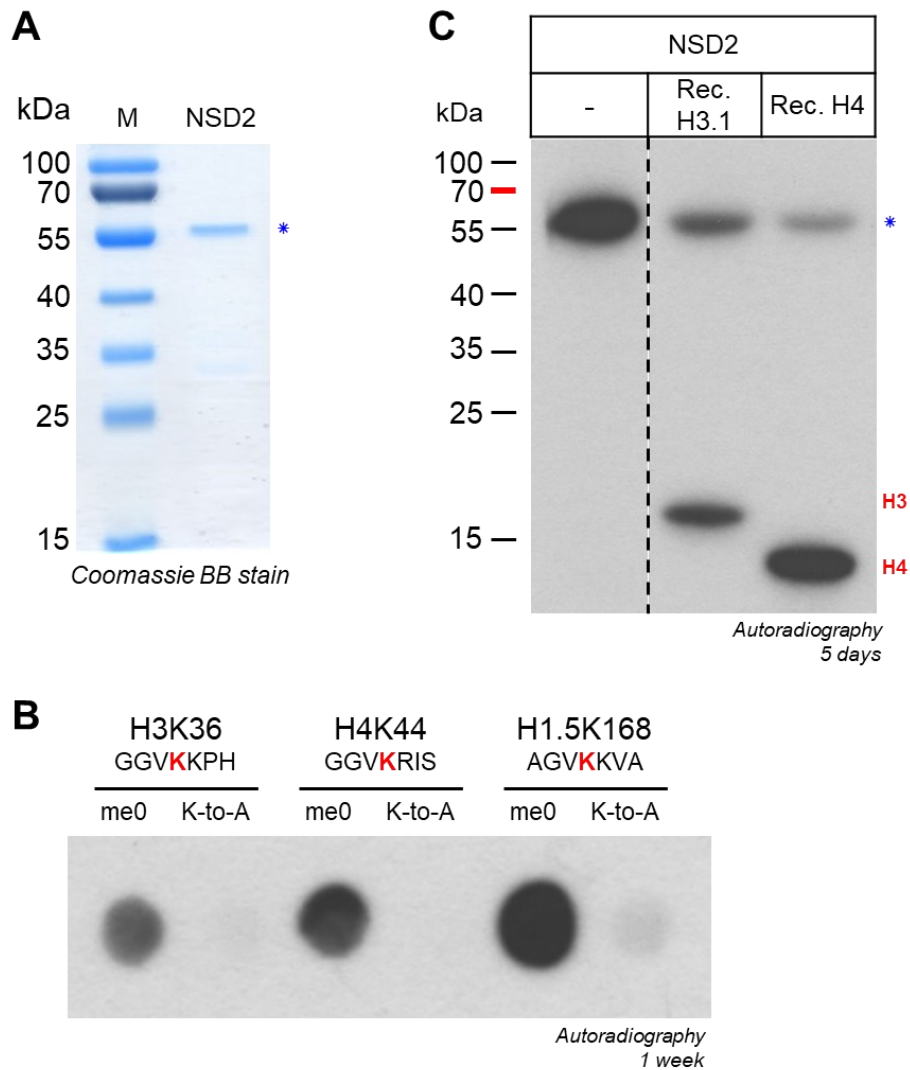
Supplementary Table 2: Sequences of the peptide SPOT array shown in Figure 2B.

Supplementary Table 3: Sequences of the peptide SPOT array shown in Figure 2C.

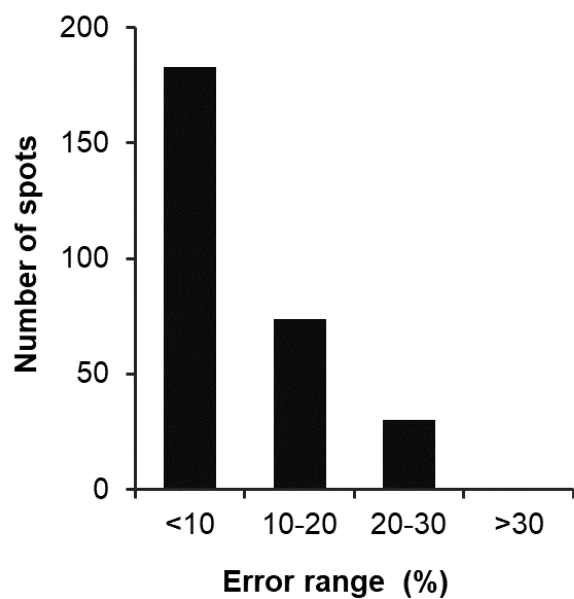
Supplementary Table 4: Sequences of the peptide SPOT array shown in Figure 6A.

Supplementary Table 5: Sequences of the peptide SPOT array shown in Figure 6B.

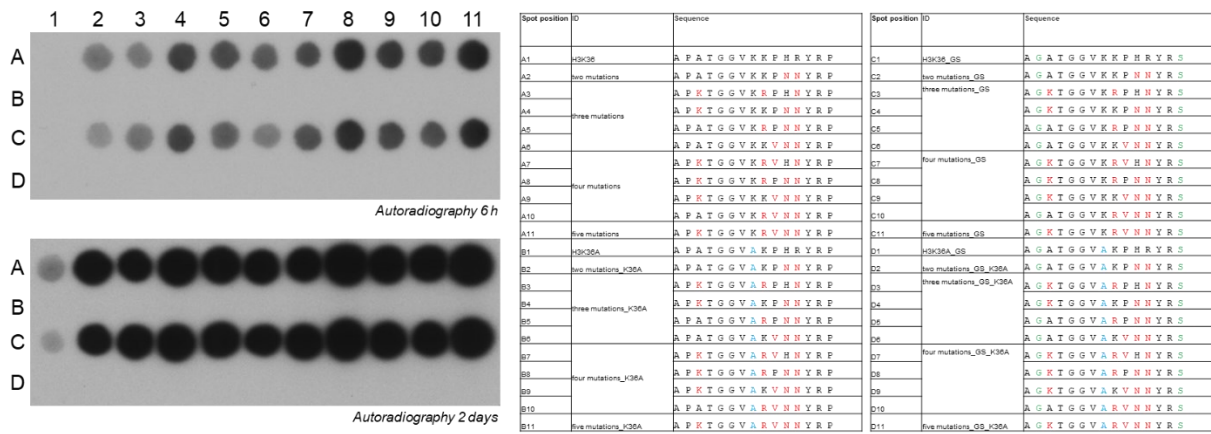
## Supplementary Figures



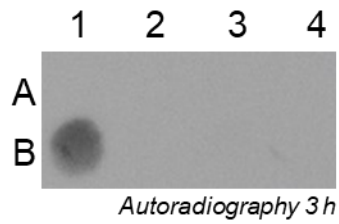
**Supplementary Figure 1. Investigation of the protein quality and methylation activity of NSD2. A)** Coomassie stained gel of the purified GST tagged NSD2 (aa 991-1240). **B)** Peptide array containing 15 aa long peptides of H3K36 (29-43), H4K44 (37-51) and H1.5K168 (161-175) and the corresponding K-to-A mutations as negative controls were incubated with NSD2 in presence of radioactively labeled AdoMet as cofactor. The autoradiographic image after exposure for one week confirmed the methylation of H3K36, H4K44, H1.5K168 and loss of methylation for the negative controls. **C)** Recombinant H3.1 and H4 were methylated by NSD2 using radioactively labeled AdoMet. To investigate potential automethylation of NSD2 same experiment was performed without substrate. Methylated samples were separated by SDS-PAGE and analyzed by autoradiography after 5 days of exposure. The corresponding bands of the methylated H3 and H4 are marked in red. Automethylation of NSD2 is labelled with a blue asterisk. Both parts were taken from the same gel image.



**Supplementary Figure 2. Additional information related to Figure 1.** The distribution of the standard errors of the mean of NSD2 activity is given for all peptides tested in the three independent substrate specificity arrays used for the averaged data shown in Figure 1B and C.

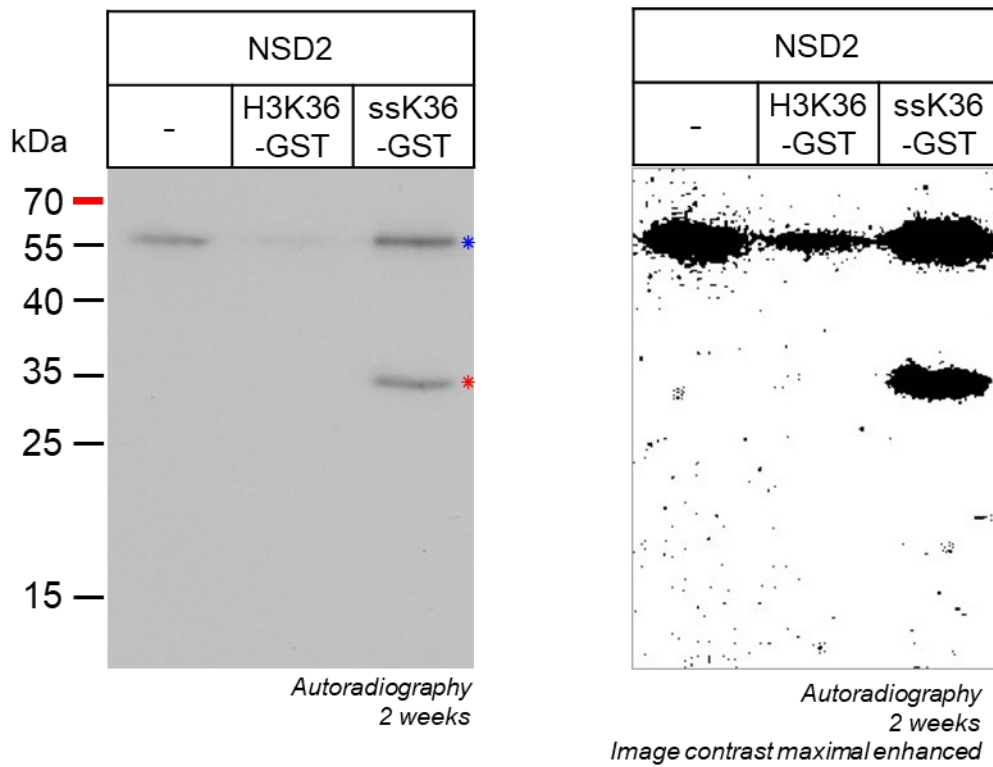


**Supplementary Figure 3. Comparison of the methylation level at peptides when P30 and P43 were replaced by G and S.** As the H3 peptide contains P at position 2 and at the end, we were concerned that this may cause artefacts. Therefore, in some of the design steps, P30 and P43 were replaced by G and S. This array shows a direct comparison of the methylation of PP and GS peptides in the design step shown in Figure 2C, indicating that this change does not affect the peptide methylation.

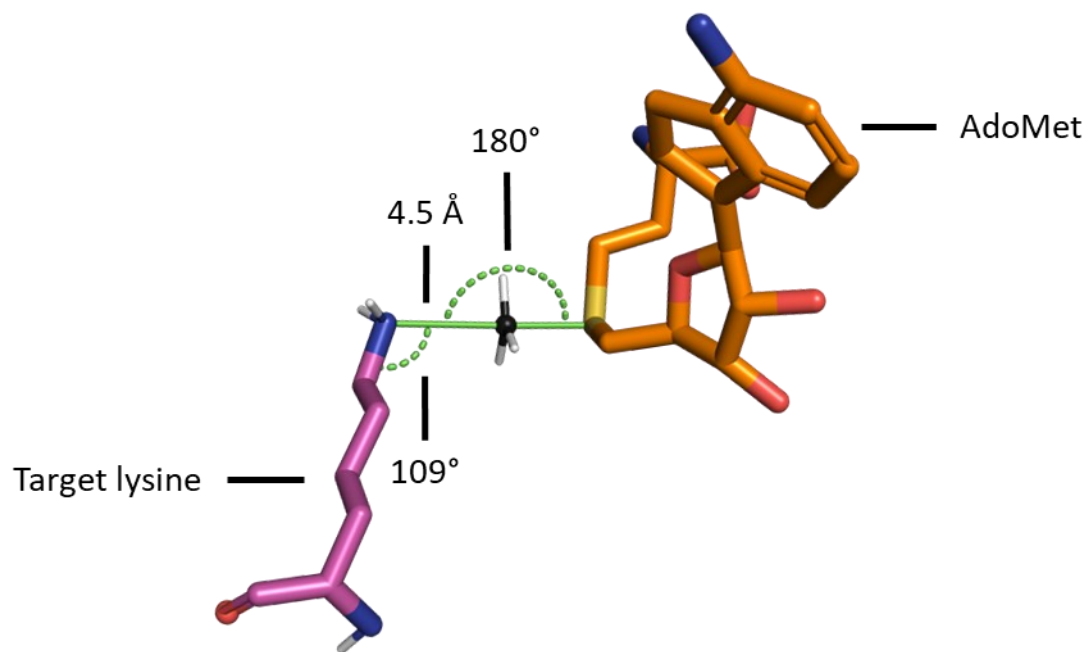


Spot position	ID	Sequence
A1	H3K36	A P A T G G V K K P H R Y R P
A2	H3K36A	A P A T G G V A K P H R Y R P
A3	H3K36_GS	A G A T G G V K K P H R Y R S
A4	H3K36A_GS	A G A T G G V A K P H R Y R S
B1	H3K36ss (NSD2)	A G K T G G V K R P N N Y R S
B2	H3K36Kss (NSD2)	A G K T G G V A R P N N Y R S
B3	H3K36ss (SETD2)	A P R F G G V K R P N R Y R P
B4	H3K36Kss (SETD2)	A P R F G G V A R P N R Y R P

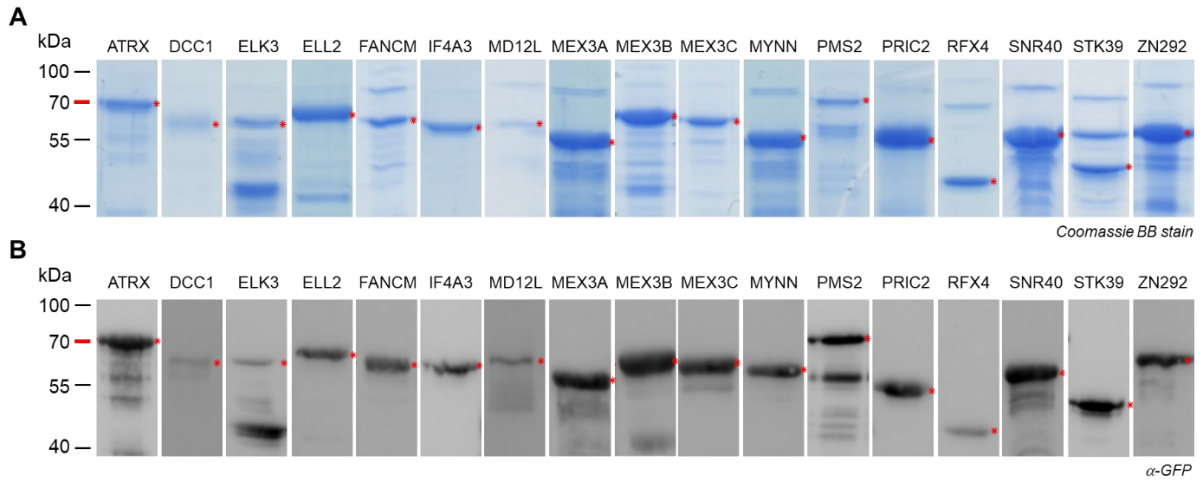
**Supplementary Figure 4. NSD2 specifically methylates the NSD2 super-substrate designed here, but not the previously designed super-substrate for SETD2.** SsK36 (NSD2) refers to the NSD2 super-substrate designed here. SsK36 (SETD2) refers to the SETD2 super-substrate designed and investigated by us previously (P. Schnee et al., 2022; Schuhmacher et al., 2020). The exchange of P30 and P43 by G and S in the SsK36 (NSD2) has no effect on enzyme activity (see Supplementary Figure 3).



**Supplementary Figure 5. Additional information related to Figure 2D.** No methylation signal for the wildtype H3K36 protein is detectable in the film shown in Figure 2D, even after a massive increase of image sensitivity.

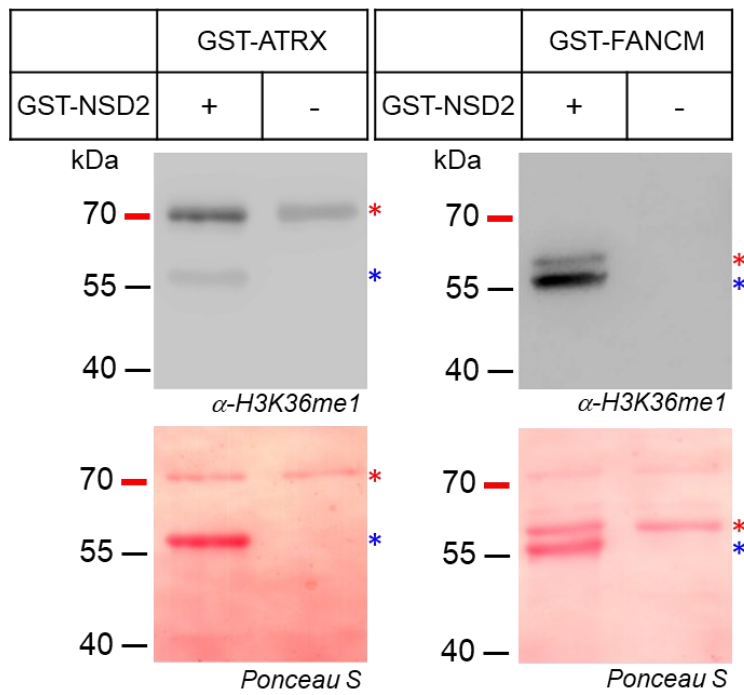


**Supplementary Figure 6. Additional information related to Figure 2.** Criteria used for definition of a successful docking event derived from the geometry of the  $S_N2$  TS-like conformation of PKMTs (P. Schnee et al., 2022).

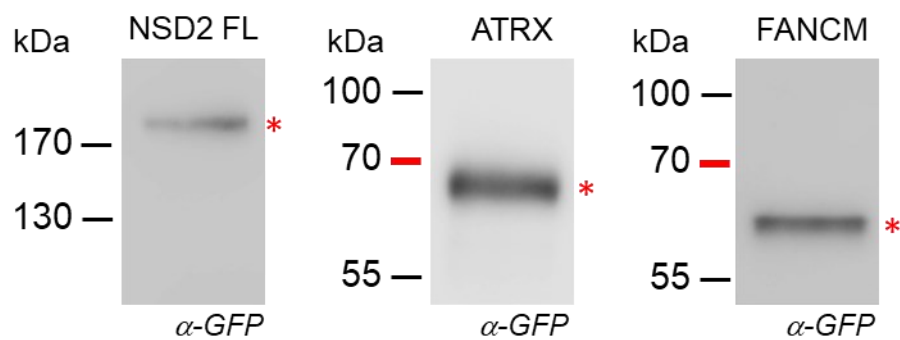


**Supplementary Figure 7. Purification and Western Blot analysis of NSD2 non-histone substrate candidate proteins. A)** Coomassie stained SDS gel of 17 potential NSD2 substrate proteins. **B)** Western Blot analysis of the purified proteins using an anti-GST antibody. The protein bands of expected size are labelled with red asterisk.





**Supplementary Figure 8. Validation of the anti-H3K36me1 antibody for detection of methylated ATRX and FANCM proteins.** ATRX and FANCM proteins were incubated with or without NSD2 in the presence of unlabeled AdoMet as methyl group donor, followed by SDS-PAGE and Western Blot analysis. Ponceau S staining of the methylated and unmethylated protein substrates is shown in the lower panel. Western blot of the transferred methylated and unmethylated protein substrates probed with the anti-H3K36me1 antibody is shown in the upper panel. The corresponding bands of the expected size are marked with a red asterisk. Automethylation of NSD2 is labelled with a blue asterisk.



**Supplementary Figure 9. Immunoblot detection of the expression of NSD2 full-length, ATRX and FANCM in HEK293 cells.** The expressed proteins were detected by probing with an anti-GFP antibody. The corresponding bands of the expected size are marked with a red asterisk.

## Supplementary Tables

**Supplementary Table 1: Sequences of the peptide SPOT array shown in Figure 2A.**

Spot position	ID	Sequence
A1	H3K36	A P A T G G V K K P H R Y R P
A2	H3K36A	A P A T G G V A K P H R Y R P
A3	H3A31G	A P G T G G V K K P H R Y R P
A4	H3A31H	A P H T G G V K K P H R Y R P
A5	H3A31K	A P K T G G V K K P H R Y R P
A6	H3A31M	A P M T G G V K K P H R Y R P
A7	H3T32G	A P A G G G V K K P H R Y R P
A8	H3T32K	A P A K G G V K K P H R Y R P
A9	H3T32F	A P A F G G V K K P H R Y R P
A10	H3T32V	A P A V G G V K K P H R Y R P
B1	H3G34F	A P A T G F V K K P H R Y R P
B2	H3V35I	A P A T G G I K K P H R Y R P
B3	H3V35L	A P A T G G L K K P H R Y R P
B4	H3K37R	A P A T G G V K R P H R Y R P
B5	H3K37Q	A P A T G G V K Q P H R Y R P
B6	H3K37I	A P A T G G V K I P H R Y R P
B7	H3K37V	A P A T G G V K V P H R Y R P
B8	H3P38I	A P A T G G V K K I H R Y R P
B9	H3P38L	A P A T G G V K K L H R Y R P
B10	H3P38V	A P A T G G V K K V H R Y R P
C1	H3H39N	A P A T G G V K K P N R Y R P
C2	H3H39G	A P A T G G V K K P G R Y R P
C3	H3H39L	A P A T G G V K K P L R Y R P
C4	H3H39S	A P A T G G V K K P S R Y R P
C5	H3R40N	A P A T G G V K K P H N Y R P
C6	H3R40H	A P A T G G V K K P H H Y R P
C7	H3R40L	A P A T G G V K K P H L Y R P
C8	H3R40K	A P A T G G V K K P H K Y R P
C9	H3K36	A P A T G G V K K P H R Y R P
C10	H3K36A	A P A T G G V A K P H R Y R P

**Supplementary Table 2: Sequences of the peptide SPOT array shown in Figure 2B.**

Spot position	ID	Sequence
A1	H3K36	A P A T G G V K K P H R Y R P
A2	H3K36A	A P A T G G V A K P H R Y R P
A3	H3K36_GS	A G A T G G V K K P H R Y R S
A4	H3K36A_GS	A G A T G G V A K P H R Y R S
A5		
A6	H3A31K_GS	A G K T G G V K K P H R Y R S
A7	H3K37R_GS	A G A T G G V K R P H R Y R S
A8	H3P38V_GS	A G A T G G V K K V H R Y R S
A9	H3H39N_GS	A G A T G G V K K P N R Y R S
A10	H3T32V_GS	A G A T G G V K K P H N Y R S
B1	two mutations GS	A G K T G G V K R P H R Y R S
B2		A G K T G G V K K V H R Y R S
B3		A G K T G G V K K P N R Y R S
B4		A G K T G G V K K P H N Y R S
B5		A G A T G G V K R V H R Y R S
B6		A G A T G G V K R P N R Y R S
B7		A G A T G G V K R P H N Y R S
B8		A G A T G G V K K V N R Y R S
B9		A G A T G G V K K V H N Y R S
B10		A G A T G G V K K P N N Y R S
C1	three mutations GS	A G K T G G V K R V H R Y R S
C2		A G K T G G V K R P N R Y R S
C3		A G K T G G V K R P H N Y R S
C4		A G K T G G V K K V N R Y R S
C5		A G K T G G V K K V H N Y R S
C6		A G K T G G V K K P N N Y R S
C7		A G A T G G V K R V N R Y R S
C8		A G A T G G V K R V H N Y R S
C9		A G A T G G V K R P N N Y R S
C10		A G A T G G V K K V N N Y R S
D1	four mutations GS	A G K T G G V K R V N R Y R S
D2		A G K T G G V K R V H N Y R S
D3		A G K T G G V K R P N N Y R S
D4		A G K T G G V K K V N N Y R S
D5		A G A T G G V K R V N N Y R S
D6		
D7	5 mutations_GS	A G K T G G V K R V N N Y R S
D8	5 mutations H3K36A_GS	A G K T G G V A R V N N Y R S
D9	five mutations	A P K T G G V K R V N N Y R P
D10	5 mutations H3K36	A P K T G G V A R V N N Y R P

**Supplementary Table 3: Sequences of the peptide SPOT array shown in Figure 2C.**

Spot position	ID	Sequence
A1	H3K36	A P A T G G V K K P H R Y R P
A2	two mutations	A P A T G G V K K P N N Y R P
A3	three mutations	A P K T G G V K R P H N Y R P
A4		A P K T G G V K K P N N Y R P
A5		A P A T G G V K R P N N Y R P
A6		A P A T G G V K K V N N Y R P
A7	four mutations	A P K T G G V K R V H N Y R P
A8		A P K T G G V K R P N N Y R P
A9		A P K T G G V K K V N N Y R P
A10		A P A T G G V K R V N N Y R P
A11	five mutations	A P K T G G V K R V N N Y R P
B1	H3K36A	A P A T G G V A K P H R Y R P
B2	two mutations_K36A	A P A T G G V A K P N N Y R P
B3	three mutations_K36A	A P K T G G V A R P H N Y R P
B4		A P K T G G V A K P N N Y R P
B5		A P A T G G V A R P N N Y R P
B6		A P A T G G V A K V N N Y R P
B7	four mutations_K36A	A P K T G G V A R V H N Y R P
B8		A P K T G G V A R P N N Y R P
B9		A P K T G G V A K V N N Y R P
B10		A P A T G G V A R V N N Y R P
B11	five mutations_K36A	A P K T G G V A R V N N Y R P

**Supplementary Table 4: Sequences of the peptide SPOT array shown in Figure 6A.**

Spot position	Swiss Prot No	ID	Protein name	Sequence	Target position
A1	P68431	H3K36	H3K36	A P A T G G V <b>K</b> K P H R Y R P	36
A2	P68431	H3K36A	H3K36A	A P A T G G V <b>A</b> K P H R Y R P	36
A3	P62805	H4K44	H4K44	L A R R G G V <b>K</b> R I S G L I Y	44
A4	P62805	H4K44A	H4K44A	L A R R G G V <b>A</b> R I S G L I Y	44
A5	Q9UKV3	ACINU	Apoptotic chromatin condensation inducer in the nucleus	P P A E H E V <b>K</b> K V T L G D T	969
A6	Q9UKV3	ACINU	Apoptotic chromatin condensation inducer in the nucleus	G I T E E C L <b>K</b> Q P S L E Q K	548
A7	Q9Y2D8	ADIP	Afadin- and alpha-actinin-binding protein	L M E N A E L <b>K</b> K V L Q Q M K	278
A8	Q8IUX7	AEBP1	Adipocyte enhancer-binding protein 1	D E E K E E L <b>K</b> K P K K E D S	340
A9	Q08117	AES	Amino-terminal enhancer of split	H K Q A E I V <b>K</b> R L N G I C A	83
A10	Q9NVM4	ANM7	Protein arginine N-methyltransferase 7	N G F S D K I <b>K</b> V I N K H S T	117
A11	Q86SG2	ANR23	Ankyrin repeat domain-containing protein 23	G G H L V I L <b>K</b> Q L L N Q G A	194
A12	Q9NX46	ARHL2	Poly(ADP-ribose) glycohydrolase ARH3	S S S E H F L <b>K</b> Q L L G H M E	213
A13	P29374	ARI4A	AT-rich interactive domain-containing protein 4A	V K R L V K V <b>K</b> V L L K Q D N	39
A14	Q9HBZ2	ARNT2	Aryl hydrocarbon receptor nuclear translocator 2	I C T N T N V <b>K</b> Q L Q Q Q Q A	439
A15	Q8IZT6	ASPM	Abnormal spindle-like microcephaly-associated protein	A L S K F T L <b>K</b> K L L L L V C	888
A16	Q8IZT6	ASPM	Abnormal spindle-like microcephaly-associated protein	Q T Y F N K L <b>K</b> K I T K T V Q	2213
A17	P46100	ATRX	Transcriptional regulator ATRX	C H F P K G I <b>K</b> Q I K N G T T	1033
A18	P54253	ATX1	Ataxin-1	V C I S L T L <b>K</b> N L K N G S V	688
A19	Q8N9N5	BANP	Protein BANP	R Q N T I V V <b>K</b> V P G Q E D S	165
A20	Q9UIF8	BAZ2B	Bromodomain adjacent to zinc finger domain protein 2B	M K Q Q E K I <b>K</b> R I Q Q I R M	948
A21	Q9P287	BCCIP	BRCA2 and CDKN1A-interacting protein	D N D Y D G I <b>K</b> K L L Q Q L F	75
B1	Q14137	BOP1	Ribosome biogenesis protein BOP1	N P L L V P V <b>K</b> V L K G H V L	708
B2	Q9H0E9	BRD8	Bromodomain-containing protein 8	A E R V E E L <b>K</b> K V I K E T Q	109
B3	Q5VTR2	BRE1A	E3 ubiquitin-protein ligase BRE1A	K K E A E I I <b>K</b> Q L K I E L K	627
B4	Q13892	BT3L3	Transcription factor BTF3 homolog 3	K K L Q F S L <b>K</b> K L Q V N N I	79

B5	O43683	BUB1	Mitotic checkpoint serine/threonine-protein kinase BUB1	D L L R Q K L K K V F Q Q H Y	1055
B6	Q8TBZ0	CC110	Coiled-coil domain-containing protein 110	I Q P Q S A L K V L Q Q Q L E	57
B7	Q8TD31	CCHCR	Coiled-coil alpha-helical rod protein 1	L E H S D S V K Q L K G Q V A	360
B8	P21127	CD2L1	PITSLRE serine/threonine-protein kinase CDC2L1	T D E I V A L K R L K M E K E	467
B9	Q9UQ88	CD2L2	PITSLRE serine/threonine-protein kinase CDC2L2	T D E I V A L K R L K M E K E	455
B10	P50613	CDK7	Cell division protein kinase 7	T N Q I V A I K K I K L G H R	41
B11	P50750	CDK9	Cell division protein kinase 9	T G Q K V A L K K V L M E N E	48
B12	O15078	CE290	Centrosomal protein of 290 kDa	S S L L V K L K K V S Q D L E	1645
B13	O15078	CE290	Centrosomal protein of 290 kDa	E N H E D E V K K V K A E V E	1681
B14	Q96EP1	CHFR	E3 ubiquitin-protein ligase CHFR	G T V I N K L K V V K K Q T C	88
B15	Q969X6	CIR1A	Cirhin	G V H V Y N V K Q L K L H C T	514
B16	Q86XI2	CNDG2	Condensin-2 complex subunit G2	T I L I D L L K K V T G E L A	421
B17	Q9P2I0	CPSF2	Cleavage and polyadenylation specificity factor subunit 2	R S D G D S I K K I I N Q M K	550
B18	O60716	CTND1	Catenin delta-1	L G A C G A L K N I S F G R D	433
B19	Q9UI36	DACH1	Dachshund homolog 1	I A E A M K V K K I K L E A M	347
B20	Q9BVC3	DCC1	Sister chromatid cohesion protein DCC1	R P K L K K L K K L L M E N P	139
B21	P27707	DCK	Deoxycytidine kinase	S S E G T R I K K I S I E G N	22
C1	Q9UNQ2	DIMT1	Probable dimethyladenosine transferase	G I G Q H I L K N P L I I N S	40
C2	O60832	DKC1	H/AACA ribonucleoprotein complex subunit 4	H G I V A K I K R V I M E R D	367
C3	Q13316	DMP1	Dentin matrix acidic phosphoprotein 1	S E E D G Q L K N I E I E S R	482
C4	Q9BZG8	DPH1	Diphthamide biosynthesis protein 1	Q I P P E I L K N P Q L Q A A	43
C5	Q15054	DPOD3	DNA polymerase delta subunit 3	S K K A E P V K V L Q K E K K	292
C6	P09884	DPOLA	DNA polymerase alpha catalytic subunit	V E R R K Q V K Q L M K Q Q D	926
C7	Q92630	DYRK2	Dual specificity tyrosinephosphorylation - regulated kinase 2	A Y R Y E V L K V I G K G S F	226
C8	Q3B7T1	EDRF1	Erythroid differentiation-related factor 1	V S D S E N I K K L L K I P Y	131
C9	Q5THR3	EFCB6	EF-hand calcium-binding domain-containing protein 6	T L S S L D V K R I L F Q K I	61
C10	P60228	EIF3E	Eukaryotic translation initiation factor 3 subunit E	Q V L K D L V K V I Q Q E S Y	279
C11	P41970	ELK3	ETS domain-containing protein Elk-3	Y Y D K N I I K K V I G Q K F	73

C12	P28324	ELK4	ETS domain-containing protein Elk-4	Y Y V K N I I K K V N G Q K F	73
C13	O00472	ELL2	RNA polymerase II elongation factor ELL2	H N K L A H I K R L I G E F D	625
C14	P60228	ENOA	Alpha-enolase	Q V L K D L V K V I Q Q E S Y	279
C15	Q14674	ESPL1	Separin	R L Q V E S L K K L G K Q A Q	515
C16	Q14674	ESPL1	Separin	T Q H L D S V K K V H L Q K G	1075
C17	Q03112	EVI1	Ecotropic virus integration site 1 protein homolog	P Q S P G E V K K L Q K G S S	543
C18	Q9UQ84	EXO1	Exonuclease 1	A N N P D I V K V I K K I G H	252
C19	Q8IYD8	FANCM	Fanconi anemia group M protein	H K K S S F I K N I N Q G S S	819
C20	Q13451	FKBP5	FK506-binding protein 5	R G V L K I V K R V G N G E E	38
C21	O15117	FYB	FYN-binding protein	K T E E K D L K K L K K Q E K	683
D1	P04150	GCR	Glucocorticoid receptor	K Y S N G N I K K L L F H Q K	770
D2	Q9HC44	GPBL1	Vasculin-like protein 1	P S K M L V I K K V S K E D P	199
D3	P78347	GTF2I	General transcription factor II-I	A G I S F I I K R P F L E P K	185
D4	P0C1H6	H2BFM	Histone H2B type F-M	P Y F P R V L K Q V H Q G L S	69
D5	Q7Z2G1	H2BWT	Histone H2B type W-T	T Y F R R V L K Q V H Q G L S	90
D6	Q9UBN7	HDAC6	Histone deacetylase 6	H Q A L L D V K N I A H Q N K	1199
D7	Q9H2X6	HIPK2	Homeodomain-interacting protein kinase 2	S S A F C S V K K L K I E P S	29
D8	A6NHT5	HMX3	Homeobox protein HMX3	K E S P F S I K N L L N G D H	34
D9	Q5SSJ5	HP1B3	Heterochromatin protein 1-binding protein 3	R P S S T V I K K P S G G S S	517
D10	P38919	IF4A3	Eukaryotic initiation factor 4A-III	A I Q Q R A I K Q I I K G R D	70
D11	Q9UKT9	IKZF3	Zinc finger protein Aiolos	I C P R D S V K V I N K E G E	440
D12	Q9UL03	INT6	Integrator complex subunit 6	Y S V I S Y L K K L S Q Q A K	457
D13	Q9NVH2	INT7	Integrator complex subunit 7	N D P R K A V K R L A I Q D L	269
D14	Q9BZ11	IRX2	Iroquois-class homeodomain protein IRX-2	E I A T S D L K Q P S L G P G	341
D15	Q8NHM5	KDM2B	Lysine-specific demethylase 2B	Q A L L E G V K N V L K E H A	549
D16	O75164	KDM4A	Lysine-specific demethylase 4A	E V K F E E L K N V K L E E E	468
D17	Q96Q89	KI20B	Kinesin-like protein KIF20B	T D A K K Q I K Q V Q K E V S	1300
D18	P46013	KI67	Antigen KI-67	M H T P P V L K K I I K E Q P	548
D19	Q2VIQ3	KIF4B	Chromosome-associated kinesin KIF4B	F Q Y Q D N I K N L E L E V I	562



D20	Q6P0N0	KNL2	Kinetochores-associated protein KNL-2 homolog	T R L L P K L K K I E N Q V A	752
D21	P50748	KNTC1	Kinetochores-associated protein 1	E K A E A L L K K L H I Q Y R	1737
E1	Q15349	KS6A2	Ribosomal protein S6 kinase alpha-2	P S Q F E L L K V L G Q G S Y	63
E2	A8MZ59	LEUTX	Putative leucine-twenty homeobox	L R E P S G I K N P G G A S A	92
E3	Q6MZP7	LIN54	Protein lin-54 homolog	K L P P Q Q I K V V T I G G R	184
E4	Q86V48	LUZP1	Leucine zipper protein 1	V I V D K D V K K I M G G S G	766
E5	Q86V48	LUZP1	Leucine zipper protein 1	K G G L D Y L K Q V E N E T R	265
E6	Q9Y2U5	M3K2	Mitogen-activated protein kinase 2	E C E I Q L L K N L L H E R I	411
E7	Q6ZN28	MACC1	Metastasis-associated in colon cancer protein 1	L V H C K N V K V I S K E Q V	615
E8	Q05195	MAD1	MAD protein	G Y S S T S I K R I K L Q D S	207
E9	Q86YW9	MD12L	Mediator of RNA polymerase II transcription subunit 12-like protein	R A Y M N L V K K L K K E L G	1604
E10	Q9NU22	MDN1	Midasin	M G E E A A L K R P E I I S T	1622
E11	Q9NU22	MDN1	Midasin	E C L K F L I K R L A K I V R	1670
E12	P35240	MERL	Merlin	S S K H N T I K K L T L Q S A	578
E13	A1L020	MEX3A	RNA-binding protein MEX3A	G P K G A T I K R I Q Q Q T N	247
E14	Q6ZN04	MEX3B	RNA-binding protein MEX3B	G P K G A T I K R I Q Q Q T H	184
E15	Q5U5Q3	MEX3C	RNA-binding protein MEX3C	G P K G A T I K R I Q Q Q T H	350
E16	Q8IWI9	MGAP	MAX gene-associated protein	P T F F V I L K Q P G N G K T	29
E17	P16455	MGMT	Methylated-DNA-protein-cysteine methyltransferase	Q V L W K L L K V V K F G E V	104
E18	Q96T58	MINT	Msx2-interacting protein	V T Q G G T V K V L T Q G I N	2930
E19	O14686	MLL2	Histone-lysine N-methyltransferase MLL2	R A S E A L L K Q L K Q E L S	4553
E20	O14686	MLL2	Histone-lysine N-methyltransferase MLL2	G R P E F V I K V I E Q G L E	5244
E21	Q9NYL2	MLTK	Mitogen-activated protein kinase MLT	Q D K E V A V K K L L K I E K	45
F1	P41218	MNDA	Myeloid cell nuclear differentiation antigen	T Q E K A P V K K I N Q E E V	102
F2	Q99549	MPP8	M-phase phosphoprotein 8	V K E T K E L K K V K K G E I	228
F3	Q9BYG7	MSTRO	Protein maestro	F Q K R E P L K N V F F I L A	47
F4	P10242	MYB	Myb proto-oncogene protein	E N G P P L L K K I K Q E V E	524
F5	P01106	MYC	Myc proto-oncogene protein	P D D E T F I K N I I I Q D C	126
F6	Q9NPC7	MYNN	Myoneurin	G N S Y T D I K N L K K H K T	512

F7	Q92794	MYST3	Histone acetyltransferase MYST3	E W I L E A I K K V K K Q K Q	18
F8	Q01538	MYT1	Myelin transcription factor 1	S S M E K N L K N I E E E N K	1040
F9	Q15742	NAB2	NGFI-A-binding protein 2	E L G G P P L K K L K Q E V G	376
F10	Q9C000	NALP1	NACHT, LRR and PYD domains-containing protein 1	I K F S R H V K K L Q L I E G	765
F11	P52298	NCBP2	Nuclear cap-binding protein subunit 2	F S K S G D I K K I I M G L D	67
F12	Q6KC79	NIPBL	Nipped-B-like protein	G S I E R I L K Q V S G G E D	1639
F13	Q9ULX3	NOB1	RNA-binding protein NOB1	W I T P S N I K Q I Q Q E L E	215
F14	Q8WTT2	NOC3L	Nucleolar complex protein 3 homolog	E M C C E A V K K L F K Q D K	386
F15	P78316	NOP14	Nucleolar protein 14	L F T P R L V K V L E F G R K	772
F16	Q8TAT6	NPL4	Nuclear protein localization protein 4 homolog	E T A A T F L K K V A K E F G	31
F17	Q15466	NR0B2	Nuclear receptor subfamily 0 group B member 2	A P V P S I L K K I L L E E P	119
F18	O75469	NR1I2	Nuclear receptor subfamily 1 group I member 2	L K F H Y M L K K L Q L H E E	331
F19	P19338	NUCL	Nucleolin	F E K A T F I K V P Q N Q N G	513
F20	Q13416	ORC2	Origin recognition complex subunit 2	P S F S A E L K Q L N Q Q Y E	288
F21	Q13177	PAK2	Serine/threonine-protein kinase PAK 2	L G Q E V A I K Q I N L Q K Q	278
G1	Q460N3	PAR15	Poly [ADP-ribose] polymerase 15	T P S L K T V K V V I F Q P E	444
G2	Q9H4Z3	PCIF1	Phosphorylated CTD-interacting factor 1	Q P S G N G V K K P K I E I P	126
G3	O60437	PEPL	Periplakin	S F L Q D K L K R L E K E R A	1099
G4	O60437	PEPL	Periplakin	A E R A K D L K N I T N E L L	550
G5	P56645	PER3	Period circadian protein homolog 3	Y A S V N K I K N L G Q Q L Y	470
G6	Q96BD5	PF21A	PHD finger protein 21A	A T P P Q P I K V P Q F I P P	214
G7	O60346	PHLP1	PH domain leucine-rich repeat-containing protein phosphatase 1	M S C E E E L K R I K Q H K A	1317
G8	O60346	PHLP1	PH domain leucine-rich repeat-containing protein phosphatase 1	H K G G G V V K V L G Q G P G	246
G9	Q9UMS5	PHTF1	Putative homeodomain transcription factor 1	E I P H F R L K K V E N I K I	585
G10	Q9UMS5	PHTF1	Putative homeodomain transcription factor 1	T I F G N R I K R V K L I S N	210
G11	Q8N3S3	PHTF2	Putative homeodomain transcription factor 2	E V P H F R L K K V Q N I K M	574
G12	Q9H307	PININ	Pinin	D P E D D D V K K P A L Q S S	108
G13	P54278	PMS2	Mismatch repair endonuclease PMS2	S S L A K R I K Q L H H E A Q	630

G14	Q12972	PO6F2	POU domain, class 6, transcription factor 2	L V Y H K H L K R V F L I D L	81
G15	Q9UBT6	POLK	DNA polymerase kappa	R T V T I K L K N V N F E V K	461
G16	Q15435	PP1R7	Protein phosphatase 1 regulatory subunit 7	D I A S N R I K K I E N I S H	287
G17	Q12972	PP1R8	Nuclear inhibitor of protein phosphatase 1	L V Y H K H L K R V F L I D L	81
G18	Q7Z3G6	PRIC2	Prickle-like protein 2	P G E K L R I K Q L L H Q L P	74
G19	P28070	PSB4	Proteasome subunit beta type-4	Y A D F Q Y L K Q V L G Q M V	109
G20	P61289	PSME3	Proteasome activator complex subunit 3	T L H D M I L K N I E K I K R	237
G21	Q5TB80	QN1	Protein QN1 homolog	P V T G E K L K Q I Q K E I Q	704
H1	Q8TDY2	RBCC1	RB1-inducible coiled-coil protein 1	E E N E N K I K K L K G E L V	893
H2	Q9Y4C8	RBM19	Probable RNA-binding protein 19	E Q A Q K A L K Q L Q G H V V	792
H3	Q96LT9	RBM40	RNA-binding protein 40	P N C R I Y V K N L A K H V Q	425
H4	P49792	RBP2	E3 SUMO-protein ligase RanBP2	P G I R K W L K Q L F H H L P	458
H5	Q9H1J1	REN3A	Regulator of nonsense transcripts 3A	E V R I K L L K K P E K G E E	285
H6	Q9BZI7	REN3B	Regulator of nonsense transcripts 3B	V N Q K N L L K K P E K G D E	285
H7	Q33E94	RFX4	Transcription factor RFX4	A K R Q G S L K K V A Q Q F L	421
H8	O14715	RGPD8	RANBP2-like and GRIP domain-containing protein 8	A A N L E Y L K N V L L Q F I	1713
H9	Q5UIP0	RIF1	Telomere-associated protein RIF1	R S G A P M I K K I A F I A W	290
H10	Q9H9A7	RMI1	RecQ-mediated genome instability protein 1	L L K P E N V K V L G G E V D	188
H11	Q92753	RORB	Nuclear receptor ROR-beta	G L D M T G I K Q I K Q E P I	187
H12	P24928	RPB1	DNA-directed RNA polymerase II subunit RPB1	A C P L R T I K R V Q F G V L	19
H13	P56182	RRP1	Ribosomal RNA processing protein 1 homolog A	M V L N E S L K V L K M Q G W	131
H14	Q9Y3Z3	SAMH1	SAM domain and HD domain-containing protein 1	F Q R L R Y I K Q L G G G Y Y	148
H15	Q9UN30	SCML1	Sex comb on midleg-like protein 1	K H S Y R L V K K L K L Q K M	115
H16	O60524	SDCG1	Serologically defined colon cancer antigen 1	G A T S C V I K N P T G E P I	578
H17	Q9P0V9	Sep 10	Septin-10	Q A K F E H L K R L H Q E E R	396
H18	Q9NQR1	SETD8	Histone-lysine N-methyltransferase SET8	A I A K Q A L K K P I K G K Q	158
H19	Q15637	SF01	Splicing factor 1	G P R G N T L K N I E K E C N	165
H20	O00141	SGK1	Serine/threonine-protein kinase Sgk1	P S D F H F L K V I G K G S F	102
H21	O95905	SGT1	Protein SGT1	W F I V Y V I K Q I T K E F P	96

I1	Q96ST3	SIN3A	Paired amphipathic helix protein Sin3a	K E D K Y K I K Q I M H H F I	813
I2	Q96GM5	SMRD1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1	L D I Q E A L K R P I K Q K R	173
I3	Q6STE5	SMRD3	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 3	V D I Q E A L K R P M K Q K R	148
I4	Q96DI7	SNR40	U5 small nuclear ribonucleoprotein 40 kDa protein	S E T G E R V K R L K G H T S	145
I5	Q9BQ15	SOSB1	SOSS complex subunit B1	K D I K P G L K N L N L I F I	15
I6	Q9H6I2	SOX17	Transcription factor SOX-17	R K Q V K R L K R V E G G F L	149
I7	O00267	SPT5H	Transcription elongation factor SPT5	P T K N N K V K V I L G E D R	1042
I8	Q96SB4	SRPK1	Serine/threonine-protein kinase SRPK1	G L P L P C V K K I I Q Q V L	190
I9	Q9UEW8	STK39	STE20/SPS1-related proline-alanine-rich protein kinase	R Q E R V A I K R I N L E K C	92
I10	Q13188	STK3	Serine/threonine-protein kinase 3	D G D F D F L K N L S L E E L	441
I11	Q8IYB8	SUV3	ATP-dependent RNA helicase SUPV3L1, mitochondrial	L L T P D M L K Q L E K E W M	749
I12	Q15022	SUZ12	Polycomb protein SUZ12	G A A V L P V K K P K M E H V	72
I13	Q5T4T6	SYC2L	Synaptonemal complex protein 2-like	K I F I I Y L K K P M I I S Y	370
I14	Q15431	SYCP1	Synaptonemal complex protein 1	E V E L E E L K K V L G E K E	437
I15	Q8NF91	SYNE1	Nesprin-1	Q D S G I V L K R V T I H L E	4833
I16	Q8WXH0	SYNE2	Nesprin-2	K L L E S Q I K Q L E H G W E	2601
I17	Q8WXH0	SYNE2	Nesprin-2	Q E Q N E L L K V V I K Q T N	3992
I18	P13984	T2FB	General transcription factor IIF subunit 2	S E N Y M R L K R L Q I E E S	128
I19	Q15573	TAF1A	TATA box-binding protein-associated factor RNA polymerase I subunit A	K Y L A K Y L K N I L M G N H	357
I20	Q6P1X5	TAF2	Transcription initiation factor TFIID subunit 2	W K H V D E L K V L K I H I N	142
I21	Q92750	TAF4B	Transcription initiation factor TFIID subunit 4B	K V A V T P V K K L A Q I G T	179
J1	Q6SJ96	TBPL2	TATA box-binding protein-like protein 2	L A C K L D L K K I A L H A K	218
J2	O14776	TCRG1	Transcription elongation regulator 1	T S T W K E V K K I I K E D P	981
J3	Q13569	TDG	G/T mismatch-specific thymine DNA glycosylase	E V F G V K V K N L E F G L Q	248
J4	Q8NFU7	TET1	Methylcytosine dioxygenase TET1	T L S P G K L K Q L I Q E R D	50
J5	Q9Y5Q9	TF3C3	General transcription factor 3C polypeptide 3	F C L R L M L K N P E N H A L	728

J6	Q01664	TFAP4	Transcription factor AP-4	H M Y P E K L K V I A Q Q V Q	189
J7	Q04725	TLE2	Transducin-like enhancer protein 2	H K Q A E I V K R L S G I C A	82
J8	Q9UK18	TLK1	Serine/threonine-protein kinase tousled-like 1	N L H I R E L K R I N N E D N	436
J9	Q86UE8	TLK2	Serine/threonine-protein kinase tousled-like 2	N L H I R E L K R I H N E D N	442
J10	Q02880	TOP2B	DNA topoisomerase 2-beta	K V G K P K V K K L Q L E E T	1226
J11	Q9ULW0	TPX2	Targeting protein for Xklp2	N L P E K K V K N V T Q I E P	585
J12	Q9Y606	TRUA	tRNA pseudouridine synthase A	H I R I L G L K R V T G G F N	184
J13	Q9UNY4	TTF2	Transcription termination factor 2	S Q W T N M L K V V A L H L K	1023
J14	O75643	U520	U5 small nuclear ribonucleoprotein 200 kDa helicase	S N S L I S I K R L T L Q Q K	2080
J15	Q13404	UB2V1	Ubiquitin-conjugating enzyme E2 variant 1	W Q N S Y S I K V V L Q E L R	118
J16	P63279	UBC9	SUMO-conjugating enzyme UBC9	W R P A I T I K Q I L L G I Q	110
J17	Q9UK80	UBP21	Ubiquitin carboxyl-terminal hydrolase 21	G L P D E R L K K L E L G R G	64
J18	Q9P275	UBP36	Ubiquitin carboxyl-terminal hydrolase 36	L V H S S N V K V V L N Q Q A	408
J19	Q9NWZ5	UCKL1	Uridine-cytidine kinase-like 1	D L I I S T L K K L K Q G K S	170
J20	O94901	UN84A	Protein unc-84 homolog A	K P T T S R L K Q P L Q G D S	374
J21	P13051	UNG	Uracil-DNA glycosylase	M C D I K D V K V V I L G Q D	147
K1	Q9Y3A2	UTP11	Probable U3 small nucleolar RNA-associated protein 11	V T N Q T G L K R I A K E R Q	189
K2	O75691	UTP20	Small subunit processome component 20 homolog	H R R A R A L K K L A K Q L M	1591
K3	O75691	UTP20	Small subunit processome component 20 homolog	S E Q D P L L K N L S Q E I I	2690
K4	P11473	VDR	Vitamin D3 receptor	I K F Q V G L K K L N L H E E	321
K5	Q9HBM0	VEZA	Vezatin	H C T V V P L K Q P T L H I A	525
K6	O75554	WBP4	WW domain-binding protein 4	K A Y Q E D L K R L G L E S E	81
K7	Q9ULM3	YETS2	YEATS domain-containing protein 2	A Q G Q Q T L K V I S G Q K T	900
K8	P59817	Z280A	Zinc finger protein 280A	F K C L S C V K V L K N I K F	305
K9	Q6P3V2	Z585A	Zinc finger protein 585A	S Q L K V H L K V L A G E K L	150
K10	Q52M93	Z585B	Zinc finger protein 585B	S Q F K V H L K V P T G E K L	150
K11	Q9P1Z0	ZBTB4	Zinc finger and BTB domain-containing protein 4	G G P E H V V K V V G G H V L	301
K12	Q9P243	ZFAT	Zinc finger protein ZFAT	K K K Y S D V K N L I K H I R	367
K13	Q86UP3	ZFHX4	Zinc finger homeobox protein 4	V S H L H K L K K V L Q E A S	1525

K14	Q9UJ78	ZMYM5	Zinc finger MYM-type protein 5	G S S N T L L K K I E G I P E	454
K15	Q9BRH9	ZN251	Zinc finger protein 251	K R Y F I H I K K I F Q E R H	663
K16	O60281	ZN292	Zinc finger protein 292	R Q K A S N L K R V N K E K N	2531
K17	Q6ZSB9	ZN509	Zinc finger protein 509	D V F H L D V K N V S G I G Q	70
K18	Q9H582	ZN644	Zinc finger protein 644	D H R R V A V K R V I K E S K	809
K19	P51814	ZNF41	Zinc finger protein 41	N N L L S H V K V L I K E R G	201
K20	P17098	ZNF8	Zinc finger protein 8	L K E Q N N L K Q L E F G L K	157

**Supplementary Table 5: Sequences of the peptide SPOT array shown in Figure 6B.**

Swiss Prot No	ID	Protein name	Sequence	Target position
Q8IZT6	ASPM	Abnormal spindle-like microcephaly-associated protein	Q T Y F N K L <b>K</b> K I T K T V Q	2213
Q8IZT6	ASPM	Abnormal spindle-like microcephaly-associated protein	Q T Y F N K L <b>A</b> K I T K T V Q	2213
P46100	ATRX	Transcriptional regulator ATRX	C H F P K G I <b>K</b> Q I K N G T T	1033
P46100	ATRX	Transcriptional regulator ATRX	C H F P K G I <b>A</b> Q I K N G T T	1033
Q9UIF8	BAZ2B	Bromodomain adjacent to zinc fingerdomain protein 2B	M K Q Q E K I <b>K</b> R I Q Q I R M	948
Q9UIF8	BAZ2B	Bromodomain adjacent to zinc fingerdomain protein 2B	M K Q Q E K I <b>A</b> R I Q Q I R M	948
Q9BVC3	DCC1	Sister chromatid cohesion protein DCC1	R P K L K K L <b>K</b> K L L M E N P	139
Q9BVC3	DCC1	Sister chromatid cohesion protein DCC1	R P K L K K L <b>A</b> K L L M E N P	139
P09884	DPOLA	DNA polymerase alpha catalytic subunit	V E R R K Q V <b>K</b> Q L M K Q Q D	926
P09884	DPOLA	DNA polymerase alpha catalytic subunit	V E R R K Q V <b>A</b> Q L M K Q Q D	926
P41970	ELK3	ETS domain-containing protein Elk-3	Y Y D K N I I <b>K</b> K V I G Q K F	73
P41970	ELK3	ETS domain-containing protein Elk-3	Y Y D K N I I <b>A</b> K V I G Q K F	73
P28324	ELK4	ETS domain-containing protein Elk-4	Y Y V K N I I <b>K</b> K V N G Q K F	73
P28324	ELK4	ETS domain-containing protein Elk-4	Y Y V K N I I <b>A</b> K V N G Q K F	73
O00472	ELL2	RNA polymerase II elongation factor ELL2	H N K L A H I <b>K</b> R L I G E F D	625
O00472	ELL2	RNA polymerase II elongation factor ELL2	H N K L A H I <b>A</b> R L I G E F D	625
Q8IYD8	FANCM	Fanconi anemia group M protein	H K K S S F I <b>K</b> N I N Q G S S	819

Q81YD8	FANCM	Fanconi anemia group M protein	H K K S S F I A N I N Q G S S	819
P38919	IF4A3	Eukaryotic initiation factor 4A-III	A I Q Q R A I K Q I I K G R D	70
P38919	IF4A3	Eukaryotic initiation factor 4A-III	A I Q Q R A I A Q I I K G R D	70
Q86YW9	MD12L	Mediator of RNA polymerase II transcription subunit 12-like protein	R A Y M N L V K K L K K E L G	1604
Q86YW9	MD12L	Mediator of RNA polymerase II transcription subunit 12-like protein	R A Y M N L V A K L K K E L G	1604
A1L020	MEX3A	RNA-binding protein MEX3A	G P K G A T I K R I Q Q Q T N	247
A1L020	MEX3A	RNA-binding protein MEX3A	G P K G A T I A R I Q Q Q T N	247
Q6ZN04	MEX3B	RNA-binding protein MEX3B	G P K G A T I K R I Q Q Q T H	184
Q6ZN04	MEX3B	RNA-binding protein MEX3B	G P K G A T I A R I Q Q Q T H	184
Q5U5Q3	MEX3C	RNA-binding protein MEX3C	G P K G A T I K R I Q Q Q T H	350
Q5U5Q3	MEX3C	RNA-binding protein MEX3C	G P K G A T I A R I Q Q Q T H	350
Q9NPC7	MYNN	Myoneurin	G N S Y T D I K N L K K H K T	512
Q9NPC7	MYNN	Myoneurin	G N S Y T D I A N L K K H K T	512
Q8N3S3	PHTF2	Putative homeodomain transcription factor 2	E V P H F R L K K V Q N I K M	574
Q8N3S3	PHTF2	Putative homeodomain transcription factor 2	E V P H F R L A K V Q N I K M	574
P54278	PMS2	Mismatch repair endonuclease PMS2	S S L A K R I K Q L H H E A Q	630
P54278	PMS2	Mismatch repair endonuclease PMS2	S S L A K R I A Q L H H E A Q	630
Q7Z3G6	PRIC2	Prickle-like protein 2	P G E K L R I K Q L L H Q L P	74
Q7Z3G6	PRIC2	Prickle-like protein 2	P G E K L R I A Q L L H Q L P	74
Q33E94	RFX4	Transcription factor RFX4	A K R Q G S L K K V A Q Q F L	421
Q33E94	RFX4	Transcription factor RFX4	A K R Q G S L A K V A Q Q F L	421
Q9NQR1	SETD8	Histone-lysine N-methyltransferase SET8	A I A K Q A L K K P I K G K Q	158
Q9NQR1	SETD8	Histone-lysine N-methyltransferase SET8	A I A K Q A L A K P I K G K Q	158



Q96DI7	SNR40	U5 small nuclear ribonucleoprotein 40 kDa protein	S E T G E R V <b>K</b> R L K G H T S	145
Q96DI7	SNR40	U5 small nuclear ribonucleoprotein 40 kDa protein	S E T G E R V <b>A</b> R L K G H T S	145
Q9H6I2	SOX17	Transcription factor SOX-17	R K Q V K R L <b>K</b> R V E G G F L	149
Q9H6I2	SOX17	Transcription factor SOX-17	R K Q V K R L <b>A</b> R V E G G F L	149
Q9UEW8	STK39	STE20/SPS1-related proline-alanine-rich protein kinase	R Q E R V A I <b>K</b> R I N L E K C	92
Q9UEW8	STK39	STE20/SPS1-related proline-alanine-rich protein kinase	R Q E R V A I <b>A</b> R I N L E K C	92
Q9Y3A2	UTP11	Probable U3 small nucleolar RNA-associated protein 11	V T N Q T G L <b>K</b> R I A K E R Q	189
Q9Y3A2	UTP11	Probable U3 small nucleolar RNA-associated protein 11	V T N Q T G L <b>A</b> R I A K E R Q	189
O60281	ZN292	Zinc finger protein 292	R Q K A S N L <b>K</b> R V N K E K N	2531
O60281	ZN292	Zinc finger protein 292	R Q K A S N L <b>A</b> R V N K E K N	2531

## Supplementary References

- Al Temimi, A., Reddy, Y. V., White, P. B., Guo, H., Qian, P., & Mecinovic, J. (2017). Lysine Possesses the Optimal Chain Length for Histone Lysine Methyltransferase Catalysis. *Sci Rep*, 7(1), 16148. <https://doi.org/10.1038/s41598-017-16128-4>
- Al Temimi, A. H. K., Teeuwen, R. S., Tran, V., Altunc, A. J., Lenstra, D. C., Ren, W., Qian, P., Guo, H., & Mecinovic, J. (2019). Importance of the main chain of lysine for histone lysine methyltransferase catalysis. *Org Biomol Chem*, 17(23), 5693-5697. <https://doi.org/10.1039/c9ob01038f>
- Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat Rev Genet*, 17(8), 487-500. <https://doi.org/10.1038/nrg.2016.59>
- An, S., Yeo, K. J., Jeon, Y. H., & Song, J. J. (2011). Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism. *J Biol Chem*, 286(10), 8369-8374. <https://doi.org/10.1074/jbc.M110.203380>
- Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res*, 21(3), 381-395. <https://doi.org/10.1038/cr.2011.22>
- Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C., & Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem*, 280(18), 17732-17736. <https://doi.org/10.1074/jbc.M500796200>
- Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., Akalin, A., & Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature*, 520(7546), 243-247. <https://doi.org/10.1038/nature14176>
- Bea, S., Valdes-Mas, R., Navarro, A., Salaverria, I., Martin-Garcia, D., Jares, P., Gine, E., Pinyol, M., Royo, C., Nadeu, F., Conde, L., Juan, M., Clot, G., Vizan, P., Di Croce, L., Puente, D. A., Lopez-Guerra, M., Moros, A., Roue, G., . . . Campo, E. (2013). Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. *Proc Natl Acad Sci U S A*, 110(45), 18250-18255. <https://doi.org/10.1073/pnas.1314608110>
- Bennett, R. L., Swaroop, A., Troche, C., & Licht, J. D. (2017). The Role of Nuclear Receptor-Binding SET Domain Family Histone Lysine Methyltransferases in Cancer. *Cold Spring Harb Perspect Med*, 7(6). <https://doi.org/10.1101/cshperspect.a026708>
- Bergemann, A. D., Cole, F., & Hirschhorn, K. (2005). The etiology of Wolf-Hirschhorn syndrome. *Trends Genet*, 21(3), 188-195. <https://doi.org/10.1016/j.tig.2005.01.008>
- Bestor, T., Laudano, A., Mattaliano, R., & Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of Molecular Biology*, 203(4), 971-983. [https://doi.org/https://doi.org/10.1016/0022-2836\(88\)90122-2](https://doi.org/https://doi.org/10.1016/0022-2836(88)90122-2)
- Biggar, K. K., & Li, S. S. (2015). Non-histone protein methylation as a regulator of cellular signalling and function. *Nat Rev Mol Cell Biol*, 16(1), 5-17. <https://doi.org/10.1038/nrm3915>
- Bock, I., Kudithipudi, S., Tamas, R., Kungulovski, G., Dhayalan, A., & Jeltsch, A. (2011). Application of Cellusspots peptide arrays for the analysis of the binding specificity of epigenetic reading domains to modified histone tails. *BMC Biochemistry*, 12(48). <https://doi.org/10.1186/1471-2091-12-48>

- Boriack-Sjodin, P. A., & Swinger, K. K. (2016). Protein Methyltransferases: A Distinct, Diverse, and Dynamic Family of Enzymes. *Biochemistry*, 55(11), 1557-1569. <https://doi.org/10.1021/acs.biochem.5b01129>
- Borkin, D., He, S., Miao, H., Kempinska, K., Pollock, J., Chase, J., Purohit, T., Malik, B., Zhao, T., Wang, J., Wen, B., Zong, H., Jones, M., Danet-Desnoyers, G., Guzman, M. L., Talpaz, M., Bixby, D. L., Sun, D., Hess, J. L., . . . Grembecka, J. (2015). Pharmacologic inhibition of the Menin-MLL interaction blocks progression of MLL leukemia in vivo. *Cancer Cell*, 27(4), 589-602. <https://doi.org/10.1016/j.ccell.2015.02.016>
- Bourc'his, D., Xu, G.-L., Lin, C.-S., Bollman, B., & Bestor, T. H. (2001). Dnmt3L and the Establishment of Maternal Genomic Imprints. *Science*, 294(5551), 2536-2539. <https://doi.org/10.1126/science.1065848>
- Brabletz, T., Jung, A., Spaderna, S., Hlubek, F., & Kirchner, T. (2005). Migrating cancer stem cells — an integrated concept of malignant tumour progression. *Nature Reviews Cancer*, 5(9), 744-749. <https://doi.org/10.1038/nrc1694>
- Brohm, A., Elsayy, H., Rathert, P., Kudithipudi, S., Schoch, T., Schuhmacher, M. K., Weirich, S., & Jeltsch, A. (2019). Somatic Cancer Mutations in the SUV420H1 Protein Lysine Methyltransferase Modulate Its Catalytic Activity. *J Mol Biol*, 431(17), 3068-3080. <https://doi.org/10.1016/j.jmb.2019.06.021>
- Bussi, G., & Parrinello, M. (2007). Accurate sampling using Langevin dynamics. *Phys Rev E Stat Nonlin Soft Matter Phys*, 75(5 Pt 2), 056707. <https://doi.org/10.1103/PhysRevE.75.056707>
- Cancer Genome Atlas, N. (2015). Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, 517(7536), 576-582. <https://doi.org/10.1038/nature14129>
- Cao, F., Townsend, E. C., Karatas, H., Xu, J., Li, L., Lee, S., Liu, L., Chen, Y., Ouillette, P., Zhu, J., Hess, J. L., Atadja, P., Lei, M., Qin, Z. S., Malek, S., Wang, S., & Dou, Y. (2014). Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol Cell*, 53(2), 247-261. <https://doi.org/10.1016/j.molcel.2013.12.001>
- Case, D. A., V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossvary, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollman. (2014). AMBER 14. *University of California, San Francisco*.
- Chang, C.-E. G., Michael K. (2004). Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. *Journal of the American Chemical Society*, 40, 9. <https://doi.org/13156>–13164
- Chen, K., Liu, J., Liu, S., Xia, M., Zhang, X., Han, D., Jiang, Y., Wang, C., & Cao, X. (2017). Methyltransferase SETD2-Mediated Methylation of STAT1 Is Critical for Interferon Antiviral Activity. *Cell*, 170(3), 492-506 e414. <https://doi.org/10.1016/j.cell.2017.06.042>
- Chen, R., Chen, Y., Zhao, W., Fang, C., Zhou, W., Yang, X., & Ji, M. (2020). The Role of Methyltransferase NSD2 as a Potential Oncogene in Human Solid Tumors. *Onco Targets Ther*, 13, 6837-6846. <https://doi.org/10.2147/OTT.S259873>

- Chen, S., Kapilashrami, K., Senevirathne, C., Wang, Z., Wang, J., Linscott, J. A., & Luo, M. (2019). Substrate-Differentiated Transition States of SET7/9-Catalyzed Lysine Methylation. *J Am Chem Soc*, 141(20), 8064-8067. <https://doi.org/10.1021/jacs.9b02553>
- Cheng, X., Collins, R. E., & Zhang, X. (2005). Structural and sequence motifs of protein (histone) methylation enzymes. *Annu Rev Biophys Biomol Struct*, 34, 267-294. <https://doi.org/10.1146/annurev.biophys.34.040204.144452>
- Cheng, X., & Zhang, X. (2007). Structural dynamics of protein lysine methylation and demethylation. *Mutat Res*, 618(1-2), 102-115. <https://doi.org/10.1016/j.mrfmmm.2006.05.041>
- Cheong, C. M., Mrozik, K. M., Hewett, D. R., Bell, E., Panagopoulos, V., Noll, J. E., Licht, J. D., Gronthos, S., Zannettino, A. C. W., & Vandyke, K. (2020). Twist-1 is upregulated by NSD2 and contributes to tumour dissemination and an epithelial-mesenchymal transition-like gene expression signature in t(4;14)-positive multiple myeloma. *Cancer Lett*, 475, 99-108. <https://doi.org/10.1016/j.canlet.2020.01.040>
- Chmiela, S., Saucedo, H. E., Muller, K. R., & Tkatchenko, A. (2018). Towards exact molecular dynamics simulations with machine-learned force fields. *Nat Commun*, 9(1), 3887. <https://doi.org/10.1038/s41467-018-06169-2>
- Chu, Y., Yao, J., & Guo, H. (2012). QM/MM MD and free energy simulations of G9a-like protein (GLP) and its mutants: understanding the factors that determine the product specificity. *PLoS One*, 7(5), e37674. <https://doi.org/10.1371/journal.pone.0037674>
- Clarke, S. G. (2013). Protein methylation at the surface and buried deep: thinking outside the histone box. *Trends Biochem Sci*, 38(5), 243-252. <https://doi.org/10.1016/j.tibs.2013.02.004>
- Cole, P. A. (2008). Chemical probes for histone-modifying enzymes. *Nature Chemical Biology*, 4(10), 590-597. <https://doi.org/10.1038/nchembio.111>
- Collins, R. E., Tachibana, M., Tamaru, H., Smith, K. M., Jia, D., Zhang, X., Selker, E. U., Shinkai, Y., & Cheng, X. (2005). In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *J Biol Chem*, 280(7), 5563-5570. <https://doi.org/10.1074/jbc.M410483200>
- Copeland, R. A., Solomon, M. E., & Richon, V. M. (2009). Protein methyltransferases as a target class for drug discovery. *Nat Rev Drug Discov*, 8(9), 724-732. <https://doi.org/10.1038/nrd2974>
- Cornett, E. M., Ferry, L., Defossez, P. A., & Rothbart, S. B. (2019). Lysine Methylation Regulators Moonlighting outside the Epigenome. *Mol Cell*, 75(6), 1092-1101. <https://doi.org/10.1016/j.molcel.2019.08.026>
- Cortopassi, W. A., Kumar, K., Duarte, F., Pimentel, A. S., & Paton, R. S. (2016). Mechanisms of histone lysine-modifying enzymes: A computational perspective on the role of the protein environment. *J Mol Graph Model*, 67, 69-84. <https://doi.org/10.1016/j.jmgm.2016.04.011>
- Couture, J. F., Collazo, E., Brunzelle, J. S., & Trievel, R. C. (2005). Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev*, 19(12), 1455-1465. <https://doi.org/10.1101/gad.1318405>
- Couture, J. F., Collazo, E., Hauk, G., & Trievel, R. C. (2006). Structural basis for the methylation site specificity of SET7/9. *Nat Struct Mol Biol*, 13(2), 140-146. <https://doi.org/10.1038/nsmb1045>

- Dasetty, S., Meza-Morales, P. J., Getman, R. B., & Sarupria, S. (2019). Simulations of interfacial processes: recent advances in force field development. *Current Opinion in Chemical Engineering*, 23, 138-145. <https://doi.org/10.1016/j.coche.2019.04.003>
- David W.H. Swenson, S. R. (2017). *Contact Map Explorer*. In [https://github.com/dwhswenson/contact\\_map](https://github.com/dwhswenson/contact_map). <https://contact-map.readthedocs.io/en/latest/index.html>
- Del Rizzo, P. A., Couture, J. F., Dirk, L. M., Strunk, B. S., Roiko, M. S., Brunzelle, J. S., Houtz, R. L., & Trievel, R. C. (2010). SET7/9 catalytic mutants reveal the role of active site water molecules in lysine multiple methylation. *J Biol Chem*, 285(41), 31849-31858. <https://doi.org/10.1074/jbc.M110.114587>
- Del Rizzo, P. A., & Trievel, R. C. (2014). Molecular basis for substrate recognition by lysine methyltransferases and demethylases. *Biochim Biophys Acta*, 1839(12), 1404-1415. <https://doi.org/10.1016/j.bbagr.2014.06.008>
- Dhayalan, A., Kudithipudi, S., Rathert, P., & Jeltsch, A. (2011). Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. *Chem Biol*, 18(1), 111-120. <https://doi.org/10.1016/j.chembiol.2010.11.014>
- DiFiore, J. V., Ptacek, T. S., Wang, Y., Li, B., Simon, J. M., & Strahl, B. D. (2020). Unique and Shared Roles for Histone H3K36 Methylation States in Transcription Regulation Functions. *Cell Rep*, 31(10), 107751. <https://doi.org/10.1016/j.celrep.2020.107751>
- Dillon, S. C., Zhang, X., Trievel, R. C., & Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol*, 6(8), 227. <https://doi.org/10.1186/gb-2005-6-8-227>
- Dirk, L. M., Flynn, E. M., Dietzel, K., Couture, J. F., Trievel, R. C., & Houtz, R. L. (2007). Kinetic Manifestation of Processivity during Multiple Methylations Catalyzed by SET Domain Protein Methyltransferases. *Biochemistry*, 46, 3905-3915. <https://doi.org/10.1021/bi6023644>
- Dukatz, M., Adam, S., Biswal, M., Song, J., Bashtrykov, P., & Jeltsch, A. (2020). Complex DNA sequence readout mechanisms of the DNMT3B DNA methyltransferase. *Nucleic Acids Res*, 48(20), 11495-11509. <https://doi.org/10.1093/nar/gkaa938>
- Dukatz, M., Dittrich, M., Stahl, E., Adam, S., de Mendoza, A., Bashtrykov, P., & Jeltsch, A. (2022). DNA methyltransferase DNMT3A forms interaction networks with the CpG site and flanking sequence elements for efficient methylation. *J Biol Chem*, 298(10), 102462. <https://doi.org/10.1016/j.jbc.2022.102462>
- Dukatz, M., Holzer, K., Choudalakis, M., Emperle, M., Lungu, C., Bashtrykov, P., & Jeltsch, A. (2019). H3K36me2/3 Binding and DNA Binding of the DNA Methyltransferase DNMT3A PWWP Domain Both Contribute to its Chromatin Interaction. *J Mol Biol*, 431(24), 5063-5074. <https://doi.org/10.1016/j.jmb.2019.09.006>
- Eastman, P., & Pande, V. S. (2010). OpenMM: A Hardware Independent Framework for Molecular Simulations. *Computing in science & engineering*, 12(4), 34-39. <https://doi.org/10.1109/MCSE.2010.27>
- Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L. P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R., & Pande, V. S. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol*, 13(7), e1005659. <https://doi.org/10.1371/journal.pcbi.1005659>

- Edmunds, J. W., Mahadevan, L. C., & Clayton, A. L. (2008). Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J*, 27(2), 406-420. <https://doi.org/10.1038/sj.emboj.7601967>
- Emperle, M., Adam, S., Kunert, S., Dukatz, M., Baude, A., Plass, C., Rathert, P., Bashtrykov, P., & Jeltsch, A. (2019). Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. *Nucleic Acids Res*, 47(21), 11355-11367. <https://doi.org/10.1093/nar/gkz911>
- Emperle, M., Bangalore, D. M., Adam, S., Kunert, S., Heil, H. S., Heinze, K. G., Bashtrykov, P., Tessmer, I., & Jeltsch, A. (2021). Structural and biochemical insight into the mechanism of dual CpG site binding and methylation by the DNMT3A DNA methyltransferase. *Nucleic Acids Res*, 49(14), 8294-8308. <https://doi.org/10.1093/nar/gkab600>
- Emperle, M., Dukatz, M., Kunert, S., Holzer, K., Rajavelu, A., Jurkowska, R. Z., & Jeltsch, A. (2018). The DNMT3A R882H mutation does not cause dominant negative effects in purified mixed DNMT3A/R882H complexes. *Sci Rep*, 8(1), 13242. <https://doi.org/10.1038/s41598-018-31635-8>
- Ezponda, T., Popovic, R., Shah, M. Y., Martinez-Garcia, E., Zheng, Y., Min, D. J., Will, C., Neri, A., Kelleher, N. L., Yu, J., & Licht, J. D. (2013). The histone methyltransferase MMSET/WHSC1 activates TWIST1 to promote an epithelial-mesenchymal transition and invasive properties of prostate cancer. *Oncogene*, 32(23), 2882-2890. <https://doi.org/10.1038/onc.2012.297>
- Faller, R., & de Pablo, J. J. (2002). Constant pressure hybrid Molecular Dynamics–Monte Carlo simulations. *The Journal of Chemical Physics*, 116(1). <https://doi.org/10.1063/1.1420460>
- Falnes, Pål Ø., Jakobsson, Magnus E., Davydova, E., Ho, A., & Małeckki, J. (2016). Protein lysine methylation by seven-β-strand methyltransferases. *Biochemical Journal*, 473(14), 1995-2009. <https://doi.org/10.1042/bcj20160117>
- Fang, D., Gan, H., Lee, J., Han, J., Wang, Z., Riester, S. M., Jin, L., Chen, J., Zhou, H., Wang, J., Zhang, H., Yang, N., Bradley, E. W., Ho, T. H., Rubin, B. P., Bridge, J. A., Thibodeau, S. N., Ordog, T., Chen, Y., . . . Zhang, Z. (2016). The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. *Science*, 352, 1344-1348. <https://doi.org/10.1126/science.aae0065>
- Fatemi, M., Hermann, A., Pradhan, S., & Jeltsch, A. (2001). The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. Edited by J. Karn. *Journal of Molecular Biology*, 309(5), 1189-1199. <https://doi.org/10.1006/jmbi.2001.4709>
- Fegan, A. W., Brian; Carlson, Jonathan C. T.; Wagner, Carston R. (2010). Chemically controlled protein assembly: techniques and applications. *Chemical reviews*, 6, 22. <https://doi.org/10.1021/cr8002888>
- Feoli, A., Viviano, M., Cipriano, A., Milite, C., Castellano, S., & Sbardella, G. (2022). Lysine methyltransferase inhibitors: where we are now. *RSC Chem Biol*, 3(4), 359-406. <https://doi.org/10.1039/d1cb00196e>
- Ferguson, A. D., Larsen, N. A., Howard, T., Pollard, H., Green, I., Grande, C., Cheung, T., Garcia-Arenas, R., Cowen, S., Wu, J., Godin, R., Chen, H., & Keen, N. (2011). Structural basis of substrate methylation and inhibition of SMYD2. *Structure*, 19(9), 1262-1273. <https://doi.org/10.1016/j.str.2011.06.011>

- Fontebasso, A. M., Schwartzenruber, J., Khuong-Quang, D. A., Liu, X. Y., Sturm, D., Korshunov, A., Jones, D. T., Witt, H., Kool, M., Albrecht, S., Fleming, A., Hadjadj, D., Busche, S., Lepage, P., Montpetit, A., Staffa, A., Gerges, N., Zakrzewska, M., Zakrzewski, K., . . . Majewski, J. (2013). Mutations in SETD2 and genes affecting histone H3K36 methylation target hemispheric high-grade gliomas. *Acta Neuropathol*, *125*(5), 659-669. <https://doi.org/10.1007/s00401-013-1095-8>
- Frank, R. (2002). The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports--principles and applications [Review]. *J Immunol Methods*, *267*(1), 13-26. [https://doi.org/10.1016/s0022-1759\(02\)00137-0](https://doi.org/10.1016/s0022-1759(02)00137-0)
- Gaughan, L., Stockley, J., Wang, N., McCracken, S. R., Treumann, A., Armstrong, K., Shaheen, F., Watt, K., McEwan, I. J., Wang, C., Pestell, R. G., & Robson, C. N. (2011). Regulation of the androgen receptor by SET9-mediated methylation. *Nucleic Acids Res*, *39*(4), 1266-1279. <https://doi.org/10.1093/nar/gkq861>
- George, R. A., Lin, K., & Heringa, J. (2005). Scooby-domain: prediction of globular domains in protein sequence. *Nucleic Acids Res*, *33*(Web Server issue), W160-163. <https://doi.org/10.1093/nar/gki381>
- Gowher, H., & Jeltsch, A. (2001). Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *J Mol Biol*, *309*(5), 1201-1208. <https://doi.org/10.1006/jmbi.2001.4710>
- Gowher, H., & Jeltsch, A. (2018). Mammalian DNA methyltransferases: new discoveries and open questions. *Biochemical Society Transactions*, *46*(5), 1191-1202. <https://doi.org/10.1042/BST20170574>
- Goyal, R., Reinhardt, R., & Jeltsch, A. (2006). Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. *Nucleic Acids Research*, *34*(4), 1182-1188. <https://doi.org/10.1093/nar/gkl002>
- Grebien, F., Vedadi, M., Getlik, M., Giambruno, R., Grover, A., Avellino, R., Skucha, A., Vittori, S., Kuznetsova, E., Smil, D., Barsyte-Lovejoy, D., Li, F., Poda, G., Schapira, M., Wu, H., Dong, A., Senisterra, G., Stukalov, A., Huber, K. V. M., . . . Superti-Furga, G. (2015). Pharmacological targeting of the Wdr5-MLL interaction in C/EBPalpha N-terminal leukemia. *Nat Chem Biol*, *11*(8), 571-578. <https://doi.org/10.1038/nchembio.1859>
- Gregory, G. D., Vakoc, C. R., Rozovskaia, T., Zheng, X., Patel, S., Nakamura, T., Canaani, E., & Blobel, G. A. (2007). Mammalian ASH1L Is a Histone Methyltransferase That Occupies the Transcribed Region of Active Genes. *Molecular and Cellular Biology*, *27*(24), 8466-8479. <https://doi.org/10.1128/MCB.00993-07>
- Hade, M. D., Suire, C. N., & Suo, Z. (2023). An Effective Peptide-Based Platform for Efficient Exosomal Loading and Cellular Delivery of a microRNA. *ACS Appl Mater Interfaces*, *15*(3), 3851-3866. <https://doi.org/10.1021/acsami.2c20728>
- Hajdu, I., Ciccia, A., Lewis, S. M., & Elledge, S. J. (2011). Wolf-Hirschhorn syndrome candidate 1 is involved in the cellular response to DNA damage. *Proc Natl Acad Sci U S A*, *108*(32), 13130-13134. <https://doi.org/10.1073/pnas.1110081108>
- Hamamoto, R., Saloura, V., & Nakamura, Y. (2015). Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nature Reviews Cancer*, *15*(2), 110-124. <https://doi.org/10.1038/nrc3884>

- Hanley, R. P., Nie, D. Y., Tabor, J. R., Li, F., Sobh, A., Xu, C., Barker, N. K., Dilworth, D., Hajian, T., Gibson, E., Szewczyk, M. M., Brown, P. J., Barsyte-Lovejoy, D., Herring, L. E., Wang, G. G., Licht, J. D., Vedadi, M., Arrowsmith, C. H., & James, L. I. (2023). Discovery of a Potent and Selective Targeted NSD2 Degrader for the Reduction of H3K36me2. *Journal of the American Chemical Society*, 145(14), 8176-8188. <https://doi.org/10.1021/jacs.3c01421>
- Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. *Neuron*, 99(6), 1129-1143. <https://doi.org/10.1016/j.neuron.2018.08.011>
- Horn, H. W., Swope, W. C., Pitner, J. W., Madura, J. D., Dick, T. J., Hura, G. L., & Head-Gordon, T. (2004). Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys*, 120(20), 9665-9678. <https://doi.org/10.1063/1.1683075>
- Hospital, A., Goni, J. R., Orozco, M., & Gelpi, J. L. (2015). Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem*, 8, 37-47. <https://doi.org/10.2147/AABC.S70333>
- Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., & Eden, A. (2008). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene*, 27(3), 404-408. <https://doi.org/10.1038/sj.onc.1210631>
- Hsu, P. L., Shi, H., Leonen, C., Kang, J., Chatterjee, C., & Zheng, N. (2019). Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS. *Mol Cell*, 76(5), 712-723 e714. <https://doi.org/10.1016/j.molcel.2019.10.013>
- Hu, P., & Zhang, Y. (2006). Catalytic Mechanism and Product Specificity of the Histone Lysine Methyltransferase SET7/9: An ab Initio QM/MM-FE Study with Multiple Initial Structures. *J Biol Chem*, 281(10), 1272-1278. <https://doi.org/10.1021/ja056153>
- Huang, H., Lin, S., Garcia, B. A., & Zhao, Y. (2015). Quantitative proteomic analysis of histone modifications. *Chem Rev*, 115(6), 2376-2418. <https://doi.org/10.1021/cr500491u>
- Huang, X. X., LeDuc, R. D., Fornelli, L., Schunter, A. J., Bennett, R. L., Kelleher, N. L., & Licht, J. D. (2019). Defining the NSD2 interactome: PARP1 PARylation reduces NSD2 histone methyltransferase activity and impedes chromatin binding. *Journal of Biological Chemistry*, 294(33), 12459-12471. <https://doi.org/10.1074/jbc.RA118.006159>
- Husmann, D., & Gozani, O. (2019). Histone lysine methyltransferases in biology and disease. *Nature Structural & Molecular Biology*, 26(10), 880-889. <https://doi.org/10.1038/s41594-019-0298-7>
- Husmann, D., & Gozani, O. (2019). Histone lysine methyltransferases in biology and disease. *Nat Struct Mol Biol*, 26(10), 880-889. <https://doi.org/10.1038/s41594-019-0298-7>
- Hyun, K., Jeon, J., Park, K., & Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Experimental & Molecular Medicine*, 49(4), e324-e324. <https://doi.org/10.1038/emm.2017.11>
- Iglesias, N., Currie, M. A., Jih, G., Paulo, J. A., Siuti, N., Kalocsay, M., Gygi, S. P., & Moazed, D. (2018). Automethylation-induced conformational switch in Clr4 (Suv39h) maintains epigenetic stability. *Nature*, 560(7719), 504-508. <https://doi.org/10.1038/s41586-018-0398-2>
- Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science*, 333(6047), 1300-1303. <https://doi.org/10.1126/science.1210597>
- Jaffe, J. D., Wang, Y., Chan, H. M., Zhang, J., Huether, R., Kryukov, G. V., Bhang, H. E., Taylor, J. E., Hu, M., Englund, N. P., Yan, F., Wang, Z., Robert McDonald, E., 3rd, Wei, L., Ma, J., Easton, J., Yu,



- Z., deBeaumont, R., Gibaja, V., . . . Stegmeier, F. (2013). Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat Genet*, *45*(11), 1386-1391. <https://doi.org/10.1038/ng.2777>
- Jambhekar, A., Dhall, A., & Shi, Y. (2019). Roles and regulation of histone methylation in animal development. *Nat Rev Mol Cell Biol*, *20*(10), 625-641. <https://doi.org/10.1038/s41580-019-0151-1>
- Janson, G., & Paiardini, A. (2021). PyMod 3: a complete suite for structural bioinformatics in PyMOL. *Bioinformatics*, *37*(10), 1471-1472. <https://doi.org/10.1093/bioinformatics/btaa849>
- Jeltsch, A., Adam, S., Dukatz, M., Emperle, M., & Bashtrykov, P. (2021). Deep Enzymology Studies on DNA Methyltransferases Reveal Novel Connections between Flanking Sequences and Enzyme Activity. *J Mol Biol*, *433*(19), 167186. <https://doi.org/10.1016/j.jmb.2021.167186>
- Jeltsch, A., & Lanio, T. (2002). Site-directed mutagenesis by polymerase chain reaction. *Methods Mol Biol*, *182*, 85-94. <https://doi.org/10.1385/1-59259-194-9:085>
- Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., & Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature*, *449*(7159), 248-251. <https://doi.org/10.1038/nature06146>
- Jones, P. L., Veenstra, G. J. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., & Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics*, *19*(2), 187-191. <https://doi.org/10.1038/561>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Junmei Wang, W. W., Peter A. Kollman and David A. Case. (2001). Antechamber, An Accessory Software Package For Molecular Mechanical Calculations. *Journal of Chemical Information and Computer Sciences*, *222*, 41.
- Jurkowska, R. Z., Anspach, N., Urbanke, C., Jia, D., Reinhardt, R., Nellen, W., Cheng, X., & Jeltsch, A. (2008). Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic Acids Res*, *36*(21), 6656-6663. <https://doi.org/10.1093/nar/gkn747>
- Jurkowska, R. Z., Jurkowski, T. P., & Jeltsch, A. (2011). Structure and function of mammalian DNA methyltransferases. *Chembiochem*, *12*(2), 206-222. <https://doi.org/10.1002/cbic.201000195>
- Jurkowska, R. Z., Rajavelu, A., Anspach, N., Urbanke, C., Jankevicius, G., Ragozin, S., Nellen, W., & Jeltsch, A. (2011). Oligomerization and binding of the Dnmt3a DNA methyltransferase to parallel DNA molecules: heterochromatic localization and role of Dnmt3L. *J Biol Chem*, *286*(27), 24200-24207. <https://doi.org/10.1074/jbc.M111.254987>
- Kang, H. B., Choi, Y., Lee, J. M., Choi, K. C., Kim, H. C., Yoo, J. Y., Lee, Y. H., & Yoon, H. G. (2009). The histone methyltransferase, NSD2, enhances androgen receptor-mediated transcription. *FEBS Lett*, *583*(12), 1880-1886. <https://doi.org/10.1016/j.febslet.2009.05.038>

- Khella, M. S., Brohm, A., Weirich, S., & Jeltsch, A. (2020). Mechanistic Insights into the Allosteric Regulation of the Clr4 Protein Lysine Methyltransferase by Autoinhibition and Automethylation. *Int J Mol Sci*, 21(22). <https://doi.org/10.3390/ijms21228832>
- Khella, M. S., Schnee, P., Weirich, S., Bui, T., Brohm, A., Bashtrykov, P., Pleiss, J., & Jeltsch, A. (2023a). The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation. *J Biol Chem*, 104796. <https://doi.org/10.1016/j.jbc.2023.104796>
- Khella, M. S., Schnee, P., Weirich, S., Bui, T., Brohm, A., Bashtrykov, P., Pleiss, J., & Jeltsch, A. (2023b). The T1150A cancer mutant of the protein lysine dimethyltransferase NSD2 can introduce H3K36 trimethylation. *J Biol Chem*, 299(6), 104796. <https://doi.org/10.1016/j.jbc.2023.104796>
- Kim, G., Lee, Y., Ha, J., Han, S., & Lee, M. (2021). Engineering exosomes for pulmonary delivery of peptides and drugs to inflammatory lung cells by inhalation. *J Control Release*, 330, 684-695. <https://doi.org/10.1016/j.jconrel.2020.12.053>
- Kim, J. Y., Kee, H. J., Choe, N. W., Kim, S. M., Eom, G. H., Baek, H. J., Kook, H., Kook, H., & Seo, S. B. (2008). Multiple-myeloma-related WHSC1/MMSET isoform RE-IIBP is a histone methyltransferase with transcriptional repression activity. *Mol Cell Biol*, 28(6), 2023-2034. <https://doi.org/10.1128/MCB.02130-07>
- Kim, S. J., Zhao, H., Hardikar, S., Singh, A. K., Goodell, M. A., & Chen, T. (2013). A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. *Blood*, 122(25), 4086-4089. <https://doi.org/10.1182/blood-2013-02-483487>
- Klose, R. J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., & Zhang, Y. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature*, 442(7100), 312-316. <https://doi.org/10.1038/nature04853>
- Ko, S., Ahn, J., Song, C. S., Kim, S., Knapczyk-Stwora, K., & Chatterjee, B. (2011). Lysine methylation and functional modulation of androgen receptor by Set9 methyltransferase. *Mol Endocrinol*, 25(3), 433-444. <https://doi.org/10.1210/me.2010-0482>
- Kong, X., Ouyang, S., Liang, Z., Lu, J., Chen, L., Shen, B., Li, D., Zheng, M., Li, K. K., Luo, C., & Jiang, H. (2011). Catalytic mechanism investigation of lysine-specific demethylase 1 (LSD1): a computational study. *PLoS One*, 6(9), e25444. <https://doi.org/10.1371/journal.pone.0025444>
- Kontaki, H., & Talianidis, I. (2010). Lysine Methylation Regulates E2F1-Induced Cell Death. *Molecular Cell*, 39(1), 152-160. <https://doi.org/https://doi.org/10.1016/j.molcel.2010.06.006>
- Kudithipudi, S., Dhayalan, A., Kebede, A. F., & Jeltsch, A. (2012). The SET8 H4K20 protein lysine methyltransferase has a long recognition sequence covering seven amino acid residues. *Biochimie*, 94(11), 2212-2218. <https://doi.org/10.1016/j.biochi.2012.04.024>
- Kudithipudi, S., & Jeltsch, A. (2014). Role of somatic cancer mutations in human protein lysine methyltransferases. *Biochim Biophys Acta*, 1846(2), 366-379. <https://doi.org/10.1016/j.bbcan.2014.08.002>
- Kudithipudi, S., & Jeltsch, A. (2016). Approaches and Guidelines for the Identification of Novel Substrates of Protein Lysine Methyltransferases. *Cell Chem Biol*, 23(9), 1049-1055. <https://doi.org/10.1016/j.chembiol.2016.07.013>
- Kudithipudi, S., Jeltsch, A. (2014). Role of somatic cancer mutations in human protein lysine methyltransferases. *Biochim Biophys Acta*, 1846(2), 366-379. <https://doi.org/10.1016/j.bbcan.2014.08.002>

- Kudithipudi, S., Kusevic, D., Weirich, S., & Jeltsch, A. (2014). Specificity analysis of protein lysine methyltransferases using SPOT peptide arrays. *J Vis Exp*(93), e52203. <https://doi.org/10.3791/52203>
- Kudithipudi, S., Lungu, C., Rathert, P., Happel, N., & Jeltsch, A. (2014). Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1. *Chem Biol*, 21(2), 226-237. <https://doi.org/10.1016/j.chembiol.2013.10.016>
- Kumar, A., & Kono, H. (2020). Heterochromatin protein 1 (HP1): interactions with itself and chromatin components. *Biophys Rev*, 12(2), 387-400. <https://doi.org/10.1007/s12551-020-00663-y>
- Kusevic, D., Kudithipudi, S., Iglesias, N., Moazed, D., & Jeltsch, A. (2017). Clr4 specificity and catalytic activity beyond H3K9 methylation. *Biochimie*, 135, 83-88. <https://doi.org/10.1016/j.biochi.2017.01.013>
- Kwon, T., Chang, J. H., Kwak, E., Lee, C. W., Joachimiak, A., Kim, Y. C., Lee, J., & Cho, Y. (2003). Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. *EMBO J*, 22(2), 292-303. <https://doi.org/10.1093/emboj/cdg025>
- Lam, U. T. F., Tan, B. K. Y., Poh, J. J. X., & Chen, E. S. (2022). Structural and functional specificity of H3K36 methylation. *Epigenetics Chromatin*, 15(1), 17. <https://doi.org/10.1186/s13072-022-00446-7>
- Lee, C. H., Yu, J. R., Granat, J., Saldana-Meyer, R., Andrade, J., LeRoy, G., Jin, Y., Lund, P., Stafford, J. M., Garcia, B. A., Ueberheide, B., & Reinberg, D. (2019). Automethylation of PRC2 promotes H3K27 methylation and is impaired in H3K27M pediatric glioma. *Genes Dev*, 33(19-20), 1428-1440. <https://doi.org/10.1101/gad.328773.119>
- Lee, M. S., & Olson, M. A. (2006). Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys J*, 90(3), 864-877. <https://doi.org/10.1529/biophysj.105.071589>
- Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. E., Kandoth, C., Payton, J. E., Baty, J., Welch, J., Harris, C. C., Lichti, C. F., Townsend, R. R., Fulton, R. S., Dooling, D. J., Koboldt, D. C., Schmidt, H., Zhang, Q., Osborne, J. R., . . . Wilson, R. K. (2010). DNMT3A Mutations in Acute Myeloid Leukemia. *New England Journal of Medicine*, 363(25), 2424-2433. <https://doi.org/10.1056/NEJMoa1005143>
- Li, J., Ahn, J. H., & Wang, G. G. (2019). Understanding histone H3 lysine 36 methylation and its deregulation in disease. *Cell Mol Life Sci*, 76(15), 2899-2916. <https://doi.org/10.1007/s00018-019-03144-y>
- Li, M., Phatnani, H. P., Guan, Z., Sage, H., Greenleaf, A. L., & Zhou, P. (2005). Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1. *Proceedings of the National Academy of Sciences*, 102(49), 17636-17641. <https://doi.org/10.1073/pnas.0506350102>
- Li, W., Tian, W., Yuan, G., Deng, P., Sengupta, D., Cheng, Z., Cao, Y., Ren, J., Qin, Y., Zhou, Y., Jia, Y., Gozani, O., Patel, D. J., & Wang, Z. (2021). Molecular basis of nucleosomal H3K36 methylation by NSD methyltransferases. *Nature*, 590(7846), 498-503. <https://doi.org/10.1038/s41586-020-03069-8>
- Li, Y., Trojer, P., Xu, C. F., Cheung, P., Kuo, A., Drury, W. J., 3rd, Qiao, Q., Neubert, T. A., Xu, R. M., Gozani, O., & Reinberg, D. (2009). The target of the NSD family of histone lysine

- methyltransferases depends on the nature of the substrate. *J Biol Chem*, 284(49), 34283-34295. <https://doi.org/10.1074/jbc.M109.034462>
- Lin, C. C., Chen, Y. P., Yang, W. Z., Shen, J. C. K., & Yuan, H. S. (2020). Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. *Nucleic Acids Res*, 48(7), 3949-3961. <https://doi.org/10.1093/nar/gkaa111>
- Lindorff-Larsen, K., Maragakis, P., Piana, S., Eastwood, M. P., Dror, R. O., & Shaw, D. E. (2012). Systematic validation of protein force fields against experimental data. *PLoS One*, 7(2), e32131. <https://doi.org/10.1371/journal.pone.0032131>
- Linscott, J. A., Kapilashrami, K., Wang, Z., Senevirathne, C., Bothwell, I. R., Blum, G., & Luo, M. (2016). Kinetic isotope effects reveal early transition state of protein lysine methyltransferase SET8. *Proc Natl Acad Sci U S A*, 113(52), E8369-E8378. <https://doi.org/10.1073/pnas.1609032114>
- Liu, L., Shi, T., Houk, K. N., & Zhao, Y. L. (2019). Understanding the R882H mutation effects of DNA methyltransferase DNMT3A: a combination of molecular dynamics simulations and QM/MM calculations. *RSC Adv*, 9(54), 31425-31434. <https://doi.org/10.1039/c9ra06791d>
- Liu, Y., Zhang, Y., Xue, H., Cao, M., Bai, G., Mu, Z., Yao, Y., Sun, S., Fang, D., & Huang, J. (2021). Cryo-EM structure of SETD2/Set2 methyltransferase bound to a nucleosome containing oncohistone mutations. *Cell Discov*, 7(1), 32. <https://doi.org/10.1038/s41421-021-00261-6>
- Lopes, P. E. M., Guvench, O., & MacKerell, A. D. (2015). Current Status of Protein Force Fields for Molecular Dynamics Simulations. In A. Kukol (Ed.), *Molecular Modeling of Proteins* (pp. 47-71). Springer New York. [https://doi.org/10.1007/978-1-4939-1465-4\\_3](https://doi.org/10.1007/978-1-4939-1465-4_3)
- Luo, M. (2012). Current chemical biology approaches to interrogate protein methyltransferases. *ACS Chem Biol*, 7(3), 443-463. <https://doi.org/10.1021/cb200519y>
- Luo, M. (2015). Inhibitors of protein methyltransferases as chemical tools. *Epigenomics*, 7(8), 1327-1338. <https://doi.org/10.2217/epi.15.87>
- Luo, M. (2018). Chemical and Biochemical Perspectives of Protein Lysine Methylation. *Chem Rev*, 118(14), 6656-6705. <https://doi.org/10.1021/acs.chemrev.8b00008>
- Mack, A., Emperle, M., Schnee, P., Adam, S., Pleiss, J., Bashtrykov, P., & Jeltsch, A. (2022). Preferential Self-interaction of DNA Methyltransferase DNMT3A Subunits Containing the R882H Cancer Mutation Leads to Dominant Changes of Flanking Sequence Preferences. *J Mol Biol*, 434(7), 167482. <https://doi.org/10.1016/j.jmb.2022.167482>
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8), 3696-3713. <https://doi.org/10.1021/acs.jctc.5b00255>
- Mallona, I., Ilie, I. M., Karemaker, I. D., Butz, S., Manzo, M., Caflich, A., & Baubec, T. (2021). Flanking sequence preference modulates de novo DNA methylation in the mouse genome. *Nucleic Acids Res*, 49(1), 145-157. <https://doi.org/10.1093/nar/gkaa1168>
- Marango, J., Shimoyama, M., Nishio, H., Meyer, J. A., Min, D. J., Sirulnik, A., Martinez-Martinez, Y., Chesi, M., Bergsagel, P. L., Zhou, M. M., Waxman, S., Leibovitch, B. A., Walsh, M. J., & Licht, J. D. (2008). The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor. *Blood*, 111(6), 3145-3154. <https://doi.org/10.1182/blood-2007-06-092122>

- Marmorstein, R. (2003). Structure of SET domain proteins: a new twist on histone methylation. *Trends Biochem Sci*, 28(2), 59-62. [https://doi.org/10.1016/S0968-0004\(03\)00007-0](https://doi.org/10.1016/S0968-0004(03)00007-0)
- Marmorstein, R. (2003). Structure of SET domain proteins: a new twist on histone methylation. *Trends in Biochemical Sciences*, 28(2), 59-62. [https://doi.org/https://doi.org/10.1016/S0968-0004\(03\)00007-0](https://doi.org/https://doi.org/10.1016/S0968-0004(03)00007-0)
- Marrink, S. J., & Tieleman, D. P. (2013). Perspective on the Martini model. *Chem Soc Rev*, 42(16), 6801-6822. <https://doi.org/10.1039/c3cs60093a>
- Mazur, P. K., Reynoird, N., Khatri, P., Jansen, P. W., Wilkinson, A. W., Liu, S., Barbash, O., Van Aller, G. S., Huddleston, M., Dhanak, D., Tummino, P. J., Kruger, R. G., Garcia, B. A., Butte, A. J., Vermeulen, M., Sage, J., & Gozani, O. (2014). SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. *Nature*, 510(7504), 283-287. <https://doi.org/10.1038/nature13320>
- McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernandez, C. X., Schwantes, C. R., Wang, L. P., Lane, T. J., & Pande, V. S. (2015). MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J*, 109(8), 1528-1532. <https://doi.org/10.1016/j.bpj.2015.08.015>
- McLean, C. M., Karemaker, I. D., & van Leeuwen, F. (2014). The emerging roles of DOT1L in leukemia and normal development. *Leukemia*, 28(11), 2131-2138. <https://doi.org/10.1038/leu.2014.169>
- Meller, J. a. (2001). Molecular Dynamics. In *Encyclopedia of Life Sciences*. Nature Publishing Group. <https://doi.org/10.1038/npg.els.0003048>
- Min, J., Feng, Q., Li, Z., Zhang, Y., & Xu, R. M. (2003). Structure of the Catalytic Domain of Human DOT1L, a Non-SET Domain Nucleosomal Histone Methyltransferase. *Cell*, 112(5), 711-723. [https://doi.org/10.1016/S0092-8674\(03\)00114-4](https://doi.org/10.1016/S0092-8674(03)00114-4)
- Mitchell, B., Thor, S., & Piper, M. (2023). Cellular and molecular functions of SETD2 in the central nervous system. *J Cell Sci*, 136(21). <https://doi.org/10.1242/jcs.261406>
- Molenaar, T. M., Malik, M., Silva, J., Liu, N. Q., Haarhuis, J. H. I., Ambrosi, C., Kwesi-Maliepaard, E. M., van Welsem, T., Baubec, T., Faller, W. J., & van Leeuwen, F. (2022). The histone methyltransferase SETD2 negatively regulates cell size. *J Cell Sci*, 135(19). <https://doi.org/10.1242/jcs.259856>
- Molenaar, T. M., & van Leeuwen, F. (2022). SETD2: from chromatin modifier to multipronged regulator of the genome and beyond. *Cell Mol Life Sci*, 79(6), 346. <https://doi.org/10.1007/s00018-022-04352-9>
- Monticelli, L., & Tieleman, D. P. (2013). Force Fields for Classical Molecular Dynamics. In L. Monticelli & E. Salonen (Eds.), *Biomolecular Simulations: Methods and Protocols* (pp. 197-213). Humana Press. [https://doi.org/10.1007/978-1-62703-017-5\\_8](https://doi.org/10.1007/978-1-62703-017-5_8)
- Morishita, M., & di Luccio, E. (2011). Structural insights into the regulation and the recognition of histone marks by the SET domain of NSD1. *Biochem Biophys Res Commun*, 412(2), 214-219. <https://doi.org/10.1016/j.bbrc.2011.07.061>
- Mzoughi, S., Tan, Y. X., Low, D., & Guccione, E. (2016). The role of PRDMs in cancer: one family, two sides. *Curr Opin Genet Dev*, 36, 83-91. <https://doi.org/10.1016/j.gde.2016.03.009>

- Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., & Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, 543(7643), 72-77. <https://doi.org/10.1038/nature21373>
- Nguyen, A. T., & Zhang, Y. (2011). The diverse functions of Dot1 and H3K79 methylation. *Genes Dev*, 25(13), 1345-1358. <https://doi.org/10.1101/gad.2057811>
- Nguyen, T. V., Yao, S., Wang, Y., Rolfe, A., Selvaraj, A., Darman, R., Ke, J., Warmuth, M., Smith, P. G., Larsen, N. A., Yu, L., Zhu, P., Fekkes, P., Vaillancourt, F. H., & Bolduc, D. M. (2019). The R882H DNMT3A hot spot mutation stabilizes the formation of large DNMT3A oligomers with low DNA methyltransferase activity. *J Biol Chem*, 294(45), 16966-16977. <https://doi.org/10.1074/jbc.RA119.010126>
- Nicholson, T. B., & Chen, T. (2009). LSD1 demethylates histone and non-histone proteins. *Epigenetics*, 4(3), 129-132. <https://doi.org/10.4161/epi.4.3.8443>
- Nimura, K., Ura, K., Shiratori, H., Ikawa, M., Okabe, M., Schwartz, R. J., & Kaneda, Y. (2009). A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome. *Nature*, 460(7252), 287-291. <https://doi.org/10.1038/nature08086>
- Nowak, R. P., Tumber, A., Johansson, C., Che, K. H., Brennan, P., Owen, D., & Oppermann, U. (2016). Advances and challenges in understanding histone demethylase biology. *Current Opinion in Chemical Biology*, 33, 151-159. <https://doi.org/https://doi.org/10.1016/j.cbpa.2016.06.021>
- Nvidia. (2010). Cuda Toolkit Documentation. *Comp. A J. Comp. Education*. <https://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html>
- O'Brien, E. C., Brewin, J., & Chevassut, T. (2014). DNMT3A: the DioNysian MonsTer of acute myeloid leukaemia. *Ther Adv Hematol*, 5(6), 187-196. <https://doi.org/10.1177/2040620714554538>
- Obenauer, J. C., Cantley, L. C., & Yaffe, M. B. (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res*, 31(13), 3635-3641. <https://doi.org/10.1093/nar/gkg584>
- Oelschlaeger, P. S., Rolf D.; Pleiss, Juergen. (2003). Modeling domino effects in enzymes: molecular basis of the substrate specificity of the bacterial metallo-beta-lactamases IMP-1 and IMP-6. *Biochemistry*, 30, 12. <https://doi.org/10.1021/bi0300332>
- Okano, M., Xie, S., & Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genetics*, 19(3), 219-220. <https://doi.org/10.1038/890>
- Oyer, J. A., Huang, X., Zheng, Y., Shim, J., Ezponda, T., Carpenter, Z., Allegretta, M., Okot-Kotber, C. I., Patel, J. P., Melnick, A., Levine, R. L., Ferrando, A., Mackerell, A. D., Jr., Kelleher, N. L., Licht, J. D., & Popovic, R. (2014). Point mutation E1099K in MMSET/NSD2 enhances its methyltransferase activity and leads to altered global chromatin methylation in lymphoid malignancies. *Leukemia*, 28(1), 198-201. <https://doi.org/10.1038/leu.2013.204>
- P. Eastman, M. S. F., J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M. R. Shirts, and V. S. Pande. (2013). OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *Journal of Chemical Theory and Computation*, 9, 9. <https://doi.org/10.1021/ct300857j>

- Pang, Y.-P. (1999). Novel Zinc Protein Molecular Dynamics Simulations: Steps Toward Antiangiogenesis for Cancer Treatment. *J. Mol. Model*, 5, 7. <https://doi.org/10.1007/s008940050119>
- Pang, Y.-P. (2001). Successful Molecular Dynamics Simulation of Two Zinc Complexes Bridged by a Hydroxide in Phosphotriesterase Using the Cationic Dummy Atom Method. *PROTEINS: Structure, Function, and Genetics*, 45, 7. <https://doi.org/10.1002/prot.1138>
- Park, I. Y., Powell, R. T., Tripathi, D. N., Dere, R., Ho, T. H., Blasius, T. L., Chiang, Y. C., Davis, I. J., Fahey, C. C., Hacker, K. E., Verhey, K. J., Bedford, M. T., Jonasch, E., Rathmell, W. K., & Walker, C. L. (2016). Dual Chromatin and Cytoskeletal Remodeling by SETD2. *Cell*, 166(4), 950-962. <https://doi.org/10.1016/j.cell.2016.07.005>
- Park, J. W., Chae, Y. C., Kim, J. Y., Oh, H., & Seo, S. B. (2018). Methylation of Aurora kinase A by MMSET reduces p53 stability and regulates cell proliferation and apoptosis. *Oncogene*, 37(48), 6212-6224. <https://doi.org/10.1038/s41388-018-0393-y>
- Patel, D. J., & Wang, Z. (2013). Readout of epigenetic modifications. *Annu Rev Biochem*, 82, 81-118. <https://doi.org/10.1146/annurev-biochem-072711-165700>
- Patnaik, D., Chin, H. G., Esteve, P. O., Benner, J., Jacobsen, S. E., & Pradhan, S. (2004). Substrate specificity and kinetic mechanism of mammalian G9a histone H3 methyltransferase. *J Biol Chem*, 279(51), 53248-53258. <https://doi.org/10.1074/jbc.M409604200>
- Pei, H., Zhang, L., Luo, K., Qin, Y., Chesi, M., Fei, F., Bergsagel, P. L., Wang, L., You, Z., & Lou, Z. (2011). MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites. *Nature*, 470(7332), 124-128. <https://doi.org/10.1038/nature09658>
- Piao, L., Nakakido, M., Suzuki, T., Dohmae, N., Nakamura, Y., & Hamamoto, R. (2016). Automethylation of SUV39H2, an oncogenic histone lysine methyltransferase, regulates its binding affinity to substrate proteins. *Oncotarget*, 7(16), 22846-22856. <https://doi.org/10.18632/oncotarget.8072>
- Pierro, J., Saliba, J., Narang, S., Sethia, G., Saint Fleur-Lominy, S., Chowdhury, A., Qualls, A., Fay, H., Kilberg, H. L., Moriyama, T., Fuller, T. J., Teachey, D. T., Schmiegelow, K., Yang, J. J., Loh, M. L., Brown, P. A., Zhang, J., Ma, X., Tsirigos, A., . . . Carroll, W. L. (2020). The NSD2 p.E1099K Mutation Is Enriched at Relapse and Confers Drug Resistance in a Cell Context-Dependent Manner in Pediatric Acute Lymphoblastic Leukemia. *Mol Cancer Res*, 18(8), 1153-1165. <https://doi.org/10.1158/1541-7786.MCR-20-0092>
- Porter, J. R., Zimmerman, M. I., & Bowman, G. R. (2019). Enspara: Modeling molecular ensembles with scalable data structures and parallel computing. *J Chem Phys*, 150(4), 044108. <https://doi.org/10.1063/1.5063794>
- Poulin, M. B., Schneck, J. L., Matico, R. E., McDevitt, P. J., Huddleston, M. J., Hou, W., Johnson, N. W., Thrall, S. H., Meek, T. D., & Schramm, V. L. (2016). Transition state for the NSD2-catalyzed methylation of histone H3 lysine 36. *Proc Natl Acad Sci U S A*, 113(5), 1197-1201. <https://doi.org/10.1073/pnas.1521036113>
- Powers, N. R., Parvanov, E. D., Baker, C. L., Walker, M., Petkov, P. M., & Paigen, K. (2016). The Meiotic Recombination Activator PRDM9 Trimethylates Both H3K36 and H3K4 at Recombination Hotspots In Vivo. *PLoS Genet*, 12(6), e1006146. <https://doi.org/10.1371/journal.pgen.1006146>
- Qian, C., & Zhou, M. M. (2006). SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell Mol Life Sci*, 63(23), 2755-2763. <https://doi.org/10.1007/s00018-006-6274-5>

- Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D., & Xu, R. M. (2011). The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. *J Biol Chem*, *286*(10), 8361-8368. <https://doi.org/10.1074/jbc.M110.204115>
- Rathert, P., Dhayalan, A., Ma, H., & Jeltsch, A. (2008). Specificity of protein lysine methyltransferases and methods for detection of lysine methylation of non-histone proteins. *Mol Biosyst*, *4*(12), 1186-1190. <https://doi.org/10.1039/b811673c>
- Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai, Y., Cheng, X., & Jeltsch, A. (2008). Protein lysine methyltransferase G9a acts on non-histone targets. *Nat Chem Biol*, *4*(6), 344-346. <https://doi.org/10.1038/nchembio.88>
- Razin, A., & Riggs, A. D. (1980). DNA Methylation and Gene Function. *Science*, *210*(4470), 604-610. <https://doi.org/10.1126/science.6254144>
- Richon, V. M., Johnston, D., Sneeringer, C. J., Jin, L., Majer, C. R., Elliston, K., Jerva, L. F., Scott, M. P., & Copeland, R. A. (2011). Chemogenetic analysis of human protein methyltransferases. *Chem Biol Drug Des*, *78*(2), 199-210. <https://doi.org/10.1111/j.1747-0285.2011.01135.x>
- Rogawski, D. S., Ndoj, J., Cho, H. J., Maillard, I., Grembecka, J., & Cierpicki, T. (2015). Two Loops Undergoing Concerted Dynamics Regulate the Activity of the ASH1L Histone Methyltransferase. *Biochemistry*, *54*(35), 5401-5413. <https://doi.org/10.1021/acs.biochem.5b00697>
- Saddic, L. A., West, L. E., Aslanian, A., Yates, J. R., 3rd, Rubin, S. M., Gozani, O., & Sage, J. (2010). Methylation of the retinoblastoma tumor suppressor by SMYD2. *J Biol Chem*, *285*(48), 37733-37740. <https://doi.org/10.1074/jbc.M110.137612>
- Sankaran, S. M., Wilkinson, A. W., Elias, J. E., & Gozani, O. (2016). A PWWP Domain of Histone-Lysine N-Methyltransferase NSD2 Binds to Dimethylated Lys-36 of Histone H3 and Regulates NSD2 Function at Chromatin. *J Biol Chem*, *291*(16), 8465-8474. <https://doi.org/10.1074/jbc.M116.720748>
- Sarno, F., Nebbioso, A., & Altucci, L. (2020). DOT1L: a key target in normal chromatin remodelling and in mixed-lineage leukaemia treatment. *Epigenetics*, *15*(5), 439-453. <https://doi.org/10.1080/15592294.2019.1699991>
- Sato, K., Kumar, A., Hamada, K., Okada, C., Oguni, A., Machiyama, A., Sakuraba, S., Nishizawa, T., Nureki, O., Kono, H., Ogata, K., & Sengoku, T. (2021). Structural basis of the regulation of the normal and oncogenic methylation of nucleosomal histone H3 Lys36 by NSD2. *Nat Commun*, *12*(1), 6605. <https://doi.org/10.1038/s41467-021-26913-5>
- Schnee, P., Choudalakis, M., Weirich, S., Khella, M. S., Carvalho, H., Pleiss, J., & Jeltsch, A. (2022). Mechanistic basis of the increased methylation activity of the SETD2 protein lysine methyltransferase towards a designed super-substrate peptide. *Communications Chemistry*, *5*(1), 139. <https://doi.org/10.1038/s42004-022-00753-w>
- Schnee, P., Pleiss, J., & Jeltsch, A. (2023). Approaching the catalytic mechanism of protein lysine methyltransferases by biochemical and simulations techniques. *Crit Rev Biochem Mol Biol.*, *7*, 1-49. <https://doi.org/10.1080/10409238.2024.2318547>
- Schoelz, J. M., & Riddle, N. C. (2022). Functions of HP1 proteins in transcriptional regulation. *Epigenetics Chromatin*, *15*(1), 14. <https://doi.org/10.1186/s13072-022-00453-8>
- Schrödinger, L. (2015). The PyMOL molecular graphics system, version 1.7. 6.6. *Schrödinger LLC*.



- Schrödinger, L. L. C. (2015). The {PyMOL} Molecular Graphics System. *Version: 2.4.1*.
- Schuhmacher, M. K., Beldar, S., Khella, M. S., Brohm, A., Ludwig, J., Tempel, W., Weirich, S., Min, J., & Jeltsch, A. (2020). Sequence specificity analysis of the SETD2 protein lysine methyltransferase and discovery of a SETD2 super-substrate. *Commun Biol*, 3(1), 511. <https://doi.org/10.1038/s42003-020-01223-6>
- Schuhmacher, M. K., Kudithipudi, S., Kusevic, D., Weirich, S., & Jeltsch, A. (2015). Activity and specificity of the human SUV39H2 protein lysine methyltransferase. *Biochim Biophys Acta*, 1849(1), 55-63. <https://doi.org/10.1016/j.bbagr.2014.11.005>
- Seervai, R. N. H., Jangid, R. K., Karki, M., Tripathi, D. N., Jung, S. Y., Kearns, S. E., Verhey, K. J., Cianfrocco, M. A., Millis, B. A., Tyska, M. J., Mason, F. M., Rathmell, W. K., Park, I. Y., Dere, R., & Walker, C. L. (2020). The Huntingtin-interacting protein SETD2/HYPB is an actin lysine methyltransferase. *Science Advances*, 6(40), eabb7854. <https://doi.org/10.1126/sciadv.abb7854>
- Sengupta, D., Zeng, L., Li, Y., Hausmann, S., Ghosh, D., Yuan, G., Nguyen, T. N., Lyu, R., Caporicci, M., Morales Benitez, A., Coles, G. L., Kharchenko, V., Czaban, I., Azhibek, D., Fischle, W., Jaremko, M., Wistuba, II, Sage, J., Jaremko, L., . . . Gozani, O. (2021a). NSD2 dimethylation at H3K36 promotes lung adenocarcinoma pathogenesis. *Mol Cell*, 81(21), 4481-4492 e4489. <https://doi.org/10.1016/j.molcel.2021.08.034>
- Sengupta, D., Zeng, L., Li, Y., Hausmann, S., Ghosh, D., Yuan, G., Nguyen, T. N., Lyu, R., Caporicci, M., Morales Benitez, A., Coles, G. L., Kharchenko, V., Czaban, I., Azhibek, D., Fischle, W., Jaremko, M., Wistuba, II, Sage, J., Jaremko, Ł., . . . Gozani, O. (2021b). NSD2 dimethylation at H3K36 promotes lung adenocarcinoma pathogenesis. *Mol Cell*, 81(21), 4481-4492.e4489. <https://doi.org/10.1016/j.molcel.2021.08.034>
- Sessa, A., Fagnocchi, L., Mastrototaro, G., Massimino, L., Zaghi, M., Indrigo, M., Cattaneo, S., Martini, D., Gabellini, C., Pucci, C., Fasciani, A., Belli, R., Taverna, S., Andreazzoli, M., Zippo, A., & Broccoli, V. (2019). SETD5 Regulates Chromatin Methylation State and Preserves Global Transcriptional Fidelity during Brain Development and Neuronal Wiring. *Neuron*, 104(2), 271-289 e213. <https://doi.org/10.1016/j.neuron.2019.07.013>
- Shah, M. Y., Martinez-Garcia, E., Phillip, J. M., Chambliss, A. B., Popovic, R., Ezponda, T., Small, E. C., Will, C., Phillip, M. P., Neri, P., Bahlis, N. J., Wirtz, D., & Licht, J. D. (2016). MMSET/WHSC1 enhances DNA damage repair leading to an increase in resistance to chemotherapeutic agents. *Oncogene*, 35(45), 5905-5915. <https://doi.org/10.1038/onc.2016.116>
- Sirinupong, N., Brunzelle, J., Doko, E., & Yang, Z. (2011). Structural insights into the autoinhibition and posttranslational activation of histone methyltransferase SmyD3. *J Mol Biol*, 406(1), 149-159. <https://doi.org/10.1016/j.jmb.2010.12.014>
- Sirinupong, N., Brunzelle, J., Ye, J., Pirzada, A., Nico, L., & Yang, Z. (2010). Crystal structure of cardiac-specific histone methyltransferase SmyD1 reveals unusual active site architecture. *J Biol Chem*, 285(52), 40635-40644. <https://doi.org/10.1074/jbc.M110.168187>
- Song, D., Lan, J., Chen, Y., Liu, A., Wu, Q., Zhao, C., Feng, Y., Wang, J., Luo, X., Cao, Z., Cao, X., Hu, J., & Wang, G. (2021). NSD2 promotes tumor angiogenesis through methylating and activating STAT3 protein. *Oncogene*, 40(16), 2952-2967. <https://doi.org/10.1038/s41388-021-01747-z>
- Stec, I., Wright, T. J., van Ommen, G. J., de Boer, P. A., van Haeringen, A., Moorman, A. F., Altherr, M. R., & den Dunnen, J. T. (1998). WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphia gene maps in the Wolf-Hirschhorn

- syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum Mol Genet*, 7(7), 1071-1082. <https://doi.org/10.1093/hmg/7.7.1071>
- Sun, X. J., Wei, J., Wu, X. Y., Hu, M., Wang, L., Wang, H. H., Zhang, Q. H., Chen, S. J., Huang, Q. H., & Chen, Z. (2005). Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *J Biol Chem*, 280(42), 35261-35271. <https://doi.org/10.1074/jbc.M504012200>
- Swaroop, A., Oyer, J. A., Will, C. M., Huang, X., Yu, W., Troche, C., Bulic, M., Durham, B. H., Wen, Q. J., Crispino, J. D., MacKerell, A. D., Jr., Bennett, R. L., Kelleher, N. L., & Licht, J. D. (2019). An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. *Oncogene*, 38(5), 671-686. <https://doi.org/10.1038/s41388-018-0474-y>
- Swenson, D. E. H., & Roet, S. (2017). Contact Map Explorer. [https://github.com/dwhswenson/contact\\_map](https://github.com/dwhswenson/contact_map). <https://contact-map.readthedocs.io/en/latest/index.html>.
- Tachibana, M., Sugimoto, K., Fukushima, T., & Shinkai, Y. (2001). Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem*, 276(27), 25309-25317. <https://doi.org/10.1074/jbc.M101914200>
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, 324(5929), 930-935. <https://doi.org/10.1126/science.1170116>
- Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., . . . Forbes, S. A. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, 47(D1), D941-D947. <https://doi.org/10.1093/nar/gky1015>
- Tate, P. H., & Bird, A. P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. *Current Opinion in Genetics & Development*, 3(2), 226-231. [https://doi.org/https://doi.org/10.1016/0959-437X\(93\)90027-M](https://doi.org/https://doi.org/10.1016/0959-437X(93)90027-M)
- Tom Darden, D. Y., and Lee Pedersen. (1993). An N ·log( N ) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98, 4. <https://doi.org/10.1063/1.464397>
- Topchu, I., Pangen, R. P., Bychkov, I., Miller, S. A., Izumchenko, E., Yu, J., Golemis, E., Karanicolas, J., & Bumber, Y. (2022). The role of NSD1, NSD2, and NSD3 histone methyltransferases in solid tumors. *Cell Mol Life Sci*, 79(6), 285. <https://doi.org/10.1007/s00018-022-04321-2>
- Triebel, R. C., Beach, B. M., Dirk, L. M. A., Houtz, R. L., & Hurley, J. H. (2002). Structure and Catalytic Mechanism of a SET Domain Protein Methyltransferase. *Cell*, 111, 91-103. [https://doi.org/10.1016/s0092-8674\(02\)01000-0](https://doi.org/10.1016/s0092-8674(02)01000-0)
- Triebel, R. C., Flynn, E. M., Houtz, R. L., & Hurley, J. H. (2003). Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSM1. *Nat Struct Mol Biol*, 10, 545-552. <https://doi.org/10.1038/nsb946>
- van Dongen, S. F., Elemans, J. A., Rowan, A. E., & Nolte, R. J. (2014). Processive catalysis. *Angew Chem Int Ed Engl*, 53(43), 11420-11428. <https://doi.org/10.1002/anie.201404848>

- Wagner, E. J., & Carpenter, P. B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nature Reviews Molecular Cell Biology*, 13(2), 115-126. <https://doi.org/10.1038/nrm3274>
- Wagner, J. R., Sørensen, J., Hensley, N., Wong, C., Zhu, C., Perison, T., & Amaro, R. E. (2017). POVME 3.0: Software for Mapping Binding Pocket Flexibility. *Journal of Chemical Theory and Computation*, 13(9), 4584-4592. <https://doi.org/10.1021/acs.jctc.7b00500>
- Wang, J., Wang, W., Kollman, P. A., & Case, D. A. (2001). Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc.*, 222(U403).
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*, 46(W1), W296-W303. <https://doi.org/10.1093/nar/gky427>
- Weil, L. E., Shmidov, Y., Kublanovsky, M., Morgenstern, D., Feldman, M., Bitton, R., & Levy, D. (2018). Oligomerization and Auto-methylation of the Human Lysine Methyltransferase SETD6. *J Mol Biol*, 430(21), 4359-4368. <https://doi.org/10.1016/j.jmb.2018.08.028>
- Weinberg, D. N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K. N., Horth, C., McGuire, J. T., Xu, X., Nikbakht, H., Lemiesz, A. E., Marchione, D. M., Marunde, M. R., Meiners, M. J., Cheek, M. A., Keogh, M. C., Bareke, E., Djedid, A., Harutyunyan, A. S., . . . Lu, C. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature*, 573(7773), 281-286. <https://doi.org/10.1038/s41586-019-1534-3>
- Weirich, S., & Jeltsch, A. (2022). Specificity Analysis of Protein Methyltransferases and Discovery of Novel Substrates Using SPOT Peptide Arrays. *Methods Mol Biol*, 2529, 313-325. [https://doi.org/10.1007/978-1-0716-2481-4\\_15](https://doi.org/10.1007/978-1-0716-2481-4_15)
- Weirich, S., & Jeltsch, A. (2022). Specificity Analysis of Protein Methyltransferases and Discovery of Novel Substrates Using SPOT Peptide Arrays. In R. Margueron & D. Holoch (Eds.), *Histone Methyltransferases: Methods and Protocols* (pp. 313-325). Springer US. [https://doi.org/10.1007/978-1-0716-2481-4\\_15](https://doi.org/10.1007/978-1-0716-2481-4_15)
- Weirich, S., Kudithipudi, S., & Jeltsch, A. (2016). Specificity of the SUV4-20H1 and SUV4-20H2 protein lysine methyltransferases and methylation of novel substrates. *J Mol Biol*, 428(11), 2344-2358. <https://doi.org/10.1016/j.jmb.2016.04.015>
- Weirich, S., Kudithipudi, S., & Jeltsch, A. (2017). Somatic cancer mutations in the MLL1 histone methyltransferase modulate its enzymatic activity and dependence on the WDR5/RBBP5/ASH2L complex. *Mol Oncol*, 11(4), 373-387. <https://doi.org/10.1002/1878-0261.12041>
- Weirich, S., Kudithipudi, S., Kycia, I., & Jeltsch, A. (2015). Somatic cancer mutations in the MLL3-SET domain alter the catalytic properties of the enzyme. *Clin Epigenetics*, 7(1), 36. <https://doi.org/10.1186/s13148-015-0075-3>
- Weirich, S., Kusevic, D., Schnee, P., Reiter, J., Pleiss, J., & Jeltsch, A. (2023). Discovery of new NSD2 non-histone substrates and design of a super-substrate. - *Publication submitted for review.*
- Weirich, S., Schuhmacher, M. K., Kudithipudi, S., Lungu, C., Ferguson, A. D., & Jeltsch, A. (2020). Analysis of the Substrate Specificity of the SMYD2 Protein Lysine Methyltransferase and Discovery of Novel Non-Histone Substrates. *ChemBiochem*, 21(1-2), 256-264. <https://doi.org/10.1002/cbic.201900582>

- West, L. E., & Gozani, O. (2011). Regulation of p53 function by lysine methylation. *Epigenomics*, 3(3), 361-369. <https://doi.org/10.2217/epi.11.21>
- Wilson, J. R., Jing, C., Walker, P. A., Martin, S. R., Howell, S. A., M., B. G., Gamblin, S. J., & Xiao, B. (2002). Crystal Structure and Functional Analysis of the Histone Methyltransferase SET7/9. *Cell*, 111(1), 105-115. [https://doi.org/10.1016/S0092-8674\(02\)00964-9](https://doi.org/10.1016/S0092-8674(02)00964-9)
- Wu, C.-Y., Hsieh, C.-Y., Huang, K.-E., Chang, C., & Kang, H.-Y. (2012). Cryptotanshinone down-regulates androgen receptor signaling by modulating lysine-specific demethylase 1 function. *International Journal of Cancer*, 131(6), 1423-1434. <https://doi.org/https://doi.org/10.1002/ijc.27343>
- Wu, H., Min, J., Lunin, V. V., Antoshenko, T., Dombrowski, L., Zeng, H., Allali-Hassani, A., Campagna-Slater, V., Vedadi, M., Arrowsmith, C. H., Plotnikov, A. N., & Schapira, M. (2010). Structural biology of human H3K9 methyltransferases. *PLoS One*, 5(1), e8570. <https://doi.org/10.1371/journal.pone.0008570>
- Wu, J., Cheung, T., Grande, C., Ferguson, A. D., Zhu, X., Theriault, K., Code, E., Birr, C., Keen, N., & Chen, H. (2011). Biochemical characterization of human SET and MYND domain-containing protein 2 methyltransferase. *Biochemistry*, 50(29), 6488-6497. <https://doi.org/10.1021/bi200725p>
- Xiang, W., He, J., Huang, C., Chen, L., Tao, D., Wu, X., Wang, M., Luo, G., Xiao, X., Zeng, F., & Jiang, G. (2015). miR-106b-5p targets tumor suppressor gene SETD2 to inactive its function in clear cell renal cell carcinoma. *Oncotarget; Vol 6, No 6*. <https://www.oncotarget.com/article/2926/text/>
- Xiao, B., Jing, C., Kelly, G., Walker, P. A., Muskett, F. W., Frenkiel, T. A., Martin, S. R., Sarma, K., Reinberg, D., Gamblin, S. J., & Wilson, J. R. (2005). Specificity and mechanism of the histone methyltransferase Pr-Set7. *Genes Dev*, 19(12), 1444-1454. <https://doi.org/10.1101/gad.1315905>
- Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, I. A., Blackburn, G. M., & Gamblin, S. J. (2003). Structure and catalytic mechanism of the human histone methyltransferase SET7/9. *Nature*, 421, 652-656. <https://doi.org/10.1038/nature01378>
- Xu, H., Liao, C., Liang, S., & Ye, B. C. (2021). A Novel Peptide-Equipped Exosomes Platform for Delivery of Antisense Oligonucleotides. *ACS Appl Mater Interfaces*, 13(9), 10760-10767. <https://doi.org/10.1021/acsami.1c00016>
- Xue, H., Yao, T., Cao, M., Zhu, G., Li, Y., Yuan, G., Chen, Y., Lei, M., & Huang, J. (2019). Structural basis of nucleosome recognition and modification by MLL methyltransferases. *Nature*, 573(7774), 445-449. <https://doi.org/10.1038/s41586-019-1528-1>
- Yan, X.-J., Xu, J., Gu, Z.-H., Pan, C.-M., Lu, G., Shen, Y., Shi, J.-Y., Zhu, Y.-M., Tang, L., Zhang, X.-W., Liang, W.-X., Mi, J.-Q., Song, H.-D., Li, K.-Q., Chen, Z., & Chen, S.-J. (2011). Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nature Genetics*, 43(4), 309-315. <https://doi.org/10.1038/ng.788>
- Yang, J., Huang, J., Dasgupta, M., Sears, N., Miyagi, M., Wang, B., Chance, M. R., Chen, X., Du, Y., Wang, Y., An, L., Wang, Q., Lu, T., Zhang, X., Wang, Z., & Stark, G. R. (2010). Reversible methylation of promoter-bound STAT3 by histone-modifying enzymes. *Proceedings of the National Academy of Sciences*, 107(50), 21499-21504. <https://doi.org/10.1073/pnas.1016147107>
- Yang, J., Huang, J., Dasgupta, M., Sears, N., Miyagi, M., Wang, B., Chance, M. R., Chen, X., Du, Y., Wang, Y., An, L., Wang, Q., Lu, T., Zhang, X., Wang, Z., & Stark, G. R. (2010). Reversible methylation of

- promoter-bound STAT3 by histone-modifying enzymes. *Proc Natl Acad Sci U S A*, 107(50), 21499-21504. <https://doi.org/10.1073/pnas.1016147107>
- Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A., & Weinberg, R. A. (2004). Twist, a Master Regulator of Morphogenesis, Plays an Essential Role in Tumor Metastasis. *Cell*, 117(7), 927-939. <https://doi.org/https://doi.org/10.1016/j.cell.2004.06.006>
- Yang, L., Rau, R., & Goodell, M. A. (2015). DNMT3A in haematological malignancies. *Nature Reviews Cancer*, 15(3), 152-165. <https://doi.org/10.1038/nrc3895>
- Yang, S., Winstone, L., Mondal, S., & Wu, Y. (2023). Helicases in R-loop Formation and Resolution. *J Biol Chem*, 299(11), 105307. <https://doi.org/10.1016/j.jbc.2023.105307>
- Yang, S., Zheng, X., Lu, C., Li, G. M., Allis, C. D., & Li, H. (2016). Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase. *Genes Dev*, 30(14), 1611-1616. <https://doi.org/10.1101/gad.284323.116>
- Yang, T., Zhang, W., Cheng, J., Nie, Y., Xin, Q., Yuan, S., & Dou, Y. (2019). Formation Mechanism of Ion Channel in Channelrhodopsin-2: Molecular Dynamics Simulation and Steering Molecular Dynamics Simulations. *Int J Mol Sci*, 20(15). <https://doi.org/10.3390/ijms20153780>
- Yang, W., Zhuang, J., Li, C., & Cheng, G. J. (2023). Unveiling the methyl transfer mechanisms in the epigenetic machinery DNMT3A-3 L: A comprehensive study integrating assembly dynamics with catalytic reactions. *Comput Struct Biotechnol J*, 21, 2086-2099. <https://doi.org/10.1016/j.csbj.2023.03.002>
- Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., & Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, 356(6337), eaaj2239. <https://doi.org/10.1126/science.aaj2239>
- Yu, H., & Dalby, P. A. (2020). A beginner's guide to molecular dynamics simulations and the identification of cross-correlation networks for enzyme engineering. *Methods Enzymol*, 643, 15-49. <https://doi.org/10.1016/bs.mie.2020.04.020>
- Yuan, G., Flores, N. M., Hausmann, S., Lofgren, S. M., Kharchenko, V., Angulo-Ibanez, M., Sengupta, D., Lu, X., Czaban, I., Azhibek, D., Vicent, S., Fischle, W., Jaremko, M., Fang, B., Wistuba, II, Chua, K. F., Roth, J. A., Minna, J. D., Shao, N. Y., . . . Gozani, O. (2021). Elevated NSD3 histone methylation activity drives squamous cell lung cancer. *Nature*, 590(7846), 504-508. <https://doi.org/10.1038/s41586-020-03170-y>
- Yun, M., Wu, J., Workman, J. L., & Li, B. (2011). Readers of histone modifications. *Cell Research*, 21(4), 564-578. <https://doi.org/10.1038/cr.2011.42>
- Zhang, J., Lee, Y. R., Dang, F., Gan, W., Menon, A. V., Katon, J. M., Hsu, C. H., Asara, J. M., Tibarewal, P., Leslie, N. R., Shi, Y., Pandolfi, P. P., & Wei, W. (2019). PTEN Methylation by NSD2 Controls Cellular Sensitivity to DNA Damage. *Cancer Discov*, 9(9), 1306-1323. <https://doi.org/10.1158/2159-8290.CD-18-0083>
- Zhang, X., & Bruce, T. (2007a). Catalytic Mechanism and Product Specificity of Rubisco Large Subunit Methyltransferase: QM/MM and MD Investigations. *Biochemistry*, 2007, 5505-5514. <https://doi.org/10.1021/bi700119p>

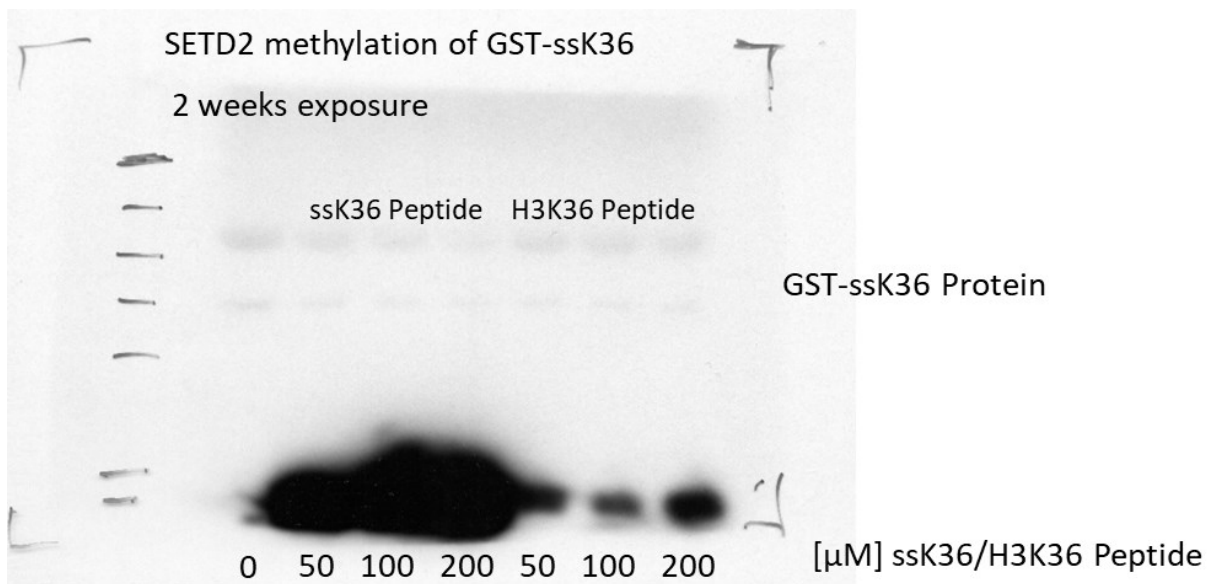
- Zhang, X., & Bruice, T. (2007b). Histone Lysine Methyltransferase SET7/9: Formation of a Water Channel Precedes Each Methyl Transfer. *Biochemistry*, *46*, 14838-14844. <https://doi.org/10.1021/bi7014579>
- Zhang, X., & Bruice, T. (2007c). A Quantum Mechanics/Molecular Mechanics Study of the Catalytic Mechanism and Product Specificity of Viral Histone Lysine Methyltransferase. *Biochemistry*, *46*, 9743-9751. <https://doi.org/10.1021/bi700515q>
- Zhang, X., & Bruice, T. (2008a). Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases. *PNAS*, *105*. <https://doi.org/10.1073/pnas.0801788105>
- Zhang, X., & Bruice, T. (2008b). Mechanism of Product Specificity of AdoMet Methylation Catalyzed by Lysine Methyltransferases: Transcriptional Factor p53 Methylation by Histone Lysine Methyltransferase SET7/9. *Biochemistry*, *47*, 2743-2748. <https://doi.org/10.1021/bi702370p>
- Zhang, X., & Bruice, T. (2008c). Product Specificity and Mechanism of Protein Lysine Methyltransferases: Insights from the Histone Lysine Methyltransferase SET8. *Biochemistry*, *47*, 6671-6677. <https://doi.org/10.1021/bi800244s>
- Zhang, X., & Bruice, T. C. (2008). Enzymatic mechanism and product specificity of SET-domain protein lysine methyltransferases. *Proc Natl Acad Sci U S A*, *105*(15), 5728-5732. <https://doi.org/10.1073/pnas.0801788105>
- Zhang, X., Huang, Y., & Shi, X. (2015). Emerging roles of lysine methylation on non-histone proteins. *Cell Mol Life Sci*, *72*(22), 4257-4272. <https://doi.org/10.1007/s00018-015-2001-4>
- Zhang, X., Tamaru, H., Khan, S. I., Horton, J. R., Keefe, L. J., Selker, E. U., & Chen, X. (2002). Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell*, *111*(1), 117-127. [https://doi.org/10.1016/s0092-8674\(02\)00999-6](https://doi.org/10.1016/s0092-8674(02)00999-6).
- Zhang, X., Tanaka, K., Yan, J., Li, J., Peng, D., Jiang, Y., Yang, Z., Barton, M. C., Wen, H., & Shi, X. (2013). Regulation of estrogen receptor alpha by histone methyltransferase SMYD2-mediated protein methylation. *Proc Natl Acad Sci U S A*, *110*(43), 17284-17289. <https://doi.org/10.1073/pnas.1307959110>
- Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., & Cheng, X. (2003). Structural basis for the product specificity of histone lysine methyltransferases. *Mol Cell*, *12*(1), 177-185. [https://doi.org/10.1016/s1097-2765\(03\)00224-7](https://doi.org/10.1016/s1097-2765(03)00224-7)
- Zhang, Y., Fang, Y., Tang, Y., Han, S., Jia, J., Wan, X., Chen, J., Yuan, Y., Zhao, B., & Fang, D. (2022). SMYD5 catalyzes histone H3 lysine 36 trimethylation at promoters. *Nat Commun*, *13*(1), 3190. <https://doi.org/10.1038/s41467-022-30940-1>
- Zhang, Y., Shan, C. M., Wang, J., Bao, K., Tong, L., & Jia, S. (2017). Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. *Sci Rep*, *7*, 43906. <https://doi.org/10.1038/srep43906>
- Zhang, Z.-M., Lu, R., Wang, P., Yu, Y., Chen, D., Gao, L., Liu, S., Ji, D., Rothbart, S. B., Wang, Y., Wang, G. G., & Song, J. (2018). Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature*, *554*(7692), 387-391. <https://doi.org/10.1038/nature25477>
- Zhao, S., Allis, C. D., & Wang, G. G. (2021). The language of chromatin modification in human cancers. *Nat Rev Cancer*, *21*(7), 413-430. <https://doi.org/10.1038/s41568-021-00357-x>

- Zheng, W., Ibanez, G., Wu, H., Blum, G., Zeng, H., Dong, A., Li, F., Hajian, T., Allali-Hassani, A., Amaya, M. F., Siarheyeva, A., Yu, W., Brown, P. J., Schapira, M., Vedadi, M., Min, J., & Luo, M. (2012). Sinefungin derivatives as inhibitors and structure probes of protein lysine methyltransferase SETD2. *J Am Chem Soc*, *134*(43), 18004-18014. <https://doi.org/10.1021/ja307060p>
- Zhu, K., Lei, P. J., Ju, L. G., Wang, X., Huang, K., Yang, B., Shao, C., Zhu, Y., Wei, G., Fu, X. D., Li, L., & Wu, M. (2017). SPOP-containing complex regulates SETD2 stability and H3K36me3-coupled alternative splicing. *Nucleic Acids Res*, *45*(1), 92-105. <https://doi.org/10.1093/nar/gkw814>

### 7.3. Appendix III

Schnee P, Jeltsch A, Weirich S (2023) Artifizielles Peptid mit PKMT-inhibitorischer Wirkung - Universität Stuttgart PCT-Patentanmeldung

Substrate-competitive peptide inhibitors were tested to reduce the methylation activity of SETD2. In the presented set up, the peptide inhibitor was methylated by the investigated PKMT (**Supplementary Fig. 1**). As expected, the ssK36 peptide is being methylated much faster compared to the H3K36 peptide. Due to the methylation, the peptide inhibitor dissociates from the PKMT and is unable to bind again, leading to a constant decrease of effective inhibitor concentration over time.



**Supplementary Figure 1: The substrate-competitive peptide inhibitors are methylated by SETD2.** Autoradiography result of *in vitro* methylation of GST-ssK36 (0.28 μM) by His<sub>6</sub>-tagged SETD2 (12 μM) using radioactively labeled SAM (0.76 μM) and increasing concentration of peptide competitor ssK36 or H3K36.