

Protein Engineering Tools to Explore the Function of Protein Post-Translational Modifications for Chromatin and Microtubule Cytoskeletal Biology

Thèse N° 9354

Présentée le 10 avril 2019

à la Faculté des sciences de base

Laboratoire de chimie biophysique des macromolécules

Programme doctoral en chimie et génie chimique

pour l'obtention du grade de Docteur ès Sciences

par

Ninad Dilip AGASHE

Acceptée sur proposition du jury

Dr A.-S. Chauvin, présidente du jury

Prof. B. Fierz, directeur de thèse

Prof. C. Janke, rapporteur

Prof. M. Lochner, rapporteur

Prof. C. Heinis, rapporteur

2019



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Abstract

Protein post-translational modifications (PTMs) play a crucial role in expanding the protein diversity and are one of the major mechanisms through which cells respond to ever changing environmental cues. The function of two of the most important cellular complexes, chromatin and microtubules, is influenced and very tightly regulated by underlying protein post-translational modifications. The focus of my thesis is to develop tools to investigate the role of protein PTMs in the context of chromatin and microtubules.

Eukaryotic DNA is organized in the form of chromatin whose basic unit is the nucleosome. The nucleosome is composed of 147 base pair of DNA wrapped around four core histones forming H2A, H2B, H3 and H4 forming an octamer. Each histone can be highly post-translationally modified, especially on their N-terminal tails protruding from the nucleosome particle. Histone post-translational modifications (PTMs) work combinatorially to establish chromatin states defined by specific gene expression status known as the Histone Code. Although, each nucleosome carries two copies of each histone, each copy in a single nucleosome can be differently modified resulting in PTM based nucleosome asymmetry. In bivalent domains, a chromatin signature prevalent in embryonic stem cells (ESCs), histone H3 methylated at lysine 4 (H3K4me3)- an activating histone PTM mark, coexists with H3K27me3- a repressive histone PTM mark, in asymmetric nucleosomes. In the first project, a general, modular and a traceless synthetic strategy to produce asymmetrically modified nucleosomes is described. Using these asymmetric nucleosomes, I show that in bivalent nucleosomes, H3K4me3 inhibits the activity of the H3K27-specific lysine methyltransferase (KMT) Polycomb Repressive Complex 2 (PRC2) solely on the same histone tail. Whereas H3K27me3 stimulates PRC2 activity via a positive feedback mechanism across tails, thereby partially overriding the H3K4me3-mediated repressive effect.

Microtubules are the largest components of the eukaryotic cytoskeleton and play an important role in maintain cellular organization, intracellular and neuronal transport, cell motility and cell division. Microtubules are dynamically assembled from $\alpha\beta$ -tubulin heterodimer, which is the basic repeating unit. The primary sequence and structure of tubulin proteins and microtubules are highly conserved in eukaryotic evolution. Despite this conservation, tubulin is subjected to heterogeneity that arises from differential expression of multiple tubulin

isotypes and the vast repertoire of tubulin PTMs, predominantly on the unstructured C-terminal tubulin tails. Together, this gives rise to the Tubulin Code. Lack of access to uniformly modified tubulin presents a major difficulty in study of PTM function. Here, I have developed a method to link synthetic, modified tails to recombinant tubulin by use of a split intein based, protein trans-splicing (PTS) approach. I demonstrate this approach by linking a fluorescently modified tubulin tail to recombinant human tubulin, preparing semisynthetic tubulin dimers. Extending the method to native PTMs will open the door for an in-depth study of tubulin modification using a chemically defined system.

Keywords: *post-translational modifications (PTMs), histone code, bivalent chromatin, Polycomb Repressive Complex 2 (PRC2), $\alpha\beta$ -tubulin heterodimer, tubulin code, split inteins, protein trans-splicing (PTS)*

Résumé

Les modifications post-traductionnelles (PTM) jouent un rôle crucial dans l'extension de la diversité des protéines et sont un des mécanismes majeurs par lequel les cellules répondent aux changements incessants de signaux environnementaux. La fonction de deux des plus importants complexes cellulaires, la chromatine et les microtubules, est influencée et précisément régulée par les PTM. Le sujet de ma thèse est de développer des outils pour étudier le rôle des PTM dans le contexte de la chromatine et des microtubules.

Chez les eucaryotes, l'ADN est organisé en chromatine dont l'unité de base est le nucléosome. Le nucléosome est constitué de 147 paires de bases d'ADN enroulées autour de quatre dimères d'histones, H2A, H2B, H3 et H4 formant ainsi un octamère. Chaque histone peut contenir de nombreuses PTM, principalement dans leur région N-terminale qui sont protubérantes par rapport à l'histone. Les PTM des histones travaillent de manière combinatoire pour établir des états distincts dans la chromatine caractérisée par des profils distincts d'expression de gènes. Ce phénomène est connu sous le nom de "Code des Histones". Bien qu'un nucléosome comporte deux copies de chaque histone, chacune d'entre elles peut être modifiée de manière indépendante, créant ainsi un nucléosome asymétrique sur la base de ses PTM. Dans les domaines bivalents, un état caractéristique de la chromatine des cellules souches embryonniques (ESC), l'histone H3 est méthylé sur la lysine 4 (H3K4me3). Cette modification est activante et elle coexiste avec une autre modification, H3K27me3, qui elle est répressive dans un même nucléosome asymétrique. Dans le premier projet, une stratégie de synthèse générale, modulaire et sans traces pour produire des nucléosomes modifiés de manière asymétrique est décrite. À l'aide de ces nucléosomes modifiés de manière asymétrique je démontre que sur les nucléosomes bivalents, la modification H3K4me3 inhibe l'activité de la lysine méthyletransférase (KMT) "Polycomb Repressive Complex 2" (PRC2) spécifique à H3K27 uniquement sur la région N-terminale de l'histone sur lequel elle est placée. Par contre, H3K27me3 stimule l'activité de PRC2 par un mécanisme de rétroaction positive entre les régions N-terminales d'histones différents, outrepassant partiellement l'effet répressif de H3K4me3.

Les microtubules sont le principal composant du cytosquelette des cellules eucaryotes et jouent un rôle important dans la maintenance de l'organisation cellulaire, le transport

intracellulaire et neuronal, la motilité et la division. Elles sont assemblées de manière dynamique à partir d'hétérodimères $\alpha\beta$ -tubuline, qui en sont les unités de base. La séquence primaire des protéines de tubuline et microtubules est hautement conservée au long de l'évolution des eucaryotes. Malgré cela, la tubuline est sujette à l'hétérogénéité par l'expression de multiples isotypes et un vaste répertoire de PTM, principalement dans la région C-terminale non-structurée. Ceci génère le "Code de la Tubuline". La difficulté d'accès à des tubulines modifiées de manière uniforme est l'une des difficultés majeures à l'étude de la fonction de leurs PTM. J'ai donc développé une méthode pour assembler des parties C-terminales synthétiques et modifiées à de la tubuline recombinante à l'aide d'une méthode basée sur le "split intéin protein trans-splicing" (PTS). Je démontre la faisabilité de cette approche en assemblant une partie C-terminale synthétique et marquée avec un fluorophore à de la tubuline humaine, préparant ainsi un hétérodimer semi-synthétique. L'extension de cette méthode à des PTM naturelles ouvrira la voie à une étude détaillée des modifications de la tubuline en utilisant un système rigoureusement et chimiquement défini.

Mots-clé: *modifications post-traductionnelles (PTM), Code de l'Histone, chromatine bivalente, Polycomb Repressive Complex 2 (PRC2), hétérodimère d' $\alpha\beta$ -tubuline, Code de la Tubuline, split inteins, protéine trans-splicing (PTS)*

Acknowledgements

I owe many thanks to several people who have contributed in many ways during the course of my PhD research.

First of all, I would like to thank my thesis director, Dr. Beat Fierz, who gave me this opportunity to work in his lab. He has been an excellent mentor and continuously supported and motivated me during some of the challenging parts of the project. The continuous discussions that we have had has helped me in improving analytical skills and develop problem solving mindset.

I also want to thank the jury members, Dr. Christian Heinis, Dr. Martin Lochner and Dr. Carsten Janke for taking some of their precious time to review this thesis. I would like to thank Dr. MER Anne-Sophie Chauvin for being the president of the jury.

I would also like to thank several other research groups who collaborated with me during the course of this project. I would like to thank Dr. Pierre Gönczy, Dr. Charlotte Aumeier and Dr. Georgios Hatzopoulos, with whom I collaborated very closely as part of NCCR chemical biology program. I would also like to thank Dr. Carsten Janke for providing tubulin genes his critical inputs during the project. I would like to thank group of Prof. Thomas Schalch who helped with setting up Baculovirus expression system in our lab.

I would like to thank all current and former members of LCBM for making this lab a wonderful working environment. I would like to thank Dr. Carolin Lechner with whom I had a very successful collaboration on asymmetric nucleosome project. I want to extend a special thanks to Dr. Eduard Ebberink who tirelessly worked with me on a very challenging project on tubulin PTMs. I want to thank Dr. Luc Reymond who helped me with the French translation of the abstract for this thesis. I would like to thank Sinan, Aurore, Louise, Andreas, and Nora who were welcoming and helped me get adjusted in the lab. I am also very happy to have found great colleagues in Iullia, Maxime, Harsh, Anne, Karthik, Luc, Ruud and Kristina. They have been very helpful during last four years of my stay at LCBM and I wish them all a very good luck in their careers. I would also like to thank Marie Munoz and Anne Lene Odegaard for their help with administrative processes and Yoann Dind for his help with computer systems.

Lastly, I would like to thank my long time mentor Prof. Haresh Kamdar, who has provided tremendous motivation for me throughout my career. I would like to thank my parents who have emboldened and supported me throughout my life and back me up through challenging times. I am grateful to Mugdha for all the affection, support and motivation she has provided throughout. The contribution of my parents and Mugdha throughout my personal and professional life is simply priceless.

Table of Contents

Abstract	I
Résumé	III
Acknowledgements	V
List of Figures and Tables	IX
1. Review of Literature	1
1.1. Protein Post-Translational Modifications (PTMs)	2
1.1.1. Repertoire and Function of Protein PTMs	4
1.1.2. Functional Importance of Protein PTMs	8
1.2. Chromatin Biology	9
1.2.1. DNA Organization in Eukaryotes	9
1.2.2. Nucleosome – Base Repeating Unit in Chromatin	10
1.2.3. Hierarchical Structure of Chromatin	11
1.2.4. Epigenetics and Histone PTMs	12
1.2.5. Histone PTMs Play an Important Role in Regulation of Gene Expression	14
1.3. Microtubule Cytoskeleton	16
1.3.1. Intracellular Organization and Function of Microtubules	16
1.3.2. Structural Characteristics of Tubulin Dimer	17
1.3.3. Microtubule Structure and Dynamics	19
1.3.3.1. Tubulin Folding into Stable Dimer	19
1.3.3.2. Microtubule Assembly and Dynamic Instability	20
1.3.4. The Tubulin Code	22
1.3.4.1. Tubulin Isozyme Diversity and Intrinsically Disordered Tubulin Tails	22
1.3.4.2. Tubulin PTMs	23
1.3.4.2. Investigating the Tubulin Code	26
1.4. Protein Engineering Tools to Introduce Protein Modifications	27
1.4.1. Genetic Code Expansion	27
1.4.2. Cysteine Chemistry	28
1.4.3. Solid Phase Peptide Synthesis (SPPS)	29
1.4.4. Native Chemical Ligation (NCL)	30
1.4.5. Expressed Protein Ligation (EPL)	32
1.4.6. Split Intein Mediated Protein Tran-Splicing (PTS)	33
1.4.7. Intein Splicing Kinetics	35
2. Objectives	37
2.1. Analyzing the Effect of Asymmetrically Modified Bivalent Nucleosome on Regulation of PRC2 Enzyme Activity	38
2.2. Exploring the Function of Microtubule Post-Translational Modifications by Semi Synthetic Tubulin	38

3. Analyzing the Effect of Asymmetrically Modified Bivalent Nucleosome on Regulation of PRC2 Enzyme Activity	40
3.1. Introduction	41
3.1.1. Embryonic Stem Cells	41
3.1.2. Bivalent Chromatin and Developmental Genes in Embryonic Stem Cells	43
3.1.3. Asymmetrically Modified Nucleosomes in Bivalent Chromatin	44
3.1.4. Polycomb and Trithorax Complexes	45
3.1.5. Installation, Maintenance and Regulation of Bivalent Domains	47
3.2. Aims	49
3.2.1. Traceless Synthesis of Asymmetrically Modified Nucleosomes	50
3.2.2. Understanding the Regulation of PRC2 Activity by Asymmetric Histone Modifications	50
3.3. Results and Discussion	51
3.3.1. Synthetic Strategy	51
3.3.2. Transient Crosslinking of Modified H3 Molecules	52
3.3.3. Reconstitution of Modified Octamers and Nucleosome Library	55
3.3.4. PRC2 Methyltransferase Assay Optimization	58
3.3.5. Effect of H3K4me3 on PRC2 Activity	59
3.3.6. Effects of H3K27me3 on PRC2 Activity	59
3.3.7. Effect of Combinatorial HeK4me3 and H3K27me3 on PRC2 Activity	60
3.4. Conclusion and Outlook	62
4. Exploring the Function of Microtubule Post-Translational Modifications by Semi Synthetic Tubulin	64
4.1. Project Background	65
4.1.1. Recombinant Expression and Purification of Functional Tubulin Dimer	66
4.1.2. Multigene Baculovirus Expression System	67
4.2. Aims of the Project	68
4.3. Results and Discussion	70
4.3.1. Characterization of Intein Splice Junction for Protein Trans-Splicing	70
4.3.1.1. Purification of Tubulin Adapter Proteins	71
4.3.1.2. Protein Trans-Splicing Assays and Kinetics	73
4.3.2. Expression of Tubulin Dimer in Insect Cells	75
4.3.3. Synthesis of Tubulin Tail Peptide	80
4.3.4. PTS Tests in Cleared Cell Lysates to Generate Fluorescein Tagged Tubulin	82
4.3.5. Purification of Functional Tubulin Dimers	83
4.3.6. Purification of Semisynthetic Tubulin Dimers	85
4.3.7. Microtubule Pelleting and Immunofluorescence Imaging	88
4.4. Conclusion and Outlook	90

5. Thesis Summary	93
6. Materials and Methods	96
6.1. Reagents and Instrumentation	97
6.2. Methods	99
6.2.1. Octamer Refolding	99
6.2.2. Nucleosome Reconstitution	99
6.2.3. PRC2 Methyltransferase Assay	100
6.2.4. Immunoprecipitation of Nucleosomes	100
6.2.5. Gibson Assembly Cloning	101
6.2.6. Cloning, Expression and Purification of Tubulin Adapter Proteins	102
6.2.7. Protein Trans-Splicing Assays	102
6.2.8. Cloning and Expression of Recombinant Tubulin Dimer Using Baculovirus Expression System	103
6.2.9. Purification of Recombinant Tubulin Dimer	104
6.2.10. Microtubule Imaging by Wide Field Fluorescence Microscopy	106
6.2.11. Tubulin Dimer Resolving SDS-PAGE	106
6.2.12. Synthesis of Tubulin Tail Peptide	107
7. References	109
8. Curriculum Vitae	128

List of Figures and Tables

Figure 1.	Central dogma in molecular biology	2
Figure 2.	Posttranslational modifications are key mechanisms to increase proteomic diversity	3
Figure 3.	Types of posttranslational modifications of	3
Table 1.	Protein Posttranslational Modifications at side chains	4
Figure 4.	Chemical Structure of Phosphorylated Amino Acids	5
Figure 5.	Protein acylation and alkylation modifications	6
Figure 6.	Mechanism of protein splicing and Intein excision	8
Figure 7.	Chromatin organization in the eukaryotic cell	10
Figure 8.	Structure of the nucleosome.	11
Figure 9.	Chromatin Structure Hierarchy	12
Figure 10.	Histone post translational modifications.	13
Figure 11.	Reader domains as histone PTM binding modules	15
Figure 12.	Analogy between Tubulin Code and Histone Code.	16
Figure 13.	Microtubule Cytoskeleton Organization and Function	17
Figure 14.	Structure of $\alpha\beta$ -tubulin dimer (PDB 1TUB)	19
Figure 15.	Tubulin Dimer Folding	20
Figure 16.	Microtubule Structure and Dynamic Instability	22
Figure 17.	Tubulin Isozyme Diversity.	23
Figure 18.	Tubulin diversity generated by post-translational modifications	24
Figure 19.	Site specific modifications of proteins by Expansion of Genetic	28
Figure 20.	Cysteine Chemistry for Protein Modifications	29
Figure 21.	General Fmoc SPPS Strategy	30
Figure 22.	Native Chemical Ligation	30
Figure 23.	Thioester formation for NCL and desulfurization of cysteine	32
Figure 24.	Expressed Protein Ligation	33
Figure 25.	Protein splicing mediated by inteins	34
Figure 26.	Mechanism of Protein Trans-Splicing	35
Figure 27.	Cell potency during embryonic development	41
Figure 28.	Potential conformations at Bivalent Domains	45
Figure 29.	Subunit Composition and interplay of PRC1 and PRC2 complexes.	47
Figure 30.	Factors Involved in Resolution of Bivalent Domains during Differentiation	49
Figure 31.	Scheme for Traceless Chemical Synthesis of Asymmetrically Modified Octamers	51
Figure 32.	Synthesis of ^{Inc*} H3K4me3	53
Figure 33.	Synthesis of ^{Inc} H3K27me3	54
Figure 34.	Synthesis of ^{xInc} H3K4me3/H3K27me3	54
Figure 35.	Refolding Asymmetrically Modified Octamers	55

Figure 36.	Octamer Library	56
Figure 37.	Reconstitution of trans-bivalent nucleosomes	57
Figure 38.	Nucleosome Library	57
Figure 39.	PRC2 Methyltransferase assay	58
Figure 40.	Effect of H3K4me3 on PRC2 activity	59
Figure 41.	Effect of H3K27me3 on PRC2 activity	60
Figure 42.	Regulation of PRC2 in asymmetric Bivalent Nucleosomes	61
Figure 43.	Model for PRC2 Regulation by Asymmetrically Modified Nucleosomes	61
Figure 44.	Expression of recombinant human $\alpha\beta$ -tubulin	67
Figure 45.	MultiBac™ Baculovirus system	68
Figure 46.	Tubulin Modification Strategy	69
Figure 47.	Selection of trans-splicing junction for tubulin	71
Figure 48.	Purification of Npu ^C -Tubulin tail-ubiquitin	72
Figure 49.	Purification of α -tubulin adapter	72
Figure 50.	Purification of β -tubulin adapter	73
Figure 51.	PTS reaction for α -tubulin adapter	74
Figure 52.	PTS reaction for β -tubulin adapter	74
Figure 53.	PTS Rate kinetics	75
Figure 54.	Design of Tubulin Expression Constructs for Baculovirus Expression	76
Figure 55.	Generation of Multigene Transfer Vectors	77
Figure 56.	Analysis of Bacmid	78
Figure 57.	Tubulin Expression in Insect Cells	79
Figure 58.	Time Course Expression Analysis of Recombinant Human $\alpha\beta$ -tubulin	79
Figure 59.	Fluorescein labelling of tubulin tail by copper catalyzed click chemistry	80
Figure 60.	Synthesis of Npu ^C -tubulin tail peptide	81
Figure 61.	Native Chemical Ligation to generate the Npu ^C -tubulin tail	81
Figure 62.	PTS is Cleared Cell Lysates	82
Figure 63.	Purification of Recombinant Tubulin Dimer	83
Figure 64.	Purification and Polymerization of Recombinant Tubulin	85
Figure 65.	Anion exchange purification for Tubulin-Ava ^N construct	86
Figure 66.	PTS is Cleared Cell Lysates	86
Figure 67.	Purification of Semisynthetic Tubulin Dimer	87
Figure 68.	Immunofluorescence Microscopy of Microtubules (from Porcine Tubulin) in the Presence of CEP135	88
Figure 69.	Immunofluorescence Microscopy of Microtubules (from Recombinant $\alpha\beta$ -tubulin).	89
Figure 70.	Project Outlook	91
Figure 71.	Thesis Summary	94

1. Review of Literature

1.1. Protein Post-Translational Modifications

The central dogma of molecular biology¹, first explained by Francis Crick in 1958 provides information regarding the flow of genetic information within the biological system. The dogma states three classes of information transfers, general transfers (occurs normally in most cells), special transfers (only under specific circumstances) and unknown transfers (postulated by the dogma to never occur) (**Figure 1**). The general transfers are further classified in three classes of flow of information, DNA information copied into DNA (DNA replication), DNA information copied into mRNA (transcription) and synthesis of proteins using information in mRNA as a template (translation). Three special transfers are RNA copied into RNA (RNA replication), synthesis of DNA using RNA as template (reverse transcription) and protein synthesis directly using DNA template without mRNA (only in cell-free systems)². Transfers that do not occur naturally are protein being copied from protein and synthesis of DNA or RNA using information from protein primary structure.¹

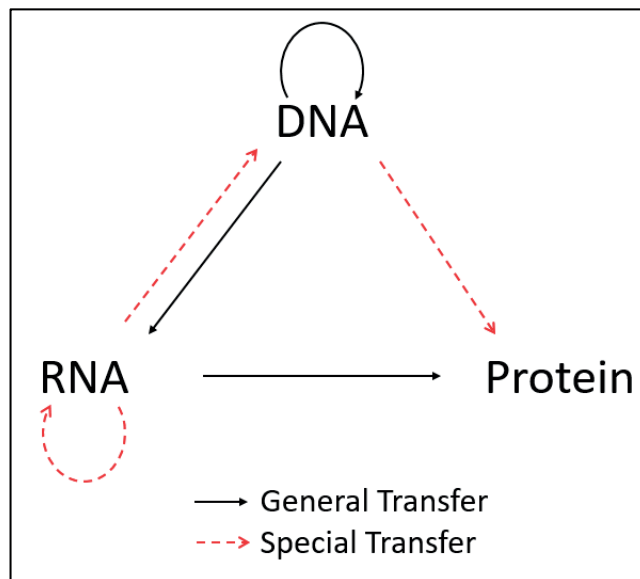


Figure 1. Central dogma in molecular biology. Solid arrows show general transfer of genetic information. Dashed arrows show special transfer of genetic information.¹

The human genome has ~25000 protein coding genes.³ The human proteome, the inventory of all proteins, is several order of magnitude more complex than the underlying genome would predict (>1000000 proteins) as shown in **Figure 2**.⁴ There are two major mechanisms that lead to diversification of proteins. The first is at the transcription level, genomic recombination, transcription initiation at alternative promoters, differential transcription termination, and alternative splicing of the transcript are mechanisms that generate different mRNA transcripts from a single gene.⁵⁻⁷ The second route of the proteome expansion is the focus of this thesis, covalent posttranslational modifications (PTM) of proteins at one or more sites. These modifications occur after DNA has been transcribed into mRNA and translated into proteins.

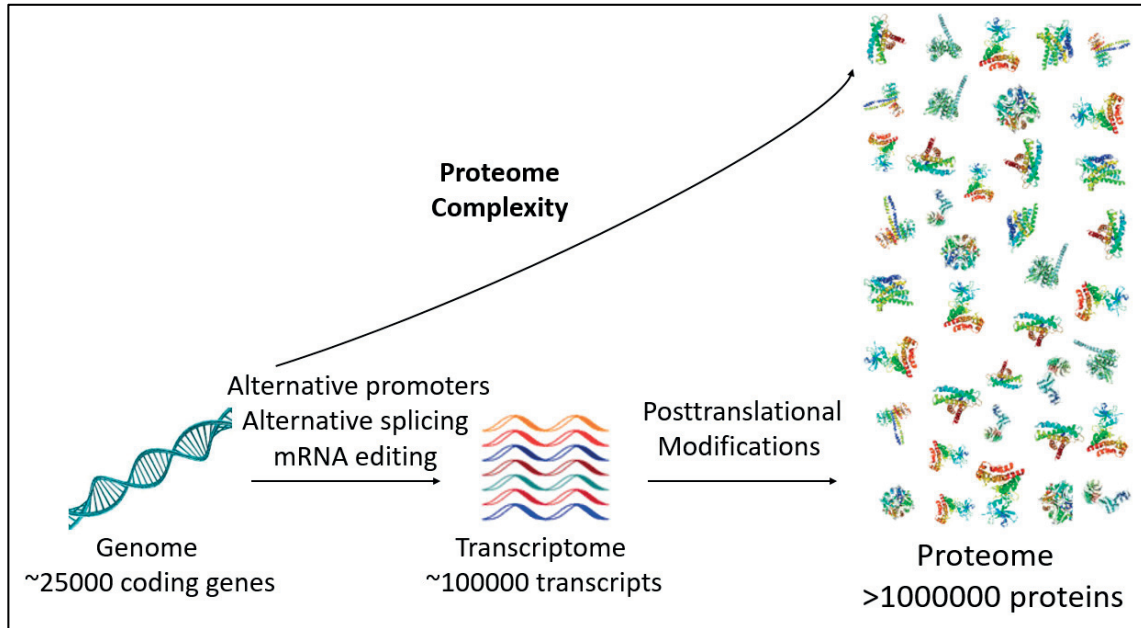


Figure 2. Posttranslational modifications are key mechanism to increase proteomic diversity. Human genome consists of ~25000 coding genes, but myriad of transcription modifications and posttranslational modifications increase the complexity of protein to >1000000.

The nascent or folded proteins are subjected to a variety of specific enzyme catalyzed modifications on the side chain or peptide backbone. About 5% of genomes of higher eukaryotes can be attributed to enzymes that carry out posttranslational modification of other proteins.⁴ There are two broad types of protein PTMs as shown in **Figure 3**. The first group includes covalent additions of a chemical group, usually an electrophilic fragment of a co-substrate, to a side chain residue in a protein. The side chain modified is usually electron rich, acting as a nucleophile in the transfer. The second category of PTM is covalent cleavage of peptide backbones in proteins either by action of proteases or, less commonly, by autocatalytic cleavage.^{4,8}

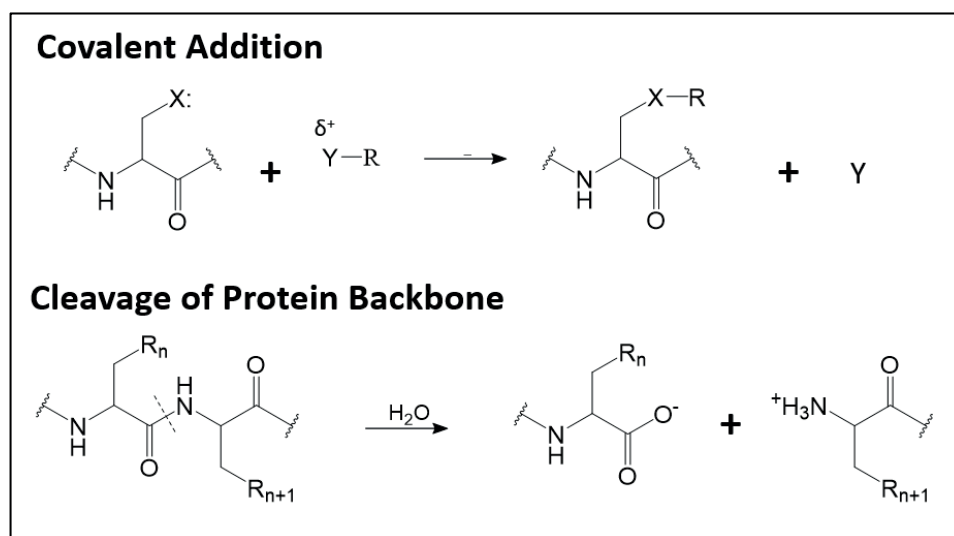


Figure 3. Types of posttranslational modifications of proteins: 1) Covalent modification of amino acid side chain, 2) Cleavage of protein backbone at specific peptide bond.

1.1.1. Repertoire and Function of Protein PTMs

Protein covalent modification are further classified in several types. One is based on the amino acid side chain that is being modified. Two, based on the fragment of co-substrate or coenzyme that is enzymatically coupled to protein. Three, based on the gain of function by the protein enabled by the covalent addition such as gain in catalytic function, change in subcellular location and mark for proteolytic degradation of protein. The following **Table 1** provides a non-exhaustive list of known protein PTMs classified as per above three classes

Residue	Modification	Example
Asp	Phosphorylation	Protein tyrosine phosphatases
Glu	Methylation, Carboxylation, Polyglycylation Polyglutamylation	Chemotaxis receptor protein Gla residues in blood coagulation tubulin
Ser	Phosphorylation, O-glycosylation, auto cleavages	Protein serine kinases and phosphatase, Notch O-glycosylation
Thr	Phosphorylation, O-glycosylation	Protein threonine kinases and phosphatases
Tyr	Phosphorylation, Sulfation, Ortho-nitration,	Tyrosine kinases and phosphatases Inflammatory response
His	Phosphorylation, N-methylation	Sensor protein kinases Methyl CoM reductase
Lys	N-methylation, N-acylation by acetyl, biotinyl, ubiquityl groups	Histone methylation Histone acetylation Histone ubiquitination
Cys	S-hydroxylation, disulfide bond formation, phosphorylation, S- acylation, Protein splicing	Sulfonate intermediates Protein in oxidizing environments PTPases Ras Intein excision
Met	Oxidation to sulfoxide	Met sulfoxide reductase
Arg	N-methylation	Histones
Asn	N-glycosylation, Protein splicing	N-glycoproteins, Intein excision step
Gln	Transglutamination	Protein cross-linking
Trp	C-mannosylation	Plasma membrane proteins
Pro	C-hydroxylation	Collagen, HIF1- α
Gly	C-Hydroxylation	C-terminal amide formation

Table 1. Protein Posttranslational Modifications at side chains ⁴

Protein PTMs by Covalent Additions

Some of the most common of the modifications and their functions are described in the following section. The most common types of covalent additions are phosphorylation, acylation, alkylation, glycosylation and oxidation and these are catalyzed by dedicated PTM enzymes.⁴

In mammalian proteins, most commonly phosphorylated residues are phosphoSer(pS), phosphoThr(pT), and phosphoTyr (pY) as shown in **Figure 4** with a split of about 90:10 (pS, pT/pY).^{4,9} In bacteria and fungi, phosphorylation of His and Asp residues (phosphoHis and phosphoAsp) have been seen in proteins in two-component signal transduction cascades.^{4,10} The superfamily of protein kinases, kinome, are dedicated to protein phosphorylation and are among the largest class of protein PTM enzymes.⁴ The human kinome consists of >500 kinases.^{4,11} The addition of a charged phosphate group induces conformational changes in the local protein environment, which often create an impetus for signal initiation such as seen in Map kinase pathways and activation of membrane receptor tyrosine kinases during auto phosphorylation.^{4,12}

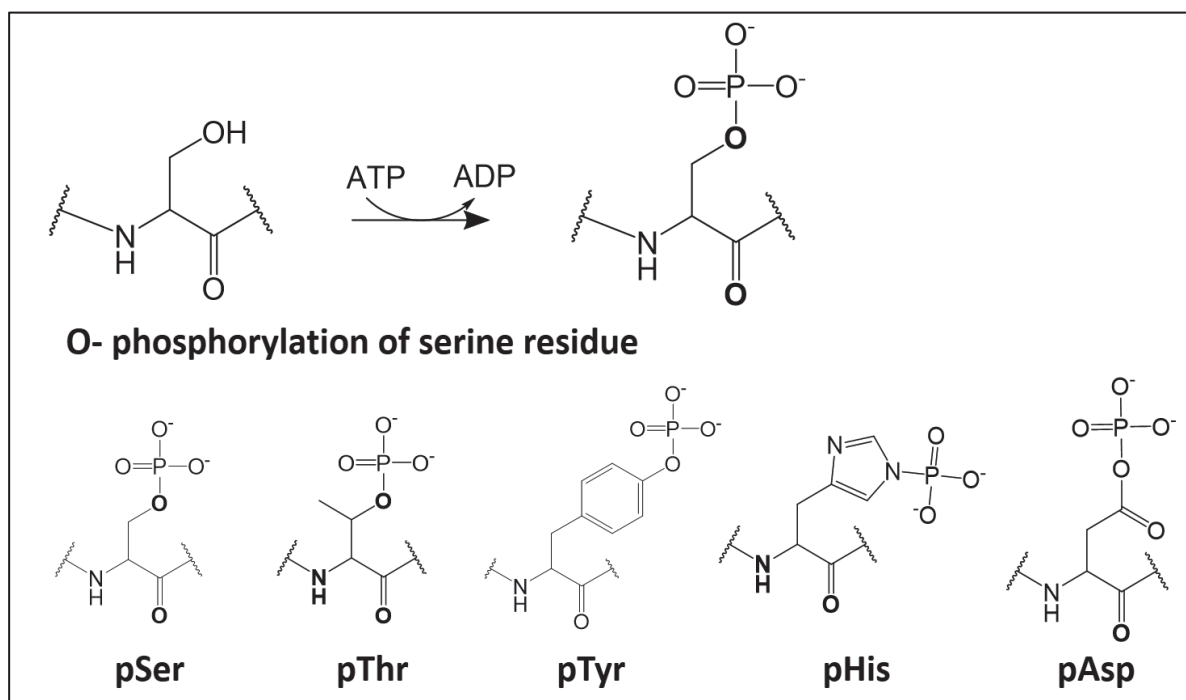


Figure 4. Chemical Structure of Phosphorylated Amino Acids. Protein phosphorylation at the serine residue and phosphorylated forms of amino acid side chains in proteins.

One of the most well-known acylation modification is the acetylation of lysine residue for which Acetyl CoA acts as donor of the acetyl group as shown in **Figure 5**. The acetylation is catalyzed by variety of protein specific acetyltransferases enzymes (such Histone Acetyltransferases (HAT) isoenzymes).⁴ Small protein such as ubiquitin (with molecular weight of 8 kDa) is also added to the protein as an acyl moiety and thus is classified into acylation modifications.^{4,13} Eukaryotic cells make use of ubiquityl-S-Protein instead of ubiquityl CoA as donor.^{14,15} The enzymatic activation involves 3 set of enzyme families, enzyme 1 (E1) to make

ubiquityl-AMP, a set of thiol containing enzymes 2 (E2), which provide active site for nucleophile capture, and enzyme 3 (E3) variants that are required for catalysis of transfer of activated ubiquityl protein to Lysine side chains on selected proteins.⁴ There are several isoforms of E3 ligases which allow for discrimination among many target proteins that are ubiquitylated.¹⁶ Protein ubiquitylation can be of two types, monoubiquitylation or polyubiquitylation. Both the types often mark designated proteins for relocation and proteolytic degradation.⁴

One of the most well-known alkylation modification is the methylation of Lys residue. Histone H3 is well known to be methylated by variety of histone methyltransferases.¹⁷ An additional layer of information complexity is added by the fact that Lys- ϵ -NH₂ groups can be progressively mono-, di-, or tri-methylated, again with distinct distributions shown by different methyltransferases as shown in **Figure 5**.⁴

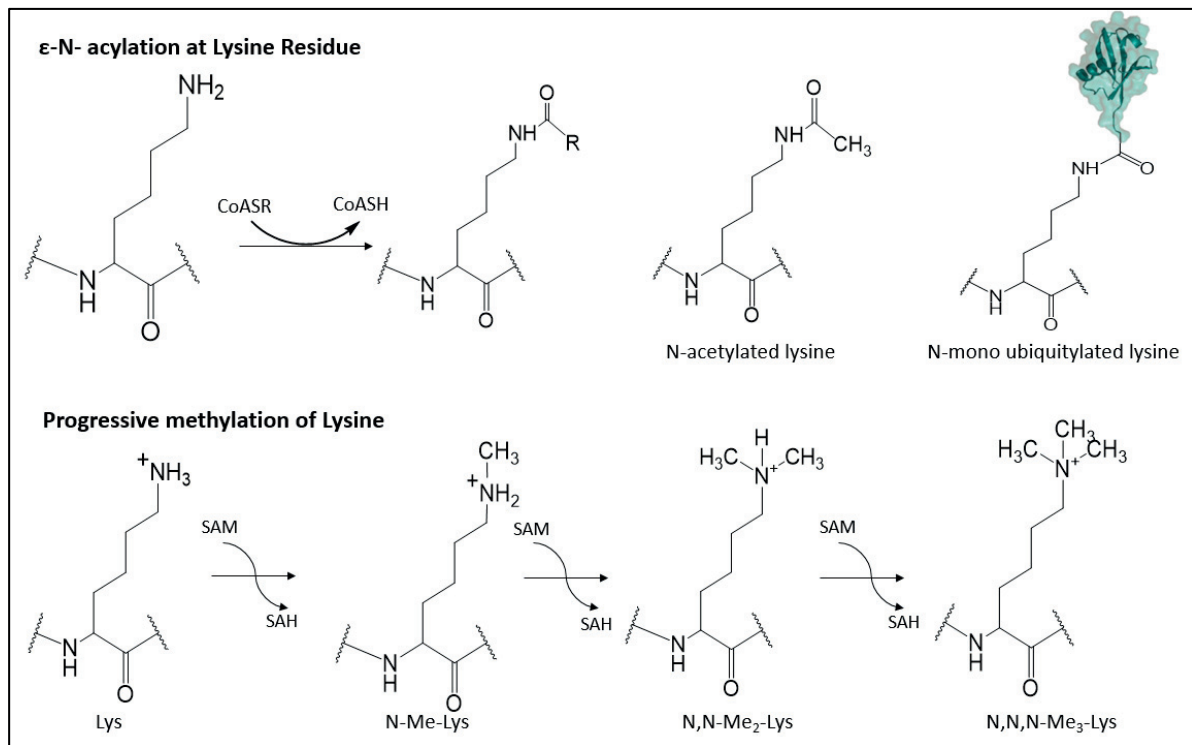


Figure 5. Protein acylation and alkylation modifications. General reaction of Lysine residue acylation with example of acetylated lysine and ubiquitylated lysine residue. Progressive mono-, di-, tri-methylation of lysine side chains.

Protein PTMs play important role in various biological processes and pathways. Depending on their role, reversibility is often an important factor for regulation of such PTMs. There are myriad of enzymes that catalyze the removal of covalent modifications. Protein kinases usually occur in a low activity “off” state in the absence of stimulus and are activated by a stimulus usually through autophosphorylation.¹² Activated kinases modify their target proteins with much higher efficiency.¹² To control the duration and intensity of signaling through phosphorylation, the modification needs to be terminated followed by removal of PO₃²⁻. There are about 150 phosphoprotein phosphatases encoded in human genome.¹⁸ Phosphoproteome content of any cell at a given time is the balance between the activity of

~500 kinases and ~150 phosphatases towards their diverse substrates, thus generating enormous diversity and complexity.¹⁹ Similar to protein phosphorylation, protein acetylation and protein ubiquitylation can be readily reversed. Histone acetyltransferase (HATs) and functionally opposed by the family of histone deacetylases (HDACs) which catalyze removal of acetyl group.^{20,21} Unlike the acetyl group, histone Lys-methylation is not susceptible to enzymatic or non-enzymatic hydrolytic removal of the methyl group, but is reversed by action of FAD-dependent methyl-Lys deaminases through an oxidative removal route.²² Ubiquitin is attached to proteins through an isopeptide bond which is resistant to normal proteases. Such linkages are cleaved by deubiquitylases (DUBs) rescuing proteins from degradation.⁴

Protein PTMs by Proteolytic Cleavage of Peptide Bonds

The life cycle of a protein is controlled by homeostatic function of proteases, which cleave peptide backbones to release the amino acids back to monomer pool.⁴ Proteins also undergo proteolytic cleavages as part of maturation processes. Essentially, every protein that enters endoplasmic reticulum (ER) undergoes cleavage of N-terminal signal sequence by signal peptidases.²³ Proteins are further processed by maturation reactions by protein convertases in Golgi and trans Golgi network of the secretory compartments.^{4,24} Folded proteins also can undergo peptide backbone rearrangements at one or more peptide bonds through autocatalytic processes. This often leads to cleavage of specific peptide bonds.⁴ One of the important examples of such modification is detyrosination and tyrosination of tubulin. Most α -tubulin genes encode for a C-terminal tyrosine residue which is removed by detyrosination and added back by tyrosination depending of physiological signal.²⁵ Tyrosination is catalyzed by TTL enzymes while the detyrosination is catalyzed by the Vasohibins family of proteins.²⁶ Following detyrosination, additional residues can further be removed to generate $\Delta 2$ and $\Delta 3$ tubulin. This process is catalyzed by the CCP family of enzymes. The $\Delta 2$ tubulin can't be tyrosinated.^{25,27,28}

Another special variant of such autocatalytic cleavage is protein splicing and excision of inteins. Intein-based methods have been extensively used in the projects of this thesis. Inteins are found in bacteria, archaea and eucarya embedded in essential proteins involved in DNA replication, transcription and maintenance.^{29,30} Intein splicing is an auto-catalytic event where an intein (intervening protein) excises itself out from a host protein by cleaving two peptide bonds and ligates the exteins (external protein) through the formation of a new peptide bond.³¹ Inteins can either be expressed as a single transcript or two transcripts (split inteins). Based on this inteins splicing can be classified in two categories; cis-splicing and trans-splicing. The cis-splicing (see **Figure 6**) is the most common type of intein, where the intein is expressed as a single transcript between two exteins. The splicing reaction is initiated by the cysteine at the N-terminus of the intein that attacks the carbonyl carbon of the C-terminal N-extein. The N- to S-acyl shift forms a linear thioester. The cysteine, which is the first residue of the C-extein attacks the thioester leading to a transthioesterification. The C-terminal asparagine of the intein cyclizes to form a succinimide and excises the intein. The new bond at the extein, rearranges through a S- to N-acyl shift to result in a peptide bond.³² In case of

split inteins, the two halves (N-intein and C-intein) are expressed separately fused to exteins. Individually they are inactive, however they associate non-covalently to form the full intein, the trans-splicing reaction proceeds same as cis-splicing reaction. Split-intein mediated tools are extensively used for protein engineering and are described later parts of this thesis in further details.

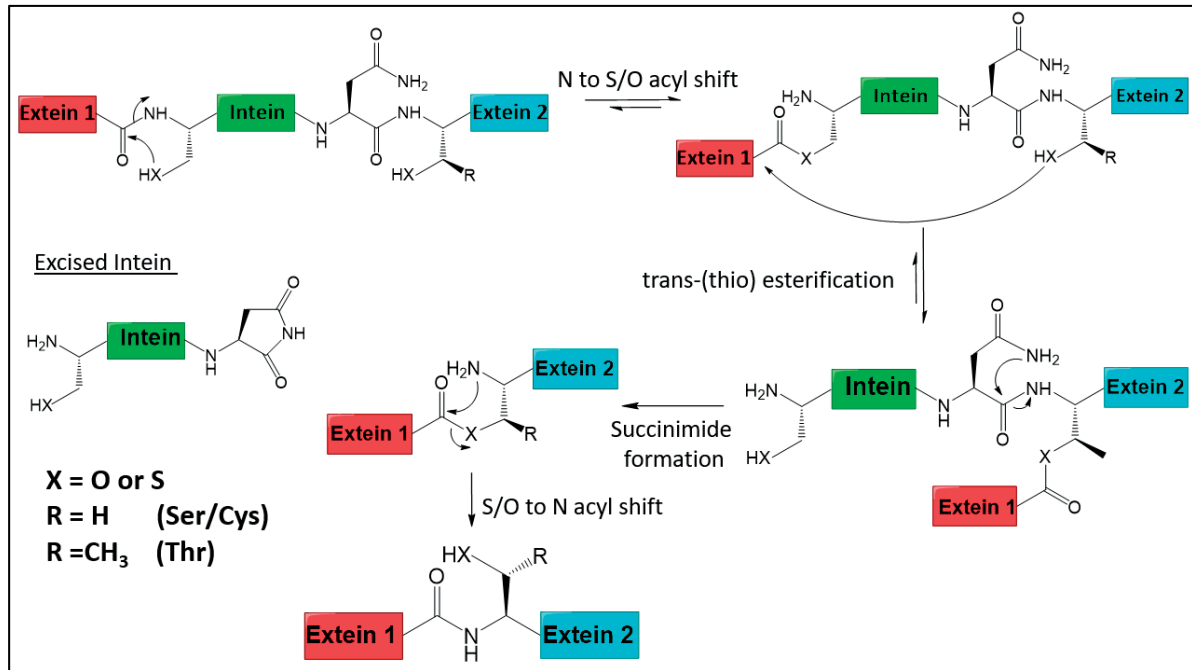


Figure 6. Mechanism of protein splicing and intein excision. The cysteine attacks the carbonyl group of the peptide bond at the C-terminal of the N-extein, leading to an N to S acyl shift, followed by transthioesterification. The C-terminal asparagine of the intein then cyclizes and excises out the intein. The peptide bond is found after a S to N acyl shift.

1.1.2. Functional Importance of Protein PTMs

Cells need to constantly sense and respond to the changes in internal and external conditions. One of the major way of relaying the changes from sensors or effectors is by the way of protein PTMs. The repertoire of protein PTMs is regulated by enzymes that add (writer domains) or remove (eraser domains) the modifications and the functional consequences are mediated through ‘reader’ domains that recognize and bind to specified PTM mark.^{33,34} PTMs influence several properties of proteins including catalytic activity, protein interactions, subcellular location and protein life cycle. PTMs themselves are highly dynamic in nature, and often cells are equipped with enzymes with opposing activities to add or remove PTMs on response to functional stimuli.³⁵

Despite the importance of PTMs in cellular biology, identification and functional investigation of specifically modified substrate remains challenging task. It is often difficult to separate modified and unmodified proteins that are harvested from cells. Also, a single

protein may be modified at various sites simultaneously thus making isolation of completely homogenous population of specifically modified protein challenging.³⁵ Thus, tools to site specifically modify proteins to understand the site-specific effects of particular modification are of utmost importance.

Chemical techniques have been extremely valuable for synthesis and identification of site-specifically modified proteins. Protein-peptide ligations for semi synthesis of modified proteins and site-specific chemical reaction for installation of modifications have expanded the toolbox to investigate PTMs function.³⁵ These techniques paired with DNA sequencing techniques, chemical reporters and MS methods have allowed identification of new modifications and array of binding partners and protein targets for drug discovery.³⁵

The main focus of my projects is to understand the role of protein posttranslational modifications and develop protein engineering tools to investigate the biological function of protein PTMs in the context of chromatin and microtubule cytoskeletal biology.

In the subsequent parts of this chapter, I will discuss about the structure of eukaryotic chromatin with focus on the nucleosome, histone PTMs and their role in the epigenetic regulation of gene expression. I will then discuss the structure and organization of the microtubule cytoskeleton, the important features of tubulin dimers (basic building block of microtubules), the repertoire of tubulin PTMs and the concept of tubulin code. In the last part of this chapter, I will describe variety of protein engineering tools that are employed to study protein PTMs.

1.2. Chromatin Biology

1.2.1. DNA Organization in Eukaryotes

The human genome has ~6 billion base pairs and when fully extended is about two meters in length.^{3,36} Yet, it must be packaged in cell nucleus approximately 10 μm in diameter. Within this context, the underlying genetic material must be transcriptionally active or repressed in dynamic and coordinated fashion in order to allow the cell to react to an ever changing environment. To achieve this compaction and coordinated accessibility, nuclear DNA is organized in the form of chromatin which dynamically controls the access to genes by transcription machinery. Chromatin has several layers of organization as shown in **Figure 7**.

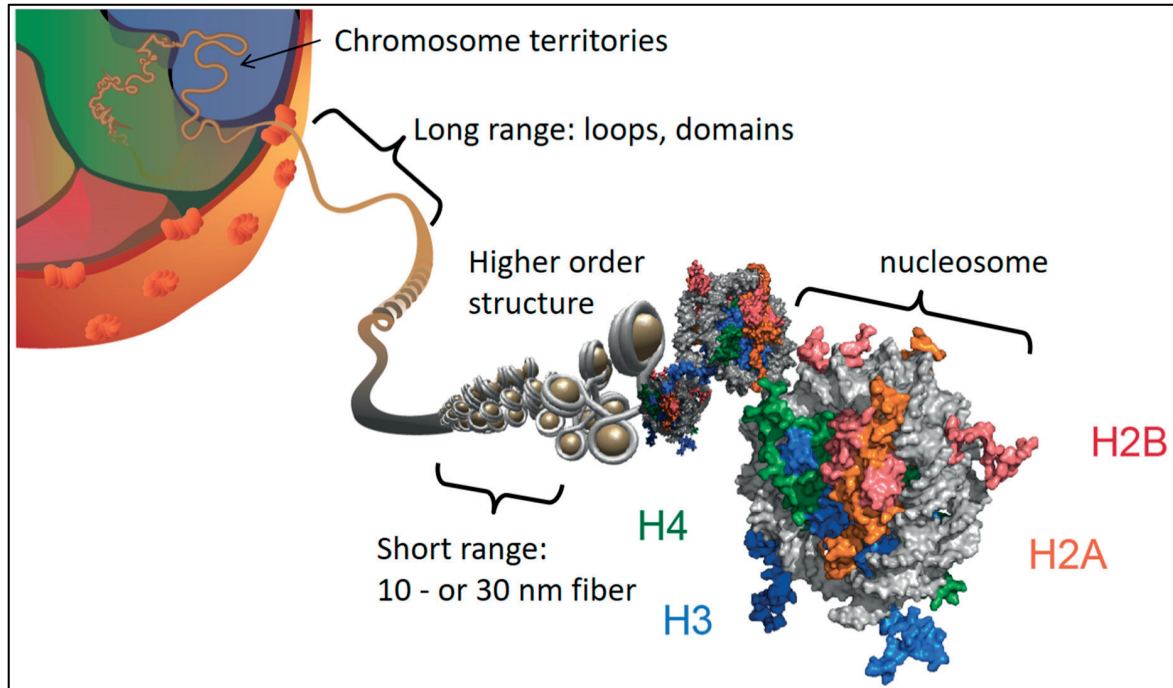


Figure 7. Chromatin organization in the eukaryotic cell nucleus. Chromatin is densely packed in the cell nucleus with the fundamental unit being the nucleosome consisting of DNA wrapped around the core histones H2A, H2B, H3 and H4. (Figure taken from Fierz & Muir, 2012)³⁷

1.2.2. Nucleosome - Base Repeating Unit in Chromatin

The nucleosome is the base repeating unit of the chromatin. It was first observed in EM analysis as “beads on string” structure of chromatin.^{38,39} The crystal structure of the nucleosome was solved by Luger et al. in 1997 and is shown in **Figure 8**.⁴⁰ The nucleosome core particle consists of 147bp of DNA wrapped around eight histone core proteins, two copies each of H2A, H2B, H3 and H4.⁴⁰ This forms a disk shaped structure with diameter of 10 nm and height of 2 nm.^{36,40} Some of the key highlights that are observed from the crystal structure of nucleosome core particle are,⁴⁰ the DNA wraps around the histone octamer in a left handed super helix. A single base pair on the dyad axis mark a pseudo 2-fold symmetry axis. The nucleosome core particle is stabilized by the interaction between the histone proteins and electrostatic interactions or hydrogen bonds between the proteins and the minor groove of DNA.⁴¹ The core histone proteins are highly positively charged and they contain a histone fold made of three alpha helices (α 2-central helix, α 1 and α 3- short helices) and two loops (L1 and L2) linking these helices.⁴⁰ Histones have N-terminal unstructured tails (H2A also has a flexible C-terminal tail). These tails protrude out of the nucleosome core and provide platform for a large number of PTMs.³⁶ Each histone binds to a complementary histone (histone H2A pairs with H2B and histone H3 pairs with H4) at the histone fold. The external surface, composed of the L1 and L2 loops and the α 1 helix, provide a strong positive charge creating an important platform for DNA interaction. The internal concave surface is composed of the α 3 and α 2 helices.^{40,42,43}

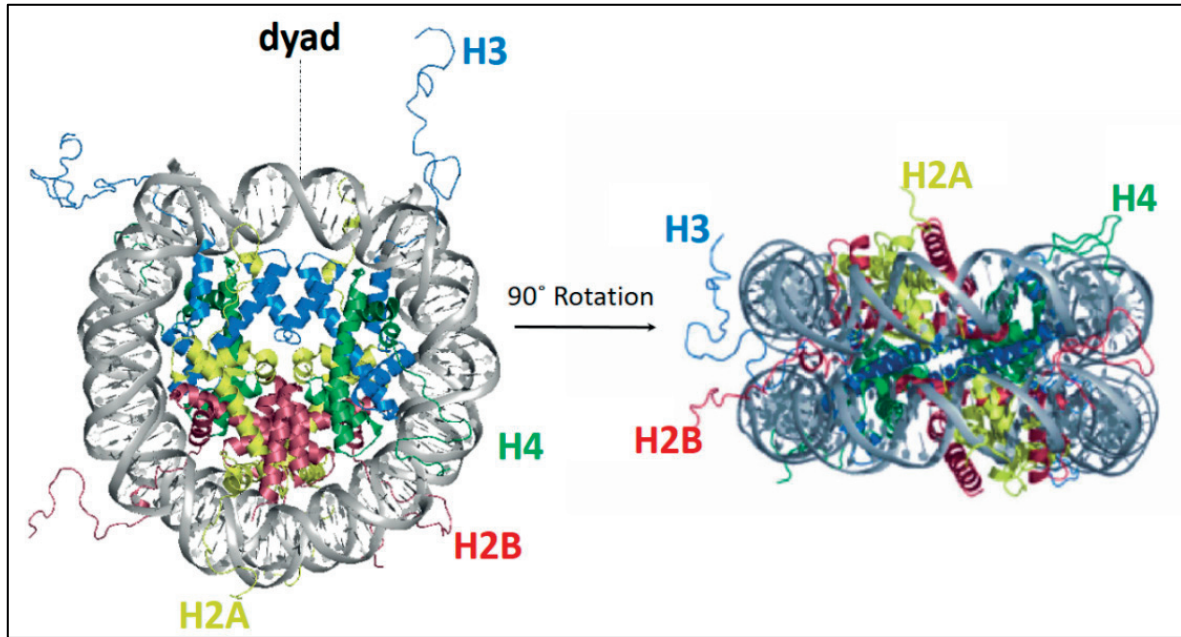


Figure 8. Structure of the nucleosome. 147bp of DNA wrap around eight histone proteins; 2 copies of H2A (yellow), 2 copies of H2B (red), 2 copies of H3 (blue) and 2 copies of H4 (green) (PDB: 1KX5).

1.2.3. Hierarchical Structure of Chromatin

Nucleosome arrays are linked by linker DNA which can vary between 10-80 base pairs in length.⁴⁴ This forms the primary structure of chromatin known as “beads on string” shown in **Figure 9(A)**.⁴⁵ The nucleosome arrays can further fold into 30 nm chromatin fibers, known as the secondary structure.⁴⁶ There are two structures proposed for 30nm fiber, a two-start helix model with zigzag conformation was initially proposed by Woodcock *et al* shown in **Figure 9(B)**⁴⁷ and the solenoid one-start helical conformation was first proposed by Finch and Klug in 1976.⁴⁶ The crystal structure of tetra nucleosome (shown in **Figure 9(C)**) and other studies support two start model of the 30 nm fiber.^{48–50} The 30 nm fiber is only a short range organizational motif, whereas over longer length scales chromatin is highly heterogeneous.⁵¹ Chromatin in the nucleus can form higher-order structures by interaction of secondary structures. Transcriptional enhancers interact with target genes by the formation of chromatin loops. Chromatin loops are constrained within larger loops called insulated neighborhoods. Clustered neighborhoods form sub-megabase scale domains contributing to topologically associated domains (TADs)⁵². Chromatin compaction changes throughout the cell cycle have been investigated by chromatin conformation capture (3C, 5C, Hi-C) type methods. In the G1 phase, TADs are insulated from one another, by occupying distinct nuclear compartments. In the S phase, where DNA replication takes places, the TADs are still present, however less insulated. In the M phase, chromatin is highly compacted with very short-range contacts between different TADs.^{53,54}

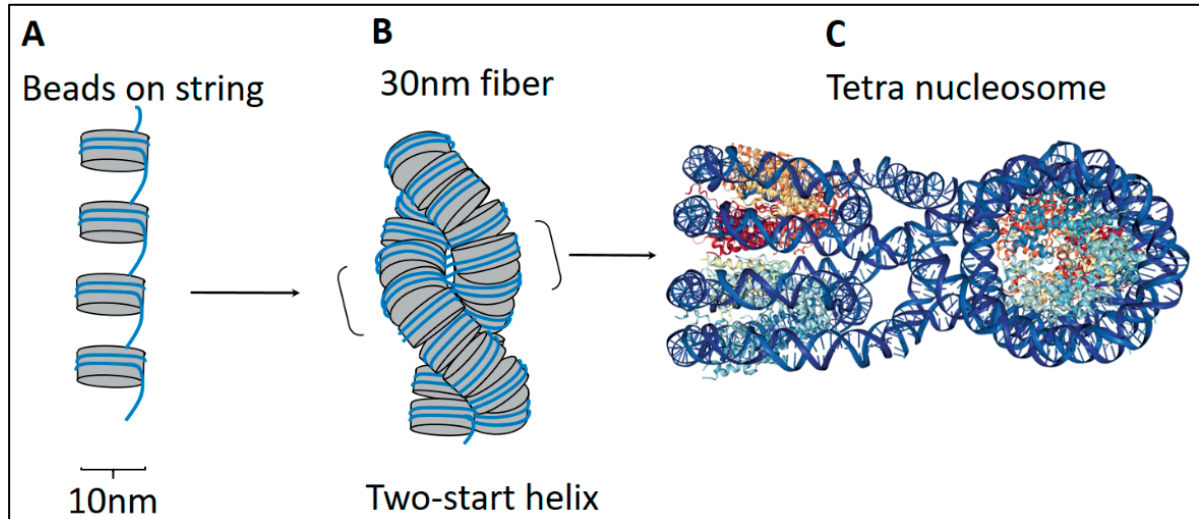


Figure 9. Chromatin Structure Hierarchy. A) The primary structure of chromatin as a 'beads on a string' structure. B) Chromatin is further compacted into a 30 nm fiber structure. C) Tetra nucleosome two-start helix crystal structure (PDB: 1ZBB).

1.2.4. Epigenetics and Histone PTMs

Histone proteins contain a globular domain and unstructured tails that extend out of the nucleosome. These tails display various post translational modification (PTMs) (**Figure 10 A**). These PTMs vary from small modifications such as methylation, acetylation, phosphorylation to larger proteins such as ubiquitin or SUMO (**Figure 10 B**).⁵⁵ Histone marks are involved in a variety of chromatin-associated events such as transcription, genome stability, chromatin compaction, DNA replication and repair.⁵⁶ Histone PTMs form one of the major layers of epigenetic regulations. The term of "epigenetics" was first used by Waddington in 1939 referring to the link between genes and development.⁵⁷ The present definition describes epigenetics as "a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence".⁵⁸ There are three main layers of epigenetic regulations, DNA methylation involved in gene silencing and gene imprinting,^{59,60} a large number of histone PTMs and RNA mediated silencing as observed with X-chromosome inactivation.⁶¹ The focus of this section is the variety of histone PTMs shown in Figure 10 and their role in epigenetic regulations.

DNA and histone modifications are labile, thus raising questions regarding their inheritance over generations as suggested in the definition of epigenetics. Evidence that the DNA methyltransferase DNMT1 is preferentially active on hemi-methylated DNA suggests the hypothesis that DNMT1 installs methyl groups on the newly synthesized DNA strand at methylated positions of the parental strand.^{62,63} DNA methylation at imprinted genes is also an example of trans generationally inherited epigenetic marks.^{59,60} Regarding histone modifications, several studies showed that methylation at Lysine 9 and 27 of histone H3

1.2.5. Histone PTMs Play an Important Role in Regulation of Gene Expression

Chromatin is organized into distinct domains such as euchromatin and heterochromatin based on the level of compaction and associated function. Euchromatin has relatively loose compaction and thus is transcriptionally permissive, whereas heterochromatin (facultative or constitutive) is condensed and transcriptionally repressive.^{69–71} The degree of chromatin compaction is regulated by a number of factors, including linker histone H1, histone variants substituting canonical histones, chromatin remodelers, histone chaperones involved in the deposition and eviction of histones and covalent modifications of DNA and histones.⁶⁹

Fundamental breakthroughs in understanding the role of histone PTMs have occurred with identification of proteins that deposit (writer), remove (eraser) and bind (reader) these histone PTMs.⁶⁹ These findings gave rise to the histone code hypothesis.^{55,72} Histone PTMs are read in a combinatorial manner by multiple effector proteins to mediate distinct biological functions associated with them.⁶⁹ Histone PTMs modulate DNA accessibility and chromatin structure in two distinct ways, directly altering DNA-histone or histone-histone interactions (cis-effect) or by recruiting non-histone proteins at specified sites marked by histone PTMs (trans-effect).⁵⁶

Some of the histone PTMs manifest their function by modifying the charge on the amino acid and thereby affecting chromatin structure. For example, acetylation neutralizes the positive charge of lysine residues, whereas phosphorylation adds negative charge (see **Figure 10B**). Acetylation of lysine 16 on H4 (H4K16ac) is thought to prevent the interaction of the H4 tail to the acidic patch of the neighboring nucleosome and leads to decompaction of chromatin fibers.⁷³ H3K56Ac, found at the entry/exit site of DNA, leads to the removal of the positive charge on lysine 56. The interaction between H3K56 and DNA is thereby lost and leads to unwrapping of DNA.⁷⁴ H3T118ph is located at the dyad axis region, increases accessibility to DNA at the dyad axis but not at the DNA entry/exit site and leads to increase in the remodeling factor SWI/SNF activity compared to unmodified nucleosomes⁷⁵. PTMs such as ubiquitylation increases steric bulk as seen in case of H2BK120Ub which prevents nucleosome compaction.^{76,77} These examples indicate the important role of histone PTMs in regulating DNA wrapping in nucleosome. Nucleosome in itself is highly dynamic structure susceptible to unwrapping and rewrapping of DNA, sliding of the octamers on the DNA, assembly and disassembly of the nucleosomes and formation or disruption of interactions between histones or histones and DNA.⁷⁸

Histone PTMs also manifest the biological function by recruiting effector protein. The cell uses small protein binding modules as a general way to decode protein PTMs.³⁴ Such interaction domains also exist for histone marks and are called reader domains. Reader domains have a binding pocket that accommodates the modified residue thus providing specificity for PTM type and state (methylation vs acetylation and mono- vs di- vs tri-

methylation). Reader domains typically make interactions with flanking residues thereby recognizing sequence context. Example of reader domains (shown in **Figure 11**) include Bromodomains that recognize acetyl lysine,⁷⁹ Chromodomains (CD), Tudor domains and Plant Homeodomains (PHD) which recognize lysine methylation.^{80–82} Methyl lysine binders also interact with residues flanking the modified amino acid to reach high sequence specificity. An example for this is provided by the chromodomains of HP1 and Polycomb (Pc). Although K9 and K27 residues are located in a similar sequence motif, the chromodomain of HP1 is specific for H3K9me3, whereas the chromodomain of Pc is specific for H3K27me3.⁸³ Reader domains are found in chromatin-associated proteins or complexes such as remodelers and modifiers. Reader domains individually interact to histone PTMs with low affinity, usually with a dissociation constant in the micromolar range.⁸⁴ However, chromatin-binding proteins are generally either composed of several binding domains that provide multiple interactions to the nucleosome or they can oligomerize to form higher order complexes such as dimers.⁸⁵

Combinatorial readout of epigenetic marks by multivalent effectors is the basis of the histone code hypothesis.^{55,72} It implies that combinations of histone marks produce distinct biological outputs via a multivalent readout. Additionally, combinations of histone marks are often associated with a particular gene expression state (silenced or transcribed), and thus form a chromatin state. These observations support the histone code hypothesis and suggest that histone PTM patterns define specific chromatin states associated with a certain gene expression status.^{86–88}

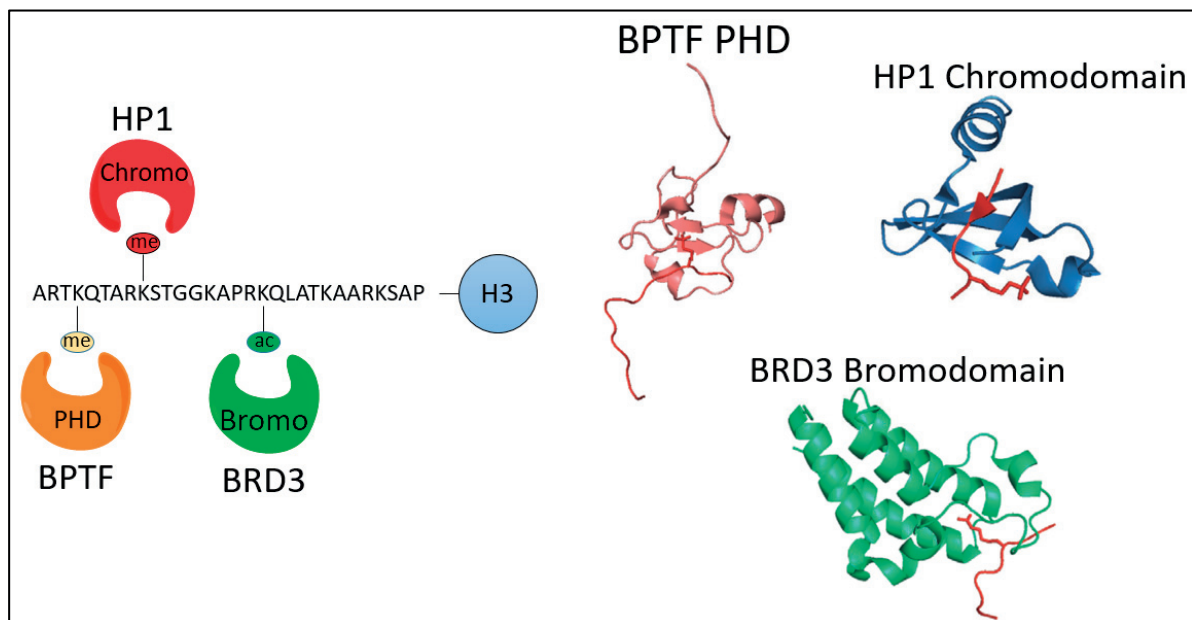


Figure 11. Reader domains as histone PTM binding modules. Three distinct reader domains binding to distinct PTM mark of H3 tail are shown. Ribbon representation of BPTF PHD bound to H3K4me3 peptide (PDB: 2FUU), HP1 chromodomain bound to H3K9me3 peptide (PDB: 1KNE) and BRD3 Bromodomain bound to H3K18ac peptide (PDB: 5HJC).

Parallels in Between Chromatin and Microtubule Cytoskeleton

Like chromatin, microtubules are also polymers composed of highly conserved subunits namely α - and β - tubulin. These assemble into a heterodimer which forms the base repeating unit of microtubules which are integral component of the cytoskeleton. Both α - and β - tubulin have unstructured C-terminal tails which contain several sites for post translational modifications. These tails are known to play an important role in interacting with microtubule associated proteins (MAPs) and these interactions are further affected by tubulin PTM landscape. The apparent parallels between these two cellular polymeric structures and role of PTMs in both cases is very striking. Tubulin PTMs and their role is the theme of the next section of this thesis.

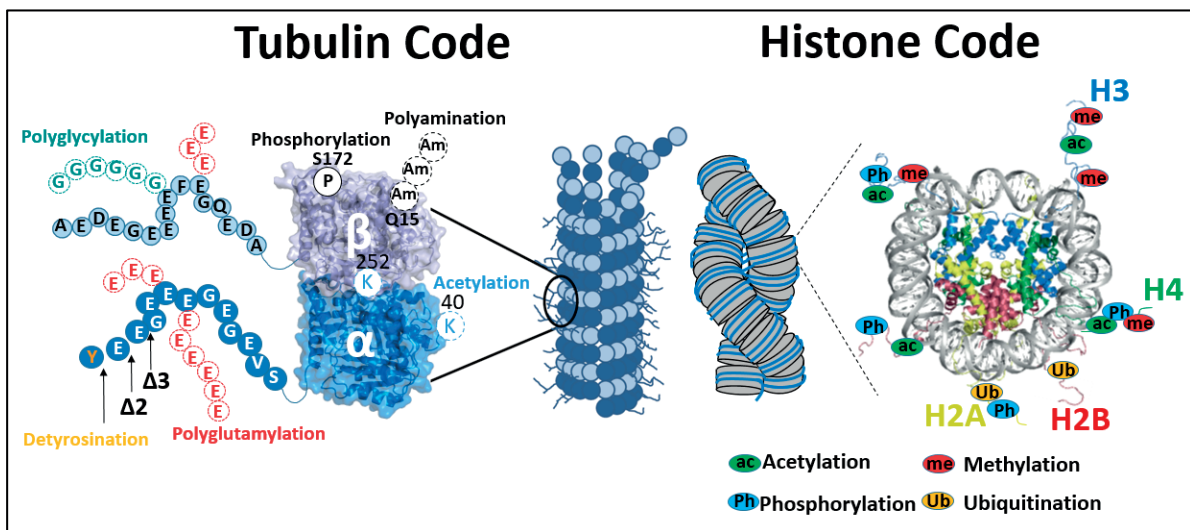


Figure 12. Analogy between Tubulin Code and Histone Code. Both chromatin and microtubule fibers are marked by histone or tubulin PTMs that mostly occur on N- or C-terminal tails respectively which are exposed at the surface of the polymer.

1.3. Microtubule Cytoskeleton

1.3.1. Intracellular Organization and Function of Microtubules

Eukaryotic cells contain three main types of cytoskeletal filaments; actin filaments, intermediate filaments and microtubules as shown in **Figure 13 A**. Cytoskeletal components play an important role in cell shape, motility, intracellular transport and cell division. Microtubules are the largest of all the cytoskeletal components. Microtubules fulfill a vast range of function by forming distinct molecular assemblies, such as mitotic spindles for chromosome segregation during mitosis, centrioles involved in organization of mitotic spindle (**Figure 13 B, E**), axonemes in cilia and flagella (**Figure 13 D**) and specialized microtubules in neurons involved in neuronal transport (**Figure 13 C**).

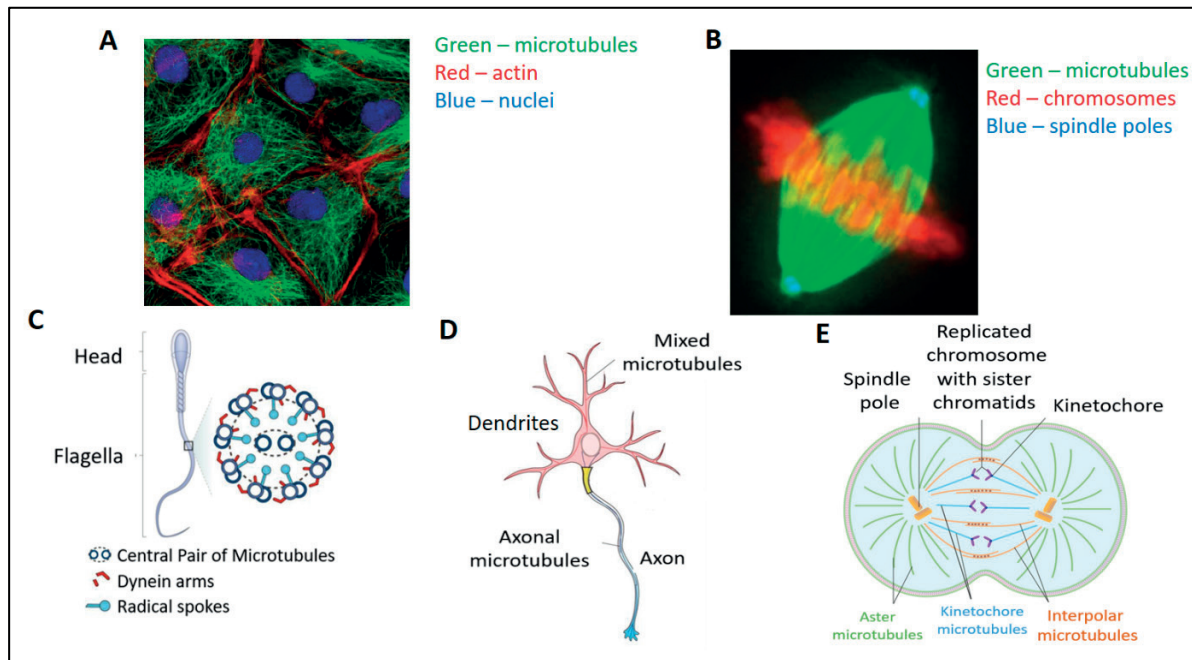


Figure 13. Microtubule Cytoskeleton Organization and Function. A), B) Fluorescence microscopy images showing components of cytoskeleton and mitosis. C) Axonemes ultrastructure in sperm flagella. D) Types of microtubules in neurons. E) Types of microtubules involved in mitosis process with other components. (Figure panel A and B adapted from Risler, 2011)⁹⁰

Microtubules (MTs) are non-covalent cylindrical polymers which are formed by $\alpha\beta$ -tubulin heterodimers as the basic repeating unit. Microtubules are highly dynamic in nature, as they undergo rapid cycles of growth and catastrophe of their ends. The versatile properties of microtubule architecture and distinct assemblies are regulated by a large number of cellular effectors which include motor proteins and microtubule-associated proteins (MAPs).⁹¹ The complexity of such regulation is further increased by the ‘tubulin code’ – which is a combination of differential expression of α - and β -tubulin isoforms and a number of PTMs that impart specific functions.^{25,91,92} Subsequent sections will further discuss different components of tubulin code with focus on tubulin PTMs and challenges involved in understanding the role of tubulin code.

1.3.2. Structural Characteristics of Tubulin Dimer

Microtubules are formed by lateral assembly of protofilaments. Each protofilament is composed of chain of stable $\alpha\beta$ -tubulin dimer which is the structural subunit of microtubules. The protofilaments are arranged in parallel to form the cylindrical microtubule wall. Both α and β tubulin bind a guanine nucleotide; GTP is nonexchangeable when bound to α , whereas GTP and GDP can exchange on β , either within the dimer or at the plus ends of microtubules, but not within the microtubule body. The structure of $\alpha\beta$ -tubulin dimer was solved in 1998 using electron crystallography on two dimensional, crystalline sheets of tubulin that form in the presence of zinc ions. The tubulin sheets used in this study were also stabilized by Taxol.

The structural model (**Figure 14**) of $\alpha\beta$ -tubulin dimer was built into a 3.7 Å density map and contains all the residues within the primary structure except for the highly charged C-termini, which are disordered in the crystals.^{93,94}

The α - and β -tubulin share 40% sequence identity, and are encoded by several genes in most organism which give rise to the several isotypes of α - and β -tubulin. These isotypes are highly similar in terms of sequence. At the structural level, both α - and β -tubulin have basically identical structure, formed by core of two beta sheets of 6 and 4 strands surrounded by 12 α -helices. The monomer structure is very compact and can be divided into three functional domains, the N-terminal domain containing the nucleotide binding region, an intermediate domain containing the Taxol-binding site, and the C-terminal domain, which constitutes the binding surface for motor proteins.⁹³ The secondary structure element and the ribbon diagram of tubulin dimer model is shown in Figure 14. The core of the structure contains two β -sheets of 6 and 4 strands, flanked by 12 α -helices. The N-terminal spans from residues 1 to 205 and forms a Rossmann fold, which is characteristic of nucleotide binding proteins. This domain consists of parallel β -strands alternate with α -helices. Helices H1 and H2 are on one side of the sheet, whereas helices H3, H4 and H5 are on the other. Strand B6 leads to the intermediate domain, spanning from residue 206 to 381, containing a mixed β -sheet and five surrounding helices. The domain starts with helices H6 and H7, followed by a long loop and helix H8 at the longitudinal interface between monomers. The loop connecting B7 and H9 is more ordered in α -tubulin, where it is involved in strong lateral contacts. the loop between B9 and B10 includes an 8-residue insertion in the α - subunit which occludes the site that in β is occupied by Taxol. The C-terminal domain is formed by helices H11 and H12. They are probably involved in the binding of MAPs and motor proteins. The loop connecting H11 and H12 is important for the interaction with the next monomer along the protofilaments.⁹³ The last C-terminal residues of each monomer are missing from the model. This segment of tubulin contains highly acidic residues in both tubulins and form unstructured tails. Most of the sequence differences among tubulin isotypes are seen in these unstructured tubulin tails.

The $\alpha\beta$ -tubulin heterodimer has two nucleotide binding sites. The N-site in α -tubulin which is buried in the intradimer interface is occupied by GTP. Whereas, the GTP at E-site in the β -subunit (forming interdimer interface) is partially exposed in dimer and thus is exchangeable with GDP, but would be buried in microtubules where it becomes non-exchangeable. The GTP bound tubulin dimer has more flexible conformation and thus it can be easily integrated into the microtubule lattice.^{93,95,96}

The structure of tubulin is very similar to that of a bacterial GTP-binding protein FtsZ, which is involved in cell division. FtsZ is evolutionary related to tubulin based on its GTP-binding function and ability to form protofilaments and sheets.⁹⁴

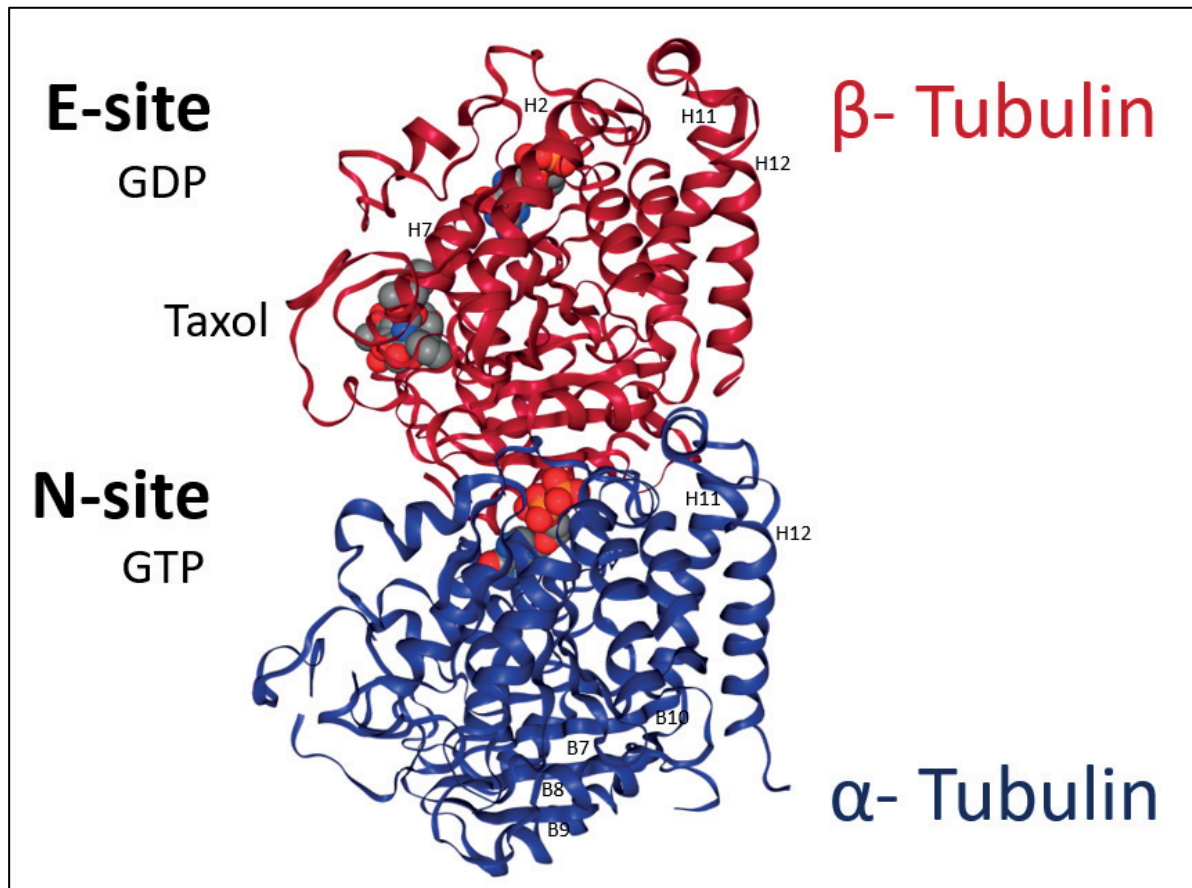


Figure 14. Structure of $\alpha\beta$ -tubulin dimer (PDB 1TUB).⁹³ The tubulin dimer structure with important structural subdomains is shown. Tubulin binds a GTP molecule at the N-site (nonexchangeable) that plays a structural role. The E-site (exchangeable) is occupied by GTP when the heterodimer is in solution. The GTP at the E-site undergoes hydrolysis within the MT lattice. Structure of GDP- bound tubulin derived by electron crystallography in presence of Taxol.

1.3.3. Microtubule Structure and Dynamics

1.3.3.1. Tubulin Folding into Stable Dimer

Tubulin requires the cytosolic chaperonin CCT for its correct folding.⁹⁷ CCT (also known as TRiC) is a hetero-oligomer formed by eight different subunits that are assembled into hexadecamer of two double rings.⁹⁸ Tubulin requires additional cofactors that bind sequentially to α - and β -monomers and lead to the formation of $\alpha\beta$ -tubulin heterodimers as shown in **Figure 15**. The folding intermediates that are released from CCT are captured by cofactor B and E for α -tubulin and cofactor A and D for β -tubulin. Out of these, the α E and β D intermediate then form a temporary pentameric subunit with cofactor C and hydrolysis of GTP in β -tubulin triggers the release of native $\alpha\beta$ -tubulin heterodimers. These folding cofactors play an important role in regulating tubulin ratios. Cofactor A and B are not needed for dimer folding but provide a reservoir of tubulin folding intermediates. The dimeric complexes α E and β D can be formed with a back reaction from tubulin heterodimers and cofactors E and D.⁹⁹

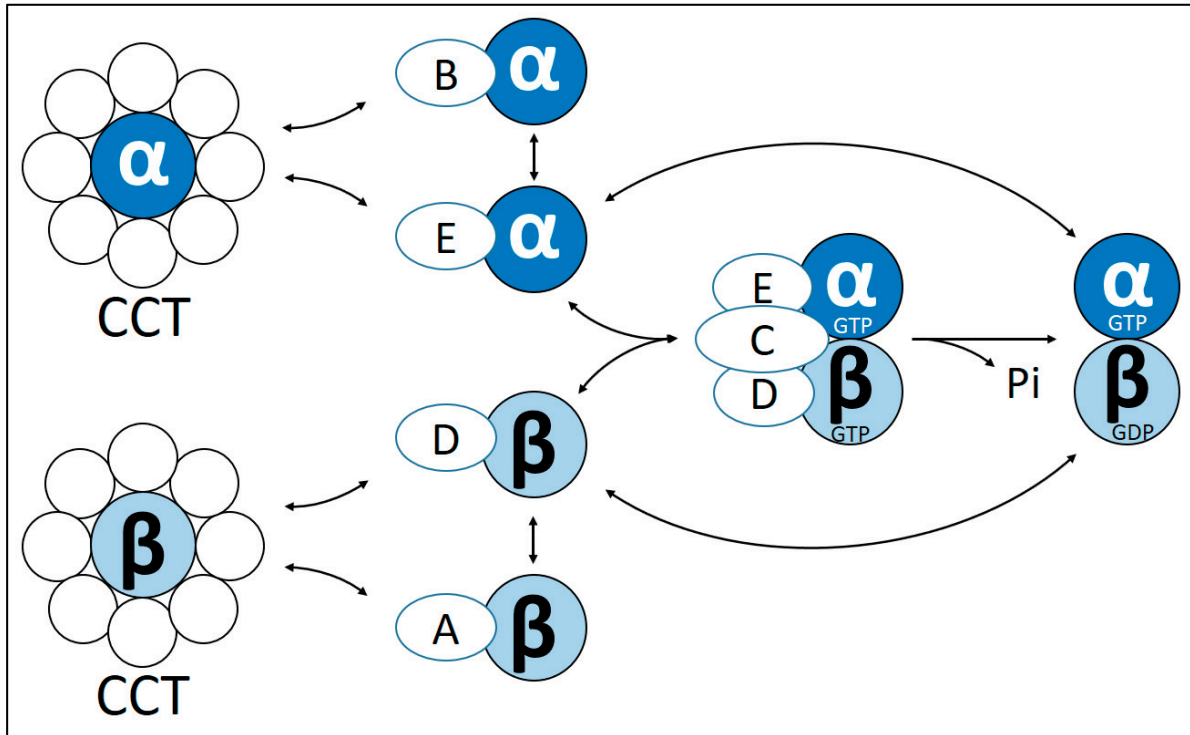


Figure 15. Tubulin Dimer Folding. Schematic diagram for tubulin interaction with chaperonin and other cofactors during folding dimer formation.⁹⁹

1.3.3.2. Microtubule Assembly and Dynamic Instability

One of the key features central to the functionality of microtubules is the ability to reorganize the microtubule network during the cell cycle in response to various environmental cues. Microtubules have a built-in instability which helps in this rapid reorganization process. Tubulin dimers constantly cycle between polymerized microtubules and the soluble dimer form as microtubules constantly switch between polymerizing and depolymerizing state. This behavior is known as dynamic instability of microtubules as shown in **Figure 16 B**.¹⁰⁰ When the concentration of tubulin dimer is sufficiently high (above critical concentration) and there is steady supply of GTP, tubulin will spontaneously nucleate to form microtubules. The microtubules will continue to grow until soluble tubulin concentration decreases to a critical concentration at which the total mass of polymerized tubulin remains steady. Even in this state, individual microtubules can continuously grow and shrink. The availability of GTP determines the constant flux of tubulin dimers between the soluble and polymerized forms.^{101–103}

Based on the structure of the $\alpha\beta$ -tubulin heterodimer^{93,104}, it is observed that the GTP bound to the α -tubulin at the nonexchangeable site (N-site) plays an important structural role and is never hydrolyzed.¹⁰⁴ This GTP binds during assembly of the dimer and remains intact until dimer dissociates or degraded. In contrast, GTP bound to β -tubulin at the exchangeable site (E-site) can be hydrolyzed. This GTP makes the tubulin dimer available for polymerization. Once the dimer is incorporated into the microtubule lattice, the GTP at the E-site can be

hydrolyzed resulting in compaction of the lattice.^{105,106} When microtubules switch from the growing to the shrinking phase, the stress stored in the lattice due to this compaction is released resulting in rapid depolymerization.

During polymerization, tubulin dimers associate longitudinally in a head-to-tail manner to form polar protofilaments. These protofilaments associate laterally to form a left handed helical microtubule polymer which may range from 11-15 protofilaments, though 13 is the most typical in cells. The protofilaments associate with a slight offset with regards to each other and the degree of offset depends on the number of protofilaments in the microtubule. The closure of protofilament sheets into a cylinder usually results in a discontinuity of the helical packing, termed the seam as shown in **Figure 16 A**. This makes the microtubule a pseudo-helical polymer.^{100,106}

The molecular mechanism that regulate the switch between growth phase to shrinking of microtubules are yet not fully understood. Currently, the GTP-cap model is used to explain this dynamic behavior.^{103,107-109} Tubulin dimers are added to the end of growing microtubule in GTP bound state. Once in the lattice, the GTP bound at the E-site is hydrolyzed. The rate of GTP hydrolysis is slower than the rate of addition of tubulin (i.e. growth rate). As the concentration of GTP bound soluble tubulin dimer decreases, the hydrolysis of GTP within the lattice catches up at the tip of the microtubule. At this point, structural changes caused by hydrolysis of GTP cause the switch from polymerization state to depolymerization state.¹⁰⁵ The strain resulting from the hydrolysis of GTP is stored in the lattice until the microtubule depolymerizes, making depolymerization a much more rapid event, and is often termed “catastrophe.” There are some GTP bound dimers left within the lattice marked by blue arrows in Figure 16 B. They are thought to be required for the rescue event and to reverse depolymerization to polymerization state.¹⁰⁰

Understanding microtubule dynamic instability is key to understanding function of microtubules. One of the key function of microtubules is to move cellular structures such as chromosomes, mitotic spindle and other organelles and cargo within the cell. This is done by attaching the ends of microtubule to the target cellular structures. As the microtubules grow and shrink, they exert push or pull to the attached structure, moving them to desired location.¹⁰⁸ To this end, the dynamic instability of microtubules make them well suited to performing various functions within the cell.

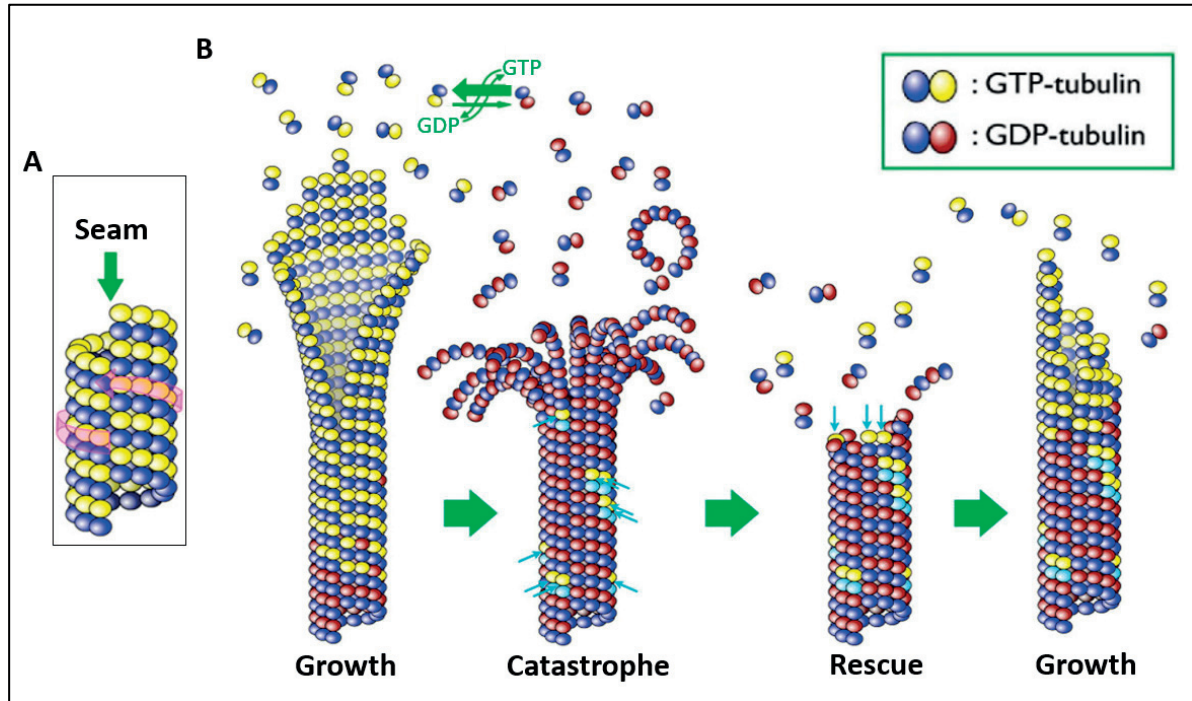


Figure 16. Microtubule Structure and Dynamic Instability. A) The structure of 13 start microtubule helix is shown with the protofilaments offset and the discontinuity at the seam. B) Microtubules cycle between growing and shrinking phases. Adapted from (Horio & Murata, 2014).¹⁰⁰

1.3.4. The Tubulin Code

1.3.4.1. Tubulin Isozyme Diversity and Intrinsically Disordered Tubulin Tails

Microtubules show very little variance throughout the eukaryotic evolutionary landscape. This is also the case with the building blocks of microtubules, the α - and β -tubulin dimers which are highly conserved.¹¹⁰ Nevertheless, both α - and β -tubulins are encoded by multiple genes in most organism giving rise to highly conserved tubulin isoforms.¹¹¹ The $\alpha\beta$ -tubulin heterodimer has a compactly folded 'body' and disordered and negatively charged C-terminal tails. Most isotype sequence variations are seen in these C-terminal tail regions. The tails are also site for a number of PTMs and interactions with various MAPs and motor proteins. Thus subtle changes at these sites can alter biophysical properties of microtubules without affecting its polymerizations interactions.¹¹² This situation is similar to the 'histone code' where most histone PTMs that regulate chromatin function are situated at the N-terminal disordered tail. By this analogy, the tubulin isotype diversity and C-terminal PTMs in combination give rise to the 'tubulin code'.^{25,112–115}

Humans have eight α -tubulin and nine β -tubulin isoforms as shown in the **Figure 17**.¹¹¹ Some of these isoforms are essential for highly specialized cells like sperm, neurons and

platelets. Many tubulin isotypes can coexist in a cell while some are expressed in tissue specific manner. Specialized microtubules such as axonemes and neuronal microtubules are selectively enriched for a specific β -tubulin isotype.²⁵ The roles of α -tubulin isotypes yet remain unclear. Mutations in human tubulin isoforms are associated with various pathological conditions. Mutations in $\beta 6$, a tubulin isoform most abundant in platelets, have been identified in patients with congenital macro thrombocytopenia.¹¹⁶ The $\beta 3$ isotype is exclusively expressed in neurons at the onset of differentiation and in certain tumors of non-neuronal origins such as lymphoma, squamous cell carcinoma and malignant melanoma where it is not found before transformation.¹¹⁷ The α -tubulin isotype $\alpha 4A$ does not feature the characteristic C-terminal tyrosine residue and thus its expression could affect the tyrosination-detyrosination cycle.²⁵

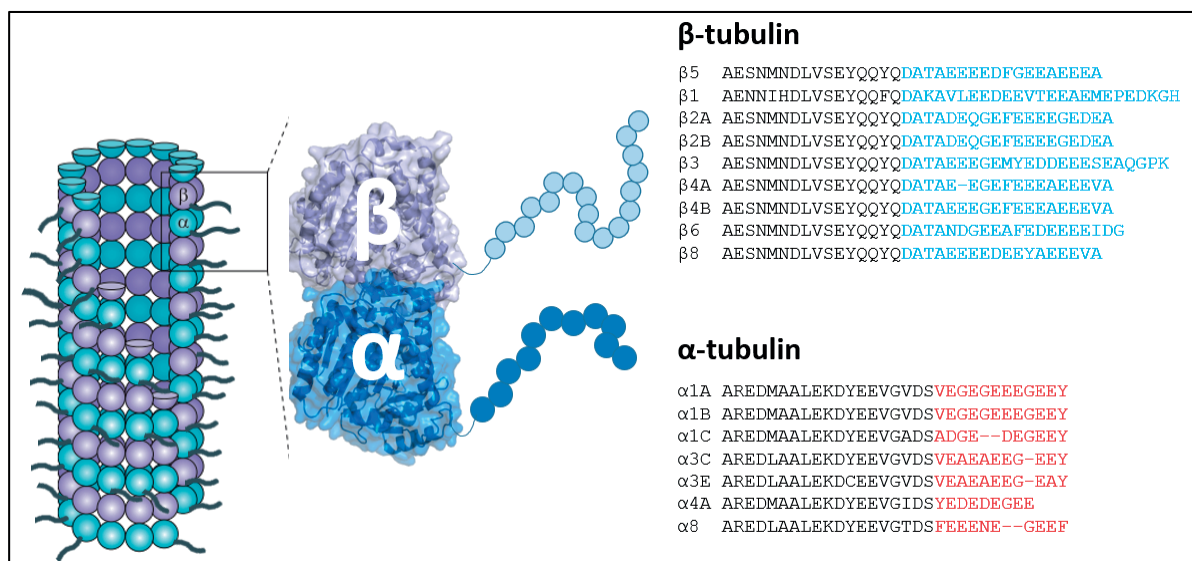


Figure 17. Tubulin Isotype Diversity. View of microtubule surface showing the C-terminal tails which decorate the microtubule shaft (figure partially adapted from (Janke & Bulinski, 2011). The Sequence variation in the C-terminal tails for number of human α - and β -tubulin isotypes is shown. (The tail residues are in blue for β -tubulin and red for α -tubulin). Dimer Structure (PDB 1TUB).

1.3.4.2. Tubulin PTMs

In addition to the isotype diversity, further chemical complexity is generated by a large number of post-translational modifications primarily at the C-terminal tails of tubulin but also some in the structured region. The modifications at the C-terminal tails include phosphorylation¹¹⁸, detyrosination/tyrosination¹¹⁹⁻¹²¹, polyglutamylolation¹²²⁻¹²⁵, polyglycylation¹²⁶ as shown in **Figure 18**. Additionally, tubulin is subjected to post-translational modifications in the structured domain such acetylation^{127,128}, phosphorylation¹²⁹ and polyamination¹³⁰. Although, tubulin modifications have been known for over forty years, it is only in the recent times we are beginning to understand the cellular function and the mechanisms that regulate tubulin PTMs.

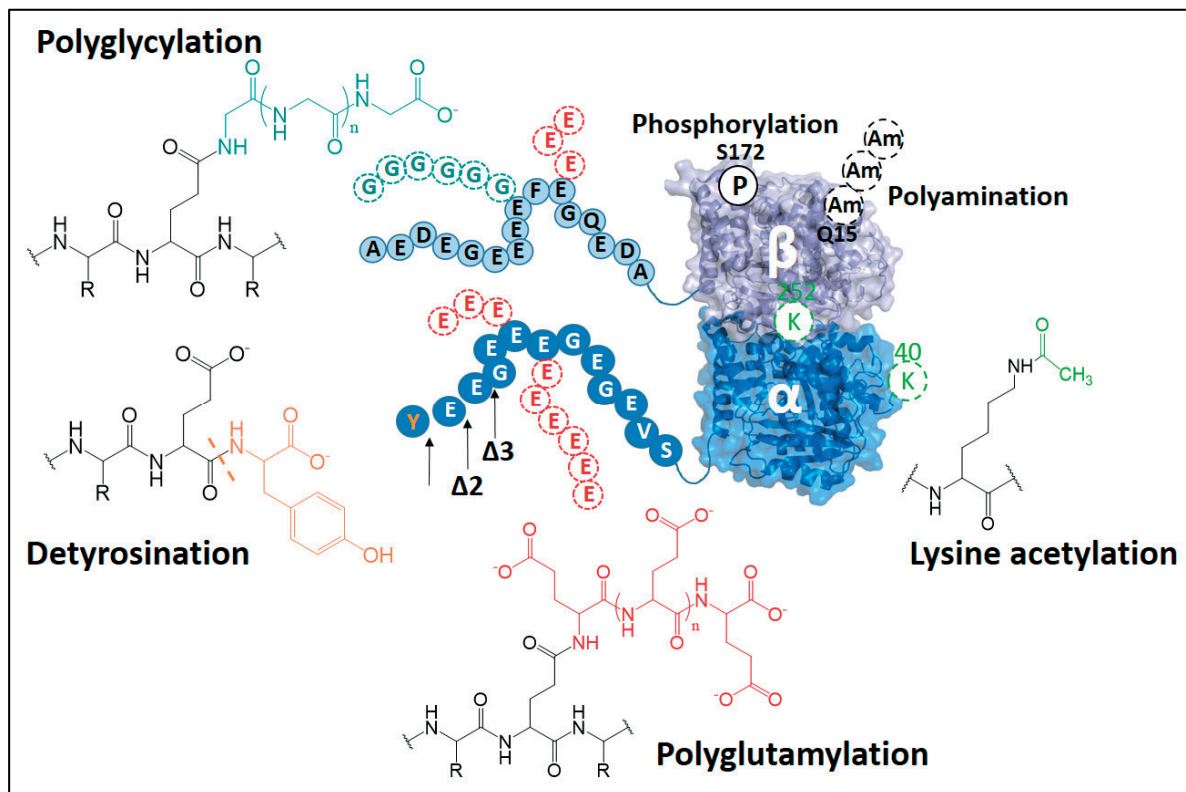


Figure 18. Tubulin diversity generated by post-translational modifications. Crystal structure of tubulin dimer (PDB 1TUB) and chemical structures of known tubulin PTMs and sites are shown in the figure. Detyrosination, $\Delta 2$, $\Delta 3$ tubulin, polyglutamylation and polyglycylation occur at the C-terminal tails whereas acetylation, phosphorylation and polyamination occur in the structured domain.

As shown in the above figure, these modifications are chemically diverse. Most of these modifications are reversible and are evolutionary conserved.⁹¹ Most importantly, at the subcellular level, post-translational modifications mark distinct subpopulations of microtubules. These PTM marks are thought to locally adapt microtubules for specific functions.¹³¹ Interphase microtubules are highly enriched in tyrosination¹³², whereas kinetochore fibers and midbody microtubules are enriched in detyrosination and glutamylation.^{133,134} Microtubules in centrioles, cilia and flagella are heavily glutamylated.^{134–136} Glycylated microtubules are predominantly found in axonemes of cilia and flagella.¹³⁷ In some cases, adjacent microtubules may show very different PTMs as seen in case of axonemes where one of the microtubule is highly glutamylated whereas the adjacent microtubule is not glutamylated but is enriched for tyrosination mark.^{138,139} Tubulin PTMs are also regulated at the developmental level as seen in case of neuronal development which shows increasing levels of glutamylation of both α - and β -tubulin but the β -tubulin glutamylation only occurs at the later stages of neuronal differentiation.¹⁴⁰

Most mammalian α -tubulin isoforms are genetically encoded with a C-terminal tyrosine that can be enzymatic removed and re-added as part of a detyrosination/tyrosination cycle. Tyrosination is catalyzed by TTL enzyme while the detyrosination is catalyzed by Vasohibins^{141,142}. Tubulin tyrosination has been seen as indicator for microtubule stability as tyrosinated microtubules have lifetime of 3-5 min whereas long-lived microtubules are

detyrosinated and have lifetime of 2-16 hr in cultured kidney cells.^{91,132} Detyrosination has been demonstrated to regulate molecular motor (kinesin 1, kinesin 2) interaction with microtubules.²⁵ Following detyrosination, the penultimate glutamate residue of the α -tubulin C-terminal tail can be further removed, producing $\Delta 2$ -tubulin. This modification is irreversible and prevents the tyrosination reaction. The $\Delta 2$ microtubules are highly stable and are seen to be enriched in axonal microtubules.¹⁴³ An additional amino acid may also be removed to produce $\Delta 3$ -tubulin.¹⁴⁴ These processes are catalyzed by CCP family of enzymes.^{145,146}

Reversible acetylation of α -tubulin at lysine-40 (α K40) is unique PTM in a sense that it occurs on the luminal surface of the microtubule.¹⁴⁷ The K40 acetylation is catalyzed by the tubulin acetyl transferase α TAT1 and removed by histone deacetylase 6 (HDAC6) or sirtuin 2 (SIRT2).¹⁴⁸⁻¹⁵¹ The K40 acetylation mark is enriched on microtubules in cilia, basal bodies and stable long-lived microtubules in cytoplasm.¹⁵² Another acetylation site has been reported on β -tubulin Lys-252 in free heterodimers. This modification inhibits the incorporation of tubulin into microtubules, most likely due to its proximity to the nucleotide-binding site on α -tubulin at the interface between the α - and β -tubulin.⁹¹

Polyglutamylation and polyglycylation are not simple ON/OFF signals like tyrosination/detyrosination, phosphorylation or acetylation. The length of the added chain can be varied. The glutamic acid chains can be as long as 20 residues, although most are between one and six.^{91,125,153,154} The first glutamate is added through an isopeptide bond between the γ -carboxyl group of tubulin's encoded glutamate residue and the amino group of the post-translationally added glutamate. The glutamates added beyond the initial branching point are linked through peptide bonds on the α -carboxyl groups.¹²² Glutamylation occurs on neuronal MTs during neuronal differentiation.^{140,155} Microtubules at the centrioles and basal bodies are highly enriched for polyglutamylation and blocking this PTM with anti-glutamylation antibodies results in the disassembly of centrioles.¹⁵⁶ Polyglutamylation is also prominent on microtubules in cilia and flagella (axonemes) and it regulates the beating behavior and integrity of these organelles.^{157,158} Microtubule glycylation is the addition of glycine residues to the side chains of glutamates on α - and β -tubulin C-terminal tails. Multiple Glycine chains of up to 34 residues have been identified.^{131,137} Monoglycylation is widely seen in ciliated tissues. Polyglycylation is important for the stability, length, and function of motile cilia, the formation and maintenance of primary cilia, and control of cell proliferation.⁹¹ Glycylation has so far been exclusively observed on axonemal microtubules and plays an important role in the mechanical stabilization of the axoneme.^{25,159} The glutamylases and glycyllases are members of TTL-like (TTLL) family of enzymes. Each enzyme has distinct preference for substrate. Separate enzymes for initiation and elongation of branched chain with preference for either α - or β -tubulin exists.¹⁵⁹⁻¹⁶² Deglutamylases are members of CCP family which again preferentially act to either shorten the polyglutamate chain or cleave the isopeptide bond at the branching point.^{146,155}

1.3.4.3. Investigating the Tubulin Code

The molecular heterogeneity of microtubules, generated through variable expression of number of tubulin isotypes and a number of tubulin PTMs, has been of great interest of scientific community for more than forty years. Until recently, research on tubulin post-translational modifications was done by using variety of antibodies that could distinguish differently modified microtubules in the cell. The field has progress substantially by the discovery of various enzymes that write or erase tubulin PTMs and many studies have been performed through knock out of these enzymes to study their role on cellular physiology.¹³¹

To understand the dynamic behavior and various physical properties of microtubules and their regulation and modulation by tubulin isotypes and PTMs, quantitative biophysical investigations are required. The most direct of these experiments would be *in vitro* measurements with purified proteins. Most biophysical work on microtubules has been performed with tubulin purified from bovine, ovine, or porcine brains. Brain tubulin is a heterogeneous mixture of different tubulin isotypes and is heavily post-translationally modified making such preparations inept for investigating the functions of tubulin heterogeneity. Thus, pure, recombinant tubulin is essential to dissect the roles of different tubulin isoforms and PTMs. Most attempts to produce recombinant, functional $\alpha\beta$ -tubulin dimers have failed due to absence of extensive folding machinery required. An alternative is to purify tubulin from HeLa cells such as to reduce isotype heterogeneity and has almost no PTMs. This can be modified in cells by expressing modifying enzymes before tubulin purification or *in vitro* modification of purified tubulin.⁹² The attempts made to express mammalian tubulin in the yeast system have not been successful but this problem was partially circumvented by expressing chimeras that consists of yeast tubulin body and mammalian tubulin tails, thus mimicking different tubulin isotypes.¹⁶³ Using this method effects of tubulin detyrosination have been studied.¹⁶⁴

Some of the most interesting modifications are the branched chain polyglutamylation and polyglycylation of microtubules. Adding a chain of negatively charged glutamates will likely have a strong effect on the electrostatic interaction with MAPs and motor protein and affect the polymerization and depolymerization kinetics of the microtubule. On the other hand, adding a long chain of glycines will increase the exclusion volume around the microtubule and act as a barrier for potential binding partners. Recent breakthroughs in purifying tubulin from cells with very low level of modifications and then modifying purified tubulin *in vitro* with homogenous active preparations of tubulin modification enzymes has made biophysical investigations possible for these modifications.^{165,166} Chimeric tubulins purified from yeast carrying a cysteine to which polyglutamate chains can be linked have been used to investigate role of polyglutamylation on mechanical properties on microtubules.¹⁶³ In this system, the glutamate chain is not linked through the glutamate branch that the native glutamylation reaction creates.

The most promising development has been the successful purification of fully functional recombinant tubulin from the baculovirus expression system.¹⁶⁷ Isotopically defined $\alpha\beta$ -tubulin dimers can be obtained using two different purification tags on α - and β -tubulin, respectively. These epitope tags are essential for purifying recombinant tubulin, they could also affect tubulin assembly or microtubule–MAP interactions. Thus, further advancements are needed to make this method traceless by removal of tags. Nonetheless, this method opens the door for a detailed biophysical investigation of tubulin PTMs. This method is central to the project described in this thesis and is further explained in Chapter 4.

1.4. Protein Engineering Tools to Introduce Protein Modifications

Considering the important role protein PTMs play in protein function and diversity, it is crucial to have access to large amounts of homogeneously modified proteins. This would help in investigating the functional role of a particular PTM and design therapeutic molecules or probes that target the PTM. Significant part of research work is devoted to developing methods to homogeneously label or modify proteins. Multiple methods are currently available out of which following ones are described in this section; expansion of genetic code, cysteine targeted conjugation, native chemical ligation (NCL), intein-mediated expressed protein ligation (EPL) and protein-trans splicing (PTS).

1.4.1. Genetic Code Expansion

The method was first described by Schultz's and Chamberlin's group in 1989, which was based on development of orthogonal aminoacyl-tRNA synthetase/tRNA pair.^{168,169} This method involves using the ribosomal machinery to encode for modified or unnatural amino acid (UAA) instead of the amber stop codon UAG.¹⁷⁰ Some bacteria use UAG to encode for pyrrolysine. By evolving the pyrrolysine-tRNA synthetases to recognize and charge CUA-codon tRNA with any UAA, such modifications can be introduced in the target protein recombinantly at one or more sites. Using this technique, acetylated lysine has been directly introduced into histone protein (**Figure 19 A**).^{74,171} The subtle differences between mono-, di-, tri-methyl lysine makes it difficult to evolve tRNA synthetases to directly introduce methyl lysine in the protein using this technique. Thus, precursors of methyl lysine are incorporated using amber suppression method which are subsequently converted to desired methyl lysine state chemically (**Figure 19 B**).^{172,173} The main challenge remains to reach orthogonality with the 20 natural aminoacyl-tRNA synthetase/tRNA pairs, which is one the major limitations of this technique, the lack of modularity. For every UAA of interest, a new pair of aminoacyl-tRNA synthetase/tRNA needs to be developed.

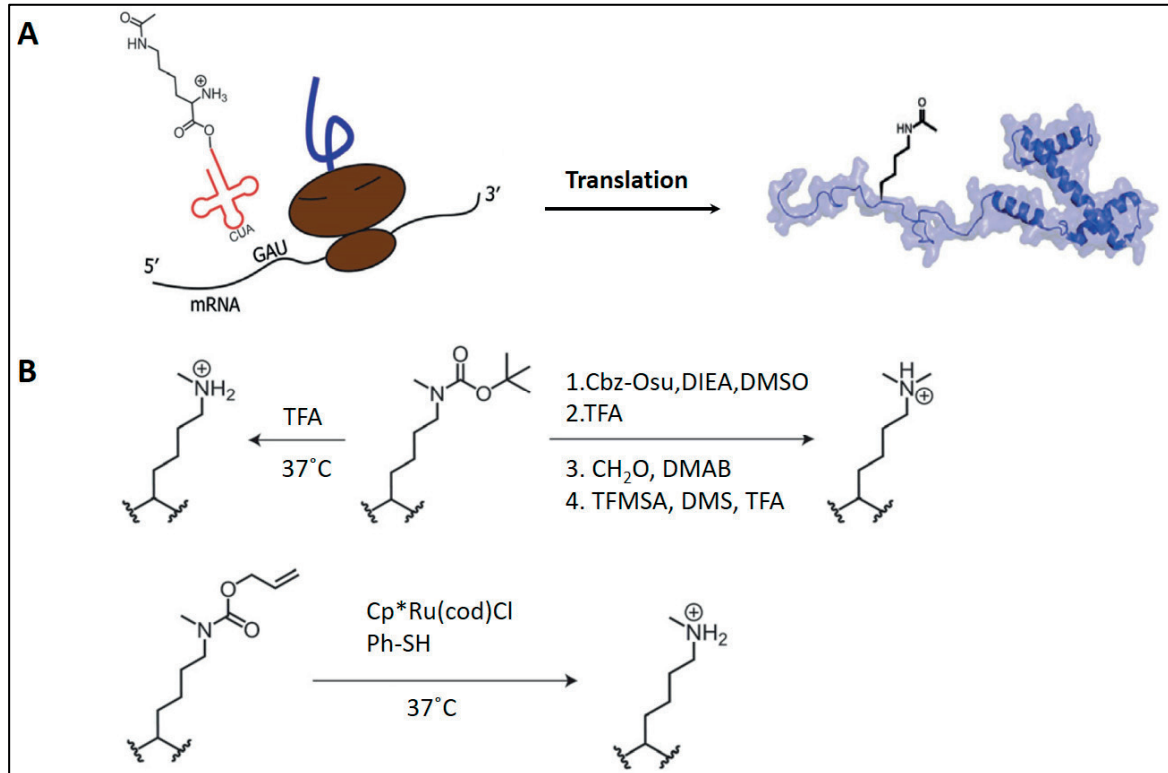


Figure 19. Site specific modifications of proteins by Expansion of Genetic Code. A) Using amber suppression codon, unnatural or modified amino acids (like acetyl lysine) can be recombinantly introduced into histone. B) Precursors can be incorporated into histones recombinantly which can be subsequently converted into mono- or dimethyl lysine.

1.4.2. Cysteine Chemistry

Cysteine provides an important handle to for a site specific reaction to introduce modifications or labels to proteins. Cysteines are found in low abundance in proteins, therefore a site specific mutation can be done to insert single Cysteine. These reactions are not traceless since the Sulphur atom replaces the native γ -carbon atom of the side chain.

Aminoethylating agents are used as analogues for variety of histone PTMs. Methyl lysine analogues are produced by alkylating cysteine residues with an electrophile ethylamine as shown in **Figure 20 A, B and C**.¹⁷⁴ Similarly, ubiquitylated histone H2B was produced using a disulfide link. The disulfide link creates a slightly longer side chain of 2.4 Å (**Figure 20 D**).^{77,175} Cysteine is also used as a handle to label proteins with variety of dyes conjugated with maleimide or iodoacetamide (**Figure 20 E, F**).¹⁷⁶ Reactions with iodoacetamides are slower and less specific. Reactions with maleimides can be potentially reversed with competitive thiols of maleimides or undergo hydrolysis that can lead to subsequent decomposition of protein conjugates.

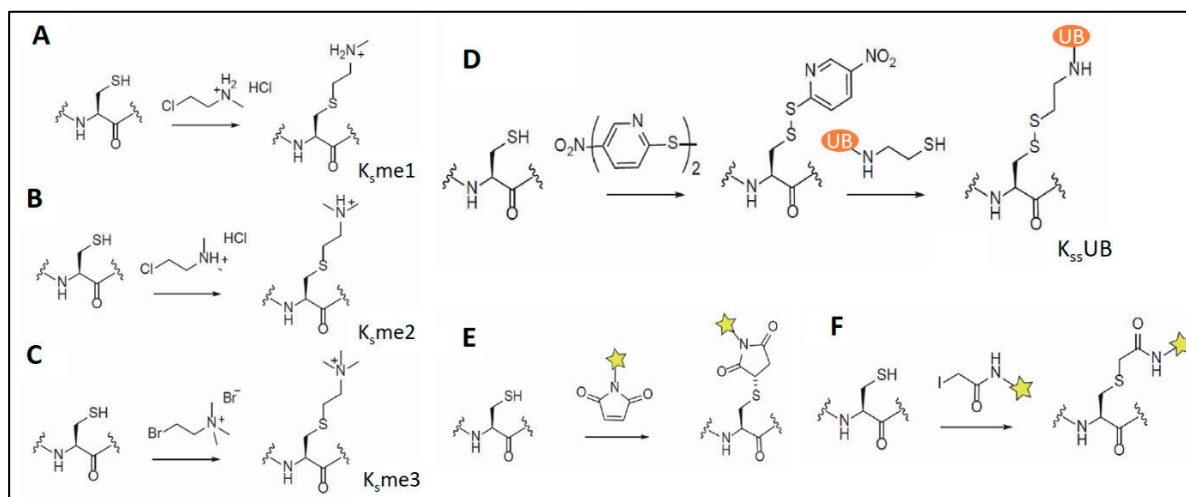


Figure 20. Cysteine Chemistry for Protein Modifications. A-C) Using electrophilic ethylamine, methyl lysine analogues can be generated by alkylation of cysteine residue. D) Disulfide linked ubiquitin. E, F) Proteins can be labeled at cysteine by using maleimide or iodoacetamide linked dyes.

1.4.3. Solid Phase Peptide Synthesis (SPPS)

The chemical synthesis of peptide allows for the introduction of desired PTM in the peptide which can be then ligated with other peptides or proteins. The SPPS was first developed by group of Robert Bruce Merrifield during the 1960s.^{177,178} The methods involves successive addition of amino acids on a solid support. The peptide is synthesized from C-terminal end to N-terminal end (**Figure 21**). An amino acid with α -carbon protected by either *tert*-butoxycarbonyl (Boc) or fluorenylmethoxycarbonyl (Fmoc) is used in order to prevent multiple additions. Prior to coupling to the peptide, the α -carbon is activated. Hexafluorophosphate Benzotriazole Tetra-methyl Uronium (HBTU) is used as coupling reagents. N, N-ethyldiisopropylamine (DIPEA) is used to as base during activation process. After washing unreacted reagents, α -carbon of amino acid is deprotected (by piperidine for Fmoc group or trifluoroacetic acid (TFA) for Boc group). The side chains of amino acids are protected with orthogonal protecting groups which can be removed after cleaving synthesized peptide of the resin (with TFA for Fmoc SPPS or hydrofluoric acid (HF) for Boc SPPS). Handling HF can be hazardous, thus most peptide synthesis is done using Fmoc strategy. A wide variety of modified or unnatural amino acids can be added to peptides using SPPS as it provides complete chemical control. The efficiency each amino acid addition step is $\sim 99\%$ using automated peptide synthesizer, thus limiting the length of peptide synthesized by this method to ~ 50 -60 amino acids as impurities build up and yield decreases during each step. To synthesize full length proteins, various strategies are devised to ligate multiple peptides.

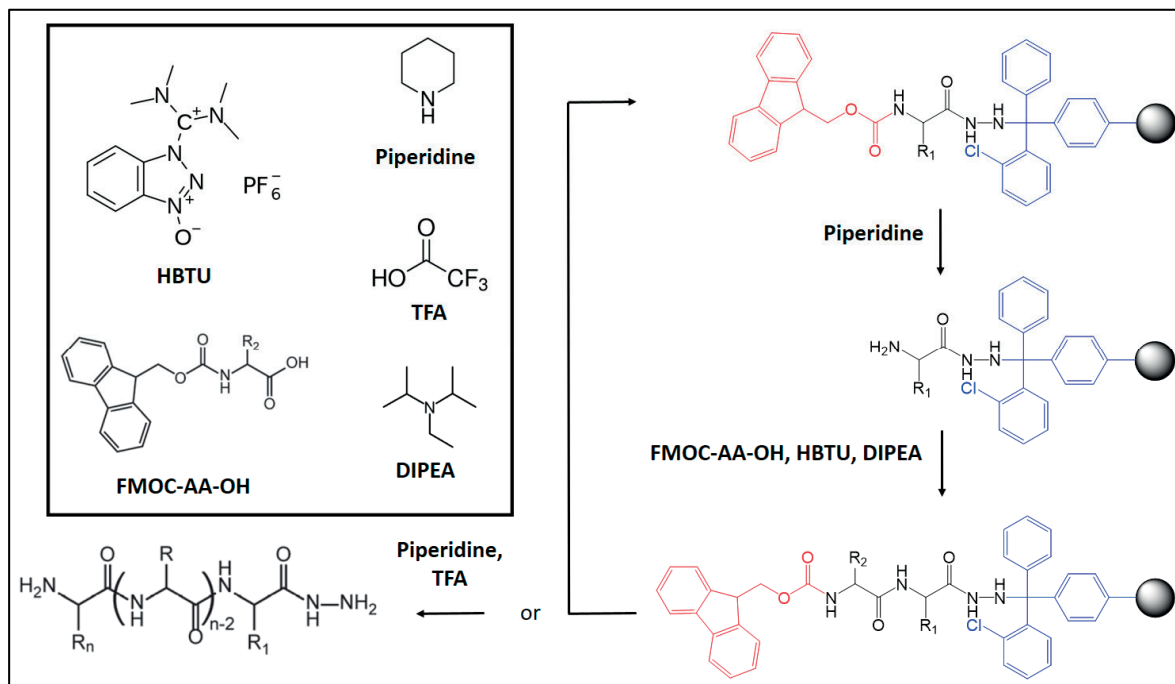


Figure 21. General Fmoc SPPS Strategy. In each cycle has following steps, Fmoc (in red) deprotection by piperidine followed by addition of Fmoc protected pre-activated amino acid. After last amino acid is coupled, the peptide is cleaved from 2-chlorotrityl (in blue) resin resulting in peptide hydrazide. Structures of other reagents in shown in the box.

1.4.4. Native Chemical Ligation (NCL)

A wide range of methods have been developed to join peptide fragments together. These include, Staudinger ligation^{179,180}, KAHA ligation¹⁸¹, salicylaldehyde ester mediated ligation¹⁸² and NCL. Out of all these, NCL has gained the most popularity due to its ability to chemoselectively join two unprotected peptide fragments in aqueous conditions to form a native amide bond.^{183–185} For NCL reaction, the N-terminal peptide fragment carries C-terminal thioester and the C-terminal peptide fragment has N-terminal cysteine. The reaction proceeds with trans-thioesterification (rate limiting step) followed by a virtually irreversible S to N acyl shift, which generates the native peptide bond as shown in **Figure 22**.

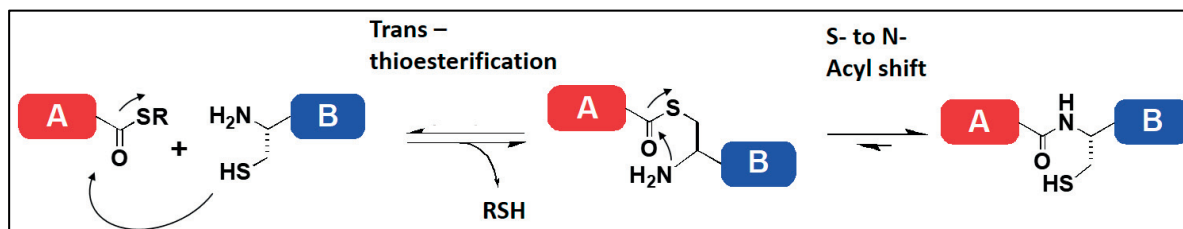


Figure 22. Native Chemical Ligation. The peptide thioester and the cysteine-peptide react to form a thioester-linked intermediate via rate limiting trans-thioesterification step. This undergoes a spontaneous rearrangement, via an intramolecular nucleophilic attack to provide a native amide bond to link two fragments.

This significantly increases the capacity to chemical synthesize peptides from 50-60 amino acids to ~100-120 amino acids. NCL has also been widely used in protein semi-synthesis where a natural protein fragment carrying N-terminal cysteine is ligated to a thioester peptide. Using this technique, the N-terminal tail synthesized by SPPS with desired PTM is ligated to globular domain of histone carrying N-terminal cysteine to produce semi-synthetic histone with specific PTM.^{186,187} Still, major barriers remain which limit the utility of NCL. One, the generation of thioester was limited to Boc SPPS which requires HF cleavage, thus only few groups could prepare required set up and components. Thus significant research efforts have been devoted to produce thioester peptides by Fmoc SPPS. One of the most facile way is to use peptide hydrazides as masked thioesters.¹⁸⁸ Treatment of chlorotrityl resins with hydrazine converts the resin to produce peptide hydrazide instead of peptide acid.¹⁸⁹ Oxidation of peptide hydrazide with sodium nitrite at low pH can generate peptide azide. This can be exchanged with a small molecule thiol compound such as 2-mercaptoethanesulfonic acid sodium (MESNa) or 4-mercaptophenylacetic acid (MPAA) to yield the desired thioester (See Figure 23).^{188,190} The second major barrier with NCL is the requirement of a cysteine at the ligation junction. This combined with the scarcity of cysteine throughout the proteome means that mutations have to be introduced (such as serine to cysteine) to use NCL as semi-synthetic approach. To overcome this barrier, desulfurization of cysteine to alanine via a thiyl radical mechanism has been introduced. The cysteine thiyl radical is generated by the addition of radical generating compound and heat. The thiyl radical then reacts with the alkylphosphine of TCEP to form a phosphoranyl radical. Finally, alanine is produced as shown in **Figure 23** from homolytic cleavage by abstracting the hydrogen of a neighboring cysteine peptide, forming a new radical and the reaction propagates.¹⁹¹ To prevent loss of material during intermediate purification steps, one-pot ligation and desulfurization is preferred. The aryl thiol MPAA is incompatible for this as it acts as radical scavenger. Thus, 2,2,2-trifluoroethanethiol (TFET) is used as efficient thiol additive that enables one-pot ligation/desulfurization. TFET thioester peptides are prone to hydrolysis. Thus Methyl thioglycolate (MTG) is used to circumvent this problem.^{192,193} Desulfurization techniques have enhanced the scope of NCL which have been used to produce traceless modified histone proteins.¹⁸⁷

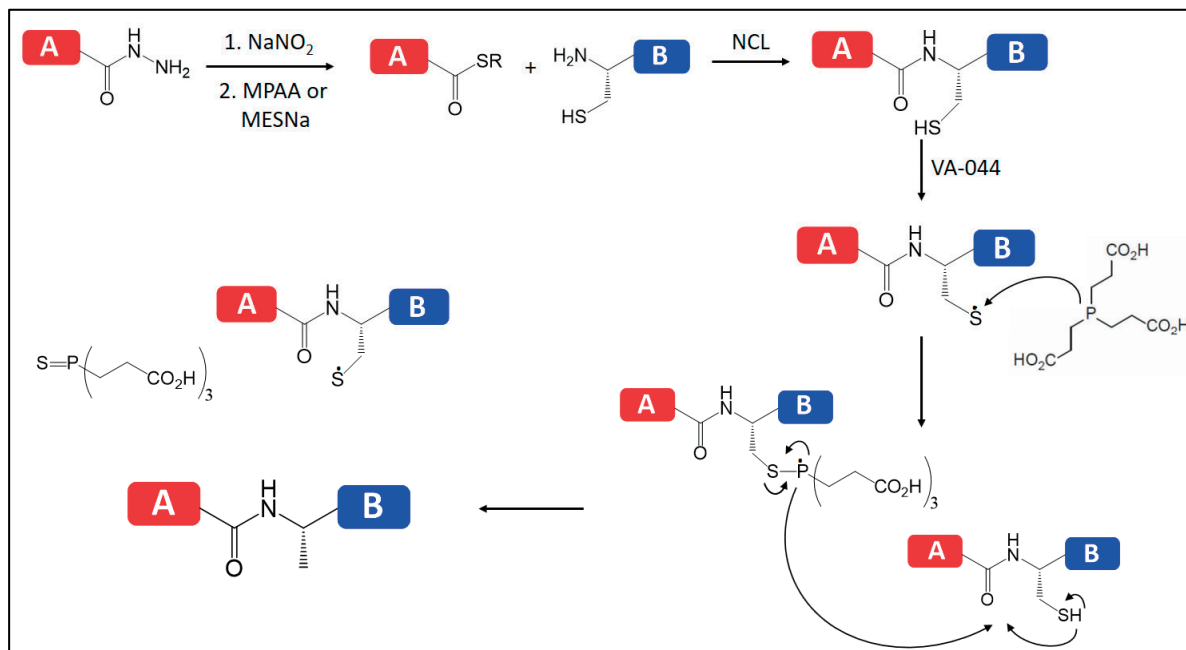


Figure 23. Thioester formation for NCL and desulfurization of cysteine. The C-terminal thioester on the N-fragment is obtained from a hydrazide moiety by reaction with NaNO_2 and a thiol (MPAA or MESNa). Once two peptides are ligated by NCL, the cysteine radical is initiated by the addition of VA-044 and then reacts with TCEP to form a phosphoranyl radical. Homolytic cleavage results in the alanine peptide and the production of a new cysteine radical.

1.4.5. Expressed protein Ligation (EPL)

NCL provides a powerful tool to ligate two synthetic peptides to generate proteins up to ~ 120 amino acids. Most proteins are larger than 120 amino acids. Multiple NCLs can be done to produce larger synthetic proteins but this results in poor yield. EPL, first developed by Muir et al., is much more flexible and modular in its approach to generate semi-synthetic modified proteins.¹⁹⁴ In EPL, either C- or N-terminal fragment can be provided by recombinant protein expression while the other can come from SPPS carrying synthetic modification, thus making the approach very modular for introducing PTMs at N- or C-terminal end of the protein. Fragments carrying N-terminal cysteine can be synthesized by SPPS or by recombinantly expressing a protein with N-terminal cysteine with or out without a tag. To generate C-terminal thioesters, the recombinant protein is fused to mutated intein, where the C-terminal asparagine is mutated to alanine to block the last step in protein splicing. Instead, the cleavage of the intein is obtained by the addition of a small molecule alkyl or aryl thiol to provide the protein thioester which is then ligated to the N-terminal peptide by NCL to provide a native peptide bond as shown in **Figure 24**.^{194,195}

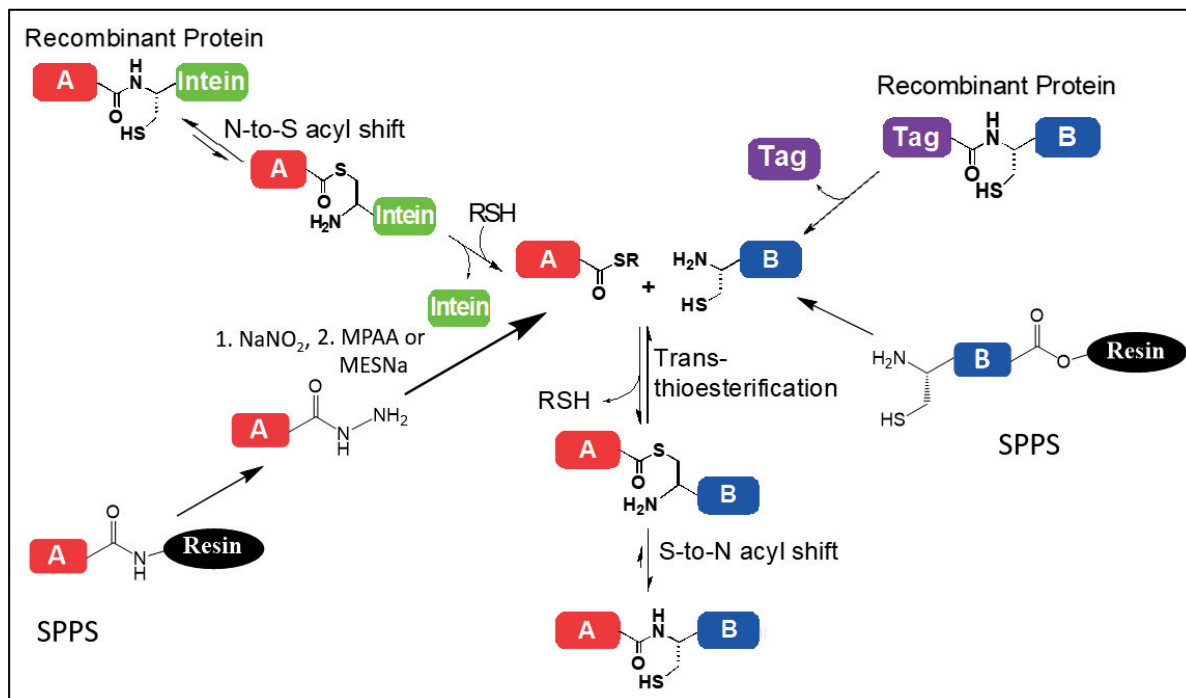


Figure 24. Expressed Protein Ligation. Multiple strategies for incorporating synthetic peptides using EPL. Any one part can be either recombinantly expressed or synthesized by SPPS and then later ligated by same route as NCL (Figure adapted from Chatterjee, 2010).⁶⁸

1.4.6. Split Intein Mediated Protein Trans-Splicing (PTS)

A small fraction (less than 5%) of identified intein genes encode a special variety called split inteins. Unlike their contiguous counterparts which are more common, split inteins are transcribed and translated as two separate polypeptides, N- and C-inteins (Intein^N and Intein^C) fused to one extein each. These two intein fragments can spontaneously and non-covalently assemble into canonical intein structure and lead to protein splicing in trans as shown in **Figure 25**.¹⁹⁶ In 1998, several groups reported that intact inteins can be artificially split and purified separately and when mixed they can reassemble and catalyze protein splicing in trans.^{197–199} This paved way for a variety of protein chemistry applications using artificially split inteins. They were first used for segmental isotopic labeling by Yamazaki et al.¹⁹⁸ An intein from the thermophile *Pyrococcus furiosus* (PI-Pful) was artificially split, and the two halves were fused to two fragments of the C-terminal domain of RNA polymerase α -subunit (α C). The two fusion constructs were separately expressed with or without ¹⁵N isotopic labeling, and full-length α C was generated by PTS. This allowed to simply NMR studies with only resonance corresponding to the labelled α C portion.^{198,200} The early artificial split inteins required unfolding, mixing and refolding steps.

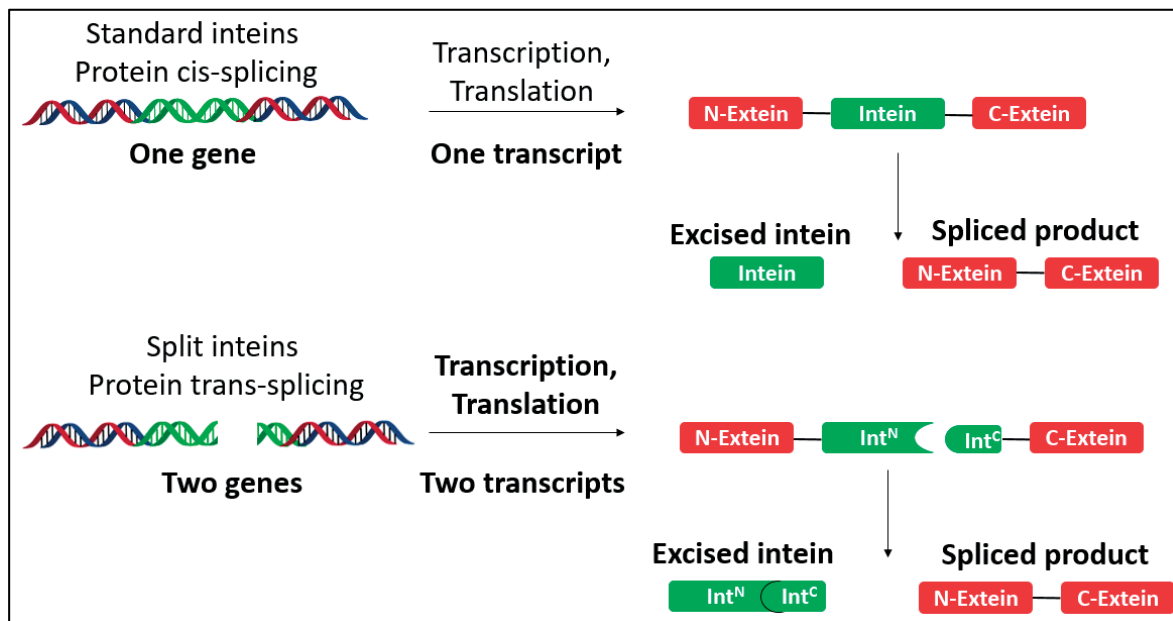


Figure 25. Protein splicing mediated by inteins. Protein cis-splicing splicing by more common contiguous inteins shown in top part and trans-splicing by rare split intein shown in bottom part.

A major breakthrough was achieved when Liu and coworkers reported that catalytic subunit of DNA polymerase III (DnaE) from the cyanobacterium *Synechocystis* sp. strain PCC6803 (Ssp) exists as discontinuous gene containing two fragments fused to sequences homologous to the N- and C-terminal regions of inteins.²⁰¹ These were then co-expressed in *E.coli* and showed that they could assemble to carry out protein trans-splicing, thus establishing existence of split inteins in nature. The naturally occurring split intein (Ssp DnaE) negated the requirement of unfolding and refolding for artificially split intein. Since these inteins can reassemble in native physiological conditions, they can be used for *in vivo* synthetic methods.²⁰² Using the Ssp DnaE split intein, Benkovic group developed a method for *in vivo* split intein-mediated circular ligation of peptides and proteins (SICLOPPS).²⁰³ This allowed for the production of cyclic proteins with enhanced biological properties, and provided a method to generate large libraries of genetically encoded cyclic peptides. Using SICLOPPS, various groups have discovered methyltransferase inhibitors, protease inhibitors and modulators of protein-protein interactions.^{200,204,205} Split intein mediated PTS is also been used for semi-synthesis of proteins. This has two major advantages, one or both the fragments can be recombinantly expressed and thus is can circumvent the size limits of SPPS and the reaction can be carried out in native conditions at low concentrations.²⁰⁰ The discovery of naturally split α subunit of the DNA polymerase III (DnaE) intein from *Nostoc punctiforme* PCC73102 (Npu) further enhanced the applications of PTS for protein semi-synthesis.²⁰⁶ The PTS reaction with Npu split intein is very fast ($t_{1/2}$ of ~ 60 sec), can be carried out at temperatures ranging from 4°C to 37°C and also in presence of 6M urea. The intein has been further evolved into variety of homologues of Intein^N fragment of Npu (Npu^N, Ava^N, Cfa, Mcht) based on the need of faster splicing in various conditions and better recombinant expression characteristics.²⁰⁷ Using Npu split intein mediated PTS, group of Tom W. Muir has developed a synthetic biology method to engineer histones carrying site-specific modifications on cellular chromatin.²⁰⁸

1.4.7. Intein Splicing Kinetics

The most well-characterized family of split inteins, the cyanobacterial DnaE inteins, have great utility for protein engineering applications. Many of these inteins can splice proteins in less than 1 min. Despite this fact, the activity of these inteins is context-dependent: certain peptide sequences surrounding their ligation junction (called local N- and C- exteins) are strongly preferred, while other sequences cause a dramatic reduction in the splicing kinetics and yield.²⁰⁹ Kinetic assays, structural molecular dynamics studies done by group of Tom W. Muir reveal that the motions of catalytic residues are constrained by the second C-extein residue, likely forcing them into an active conformation that leads to rapid protein splicing as shown in the **Figure 26**.

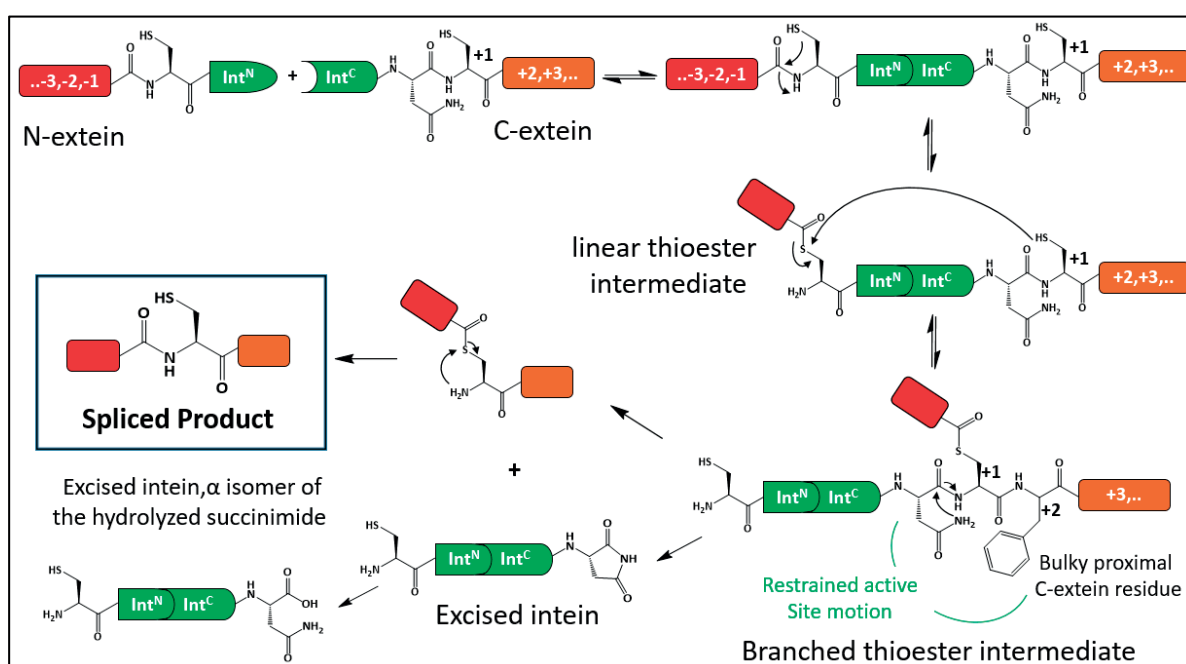


Figure 26. Mechanism of Protein Trans-Splicing. Various intermediate steps and species of trans-splicing reaction are shown in the figure. The +2 extein residue constrains the catalytic site leading to resolution of branched intermediate towards excision of intein and protein splicing.

The context dependent kinetics of split inteins protein trans splicing limit the applicability of them for protein engineering. Significant research has been done to artificially engineer inteins to enhance their tolerance to non-native extein sequence. The engineered split inteins Np_{UGEP} and Cfa_{GEP} offer greater promiscuity for splicing nonnative sequences.^{207,210,211} Most naturally occurring split inteins have a long Intein^N (~120 aa) and short Intein^C (~35 aa). These can be readily used for attaching synthetic moiety to a C-terminal part of protein but not to N-terminal part as Intein^N is out of scope of efficient SPPS. To overcome this limitation, artificial and natural split inteins have been discovered and characterize with shorter Intein^N and longer Intein^C.²¹² Having distinct inteins that could catalyze PTS under unique conditions add enormous value to existing array of protein ligation tools thus significant research efforts are devoted to this by several research groups.

2. Objectives

The nucleosome is the base repeating unit of chromatin fiber is composed of DNA wrapped around core histone proteins. The histones carry N-terminal unstructured tails which are sites for large number of PTMs. Histone PTMs work in synergy to establish defined and dynamic chromatin state. Like chromatin, microtubules are also polymers composed of highly conserved subunits namely α - and β - tubulin which exhibit unstructured C-terminal tails that are exposed to the outside of microtubules (MTs) and are decorated with sites for multiple PTMs. Developing protein engineering tools to generate modified nucleosomes or tubulin dimers to investigate roles of PTMs is the focus of my thesis.

2.1. Analyzing the Effect of Asymmetrically Modified Bivalent Nucleosome on Regulation of PRC2 Enzyme Activity.

In embryonic stem cells (ESCs), two antagonistic histone PTMs, H3K4me3 (an activating mark) – installed by Mixed Lineage Leukemia (MLL) family, and H3K27me3 (a repressive mark) – installed by Polycomb Repressive Complex 2 (PRC2), are seen to coexist on subset of promoters known as bivalent domains which mark key developmental genes. Bivalency is thought to keep these developmental genes in poised state in ESCs. Nucleosomes in bivalent promoter regions were found to be asymmetrically modified for H3K4me3 and h3K27me. To understand the regulatory processes that establish these bivalent domains and effect of nucleosome asymmetry on PRC2 and MLL enzyme activity, nucleosomes carrying chemically defined modifications are needed. In this collaborative project, we aim to establish a chemical method that will provide direct control of assembly of asymmetrically modified nucleosome without any residual purification tags. These nucleosomes are then used to probe the impact of nucleosome asymmetric modification on catalytic activity of PRC2.

2.2. Exploring the Function of Microtubule PTMs by Semi Synthetic Tubulin

Microtubules are made base repeating unit comprising of α - and β -tubulin heterodimer. Both α - and β -tubulin exhibit unstructured C-terminal tails that are decorated with sites for PTMs, such as phosphorylation, detyrosination, poly-glutamylation and poly-glycylation. Tubulin PTMs are thought to modulate MAP interactions, MT stability, activity of motor proteins and contribute to assembly of specialized structures such as centrioles. A detailed investigation of the effects of tubulin PTMs on cellular processes has proven to be difficult task due to absence of uniformly modified and isotypically defined tubulin preparation. Here, based on recombinant expression of tubulin dimer, we aim to develop a strategy to generate semi-synthetic tubulin with chemically defined PTM. To this end, we will use a split-intein mediated PTS approach to ligate a tubulin tail peptide carrying desired PTM or combination of PTMs to recombinantly expressed $\alpha\beta$ -tubulin dimer. This is part of a larger project under NCCR Chemical Biology program to identify, quantify tubulin PTMs, investigate the role of tubulin PTMs on MT dynamics and assembly of centrioles.

3. Analyzing the Effect of Asymmetrically Modified Bivalent Nucleosome on Regulation of PRC2 Enzyme Activity.

3.1. Introduction

3.1.1. Embryonic Stem Cells

Stem cells have three important features that differentiate them from other somatic cells, the ability to differentiate into any cell type (pluripotency), the ability to proliferate indefinitely and the ability to remain in undifferentiated state (self-renewal).²¹³ During the embryonic development, the cell pluripotency progressively decreases.²¹⁴ The zygote undergoes several cell divisions to form the morula embryo(16 cell state). At the stage of morula, all cells are totipotent and can give rise to embryonic and extra embryonic tissue.²¹⁵ Morula further develops into a blastocyst. The outer layer of blastocyst is known as trophectoderm which forms the extra embryonic tissue such as placenta. The inner cell mass (ICM) of blastocyst is composed of pluripotent cells that have the ability to differentiate into any somatic cells or an organism.²¹⁶ The ICM produces three types of germ layers, ectoderm, endoderm and mesoderm. These are composed of multipotent cells that have ability to differentiate into limited types of cells. During the progress of embryonic development, the cell potency further decreases as shown in the **Figure 27**.

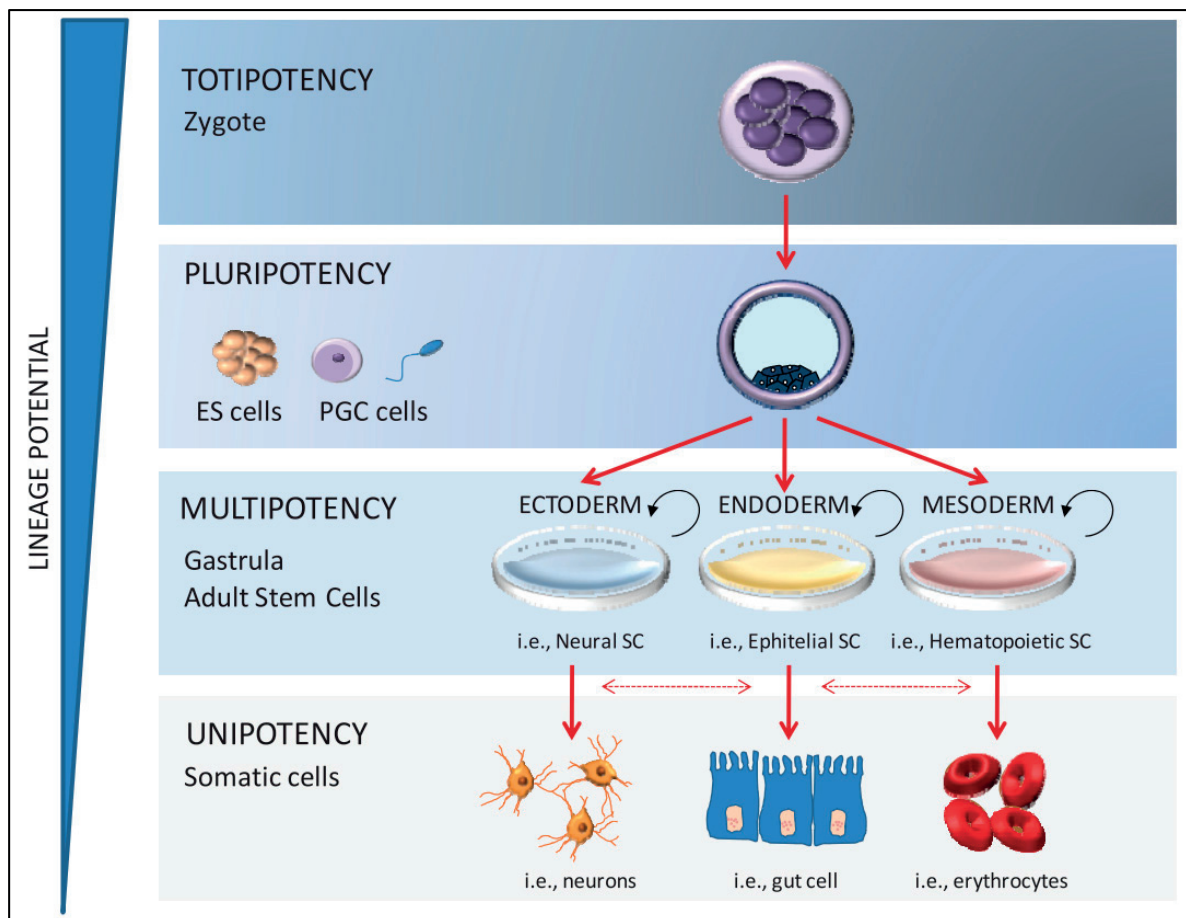


Figure 27. Cell potency during embryonic development. From the totipotent zygote to the somatic cells of an adult organism, cell potency decreases (blue triangle). The figure is adapted from Berdasco & Esteller, 2011.²¹⁴

Embryonic stem cells (ESCs) are pluripotent self-renewing cells derived from inner cell mass of developing blastocyst. ESC pluripotency and self-renewing ability are important aspects of scientific investigation to understand developmental processes. Developmental irregularities cause many diseases and thus these studies are crucial to understand mechanisms in pluripotent cells and develop new therapies for these diseases. ESCs have gene expression patterns that allow them to self-renew and maintain pluripotency but at same time keep the cells poised to differentiate into cell lineages in response to developmental cues. The regulators of gene expression can participate in gene activation, establish a poised state, or contribute to gene silencing. The pluripotency and self-renewing 'state' of ESCs is a combined product of regulatory inputs coming from transcription factors, chromatin state and gene expression and chromatin remodelers acting in a controlled manner.²¹⁷

In mammals, transcription factors form the single largest class of proteins encoded in the genome, ~10% of all protein-coding genes.²¹⁸ Transcription factors can activate gene expression by recruiting transcription apparatus, stimulating release of RNA polymerase II from pause site and by recruiting chromatin regulators to promoter regions that enhance the access of underlying DNA.^{219,220} In ESCs, the pluripotent state is largely governed by the core transcription factors Oct4 (Octamer-binding transcription factor4 or Pou5f1), Sox2 (Sry box-containing gene 2), and Nanog.²¹⁷ Oct4, Sox2 and NANOG (abbreviated O/S/N) are highly expressed in the ICM and knockouts of any of them in mouse embryos lead to development failure.²²¹⁻²²³ The O/S/N core transcription factors function together to positively regulate their own promoters, thus forming interconnected auto regulatory loop. These factors activate expression of genes necessary to maintain ESC state, and also contribute to repression of genes encoding lineage-specific transcription factors. The loss of these core factors leads to induction of wide spectrum of genes encoding lineage specific regulators.²¹⁷

The instructive nature of signaling pathways in governing cell state transitions suggests that signaling effectors must converge on chromatin, affecting changes in chromatin structure and gene expression. Investigations in genome-wide charting of DNA and histone modifying enzymes and their associated modifications in stem cells and differentiating tissues have provided with wealth of information regarding chromatin modification 'maps'. These aid in identifying genetic and epigenetic elements of regulatory importance. This wealth of epigenetic information on the chromatin is further organized spatially, in a three-dimensional manner in relation to the nuclear structure and also characterizes several epigenetic features of ESCs.²²⁴

3.1.2. Bivalent Chromatin and Developmental Genes in Embryonic Stem Cells

Two prominent chromatin modifying systems are the Trithorax group of proteins (TrxG) and the Polycomb group of proteins (PcG). TrxG proteins were discovered as activators of *Hox* genes in *Drosophila*.²²⁵ A subset of TrxG protein complexes catalyze trimethylation of H3 Lys 4 (H3K4me3). This mark is associated with active transcription. In mammals, the enzymes responsible for H3K4me3 modifications are SET1A, SET1B, and mixed lineage leukemia (MLL) proteins 1–4.²²⁶ The PcG proteins were identified as silencers of *Hox* genes in *Drosophila*.²²⁵ In vertebrates and flies, PcG proteins form the multisubunit Polycomb-repressive complexes (PRCs) 1 and 2.²²⁷ PRC2 catalyzes H3K27me3, a pivotal mark in the establishment of repressive chromatin in both early development and adult organisms. Some PRC1 complexes catalyze ubiquitination of H2A Lys 119 (H2AK119ub), whereas others likely act by directly compacting chromatin.²²⁸ Genome-wide mapping studies of chromatin modifications in ES cells show that H3K4me1 and acetylation of H3K27 (H3K27ac) are present within active enhancers and H3K4me3 and H3K27me3 within active and repressed promoters, respectively.²²⁹ These findings also true for all mammalian cells. But certain chromatin features appear more specifically in ES cells. There are a large number of developmental gene promoters that are simultaneously decorated by both activating H3K4me3 and repressive H3K27me3 marks. These patterns of opposing histone marks on the same promoter are referred to as “bivalent” domains, a term first coined by Bernstein et al. in 2006.^{228,230} By exhibiting both active and repressive marks, bivalent genes are kept in poised state, thus they can be rapidly activated upon developmental cues.

Sequential Chromatin immunoprecipitation (ChIP) studies were used to study the pattern of H3K4me3 and H3K27me3 marks in mouse ES cells. These studies showed the transcription start sites (TSS) for select genes were marked by both modifications. Most of these genes encoded for transcription factors important for developmental processes. Also, despite presence of H3K4me3 (activating mark), the bivalent genes were expressed at very low levels. Upon differentiation to particular cell lineage, some genes lost H3k27me3 mark and were expressed while some lost H3K4me3 marked and were silenced.^{228,230} Bivalent domains were also detected in more differentiated states and in cancer cells, although at lower levels.^{228,231–233} by 2013, ~4000 genes in human ESCs and ~3000 genes in mouse ESCs were reported to be bivalent are further numbers are added to the database every year.²³⁴ Bivalent domains were not only seen in cultured ES cells (which are kept in artificially in state of pluripotency) but also in developing mouse embryonic tissue.^{228,235} Bivalent domains were further classified based on occupancy of PRC1 at these domains.²³⁶ Ku et. al. found that PRC2 occupies essentially all bivalent genes, but PRC1 occupies a conserved subset of them. The PRC1 occupied bivalent genes are better at retaining H3K27me3 upon differentiation, compared to the PRC1-negative bivalent genes.²³⁶

3.1.3. Asymmetrically Modified Nucleosomes in Bivalent Chromatin

The discovery of bivalent chromatin raised questions regarding molecular conformation and distribution of H3K4me3 and H3K27me3 marks in the bivalent domains. Are these marks present on the same nucleosomes? Within a single nucleosome, are these marks present on same H3 molecule or opposite H3 molecules? What is the distribution of these marks in bivalent domain? ChIP methods have been widely used to map histone modification genome wide. In this method, fragmented chromatin is immunoprecipitated using antibodies for specific protein or histone modification mark. The underlying DNA is further analyzed by variety of methods such as quantitative PCR (qPCR), microchip or sequencing. Sequential ChIP is used to address limitations of single ChIP which can't answer if the enriched mark originated from same loci but different cell subpopulation or different allele. In Sequential ChIP, chromatin recovered from first ChIP is immunoprecipitated for second mark. Bernstein et al. used sequential ChIP to show the coexistence of H3K4me3 and H3K27me3 at the *Irf2* promoter.²³⁰ Further, using ChIP-seq methods with semisynthetic modified nucleosomes as internal standards, Grzybowski et. al. measured the actual histone modification densities in bivalent chromatin.²³⁷ Their observation revealed that the sum of H3K4me3 and H3K27me3 densities were more than 100% for transcriptionally silent genes, thus suggesting that the both the marks present on single nucleosome. They also found that, among bivalent genes, transcriptionally silent genes carried 70-100% H3K27me3 and 1-60% H3K4me3 at their TSS. Whereas, for transcriptionally active genes, TSS were decorated with nearly 100% H3K4me3 and 10-40% H3K27me3. These findings showed that dominant modification ultimately manifested the transcription state of bivalent chromatin.²³⁷

To accommodate both H3K4me3 and H3K27me3 in a bivalent configuration, promoters can adapt any of the following configuration, the active and repressive histone marks may be present on adjacent nucleosomes, the same nucleosome, or same copy of H3 within a nucleosome as shown in Figure 28. Using immunoaffinity purification of native nucleosomes followed by MS analysis Voigt et. al. showed that both H3K4me3 and H3K27me3 marks are present on same nucleosome and on opposite H3 tails (**Figure 28 B**).²³³ Using fluorescently labelled antibodies and single molecule total internal reflection (TIRF) microscopy approach, Shema et. al. observed that majority (~94%) bivalent nucleosomes contains H3K4Me3 and H3K27me3 marks on opposite H3 tails, but a small portion (~6%) bivalent nucleosomes carry both the marks on same H3 tail.²³⁸

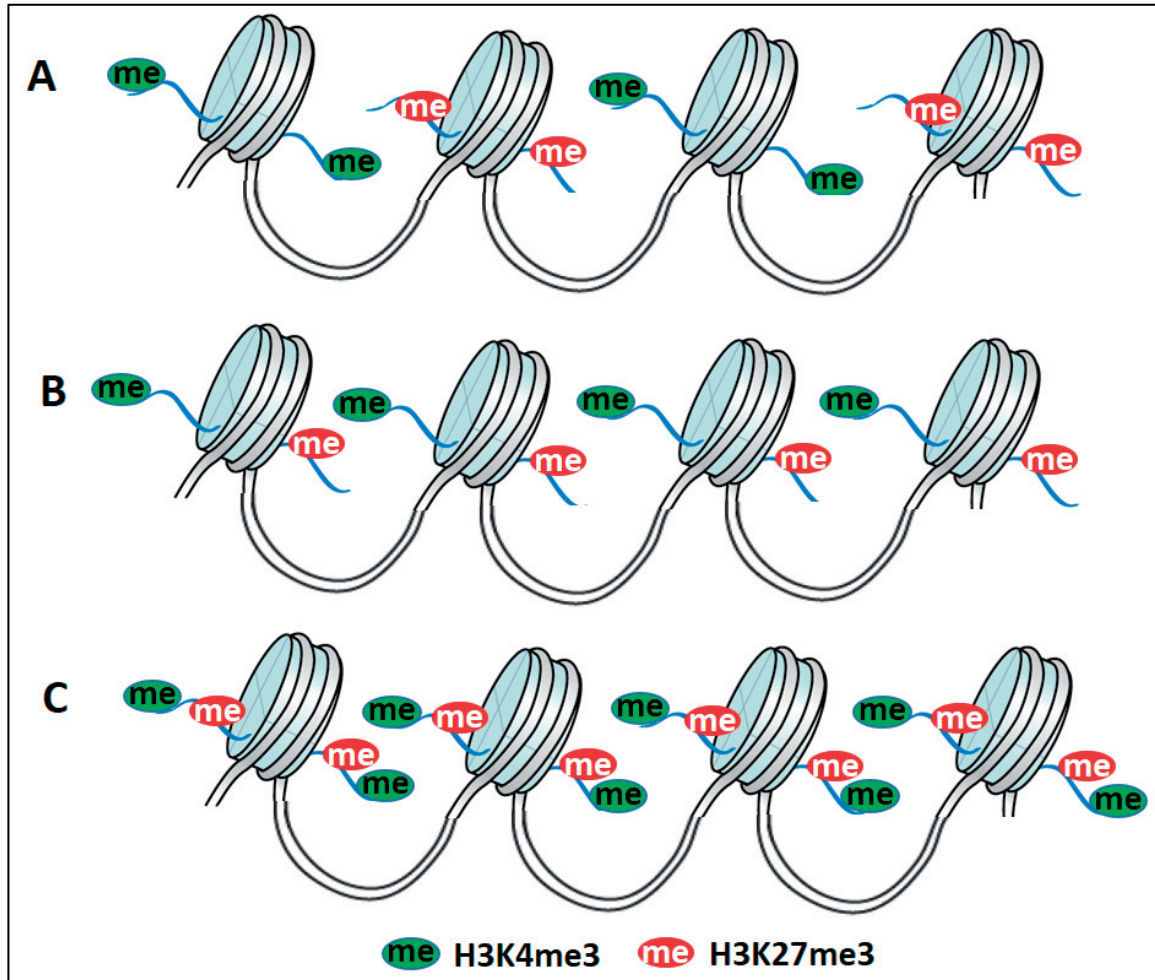


Figure 28. Potential conformations at Bivalent Domains. As illustrated in A) H3K4me and H3K27me3 may occupy neighboring nucleosomes in vicinity of transcription start site (TSS). Alternatively, B) both marks may feature on same nucleosome in asymmetric manner (differentially modified H3 copies) or C) in symmetric manner with H3 molecules carrying both modifications simultaneously.

3.1.4. Polycomb and Trithorax Complexes

Polycomb Repressive Complex

PRC2 catalyzes mono-, di-, tri- methylation of H3K27 residue. (H3K27me1, H3K27me2, H3K27me3).^{239,240} H3K27me1 mark is localized at active gene bodies, H3K27me2 is localized at intergenic regions, whereas H3K27me3 is associated with gene silencing.^{227,241–243} EZH2 (or its homolog EZH1) is the H3K27- specific methyltransferase unit contain SET domain. EZH2 along with EED, RBAP46/48 and SUZ12 forms the core of od PRC2 complex.^{225,227,244} The WD40 repeat protein EED is important to bind to repressive marks including H3K27me3 thus leading to allosteric activation of PRC2. This finding suggests a positive feedback loop for PRC2 activity and local spreading of H3K27me3 mark by PRC2.^{227,245} SUZ12 and RBAP46/48 together form the minimal unit for nucleosome binding and are important for stability of the complex.^{246–248} Apart from the core complex, non-core proteins such as JARID2, AEBP2 and PCL are associated

with PRC2 activity regulation and targeting.²⁴⁴ JARID2 is important for enzymatic regulation and complex recruitment to target genes.^{249,250} AEBP2 is a zinc finger protein and is involved in complex recruitment and stimulates PRC2 activity, whereas PCL protein modulates PRC2 targeting.^{246,251–253}

PRC1 complex is composed of RING1A/B core, a H2AK119-specific E3 ubiquitin ligase, and other non-core proteins. H2K119Ub mark is associated with transcription repression and the RING1A/B is known to repress developmental regulators in mouse ESCs.^{254,255} Depending on subunit composition of non-core protein, PRC1 complex is classified as canonical (containing PCGF4/2, CBX2/4/6/7/8, PHC1/2/3 and RING1A/B) and non-canonical (containing RYBP, KDM2B, PCGF1 and RING1A/B).²⁴⁴ KDM2B is thought to recruit non-canonical PRC1 to unmethylated CpG islands.^{256,257} In canonical PRC1, protein PCGF4 enhances ubiquitin ligase activity and PHC proteins promote PRC1 clustering.^{258,259} In *Drosophila* PRC1, Pc protein, equivalent to mammalian CBX protein, recruits the canonical complex by binding to H3K27me3 through its chromodomain. The human CBX proteins have lower affinity for H3K27me3 compared to *Drosophila* Pc protein, thus such recruitment in human is questionable.²⁶⁰ Also, canonical PRC1 is recently being observed to carry out H2AK199Ub independent of PRC2.^{261,262} AEBP2 and JARID2 in PRC2 complex are seen to be binding to H2AK199Ub installed by non-canonical PRC1, thus recruiting PRC2 to install H3K27me3 mark.^{263–265}

MLL Complexes

The Trithorax family of protein includes chromatin modifiers and remodelers. This group encompasses the NURF and SWI/SNF complexes as well as the COMPASS family.²⁶⁶ In yeast, SET1 is the TrxG protein which installs H3K4 methylation mark. SET1 is found within a multi protein complex called COMPASS (complex of proteins associated with SET1). In *Drosophila*, the COMPASS family consists of 3 multimeric complexes, SET1, Trithorax (Trx) and Trithorax-related (Trr). In mammals, this family contains 6 members called MLL complexes.²⁴⁴ The main catalytic subunit of mammalian MLL complexes comprises of SET-domain containing protein (SET1A/B or MLL1/2/3/4), which has H3K4-specific methyltransferase activity. H3K4me2/me3 are mainly associated with TSS whereas H3K4me1 is associated with enhancers. All three states are involved in transcription activation.^{267–269} The MLL complexes have non-redundant roles. SET1A/B is thought to be default methyltransferase installing di- and tri- methyl marks at active promoters.²⁷⁰ MLL1/2 is more specific with *Hox* genes, MLL3/4 installs H3K4me1 mark at poised enhancers.^{270,271}

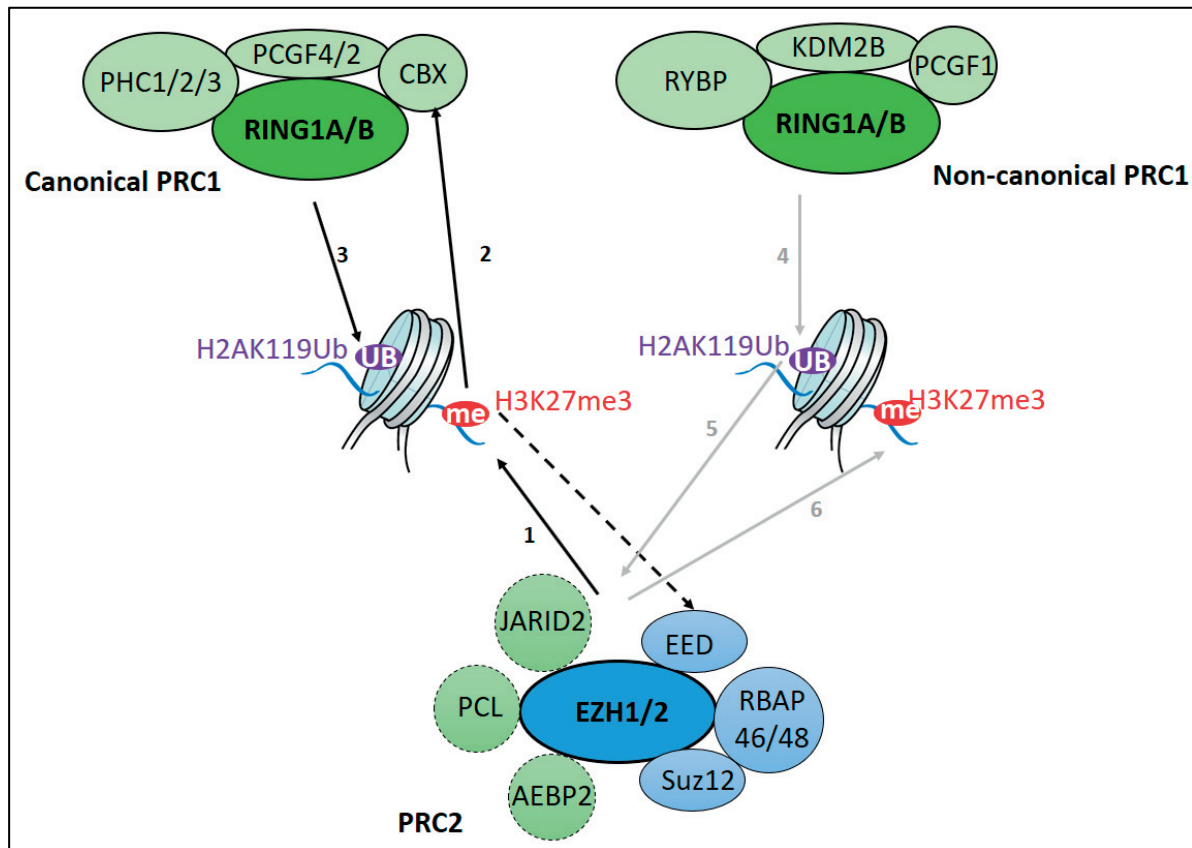


Figure 29. Subunit Composition and interplay of PRC1 and PRC2 complexes. The classical model involves methylation of H3K27 by PRC2 (Black arrow 1), local spreading by EED-mediated recruitment of PRC2 to H3K27me3 nucleosomes (black dotted arrow), canonical PRC1 recruitment via CBX proteins, which bind to H3K27me3 (black arrow 2) and subsequent H2AK119 ubiquitination by canonical PRC1 RING1A/B (black arrow 3.). The alternative model involves H2AK119 ubiquitination by non-canonical PRC1 RING1A/B (grey arrow 4), recruitment of PRC2 to H2AK119Ub nucleosomes (grey arrow 5) and subsequent H3K27 methylation by PRC2 (grey arrow 6).

3.1.5. Installation, Maintenance and Regulation of Bivalent Domains

The discovery of bivalent domains in ESCs^{230,272} suggested an elegant concept that bivalent domains keep genes in poised state maintaining low expression levels in ES cells, and also allowing for rapid activation or stable silencing upon differentiation. This was supported by studies showing resolution of bivalent genes and associated changes in expression levels upon differentiation.^{273,274} Yet, questions remained about how bivalency is installed by PcG and TrxG complexes at the promoters of developmental genes? How are the bivalent domains maintained in ESCs and resolved upon differentiation? How does asymmetric nature of nucleosome modifications present in bivalent domains influence the enzymatic activity and binding of chromatin remodelers?

To directly probe the role of bivalency in poising gene expression, it would be needed to specifically ablate PcG and TrxG protein targeting to bivalent loci without affecting the

expression of other genomic features normally bound by these complexes.²²⁸ With genomic editing tools, such studies are now possible. Initial observations from PcG and TrxG knockouts provide key information about these complexes in regulating bivalent domains. PcG mutant ES cells show upregulation of developmental genes. EED depletion in ES cells show premature expression of bivalent genes.^{272,275} Similarly Suz12 depletion show higher expression of lineage specific genes.^{248,276} Yet, cell viability and self-renewal are not compromised in PRC2-deficient ES cells.²⁷⁶ This can be explained by PRC1 mediated compensatory effects and absence of TFs that could robustly activate affected genes in undifferentiated cells. Simultaneous depletion of RING1B and EED in ES cells provokes an even stronger inclination toward differentiation.²⁷⁷ Based on the observations from these knockout models, it can be demonstrated that PcG complexes, through control of bivalent target genes encoding developmental factors, are vital for proper differentiation.²²⁸

Histone modifications, introduced by PcG and TrxG play an important role in maintenance and regulation of bivalency. Many proteins, binding to H3K4me3 serve as effectors and are associated with active transcription. The PHD finger of the TAF3, subunit of TFIID, recognizes H3K4me3²⁷⁸. TAF3–H3K4me3 interaction directly contributes to preinitiation complex formation in a reconstituted transcription system and to TFIID recruitment in vivo.²⁷⁹ H3K4me3 interferes with the recruitment of the de novo DNA methyltransferases DNMT3A and DNMT3B.^{228,280} Avoiding of DNA methylation is requirement for bivalent genes, as they are required to retain plasticity for subsequent activation or repression.²²⁸ In the case of PRC2, the deposition of H3K27me3 prevents subsequent deposition of H3K36me3.^{228,233,281} H3K27me3 further serves to recruit PRC1–CBX complexes, consolidating repressed gene state.²²⁸

Bivalency represents a dynamic equilibrium between activation and repression. In the bivalent state, activating stimuli and repressive complexes balance each other in a metastable equilibrium. During ES cell differentiation, a significant share of bivalent loci undergoes either activation or silencing. Activating TFs, H3K27 demethylases and H2A-deubiquitinating enzymes (DUBs) shift the equilibrium toward activation, displacing repressive factors and converting bivalent loci to active ones. Both UTX and JMJD3 are capable of demethylating H3K27me3, and both proteins are required for proper differentiation.²⁸² Removal of H3K27me3 makes bivalent promoters more amenable for transcription, in part due to the dissociation of CBX-containing PRC1 complexes and reduced compaction.²²⁸ Moreover, removal of H3K27me3, will destabilize binding of PRC2 itself, thus initiating a negative feedback loop further strengthened by the H3K4me3 when transcription commences. During differentiation, a sizable number of bivalent domains also undergo silencing. In this case, removal of the activating stimuli that are present at low levels at bivalent domains shifts genes to a repressed state. This requires H3K4 demethylase activity (KDM5). H3K9 methylation and DNA methylation may further consolidate the repressed state.²²⁸ The balance of these numerous factors as shown in the **Figure 30** determines ultimately the fate of genes marked by bivalent promoters in ES cells.

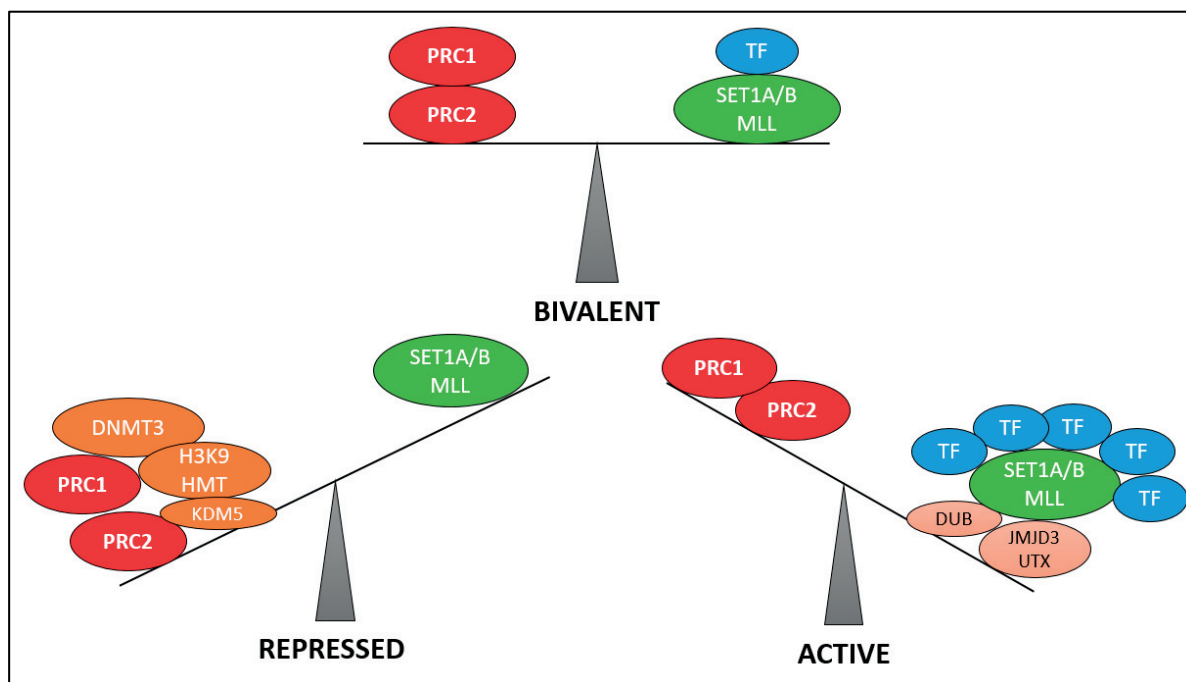


Figure 30. Factors Involved in Resolution of Bivalent Domains during Differentiation. In Bivalent state, activating stimuli and repressive complex counter balance each other establishing a metastable equilibrium. Upon differentiation, activating TF, H3K27 demethylases and H2A deubiquitinating enzymes drive equilibrium towards activation and displace repressive complex. Conversely, removal of activating stimuli, H3K4 demethylases and DNA methylation push the equilibrium towards repression.

3.2. Aims

Nucleosomes are the base repeating units of chromatin and are composed of two copies each of H2A, H2B, H3 and H4 histones assembled in an octamer, around which 147bp of DNA is wrapped. Nucleosomes display two pseudo symmetrical halves each possessing one copy of each histone.^{40,55} All the histones are post translationally modified and these modifications influence chromatin regulation and signaling. Individual copy of each histone can be differentially modified thus imparting PTM-based asymmetry to the nucleosome.^{233,283} In embryonic stem cells (ESCs), two antagonistic histone PTMs, H3K4me3 (an activating mark) – installed by Mixed Lineage Leukemia (MLL) family, and H3K27me3 (a repressive mark) – installed by Polycomb Repressive Complex 2 (PRC2), are seen to coexist on subset of promoters known as bivalent domains.²²⁵ In normal somatic cells, these two marks are seen in distinct chromatin regions. Nucleosomes in bivalent promoter regions were found to be asymmetrically modified for H3K4me3 and h3K27me (trans-bivalent).²³³ It is not yet fully understood how these asymmetric patterns are established and maintained in ESCs. To understand the underlying biochemical process that govern the chromatin asymmetry, nucleosomes carrying chemically defined modifications are needed. Existing methods are based on statistical histone octamer assembly from histone mixtures including modified

histones. Subsequently, asymmetrically modified species are enriched by sequential affinity pulldown steps with two orthogonal peptide tags.^{233,284,285} These methods were limited to assembling nucleosomes containing single modified histones and also suffer from prohibitively low yields. Also, the remaining affinity peptide tags cast might interfere with biological and enzymatic assays on these nucleosomes. Nonetheless, using tagged and affinity enriched nucleosomes, it was found that PRC2 enzymatic activity is inhibited by H3K4me3. On the contrary, PRC2 activity is stimulated by pre-existing H3K27me3 mark.²³³

In this collaborative project, we aim to establish a chemical method that will provide direct control of assembly of asymmetrically modified nucleosome without any residual purification tags. These nucleosomes are then used to probe the impact of nucleosome asymmetric modification on catalytic activity of PRC2. The project has two parts:

3.2.1. Traceless Synthesis of Asymmetrically Modified Nucleosomes

(Done by Dr. Carolin C. Lechner)

In this part, Dr. Carolin Lechner developed a strategy to synthesize chemically pure asymmetrically modified nucleosomes carrying defined combinations of PTMs. The method developed is modular and traceless. Using this technique, she synthesized a library of asymmetrically modified histone octamers with all possible combinations with regards to H3K4me and H3K27me3 marks.

3.2.2. Understanding the Regulation of PRC2 Activity by Asymmetric Histone Modification

When I joined this project, I assisted Dr. Carolin Lechner during assembly and analysis of modified octamers library. Using the library of asymmetrically modified histone octamers, I assembled a library of nucleosomes. This library is then used to probe the catalytic activity of recombinant PRC2 complex using series of methyltransferase assays. Based on these observations, I aim to understand how the asymmetric nucleosome modification impact PRC2 activity thus gaining insights into mechanistic ability of PRC2 to modify and maintain bivalent chromatin.

3.3. Results and Discussion

This chapter is based on following research article of which I am one of the authors as highlighted below.

Lechner CC., Agashe N., Fierz B., Traceless synthesis of asymmetrically modified bivalent nucleosomes, *Angew Chem Int Ed Engl* 2016, 55(8):2903-6

Contribution by other members are indicated in the text below at appropriate places as well in the figure legend.

3.3.1. Synthetic Strategy

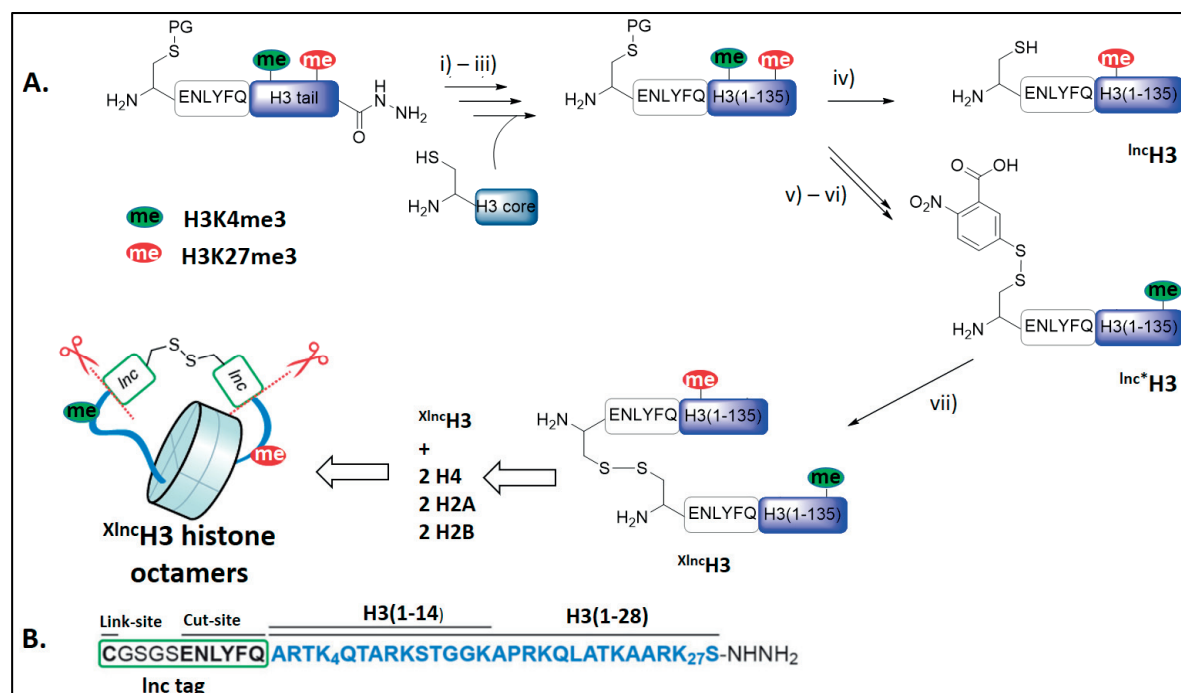


Figure 31. Scheme for Traceless Chemical Synthesis of Asymmetrically Modified Octamers. A) General synthetic route for the preparation of asymmetrically modified histone octamers. i) Conversion of Inc-peptide hydrazide into the TFET thioester, ii) ligation of the Inc-peptide thioester to truncated H3 protein, iii) in situ desulfurization, iv) removal of AcM protecting group (PG) at the N-terminal cysteine or v, vi) AcM removal and in situ modification of cysteine with DTNB, vii) formation of the hetero-disulfide. B) Design of Inc-peptides for the chemical synthesis of cross-linkable H3 proteins. (Figure adapted from Lechner, Agashe, & Fierz, 2016)¹⁸⁷

The strategy is based on transiently crosslinking two H3 molecules, each carrying distinct PTMs. This forces the cross-linked H3 molecules into same octamer assembly. This histone octamer would be used to form asymmetric nucleosome with regards to H3 PTMs. The design is such that, all the additional residues used in the cross-linking process will be

removed after assembly of octamer or nucleosome, thus restoring native architecture of the nucleosome as shown in **Figure 31 A**. To fulfill these design parameters, a “link and cut” tag (*Inc*-tag) was designed which carries a cysteine residue used to crosslink two H3 molecules via a disulfide bond and a tobacco etch virus (TEV) protease site as shown in **Figure 31 B**. In the first step, both the *Inc*-tag, containing a protected cysteine sidechain, and the desired PTMs are incorporated into semisynthetic H3 through expressed protein ligation (EPL).^{194,286} This is followed by the desulfurization of cysteine at the linkage site to alanine to generate the native H3 sequence.²⁸⁷ The cysteine in one of the two molecular species ($^{Inc}H3$) is deprotected followed by activation of the *Inc*-tag in a second H3 molecule ($^{Inc*}H3$). These intermediates can be stored for later combinatorial assembly. Combination of activated molecules with $^{Inc}H3$ molecules results in the formation of hetero-disulfides crosslinks which can be incorporated into histone octamers. Finally, the *Inc*-tag motif is removed by protease treatment thus making the asymmetric nucleosomes devoid of any synthetic scars.

3.3.2. Transient Crosslinking Modified H3 Molecules

Individual histone proteins (H2A, H2B and H4) were recombinantly expressed and purified. Site specifically modified histone H3 molecules were generated using EPL.¹⁹⁴ Two truncated variants of human H3, H3(Δ 1-14) A15C and H3(Δ 1-28) A29C were recombinantly expressed with N-terminal His⁶-SUMO tag. Proteins were purified using Ni-NTA affinity chromatography and tags were cleaved using SUMO protease Ulp1. Four variants of N-terminal tail peptide of H3, carrying trimethylation mark at K4, K27, without any modification and K4 and K27 both were synthesized. The peptides were synthesized removable *Inc*-tag and C-terminal hydrazides which are converted to thioesters in situ following oxidation with sodium nitrite followed by thiol addition. The cysteine in *Inc*-tag was protected using S-acetamidomethyl (Acm) during desulfurization which is required for crosslinking. Purified proteins and peptides were lyophilized and stored at -20°C for later use.

The *Inc*-tag K4me3 H3 tail was ligated with H3(Δ 1-14) A15C using one pot ligation desulfurization reaction.¹⁹² The peptide hydrazide was converted to trifluoroethanethiol (TFET) thioester under denaturing condition. The ligation was carried out by addition of truncated H3(Δ 1-14) A15C which carries a N-terminal cysteine. This product was desulfurized and HPLC purified (yield 51%). Silver-acetate (AgOAc) mediated deprotection of N-terminal cysteine followed by modification by 5,5'-Dinitrobis (2-nitrobenzoic acid) (DTNB) resulted in $^{Inc*}H3K4me3$ as shown in **Figure 32**. This was purified by HPLC (yield 73%).

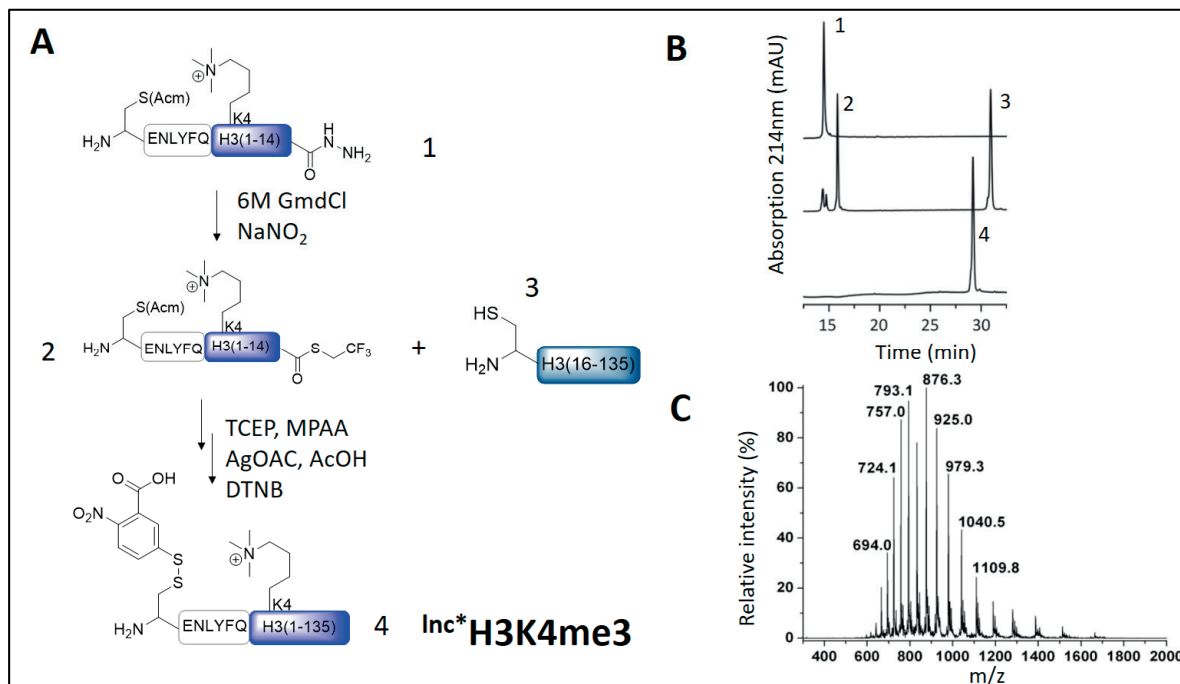


Figure 32. Synthesis of $^{Inc*}H3K4me3$. A) Schematics of the reaction, one pot ligation desulfurization followed by activation of N-terminal cysteine by DTNB. B) Analytical HPLC analysis of each reaction species. C) MS of $^{Inc*}H3K4me3$, Calculated MW: 16635 Da, observed MW: 16631 Da. (Done by Dr. Carolin C. Lechner) (Figure adapted from Lechner, Agashe, & Fierz, 2016)¹⁸⁷

Using a similar procedure but omitting the final activation of cysteine with DTNB, $^{Inc}H3K27me3$ was synthesized as shown in **Figure 33**. For this, the Inc -tag K27me3 H3 tail peptide was ligated with H3(Δ 1-28) A29C, followed by desulfurization of cysteine at the ligation junction to alanine and removal of AcM protecting group from cysteine in Inc -tag. By mixing $^{Inc*}H3K4me3$ (in 20% excess) and $^{Inc}H3K27me3$ under denaturing and slightly acidic condition, $^{xInc}H3K4me3/H3K27me3$ (40% yield of purified product) was formed as shown in **Figure 34**. This was stored and lyophilized for further use. Similarly, following combinations of modified H3 molecules were created: $^{xInc}H3K27me3/H3K27me3$, $^{xInc}H3K4me3/H3K4me3$, $^{xInc}H3K27me3/H3$, $^{xInc}H3K4me3/H3$ and $^{xInc}H3K4me3K27me3/H3$. The advantage of this approach is in its modularity allowing us to generate all possible combinations of H3K4me3 and H3K27me3 marks. We have also maximized the use of one-pot reactions thus the number of HPLC purification steps required are kept at a minimum.

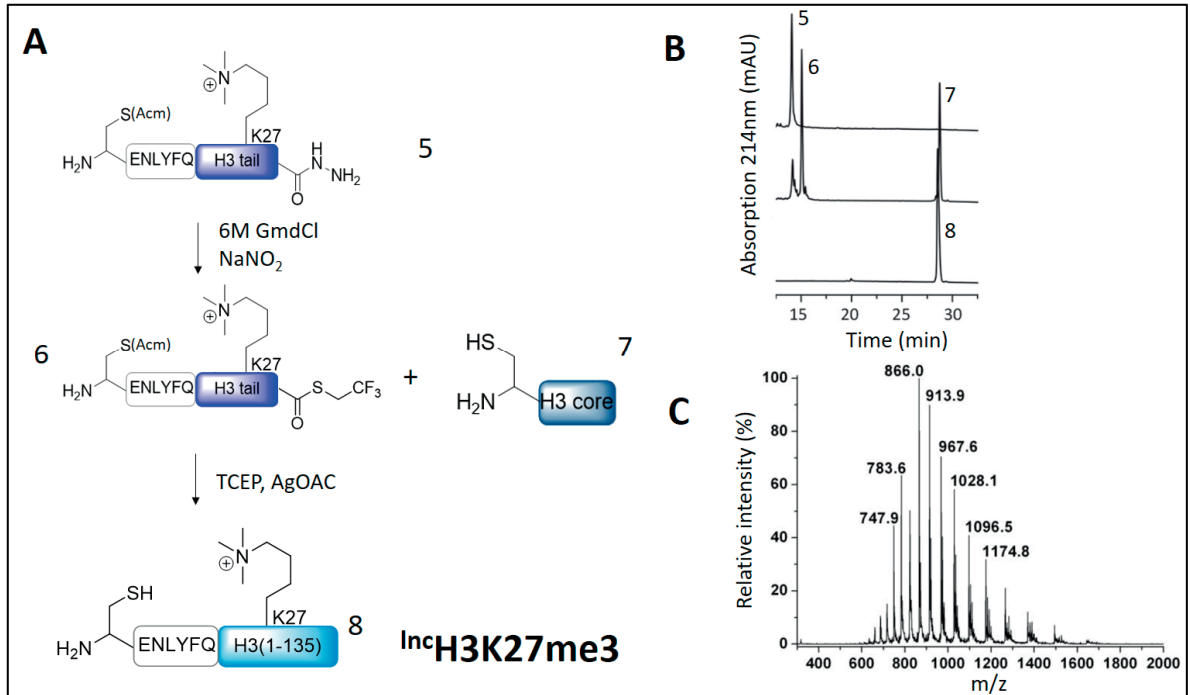


Figure 33. Synthesis of $^{Inc}H3K27me3$. A) Schematics of the reaction, one pot ligation desulfurization followed by Acm deprotection of N-terminal. B) Analytical HPLC analysis of each reaction species. C) MS of $^{Inc}H3K4me3$, Calculated MW: 16438 Da, observed MW: 16433 Da. (Done by Dr. Carolin C. Lechner) (Figure adapted from Lechner, Agashe, & Fierz, 2016)¹⁸⁷

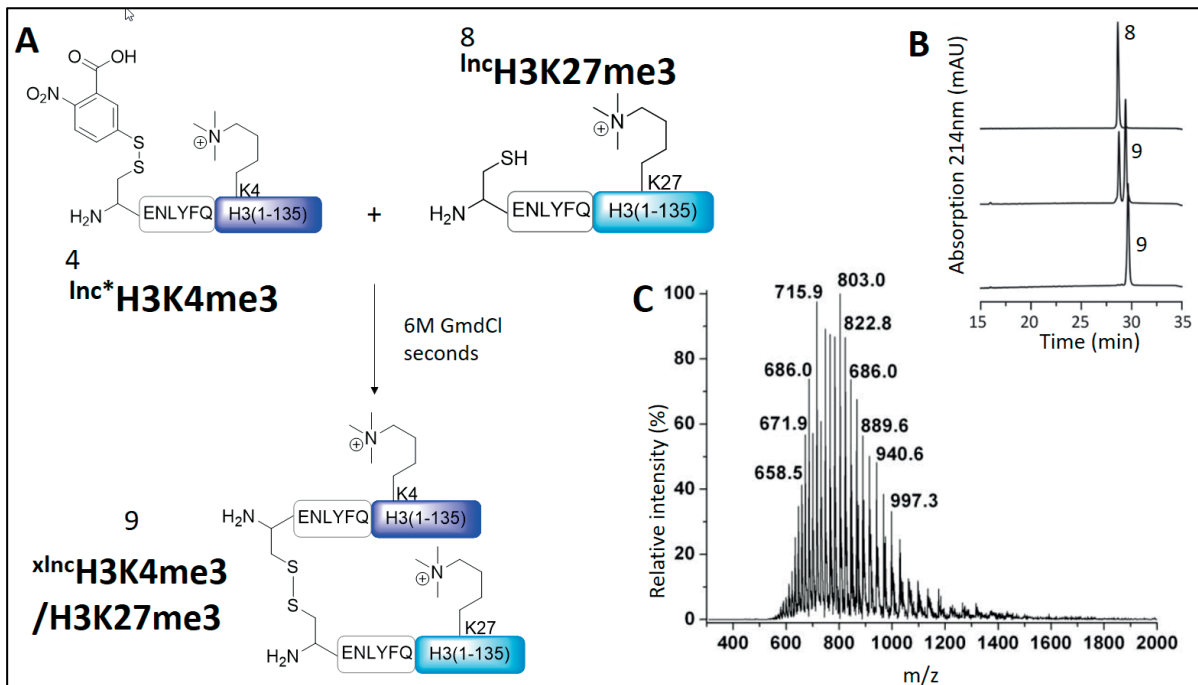


Figure 34. Synthesis of $^{xInc}H3K4me3/H3K27me3$. A) Schematics of the reaction, $^{Inc}H3K4me3$ was added in 20% excess with $^{Inc}H3K27me3$ under denaturing conditions. B) Analytical HPLC analysis of each reaction species. C) MS of $^{xInc}H3K4me3/H3K27me3$, Calculated MW: 32874 Da, observed MW: 32877 Da. (Done by Dr. Carolin C. Lechner) (Figure adapted from Lechner, Agashe, & Fierz, 2016)¹⁸⁷

3.3.3. Reconstitution of Modified Histone Octamers and Nucleosome Library

Having validated our strategy to crosslink asymmetrically modified H3 molecules, we then proceeded to refold the histone octamers with ^xlncH3 and other core histones (H2A, H2B and H4). Octamers were refolded by dissolving stoichiometric quantities of lyophilized histone in denaturing buffer followed by dialysis in 2M NaCl containing buffer. The crude octamers from dialysis were further purified by gel filtration on S200 10/300GL column. Fractions containing pure octamers were pooled and concentrated to 20-80 μM octamer concentration. The pooled fraction is homogenous in nature where each octamers contains one ^xlncH3 molecule as can be judged by the symmetric gel filtration elution profile shown in Figure 35 A. At this point, the disulfide link in ^xlncH3 was still intact which can be reduced with a reducing agent as shown from SDS PAGE analysis of octamers in Figure 35 B. The ^xlncH3K4me3/H3K27me3 trans-bivalent octamers were kept in oxidized and stored in 50% glycerol at -20°C.

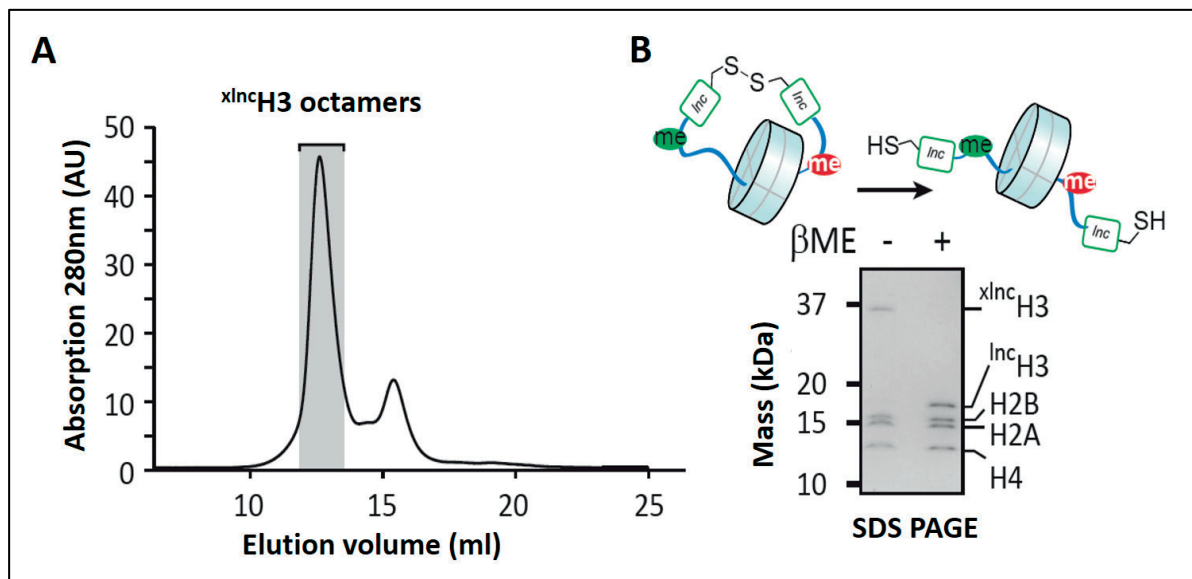


Figure 35. Refolding Asymmetrically Modified Octamers. A) Asymmetric octamers containing ^xlncH3K4me3/H3K27me3 were refolded and purified by gel filtration. Fractions within shaded region were pooled. The peak at 16 ml corresponds to H2A-H2B dimers (added in excess). B) Reduction of disulfide crosslink as analyzed by SDS PAGE and Coomassie brilliant blue (CBB) staining with and without reducing agent. (Done by Dr. Carolin C. Lechner) (Figure adapted from Lechner, Agashe, & Fierz, 2016)¹⁸⁷

Having established a method to refold trans-bivalent octamers (A) using ^xlncH3K4me3/H3K27me3, we proceeded to generate an octamers library carrying all possible combination of H3K4me3 and H3K27me3 mark. We refolded unmodified (B), symmetric s-H3K4me3 (C), asymmetric as-H3K4me3 (D), symmetric s-H3K27me3 (E), asymmetric as-H3K27me3 (F) and cis-bivalent octamers. I analyzed the Octamers containing ^xlncH3 (A, D, F and G) by SDS PAGE under non reducing and reducing conditions to validate reduction of disulfide crosslink as shown in Figure 36.

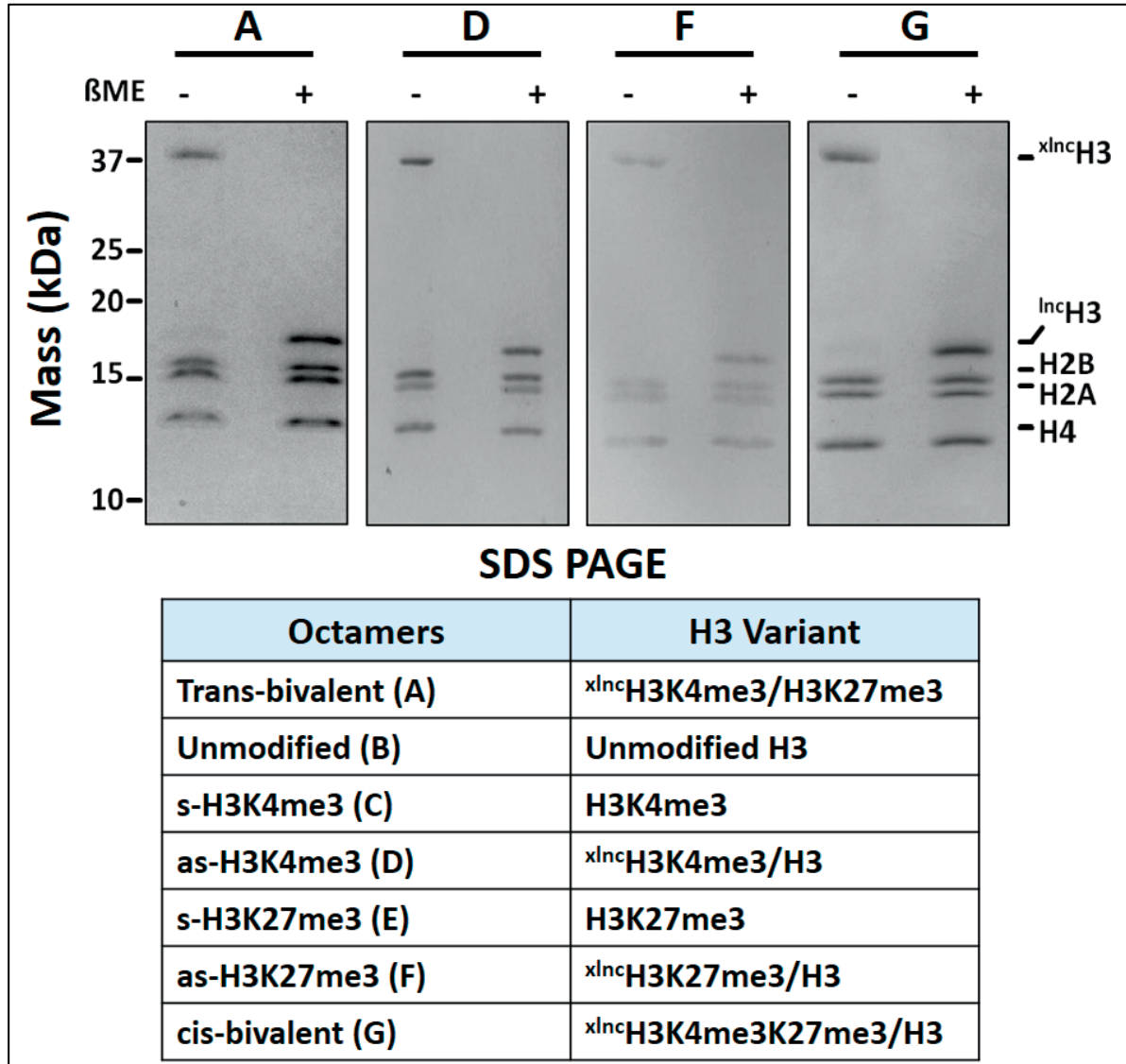


Figure 36. Octamer Library. SDS-PAGE analyses of purified $x^{Inc}H3$ octamers A, D, F and G under reducing (+ βME) and non-reducing conditions (- βME). The library of refolded octamers is described in the table.

We have established a robust protocol to generate asymmetrically modified histone octamers with regards to H3K4me3 and H3K27me3 mark. Using asymmetrically modified octamers, I first reconstituted trans-bivalent nucleosomes by mixing $x^{Inc}H3K4me3/H3K27me3$ with 153bp fragment of “601” nucleosome positioning DNA sequence²⁸⁸ followed by dialysis from high salt (1.4M KCl) to low salt (10mM KCl) buffer. The disulfide link in $x^{Inc}H3$ molecule continued to remain intact during the process of nucleosome formation. After dialysis, the disulfide bond was reduced using 1mM dithiothreitol (DTT) and the Inc-tag was cleaved using TEV protease treatment for 6h at 4°C. This resulted in trans-bivalent nucleosomes devoid of any synthetic scars. The nucleosomes were analyzed by native gel electrophoresis (0.8% agarose) and SDS PAGE (15% polyacrylamide gel) for removal of Inc-tag as shown in **Figure 37 A**. We also analyzed the nucleosomes by MS to find the expected mass of H3K4m3/H3K27me3 nucleosomes (15252Da) seen in **Figure 37 B**. Having established a robust method, I generated

a library of nucleosomes from the library of octamers as shown in **Figure 38 A, B**. I further validated the trans-bivalent nature of nucleosomes by immunoprecipitation (IP) with an antibody against H3K27me3 and detecting H3K4me3 mark by immunoblotting **Figure 38 C**.

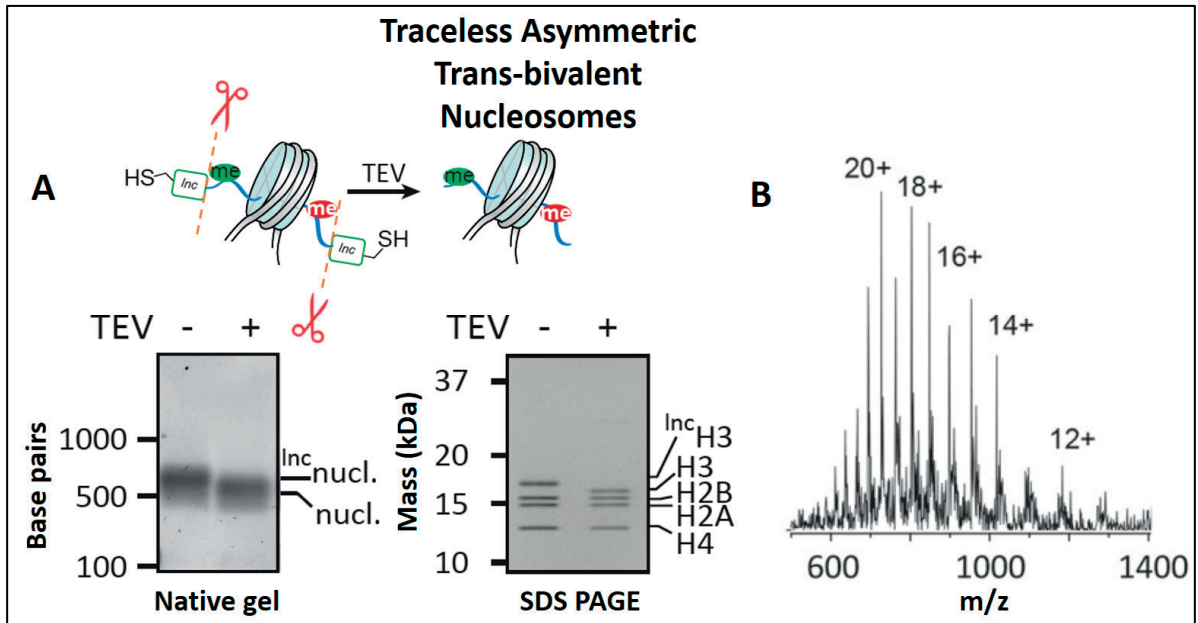


Figure 37. Reconstitution of trans-bivalent nucleosomes. A) *Inc*-tag is removed by TEV protease treatment to yield native trans-bivalent asymmetrically modified nucleosome. B) MS analysis of H3K4me3/H3K27me3 nucleosomes (calculated MW: 15252 Da, Observed MW: 15249 +/- 5 Da). (Mass Spec analysis done by Dr. Carolin Lechner) (Figure adapted from Lechner, Agashe, & Fierz, 2016)¹⁸⁷

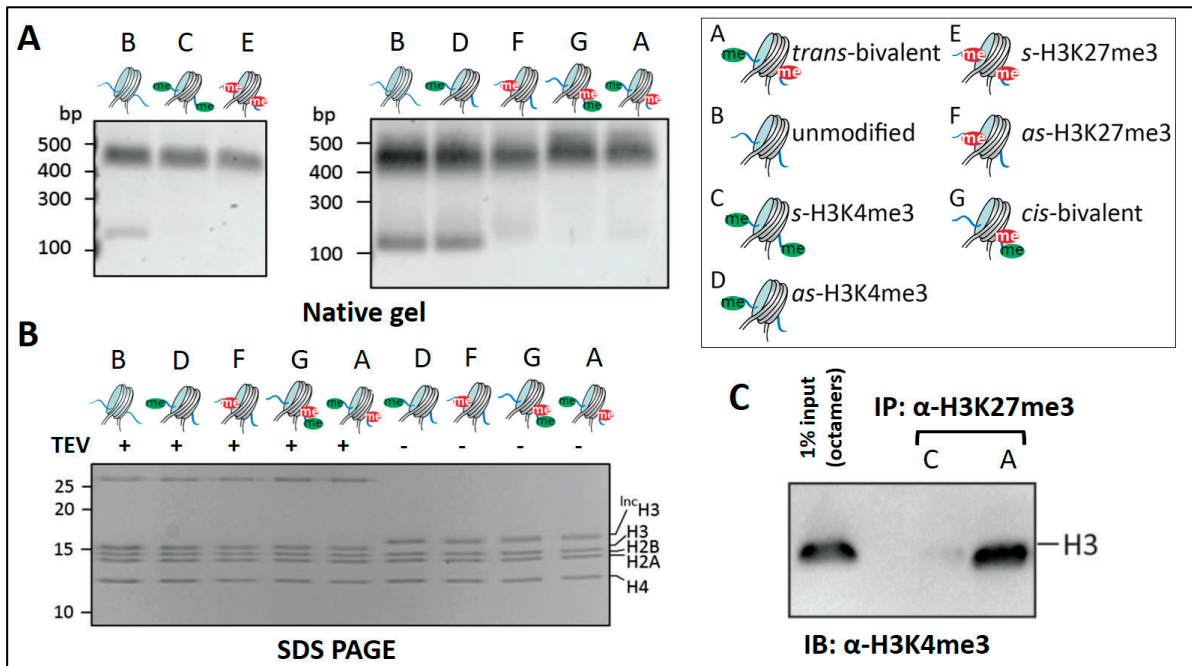


Figure 38. Nucleosome Library. A) Native agarose gel analysis of reconstituted nucleosomes. B) SDS-PAGE analysis of removal of *Inc*-tag by TEV protease treatment. C) IP of nucleosomes to show trans-bivalent state. (Figure adapted from Lechner, Agashe, & Fierz, 2016)¹⁸⁷

3.3.4. PRC2 Methyltransferase Assay Optimization.

A particular goal of here was to understand the impact of existing histone PTM marks on the PRC2 activity. I have generated a robust synthetic method to generate a nucleosome library carrying combinations of H3K4me3 and H3K27me3 marks in symmetric and asymmetric manner. Using this library, I investigated the PRC2 activity on the bivalent chromatin domains. To investigate this, I performed a series of methyltransferase assays using gel based fluorography technique. As a substrate, I used S-adenosyl methionine (SAM) labelled with tritium ^3H -containing-methyl group, which then could be detected X-ray film using fluorography. Thus, using band densitometry, I detected methylation levels by PRC2 on the H3 histone. Recombinant PRC2 was purchased from Active Motif (CatLog number 31387) which includes full length EZH2, SUZ12, EED and RbAp46/48. To establish optimal assay conditions, I incubated unmodified (B) octamers and nucleosomes, PRC2 enzyme of stated quantity with a constant amount of ^3H -SAM for 2h at room temperature. This was analyzed by SDS PAGE and then the gel was soaked in Amersham Amplify Fluorographic Reagent (GE healthcare: NAMP1000) and dried at 80°C for 2hrs. X-Ray film is exposed against the gel at -80°C and developed to detect signal. Based on these results shown in **Figure 39**, I performed all the further assays using 15pmol nucleosomes and 1pmol PRC2 enzyme.

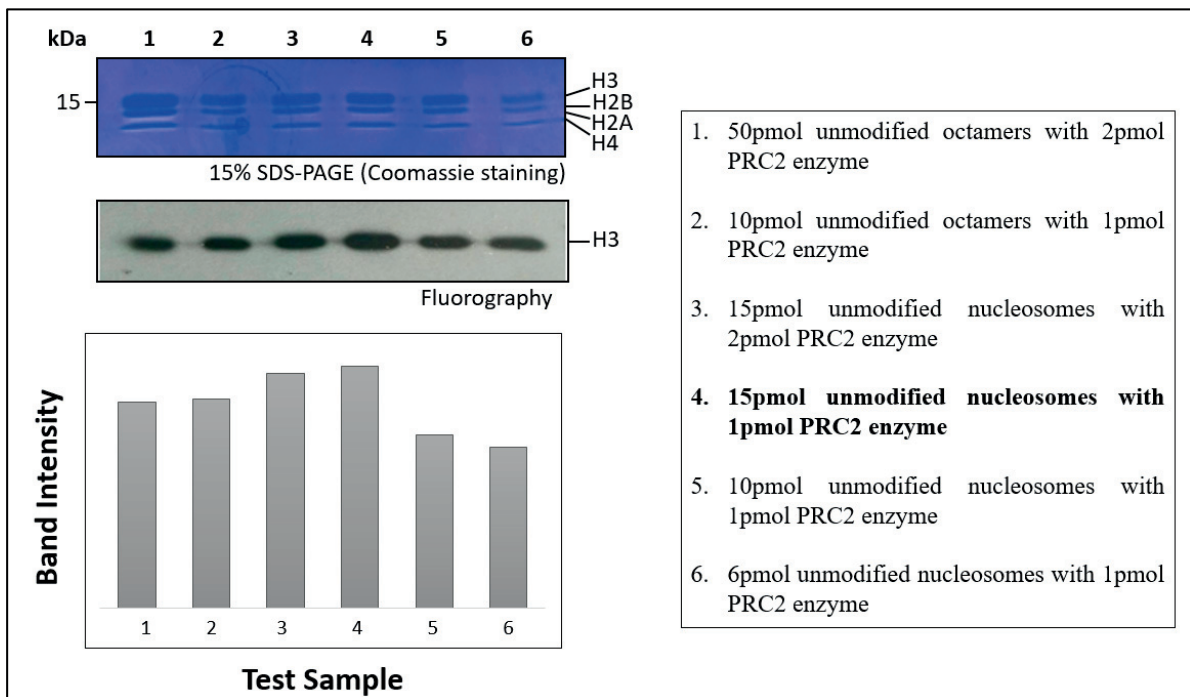


Figure 39. PRC2 Methyltransferase assay. Varying quantities of substrate and enzymes were tested to find the optimal condition. The methyltransferase activity was measure by band densitometry.

3.3.5. Effect of H3K4me3 on PRC2 activity

Having established assay conditions, I investigated the role of each PTM mark individually on PRC2 activity. To investigate the effect of existing H3K4Me3 marks on PRC2 activity, I carried out methyltransferase assay as described in section 3.7 on unmodified (B), and s-H3K4me3 (C) and as-H3K4me3 (D) nucleosomes. I observed that PRC2 activity is completely abolished by s-H3K4me3 whereas it is reduced to 50% by as-H3K4me3 as shown in **Figure 40**.

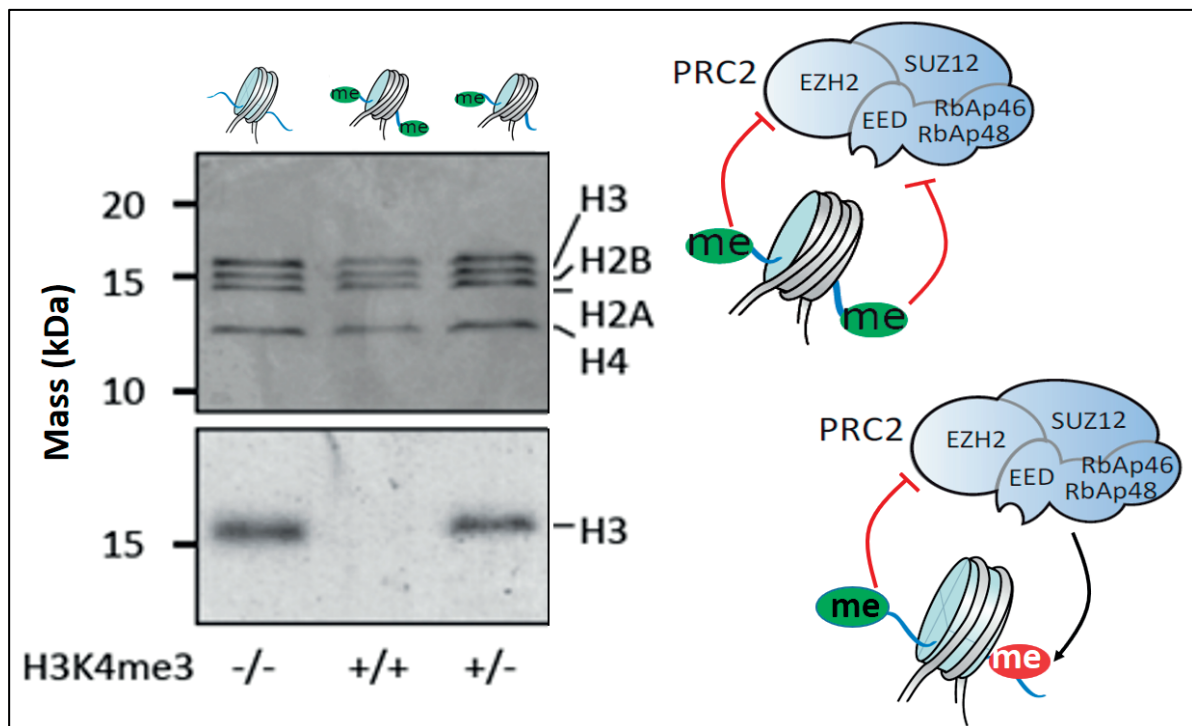


Figure 40. Effect of H3K4me3 on PRC2 activity. The s-H3K4me inhibits PRC2 activity as shown by red arrows whereas as-H3K4me3 reduces PRC2 activity to 50% shown by black arrow (Normal K27 methyltransferase activity of PRC2). (Figure adapted from Lechner, Agashe, & Fierz, 2016)¹⁸⁷

3.3.6. Effect of H3K27me on PRC2 Activity

Having analyzed the effects of H3K4me3 mark on PRC2 activity, I performed the assays using unmodified and s-H3K27me3 and as-H3K27me3 nucleosomes as shown in **Figure 41**. PRC2 did not methylate s-H3K27me3 nucleosome. This shows the specificity of PRC2 enzyme towards K27 position as in s-H3K27me3 nucleosomes, all K27 positions are already modified. In contrast, PRC2 enzyme activity was stimulated by ~2.4 fold by as-H3K27me3 factoring in only one site per nucleosome is available for methylation by PRC2. These results validate the positive feedback mechanism of PRC2 regulation as postulated in literature.^{224,228,233}

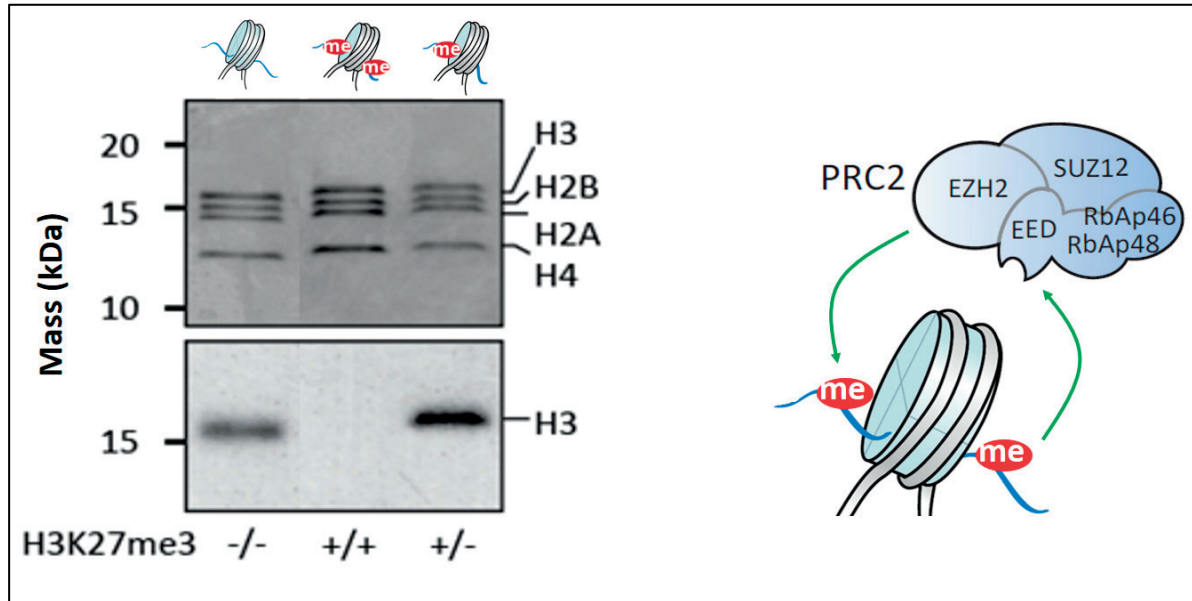


Figure 41. Effect of H3K27me3 on PRC2 activity. PRC2 enzyme does not methylate *s*-H3K27me3 nucleosome, but its activity is stimulated by *as*-H3K27me3 nucleosome as shown by green arrow. (Figure adapted from Lechner, Agashe, & Fierz, 2016)¹⁸⁷

3.3.7. Effect of Combinatorial H3K4me3 and H3K27me3 on PRC2

Further mechanistic understanding of PRC2 regulations were gained by performing methyltransferase assays on bivalent nucleosomes as shown in **Figure 42**. Firstly, PRC2 activity on *cis*-bivalent (G) nucleosomes is increased compared to unmodified, in spite of the presence of the inhibitory H3K4me3 mark. This leads to two inferences. Firstly, the inhibitory effect of H3K4me3 is of local nature and is not extended to other H3 molecule in the same nucleosome and secondly, the effect of stimulation of PRC2 by H3K27me3 has a dominant effect. By contrast, *trans*- bivalent nucleosomes (A) showed low but detectable activity for PRC2. This infers that the stimulatory function of H3K27me3 across the nucleosome can partially override the local inhibition of H3K4me3 as shown in **Figure 42 B**. To further validate the robustness of our assays, I assembled nucleosomes from a mixture of *s*-H3K4me3 and *s*-H3K27me3 octamers **Figure 44 E**. These nucleosomes showed no PRC2 activity as shown in **Figure 44 C**. This leads to three major inferences; 1) during nucleosome assembly no histone exchange occurs between octamers, 2) nucleosomes remain stable during the enzymatic assays and, 3) stimulation of PRC2 by H3K27me3 does not occur in *trans* between individual nucleosomes. These results demonstrate that within a nucleosome, the transcription activating mark, H3K4me3, blocks PRC2 activity only on the same histone H3 tail. In contrast, H3K27me3 has a long range effect and thus is able to override the blocking effect of H3K4Me3 present on the other tail at least partially shown in **Figure 45**. I postulate that this effect may be due to the role H3K27Me3 in recruitment of PRC2 and allosteric enzyme activation. This can be investigated by structural studies regarding organization of PRC2 complex on nucleosomes and understanding molecular interaction driving recruitment and regulation of the enzyme.

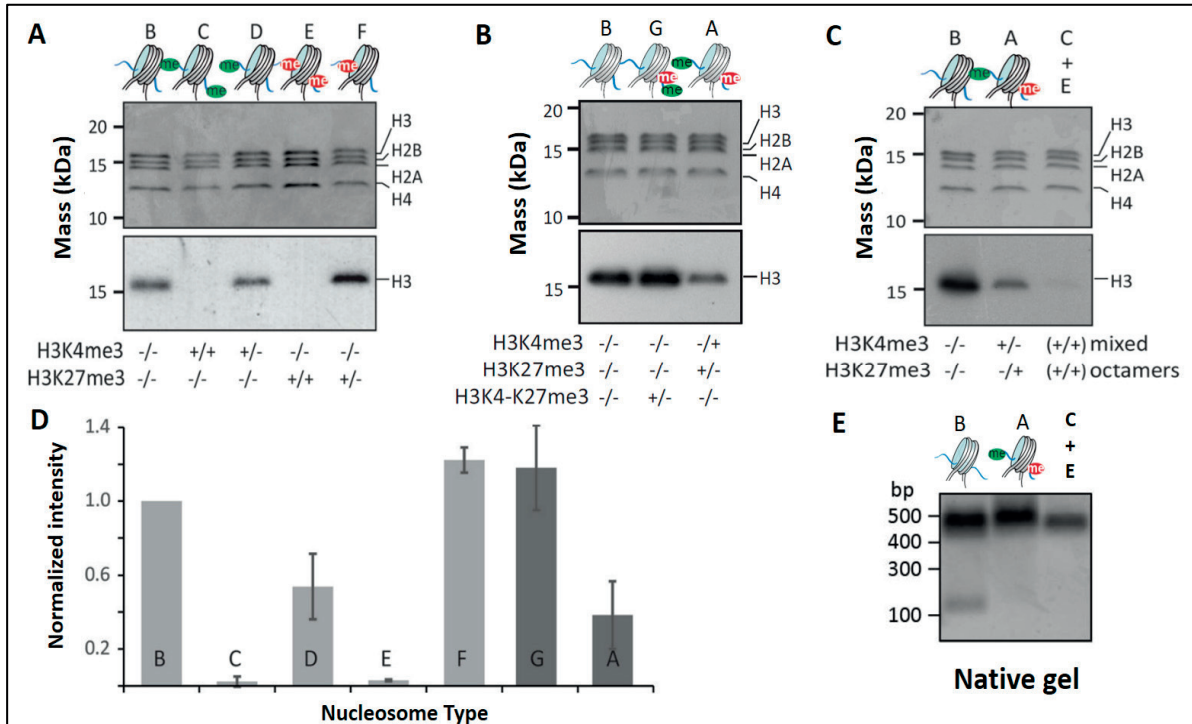


Figure 42. Regulation of PRC2 in asymmetric Bivalent Nucleosomes. A) The indicated symmetric or asymmetric nucleosomes were used for methyltransferase assay. SDS PAGE (upper panel) and fluorography (lower panel) analysis for the assay. B) Cis-bivalent and trans-bivalent nucleosomes were used as substrates for the methyltransferase assay. C) Methyltransferase assays on trans-bivalent nucleosomes and nucleosomes made from 1:1 mixture of *s*-H3K4me3 and *s*-H3K27me3 octamers. D) Incorporation of ³H-SAM by band densitometry over multiple experiments (n=3-4). E) Native gel analysis for nucleosome formation of trans-bivalent and nucleosomes made from 1:1 mixture of *s*-H3K4me3 and *s*-H3K27me3 octamers. (Figure adapted from Lechner, Agashe, & Fierz, 2016)¹⁸⁷

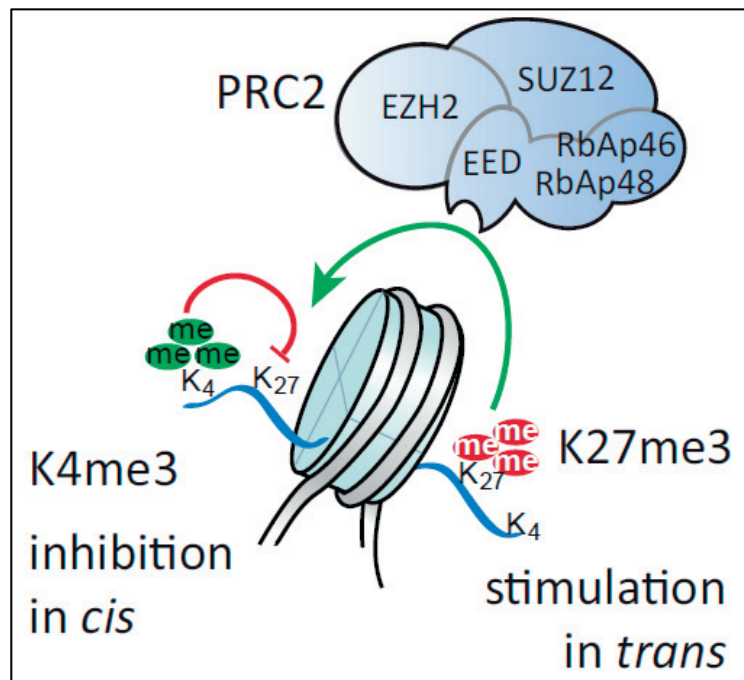
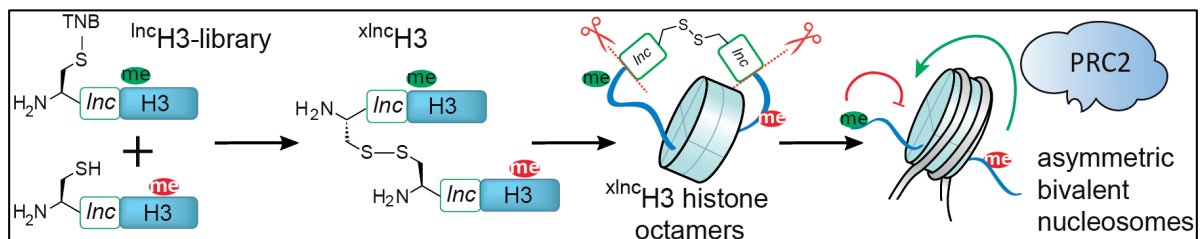


Figure 43. Model for PRC2 Regulation by Asymmetrically Modified Nucleosomes. PRC2 activity is inhibited by H3K4me3 mark, but the inhibitory effect is limited to same H3 molecule. In contrast, PRC2 activity is stimulated by its own mark, H3K27me3. The stimulation effect is long range and can partially override the inhibitory effect of H3K4me3 on other H3 molecule. (Figure adapted from Lechner, Agashe, & Fierz, 2016)¹⁸⁷

3.4. Conclusion and Outlook

We have developed a robust and modular synthetic method to produce asymmetrically modified nucleosomes. To this end, we used a method that transiently linked H3 precursor molecules, carrying distinct PTMs. These were then forced into same octamer molecule. The link was cleaved and the nucleosomes are formed in native state devoid of any synthetic scars. Such asymmetrically modified nucleosomes have wide range of applications in epigenetic research. We used these asymmetrically modified nucleosomes to understand the regulation of PRC2 activity. Our findings show that pre-existing H3K4me3 on one histone tail inhibits PRC2 locally, that is on the same histone tail, but this inhibitory effect can be partially overridden by nearby H3K27me3 which can stimulate PRC2 methyltransferase activity in trans. We thus speculate that bivalent domains have to be constantly maintained in ESCs by restricting PRC2 activity or local reversal of H3K27 methylation.



The asymmetrically modified nucleosomes may be used in mechanistic or high-throughput studies for understanding histone PTM crosstalk²⁸⁹ or in quantitative ChIP experiments.²³⁷ In future, this technique can be further engineered to generate other asymmetrically modified nucleosomes. In our lab, such effort has already been undertaken to generate asymmetrically modified nucleosomes with regards to H4 PTMs and understand the impact of such asymmetric modification on enzymatic activity of Set8.²⁹⁰ These techniques may further be used to understand regulation of variety of enzymes like SET1A, SET1B and MLL. Such understanding would be valuable in understanding the regulation of bivalent domains in embryonic stem cells and cancer cells.

4. Exploring the Function of Microtubule PTMs by Semi Synthetic Tubulin

4.1. Project Background

Microtubules are formed by spontaneous polymerization of tubulin heterodimers which are composed of α - and β - tubulin. The structure of microtubules as well as the sequence of α - and β - tubulin is highly conserved throughout the evolution of eukaryotes. Despite the high level of conservation, microtubules play an important role in diverse range of cellular function involving intracellular organization, formation of mitotic spindle for cell division, cell motility, and neuronal transport. The proper organization and function of microtubules depends on dynamic instability, the stochastic transitions of the microtubule between growth and shrinkage phase.¹⁰² Microtubule functions are finely regulated by their interaction with a large number of Microtubule Associated Proteins (MAPs) and motor proteins (kinesins and dyneins) which step along microtubules to transport cellular cargos as well as alter microtubule stability and dynamic properties.²⁹¹ The complexity of microtubule network regulation is further increased by the inherent 'tubulin code' – which is a combination of differential expression of α - and β -tubulin isotypes and a number of PTMs that impart specific functions.¹¹³

Both α - and β - tubulin consist of a well-defined and highly conserved globular domain (the tubulin body) that contribute a large molecular surface for protomer-protomer interface in the microtubule lattice, while the outside of microtubule shaft is decorated by highly negatively charged unstructured C-terminal region (the tubulin tails).¹³¹ Most of the sequence variations among α - and β - tubulin isotypes occur in these C-terminal tails as described in section 1.3.4. Importantly, the tubulin C-terminal tails contain sites for multiple post-translational modifications (PTMs), including tyrosine phosphorylation, detyrosination, $\Delta 2$ and $\Delta 3$ tubulin, glutamate polyglutamylation and polyglycylation apart from few exceptions such as acetylation (K40) seen in the globular domain. These modifications can modulate MAP interactions, alter MT stability and influence the activity of motor proteins. They are also thought to play an important role in orchestrating the assembly of tubulin-containing complexes such as centrioles.^{292,293} From evolutionary view, **the tubulin code** is very logical for an essential polymer such as microtubules where additional regulatory elements (PTMs and sequence variations) are added to the regions of tubulin that do not participate in lattice interactions. Thus, these modifications are less likely to affect the viability and can be tuned dynamically based on environmental cues.¹³¹ This situation is analogous to myriad of histone PTMs at the N-terminal unstructured histone tails which regulate chromatin dynamics and accessibility (**the histone code**).

To better understand the role of microtubule PTMs in regulating microtubule dynamics and function, *in vitro* functional analysis through biochemical and biophysical assays are required. This has been however hampered by the heterogeneity of tubulin preparations from natural sources such as porcine brain. **For detail investigation of tubulin PTMs, isotypically pure tubulin dimer carrying chemically defined PTMs is required.** This has been a great

challenge as recombinant expression of tubulin in bacteria has been impossible as they do not have the extensive chaperon and cofactor machinery required for the folding of functional tubulin dimer. The most significant development has been the successful purification of fully functional recombinant tubulin using the baculovirus expression system.¹⁶⁷ This system has been used to elucidate the role of β -tubulin isotypes and key mutations in the $\beta 3$ tubulin genes on microtubule dynamics.^{294,295} The α -tubulin acetylated in the globular domain, a PTM that confers mechanical resistance to microtubules could also be recombinantly produced.²⁹⁶ Tubulin tail modifications, especially the branched chain modifications such as polyglutamylation (highly enriched in centrioles) and polyglycylation have not yet been directly investigated. The effect of these modifications is manifested by length of glutamate or glycine chains. Thus, to elucidate the effect of defined tubulin PTMs, in this project I aim to develop a semisynthetic strategy combining recombinant expression of tubulin and protein trans-splicing approach to gain chemical access to the tubulin C-terminal tails. The baculovirus expression and purification of functional tubulin dimer is explained in the following parts of this section. The specific aims and results for generating semisynthetic modified tubulin are presented in the following sections.

4.1.1. Recombinant Expression and Purification of Functional Human Tubulin Dimer

In order to generate semisynthetic tubulin, I have to express $\alpha\beta$ -tubulin dimer lacking C-terminal tails but fused to N-terminal half of the split-intein. To this end, I will express tubulin dimer using baculovirus expression system which was first described by in 2013 by Minoura et al.¹⁶⁷ In this work, they expressed human $\alpha 1$ tubulin (TUBA1B; NP_006073) fused to C-terminal His-tag and human $\beta 3$ tubulin (TUBB3; NP_006077) fused to C-terminal FLAG-tag as shown in **Figure 44**. To improve the yield, L21 lead sequence was added just before the start codon of each gene. The $\alpha 1$ sequence was inserted after the polyhedrin (PH) promotor and the $\beta 3$ sequence was inserted after the P10 promotor in the transfer plasmids. The insect cells in their exponential growth phase were infected with the virus at a multiplicity of infection (MOI) of 20 and cells were harvested after 72 hr. The functional tubulin dimer was purified using sequential anion exchange chromatography, metal affinity chromatography and FLAG affinity chromatography and resultant preparation was concentrated to 3-4 mg/ml concentration and polymerized. The typical yield of $\alpha 1\beta 3$ -microtubule was 5 mg per liter of culture.

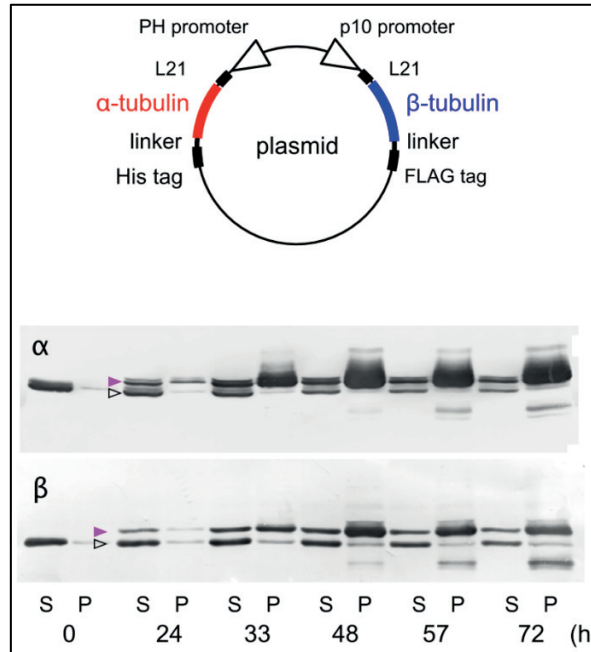


Figure 44. Expression of recombinant human $\alpha 1\beta 3$ -tubulin. The Design of the transfer plasmid is shown in upper panel. Time-course of recombinant tubulin expression after viral infection into cells (shown by magenta arrows), which migrates slower compared to endogenous tubulin. Tubulin in the cell lysate supernatant (S) and precipitate (P) was detected by Western blotting using DM1A (anti- α -tubulin) and ab6046 (anti- β -tubulin) antibodies. (Figure taken from Minoura et. al 2013)¹⁶⁷

4.1.2. Multigene Baculovirus Expression System

As described previously, it is possible to recombinantly express human tubulin dimer in insect cells. For the expression of recombinant tubulin, I used MultiBac™ system from Geneva Biotech which is explained in this section.²⁹⁷ The system utilizes novel transfer vectors enabling modular multigene applications.^{298,299} The system consists of two types of acceptor vectors each with polyhedrin (P_H) or P10 promoter, Gentamycin resistance gene (selection marker), LoxP recombination site and Tn7 transposition sequence (Tn7L and Tn7R) encompassing expression cassettes and a gentamycin resistance marker as shown in **Figure 45 A**. Additionally, system comprises of donor plasmids (**Figure 45 B**) which carry distinct resistance marker and can be fused with acceptor vector or other donor vectors to create modular multigene assembly through Cre-recombinase reaction using LoxP sites. The donor vectors need to be transformed into *E. coli* strains expressing *pir* gene such as BW23473 (due to introduction of R6Ky origin of replication). The resultant acceptor strain is used to introduce the target constructs into the MultiBac™ baculovirus genome (Bacmid) via transposition into the mini Tn7 attachment site. Successful integration results in disruption of the lacZ α subunit-coding sequence. As a consequence, clones carrying inserted DNA will appear white. The schematic view of the cloning steps and expression system are shown in **Figure 45 C**.

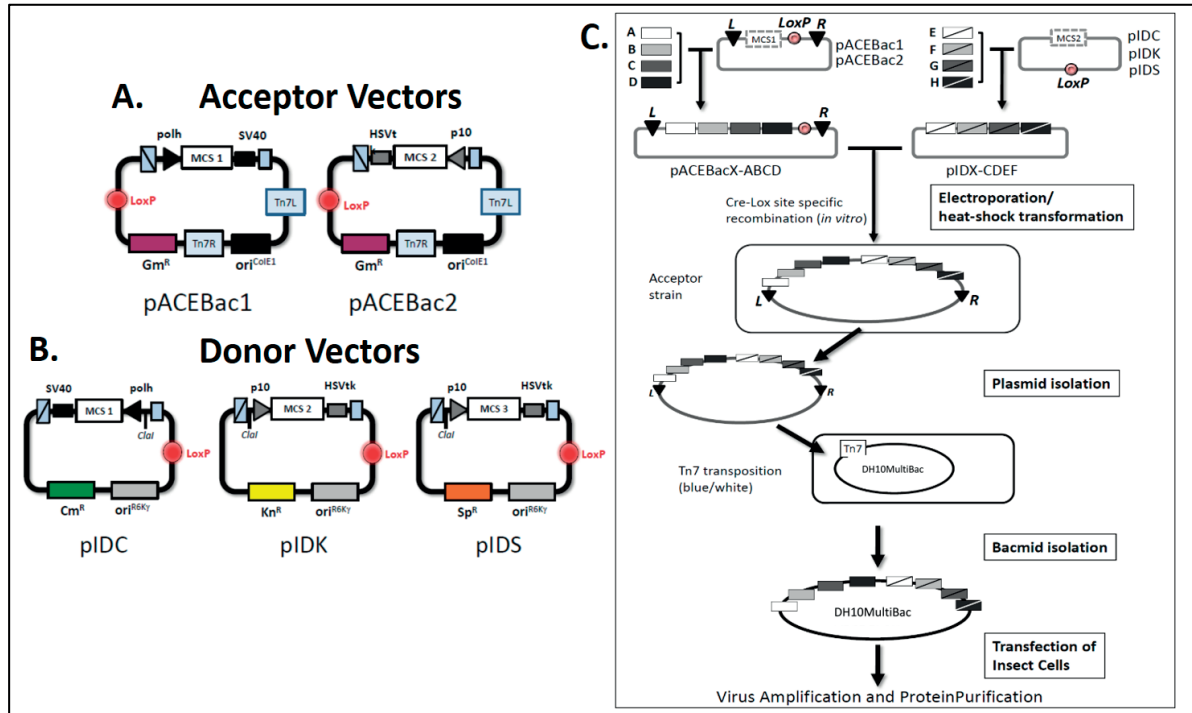


Figure 45. MultiBac™ Baculovirus system. A), B) The design of the acceptor and the donor plasmids used for multigene assembly. C) Schematic overview of the steps for expression of multi-protein constructs using MultiBac™. (Figure adapted from MultiBac™ User manual)²⁹⁷.

4.2. Aims of the Project

In this project, I aim to devise a protein engineering protocol to generate isotypically defined semisynthetic tubulin dimer carrying various chemically defined PTMs. The recombinant expression of human tubulin dimer in insect cells using baculovirus expression system allows to put synthetic handles to tubulin proteins which can be exploited for synthetic exploration of tubulin PTM landscape. To this end, I used split-intein mediated protein trans-splicing approach. In this method, I aim to expressed functional tubulin dimer lacking their C-terminal tails but fused to N-terminal half of a split intein. The C-terminal tail, carrying a defined PTM fused to the C-terminal half of the split intein can be synthetically made using SPPS. The synthetically modified tubulin tail can be ligated to tubulin by protein trans-splicing reaction described in section 1.4.6 and shown in the **Figure 46**. This approach would allow the reaction to be carried out splicing under mild conditions and perhaps even in cell lysates or in future *in vivo*. This also does not need purification of intermediate thioester and refolding steps, which in case of tubulin dimer would be impossible to do *in vitro* considering the extensive chaperon and cofactor machinery required for this purpose. This method would also allow me to introduce branched peptide (polyglutamylation and polyglycylation) with defined length of the PTM and also introduce multiple well defined PTM combinations.

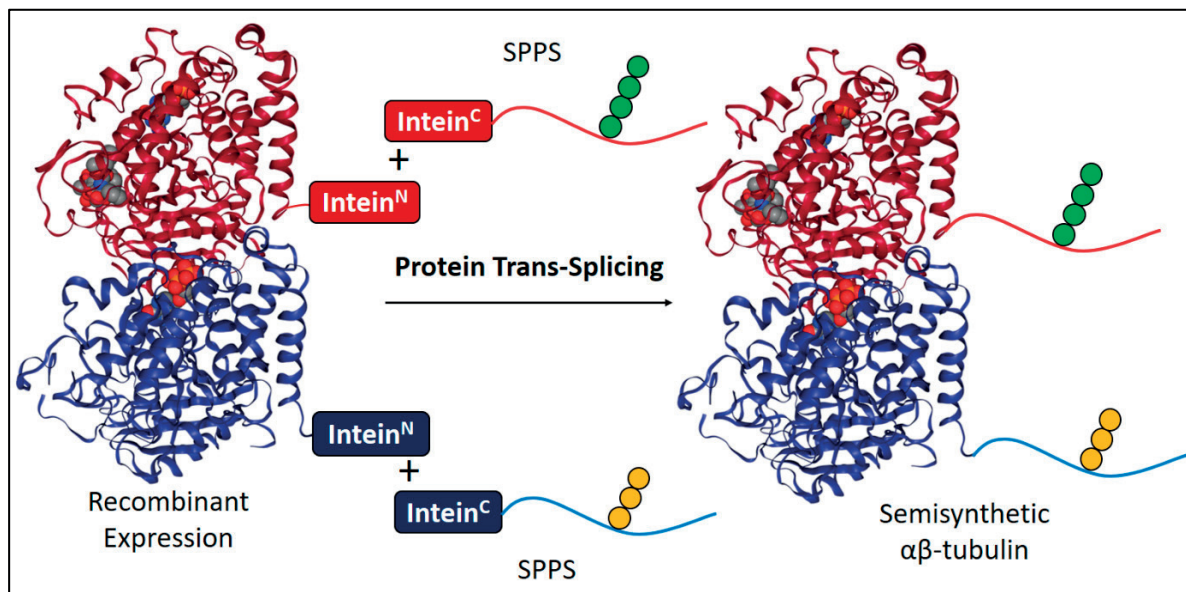


Figure 46. Tubulin Modification Strategy. Chemically defined C-terminal tail modifications can be introduced using a split intein mediated protein trans-splicing approach. (Tubulin dimer structure PDB: 1TUB)

To this end, I have designed following specific aims

1. Design a splice junction within tubulin sequence such that it fulfills the splicing kinetic requirements for the chosen split inteins and leaves behind minimal/no chemical scar.
2. Verify the splice junction to optimize splicing conditions.
3. Design, clone and express human tubulin dimer (unmodified and fused to N-terminal half of split intein).
4. Design and Synthesize modified tail tubulin tail peptides fused to C-terminal half of split intein. This was done by Dr. Eduard Ebberink. To demonstrate the validity of our strategy, we will make a fluorophore labelled tubulin tail peptide mimicking a modification. Also, approaches to synthesize the polyglutamylated tubulin tails are also discussed in the following section.
5. Purify functional semi-synthetically modified tubulin dimers.

This is part of a larger project under NCCR Chemical Biology program to identify, quantify tubulin PTMs, investigate the role of tubulin PTMs on MT dynamics and assembly of centrioles. This involves designing proteomic approaches to identify and quantify tubulin isotypes and PTMs seen in centriolar microtubules (being done by Dr. Eduard Ebberink), characterizing the effects of polyglutamylation on MT dynamics using TIRF microscopy based biophysical measurements and investigate the role of tubulin PTMs in centriole assembly along with centriolar proteins such as SAS6 and CEP135.

4.3. Results and Discussion

4.3.1. Characterization of Splice Junction for Protein Trans-Splicing

As discussed in section 1.4.7., split-intein mediated trans-splicing requires a specific context with regards to the flanking extein residues. Thus, I have to identify a splice junction within tubulin sequence that will adhere to following requirements. Firstly, the splice junction must be sufficiently far from globular domain of tubulin in the unstructured tubulin tails, thereby not interfering with tubulin folding process. Secondly, it must match closely with the natural flanking extein residues that are required for the intein splicing. For this project, I have selected Ava^N (a fast splicing homologue of Npu^N intein)- Npu^C split intein pair.^{207,210} The Ava^N is larger N-terminal half of the split intein (100 aa) and would be fused with the tubulin to recombinantly expressed, while the Npu^C would be synthesized along with modified tubulin tail using SPPS. The intein pair requires 'CFN' as the first three C-extein splicing but can also splice with certain promiscuity in presence of other amino acids. In order to select an appropriate junction in the context of tubulin, I analyzed the mammalian tubulin isotypes by sequence alignment of their C-terminal tail residues as shown in **Figure 47 A, B**. Since, the recombinant expression of α -tubulin A1B and β -tubulin B3 isotype was established by Minoura et al.¹⁶⁷, I picked these for the project. The splice junction chosen are shown by black arrows. In case of α -tubulin, the splicing would leave a scar of 'CF' instead of 'V' residue. Phenylalanine (F) is not uncommon in the tail sequence and is present at this particular position in isotype $\alpha 8$. This is also the case with the splice junction chosen for β -tubulin. I decided not to add asparagine (N) at +3 site but leave the native glutamate (E) as it is. The cysteine (C) is required for the splicing which is uncommon in the unstructured tails, but has been used in the past to link polyglutamate chains via maleimide chemistry to chimeric yeast tubulin with human tubulin isotype tails.¹⁶³ To analyze our splice junction, I designed three constructs mimicking the above mentioned conditions as shown in **Figure 47 C**. The α - and β -tubulin adapter are composed of the last helix (H12) which provides the exact N-terminal extein flanking residues. These were fused with thioredoxin, a soluble protein that can be easily overexpressed. The C-terminal tail was designed with appropriate changes to extein residues that are required for efficient splicing and fused to ubiquitin, another soluble structured protein. All the three constructs can be expressed in the quickly in *E. coli* and thus be tested with trans-splicing assays to estimate splicing kinetics in the context of tubulin proteins.

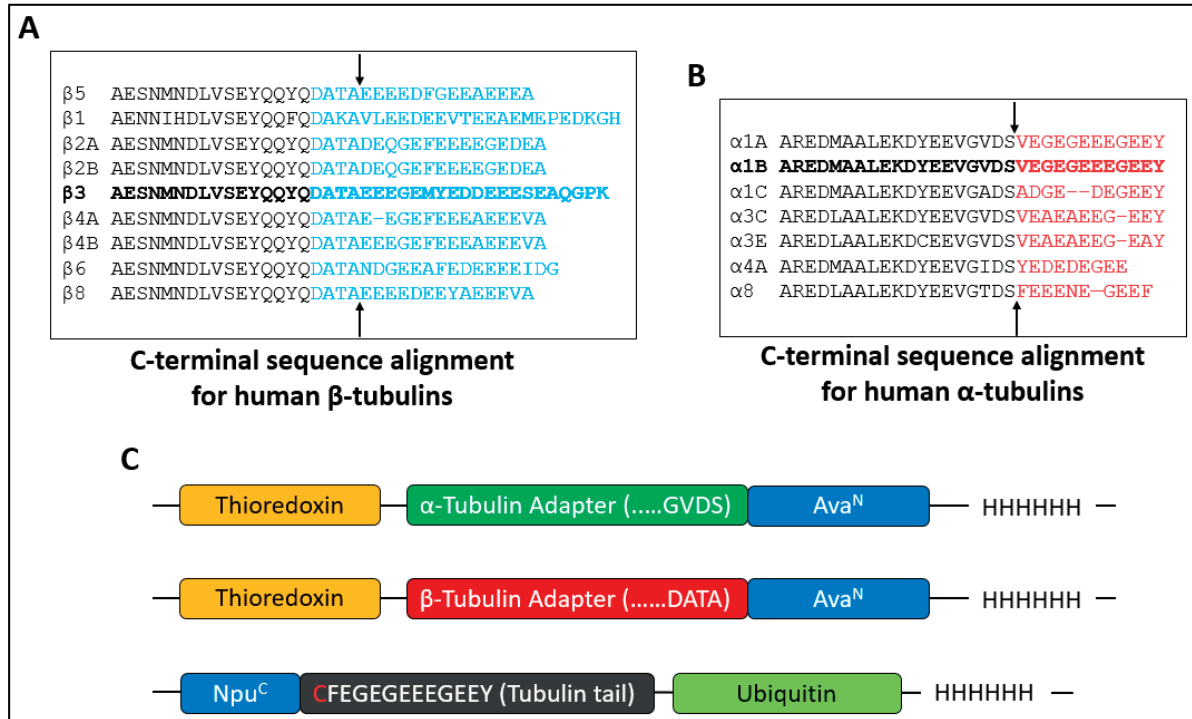


Figure 47. Selection of trans-splicing junction for tubulin. A-B) Sequence alignment of C-terminal residues for human α - and β - isotypes. The chosen one $\alpha 1B$ and $\beta 3$ and shown in bold. The tails residues are shown in color (red for α -tubulin and blue for β -tubulin). The splice junction is marked by a black arrow. C) Design of adapter proteins which will be used to analyze the splice junction and intein splicing kinetics in the context of tubulin.

4.3.1.1. Purification of Tubulin Adapter Proteins

The first step was to purify the individual adapter proteins to set up trans-splicing assays in the context of tubulin. I cloned individual constructs into pET15 vector containing ampicillin resistance marker and IPTG inducible promoter using by Gibson cloning. They were expressed in *BL21 PLYS E. coli* at 37°C for 3hr. Each protein was purified by metal affinity Ni-NTA chromatography followed by size-exclusion chromatography and concentrated to 30-50 μ M concentration range. The Npu^C-Tubulin tail-ubiquitin protein was purified by S75 10/300GL gel-filtration column as shown in **Figure 48 A, B**, while other two were purified by s200 10/300GL column as shown in **Figure 49 A, B and 50 A, B**. Each purified protein was analyzed by RP-HPLC and MS, flash frozen and stored at -80°C for further use.

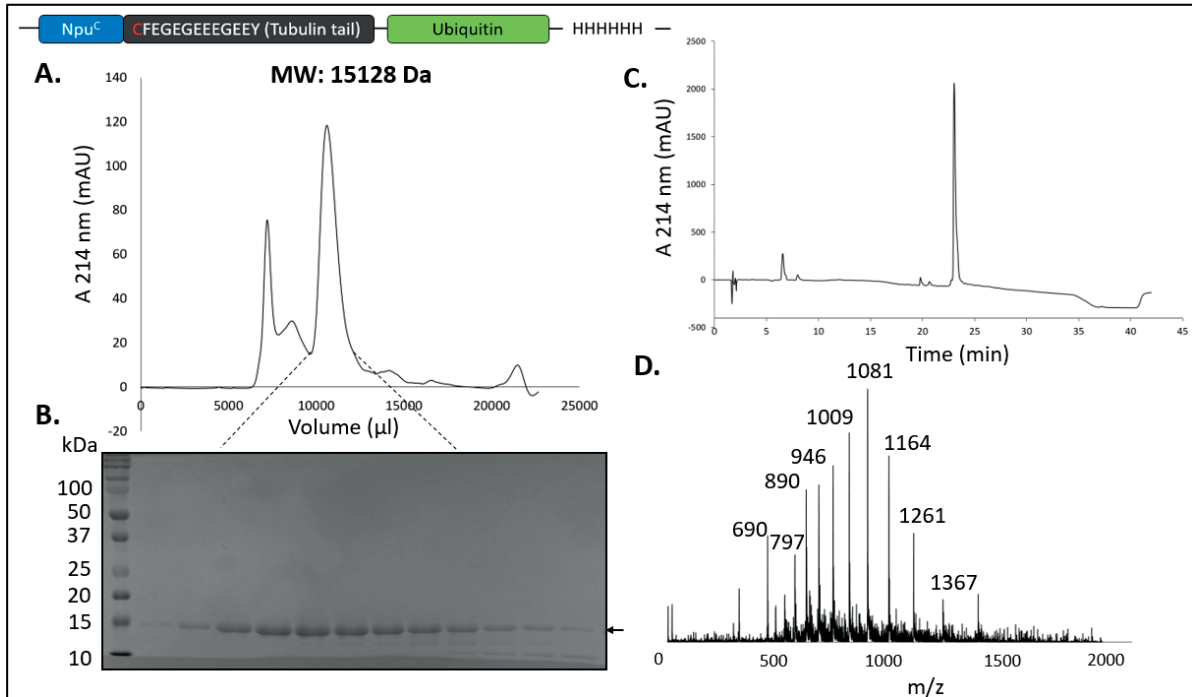


Figure 48. Purification of Npu^C-Tubulin tail-ubiquitin. A) Size-exclusion purification profile. B) The size-exclusion purification fractions corresponding to the peak marked by dashed lines were analyzed by 15% SDS-PAGE. C-D) The RP-HPLC and MS analysis of purified protein. (expected MW: 15128 Da, observed MW: 15129.3 Da)

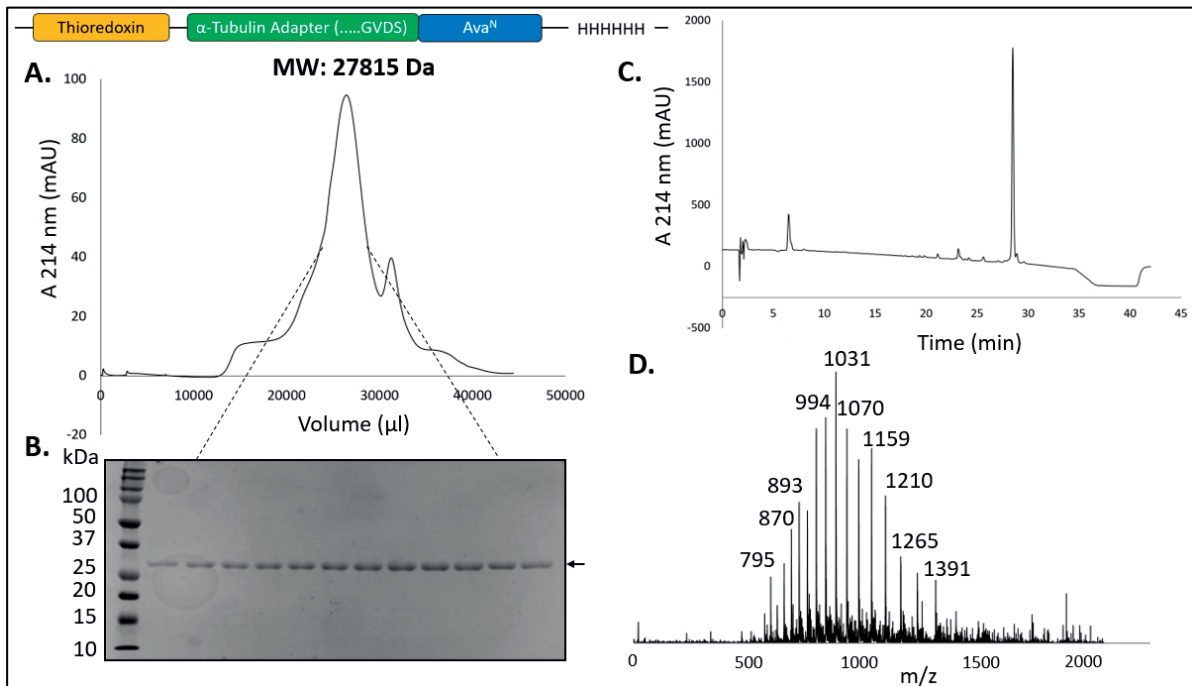


Figure 49. Purification of α-tubulin adapter. A) Size-exclusion purification profile. B) The size-exclusion purification fractions corresponding to the peak marked by dashed lines were analyzed by 15% SDS-PAGE. C-D) The RP-HPLC and MS analysis of purified protein. (expected MW: 27815 Da, observed MW: 27816 Da)

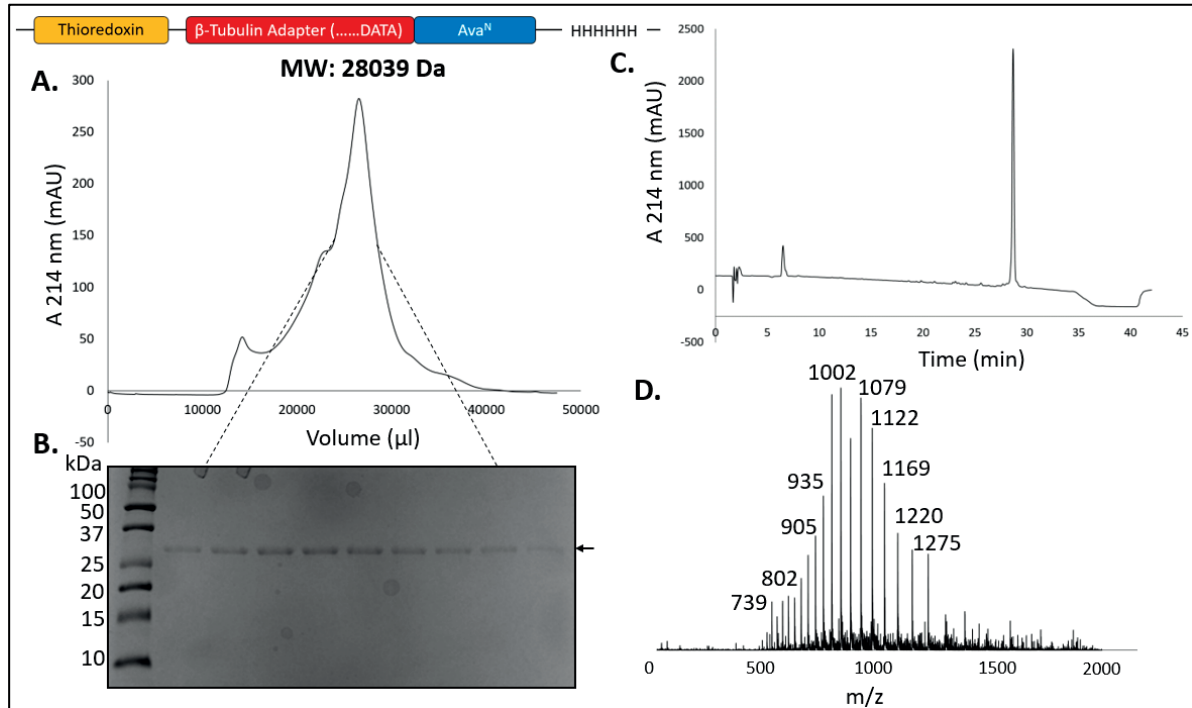


Figure 50. Purification of β -tubulin adapter. A) Size-exclusion purification profile. B) The size-exclusion purification fractions corresponding to the peak marked by dashed lines were analyzed by 15% SDS-PAGE. C-D) The RP-HPLC and MS analysis of purified protein. (expected MW: 28039 Da, observed MW: 28037 Da)

4.3.1.2. Protein Trans-Splicing Assays and Kinetics

I used these adapter proteins to determine the splicing kinetics of the chosen split intein pair in the context of tubulin splice site. For this, I mixed the α - or β -tubulin adapter protein with the Npu^C-Tubulin tail adapter, which was added in 2-fold excess. The reaction was quenched at regular time intervals by adding SDS-gel loading buffer and boiled or by adding 6M guanidine buffer with 4% TFA. The reaction was carried out both at room temperature as well as at 4°C. I specifically wanted to check the reaction kinetics at 4°C, as in case of folded tubulin dimers, I would need to carry out the same reaction in similar conditions. Keeping sample at 4°C state is important during tubulin purification process. As shown in **Figure 51 B and C** and **Figure 52 B and C**, splicing is achieved for both α - and β -tubulin adapters at room temperature as well as at 4°C. At room temperature, the reaction reaches to completion (steady state) in approximately 1hr for both α - and β - tubulin adapters, but the β -tubulin adapter splices slightly faster. This is also well corroborated by the HPLC as shown in **Figure 51 C** and **Figure 52 C**. The HPLC chromatogram shown the disappearance of adapter proteins as they are consumed in the splicing reaction as appearance of released intein a ligated exteins. A double peak can be seen at the earlier time points, which corresponds to the branched intermediate (see section 1.4.7. and Figure 26) which resolves

into the final spliced product. The reaction at 4°C is slower and completed in approximately 2.5 hr. The overall splicing reaction outline is shown in **Figure 51 A** and **Figure 52 A**.

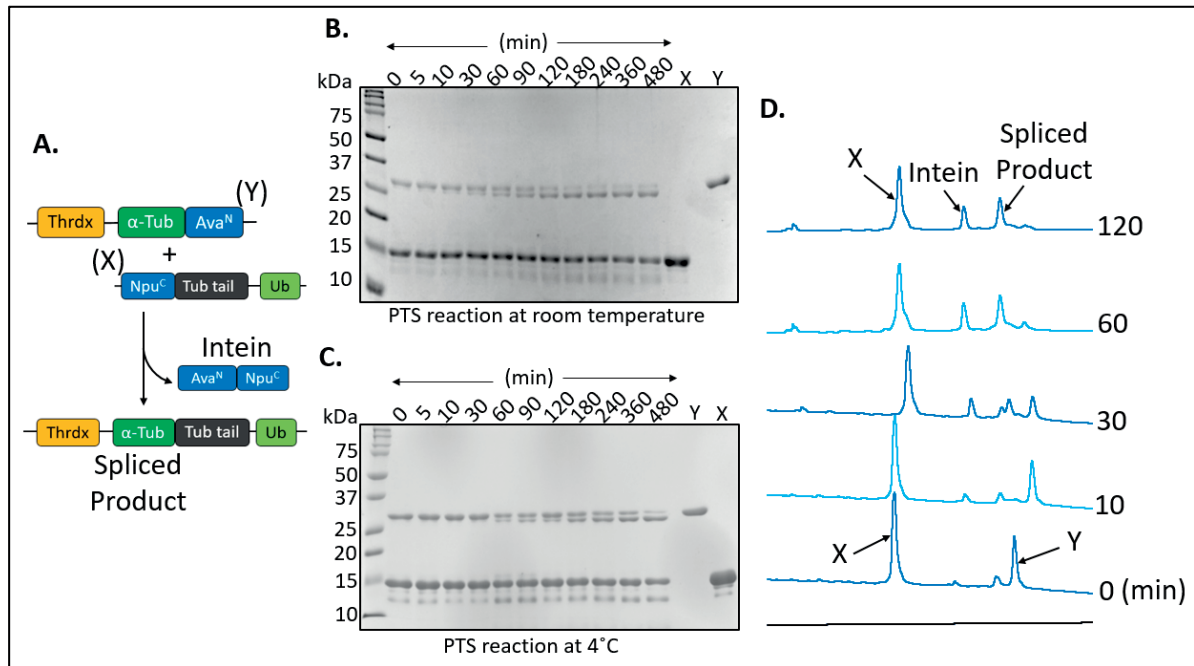


Figure 51. PTS reaction for α -tubulin adapter. A) The reaction outline for PTS reaction. B-C) The progress of PTS reaction at room temperature and at 4°C as seen by SDS-PAGE. D) The reaction progress can also be followed by HPLC which shows the resolution of branched intermediate to final spliced product.

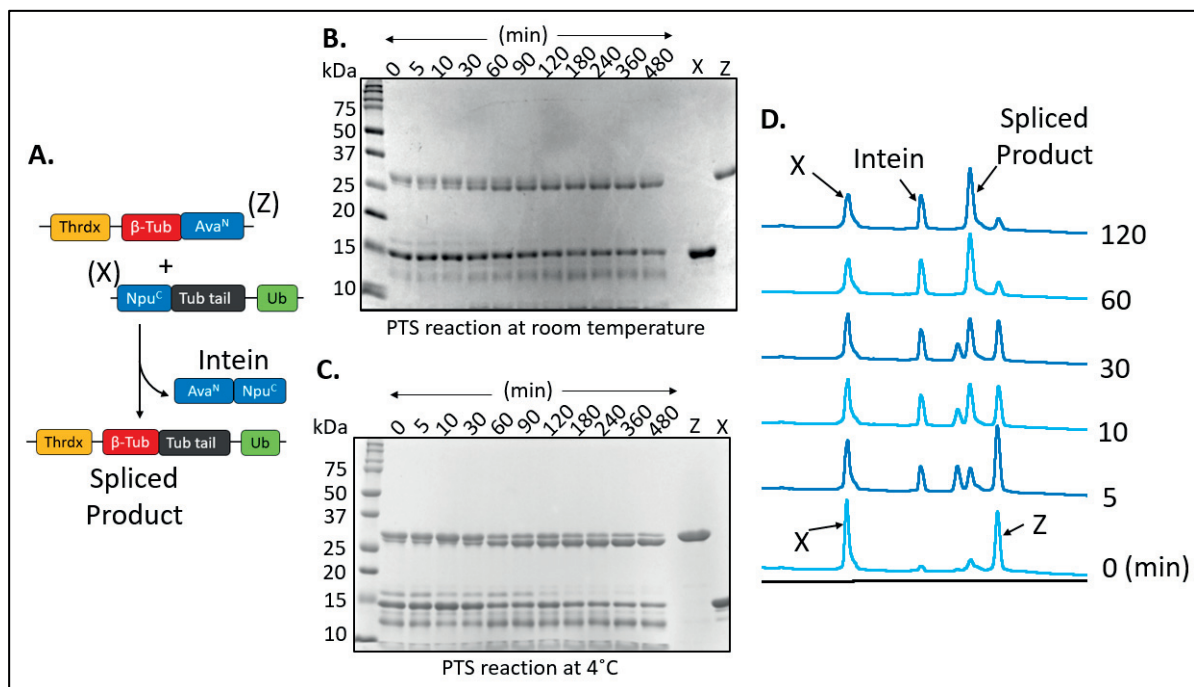


Figure 52. PTS reaction for β -tubulin adapter. A) The reaction outline for PTS reaction. B-C) The progress of PTS reaction at room temperature and at 4°C as seen by SDS-PAGE. D) The reaction progress can also be followed by HPLC which shows the resolution of branched intermediate to final spliced product.

The splicing kinetics for the above two conditions are shown in the **Figure 53**. I plotted the band densities for the emergence of spliced product band and the disappearance of N-intein adapter protein band. Based on this observation, I concluded that the selected splice junction would be appropriate to generate semi-synthetic tubulin by the PTS reaction. The assay used here is very robust and can be used to determine splicing rates for other split-intein pairs in future. In the context of this project, this allowed me to verify the PTS strategy based on which I could proceed to expressing the tubulin constructs.

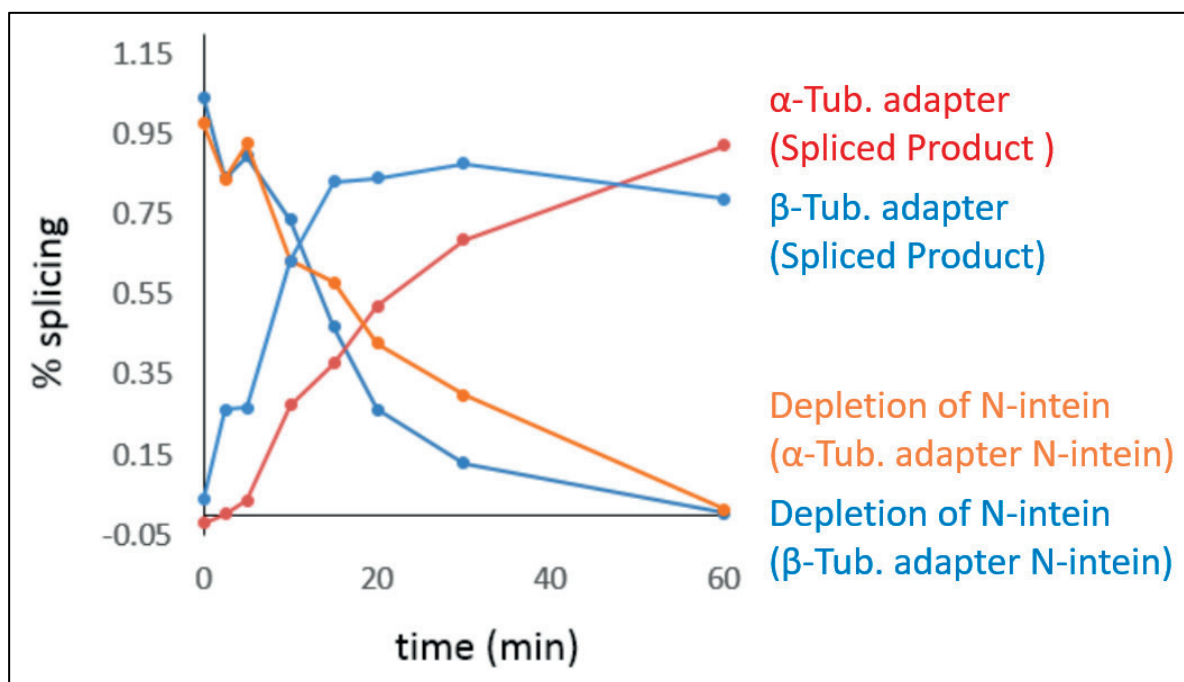


Figure 53. PTS Rate kinetics. The plot shows the rate of emergence of spliced products and consumption of N-intein in the reaction. The kinetics shown here are for the reaction at room temperature.

4.3.2. Expression of Tubulin Dimer in Insect Cells

Having identified the splice site for PTS, I proceeded to express tubulin fused to Ava^N intein using baculovirus expression system as described in 4.1.2. To start with, I decided to clone α -tubulin constructs (with and without intein) in the pACEBac1 acceptor vector while the β -tubulin constructs (with and without intein) in the pIDK donor vector. The plasmids were kindly provided by the group of Prof. Thomas Schalch. I added a N-terminal 6x-HIS tag to α -tubulin and a N-terminal FLAG tag to β -tubulin. I chose to add the tags to the N-termini as it would allow me the flexibility to purify the intein fused tubulin dimer and carry out PTS reaction on purified dimer, but also, if need be, to carry out PTS reaction in cell lysate and purify semisynthetic tubulin dimer. The choice to put purification tags at the N-termini was based of the consideration that the tubulin may not tolerate the addition of non-native C-terminal intein (100 aa long) and thus may misfolded and become insoluble. By carrying out PTS reaction in lysate, the tubulin can be returned to native form before it is purified.

For better expression, the L21 leader sequence from a lobster tropomyosin cDNA³⁰⁰ was added at the beginning of constructs to enhance expression. This is a 5'UTR sequence with the optimized Kozak consensus sequence for insect cell expression. The design of the vectors and the constructs is shown in the **Figure 54 A–C**. The constructs 3 (Tub.A1B) and 4 (Tub.A1BAva^N) were cloned in to the acceptor vector (construct 1) and 5 (Tub.B3) and 6 (Tub.B3Ava^N) were cloned in to the donor vector (construct 2) using Gibson cloning. The tubulin genes were kindly provided by group of Dr. Carsten Janke.

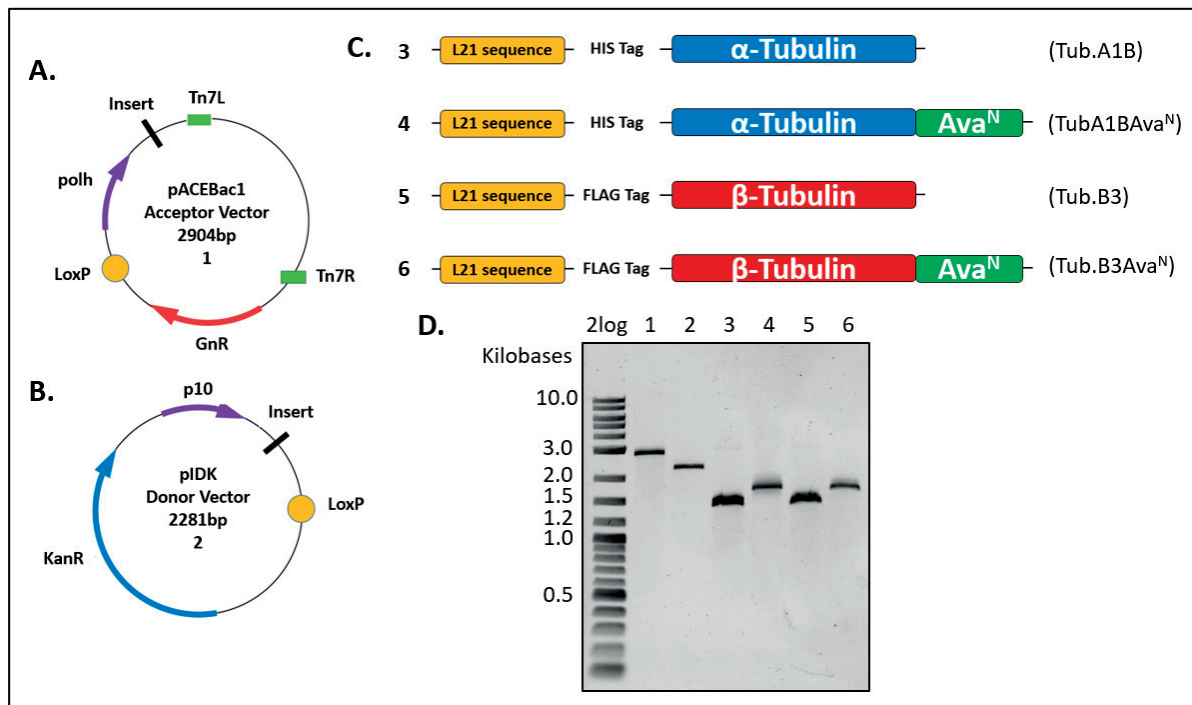


Figure 54. Design of Tubulin Expression Constructs for Baculovirus Expression. A-C) The schematic design of the acceptor and donor vector and the individual gene constructs. (GnR- Gentamycin resistance, KanR- Kanamycin resistance). D) The PCR purified fragments of linearized vectors (1 and 2) and individual gene fragments (3-6) analyzed by 1% agarose gel electrophoresis.

Having cloned the tubulin constructs in appropriate vectors, the acceptor vectors were used to transform *Dh5α E. coli* cells and colonies were selected on a Gentamycin (7 µg/ml) containing agar plate, while the donor vector were used to transform *BW27343 E. coli* cells (which express *pir* gene) and colonies were selected on a Kanamycin (50 µg/ml) containing agar plate.

To generate the multigene vector, I combined the acceptor and donor plasmids using *Cre recombinase*, which recombines the vectors through LoxP sites as shown in **Figure 55 A**. The resultant vectors have a dual resistance selection marker of Gentamycin and Kanamycin. The recombined vector was used to transform *Dh5α E. coli* cells and the colonies were selected on a dual antibiotic agar plate. To further screen out for any false positives, I screened 6 colonies using colony PCR to screen for presence of both α- and β-tubulin genes in the same construct as shown in **Figure 55 C**. The colonies that positively showed presence of both genes were used to harvest the resultant multigene plasmid which will be used to generate Bacmid.

I particularly generated 3 multigene constructs, Tub.A1B-Tub.B3, Tub.A1BAva^N-Tub.B3 and Tub.A1B-Tub.B3Ava^N. I also generated bacmids using the acceptor vector containing only α -tubulin genes. The list of constructs is shown in Figure 57 B. In the future, I can use these vectors to clone additional tubulin isoforms to create multiple combinations of α - and β -tubulin heterodimers. Thus, the modular nature of this system is very advantageous.

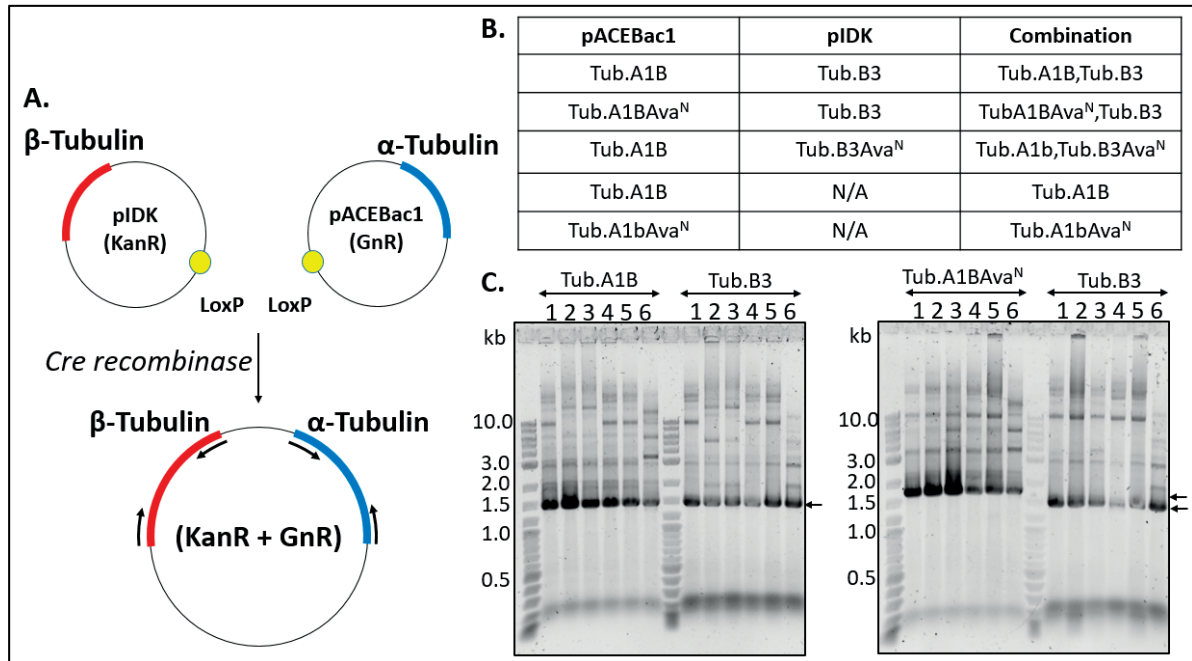


Figure 55. Generation of Multigene Transfer Vectors. A) The scheme of Cre-Lox recombination to generate multigene vectors. The donor plasmid containing β -tubulin or β -Tubulin-Ava^N can be combined with acceptor vector containing α -tubulin or α -Tubulin-Ava^N. B) The list of all the multigene combinations created. C) Colony PCR screen to screen for presence of both genes in combined vector using primers as shown by arrows in panel A (lower part). The gene of interest is marked by arrow next to the gel (1% agarose gel).

The multigene transfer vectors were used to transform DH10MultiBac cells which contain the Bacmid backbone. In the MultiBacTM baculovirus system, two baculoviral genes, *v-cath* and *chiA*, have been disrupted as shown in **Figure 56 A**, which leads to improved maintenance of cellular compartments during infection and protein production.²⁹⁷ The transformed cells were plated on LB agar plates containing 50 μ g/mL kanamycin, 7 μ g/mL gentamicin, 10 μ g/mL tetracycline, 100 μ g/mL X-gal, and 40 μ g/mL IPTG. The genes of interest were transferred to the bacmid via transposition into the mini Tn7 attachment site. Successful integration results in disruption of the lacZ subunit-coding sequence as shown in Figure 56 A. As a consequence, clones carrying inserted DNA will appear white. The recombinant bacmid was isolated by isopropanol precipitation. The bacmid is \sim 135kb in size and thus restriction analysis of entire Bacmid is not possible. To verify successful transposition of insert, I performed PCR analysis using pUC/M13 forward and reverse primers that hybridize to sites flanking the mini-attTn7 site within the lacZ-complementation region as shown in **Figure 56 B**. If the transposition did not happen, the PCR product is \sim 300bp size (as seen in **Figure 56 E** lane 3) while if the transposition is successful, the PCR product would be \sim 4kbp for single gene and

~8kbp for dual gene inserts as shown in **Figure 56 F**. The positive screens are shown in **Figure 56 C, D, E** for indicated constructs. In the lane 1 of the gel in **Figure 58 E**, a positive band at ~4kbp and an impurity at ~300bp is seen. Such constructs are rejected for transfection.

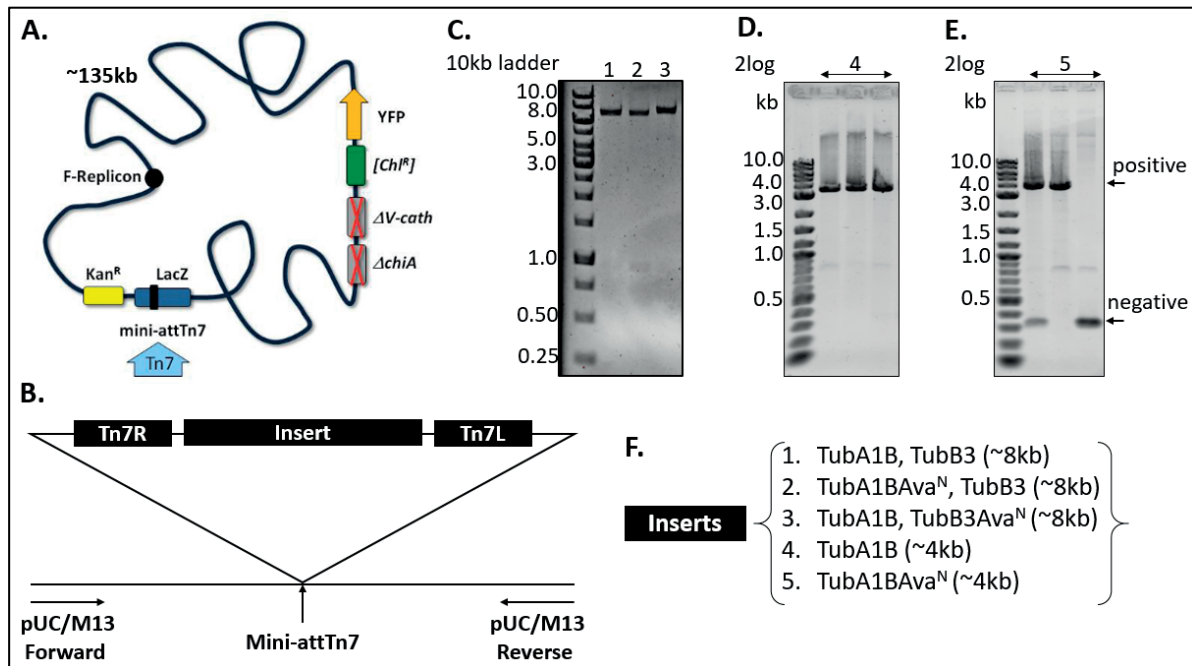


Figure 56. Analysis of Bacmid. A) The viral genome is shown where the Tn7 attachment site is located within a LacZ gene; insertion of Tn7 elements from pACEBac derivatives therefore produces a white phenotype when plated on agar containing BluoGal and IPTG. The viral genes *v-cath* and *chiA* are disrupted. B) The scheme for PCR analysis by pUC/M13 primers is shown. C-E) The analysis of PCR products on 1% agarose gel. F) The type of insert generated and expected PCR product size for each insert by pUC/M13 primers.

Having isolated and verified recombinated Bacmid, I proceeded to transfect insect cells with the Bacmid. I used Cellfectin™ II reagent to transfect SF9 insect cells with the bacmid using standard protocols. After 72 hr. infection, the virus (produced by infected cells) was harvested and amplified to P3 stock (>10⁸ pfu/ml). The P3 stock of virus was used for large scale expression. I infected HighFive insect cells (Life Technologies) growing in the exponential phase (2.0-2.5x10⁶ cells/ml) with viruses at a multiplicity of infection of 20. The cells were cultured in suspension for 72hr at 27°C. The overexpression of tubulin constructs was analyzed by SDS-PAGE as shown in the **Figure 57**. I did not see any overexpression of Tub.B3Ava^N construct on SDS-PAGE or by Western blot analysis using anti-HIS and anti-FLAG antibodies. At this point, I decided to shelve this construct and proceed with the Tub.A1BAva^N construct to modify α-tubulin in the semisynthetic strategy to establish the proof of concept. Though I expressed individual α-tubulin (with and without intein), which will form dimer with insect β-tubulin (construct 1 and 2), I decided to only proceed with the recombinant human αβ-tubulin dimer. I further analyzed the time course expression of constructs 3 and 4 by Western blot analysis. I observed that the amount of non-degraded recombinant tubulin in the soluble fraction reached a maximum level at 33-36hr. Further expression resulted in accumulation of degraded β3-tubulin product as seen in **Figure 58**. Thus, I decided to harvest cells at 36hr after

infection. A large fraction of recombinant tubulin also appeared in the insoluble fraction, probably due to a limited capacity for folding and dimerization.

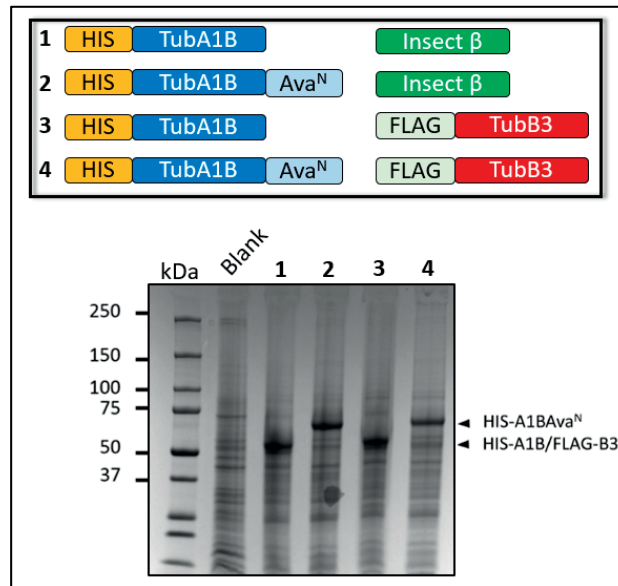


Figure 57. Tubulin Expression in Insect Cells. Number of tubulin constructs shown in upper panel were expressed using baculovirus expression system and the overexpression can be seen on 15% SDS-PAGE.

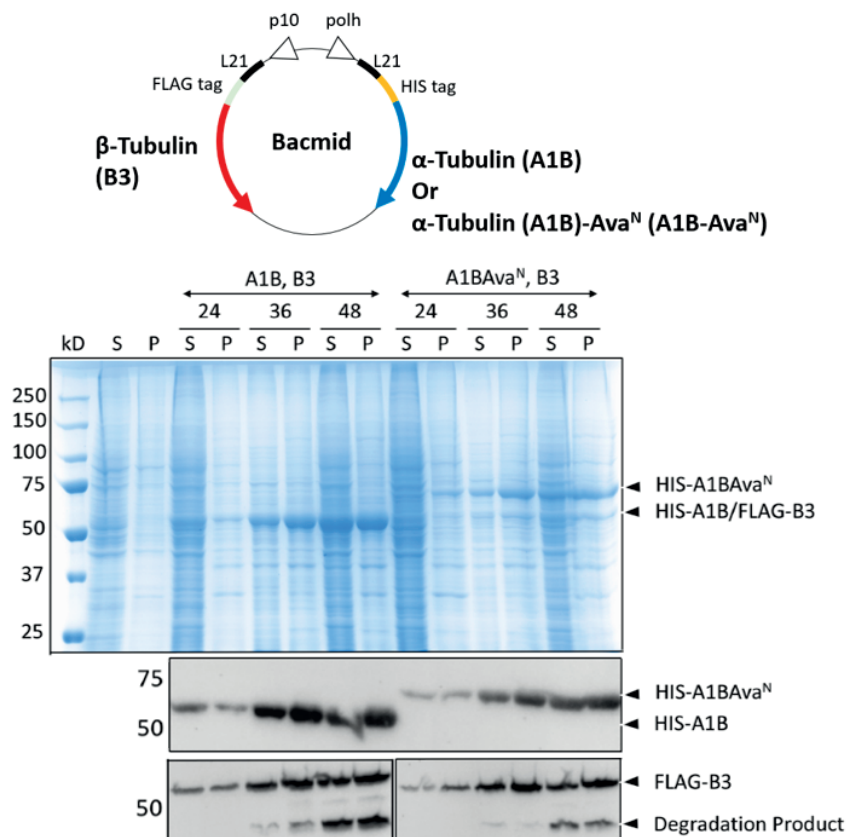


Figure 58. Time Course Expression Analysis of Recombinant Human $\alpha\beta$ -tubulin. The design of the Bacmid with critical expression elements marked. Tubulin in the cell lysate supernatant (S) and precipitate (P) was detected by Western blotting using anti-HIS and anti-FLAG antibodies

4.3.3. Synthesis of Tubulin Tail Peptide

Having established the expression of human $\alpha\beta$ -tubulin dimer using baculovirus system, the next step was to synthesize the tubulin tail peptide fused to the Npu^C intein. The peptide synthesis was done by my colleague Dr. Eduard Ebberink. Solid phase peptide synthesis of the full length 51 aa peptide (Npu^C-tubulin tail) seemed unfeasible. Thus, it was synthesized in two parts and then ligated using Native Chemical Ligation (NCL) as described in section 1.4.4. This required the insertion of a cysteine residue within the intein which upon splicing with the tubulin dimer would be released together with the intein, thus not affecting the tubulin sequence. As a proof of concept of PTS approach to generate semi-synthetic tubulin, we decided to attach fluorescein dye to the tubulin tail sequence. This would allow us to assess the splicing in the cell lysate by SDS-PAGE.

To this end, peptide 1 was synthesized on a Tentagel rink amide. Glycine residues were added at the C-terminal end of the peptide which will be used to link the fluorescein dye. To introduce a modular strategy of C-terminal tagging, we synthesized a peptide tail an alkyne group for click chemistry of various tags. By use of copper-catalyzed azide-alkyne cycloaddition, it would be possible to conjugate a biotin and fluorescein group to the C-terminal construct. As shown in **Figure 59**, the peptide 1 was functionalized with fluorescein and purified.

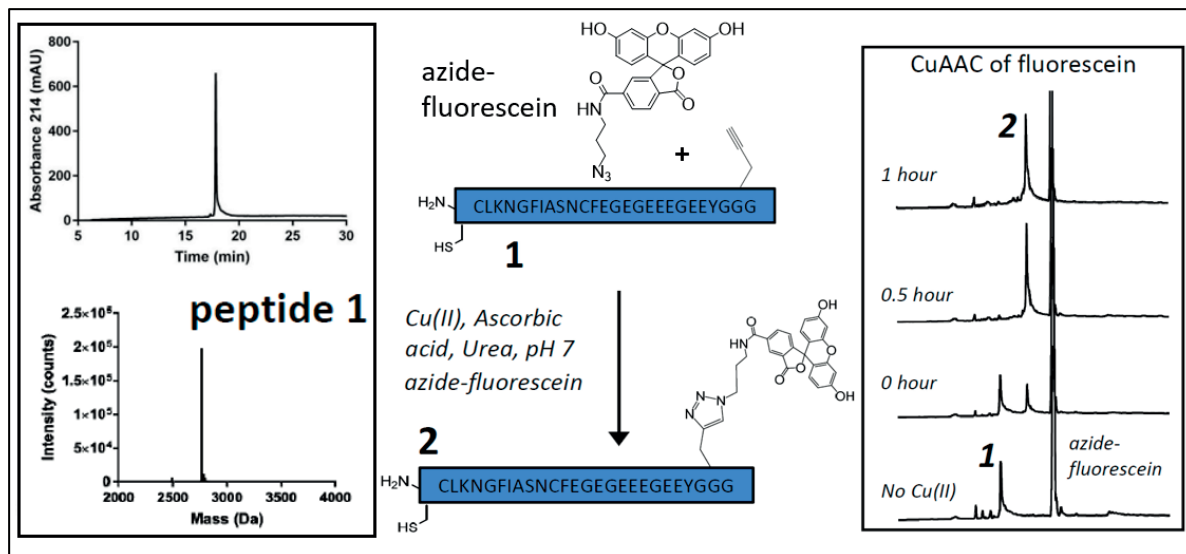


Figure 59. Fluorescein labelling of tubulin tail by copper catalyzed click chemistry. The peptide 1 was synthesized N-terminal cysteine and a propargylglycine which was used to link fluorescein dye via copper catalyzed click chemistry. The RP-HPLC and MS analysis of peptide 1 is shown in left panel (MW: 2762.8 Da). Analysis of the fluorescein labelling reaction by RP-HPLC is shown in right panel. (Done by Dr. Eduard Ebberink)

The other half of the Npu^C-tubulin tail peptide was synthesized on 2-chlorotrityl resin and purified as hydrazide peptide 3. This was then converted to thioester *in situ* using methyl thioglycolate (MTG), described as peptide 4 as shown in **Figure 60**. The peptide 2 and peptide 4 were ligated by NCL resulting in final Npu^C-tubulin tail peptide labelled with fluorescein as

shown in **Figure 61**. The alanine to cysteine mutation introduced to facilitate native chemical ligation did not affect the trans splicing reaction when tested with N-terminal tubulin adapters. It would also not affect the tubulin sequence as the mutation is in intein part which would be eliminated. Thus, the cysteine was not desulfurized back to alanine. This strategy also forms the basis for synthesis of other peptides carrying variety of tubulin tail PTMs.

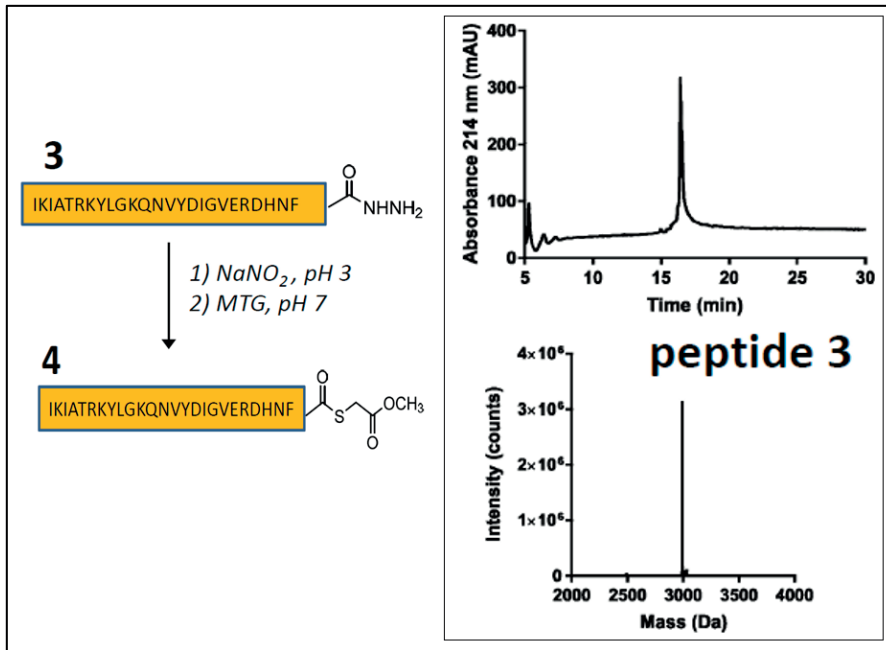


Figure 60. Synthesis of NpuC-tubulin tail peptide. The peptide 3 was synthesized and purified as hydrazide which was converted to C-terminal thioester using MTG. The RP-HPLC and MS analysis of peptide 3 is shown (MW: 2992.1 Da). (Done by Dr. Eduard Ebberink)

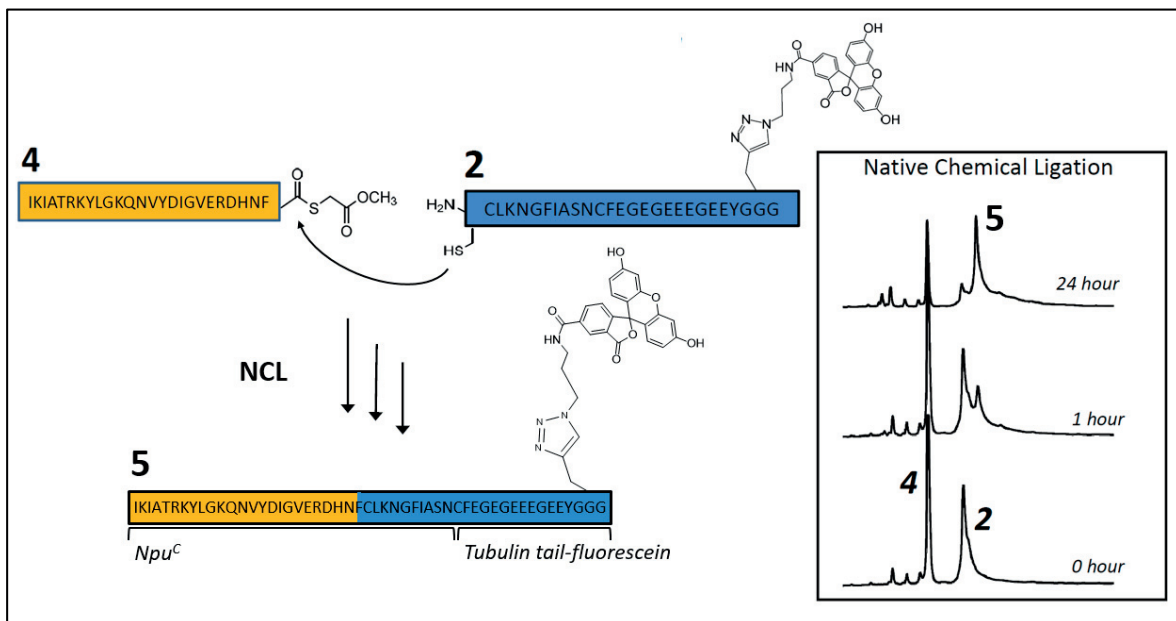


Figure 61. Native Chemical Ligation to generate the NpuC-tubulin tail. Peptide 2 (fluorescein labelled) carrying N-terminal cysteine and peptide 4 carrying C-terminal thioester were ligated and the reaction was analyzed by RP-HPLC (Ligated product MW: 5723.3 Da). (Done by Dr. Eduard Ebberink)

4.3.4. PTS Tests in Cleared Cell Lysates to Generate Fluorescein Tagged Tubulin

Having synthesized Npu^C-Tubulin tail (fluorescein labelled), I decided to validate the strategy of PTS to generate semisynthetic tubulin. To validate the strategy, I decided to investigate the PTS reaction in cell lysates expressing $\alpha\beta$ -tubulin dimer fused to Ava^N. I lysed the cells and centrifuged the lysate at 20,000 x G for 25 min to separate soluble protein fraction from cellular debris and protein aggregates. I added the 5 μ M Npu^C-Tubulin tail (fluorescein labelled) peptide to cleared cell lysates and quenched the reaction at specified time points by SDS-PAGE loading buffer. The reaction was carried out at 4°C. By analyzing the reaction on SDS-PAGE, I observed a fluorescein labelled band appearing at ~50 kDa weight corresponding to tubulin dimer as shown in **Figure 62 A**. I further validated this by Western blot which showed that the fluorescein labelled band corresponds with HIS-tag α -tubulin as shown in **Figure 62 B**. Here, I also observed a HIS-tag α -tubulin Ava^N at ~70 kDa which corresponds to a fraction that did not undergo PTS. A potential explanation is that this corresponds to aggregated protein which were not separated by 200,00 x G centrifugation, and thus I cleared the lysate at 200000G ultracentrifugation prior to splicing. This proved to be a better option as I was able to remove of non-reactive protein aggregates completely as shown in **Figure 62 C**.

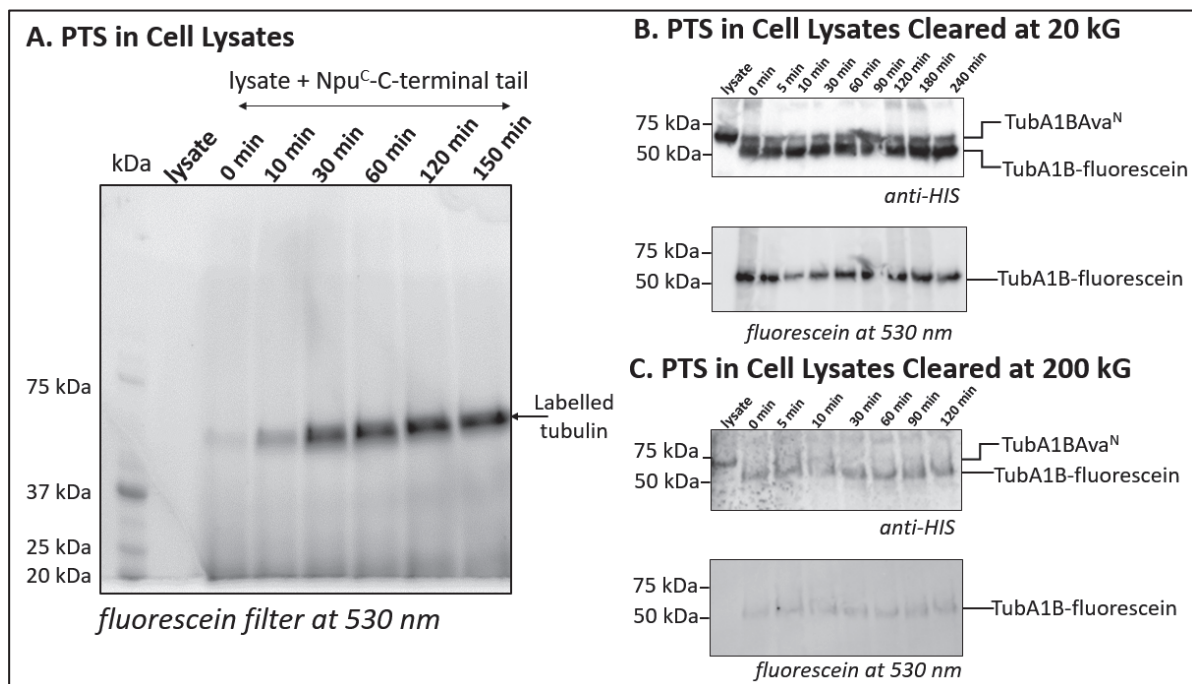


Figure 62. PTS is Cleared Cell Lysates. A) The progress of PTS reaction to generate fluorescein labelled tubulin analyzed by SDS-PAGE. The gel was scanned at 530 nm. B) Western blot analysis of PTS reaction in cell lysates cleared by 20000 G. C) Western blot analysis of PTS reaction in cell lysates cleared by 200,000 x G.

4.3.5. Purification of Functional Tubulin Dimers

I have established that it is possible to generate semisynthetic fluorescein labelled tubulin in cell lysate using PTS. The next aim was to purify functional tubulin dimers. All the purification steps were carried out at 4°C. I first cleared the cell lysate using high speed centrifugation at 200000 G to remove cell debris and misfolded and aggregated proteins. I then fractionated the cleared cell lysate through DEAE-Sepharose (diethylaminoethanol covalently linked to Sepharose) weak anion-exchange chromatography. This is a crude purification step and the DEAE elution has four tubulin species, endogenous α -tubulin, recombinant α -tubulin, endogenous β -tubulin, and recombinant β -tubulin. From this DEAE eluent, recombinant α -tubulin containing dimer was enriched using the immobilized metal affinity chromatography (IMAC) using TALON metal affinity resin (a tetra-dentate chelator charged with cobalt Co^{2+}). The recombinant $\alpha\beta$ -tubulin was subsequently enriched by using the FLAG-affinity chromatography using ANTI-FLAG M2 affinity gel (purified murine IgG1 monoclonal antibody covalently attached to agarose by hydrazide linkage) as shown in Figure 63. I observed that the yield from FLAG purification was very poor as most of the recombinant α -tubulin was lost as dimer with endogenous β -tubulin during FLAG purification as seen by western blot analysis (Figure 63 C).

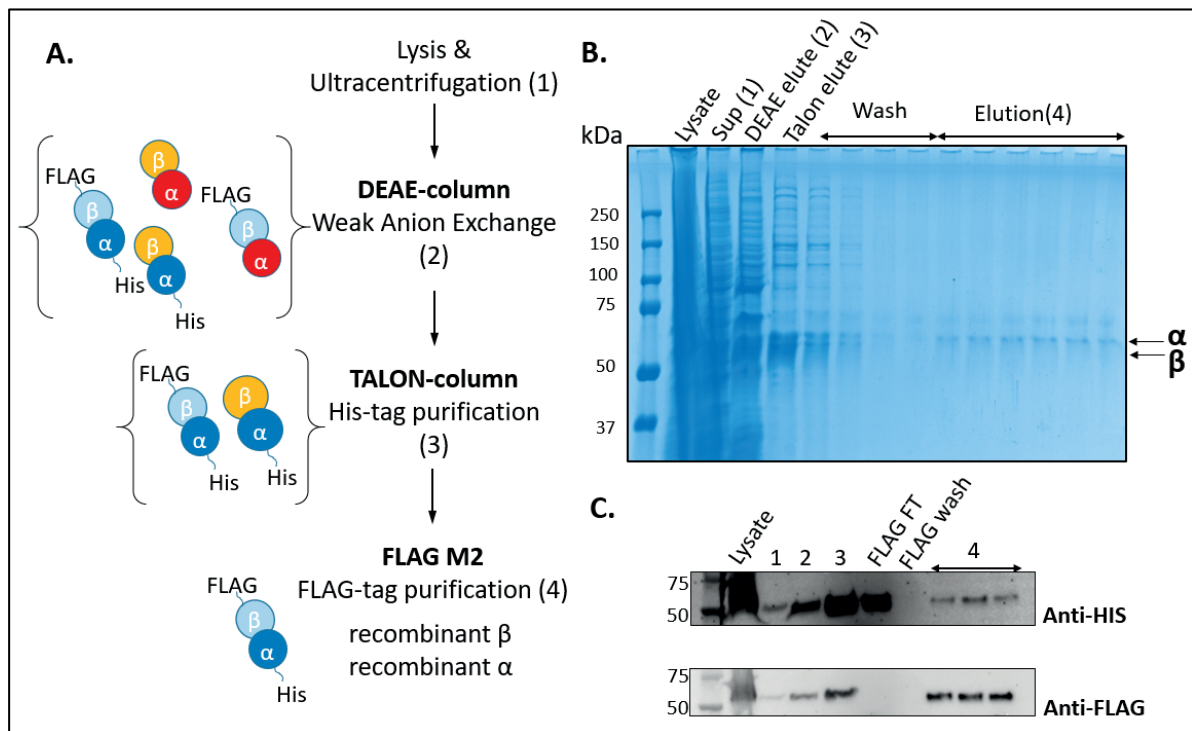


Figure 63. Purification of Recombinant Tubulin Dimer. A) The overall scheme to purify the recombinant tubulin dimer. The types of tubulin species in each step are indicated. B) SDS-PAGE (7.5%) analysis of the anion exchange, HIS tag purification and FLAG tag purification steps. C) Western blot analysis of purified recombinant tubulin. (endogenous β , recombinant β , endogenous α , recombinant α)

Despite several attempts, I continued to observe very poor yield from FLAG purification step (less than 0.5 mg of total protein from starting 1-liter culture). This is perhaps due to poor incorporation of recombinant β -tubulin in the dimer as I saw that most HIS tagged recombinant α -tubulin was lost in the FLAG flow through in the form of dimer with endogenous insect β -tubulin. Deeper investigation into this matter indicated that the placement of N-terminal FLAG tag on recombinant β -tubulin may interfere with GTP binding at the E-site. I observed from literature that the attempts to place the N-terminal biotin tag on β -tubulin in mammalian tubulin resulted in poor yield³⁰¹ and the addition of N-terminal GFP on β -tubulin for *in vivo* imaging experiments³⁰² had been troublesome. For this proof of concept project, my attempt is to modify recombinant α -tubulin using split intein mediated PTS strategy, I decided to skip the FLAG purification step and work with a mixture of recombinant and endogenous β -tubulin containing dimer. I also analyzed the sequence identity between endogenous insect tubulin and human tubulin which showed >90% sequence conservation. Such changes to the purification strategy were made by group of Prof. Tarun Kapoor to circumvent the problems with low yield.²⁹⁴ Thus, in the reworked two-step strategy, I decided to purify tubulin using anion exchange and HIS-affinity tag purification. The tubulin in this fraction would be further enriched by sequential of polymerization and depolymerization steps to generate pure tubulin dimer. This is widely used strategy to purify tubulin from porcine or bovine brain.³⁰³

To this end, I purified the tubulin dimer following same procedure till IMAC purification, which was then concentrated and to 3-6 mg/ml concentration range in BRB80 buffer supplemented with 10% glycerol and 1mM GTP. Protein aggregates were removed from the concentrated protein fraction by high speed (400,000 x G) spin and the resulting supernatant was used for polymerization. The tubulin was polymerized at 30°C for up to 1 hr. and were fractionated into polymerized microtubules and non-polymerized dimers by high speed spin (250,000 x G) for 15 min. I observed that the purified tubulin dimer remained stable during purification, buffer exchange and concentration steps and was not lost as aggregates. By avoiding the FLAG purification step, the yield was significantly higher (4-5 mg of protein from initial starting culture of 1-liter). But, a significant portion of protein did not polymerize and remained in supernatant as seen in Figure 64 B. This was also corroborated by Western blot as shown in **Figure 64 C**. The polymerization of tubulin can further be improved by addition of microtubule stabilizing agent like Taxol and increasing the temperature for polymerization to 37°C. The microtubules can be stored in a GTP containing buffer. Alternatively, microtubules may be depolymerized and resulting dimer can be concentrated to 5 mg/ml, flash frozen with 30% glycerol and stored in -80°C for future use.

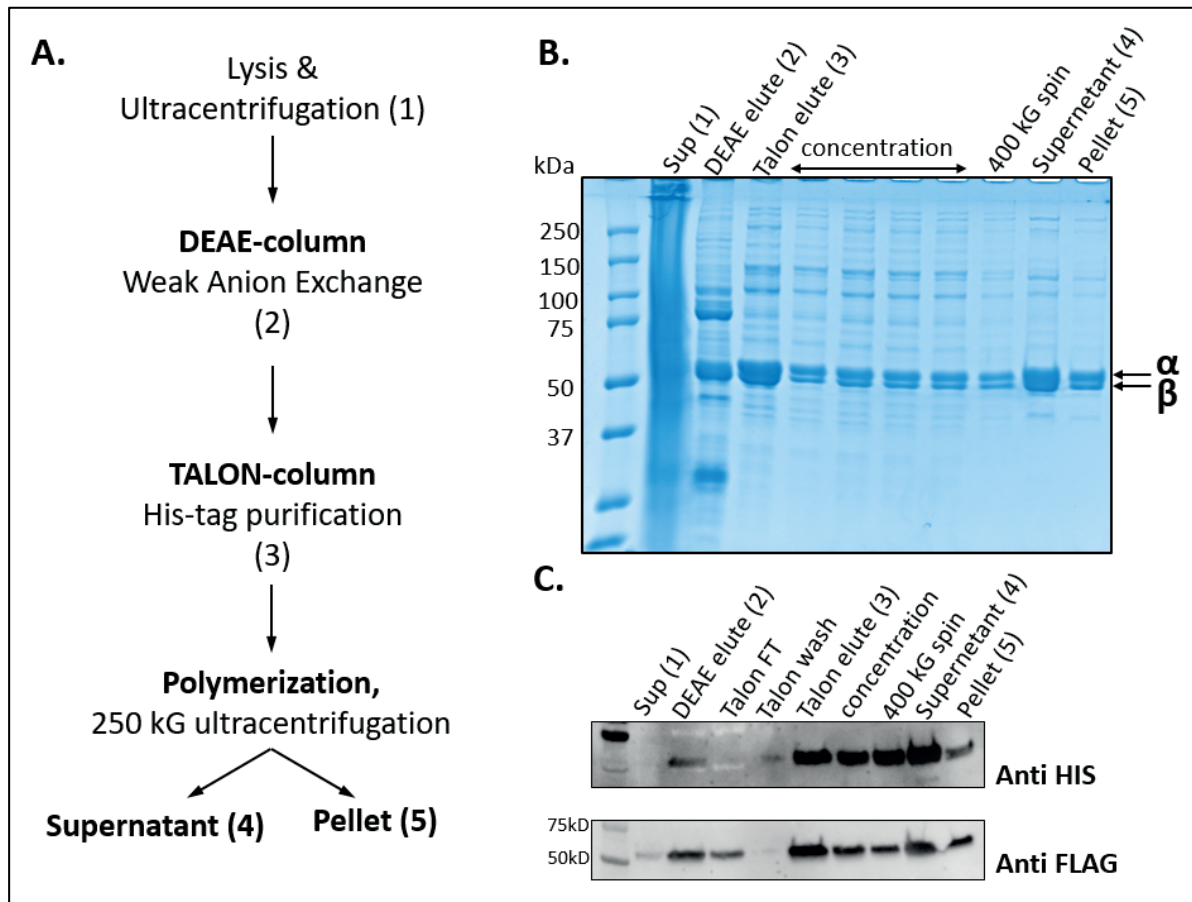


Figure 64. Purification and Polymerization of Recombinant Tubulin. A) The reworked scheme to purify the recombinant tubulin dimer. B) SDS-PAGE (7.5%) analysis of the anion exchange, HIS tag purification and polymerization steps. D) Western blot analysis of purified and polymerized recombinant tubulin.

4.3.6. Purification of Semisynthetic Tubulin Dimer

Having established a strategy to purify recombinant tubulin dimers, I turned my attention to purifying fluorescein tagged semisynthetic tubulin dimers generated by PTS. To this end, I first attempted to purify recombinant tubulin dimer with α -tubulin lacking the C-terminal tail but fused to Ava^N split intein. I followed the same strategy as illustrated in previous section. Unfortunately, tubulin lacking C-terminal tail but fused to Ava^N did not bind to the weak anion exchange column and thus was lost in flow through as seen in **Figure 65**. This can be attributed to loss of multiple negatively charged residue that are present on the tubulin tails due to truncation. Also, a significant portion of α -tubulin fused to Ava^N was lost in initial clearing step during 200000 G ultracentrifugation. The presence of non-native Ava^N may cause protein to misfold and aggregate. Thus, I decided to first generate semisynthetic dimers by PTS in clarified cell lysates and then purify the dimers by anion exchange and HIS-tag purification as mentioned before.

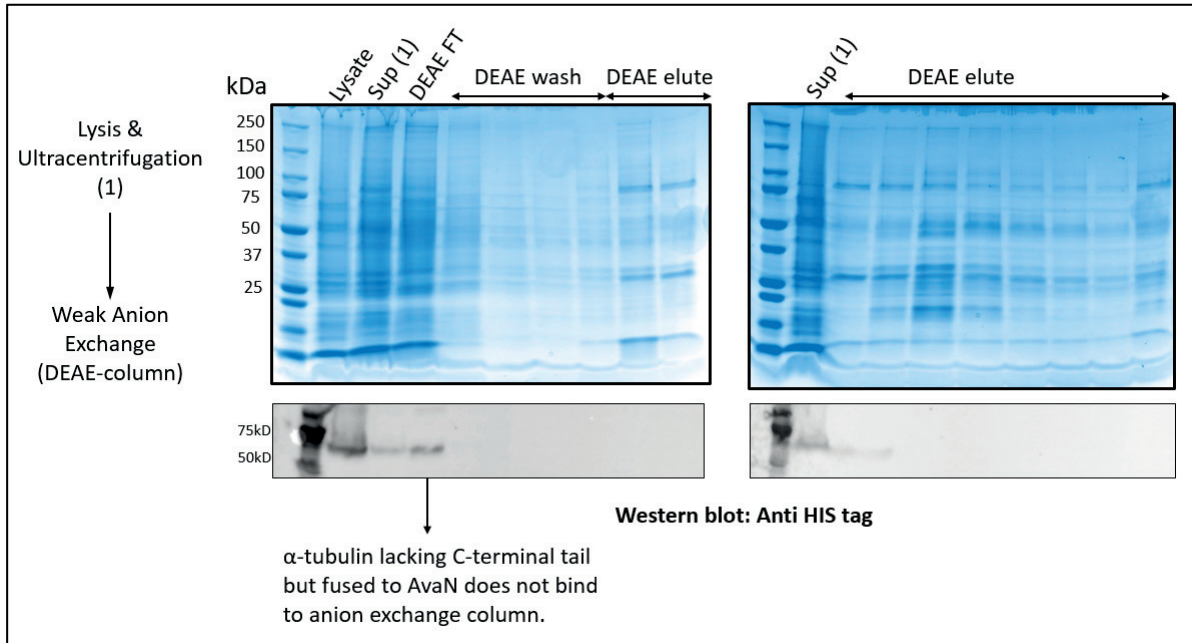


Figure 65. Anion exchange purification for Tubulin-Ava^N construct. The SDS-PAGE gel and western blot analysis for anion exchange purification. The protein does not bind to DEAE column and is lost in the flow through.

To purify semisynthetic tubulin dimer where α -tubulin is modified at C-terminal (labelled with fluorescein), I used twice the amount of cells as compared to purification of unmodified recombinant tubulin dimer. The cell lysate was centrifuged at 200000 G and supernatant was supplemented with 10% glycerol. I added 5 μ M of Npu^C-tubulin tail (fluorescein labelled) peptide to cleared lysates and analyzed the PTS reaction at specified time points for 60 min as shown in **Figure 66**. I was able to observe a fluorescein labelled band appear at \sim 50 kDa, which then saturated in intensity with time.

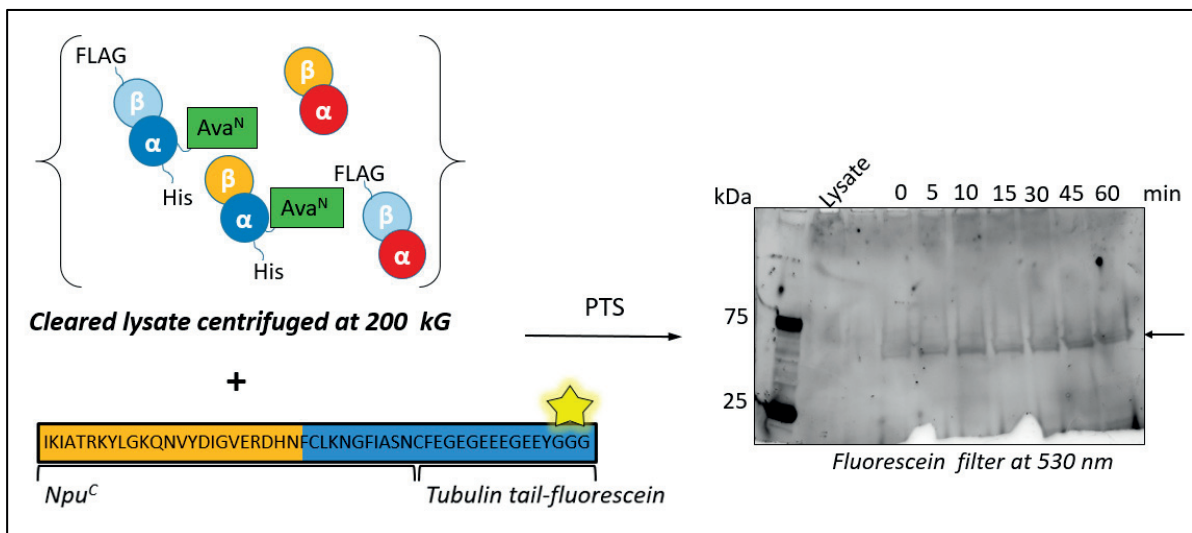


Figure 66. PTS is Cleared Cell Lysates. Types of tubulin dimer species present in cell lysate is illustrated. The progress of PTS reaction to generate fluorescein labelled tubulin analyzed by SDS-PAGE. The gel was scanned at 530 nm. (endogenous β , recombinant β , endogenous α , recombinant α)

The semisynthetic fluorescein labelled tubulin dimer was purified sequential anion exchange and IMAC purification as shown in **Figure 67**. Having spliced the fluorescein labelled

tail on the semisynthetic dimer, it was able to bind to the weak anion exchange column and only a very small fraction of the labelled was lost during the flow through and wash.

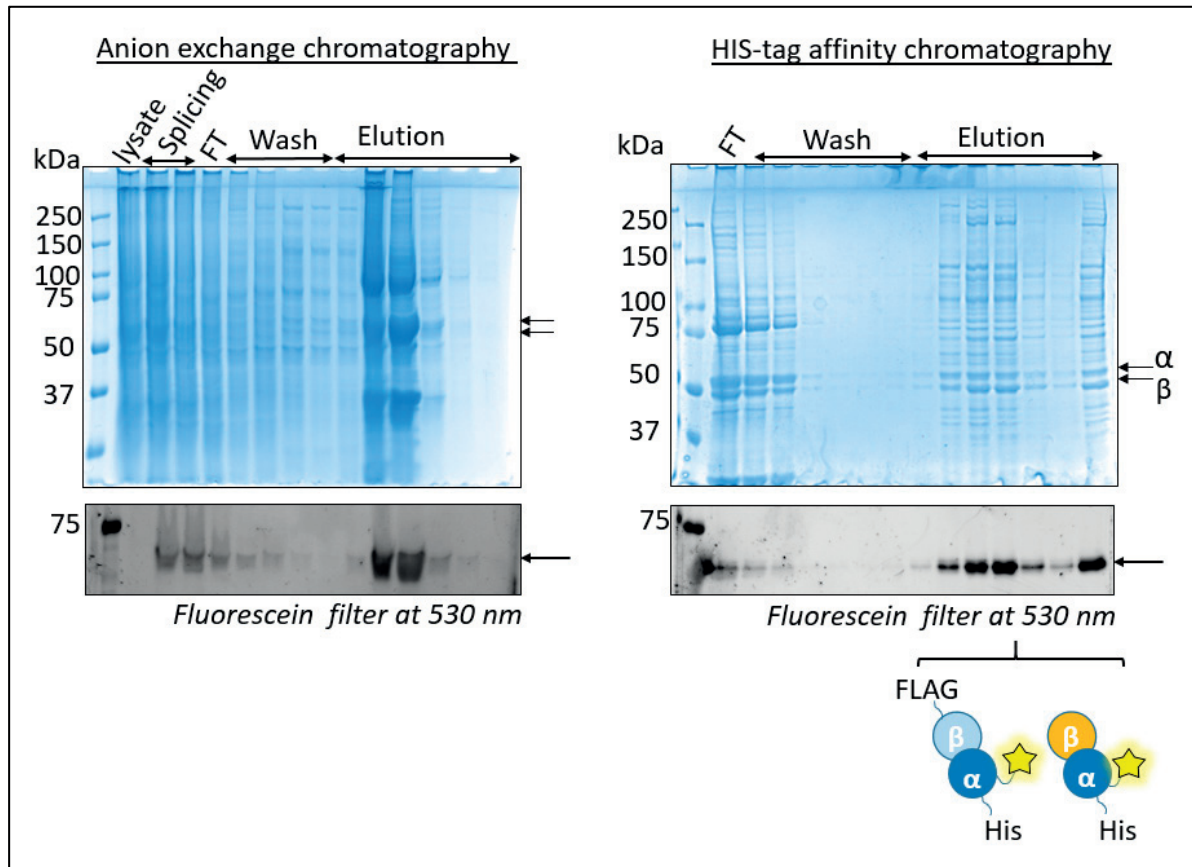


Figure 67. Purification of Semisynthetic Tubulin Dimer. SDS-PAGE (7.5%) analysis of the anion exchange, IMAC purification. The gel was scanned at 530 nm and stained with Coomassie brilliant blue. (endogenous β , recombinant β , recombinant α)

Analyzing the purification profile, I observed enrichment of fluorescein labelled semisynthetic tubulin dimer after IMAC purification. This contains mixture of dimers between formed by recombinant α -tubulin (labelled) and endogenous as well as recombinant β -tubulin. A large fraction of tubulin dimer is lost during IMAC purification in flow through and wash steps. But only a tiny fraction of labelled tubulin is lost as seen from 530 nm scans during both chromatography steps. Due to presence of unnatural intein at C-terminus, the incorporation of recombinant α -tubulin may be less efficient. This also proves as an impediment to purification as the yield and purity is poorer compared to unmodified recombinant tubulin dimer. Nonetheless, it is shown here for the first time that a isotypically uniform tubulin (for α -tubulin), carrying a chemically defined C-terminal modification (fluorescein label) can be generated through split intein mediated PTS.

This result, in its own right, is a breakthrough and opens a door to investigate the roles of specific tubulin PTMs. A further optimization is required to improve yield and purity towards which future steps are discussed in the outlook section.

4.3.7. Microtubule Pelleting and Immunofluorescence Imaging

In order to assess the functionality of the purified functional dimers, I conducted immunofluorescence experiments on microtubules. For this, I used tubulin from porcine brain (purchased from Cytoskeloeton.inc) as a control sample. Microtubules from porcine tubulin were polymerized and spun down on a coverslip. The microtubules were polymerized in the presence of CEP135 (provided by Prof. Pierre Gonczy lab) protein and stained with antibodies against α - tubulin and CEP135 (provided by Prof. Pierre Gonczy lab) as seen in **Figure 68**. The CEP135 is an important centriolar protein and is thought to stabilize centrioles by through interaction with cartwheel and microtubule walls.³⁰⁴ I observed that microtubules polymerized in presence of CEP135 were cross-linked and bundled similar to results obtained by group of Prof. Michel O. Steinmetz.³⁰⁴

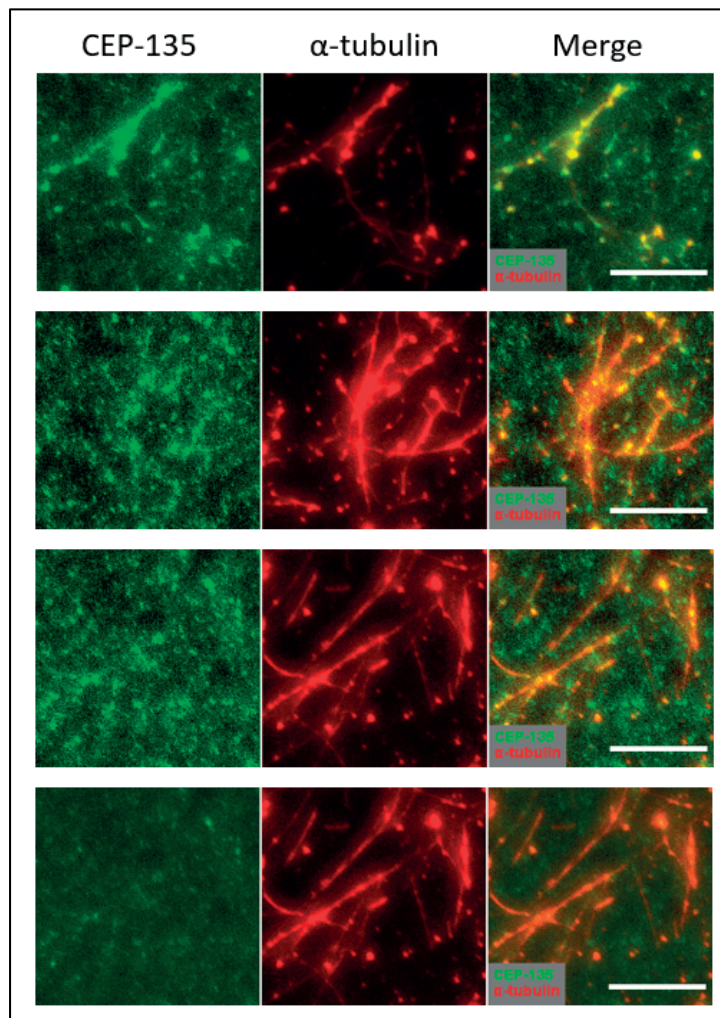


Figure 68. Immunofluorescence Microscopy of Microtubules (from Porcine Tubulin) in the Presence of CEP135. Fluorescence images were taken by staining microtubules with α -tubulin (labelled with Alexa 568-red channel) and CEP135 (labelled with Alexa 488- green channel) antibodies. Scale bars, 10 μ m.

While polymerizing porcine tubulin, I added CEP135 which would bundle and stabilize microtubules thus making it easy to pellet and visualize. To assess the functionality of recombinant tubulin dimer, I decided to do the assay in the absence of CEP135. Microtubules from the purified recombinant tubulin dimer were polymerized and spun down on a coverslip and stained with antibodies against α - and β - tubulin. I was able to observe several microtubules but they appeared less bundled (**Figure 69**) as compared to porcine microtubules polymerized in the presence of CEP135. At this moment, I was not able to achieve sufficient purity and concentration of semisynthetic fluorescein labelled tubulin dimer to perform immunofluorescence microscopy. Thus, next step in that direction would be to improve the purification of semisynthetic tubulin dimer using some alternative strategies that are indicated in the outlook section.

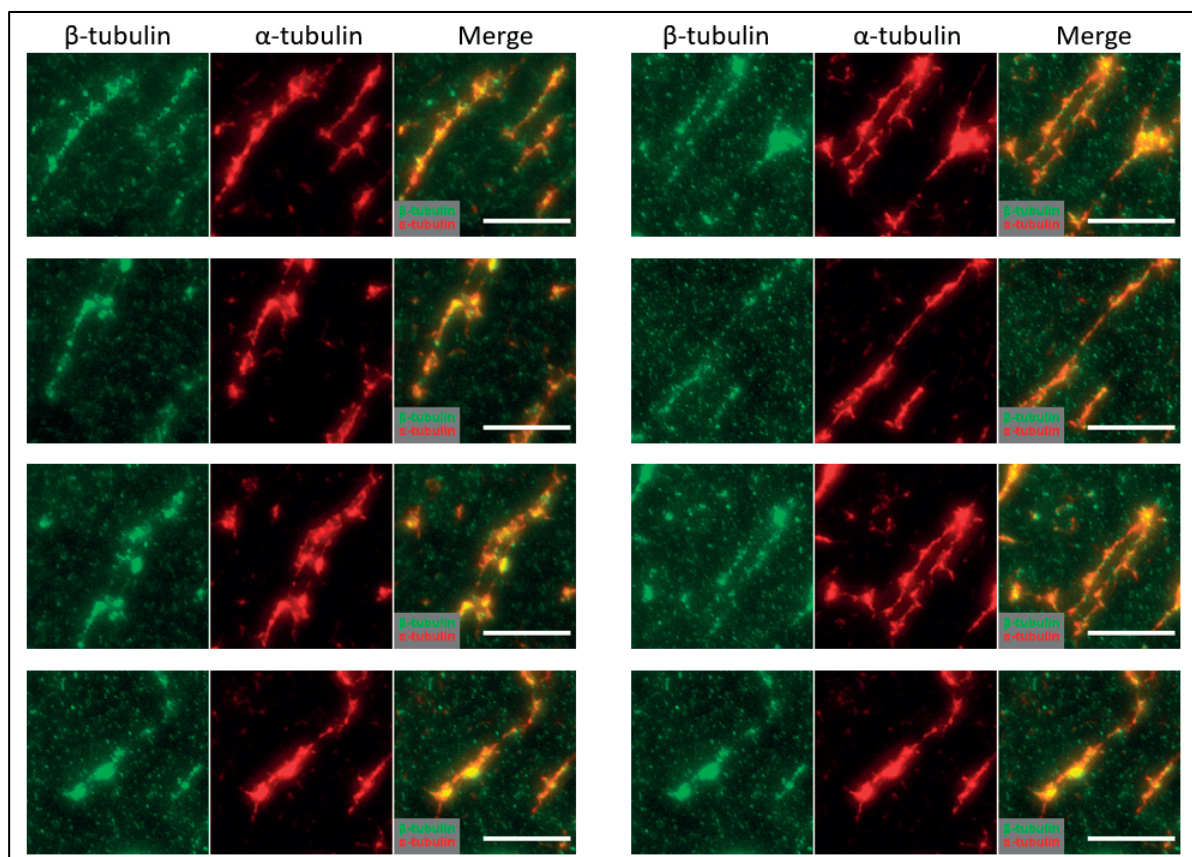


Figure 69. Immunofluorescence Microscopy of Microtubules (from Recombinant $\alpha\beta$ -tubulin). Fluorescence images were taken by staining microtubules with α -tubulin (labelled with Alexa 568-red channel) and β -tubulin (labelled with Alexa 488- green channel) antibodies. Scale bars, 10 μ m.

This is a very quick assay to assess the functionality of purified tubulin dimers and will be used in future, along with other assays, to evaluate the effect tubulin PTMs introduced using semisynthetic method. For semisynthetic tubulin dimer labelled with fluorescein, a further improvement in yield and purity is required and then they can be tested for functionality by polymerization assays.

4.4. Conclusion and Outlook

Both α - and β - tubulin exhibit unstructured C-terminal tails that contain sites for a number of tubulin PTMs. The role of tubulin isoforms and C-terminal PTMs in modulating microtubule function *in vivo* and activity of MAPs remain poorly understood. A detailed investigation of tubulin PTMs is hampered by heterogeneity of tubulin preparation from natural sources.

In this project, I have made significant progress towards the development of a method to chemically access the tubulin C-terminal tails. To this end, I have used a split-intein mediated PTS approach to generate semisynthetic tubulin dimers. I have established a simple PTS assay that allows to verify intein splicing in the context of tubulin without actually expressing full length tubulin dimers which is a cumbersome process. This allows from validation and refinement of method at early stages. Using this, I established a splicing junction in the context of tubulin for given split-intein. Other orthogonal split-inteins can be tested using this assay and can be used to differentially modify individual α - and β -tubulin monomers in the same dimer. To generate semi-synthetically modified tubulin dimers, I have expressed recombinant human tubulin dimer (with and without N-terminal half of split-intein) in insect cells using baculovirus expression system. The MultiBac™ system used for this purpose is modular in nature and thus can be used to express variety of tubulin dimers containing combinations of multiple tubulin isoforms. I have successfully purified the recombinant tubulin dimer using either two or three step process. However, the yield with FLAG-tag affinity purification have been low. The purification of semisynthetic tubulin labelled with fluorescein has been more challenging. Nonetheless, from the results, I have shown for the first time, the generation of the semi-synthetic tubulin dimer containing defined tubulin isoform and carrying a chemically defined modification. A further improvement is required with regards to yield and purity of semisynthetic tubulin dimer.

One of the major problem encountered during the purification is with the N-terminal purification tags on tubulin. This is required in order to purify the tubulin after trans-splicing. Though α -tubulin may tolerate the presence of N-terminal tag; this appears to be a significant issue for β -tubulin due to possible interference with the GTP binding E-site on β -tubulin. This may lead to less efficient incorporation of β -tubulin into the dimer as well as microtubule polymerization. I also observed degradation of β -tubulin over time (see Figure 58) and thus I was required to harvest cells at 36 hrs. instead of 72 hrs. This may explain the low yields with FLAG-tag purification as I have less than optimal starting material. Thus, in our group, we are now working of expressing tubulin dimer with placement of the purification tags on an internal loop as shown in **Figure 69 A**. We will also express β -tubulin with a protease cleavable C-terminal STREP-tag as described recently by group of Prof. Tarun Kapoor.²⁹⁵ Having generated fluorescein labelled semi-synthetic tubulin dimers, the next logical step will be to generate polyglutamylated tubulin dimer with a defined number of glutamates. To this end, we will

generate a polyglutamylated tubulin tail peptide (**Figure 69 B**) using strategy described by Tegge et al.³⁰⁵ We will use these modified tubulin to analyze the role of polyglutamylation in microtubule stability, polymerization-depolymerization dynamics and binding interactions with number of binding partners employing Total Internal Reflection Microscopy (TIRF) (**Figure 69 C**).

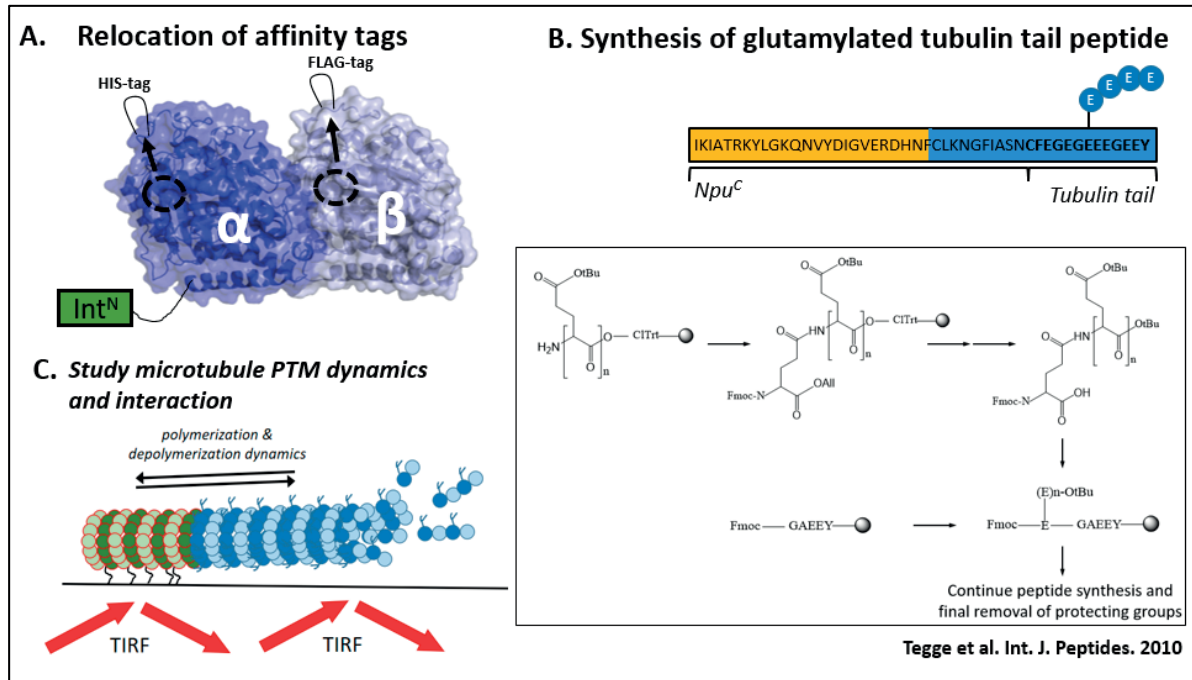


Figure 70. Project Outlook. A) Recombinant expression of tubulin dimers with relocated affinity tags to an internal loop. B) Synthesis of branched polyglutamylated tubulin tail peptide. C) Schematic representation of TIRF microscopy experiment to measure microtubule polymerization and depolymerization dynamics.

5. Thesis Summary

The function of two of the most important cellular complexes, chromatin and microtubules, is influenced and very tightly regulated by underlying protein PTMs. The focus of my thesis was to develop tools to investigate the role of protein PTMs.

In bivalent domains, a chromatin signature prevalent in embryonic stem cells (ESCs), histone H3 methylated at lysine 4 (H3K4me3)- an activating histone PTM mark, coexists with H3K27me3- a repressive histone PTM mark, in asymmetric nucleosomes. In the first collaborative project, my colleague, Dr. Carolin Lechner established a synthetic strategy to generate asymmetrically modified nucleosome. Using these asymmetrically modified nucleosome, I probed the effect of bivalent nucleosome modifications in regulating the activity of a key developmental enzyme PRC2. Towards this end, I optimized a protocol based on gel based fluorography method using radiolabeled (tritium) substrate. I show that in bivalent nucleosomes, H3K4me3 inhibits the activity of the H3K27-specific lysine methyltransferase (KMT) Polycomb Repressive Complex 2 (PRC2) solely on the same histone tail. Whereas H3K27me3 stimulates PRC2 activity via a positive feedback mechanism across tails, thereby partially overriding the H3K4me3-mediated repressive effect.

Both α - and β - tubulin consist of a compactly folded tubulin body and highly negatively charged unstructured C-terminal region - the tubulin tails. The tubulin C-terminal tails contain sites for multiple PTMs which play important role in microtubule dynamics and functions. For detail investigation of tubulin PTMs, isotypically pure tubulin dimer carrying chemically defined PTMs is required. In the second project, I have developed a protein engineering approach to link synthetic, modified tails to recombinant tubulin by use of a split intein based, protein trans-splicing (PTS) approach. I have demonstrated the validity of this approach by linking a fluorescently modified tubulin tail to recombinant human tubulin, preparing semisynthetic tubulin dimers. Extending the method to native PTMs will open the door for an in-depth study of tubulin modification using a chemically defined system.

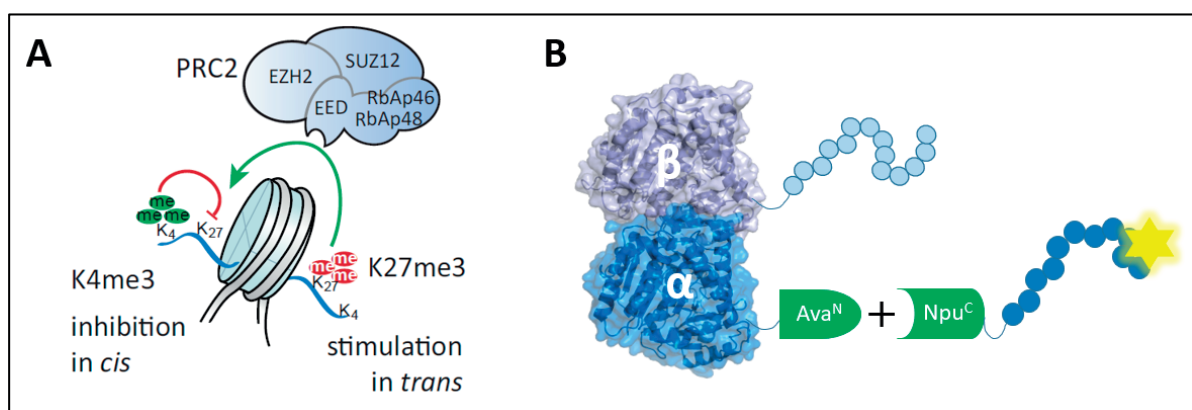


Figure 71. Thesis Summary. A. Model for PRC2 Regulation by Asymmetrically Modified Nucleosomes. **B.** An approach to generate semisynthetic tubulin dimer carrying chemically defined C-terminal modification of α -tubulin using split-intein based protein splicing strategy.

Protein PTMs play important role in regulating many cellular processes. But, functional investigation of PTMs remains a challenging task. During these projects, I have made contributions towards developing approaches to study the functional role of PTMs. These protein engineering approaches have broad range of applications for variety of molecular systems. We have established a novel and robust approach to generate asymmetrically modified nucleosomes which play an important role in the bivalent domains seen in embryonic stem cells and cancer cells. The synthetic approach and the biological *in vitro* assay established during this project can help us understand the regulation of writer enzymes at the bivalent domains. The approach has since been extended to generate various combinations of asymmetric nucleosome modifications. On the other hand, I have successfully demonstrated application of intein-based protein splicing approach to modify tubulin dimer. The complex and dynamic nature of tubulin combined with difficulties in recombinant expression of tubulin dimer posed significant challenges to investigate the role of tubulin PTMs. Recombinant expression of human tubulin dimer fused to split-intein opens the door to generate chemical access to tubulin C-terminal tails. By demonstrating the validity of this approach by generating fluorescently labelled semisynthetic tubulin dimers, I have set up a method that would make significant impact to the tubulin PTM field. We continue to further optimize the method to improve the yield of semisynthetic tubulin. The next big challenge in this approach would be to generate a library of polyglutamylated tubulin dimers. This would be a very important step in understanding the regulation of centrioles in which the microtubules are highly polyglutamylated but the exact nature and function of this modification remains unknown. Due to the modular nature of this method, it has a large potential to explore the interplay between large number of tubulin isotypes and combinations of tubulin PTMs, thus allowing for systematic investigation of the Tubulin Code.

6. Materials and Methods

6.1. Reagents and Instrumentation

All commonly used chemical reagents and solvents were purchased from Sigma-Aldrich Chemical Company (Steinheim, Germany), Fischer Scientific (Fair lawn, NJ, US/Loughborough, UK) or Applichem (Darmstadt, Germany).

Chemically competent DH5 α , Rosetta(DE3), BL21(DE3) and BL21(DE3) plysS were from Novagen (Darmstadt, Germany) and used to generate stocks of competent cells.

Ampicillin, phenylmethylsulfonyl fluoride (PMSF), lysozyme and isopropyl- β -D-thiogalactoside (IPTG) were purchased from BioChemica Applichem (Darmstadt, Germany). Kanamycin was purchased from Roth (Karlsruhe, Germany) or Acros (Geel, Belgium).

T4 DNA ligase, restriction enzymes, Phusion DNA polymerase, Q5 DNA polymerase, Cre-recombinase, Taq ligase, T5 exonuclease, DNA ladders, DNA loading dyes and dNTPs were purchased from New England Biolabs (Ipswich, MA, US). Distributed through BioConcept (Allschwil, Switzerland).

Primers were ordered from and synthesized by Integrated DNA technologies (Leuven, Belgium). Gene sequencing was performed at GATC Biotech (Constance, Germany). Single-point mutations were done using a QuikChange II XL site-directed mutagenesis kit from Agilent (Basel, Switzerland).

Materials for hand-casting of agarose and SDS-PAGE gels (agarose, TEMED, APS, acrylamide), analysis of SDS-PAGE gels (Precision Plus ProteinTM All blue standards and Precision Plus ProteinTM dual color standards) and pre-cast Criterion 5% TBE gels were purchased from BioRad (Hercules, CA, US). Immuno-Blot[®] polyvinylidene difluoride (PVDF) Membrane for Protein Blotting and ClarityTM Western ECL Substrate were purchased from Bio-Rad (Hercules, CA, USA).

Slide-A-LyzerTM dialysis cassettes, Slide-A-LyzerTM Mini Dialysis buttons (3500 Da MWCO), Slide-A-LyzerTM MINI dialysis devices (MWCO 10'000 Da) and protein G magnetic beads were from Thermo Scientific (Rockford, IL, USA). The centrifugal concentrators vivaspin 500 (10'000 MWCO, PES membrane) were purchased from Sartorius Stedim (Goettingen, Germany).

QIAquick spin column for PCR purification, gel extraction and nucleotide removal as well as QiaPrep spin columns for miniprep plasmid purification were from Qiagen (Hilden, Germany). The Ni-NTA resin and QIAquick PCR Purification Kit were purchased from Qiagen (Hombrechtikon, Switzerland). Pure Link RNase A were purchased from Invitrogen (Thermo Scientific, Rockford, IL, USA).

The IgG, anti- H3K4me3 and anti-H3K27me3 antibodies used for CHIP as well as the anti-Histone H3 antibody were from Abcam (Cambridge, UK). The anti-rabbit HRP conjugate was from Perkin Elmer (Zürich, Switzerland).

Penicillin/streptomycin was purchased from Bioconcept (Allschwil, Switzerland). Express five SFM insect cell medium, SF900 II SFM insect cell media, High Five insect cells, Sf9 insect cells, L-glutamine (200mM) stock were purchased from Life Technologies (Thermo Scientific, Rockford, IL, USA).

DEAE Sepharose Fast Flow resin was purchased from GE Healthcare. TALON Metal Affinity resin was purchased from Clontech. Anti-FLAG M2 Affinity Gel was purchased from Sigma-Aldrich. Centrifugal filter (Amicon Ultra 15 - 30k, Millipore) was purchased from Millipore.

Amino acid derivatives, 2-chlorotriyl chloride resin and 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) were purchased from Novabiochem, Merck (Darmstadt, Germany). O-(Benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HBTU) was purchased from Protein Technologies (Manchester, United Kingdom). N,N-Dimethylformamide (DMF), N,N-diisopropylethylamine (DIPEA) and piperidine were from Acros Organics (Geel, Belgium).

Bovine serum albumin (BSA), D,L-Dithiothreitol (DTT), trifluoroacetic acid (TFA), tris(2-carboxyethyl)phosphine hydrochloride (TCEP), sodium nitrite (NaNO₂), L-Glutathione reduced, Triton X-100, sodium fluoride (NaF), IGEPAL CA-630, β-mercaptoethanol, triisopropylsilane (TIS), 2,2,2-Trifluoroethanethiol (TFET), anti-Flag monoclonal M2 antibody mouse, anti-mouse IgG peroxidase Conjugate were from Sigma Aldrich (Taufkirchen, Germany). 2,2'-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) was purchased from Wako Pure Chemical Industries, Ltd. (Osaka, Japan).

Bacterial cells for recombinant protein expression were grown in an HT infors AG incubator, recovered with an Avanti J-20 XPI centrifuge from Beckman Coulter and sonicated using a Vibra-cell VCX 750 Sonics & Materials sonicator.

Size exclusion chromatography and ion exchange purification was performed on an AKTA Pure FPLC system from GE Healthcare. Size-exclusion were done using a S200 10/300GL or S75 10/300 column from GE Healthcare. Cation exchange and anion exchange purification was done using HiTrap SP HP (5 ml) and HiTrap Q FF (1 ml) traps from GE Healthcare.

Analytical reversed-phase HPLC (RP-HPLC) was performed on an Agilent 1260 series instrument with an Agilent Zorbax C18 column (5 μm, 4.6 x 150 mm), employing 0.1% TFA in water (RP-HPLC solvent A), and 90% acetonitrile, 0.1% TFA in water (RP-HPLC solvent B), as the mobile phases. Typical analytical gradients were 0-70% solvent B over 30 min at a flow rate of 1 ml/min. Preparative scale purifications were conducted on an Agilent 1260 preparative HPLC system. A Zorbax C18 preparative column (7 μm, 21.2 x 250 mm) or a semi-preparative column (5 μm, 9.4 x 250 mm) was employed at a flow rate of 20 mL/min or 4 mL/min, respectively. ESI-MS analysis was conducted on a Shimadzu MS2020 single quadrupole instrument connected to a Nexera UHPLC system. Absorbance spectra were recorded with an Agilent 8453 UV-Vis spectrophotometer.

SDS-PAGE, native PAGE and agarose gels were imaged using a ChemiDoc MP imaging system from BioRad. Manual peptide synthesis and reactions on solid-phase were carried out in reaction vessels from Peptides International and automated peptide synthesis done on a Tribute instrument from Peptides International Inc.

6.2. Methods

Section 5.2.1. to 5.2.4. are based on following research article of which I am one of the authors as highlighted below,

Lechner CC., Agashe N., Fierz B., *Traceless synthesis of asymmetrically modified bivalent nucleosomes*, *Angew Chem Int Ed Engl* 2016, 55(8):2903-6

6.2.1. Octamer Refolding

- Lyophilized recombinant histone proteins (H2A, H2B and H4) as well as the synthetic H3K4me3, H3K27me3 and the disulfide-cross linked xlnCH3 heterodimeric proteins were dissolved in unfolding buffer (6 M GuHCl, 10 mM Tris-HCl, pH 7.5).
- For refolding of xlnCH3-octamers, 1 equivalent of xlnCH3 dimer was mixed along with 2 equivalents of H4 and 2.2 equivalents of H2A and H2B at a final concentration of 0.5 mg/mL in unfolding buffer.
- For refolding of unmodified octamers and octamers containing synthetic H3K4me3 and H3K27me3, equimolar amounts of the H3 variant and hH4 were mixed along with 1.1 equivalents of H2A and H2B at a final concentration of 1 mg/mL in unfolding buffer.
- The protein solution was transferred to a Slide-A-Lyzer™ dialysis cassette (7k MWCO) and dialyzed against refolding buffer (2 M NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 7.5).
- The refolded octamers were removed from the dialysis cassette and subsequently purified by size exclusion chromatography on a Superdex S200 10/300GL column.
- Collected fractions were analyzed by SDS-PAGE, octamer containing fractions were pooled and concentrated to 20-80 μM octamer concentration. After addition of glycerol to a final concentration of 50 %, octamers were stored at -20 °C.
- Purified xlnCH3 octamers were finally analyzed via SDS-PAGE under reducing and non-reducing conditions.

6.2.2. Nucleosome Reconstitution

- Mononucleosomes were typically reconstituted in 50 μL sample volume by mixing 75 pmol of the 153 base pair fragment of the '601' nucleosome positioning DNA sequence²⁸⁸ with the respective histone octamer at high salt conditions followed by gradient dialysis into low salt conditions (10 mM KCl, 10 mM Tris-HCl, 0.1 mM EDTA, pH 7.5).
- Dialysis was carried out in Slide-A-Lyzer™ MINI dialysis devices using a two channel peristaltic pump (1 mL/min) over 16 hr at 4 °C.
- Upon reduction of xlnCH3 in the formed nucleosomes with DTT, TEV-protease was added (1 μg per 60 pmol nucleosomes) for 6 h at 4 °C to remove the *Inc*-tag.
- Nucleosome concentrations were determined by UV quantification and the quality of nucleosome reconstitution was assessed by native gel electrophoresis (0.8 % agarose).
- Removal of the *Inc*-tag by TEV protease treatment was analyzed by SDS-PAGE (15 % polyacrylamide gel).

6.2.3. PRC2 Methyltransferase Assay

- In a typical histone methyltransferase assay, 1 pmol of recombinant PRC2 complex (EZH2, SUZ12, EED, RbAp46/48) was incubated with 15 pmol of the respective nucleosome substrate and 3H-SAM (40 μ Ci of 3H) in 50 mM Tris-HCl, 10 mM NaCl, 2 mM MgCl₂, 1 mM EDTA, 5 mM DTT, pH 8.6 at room temperature for 2 h.
- Subsequently, proteins were separated by SDS-PAGE (13 % polyacrylamide gel) and the gel was stained with Coomassie Brilliant Blue G250.
- The gel was then soaked in Amersham Amplify Fluorographic Reagent for 30 min and dried at 80 °C for 2 h.
- Incorporation of ³H in histone substrates was detected by exposure to an X-ray film at -80 °C. PRC2 methyltransferase activity was assessed by densitometry analysis of bands.

6.2.4. Immunoprecipitation of Nucleosomes

- Immunoprecipitation of reconstituted mononucleosomes was performed with Dynabeads[®] Protein G (0.5 mg Dynabeads per sample). Dynabeads were transferred to a 1.5 mL tube and supernatant was separated from the beads with the help of a magnet.
- The beads were washed twice with PBS + 0.02 % Tween-20, pH 7.4 and the buffer was removed from the beads. 4 μ L of α -H3K27me3 antibody (1 mg/mL) were diluted into 200 μ L PBS + 0.02 % Tween-20, pH 7.4, added to the beads and incubated for 1 h at room temperature with rotation to allow binding of the antibody to Dynabeads[®] Protein G.
- Subsequently, the supernatant was removed and the beads were washed 3 times with PBS + 0.02 % Tween-20, pH 7.4. The beads were resuspended in 0.2 M triethanolamine in PBS (pH 8.2) and incubated for 10 min at room temperature with rotation to equilibrate the beads.
- Cross-linking of α -H3K27me3 antibody to Dynabeads[®] Protein G was achieved by addition of a solution of 25 mM dimethyl pimelinediimidate dihydrochlorid (DMP) in 0.2 M triethanolamine in PBS, pH 8.2. After 30 min incubation with rotation, beads were washed twice with 0.2 M triethanolamine in PBS, pH 8.2.
- The crosslinking procedure was repeated two more times. Possibly residual DMP was quenched by addition of 50 mM ethanolamine in PBS for 5 min followed by washing twice with PBS.
- The quenching procedure was repeated one more time. Finally, the beads were washed twice with 1 M glycine, pH 3 for 10 min with rotation in order to remove not-cross-linked antibody.

- For immunoprecipitation, the beads were equilibrated in 10 mM Tris-HCl, 100 mM NaCl, 0.02 % Tween-20, pH 7.5 for at least 30 min. A sample of the respective nucleosomes (40 pmol in 150 μ L buffer) were added to the prepared α -H3K27me3-Dynabeads and incubated for 18 h at 4 °C with rotation.
- The flow through was collected and the beads were washed 6 times with 10 mM Tris-HCl, 100 mM NaCl, 0.02 % Tween-20, pH 7.5. Elution of bound nucleosomes from the α -H3K27me3-Dynabeads was achieved with 15 μ L elution buffer (50 mM Tris-HCl, pH 6.8, 25 % glycerol, 2 % SDS, 0.01 % bromophenol blue) at 70 °C for 10 min.
- The elution fractions were collected from the beads and directly used for SDS-PAGE (13 % polyacrylamide gel) separation of proteins.
- Western blot analysis was performed using a discontinuous buffer S27 system with the Trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell. The PVDF membrane was soaked in methanol for 2 min and subsequently equilibrated in anode buffer (60 mM Tris, 40 mM CAPS, pH 9.6; 30 % MeOH) for 30-45 min.
- The polyacrylamide gel was equilibrated in cathode buffer (60 mM Tris, 40 mM CAPS, pH 9.6; 0.1 % SDS) for 5-10 min.
- Transfer was prepared by assembling membrane and gel in a sandwich of filter paper soaked in anode and cathode buffer, respectively. Proteins were transferred to the membrane at 20 V for 12 min.
- After transfer, the membrane was blocked with 5 % BSA in TBS-T (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 % Tween-20) for 2 h at room temperature followed by incubation with α -H3K4me3 antibody (1:2000 in 5 % BSA in TBS-T) over night.
- The membrane was washed three times for 3 min with TBS-T and incubated with α -mouse-peroxidase antibody (1:6000 in 5 % BSA in TBS-T) for 1 hr. at RT.
- After washing with TBS-T (3x 3 min), the membrane was incubated in Clarity™ Western ECL Substrate for 5 min and imaged using the ChemiDoc MP imaging system from BioRad.

6.2.5. Gibson Assembly Cloning

- Gibson assembly mix for isothermal cloning was prepared as described in literature.³⁰⁶ To this end, 5x isothermal buffer (0.5 M Tris, 1 mM each dNTP, 50 mM MgCl₂, 50 mM DTT, 5 mM NAD, 0.025% (w/v) PEG6000), 167 μ L water, 3.0 U T5 DNA exonuclease, 18.5 U Phusion DNA polymerase and 2000 U Taq DNA ligase were mixed and fractionated into 15 μ L aliquots, flash frozen in liquid nitrogen and stored at -20°C.
- 50-100 ng of linearized vector backbone and equimolar amounts of insert were to 10 μ L water and 5 μ L was added to 15 μ L Gibson assembly aliquot.
- This was incubated for 60 min at 50°C and 5 μ L of mixture was for transformation of appropriate bacterial strain.

6.2.6. Cloning, Expression and Purification of Tubulin Adapter Proteins

- The gene fragments for N-terminal and C-terminal adapter proteins used to validate PTS site were generated by PCR.
- The gene fragment of AvaN split intein was ordered as G-block fragment from Integrated DNA Technologies (IDT, Leuven, Belgium).
- The fragments were ligated and inserted into pET15 vector (contain Ampicillin resistance selection marker) using Gibson assembly cloning.
- The proteins were expressed in *E. coli* PLYS cells grown in LB media by the addition of 0.5 mM IPTG.
- In general, cells were induced at 37 °C for 4 h. Cell pellets were resuspended in lysis buffer consisting of a solution of 50 mM NaH₂PO₄ Ph 7.5, 300 mM NaCl, 10 mM imidazole, 1 % v/v IGEPAL CA-630, 10 % w/v glycerol, 0.5 mM PMSF, 0.5 mM β-mercaptoethanol, 1x protease inhibitor cocktail.
- Cells were then lysed by addition of lysozyme to a final concentration of 1 mg/mL followed by sonication. The lysates were clarified by centrifugation (20000 G for 25 min) and the supernatants were applied to Ni-NTA affinity purification.
- The resin-bound proteins were washed three times with a wash buffer consisting of 50 mM NaH₂PO₄ pH 7.5, 300 mM NaCl, 50 mM imidazole, 0.5 mM β-mercaptoethanol. The proteins were eluted with a buffer consisting of 50 mM NaH₂PO₄ pH = 7.5, 150 mM NaCl, 250 mM imidazole, 1 mM β mercaptoethanol.
- The Ni-NTA elution fractions were combined and further purified by size exclusion chromatography on a Superdex S200 10/300GL column and the protein were eluted in splicing buffer (50 mM Tris-HCl, pH 7.0, 300 mM NaCl, 1 mM DTT and 10% glycerol).
- Collected fractions were analyzed by SDS-PAGE, octamer containing fractions were pooled and concentrated to 50 μM octamer concentration, flash frozen in liquid nitrogen and stored in -80°C.

6.2.7. Protein Trans Splicing Assays with Tubulin Adapter Proteins

- The N-intein and C-intein adapter proteins were diluted to concentration of 15 μM and 20 μM respectively in splicing buffer and mixed together at room temperature or 4°C.
- The reaction was quenched at specified time points by mixing 20 μl of reaction with 3X SDS-PAGE gel loading buffer and samples were boiled for 2 min.
- The gel samples were analyzed on by SDS-PAGE and stained with Coomassie Brilliant Blue G250.
- Alternatively, the reaction was quenched by removing mixing sample at specified time points with 3:1 (v/v) quenching solution (8 M guanidine hydrochloride and 4% trifluoroacetic acid) and analyzed by analytical RP-HPLC on C18 column recording absorbance at 214 nm.

6.2.8. Cloning and Expression of Recombinant Tubulin Dimer using Baculovirus Expression System

- The tubulin genes were kindly provided by group of Dr. Carsten Janke. The baculovirus expression vectors (Acceptor vectors, donor vectors and DH10MultiBac™ cells were kindly provided by group of Prof. Thomas Schalch.
- The α -tubulin gene (with and without Ava^N split intein) was cloned into pACEBac1 (Gentamycin resistance marker) acceptor plasmid using Gibson assembly cloning and was used to transform *DH5 α* *E. Coli* cells.
- The β -tubulin gene (with and without Ava^N split intein) was cloned into pDK (kanamycin resistance marker) donor plasmid using Gibson assembly cloning and was used to transform BW23473 *E. coli* strains expressing *pir* gene.
- The plasmids were recovered using QiaPrep miniprep kits.
- To generate dual gene assembly, acceptor and donor plasmids were recombined using Cre-lox recombination where 50 ng of each plasmid was mixed with 10X Cre recombinase buffer and 1 μ L of Cre Recombinase enzyme.
- The mixture was incubated at 37°C for 30 min and then was used to transform *DH5 α* *E. Coli* cells selected on LB agar plate containing gentamycin and kanamycin antibiotics.
- The recovered dual gene transfer vector was used to transform DH10 MultiBac™ carrying Bacmid.
- The genes were inserted into the Bacmid through Tn7 transposition sites. The successful transposition of inset leads to disruption of LacZ gene in the Bacmid. The transformed DH10 MultiBac™ cells were plated on LB agar plate containing 50 μ g/mL kanamycin, 7 μ g/mL gentamicin, 10 μ g/mL tetracycline, 100 μ g/mL Bluo-gal, and 40 μ g/mL IPTG and incubated at 37°C for 48 hrs.
- White colonies were selected and single colonies were restreaked on same plates for verification. The Bacmid was isolated by isopropanol precipitation.
- The positive inserts were verified using PCR using pUC M13 primers as described in MultiBac user manual.²⁹⁷
- To generate expression virus, SF9 insect cells were grown to density of 1.5-2x10⁶ cells/ml in Sf 900 II SFM insect cell media without any antibiotics.
- In a 6 well plate, 8x10⁵ cell were seeded in each well in Grace's insect cell media supplemented with 10% FBS.
- For a single well, 8 μ L Cellfectin II™ was mixed with 100 μ L Grace's insect cell media and kept for 30 min at room temperature to allow micelles to form.
- 1 μ g of Bacmid DNA was diluted in 100 μ L Grace's insect cell media and mixed with diluted Cellfectin II™ and kept for 30 min at room temperature.

- 210 μ L of DNA-lipid mixture was added dropwise to each well to transfect the cells and incubated for 5 hrs.
- The transfection media was removed and 2 ml Sf 900 II SFM insect cell media was added to each well and cells were incubated at 27°C for 72hrs.
- The P1 viral stock was isolated by spinning down cells 2000 G for 5 min. The viral stock was amplified by infecting Sf9 cells in suspension culture to generate P2 and P3 stock and were stored at 4°C in dark for up to 6 months.
- The viral titer was determined by a plaque assay (5×10^8 pfu/ml).²⁹⁷
- To express recombinant tubulin dimer, HighFive insect cells were grown in Express Five SFM media supplemented with 18 mM L-Glutamine to a density of $1.5-2 \times 10^6$ and were infected with P3 viral stock with MOI 20.
- The cells were incubated in suspension culture at 27°C at 100 RPM for 36 hrs. The cells were harvested by centrifugation at 2000 G for 5 min.
- The expression of recombinant HIS- and FLAG- tagged tubulin was determined by analyzing overexpression on 15% SDS-PAGE gels and Western blot.

6.2.9. Purification of Recombinant Tubulin Dimer

For purification of recombinant tubulin dimer, following buffers are required

1. HBS: 30 mM HEPES-KOH (pH 7.5), 150 mM NaCl
2. PMI: 100 mM PIPES-KOH (pH 6.8), 10 mM MgSO₄, 2 mM EGTA
3. PMI with glycerol: 100 mM PIPES-KOH (pH 6.8), 10 mM MgSO₄, 4 mM EGTA, 10 % (v/v) glycerol, 0.04 % (v/v) NP-40
4. DEAE WASH: 100 mM PIPES-KOH (pH 6.8), 60 mM NaCl, 10 mM MgSO₄, 10 % (v/v) glycerol
5. DEAE ELUTION: 100 mM PIPES-KOH (pH 6.8), 400 mM NaCl, 10 mM MgSO₄, 10 % (v/v) glycerol
6. TALON WASH: 100 mM PIPES-KOH (pH 7.0), 300 mM NaCl, 5 mM MgSO₄, 10 % (v/v) glycerol
7. TALON ELUTION: 100 mM PIPES-KOH (pH 7.0), 300 mM NaCl, 5 mM MgSO₄, 10 % (v/v) glycerol, 250 mM imidazole-HCl (pH 7.0)
8. FLAG WASH: 100 mM PIPES-KOH (pH 7.0), 150 mM NaCl, 5 mM MgSO₄, 2 mM EGTA, 10 % (v/v) glycerol, 0.02 % (v/v) NP-40
9. FLAG Elution: 100 mM PIPES-KOH (pH 7.0), 150 mM NaCl, 5 mM MgSO₄, 2 mM EGTA, 10 % (v/v) glycerol, 0.02 % (v/v) NP-40, 0.2 mg/ml FLAG peptide
10. Glycine-HCl: Dissolve glycine powder in H₂O. Adjust pH to 3.5 with HCl. Then, adjust the volume to 100 mM glycine (12 ml).
11. BRB-2Mg: 80 mM PIPES-KOH (pH 6.8), 2 mM MgCl₂, 1 mM EGTA

- HighFive insect cells were grown in suspension culture to a density of $2\text{-}2.5 \times 10^6$ cells/ml and were infected with P3 viral stock with MOI 20. Cells were harvested at 36 hrs. after infection. All the purification steps were done at 4°C with pre-chilled buffers unless mentioned otherwise.
- To lyse cells, two extraction solutions were prepared, extraction solution A (PMI supplemented with 2 mM ATP, 2 mM GTP, 2 mM DTT and protease inhibitor cocktail) and extraction solution B (PMI supplemented with 1% CHAPS and 1M NDSB201).
- 10 gm of cell pellet was resuspended in 30 ml extraction solution A and then 30 ml extraction solution B was added to lyse cells for 30 min.
- The lysate was centrifuged at 200000 G for 25min and the supernatant was supplemented with 10% glycerol.
- This was applied to 10 ml DEAE Sepharose fast flow resin equilibrated with 100 mM PIPES-KOH (pH 6.8) and incubated for 60 min.
- The resin was washed with four column volumes of DEAE wash buffer supplemented with 1 mM ATP, 1 mM GTP and protease inhibitor cocktail and the bound fraction was eluted in DEAE elution buffer supplemented 1 mM ATP and 1 mM GTP.
- The elution fraction with total protein concentration >0.7 mg/ml were pooled together and applied to 2 ml TALON resin coated with BSA and equilibrated with TALON wash buffer for 60 min.
- The unbound proteins were washed with 10 column volumes TALON wash buffer supplemented with 1 mM ATP and 1 mM GTP and protein was eluted by TALON elution buffer with 1 mM ATP and 1 mM GTP.
- Elution fraction with concentration >0.2 mg/ml were pooled to FLAG affinity resin coated with BSA and equilibrated with FLAG wash buffer for 60 min.
- The unbound proteins were washed with 10 column volumes FLAG wash buffer supplemented with 1 mM ATP and 1 mM GTP and protein was eluted by FLAG elution buffer with 1 mM GTP.
- Alternatively, the selected TALON elution fractions were concentrated in Centrifugal filter (Amicon Ultra 15 - 30k, Millipore) and transferred into BRB80-2Mg buffer supplemented with 1 mM GTP and 10% glycerol to a concentration of 3-8 mg/ml.
- Concentrated tubulin was centrifuged at 400000 G for 10 min to remove denatured and aggregated protein.
- The purified tubulin was polymerized using microtubule pelleting assay.³⁰⁷ In a typical polymerization set up, 800 μl of 1 mg/ml tubulin sample in BRB80 buffer with 2 mM GTP and 0.1 mM DTT was taken.
- It was then warmed to 37°C for 10 min and Taxol was added in steps (20 μM (4 μL), 200 μM (4 μL) 2 mM (4 μL)) separated by 5 min incubation at 37°C .
- A Taxol glycerol cushion was prepared with 1ml of 2X BRB80, 0.8ml glycerol (100%) and 20ul Taxol (2mM).

- To centrifuge polymerized microtubules, 150 μL of cushion was added and polymerization mixtures was gently pipetted on top and spun at 80000 RPM for 15 min.
- The pellet was resuspended gently in 200 μL BRB80 buffer with 1 mM GTP and used for wide field fluorescence microscopy imaging.

6.2.10. Microtubule Imaging by Wide Field Fluorescence Microscopy

- For immunofluorescence analysis³⁰⁴, 10 μM Taxol-stabilized microtubules were spun down on a coverslip for 10 min at 10000 G.
- Samples were then fixed in ice-cold methanol for 5min followed by a wash in PBS and incubation with blocking solution (1%, w/v, BSA in PBS, 0.5% Tween 20 [PBT]) for 30 min.
- Primary and secondary antibodies were diluted in PBT as follows: mouse anti- α -tubulin (1/1,000) rabbit anti- β -tubulin (1/500), goat anti-rabbit Alexa Fluor 488 and goat anti-mouse Alexa Fluor 568 (both Life Technologies): both 1/1,000.
- Primary antibodies were incubated at room temperature for 2 hrs. and secondary antibodies at room temperature for 1 hr. and the slides were washed with PBT.
- All images were processed with ImageJ. The antibodies were kindly provided by Prof. Pierre Gonczy lab and the imaging was done with help of Dr. Georgios Hatzopoulos.

6.2.11. Tubulin Dimer Resolving SDS-PAGE

- Both α - and β - tubulin monomers, present in equimolar amounts per sample and being of approximately equal molecular weight (50 kDa), are difficult to separate adequately on polyacrylamide minigels.
- Taking advantage of the disparate detergent binding of the α - and β -subunits under denaturing conditions, electrophoretic separation of the heterodimers has been achieved.³⁰⁸
- Acrylamide minigels were freshly cast for all experiments. A stock solution of 30% acrylamide and 0.8% (w/v) N, N'-methylenebisacrylamide was prepared.
- The stacking layer (6% acrylamide final concentration) was prepared by mixing 2 ml of acrylamide/bis-acrylamide stock solution, 5.4 ml of water (ddH₂O), 2.5 ml of stacking buffer (0.5 M Tris-OH, pH 6.8), and 100 μL of 10% SDS.
- Polymerization was initiated by adding 50 μL of freshly prepared 10% APS and 10 μL of TEMED.
- For resolving gel, 5 ml of acrylamide/bis-acrylamide stock solution (7.5% acrylamide final concentration), 8 ml of ddH₂O, 5 ml of resolving buffer made with 1.5 M Tris-OH (pH 9.8), and 200 μL of 10% SDS.
- Polymerization was initiated with 250 μL of 10% APS and 16 μL of TEMED. Tubulin gel loading samples were prepared in 3X- Laemmli sample buffer.
- For optimal separation, gels were run for 90 min at 120V.

6.2.12. Synthesis of Tubulin Tail Peptides

- The Npu^C-tubulin tail (fluorescein labelled) tail peptide (51 aa long) was synthesized in two parts as described in 4.3.3. The peptide 1 was synthesized on a Tentagel rink amide resin using standard Fmoc synthesis on an automated peptide synthesizer.
- After the synthesis, the resin was washed with DMF, dichloromethane and methanol and dried under vacuum.
- Following cleavage from the resin by a mixture of 95 % trifluoroacetic acid (TFA), 2.5 % water and 2.5 % triisopropylsilane (TIS), crude peptide was washed at least twice with cold diethyl ether and recovered by centrifugation.
- The peptide 1 was purified by RP-HPLC on a semi-preparative scale using linear gradients of 0% - 70% solvent B over 40 min and characterized by analytical RP-HPLC and ESI-MS and lyophilized (~ 2mg).
- Part of lyophilized peptide was taken and dissolved in (100 mM NaH₂PO₄, 4 M urea, pH 7.0) to a concentration of 500 μM and 2 eq. of azide fluorescein was added with 1 mM CuSO₄ and 50 mM sodium ascorbate.
- The labelling reaction was carried out for 60 min and fluorescein labelled peptide 2 was purified on a semi-preparative scale using linear gradients of 0% - 70% solvent B over 40 min and characterized by analytical RP-HPLC and ESI-MS and lyophilized.
- For synthesis of peptide 3, 2-Chlorotriyl resin was first reacted with hydrazine. For this, 0.5g 2-chloro-trityl chloride resin was swollen in 3mL DMF for 15min and cooled to 4°C. A solution of 350μL DIPEA (2.01mmol, 3 eq.), 65μL hydrazine monohydrate (1.33mmol, 2 eq.) and 585μL DMF was added dropwise to the resin.
- The reaction was stirred for 1h at room temperature (RT). The peptide 3 was synthesized on using standard Fmoc synthesis on an automated peptide synthesizer.
- After the synthesis, the resin was washed with DMF, dichloromethane and methanol and dried under vacuum.
- Following cleavage from the resin by a mixture of 95 % trifluoroacetic acid (TFA), 2.5 % water and 2.5 % triisopropylsilane (TIS), crude peptide was washed at least twice with cold diethyl ether and recovered by centrifugation.
- The peptide 3 was purified as peptide hydrazide by RP-HPLC on a semi-preparative scale using linear gradients of 0% - 70% solvent B over 40 min and characterized by analytical RP-HPLC and ESI-MS and lyophilized (~ 3mg).
- The lyophilized peptide was dissolved in (200 mM NaH₂PO₄, 6 M GuHCl, pH 3.0) and 10 mM NaNO₂ was added at 20°C for 20 min followed by 75 mM MTG to convert hydrazine to thioester as peptide 4.
- To this, 25 mM TCEP in same buffer was added. The peptide 4 was ligated to 0.8 eq. of fluorescein labelled peptide 2 using native chemical ligation and the final peptide 5 was purified by semi-preparative RP-HPLC, analyzed by analytical RP-HPLC and ESI-MS and lyophilized.

References

1. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–3 (1970).
2. McCarthy, B. J. & Holland, J. J. Denatured DNA as a direct template for in vitro protein synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **54**, 880–886 (1965).
3. J. Craig Venter, Josep F. Abril & Roderic Guigó. The Sequence of the Human Genome. (2001).
4. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chemie - Int. Ed.* **44**, 7342–7372 (2005).
5. Black, D. L. Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
6. Black, D. L. Protein diversity from alternative splicing: A challenge for bioinformatics and post-genome biology. *Cell* **103**, 367–370 (2000).
7. Maniatis, T. & Tasic, B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236–243 (2002).
8. Lodish, H. F. Post- translational modification of proteins. *Enzym. Microb. Technol.* **3**, 178–188 (1981).
9. Mann, M. *et al.* Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol.* **20**, 261–8 (2002).
10. Barrett, J. F. & Hoch, J. A. Two-component signal transduction as a target for microbial anti-infective therapy. *Antimicrob. Agents Chemother.* **42**, 1529–1536 (1998).
11. C. Manning, D.B. Whyte, R. Martinez, T. hunter, S. S. The Protein Kinase Complement of the Human Genome: EBSCOhost. *Sciencemag.Org* **298**, 1912–1934 (2002).
12. Johnson, L. N. & Lewis, R. J. Structural basis for control by phosphorylation. *Chem. Rev.* **101**, 2209–2242 (2001).
13. Brooks, C. L. & Gu, W. Ubiquitination, phosphorylation and acetylation: The molecular basis for p53 regulation. *Curr. Opin. Cell Biol.* **15**, 164–171 (2003).
14. Pickart, C. M. Echanisms nderlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–33 (2001).
15. Pickart, C. M. & Cohen, R. E. Proteasomes and their kin: Proteases in the machine age. *Nat. Rev. Mol. Cell Biol.* **5**, 177–187 (2004).
16. Cardozo, T. & Pagano, M. The SCF ubiquitin ligase: Insights into a molecular machine. *Nat. Rev. Mol. Cell Biol.* **5**, 739–751 (2004).
17. Khorasanizadeh, S. The Nucleosome. *Cell* **116**, 259–272 (2004).
18. Jackson, M. D. & Denu, J. M. Molecular reactions of protein phosphatases - Insights from structure and chemistry. *Chem. Rev.* **101**, 2313–2340 (2001).

19. Ferguson, P. L. & Smith, R. D. P Roteome a Nalysis By M Ass S Pectrometry *. 399–424 (2003). doi:10.1146/annurev.biophys.32.110601.141854
20. Grozinger, C. M. & Schreiber, S. L. Deacetylase enzymes: Biological functions and the use of small-molecule inhibitors. *Chem. Biol.* **9**, 3–16 (2002).
21. Hui Ng, H. & Bird, A. Histone deacetylases: Silencers for hire. *Trends Biochem. Sci.* **25**, 121–126 (2000).
22. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
23. Paetzel, M., Karla, A., Strynadka, N. C. J. & Dalbey, R. E. Signal peptidases. *Chem. Rev.* **102**, 4549–4579 (2002).
24. Rockwell, N. C., Krysan, D. J., Komiyama, T. & Fuller, R. S. Precursor processing by Kex2/Furin proteases. *Chem. Rev.* **102**, 4525–4548 (2002).
25. Gadadhar, S., Bodakuntla, S., Natarajan, K. & Janke, C. The tubulin code at a glance. *J. Cell Sci.* **130**, 1347–1353 (2017).
26. Nieuwenhuis, J. *et al.* Vasohibins encode tubulin detyrosinating activity. **1456**, 1453–1456 (2017).
27. Chakraborti, S., Natarajan, K., Curiel, J., Janke, C. & Liu, J. The emerging role of the tubulin code: From the tubulin molecule to neuronal function and disease. *Cytoskeleton* **73**, 521–550 (2016).
28. Magiera, M. M., Singh, P. & Janke, C. SnapShot: Functions of Tubulin Posttranslational Modifications. *Cell* **173**, 1552–1552.e1 (2018).
29. Perler, F. B. Protein splicing of inteins and hedgehog autoproteolysis: Structure, function, and evolution. *Cell* **92**, 1–4 (1998).
30. Perler, F. B., Olsen, G. J. & Adam, E. Compilation and analysis of intein sequences. *Nucleic Acids Res.* **25**, 1087–1093 (1997).
31. Xu, M. Q. & Perler, F. B. The mechanism of protein splicing and its modulation by mutation. *EMBO J.* **15**, 5146–5153 (1996).
32. Vila-Perelló, M. & Muir, T. W. Biological Applications of Protein Splicing. *Cell* **143**, 191–200 (2010).
33. Beltrao, P., Bork, P., Krogan, N. J. & Van Noort, V. Evolution and functional cross-talk of protein post-translational modifications. *Mol. Syst. Biol.* **9**, 1–13 (2013).
34. Seet, B. T., Dikic, I., Zhou, M. M. & Pawson, T. Reading protein modifications with interaction domains. *Nat. Rev. Mol. Cell Biol.* **7**, 473–483 (2006).
35. Chuh, K. N., Batt, A. R. & Pratt, M. R. Chemical Methods for Encoding and Decoding of Posttranslational Modifications. *Cell Chem. Biol.* **23**, 86–107 (2016).
36. McGinty, R. K. & Tan, S. Nucleosome structure and function. *Chem. Rev.* **115**, 2255–2273 (2015).

37. Fierz, B. & Muir, T. W. Chromatin as an expansive canvas for chemical biology. *Nat. Chem. Biol.* **8**, 417–427 (2012).
38. Woodcock, C. L. F., Safer, J. P. & Stanchfield, J. E. Structural repeating units in chromatin: I. Evidence for their general occurrence. *Exp. Cell Res.* **97**, 101–110 (1976).
39. Hewish, D. R. & Burgoyne, L. A. Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.* **52**, 504–510 (1973).
40. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
41. Chua, E. Y. D., Vasudevan, D., Davey, G. E., Wu, B. & Davey, C. A. The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res.* **40**, 6338–6352 (2012).
42. Arents, G., Burlingame, R. W., Wang, B. C., Love, W. E. & Moudrianakis, E. N. The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci.* **88**, 10148–10152 (1991).
43. Arents, G. & Moudrianakis, E. N. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc. Natl. Acad. Sci.* **92**, 11170–11174 (1995).
44. Widom, J. A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. *Proc. Natl. Acad. Sci.* **89**, 1095–1099 (1992).
45. Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: An ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* **13**, 436–447 (2012).
46. Finch, J. T. & Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci.* **73**, 1897–1901 (1976).
47. Woodcock, C. L. F., Frado, L. L. Y. & Rattner, J. B. The higher-order structure of chromatin: Evidence for a helical ribbon arrangement. *J. Cell Biol.* **99**, 42–52 (1984).
48. Bednar, J. *et al.* Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc. Natl. Acad. Sci.* **95**, 14173–14178 (1998).
49. Dorigo, B. Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber. *Science (80-)*. **306**, 1571–1573 (2004).
50. Schalch, T., Duda, S., Sargent, D. F. & Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **436**, 138–141 (2005).
51. Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science (80-)*. **357**, (2017).
52. Hnisz, D., Day, D. S. & Young, R. A. Insulated Neighborhoods: Structural and Functional

- Units of Mammalian Gene Control. *Cell* **167**, 1188–1200 (2016).
53. Barrington, C., Pezic, D. & Hadjur, S. Chromosome structure dynamics during the cell cycle: a structure to fit every phase. *EMBO J.* **36**, 2661–2663 (2017).
 54. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
 55. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).
 56. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).
 57. Holliday, R. Epigenetics: A historical overview. *Epigenetics* **1**, 76–80 (2006).
 58. Berger, S. L., Kouzarides, T., Shiekhattar, R. & Shilatifard, A. An operational definition of epigenetics An operational definition of epigenetics. 781–783 (2009). doi:10.1101/gad.1787609
 59. Razin, A. & Cedar, H. DNA methylation and genomic imprinting. *Cell* **77**, 473–476 (1994).
 60. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
 61. Plath, K., Mlynarczyk-Evans, S., Nusinow, D. A. & Panning, B. *Xist* RNA and the Mechanism of X Chromosome Inactivation. *Annu. Rev. Genet.* **36**, 233–278 (2002).
 62. Bestor, T. H. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J.* **11**, 2611–7 (1992).
 63. Bacolla, A., Pradhan, S., Roberts, R. J. & Wells, R. D. Recombinant Human DNA (Cytosine-5) Methyltransferase. *J. Biol. Chem.* **274**, 33002–33010 (1999).
 64. Ragnathan, K., Jih, G. & Moazed, D. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science (80-.).* **348**, (2015).
 65. Audergon, P. N. C. B. *et al.* Restricted epigenetic inheritance of $\{H3K9\}$ methylation. *Science (80-.).* **348**, 132–135 (2015).
 66. Morgan, H. D., Sutherland, H. G. E., Martin, D. I. K. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* **23**, 314–318 (1999).
 67. Greer, E. L. *et al.* Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. *Nature* **479**, 365–371 (2011).
 68. Chatterjee, C. Chemical Approaches for Studying Histone Modifications. *J. Biol. Chem.* **285**, 11045–11050 (2010).
 69. Rothbart, S. B. & Strahl, B. D. Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839**, 627–643 (2014).
 70. Luger, K. & Hansen, J. C. Nucleosome and chromatin fiber dynamics. *Curr. Opin. Struct. Biol.* **15**, 188–196 (2005).
 71. Li, G. & Reinberg, D. Chromatin higher-order structures and gene regulation. *Curr.*

- Opin. Genet. Dev.* **21**, 175–186 (2011).
72. Jenuwein, T. & Allis, C. D. Allis: Translating the histone code. *Science (80-.)*. **293**, 1074–1080 (2001).
 73. Bradley, R. S. *et al.* Histone H4-K16 Acetylation. **16**, 844–848 (2006).
 74. Neumann, H. *et al.* A Method for Genetically Installing Site-Specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation. *Mol. Cell* **36**, 153–163 (2009).
 75. North, J. A. *et al.* Phosphorylation of histone H3(T118) alters nucleosome dynamics and remodeling. *Nucleic Acids Res.* **39**, 6465–6474 (2011).
 76. Kilic, S., Boichenko, I., Lechner, C. C. & Fierz, B. A bi-terminal protein ligation strategy to probe chromatin structure during DNA damage. *Chem. Sci.* **9**, 3704–3709 (2018).
 77. Fierz, B. *et al.* Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nat. Chem. Biol.* **7**, 113–119 (2011).
 78. Bowman, G. D. & Poirier, M. G. Post-Translational Modifications of Histones That Influence Nucleosome Dynamics. *Chem. Rev.* **115**, 2274–2295 (2015).
 79. Dhalluin, C. *et al.* Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491–496 (1999).
 80. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
 81. Min, J., Zhang, Y. & Xu, R. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27 Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. 1823–1828 (2003). doi:10.1101/gad.269603
 82. Li, H. *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91–95 (2006).
 83. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870–1881 (2003).
 84. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**, 1025–1040 (2007).
 85. Ruthenburg, A. J., Li, H., Patel, D. J. & David Allis, C. Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **8**, 983–994 (2007).
 86. Ernst, J. & Kellis, M. Discovery and Characterization of Chromatin States for Systematic Annotation of the Human Genome. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **6577 LNBI**, 53 (2011).
 87. Fillion, G. J. *et al.* Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell* **143**, 212–224 (2010).

88. Consortium, T. *et al.* Identification of Functional Elements and Regulatory Circuits by *Drosophila* modENCODE. *October* **330**, 1787–1797 (2011).
89. Mayer, T. & Biology, C. Chemical tools for exploring cell division. 1–4
90. Risler, T. Cytoskeleton and Cell Motility. *Compr. Biotechnol. Second Ed.* **1**, 191–204 (2011).
91. Yu, I., Garnham, C. P. & Roll-Mecak, A. Writing and reading the tubulin code. *J. Biol. Chem.* **290**, 17163–17172 (2015).
92. Janke, C. The tubulin code: Molecular components, readout mechanisms, functions. *J. Cell Biol.* **206**, 461–472 (2014).
93. Nogales, E., Wolf, S. G. & Downing, K. H. Structure of the $\alpha\beta$ tubulin dimer by electron crystallography. *Nature* **391**, 199–203 (1998).
94. Downing, K. H. & Nogales, E. Tubulin structure: Insights into microtubule properties and functions. *Curr. Opin. Struct. Biol.* **8**, 785–791 (1998).
95. Mandelkow, E. & Mandelkow, E. M. Microtubules and microtubule-associated proteins. *Curr. Opin. Cell Biol.* **7**, 72–81 (1995).
96. Mandelkow, E. & Mandelkow, E. M. Microtubular structure and tubulin polymerization. *Curr. Opin. Cell Biol.* **2**, 3–9 (1990).
97. Lewis, S. A., Tian, G., Vainberg, I. E. & Cowan, N. J. Chaperonin-mediated folding of actin and tubulin. *J. Cell Biol.* **132**, 1 LP-4 (1996).
98. Llorca, O. *et al.* The ‘sequential allosteric ring’ mechanism in the eukaryotic chaperonin-assisted folding of actin and tubulin. *EMBO J.* **20**, 4065–4075 (2001).
99. Nogales, E. STructural INsights Into M IcroTubule F Unction. 277–302 (2000).
100. Horio, T. & Murata, T. The role of dynamic instability in microtubule organization. *Front. Plant Sci.* **5**, 1–10 (2014).
101. Mitchison, T. & Kirschner, M. Dynamic instability of microtubule growth. *Nature* **312**, 237 (1984).
102. Desai, A. & Mitchison, T. J. Microtubule Polymerization Dynamics. *Annu. Rev. Cell Dev. Biol.* **13**, 83–117 (1997).
103. Mitchison, T. J. The engine of microtubule dynamics comes into focus. *Cell* **157**, 1008–1010 (2014).
104. Löwe, J., Li, H., Downing, K. H. & Nogales, E. Refined structure of $\alpha\beta$ -tubulin at 3.5 Å resolution. *J. Mol. Biol.* **313**, 1045–1057 (2001).
105. Zhang, R., Alushin, G. M., Brown, A. & Nogales, E. Mechanistic origin of microtubule dynamic instability and its modulation by EB proteins. *Cell* **162**, 849–859 (2015).
106. Alushin, G. M. *et al.* High-Resolution microtubule structures reveal the structural transitions in $\alpha\beta$ -tubulin upon GTP hydrolysis. *Cell* **157**, 1117–1129 (2014).
107. Bowne-Anderson, H., Zanic, M., Kauer, M. & Howard, J. Microtubule dynamic

- instability: A new model with coupled GTP hydrolysis and multistep catastrophe. *BioEssays* **35**, 452–461 (2013).
108. Howard, J. & Hyman, A. A. Dynamics and mechanics of the microtubule plus end. *Nature* **422**, 753–758 (2003).
 109. Akhmanova, A. & Steinmetz, M. O. Control of microtubule organization and dynamics: Two ends in the limelight. *Nat. Rev. Mol. Cell Biol.* **16**, 711–726 (2015).
 110. Luduena, R. F. A hypothesis on the origin and evolution of tubulin. *Int. Rev. Cell Mol. Biol.* **302**, 41–185 (2013).
 111. Ludueña, R. F. & Banerjee, A. The Isoforms of Tubulin. in *The Role of Microtubules in Cell Biology, Neurobiology, and Oncology* (ed. Fojo, T.) 123–175 (Humana Press, 2008). doi:10.1007/978-1-59745-336-3_6
 112. Roll-Mecak, A. Intrinsically disordered tubulin tails: Complex tuners of microtubule functions? *Semin. Cell Dev. Biol.* **37**, 11–19 (2015).
 113. Verhey, K. J. & Gaertig, J. The tubulin code. *Cell Cycle* **6**, 2152–2160 (2007).
 114. Strzyz, P. Post-translational modifications: Extension of the tubulin code. *Nat. Rev. Mol. Cell Biol.* **17**, 609 (2016).
 115. Raunser, S. & Gatsogiannis, C. Deciphering the tubulin code. *Cell* **161**, 960–961 (2015).
 116. Kunishima, S., Kobayashi, R., Itoh, T. J., Hamaguchi, M. & Saito, H. Mutation of the β 1-tubulin gene associated with congenital macrothrombocytopenia affecting microtubule assembly. *Blood* **113**, 458–461 (2009).
 117. Burgoyne, R. D., Cambray-Deakin, M. A., Lewis, S. A., Sarkar, S. & Cowan, N. J. Differential distribution of beta-tubulin isoforms in cerebellum. *EMBO J.* **7**, 2311–2319 (1988).
 118. Eipper, B. A. Rat Brain Microtubule Protein: Purification and Determination of Covalently Bound Phosphate and Carbohydrate. *Proc. Natl. Acad. Sci.* **69**, 2283–2287 (1972).
 119. Barra, H. S., Rodriguez, J. A., Arce, C. A. & Caputto, R. A SOLUBLE PREPARATION FROM RAT BRAIN THAT INCORPORATES INTO ITS OWN PROTEINS [14C]ARGININE BY A RIBONUCLEASE-SENSITIVE SYSTEM AND [14C]TYROSINE BY A RIBONUCLEASE-INSENSITIVE SYSTEM. *J. Neurochem.* **20**, 97–108
 120. ARCE, C. A., RODRIGUEZ, J. A., BARRA, H. S. & CAPUTTO, R. Incorporation of L-Tyrosine, L-Phenylalanine and L-3,4-Dihydroxyphenylalanine as Single Units into Rat Brain Tubulin. *Eur. J. Biochem.* **59**, 145–149
 121. Raybin, D. & Flavin, M. An enzyme tyrosylating α -tubulin and its role in microtubule assembly. *Biochem. Biophys. Res. Commun.* **65**, 1088–1095 (1975).
 122. Redeker, V., Le Caer, J.-P., Rossier, J. & Prome, J.-C. Structure of the polyglutamyl side chain posttranslationally added to α -tubulin. *J. Biol. Chem.* **266**, 23461–23466 (1991).
 123. Redeker, V., Melki, R., Promé, D., Caer, J.-P. Le & Rossier, J. Structure of tubulin C-terminal domain obtained by subtilisin treatment The major α and β tubulin isoforms

- from pig brain are glutamylated. *FEBS Lett.* **313**, 185–192 (1992).
124. Alexander, J. E. *et al.* Characterization of posttranslational modifications in neuron-specific class III beta-tubulin by mass spectrometry. *Proc. Natl. Acad. Sci.* **88**, 4685–4689 (1991).
 125. Edde, B. *et al.* Posttranslational glutamylation of alpha-tubulin. *Science (80-.)*. **247**, 83–85 (1990).
 126. Redeker, V. *et al.* Polyglycylation of tubulin: A posttranslational modification in axonemal microtubules. *Science (80-.)*. **266**, 1688–1691 (1994).
 127. Greer, K., Maruta, H., L'Hernault, S. W. & Rosenbaum, J. L. Alpha-tubulin acetylase activity in isolated *Chlamydomonas* flagella. *J. Cell Biol.* **101**, 2081–2084 (1985).
 128. L'Hernault, S. W. & Rosenbaum, J. L. *Chlamydomonas* alpha-tubulin is posttranslationally modified by acetylation on the epsilon-amino group of a lysine. *Biochemistry* **24**, 473–478 (1985).
 129. Fourest-Lieuvin, A. *et al.* Microtubule Regulation in Mitosis: Tubulin Phosphorylation by the Cyclin-dependent Kinase Cdk1. *Mol. Biol. Cell* **17**, 1041–1050 (2006).
 130. Song, Y. *et al.* Transglutaminase and Polyamination of Tubulin: Posttranslational Modification for Stabilizing Axonal Microtubules. *Neuron* **78**, 109–123 (2013).
 131. Garnham, C. P. & Roll-Mecak, A. The chemical complexity of cellular microtubules: Tubulin post-translational modification enzymes and their roles in tuning microtubule functions. *Cytoskeleton* **69**, 442–463 (2012).
 132. Webster, D. R., Gundersen, G. G., Bulinski, J. C. & Borisy, G. G. Differential turnover of tyrosinated and detyrosinated microtubules. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 9040–9044 (1987).
 133. Bobinnec, Y. *et al.* Glutamylation of centriole and cytoplasmic tubulin in proliferating non-neuronal cells. *Cell Motil. Cytoskeleton* **39**, 223–232 (1998).
 134. Lacroix, B. *et al.* Tubulin polyglutamylation stimulates spastin-mediated microtubule severing. *J. Cell Biol.* **189**, 945–954 (2010).
 135. Mary, J., Redeker, V., Le Caer, J. P., Rossier, J. & Schmitter, J. M. Posttranslational modifications in the C-terminal tail of axonemal tubulin from sea urchin sperm. *J. Biol. Chem.* **271**, 9928–9933 (1996).
 136. Mary, J., Redeker, V., Le Caer, J.-P., Rossier, J. & Schmitter, J.-M. Posttranslational Modifications of Axonemal Tubulin. *J. Protein Chem.* **16**, 403–407 (1997).
 137. Bre, M. H. *et al.* Axonemal tubulin polyglycylation probed with two monoclonal antibodies: widespread evolutionary distribution, appearance during spermatozoan maturation and possible function in motility. *J. Cell Sci.* **109**, 727–738 (1996).
 138. Kubo, T., Yanagisawa, H. aki, Yagi, T., Hirono, M. & Kamiya, R. Tubulin Polyglutamylation Regulates Axonemal Motility by Modulating Activities of Inner-Arm Dyneins. *Curr. Biol.* **20**, 441–445 (2010).

139. Suryavanshi, S. *et al.* Tubulin Glutamylation Regulates Ciliary Motility by Altering Inner Dynein Arm Activity. *Curr. Biol.* **20**, 435–440 (2010).
140. Audebert, S. *et al.* Reversible polyglutamylation of alpha- and beta-tubulin and microtubule dynamics in mouse brain neurons. *Mol. Biol. Cell* **4**, 615–626 (1993).
141. MUROFUSHI, H. Purification and Characterization of Tubulin-Tyrosine Ligase from Porcine Brain1. *J. Biochem.* **87**, 979–984 (1980).
142. Schroder, H., Wehland, J. & Weber, K. Purification of Brain Tubulin-Tyrosine Ligase by Biochemical and Immunological Methods. **100**, 276–281 (1985).
143. Paturle-lafanechère, L. *et al.* Accumulation of delta 2-tubulin , a major tubulin variant that cannot be tyrosinated , in neuronal tissues and in stable microtubule assemblies. **1543**, 1529–1543 (1994).
144. Aillaud, C. *et al.* Evidence for new C-terminally truncated variants of α - and β - tubulins. (2016). doi:10.1091/mbc.E15-03-0137
145. Kalinina, E. *et al.* A novel subfamily of mouse cytosolic carboxypeptidases. (2018). doi:10.1096/fj.06-7329com
146. Kimura, Y. *et al.* Identification of Tubulin Deglutamylase among *Caenorhabditis elegans* and Mammalian Cytosolic Carboxypeptidases (CCPs) * □. **285**, 22936–22941 (2010).
147. Hernault, S. W. L. & Rosenbaum, J. L. Chlamydomonas α -Tubulin Is Posttranslationally Modified by Acetylation e-Amino Group of a Lysine \wedge . 473–478 (1985). doi:10.1021/bi00323a034
148. Akella, J. S. *et al.* MEC-17 is an α -tubulin acetyltransferase. *Nature* **467**, 218–222 (2010).
149. Shida, T., Cueva, J. G., Xu, Z., Goodman, M. B. & Nachury, M. V. promotes rapid ciliogenesis and efficient mechanosensation. **4**, (2010).
150. Hubbert, C., Guardiola, A. & Shao, R. HDAC6 is a microtubule-associated deacetylase. **417**, 455–458 (2002).
151. North, B. J. *et al.* Is an NAD $\%$ -Dependent Tubulin Deacetylase. **11**, 437–444 (2003).
152. Bulinski, J. C., Richards, J. E. & Piperno, G. Posttranslational Modifications of α -Tubulin : Detyrosination and Acetylation Differentiate Populations of Interphase Microtubules in Cultured Cells. **106**, 1213–1220 (1988).
153. Geimer, S., Clees, J., Melkonian, M. & Lehtreck, K. F. A novel 95-kD protein is located in a linker between cytoplasmic microtubules and basal bodies in a green flagellate and forms striated filaments in vitro. *J. Cell Biol.* **140**, 1149–1158 (1998).
154. Janke, C., Rogowski, K. & van Dijk, J. Polyglutamylation: A fine-regulator of protein function? ‘Protein Modifications: Beyond the Usual Suspects’ Review Series. *EMBO Rep.* **9**, 636–641 (2008).
155. Rogowski, K. *et al.* A family of protein-deglutamylating enzymes associated with neurodegeneration. *Cell* **143**, 564–578 (2010).

156. Bobinnec, Y., Khodjakov, A., Mir, L. M., Rieder, C. L. & Bornens, M. Centriole Disassembly In Vivo and Its Effect on Centrosome Structure and Function in Vertebrate Cells. **143**, 1575–1589 (1998).
157. Bré, M. H., de Néchaud, B., Wolff, A. & Fleury, A. Glutamylated tubulin probed in ciliates with the monoclonal antibody GT335. *Cell Motil.* **27**, 337–349
158. Konno, A., Setou, M. & Ikegami, K. Chapter three - Ciliary and Flagellar Structure and Function—Their Regulations by Posttranslational Modifications of Axonemal Tubulin. in (ed. Jeon, K. W.) **294**, 133–170 (Academic Press, 2012).
159. Levilliers, N. *et al.* Article TTL3 Is a Tubulin Glycine Ligase that Regulates the Assembly of Cilia. 867–876 (2009). doi:10.1016/j.devcel.2009.04.008
160. Janke, C., Rogowski, K. & Wloga, D. Tubulin Polyglutamylase Enzymes Are Members of the TTL Domain Protein Family. **308**, 1758–1763 (2005).
161. Dijk, J. Van *et al.* Polyglutamylation Is a Post-translational Modification with a Broad Range of Substrates * □. **283**, 3915–3922 (2008).
162. Janke, C. *et al.* Article A Targeted Multienzyme Mechanism for Selective Microtubule Polyglutamylation. 437–448 (2007). doi:10.1016/j.molcel.2007.04.012
163. Sirajuddin, M., Rice, L. M. & Vale, R. D. Regulation of microtubule motors by tubulin isoforms and post-translational modifications. *Nat. Cell Biol.* **16**, (2014).
164. Boscheron, C. *et al.* Suppression of nuclear oscillations in *Saccharomyces cerevisiae* expressing Glu tubulin. **101**, (2004).
165. Vemu, A., Garnham, C. P., Lee, D. Y. & Roll-Mecak, A. *Generation of differentially modified microtubules using in vitro enzymatic approaches. Methods in Enzymology* **540**, (Elsevier Inc., 2014).
166. Vemu, A. *et al.* Structure and Dynamics of Single-isoform Recombinant Neuronal Human Tubulin. *J. Biol. Chem.* **291**, 12907–15 (2016).
167. Minoura, I. *et al.* Overexpression, purification, and functional analysis of recombinant human tubulin dimer. *FEBS Lett.* **587**, 3450–3455 (2013).
168. Noren, C. J., Anthony-cahill, S. J., Griffith, M. C., Schultz, P. G. & April, I. A General Method for Site-Specific Incorporation of Unnatural Amino Acids into Proteins. *Science (80-)*. **244**, 182–188 (1989).
169. Bain, J. D., Glabe, C. G., Dix, T. A., Chamberlin, A. R. & Diala, E. S. Biosynthetic Site-Specific Incorporation of a Non-Natural Amino Acid into a Polypeptide. *J. Am. Chem. Soc.* **111**, 8013–8014 (1989).
170. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
171. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding N ϵ -acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4**, 232–234 (2008).
172. Nguyen, D. P., Alai, M. M. G., Virdee, S. & Chin, J. W. Genetically directing ϵ -N, N-dimethyl-l-lysine in recombinant histones. *Chem. Biol.* **17**, 1072–1076 (2010).

173. Ai, H. W., Lee, J. W. & Schultz, P. G. A method to site-specifically introduce methyllysine into proteins in *E. coli*. *Chem. Commun.* **46**, 5506–5508 (2010).
174. Simon, M. D. *et al.* The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **128**, 1003–1012 (2007).
175. Chatterjee, C., McGinty, R. K., Fierz, B. & Muir, T. W. Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. *Nat. Chem. Biol.* **6**, 267–269 (2010).
176. Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **5**, 1–14 (2014).
177. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).
178. Merrifield, R. B. Solid Phase Peptide Synthesis. II. The Synthesis of Bradykinin. *J. Am. Chem. Soc.* **85**, 2149–2154 (1964).
179. Saxon, E., Armstrong, J. I. & Bertozzi, C. R. A ‘traceless’ Staudinger ligation for the chemoselective synthesis of amide bonds. *Org. Lett.* **2**, 2141–2143 (2000).
180. Nilsson, B. L., Kiessling, L. L. & Raines, R. T. Staudinger Ligation: A Peptide from a Thioester and Azide. *Org. Lett.* **2**, 1939–1941 (2000).
181. Bode, J. W., Fox, R. M. & Baucom, K. D. Chemoselective amide ligations by decarboxylative condensations of N-alkylhydroxylamines and α -ketoacids. *Angew. Chemie - Int. Ed.* **45**, 1248–1252 (2006).
182. Zhang, Y., Xu, C., Lam, H. Y., Lee, C. L. & Li, X. Protein chemical synthesis by serine and threonine ligation. *Proc. Natl. Acad. Sci.* **110**, 6657–6662 (2013).
183. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science (80-.)*. **266**, 776 LP-779 (1994).
184. Hackenberger, C. P. R. & Schwarzer, D. Chemoselective ligation and modification strategies for peptides and proteins. *Angew. Chemie - Int. Ed.* **47**, 10030–10074 (2008).
185. Bondalapati, S., Jbara, M. & Brik, A. Expanding the chemical toolbox for the synthesis of large and uniquely modified proteins. *Nat. Chem.* **8**, 407–418 (2016).
186. Kilic, S., Bachmann, A. L., Bryan, L. C. & Fierz, B. Multivalency governs HP1 α association dynamics with the silent chromatin state. *Nat. Commun.* **6**, 1–11 (2015).
187. Lechner, C. C., Agashe, N. D. & Fierz, B. Traceless Synthesis of Asymmetrically Modified Bivalent Nucleosomes. *Angew. Chemie Int. Ed.* **55**, 2903–2906 (2016).
188. Fang, G. M. *et al.* Protein chemical synthesis by ligation of peptide hydrazides SI. *Angew. Chemie - Int. Ed.* **50**, 7645–7649 (2011).
189. Stavropoulos, G., Gatos, D., Magafa, V. & Barlos, K. Preparation of polymer-bound trityl-hydrazines and their application in the solid phase synthesis of partially protected peptide hydrazides. *Lett. Pept. Sci.* **2**, 315–318 (1996).

190. Zheng, J. S., Tang, S., Qi, Y. K., Wang, Z. P. & Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **8**, 2483–2495 (2013).
191. Wang, Z., Rejtar, T., Zhou, Z. S. & Karger, B. L. Desulfurization of cysteine-containing peptides resulting from sample preparation for protein characterization by mass spectrometry. *Rapid Commun. Mass Spectrom.* **24**, 267–275 (2010).
192. Thompson, R. E. *et al.* Trifluoroethanethiol: An additive for efficient one-pot peptide ligation - Desulfurization chemistry. *J. Am. Chem. Soc.* **136**, 8161–8164 (2014).
193. Huang, Y. C. *et al.* Synthesis of l-and d-Ubiquitin by One-Pot Ligation and Metal-Free Desulfurization. *Chem. - A Eur. J.* **22**, 7623–7628 (2016).
194. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6705–10 (1998).
195. Vila-Perelló, M. *et al.* Streamlined expressed protein ligation using split inteins. *J. Am. Chem. Soc.* **135**, 286–292 (2013).
196. Shah, N. H. & Muir, T. W. Inteins : nature ' s gift to protein chemists. *Chem. Sci.* 446–461 (2014). doi:10.1039/c3sc52951g
197. Southworth, M. W. *et al.* Control of protein splicing by intein fragment reassembly. **17**, 918–926 (1998).
198. Yamazaki, T., Otomo, T. & Oda, N. Segmental Isotope Labeling for Protein NMR Using Peptide Splicing. **7863**, 5591–5592 (1998).
199. Lew, B. M., Mills, K. V, Paulus, H., Natl, H. P. & Sci, A. Protein Splicing in Vitro with a Minimal Intein *. 15887–15891 (1998).
200. Shah, N. H. & Muir, T. W. Split Inteins : Nature ' s Protein Ligases. *Isr. J. Chem.* **51**, 854–861 (2013).
201. Wu, H., Hu, Z. & Liu, X.-Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci.* **95**, 9226–9231 (1998).
202. Iwai, H., Züger, S., Jin, J. & Tam, P. H. Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett.* **580**, 1853–1858 (2006).
203. Scott, C. P., Abel-santos, E., Wall, M., Wahnnon, D. C. & Benkovic, S. J. Production of cyclic peptides and proteins in vivo. *Pnas* (1999).
204. Young, T. S. *et al.* Evolution of cyclic peptide protease inhibitors. **108**, 11052–11056 (2011).
205. Naumann, T. A., Tavassoli, A. & Benkovic, S. J. Genetic Selection of Cyclic Peptide Dam Methyltransferase Inhibitors. 194–197 (2008). doi:10.1002/cbic.200700561
206. Zettler, J., Schütz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett.* **583**, 909–914 (2009).
207. Stevens, A. J. *et al.* Design of a Split Intein with Exceptional Protein Splicing Activity. *J.*

- Am. Chem. Soc.* **138**, 2162–2165 (2016).
208. David, Y., Vila-Perelló, M., Verma, S. & Muir, T. W. Chemical tagging and customizing of cellular chromatin states using ultrafast trans-splicing inteins. *Nat. Chem.* **7**, 394–402 (2015).
 209. Shah, N. H., Eryilmaz, E., Cowburn, D. & Muir, T. W. Extein Residues Play an Intimate Role in the Rate-Limiting Step of Protein *Trans* -Splicing. *J. Am. Chem. Soc.* **135**, 5839–5847 (2013).
 210. Stevens, A. J. *et al.* A promiscuous split intein with expanded protein engineering applications. *Proc. Natl. Acad. Sci.* **114**, 8538–8543 (2017).
 211. Lockless, S. W. & Muir, T. W. Traceless protein splicing utilizing evolved split inteins. *Pnas* **2009**, (2009).
 212. Thiel, I. V., Volkmann, G., Pietrokovski, S. & Mootz, H. D. An atypical naturally split intein engineered for highly efficient protein labeling. *Angew. Chemie - Int. Ed.* **53**, 1306–1310 (2014).
 213. Brunauer, R. H Uman M Esenchymal S Tem C Ells : (2006).
 214. Berdasco, M. & Esteller, M. DNA methylation in stem cell renewal and multipotency. *Stem Cell Res. Ther.* **2**, (2011).
 215. Condic, M. L. Totipotency: What It Is and What It Is Not. *Stem Cells Dev.* **23**, 796–812 (2014).
 216. Marikawa, Y. & Alarcón, V. B. Establishment of trophoctoderm and inner cell mass lineages in the mouse embryo. *Mol. Reprod. Dev.* **76**, 1019–1032 (2009).
 217. Young, R. A. Control of the embryonic stem cell state. *Cell* **144**, 940–954 (2011).
 218. Levine, M. Control of the embryonic stem cell state. (2003). doi:10.1038/nature01763
 219. Fuda, N. J., Ardehali, M. B. & Lis, J. T. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186–192 (2009).
 220. Li, B., Carey, M. & Workman, J. L. The Role of Chromatin during Transcription. *Cell* **128**, 707–719 (2007).
 221. Nichols, J. *et al.* Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4 Jennifer. **95**, 379–391 (1998).
 222. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126–140 (2003).
 223. Mitsui, K. *et al.* Nanog is required for maintenance of pluripotency in ES cells.pdf. **113**, 631–642 (2003).
 224. Tee, W.-W. & Reinberg, D. Chromatin features and the epigenetic regulation of pluripotency states in ESCs. *Development* **141**, 2376–2390 (2014).
 225. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Giacomo Cavalli1, *. Genome Regulation by Polycomb and Trithorax Proteins. *Cell* **128**, 107–110 (2007).

226. Shilatifard, A. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu. Rev. Biochem.* **81**, 65–95 (2012).
227. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
228. Voigt, P., Tee, W. W. & Reinberg, D. A double take on bivalent promoters. *Genes Dev.* **27**, 1318–1338 (2013).
229. Calo, E. & Wysocka, J. Modification of Enhancer Chromatin: What, How, and Why? *Mol. Cell* **49**, 825–837 (2013).
230. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315–326 (2006).
231. Lin, B. *et al.* Global analysis of H3K4me3 and H3K27me3 profiles in glioblastoma stem cells and identification of SLC17A7 as a bivalent tumor suppressor gene. *Oncotarget* **6**, 5369–5381 (2015).
232. Cui, K. *et al.* Chromatin Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of Bivalent Genes during Differentiation. *Cell Stem Cell* **4**, 80–93 (2009).
233. Voigt, P. *et al.* Asymmetrically modified nucleosomes. *Cell* **151**, 181–193 (2012).
234. Li, Q., Lian, S., Dai, Z., Xiang, Q. & Dai, X. BGDB: A database of bivalent genes. *Database* **2013**, 1–7 (2013).
235. Rugg-Gunn, P. J., Cox, B. J., Ralston, A. & Rossant, J. Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. *Proc. Natl. Acad. Sci.* **107**, 10783–10790 (2010).
236. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, (2008).
237. Grzybowski, A. T., Chen, Z. & Ruthenburg, A. J. Calibrating ChIP-Seq with Nucleosomal Internal Standards to Measure Histone Modification Density Genome Wide. *Mol. Cell* **58**, 886–899 (2015).
238. Shema, E. *et al.* Single-Molecule Decoding of Combinatorially-Modified Nucleosomes. *Hhs* **352**, 717–721 (2016).
239. Montgomery, N. D. *et al.* The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr. Biol.* **15**, 942–947 (2005).
240. Aranda, S., Mas, G. & Di Croce, L. Regulation of gene transcription by Polycomb proteins. *Sci. Adv.* **1**, 1–16 (2015).
241. Papp, B. & Müller, J. Histone trimethylation and the maintenance of transcriptional {ON} and {OFF} states by {trxG} and {PcG} proteins. *Genes Dev.* **20**, 2041–2054 (2006).
242. Schwartz, Y. B. *et al.* Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* **38**, 700–705 (2006).

243. Ferrari, K. J. *et al.* Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Mol. Cell* **53**, 49–62 (2014).
244. Piunti, A. & Shilatifard, A. Epigenetic balance of gene expression by polycomb and compass families. *Science* (80-.). **352**, (2016).
245. Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762–767 (2009).
246. Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Mol. Cell* **15**, 57–67 (2004).
247. Nekrasov, M., Wild, B. & Müller, J. Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. *EMBO Rep.* **6**, 348–353 (2005).
248. Pasini, D., Bracken, A. P., Jensen, M. R., Denchi, E. L. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* **23**, 4061–4071 (2004).
249. Pasini, D. *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306–310 (2010).
250. Peng, J. C. *et al.* Jarid2/Jumonji Coordinates Control of PRC2 Enzymatic Activity and Target Gene Occupancy in Pluripotent Cells. *Cell* **139**, 1290–1302 (2009).
251. Cao, R. *et al.* Role of hPHF1 in H3K27 Methylation and Hox Gene Silencing. *Mol. Cell Biol.* **28**, 1862–1872 (2008).
252. Walker, E. *et al.* Polycomb-like 2 Associates with PRC2 and Regulates Transcriptional Networks during Mouse Embryonic Stem Cell Self-Renewal and Differentiation. *Cell Stem Cell* **6**, 153–166 (2010).
253. Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V. & Reinberg, D. Ezh2 Requires PHF1 To Efficiently Catalyze H3 Lysine 27 Trimethylation In Vivo. *Mol. Cell Biol.* **28**, 2718–2731 (2008).
254. Endoh, M. *et al.* Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. *PLoS Genet.* **8**, (2012).
255. Endoh, M. *et al.* Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135**, 1513–1524 (2008).
256. Farcas, A. M. *et al.* KDM2B links the polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *Elife* **2012**, 1–26 (2012).
257. He, J. *et al.* Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat. Cell Biol.* **15**, 373–384 (2013).
258. Wei, J., Zhai, L., Xu, J. & Wang, H. Role of Bmi1 in H2A ubiquitylation and Hox gene silencing. *J. Biol. Chem.* **281**, 22537–22544 (2006).
259. Isono, K. *et al.* SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* **26**, 565–577 (2013).




260. Kaustov, L. *et al.* Recognition and specificity determinants of the human Cbx chromodomains. *J. Biol. Chem.* **286**, 521–529 (2011).
261. Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664–678 (2012).
262. Schoeftner, S. *et al.* Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* **25**, 3110–3122 (2006).
263. Blackledge, N. P. *et al.* Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* **157**, 1445–1459 (2014).
264. Cooper, S. *et al.* Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment. *Cell Rep.* **7**, 1456–1470 (2014).
265. Kalb, R. *et al.* Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* **21**, 569–571 (2014).
266. Schuettengruber, B., Martinez, A. M., Iovino, N. & Cavalli, G. Trithorax group proteins: Switching genes on and keeping them active. *Nat. Rev. Mol. Cell Biol.* **12**, 799–814 (2011).
267. Byrd, K. N. & Shearn, A. ASH1, a Drosophila trithorax group protein, is required for methylation of lysine 4 residues on histone H3. *Pnas* **100**, 11535–40 (2003).
268. Dou, Y. *et al.* Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* **121**, 873–885 (2005).
269. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
270. Wu, M. *et al.* Molecular Regulation of H3K4 Trimethylation by Wdr82, a Component of Human Set1/COMPASS. *Mol. Cell. Biol.* **28**, 7337–7344 (2008).
271. Hu, D. *et al.* The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. *Mol. Cell. Biol.* **33**, 4745–4754 (2013).
272. Azuara, V. *et al.* Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8**, 532–538 (2006).
273. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
274. Pan, G. *et al.* Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. *Cell Stem Cell* **1**, 299–312 (2007).
275. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–353 (2006).
276. Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M. & Helin, K. The Polycomb Group Protein Suz12 Is Required for Embryonic Stem Cell Differentiation. *Mol. Cell. Biol.* **27**, 3769–3779 (2007).

277. Leeb, M. *et al.* Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* **24**, 265–276 (2010).
278. Vermeulen, M. *et al.* Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* **131**, 58–69 (2007).
279. Lauberth, S. M. *et al.* H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152**, 1021–1036 (2013).
280. Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714–717 (2007).
281. Schmitges, F. W. *et al.* Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol. Cell* **42**, 330–341 (2011).
282. Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731–734 (2007).
283. Rhee, H. S., Bataille, A. R., Zhang, L. & Pugh, B. F. Subnucleosomal structures and nucleosome asymmetry across a genome. *Cell* **159**, 1377–1388 (2014).
284. Li, S. & Shogren-Knaak, M. A. Cross-talk between histone H3 tails produces cooperative nucleosome acetylation. *Proc. Natl. Acad. Sci.* **105**, 18243–18248 (2008).
285. Munari, F. *et al.* Methylation of lysine 9 in histone H3 directs alternative modes of highly dynamic interaction of heterochromatin protein hHP1 β with the nucleosome. *J. Biol. Chem.* **287**, 33756–33765 (2012).
286. Yan, L. Z. & Dawson, P. E. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J. Am. Chem. Soc.* **123**, 526–533 (2001).
287. Wan, Q. & Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: A powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew. Chemie - Int. Ed.* **46**, 9248–9252 (2007).
288. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
289. Nguyen, U. T. T. *et al.* Accelerated chromatin biochemistry using dnA-barcoded nucleosome libraries. *Nat. Methods* **11**, 834–840 (2014).
290. Guidotti, N., Lechner, C. C. & Fierz, B. Controlling the supramolecular assembly of nucleosomes asymmetrically modified on H4. *Chem. Commun.* **53**, 10267–10270 (2017).
291. Subramanian, R. & Kapoor, T. M. Building Complexity: Insights into Self-Organized Assembly of Microtubule-Based Architectures. *Dev. Cell* **23**, 874–885 (2012).
292. Janke, C. & Bulinski, J. C. Post-translational regulation of the microtubule cytoskeleton: Mechanisms and functions. *Nat. Rev. Mol. Cell Biol.* **12**, 773–786 (2011).
293. Gönczy, P. Towards a molecular architecture of centriole assembly. *Nat. Rev. Mol. Cell Biol.* **13**, 425–435 (2012).

294. Ti, S. C. *et al.* Mutations in Human Tubulin Proximal to the Kinesin-Binding Site Alter Dynamic Instability at Microtubule Plus- and Minus-Ends. *Dev. Cell* **37**, 72–84 (2016).
295. Ti, S.-C., Alushin, G. M. & Kapoor, T. M. Human β -Tubulin Isoforms Can Regulate Microtubule Protofilament Number and Stability. *Dev. Cell* **47**, 175–190.e5 (2018).
296. Kleiner, R. E., Ti, S. C. & Kapoor, T. M. Site-specific chemistry on the microtubule polymer. *J. Am. Chem. Soc.* **135**, 12520–12523 (2013).
297. Berger, I. & Craig, A. MultiBac Expression System User Manual. (2010). doi:10.13140/2.1.2563.6645
298. Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F. & Berger, I. New baculovirus expression tools for recombinant protein complex production. *J. Struct. Biol.* **172**, 45–54 (2010).
299. Vijayachandran, L. S. *et al.* Robots, pipelines, polyproteins: Enabling multiprotein expression in prokaryotic and eukaryotic cells. *J. Struct. Biol.* **175**, 198–208 (2011).
300. Sano, H., Nakamura, A. & Kobayashi, S. Identification of a transcriptional regulatory region for germline-specific expression of vasa gene in *Drosophila melanogaster*. *Mech. Dev.* **112**, 129–139 (2002).
301. Yu, N. *et al.* Isolation of Functional Tubulin Dimers and of Tubulin-Associated Proteins from Mammalian Cells. *Curr. Biol.* **26**, 1728–1736 (2016).
302. Joe, P. A., Banerjee, A. & Ludueña, R. F. Roles of β -tubulin residues Ala428 and ThrA429 in microtubule formation in vivo. *J. Biol. Chem.* **284**, 4283–4291 (2009).
303. Castoldi, M. & Popov, A. V. Purification of brain tubulin through two cycles of polymerization- depolymerization in a high-molarity buffer. *Protein Expr. Purif.* **32**, 83–88 (2003).
304. Kraatz, S. *et al.* The Human Centriolar Protein CEP135 Contains a Two-Stranded Coiled-Coil Domain Critical for Microtubule Binding. *Structure* **24**, 1358–1371 (2016).
305. Tegge, W., Bonafe, C. F. S., Teichmann, A. & Erck, C. Synthesis of peptides from α - And β -tubulin containing glutamic acid side-chain linked oligo-glu with defined length. *Int. J. Pept.* **2010**, (2010).
306. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. **6**, 12–16 (2009).
307. Campbell, J. N. & Slep, K. C. $\alpha\beta$ -Tubulin and Microtubule-Binding Assays. in *Microtubule Dynamics: Methods and Protocols* (ed. Straube, A.) 87–97 (Humana Press, 2011). doi:10.1007/978-1-61779-252-6_6
308. High-resolution separation of tubulin monomers on polyacrylamide minigels. *Anal. Biochem.* **402**, 194–196 (2010).

Curriculum Vitae

Ninad Agashe

St-Sulpice, Switzerland 
+41(0)78 831 64 95 
mr.ninad.agashe@gmail.com 

Education

OCTOBER 2014 – MARCH 2019

PhD (Chemical Biology)

École polytechnique fédérale de Lausanne (EPFL), Switzerland

AUGUST 2012 – JULY 2014

Masters of Science M.S. (Biochemistry and Molecular Biology)

University of Southern California (USC), Los Angeles, USA

AUGUST 2008 – MAY 2012

Bachelor of Engineering B.E. (Biotechnology)

University of Mumbai, India

Research Experience

OCTOBER 2014 – MARCH 2019

Doctoral Assistant / EPFL, Lausanne, Switzerland

- Developing chemical biology and protein engineering tools for chromatin and microtubules
- Developing model to elucidate mechanism of enzyme regulation
- Characterization of intein splicing assays and kinetics
- Baculovirus protein expression, protein purification strategies and protein semi-synthesis of tubulin
- Immunofluorescence microscopy of microtubules

Publication

- Lechner CC., Agashe N., Fierz B., Traceless synthesis of asymmetrically modified bivalent nucleosomes, *Angew Chem Int Ed Engl* 2016, 55(8):2903-6
- Agashe N., Ebberink E., Fierz B., Exploring the function of microtubule PTMs by semi synthetic tubulin, *manuscript in preparation*

JANUARY 2013 – JULY 2014

Research Assistant / USC, Los Angeles, USA

- Structural biology and analysis of molecular dynamics of amyloid and membrane proteins
- Structural characterization of early amyloid oligomers using circular dichroism (CD) spectroscopy, electron microscopy, dynamic light scattering (DLS) and solid-state NMR

Awards

- Best Poster Prize, NCCR Chemical Biology Symposium (June 2016)
- J N TATA Scholarship (\$20000) for Masters students (August 2012)

