2010

# Mechanistic Insights into the Stimulation of Dot1L-Mediated Methylation of Histone H3 by Semisynthetically Ubiquitylated Histone H2b

Robert Kendall McGinty


# MECHANISTIC INSIGHTS INTO THE STIMULATION OF DOT1L-MEDIATED METHYLATION OF HISTONE H3 BY SEMISYNTHETICALLY UBIQUITYLATED HISTONE H2B

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Robert Kendall McGinty

June 2010

# MECHANISTIC INSIGHTS INTO THE STIMULATION OF DOT1L-MEDIATED METHYLATION OF HISTONE H3 BY SEMISYNTHETICALLY UBIQUITYLATED HISTONE H2B

Robert Kendall McGinty, Ph.D.

The Rockefeller University, 2010

Post-translational modification of histones plays an integral role in regulation of chromatin-templated processes through modulation of chromatin structure and function. One such modification, ubiquitylation of histone H2B on lysine 120 (uH2B), has been correlated with enhanced methylation of lysine 79 (K79) of histone H3 by K79-specific methyltransferase, disruptor of telomeric silencing-like (Dot1L/KMT4). However, the specific function of uH2B in this crosstalk pathway was not understood, in part due to the challenges associated with isolating or generating homogeneously ubiquitylated H2B for use in biochemical studies. As both modifications are integral to transcriptional regulation and DNA damage repair, full elucidation of their functions is critical to understanding their roles in development and disease. In this thesis, a chemical strategy is presented for the preparation of native uH2B. Two traceless orthogonal expressed protein ligation (EPL) reactions were used for this purpose, one employing a photolytically removable ligation auxiliary, and the other, a cysteine-mediated ligation followed by a desulfurization to restore the native sequence. Reconstitution of semisynthetic uH2B into chemically defined nucleosomes, followed by biochemical analysis, revealed a direct role for uH2B in the stimulation of Dot1L-mediated methylation of H3K79. Although

recruitment of Dot1L to the nucleosomal surface by uH2B could be excluded, comprehensive mechanistic analysis was precluded by systematic limitations in the ability to generate native uH2B in large-scale. To overcome this shortcoming, a highly optimized synthesis of ubiquitylated H2B bearing a Gly76Ala point mutation (uH2BG76A) was developed, yielding tens of milligrams of ubiquitylated protein. This mutant was indistinguishable from native uH2B by Dot1L, allowing for detailed studies of the resultant *trans*-histone crosstalk pathway. Kinetic and structure activity relationship analyses using uH2BG76A suggest a non-canonical role for ubiquitin in the enhancement of the chemical step of H3K79 methylation. This enhancement likely results from an allosteric change in the nucleosome and/or Dot1L following H2B ubiquitylation. Current work is aimed at further elucidation of the molecular mechanism of uH2B-mediated stimulation of Dot1L and the role of uH2B in other chromatin templated-processes.

This thesis is dedicated to my late father, Dr. Dean McGinty, for his legacy of compassion and service and his commitment to education.

**Acknowledgements**

Firstly, I would like to thank my advisor and mentor, Professor Tom W. Muir. This thesis is a testament to his intoxicating curiosity and scientific rigor.

I would like to thank the members of my Faculty Advisory Committee, Drs. C. David Allis, Robert G. Roeder, and Scott C. Blanchard, for their thoughtful suggestions and guidance throughout my thesis work. I would also like to recognize Dr. Hidde L. Ploegh from the Whitehead Institute for Biomedical Research for serving as an external examiner for my thesis defense.

I would like to thank all current and former members of the Laboratory of Synthetic Protein Chemistry. Without their insights, selflessness, and patience, none of this would have been possible. I would like to extend a special acknowledgement to Dr. Champak Chatterjee for sharing his expertise in solution phase, peptide, and protein chemistry. His contributions launched the work in this thesis and my career as a scientist. I would also like to thank several other members of the laboratory, Drs. Kyle P. Chiang, Matthew R. Pratt, Edmund C. Schwartz, Robert R. Flavell, and Michael E. Hahn, whose ideas and encouragement factored heavily in both my thesis work and my growth as a scientist. I would like to thank Dr. Maja Köhn and Melanie Yuen for their contributions to projects included in this thesis.

None of the work in this thesis would have been possible without the assistance of superb collaborators. Dr. Jaehoon Kim in the Laboratory of Biochemistry and Molecular Biology was an ideal collaborator, a model that I strive to emulate. His willingness to share reagents and advice embodies the scientific spirit. I would also like to thank a group of collaborators in the Laboratory of Chromatin Biology and Epigenetics, including Drs. Alex J. Ruthenburg, Jason C. Tanny, and Sung-Hee Ahn-Upton and Ms. Sarah J. Whitcomb. Together with Dr. Kim, their guidance helped me to bridge the gap between protein chemistry and chromatin biology. I would like to thank Mr. Matthew J. Bick in the Laboratory of Molecular Biophysics, and his advisor Dr. Seth A. Darst for their assistance in crystallographic studies and Mr. Brad R. Rosenberg for assistance with phosphorimaging analysis. I would like to thank Drs. Haiteng Deng, Joseph P. Fernandez, and Haiqiang Yu at the Rockefeller University Proteomics Resource Center for their assistance in the mass spectrometric analysis of biochemical assays.

I would like to thank the Tri-Institutional MD-PhD Program and the Dean's Office at the Rockefeller University for the opportunity to perform this work. A special thanks is due to Dr. Olaf S. Andersen, Director of the MD-PhD program, for his guidance and his dedication to improving the program.

I could not have done any of this without the constant support and understanding of my mother, Mrs. Christine McGinty, and my brother, Mr. Mitchell McGinty, and my in-laws Mrs. Ilene Herz and Mr. Jerry Herz. And above all, I would like to thank my wife and partner, Dr. Katrina McGinty. Her unwavering encouragement and patience have made this all possible. I cannot begin to describe my appreciation and love for her.

## TABLE OF CONTENTS

## LIST OF FIGURES

**Figure** **Page**

### Chapter 1



### Chapter 2

**Chapter 3**



**Chapter 4**

**Chapter 5**



**Chapter 6**

## Chapter 1: Introduction

Under the central dogma of modern biology, deoxyribonucleic acid (DNA) is charged with the responsibility of storing the genetic material that is essential to the maintenance of life (Crick, 1958). Each time a cell divides, the complete sequence of its DNA, its genome, is propagated by semi-conservative replication (Meselson and Stahl, 1958), thus bestowing its entire genetic potential on each daughter cell. Through this framework, the genetic makeup of a cell or organism, its genotype, determines the observable and inheritable characteristics of that cell or organism, its phenotype. Mutations in gene sequences can cause phenotypic changes, leading to the variation essential for speciation and evolution through natural selection.

However, for nearly a century, inheritable phenotypic changes have been observed in the absence of genotypic changes (Allis et al., 2007). The discovery of position effect variegation (PEV) in *Drosophila* represents one of the early descriptions of such a phenomenon. In 1930, Muller described inheritable phenotypic changes in the eyes of *Drosophila*, following X-ray induced DNA translocations (Muller, 1930). The dosage of each gene was appropriate but the chromosomal locations of some genes changed. Over the next several decades, it was demonstrated that the translocation placed the a gene coding for eye color next to transcriptionally-silenced condensed chromatin, resulting in its variegated expression (Hannah, 1951). This phenomenon, as well as others, led to the inception of the field of epigenetics – the study of changes in gene functions that cannot be attributed to changes to the DNA sequence of genes (Allis et al., 2007).

### 1.1. Chromatin: the physiologic state of the eukaryotic genome

It is impossible to understand epigenetics without first considering the physiologic state of the genome. In organisms ranging from yeast to humans, DNA is packaged inside the nucleus of cells in a polymeric complex called chromatin (Figure 1.1). The fundamental unit of chromatin is the nucleosome, containing an octamer with two copies of each of the core histone proteins, histone H2A (H2A), histone H2B (H2B), histone H3 (H3), and histone H4 (H4), wrapped by 146 base pairs (bps) of DNA (Luger et al., 1997a). Although the composition of the nucleosome has long since been established (Kornberg, 1977), more recent X-ray crystallographic investigation of the nucleosome has painted a high-resolution picture of the histone octamer and its interactions with DNA (Davey et al., 2002; Luger et al., 1997a; Schalch et al., 2005). Each histone is comprised of a uniquely structured α-helical histone fold domain, which contributes to the central core of the octamer, from which an N-terminal tail protrudes (Luger et al., 1997a). DNA is wrapped in 1.65 turns of a left-handed superhelix around this central protein core (Luger et al., 1997a).

The nucleosomal unit repeats throughout the genome every 160-240 bp (McGhee and Felsenfeld, 1980), with each nucleosome connected to an adjacent nucleosome through a stretch of intervening DNA. This intervening DNA, or linker DNA, is often found in complex with a linker histone – for example, mammalian cells have up to six subtypes of linker histone H1 (Georgel and Hansen, 2001). Linker histones may function in heterochromatic condensation, but whether or not they are essential remains elusive due to compensation between redundant subtypes (Georgel and Hansen, 2001).

Arrays of nucleosomes are progressively condensed through a hierarchy of higher-order structures, ranging from extended stands to condensed mitotic chromosomes (Allis et al., 2007).

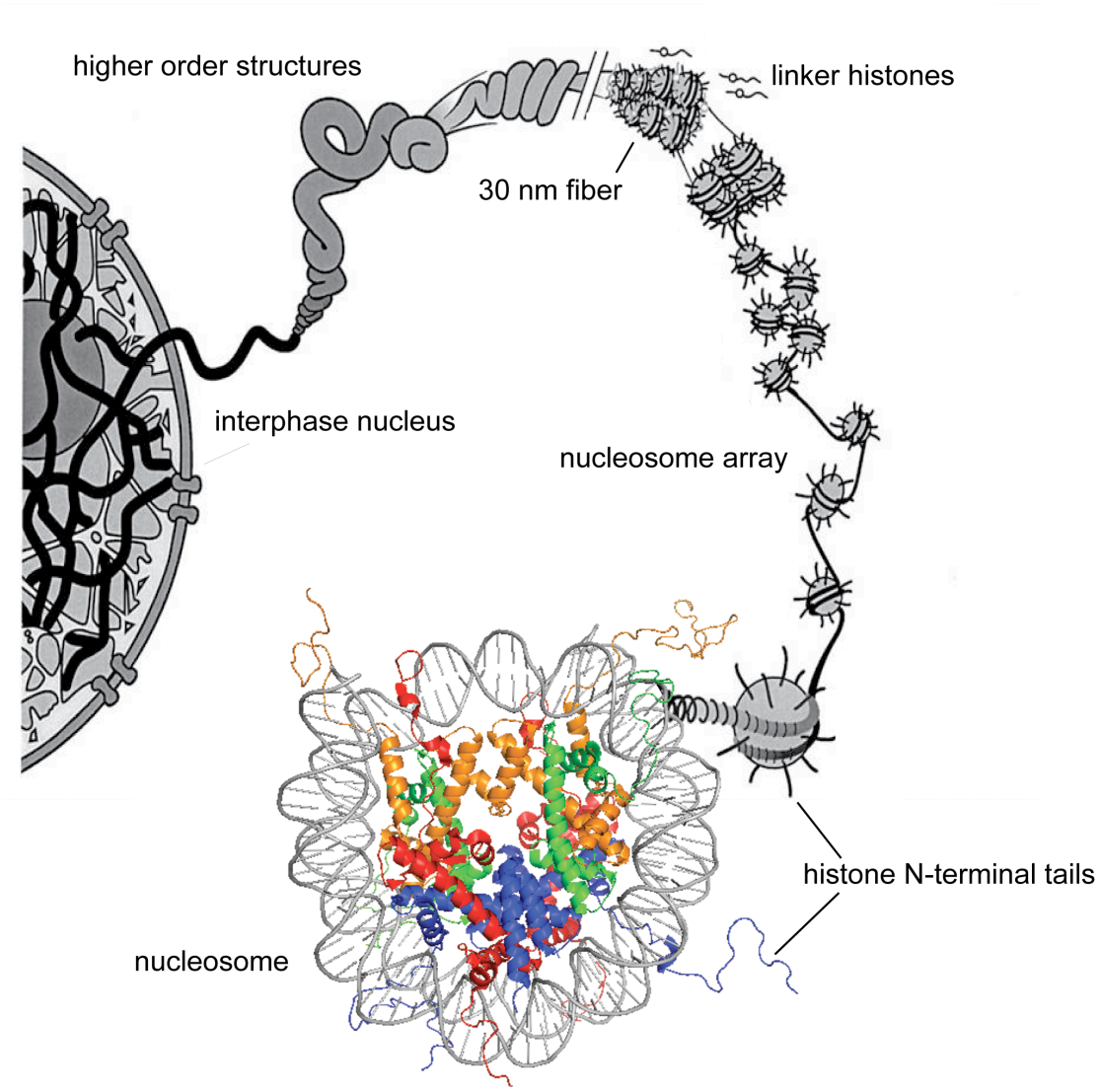



**Figure 1.1. The nucleosome is the fundamental unit of chromatin.** The nucleosomal structure (1KX5) (Davey et al., 2002) rendered with Pymol. H2A, H2B, H3, and H4 are displayed in red, blue, orange, and green, respectively. Chains of nucleosomes are condensed with linker histones into higher order chromatin structures. Figure adapted from a previously published illustration (Hansen, 2002).

Much remains to be elucidated regarding the composition of these higher-order chromatin structures. This is exemplified by the 30 nm fiber, the first level of chromatin compaction. While the existence of the 30 nm fiber is generally agreed upon, the structural composition of the fiber remains controversial (Robinson and Rhodes, 2006). Two models have been proposed: a one-start fiber formed from a simple solenoid pattern of nucleosomes (Huynh et al., 2005; Robinson et al., 2006), and a two-start fiber with interdigitating nucleosomes (Dorigo et al., 2004; Schalch et al., 2005). It is formally possible that both could co-exist in different environments *in vivo*. Further compaction, beyond that of the 30 nm fiber, is even less well understood. This relative ignorance is not a product of lack of investigation; to the contrary, it truly underscores the complexity of even the simplest forms of chromatin structure.

From microscopic investigation, chromatin within the interphase nucleus has been subdivided into two morphological states, heterochromatin and euchromatin. Heterochromatin consists of densely compacted chromatin that is traditionally considered inaccessible and therefore transcriptionally inactive. In contrast, euchromatin is more loosely compacted, and consequently highly accessible and transcriptionally active (Allis et al., 2007). In recent years, the inaccessibility of heterochromatin has come under scrutiny. While highly condensed and transcriptionally silenced, heterochromatin is not inaccessible. In fact, proteins required for the maintenance of condensed chromatin can freely diffuse into and out of heterochromatic areas (Phair et al., 2004). A heightened understanding of higher-order chromatin structures will be critical to the understanding of epigenetic mechanisms.

## 1.2. Epigenetic phenomena

The field of epigenetics formed around a series of phenotypic phenomena observed in the absence of genotypic changes, beginning as early as 1930 with the first report of PEV (Muller, 1930). This phenomenon was not restricted to phenotypes involving eye color; rather, it was a general property of the relative location of any gene within the linear sequence of chromosomes (Hannah, 1951). Similar phenotypic changes were reported in maize, caused by the movement of transposable elements. These segments of DNA have the intrinsic ability to jump between different loci within the genome, causing heritable, but often reversible changes to the expression of neighboring genes (McClintock, 1950).

Inheritance of sex chromosomes presents another epigenetic quandary. While females inherit two X chromosomes, one from each parent, males only inherit a maternal X chromosome. The X chromosome contains many genes required for sex determination, but it also carries many housekeeping genes that if unregulated, could have disastrous consequences for a cell (Allis et al., 2007). Therefore, an equalization of the expression of genes residing in the X chromosome between males and females, despite a different copy number, is required – a process called dosage compensation. Dosage compensation necessitates two functions: 1) the ability to count X chromosomes, and 2) the ability to adjust the level of expression accordingly. Surprisingly, three different mechanisms of dosage compensation have been observed in different model organisms: two-fold up regulation of X-linked genes in the male worm, two-fold suppression of X-linked genes in the female fly, and random inactivation of a single X chromosome in the female mammal (Chow and Heard, 2009; Ercan and Lieb, 2009; Gelbart and Kuroda, 2009). Importantly, the mechanisms are reversible, so that following reproduction, the correct

compensation can be applied anew to the next generation, based on the genotype of the progeny (Chow and Heard, 2009).

Contrary to genes on sex chromosomes, most genes located on autosomal chromosomes are expressed equally, regardless of their parental origins. However, the expression of a small subset of genes is restricted to either the maternal- or paternal-inherited gene in both males and females, through a process call imprinting (Allis et al., 2007). One of the earliest observations of imprinting involved mice with a genetic deletion on chromosome 17 resulting in a "hairpin" tail. Progeny inheriting the deletion were viable only when the deletion was inherited from the father and died during embryogenesis when the deletion was inherited from the mother (Johnson, 1974). This led to the suggestion that the deleted region of chromosome 17 may be active only on the maternally inherited chromosome (McClaren, 1979). While affecting a small fraction of the genome, imprinting is of great consequence to human medical genetics as a number of disorders are inherited through imprinted genes, such as Prader Willi and Angelman syndromes (Allis et al., 2007).

Perhaps much of the attraction of epigenetics arises from the mysteries of development of multi-cellular organisms. Through exponential rounds of division without changes to the DNA sequence itself, how can one cell lead to the formation of a complex organism, complete with organ systems, and nearly countless cell morphologies and functions? Recent research in epigenetics has focused on elucidating the mechanisms underlying these and other related phenomena. Not surprisingly, an overlapping set of fundamental molecular mechanisms have been discovered that form the foundation for choreographing the complex expression of the genome.

### 1.3. Epigenetic mechanisms

Most epigenetic phenomena result from a complex interplay between a small set of molecular mechanisms, including post-translational modification of histones, substitution of canonical histones for sequence variants, nucleosomal remodeling, DNA methylation, and siRNA-mediated transcriptional silencing (Allis et al., 2007). These individual mechanisms act in concert to control gene expression, with each mechanism contributing to the initiation and/or maintenance of the epigenetic signal. One of the oldest epigenetic molecular mechanisms, found in organisms ranging from yeast to vertebrates, is the post-translational modification (PTM) of histones (Allis et al., 2007; Kouzarides, 2007). PTMs are chemical modifications to proteins subsequent to ribosomal synthesis that increase the chemical complexity afforded by the standard 20 amino acids (Walsh, 2005). As most PTMs are reversible, they allow transient regulation of protein activity and are used to control complex biochemical systems. Histones harbor an extraordinary density of PTMs, especially within their N-terminal tails, including lysine and arginine methylation, lysine acetylation, serine and threonine phosphorylation, and lysine ubiquitylation and sumoylation (Allis et al., 2007; Kouzarides, 2007). Functions of specific histone PTMs will be discussed in detail in the following sections.

In addition to PTM of histones, the chemical composition of chromatin can be reversibly controlled through the incorporation of histone variant proteins (Allis et al., 2007). These proteins are sequence variants of the canonical histone sequences that can replace their respective core histones within the nucleosome. Histone variants can occupy specific locations within the linear sequence of chromosomes, demarcate specific regions of active and inactive chromatin, and signal in diverse chromatin-templated processes (Allis et al., 2007). For example, H3.3 and H2A.Z are associated with actively transcribed genes, localized to the open-reading

frames and promoters, respectively, suggesting a role in transcriptional regulation (Henikoff and Ahmad, 2005; Kamakaka and Biggins, 2005); conversely, macroH2A is found in the inactive X chromosome of mammals, possibly playing a role in dosage compensation (Costanzi and Pehrson, 1998).

Collectively, histone variants present chemical changes that can potentially be interpreted similarly to histone PTMs. In fact, incorporation of variants has been proposed as a mechanism for erasing a record of histone PTMs (Schwartz and Ahmad, 2005). Furthermore, histone variants harbor a set of PTMs that are partially distinct from those found in canonical histones (Allis et al., 2007). This is exemplified by the phosphorylation of H2A.X coincident with DNA double strand breaks (Thambirajah et al., 2009).

The role of incorporating histone variants into chromatin falls to nucleosomal remodeling complexes (Clapier and Cairns, 2009). These molecular machines can assemble and disassemble nucleosomes, and reposition nucleosomes translationally along DNA. As the nucleosome is the first barrier to DNA for many DNA binding proteins, nucleosomal composition and positioning is relevant to almost all chromatin-templated activities. This is exemplified in transcriptional activation, where significant nucleosomal ejection and sliding occurs surrounding the transcription start site, the promoter, and throughout the coding region (Clapier and Cairns, 2009), allowing the polymerase complex to assemble and proceed through a gene. The function of many chromatin remodelers is likely to be integrally linked to specific patterns of histone PTMs. This is reflected in the prevalence of modification-specific histone binding modules within remodeling complexes (Clapier and Cairns, 2009).

DNA methylation also plays an integral role in many epigenetic phenomena. Position 5 of the cytidine pyrimidine ring is methylated by dedicated DNA methyltransferases. Generally,

DNA methylation is found in CpG rich regions of repeating elements and CpG islands in promoters, leading to silencing in both cases (Allis et al., 2007). Two types of DNA methyltransferases are required for placement and inheritance of methylation patterns. *De novo* DNA methyltransferases modify unmethylated loci within the genome, while maintenance DNA methyltransferases preferentially act on hemimethylated loci, allowing for inheritance of the mark following replication (Bestor and Ingram, 1983; Okano et al., 1999). Asymmetric methylation of maternally and paternally inherited genes assists in the establishment of imprinting (Allis et al., 2007).

Methylation of DNA can regulate the binding of DNA interacting proteins both positively and negatively. Certain transcription factors are unable to interact with methylated promoters, thereby passively repressing expression of methylated genes (Watt and Molloy, 1988). Conversely, a family of proteins harboring methyl CpG binding domains (MBDs) interact with methylated DNA, often in a sequence-specific manner, to actively repress transcription (Allis et al., 2007). As described for remodeling complexes, DNA methyltransferases rely heavily on histone PTMs. In mammals, deletion of methyltransferases establishing either H3K9me or K27me disrupts DNA methylation patterns (Lehnertz et al., 2003; Vire et al., 2006). Furthermore, many histone remodeling complexes, including ATRX, assist in maintaining appropriate DNA methylation patterns (Gibbons et al., 2000).

Recently, small non-coding RNAs have been implicated in the maintenance of silenced chromatin. These non-coding RNAs are processed into 21-27 bp small interfering RNAs (siRNAs) by Dicer and can mediate both post-transcriptional and transcriptional gene silencing (TGS) (Allis et al., 2007). siRNA-mediated TGS employs the RNA-Induced Transcriptional Silencing (RITS) complex. The RITS complex, through complementary interaction with a

nascent transcript (Buhler et al., 2006), or possibly through complementarity with the genomic DNA itself, targets repressive complexes to genomic loci to initiate the formation of heterochromatin. This process is best understood in *S. pombe*, where the RITS complex recruits H3K9 methyltransferase, Clr4, and HP1 homolog, Swi6, which binds H3K9me, and so facilitates the spread of heterochromatin (Hall et al., 2002; Volpe et al., 2002). In humans, siRNA-mediated TGS acts cooperatively with histone PTMs and DNA methylation to establish dosage compensation and gene imprinting (Allis et al., 2007).

These epigenetic mechanisms alter the structure of chromatin through nucleosomal positioning and control of chromatin condensation, and fine-tune the recruitment of chromatin-associated proteins. These combined effects allow the control of gene expression programs in embryogenesis and cell fate determination, and when disrupted, can result in the pathogenesis of many disease states.

### 1.4. Post-translational modification of histones

At least nine distinct types of histone PTMs have been observed. Certain types have been well characterized, such as acetylation, methylation of lysines and arginines, and phosphorylation, while current understanding of other types, including ubiquitylation, sumoylation, ADP ribosylation, deimination, proline isomerization, and proteolysis is incomplete (Duncan et al., 2008; Kouzarides, 2007). While acetylation is synonymous with transcriptionally active chromatin, other modifications, including methylation (Martin and Zhang, 2005) and ubiquitylation (Shilatifard, 2006) correlate with active and repressed chromatin in a position-specific manner. In addition to transcriptional regulation, histone PTMs function in diverse processes, including DNA replication, DNA repair, mitosis, and apoptosis (Allis et al., 2007).

Although the existence of histone PTMs has been established since the late 1960s (Allfrey et al., 1964), the mid-1990s brought the discovery of the first enzymes involved in the modification of histones. Isolation of a histone lysine acetyltransferase (HAT), Gcn5, and a histone lysine deacetylase (HDAC), Rpd3p, set the stage for reversible regulation of histone PTMs (Brownell et al., 1996; Taunton et al., 1996). Over the last decade, astonishing progress has been made in the identification and characterization of the enzymes that establish and reverse the vast array of known histone modifications (Kouzarides, 2007).

What is most compelling regarding histone PTMs is their sheer density. More than 60 amino acid side chains have been shown to carry PTMs, representing over two-thirds of the modifiable positions in the nucleosome (Kouzarides, 2007; Taverna et al., 2007). However, this number is an underestimate of the combinatorial potential of histone PTMs because some lysines can be modified by either methylation or acetylation. Furthermore, lysine methylation exists in three distinct forms: mono-, di-, and trimethylation; arginine methylation is also present is three forms: mono-, asymmetric di-, and symmetric dimethylation. In some cases each methylation state is thought to have a distinct function (Martin and Zhang, 2005), and in other cases the different states seem to be redundant (Frederiks et al., 2008). Despite this extraordinary combinatorial diversity, a smaller subset of the sum of possible PTM combinations may exist and be functionally meaningful *in vivo*. Many gene promoters in a pool of CD4+ T cells show evidence of a common set of 17 acetylations and methylations (Wang et al., 2008). Whether these modifications reside in a nucleosome simultaneously is yet to be verified. However, it underscores the complexity and combinatorial power of histone PTMs, as well as the tendency for certain patterns to arise, parallel to genomic functions.

Histone PTMs, often in a site-specific context, facilitate chromatin-templated processes through three proposed mechanisms: 1) histone PTMs can alter charge density, thereby assisting or impeding intra- and internucleosomal contacts and changing chromatin accessibility as a result; 2) histone PTMs can recruit position- and modification-specific binding modules and any associated activities; and, 3) conversely, histone PTMs can prevent the binding of chromatin associated proteins, excluding any associated activities (Figure 1.2) (Allis et al., 2007). Examples of each of the three mechanistic roles of histone PTMs will be discussed below.

Chromatin is a highly charged polymer. As such, its electrostatic surface plays a decisive role in its conformational state and oligomeric potential. PTMs, including acetylation and phosphorylation, alter intrinsic electrostatic properties of amino acid side chains and thus, have the potential to tune the conformational state of chromatin (Figure 1.2, top panel). A number of different lysines on each of the core histones are acetylated, mostly residing within their N-terminal tails (Allis et al., 2007). Hyperacetylation of histone tails controls chromatin conformation on multiple levels, weakening interactions between histone tails and nucleosomal DNA, decreasing compaction of the 30 nm fiber, and reducing oligomerization of nucleosomal arrays (Wolffe and Hayes, 1999). Furthermore, the acetylation of H4K16 alone precludes the formation of 30 nm fibers *in vitro* (Robinson et al., 2008; Shogren-Knaak et al., 2006).

Ubiquitylation has also been proposed to prevent higher order compaction of chromatin structures (Henry and Berger, 2002; Sun and Allis, 2002). However, this is not due to charge neutralization; rather, it has been speculated that given its sheer size (76 amino acids compared to 102-135 amino acids for each of the histones), ubiquitin may act as a wedge between adjacent nucleosomes, preventing chromatin condensation. Local control of chromatin conformation by

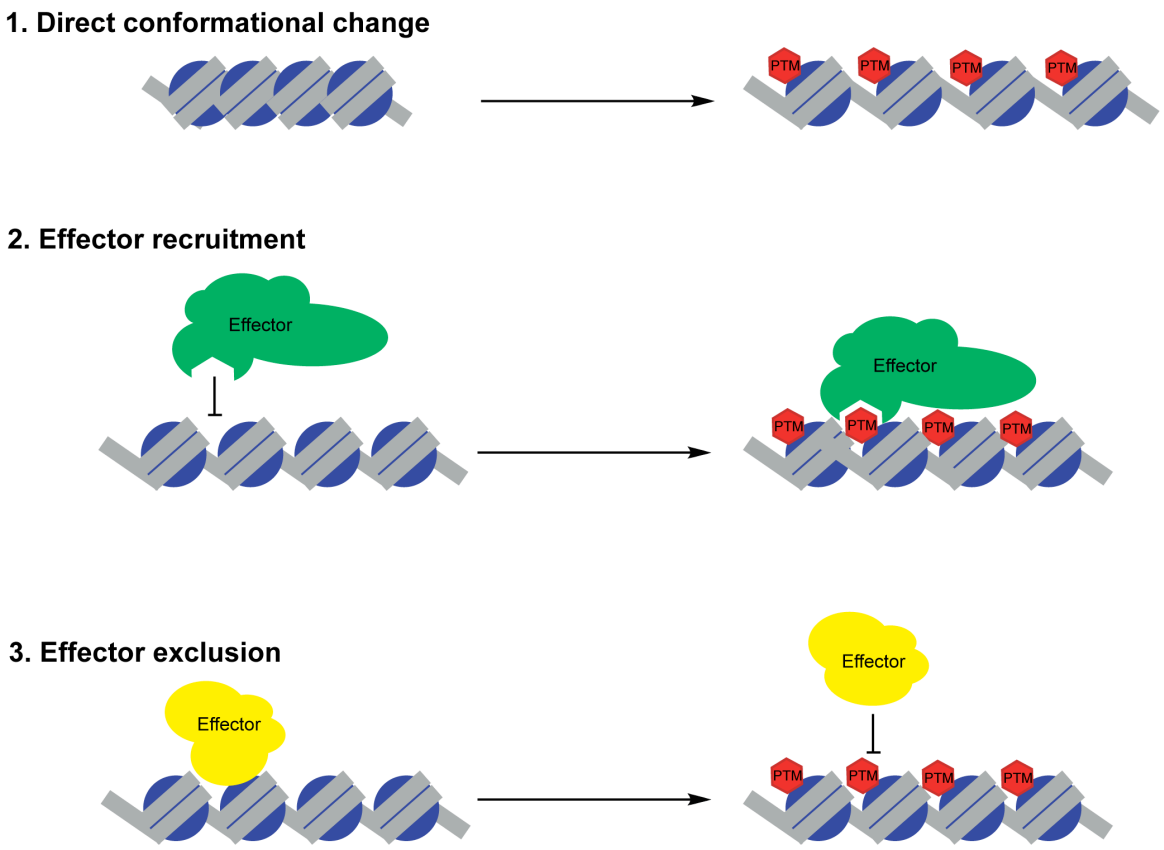



**Figure 1.2. Mechanisms of action of histone PTMs.** Histone PTMs can function through three distinct mechanisms: 1) direct conformational changes to nucleosomes and chromatin; 2) recruitment of effectors; and 3) exclusion of effectors.

PTM of histones through electrostatic or steric mechanisms presents a mechanism for regulating access of specific genomic regions to effectors of chromatin-templated processes.

Histone PTMs can also regulate chromatin indirectly through the recruitment of position- and modification-specific binding modules (Figure 1.2, middle panel). Over the past decade, a vast set of these modules have been reported that can act as intermediaries between modifications and downstream events (Taverna et al., 2007). Examples include bromodomains, the Royal superfamily of domains – collectively comprised of chromodomains, tudor domains, and MBT repeats – and 14-3-3 domains, which recognize acetyllysines, methyllysines, and phosphoserines, respectively (Dhalluin et al., 1999; Macdonald et al., 2005; Ruthenburg et al., 2007a; Zeng and Zhou, 2002). Histone PTMs can be interpreted and translated into a diverse set of biological responses by linking specific modules with an array of effectors, including transcriptional coactivators and corepressors, chromatin remodelers, and histone modifying enzymes (Ruthenburg et al., 2007b).

The modular nature of interpretation of histone PTMs permits the combinatorial recognition of multiple PTMs, potentially within one histone tail, on separate histones within a nucleosome, or in neighboring nucleosomes (Ruthenburg et al., 2007b). In the first example of cooperative recognition of histone PTMs, the tandem bromodomains of TFIID were shown to have enhanced affinity for doubly acetylated H4 peptides over either singly acetylated species, suggesting bivalent recognition within a single histone tail (Jacobson et al., 2000). Given the prevalence of multiple PTM recognition modules contained within single polypeptides or within complexes of chromatin-associated proteins, this is likely to be a universal strategy for the interpretation of multiple histone PTMs (Ruthenburg et al., 2007b).

Histone PTMs can also function through the exclusion of effectors (Figure 1.2, bottom panel). This is best exemplified by the observation of binary switches between adjacent modifications. For example, the phosphorylation of H3S10 prevents the binding of HP1 to methylated H3K9, consequently disrupting heterochromatin formation (Fischle et al., 2005; Hirota et al., 2005). A similar switch exists in H2B in yeast, where deacetylation of K11 is required for phosphorylation of S10 during apoptotic chromatin condensation (Ahn et al., 2006).

The diversity and scope of histone PTMs has led to the proposal of a histone code, through which combinatorial site-specific PTMs provide the blueprint for choreographing complex genome-templated processes (Strahl and Allis, 2000). Alternatively, the large set of histone PTMs may be more reminiscent of cytosolic PTM-regulated signaling cascades of cellular metabolism (Allis et al., 2007). The dynamic and complex nature of histone PTMs compared to the stable, four-letter genetic code, makes the histone code hypothesis difficult to validate – studying histone PTMs is more like interpreting a conversation than translating a book. However, the position- and modification-specific recruitment of binding modules, the presence of binary PTM-based switches, and the conservation of complex PTM patterns at specific functional loci lend credibility to the histone code hypothesis. Extraordinary progress has been made in the identification of novel histone PTMs, the isolation of enzymatic activities responsible for their deposition and removal, and in deciphering the mechanisms underlying their physiologic functions. Yet, much remains to be discovered to broaden our understanding of histone PTMs and their epigenetic roles in development and disease.

### 1.4.1. Ubiquitylation of histones

Monoubiquitylation of proteins was discovered in the context of H2A (Goldknopf and Busch, 1977; Hunt and Dayhoff, 1977). Soon after, monoubiquitylation of H2B was established (West and Bonner, 1980). Unlike small PTMs, including acetylation, methylation, and phosphorylation, ubiquitylation results from the acylation of a lysine side chain with the 76 amino acid protein ubiquitin. The C-terminal glycine (G76) of ubiquitin is attached to the ε-$NH_2$ of a lysine through an amide bond, forming a branched isopeptide linkage (Hershko et al., 2000). Due to the size of ubiquitin relative to small chemical PTMs, ubiquitylation is considered an information-rich modification (Walsh, 2005).

Ubiquitin is expressed as a head-to-tail repeat of five ubiquitin sequences (Figure 1.3). Ubiquitin monomers can be mobilized from the repeats through the action of deubiquitylating enzymes (DUBs) and ubiquitin C-terminal hydrolases (UCHs) (Walsh, 2005). The ubiquitin monomers are then activated and transferred to substrates through the concerted effort of ubiquitin activating enzymes (E1s), ubiquitin-conjugating enzymes (E2s), and ubiquitin protein ligases (E3s) (Hershko et al., 2000; Pickart, 2001). E1 activates the C-terminus of ubiquitin as an ubiquityl-*S*-E1 thioester through an ubiquityl-AMP intermediate. The ubiquityl moiety is then transferred from the cysteine side chain of E1 to a cysteine side chain of a specific E2, forming an ubiquityl-*S*-E2 thioester. Finally, often cooperatively with an E3, the ubiquityl moiety is conjugated to the lysine side chain of a target protein. Most of the substrate specificity of the ubiquitylation machinery is encoded within E3-substrate interactions. Ubiquitin can be removed from lysine side chains by DUBs and UCHs, allowing for reversible signaling, as well as the recycling of ubiquitin monomers (Walsh, 2005).

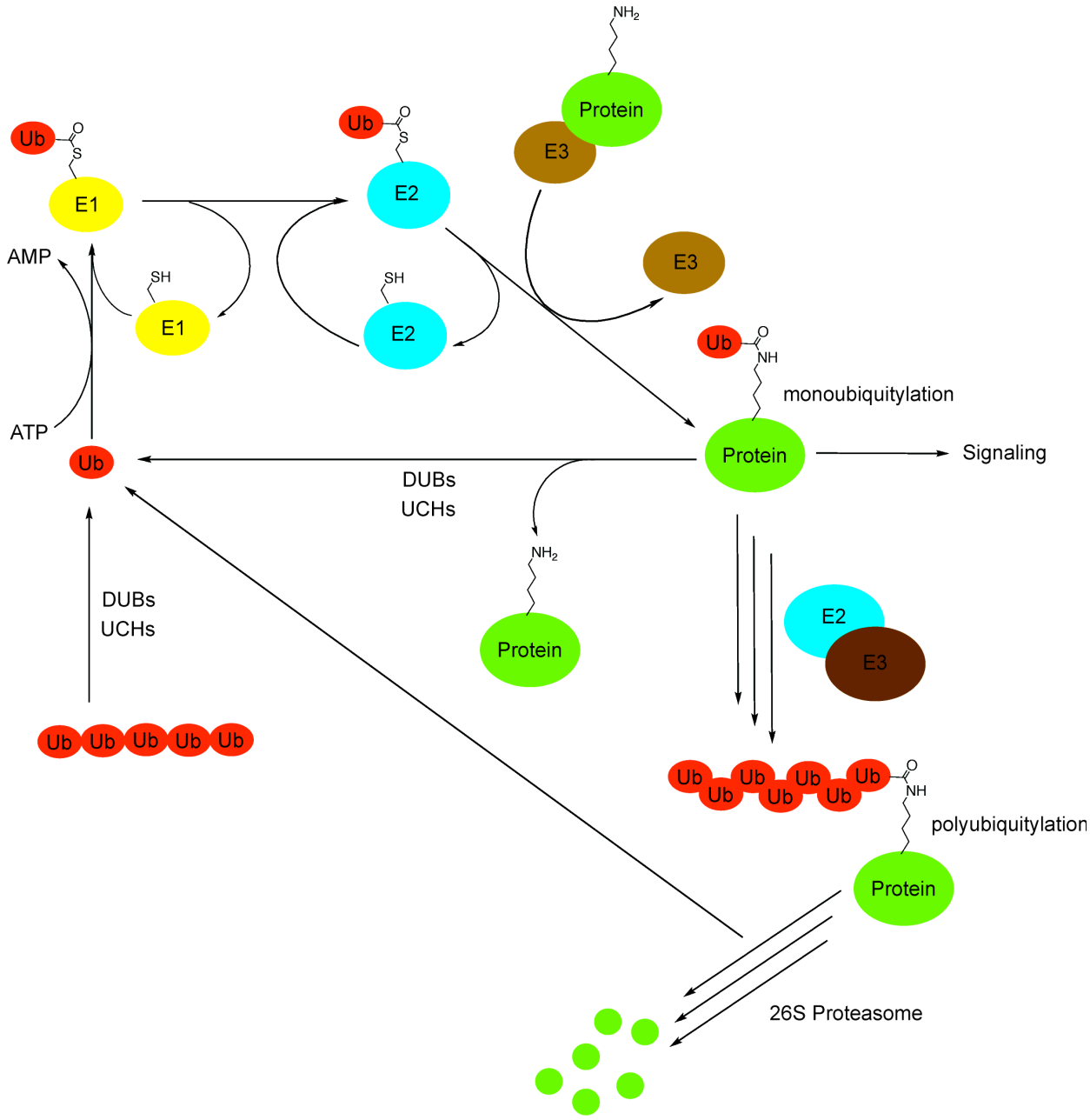



**Figure 1.3. The ubiquitin cycle.** Ubiquitin is attached to proteins through the concerted action of E1, E2, and E3 ubiquitin conjugating enzymes. Ubiquitin is recycled through the actions of DUBs and UCHs. Monoubiquitylation functions in cell signaling, while polyubiquitylation targets proteins for proteasomal degradation.

Early investigation of ubiquitin elucidated the role of polyubiquitylation in proteolytic degradation of proteins (Hershko et al., 2000). Polyubiquitin chains are linked through isopeptide bonds between ubiquitin lysines and successive ubiquityl-α-carboxylic acids; these chains can be built through K6, K11, K27, K29, K33, K48, and K63 – each linkage correlated with divergent functions (Xu et al., 2009). For proteasomal targeting, a protein must carry a minimum of four K48-linked ubiquitins (Walsh, 2005). While polyubiquitylation typically leads to proteasomal destruction of target proteins, recent investigation shows that monoubiquitylation signals in non-proteasomal pathways, including receptor internalization and vesicular trafficking (Katzmann et al., 2002; Raiborg et al., 2003), gene regulation (Shilatifard, 2006), and DNA damage repair (Huang and D'Andrea, 2006).

Most ubiquitin-based signaling pathways employ specific ubiquitin binding motifs. *In vivo*, ubiquitin is recognized by more than sixteen families of ubiquitin binding motifs (Hurley et al., 2006). These motifs are primarily α-helical, from the single α-helix of the ubiquitin interacting motif (UIM) to the small helix bundle of the ubiquitin associated (UBA) domain. Although interaction surfaces of ubiquitin binding motifs vary widely, most center on a hydrophobic patch of ubiquitin's surface including I44 (Hurley et al., 2006). Other ubiquitin surfaces involved in ubiquitin signaling include the C-terminal Gly-Gly region and a charged patch including D58 (Hurley et al., 2006).

Monoubiquitylated forms of H2A (uH2A) and H2B (uH2B) most often perform opposite functions in gene regulation. H2A is ubiquitylated by the polycomb repressor complex 1 (PRC1) and is correlated with methylation of H3K27, silencing of homeotic genes, and X inactivation (Shilatifard, 2006). In contrast, uH2B is implicated in transcriptional elongation and co-localizes with active histone PTMs, including methylated H3K4 (Weake and Workman, 2008). uH2B is

also critical to expression of homeotic genes, suggesting that an interplay between ubiquitylation of H2A and H2B in gene expression programs responsible for developmental patterning. In addition to roles in regulation of gene expression, uH2A and uH2B signal in DNA damage repair pathways (Game and Chernikova, 2009; Huang and D'Andrea, 2006). The function of uH2B in these processes is detailed in the following sections.

### 1.4.2. Ubiquitylation of H2B

H2B is ubiquitylated on K120 in vertebrates (Thorne et al., 1987), corresponding to K123 in *S. cerevisiae* (Robzyk et al., 2000) and K119 in *S. pombe* (Tanny et al., 2007). In higher organisms, uH2B represents 1% of the total cellular H2B (West and Bonner, 1980); this number may reach 10% in *S. cerevisiae* (Robzyk et al., 2000). The E2 and E3 enzymes responsible for H2B ubiquitylation were first discovered in budding yeast as Rad6 and Bre1, respectively (Hwang et al., 2003; Robzyk et al., 2000; Wood et al., 2003a). Orthologs of yeast Rad6, RAD6A and RAD6B, and yeast Bre1, BRE1A and BRE1B (RNF20/40), can efficiently ubiquitylate H2BK120 in human cells (Kim et al., 2009; Kim et al., 2005; Zhu et al., 2005). BRE1A and BRE1B interact to form a complex, which in turn, interacts with RAD6A or RAD6B. RNAi-mediated knockdown of either BRE1A or BRE1B, or RAD6A drastically reduces uH2B in human cells (Kim et al., 2009; Kim et al., 2005; Zhu et al., 2005). RAD6B knockdown has only a minimal effect on uH2B accumulation, suggesting a dominant role for RAD6A *in vivo* (Kim et al., 2009). Rad6, in concert with E3s other than Bre1, also ubiquitylates other proteins in DNA damage and protein degradation pathways (Osley, 2006).

In addition to RAD6A/B, other E2s have been implicated in ubiquitylation of H2BK120, including UbcH6 (Zhu et al., 2005), though their physiological roles are controversial (Kim et

al., 2009). Additionally, Mdm2 and BRCA1, both E3 ubiquitin ligases, can ubiquitylate H2B *in vitro*, but the *in vivo* relevance and site-specificity remain unclear (Chen et al., 2002; Mallery et al., 2002; Minsky and Oren, 2004; Xia et al., 2003).

In *S. cerevisiae,* ubiquitin is removed from H2B by two DUBs, Ubp8 and Ubp10 (Daniel et al., 2004; Emre et al., 2005; Gardner et al., 2005; Henry et al., 2003). Ubp8 is a component of the SAGA complex containing HAT, Gcn5 (Daniel et al., 2004; Henry et al., 2003). While Ubp8 is not necessary for SAGA function, two SAGA components, Sus1 and Sgf1, are required for association of Ubp8 with SAGA and the resultant removal of ubiquitin from H2B (Henry et al., 2003; Kohler et al., 2006). Recently, the human ortholog of Ubp8, ubiquitin specific protease 22 (USP22), was reported to be a part of the human SAGA-like complex, STAGA (Zhang et al., 2008; Zhao et al., 2008). USP22 can remove ubiquitin from H2B and its incorporation into the STAGA complex, necessitates human orthologs of Sgf1 and Sus1, ATXN7L3 and ENY2, respectively (Zhao et al., 2008). In addition to Ubp8, Ubp10 carries out the deubiquitylation of H2B in *S. cerevisiae* (Emre et al., 2005; Gardner et al., 2005). Ubp10 associates with Sir4 and is implicated in telomeric silencing. Human DUBs, USP7 and USP3 may also deubiquitylate H2B, but their physiologic roles are currently unclear (Nicassio et al., 2007; van der Knaap et al., 2005).

Ubiquitylation and deubiquitylation of H2B has been correlated with many chromatin templated processes, including transcriptional activation and elongation, *trans*-histone methylation of H3, and DNA damage repair (Figure 1.4) (Game and Chernikova, 2009; Weake and Workman, 2008). The role of uH2B in these processes will be discussed in detail below.

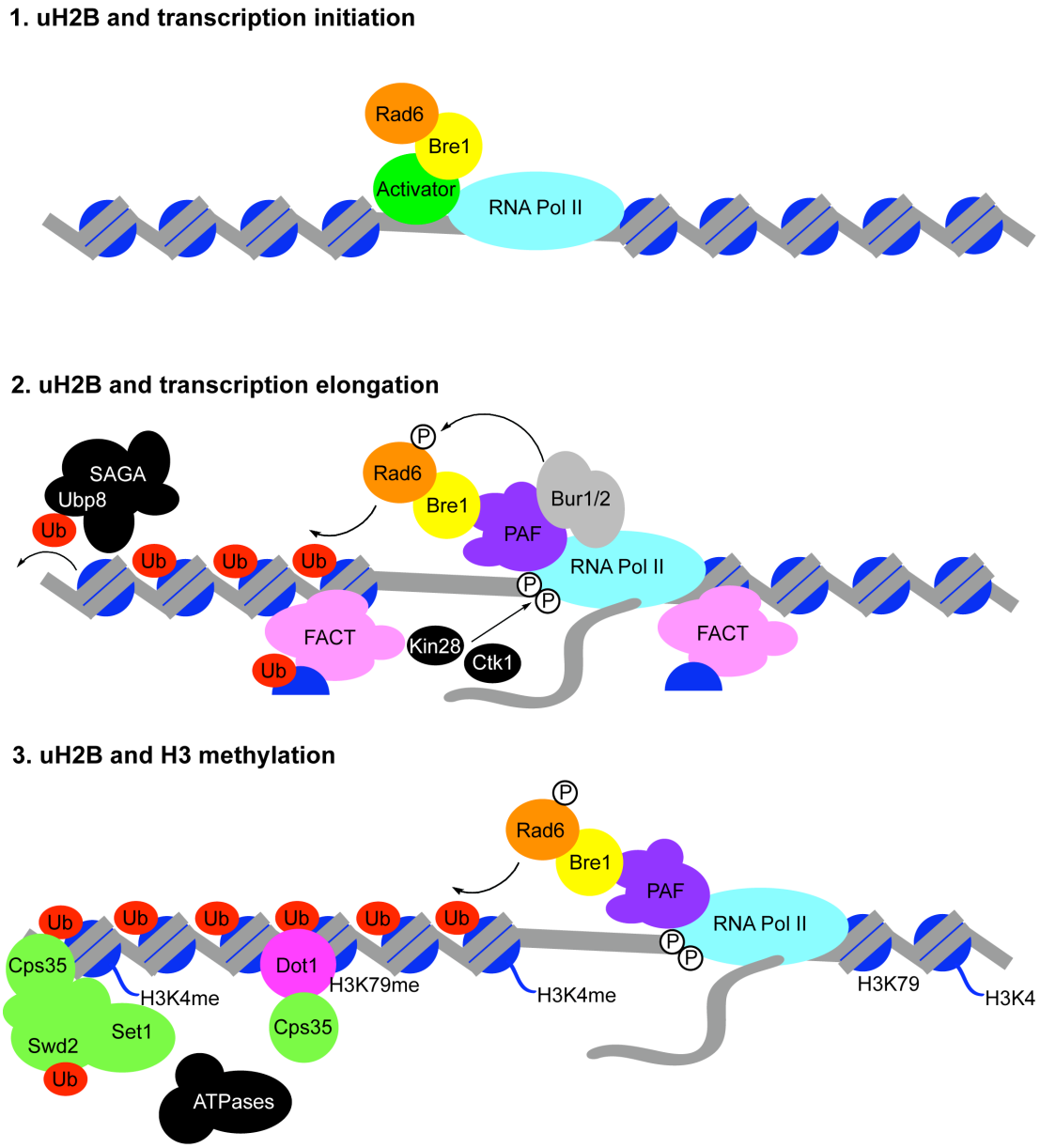



**Figure 1.4. The role of uH2B in transcription and H3 methylation.** 1. Rad6 and Bre1 are recruited during transcription initiation by Bre1-activator interactions. 2. A cycle of H2B ubiquitylation is correlated with transcriptional elongation. The Pol II CTD is phosphorylated by Kin28 and Ctk. The PAF and Bur1/Bur2 complexes mediate the movement of ubiquitin ligase activities of Rad6 and Bre1 with this elongating form of Pol II through the open reading frame of a gene. The resultant uH2B formation is shown behind the elongating polymerase for clarity, but this spatial localization is unknown. Remodeling complex FACT may participate in the disassembly and reassembly of nucleosomes as Pol II processes through the chromatin template. DUB activity of Ubp8 in the SAGA complex subsequently removes ubiquitin from H2B. 3. uH2B is required for efficient methylation of H3K4 and K79 by Set1 and Dot1, respectively. In both cases Cps35 and proteasomal ATPases may bridge uH2B and methylation. Cps35 ubiquitylation is also implicated in H3K4 methylation.

#### 1.4.2.1 uH2B and transcription

During transcriptional activation, Pol II is recruited to promoters of genes and transcription is initiated by cooperative functions of activators and coactivators. Subsequent promoter clearance allows Pol II to exit the promoter region and move into the open reading frame of the gene. Transcription elongation assisted by elongation factors continues until the nascent mRNA is complete and transcription is terminated. Considerable evidence suggests a role for the ubiquitylation cycle of H2B in transcriptional elongation (Laribee et al., 2007; Weake and Workman, 2008). Mutations affecting ubiquitylation of H2B in *S. cerevisiae* and *S. pombe* decrease accumulation of transcripts (Tanny et al., 2007; Xiao et al., 2005). In *S. cerevisiae*, Rad6 associates with the elongating form of Pol II, which is phosphorylated at Ser2 and Ser5 in its C-terminal domain (CTD) (Figure 1.4, middle panel) (Xiao et al., 2005). Interestingly, deletion of Kin28, responsible for phosphorylation of Ser5 in the Pol II CTD, abolishes ubiquitylation of H2B, while deletion of Ctk1, responsible for phosphorylation of Ser2 in the Pol II CTD, has no effect on H2B ubiquitylation levels (Xiao et al., 2005).

The interaction of Rad6 and elongating Pol II is mediated through Bre1 and the Pol II associated factor (PAF) complex. Bre1 in yeast and humans is recruited to promoters through interactions with activators, including p53 and Gal4 (Figure 1.4, top panel) (Kao et al., 2004; Kim et al., 2005; Wood et al., 2003a). This recruitment is independent of Rad6 and, in turn, is required for Rad6 localization to promoter regions (Kao et al., 2004; Wood et al., 2003a; Wood et al., 2003b). However, Bre1 and Rad6 recruitment to promoters is insufficient for H2B ubiquitylation, which requires additional factors, most notably the PAF complex and cyclin-dependent kinase complex, Bur1/Bur2 (Figure 1.4, middle panel) (Kim and Roeder, 2009; Laribee et al., 2005; Ng et al., 2003b; Wood et al., 2003b; Wood et al., 2005; Xiao et al., 2005).

Mutations in two components of the PAF complex, Paf1 and Rtf1, disrupt H2B ubiquitylation without altering Rad6 recruitment to promoters (Ng et al., 2003b; Wood et al., 2003b; Xiao et al., 2005). While properly recruited to the promoter, Rad6 fails to follow Pol II into the open reading frame in *rtf1Δ* cells (Xiao et al., 2005). This suggests that the ubiquitylation machinery is recruited to the promoter through Bre1-activator interactions, and then it is handed off to PAF to follow Pol II into the open reading frame. The Bur1/Bur2 complex is not only required for recruitment of the PAF complex, but also phosphorylates Ser120 of Rad6 (Figure 1.4, middle panel) (Laribee et al., 2005; Wood et al., 2005). Similar phosphorylation of human RAD6A by human cyclin-dependent kinase 2 (CDK2) enhances H2B ubiquitylation *in vitro* (Sarcevic et al., 2002).

Coincident with Pol II, the PAF complex, Rad6, and uH2B are found throughout the coding region of active genes (Kao et al., 2004; Minsky et al., 2008; Xiao et al., 2005). Interestingly, the presence of Rad6 in the coding regions of activated genes is transient compared to Pol II (Kao et al., 2004; Xiao et al., 2005). Thus, Rad6-mediated ubiquitylation of H2B may be involved in pioneering transcription of a newly activated gene that brings about stable changes in the chromatin template allowing subsequent rounds of transcription without Rad6. Similarly to H2B ubiquitylation, H2B deubiquitylation is required for optimal transcriptional efficiency. Reminiscent of Rad6, Ubp8 associates with elongating Pol II in the open reading frame of SAGA-regulated genes (Figure 1.4, middle panel), which have decreased expression in *ubp8Δ* strains (Daniel et al., 2004; Henry et al., 2003; Mutiu et al., 2007). Additionally, Ubp8 is a prerequisite for Ctk1-mediated phosphorylation of Ser2 of the Pol II CTD (Wyce et al., 2007). A double knockout of Ubp8 and Bre1 reestablishes Ser2 phosphorylation, suggesting that ubiquitylation of H2B blocks Ctk1 recruitment. Supporting this notion, Ctk1 associates with

H2A and H2B but shows no affinity for uH2B (Wyce et al., 2007). This evokes a model in which a cycle of ubiquitylation and deubiquitylation coordinated with Pol II CTD phosphorylation is required during transcriptional elongation of certain genes (Weake and Workman, 2008). However, minimal evidence for this model has been reported.

While much is known about how ubiquitylation of H2B is coordinated with transcription elongation, far less is understood regarding any mechanistic role of uH2B in the process. Ubiquitylation of H2B is required for efficient methylation of H3K4 and H3K79 in an evolutionarily conserved pathway (Figure 1.4, bottom panel) (Briggs et al., 2002; Kim et al., 2005; Ng et al., 2002b; van Leeuwen et al., 2002; Zhu et al., 2005). It is plausible that these methylation events could translate ubiquitylation of H2B into transcriptional activity through the recruitment of chromatin remodelers or other elongation factors. However, evidence suggests that the impact of H2B ubiquitylation on transcriptional efficiency is independent of H3K4 and K79 methylation (Fleming et al., 2008; Shukla and Bhaumik, 2007; Tanny et al., 2007). A second model has been proposed to explain methylation-independent functions of uH2B, implicating chromatin remodeling and assembly. uH2B may stimulate the activity of a chromatin remodeling complex, "facilitates chromatin transcription" or FACT, thus easing the nucleosomal barrier to transcription (Figure 1.4, middle panel) (Laribee et al., 2007). FACT is proposed to remove an H2A/H2B dimer from a nucleosome during transcription and has also been implicated in nucleosome reassembly behind elongating Pol II (Fleming et al., 2008; Kaplan et al., 2003; Mason and Struhl, 2003; Reinberg and Sims, 2006). While FACT permits low levels of transcription of a chromatinized template *in vitro*, transcriptional efficiency is greatly enhanced by the addition of PAF, UbcH6, and BRE1A and BRE1B (Pavri et al., 2006). Unfortunately, although the chromatin template was ubiquitylated in this work, H2B was never verified as the

substrate for ubiquitylation. As a result of the contentious role of UbcH6 in H2B ubiquitylation, this conclusion remains controversial. While much has been discovered regarding the requirements for ubiquitylation of H2B, future work will be required to elucidate the mechanism linking the H2B ubiquitylation cycle to efficient transcription elongation.

#### 1.4.2.2. uH2B and methylation of histone H3

Soon after its discovery in *S. cerevisiae*, uH2B was demonstrated to be required for efficient methylation of H3K4 and H3K79, establishing a *trans*-histone regulatory pathway (Figure 1.4, bottom panel) (Briggs et al., 2002; Dover et al., 2002; Ng et al., 2002b; Sun and Allis, 2002). Further investigation revealed that uH2B is required for di- and trimethylation of these positions, while elimination of H2B ubiquitylation has no effect on monomethylation (Shahbazian et al., 2005). Similar observations have been reported in human cell lines. Knockdown of RAD6A, BRE1A, or BRE1B results in decreased di- and trimethylation H3K4, as well as decreased dimethylation H3K79, to a degree that correlates well with reduction of uH2B (Kim et al., 2009; Kim et al., 2005). Monomethylation of both positions is unaffected. Overexpression of BRE1A results in increased trimethylation and dimethylation of H3K4 and K79, respectively; however, contradicting results have been observed regarding the levels of H3K4 mono- and dimethylation (Kim et al., 2005; Zhu et al., 2005).

It would be expected that disruption of H2B deubiquitylation machinery would have an effect similar to overexpression of ubiquitylation machinery. Indeed, promoter-specific analysis reveals an increase in H3K4 trimethylation in *ubp8Δ* strains compared to wild-type background following gene induction, with negligible changes to mono- and dimethylation (Henry et al., 2003). Surprisingly, in a separate study deletion of *ubp8* resulted in minimal decreases in global

H3K4 di- and trimethylation, simultaneous with a drastic increase in monomethylation (Daniel et al., 2004). This apparent discrepancy is unresolved.

The mechanistic role of uH2B in stimulating methylation is unclear. It has been proposed that uH2B may induce H3 methylation directly, either by altering chromatin structure and therefore nucleosomal accessibility, or through the recruitment of enzymatic function (Henry and Berger, 2002; Sun and Allis, 2002). However, the possibility that one or more additional factors may be required to translate the effect of uH2B into heightened methyltransferase activity is equally likely. For example, proteasomal ATPases and Cps35, a component of the Set1 H3K4 methyltransferase complex, have been implicated in bridging uH2B and H3 methylation (Figure 1.4, bottom panel) (Ezhkova and Tansey, 2004; Lee et al., 2007). Additionally, it has been suggested that Cps35 is itself ubiquitylated, leading to efficient methylation of H3K4 (Vitaliano-Prunier et al., 2008). These proposed mechanisms are not mutually exclusive and further investigation will elucidate the possible role of each in this *trans*-histone biochemical network.

### 1.4.3. Methylation of histone H3K4

In organisms ranging from yeast to humans, trimethylated H3K4 occupies the 5' region of actively transcribed genes, where it correlates with histone acetylation and Pol II occupancy and is critical to regulation of homeotic genes (Ruthenburg et al., 2007a). Functions of mono- and dimethylated H3K4 are less clear and may vary across organisms and even in different cell types within organisms (Ruthenburg et al., 2007a). Set1 is the sole H3K4 methyltransferase in *S. cerevisiae*. Increased complexity is found in vertebrates, where up to nine H3K4 methyltransferases have been reported, most included in the conserved MLL family (Ruthenburg et al., 2007a). While each MLL family methyltransferase is specific for H3K4, they seem to

occupy non-redundant niches (Glaser et al., 2006; Lee et al., 2006; Yu et al., 1995). Each MLL methyltransferase exists in a large multi-subunit complex, and while the common core subunits are conserved, other components are unique and may hold clues to the regulated actions of individual methyltransferases (Dou et al., 2006).

### 1.4.4. Methylation of histone H3K79

Methylation of H3K79 is implicated in maintenance of telomeric silencing, DNA damage repair pathways, and in the pathogenesis of human disease (van Leeuwen et al., 2002). H3K79 exists in mono-, di-, and trimethylated states in yeast and mono- and dimethylated states in humans (Garcia et al., 2007) and is closely correlated with transcriptionally active chromatin (Ng et al., 2003a). While the trimethylated state globally dominates *S. cerevisiae* chromatin, the distribution in vertebrates centers around unmodified and monomethylated states (Garcia et al., 2007). The various levels of methylation function redundantly in telomeric silencing in yeast (Frederiks et al., 2008), but distinct roles for methylation states in human cells have not been investigated.

To date, a single H3K79 methyltransferase, disruptor of telomeric silencing-like (Dot1L), has been reported in humans (Feng et al., 2002), analogous to the lone H3K79 methyltransferase in yeast, Dot1 (Lacoste et al., 2002; Ng et al., 2002a; van Leeuwen et al., 2002). Unlike most histone lysine methyltransferases, Dot1 lacks a characteristic Set domain and its catalytic domain bears resemblance to protein arginine methyltransferases (Dlakic, 2001; Min et al., 2003; Sawada et al., 2004). While the catalytic domain of Dot1 is evolutionarily conserved, the domain organization of yeast and human homologs is markedly different, suggesting divergent regulation or mechanism of action (Figure 1.5) (Sawada et al., 2004). Except for a lysine rich stretch

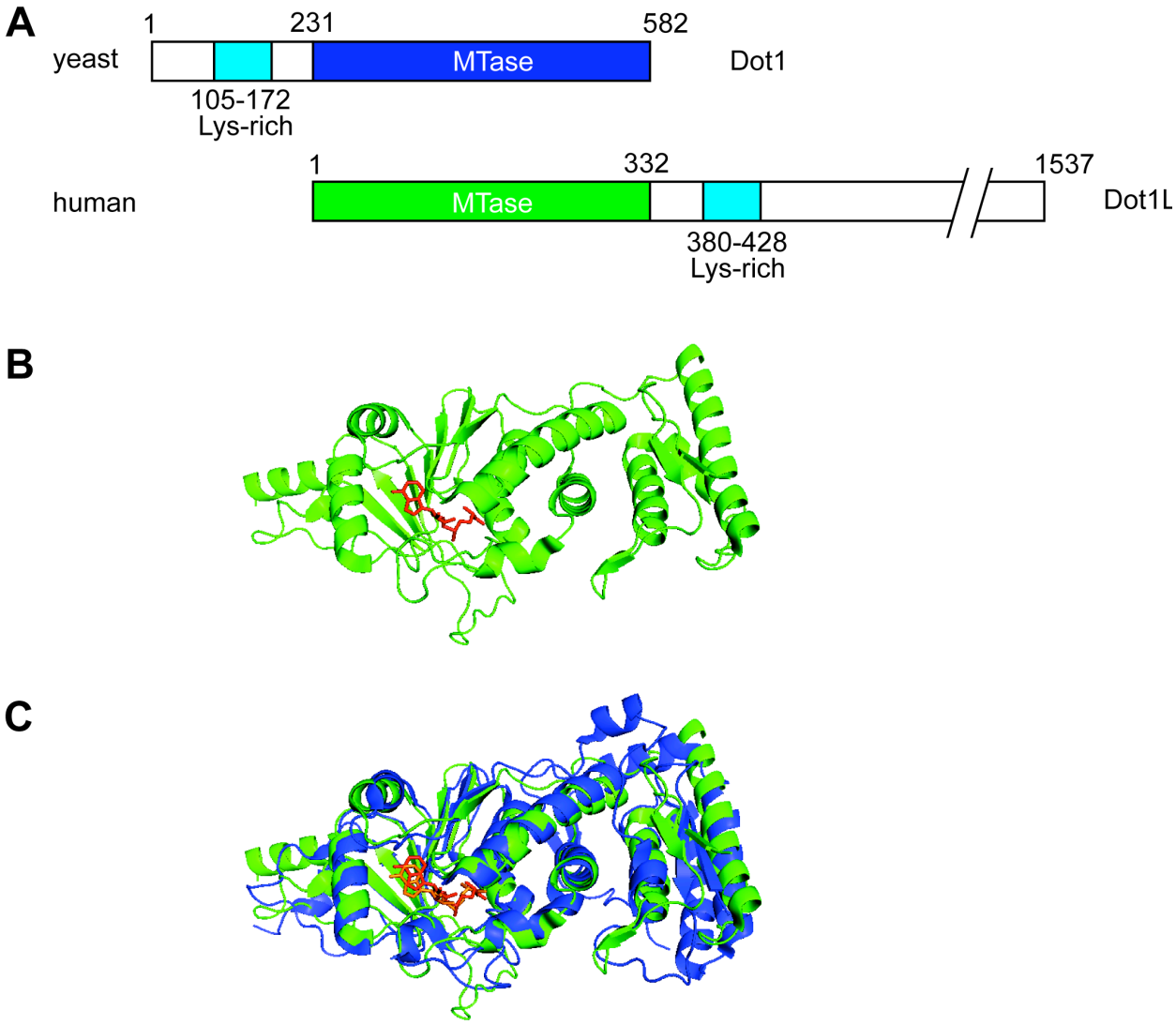



**Figure 1.5. Comparison of yeast Dot1 and human Dot1L methyltransferases.** A) Domain alignment of Dot1 and Dot1L. Lysine-rich domains implicated in DNA binding are labeled. MTase = methyltransferase domain. B) Cartoon representation of human Dot1L catalytic domain (1NW3) (Min et al., 2003). *S*-adenosylmethionine (SAM) is shown in red. C) Structural alignment of human Dot1L (green) and yeast Dot1 (blue, 1U2Z) (Sawada et al., 2004) catalytic domains. SAM bound to the human and yeast methyltransferases is shown in red and orange, respectively.

involved in DNA binding (Min et al., 2003; Sawada et al., 2004), the functions of the evolutionarily distinct domains of Dot1 have not been characterized.

#### 1.4.4.1. Methylation of histone H3K79 and maintenance of silencing

Dot1 was discovered in *S. cerevisiae* based on its role in telomeric silencing. Both overexpression and deletion of *dot1* results in a weakening of telomeric silencing (Singer et al., 1998). This seems counterintuitive for a methyltransferase marking active chromatin. Indeed, methylation of H3K79 functions in telomeric silencing by demarcating active chromatin (van Leeuwen et al., 2002). Nucleosomes containing methylated H3K79 disallow the binding of Sir proteins required for telomeric silencing. Thus, overexpression of Dot1 results in spreading of methylation of H3K79 into telomeric regions, preventing silencing; conversely, deletion of *dot1* eliminates H3K79 methylation allowing Sir proteins to bind to active chromatin, effectively diluting these proteins at telomeres below the threshold concentration required for silencing (Ng et al., 2003a; Shilatifard, 2006; Yang et al., 2008). This is in part facilitated through competition between Sir3 binding and Dot1 activity. Both Dot1 and Sir3 recognize a region of the N-terminal tail of H4 including the basic patch spanning residues 16-20. H4K16Q, a mutation mimicking acetylation of H4K16, disrupts Sir3 binding, favoring Dot1 activity in *in vitro* competition experiments. Additionally, the surface of the nucleosome surrounding H3K79 is critical for silencing (Park et al., 2002). Sir3 recognizes H3K79 peptides in a methylation dependent manner, suggesting that Sir3 binding *in vivo* may be disrupted by Dot1-mediated methylation (Altaf et al., 2007). The role of Dot1 in silencing is evolutionarily conserved. Dot1L-deficient mouse embryonic stem cells, exhibit reduced heterochromatic histone PTMs at telomeres and

centromeres along with elongated telomeres and chromosomal segregation defects (Jones et al., 2008).

#### 1.4.4.2. Methylation of histone H3K79 and the DNA damage response

To ensure the fidelity of the genome, various pathways have evolved to repair DNA damage. Two pathways exist for the repair of DNA double strand breaks (DSBs): non-homologous end-joining (NHEJ) and homologous recombination repair (HRR) (Aylon and Kupiec, 2004). HRR employs homologous DNA sequences to reconstruct DSB junctions and is the primary DSB repair mechanism in *S. cerevisiae*. Quality control measures, or cell-cycle checkpoints, exist throughout the replicative cycle of the cell and are activated when DNA damage is recognized, to prevent further cell-cycle progression until the integrity of the genome is restored (Wahl and Carr, 2001).

The *trans*-histone pathway linking Rad6/Bre1-mediated ubiquitylation of H2B and Dot1-mediated methylation of H3K79 is intimately linked with cell-cycle checkpoint control and DNA damage repair in *S. cerevisiae* (Game and Chernikova, 2009). Decreased survival following ionizing radiation (IR) -induced DSBs is observed in *dot1Δ* strains (Game et al., 2005). This mutation is epistatic with *bre1Δ* and H3K79 mutations, implicating the uH2B-H3K79me axis in DSB repair (Game et al., 2005; Game et al., 2006). This sensitivity is due to defective HRR, as *dot1Δ* on the background of *rad50Δ* or *rad51Δ*, both deficient in HRR, exhibit no increased sensitivity to IR over *rad50Δ* and *rad51Δ* strains alone (Game et al., 2006). The reported dysfunction in the HRR pathway likely results from the abrogation of checkpoint control. *dot1Δ* strains fail to activate G1 and S-phase cell-cycle checkpoints following ultraviolet- or IR-induced DNA damage as measured by activation of Rad53, a checkpoint regulator of all cell-

cycle stages in *S. cerevisiae* (Giannattasio et al., 2005; Wysocki et al., 2005). Similar defects are observed in *bre1Δ* strains, and in strains with mutations of H3K79 or H2BK123 (Giannattasio et al., 2005; Wysocki et al., 2005).

The discovery of H3K79me-specific binding modules in checkpoint control pathways presented a mechanistic role for Dot1 in HRR (Huyen et al., 2004). Mutations within the tandem tudor domain of 53BP1, or its *S. cerevisiae* ortholog, Rad9, disrupt interaction with H3K79me *in vitro* and prevent the formation of 53BP1/Rad9 foci at DSBs *in vivo* (Grenon et al., 2007; Huyen et al., 2004). Knockdown of Dot1L also abolishes the establishment of 53BP1 foci and the resultant cell-cycle delay and initiation of HRR (Huyen et al., 2004). No difference is observed in either G2 IR-sensitivity or G1 checkpoint failure between Rad9 Tudor domain mutants and *dot1Δ* strains, verifying its significance in yeast (Grenon et al., 2007). Recent conflicting evidence implicates methylation of H4K20, instead of H3K79, in 53BP1 binding, bringing this model into question in mammals (Botuyan et al., 2006).

#### 1.4.4.3. Methylation of histone H3K79 and human disease

Methylation of H3K79 is pathogenically linked with a subset of human leukemias, MLL-rearranged leukemias. MLL-rearranged leukemias result from translocations, fusing regions of the H3K4 methyltransferase, MLL, which typically lack the catalytic Set domain, to greater than 50 fusion partners and can present as acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), or mixed lineage leukemia (MLL). These fusions are present in roughly 10% of all human leukemias and in excess of 70% of infant leukemias, and carry poor prognoses compared to non MLL-rearranged leukemias, underscoring their importance to human disease (Krivtsov and Armstrong, 2007). Of the myriad of fusion partners, greater than 80% of

pathogenic MLL fusions occur with a small subset, including elongation factors, AF4, AF10, AF9, AF6, ENL, and ELL (Krivtsov and Armstrong, 2007).

Model leukemogenic systems induced by MLL fusions have unequivocally demonstrated a role for aberrant Dot1L-mediated H3K79 methylation in the majority of cases. AF4 and AF10 interact with Dot1L, allowing the recruitment of Dot1L activity to genomic loci targeted by the associated MLL fusions (Bitoun et al., 2007; Okada et al., 2005). This recruitment alone is leukemogenic, as MLL-Dot1L fusions can transform cell lines (Okada et al., 2005). Genome-wide expression analysis and chromatin immunoprecipitation (ChIP) reveal aberrant expression of developmentally regulated genes, including genes involved in hematopoesis and self-renewal, and associated regions of H3K79 methylation in MLL-AF4 and AF10 models (Guenther et al., 2008; Krivtsov et al., 2008). These patterns correlate with observations in gene expression and H3K79 methylation in MLL-rearranged leukemic patients. In fact, H3K79 methylation patterns alone can distinguish MLL-rearranged from non MLL-rearranged leukemia (Krivtsov et al., 2008). Furthermore, overexpression of a catalytically inactive mutant of Dot1L or RNAi knockdown of Dot1L abolish oncogenicity of MLL-AF10 fusion cell lines (Okada et al., 2005; Zhang et al., 2006); similarly, knockdown of Dot1L reduces H3K79 methylation and aberrant transcription of developmentally regulated genes in MLL-AF4 cell lines (Krivtsov et al., 2008). This invokes the possibility of exploiting Dot1L as a much needed therapeutic target in MLL-rearranged leukemia.

### 1.5. Semisynthetic strategies to generate post-translationally modified histones

The last fifteen years has brought somewhat of an enlightenment period in the investigation of histone PTMs. Many new types of histone PTMs have been discovered.

Advances in antibody- and mass spectrometry-based technologies have unearthed nearly countless modification sites. Clever biochemistry guided by homology searches, only possible in the post-genomic era, have led to the isolation of the proteins responsible for establishing, erasing, and interpreting many of these histone PTMs in a wide range of organisms. Furthermore, ChIP protocols, coupled with microarray technology (ChIP-chip) or massively parallel sequencing (ChIP-seq) have been used to map the genome-wide localization of these histone PTMs and chromatin associated proteins at the resolution of a single nucleosome. The breadth of knowledge gained has been extraordinary, the correlative potential between structure and function, compelling; yet, as the dust settles, it is clear that the mechanisms of action through which histone PTMs function in many chromatin-templated processes are vague, indirect, or completely obscure. This naiveté stems from the difficulty in isolating the actions of single components within complex multi-component systems, often equipped with redundant and compensatory mechanisms. In order to interrogate the mechanisms of PTMs in these processes, it is necessary to develop defined biochemical models *in vitro*. A few exceptions aside, a wide chasm exists between current ability for bottom-up reconstitution and top-down observation of these chromatin-templated processes. Advancing our understanding of epigenetics will require the narrowing of this gap.

Dissecting the functions of histone PTMs in chromatin-templated processes necessitates the preparation of homogeneously modified histones. This allows the roles of modifications to be isolated from distinct functions associated with the enzymes responsible for establishing the modifications. This can be accomplished enzymatically; however, the relative promiscuity of many histone modifying enzymes *in vitro* and challenges associated with separating modified from unmodified histones, detracts from this strategy. Recently, three semisynthetic protein

chemistry technologies have been employed to generate modified histones, ensuring chemical homogeneity and bypassing the use of complex enzymatic machinery. These technologies, expressed protein ligation, ribosomal incorporation of unnatural amino acids, and alkylated methyllysine analogs, and their applications to histones will be explored in the following sections.

### 1.5.1. Expressed protein ligation and applications to modified histones

Native chemical ligation (NCL) enables the regioselective ligation of two polypeptides when one contains an N-terminal cysteine and the other, a C-terminal thioester, forming a native amide bond at the junction (Dawson et al., 1994). As robust methodologies exist for the generation of C-terminal thioesters and N-terminal cysteines in both recombinant proteins and synthetic peptides, NCL can be extended to the ligation of recombinant proteins with synthetic peptides in an application known as expressed protein ligation (EPL) (Figure 1.6) (Muralidharan and Muir, 2006). The power of EPL lies in the ability to apply the vast toolkit of peptide chemistry to proteins otherwise too large to be accessible synthetically, using modern solid-phase methodologies. EPL has been applied to structure-function analysis in a diverse set of proteins through the site-specific introduction of post-translational modifications, unnatural amino acids, optical probes, and isotopic labels (Flavell and Muir, 2009; Muralidharan and Muir, 2006; Pellois and Muir, 2006). Due to the length limitations of solid phase peptide synthesis (SPPS), the N- and C-terminal regions of proteins are most accessible to EPL. Chemical access to internal regions of large proteins by EPL requires sequential ligation of a minimum of three fragments.

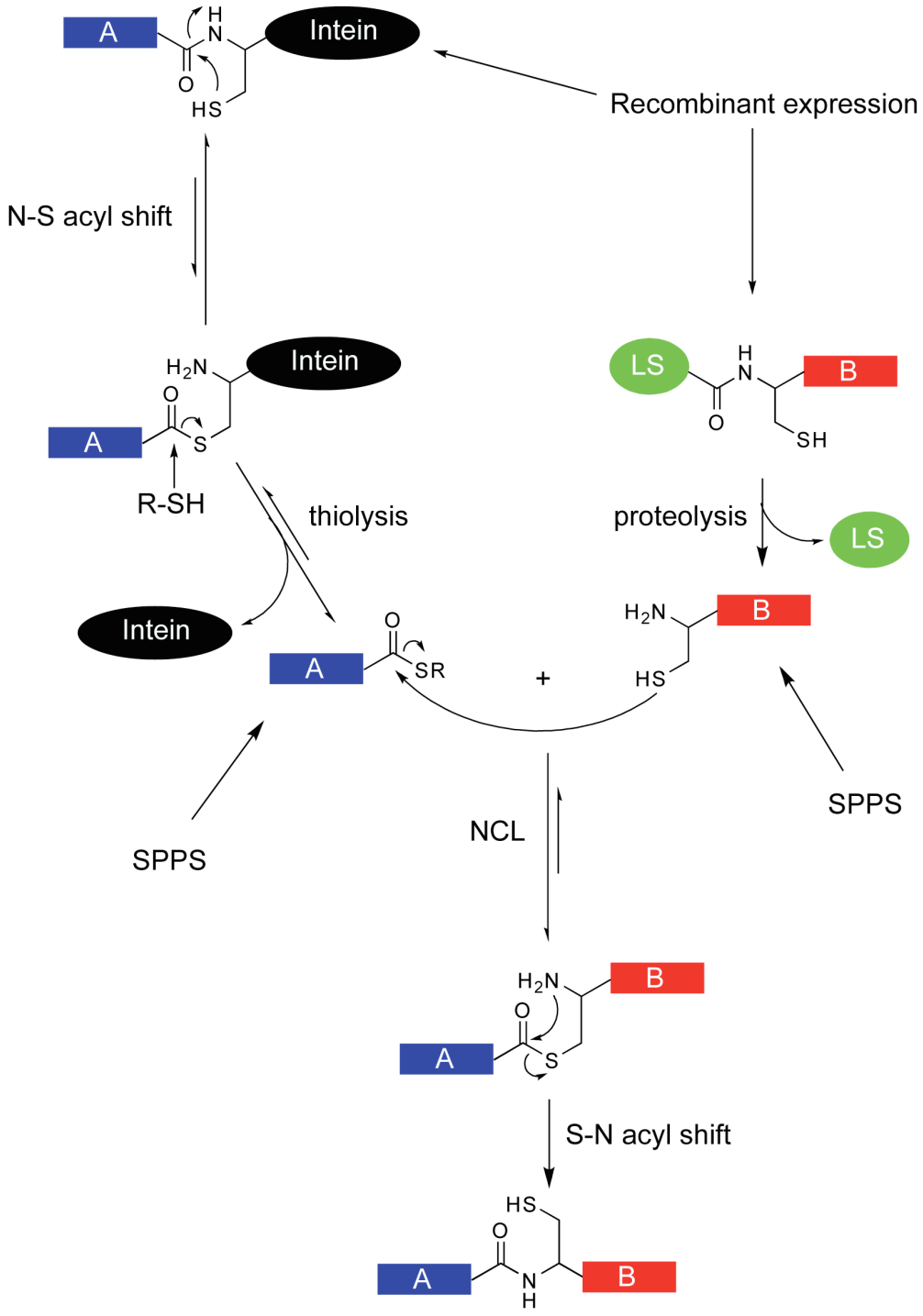



**Figure 1.6. Expressed protein ligation.** EPL allows the ligation of two polypeptides of synthetic and recombinant origins, when one contains a C-terminal thioester, and the other, an N-terminal cysteine. Protein C-terminal thioesters can be generated through thiolysis of recombinant intein fusion proteins. Protein N-terminal cysteines can be revealed following proteolysis of leader sequences (LS). Peptides bearing C-terminal thioesters or N-terminal cysteines can be prepare using SPPS.

Not surprisingly, the generation of post-translationally modified histones is an attractive application of EPL, owing to the density of PTMs located within histone N-terminal tails. H2B, H3, and H4 have been prepared bearing acetylations, lysine and arginine methylations, and phosphorylations within their N-terminal tails (Chiang et al., 2009; Ferreira et al., 2007; He et al., 2003; Li and Shogren-Knaak, 2008; Li and Shogren-Knaak, 2009; Shogren-Knaak et al., 2003; Shogren-Knaak and Peterson, 2004). Additionally, acetylations have been introduced near the C-terminus of H3 (Manohar et al., 2009). Combinations of these PTMs can be easily introduced simultaneously into a single histone N- or C-terminal tail using EPL. To date, semisynthetic histones generated by EPL have been used to gain insight into a direct role for H4K16 acetylation in chromatin conformation (Shogren-Knaak et al., 2006), to reveal cooperativity in SAGA-mediated acetylation of histones within nucleosomes and in nucleosomal arrays (Li and Shogren-Knaak, 2008; Li and Shogren-Knaak, 2009), and to dissect the effect of acetylation of histones on activity of chromatin remodeling complexes (Ferreira et al., 2007).

### 1.5.2. Ribosomal methodologies for the incorporation of PTMs into histones.

Nonsense suppression mutagenesis allows for ribosomal synthesis of polypeptides containing unnatural amino acids (Wang et al., 2006). Typically, an amber nonsense tRNA is hijacked to code for an unnatural amino acid. This amber tRNA can be charged with an unnatural amino acid *in vitro*, or an orthogonal tRNA/tRNA synthetase pair can be evolved to charge the tRNA *in vivo* (Dougherty, 2000; Wang et al., 2006). Recently, a bacterial pyrrolysyl tRNA synthetase/amber tRNA pair was evolved to incorporate acetyllysine into myoglobin (Neumann et al., 2008). Unlike EPL, nonsense suppression mutagenesis is not limited to terminal

protein sequences. Research is underway to adapt this technology to the generation of acetylated histones.

In vitro genetic reprogramming has also been exploited for the ribosomal synthesis of histone peptides bearing acetyl- and methyllysines (Kang et al., 2008). In this work, codons from unused amino acids were reassigned to acetyl-, mono-, di-, and trimethylated lysines. A ribozyme-based tRNA acylating catalyst, dinitro-flexizyme (Murakami et al., 2006), was used to charge the corresponding reassigned tRNAs. mRNAs containing reassigned codons were translated in vitro using a system lacking the unused amino acids, allowing the generation of modified histone peptides. The extension of this technology to full-length histone proteins, where this application will be more limited due to fewer unused codons, has not been reported. Presumably, cysteine and tryptophan, which are either absent or found in low abundance in canonical histone sequences, can be hijacked for the generation of full-length modified histones.

### 1.5.3. Methyllysine analogs and applications to modified histones

Recently, the unique nucleophilic properties of cysteine were exploited to incorporate methyllysine analogs (MLAs) into histones in a site-specific fashion (Simon et al., 2007). Under stringent pH control, the thiol of a cysteine side chain can be directly targeted in an alkylation reaction, leaving all other nucleophiles within a protein untouched. In this fashion, a cysteine can be alkylated with an electrophilic ethylamine, to generate aminoethylcysteine (Kenyon and Bruice, 1977). Aminoethylcysteine is structurally similar to lysine, replacing the γ-methylene of lysine with a sulfide. This substitution has minor consequences on the structural and chemical properties of lysine, lengthening the side chain by 0.28 Å and decreasing the pKa of the ε-$NH_2$

group by 1.1 (Gloss and Kirsch, 1995). By replacing the electrophile with the corresponding mono-, di-, and trimethylated amines, cysteines can be converted to MLAs (Simon et al., 2007).

Only one or two cysteines exists in the canonical core histone proteins. These cysteines in the histone fold of H3 can be mutated to an alanine without significant effects on nucleosomal reconstitution or chromatin function. Thus, orthogonal cysteines can be genetically incorporated into recombinant histones in a site-specific manner. These cysteines can be alkylated to generate histones bearing MLAs. Using this approach, MLAs have been introduced in the globular domains of histones with similar ease as N-terminal tails (Simon et al., 2007). Despite minor chemical changes compared to native methylated counterparts, histones bearing MLAs are recognized by site- and modification-specific antibodies and binding modules, and can function in enzymatic assays. However, in some cases, slightly decreased performance is observed compared to natively modified histones (Simon et al., 2007). To date, MLAs have been used to determine methylation-state preferences of lysine demethylases (Lee et al., 2009; Lin et al., 2008), to examine histone PTM crosstalk (Hung et al., 2009; Li et al., 2009), and to investigate structural perturbation of nucleosomes and nucleosome arrays carrying lysine methylations (Lu et al., 2008). This strategy has yet to be successfully applied to the installment of other types of histone PTMs.

A similar approach has been reported for the incorporation of acetyl- and methyllysine analogs into histones (Guo et al., 2008). In this case, nonsense suppression mutagenesis was used to encode phenylselenocysteine, which was subsequently eliminated to form dehydroalanine. Michael addition of *N*-acetylated or *N*-methylated cysteamine yields the corresponding acetyl- or methyllysine analog. However, Michael addition with dehydroalanine results in racemization of the $\alpha$-carbon, greatly limiting the use of this technology for biochemical applications.

## 1.6. Summary

Complex epigenetic phenomena are choreographed through the cooperative actions of a small subset of molecular mechanisms, including histone PTMs, histone variants, chromatin remodeling, DNA methylation, and siRNA-mediated TGS. Histone PTMs densely populate chromatin and influence its function through three mechanisms: 1) intra- and internucleosomal conformational changes; 2) site- and modification-specific recruitment of effectors; and 3) exclusion of effectors. One such modification, uH2B, is correlated with diverse processes including transcriptional elongation, *trans*-histone biochemical pathways, and the response to DNA DSBs.

Full elucidation of the role of uH2B and other histone PTMs requires their isolation from complex cellular processes. Robust semisynthetic methodologies have been reported for the incorporation of PTMs into histones to isolate their functions *in vitro*. Early mechanistic studies using semisynthetically modified histones have revealed the power of these technologies. A combination of semisynthetic protein chemistry to create engineered chromatin substrates, and their subsequent biochemical analysis, will play an important role in the hypothesis-driven dissection of the mechanisms underlying histone PTMs and their potential function in a histone code.

The goals of this thesis are two-fold: 1) to develop EPL-based technology for the site-specific ubiquitylation H2B; and 2) to interrogate the role of uH2B in Dot1L-mediated methylation of H3K79. In chapter 2, a generalizable strategy for the ubiquitylation of peptides using a photolytically removable ligation auxiliary is presented. This strategy is then extended to generate full-length uH2B. In chapter 3, a direct stimulation of intranucleosomal Dot1L-mediated methylation of H3K79 by uH2B is established using semisynthetic uH2B. In chapter 4,

a highly optimized semisynthesis of uH2B bearing a single G76A ubiquitin point mutation is presented. Semisynthetic u(G76A)H2B is indistinguishable from uH2B by Dot1L, allowing the detailed kinetic and structure activity relationship analyses presented in chapter 5.

## Chapter 2: Semisynthesis of native ubiquitylated H2B[1]

The ability to generate homogeneously ubiquitylated H2B is paramount to understanding its functions in transcriptional elongation, methylation of H3K4 and K79, and DNA damage repair. Two biological solutions exist to accomplish this: 1) purification from endogenous sources (Figure 2.1, top panel), and 2) *in vitro* modification using purified enzymes (Figure 2.1, middle panel). Because of its natural abundance uH2B can be purified directly from cells (Davies and Lindsey, 1994; West and Bonner, 1980). However, heterogeneity due to the coexistence of other PTMs complicates biochemical analysis. Additionally, RAD6A or UbcH6 can be combined with BRE1A/B to ubiquitylate H2B *in vitro*, though yields are limiting (Kim et al., 2009; Zhu et al., 2005). Therefore, we decided to employ EPL to regioselectively ubiquitylate H2B, thus bypassing the requirement for the complex cellular ubiquitylation machinery and ensuring chemical homogeneity (Figure 2.1, bottom panel).

While robust methodologies exist for the installment of small PTMs into proteins using EPL, including phosphorylations (Flavell et al., 2002; Muir et al., 1998; Wu et al., 2001), acetylations (He et al., 2003; Shogren-Knaak and Peterson, 2004), methylations (He et al., 2003; Shogren-Knaak and Peterson, 2004), lipidations (Alexandrov et al., 2002; Pylypenko et al., 2006; Rak et al., 2003), and glycosylations (Macmillan and Bertozzi, 2000; Macmillan and Bertozzi, 2004; Tolbert and Wong, 2000), all rely on the incorporation of pre-modified amino acid building blocks into a synthetic peptide or modest chemical derivatization of a protected peptide. The application of these strategies to ubiquitylation is not feasible. The total chemical synthesis of ubiquitin has been achieved using both stepwise solid phase peptide synthesis (SPPS)

[1] The work in this chapter was performed as part of a close collaboration with Dr. Champak Chatterjee in the Laboratory of Synthetic Protein Chemistry at the Rockefeller University.

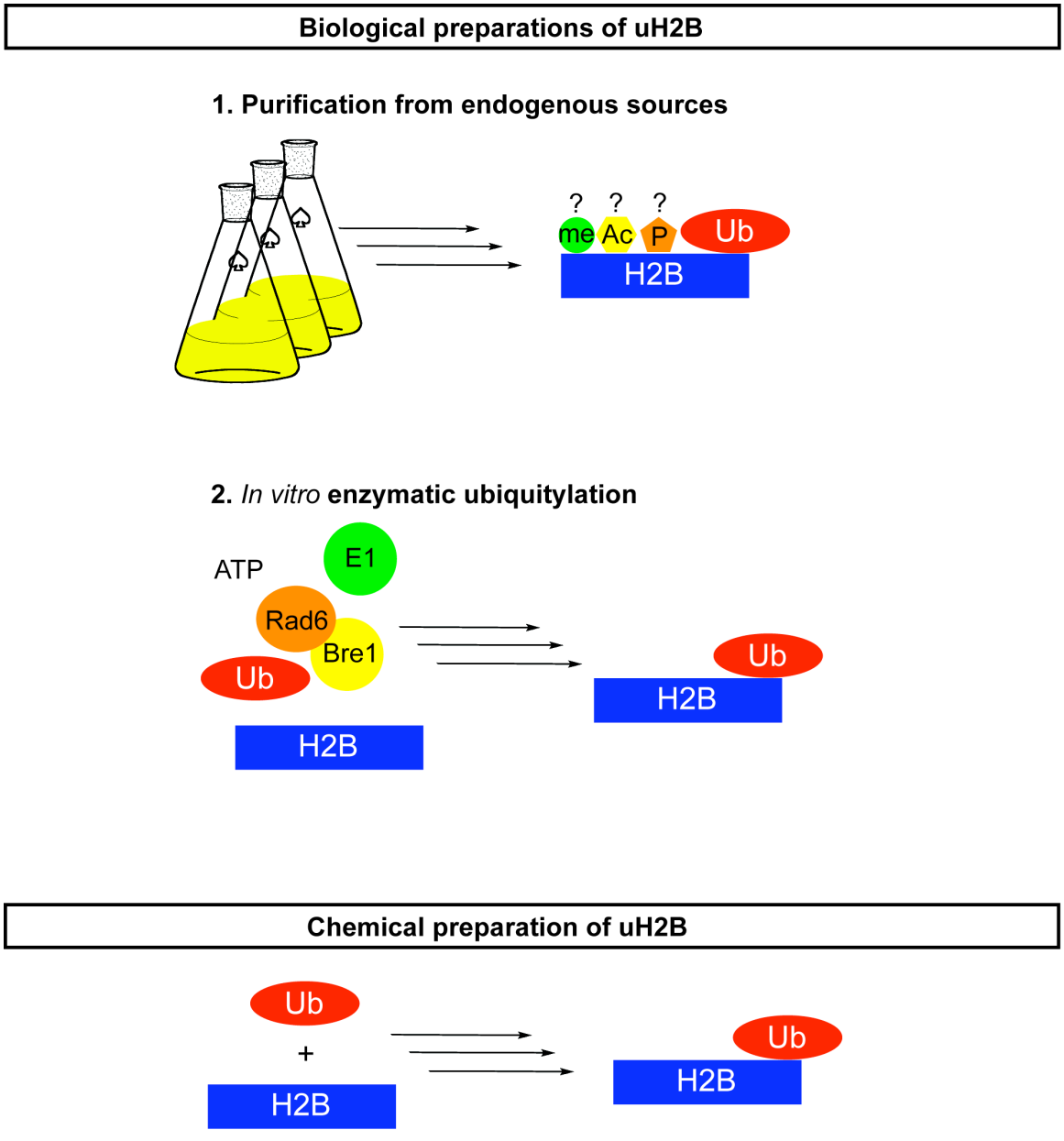



**Figure 2.1. Biological and chemical approaches to the preparation of uH2B.** Two biological approaches exist for the preparation of uH2B: 1) purification from endogenous sources (top panel), and 2) *in vitro* enzymatic ubiquitylation (middle panel). The former is complicated by co-purification of other PTMs, while the latter only leads to modest yields. We decided to develop a chemical approach for the generation of uH2B (bottom panel).

(Ramage et al., 1994) and NCL (Bang et al., 2005). This has allowed conjugation of ubiquitin to the ε-$NH_2$ of a single lysine residue through an isopeptide bond (Layfield et al., 1999). However, extension of these approaches to the site-specific ubiquitylation of peptides and proteins is nontrivial. In thinking about this problem, we were influenced by the elegant semisynthesis of ubiquitin analogs containing C-terminal electrophiles (Borodovsky et al., 2002). This is accomplished through the direct aminolysis of recombinant ubiquitin-α-thioesters generated by thiolysis of the corresponding ubiquitin-intein fusions. We wondered whether a similar approach could be exploited for the site-specific ubiquitylation of a lysine in a peptide or protein using EPL. Due to the absence of native cysteines in ubiquitin, such a scheme requires the use of a traceless ligation strategy to retain the native isopeptide linkage. In principle, use of a ligation auxiliary provides a solution to this problem. Several thiol-bearing auxiliaries have been reported that can facilitate an EPL reaction, acting as N-terminal cysteine surrogates, and following ligation can be removed under acidic (Low et al., 2001) or photolytic conditions (Kawakami and Aimoto, 2003). However, their general applicability suffers from the common requirement for a sterically undemanding ligation site (Hackenberger and Schwarzer, 2008; Marinzi et al., 2001; Marinzi et al., 2004; Offer et al., 2002). The two C-terminal residues of ubiquitin are Gly-Gly, which constitutes an ideal junction for an auxiliary-mediated ligation. Thus, an ubiquitin-α-thioester lacking the C-terminal glycine could be ligated to a peptide containing a ligation auxiliary linked to a lysine side-chain through a glycyl linker (Figure 2.2). Following protein ligation, the auxiliary could be removed leaving the native sequence surrounding the isopeptide linkage (Chatterjee et al., 2007). In this chapter, this strategy is used in the semisynthesis of an ubiquitylated H2B peptide and the full-length native uH2B.

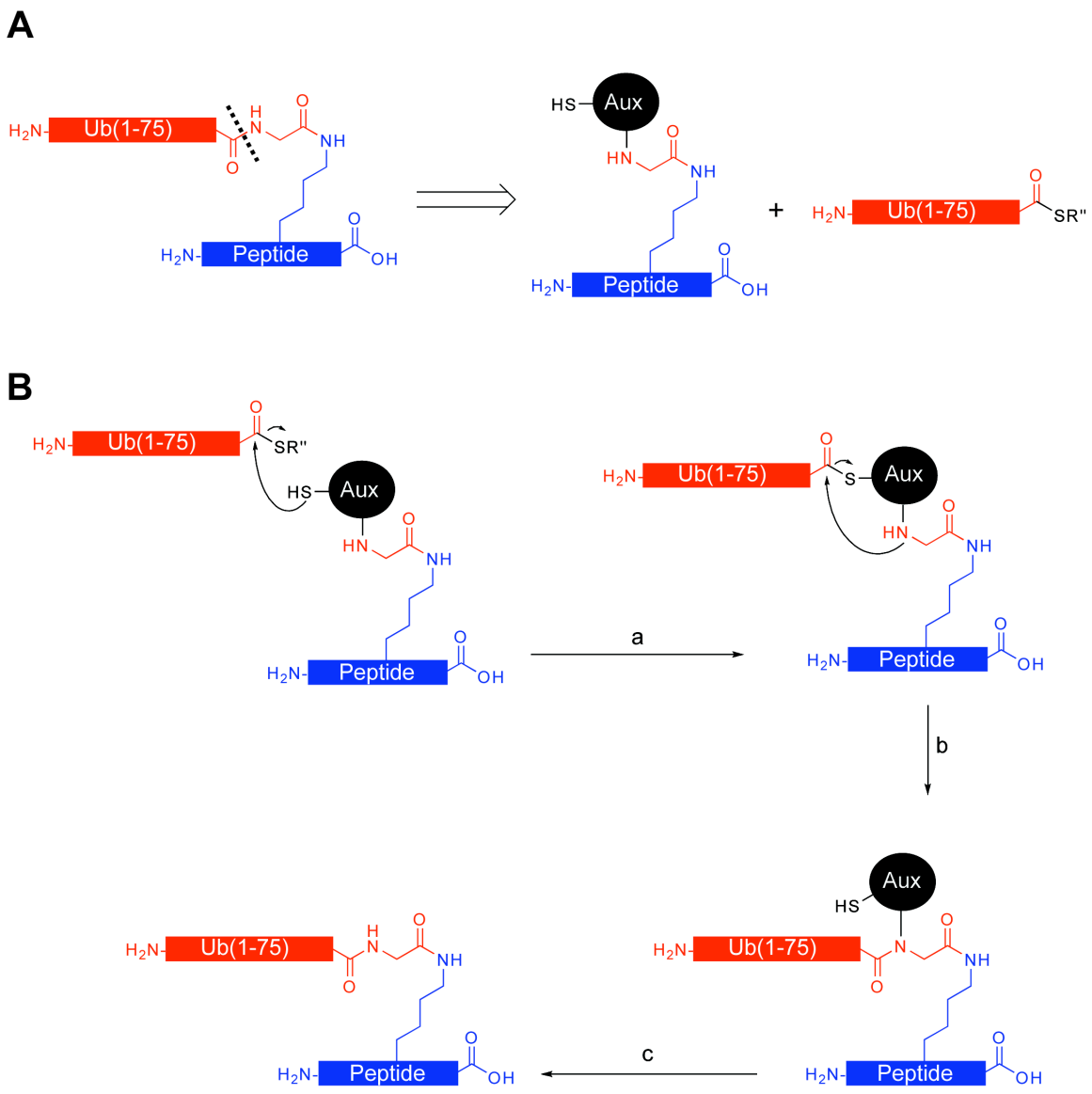



**Figure 2.2. Scheme of auxiliary-mediated peptide ubiquitylation.** A) Retrosynthetic scheme of ligation strategy. A ubiquitylated peptide is generated by ligation of ubiquitin(1-75)-α-thioester to a peptide conjugated to a ligation auxiliary through a glycyl linker. A dashed line represents the retrosynthetic disconnection. B) Synthetic scheme for peptide ubiquitylation. Auxiliary-mediated NCL reaction proceeds in two steps: (a) trans-thioesterification followed by (b) an *S*-to-*N* acyl shift. In step (c), the ligation auxiliary is removed, leaving the native ubiquitin sequence.

## 2.1. Synthesis of photolytically removable ligation auxiliary

To perform the traceless ligation reaction, we selected a derivative of a previously reported photolytically cleavable ligation auxiliary, **1** (Kawakami and Aimoto, 2003; Marinzi et al., 2004; Pellois and Muir, 2005). Compound **1** was synthesized in solution using eight steps from vanillin, **2** (Figure 2.3) (Chatterjee et al., 2007; Pellois and Muir, 2005). In the first step, the phenolic-OH of vanillin was alkylated with methyl 4-chlorobutanoate, in the presence of potassium carbonate and tetrabutyl ammonium iodide in tetrahydrofuran (THF), forming ester **3**. Following extraction, compound **3** was subjected to electrophilic aromatic substitution using nitric acid, affording nitrated product **4**. Compound **4** was purified by extraction and crystallization. Wittig reaction of **4** with methyltriphenylphosphonium bromide and sodium hexamethyldisilazide in THF, yielded styrene **5**. Compound **5** was isolated following flash column chromatography. Asymmetric aminohydroxylation of **5** with tert-butylcarbamate, $(DHQ)_2PHAL$, $K_2OsO_2(OH)_4$, and tert-butyl hypochlorite in sodium hydroxide, followed by flash column chromatography, yielded alcohol **6**. Regioselectivity of the asymmetric aminohydroxylation reaction was accomplished using a mixture of n-propanol and water, while enantioselectivity was established through the use of chiral compound, $(DHQ)_2PHAL$. $K_2OsO_2(OH)_4$ functioned as a reaction catalyst (Reddy and Sharpless, 1998). A Mitsunobu reaction, employing thioacetic acid, triphenylphosphine and diisopropyl azodicarboxylate (DIAD) in THF, was used to convert compound **6** into thioester **7** (Rozwadowska, 1997). Following flash column chromatography, **7** underwent simultaneous saponification and tert-butyl disulfide formation, in the presence of sodium methoxide and 2-methyl-2-propane thiol in aqueous sodium hydroxide with constant oxygen bubbling, yielding disulfide **8**.

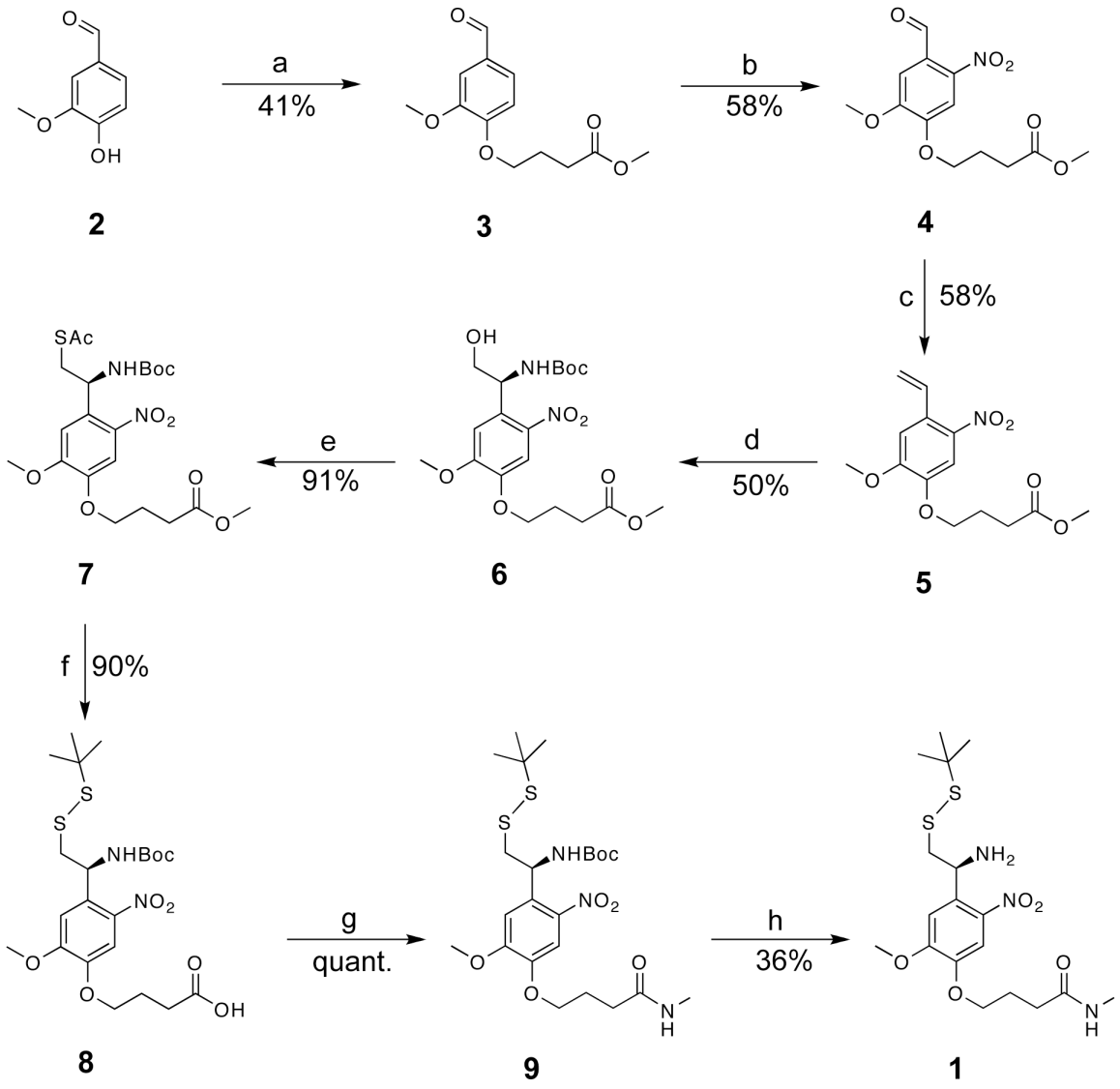



**Figure 2.3. Synthesis of photolytically removable ligation auxiliary, 1.** (a) $ClCH_2CH_2CH_2C(O)OCH_3$, $K_2CO_3$, $(CH_3CH_2CH_2CH_2)_4NI$, $CH_3CN$; (b) $HNO_3$, AcOH; (c) $Ph_3PCH_3Br$, sodium hexamethyldisilazide, THF; (d) $K_2OsO_2(OH)_4$, $(DHQ)_2PHAL$, *n*-prOH, *t*-Bu hypochlorite, *t*-Bu carbamate, NaOH, $H_2O$; (e) AcSH, $Ph_3P$, DIAD, THF; (f) *t*-BuSH, sodium methoxide, NaOH, MeOH; (g) $CH_3NH_2$, PyBOP, DIEA, DCM; (h) TFA, TIS, $H_2O$.

Nucleophilicity of the carboxylic acid in **8** during conjugation to peptides necessitated the elimination of this reactive group. High temperatures required for acid chloride formation with **8** led to decomposition of the compound. Resultantly, compound **8** was treated with methylamine and PyBOP in the presence of diisopropylethylamine (DIEA) in dichloromethane (DCM) to form amide **9**. Following removal of solvent *in vacuo*, compound **9** was dissolved in a 95:2.5:2.5 mixture of trifluoroacetic acid (TFA), triisopropylsilane (TIS), and water to effect removal of the $N^{\alpha}$-tert-butoxycarbonyl (Boc) protecting group, affording ligation auxiliary, **1**. Compound **1** was purified by process scale reverse-phase high performance liquid chromatography (RP-HPLC) and its identity was verified by $^{1}H$ and $^{13}C$ nuclear magnetic resonance (NMR) and high-resolution mass spectrometry (Figure 2.4).

**2.2. Site-specific peptide ubiquitylation**

As an intermediate step toward the semisynthesis of uH2B, we established a robust methodology for the site-specific ubiquitylation of a peptide using the C-terminal peptide of H2B as a model (Figure 2.5) (Chatterjee et al., 2007). To this end, peptidyl resin **10**, corresponding to residues 115-125 of human H2B and containing an A117C mutation, was synthesized by SPPS using the 9-fluorenylmethoxylcarbonyl (Fmoc) $N^{\alpha}$ protection strategy. The cysteine was included to allow further derivatization of the ubiquitylated product. Orthogonal protection of K6 of peptidyl resin **10**, representing K120 in the full-length H2B, with a hydrazine-labile ivDde group, allowed the regioselective installment of the ligation auxiliary. Following hydrazine treatment to give peptidyl resin **11**, bromoacetic acid was coupled to the ε-$NH_2$ of K6 with diisopropylcarbodiimide (DIC), affording peptidyl resin **12**. Reaction of the ligation auxiliary, **1**, with resin **12** in the presence of DIEA and 1,8-diazobicycloimide (DBU) in *N*,*N*-

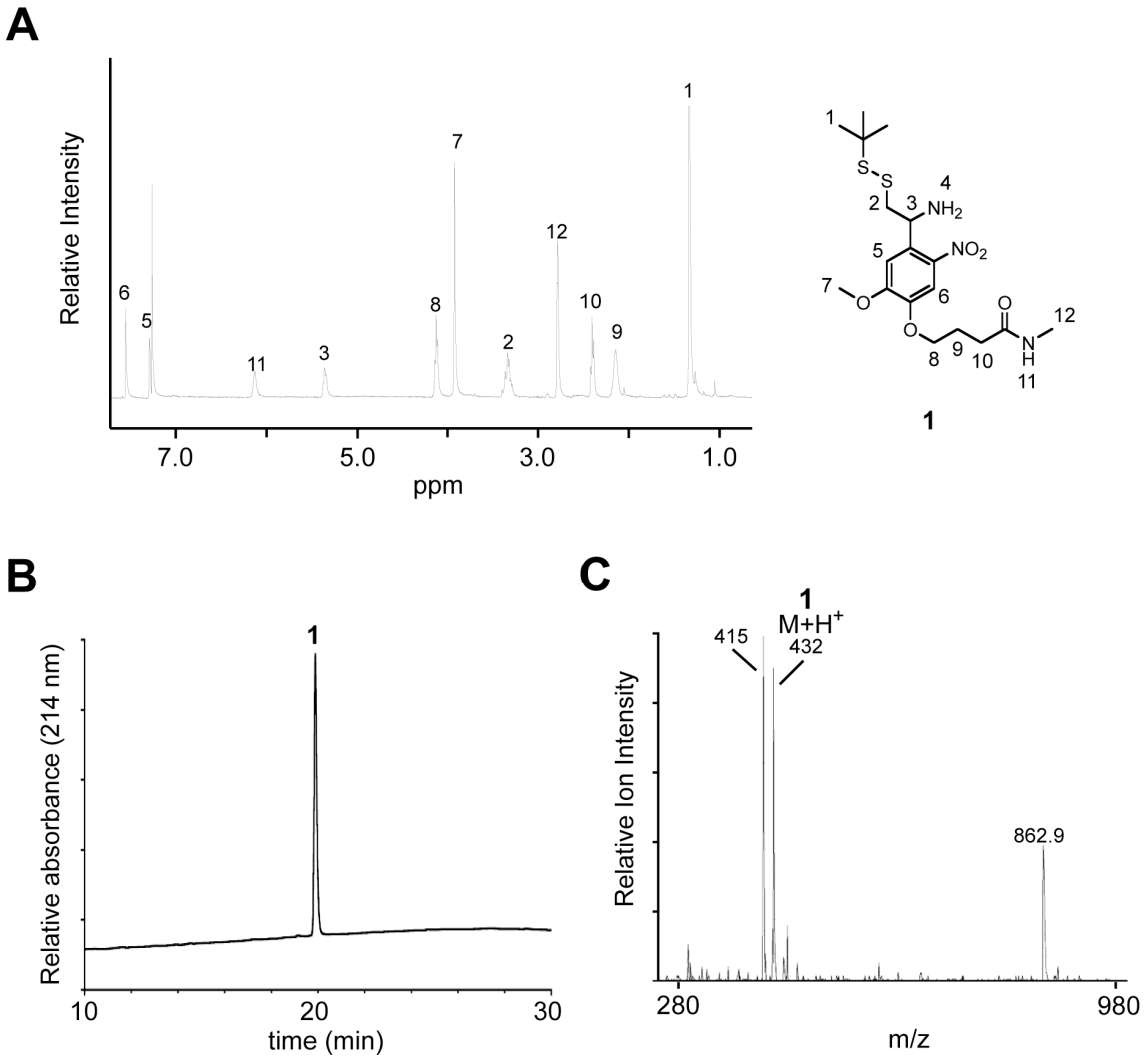



**Figure 2.4. Characterization of ligation auxiliary 1.** A) $^{1}$H NMR spectrum of **1**. Peaks are labeled relative to diagram of **1** at right. B) RP-HPLC chromatogram of compound **1**. C) ESI-MS spectrum of compound **1**. Peaks at 415 and 862.9 represent photolysis of CH-$NH_2$ bond and dimeric form of **1**, respectively.

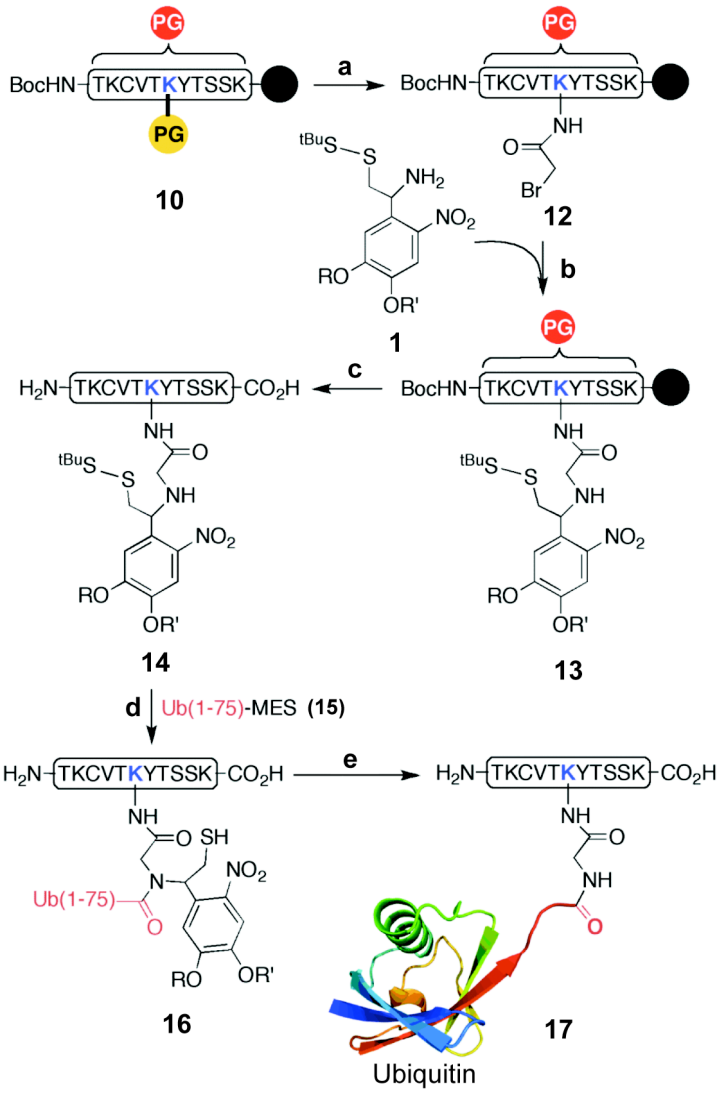



**Figure 2.5. Synthesis of auxiliary conjugated peptide and ligation to ubiquitin.** (a) 1. 2% $NH_2NH_2$, DMF; 2. $BrCH_2CO_2H$, DIC, DMF; (b) auxiliary **1**, DIEA, DBU, DMF; (c) TFA, TIS, anisole, $H_2O$; (d) 300 mM $NaP_i$, 50 mM MESNa, 25 mM TCEP, 3 M guanidinium HCl, pH 7.5; (e) He-Cd laser 325 nm. PG = protecting group, R = $CH_3$, R' = $CH_2CH_2CH_2C(O)NHCH_3$, MES = 2-mercaptoethanesulfonic acid.

dimethylformamide (DMF) yielded resin **13**. Auxiliary-conjugated peptide **14**, was purified by RP-HPLC following simultaneous side chain deprotection and cleavage from the resin using a 92.5:2.5:2.5:2.5 mixture of TFA, TIS, anisole, and water (Figure 2.6).

Peptide **14** was reduced with tris(2-carboxyethyl)phosphine (TCEP) to remove the *S*-tert-butyl protecting group on the ligation auxiliary. This deprotected peptide was ligated to ubiquitin(1-75)-α-thioester, **15**, generated by thiolysis of the corresponding intein fusion using mercaptoethanesulfonic acid (MES) (Figure 2.6). The time course of the EPL reaction was followed by RP-HPLC and was significantly slower than a typical cysteine-mediated EPL reaction, likely owing from the ligation onto a sterically hindered secondary amine (Figure 2.7). Resultantly, the reaction was allowed to proceed for 5 days prior to purification of intermediate product, **16**, by RP-HPLC. Removal of the ligation auxiliary was achieved by irradiation with a He-Cd laser at 325 nm. Three to four pulses of five seconds each led to nearly complete conversion of **16**, to ubiquitylated peptide **17**. Progression of the photolysis reaction was monitored by RP-HPLC - removal of the auxiliary resulted in the diminished absorption of the ubiquitylated peptide at 350 nm (Figure 2.8). Ubiquitylated peptide **17** was purified by RP-HPLC and its identity was confirmed by electrospray ionization mass spectrometry (ESI-MS).

To verify the site-specificity of protein ligation, we subjected protein **17** to an enzymatic ubiquitin hydrolysis reaction with ubiquitin C-terminal hydrolase L3 (UCH-L3). In addition to processing linear ubiquitin fusions, UCH-L3 is able to hydrolyze certain isopeptide conjugates of ubiquitin (Larsen et al., 1998; Misaghi et al., 2005; Tirat et al., 2005). Incubation of ubiquitylated peptide **17** with UCH-L3 resulted in almost complete conversion to ubiquitin(1-76) and the H2B

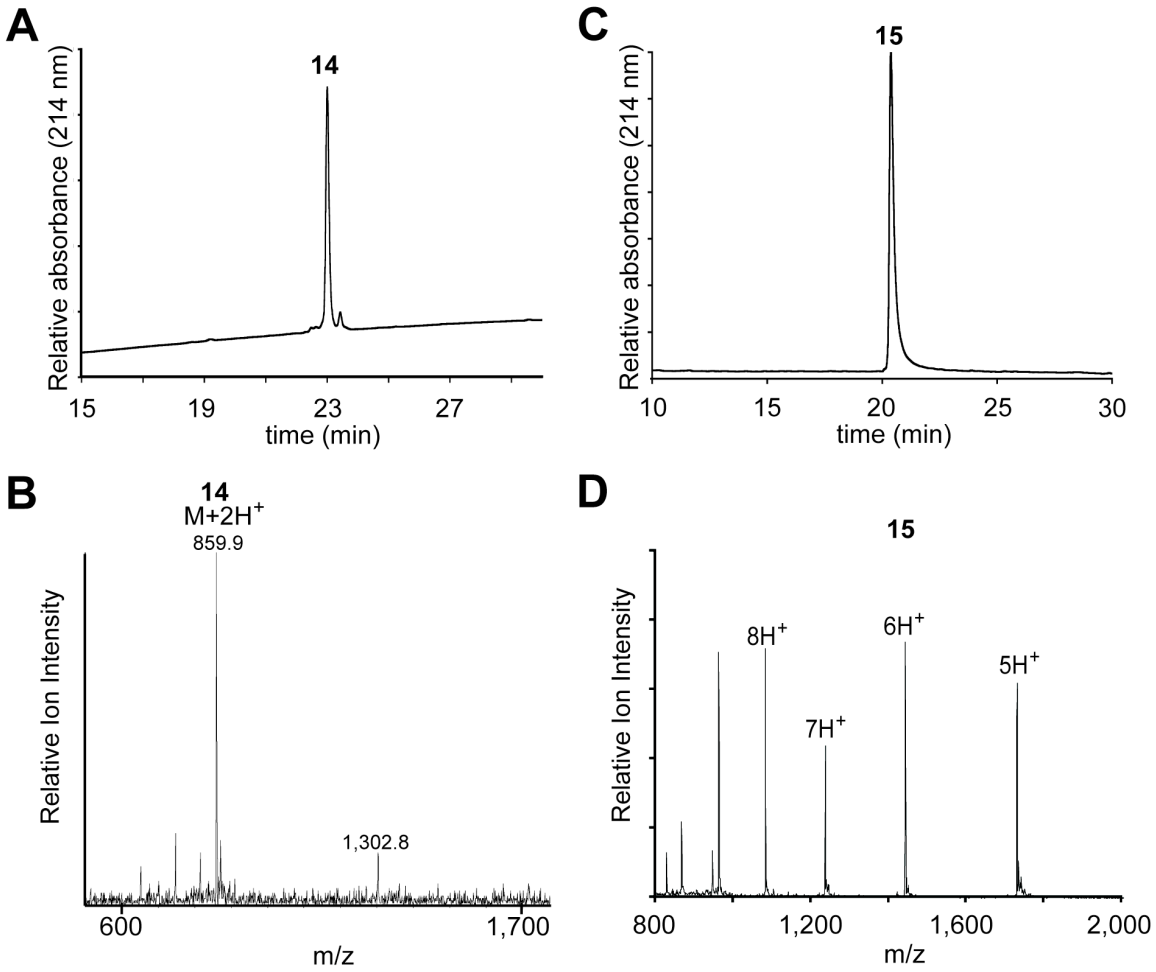



**Figure 2.6. Characterization of peptide 14 and ubiquitin(1-75)-α-thioester, 15.** A) RP-HPLC chromatogram, and B) ESI-MS spectrum of peptide **14**. $[(M+2H)^{2+}$ observed = 859.9 Da. $(M+2H)^{2+}$ expected = 859.5 Da.] 1,302.8 represents MS fragmentation of parent ion. C) RP-HPLC chromatogram, and D) ESI-MS spectrum of ubiquitin(1-75)-α-thioester, **15**. $[(M+H)^{+}$ observed = 8,632 ± 2 Da (s.d.). $(M+H)^{+}$ expected = 8,632 Da.] Charge states are labeled.

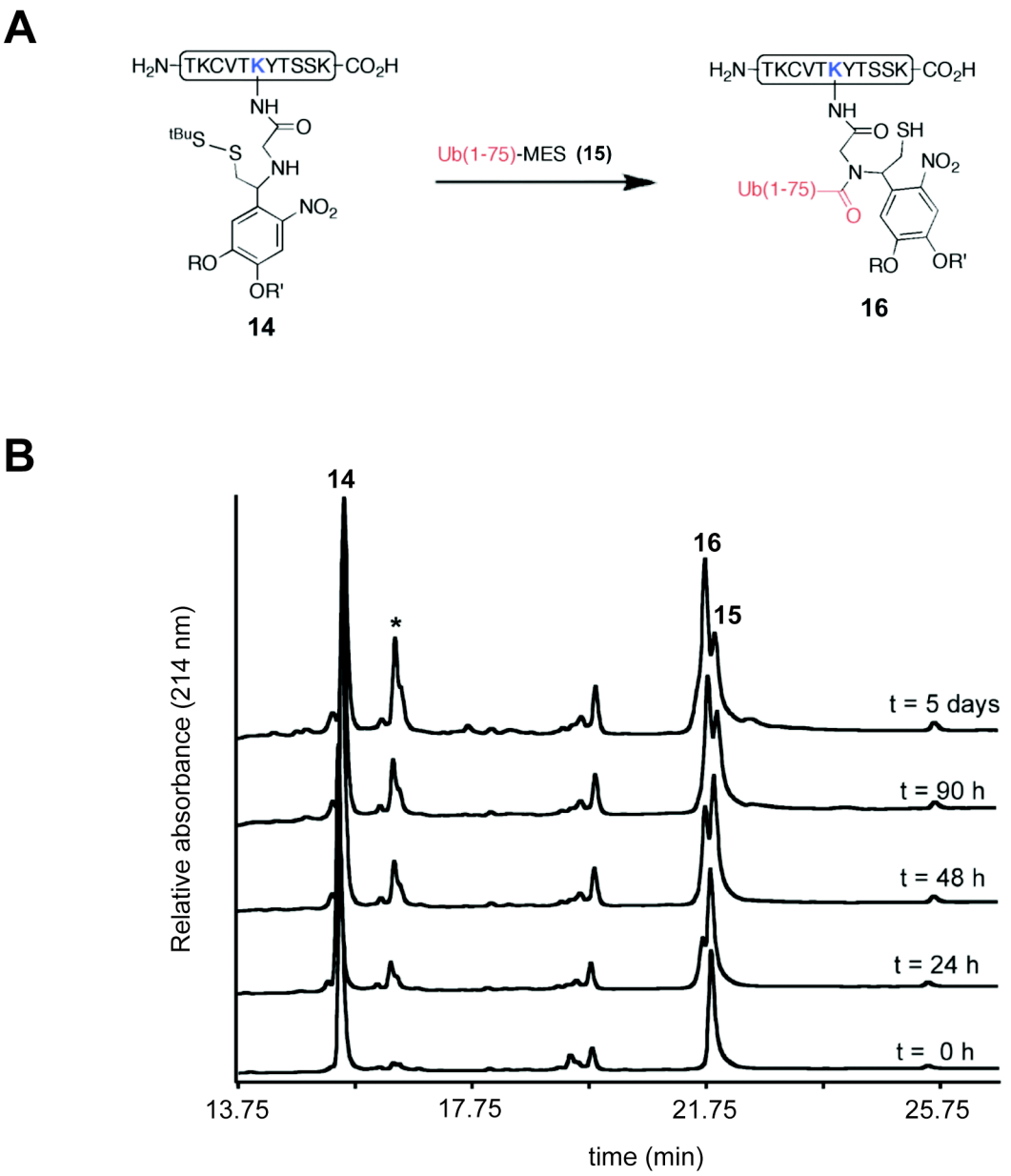



**Figure 2.7. EPL of peptide 14 and ubiquitin(1-75)-α-thioester, 15.** A) Schematic of auxiliary-mediated EPL linking ubiquitin(1-75)-α-thioester, **15**, with peptide **14**. B) Superimposed RP-HPLC chromatograms showing accumulation of ligation product **16** at 0, 24, 48, 90 h and 5 days. An asterisk indicates formation of a mixed disulfide between peptide **14** and MES.

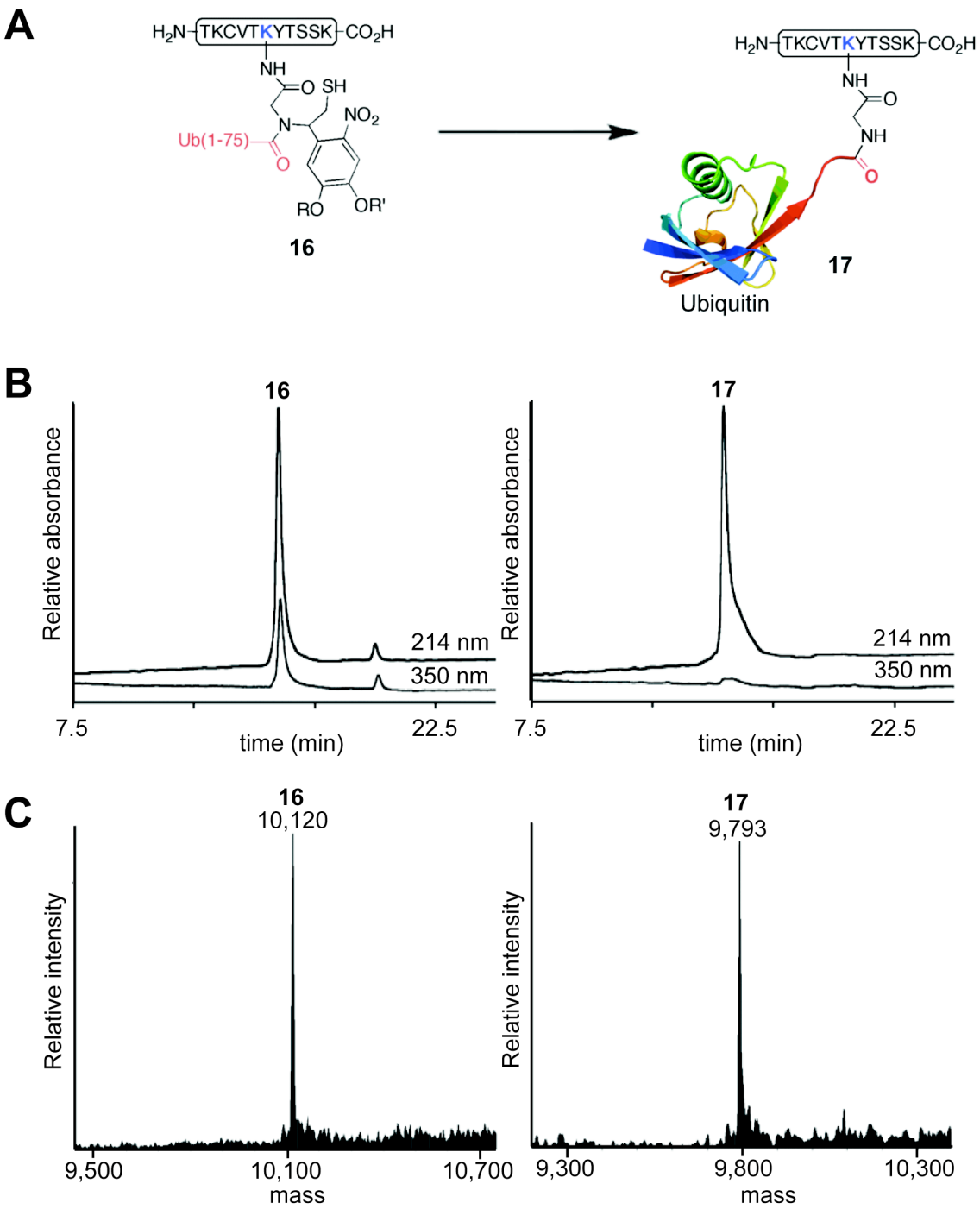



**Figure 2.8. Photolytic removal of ligation auxiliary forming ubiquitylated peptide.** A) Irradiation of protein **16** with a He-Cd laser at 325 nm effected removal of the ligation auxiliary, yielding ubiquitylated peptide **17**. B) RP-HPLC chromatograms of ligation product **16** and crude photolysis mixture containing ubiquitylated peptide **17**. C) Corresponding deconvoluted ESI-MS spectra of **16** [$(M+H)^+$ observed = 10,120 ± 2 Da (s.d.). $(M+H)^+$ expected = 10,120 Da.] and **17** [$(M+H)^+$ observed = 9,793 ± 4 Da. $(M+H)^+$ expected = 9,793 Da.]

peptide (Figure 2.9). As these products can only be generated from the expected protein ligation, this result verifies the site-specificity of our strategy. Additionally, successful ubiquitin hydrolysis indicates that the ubiquityl moiety in peptide conjugate **17** exhibits a native fold, validating this semisynthetic approach for the generation of biologically active ubiquitylated peptides.

In addition to ubiquitin, proteins are modified by various ubiquitin-like modifiers (Ubls), most notably, small ubiquitin-like modifier (SUMO) (Walsh, 2005). Typically, a single Ubl is conjugated to a target protein, analogous to protein monoubiquitylation. As most Ubls also terminate in Gly-Gly sequences, our semisynthetic strategy can be applied to the site-specific conjugation of Ubls to peptides. To demonstrate this generality, peptide **14** was conjugated with yeast SUMO ortholog, Smt3 (Figure 2.10). Peptide **14** was ligated to HA-tagged Smt3(2-97)-α-thioester, **18**, yielding intermediate product **19**. Irradiation of protein **19** as described above, afforded sumoylated peptide **20**. Thus, this semisynthesis facilitates the general site-specific modification of peptides with ubiquitin and Ubls.

### 2.3. Site-specific ubiquitylation of full-length H2B

Adapting this approach to the production of an ubiquitylated protein, in this case uH2B, requires the ligation of three polypeptides (Figure 2.11A) (McGinty et al., 2008). As H2B is devoid of cysteines, much like ubiquitin, this strategy necessitates the incorporation of additional functionality to facilitate a second traceless EPL reaction. With this in mind, peptide **21** was synthesized corresponding to residues 117-125 of *Xenopus* H2B bearing both the ligation auxiliary, **1**, attached to the ε-$NH_2$ of K120, and an A117C mutation (Figure 2.12). Orthogonal side-chain protection of K120 with the acid-labile 4-methyltritryl (Mtt) protecting group allowed the ligation auxiliary to be coupled to this amino group as described above. Also critical to the

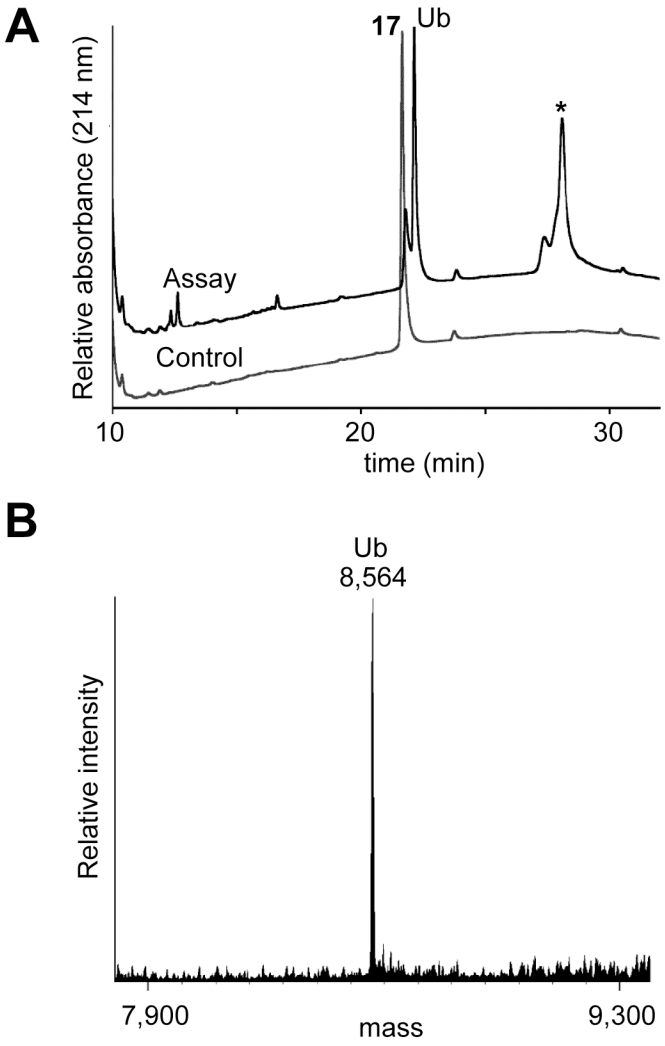



**Figure 2.9. Ubiquitin hydrolysis with UCH-L3.** A) Superimposed RP-HPLC chromatograms of **17** after incubation at 37 °C for 8 h with (assay) and without (control) UCH-L3. An asterisk indicates UCH-L3. B) Deconvoluted ESI-MS spectrum of hydrolyzed ubiquitin [$(M+H)^+$ observed = 8,564 ± 3 Da. $(M+H)^+$ expected = 8,566 Da.]

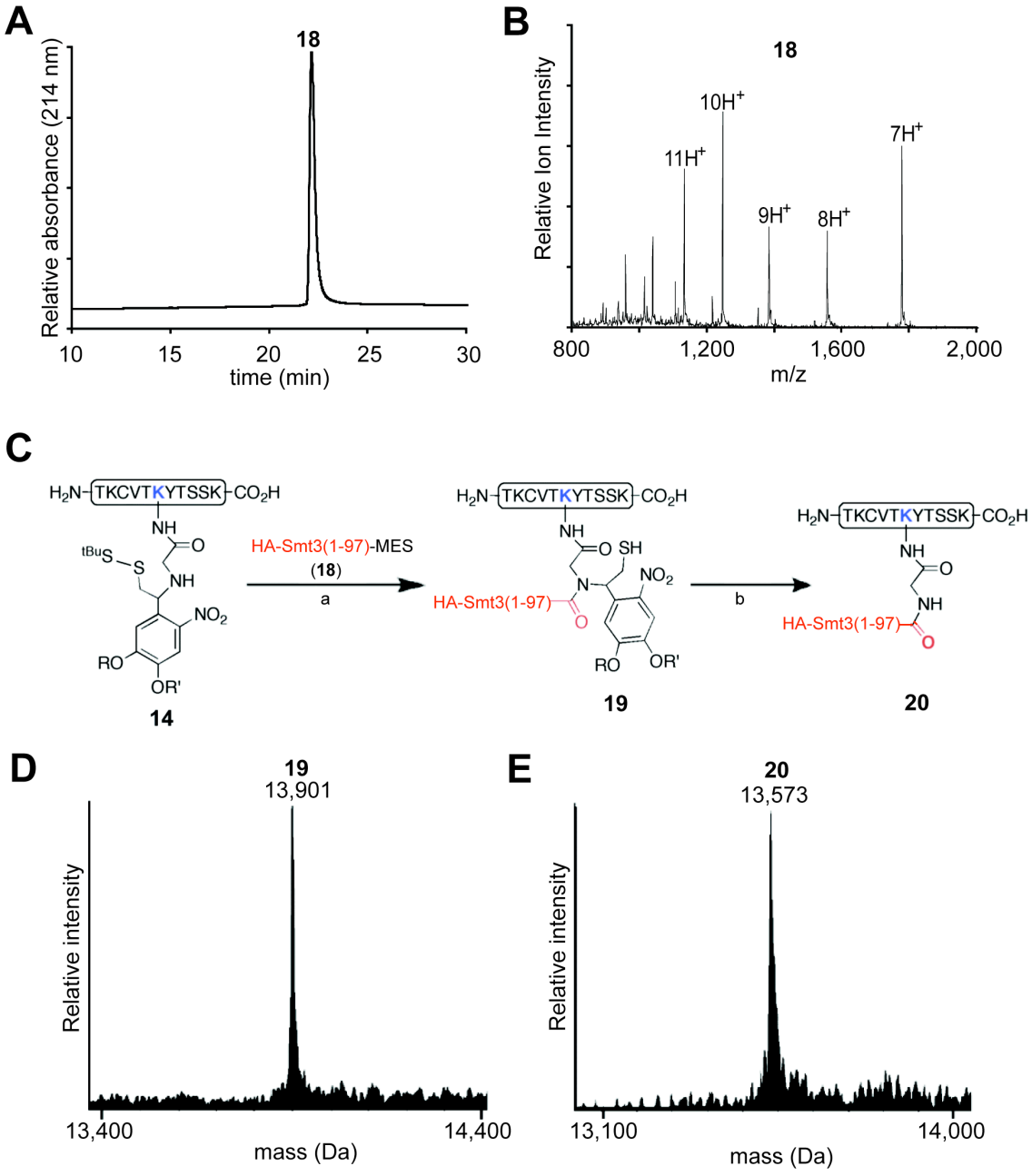



**Figure 2.10. Peptide sumoylation.** A) RP-HPLC chromatogram and B) ESI-MS spectrum of HA-Smt3(2-97)-α-thioester, **18.** [$(M+H)^+$ observed = 12,414 ± 3 Da (s.d.). $(M+H)^+$ expected = 12,414 Da.] C) (a) 300 mM $NaP_i$, 50 mM MESNa, 25 mM TCEP, 3 M guanidinium HCl, pH 7.5; (b) He-Cd laser 325 nm. D) Deconvoluted ESI-MS spectrum of ligation product **19**. [$(M+H)^+$ observed = 13,901 ± 5 Da. $(M+H)^+$ expected = 13,900 Da.] E) Deconvoluted ESI-MS spectrum of photolysis product **20**. [$(M+H)^+$ observed = 13,573 ± 4 Da. $(M+H)^+$ expected = 13,573 Da.]

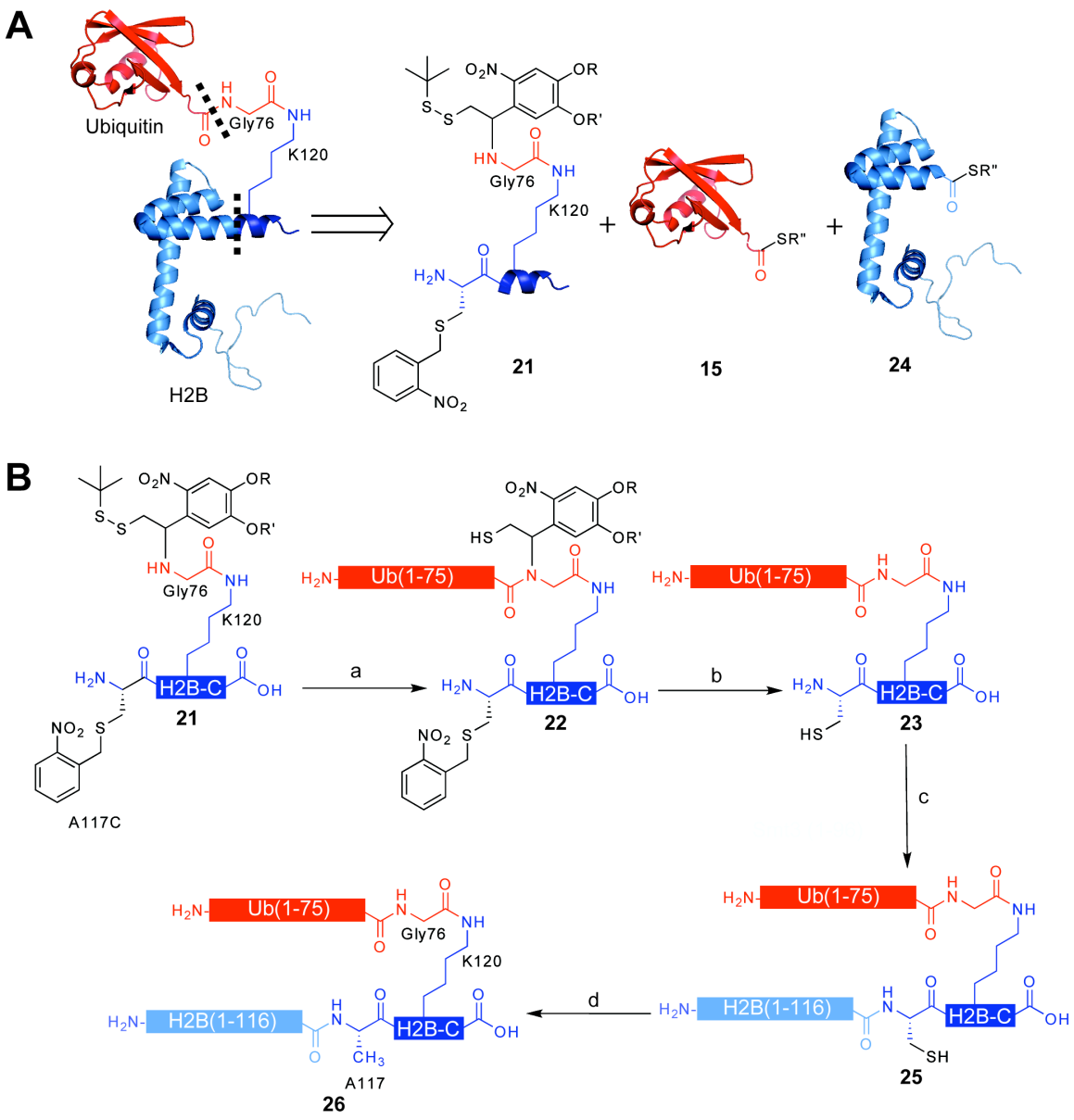



**Figure 2.11. Semisynthesis of ubiquitylated H2B.** A) Retrosynthetic analysis of uH2B synthesis. uH2B was generated via a 3-piece ligation with the following polypeptides: auxiliary linked synthetic peptide containing residues 117-125 of H2B and bearing an A117C mutation, H2B-C, **21**; recombinant ubiquitin(1-75)-α-thioester **15**; and recombinant H2B(1-116)-α-thioester, **24**. Dashed lines indicate junctions formed by EPL reactions. B) Synthetic scheme for the ubiquitylation of H2B. (a) EPL was used to ligate peptide **21** to protein **15**, forming branched protein **22**. (b) Ligation product **22** was irradiated with 365 nm light, yielding protein **23**. (c) Ligation of protein **23** to protein **24**, forming uH2BA117C, **25**. (d) Raney nickel desulfurization of protein **25**, forming uH2B, **26**. R = $CH_2CH_2CH_2C(O)NHCH_3$; R' = $CH_3$; R'' = $CH_2CH_2SO_3H$.

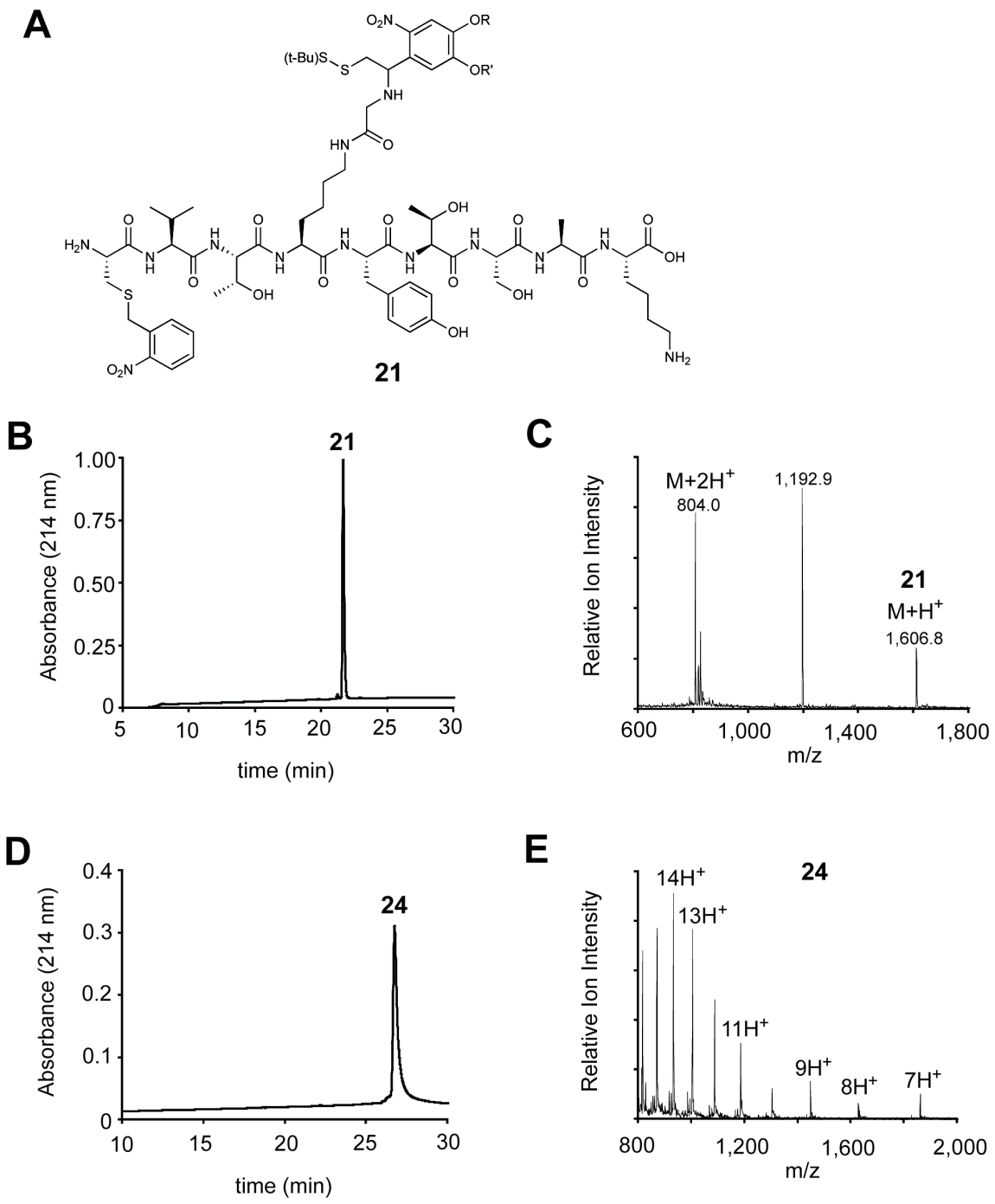



**Figure 2.12. Characterization of peptide 21 and H2B(1-116)-α-thioester, 24.** A) Structure of auxiliary conjugated peptide **21**. B) RP-HPLC chromatogram and C) ESI-MS spectrum of peptide **21**. Peak at 1,192.9 represents MS fragmentation product. [$(M+H)^+$ observed = 1,606.8 Da. $(M+H)^+$ expected = 1,607.9 Da.] D) RP-HPLC chromatogram and E) ESI-MS spectrum of H2B(1-116)-α-thioester, **24**. Representative charge states are labeled [$(M+H)^+$ observed = 12,991 ± 3 Da (s.d.). $(M+H)^+$ expected = 12,991 Da.] R = $CH_2CH_2CH_2C(O)NHCH_3$; R' = $CH_3$.

synthetic design, was transient protection of the N-terminal cysteine in **21**, which precluded unwanted double ligation of ubiquitin. After examining several possibilities, we settled on the photoremovable *S*-(*o*-nitrobenzyl) group for this purpose, as this proved to be completely stable during the course of the first ligation reaction and was easily removed by photolysis.

In the first step of the synthesis, the disulfide in peptide **21** was reduced with TCEP followed by ligation to ubiquitin(1-75)-α-thioester, **15**, to give ubiquitylated peptide, **22** (Figure 2.13A, panel i and Figure 2.13B). Protein **22** was purified by RP-HPLC and subsequently irradiated at 365 nm, resulting in efficient removal of both the ligation auxiliary and the cysteine protecting group to give deprotected branched protein **23** (Figure 2.13A, panel ii, and Figure 2.13C). In this case, irradiation was performed with a collimated light source equipped with an inline filter, permitting larger scale photolysis reactions. Purified intermediate product, **23**, was then ligated to recombinant H2B(1-116)-α-thioester, **24** (Figure 2.12), to give uH2BA117C, **25** (Figure 2.13A, panel iii and Figure. 2.13D). In the final step, Raney nickel-mediated desulfurization (Yan and Dawson, 2001) was used to convert the single cysteine residue in branched protein **25** to the native alanine residue present in uH2B (Figure 2.13A, panel iv, and Figure 2.13E). Under optimized conditions, this reduction was found to be highly specific for cysteine desulfurization over methionine, thereby affording native ubiquitylated H2B, **26**. Excessively long incubations with Raney nickel led to a second desulfurization reaction involving methionine. The overall yield of the semisynthesis (i.e. steps a-d in Figure 2.11B) was excellent (20%).

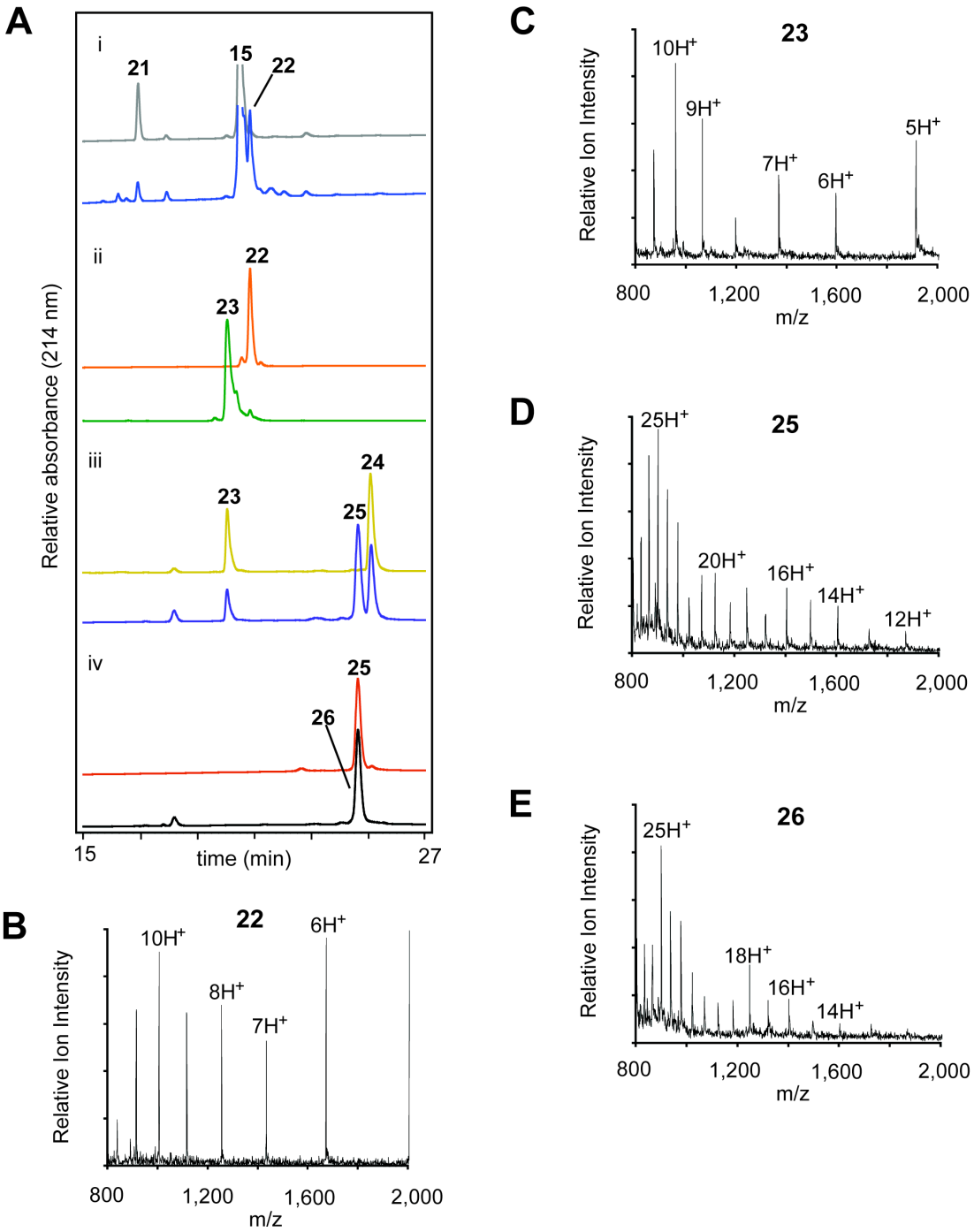



**Figure 2.13. Semisynthesis of uH2B.** A) Superimposed RP-HPLC chromatograms of starting materials, top trace, and products, bottom trace of each pair, for reactions a-d described in Figure 2.11. ESI-MS spectrum of B) ligation product **22** [$(M+H)^+$ observed = 10,008 ± 1 Da (s.d.). $(M+H)^+$ expected = 10,009 Da.], C) photolysis product **23** [$(M+H)^+$ observed = 9,547 ± 2 Da. $(M+H)^+$ expected = 9,548 Da.], D) ligation product **25** [$(M+H)^+$ observed = 22,395 ± 3 Da. $(M+H)^+$ expected = 22,396 Da.], and E) desulfurization product **26** [$(M+H)^+$ observed = 22,366 ± 4 Da. $(M+H)^+$ expected = 22,365 Da.]. Representative charge states are labeled.

## 2.4 Summary

The site-specific ubiquitylation and sumoylation of a peptide was accomplished using a photolytically removable ligation auxiliary. This strategy was extended to the semisynthesis of native uH2B. To accomplish this, two orthogonal traceless EPL reactions were performed – the first using the auxiliary, and the second, a cysteine-mediated ligation coupled with a chemical desulfurization to convert the cysteine to the native alanine. Importantly, these methodologies are applicable to the conjugation of ubiquitin and Gly-Gly containing Ubls to other histone, as well as non-histone peptides and proteins. Mechanistic investigations into the dependence of Dot1L on H2B ubiquitylation using semisynthetic uH2B are presented in the following chapter.

## Chapter 3: uH2B directly stimulates Dot1L-mediated methylation[2]

Three models have been proposed to explain the dependency of H3K79 methylation on H2B ubiquitylation (Figure 3.1). Under the first model, ubiquitin, being a large modification relative to the nucleosome, acts as a wedge between adjacent nucleosomes, thereby favoring a less condensed chromatin conformation (Henry and Berger, 2002; Sun and Allis, 2002). In so doing, uH2B increases the accessibility of the nucleosomal surface, including H3K79 to Dot1. Alternatively, the second model suggests that Dot1 recognizes elements of ubiquitin and the nucleosome and as such, is selectively recruited to ubiquitylated nucleosomes (Henry and Berger, 2002). Moreover, because all evidence linking uH2B to H3K79 methylation has been observed following genetic perturbations *in vivo*, it is equally possible that one or more additional factors may be required to translate the effect of uH2B into heightened Dot1 methyltransferase activity. A number of factors have been implicated in carrying out this function in yeast, including proteasomal ATPases and a component of the Set1 complex, Cps35. Mutations in two proteasomal ATPases, Rpt4 and Rpt6, disrupt H3K79 methylation in *S. cerevisiae* leaving ubiquitylation of H2B intact (Ezhkova and Tansey, 2004). A Dot1-Cps35 interaction has been demonstrated in yeast by co-immunoprecipitation (Lee et al., 2007). Strains bearing temperature sensitive alleles of Cps35 show decreased trimethylation of H3K79 at restrictive temperatures. How these factors influence Dot1 activity is unknown. Furthermore, the role of these factors in H3K79 methylation in higher organisms has not been investigated. In this chapter, semisynthetic uH2B (Chapter 2) is used to demonstrate a direct stimulation of Dot1L-mediated intranucleosomal methylation of H3K79 (McGinty et al., 2008).

---

[2] The work in this chapter was performed as part of a close collaboration with Dr. Jaehoon Kim in the Laboratory of Biochemistry and Molecular Biology at the Rockefeller University.

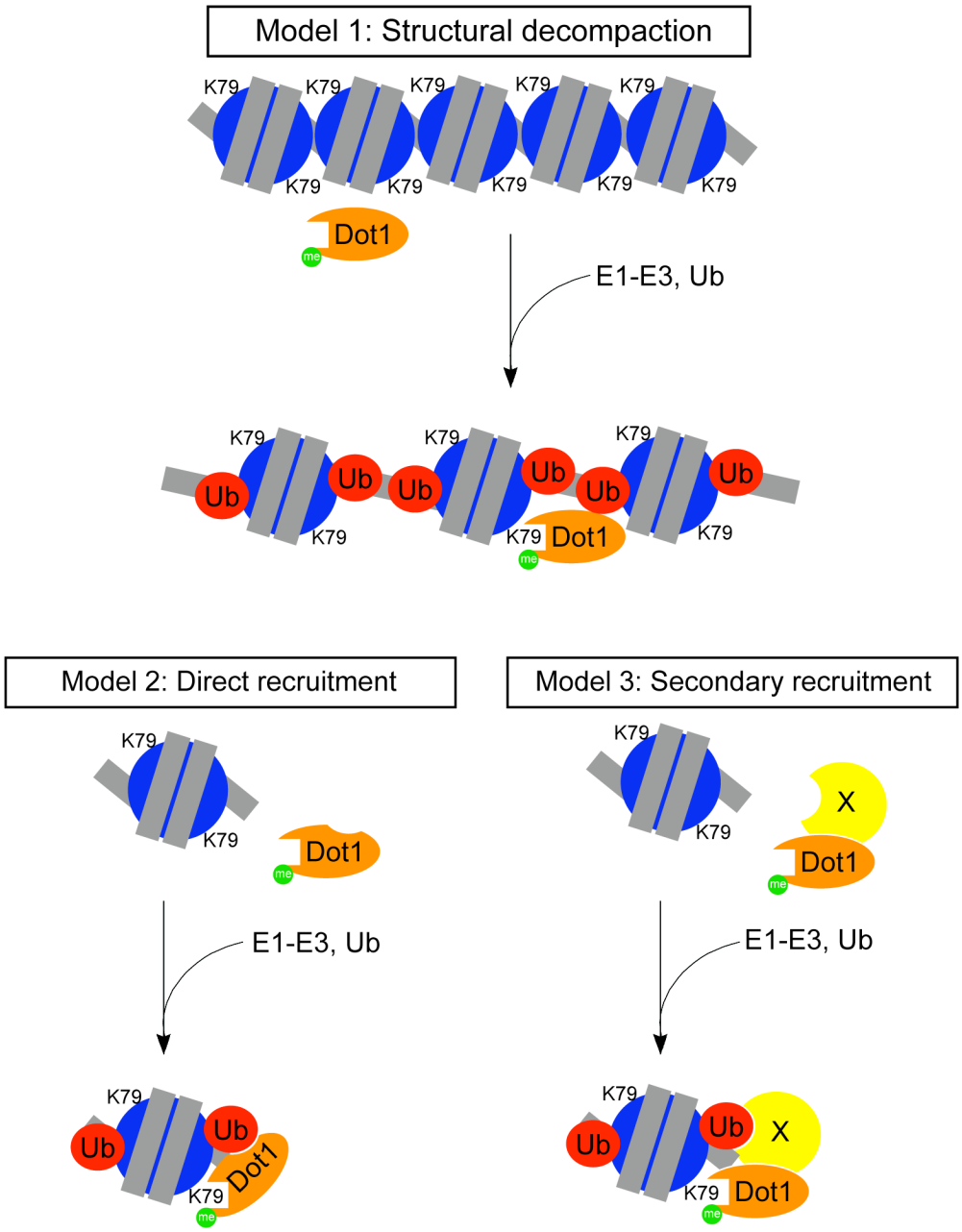



**Figure 3.1. Proposed mechanisms for uH2B-mediated stimulation of Dot1.** Model 1: ubiquitin acts as a wedge between adjacent nucleosomes, loosening chromatin structure and increasing accessibility of Dot1 to H3K79. Model 2: Dot1 recognizes ubiquitin within the context of the nucleosome, allowing recruitment specifically to ubiquitylated nucleosomes. Model 3: One or more intermediate factors recruited by uH2B facilitates the secondary recruitment of Dot1.

### 3.1. Reconstitution of ubiquitylated octamers and nucleosomes

The interrogation of the role of uH2B in H3K79 methylation minimally requires the preparation of histone octamers and nucleosomes. A robust methodology exists for the reconstitution of octamers from individual recombinant histones (Dyer et al., 2004). In fact, the preparation of histone octamers containing purified uH2B has been reported (Davies and Lindsey, 1994). Recombinant *Xenopus* H2A, H2B, H3, H4, and H3K79R were purified from *E. coli* inclusion bodies using size-exclusion chromatography followed by process scale RP-HPLC (Figure 3.2). To form octamers, purified recombinant histones and semisynthetic uH2B, **26**, were dissolved in a buffer containing 7 M guanidinium HCl and combined in equal molar concentrations. The resultant mixture was dialyzed against a 2 M NaCl solution lacking chaotropic agents, allowing the spontaneous formation of octamers. Octamers were purified by size-exclusion chromatography and appropriate constituents were verified by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 3.3). No significant difference in the efficiency of octamer reconstitution was observed when semisynthetic uH2B was substituted for recombinant H2B.

Nucleosomes can be reconstituted from preformed octamers and DNA by dialysis or dilution (Dyer et al., 2004; Owen-Hughes et al., 1999). Both affect nucleosome assembly through step-wise decreases in monovalent salt concentration. Nucleosomes were prepared by dilution using octamers containing uH2B and/or H3K79R and a PCR amplified DNA fragment containing 147 bp of the Widom 601 nucleosomal positioning sequence (Figure 3.4A). With optimized ratios of octamer and DNA, nucleosomes were reconstituted efficiently as evidenced by a single species observed following native-PAGE (Figure 3.4B and C). Arrays of nucleosomes and chromatinized plasmids can also be prepared using dialysis or ATP-dependent

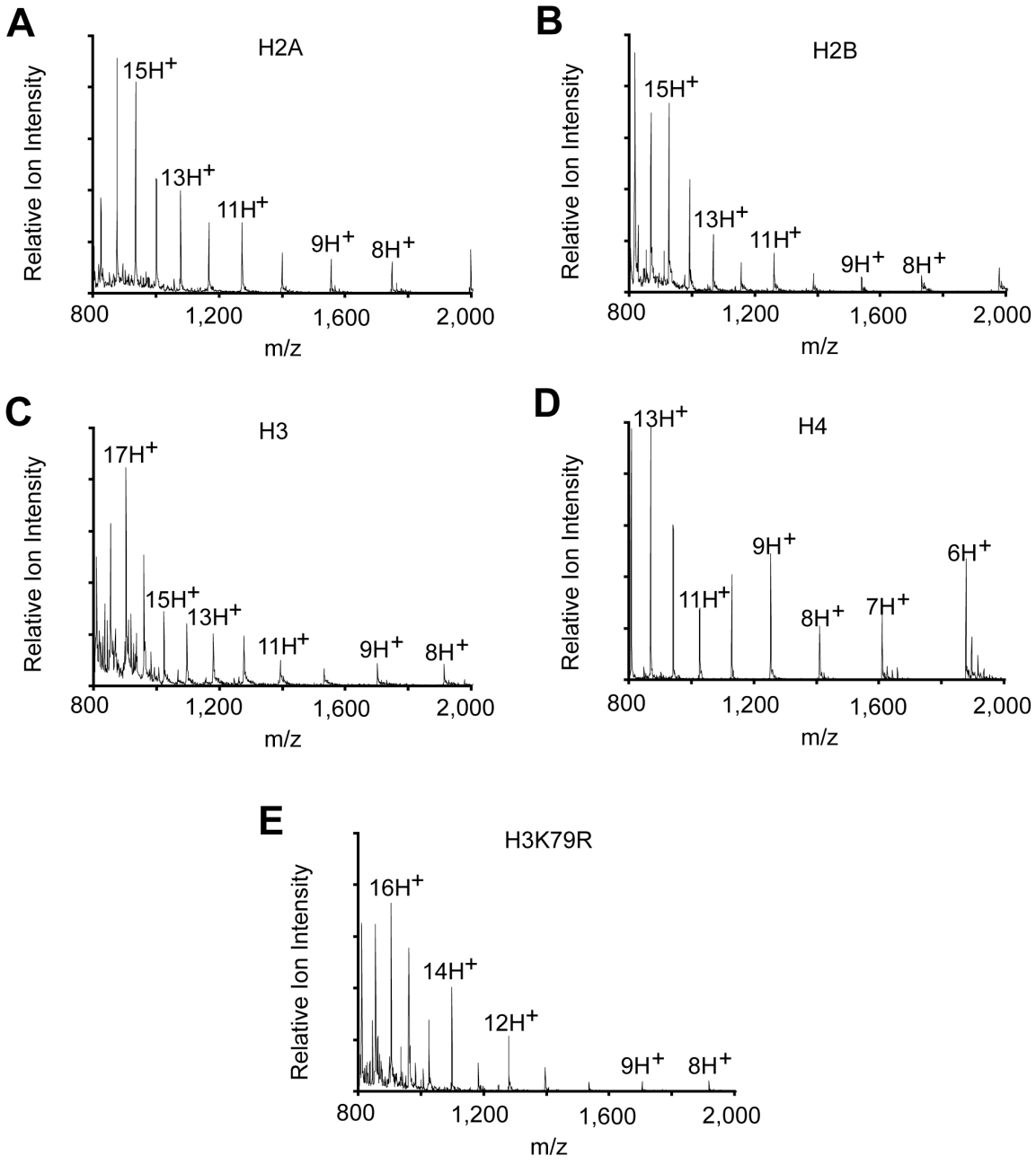



**Figure 3.2. Characterization of recombinant histone proteins.** ESI-MS spectra of purified recombinant histones. A) H2A [$(M+H)^+$ observed = 13,949 ± 1 Da (s.d.). $(M+H)^+$ expected = 13,951 Da.], B) H2B [$(M+H)^+$ observed = 13,814 ± 3 Da. $(M+H)^+$ expected = 13,818 Da.], C) H3 [$(M+H)^+$ observed = 15,270 ± 1 Da. $(M+H)^+$ expected = 15,272 Da.], D) H4 [$(M+H)^+$ observed = 11,236 ± 1 Da. $(M+H)^+$ expected = 11,237 Da.], and E) H3K79R [$(M+H)^+$ observed = 15,300 ± 3 Da. $(M+H)^+$ expected = 15,300 Da.].

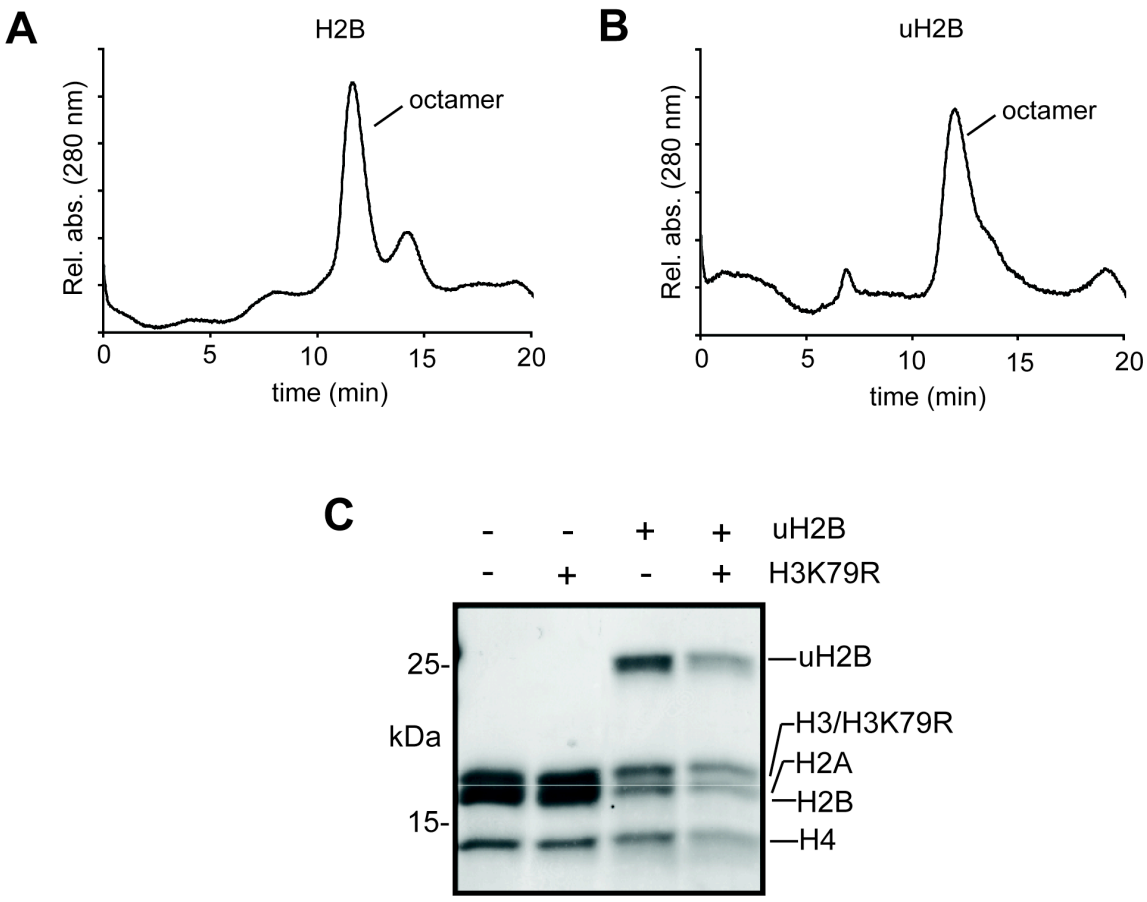



**Figure 3.3. Reconstitution of octamers containing uH2B.** Size exclusion chromatograms of octamers containing A) H2B or B) uH2B. Octamer peaks are labeled. C) Analysis of octamers containing uH2B and/or H3K79R by 15% SDS-PAGE and Coomassie staining.

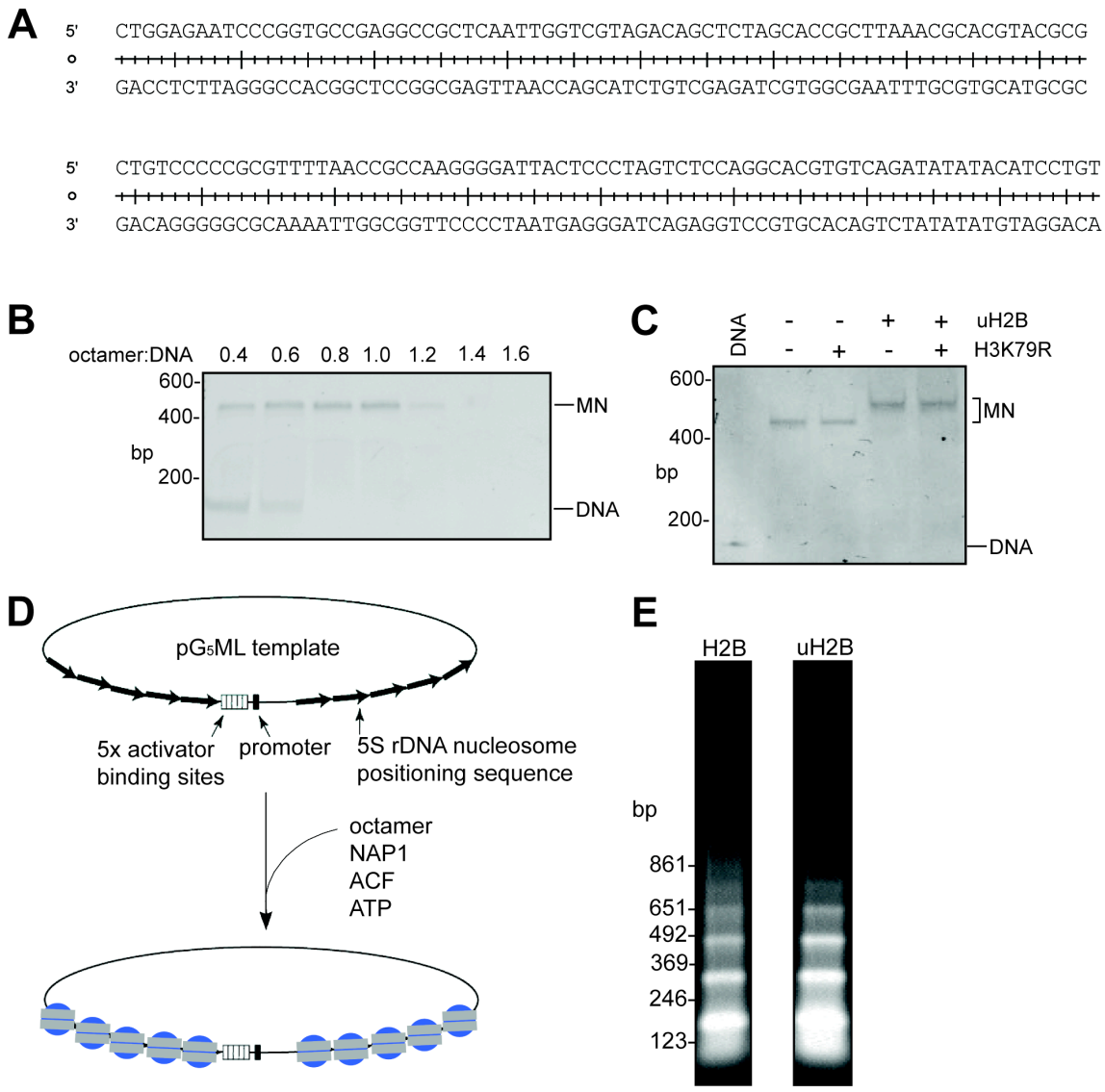



**Figure 3.4. Formation of nucleosomes and chromatinized plasmids with uH2B.** A) 147 bp Widom 601 positioning sequence used in nucleosome formations. B) Representative optimization of octamer:DNA ratios for nucleosome formation. Nucleosomes analyzed by 5% native-PAGE followed by ethidium bromide staining are shown. C) Ethidium bromide stained 5% gel of nucleosomes reconstituted with uH2B and/or H3K79R. D) Schematic of plasmid chromatinization with ATP dependent remodelers, ACF and NAP1. E) Micrococcal nuclease digestion of plasmids chromatinized with octamers containing H2B or uH2B. An agarose gel of the digestion products stained with ethidium bromide is shown. Diagram in panel D adapted from a previously published figure (An and Roeder, 2004).

chromatin assembly factors. Using the latter strategy, a plasmid was chromatinized with octamers containing either recombinant H2B or semisynthetic uH2B (Figure 3.4D). Micrococcal nuclease digestion of the resultant plasmids verified chromatin formation (Figure 3.4E).

### 3.2. uH2B directly stimulates Dot1L-mediated methylation

With modified nucleosomes in hand, we next turned our attention to exploring the effect of uH2B on Dot1L methyltransferase activity. Full-length human FLAG-Dot1L (herein referred to as Dot1L) was isolated from an insect cell overexpression system using M2 agarose (Figure 3.5A). Dot1L was able to methylate unmodified histone octamers in an assay containing $^{3}$H *S*-adenosyl methionine (SAM). However, no activity was observed on H3 alone, $(H3/H4)_2$ tetramers, or on mononucleosomes and chromatinized plasmids assembled with recombinant unmodified histones (Figure 3.5B and Figure 3.6B). This suggests that the presence of nucleosomal DNA represses the activity of Dot1L *in vitro*. This is in stark contrast to yeast methyltransferase, Dot1, which exhibits greater activity on recombinant nucleosomes than octamers (Sawada et al., 2004). We predicted that ubiquitylation of H2B would stimulate Dot1L methyltransferase activity in the context of nucleosomes. To test this hypothesis, a $^{3}$H SAM methyltransferase assay was performed using Dot1L and chemically defined nucleosomes. Robust methyltransferase activity was observed on mononucleosomes containing uH2B, whereas no radioactivity was detected on unmodified mononucleosomes (Figure 3.6A). This activity was specific for H3K79 as Dot1L was incapable of methylating nucleosomes containing both uH2B and an H3K79R mutant. Importantly, the level of methyltransferase activity observed on ubiquitylated mononucleosomes was far greater than that observed on unmodified histone octamers (Figure 3.6C). Dot1L also methylated a chromatinized plasmid containing uH2B

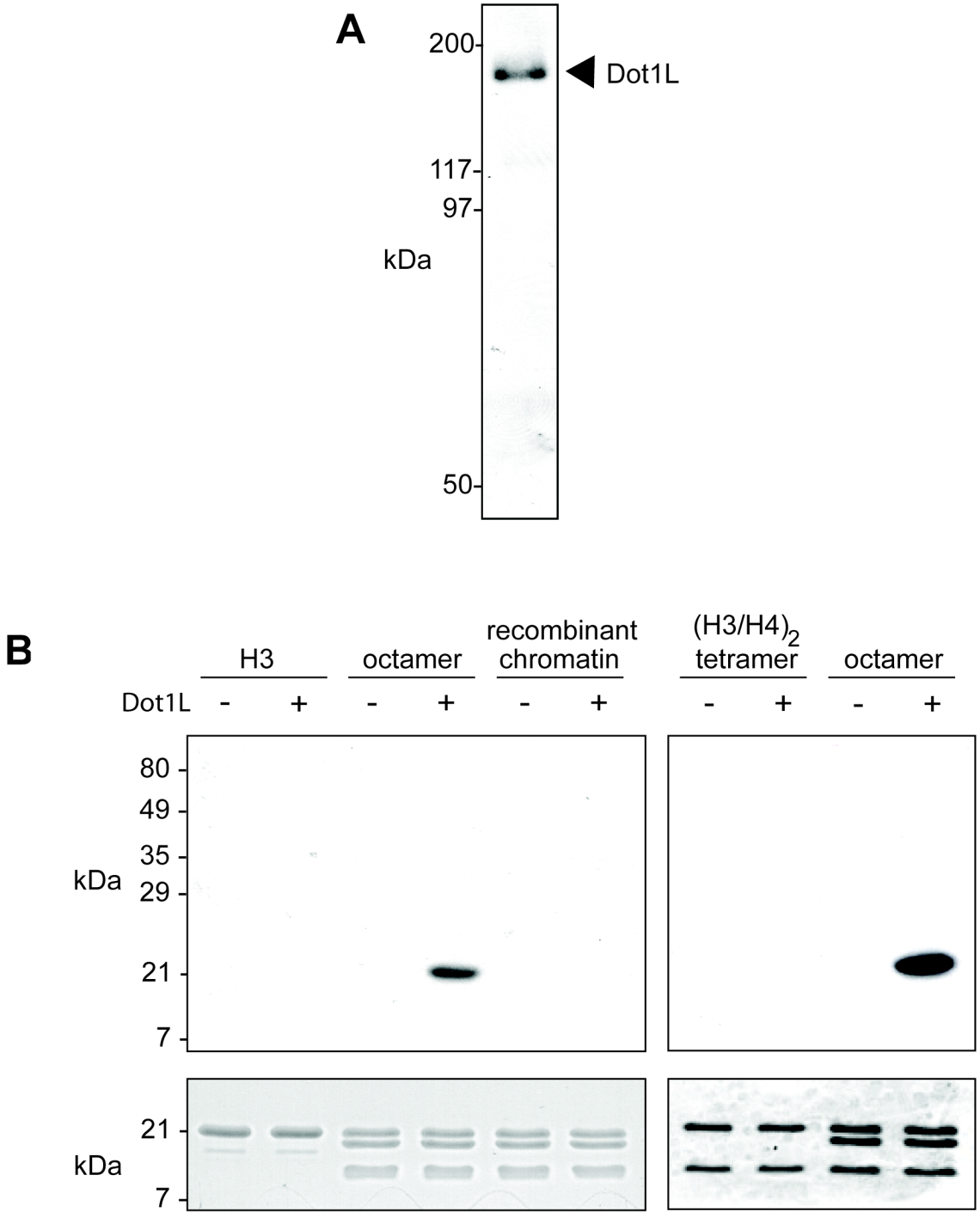



**Figure 3.5. Activity of Dot1L on unmodified substrates.** A) Coomassie stained SDS-PAGE gel of purified FLAG-tagged Dot1L. B) $^{3}$H SAM methyltransferase assay with Dot1L and indicated substrates separated by SDS-PAGE and stained with Coomassie (bottom panels) prior to probing for $^{3}$H methyl incorporation by fluorography (top panels).

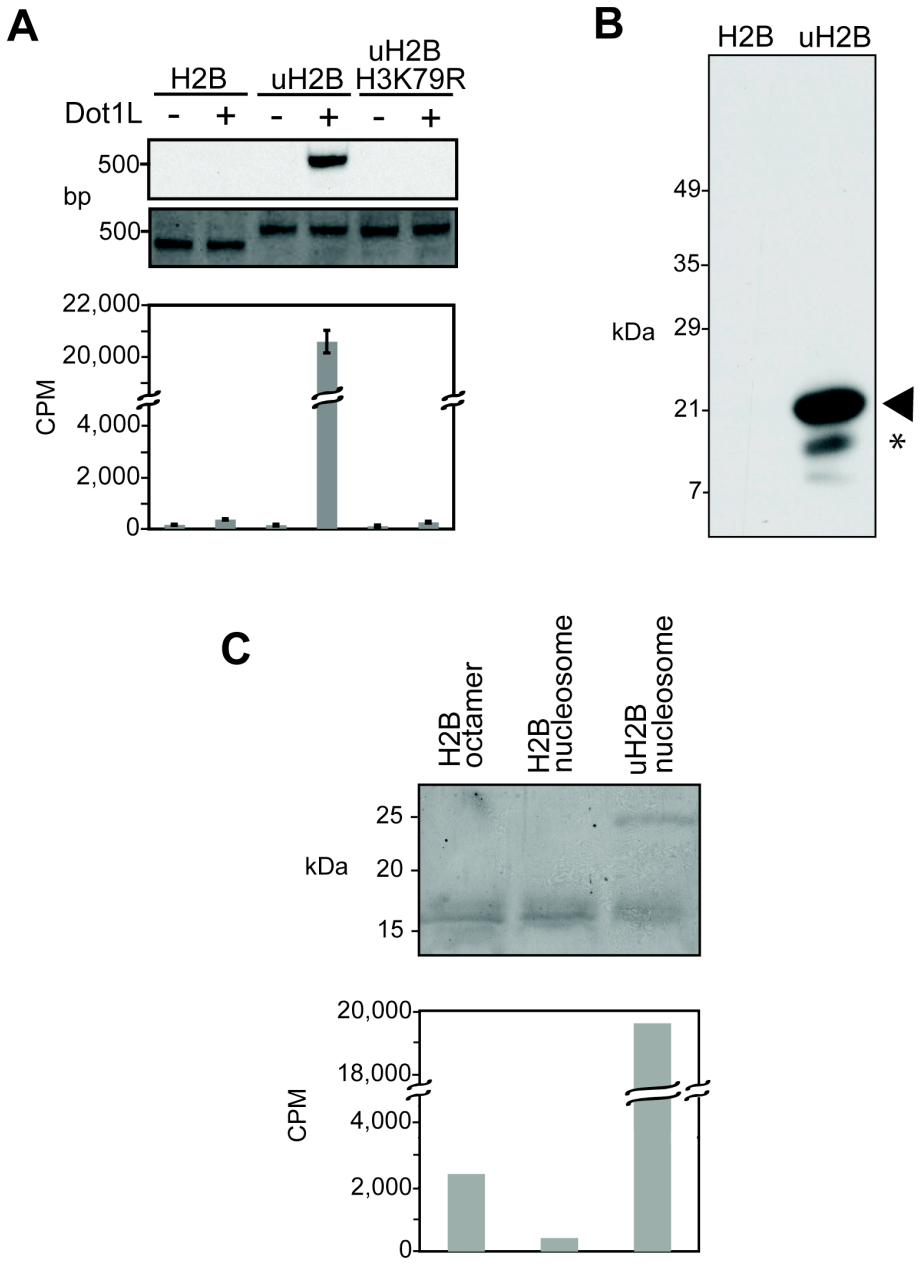



**Figure 3.6. Effects of uH2B on Dot1L-mediated methylation.** A) Dot1L methyltransferase assay using chemically defined nucleosomes. Assays performed on mononucleosomes with $^{3}$H SAM and Dot1L were separated on a native 5% gel and stained with ethidium bromide (middle panel) prior to probing for $^{3}$H methyl incorporation by fluorography (top panel). Quantification was performed by liquid scintillation counting (bottom panel). B) $^{3}$H SAM methyltransferase assay using indicated chromatinized plasmids. Assay samples were separated by 15% SDS-PAGE and fluorography was performed. Arrow marks H3. An asterisk indicates an H3 proteolytic product. C) $^{3}$H SAM assay with indicated substrates separated by 15% SDS-PAGE and stained with Coomassie (top panel). Quantification was performed by liquid scintillation counting (bottom panel). Error bars indicate s.e.m. (n = 4-7).

(Figure 3.6B). These results establish a direct biochemical connection between ubiquitylated H2B and H3K79 methylation by Dot1L.

Although H3K79 and H2BK120 reside in separate polypeptides, their side-chains are located in close proximity (~31 Å) to one another on the face of the nucleosome (Luger et al., 1997a), posing a structural basis for a crosstalk pathway (Figure 3.7A). Indeed, docking of the structure of the catalytic domain of Dot1L onto the mononucleosome structure, positions the catalytic domain adjacent to the site of H2B ubiquitylation (Min et al., 2003). Therefore, a GST fusion of the catalytic domain of Dot1L, containing residues 1-416 (herein referred to as $Dot1L_{cat}$), was purified from an *E. coli* expression system in order to interrogate its role in uH2B-dependent H3 methylation (Figure 3.7B). Similar to our observations with Dot1L, a significant enhancement in activity of $Dot1L_{cat}$ was measured on mononucleosomes containing uH2B, when compared to unmodified mononucleosomes (Figure 3.7C). However, unlike the full-length enzyme, $Dot1L_{cat}$ also exhibited some, albeit minimal, methyltransferase activity on unmodified mononucleosomes.

H2B ubiquitylation has been correlated with increased levels of di- and trimethylation of H3 K79 in humans (Kim et al., 2005; Zhu et al., 2005) and yeast (Shahbazian et al., 2005), respectively. Therefore, we examined the degree of methylation occurring in our assays. In-gel trypsin digestion of H3, followed by MALDI-mass spectrometry showed some monomethylation and no di- and trimethylation of unmodified mononucleosomes, in assays performed with either Dot1L or the catalytic domain alone (Figure 3.8A and B, top panels). However, robust mono- and dimethylation of nucleosomes containing uH2B was observed in both cases (Figure 3.8A and B, bottom panels). No evidence of trimethylation was observed is our assays, which is

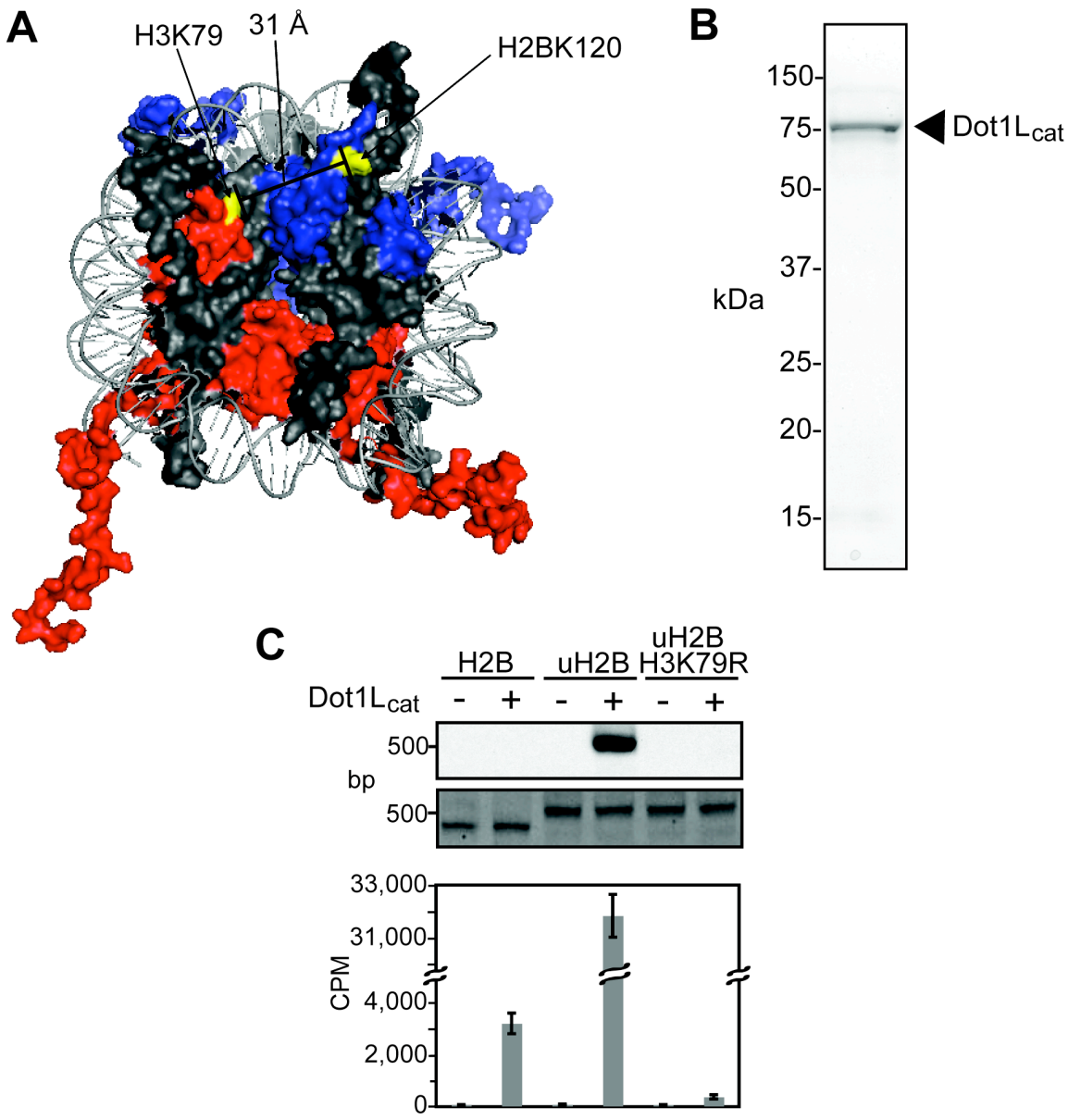



**Figure 3.7. Effects of uH2B on Dot1L$_{cat}$ activity.** A) Surface rendering of nucleosome (1KX5) (Davey et al., 2002). H2B and H3 shown in blue and red, respectively. H2BK120 and H3K79 are colored yellow and labeled. B) Coomassie stained 15% SDS-PAGE gel of purified Dot1L$_{cat}$. C) Dot1L$_{cat}$ methyltransferase assay using chemically defined nucleosomes. Assays performed on mononucleosomes with $^{3}$H SAM and Dot1L$_{cat}$ were separated on a native 5% gel and stained with ethidium bromide (middle panel) prior to probing for $^{3}$H methyl incorporation by fluorography (top panel). Quantification was performed by liquid scintillation counting (bottom panel). Error bars indicate s.e.m. (n = 4-7).

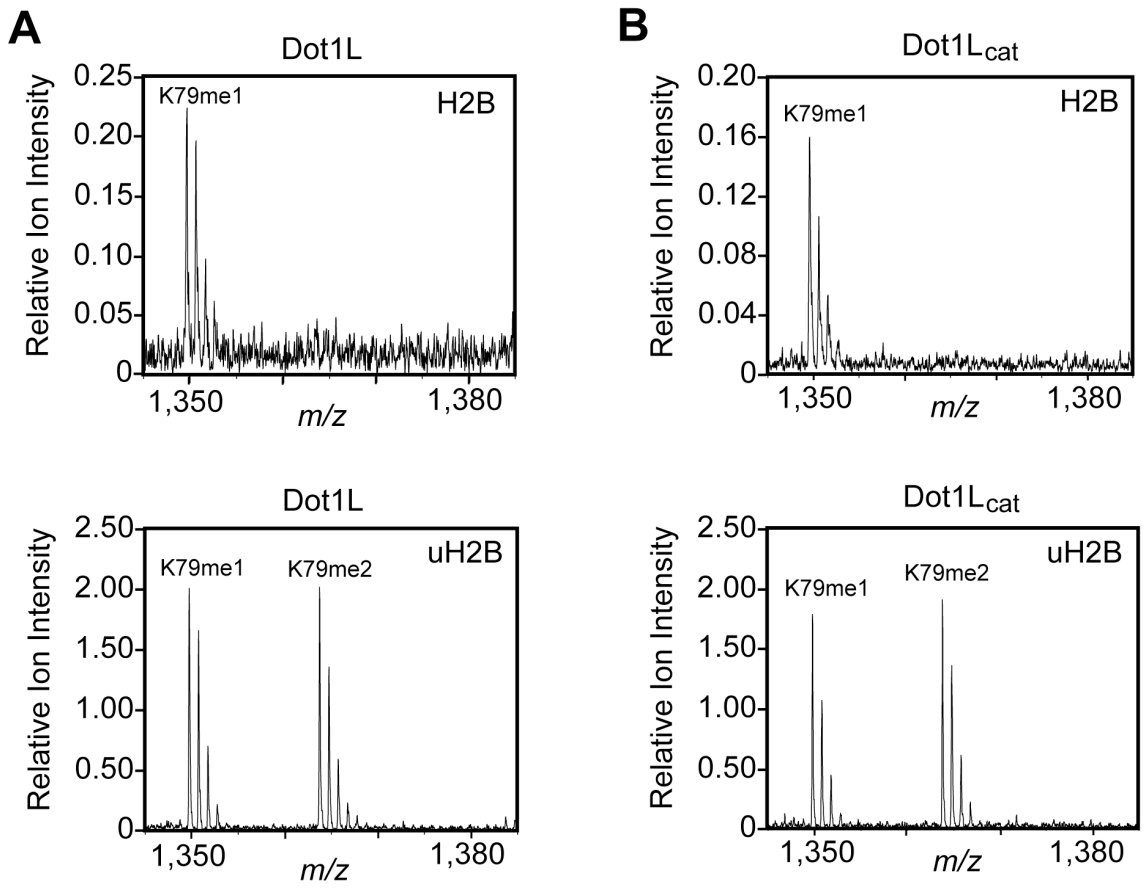
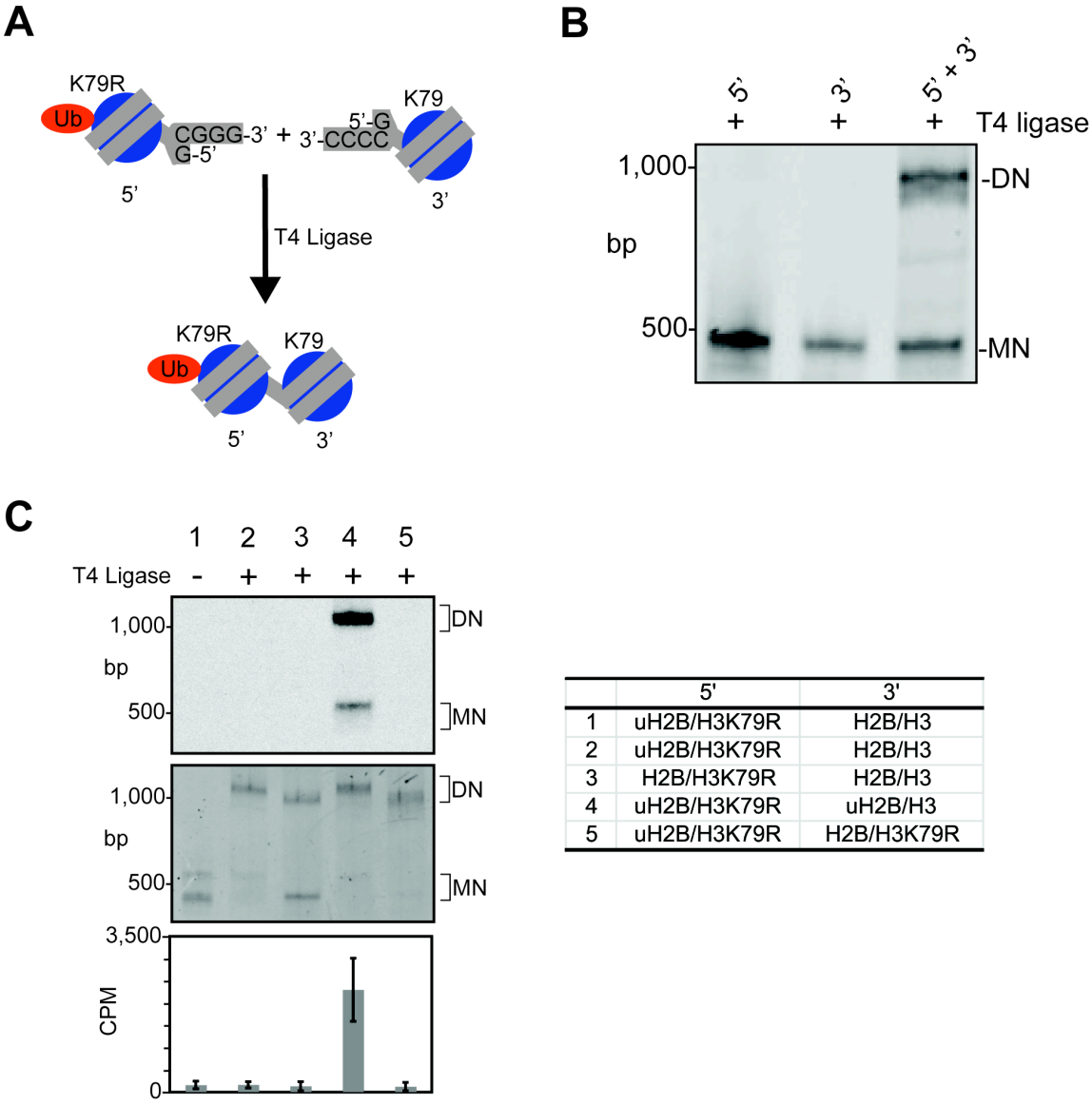



**Figure 3.8. Characterization of degree of methylation using mass spectrometry.** In gel trypsin digestion of H3 from methyltransferase assays using A) Dot1L or B) $Dot1L_{cat}$ followed by MALDI-mass spectrometry. Ion intensities of methylated peptides (residues 73-83) are scaled relative to an internal standard. Assays were performed using unmodified nucleosomes (top panels) and nucleosomes containing uH2B (bottom panels).

consistent with mass spectrometry analysis of H3K79 methylation in human cell lines (Garcia et al., 2007).

### 3.3. uH2B-H3K79me crosstalk is strictly intranucleosomal

We have demonstrated that uH2B is required for effective methylation of mononucleosomes by Dot1L. However, *in vivo* levels of H3K79 methylation vastly exceed those of H2B ubiquitylation (Garcia et al., 2007; West and Bonner, 1980). This raises the question: can uH2B enhance Dot1L-mediated methylation in an internucleosomal fashion? To address this possibility, a strategy for the creation of asymmetric dinucleosomes was adopted (Zheng and Hayes, 2003). Dinucleosomes with asymmetric incorporation of uH2B and H3K79R were produced by ligating uniquely assembled 5' and 3' mononucleosomes with complementary DNA overhangs (Figure 3.9A). The use of non-palindromic complementary overhangs, resulted in only the desired heterodinucleosome products being formed (Figure 3.9B). Unmodified, wild-type nucleosomes ligated to nucleosomes containing uH2B and H3K79R showed no increased Dot1L-mediated methylation when compared to unmodified, wild-type nucleosomes ligated to nucleosomes bearing H3K79R alone (Figure 3.9C, lanes 2 and 3). However, Dot1L was capable of methylating dinucleosomes, as evidenced by methylation of nucleosomes containing uH2B ligated to nucleosomes containing uH2B and H3K79R (Figure 3.9C, lane 4). It could be demonstrated that no significant histone shuffling was occurring during these assays because an unligated mixture of nucleosomes containing both uH2B and H3K79R and unmodified, wild-type nucleosomes showed no methylation above that observed with dinucleosomes containing only H3K79R (Figure 3.9C, lanes 1 and 5). These results strongly suggest that efficient




| | 5' | 3' |
|---|---|---|
| 1 | uH2B/H3K79R | H2B/H3 |
| 2 | uH2B/H3K79R | H2B/H3 |
| 3 | H2B/H3K79R | H2B/H3 |
| 4 | uH2B/H3K79R | uH2B/H3 |
| 5 | uH2B/H3K79R | H2B/H3K79R |

**Figure 3.9. Methyltransferase assays on dinucleosomes using Dot1L.** A) 5' and 3' mononucleosomes (MN) were reconstituted separately using DNA containing complementary overhangs. Assembly of 5' and 3' nucleosomes with different histones, followed by ligation, resulted in asymmetric dinucleosome (DN) formation. B) Ligations of unmodified nucleosomes reconstituted with indicated DNA fragments were separated on a 5% native gel and stained with ethidium bromide. C) $^{3}$H SAM methyltransferase assays performed with Dot1L and dinucleosomes were separated on a 5% native gel and stained with ethidium bromide (middle panel) prior to $^{3}$H methyl detection by fluorography (top panel). Liquid scintillation counting was used to quantify $^{3}$H methyl incorporation (bottom panel). Nucleosomes are indicated in table at right. Error bars indicate s.e.m. (n = 8).

methylation of H3K79 requires the presence of uH2B in the same nucleosome. Therefore, we propose that the majority of H3K79 methylated nucleosomes *in vivo*, are likely to have at one time carried uH2B, which has since been removed by deubiquitylating enzymes or histone replacement.

### 3.4. Dot1L-nucleosome interaction is independent of ubiquitylation

The simplest explanation that accounts for the intranucleosomal stimulatory effect of uH2B on Dot1L-mediated methylation is that uH2B recruits Dot1L to nucleosomes. However, the presence of a large excess of free ubiquitin had only a modest effect on the extent of Dot1L-mediated H3K79 methylation in uH2B containing nucleosomes (Figure 3.10A). The catalytic domain of Dot1L has previously been demonstrated to bind to unmodified nucleosomes in electrophoretic mobility shift assays (Min et al., 2003). We observed no significant difference in the recruitment of $Dot1L_{cat}$ to mononucleosomes as a function of ubiquitylation, based on gel-shift analysis, and only a modest difference in gel-shift competition assays, over a broad range of concentrations (Figure 3.10B and C). This establishes that $Dot1L_{cat}$ is able to bind to both ubiquitylated and unmodified nucleosomes during methyltransferase assays, even though efficient methyltransferase activity is only observed on ubiquitylated nucleosomes. Thus, it is likely that uH2B, possibly through an allosteric mechanism, allows Dot1L to bind nucleosomes in a catalytically competent manner. Additional investigations, perhaps involving structural analysis, will be necessary to shed further light on this phenomenon.

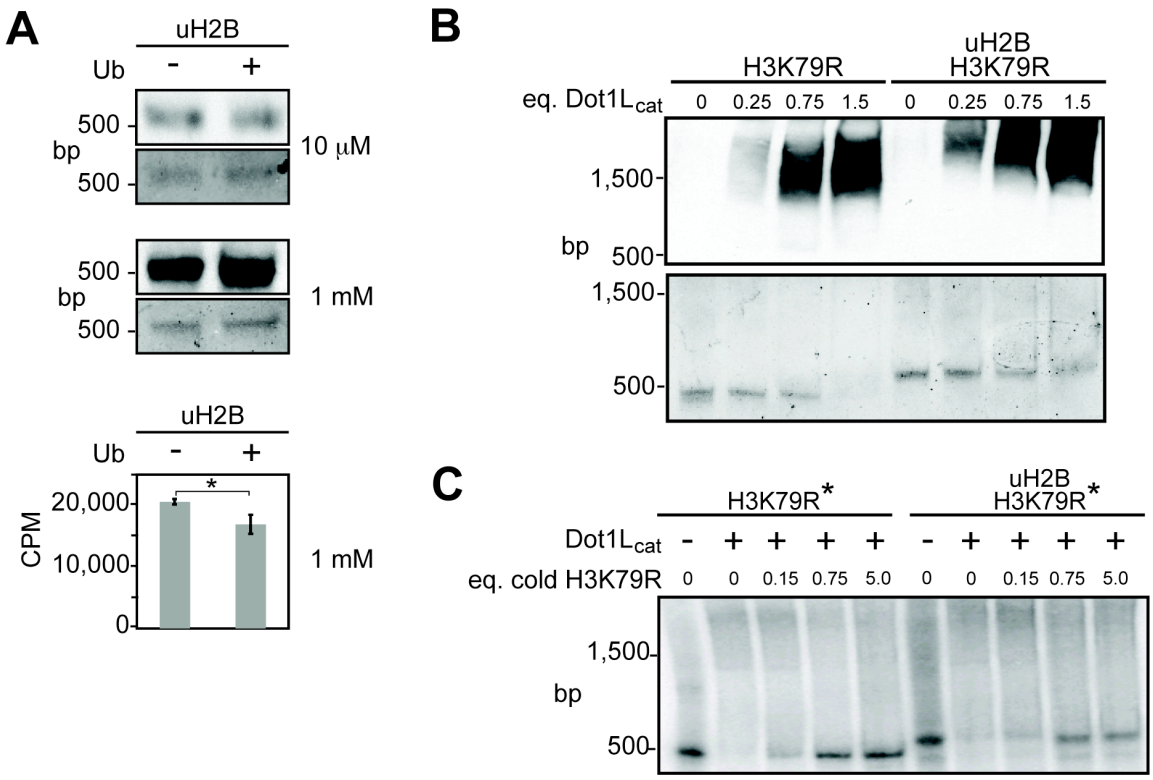



**Figure 3.10. Ubiquitin competition and electrophoretic mobility shift assays.** A) $^{3}$H SAM methyltransferase assay performed with mononucleosomes and Dot1L. 10 μM and 1 mM ubiquitin were added as indicated. Assay samples were separated on a 5% gel and stained with ethidium bromide (bottom panel of each pair) prior to fluorography, (top panel). Liquid scintillation counting was used to quantify methyltransferase activity in the presence and absence of 1 mM ubiquitin. Error bars indicate s.e.m. An asterisk indicates $p < 0.05$ using a two-tailed unpaired Student's t-test. (t-score = 2.28; 10 degrees of freedom; n = 6) B) Electrophoretic-mobility shift assay performed with Dot1L$_{cat}$ and indicated nucleosomes. Nucleosomes were incubated with increasing amounts of Dot1L$_{cat}$, separated by 5% native-PAGE, and stained with ethidium bromide (bottom panel) and a Western blot was performed against Dot1L$_{cat}$ using an α-GST antibody (top panel). C) Indicated $^{32}$P-labelled nucleosomes (an asterisk denotes radiolabelling) were combined with 5 equivalents of Dot1L$_{cat}$ and increasing quantities of cold H3K79R nucleosomes to compete the gel-shift of radiolabelled nucleosomes. Samples were separated on a 5% native gel and analyzed by phosphorimaging. H3K79R was used in these assays to eliminate potential affinity differences introduced by disproportionate methylation. eq. = molar equivalent

### 3.5. Summary

In this chapter, semisynthetic uH2B was reconstituted into octamers and nucleosomes with recombinant histones and 147 bp of a strong nucleosomal positioning sequence. At the level of the nucleosome, uH2B directly stimulates Dot1L-mediated methylation of H3K79. Although this does not rule out a role for uH2B in higher order chromatin structure (Figure 3.1, model 1), another mechanism must be responsible for the observations in this chapter. Moreover, while other factors may further enhance Dot1L activity (Figure 3.1, model 3), they are not necessary for uH2B-mediated stimulation of Dot1L. This stimulation is facilitated through the catalytic domain of Dot1L, leading to increases in both mono- and dimethylation of H3K79. Asymmetric nucleosome ligations demonstrated that the stimulatory role of uH2B is purely intranucleosomal, as the effect is not transferred to an adjacent nucleosome. Finally, two lines of evidence invalidate an uH2B-mediated, Dot1L recruitment model (Figure 3.1, model 2): 1) ubiquitin added in *trans* fails to diminish Dot1L activity on uH2B containing nucleosomes, and 2) $Dot1L_{cat}$ binds nucleosomes independently of the ubiquitylation state of H2B. Rather, it seems more likely that ubiquitylation of H2B may lead to an allosteric change in the nucleosome or Dot1L, either increasing the accessibility of H3K79 to Dot1L or the intrinsic catalytic activity of Dot1L itself (Figure 3.11).

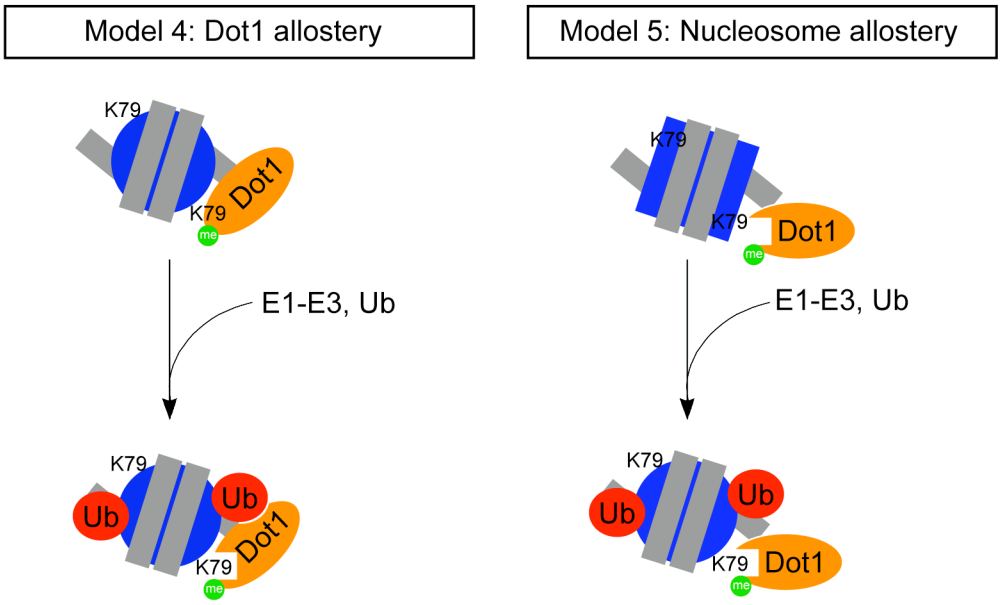



**Figure 3.11. Updated models of uH2B-mediated stimulation of Dot1.** Dot1 interacts with the nucleosome regardless of its ubiquitylation state. Model 4: Following H2B ubiquitylation, Dot1 is allosterically activated. Model 5: H2B ubiquitylation causes a conformational change in the nucleosome, which allows Dot1 access to H3K79.

## Chapter 4: Highly optimized semisynthesis of u(G76A)H2B[3]

In chapter 2, the semisynthesis of native uH2B was described using two orthogonal EPL reactions, ensuring chemical homogeneity and bypassing the cellular ubiquitylation machinery. Due to the absence of cysteines in uH2B, maintaining the native sequence necessitates the combination of three polypeptides using two traceless protein ligation strategies: one employing a photolytically removable ligation auxiliary and the other, a chemical desulfurization (Figure 2.11). In chapter 3, semisynthetic uH2B was used to demonstrate a direct stimulation of Dot1L-mediated intranucleosomal methylation of H3K79. Surprisingly, the stimulatory effect of uH2B is not a result of recruitment of Dot1L to the nucleosomal surface, as Dot1L is able to bind to nucleosomes in the absence of ubiquitylation, suggesting that uH2B may stimulate Dot1L activity through allosteric regulation of Dot1L or the nucleosome itself. However, detailed biochemical studies were precluded by limitations in scalability of the native semisynthetic strategy. In this chapter, a highly optimized semisynthesis of uH2B bearing the single G76A point mutation at the C-terminus of ubiquitin is described, allowing the rapid preparation of tens of milligrams of ubiquitylated protein. Nucleosomes containing u(G76A)H2B were indistinguishable by Dot1L from those containing uH2B, permitting both a comprehensive kinetic analysis of the role of uH2B in H3K79 methylation and the first structure activity relationship investigation of this complex system, which will be presented in chapter 5.

[3] The work in this chapter was performed as part of a close collaboration with Dr. Maja Köhn during a sabbatical in the Laboratory of Synthetic Protein Chemistry at the Rockefeller University.

### 4.1. Semisynthesis of u(G76A)H2B

We sought to design a new uH2B semisynthesis, optimizing for efficiency and scalability, while minimizing changes to the native sequence. Central to the inefficiency and low yield of the uH2B semisynthesis is the use of a photolytically removable ligation auxiliary (Figure 4.1). Firstly, the preparation of the auxiliary requires a complex eight-step solution phase synthesis (Figure 2.3), ultimately limiting the quantity that can be incorporated into a peptide to be used in EPL (Chatterjee et al., 2007; Marinzi et al., 2004; Pellois and Muir, 2005). Furthermore, the auxiliary-mediated ligation reaction requires a minimum of five days to reach 60% completion, due to the sterically hindered ligation onto a disubstituted amine (Chatterjee et al., 2007; McGinty et al., 2008). Finally, photolytic removal of the ligation auxiliary, concomitant with the cysteine-protecting group, is poorly scalable and cannot be performed in parallel without multiple irradiation sources. We reasoned that a traditional cysteine-mediated ligation could be substituted for the auxiliary-mediated ligation to attach ubiquitin to an H2B C-terminal peptide, thus alleviating the costly constraints imposed by the ligation auxiliary (Figure 4.1, blue). Further replacement of the photolytically removable *S*-(*o*-nitrobenzyl) cysteine-protecting group in **21** with a chemically labile group would eliminate the need for an irradiation step entirely (Figure 4.1, green). Following orthogonal ligation to the N-terminal sequence of H2B, the ligation site cysteines could be reduced to alanines, leaving a G76A substitution at the C-terminus of ubiquitin as the single sacrifice of the optimized semisynthetic strategy (Figure 4.1, yellow).

To this end, peptide **27** corresponding to residues 117-125 of H2B, bearing an A117C substitution, was synthesized (Figure 4.2). Orthogonal side-chain protection of K120 allowed the selective installment of a cysteine through an isopeptide bond with the ε-$NH_2$ of K120. This

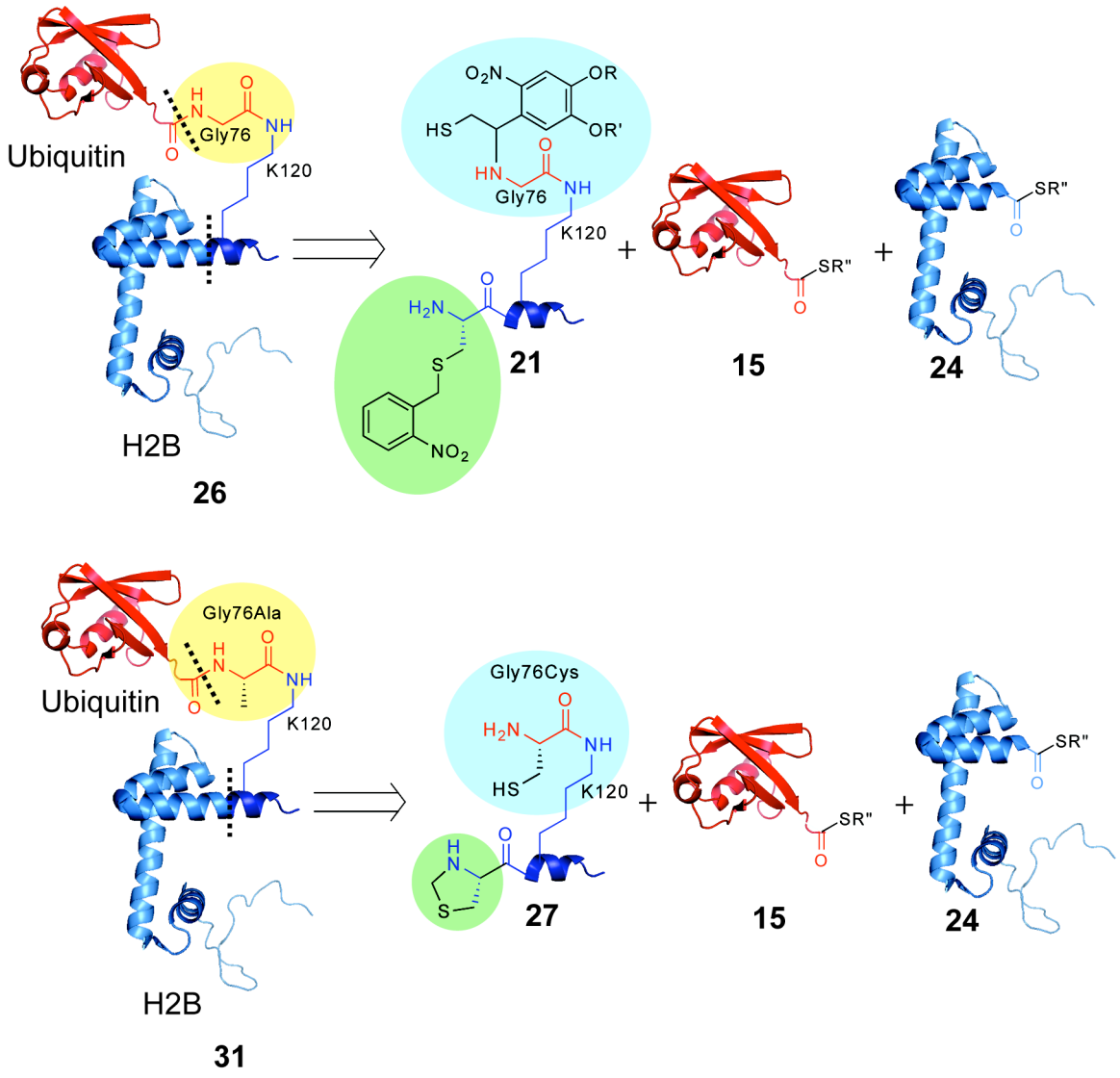



**Figure 4.1. New semisynthetic strategy for generation of ubiquitylated H2B.** Retrosynthetic comparison of uH2B (top) and u(G76A)H2B (bottom) syntheses. Both were generated via a 3-piece ligation strategy with the following polypeptides: synthetic peptide containing residues 117-125 of H2B and bearing an A117C mutation, H2B-C, **21** and **27**; recombinant ubiquitin(1-75)-α-thioester **15**; and recombinant H2B(1-116)-α-thioester **24**. For the semisynthesis of u(G76A)H2B, the ligation auxiliary was replaced with a cysteine (blue) and the photolytically removable cysteine-protecting group was replaced with a thiazolidine (green). The resultant ubiquitylated proteins differ only at position 76 of ubiquitin (yellow). Dashed lines indicate junctions formed by EPL reactions. R = $CH_2CH_2CH_2C(O)NHCH_3$; R' = $CH_3$; R'' = $CH_2CH_2SO_3H$.

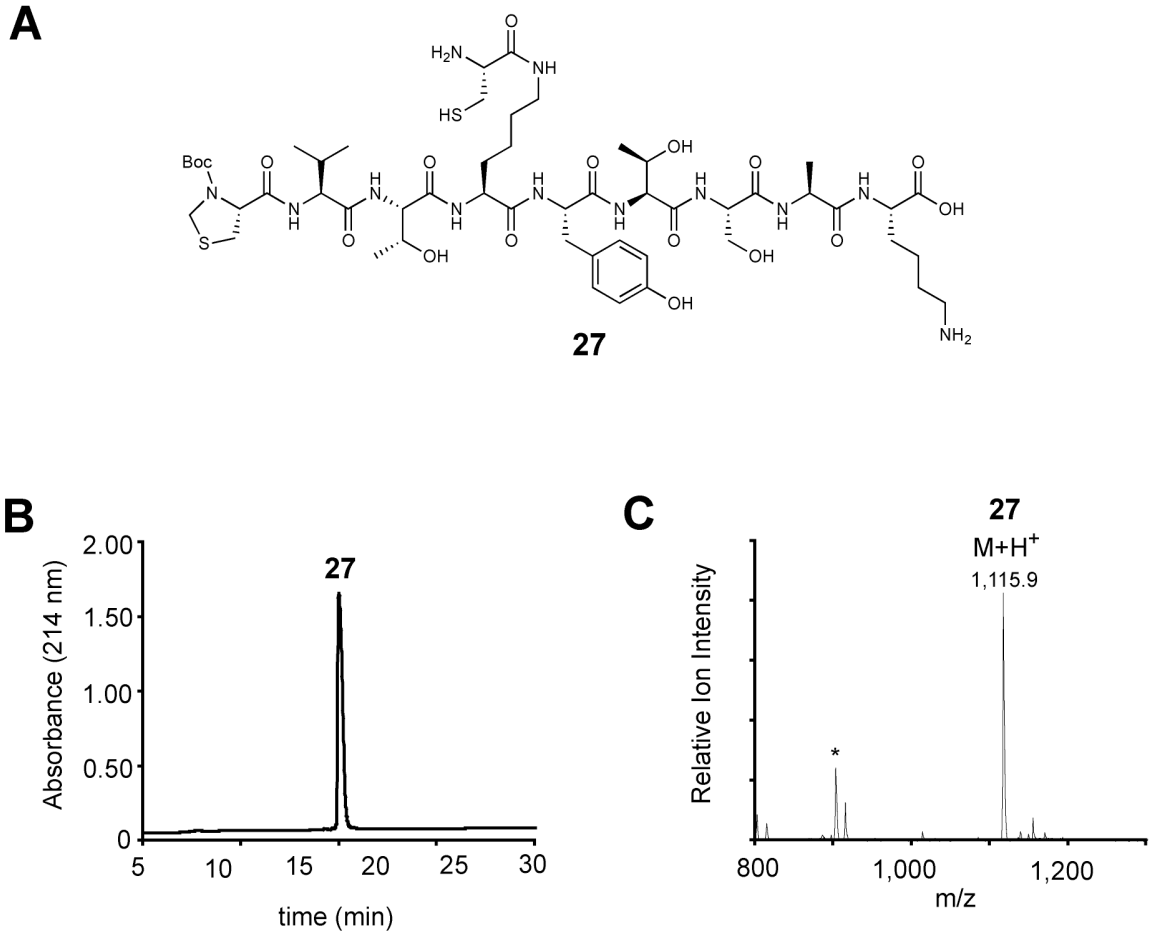



**Figure 4.2. Characterization of peptide 27.** A) Structure of peptide **27**. B) RP-HPLC chromatogram and C) ESI-MS spectra of purified peptide **27**. [$(M+H)^+$ observed = 1,115.9 Da. $(M+H)^+$ expected = 1,115.3 Da.]. An asterisk represents MS fragmentation of peptide **27**.

cysteine facilitates ligation to ubiquitin and following desulfurization results in the G76A mutation. Transient protection of the N-terminal cysteine in **27** as a thiazolidine restricts ubiquitin ligation to the desired site.

In the first step of the synthesis, an excess of peptide **27** was ligated to recombinant ubiquitin(1-75)-α-thioester, **15** (Figure 2.6), generated by thiolysis of the corresponding intein fusion protein (Figure 4.3, step a). Thioester **15** was quantitatively converted into ubiquitylated peptide **28** within two hours (Figure 4.4A). Treatment with methoxylamine *in situ* at pH 5 for 12 h led to complete conversion of the thiazolidine to a cysteine, affording branched protein **29** (Figure 4.3, step a and 4.4B). Notably, purification was unnecessary between the first two synthetic steps, further increasing the overall efficiency of the semisynthetic strategy and limiting losses. Protein **29** was then ligated to an excess of H2B(1-116)-α-thioester, **24** (Figure 2.12), yielding full-length ubiquitylated H2B bearing a cysteine at each of the ligation junctions, **30** (Figure 4.3, step c and 4.4C). Raney nickel (Yan and Dawson, 2001) or radical-initiated (Wan and Danishefsky, 2007) desulfurization smoothly converted both cysteines in **30** to alanines, leaving the native A117 in H2B and a G76A mutation in ubiquitin, thus generating u(G76A)H2B, **31** (Figure 4.4, step d and 4.4D). With this revised semisynthetic route, tens of milligrams of ubiquitylated protein were routinely generated, without the need for specialized instrumentation.

### 4.2. Validation of u(G76A)H2B as a surrogate for uH2B

We envisioned that a G76A mutation in ubiquitin would be tolerated by any biochemical processes that recognize elements of ubiquitin and its target protein without specifically engaging the isopeptide bond between them. In these cases, the ubiquitylated protein with the

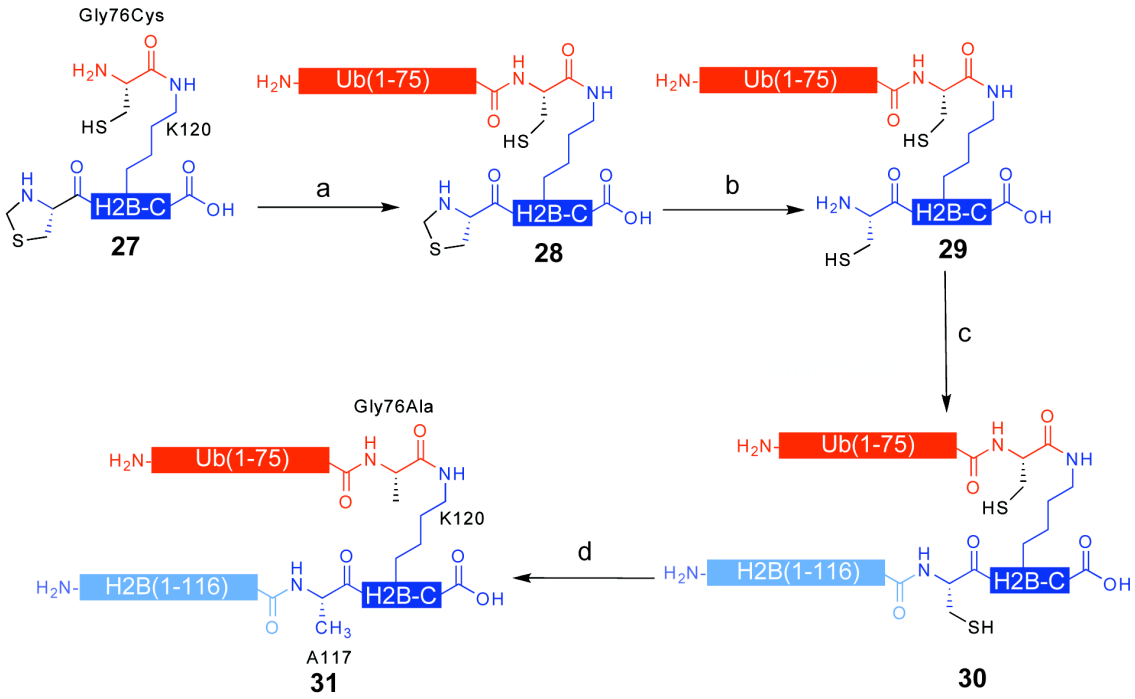



**Figure 4.3. Semisynthesis of u(G76A)H2B.** Semisynthetic scheme for the generation of u(G76A)H2B. (a) EPL was used to ligate peptide **27** to ubiquitin(1-75)-α-thioester, **15**, forming branched protein **28**; (b) Ligation product **28** was treated with methoxylamine at pH 5, affording **29**. (c) Ligation of protein **29** to H2B(1-116)-α-thioester, **24**, forming uH2BA117C/G76A, **30**. (d) Raney nickel or radical-initiated desulfurization of protein **30**, forming u(G76A)H2B, **31**.

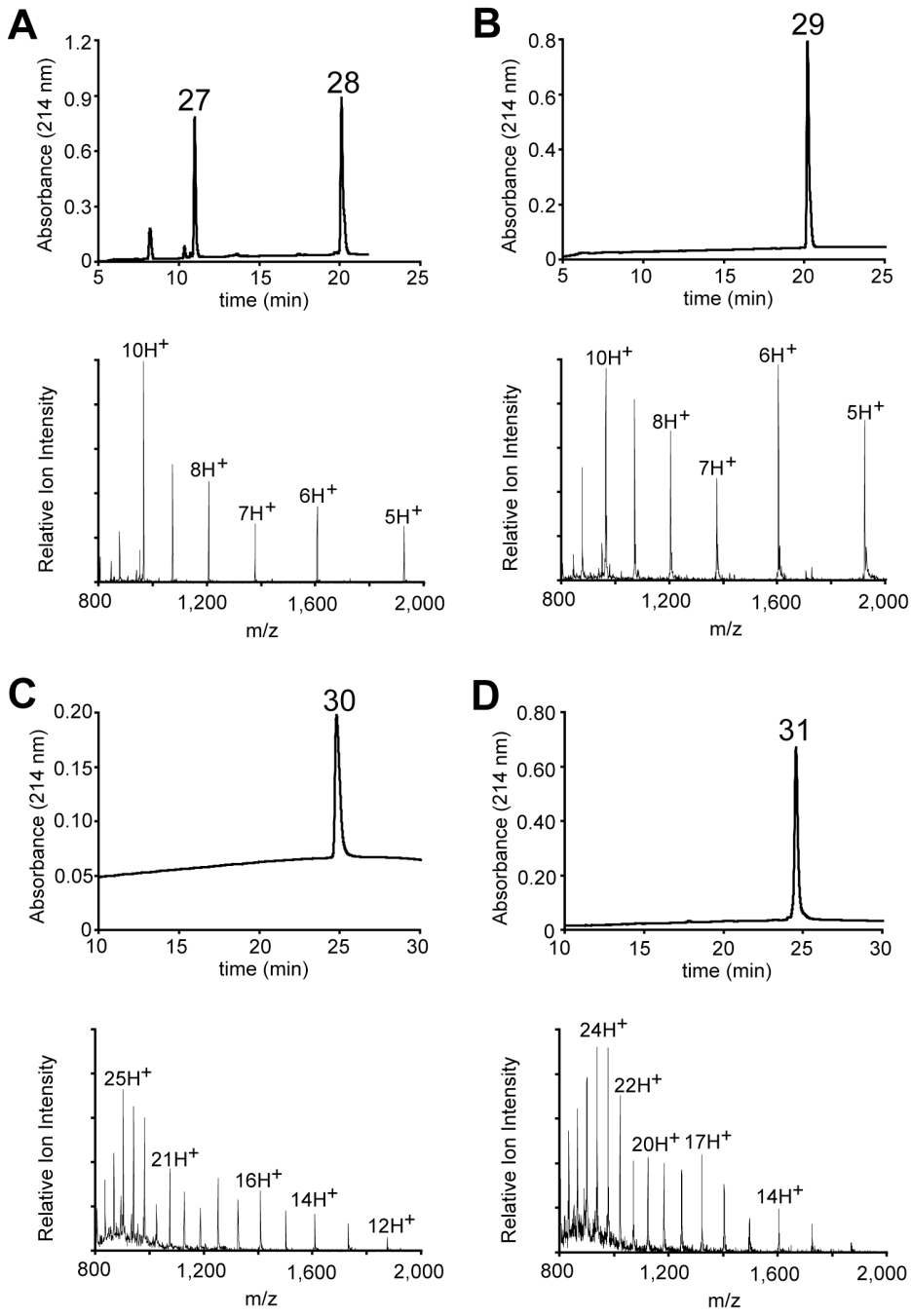



**Figure 4.4. Characterization of products in the semisynthesis of u(G76A)H2B.** RP-HPLC chromatograms (top panels) and ESI-MS spectra (bottom panels) of the following intermediate products: A) ligation product, **28** [$(M+H)^+$ observed = 9,609 ± 2 Da (s.d.). $(M+H)^+$ expected = 9,606 Da.]; B) deprotected ligation product, **29** [$(M+H)^+$ observed = 9,593 ± 2 Da. $(M+H)^+$ expected = 9,594 Da]; C) ligation product, **30** [$(M+H)^+$ observed = 22,444 ± 5 Da. $(M+H)^+$ expected = 22,443 Da.]; and D) desulfurization product, u(G76A)H2B, **31** [$(M+H)^+$ observed = 22,380 ± 4 Da. $(M+H)^+$ expected = 22,379 Da.]. Charge states are labeled. Crude RP-HPLC spectrum is shown for protein in panel A.

G76A mutation would likely be indistinguishable from the native protein. Conversely, we reasoned that processes recognizing the isopeptide bond directly might be disrupted by the mutation. Indeed, the ubiquitin G76A mutant is efficiently mobilized and transferred to target proteins *in vivo*, but the conjugated product is resistant to many deubiquitylating enzymes (DUBs) (Hodgins et al., 1992; Pickart et al., 1994). Surprisingly, u(G76A)H2B was recognized and deubiquitylated by ubiquitin C-terminal hydrolase, UCH-L3, to a similar extent as the native control (Figure 4.5A). Cleaved ubiquitin and ubiquitin G76A were identified by liquid chromatography-mass spectrometry (LC-MS), verifying hydrolysis of the isopeptide bond (Figure 4.5B and C). Furthermore, recognition of uH2B by a linkage-specific anti-uH2B antibody was only slightly decreased by introduction of the G76A mutation (Figure 4.6A). Key to our goals was the effect of the u(G76A)H2B mutation on the stimulation of Dot1L. In order to test this, native uH2B and u(G76A)H2B were reconstituted into nucleosomes with recombinant histone proteins (Figure 3.2) and 147 bp of the Widom 601 positioning sequence (Lowary and Widom, 1998) and submitted to radioactive methyltransferase assays with Dot1L. u(G76A)H2B stimulated methyltransferase activity to levels comparable to uH2B (Figure 4.6B), validating this mutant as an equally acceptable surrogate for the native sequence in biochemical analysis.

### 4.3. Summary

In summary, we have developed a robust, scalable semisynthesis of ubiquitylated proteins to overcome shortcomings of the strategy presented in chapter 2. By replacing an auxiliary-mediated ligation with a traditional cysteine-mediated ligation, we efficiently generated u(G76A)H2B on a scale of multiple tens of milligrams – an increase in excess of ten-fold over the previous methodology. This streamlined semisynthetic strategy will also allow the facile and

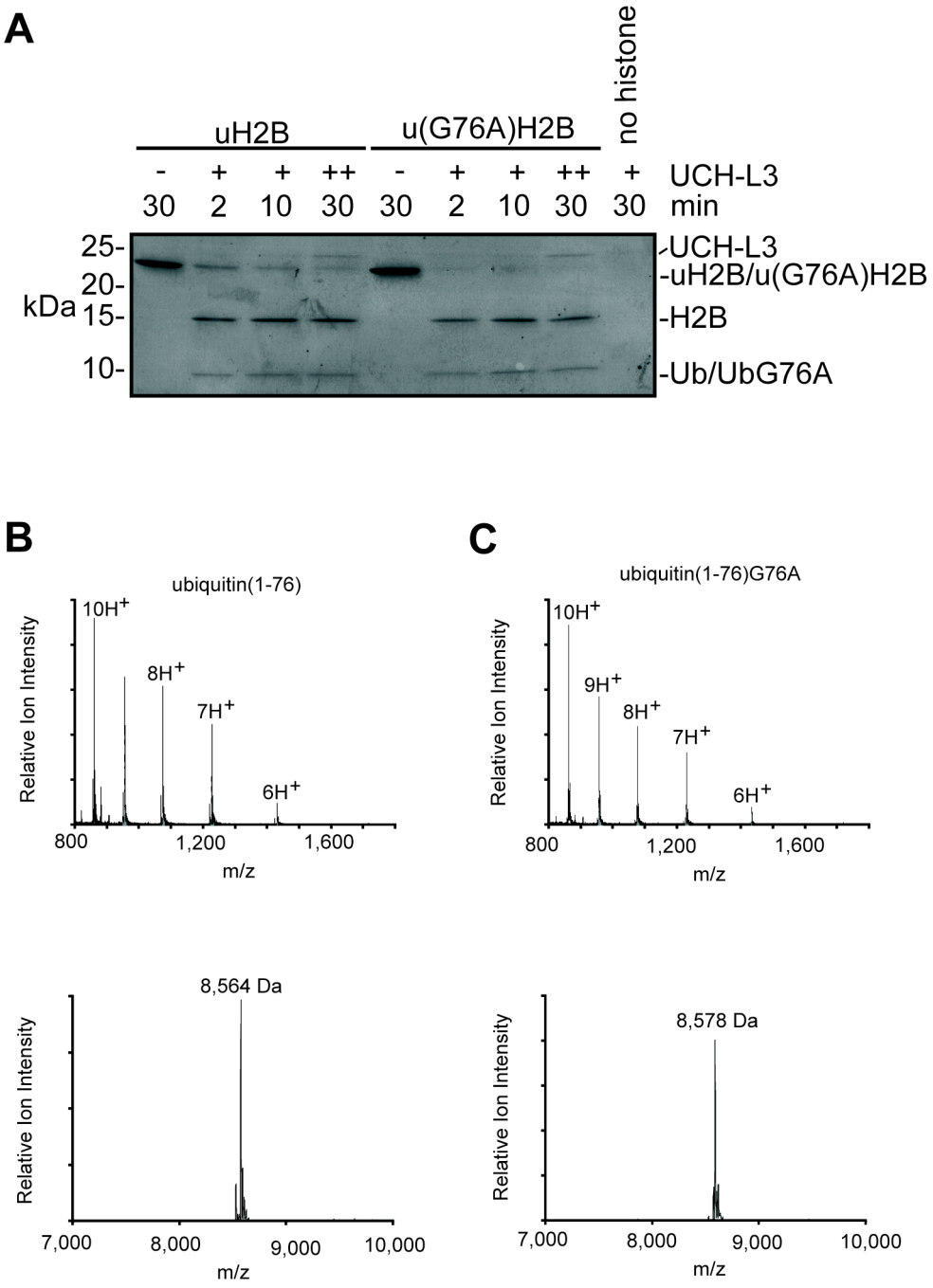



**Figure 4.5. Ubiquitin hydrolysis assay of uH2B and u(G76A)H2B.** A) UCH-L3-mediated hydrolysis of uH2B and u(G76A)H2B. Coomassie stained gel of assay samples quenched after indicated times. + indicates addition of UCH-L3 at 0 min; ++ indicates addition of UCH-L3 at 0 and 10 min. B) and C) LC-MS spectrum of ubiquitin hydrolyzed from B) uH2B and C) u(G76A)H2B by UCH-L3 (top panels) and the resultant deconvoluted masses (bottom panels). Expected masses for ubiquitin and ubiquitinG76A are 8,566 and 8,580 Da, respectively. Charge states are labeled.

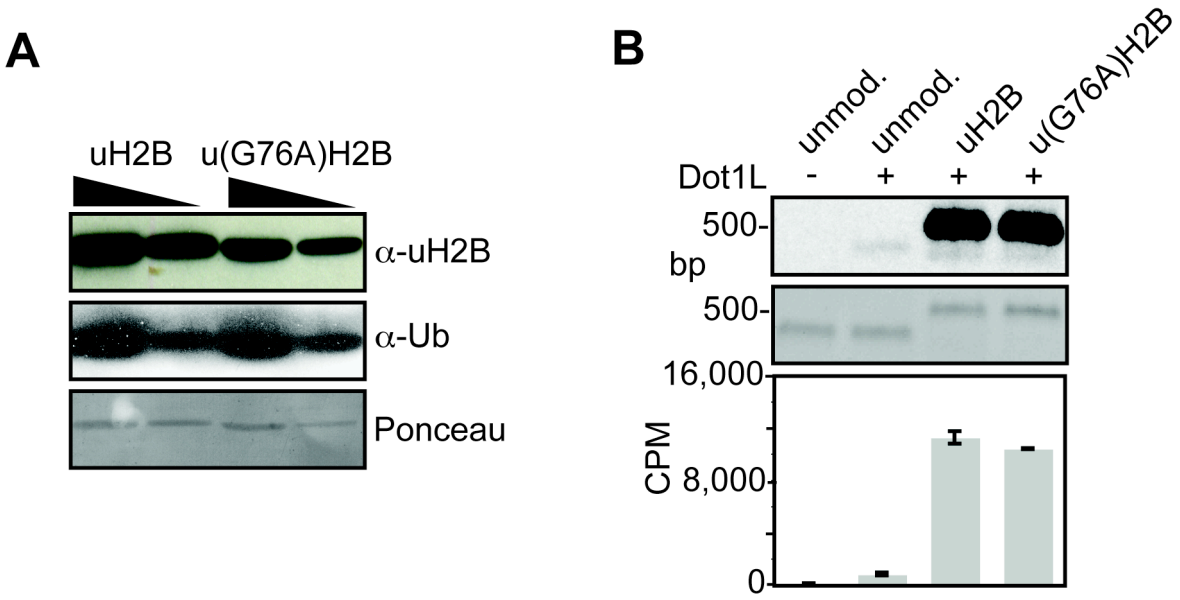



**Figure 4.6. Validation of u(G76A)H2B as an uH2B surrogate.** A) Western blot of uH2B and u(G76A)H2B with linkage specific α-uH2B antibody (top panel). Western blot with α-ubiquitin antibody (middle panel) and ponceau stain (bottom panel) represent loading controls. B) Dot1L methyltransferase assay on uH2B and u(G76A)H2B containing nucleosomes. Nucleosomes methylated with $^{3}$H SAM were separated on a 5% native gel and stained with ethidium bromide (middle panel) prior to probing for $^{3}$H methyl incorporation by fluorography (top panel). Quantification of methyltransferase activity was performed by liquid scintillation counting (bottom panel). Error bars represent one s.d. (n = 3).

rapid preparation of other ubiquitylated histones, and more broadly, other ubiquitylated proteins. The G76A mutation introduced into ubiquitin by the semisynthetic strategy may impose limitations in certain applications. Surprisingly, there was no observable effect of the mutation on ubiquitin hydrolysis by UCH-L3. This may be the exception rather than the rule and in some cases, the native ubiquitylated protein may be a more appropriate reagent for the biochemical characterization of DUBs. Most importantly, the stimulation of Dot1L by uH2B was unaffected by the G76A mutation, permitting the first systematic mechanistic investigation into the methylation process, which is presented in the following chapter.

## Chapter 5: Structure activity relationship analysis of Dot1L stimulation

The ability to rapidly prepare u(G76A)H2B on a scale of tens of milligrams (presented in chapter 4) removes limitations on detailed biochemical characterization of the roles of H2B ubiquitylation in chromatin-templated processes. Because u(G76A)H2B is indistinguishable from native uH2B by Dot1L, u(G76A)H2B can act as a surrogate in mechanistic analysis of the resultant *trans* histone pathway. In this chapter, mechanistic insights from comprehensive kinetic and structure activity relationship analyses will be discussed. Additionally, early attempts to crystallize nucleosomes containing u(G76A)H2B will be presented.

### 5.1. Kinetic analysis of Dot1L-mediated monomethylation of nucleosomes

Disruption of H2B ubiquitylation in organisms from yeast to humans decreases levels of higher order methylation of H3K79 without altering levels of monomethylation (Kim et al., 2005; Shahbazian et al., 2005). However, *in vitro* ubiquitylation of H2B drastically increases the level of Dot1L-mediated monomethylation and dimethylation in nucleosomes (Figure 3.8). This pattern is observed with full-length Dot1L as well as the catalytic domain alone. We predicted that analysis of steady-state kinetics would provide insight into the effect of uH2B on the ability of Dot1L to perform the initial methylation of H3K79. Accordingly, nucleosomes containing H2B or u(G76A)H2B were prepared with the 146 bp α-satellite palindromic sequence by stepwise dialysis followed by purification by preparative native-PAGE (Figure 5.1) (Dyer et al., 2004; Luger et al., 1997a). Initial rates of nucleosome methylation by $Dot1L_{cat}$ were obtained by measuring transfer of a radioactive methyl group from SAM, illustrating a vast rate enhancement due to ubiquitylation (Figure 5.2A). At early time points, only monomethylation of H3K79 was detected, allowing for the direct steady-state velocity analysis of the first methylation event

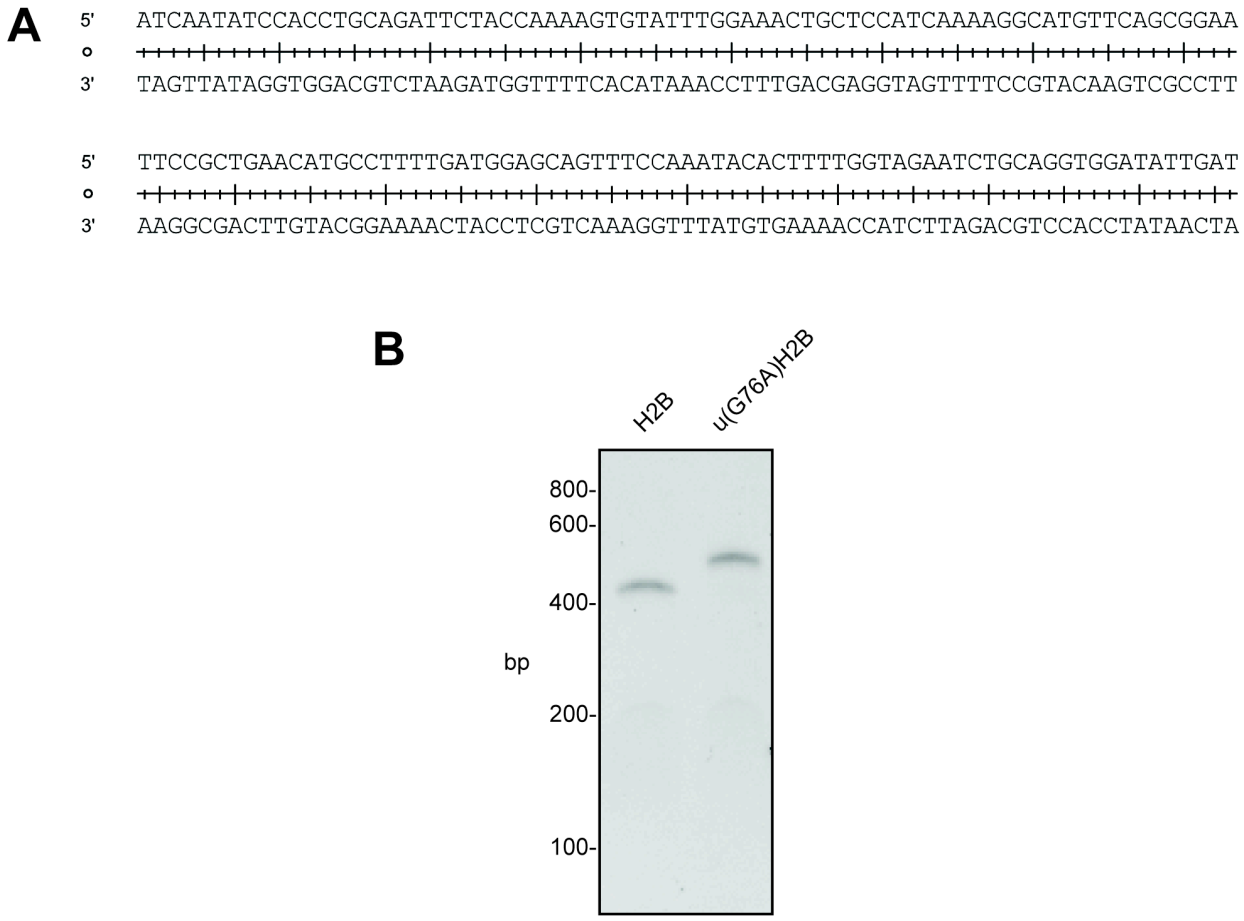



**Figure 5.1. Nucleosome formation with 146 bp palindromic α-satellite sequence.** A) 146 bp palindromic α-satellite sequence. B) Ethidium bromide stained 5% native gel of nucleosomes formed with this sequence and containing H2B or u(G76A)H2B.

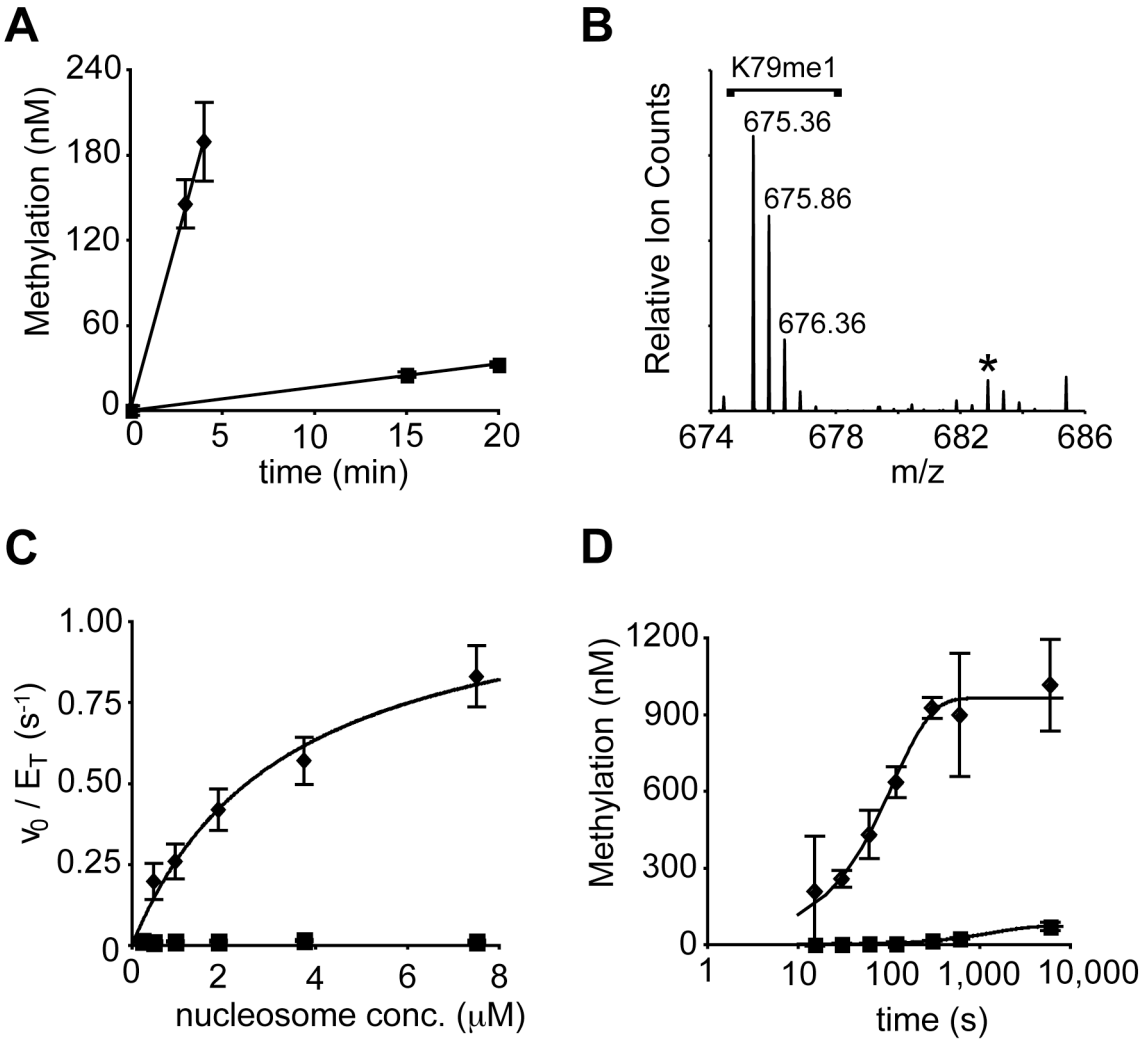



**Figure 5.2. Kinetic analysis of Dot1L-mediated monomethylation of nucleosomes.** A) Representative time course of $Dot1L_{cat}$-mediated methylation of 0.9 μM nucleosome bearing H2B (squares) or u(G76A)H2B (diamonds). Linear regression models of the data are shown. Error bars represent one s.d. (n = 2). B) ESI-MS chromatogram showing K79 monomethylation (K79me1) in the absence of dimethylation for u(G76A)H2B nucleosomes at 1.8 μM. $[M+2H]^{2+}$ charge states are labeled. An asterisk marks a contaminant that is not dimethylated K79. C) Steady-state kinetic analysis of $Dot1L_{cat}$ activity of nucleosomes bearing H2B (squares) or u(G76A)H2B (diamonds). Fit of u(G76A)H2B data to the Michaelis-Menten model is shown. Error bars represent one s. d. (n = 6). D) Kinetic analysis of methylation of nucleosomes bearing H2B (squares) or u(G76A)H2B (diamonds) by excess $Dot1L_{cat}$. X-axis is shown in $\log_{10}$ scale to emphasize early time points. Single exponential modeling of the data is shown. Error bars represent one s.d. (n = 3).

(Figure 5.2B). Under steady-state conditions (Figure 5.2C), Dot1L$_{cat}$ clearly followed Michaelis-Menten kinetics in the monomethylation of ubiquitylated nucleosomes ($K_m$ of 3.3 ± 0.6 μM and a $k_{cat}$ of 1.16 ± 0.09 $s^{-1}$). However, under all reasonable conditions tested, Dot1L$_{cat}$ failed to show any concentration dependence toward unmodified nucleosomes. This could result from an exorbitantly high $K_m$ or an immeasurably low $k_{cat}$. The latter could reflect an extremely slow chemical step and/or a decreased rate of product release. To probe this further, we performed Dot1L$_{cat}$ methyltransferase reactions under single turnover-type conditions (Figure 5.2D). Even at saturating levels, Dot1L$_{cat}$ was far more efficient in methylation of ubiquitylated nucleosomes, suggesting an increase in the velocity of the chemical step of the first methylation.

**5.2. Structure activity relationship analysis of the stimulation of Dot1L by uH2B**

A myriad of small chemical modifications to histones, such as acetylation, phosphorylation, and methylation, have evolved to control accessibility of genomic loci and to tune the recruitment and activity of effectors of genomic expression (Kouzarides, 2007; Taverna et al., 2007). Given these established mechanisms, why would nature select such a large modification in ubiquitin, when a smaller modification might suffice? To delve into this question, we performed a structure activity relationship analysis of the role of ubiquitin in Dot1L stimulation.

We first wondered whether the ubiquitin fold was essential to stimulate methyltransferase activity. Could the mere acylation of H2BK120 be sufficient? We synthesized a truncated uH2B (tr-uH2B) lacking the globular domain of ubiquitin (Figure 5.3A and B). First, peptide **32** was synthesized corresponding to residues 117-125 of H2BA117C, with residues 72-76 of ubiquitin attached to the ε-$NH_2$ of K120 (Figure 5.4). Similar to peptides **21** and **27**, orthogonal protection

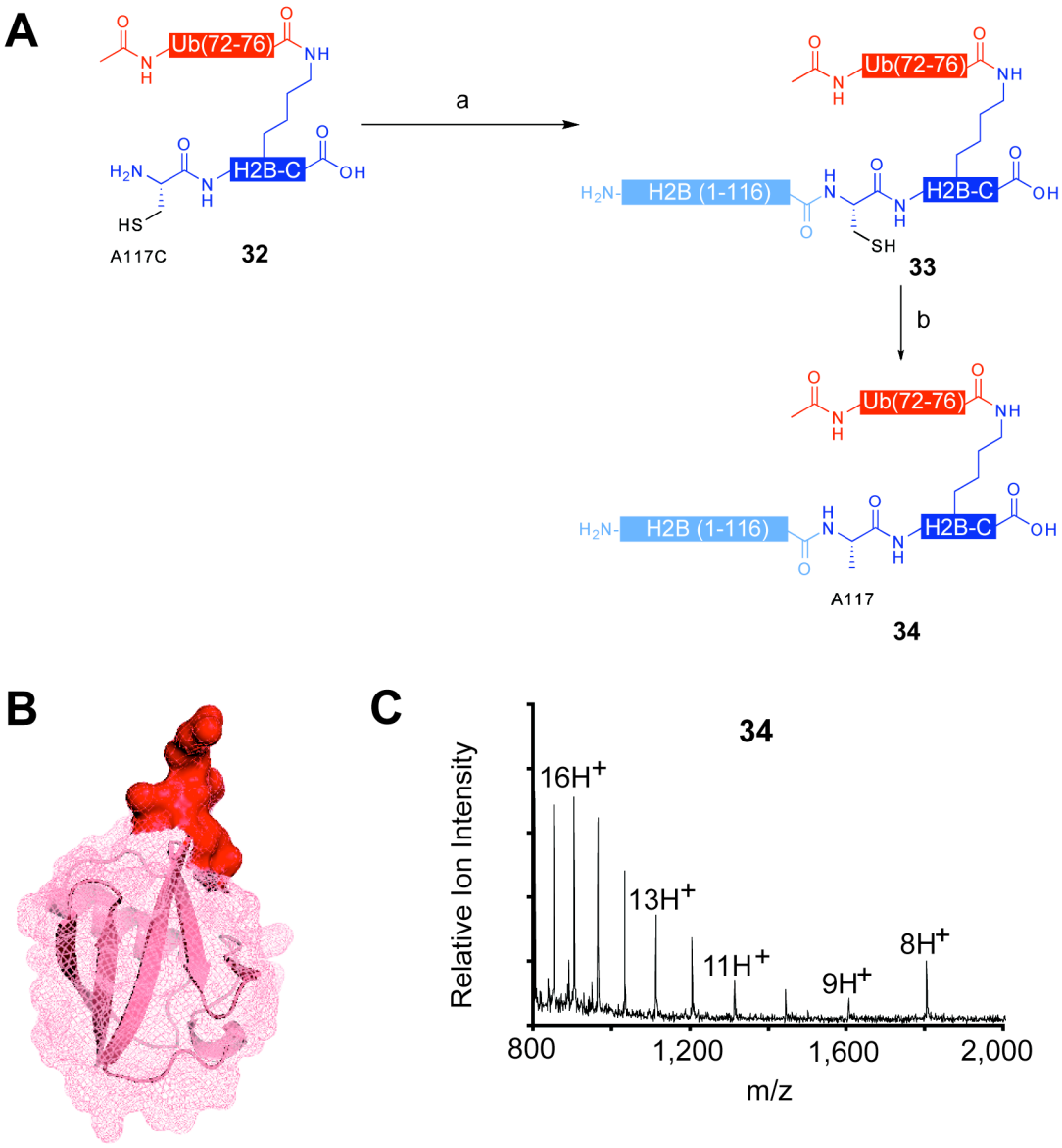



**Figure 5.3. Semisynthesis of tr-uH2B.** A) Schematic of tr-uH2B semisynthesis. (a) EPL was used to ligate peptide **32** to H2B(1-116)-α-thioester, **24**, to give tr-uH2BA117C, **33**. (b) Desulfurization of protein **33**, yielding tr-uH2B, **34**. B) Ubiquitin structure (1UBQ) (Vijay-Kumar et al., 1987) represented by superimposed ribbon and mesh diagrams. The surface of the five residues of ubiquitin included in tr-uH2B are shown in red. C) ESI-MS spectra of isolated tr-uH2B, **34** [$(M+H)^+$ observed = 14,400 ± 3 Da (s.d.). $(M+H)^+$ expected = 14,399 Da.]

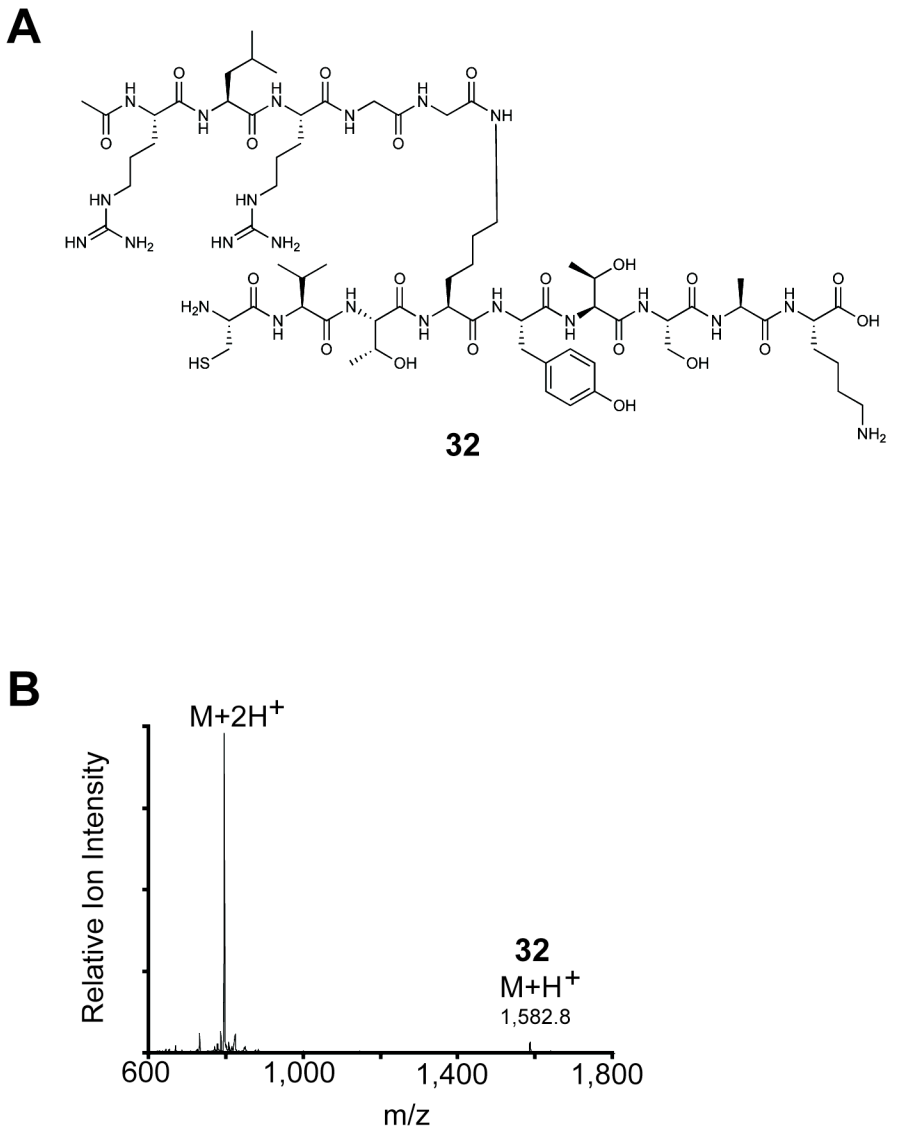



**Figure 5.4. Characterization of peptide 32.** A) Structure of peptide **32**. B) ESI-MS spectrum of peptide **32**. Charge states are labeled. [$(M+H)^+$ observed = 1,582.8 Da. $(M+H)^+$ expected = 1,582.8 Da.]

of K120 allowed the C-terminal residues of ubiquitin to be installed through an isopeptide bond with K120. The N-terminus of the ubiquitin branch was acetylated to prevent interference during EPL. Ligation of **32** to H2B(1-116)-α-thioester, **24**, yielded intermediate product **33** (Figure 5.3A, step a). The cysteine in protein **33** was desulfurized using Raney nickel, affording tr-uH2B, **34** (Figure 5.3A, step b and 5.3C). Nucleosomes assembled with 147 bp of Widom 601 sequence containing tr-uH2B were assayed against Dot1L. However, the presence of tr-uH2B did not lead to an increase in methyltransferase activity over that observed for unmodified nucleosomes (Figure 5.5A and 5.5B), suggesting that the ubiquitin fold is necessary for Dot1L stimulation.

We next asked whether a specific surface of ubiquitin was required, or whether a protein of similar size and shape could substitute. We replaced ubiquitin with Smt3, the yeast homolog of human small ubiquitin-like modifier 1 (SUMO1) (Figure 5.6A). Smt3 has a similar overall fold to that of ubiquitin (Figure 5.6B), but only a 17% sequence identity (Mossessova and Lima, 2000). The HA-Smt3(2-97)-α-thioester, **18** (Figure 2.10), was substituted for the ubiquitin thioester, **15**, in the semisynthetic strategy. Ligation of peptide **27** to protein **18** afforded branched protein **35** (Figure 5.6A, step a). Subsequent cysteine deprotection yielded intermediate product **36**, which was ligated to H2B(1-116)-a-thioester, **24**, to give sH2B(cys), **37** (Figure 5.6A, steps b and c, and 5.6C). However, no protein was recovered after incubation with Raney nickel. Therefore, the pre-desulfurization product, sH2B(cys), bearing an A117C mutation in H2B and a G98C mutation in Smt3, was incorporated into nucleosomes. Unlike nucleosomes containing the non-desulfurized uH2B(cys), **30**, sH2B(cys) nucleosomes failed to stimulate Dot1L to significant levels (Figure 5.5A and 5.5B). This result suggests that a surface of ubiquitin is recognized specifically.

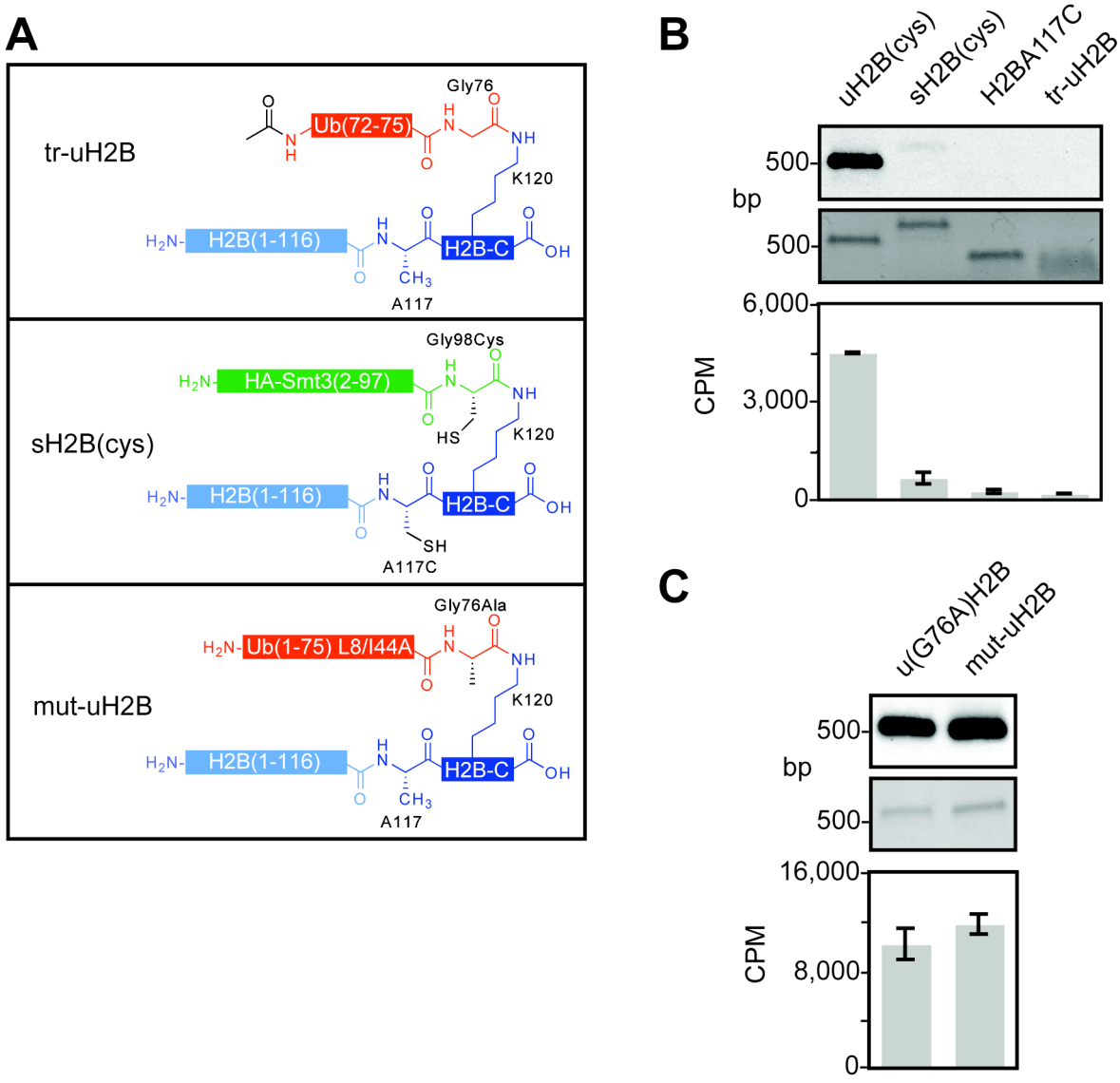



**Figure 5.5. Structure activity relationship analysis of uH2B.** A) Schematic of semisynthetic uH2B structural variants. Top panel: H2B bearing the terminal five residues of ubiquitin attached to K120 (tr-uH2B); middle panel: H2B modified with HA-Smt3 with A117C/G98C mutations (sH2B(cys)); bottom panel: u(G76A)H2B with L8A/I44A double mutation (mut-uH2B). B) Dot1L methyltransferase assay on uH2B structural variants. Nucleosomes containing uH2B(cys), **30**, sH2B(cys), H2BA117C, or tr-uH2B methylated with Dot1L and $^{3}$H SAM were separated on a 5% native gel and stained with ethidium bromide (middle panel) prior to probing for $^{3}$H methyl incorporation by fluorography (top panel). Quantification of methyltransferase activity was performed by liquid scintillation counting (bottom panel). C) Dot1L assay as in panel B comparing u(G76A)H2B and mut-uH2B. Error bars represent one s.d. (n = 3-4).

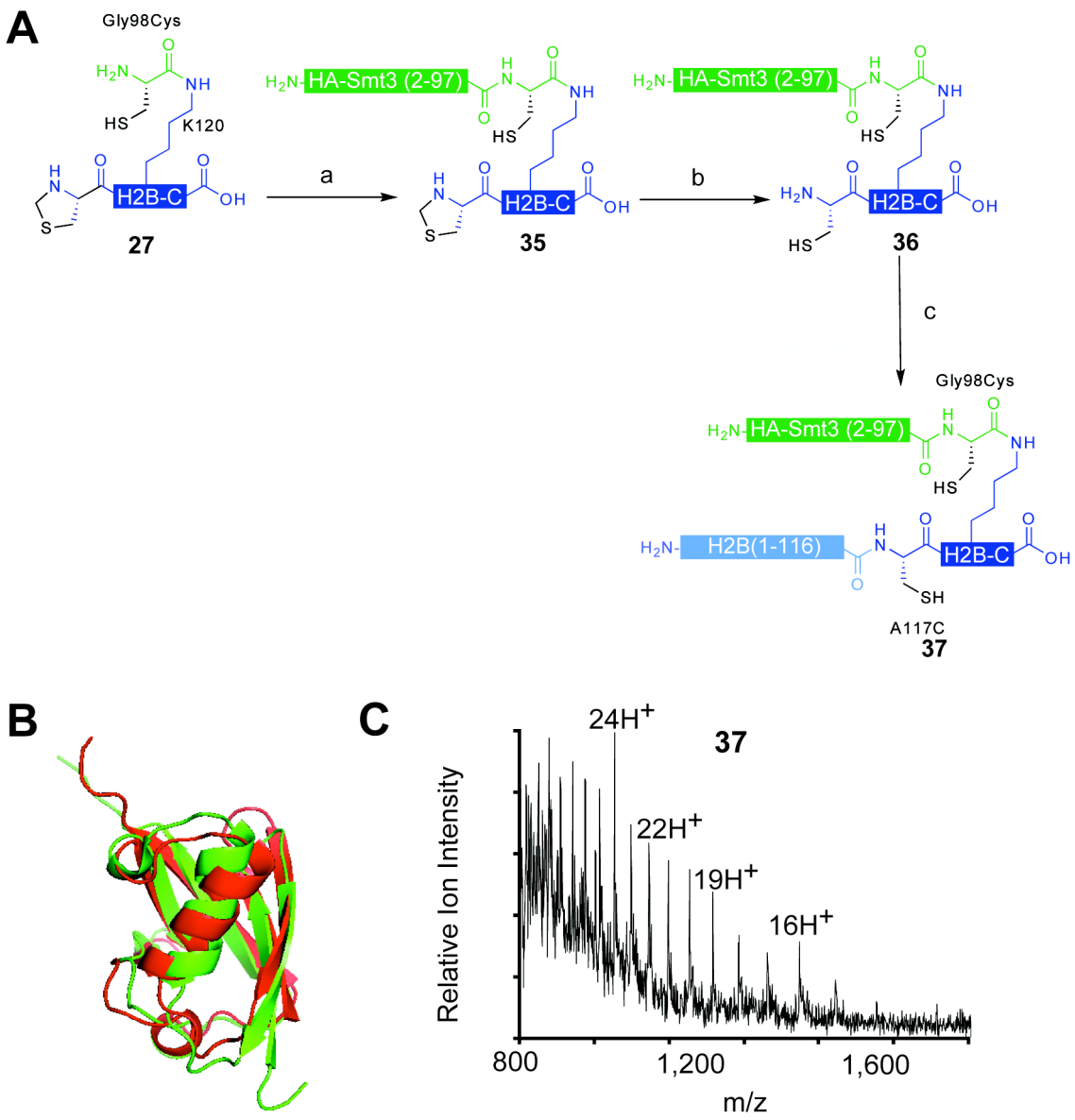



**Figure 5.6. Semisynthesis of sH2B(cys).** A) Schematic of sH2B(cys) semisynthesis. (a) EPL was used to ligate peptide **27** to HA-Smt3(2-98)-α-thioester, **18**, to give protein **35**. (b) Ligation product **35** was treated with methoxylamine at pH 5, affording protein **36**. (c) Ligation of protein **36** to protein **24**, forming sH2BA117C/G98C (sH2B(cys)), **37**. B) Structural alignment of ubiquitin (1UBQ, red) (Vijay-Kumar et al., 1987) and Smt3 (1EUV, green) (Mossessova and Lima, 2000) shown in ribbon diagram. Structures and alignment rendered with PyMol. C) ESI-MS spectra of isolated sH2B(cys), **37** [$(M+H)^+$ observed = 26,228 ± 9 Da (s.d.). $(M+H)^+$ expected = 26,224 Da.].

We turned our attention to the interaction of ubiquitin with ubiquitin binding proteins. In almost all cases reported, ubiquitin interactions involve a hydrophobic patch including leucine 8 (L8) and isoleucine 44 (I44) (Figure 5.7A) (Hurley et al., 2006). Mutation of one or both residues to alanine typically abrogates binding (Beal et al., 1996; Kang et al., 2003; Shih et al., 2002). We generated mut-uH2B containing the triple alanine mutant L8A/I44A/G76A by replacing thioester **15**, with the corresponding thioester bearing the L8A/I44A double mutation (Figure 5.7C and 5.8). Ligation of peptide **27** to mutant thioester **38** afforded ubiquitylated peptide **39** (Figure 5.7C, step a). Deprotection of **39**, giving **40**, followed by ligation to protein **24**, yielded intermediate product **41** (Figure 5.7C, steps b and c). Radical-initiated desulfurization gave mut-uH2B, **42** (Figure 5.7C, step d and 5.7B). Dot1L was able to methylate nucleosomes bearing mut-uH2B to a similar degree as those bearing u(G76A)H2B (Figure 5.5A and C). Therefore, a non-canonical interaction surface of ubiquitin seems to be involved in stimulation of Dot1L, likely by binding to the nucleosome, Dot1L, or both.

If ubiquitin binds Dot1L leading to allosteric activation, ubiquitin added in *trans* might be able to increase methylation of unmodified nucleosomes. However, at 1 mM exogenous ubiquitin, no stimulation of Dot1L was observed (Figure 5.9). In contrast, if ubiquitin binds to the nucleosome, mutations on the nucleosomal surface surrounding the site of ubiquitin attachment might prevent stimulation of Dot1L activity by decoupling ubiquitylation with changes to the nucleosomal structure. Previously, alanine scanning of the yeast nucleosome revealed two mutants in the acidic patch of H2A that prevent H3K4 methylation, which is also directly stimulated by uH2B (Kim et al., 2009), without an observable effect on ubiquitylation of H2B (Nakanishi et al., 2008). We reasoned that if H3K4 methylation is coupled to uH2B through a similar mechanism as H3K79 methylation, these mutations might disrupt Dot1L stimulation in

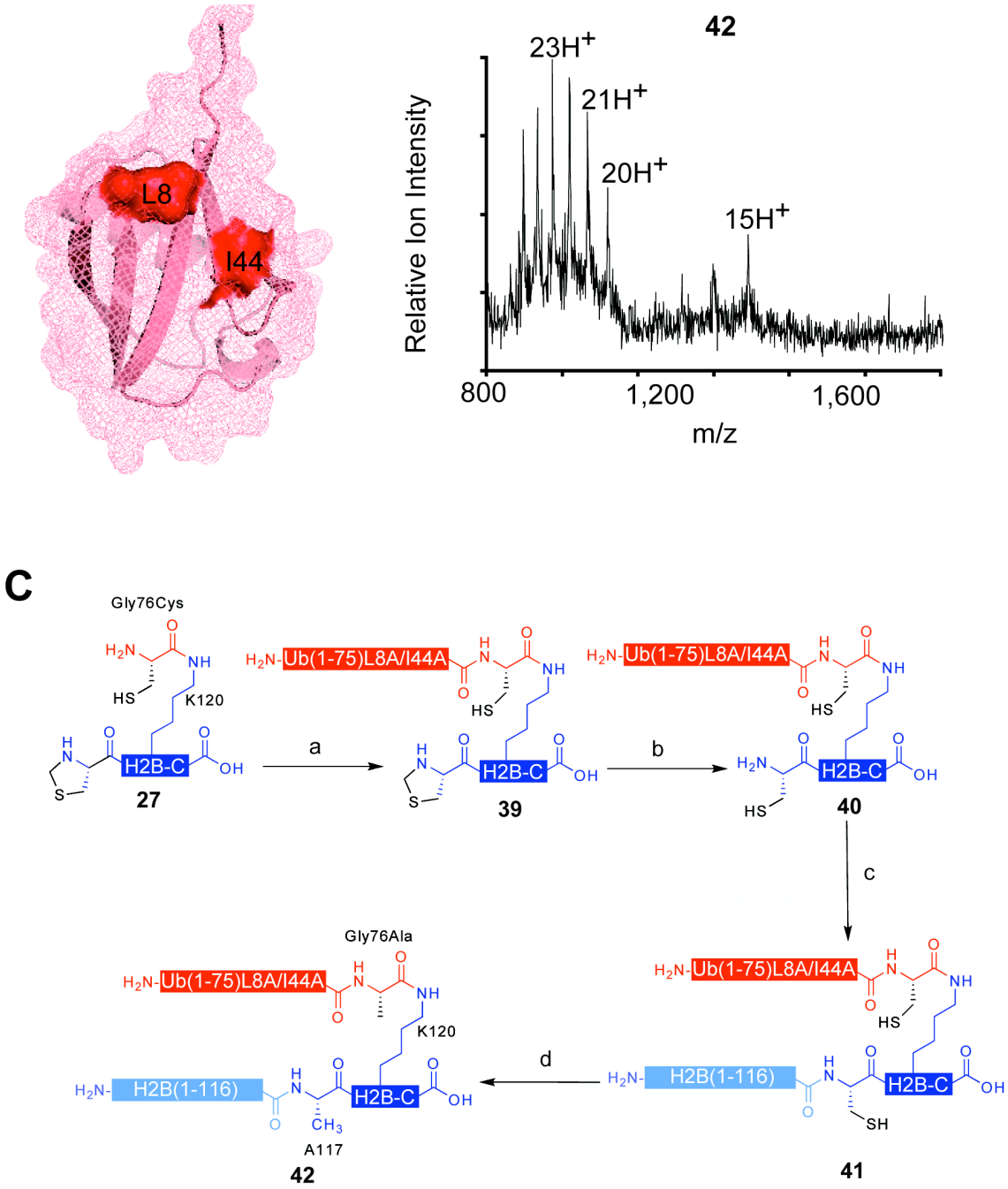



**Figure 5.7. Semisynthesis of mut-uH2B.** A) Structure of ubiquitin (1UBQ) (Vijay-Kumar et al., 1987) displayed as ribbon and mesh. The surfaces of L8 and I44 are emphasized in red. B) ESI-MS spectra of isolated mut-uH2B, **42** [$(M+H)^+$ observed = 22,293 ± 7 Da (s.d.) $(M+H)^+$ expected = 22,295 Da.]. C) Schematic of mut-uH2B semisynthesis. (a) EPL was used to ligate peptide **27** to ubiquitin(1-75)L8A/I44A-α-thioester, **38**, to give protein **39**. (b) Ligation product **39** was treated with methoxylamine at pH 5, affording protein **40**. (c) Ligation of protein **40** to protein **24**, forming intermediate product **41**. (d) Radical-initiated desulfurization of ligation product **41**, affording mut-uH2B, **42**.

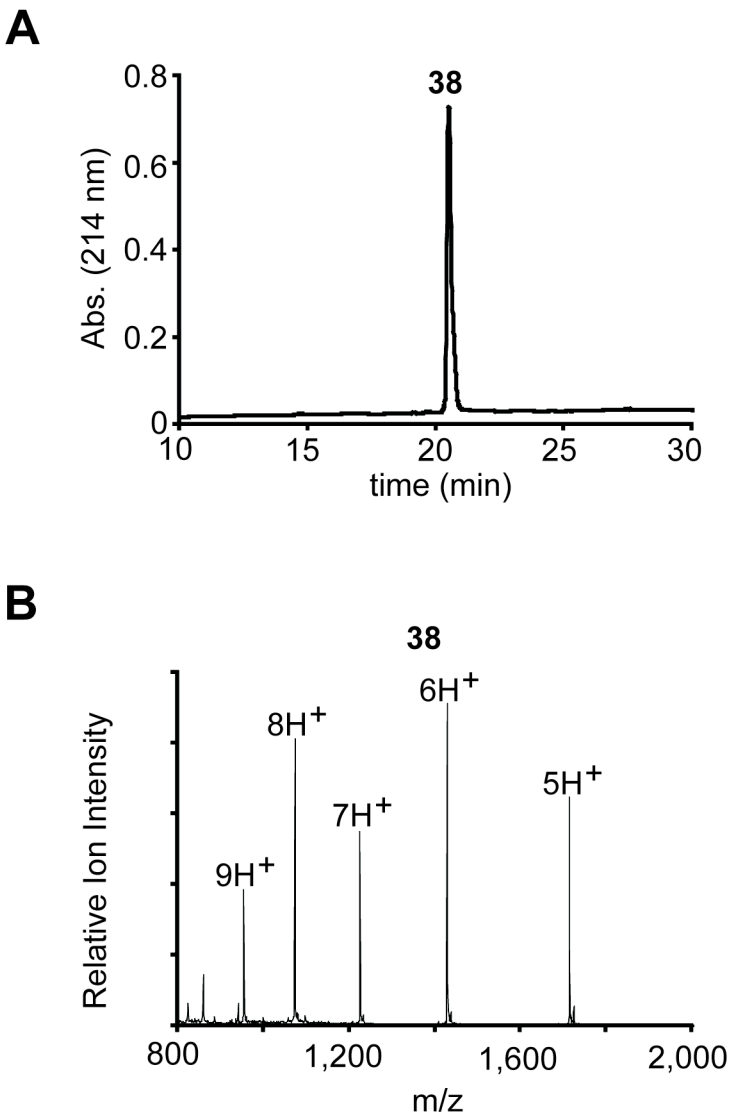



**Figure 5.8. Characterization of ubiquitin(1-75)L8A/I44A-α-thioester, 38.** A) RP-HPLC chromatogram and B) ESI-MS spectrum of ubiquitin(1-75)L8A/I44A-α-thioester, **38**. Charge states are labeled. [$(M+H)^+$ observed = 8,548 ± 1 Da (s.d.). $(M+H)^+$ expected = 8,548 Da.]

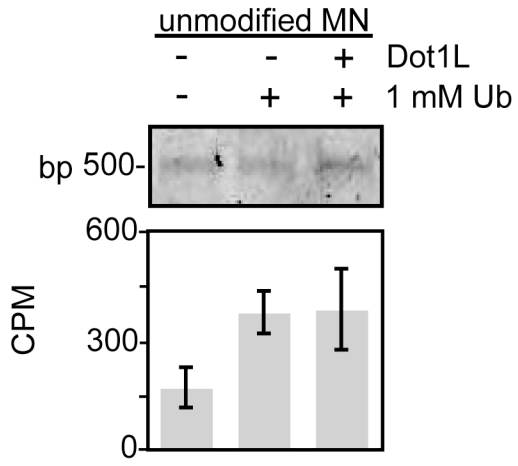



**Figure 5.9. Dot1L assay with unmodified nucleosomes and ubiquitin in *trans*.** Dot1L assay with $^{3}$H SAM and unmodified nucleosomes with and without 1 mM exogenous ubiquitin. Assay samples were separated by 5% native gel electrophoresis and visualized with ethidium bromide staining (top panel). Quantification of $^{3}$H methyl incorporation was performed by liquid scintillation counting (bottom panel). Error bars represent one s.d. (n = 4).

our system. We prepared nucleosomes containing unmodified H2B or u(G76A)H2B in combination with each of the mutations, H2AE64A and H2AN68A (Figure 5.10 and 5.11A). Neither mutation altered Dot1L stimulation by uH2B (Figure 5.11B), indicating divergent roles for uH2B in methylation of H3K4 and K79.

The basic patch of H4 is critical to the action of yeast Dot1 in telomeric silencing (Altaf et al., 2007). Simultaneous mutation of two arginines in H4 at positions 17 and 19 to alanines (H4R17/19A) nearly abolishes H3K79 methylation in yeast (Fingerman et al., 2007). We wondered if the same pattern would hold true for human Dot1L and if ubiquitylation could overcome any deficit in methylation due to the mutations. Nucleosomes were prepared containing H4R17/19A (Figure 5.10 and 5.11A) and either unmodified H2B or u(G76A)H2B. The H4R17/19A mutation decreased Dot1L activity on unmodified nucleosomes (Figure 5.11C). A similar decrease was observed with nucleosomes containing u(G76A)H2B (Figure 5.11B). Therefore, while conserved from yeast to humans, mutation of the basic patch seems to disrupt Dot1 activity independently of ubiquitylation.

### 5.3. One ubiquitylation stimulates methylation of one nucleosomal H3

Thus far, Dot1L activity has only been studied in the context of homogenously ubiquitylated nucleosomes, where each copy of H2B is ubiquitylated. Given a 1% prevalence of uH2B *in vivo* (West and Bonner, 1980), statistically, it is likely that heterogeneously ubiquitylated nucleosomes are predominant. This raises the question: can a nucleosome bearing one ubiquitylated H2B stimulate Dot1L? If so, will the methyltransferase activity be confined to one copy of H3, or is one uH2B sufficient to drive methylation of H3 on both sides of the nucleosome? To address these questions, we reconstituted histone octamers, varying the ratio of

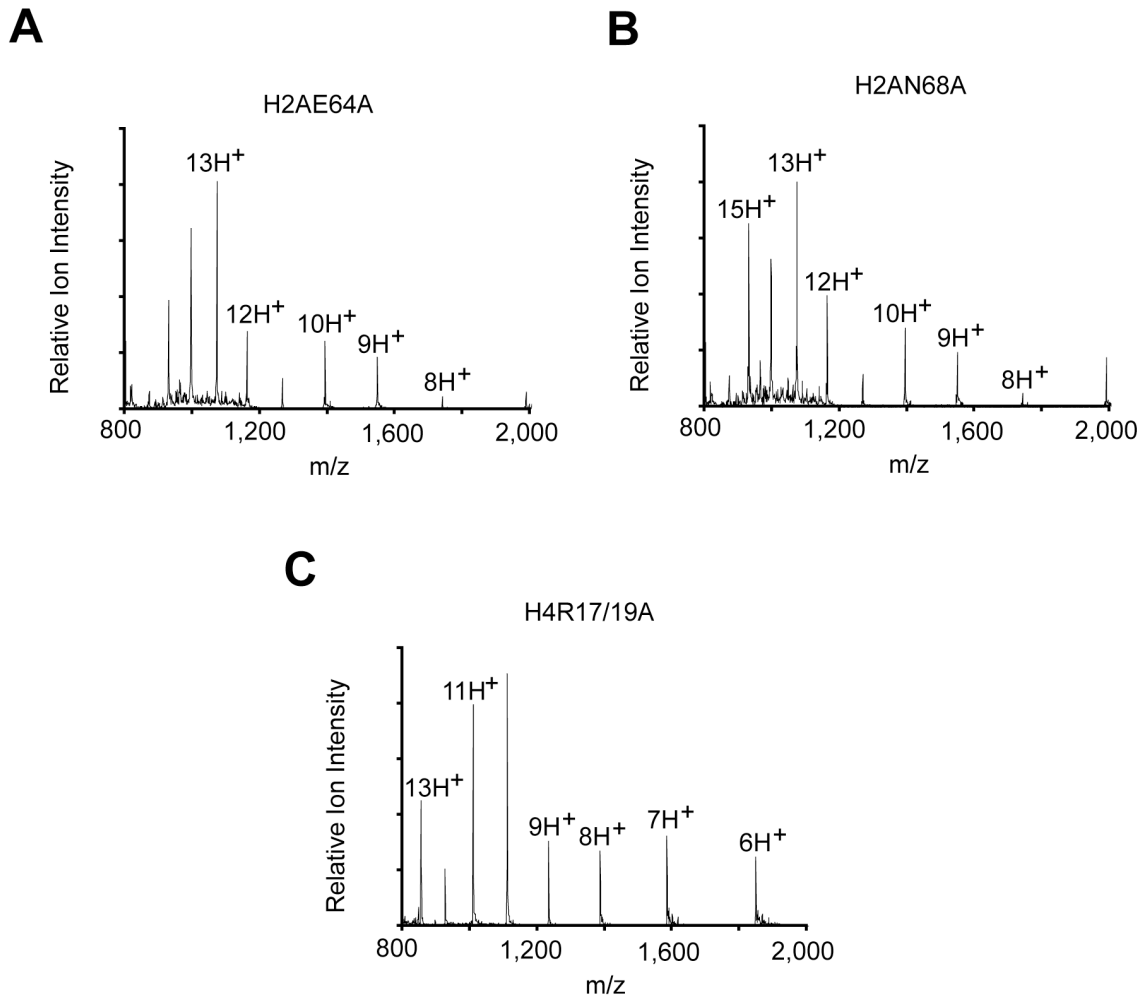



**Figure 5.10. Characterization of histone mutants.** ESI-MS spectra of the following purified recombinant histone mutants: A) H2AE64A [$(M+H)^+$ observed = 13,894 ± 3 Da (s.d.). $(M+H)^+$ expected = 13,893 Da.]; B) H2AN68A [$(M+H)^+$ observed = 13,907 ± 3 Da. $(M+H)^+$ expected = 13,908 Da.]; and C) H4R17/19A [$(M+H)^+$ observed = 11,067 ± 3 Da. $(M+H)^+$ expected = 11,067 Da.]. Charge states are labeled.

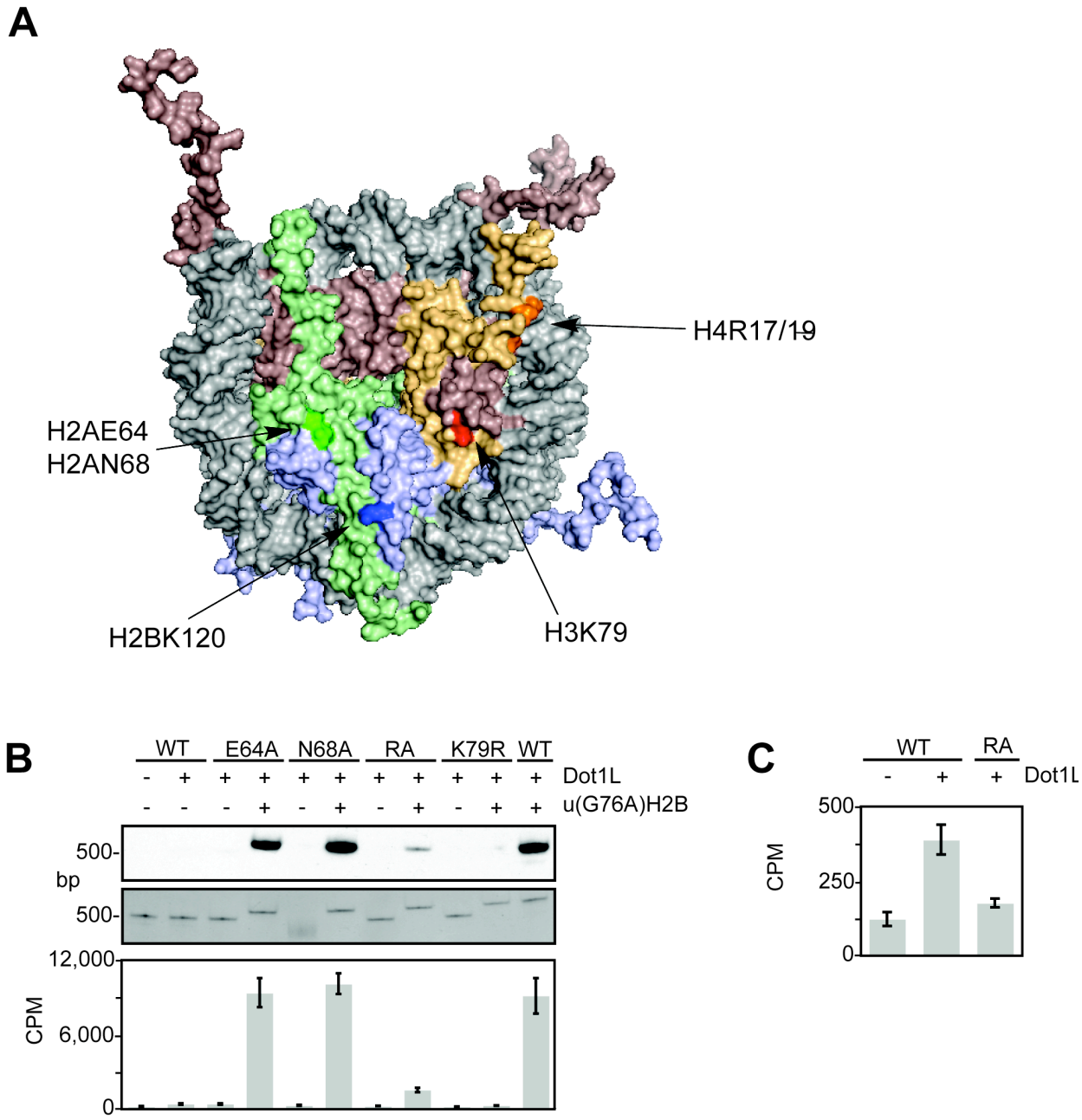



**Figure 5.11. Dot1L assays on mutant nucleosomes.** A) Surface representation of nucleosome structure (1KX5) (Davey et al., 2002) with H2A, H2B, H3, and H4 shown in green, blue, red, and orange, respectively. Amino acids in each histone are emphasized by darker shades. B) Dot1L methyltransferase assay on mutant nucleosomes containing H2B or u(G76A)H2B. Nucleosomes methylated with $^{3}$H SAM were separated by 5% native-PAGE and stained with ethidium bromide (middle panel) prior to probing for $^{3}$H methyl incorporation by fluorography (top panel). Quantification of methyltransferase activity was performed by liquid scintillation counting (bottom panel). Error bars represent one s.d. (n = 3). C) Representation of liquid scintillation counting data in panel B, illustrating the effect of the H4R17/19A (RA) mutant on Dot1L-mediated methylation of unmodified nucleosomes.

H2B to u(G76A)H2B. Electrophoretic analysis of the purified octamers confirmed incorporation of appropriate levels of unmodified and ubiquitylated H2B (Figure 5.12A). Separation of nucleosomes formed with each octamer sample by native gel electrophoresis verified independent assortment of the two forms of H2B, with singly ubiquitylated nucleosomes exhibiting an intermediate electrophoretic mobility when compared to doubly ubiquitylated and unmodified nucleosomes (Figure 5.12B, bottom panel). Methyltransferase assays showed a linear relationship between the level of u(G76A)H2B in a nucleosome and the activity of Dot1L, suggesting that one ubiquitylation stimulates methylation of only one H3K79 (Figure 5.12B and C). However, whether uH2B stimulates methylation of the same nucleosomal surface or the opposite surface cannot be determined from this set of experiments (Figure 5.12D). As further validation of this result, a pool of nucleosomes reconstituted from octamers containing an equal mixture of unmodified and ubiquitylated H2B, have two singly ubiquitylated nucleosomes for every one doubly ubiquitylated nucleosome; yet, half of the methyltransferase activity observed by fluorography occurred on the doubly ubiquitylated nucleosomes (Figure 5.12B).

### 5.4. Forays into crystallography of nucleosomes containing u(G76A)H2B[4]

High-resolution structural characterization of ubiquitylated nucleosomes would provide indisputable evidence for or against an allosteric change in the nucleosome resulting from H2B ubiquitylation. An assortment of nucleosomal structures have been solved using X-ray crystallography since the first high resolution structure was reported in 1997 (Luger et al., 1997a). These structures include nucleosomes bearing histones from different organisms (Clapier et al., 2008; Tsunaka et al., 2005; White et al., 2001), varying DNA sequences

[4] This work was performed as part of a close collaboration with Matthew Bick in the Laboratory of Molecular Biophysics at the Rockefeller University.

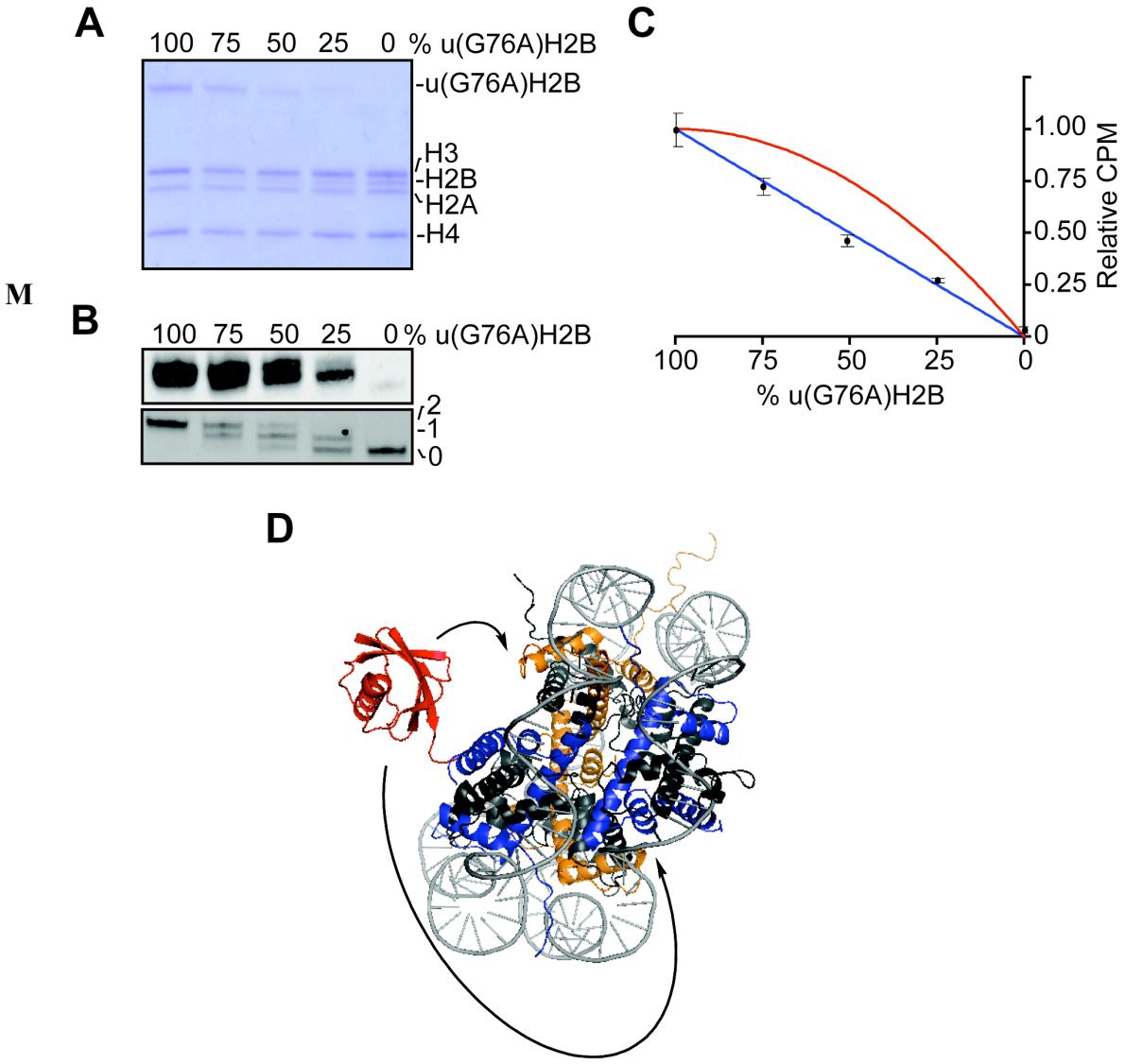



**Figure 5.12. Investigation of ubiquitylation/methylation stoichiometry.** A) Octamers formed with varying ratios or u(G76A)H2B and H2B were separated on a 15% SDS-polyacrylamide gel and stained with Coomassie. B) Nucleosomes formed with octamers in panel A, were methylated with Dot1L and $^{3}$H SAM. Assay samples were separated by native electrophoresis and stained with ethidium bromide (bottom panel) followed by probing for $^{3}$H methyl incorporation by fluorography (top panel). The position of nucleosomes bearing 0, 1, or 2 ubiquitins is indicated. C) Quantification of assay samples by liquid scintillation counting, plotted relative to the fully ubiquitylated sample. The blue line (rel. CPM = ($f_{ub}$); where $f_{ub}$ is the fraction of u(G76A)H2B) represents the case where each u(G76A)H2B stimulates methylation of only one H3K79 side chain in the same nucleosome. The red line (rel. CPM = $-(f_{ub})^2 + 2(f_{ub})$) represents a case where each u(G76A)H2B stimulates methylation of both H3K79 side chains in the same nucleosome. Error bars represent one s.d. (n = 4). D) Superimposed structures of the nucleosome (1KX5) (Davey et al., 2002) and ubiquitin (1UBQ, red) (Vijay-Kumar et al., 1987). H2B and H3 are colored blue and orange respectively. Methylation is either targeted to the same nucleosomal face as ubiquitin or the opposite face.

(Bao et al., 2006; Davey et al., 2002; Ong et al., 2007), histone variants (Suto et al., 2000) and methyllysine analogs (MLAs) (Lu et al., 2008), as well as complexes between nucleosomes and small ligands (Barbera et al., 2006; Edayathumangalam et al., 2004; Suto et al., 2003; Wu et al., 2008) and higher order chromatin structures (Schalch et al., 2005). However, despite diverse structural differences, an overwhelming similarity in optimized crystallography conditions is observed. Invariably, nucleosomes have been crystallized in potassium cacodylate at pH 6.0 with 25-40 mM KCl and 35-50 mM $MnCl_2$ (or in one case $MgCl_2$). We hoped that these conditions would also allow the crystallization of nucleosomes containing u(G76A)H2B.

The 146 bp palindromic α-satellite sequence (Figure 5.1A) was selected for u(G76A)H2B nucleosome formation. As this sequence has been used repeatedly in nucleosome crystallization and structure determination (Barbera et al., 2006; Lu et al., 2008; Luger et al., 1997a; Muthurajan et al., 2004; Suto et al., 2000; Suto et al., 2003; White et al., 2001), it offers a suitable starting point for u(G76A)H2B nucleosome crystallization and would best facilitate a molecular replacement structural solution. Sixty-four copies of half of the 146 bp α-satellite palindrome were inserted into a vector[5] and amplified in *E. coli*. The sequence was isolated following digestion from the vector, allowing formation of multiple milligrams of the mature palindrome in high purity (Figure 5.13).

Nucleosome formation with this DNA and octamers containing H2B or u(G76A)H2B was then optimized. Even with strong positioning sequences, nucleosomes formed by dialysis exist in multiple translational states relative to the DNA (Dyer et al., 2004). Typically, this variability can be resolved by heating. Therefore nucleosomes were formed using different ratios of octamer to DNA and heated to 37 °C or 55 °C for 2 h (Figure 5.14A and B). Conditions

[5] This was completed by Dr. Kyle P. Chiang in the Laboratory of Synthetic Protein Chemistry at the Rockefeller University.

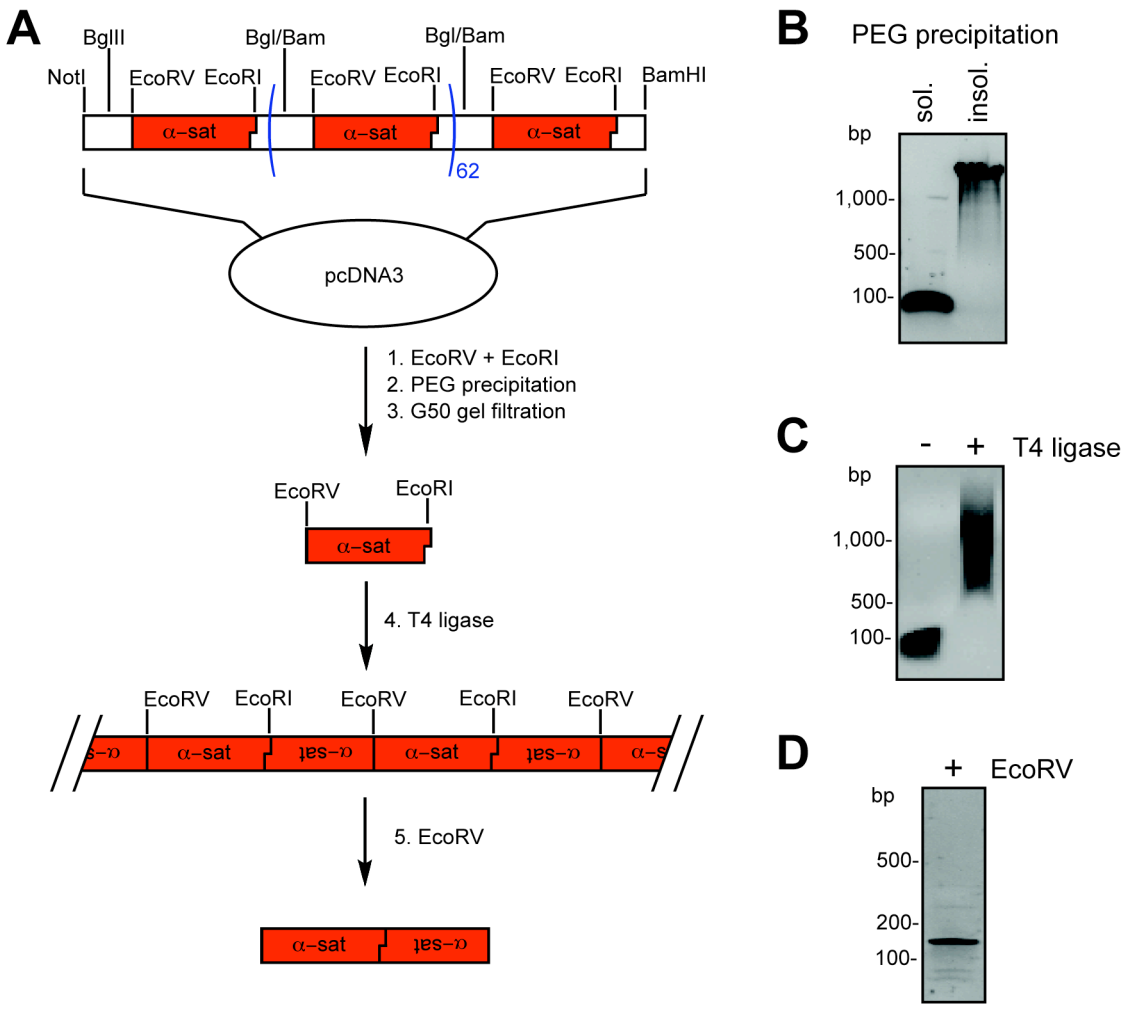



**Figure 5.13. Preparation of 146 bp palindromic α-satellite DNA sequence.** A) Schematic of 146 bp α-satellite palindromic DNA preparation. 64 copies were digested from a vector and purified by PEG precipitation and G50 resin gel filtration. Fragments were ligated with T4 ligase and subsequently digested with EcoRV to generate the mature palindromic sequence. B) Ethidium bromide stained agarose gel of PEG precipitation. Desired fragment remains soluble, while vector precipitates. C) Ethidium bromide stained gel of ligation showing laddering of DNA. D) Ethidium bromide stained gel of EcoRV digested mature palindromic DNA.

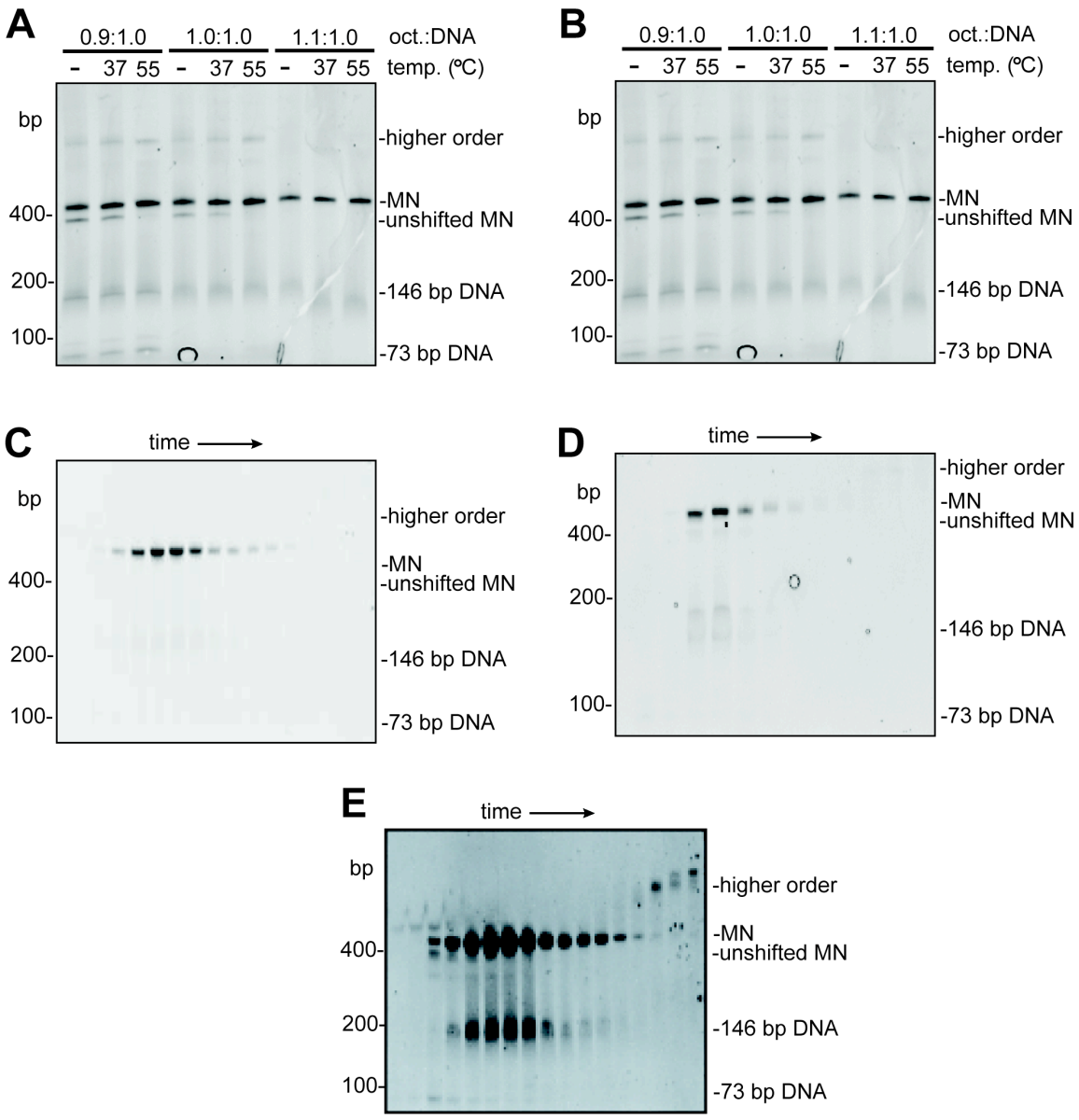



**Figure 5.14. Optimization of nucleosome formation and purification.** Small-scale nucleosome (MN) formation using indicated ratios of octamers containing A) H2B or B) u(G76A)H2B to DNA. Following reconstitution, nucleosomes were heated for 2 h at indicated temperatures and heat shifted nucleosomes were separated by 5% native-PAGE and stained with ethidium bromide. In both cases, a 1.0:1.0 octamer:DNA ratio and 55 °C incubation were selected as optimal conditions. Nucleosomes containing C) H2B and D) u(G76A)H2B were reconstituted on a large-scale and purified by 5% preparative native-PAGE. Ethidium bromide stained 5% native gels of fractions are shown. E) Overexposed image of gel in panel C to highlight purification from contaminants. DNA observed in pure nucleosome fractions results from decomposition during electrophoresis.

resulting in the greatest homogenous yields of nucleosome with minimal excess DNA were selected. Nucleosomes were formed on large-scale using these optimized conditions and purified by preparative native-PAGE (Figure 5.14C, D, and E). These purified nucleosomes were concentrated for use in crystallization trials.

Three separate crystallization screens were performed using wild type, unmodified nucleosomes and a range previously optimized crystallization conditions (Figure 5.15). Several conditions resulted in formation of crystals (Figure 5.16A). These crystals exhibited low-resolution diffraction patterns and contained appropriate ratios of histone proteins (Figure 5.16B and C). This validates the quality of the nucleosome preparations for use in crystallographic studies. Unfortunately, under the same conditions, no crystallization of nucleosomes containing u(G76A)H2B was observed (Figure 5.17). As ubiquitylation is a larger change than any reported previously in a nucleosome structure, different conditions might be required to crystallize ubiquitylated nucleosomes. As such, a screen was performed using the Qiagen Nucleix kit optimized for the crystallization of protein-nucleic acid complexes. This screen was also unsuccessful. Further screening is required to determine optimal conditions for crystallization of nucleosomes containing u(G76A)H2B; however, the notion that this line of investigation can even be considered is a testament to the power of the semisynthetic approach presented in chapter 4.

### 5.5. Summary

In this chapter, insights into the mechanism of the stimulation of Dot1L by uH2B have been presented. Kinetic analysis shows an extraordinary enhancement in Dot1L activity over a wide range of nucleosome concentrations. This enhancement owes from an increase in the

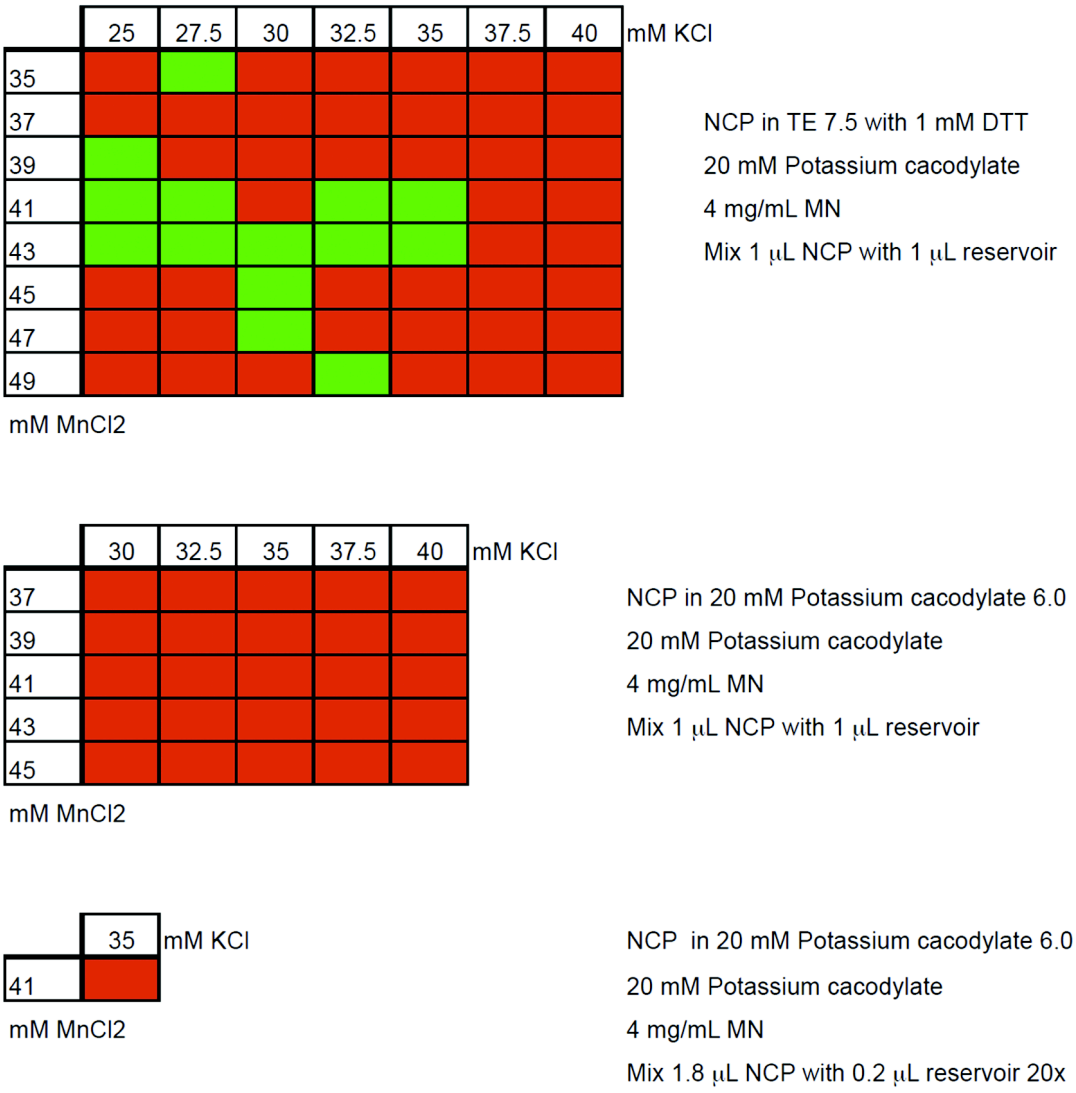



**Figure 5.15. Crystal screens with unmodified nucleosomes.** Three crystal screens were performed with unmodified nucleosomes using a range of conditions previously optimized for nucleosome crystallization. The concentrations of KCl and $MnCl_2$ in the reservoir were varied as indicated. Green and red boxes denote presence and absence of crystals, respectively, after three weeks. MN = mononucleosome; NCP = nucleosome core complex. These terms are used interchangeably.

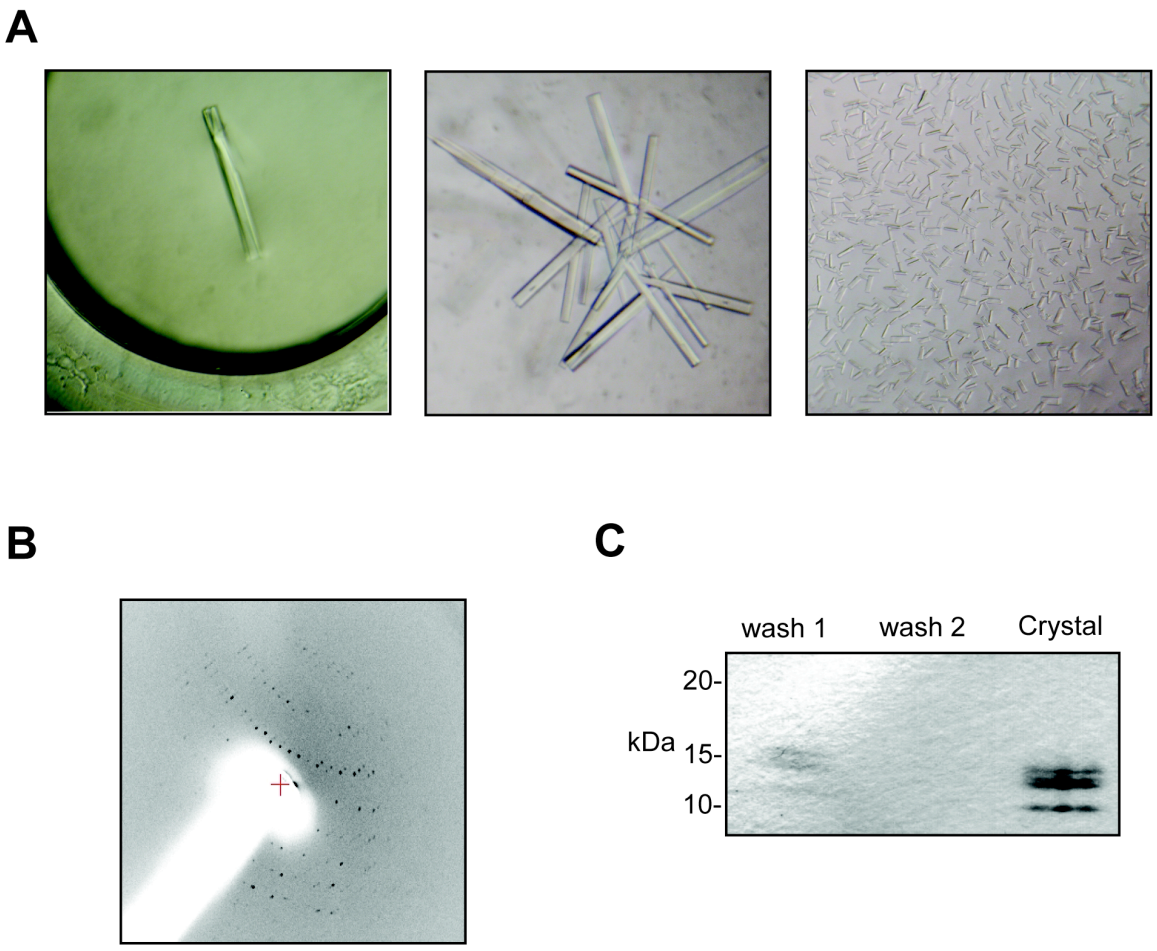



**Figure 5.16. Characterization of crystals of unmodified nucleosomes.** A) Three different crystal morphologies were observed. B) Diffraction pattern generated from crystal in panel A, left. C) Crystals in panel A, center, were washed twice with reservoir solution and dissolved in SDS-PAGE loading buffer and separated by 8-25% SDS-PAGE and stained with Coomassie.

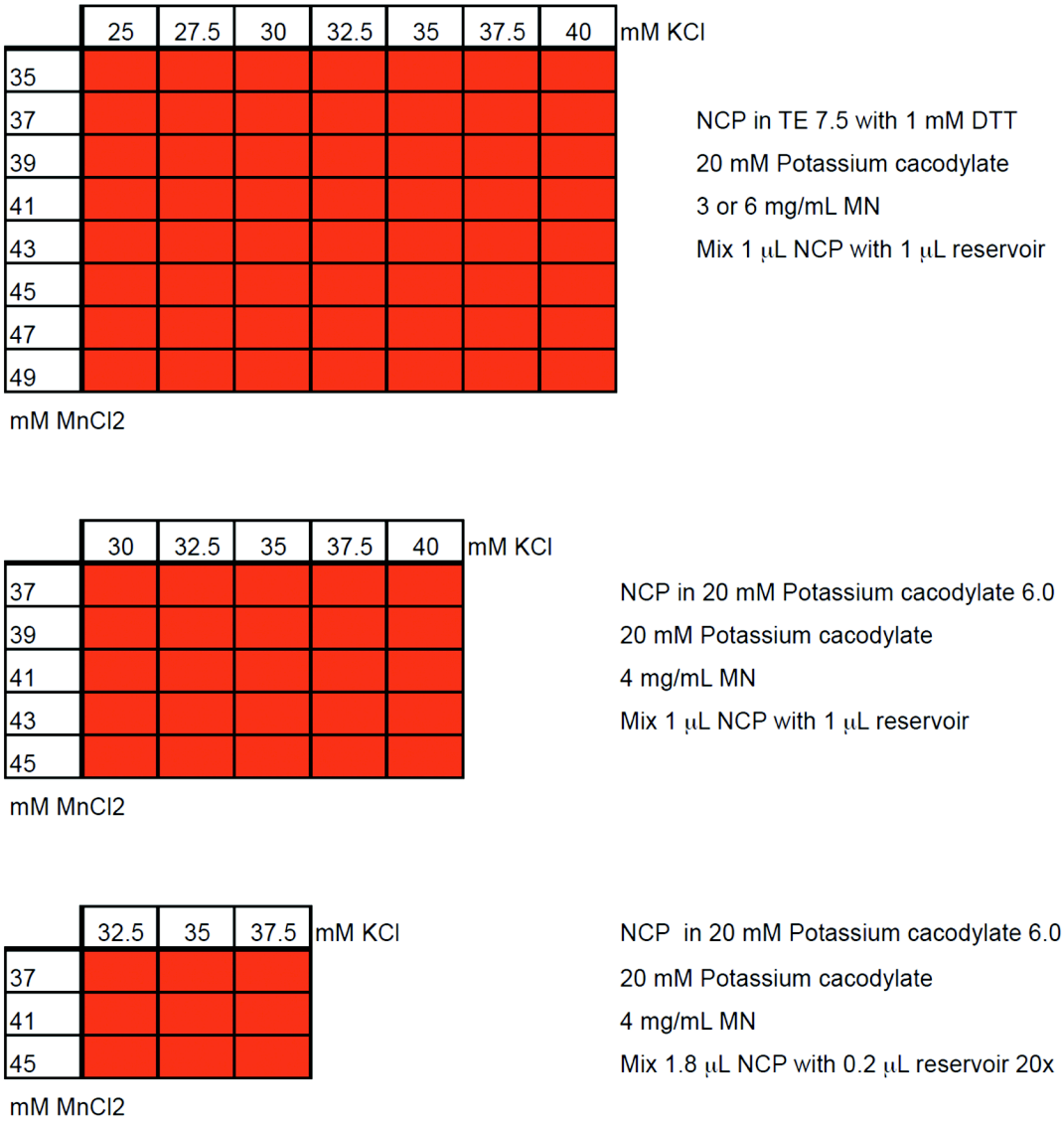



**Figure 5.17. Crystal screens with nucleosomes containing u(G76A)H2B.** Three crystal screens were performed with u(G76A)H2B nucleosomes using a range of conditions previously optimized for nucleosome crystallization. The concentrations of KCl and $MnCl_2$ in the reservoir were varied as indicated. Red denotes absence of crystals after three weeks. MN = mononucleosome; NCP = nucleosome core complex. These terms are used interchangeably.

chemical step of methylation, which must result from a change in the structure of the Dot1L-H3K79 active site complex. While the ubiquitin fold was required to translate ubiquitylation of H2B into heightened H3K79 methylation, mutation of the canonical ubiquitin binding surface had no effect on Dot1L activity, suggesting a role for a non-canonical surface. Additionally, it was demonstrated that each uH2B within the nucleosome stimulates methylation of one H3K79. Further investigation, possibly structural in nature, will be necessary to fully elucidate the mechanisms linking uH2B and Dot1L-mediated H3K79 methylation. Thus far, efforts to crystallize nucleosomes containing u(G76A)H2B have been unsuccessful.

## Chapter 6: Discussion

### 6.1. Semisynthesis of ubiquitylated H2B

In this thesis, two semisynthetic strategies for the regioselective ubiquitylation of H2B have been presented. Both rely on a pair of orthogonal EPL reactions employing a branched synthetic peptide and two protein α-thioesters to construct the ubiquitylated proteins. The first strategy, presented in chapter 2, allowed the preparation of native uH2B (Figure 2.11). Due to the absence of cysteines in both H2B and ubiquitin, two traceless protein ligations were required to generate the native branched protein. A photolytically removable ligation auxiliary facilitated the ligation of a C-terminal H2B peptide to the ubiquitin sequence and following ligation was removed upon UV irradiation, leaving the native isopeptide junction. A traditional cysteine-mediated ligation was then used to add the remainder of H2B to the ubiquitylated peptide. To accomplish this, an A117C mutation was incorporated into the synthetic peptide. Following ligation, C117 was reduced to the native alanine using Raney nickel, affording native uH2B in low milligram quantities.

In chapter 4, a scalable semisynthesis of u(G76A)H2B was presented to overcome shortcomings of the native strategy (Figure 4.3). By replacing the auxiliary-mediated ligation with a traditional cysteine-mediated ligation, we efficiently generated u(G76A)H2B on a scale of multiple tens of milligrams – an increase in excess of ten-fold over the native strategy. Moreover, by eliminating the sluggish auxiliary-mediated ligation and subsequent purification, the time required to generate ubiquitylated H2B was halved (not including time invested in synthesis of the ligation auxiliary). Eliminating the need for photolytic removal of protecting groups not only further enhanced the scalability of the semisynthesis, but also negated the need for specialized optics. Most importantly, all materials required for the solid phase synthesis of

peptide **27** (Figure 4.2) are commercially available, obviating time-intensive solution phase organic synthesis and significantly increasing the accessibility of this methodology.

Both strategies will allow the preparation of other ubiquitylated histones, and more broadly, other ubiquitylated proteins. This is especially important in systems where the natural abundance of ubiquitylated proteins is low or where ubiquitin conjugating enzymes are unknown or unavailable. Additionally, as we have illustrated in this thesis, chemical control of both ubiquitin and the target protein permits detailed analysis of the mechanisms underlying signaling through monoubiquitylation. Currently, work is underway to prepare H2A ubiquitylated at K119 to interrogate its role in heterochromatic gene silencing.[6]

The native and G76A ligation strategies should be considered complementary approaches to the study of protein monoubiquitylation. In discovery and validation experiments it is often ideal to have the native sequence surrounding the isopeptide bond to most closely mimic cellular processes. However, the quantity of material required for mechanistic investigations, including kinetics and structural biology, are inaccessible by the native ligation strategy. Thus, it is reasonable to use the G76A mutant if it can be shown to be an acceptable surrogate for the native sequence. For quantitative analyses, it is important that the mutation has no significant effect on downstream function, but for qualitative analyses, a similar trend is sufficient to warrant the substitution. In contrast to most ubiquitin binding domains that recognize defined regions of the ubiquitin fold (Hurley et al., 2006), proteins that specifically engage the isopeptide bond are likely to exhibit detrimental effects on activity as a result of the G76A mutation. For example, the ubiquitin G76A mutant is efficiently mobilized and transferred to target proteins *in vivo*, but the conjugated product has increased stability, likely due to decreased susceptibility to

[6] This is an ongoing collaboration with Sarah Whitcomb in the Laboratory of Chromatin Biology and Epigenetics at the Rockefeller University.

deubiquitylating enzymes (Hodgins et al., 1992; Pickart et al., 1994). We expected that the G76A mutation would render the ligation product increasingly stable to ubiquitin hydrolysis *in vitro*. Surprisingly, there was no observable effect of the mutation on ubiquitin hydrolysis by UCH-L3 (Figure 4.5). This may be the exception rather than the rule, and native ubiquitylated proteins may be more appropriate reagents for the biochemical characterization of DUBs.

There are also situations where significant changes to the isopeptide linkage are advantageous. One example is in structure activity relationship analyses, such as those presented in chapter 5. Another example includes fishing experiments in cell lysates for interacting partners of monoubiquitylated proteins. In the latter case, an isopeptide bond that is stable to the UCHs and DUBs present in lysates is desirable – preventing unwanted hydrolysis of the semisynthetic ubiquitylated protein. While the G76A mutation may increase stability in some cases, it is insufficient for this purpose. Fortunately, the G76A ligation strategy can be extended to incorporate further changes into the ubiquitin C-terminal sequence. The G75A/G76P double mutation is a reasonable option, as the G76P mutation alone prevents mobilization of ubiquitin from linear fusions and is likely to result in the near complete stability of the isopeptide bond (Bachmair et al., 1986; Varshavsky, 2005). The G75A/G76P double mutant could be prepared through ligation of an ubiquitin(1-74)-α-thioester and a peptide or protein with the Cys-Pro sequence coupled to the ε-$NH_2$ of a lysine side chain. To complete the semisynthesis, the cysteine at position 75 of ubiquitin could be reduced to an alanine.

EPL-based strategies for the preparation of ubiquitylated proteins are best suited to ubiquitylation sites in N- and C-terminal regions of proteins. Preparation of a protein ubiquitylated at an internal site requires a third orthogonal ligation. Recently, we have developed a disulfide-directed approach for the ubiquitylation of H2B to circumvent this limitation of EPL

(Figure 6.1).[7] First, an ubiquitin(1-76)-intein fusion was reacted with 2-aminoethanethiol to generate full-length ubiquitin bearing a C-terminal thiol. This protein was subsequently reacted with a recombinant H2BK120C mutant with a disulfide-activated thiol to generate an uH2B mimic, uH2Bss. Replacement of a methylene in K120 of uH2B with a disulfide through this strategy lengthens the side chain by one atom (corresponding to ~2.4 Å) (Figure 6.1, bottom panel). However, nucleosomes containing uH2Bss stimulate Dot1L activity to levels comparable with nucleosomes containing uH2B or u(G76A)H2B. This strategy could be readily applied to the ubiquitylation of proteins at internal sequences, provided an orthogonal cysteine can be introduced into the protein.

**6.2. Semisynthesis of polyubiquitylated proteins**

An obvious question emerging from the semisynthesis of a monoubiquitylated protein is whether or not the technology can be extended to the preparation of polyubiquitylated proteins. This is a non-trivial problem. However, three possible routes to the polyubiquitylation of target proteins can be envisioned. The first route, involves the use of E1/E2/E3 ubiquitylation machinery to extend ubiquitin chains on a semisynthetic ubiquitylated protein (Figure 6.2, top panel).[8] Different lengths of ubiquitin chains could be separated by preparative SDS-PAGE and/or size exclusion chromatography. Alternatively, the enzymatic ubiquitin polymerization could be performed on an ubiquitin(1-75)-α-thioester. The resultant polyubiquitin chains could be separated as described above, and the purified polyubiquitin thioesters used in EPL with a desired target protein (Figure 6.2, bottom panel). If the thioester linkage is not stable during the

[7] This work was completed by Dr. Champak Chatterjee in the Laboratory of Synthetic Protein Chemistry at the Rockefeller University.
[8] A collaboration to do this is currently underway with Dr. Ray Deshaies' laboratory at the California Institute of Technology.

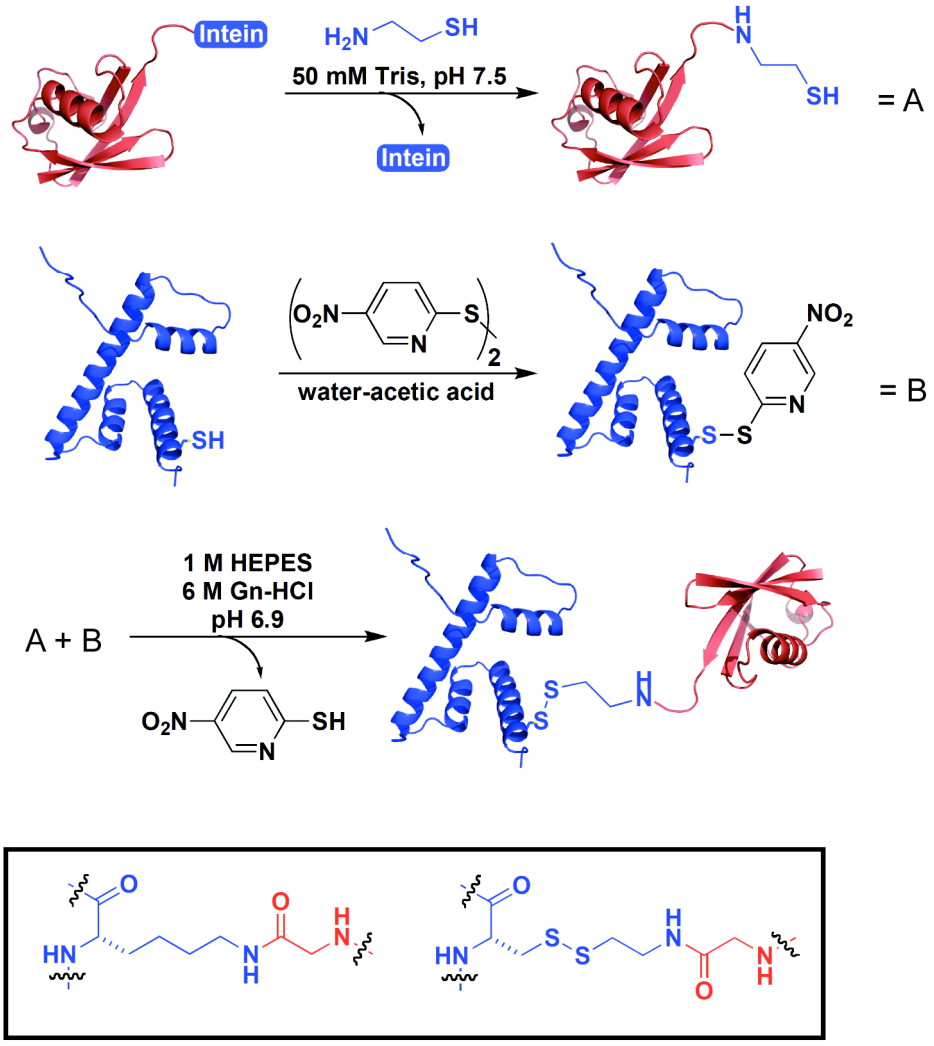



**Figure 6.1. Disulfide-directed ubiquitylation.** The recombinant ubiquitin(1-76)-intein fusion was reacted with 2-aminoethanethiol, affording ubiquitin with a C-terminal thiol, A (top panel). The thiol of H2BK120C was activated with 2,2'-dithiobis(5-nitropyridine), yielding mixed disulfide, B (second panel). Reaction of A with B led to formation of the disulfide-conjugated uH2B, uH2Bss (third panel). Inset at bottom illustrates the difference between uH2B and uH2Bss. Gn = guanidinium.

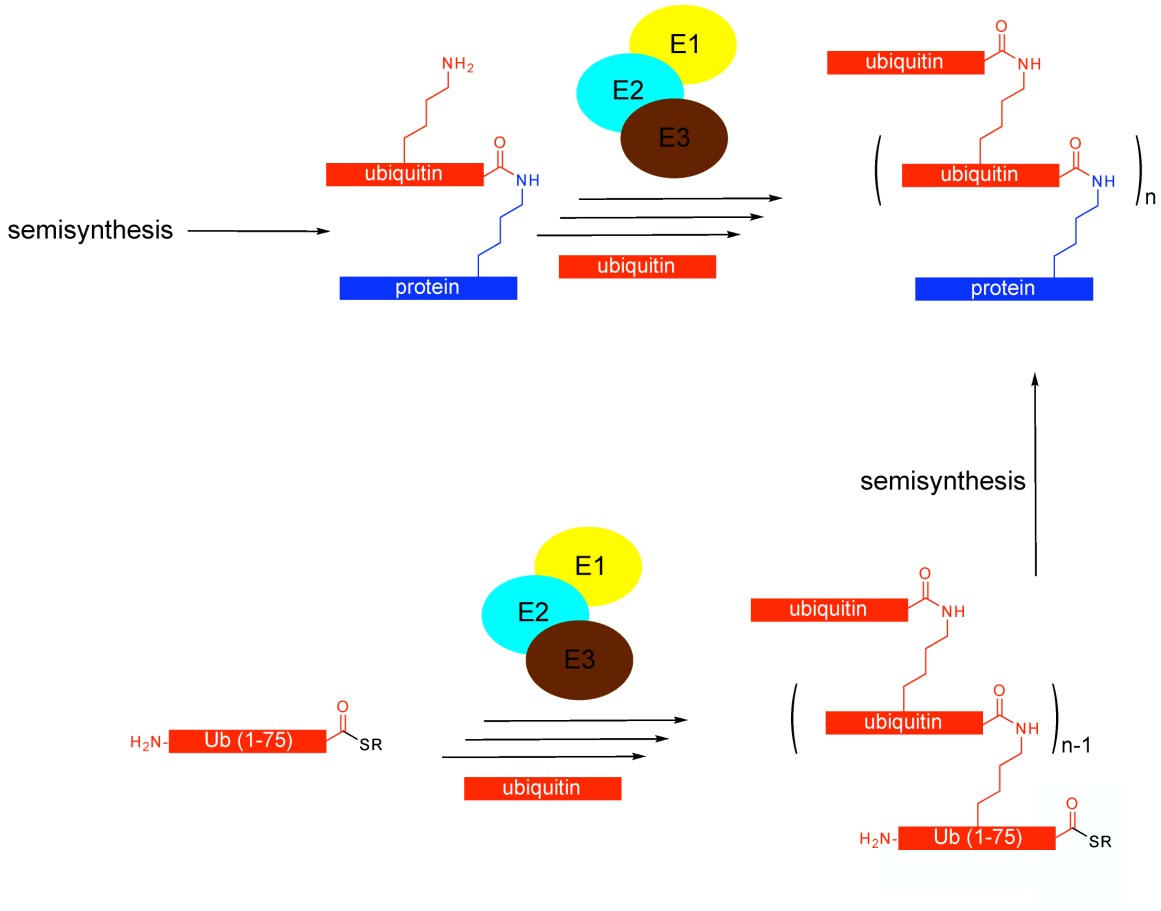



**Figure 6.2. Enzymatic approaches for the preparation of a polyubiquitylated protein.** Two possible approaches for the preparation of polyubiquitylated proteins can be envisioned harnessing E1 and specific E2s and E3s. A semisynthetic ubiquitylated protein can be used as a substrate for E1/E2/E3-mediated ubiquitylation (top panel). Alternatively, the ubiquitin(1-75)-α-thioester can be used as a substrate (bottom panel). Following ubiquitin polymerization, the polyubiquitin thioester can be used in an EPL reaction with a target protein.

ubiquitylation reaction or the ubiquitin(1-75)-α-thioester is used by the ubiquitylation machinery, a split intein at the C-terminus of ubiquitin(1-75) could be employed. These approaches rely on the specificity of the ubiquitylation machinery to generate a given ubiquitin linkage. This could be bypassed using a purely synthetic system (Figure 6.3). To accomplish this, a peptide containing residues 46-75, an A46C mutation, and a protected cysteine linked through an isopeptide bond with K48 or K63, could be synthesized with a safety catch linker. Following ligation to a synthetic ubiquitin(1-45)-α-thioester, the safety catch linker could be used to generate the branched ubiquitin(1-75)-α-thioester. This thioester could be ligated to a semisynthetic protein bearing an isopeptide-linked cysteine. Subsequent rounds of deprotection of the isopeptide-linked cysteine and ligation to the branched ubiquitin thioester would allow ubiquitin polymerization. The terminal ligation could be performed with the linear ubiquitin(1-75)-α-thioester. Desulfurization of this intermediate product would afford the polyubiquitylated protein with G76A mutations in each ubiquitin monomer. This is a challenging undertaking, but would allow direct chemical control of the ubiquitin branch points.

### 6.3. Mechanistic insights into the stimulation of Dot1L by uH2B

Prior to the initiation of the work in this thesis, three models had been proposed to explain the stimulatory effect of uH2B on methylation of H3K79 observed *in vivo* (Figure 3.1). In the first model, ubiquitin acts as a wedge to open up chromatin structure and increase accessibility of H3K79 to Dot1L. In chapter 2, native uH2B was used to demonstrate a direct stimulation of Dot1L-mediated methylation of H3K79 within a nucleosome (Figure 3.6). This does not rule out a role for the wedge model, but minimally suggests that another mechanism contributes to uH2B-mediated Dot1L activation. Interestingly, using the disulfide-directed

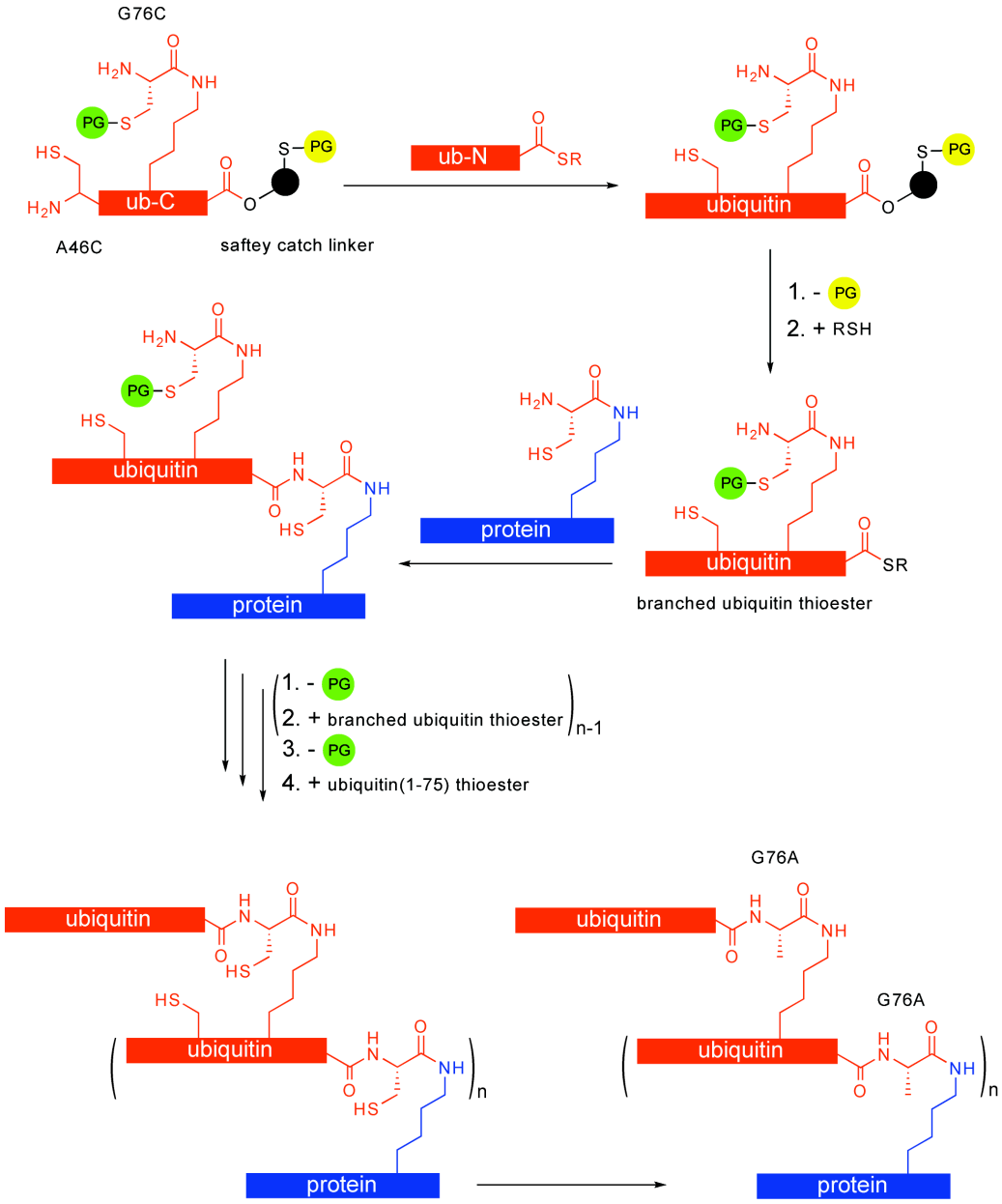



**Figure 6.3. Synthetic approach for the preparation of a polyubiquitylated protein.** A peptide containing residues 46-75 of ubiquitin, an A46C mutation, and an isopeptide linked protected cysteine conjugated to K48 or K63 is synthesized with a safety catch linker (ub-C). Ligation to a synthetic peptide-α-thioester containing residues 1-45 of ubiquitin (ub-N) is followed by removal of the linker protecting group and formation of the branched ubiquitin(1-75)-α-thioester. This thioester is ligated to a semisynthetic protein bearing an isopeptide-linked cysteine. Sequential deprotection of the isopeptide-linked cysteine and ligation to the branched ubiquitin thioester allows ubiquitin polymerization. The final ubiquitin is ligated using the linear ubiquitin(1-75)-α-thioester. Desulfurization converts C46 of each ubiquitin monomer to the native alanine and each C76 to the G76A mutant affording the G76A polyubiquitylated protein.

ubiquitylation strategy, uH2Bss was incorporated into nucleosome arrays and shown to prevent $Mg^{+2}$-dependent compaction of the 30 nm fiber analogous to acetylated H4K16.[9] Ongoing experiments are underway to investigate the effect of this structural change on the accessibility of H3K79 to Dot1L.

The second model proposes that Dot1L recognizes ubiquitin in the context of the nucleosome and is thereby recruited specifically to nucleosomes containing uH2B. However, in electrophoretic mobility shift assays, $Dot1L_{cat}$ interacts with the nucleosome regardless of its ubiquitylation state (Figure 3.10). This interaction was observed with nucleosome concentrations at and below those used in methyltransferase assays. Additionally, the interaction of $Dot1L_{cat}$ with uH2B containing nucleosomes was easily competed with unmodified nucleosomes. This is consistent with *in vivo* results demonstrating that Dot1 associates with chromatin irrespective of H2B ubiquitylation in *S. cerevisiae* (Shahbazian et al., 2005). This suggests that there is not a significant difference in the affinity of Dot1L for ubiquitylated- relative to unmodified nucleosomes. It is formally possible that in the absence of uH2B, Dot1L forms a non-productive complex with the nucleosome at a site or an ensemble of sites distinct from the productive site including H3K79. When H2B is ubiquitylated, the productive interaction could become increasingly favored. However, it seems unlikely that the distinct interaction surfaces would act competitively in a gel-shift assay.

The third model invokes one or more secondary factors responsible for translating the effect of uH2B into methylation of H3K79. Given that Dot1L and $Dot1L_{cat}$ (Figures 3.6 and 3.7), prepared from Sf9 and *E. coli* cells, respectively, are both stimulated by uH2B, it is extremely unlikely that an extra factor present in both enzyme preparations is responsible for linking uH2B

[9] This experiment was performed by Dr. Beat Fierz in the Laboratory of Synthetic Protein Chemistry at the Rockefeller University.

to H3K79 methylation. Therefore, while other proteins including Cps35 and proteasomal ATPases may enhance the stimulation of Dot1L by uH2B, they are not required. It would be interesting to test whether or not the presence of Cps35 enhances activity of Dot1L *in vitro* in an uH2B-dependent manner. Thus far, efforts to express the human ortholog of Cps35, WDR82, in a soluble form in *E. coli* have been unsuccessful. Expression using baculovirus in Sf9 cells may be more fruitful. Furthermore, WDR82 is a component of the human Set1-containing H3K4 methyltransferase complex, that we have also shown to be stimulated by semisynthetic uH2B *in vitro* (Kim et al., 2009). It is possible that the Set1 complex and Dot1L work synergistically, through WDR82, in performing simultaneous methylation of H3K4 and K79. This is readily testable in our *in vitro* system.

As each of the previously proposed models is insufficient to explain the stimulation of Dot1L by semisynthetic uH2B in our experiments, new models are required. It seems likely that ubiquitylation of H2B results in an allosteric change in the nucleosome or Dot1L, either increasing the accessibility of H3K79 to Dot1L or the intrinsic catalytic activity of Dot1L itself (Figure 3.11). The activity of Dot1L was unaffected by the G76A mutation. Therefore, we used nucleosomes containing u(G76A)H2B to perform the first systematic mechanistic investigation into the methylation process. Kinetic experiments performed with a large excess of Dot1L revealed an impressive enhancement in the chemical step of the methylation reaction due to the presence of ubiquitylated H2B. This suggests a conformational change in the active site complex of Dot1L bound to H3K79. Based on structure activity relationship analysis, we propose that the interaction of a non-canonical surface of ubiquitin (i.e. not including I44) with the catalytic domain of Dot1L or the nucleosome itself, initiates this allosteric change, favoring methylation (Figure 3.11). From our experiments it cannot be concluded if the allostery occurs within the

nucleosome, Dot1L, or both. If ubiquitylation of H2B allosterically regulates Dot1L, a composite interaction surface of ubiquitin and the nucleosome is likely to be required, as ubiquitin alone is insufficient to stimulate methyltransferase activity in *trans*. In this case, ubiquitylation of H2B will likely stimulate methylation on the same nucleosomal surface. If ubiquitylation alters the nucleosomal structure, this change is only propagated to one copy of H3 within the nucleosome and could reasonably stimulate methylation of either nucleosomal surface. Structural investigation of this complex system, aided by semisynthetic u(G76A)H2B, may shed further light on this aspect of the mechanism.

Current work is ongoing to further elucidate the structural requirements of ubiquitin for Dot1L stimulation. Using the disulfide-directed strategy, two different Ubls, Nedd8 and Hub1, which have 55 and 26% sequence identity to ubiquitin, respectively, were attached to H2BK120.[10] While Nedd8 shows nearly complete stimulation of Dot1L relative to ubiquitin, Hub1 fails to stimulate Dot1L. In conjunction with the structure activity relationship analysis presented in chapter 5, this allows the residues of ubiquitin that are potentially required for Dot1L stimulation to be focused – that is, residues shared between ubiquitin and Nedd8 that are not shared with Smt3 or Hub1 (Figure 6.4). Future work, involving alanine-scanning mutagenesis coupled to methyltransferase assays of these remaining residues, will define the required binding surface. It is surprising that disruption of the canonical binding surface of ubiquitin fails to diminish Dot1L activity (Figure 5.5). This raises the possibility that multiple surfaces of ubiquitin are interpreted simultaneously by parallel pathways, thereby necessitating the recognition of non-canonical surfaces. This could explain the evolutionary need for such a large modification in this instance. Further investigation of the role of the surfaces of uH2B in

[10] This work was performed by Dr. Champak Chatterjee in the Laboratory of Synthetic Protein Chemistry at the Rockefeller University.

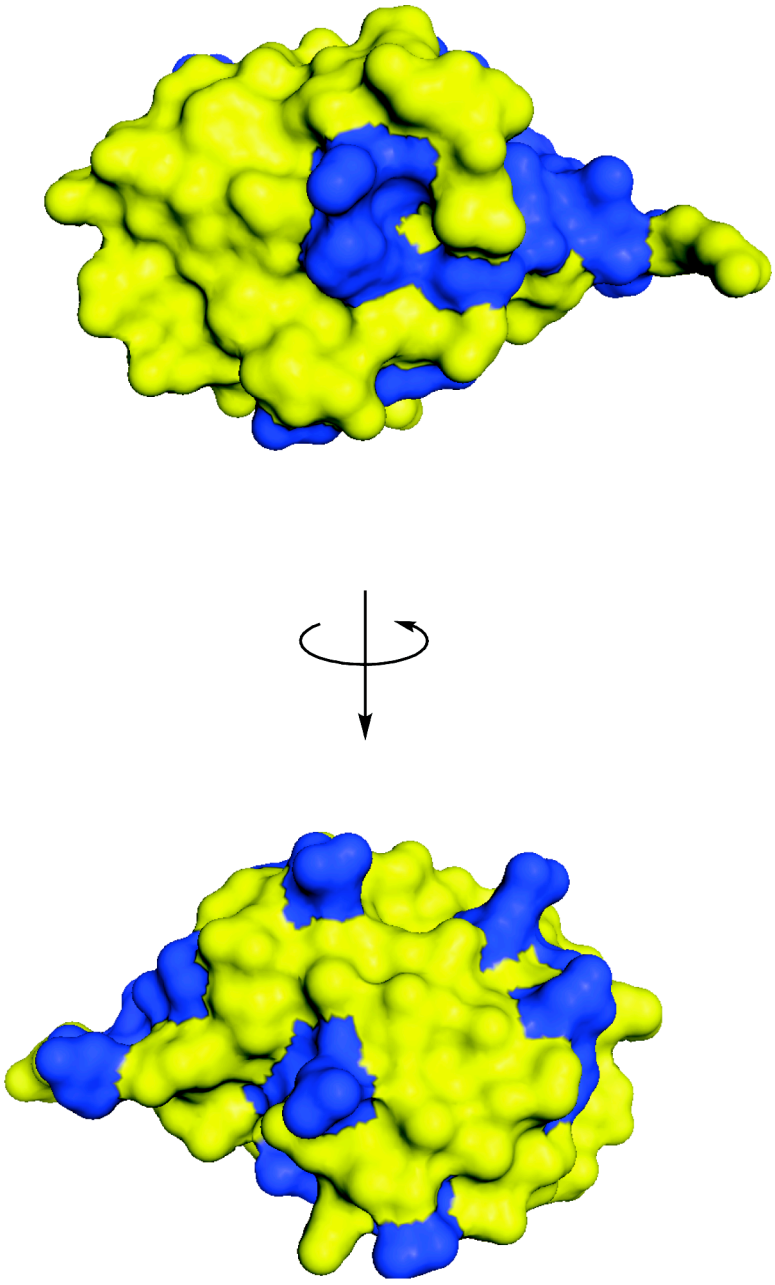


**Figure 6.4. Elements of ubiquitin potentially required for Dot1L stimulation.** A surface representation of ubiquitin (1UBQ) (Vijay-Kumar et al., 1987). Amino acids conserved between ubiquitin and Nedd8 (not including L8 and I44) and not conserved between ubiquitin and Smt3 or Hub1 based on structural alignment, are colored blue. These residues are potentially required for Dot1L stimulation. All other amino acids are colored yellow and unlikely to be required for Dot1L stimulation. Conservation based on homology.

H3K79 methylation and its myriad of other downstream processes, using the structural variants described above, will put this hypothesis to the test.

In the case of H3K79 methyltransferase activity, the non-canonical surface of ubiquitin could be required for interactions with the nucleosome, Dot1L, or both. An interaction with the nucleosome could be investigated by crystallography as attempted in chapter 5. Alternatively, NMR using a segmentally-labelled uH2B could be employed. Chemical shift perturbations in ubiquitin attached to the nucleosome relative to free ubiquitin or free uH2B would be suggestive of an intranucleosomal interaction. Moreover, because the chemical shifts of ubiquitin are assigned, NMR results could be directly compared with mutagenesis results. In contrast, an interaction of ubiquitin or uH2B with Dot1L is more difficult to interrogate. It could be studied qualitatively using pull-down experiments or quantitatively by surface plasmon resonance (SPR), fluorescence anisotropy, or isothermal titration calorimetry (ITC). However, given that 1 mM exogenous ubiquitin fails to diminish Dot1L activity toward ubiquitylated nucleosomes (Figure 5.9), it is likely that any direct interaction with ubiquitin will be of low affinity and difficult to detect. Any potential interaction with Dot1L may only be productive when ubiquitin is constrained to the nucleosomal surface. In fact, introduction of three extra glycine residues at the C-terminus of ubiquitin using the disulfide-directed technology significantly diminishes stimulation of Dot1L, suggesting a strong entropic penalty in the system.[11] As such, it may be best to interrogate a potential Dot1L-ubiquitin interaction within the context of the nucleosome. A $Dot1L_{cat}$-nucleosome co-crystal structure would be most revealing in this case, though perhaps the most challenging undertaking. It is possible that monitoring NMR chemical shift

[11] This experiment was performed by Dr. Champak Chatterjee in the Laboratory of Synthetic Protein Chemistry at the Rockefeller University.

perturbations while titrating $Dot1L_{cat}$ into a segmentally-labelled uH2B nucleosome could also give insight into this potential mechanism.

### 6.4. uH2B and the degree of H3K79 methylation

uH2B increases Dot1L- and $Dot1L_{cat}$-mediated mono- and dimethylation of H3K79 *in vitro* (Figure 3.8). This is contradictory to *in vivo* results in which monomethylation of H3K79 is unaffected by disruption of H2B ubiquitylation (Kim et al., 2005; Shahbazian et al., 2005). Steady-state kinetic analysis of Dot1L monomethyltransferase activity on ubiquitylated nucleosomes allowed for the extrapolation of kinetic parameters from the Michaelis-Menten model. The observed $K_m$ is similar to values calculated for other histone lysine methyltransferases acting on peptide and protein substrates (Chin et al., 2006; Collazo et al., 2005; Dirk et al., 2007; Eskeland et al., 2004; Patnaik et al., 2004; Trievel et al., 2002); however, the $k_{cat}$ of Dot1L on ubiquitylated nucleosomes is 10 to 100 times greater than reported values for these other reactions, suggesting a rate enhancement on nucleosomal substrates. Similar to yeast Dot1 (Frederiks et al., 2008), Dot1L appears to be a distributive enzyme, as evidenced by its ability to monomethylate an excess of substrate without measurable accumulation of dimethylation. Although a direct comparison of kinetic parameters of Dot1L on ubiquitylated and unmodified nucleosomes was not possible, it is clear that Dot1L is able to monomethylate uH2B bearing nucleosomes to a much greater extent than unmodified nucleosomes. Again, this seems contradictory to *in vivo* experiments where levels of H3K79 monomethylation are apparently unaffected by genetic disruption of H2B ubiquitylation (Kim et al., 2005; Shahbazian et al., 2005). We propose that accumulation over time or parallel pathways may compensate

specifically for any deficit in monomethylation of H3K79 in the absence of H2B ubiquitylation *in vivo*.

Dimethylation was never observed on unmodified nucleosomes in our assays. This could be a consequence of the low activity of Dot1L. In this distributive system, monomethylation must build up to significant levels before dimethylation can be detected. Alternatively, it is possible that conversion of monomethylated H3K79 to the dimethylated state is precluded in the absence of uH2B. This could be tested directly using nucleosomes bearing chemically monomethylated H3K79. The cysteine-directed methyllysine analog strategy (Simon et al., 2007) would be best suited for this purpose.

**6.5. Asymmetric nucleosome formation**

By titrating the level of u(G76A)H2B within nucleosomes, it was demonstrated that one ubiquitylated H2B leads to methylation of only one nucleosomal H3K79 (Figure 5.12). As described above, elucidating the side of the nucleosome that is methylated, relative to ubiquitin, may provide insight into the mechanism underlying Dot1L stimulation. This is a surprisingly difficult question to answer. While affinity purification strategies exist to assemble nucleosomes simultaneously bearing two different copies of a given histone protein (Li and Shogren-Knaak, 2008), these strategies cannot direct the relative position of two histones within the nucleosome. But perhaps, complementary mutations could be introduced into histone-histone interfaces to ensure that two or more given histones are localized to the same surface of the nucleosome – not unlike the elegant bump-hole approach to generate orthogonal kinase-ATP pairs (Liu et al., 1998; Shah et al., 1997). For example, one could imagine the introduction of mutations in H3 and H4 which alone prevent H3-H4 interaction, but when combined restore this interaction

(Figure 6.5). The reconstitution of nucleosomes with a mixture of the mutant and wild-type proteins would result in the defined placement of the mutants relative to each other within the nucleosome. This could potentially be performed throughout the nucleosome until a system existed where the position of each histone protein within the nucleosome could be directly predicted, effectively removing any stoichasticity from nucleosomal assembly. This would allow the placement of uH2B on the same or opposite surface relative to an H3K79R mutant to definitely illustrate the spatial nature of uH2B/H3K79me stoichiometry.

While this may seem a lofty goal, nature has provided evidence of its practicality; the histone variant, macroH2A preferentially forms asymmetric nucleosomes with one copy of macroH2A and one copy of the canonical H2A in *in vitro* reconstitution experiments (Chakravarthy and Luger, 2006). A priori, it is difficult to predict the magnitude of mutations necessary to establish such a complementation system. A stretch of four amino acids is critical in the case of macroH2A-H2A hybrid nucleosomes, but this is likely to be interface dependent. Although these mutations could be designed rationally through close inspection of the nucleosome structure, a high throughput screen might be more promising. This screen could be undertaken using the histone shuffle system in *S. cerevisiae* (Ahn et al., 2005). This system readily allows the replacement of the canonical histone sequences with mutant sequences. Targeted mutations in one histone at a desired interface could be made and inviable mutants screened for complementation with a library of mutants of the interacting histone. Complementary mutant pairs with observable phenotypes could be excluded. The remaining pairs could be sequenced and the behavior of the pairs containing the subtlest changes could be tested *in vitro*. Ideally, complementary mutations would function without significant effects on nucleosome stability and function.

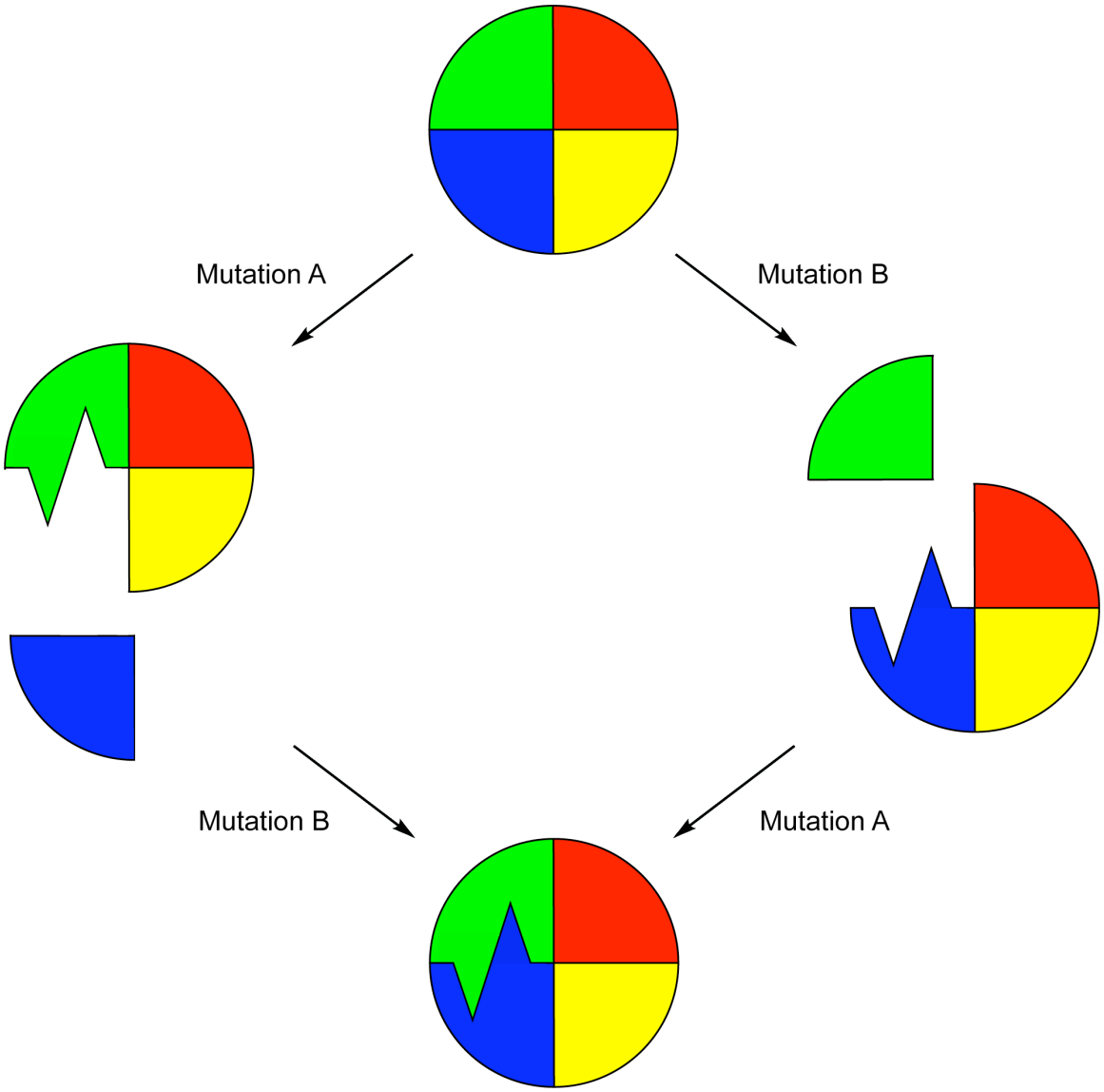



**Figure 6.5. Complementary mutants for preparing asymmetric nucleosomes.** Complementary mutations are introduced into a histone-histone interface. Each mutant alone prevents nucleosome formation. When these mutants are introduced simultaneously, the requisite interaction surface is restored.

### 6.6. Preparation of custom nucleosomal arrays

A strategy allowing the asymmetric ligation of nucleosomes was presented in chapter 3. Structurally distinct nucleosomes were reconstituted bearing DNA linkers with complementary non-palindromic overhangs (Figure 3.9). The non-palindromic nature of the DNA overhangs, allowed ligation of preassembled nucleosomes to afford dinucleosomes with defined components at each position. This approach was used to demonstrate that the stimulation of Dot1L by uH2B is strictly intranucleosomal.

This technology could be readily applied to the preparation of longer arrays of nucleosomes. This could be accomplished using two approaches. The first would involve the simultaneous ligation of multiple nucleosomes each bearing orthogonal overhangs to define nucleosome order (Figure 6.6, top panel). The advantage of this approach is expediency, ensuring nucleosomal stability during ligation. However, obtaining homogeneous samples may be challenging due to incomplete ligations. A second approach, invoking sequential chain assembly on a solid support, analogous to solid phase peptide synthesis, could overcome this limitation (Figure 6.6, bottom panel). Chain assembly begins with a nucleosome biotinylated on one side of the DNA. Subsequent rounds of digestion and ligation with intervening washes would allow the sequential assembly of nucleosomes. An enzyme such as SapI, which contains a recognition sequence completely removed by digestion could be used universally, similar to Fmoc- or Boc-based protection strategies in peptide synthesis. Following completion of chain assembly, the nucleosome array could be used directly on the solid support, or cleaved using a unique restriction site between the first nucleosome and the biotin handle. This strategy would require more time for array synthesis, but under appropriately low salt conditions, histone rearrangement could be limited.

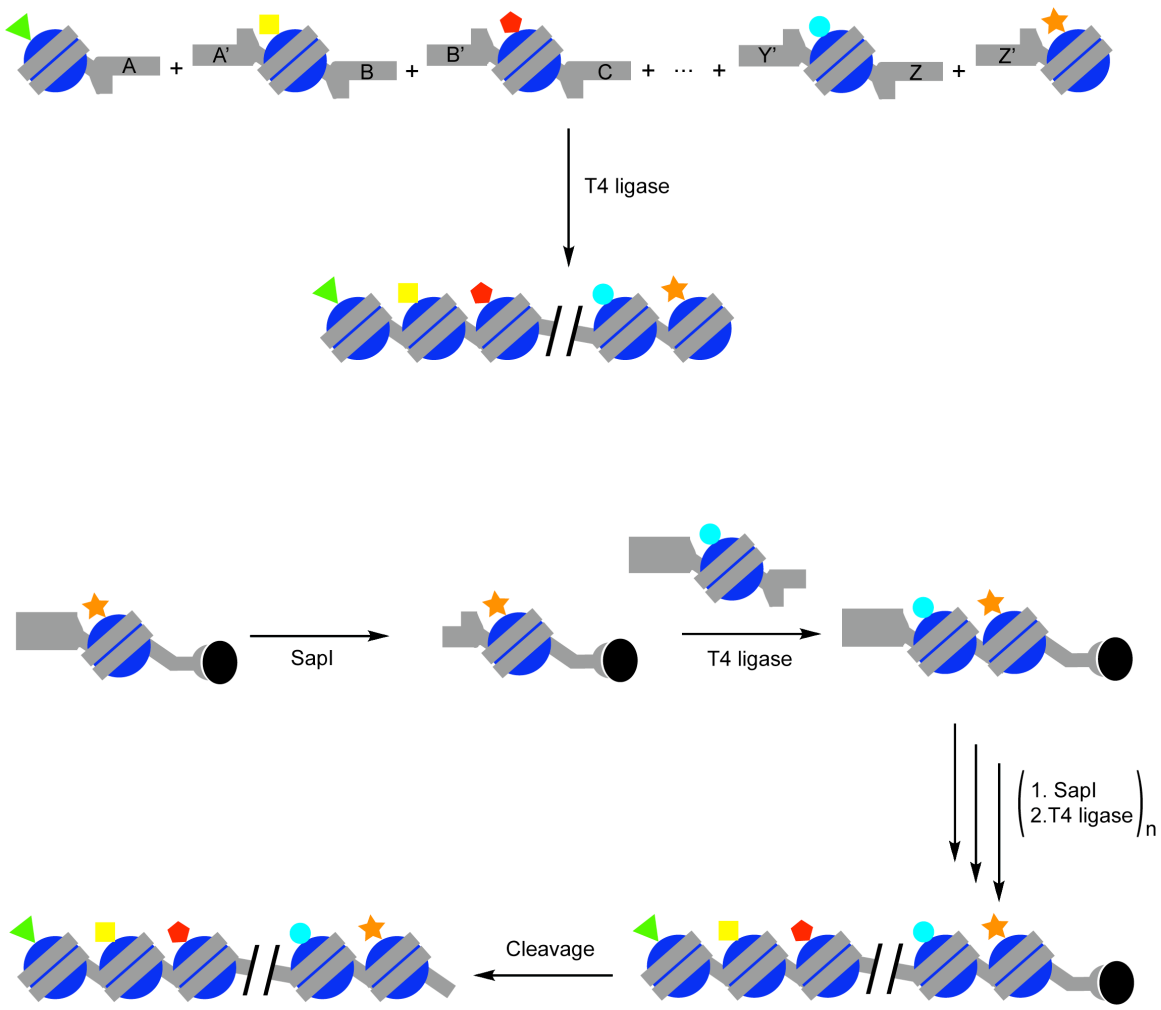



**Figure 6.6. Strategies for the preparation of custom nucleosome arrays.** Nucleosome arrays with position-specific structural differences could be prepared using two different strategies. In a one-pot method (top panel), nucleosomes containing non-palindromic, orthogonal, complementary DNA overhangs could be ligated simultaneously. X and X' represent complementary sequences. Alternatively, a sequential chain assembly could be used (bottom panel). Rounds of digestion and ligation on a solid support would allow the stepwise assembly of a nucleosome array. Following array completion, cleavage of the DNA sequence between the solid support and the first nucleosome releases the array into solution.

Combined with the bump hole strategy for asymmetric nucleosome formation, this would allow the preparation of nucleosome arrays with unprecedented chemical control. These arrays could be used to interrogate the roles of histone PTMs in chromatin-templated processes in a position-specific manner. As understanding of histone PTMs at the nucleosome level grows, this positional chemical control will become increasingly valuable.

### 6.7. Perspective

Using novel semisynthetic strategies to prepare ubiquitylated H2B in combination with biochemical analysis, a direct role for uH2B in the intranucleosomal stimulation of Dot1L-mediated H3K79 methylation was demonstrated. However, while previously proposed mechanisms were ruled out, much remains unclear about the specific role of uH2B in this complex system. Experiments proposed above will help narrow the remaining possibilities, heightening mechanistic understanding of this *trans*-histone crosstalk pathway. As both H2B ubiquitylation and H3K79 methylation are integral to the control of gene expression and DNA damage repair, a full elucidation of their functions serves to advance our understanding of the roles of histone PTMs in development and disease. Moreover, given the pathogenic role of H3K79 methylation in a large subset of human leukemias, a more complete understanding of this pathway will undoubtedly unearth potential therapeutic approaches to combat MLL-fusion leukemias.

More broadly, we propose that the application of the semisynthetic technologies described within this thesis presents an invaluable tool in the isolation of functions of individual histone PTMs. As more is elucidated at the single PTM level, combinations of PTMs can be introduced simultaneously to interrogate cooperative roles in chromatin-templated processes. We

anticipate this combinatorial approach to engineer defined chromatin substrates with increasing complexity, will play a crucial role in the hypothesis-driven dissection of the mechanisms underlying a potential histone code (Strahl and Allis, 2000).

## Methods

### General methods

Amino acid derivatives unless otherwise noted, pre-loaded Wang resin, and coupling reagents were purchased from Novabiochem. *E. coli* BL21(DE3) and pLysS cells were purchased from Novagen. $^{3}$H *S*-adenosyl methionine, was obtained from GE Healthcare. Restriction enzymes, T4 ligase, pTXB1 vector, and chitin resin were obtained from New England Biolabs. Criterion 15% and 4-20% Tris HCl, and Criterion 5% TBE gels were purchased from Biorad. Centricons were purchased from Sartorius. PCR purification and gel extraction kits were purchased from Qiagen. All other chemical reagents unless otherwise noted were purchased from Sigma-Aldrich or Fisher Scientific. Analytical and semi-preparative scale reverse-phase HPLC (RP-HPLC) were performed on a Hewlett-Packard 1100 series instrument using Vydac C18 columns (4 x 150 mm; 10 x 250 mm) at 1 and 4 mL/min, respectively. Unless otherwise noted, all analytical gradients were 0-73% B over 30 min (A: 0.1% trifluoroacetic acid (TFA) in water; B: 90% acetonitrile, 0.1% TFA in water). Preparative and process scale RP-HPLC were performed on a Waters DeltaPrep 4000 system connected to a Waters 486 tunable detector using Vydac C18 columns (22 x 250 mm; 50 x 250 mm) at 15 and 30 mL/min, respectively. Size-exclusion, ion exchange, and GSH affinity chromatography were performed on an AKTA FPLC system from GE Healthcare equipped with a P-920 pump and a UPC-900 monitor. ESI-MS was performed on a SciexAPI-100 single quadrupole mass spectrometer. Linear peptide synthesis was performed on a Liberty synthesizer equipped with a Discovery microwave module from CEM. Primer synthesis and DNA sequencing were performed by Integrated DNA Technologies and Genewiz, respectively.

**Synthesis of photolytically removable ligation auxiliary**

The ligation auxiliary, **1**, was synthesized as previously described with minor modifications (McMinn and Greenberg, 1996; Pellois and Muir, 2005).

***4-(4-Formyl-2-methoxyphenoxy)-butyric acid methyl ester (3).*** To a solution of vanillin, **2** (22.8 g, 150 mmol), in acetonitrile (400 mL), was added potassium carbonate (49.8 g, 360 mmol), tetrabutylammonium iodide (11.1 g, 30 mmol), and 4-chlorobutyric acid methyl ester (21.95 mL, 180 mmol) and the mixture was refluxed under argon for 20 h. The mixture was filtered and the precipitate was washed with ethyl acetate (4 x 100 mL). Combined organic layers were concentrated *in vacuo* and redissolved in diethyl ether (300 mL), washed with water (2 x 50 mL), and dried *in vacuo* to afford ester **3** (15.5 g, 62 mmol, 41%). Compound **3**: $R_f$ = 0.38 (silica gel, 33% ethyl acetate in hexanes); $^1$H NMR (400 MHz, $CDCl_3$): δ 9.86 (s, 1H, C(O)H), 7.45 (s, 1H, $H_{ar}$), 7.42 (s, 1H, $H_{ar}$), 7.00 (s, 1H, $H_{ar}$), 4.18 (t, 2H, $OCH_2$), 3.93 (s, 3H, $OCH_3$), 3.71 (s, 3H, $COOCH_3$), 2.56 (m, 2H, $CH_2CO$), 2.13 (m, 2H, $CH_2$).

***4-(4-Formyl-2-methoxy-5-nitrophenoxy)-butyric acid methyl ester (4).*** To a stirred mixture of glacial acetic acid (125 mL) and fuming nitric acid (40 mL) on ice, was added dropwise 4-(4-formyl-2-methoxyphenoxy)-butyric acid methyl ester, **3** (15.5 g, 62 mmol), in glacial acetic acid (35 mL). The mixture was stirred on ice for 10 min and at room temperature for an additional 6 h, after which the mixture was poured over ice (150 mL), extracted with diethyl ether (4 x 150 mL), washed with saturated sodium bicarbonate (~10 x 300 mL), brine (2 x 300 mL), dried ($MgSO_4$), and concentrated *in vacuo*. The residue was crystallized from diethyl ether, affording nitrated product **4** (10.7 g, 36 mmol, 58%). Compound **4**: $R_f$ = 0.29 (silica gel, 30% ethyl acetate in hexanes); $^1$H NMR (400 MHz, $CDCl_3$): δ 9.86 (s, 1H, C(O)H), 7.45 (s, 1H, $H_{ar}$), 7.42 (s, 1H, $H_{ar}$), 7.00 (s, 1H, $H_{ar}$), 4.18 (t, 2H, $OCH_2$), 3.93 (s, 3H, $OCH_3$), 3.71 (s, 3H, $COOCH_3$), 2.56 (m,

2H, $CH_2CO$), 2.13 (m, 2H, $CH_2$).

***4-(2-Methoxy-5-nitro-4-vinyl-phenoxy)-butyric acid methyl ester (5).*** To a suspension of methyltriphenylphosphonium bromide (6.5 g, 18.2 mmol) in THF (30 mL), a solution of sodium hexamethyldisilazide (18.2 mL, 1 M in THF, 18.2 mmol) was added dropwise over 30 min at 0 °C and stirred for 1 h. A solution of 4-(4-formyl-2-methoxy-5-nitrophenoxy)-butyric acid methyl ester, **4** (4.2 g, 14.0 mmol), in THF (40 mL) was then added dropwise and the mixture stirred at room temperature for an additional 16 h. The solvent was then evaporated and the residue dissolved in chloroform (25 mL). The organic phase was washed with a saturated solution of ammonium chloride (2 x 25 mL), brine (2 x 300mL), dried ($Na_2SO_4$) and concentrated *in vacuo*. Flash column chromatography of the residue (silica gel, 30% ethyl acetate in petroleum ether) yielded styrene **5** (2.4 g, 8.1 mmol, 58%). Compound **5**: $R_f$ = 0.54 (silica gel, 30% ethyl acetate in petroleum ether); $^1$H NMR (400 MHz, DMSO-$d_6$): δ 7.56 (s, 1H, $H_{ar}$), 7.19 (s, 1H, $H_{ar}$), 7.09 (dd, $J$ = 17.3, 11 Hz, 1H, $H_2$), 5.88 (d, $J$ = 17.3 Hz, 1H, $H_{1b}$), 5.45 (d, $J$ = 11.1 Hz, 1H, $H_{1a}$), 4.14 (t, $J$ = 6.3 Hz, 2H, $OCH_2$), 3.94 (s, 3H, $OCH_3$), 3.61 (s, 3H, $COOCH_3$), 2.48 (m, 2H, $CH_2CO$), 2.01 (m, 2H, $CH_2$). $^{13}$C NMR (400 MHz, DMSO-$d_6$): δ 173.75, 153.94, 148.17, 140.78, 133.02, 128.00, 119.15, 110.64, 109.35, 68.74, 57.18, 52.22, 30.69, 24.83. HRMS (FAB); m/z: calculated for $C_{14}H_{18}O_6N$: 296.1134; found: 296.1135 ([$MH^+$]).

***4-[4-(1-tert-Butoxycarbonylamino-2-hydroxy-ethyl)-2-methoxy-5-nitro-phenoxy]-butyric acid methyl ester (6).*** To a stirred solution of tert-butylcarbamate (2.0 g, 16.8 mmol) in n-propanol (20 mL), sodium hydroxide (33 mL 0.5 M in water, 16.5 mmol) and tert-butyl hypochlorite (1.9 mL) were added and the resulting mixture was stirred for 5 min. The flask was then put in a

slurry of ice water and $(DHQ)_2PHAL$ (211 mg, 0.27 mmol, in 20 mL of n-propanol), styrene **5** (1.6 g, 5.4 mmol, in 80 mL of n-propanol), and $K_2OsO_2(OH)_4$ (70 mg, 0.2 mmol) were sequentially added. The resulting mixture was stirred for 16 h at 4 °C. The product was extracted with ethyl acetate (2 x 100 mL), the combined organic extracts washed with brine (2 x 500 mL), dried ($MgSO_4$), and evaporated. Flash column chromatography of the residue (silica gel, 5% methanol in dichloromethane (DCM)) followed by crystallization in chloroform/hexanes afforded alcohol **6** (1.2 g, 2.7 mmol, 50%). Compound **6**: $R_f$ = 0.38 (silica gel, 5% methanol in DCM); $^1$H NMR (400 MHz, $CDCl_3$): δ 7.63 (s, 1H, $H_{ar}$), 7.02 (s, 1H, $H_{ar}$), 5.62 (s, 1H, NH), 5.46 (m, 1H, CHNH), 4.12 (t, *J* = 6.0 Hz, 2H, $OCH_2$), 3.96 (m, 2H, $CH_2OH$), 3.95 (s, 3H, $OCH_3$), 3.71 (s, 3H, $COOCH_3$), 3.50 (s, 1H, OH), 2.56 (t, *J* = 7.2 Hz, 2H, $CH_2CO$), 2.19 (q, *J* = 6.7 Hz, 2H, $CH_2$), 1.42 (s, 9H, tBu). $^{13}$C NMR (400 MHz, $CDCl_3$): δ 173.74, 155.81, 154.16, 147.58, 68.66, 65.39, 56.80, 53.68, 52.12, 51.28, 30.76, 28.69, 24.66. HRMS (FAB); m/z: calculated for $C_{19}H_{29}O_9N_2$: 429.1873; found: 429.1872 ([$MH^+$]).

***4-[4-(2-Acetylsulfanyl-1-tert-butoxycarbonylaminoethyl)-2-methoxy-5-nitro-phenoxy]-butyric acid methyl ester (7).*** Diisopropyl azodicarboxylate (0.92 mL, 4.7 mmol) was added to an ice-cold solution of triphenylphosphine (1.2 g, 4.7 mmol) in THF (30 mL). The mixture was stirred under nitrogen for 30 min, during which time a white precipitate formed. A solution of alcohol **6** (1.0 g, 2.3 mmol) in THF (20 mL) and thioacetic acid (0.33 mL, 4.7 mmol) was added dropwise and the mixture was stirred for 17 h at room temperature. The solvent was evaporated, the residue dissolved in methanol (200 mL) and left to stand at 0 °C for 16 h. The precipitate was filtered and the solvent evaporated. Flash column chromatography of the residue (silica gel, 40% ethyl acetate in petroleum ether) afforded thioester **7** (1.0 g, 2.1 mmol, 91%). Compound **7**: $R_f$ = 0.50 (40% ethyl acetate in petroleum ether); $^1$H NMR (400 MHz, $CDCl_3$): δ 7.63 (s, 1H, $H_{ar}$),

7.01 (s, 1H, $H_{ar}$), 5.69 (s, 1H, NH), 5.00 (m, 1H, CHNH), 4.12 (t, *J* = 6.0 Hz, 2H, $OCH_2$), 3.98 (s, 3H, $OCH_3$), 3.71 (s, 3H, $COOCH_3$), 3.35 (m, 2H, $CH_2S$), 2.57 (t, *J* = 7.2 Hz, 2H, $CH_2CO$), 2.40 (s, 3H, $SCOCH_3$), 2.20 (q, *J* = 6.7 Hz, 2H, $CH_2$), 1.41 (s, 9H, tBu). $^{13}C$ NMR (400 MHz, $CDCl_3$): δ 197.78, 173.76, 155.81, 154.34, 147.73, 139.37, 70.53, 68.65, 56.88, 52.13, 33.93, 30.94, 30.76, 28.67, 24.66, 22.34. HRMS (FAB); m/z: calculated for $C_{21}H_{31}O_9N_2S$: 487.1750; found: 487.1749 ([$MH^+$]).

***4-[4-(1-tert-Butoxycarbonylamino-2-tert-butyldisulfanylethyl)-2-methoxy-5-nitro-phenoxy]-butyric acid (8).*** To a solution of thioester **7** (500 mg, 1.0 mmol) in methanol (30 mL) was added sodium methoxide (2.0 mL, 0.5 M in methanol, 1.0 mmol). The mixture was stirred under nitrogen for 15 min. 2-methyl-2-propanethiol (3.4 mL, 30 mmol) was added and the mixture stirred under constant oxygen bubbling overnight. A solution of sodium hydroxide (15 mL, 1.0 M in water, 15 mmol) was then added and the mixture stirred for another 3 h. The solution was acidified with hydrochloric acid (20 mL, 1.0 M in water, 20 mmol) and extracted with ethyl acetate (2 x 50 mL). The organic phase was washed with brine (2 x 25 mL), dried ($MgSO_4$), and concentrated *in vacuo*. Flash column chromatography of the residue (silica gel, 40% ethyl acetate in petroleum ether) afforded disulfide **8** (467 mg, 0.9 mmol, 90%). Compound **8**: $R_f$ = 0.26 (silica gel, 40% ethyl acetate in petroleum ether); $^1H$ NMR (400 MHz, $CDCl_3$): δ 7.62 (s, 1H, $H_{ar}$), 7.03 (s, 1H, $H_{ar}$), 5.56 (m, 1H, CHNH), 4.14 (t, *J* = 6.0 Hz, 2H, $OCH_2$), 3.95 (s, 3H, $OCH_3$), 3.30 (d, *J* = 13.0 Hz, 2H, $CH_2S$), 2.60 (t, *J* = 7.2 Hz, 2H, $CH_2CO$), 2.20 (m, 2H, $CH_2$), 1.41 (s, 9H, tBu), 1.34 (s, 9H, *S*-tBu). $^{13}C$ NMR (400 MHz, $CDCl_3$): δ 177.70, 154.26, 147.56, 140.72, 132.62, 110.48, 100.00, 68.55, 56.84, 51.19, 48.73, 30.60, 30.24, 30.18, 28.66, 24.66. HRMS (FAB); m/z: calculated for $C_{22}H_{35}O_8N_2S_2$: 519.1835; found: 519.1837 ([$MH^+$]).

***4-[4-(1-Amino-2-tert-butyldisulfanyl-ethyl)-2-methoxy-5-nitro-phenoxy]-N-methyl-butyramide (1).*** To a stirred solution of thioester **8** (438 mg, 0.844 mmol) in DCM (10 mL) at room temperature was added PyBOP (439 mg, 0.844 mmol) and *N,N*-diisopropylethylamine (DIEA) (0.43 mL, 2.53 mmol). After an additional 5 min of stirring, methylamine (4.22 mL, 8.44 mmol, 2M in THF) was added to the mixture and the reaction was allowed to proceed for a further 10 h at room temperature, affording amide **9**. At this stage, the solvent was removed *in vacuo* and the residue dissolved in a mixture of TFA (9.5 mL), triisopropylsilane (TIS) (0.25 mL), and water (0.25 mL). The resulting solution was stirred at room temperature for 1 h to effect complete removal of the Boc-protecting group. The removal of water by lyophilization yielded crude amide **1**, which was purified by preparative RP-HPLC using a 20-45% B gradient over 60 min, to obtain the pure compound in 36% isolated yield over two steps. $^{1}$H NMR (400 MHz, $CDCl_3$): δ 7.55 (s, 1H, $H_{ar}$), 7.29 (s, 1H, $H_{ar}$), 6.13 (m, 1H, $NHCH_3$), 5.35 (m, 1H, $CHNH_2$), 4.12 (t, 2H, $OCH_2$), 4.11 (s, 3H,$OCH_3$), 3.33 (d, 2H, $CH_2S$), 2.78 (d, 3H, $CH_3NH$), 2.41 (t, 2H, $CH_2CO$), 2.15 (m, 2H,$CH_2$), 1.33 (s, 9H, *S*-tBu). $^{13}$CNMR(400 MHz, $CDCl_3$): δ 173.90, 154.51, 148.62, 141.78, 124.74, 111.11, 109.78, 69.06, 57.08, 49.06, 42.46, 32.96, 30.12, 26.72, 25.12, 22.29. HRMS (ESI-FT); m/z calculated for $C_{18}H_{29}O_5N_3S_2$: 432.1622, found 432.1609 ([$MH^+$]).

**Solid phase peptide synthesis**

***Synthesis of peptide 14***. The peptide corresponding to the C-terminus of human H2B with an A117C mutation, TKCVTKYTSSK, was manually synthesized on a 0.5 mmol scale employing the 9-fluorenylmethoxycarbonyl (Fmoc) $N^{\alpha}$-protection strategy. Starting with the pre-loaded Fmoc-Lys(Boc)-Wang resin (806 mg, 0.62 mmol/g), each successive amino acid was coupled in 4.4 molar excess. Deprotection of the Fmoc group was achieved with 20% piperidine in *N,N*-dimethylformamide (DMF) (5 mL, 3 x 5 min). Coupling reactions were undertaken for a minimum of

1 h with a mixture of Fmoc-amino acid (2.2 mmol), 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) (2 mmol), and DIEA (4.4 mmol) in DMF. The ε-$NH_2$ of Lys6, which represents Lys120 in full-length H2B, was orthogonally protected with the 1-(4,4-dimethyl-2,6-dioxocyclohexylidine)-3-methylbutyl (ivDde) group. Thr1 was coupled as Boc-Thr(tBu)-OH to preclude deprotection of the α-amino group during removal of the ivDde group. The N-terminal residues Thr1 and Lys2 were not incorporated efficiently upon single coupling and required multiple reactions with aggressive coupling reagents such as benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP), or, diisopropylcarbodiimide (DIC) in the presence of 1-hydroxybenzotriazole (HOBt). $N^{\alpha}$-deprotection times for these residues were also increased to 2 x 20 min followed by 1 x 5 min.

Peptidyl resin **10** was treated with 2% hydrazine in DMF (v/v) for 5 x 30 min, followed by a longer incubation for 8 h to effect removal of the ivDde group from Lys6. The resultant peptidyl resin **11** (550 mg, 0.33 mmol) was then coupled to bromoacetic acid (366.8 mg, 2.64 mmol) with DIC (333 mg, 2.64 mmol) in DMF (4 mL) for 30 min at room temperature. The coupling was repeated once, affording bromoacetylated peptidyl resin **12**, and the resin was dried. After swelling in 1.4 mL DMF, peptidyl resin **12** (250 mg, 0.15 mmol) was reacted with 1.75 equivalents of auxiliary **1** (112.6 mg, 0.261 mmol) in the presence of DIEA (252.3 mg, 1.95 mmol) and 1,8-diazobicycloimide (DBU) (10.2 mg, 0.07 mmol) for 96 h at room temperature with the exclusion of light, yielding peptidyl resin **13**. The reaction progress was followed by test cleavages of 5-10 mg of resin after 48 h, 72 h, and 96 h. After conjugation to the ligation auxiliary, simultaneous cleavage and deprotection in 10 mL TFA:TIS:$H_2O$:Anisole (92.5:2.5:2.5:2.5) for 2 h, afforded peptide **14**, which was purified by semi-preparative RP-HPLC to yield 10.2 mg. Auxiliary-conjugated peptide **14** was verified by ESI-MS [$(M+H)^{2+}$ observed 860 Da. $(M+H)^{2+}$ expected 859.5 Da.].

***Synthesis of peptide 21.*** The sequence corresponding to residues 117-125 of *Xenopus* H2B with an A117C replacement was synthesized on a pre-loaded Wang resin using manual solid-phase peptide synthesis as described above. Standard $^{t}$butyl side-chain protection was used throughout with the following exceptions; the ε-amino group of K120 was protected with the 4-methyltrityl (Mtt) group, and the thiol group of C117 was protected with an o-nitrobenzyl group. $N^{\alpha}$-($^{t}$butoxycarbonyl)-*S*-(*o*-nitrobenzyl)-L-cysteine (Boc-Cys(ONB)) used in the peptide synthesis was prepared as previously described (Smith et al., 2002; Wang et al., 1977). The glycyl linker and ligation auxiliary were installed on the solid phase as described above following Mtt deprotection. Briefly, the Mtt group on K120 was deprotected by successive incubations of the peptidyl-resin with 1% TFA in DCM containing 1% TIS for 10 min intervals, until no yellow color evolved. Bromoacetic acid (222 mg, 1.6 mmol) was triple coupled to the ε-$NH_2$ of K120 with DIC (1.6 mmol) for 3 x 1 h. Ligation auxiliary **1** (32.6 mg, 76 mmol) and DIEA (1.4 mmol) were added to this resin (120 mg, 76 mmol) suspended in 600 μL DMF. The alkylation was allowed to proceed for 72 h at room temperature, after which, the resin was dried. Following simultaneous deprotection and cleavage from the resin with TFA:TIS:$H_2O$ (95:2.5:2.5) for 3 h, peptide **21** was purified by RP-HPLC on a preparative scale using a 28-38% B gradient over 45 min, yielding 10.8 mg peptide. Peptide **21** was characterized by ESI-MS [$(M+H)^+$ observed: 1,606.8 Da. $(M+H)^+$ expected: 1,607.9 Da.].

***Synthesis of peptides 27 and 32.*** For peptides **27** and **32**, the sequence corresponding to residues 117-125 of *Xenopus* H2B with an A117C substitution was synthesized on a pre-loaded Wang resin using automatic solid-phase peptide synthesis. Standard $^{t}$butyl side-chain protection was used throughout with the following exceptions; the ε-$NH_2$ of K120 was protected with the Mtt group, and the thiol group of C117 was installed as an $N^{\alpha}$-($^{t}$butoxycarbonyl)-thiazolidine (Boc-

thiazolidine, Bachem). For peptide **27**, the Mtt group on K120 was deprotected as described above followed by coupling with Boc-*S*-trityl-cysteine. Simultaneous deprotection and cleavage from the resin (as described above for peptide **21**), afforded peptide **27**, which was purified by RP-HPLC on a process scale using a 6-18% B gradient over 60 min, yielding 56 mg peptide from 0.25 mmol resin. Peptide **27** was characterized by ESI-MS [$(M+H)^+$ observed: 1,115.9 Da. expected: 1,115.3 Da.].

For peptide **32**, the sequence corresponding to residues 72-76 of ubiquitin was synthesized on the ε-$NH_2$ of K120 following Mtt deprotection. This ubiquitin branch was *N*-α-acetylated by incubation with acetic anhydride:DIEA:DMF (15:15:70) for 10 min. Following cleavage from the resin (as above), the resulting crude mixture was dissolved in 0.5 M aqueous methoxylamine, pH 5, for 3.5 h at room temperature to deprotect the N-terminal cysteine. Deprotected peptide **32** was purified by RP-HPLC on a preparative scale using a 10-25% B gradient over 45 min. Peptide **32** was characterized by ESI-MS [$(M+H)^+$ observed: 1,582.8 Da. $(M+H)^+$ expected: 1,582.8 Da.].

**Preparation of protein α-thioesters**

***Construction of ubiquitin(1-75) expression plasmid.*** The first 75 residues of the human ubiquitin gene were amplified by PCR using the primers Hub13-FP (5'-GGGAATTCCATATGCAGATCTTCGTGAAGACTC-3') and Hub13-RP (5'-GAATATATGCTCTTCCGCAACCTCTGAGACGGA-3') with the plasmid pFIR-CMV-ub-IRES-lnGFP (Ho lab, Rockefeller University) as the template DNA. The PCR product was purified, digested with NdeI and SapI restriction enzymes, and ligated into the identically digested pTXB1 vector. The resulting plasmid, pHub(1-75), which encodes ubiquitin(1-75) fused at its C-terminus to the GyrA intein and a chitin binding domain, was verified by DNA sequencing.

***Construction of HA-Smt3(2-97) expression plasmid.*** DNA encoding residues 2-97 of the yeast SUMO protein, Smt3, was amplified by PCR from the plasmid JM11a-4 (Muir lab, Rockefeller University) employing the primer HASmt3-FP (5'-GGAATTCCATATGTACCCGTATGATGTCCC AGA-3') bearing the *NdeI* restriction site and a sequence encoding the HA epitope (MYPYDVPDYA) and the primer HASmt3-RP (5'-GGTGGTTGCTCTTCCGCAACCAATCTGTT CTCT-3') bearing the *SapI* restriction site. The purified PCR product was digested and ligated into the pTXB1 as described above and the resultant plasmid, pHASmt3(2-97), which encodes HA-Smt3(2-97) fused at its C-terminus to the GyrA intein and a chitin binding domain, was verified by DNA sequencing.

***Construction of H2B(1-116) expression plasmid.*** A truncated *Xenopus* H2B gene, containing residues 1-116, was PCR amplified using primers H2B-FP (5'-GGAATTCCATATGCCTGAGC CAGCCAAGTCCGCTCCAGCCCCG-3') and H2B116-RP (5'- GGTGGTTGCTCTTCCGCAC TTGGTGCCCTCGGACAC-3') and a *Xenopus* H2B expression plasmid (Luger et al., 1997b) as a template. Following digestion by NdeI and SapI, the fragment was ligated into a similarly digested pTXB1 vector, and the resulting plasmid, pRMH2B-N, which encodes H2B(1-116) fused at its C-terminus to the GyrA intein and a chitin binding domain, was verified by DNA sequencing.

***Construction of ubiquitin(1-75)L8A/I44A expression plasmid.*** Ubiquitin(1-75) containing the L8A/I44A double mutation was prepared with the QuikChange Multi Site-directed Mutagenesis kit (Stratagene) with the following primers: UbL8A (5'-GCAGATCTTCGTGAAGACTGCGAC TGGTAAGACCATCACT-3') and UbI44A (5'- CTGACCAGCAGAGGTTGGCCTTTGCTGG GAAACAGC-3') and pHub(1-75) as a template, according to manufacturer's instructions, generating pHub(1-75)L8A/I44A, which was verified by DNA sequencing.

***Preparation of ubiquitin(1-75)-α-thioester.*** Ubiquitin(1-75)-α-thioester, **15**, was prepared on a small scale using shaking flasks and on a large scale using a fermenter. *E. coli* BL21(DE3) cells transformed with the plasmid pHub(1-75), were grown in 6 L Luria-Bertani medium supplemented with 100 μg/mL ampicillin at 37 °C with shaking at 250 rpm until an $OD_{600}$ of 0.6-0.8 was reached. Overexpression was induced by the addition of 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and the cells were grown for an additional 6 h at 25 °C. The cells were then harvested by centrifugation at 10 kg for 30 min and the cell pellet was resuspended in buffer A (50 mM Tris HCl, 200 mM NaCl, 1 mM EDTA, pH 7.2). The cells were lysed by passage through a French Press and the soluble fraction separated from insoluble cellular debris by centrifugation at 18.5-20 kg for 20 min. After filtration through a 0.45 μm filter, supernatants were bound to a 20 mL chitin column, pre-equilibrated with 200 mL of buffer A, for 2 h at 4 °C. The resin was washed with 700 mL of buffer A, followed by 60 mL buffer B (50 mM Tris HCl, 200 mM NaCl, and 1 mM EDTA, pH 7.5). Ubiquitin(1-75) was cleaved from the intein-CBD fusion by incubation with 30 mL of cleavage buffer (buffer B containing 100 mM of the sodium salt of 2-mercaptoethanesulfonic acid (MESNa)) for 87 h. The eluted ubiquitin(1-75)-α-thioester was subsequently purified by semi-preparative RP-HPLC, employing a gradient of 30-45% B over 45 min. This yielded 10.2 mg of the ubiquitin(1-75)-α-MES. The identity of thioester **15** was verified by ESI-MS [$(M+H)^+$ observed 8,632 ± 2 Da (s.d.). $(M+H)^+$ expected 8,632 Da.].

Ubiquitin(1-75)-α-MES was also prepared on large scale using a BioFlo 300 Benchtop Fermenter operating in fermentation mode (New Brunswick Scientific). The fermenter was inoculated with *E. coli* BL21(DE3) transformed with pHub(1-75). Cells were grown at 37 °C in Superbroth (MP Biomedicals) until an $OD_{600}$ of 21 was reached. Protein expression was induced with 0.5 mM IPTG at 25 °C for 6 h under continuous supplementation with 50% glucose and 3x Superbroth and oxygen enrichment. pH was maintained at 7.0 with 5 N NaOH. After

centrifugation at 6.8 kg for 15 min, the cell pellet was resuspended in 500 mL buffer B supplemented with 0.1 mM phenylmethylsulfonylfluoride (PMSF) and 0.2 mg/mL deoxyribonuclease (DNase) and lysed by passage through a French press. The insoluble material was removed by centrifugation at 26 kg for 30 min and the supernatant was incubated overnight at 4 °C with chitin resin (300 mL) pre-equilibrated in lysis buffer. The resin was washed with 2 L of buffer A and then incubated with cleavage buffer for 74 h, resulting in thiolysis of the intein fusion, forming ubiquitin(1-75)-α-MES, **15**. The column was eluted and the resin was washed with 200 mL cleavage buffer. The thiolysis reaction was repeated and the combined elution fractions further purified by process RP-HPLC using a 30-45% B gradient over 60 min, yielding 300 mg of lyophilized protein.

***Preparation of HA-Smt3(2-97)-α-thioester.*** *E. coli* BL21(DE3) cells transformed with the plasmid pHASmt3(2-97), were grown in 6 L Luria-Bertani medium and purified as described above for ubiquitin(1-75). Cleavage was performed for 42 h and eluted HA-Smt3(2-97)-α-MES, **18**, was purified by semi-preparative RP-HPLC, employing a 30-60% B gradient over 45 min. The identity of thioester **18** was verified by ESI-MS [$(M+H)^+$ observed 12,414 ± 3 Da (s.d.). $(M+H)^+$ expected 12,414 Da.].

***Preparation of H2B(1-116)-α-thioester.*** *E. coli* BL21(DE3) cells transformed with pRMH2B-N were grown in 6L Luria-Bertani medium at 37 °C until an $OD_{600}$ of 0.6 was reached, and protein expression was induced by the addition of 0.5 mM IPTG and allowed to continue at 25 °C for 16 h. Cells were harvested and lysed as described above. The cleared cell lysate was filtered and incubated overnight at 4 °C with chitin resin (35 mL) pre-equilibrated in buffer A. The resin was washed with 200 mL of buffer A and 700 mL buffer C (50 mM Tris HCl, 200 mM NaCl, 1 mM EDTA, pH 7.4). The resin was then incubated with cleavage buffer 2 (50 mM Tris HCl, 200 mM

NaCl, 1 mM EDTA, 100 mM MESNa, pH 7.4) for 70 h, resulting in thiolysis of the intein fusion, forming H2B(1-116)-α-MES, **24**. The column was eluted and the resin was washed with 2 x 25 mL cleavage buffer 2. The thiolysis reaction was repeated and the combined elution fractions further purified by preparative RP-HPLC using a 42-52% B gradient over 45 min, yielding 25 mg protein. The identity of thioester **24** was verified by ESI-MS [$(M+H)^+$ observed 12,991 ± 3 Da (s.d.). $(M+H)^+$ expected 12,991 Da.]. Mass spectrometry indicates that the non-native N-terminal methionine used for expression of **24** was processed during recombinant expression, leaving the native N-terminal sequence.

***Preparation of ubiquitin(1-75)L8A/I44A-α-thioester.*** *E. coli* BL21(DE3) cells transformed with pHub(1-75)L8A/I44A were grown in 6 L Luria-Bertani medium at 37 °C until an $OD_{600}$ of 0.6 was reached, and protein expression was induced by the addition of 0.5 mM IPTG and allowed to continue at 25 °C for 8 h. The cells were harvested, lysed, and cleared as described above. Cleared lysates were incubated overnight at 4 °C with chitin resin (20 mL) pre-equilibrated in buffer A. The resin was washed with 100 mL of buffer A followed by 200 mL buffer C. The resin was then incubated with 20 mL cleavage buffer 2 for 48 h, resulting in thiolysis of the intein fusion, forming ubiquitin(1-75)L8A/I44A-α-MES, **38**. The column was eluted and the resin was washed with 20 mL cleavage buffer 2. The thiolysis reaction was repeated twice and the combined elution fractions further purified by preparative RP-HPLC using a 25-55% B gradient over 45 min. The identity of thioester **38** was verified by ESI-MS [$(M+H)^+$ observed: 8,548 ± 1 Da (s.d.). $(M+H)^+$ expected: 8,548 Da.].

**Modification of peptides with ubiquitin and Ubls**

***Ligation of peptide 14 to ubiquitin(1-75)-α-thioester, 15.*** In a typical reaction, purified peptide **14** (1 mg, 0.58 μmol) was dissolved in 70 μL of a buffer containing 3 M guanidinium HCl, 300 mM sodium phosphate pH 7.8, and 50 mM tris(2-carboxyethyl)phosphine (TCEP), and incubated for 35 min at 25 °C to remove the *S*-$^{t}$butyl protection group on the ligation auxiliary. To the solution of reduced peptide was then added thioester **15** (1 mg, 0.12 μmol) dissolved in 70 μL of ligation buffer 1 (3 M guanidinium chloride, 300 mM sodium phosphate, 100 mM MES, pH 7.8). Upon mixing, the pH of the resultant solution was found to be ~7.5, and the ligation was allowed to proceed with gentle shaking for 5 days at 4 °C. The extent of reaction was checked periodically by withdrawing 1 μL aliquots of the ligation mixture, dilution into 50 μL of 25% B and analysis by analytical RP-HPLC. At the end of 5 days, the ligation product was reduced with TCEP and purified by analytical RP-HPLC using a 25-55% B gradient over 30 min to yield 0.5 mg of ligation product **16**. The identity of ligation product **16** was verified by ESI-MS [$(M+H)^+$ observed: 10,120 ± 2 Da (s.d.). $(M+H)^+$ expected: 10,120 Da.].

***Ligation of peptide 14 to HA-Smt3(2-97)-α-thioester, 18.*** Purified peptide **14** (1 mg, 0.58 μmol) was dissolved in 70 μL of reduction buffer described above to remove the *S*-$^{t}$butyl protection group on the ligation auxiliary. To the solution of reduced peptide was then added **18** (1.4 mg, 0.11 μmol) dissolved in 70 μL of ligation buffer 1. Ligation was allowed to proceed with gentle shaking for 7 days and the extent of reaction was analyzed by analytical RP-HPLC. After 7 days, the ligation product was reduced with TCEP and purified by analytical RP-HPLC employing a gradient of 34-47% B over 30 min to yield 0.7 mg of pure **19**. The identity of ligation product **19** was verified by ESI-MS [$(M+H)^+$ observed: 13,901 ± 5 Da (s.d.). $(M+H)^+$ expected: 13,900 Da.].

***Photolysis of ubiquitylated ligation product 16.*** For large-scale photolysis, the purified and lyophilized ligation product **16** was taken up in a photolysis buffer (50 mM Tris HCl, 150 mM NaCl,

1 mM DTT, pH 7.5) at a final concentration of 80 mM and mixed by vortexing prior to photolysis. Alternatively, a RP-HPLC fraction corresponding to the ligation product could also be employed in photolysis experiments with no observable changes in the product pattern. A 50-100 μL solution of **16** was briefly exposed to a He-Cd laser at 325 nm for 3-4 bursts of light of ~5 s duration with vortexing between exposures. The photolyzed ligation product **17** was purified by analytical RP-HPLC using a 0-73% B gradient over 30 min. The identity of photolysis product **17** was verified by ESI-MS [$(M+H)^+$ observed: 9,793 ± 4 Da (s.d.). $(M+H)^+$ expected: 9,793 Da.].

***Photolysis of sumoylated ligation product 19.*** The photolysis of ligation product **19** was performed as described above for the ubiquitylated peptide, **16**. The identity of photolysis product **20** was verified by ESI-MS [$(M+H)^+$ observed: 13,573 ± 4 Da (s.d.). $(M+H)^+$ expected: 13,573 Da.].

**Semisynthesis of uH2B**

***Ligation of peptide 21 to ubiquitin(1-75)-α-thioester, 15.*** The ligation reaction between peptide, **21**, and ubiquitin(1-75)-α-MES, **15**, was performed using conditions similar to those described above. In a typical reaction, purified peptide **21**, (1.1 mg, 0.69 μmol) was dissolved in 180 μL of buffer containing 3 M guanidinium chloride, 300 mM sodium phosphate pH 7.8, 50 mM TCEP, and incubated at room temperature for 30 min to remove the *S*-$^t$butyl protection group on the ligation auxiliary. The resulting reduced peptide was added to ubiquitin(1-75)-α-MES, **15** (17.1 mg, 1.98 μmol), dissolved in ligation buffer 1. The reaction volume was increased to 950 μL with ligation buffer 1 and the pH was adjusted to 7.8 using NaOH. After 120 h at 4 °C, the reaction was quenched with 1 mL of 50% HPLC buffer B containing 100 mM TCEP. Ligation product **22** was purified using preparative HPLC with a 32-42% B gradient over 45 min, yielding

4.0 mg of lyophilized protein. ESI-MS was used to verify the identity of the ligation product [$(M+H)^+$ observed: 10,008 ± 1 Da (s.d.). $(M+H)^+$ expected: 10,009 Da.].

***Photolysis of ubiquitylated ligation product 22.*** In a typical reaction, ligation product **22** (3.5 mg, 0.35 μmol) was dissolved in 1.5 mL of photolysis buffer (25% HPLC buffer B containing 10 mM semicarbazide, 10 mM DTT, 10 mM ascorbic acid, and 2 mM cysteine) (Smith et al., 2002). The resulting solution was irradiated at 365 nm for 4 h using a collimated light source (Oriel) equipped with a 200 W Hg lamp. Irradiance (4 $mW/cm^2$) was measured by using a model 840-c monochromic photometer. Selective irradiation at 365 nm was achieved by using an analytical line filter (9.4 nm bandwidth, Oriel). Irradiation effected removal of the ligation auxiliary and the *S*-*o*-nitrobenzyl protecting group, forming branched protein **23**. Semi-preparative RP-HPLC purification of protein **23** was accomplished using a 0-73% B gradient over 45 min, yielding 2.5 mg of lyophilized protein **23**. Removal of the two photolytic groups was verified by ESI-MS [$(M+H)^+$ observed: 9,547 ± 2 Da (s.d.). $(M+H)^+$ expected: 9,548 Da.].

***Ligation of ubiquitylated peptide 23 to H2B(1-116)-α-thioester, 24.*** In a typical reaction, photolysis product, **23** (2.5 mg, 0.25 μmol), and H2B(1-116)-α-MES, **24** (5.0 mg, 0.38 μmol), were dissolved in ligation buffer 1 to a final volume of 225 μL. The pH of the resulting solution was increased to 7.8 with NaOH and the reaction was allowed to proceed for 78 h prior to quenching with 225 μL of 50% HPLC buffer B containing 100 mM TCEP. The ligation product, **25**, was purified using semi-preparative RP-HPLC with a 42-52% B gradient over 45 min, yielding 3.0 mg of uH2BA117C, **25**. The identity of the ligation product was verified by ESI-MS [$(M+H)^+$ observed: 22,395 ± 3 Da (s.d.). $(M+H)^+$ expected: 22,397 Da.].

***Raney nickel mediated desulfurization of ligation product 25.*** Raney nickel reduction was used to convert C117 of protein **25** to the native alanine (Pentelute and Kent, 2007; Yan and Dawson,

2001). In a typical reaction, uH2BA117C, **25**, (1.3 mg, 58 nmol) was dissolved in 3 mL desulfurization buffer 1 (6 M guanidinium chloride, 200 mM sodium phosphate, 35 mM TCEP, pH 7.0). Raney nickel was prepared by adding 200 mg $NaBH_4$ to a stirred solution of 1.2 g of nickel acetate in 6 mL of water. After 5 min, the Raney nickel was filtered, washed with 200 mL water and added to the solution of uH2BA117C, **25**. The reaction progress was followed by RP-HPLC and ESI-MS. An identical amount of fresh Raney nickel was added after 6 h and the reaction was found to be complete at 8.5 h. (Note, much longer reaction times (over 24 hours) led to a second desulfurization reaction on methionine.) The Raney nickel was pelleted by centrifugation and washed with 4 x 0.5 mL desulfurization buffer. The reaction supernatant and washes were combined, added to an equivalent volume of 50% HPLC buffer B, and purified using semi-preparative RP-HPLC with a 42-52% B gradient over 45 min, yielding 1.1 mg of uH2B, **26**. uH2B was characterized by ESI-MS [$(M+H)^+$ observed: 22,366 ± 4 Da (s.d.). $(M+H)^+$ expected: 22,365 Da.].

**Semisynthesis of u(G76A)H2B**

***Ligation of peptide 27 to ubiquitin(1-75)-α-thioester, 15.*** The ligation reaction between peptide, **27**, and ubiquitin(1-75)-α-MES, **15**, was performed using conditions similar to those previously optimized for the corresponding auxiliary-mediated reaction. In a typical reaction, purified peptide **27**, (19.9 mg, 17.9 mmol) and protein **15** (45.7 mg, 5.3 mmol) were combined in 2.5 mL of ligation buffer 2 (6 M guanidinium chloride, 300 mM sodium phosphate, 100 mM MESNa, 50 mM TCEP, pH 7.8). The pH was adjusted to 7.8 using 5 N NaOH and the reaction was allowed to proceed for 4 h at room temperature. After 2 h, reaction completion, forming intermediate

product **28,** was verified by RP-HPLC and ESI-MS [$(M+H)^+$ observed = 9,609 ± 2 Da (s.d.). $(M+H)^+$ expected = 9,606 Da.].

***Cysteine deprotection of ubiquitylated ligation product 28.*** To crude ligation product **28** was added 2.5 mL 50% HPLC buffer B and 0.72 mL 4 M methoxylamine, to reach a final concentration of 0.5 M. The pH of the resulting solution was increased to 5 with 5 N NaOH. Following incubation for 12 h at room temperature to allow complete deprotection of the cysteine, product **29**, was purified by preparative RP-HPLC using a 25-55% B gradient over 45 min, yielding 27.3 mg purified protein. The identity of the purified protein **29** was verified by ESI-MS [$(M+H)^+$ observed = 9,593 ± 2 Da (s.d.). $(M+H)^+$ expected = 9,594 Da.].

***Ligation of ubiquitylated peptide 29 to H2B(1-116)-α-thioester, 24.*** In a typical reaction, protein **29** (29.9 mg, 3.1 mmol), and H2B(1-116)-α-MES, **24** (52.6 mg, 4.0 mmol), were dissolved in ligation buffer 2 to a final volume of 1.6 mL. The pH of the resulting solution was increased to 7.8 with 5 N NaOH and the reaction was allowed to proceed for 48 h. Fresh TCEP was then added to a final concentration of 50 mM. The ligation product, **30**, was purified using semi-preparative RP-HPLC with a 35-52% B gradient over 45 min, yielding 38.4 mg of protein **30**. The identity of the ligation product was verified by ESI-MS [$(M+H)^+$ observed = 22,444 ± 5 Da (s.d.). $(M+H)^+$ expected = 22,443 Da.].

***Desulfurization of ligation product 30.*** The 38.4 mg of protein **30** was converted to 30 mg u(G76A)H2B, **31**, using two complementary desulfurization methodologies. Raney nickel reduction was used to convert C117 and C76 of protein **30** to alanines (Pentelute and Kent, 2007; Yan and Dawson, 2001). In a typical reaction, protein **30** (8.0 mg, 0.36 mmol) was dissolved in 3 mL desulfurization buffer 1. Raney nickel was prepared as described above and added at 0 h, 6.5 h, 10 h, and 13.5 h. The reaction progress was followed by RP-HPLC and ESI-MS. After 22 h,

the Raney nickel was pelleted by centrifugation and washed. The reaction supernatant and washes were combined, added to an equivalent volume of 50% HPLC buffer B, and purified using semi-preparative RP-HPLC with a 42-52% B gradient over 45 min, yielding 4.0 mg of u(G76A)H2B, **31**. Protein **31** was characterized by ESI-MS.

Protein **30** was also desulfurized by the previously reported radical-initiated methodology (Chiang et al., 2009; Wan and Danishefsky, 2007). In a typical reaction, protein **30** (9.0 mg, 0.40 mmol) was dissolved in 500 μL desulfurization buffer 2 (6 M guanidinium chloride, 200 mM sodium phosphate, pH 7.0). To this solution was added 10 μl ethane thiol, 750 μl 0.5 M TCEP (dissolved in desulfurization buffer followed by pH adjustment to 7.0), 50 μl 2-methyl-2-propane thiol, and 12.5 μl 0.2 M VA-061 (Wako Chemicals) (dissolved in methanol), and the mixture was incubated in a 37 °C water bath. The reaction progress was followed by RP-HPLC and ESI-MS. After 24 h, the product was purified using semi-preparative RP-HPLC with a 42-52% B gradient over 45 min, yielding 7.6 mg of u(G76A)H2B, **31**. Protein **31** was characterized by ESI-MS [$(M+H)^+$ observed = 22,380 ± 4 Da (s.d.). $(M+H)^+$ expected = 22,379 Da.].

**Semisynthesis of uH2B structural variants**

***Synthesis of tr-uH2B, 34.*** Peptide **32** (0.6 mg, 0.38 μmol) and H2B(1-116)-α-MES, **24** (2.1 mg, 0.16 μmol), were dissolved in ligation buffer 2 to a final volume of 160 μL. The pH of the resulting solution was increased to 7.8 with 5 N NaOH and the reaction was allowed to proceed for 19 h prior to adding fresh TCEP to a final concentration of 25 mM. The ligation product, **33**, was purified using semi-preparative RP-HPLC with a 42-52% B gradient over 45 min, yielding 1.1 mg of protein. Protein **33** (1.1 mg, 76 nmol) was desulfurized using Raney nickel as described above and purified using semi-preparative RP-HPLC with a 42-52% B gradient over

45 min, yielding 0.3 mg of tr-uH2B, **34**. Protein **34** was characterized by ESI-MS [$(M+H)^+$ observed = 14,400 ± 3 Da (s.d). $(M+H)^+$ expected = 14,399 Da.].

***Synthesis of sH2B(cys), 37.*** To generate sH2B(cys), HA-Smt3(2-97)-α-MES, **18**, was substituted for ubiquitin(1-75)-α-MES in the first step of the synthesis. Peptide **27** (5.7 mg, 5.1 mmol) and protein **18** (8.4 mg, 0.68 mmol) were combined in 338 μL of ligation buffer 2. The pH was adjusted to 7.8 using 5 N NaOH and the reaction was allowed to proceed for 4 h at room temperature, forming ligation product **35**. To this solution was added 338 μL 50% HPLC buffer B and 169 μL 4 M methoxylamine and the pH was adjusted to 5. After 12 h at room temperature the deprotected protein, **36**, was purified by semipreparative RP-HPLC using a 35-60% B gradient, yielding 8.1 mg of product. Protein **36** (6.8 mg, 0.51 mmol) and protein **24** (5.6 mg, 0.43 mmol) were combined in 254 μL of ligation buffer 2. After 52 h at room temperature, fresh TCEP was added to a final concentration of 20 mM and the ligation product was purified by semi-preparative RP-HPLC using a 42-52% B gradient over 45 min, yielding 5.5 mg sH2B(cys), **37** [$(M+H)^+$ observed = 26,228 ± 9 Da (s.d.). $(M+H)^+$ expected = 26,224 Da.]. No protein was recovered following Raney nickel desulfurization of protein **37** using procedures detailed above.

***Synthesis of mut-uH2B, 42.*** To generate mut-uH2B, ubiquitin(1-75)L8A/I44A-α-MES, **38**, was substituted for ubiquitin(1-75)-α-MES in the first step of the synthesis. Peptide **27** (0.6 mg, 0.54 mmol) and protein **38** (5.0 mg, 0.58 mmol) were combined in 160 μL of ligation buffer 2. The pH was adjusted to 7.8 using 5 N NaOH and the reaction was allowed to proceed for 2 h at room temperature, forming ligation product **39**. To this solution was added 160 μL 50% HPLC buffer B and 46 μL 4 M methoxylamine and the pH was adjusted to 5. After 12 h at room temperature the deprotected protein, **40**, was purified by semipreparative RP-HPLC using a 32-42% B gradient, yielding 2.4 mg of product. Protein **40** (0.8 mg, 80 nmol) and protein **24** (1.2 mg, 90

nmol) were combined in 50 μL of ligation buffer. After 41 h at room temperature, fresh TCEP was added to a final concentration of 20 mM and the ligation product was purified by semipreparative RP-HPLC using a 42-52% B gradient over 45 min, yielding 0.2 mg protein **41.** Radical initiated desulfurization was performed on protein **41** (200 μg, 9.0 nmol) as described above using 1/5$^{th}$ the total reaction volume and the product was purified by analytical RP-HPLC using a 0-73% B gradient over 30 min, yielding 150 μg mut-uH2B, **42** [$(M+H)^+$ observed = 22,293 ± 7 Da (s.d.). $(M+H)^+$ expected = 22,295 Da.].

**Preparation of recombinant histones**

***Cloning of full-length H2B and histone mutants.*** The three N-terminal residues of H2B (PEP) were added to the *Xenopus* H2B expression plasmid (Luger et al., 1997b). The DNA encoding residues 1-125 of H2B was PCR amplified using primers H2B-FP described above and H2B-RP (5'-CGGGATCCTTACTTGGCGCTGGTGTACTTG-3') and the H2B plasmid described above as a template. Following digestion with NdeI and BamHI, the H2B gene was ligated into a similarly digested pET vector from the *Xenopus* H2A expression plasmid (Luger et al., 1997b). The H3K79R, H2AE64A, H2AN68A, and H4R17/19A point mutations were generated with a QuikChange II XL kit (Stratagene) using the following primers: H3K79R-FP (5'-GCTCAGGAC TTCAGGACCGACCTGCGC-3'), H3K79R-RP (5'-GCGCAGGTCGGTCCTGAAGTCCTGA GC-3'); H2AE64A-FP (5'-CTGACCGCTGAGATTTTGGCATTGGCCGGGAATG-3'), H2AE64A-RP (5'-CATTCCCGGCCAATGCCAAAATCTCAGCGGTCAG-3'); H2AN68A-FP (5'-TTTGGAATTGGCCGGGGCTGCGGCCCGTGATAAC-3'), H2AN68A-RP (5'-GTTATC ACGGGCCGCAGCCCCGGCCAATTCCAAA-3'); H4R17/19A-FP (5'-CTGGGTAAAGGTG GTGCTAAAGCTCACGCTAAAGTTCTGCGTGACAACA-3') and H4R17/19A-RP (5'-

TGTTGTCACGCAGAACTTTAGCGTGAGCTTTAGCACCACCTTTACCCAG-3') and the corresponding *Xenopus* histone expression plasmids as templates. The resulting mutations were verified by DNA sequencing.

***Expression and purification of histones.*** Recombinantly expressed *Xenopus* histones H2A, H2B, H3, H4, H3K79R, H2AE64A, H2AN68A, and H4R17/19A were prepared similarly to previously reported (Luger et al., 1999). For protein expression, *E. coli* BL21(DE3)pLysS cells transformed with the appropriate histone expression plasmid, were grown in 3-6 L 2xTY media at 37 °C until an $OD_{600}$ of 0.6 was reached, and protein expression was induced by the addition of 0.5 mM IPTG at 37 °C for 2-3 h. Cells were harvested and lysed as described above for protein α-thioesters. The insoluble fractions of the bacterial lysates were washed twice with 20 mL wash buffer (20 mM Tris HCl, 200 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, pH 7.5) and once with 20 mL of triton wash buffer (20 mM Tris HCl, 200 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, 1% triton, pH 7.5). DMSO (1 mL) was then added to the pellets and after 15 min, 15-50 mL of extraction buffer (7 M guanidinium chloride, 20 mM Tris HCl, 200 mM NaCl, 1 mM EDTA, 1 mM 2-mercaptoethanol, pH 7.5) was added and the suspension stirred for a further 15-30 min before clearing by centrifugation at 26 kg for 30 min. The supernatants were purified using a Sephacryl S-200 column (approximately 1 L bed volume), eluting with extraction buffer. Purified fractions were combined, dialyzed into water containing 2 mM DTT, and lyophilized. Histones were further purified using process RP-HPLC with a 40-60% B gradient over 60 min (with the exception of H3K79R for which a 45-65% B gradient was used). The identity of all purified histones was verified by ESI-MS.

**Ubiquitin hydrolysis assay**

Purified ubiquitylated H2B peptide, **17,** was dissolved in 200 μL of ubiquitin hydrolase assay buffer (50 mM Tris HCl, 150 mM NaCl, 1 mM DTT, pH 7.5) to a final concentration of ~100 μM and reacted with recombinant human ubiquitin C-terminal hydrolase L3 (UCH-L3, Boston Biochem). Typically, 4.5 μg of UCH-L3 (7 μL of 25 μM stock solution) were incubated in reduction buffer (50 mM Tris HCl, 150 mM NaCl, 12 mM DTT, pH 8.0) for 15-20 min at room temperature. To the reduced UCH-L3 was then added **17** in 78 μL of assay buffer. The mixture was incubated for 8 h at 37 °C. A similar control assay was also conducted, which included all the assay components except UCH-L3. The reactions were quenched after 8 h by the addition of 1 μL TFA and analyzed by analytical RP-HPLC in order to identify the hydrolysis products. Ubiquitin(1-76) was identified by ESI-MS [$(M+H)^+$ observed 8564 ± 3 Da (s.d.). $(M+H)^+$ expected 8566 Da.] The released cH2B peptide was identified using a LTQ ion trap mass spectrometer (ThermoFinnigan) [$(M+H)^{2+}$ observed 623.33 Da. $(M+H)^{2+}$ expected 623.30 Da.]. The correct sequence was further confirmed by MS-MS analysis, which generated 9b- and 10y-type ions.

For ubiquitin hydrolysis from ubiquitylated proteins, UCH-L3 was diluted to 5 μM in reduction buffer and incubated at room temperature for 15 min. Reduced UCH-L3 (1 μM) was then combined with either uH2B, **26**, or u(G76A)H2B, **31**, (both 10 μM), respectively, in 25 μL ubiquitin hydrolase assay buffer and transferred to a 37 °C water bath. After 2 and 10 min, 5 μL of each of the samples was removed for analysis by SDS-PAGE. At 10 min, the concentration of UCH-L3 was doubled and a final 5 μL sample was removed after 30 min. Controls were performed without histone and without UCH-L3. Samples were separated on a Criterion 15% Tris HCl gel and stained with Coomassie. An equivalent reaction was performed on 2.5x scale and submitted to LC-MS analysis using a QStar QQTOF LC-MS (Applied Biosystems) to verify hydrolysis of the isopeptide bond.

**α-uH2B Western blot**

250 and 125 ng of uH2B and u(G76A)H2B were separated on a Criterion 15% Tris HCl gel and transferred to a PVDF membrane using Towbin's buffer supplemented with SDS (192 mm glycine, 25 mm Tris-HCl, 10 % methanol, 0.2% SDS, pH 8.3). A Western blot was performed using a linkage specific α-uH2B antibody (cat. # MM-0029, Medimabs) at 1:3,000. The blot was stripped and probed with an α-ubiquitin antibody (cat. # MAB701, R & D Systems) at 1:1,000. The blot was then stained with Ponceau.

**Octamer reconstitution**

Histone octamers were formed as previously described (Luger et al., 1999). Briefly, individual histones were dissolved in unfolding buffer (7 M guanidinium chloride, 20 mM Tris HCl, 10 mM DTT, Complete EDTA-free Protease Inhibitor Cocktail (Roche), pH 7.5) at 2-4 mg/mL. Histones were combined in equimolar amounts based on UV quantitation (combined protein ranged from 0.75 to 18 mg), and the solution was diluted to 1 mg/mL with unfolding buffer. The resulting mixture was dialyzed into refolding buffer (2 M NaCl, 10 mM Tris HCl, 1 mM EDTA, 1 mM DTT, pH 7.5) (3 changes of 2 L each). Crude octamer assemblies were concentrated in Vivaspin 2 and 20 centricons (3-10 kDa MWCO) and purified using a Superdex 200 size exclusion column (10/300 column for <10 mg, 16/60 column for >10 mg), eluted with refolding buffer supplemented with Complete EDTA-free Protease Inhibitor Cocktail. Octamer quality was verified by 15% or 18% SDS-PAGE, followed by Coomassie staining. Octamer samples were stored at -80 °C in 20% glycerol.

**Preparation of DNA for nucleosome formation**

***Preparation of 147 bp of the Widom 601 sequence.*** A plasmid containing 12 copies of 177 bp of the 601 sequence (Dorigo et al., 2003) was purified using a Plasmid Giga kit (Qiagen). The 177 bp repeat was digested from the vector using EcoRV sites flanking the segment. The desired segment was selectively precipitated by incrementally increasing the concentration of PEG-6000 from 4% to 8.5%, followed by centrifugation at 26 kg to separate precipitated DNA. ScaI digestion and gel purification using a QIAquick Gel Extraction Kit (Qiagen) afforded a 177 bp fragment of the 601 sequence, 1_177_601. PCR amplification of the central 147 bp region of 1_177_601 was accomplished using primers 147-FP (5'-CTGGAGAATCCCGGTGCCGAGG-3') and 147-RP (5'-ACAGGATGTATATATCTGACACG-3'), and gel-purified 1_177_601 as a template. Purification of the PCR product using a QIAquick PCR purification kit (Qiagen) yielded 1_147_601. This PCR product served as template in further PCR amplification of the sequence.

***Preparation of Widom 601 DNA with complementary overhangs.*** 5' and 3' DNA fragments for dinucleosome formation were generated using PCR amplification as described above. The 5' and 3' fragments were designed to contain 3' and 5' DraIII restriction sites, respectively, allowing for the formation of complementary overhangs. The DNA segment for 5' mononucleosome formation, 601-5', was generated using the primers 147-FP and 5'-RP (5'-ATTGAGCACCCCGTGGGATCTTACATGCACAGGATG-3') and 1_177_601 DNA as the template. The DNA segment for 3' mononucleosome formation, 601-3', was generated using 3'-FP (5'-ATTGAGCACGGGGTGCGGCCGCCCTGGAG-3') and 147-RP primers. The PCR amplified fragments (110 μg) were digested with DraIII (700 U) using NEB Buffer 3 (1 mL) for 6 h at 37 °C, followed by purification using a QIAquick PCR Purification kit. When ligated, the

resulting 601-5' and 601-3' sequences connect two identical 147 bp regions of the 601 sequence, separated by a 30 bp linker.

***Preparation of 146 bp of the α-satellite palindromic sequence.*** For large-scale nucleosome reconstitutions (for kinetics and crystallography), the palindromic α-satellite sequence was prepared similarly to previously reported (Dyer et al., 2004). Briefly, half of the α-satellite palindromic sequence was amplified by PCR using template DNA sequences α-sat-temp-forward (5'-ATCAATATCCACCTGCAGATTCTACCAAAAGTGTATTTGGAAACTGCTCC ATCAAAAGGCATGTTCAGCGGAATTC-3') and α-sat-temp-reverse (5'-GAATTCCGCTGA ACATGCCTTTTGATGGAGCAGTTTCCAAATACACTTTTGGTAGAATCTGCAGGTGGA TATTGAT-3') and primers α-sat-forward (5'-CGGGATCCGATATCAATATCCACCTGCAG-3') and α-sat-reverse (5'-AAGGAAAAAAGCGGCCGCAGATCTGAATTCCGCTGAACATG CCTTTTG-3'). The PCR product was digested with BamHI and NotI and ligated into a similarly digested pCDNA3 vector. The resultant vector was digested with BamHI and NotI, allowing the α-satellite sequence to be isolated. The vector was also digested with BglII and NotI, allowing the pCDNA3 vector containing one copy of the α-satellite sequence to be isolated. These digestion products were ligated together, generating a vector with two head to tail copies of the α-satellite sequence, while obliterating the internal restriction site. This process was repeated until a vector containing 64 head to tail copies of the α-satellite sequence, pαSat64 was generated.

*E. coli* DH5α cells (Invitrogen) transformed with pαSat64 were amplified in the fermenter as described above for 15 h at 37 °C. Harvested cells were purified using a Plasmid Giga kit according to manufacturer's instructions to yield in excess of 40 mg of plasmid. Purified pαSat64 was digested with EcoRI and EcoRV to excise the α-satellite sequence. The vector was

removed by precipitation with 7.5% PEG-6000. The α-satellite sequence was ethanol precipitated and further purified away from the small fragments of DNA between the EcoRI and EcoRV restriction sites by gel filtration using a G50 sepharose (GE Healthcare). This sequence was self ligated with T4 ligase and then digested once more with EcoRV to generate ~6 mg of the mature 146 bp α-satellite palindromic sequence following purification by chloroform extraction and ethanol precipitation.

**Nucleosome formation**

***Small-scale nucleosome reconstitution by dilution.*** Mononucleosomes were formed using a previously described step-wise dilution procedure (Owen-Hughes et al., 1999). Briefly, octamers and DNA (1_147_601, 601-5', or 601-3') were combined in 10 μL high salt refolding buffer to a final concentration of 3-5 μM. After incubation at 37 °C for 15 min, 3.3 μL of dilution buffer 1 (10 mM HEPES, 1 mM EDTA, 0.5 mM PMSF, pH 7.9) was added and the temperature was dropped to 30 °C. Further dilutions of 6.7, 5, 3.6, 4.7, 6.7, 10, 30, and 20 μL, respectively, were then performed every 15 min. A final dilution with 100 μL of dilution buffer 2 (10 mM Tris HCl, 1 mM EDTA, 0.1% NP-40, 5 mM DTT, 0.5 mM PMSF, 20% glycerol, pH 7.5) was carried out. After an additional 15 min, the nucleosomes were concentrated using Vivaspin 500 centricons (3-10 kDa MWCO) at 4 °C. Nucleosome formation was verified by separation on a Criterion 5% TBE gel run in 0.5x TBE, followed by ethidium bromide staining. Nucleosomes were stored at -80 °C. Nucleosomes formed in this manner were used for all experiments except kinetics and crystallography.

***Large-scale nucleosome reconstitution by dialysis.*** For kinetics experiments, large-scale nucleosome reconstitution was performed with the 146 bp α-satellite palindromic DNA sequence

following procedures similar to those previously reported (Dyer et al., 2004). Briefly, octamers were mixed with α-satellite DNA in ratios optimized using the small-scale reconstitution protocol described above, at 5 μM DNA concentration, and the NaCl concentration was adjusted to 2 M. The resultant solution was dialyzed against 400 mL initial buffer (1.4 M KCl, 10 mM Tris HCl, 0.1 mM EDTA, 1 mM DTT, pH 7.5) at 4 °C for 70 min. The dialysate was diluted with final buffer (10 mM KCl, 10 mM Tris HCl, 0.1 mM EDTA, 1 mM DTT) to reach a final KCl concentration of 1.2 M and dialysis was continued for 70 min. Subsequent dialysis steps of at least 2 h in 1.0 M KCl, 70 min in 0.8 M KCl, and 70 min in 0.6M KCl, were followed by two dialysis steps of at least 3 h each in final buffer. Nucleosomes were concentrated to approximately 200 μL in Vivaspin 2 centricons and purified using a Model 491 Prep Cell from Biorad as previously described (Dyer et al., 2004). Fractions deemed pure on a 5% TBE native gel were combined and concentrated to 30 μM using Vivaspin 500 centricons. This material was used directly in crystallography screens, described below. Aliquots for kinetics assays were stored in 20% glycerol at 15 μM concentrations at -80 °C.

**Dinucleosome formation**

Dinucleosome ligations were performed as previously described (Zheng and Hayes, 2003). In a typical ligation reaction, 5' and 3' mononucleosomes (6.25 pmol) were combined in 200 μL of 1x ligation buffer (New England Biolabs) and 3,600 U of T4 ligase were added. After 1 h at room temperature, the ligated nucleosomes were concentrated using Vivaspin 500 centricons (3-10 kDa MWCO) and dinucleosome formation was verified using a 5% TBE gel as described above.

**Formation of chromatinized plasmid**

Chromatinized plasmids were assembled using a previously described procedure (Ito et al., 1999) and plasmid (An et al., 2004). Micrococcal nuclease assays were used to verify equivalent chromatin reconstitution with uH2B and H2B.

**Preparation of Dot1L**

The cDNA for Dot1L was obtained as a gift from Yi Zhang. After modification to include an N-terminal FLAG-epitope, the gene was subcloned into the pFASTBAC1 vector (Invitrogen) and the baculovirus was generated according to the manufacturer's instructions. Sf9 cells were infected with the baculovirus and the resulting cell extracts were subjected to standard purification procedures as described below. Sf9 cells (700 mL at a density of 1 million cells per mL) were infected with fresh virus (7.5 mL) three days prior to collection by centrifugation at 430 g for 5 min. Cells were washed with phosphate buffered saline and centrifuged again. Washed cells were resuspended in 18 mL lysis buffer (20 mM Tris HCl, 500 mM NaCl, 4 mM $MgCl_2$, 0.4 mM EDTA, 20% glycerol, 1 mM PMSF, 2 mM DTT, pH 7.9), and disrupted with 3x10 strokes with a dounce homogenizer (A pestle) with a 10 min break between sets of 10. After removal of cell debris by centrifugation at 22.5 kg for 15 min, the supernatant was adjusted to 0.1% NP-40 and 300 mM NaCl, by dilution with 20 mM Tris HCl containing 10% glycerol, and the resulting solution was incubated with 400 μL M2-agarose (Sigma) for 3.5 h at 4 °C. The agarose was washed four times with 10 mL wash buffer (20 mM Tris HCl, 150 mM NaCl, 2 mM $MgCl_2$, 0.2 mM EDTA, 15% glycerol, 0.1% NP-40, 1 mM PMSF, 1 mM DTT, pH 7.9). Each wash was followed by centrifugation at 400 g for 1 min. The third wash was left on a rotator for

10 min at 4 °C. Bound Dot1L was eluted with 3x100 μL wash buffer containing 0.5 mg/mL 3x FLAG peptide (Sigma).

**Preparation of Dot1L$_{cat}$**

The catalytic domain of Dot1L encoding residues 1-416 was amplified from MSCN-hDot1L (gift from Y. Zhang, UNC School of Medicine) using primers 416-FP (5'-CGGGATCCCATCACCATCATCATCACATGGGGGAGAAGCTGGAGCTG-3') and 416-RP (5'-GGAATTCCTACTTCTTGGGGCGCCCGCGC-3'). The resulting sequence was digested with BamHI and EcoRI and ligated into a similarly digested pGexTEV vector (pGex2T with the thrombin recognition sequence replaced with a TEV sequence). Ligated vector, pGexTEV-hDot1L416, was verified by DNA sequencing. *E. coli* BL21(DE3) cells transformed pGexTEV-hDot1L416, were grown in 6 L Luria-Bertani medium at 37 °C until mid-log phase, and protein expression of the Dot1L$_{cat}$ GST fusion protein was induced by the addition of 0.5 mM IPTG at 18 °C for 18 h. Cells were harvested and lysed as described above. Cleared cell lysates were continuously loaded onto a GST-Prep FF 16/10 column (GE Healthcare) at 1 mL/min for 3 h. The column was washed with 200 mL of 20 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 2 mM DTT, pH 7.5 and eluted with 200 mL of the same buffer containing 5 mM glutathione. Fractions deemed pure by SDS-PAGE were stored in 20% glycerol at -80 °C. Impure fractions were diluted with an equal volume of start buffer (20 mM Tris HCl, 1 mM EDTA, 2 mM DTT, pH 7.5) and loaded onto a HiTrap SP HP 1 mL column (GE Healthcare). After washing with 10 mL start buffer, Dot1L$_{cat}$ was eluted with a 10 mL linear gradient from start buffer to end buffer (20 mM Tris HCl, 1 mM EDTA, 2 mM DTT, 1 M NaCl, pH 7.5). Pure fractions were stored at -80 °C in 20% glycerol. For single-turnover type reactions, Dot1L$_{cat}$ was concentrated using

Vivaspin 500 centricons to greater than 5 μM as determined by quantification against BSA standards by SDS-PAGE.

**Methyltransferase assays**

***Mono- and dinucleosome assays.*** Mono- and dinucleosomes (4.25 pmol) were combined with Dot1L (0.12 pmol) or $Dot1L_{cat}$ (0.18 pmol) in 25 μL of assay buffer (20 mM Tris HCl, 4 mM EDTA, 0.5 mM DTT, 1 mM PMSF, pH 7.9). To initiate the reaction, $^{3}H$ *S*-adenosyl methionine (SAM) (1 mCi, 14.25 pmol) was added and the reaction was allowed to proceed at 30 °C for 1 h. After 1 h, 20 μL of each sample were separated on a 5% TBE gel as described above and stained with ethidium bromide. After visualization, the gel was fixed with 25% isopropanol, 10% acetic acid in water, followed by a 30 min incubation with Amplify solution (GE Healthcare). Dried gels were visualized by fluorography. The remaining 5 μL of reaction mixture were spotted on Whatman p81 filter paper, washed 3 x 10 min with sodium carbonate, pH 9, and dried. Filter papers were added to 5 mL Ready Safe Liquid Scintillation Cocktail (Beckman Coulter). The samples were vortexed for 10 sec and counted using a LKB Wallac 1209 RackBeta Primo Liquid scintillation counter.

***Chromatinized template assays.*** 350 ng chromatinized template formed with octamers containing H2B or uH2B were incubated with 15 ng Dot1L in 40 μl assay buffer containing $^{3}H$ SAM at 30 °C for 1 h. Assay samples were separated by 15% SDS-PAGE and fluorography was performed as described above.

***Mass spectrometric analysis of methyltransferase assays.*** Methyltransferase assays were performed similar to that described above but on twice the scale using cold SAM (1 mM). Assays were incubated at 30 °C for 3 h, concentrated in a Vivaspin 500 centricon (10 kDa

MWCO), separated on a Criterion 15% Tris HCl gel, and Coomassie stained. The Coomassie-stained H3 bands were excised from the gel and destained with 200 mM ammonium bicarbonate. Tryptic digestion was initiated with the addition of 25 ng/μL Sequence Grade Modified Trypsin (Promega) in ammonium bicarbonate buffer. The protein was digested for at least 16 h at 37 °C. The digestion products were mixed with 0.5 μL of 10 mg/mL α-cyano-4-hydroxysuccinnamic acid in 50% acetonitrile containing 0.1% TFA, and applied to a MALDI plate. MALDI mass spectra were recorded with a PerSeptive Voyager-DE STR MALDI time-of-flight mass spectrometer operated in the reflectron mode. Ion intensity of the tryptic fragment of H3 containing residues 40-49 was used to scale the intensities of the methylated peptides for comparison of separate samples.

***Kinetics assays.*** Serial two-fold dilutions of nucleosomes from 7.5 μM to 230 nM were preformed in methyltransferase assay buffer (20 mM Tris HCl, 4 mM EDTA, 1 mM DTT, pH 7.9) supplemented with 1.6 μM $^{3}$H SAM at 4 °C. Each solution was aliquoted into individual PCR tubes. The reaction was initiated by the addition of Dot1L$_{cat}$ (2 nM, 4 nM, and 6 nM) and subsequent transfer to a 30 °C water bath. For nucleosomes containing u(G76A)H2B, 5 μL assay samples were quenched at 3 min by the addition of 10 μL 0.2 % TFA and spotted on filter paper for liquid scintillation counting as described above. For unmodified nucleosomes, 10 μL assay samples were quenched at 15 min and quantified as above. Each concentration of nucleosome was assayed against each concentration of Dot1L$_{cat}$ in duplicate. Counts from reactions containing no Dot1L$_{cat}$ were subtracted from each sample. Conversion of CPM to moles of methylation was performed using LC-MS/MS with isotopically-labelled synthetic standards. Unmodified and monomethylated peptides containing residues 73-83 of H3 and a $d_5$ isotopically-labelled Phe78 (Fmoc-Phe(d5)-OH from Cambridge Isotopes) were added into gel slices

containing assay samples. Propionylation, trypsin digestion, and propionylation were performed as previously described (Garcia et al., 2007). Relative quantities were determined by standardizing spectral counts of modified tryptic peptides to synthetic isotopically-labelled peptides following LC-MS/MS using a Dionex U3000 capillary/nano-HPLC system (Dionex) directly interfaced with the Thermo-Fisher LTQ-Orbitrap mass spectrometer (Thermo Fisher). In a typical experiment, a conversion factor ($F$) of $2.15*10^{-16}$ mol/CPM was calculated. Values of $v_o/E_T$ were calculated using equation (1),

$$\frac{v_o}{E_T} = \frac{F(CPM_{obs} - CPM_{background})}{t * E_T} \quad (1)$$

where $v_o$ is the initial velocity, $E_T$ is the molar quantity of $Dot1L_{cat}$ in the volume quantified, and $t$ is time in seconds. Kinetic parameters for the methylation of ubiquitylated nucleosomes were obtained by fitting to the non-linear Michaelis-Menten equation (2) using profit 6.1.1 with the Levenberg-Marquardt algorithm:

$$\frac{v_o}{E_T} = \frac{k_{cat}[S]}{K_m + [S]} \quad (2)$$

where [S] is the concentration of nucleosome. Time courses were fit to linear regression models to verify steady-state conditions.

Assays performed with an excess of $Dot1L_{cat}$ were set up as follows: Nucleosome and $Dot1L_{cat}$ were combined in methyltransferase assay buffer supplemented with 1 μM $^3$H SAM and 4 μM cold SAM, at 250 nM and 5 μM concentrations, respectively. The reaction was initiated by transferring to a 30 °C water bath. Samples were removed, quenched, and counted as described

above. The observed CPM value at 0 sec was subtracted from the observed CPM value after 15, 30, 60, 120, 300, 600, and 6,000 sec. The corrected value was multiplied by 5 to account for dilution with cold SAM and divided by the number of microliters quantified. The appropriate conversion factor (*F*) was applied to calculate methyltransferase activity at each time point. The resultant curves were fit to a single exponential model (3) using profit 6.1.1.

$$\text{Activity} = ae^{-bt} + c \tag{3}$$

For mass spectrometric analysis of methylation levels, assays were submitted to trypsin digestion without propionylation as described above. Tryptic peptides were analyzed by LC-MS/MS as described above.

***Asymmetric mononucleosome assays.*** Octamers were prepared as described above using mixtures of u(G76A)H2B and H2B. Incorporation of the appropriate levels of ubiquitylated and unmodified H2B was verified by separation on a Criterion 15% Tris HCl gel, followed by Coomassie staining. Nucleosomes were reconstituted on a small scale as described above and analyzed on a Criterion 5% TBE gel. Each nucleosome sample was subjected to methyltransferase assays as described above. Results, relative to homogenously ubiquitylated nucleosomes, were compared to theoretical values for each model tested, assuming negligible methylation of unmodified nucleosomes. If each u(G76A)H2B stimulates methylation of one H3K79 within the nucleosome, the simple linear model (4) is expected.

$$\text{Rel. CPM} = \text{f}_{\text{ub}} \tag{4}$$

where $f_{ub}$ is fraction of H2B that is ubiquitylated. Assuming independent assortment of u(G76A)H2B and H2B, if each u(G76A)H2B stimulates methylation of both H3K79 within the nucleosome, the non-linear model (5) is expected.

$$\text{Rel. CPM} = f_{ub}^{\ 2} + 2(f_{ub})(1 - f_{ub}) \quad (5)$$

which simplifies to equation (6).

$$\text{Rel. CPM} = -f_{ub}^{\ 2} + 2f_{ub} \quad (6)$$

**Electrophoretic mobility shift assays**

Mononucleosomes (3.0 pmol), $Dot1L_{cat}$ (0-4.5 pmol) and cold SAM (10 nmol) were combined in 30 μL of assay buffer. After 1 h at 30 °C, samples were separated on a 5% TBE gel as described above and visualized with ethidium bromide staining. Proteins were transferred to a PVDF membrane as described above and a Western blot was performed against GST using goat α-GST (1: 1000, cat # 27-4577-50, GE Healthcare) and HRP-conjugated rabbit α-goat (1:2000, cat # 0449, Dako).

For radiolabelled gel-shift competition experiments, 1_147_601 (10 μL at 250 μM) was end-labelled with $^{32}P$ using $^{32}P$-ATP (8 μL) (Perkin Elmer; Waltham, MA) and polynucleotide kinase (50 units) (New England Biolabs) in 50 μL 1x PNK buffer for 3 h at 37 °C. Radiolabelled nucleosomes were prepared as described above by dilution at 2 μM starting concentrations. Radiolabelled nucleosomes (0.64 pmol), $hDot1L_{cat}$ (3.2 pmol), and cold nucleosomes (0-3.2

pmol) were incubated and separated as described above. Dried gels were imaged by phosphorimaging on a Typhoon 8400 (GE Healthcare).

**Crystal screening**

***Crystal screen 1.*** Preparative gel fractions containing pure nucleosomes reconstituted with either H2B or u(G76A)H2B and α-satellite DNA were concentrated to 4 and 6 mg/mL, respectively, using Vivaspin 500 centricons. A fraction of the u(G76A)H2B nucleosome was diluted to 3 mg/mL in preparative gel elution buffer. Reservoirs were prepared as shown in Figures 5.15 and 5.17, containing 20 mM potassium cacodylate, 25-40 mM KCl, and 35-49 mM $MnCl_2$, pH 6.0. Hanging drops were prepared containing 1 μL of each nucleosome sample and 1 μL of each reservoir solution. Crystal growth was followed for three weeks.

***Crystal screen 2.*** Preparative gel fractions containing pure nucleosomes reconstituted with either H2B or u(G76A)H2B and α-satellite DNA were dialyzed against three changes of 20 mM potassium cacodylate, pH 6.0. Dialyzed nucleosomes were concentrated to 4 mg/mL in Vivaspin 500 centricons. Reservoirs were prepared as shown in Figures 5.15 and 5.17, containing 20 mM potassium cacodylate, 25-40 mM KCl, and 35-49 mM $MnCl_2$, pH 6.0. Hanging drops were prepared containing 1 μL of each nucleosome sample and 1 μL of each reservoir solution. Crystal growth was followed for three weeks.

***Crystal screen 3.*** Preparative gel fractions containing pure nucleosomes reconstituted with either H2B or u(G76A)H2B and α-satellite DNA were dialyzed against three changes of 20 mM potassium cacodylate, pH 6.0. Dialyzed nucleosomes were concentrated to 4 mg/mL in Vivaspin 500 centricons. Reservoirs were prepared as shown in Figures 5.15 and 5.17, containing 20 mM potassium cacodylate, 32.5-37.5 mM KCl, and 37-45 mM $MnCl_2$. pH 6.0. Hanging drops were

prepared containing 1.8 μL of each nucleosome sample and 0.2 μL of each reservoir solution at a 20x concentration. Crystal growth was followed for three weeks.

***Crystal screen 4.*** Preparative gel fractions containing pure nucleosomes reconstituted with u(G76A)H2B and α-satellite DNA were dialyzed against three changes of 10 mM Tris HCl, 10 mM NaCl, 0.1 mM EDTA, 1 mM DTT, and 0.1 mM potassium cacodylate, pH 7.5. The dialyzed nucleosome sample was concentrated to 5 mg/mL using a Vivaspin 500 centricon. Sitting drops were prepared containing 1 μL of nucleosome and 1 μL of each reservoir sample of the Nucleix kit (Qiagen). Crystal growth was followed for three weeks.

***Crystal diffraction.*** A crystal from screen 1 was looped and soaked in reservoir solution supplemented with 24% 2-methyl-2,4-propanediol (Qiagen) and 2% trehalose, prior to freezing in liquid nitrogen. Diffraction data was collected using a Rigaku/MSC microMax 007HF generator equipped with Varimax optics, an X-stream cryosystem and a RaxisIV$^{++}$ detector.

***SDS-PAGE analysis.*** The solution surrounding a cluster of crystals was wicked away. The cluster was washed with 2 x 2 μl reservoir solution. Washes and crystals dissolved in SDS loading dye were analyzed by SDS-PAGE using an 8-25% Fast gel (GE Healthcare) followed by Coomassie staining.

**References**

Ahn, S. H., Cheung, W. L., Hsu, J. Y., Diaz, R. L., Smith, M. M., and Allis, C. D. (2005). Sterile 20 kinase phosphorylates histone H2B at serine 10 during hydrogen peroxide-induced apoptosis in S. cerevisiae. Cell *120*, 25-36.

Ahn, S. H., Diaz, R. L., Grunstein, M., and Allis, C. D. (2006). Histone H2B deacetylation at lysine 11 is required for yeast apoptosis induced by phosphorylation of H2B at serine 10. Mol Cell *24*, 211-220.

Alexandrov, K., Heinemann, I., Durek, T., Sidorovitch, V., Goody, R. S., and Waldmann, H. (2002). Intein-mediated synthesis of geranylgeranylated Rab7 protein in vitro. J Am Chem Soc *124*, 5648-5649.

Allfrey, V. G., Faulkner, R., and Mirsky, A. E. (1964). Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. Proc Natl Acad Sci U S A *51*, 786-794.

Allis, C. D., Jenuwein, T., and Reinberg, D. (2007). Epigenetics (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Altaf, M., Utley, R. T., Lacoste, N., Tan, S., Briggs, S. D., and Cote, J. (2007). Interplay of chromatin modifiers on a short basic patch of histone H4 tail defines the boundary of telomeric heterochromatin. Mol Cell *28*, 1002-1014.

An, W., Kim, J., and Roeder, R. G. (2004). Ordered cooperative functions of PRMT1, p300, and CARM1 in transcriptional activation by p53. Cell *117*, 735-748.

An, W., and Roeder, R. G. (2004). Reconstitution and transcriptional analysis of chromatin in vitro. Methods Enzymol *377*, 460-474.

Aylon, Y., and Kupiec, M. (2004). DSB repair: the yeast paradigm. DNA Repair (Amst) *3*, 797-815.

Bachmair, A., Finley, D., and Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. Science *234*, 179-186.

Bang, D., Makhatadze, G. I., Tereshko, V., Kossiakoff, A. A., and Kent, S. B. (2005). Total chemical synthesis and X-ray crystal structure of a protein diastereomer: [D-Gln 35]ubiquitin. Angew Chem Int Ed Engl *44*, 3852-3856.

Bao, Y., White, C. L., and Luger, K. (2006). Nucleosome core particles containing a poly(dA.dT) sequence element exhibit a locally distorted DNA structure. J Mol Biol *361*, 617-624.

Barbera, A. J., Chodaparambil, J. V., Kelley-Clarke, B., Joukov, V., Walter, J. C., Luger, K., and Kaye, K. M. (2006). The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. Science *311*, 856-861.

Beal, R., Deveraux, Q., Xia, G., Rechsteiner, M., and Pickart, C. (1996). Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. Proc Natl Acad Sci U S A *93*, 861-866.

Bestor, T. H., and Ingram, V. M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc Natl Acad Sci U S A *80*, 5559-5563.

Bitoun, E., Oliver, P. L., and Davies, K. E. (2007). The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. Hum Mol Genet *16*, 92-106.

Borodovsky, A., Ovaa, H., Kolli, N., Gan-Erdene, T., Wilkinson, K. D., Ploegh, H. L., and Kessler, B. M. (2002). Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. Chem Biol *9*, 1149-1159.

Botuyan, M. V., Lee, J., Ward, I. M., Kim, J. E., Thompson, J. R., Chen, J., and Mer, G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell *127*, 1361-1373.

Briggs, S. D., Xiao, T., Sun, Z. W., Caldwell, J. A., Shabanowitz, J., Hunt, D. F., Allis, C. D., and Strahl, B. D. (2002). Gene silencing: trans-histone regulatory pathway in chromatin. Nature *418*, 498.

Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., and Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843-851.

Buhler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell *125*, 873-886.

Chakravarthy, S., and Luger, K. (2006). The histone variant macro-H2A preferentially forms "hybrid nucleosomes". J Biol Chem *281*, 25522-25531.

Chatterjee, C., McGinty, R. K., Pellois, J. P., and Muir, T. W. (2007). Auxiliary-mediated site-specific peptide ubiquitylation. Angew Chem Int Ed Engl *46*, 2814-2818.

Chen, A., Kleiman, F. E., Manley, J. L., Ouchi, T., and Pan, Z. Q. (2002). Autoubiquitination of the BRCA1*BARD1 RING ubiquitin ligase. J Biol Chem *277*, 22085-22092.

Chiang, K. P., Jensen, M. S., McGinty, R. K., and Muir, T. W. (2009). A Semisynthetic Strategy to Generate Phosphorylated and Acetylated Histone H2B. Chembiochem.

Chin, H. G., Patnaik, D., Esteve, P. O., Jacobsen, S. E., and Pradhan, S. (2006). Catalytic properties and kinetic mechanism of human recombinant Lys-9 histone H3 methyltransferase SUV39H1: participation of the chromodomain in enzymatic catalysis. Biochemistry *45*, 3272-3284.

Chow, J., and Heard, E. (2009). X inactivation and the complexities of silencing a sex chromosome. Curr Opin Cell Biol *21*, 359-366.

Clapier, C. R., and Cairns, B. R. (2009). The biology of chromatin remodeling complexes. Annu Rev Biochem *78*, 273-304.

Clapier, C. R., Chakravarthy, S., Petosa, C., Fernandez-Tornero, C., Luger, K., and Muller, C. W. (2008). Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. Proteins *71*, 1-7.

Collazo, E., Couture, J. F., Bulfer, S., and Trievel, R. C. (2005). A coupled fluorescent assay for histone methyltransferases. Anal Biochem *342*, 86-92.

Costanzi, C., and Pehrson, J. R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature *393*, 599-601.

Crick, F. H. (1958). On protein synthesis. Symp Soc Exp Biol *12*, 138-163.

Daniel, J. A., Torok, M. S., Sun, Z. W., Schieltz, D., Allis, C. D., Yates, J. R., 3rd, and Grant, P. A. (2004). Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription. J Biol Chem *279*, 1867-1871.

Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W., and Richmond, T. J. (2002). Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol *319*, 1097-1113.

Davies, N., and Lindsey, G. G. (1994). Histone H2B (and H2A) ubiquitination allows normal histone octamer and core particle reconstitution. Biochim Biophys Acta *1218*, 187-193.

Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994). Synthesis of proteins by native chemical ligation. Science *266*, 776-779.

Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., and Zhou, M. M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. Nature *399*, 491-496.

Dirk, L. M., Flynn, E. M., Dietzel, K., Couture, J. F., Trievel, R. C., and Houtz, R. L. (2007). Kinetic manifestation of processivity during multiple methylations catalyzed by SET domain protein methyltransferases. Biochemistry *46*, 3905-3915.

Dlakic, M. (2001). Chromatin silencing protein and pachytene checkpoint regulator Dot1p has a methyltransferase fold. Trends Biochem Sci *26*, 405-407.

Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T. J. (2003). Chromatin fiber folding: requirement for the histone H4 N-terminal tail. J Mol Biol *327*, 85-96.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R. R., and Richmond, T. J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science *306*, 1571-1573.

Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D., and Roeder, R. G. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat Struct Mol Biol *13*, 713-719.

Dougherty, D. A. (2000). Unnatural amino acids as probes of protein structure and function. Curr Opin Chem Biol *4*, 645-652.

Dover, J., Schneider, J., Tawiah-Boateng, M. A., Wood, A., Dean, K., Johnston, M., and Shilatifard, A. (2002). Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J Biol Chem *277*, 28368-28371.

Duncan, E. M., Muratore-Schroeder, T. L., Cook, R. G., Garcia, B. A., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2008). Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation. Cell *135*, 284-294.

Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M., and Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods Enzymol *375*, 23-44.

Edayathumangalam, R. S., Weyermann, P., Gottesfeld, J. M., Dervan, P. B., and Luger, K. (2004). Molecular recognition of the nucleosomal "supergroove". Proc Natl Acad Sci U S A *101*, 6864-6869.

Emre, N. C., Ingvarsdottir, K., Wyce, A., Wood, A., Krogan, N. J., Henry, K. W., Li, K., Marmorstein, R., Greenblatt, J. F., Shilatifard, A., and Berger, S. L. (2005). Maintenance of low histone ubiquitylation by Ubp10 correlates with telomere-proximal Sir2 association and gene silencing. Mol Cell *17*, 585-594.

Ercan, S., and Lieb, J. D. (2009). C. elegans dosage compensation: a window into mechanisms of domain-scale gene regulation. Chromosome Res *17*, 215-227.

Eskeland, R., Czermin, B., Boeke, J., Bonaldi, T., Regula, J. T., and Imhof, A. (2004). The N-terminus of Drosophila SU(VAR)3-9 mediates dimerization and regulates its methyltransferase activity. Biochemistry *43*, 3740-3749.

Ezhkova, E., and Tansey, W. P. (2004). Proteasomal ATPases link ubiquitylation of histone H2B to methylation of histone H3. Mol Cell *13*, 435-442.

Feng, Q., Wang, H., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002). Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Curr Biol *12*, 1052-1058.

Ferreira, H., Flaus, A., and Owen-Hughes, T. (2007). Histone modifications influence the action of Snf2 family remodelling enzymes by different mechanisms. J Mol Biol *374*, 563-579.

Fingerman, I. M., Li, H. C., and Briggs, S. D. (2007). A charge-based interaction between histone H4 and Dot1 is required for H3K79 methylation and telomere silencing: identification of a new trans-histone pathway. Genes Dev *21*, 2018-2029.

Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J., Hunt, D. F., Funabiki, H., and Allis, C. D. (2005). Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature *438*, 1116-1122.

Flavell, R. R., Huse, M., Goger, M., Trester-Zedlitz, M., Kuriyan, J., and Muir, T. W. (2002). Efficient semisynthesis of a tetraphosphorylated analogue of the Type I TGFbeta receptor. Org Lett *4*, 165-168.

Flavell, R. R., and Muir, T. W. (2009). Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. Acc Chem Res *42*, 107-116.

Fleming, A. B., Kao, C. F., Hillyer, C., Pikaart, M., and Osley, M. A. (2008). H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. Mol Cell *31*, 57-66.

Frederiks, F., Tzouros, M., Oudgenoeg, G., van Welsem, T., Fornerod, M., Krijgsveld, J., and van Leeuwen, F. (2008). Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states. Nat Struct Mol Biol *15*, 550-557.

Game, J. C., and Chernikova, S. B. (2009). The role of RAD6 in recombinational repair, checkpoints and meiosis via histone modification. DNA Repair (Amst) *8*, 470-482.

Game, J. C., Williamson, M. S., and Baccari, C. (2005). X-ray survival characteristics and genetic analysis for nine Saccharomyces deletion mutants that show altered radiation sensitivity. Genetics *169*, 51-63.

Game, J. C., Williamson, M. S., Spicakova, T., and Brown, J. M. (2006). The RAD6/BRE1 histone modification pathway in Saccharomyces confers radiation resistance through a RAD51-dependent process that is independent of RAD18. Genetics *173*, 1951-1968.

Garcia, B. A., Hake, S. B., Diaz, R. L., Kauer, M., Morris, S. A., Recht, J., Shabanowitz, J., Mishra, N., Strahl, B. D., Allis, C. D., and Hunt, D. F. (2007). Organismal differences in post-translational modifications in histones H3 and H4. J Biol Chem *282*, 7641-7655.

Gardner, R. G., Nelson, Z. W., and Gottschling, D. E. (2005). Ubp10/Dot4p regulates the persistence of ubiquitinated histone H2B: distinct roles in telomeric silencing and general chromatin. Mol Cell Biol *25*, 6123-6139.

Gelbart, M. E., and Kuroda, M. I. (2009). Drosophila dosage compensation: a complex voyage to the X chromosome. Development *136*, 1399-1410.

Georgel, P. T., and Hansen, J. C. (2001). Linker histone function in chromatin: dual mechanisms of action. Biochem Cell Biol *79*, 313-316.

Giannattasio, M., Lazzaro, F., Plevani, P., and Muzi-Falconi, M. (2005). The DNA damage checkpoint response requires histone H2B ubiquitination by Rad6-Bre1 and H3 methylation by Dot1. J Biol Chem *280*, 9879-9886.

Gibbons, R. J., McDowell, T. L., Raman, S., O'Rourke, D. M., Garrick, D., Ayyub, H., and Higgs, D. R. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nat Genet *24*, 368-371.

Glaser, S., Schaft, J., Lubitz, S., Vintersten, K., van der Hoeven, F., Tufteland, K. R., Aasland, R., Anastassiadis, K., Ang, S. L., and Stewart, A. F. (2006). Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development *133*, 1423-1432.

Gloss, L. M., and Kirsch, J. F. (1995). Decreasing the basicity of the active site base, Lys-258, of Escherichia coli aspartate aminotransferase by replacement with gamma-thialysine. Biochemistry *34*, 3990-3998.

Goldknopf, I. L., and Busch, H. (1977). Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24. Proc Natl Acad Sci U S A *74*, 864-868.

Grenon, M., Costelloe, T., Jimeno, S., O'Shaughnessy, A., Fitzgerald, J., Zgheib, O., Degerth, L., and Lowndes, N. F. (2007). Docking onto chromatin via the Saccharomyces cerevisiae Rad9 Tudor domain. Yeast *24*, 105-119.

Guenther, M. G., Lawton, L. N., Rozovskaia, T., Frampton, G. M., Levine, S. S., Volkert, T. L., Croce, C. M., Nakamura, T., Canaani, E., and Young, R. A. (2008). Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. Genes Dev *22*, 3403-3408.

Guo, J., Wang, J., Lee, J. S., and Schultz, P. G. (2008). Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. Angew Chem Int Ed Engl *47*, 6399-6401.

Hackenberger, C. P., and Schwarzer, D. (2008). Chemoselective ligation and modification strategies for peptides and proteins. Angew Chem Int Ed Engl *47*, 10030-10074.

Hall, I. M., Shankaranarayana, G. D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S. I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232-2237.

Hannah, A. (1951). Localization and function of heterochromatin in Drosophila melanogaster. Adv Genet *4*, 87-125.

Hansen, J. C. (2002). Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions. Annu Rev Biophys Biomol Struct *31*, 361-392.

He, S., Bauman, D., Davis, J. S., Loyola, A., Nishioka, K., Gronlund, J. L., Reinberg, D., Meng, F., Kelleher, N., and McCafferty, D. G. (2003). Facile synthesis of site-specifically acetylated and methylated histone proteins: reagents for evaluation of the histone code hypothesis. Proc Natl Acad Sci U S A *100*, 12033-12038.

Henikoff, S., and Ahmad, K. (2005). Assembly of variant histones into chromatin. Annu Rev Cell Dev Biol *21*, 133-153.

Henry, K. W., and Berger, S. L. (2002). Trans-tail histone modifications: wedge or bridge? Nat Struct Biol *9*, 565-566.

Henry, K. W., Wyce, A., Lo, W. S., Duggan, L. J., Emre, N. C., Kao, C. F., Pillus, L., Shilatifard, A., Osley, M. A., and Berger, S. L. (2003). Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. Genes Dev *17*, 2648-2663.

Hershko, A., Ciechanover, A., and Varshavsky, A. (2000). Basic Medical Research Award. The ubiquitin system. Nat Med *6*, 1073-1081.

Hirota, T., Lipp, J. J., Toh, B. H., and Peters, J. M. (2005). Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. Nature *438*, 1176-1180.

Hodgins, R. R., Ellison, K. S., and Ellison, M. J. (1992). Expression of a ubiquitin derivative that conjugates to protein irreversibly produces phenotypes consistent with a ubiquitin deficiency. J Biol Chem *267*, 8807-8812.

Huang, T. T., and D'Andrea, A. D. (2006). Regulation of DNA repair by ubiquitylation. Nat Rev Mol Cell Biol *7*, 323-334.

Hung, T., Binda, O., Champagne, K. S., Kuo, A. J., Johnson, K., Chang, H. Y., Simon, M. D., Kutateladze, T. G., and Gozani, O. (2009). ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. Mol Cell *33*, 248-256.

Hunt, L. T., and Dayhoff, M. O. (1977). Amino-terminal sequence identity of ubiquitin and the nonhistone component of nuclear protein A24. Biochem Biophys Res Commun *74*, 650-655.

Hurley, J. H., Lee, S., and Prag, G. (2006). Ubiquitin-binding domains. Biochem J *399*, 361-372.

Huyen, Y., Zgheib, O., Ditullio, R. A., Jr., Gorgoulis, V. G., Zacharatos, P., Petty, T. J., Sheston, E. A., Mellert, H. S., Stavridi, E. S., and Halazonetis, T. D. (2004). Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature *432*, 406-411.

Huynh, V. A., Robinson, P. J., and Rhodes, D. (2005). A method for the in vitro reconstitution of a defined "30 nm" chromatin fibre containing stoichiometric amounts of the linker histone. J Mol Biol *345*, 957-968.

Hwang, W. W., Venkatasubrahmanyam, S., Ianculescu, A. G., Tong, A., Boone, C., and Madhani, H. D. (2003). A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. Mol Cell *11*, 261-266.

Ito, T., Levenstein, M. E., Fyodorov, D. V., Kutach, A. K., Kobayashi, R., and Kadonaga, J. T. (1999). ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. Genes Dev *13*, 1529-1539.

Jacobson, R. H., Ladurner, A. G., King, D. S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. Science *288*, 1422-1425.

Johnson, D. R. (1974). Hairpin-tail: a case of post-reductional gene action in the mouse egg. Genetics *76*, 795-805.

Jones, B., Su, H., Bhat, A., Lei, H., Bajko, J., Hevi, S., Baltus, G. A., Kadam, S., Zhai, H., Valdez, R.*, et al.* (2008). The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. PLoS Genet *4*, e1000190.

Kamakaka, R. T., and Biggins, S. (2005). Histone variants: deviants? Genes Dev *19*, 295-310.

Kang, R. S., Daniels, C. M., Francis, S. A., Shih, S. C., Salerno, W. J., Hicke, L., and Radhakrishnan, I. (2003). Solution structure of a CUE-ubiquitin complex reveals a conserved mode of ubiquitin binding. Cell *113*, 621-630.

Kang, T. J., Yuzawa, S., and Suga, H. (2008). Expression of histone H3 tails with combinatorial lysine modifications under the reprogrammed genetic code for the investigation on epigenetic markers. Chem Biol *15*, 1166-1174.

Kao, C. F., Hillyer, C., Tsukuda, T., Henry, K., Berger, S., and Osley, M. A. (2004). Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. Genes Dev *18*, 184-195.

Kaplan, C. D., Laprade, L., and Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. Science *301*, 1096-1099.

Katzmann, D. J., Odorizzi, G., and Emr, S. D. (2002). Receptor downregulation and multivesicular-body sorting. Nat Rev Mol Cell Biol *3*, 893-905.

Kawakami, T., and Aimoto, S. (2003). A photoremovable ligation auxiliary for use in polypeptide synthesis. Tetrahedron Lett *44*, 6059-6061.

Kenyon, G. L., and Bruice, T. W. (1977). Novel sulfhydryl reagents. Methods Enzymol *47*, 407-430.

Kim, J., Guermah, M., McGinty, R. K., Lee, J. S., Tang, Z., Milne, T. A., Shilatifard, A., Muir, T. W., and Roeder, R. G. (2009). RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. Cell *137*, 459-471.

Kim, J., Hake, S. B., and Roeder, R. G. (2005). The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. Mol Cell *20*, 759-770.

Kim, J., and Roeder, R. G. (2009). Direct Bre1-Paf1 Complex Interactions and RING Finger-independent Bre1-Rad6 Interactions Mediate Histone H2B Ubiquitylation in Yeast. J Biol Chem *284*, 20582-20592.

Kohler, A., Pascual-Garcia, P., Llopis, A., Zapater, M., Posas, F., Hurt, E., and Rodriguez-Navarro, S. (2006). The mRNA export factor Sus1 is involved in Spt/Ada/Gcn5 acetyltransferase-mediated H2B deubiquitinylation through its interaction with Ubp8 and Sgf11. Mol Biol Cell *17*, 4228-4236.

Kornberg, R. D. (1977). Structure of chromatin. Annu Rev Biochem *46*, 931-954.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Krivtsov, A. V., and Armstrong, S. A. (2007). MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer *7*, 823-833.

Krivtsov, A. V., Feng, Z., Lemieux, M. E., Faber, J., Vempati, S., Sinha, A. U., Xia, X., Jesneck, J., Bracken, A. P., Silverman, L. B*., et al.* (2008). H3K79 methylation profiles define murine and human MLL-AF4 leukemias. Cancer Cell *14*, 355-368.

Lacoste, N., Utley, R. T., Hunter, J. M., Poirier, G. G., and Cote, J. (2002). Disruptor of telomeric silencing-1 is a chromatin-specific histone H3 methyltransferase. J Biol Chem *277*, 30421-30424.

Laribee, R. N., Fuchs, S. M., and Strahl, B. D. (2007). H2B ubiquitylation in transcriptional control: a FACT-finding mission. Genes Dev *21*, 737-743.

Laribee, R. N., Krogan, N. J., Xiao, T., Shibata, Y., Hughes, T. R., Greenblatt, J. F., and Strahl, B. D. (2005). BUR kinase selectively regulates H3 K4 trimethylation and H2B ubiquitylation through recruitment of the PAF elongation complex. Curr Biol *15*, 1487-1493.

Larsen, C. N., Krantz, B. A., and Wilkinson, K. D. (1998). Substrate specificity of deubiquitinating enzymes: ubiquitin C-terminal hydrolases. Biochemistry *37*, 3358-3368.

Layfield, R., Franklin, K., Landon, M., Walker, G., Wang, P., Ramage, R., Brown, A., Love, S., Urquhart, K., Muir, T.*, et al.* (1999). Chemically synthesized ubiquitin extension proteins detect distinct catalytic capacities of deubiquitinating enzymes. Anal Biochem *274*, 40-49.

Lee, J. S., Shukla, A., Schneider, J., Swanson, S. K., Washburn, M. P., Florens, L., Bhaumik, S. R., and Shilatifard, A. (2007). Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. Cell *131*, 1084-1096.

Lee, N., Erdjument-Bromage, H., Tempst, P., Jones, R. S., and Zhang, Y. (2009). The H3K4 demethylase lid associates with and inhibits histone deacetylase Rpd3. Mol Cell Biol *29*, 1401-1410.

Lee, S., Lee, D. K., Dou, Y., Lee, J., Lee, B., Kwak, E., Kong, Y. Y., Lee, S. K., Roeder, R. G., and Lee, J. W. (2006). Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. Proc Natl Acad Sci U S A *103*, 15392-15397.

Lehnertz, B., Ueda, Y., Derijck, A. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A. H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol *13*, 1192-1200.

Li, B., Jackson, J., Simon, M. D., Fleharty, B., Gogol, M., Seidel, C., Workman, J. L., and Shilatifard, A. (2009). Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. J Biol Chem *284*, 7970-7976.

Li, S., and Shogren-Knaak, M. A. (2008). Cross-talk between histone H3 tails produces cooperative nucleosome acetylation. Proc Natl Acad Sci U S A *105*, 18243-18248.

Li, S., and Shogren-Knaak, M. A. (2009). The Gcn5 bromodomain of the SAGA complex facilitates cooperative and cross-tail acetylation of nucleosomes. J Biol Chem *284*, 9411-9417.

Lin, C. H., Li, B., Swanson, S., Zhang, Y., Florens, L., Washburn, M. P., Abmayr, S. M., and Workman, J. L. (2008). Heterochromatin protein 1a stimulates histone H3 lysine 36 demethylation by the Drosophila KDM4A demethylase. Mol Cell *32*, 696-706.

Liu, Y., Shah, K., Yang, F., Witucki, L., and Shokat, K. M. (1998). Engineering Src family protein kinases with unnatural nucleotide specificity. Chem Biol *5*, 91-101.

Low, D. W., Hill, M. G., Carrasco, M. R., Kent, S. B., and Botti, P. (2001). Total synthesis of cytochrome b562 by native chemical ligation using a removable auxiliary. Proc Natl Acad Sci U S A *98*, 6554-6559.

Lowary, P. T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol *276*, 19-42.

Lu, X., Simon, M. D., Chodaparambil, J. V., Hansen, J. C., Shokat, K. M., and Luger, K. (2008). The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. Nat Struct Mol Biol *15*, 1122-1124.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997a). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M., and Richmond, T. J. (1997b). Characterization of nucleosome core particles containing histone proteins made in bacteria. J Mol Biol *272*, 301-311.

Luger, K., Rechsteiner, T. J., and Richmond, T. J. (1999). Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol Biol *119*, 1-16.

Macdonald, N., Welburn, J. P., Noble, M. E., Nguyen, A., Yaffe, M. B., Clynes, D., Moggs, J. G., Orphanides, G., Thomson, S., Edmunds, J. W*., et al.* (2005). Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3. Mol Cell *20*, 199-211.

Macmillan, D., and Bertozzi, C. (2000). New Directions in Glycoprotein Engineering. Tetrahedron *56*, 9515-9525.

Macmillan, D., and Bertozzi, C. R. (2004). Modular assembly of glycoproteins: towards the synthesis of GlyCAM-1 by using expressed protein ligation. Angew Chem Int Ed Engl *43*, 1355-1359.

Mallery, D. L., Vandenberg, C. J., and Hiom, K. (2002). Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiquitin chains. Embo J *21*, 6755-6762.

Manohar, M., Mooney, A. M., North, J. A., Nakkula, R. J., Picking, J. W., Edon, A., Fishel, R., Poirier, M. G., and Ottesen, J. J. (2009). Acetylation of Histone H3 at the Nucleosome Dyad Alters DNA-Histone Binding. J Biol Chem *284*, 23312-23321.

Marinzi, C., Bark, S. J., Offer, J., and Dawson, P. E. (2001). A new scaffold for amide ligation. Bioorg Med Chem *9*, 2323-2328.

Marinzi, C., Offer, J., Longhi, R., and Dawson, P. E. (2004). An o-nitrobenzyl scaffold for peptide ligation: synthesis and applications. Bioorg Med Chem *12*, 2749-2757.

Martin, C., and Zhang, Y. (2005). The diverse functions of histone lysine methylation. Nat Rev Mol Cell Biol *6*, 838-849.

Mason, P. B., and Struhl, K. (2003). The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. Mol Cell Biol *23*, 8323-8333.

McClaren, A. (1979). Maternal effects in development: The fourth symposium of the British Societ for Developmental Biology (Cambridge, UK: Cambridge University Press).

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A *36*, 344-355.

McGhee, J. D., and Felsenfeld, G. (1980). Nucleosome structure. Annu Rev Biochem *49*, 1115-1156.

McGinty, R. K., Kim, J., Chatterjee, C., Roeder, R. G., and Muir, T. W. (2008). Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. Nature *453*, 812-816.

McMinn, D., and Greenberg, M. (1996). Novel solid phase synthesis supports for the preparation of oligonucleotides containing 3'-alkyl amines. Tetrahedron *52*, 3827-3840.

Meselson, M., and Stahl, F. W. (1958). The Replication of DNA in Escherichia Coli. Proc Natl Acad Sci U S A *44*, 671-682.

Min, J., Feng, Q., Li, Z., Zhang, Y., and Xu, R. M. (2003). Structure of the catalytic domain of human DOT1L, a non-SET domain nucleosomal histone methyltransferase. Cell *112*, 711-723.

Minsky, N., and Oren, M. (2004). The RING domain of Mdm2 mediates histone ubiquitylation and transcriptional repression. Mol Cell *16*, 631-639.

Minsky, N., Shema, E., Field, Y., Schuster, M., Segal, E., and Oren, M. (2008). Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. Nat Cell Biol *10*, 483-488.

Misaghi, S., Galardy, P. J., Meester, W. J., Ovaa, H., Ploegh, H. L., and Gaudet, R. (2005). Structure of the ubiquitin hydrolase UCH-L3 complexed with a suicide substrate. J Biol Chem *280*, 1512-1520.

Mossessova, E., and Lima, C. D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol Cell *5*, 865-876.

Muir, T. W., Sondhi, D., and Cole, P. A. (1998). Expressed protein ligation: a general method for protein engineering. Proc Natl Acad Sci U S A *95*, 6705-6710.

Muller, H. J. (1930). Types of visible variations induced by X-rays in Drosophila. J Genet *22*, 299-334.

Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006). A highly flexible tRNA acylation method for non-natural polypeptide synthesis. Nat Methods *3*, 357-359.

Muralidharan, V., and Muir, T. W. (2006). Protein ligation: an enabling technology for the biophysical analysis of proteins. Nat Methods *3*, 429-438.

Muthurajan, U. M., Bao, Y., Forsberg, L. J., Edayathumangalam, R. S., Dyer, P. N., White, C. L., and Luger, K. (2004). Crystal structures of histone Sin mutant nucleosomes reveal altered protein-DNA interactions. Embo J *23*, 260-271.

Mutiu, A. I., Hoke, S. M., Genereaux, J., Liang, G., and Brandl, C. J. (2007). The role of histone ubiquitylation and deubiquitylation in gene expression as determined by the analysis of an HTB1(K123R) Saccharomyces cerevisiae strain. Mol Genet Genomics *277*, 491-506.

Nakanishi, S., Sanderson, B. W., Delventhal, K. M., Bradford, W. D., Staehling-Hampton, K., and Shilatifard, A. (2008). A comprehensive library of histone mutants identifies nucleosomal residues required for H3K4 methylation. Nat Struct Mol Biol *15*, 881-888.

Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2008). Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. Nat Chem Biol *4*, 232-234.

Ng, H. H., Ciccone, D. N., Morshead, K. B., Oettinger, M. A., and Struhl, K. (2003a). Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation. Proc Natl Acad Sci U S A *100*, 1820-1825.

Ng, H. H., Dole, S., and Struhl, K. (2003b). The Rtf1 component of the Paf1 transcriptional elongation complex is required for ubiquitination of histone H2B. J Biol Chem *278*, 33625-33628.

Ng, H. H., Feng, Q., Wang, H., Erdjument-Bromage, H., Tempst, P., Zhang, Y., and Struhl, K. (2002a). Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. Genes Dev *16*, 1518-1527.

Ng, H. H., Xu, R. M., Zhang, Y., and Struhl, K. (2002b). Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79. J Biol Chem *277*, 34655-34657.

Nicassio, F., Corrado, N., Vissers, J. H., Areces, L. B., Bergink, S., Marteijn, J. A., Geverts, B., Houtsmuller, A. B., Vermeulen, W., Di Fiore, P. P., and Citterio, E. (2007). Human USP3 is a chromatin modifier required for S phase progression and genome stability. Curr Biol *17*, 1972-1977.

Offer, J., Boddy, C. N., and Dawson, P. E. (2002). Extending synthetic access to proteins with a removable acyl transfer auxiliary. J Am Chem Soc *124*, 4642-4646.

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V. M., Su, L., Xu, G., and Zhang, Y. (2005). hDOT1L links histone methylation to leukemogenesis. Cell *121*, 167-178.

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Ong, M. S., Richmond, T. J., and Davey, C. A. (2007). DNA stretching and extreme kinking in the nucleosome core. J Mol Biol *368*, 1067-1074.

Osley, M. A. (2006). Regulation of histone H2A and H2B ubiquitylation. Brief Funct Genomic Proteomic *5*, 179-189.

Owen-Hughes, T., Utley, R. T., Steger, D. J., West, J. M., John, S., Cote, J., Havas, K. M., and Workman, J. L. (1999). Analysis of nucleosome disruption by ATP-driven chromatin remodeling complexes. Methods Mol Biol *119*, 319-331.

Park, J. H., Cosgrove, M. S., Youngman, E., Wolberger, C., and Boeke, J. D. (2002). A core nucleosome surface crucial for transcriptional silencing. Nat Genet *32*, 273-279.

Patnaik, D., Chin, H. G., Esteve, P. O., Benner, J., Jacobsen, S. E., and Pradhan, S. (2004). Substrate specificity and kinetic mechanism of mammalian G9a histone H3 methyltransferase. J Biol Chem *279*, 53248-53258.

Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. Cell *125*, 703-717.

Pellois, J. P., and Muir, T. W. (2005). A ligation and photorelease strategy for the temporal and spatial control of protein function in living cells. Angew Chem Int Ed Engl *44*, 5713-5717.

Pellois, J. P., and Muir, T. W. (2006). Semisynthetic proteins in mechanistic studies: using chemistry to go where nature can't. Curr Opin Chem Biol *10*, 487-491.

Pentelute, B. L., and Kent, S. B. (2007). Selective desulfurization of cysteine in the presence of Cys(Acm) in polypeptides obtained by native chemical ligation. Org Lett *9*, 687-690.

Phair, R. D., Scaffidi, P., Elbi, C., Vecerova, J., Dey, A., Ozato, K., Brown, D. T., Hager, G., Bustin, M., and Misteli, T. (2004). Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. Mol Cell Biol *24*, 6393-6402.

Pickart, C. M. (2001). Mechanisms underlying ubiquitination. Annu Rev Biochem *70*, 503-533.

Pickart, C. M., Kasperek, E. M., Beal, R., and Kim, A. (1994). Substrate properties of site-specific mutant ubiquitin protein (G76A) reveal unexpected mechanistic features of ubiquitin-activating enzyme (E1). J Biol Chem *269*, 7115-7123.

Pylypenko, O., Rak, A., Durek, T., Kushnir, S., Dursina, B. E., Thomae, N. H., Constantinescu, A. T., Brunsveld, L., Watzke, A., Waldmann, H.*, et al.* (2006). Structure of doubly prenylated Ypt1:GDI complex and the mechanism of GDI-mediated Rab recycling. Embo J *25*, 13-23.

Raiborg, C., Rusten, T. E., and Stenmark, H. (2003). Protein sorting into multivesicular endosomes. Curr Opin Cell Biol *15*, 446-455.

Rak, A., Pylypenko, O., Durek, T., Watzke, A., Kushnir, S., Brunsveld, L., Waldmann, H., Goody, R. S., and Alexandrov, K. (2003). Structure of Rab GDP-dissociation inhibitor in complex with prenylated YPT1 GTPase. Science *302*, 646-650.

Ramage, R., Green, J., Muir, T. W., Ogunjobi, O. M., Love, S., and Shaw, K. (1994). Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. Biochem J *299 ( Pt 1)*, 151-158.

Reddy, K., and Sharpless, K. (1998). From styrenes to enantiopure alpha-arylglycines in two steps. J Am Chem Soc *120*, 1207-1217.

Reinberg, D., and Sims, R. J., 3rd (2006). de FACTo nucleosome dynamics. J Biol Chem *281*, 23297-23301.

Robinson, P. J., An, W., Routh, A., Martino, F., Chapman, L., Roeder, R. G., and Rhodes, D. (2008). 30 nm chromatin fibre decompaction requires both H4-K16 acetylation and linker histone eviction. J Mol Biol *381*, 816-825.

Robinson, P. J., Fairall, L., Huynh, V. A., and Rhodes, D. (2006). EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. Proc Natl Acad Sci U S A *103*, 6506-6511.

Robinson, P. J., and Rhodes, D. (2006). Structure of the '30 nm' chromatin fibre: a key role for the linker histone. Curr Opin Struct Biol *16*, 336-343.

Robzyk, K., Recht, J., and Osley, M. A. (2000). Rad6-dependent ubiquitination of histone H2B in yeast. Science *287*, 501-504.

Rozwadowska, M. (1997). Reaction of *(lS,2S)*-2-amino-l-phenyl-l,3-propanediol with thioacids under the
Mitsunobu reaction condition Tetrahedron *53*.

Ruthenburg, A. J., Allis, C. D., and Wysocka, J. (2007a). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Mol Cell *25*, 15-30.

Ruthenburg, A. J., Li, H., Patel, D. J., and Allis, C. D. (2007b). Multivalent engagement of chromatin modifications by linked binding modules. Nat Rev Mol Cell Biol *8*, 983-994.

Sarcevic, B., Mawson, A., Baker, R. T., and Sutherland, R. L. (2002). Regulation of the ubiquitin-conjugating enzyme hHR6A by CDK-mediated phosphorylation. Embo J *21*, 2009-2018.

Sawada, K., Yang, Z., Horton, J. R., Collins, R. E., Zhang, X., and Cheng, X. (2004). Structure of the conserved core of the yeast Dot1p, a nucleosomal histone H3 lysine 79 methyltransferase. J Biol Chem *279*, 43296-43306.

Schalch, T., Duda, S., Sargent, D. F., and Richmond, T. J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138-141.

Schwartz, B. E., and Ahmad, K. (2005). Transcriptional activation triggers deposition and removal of the histone variant H3.3. Genes Dev *19*, 804-814.

Shah, K., Liu, Y., Deirmengian, C., and Shokat, K. M. (1997). Engineering unnatural nucleotide specificity for Rous sarcoma virus tyrosine kinase to uniquely label its direct substrates. Proc Natl Acad Sci U S A *94*, 3565-3570.

Shahbazian, M. D., Zhang, K., and Grunstein, M. (2005). Histone H2B ubiquitylation controls processive methylation but not monomethylation by Dot1 and Set1. Mol Cell *19*, 271-277.

Shih, S. C., Katzmann, D. J., Schnell, J. D., Sutanto, M., Emr, S. D., and Hicke, L. (2002). Epsins and Vps27p/Hrs contain ubiquitin-binding domains that function in receptor endocytosis. Nat Cell Biol *4*, 389-393.

Shilatifard, A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu Rev Biochem *75*, 243-269.

Shogren-Knaak, M., Ishii, H., Sun, J. M., Pazin, M. J., Davie, J. R., and Peterson, C. L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844-847.

Shogren-Knaak, M. A., Fry, C. J., and Peterson, C. L. (2003). A native peptide ligation strategy for deciphering nucleosomal histone modifications. J Biol Chem *278*, 15744-15748.

Shogren-Knaak, M. A., and Peterson, C. L. (2004). Creating designer histones by native chemical ligation. Methods Enzymol *375*, 62-76.

Shukla, A., and Bhaumik, S. R. (2007). H2B-K123 ubiquitination stimulates RNAPII elongation independent of H3-K4 methylation. Biochem Biophys Res Commun *359*, 214-220.

Simon, M. D., Chu, F., Racki, L. R., de la Cruz, C. C., Burlingame, A. L., Panning, B., Narlikar, G. J., and Shokat, K. M. (2007). The site-specific installation of methyl-lysine analogs into recombinant histones. Cell *128*, 1003-1012.

Singer, M. S., Kahana, A., Wolf, A. J., Meisinger, L. L., Peterson, S. E., Goggin, C., Mahowald, M., and Gottschling, D. E. (1998). Identification of high-copy disruptors of telomeric silencing in Saccharomyces cerevisiae. Genetics *150*, 613-632.

Smith, A. B., 3rd, Savinov, S. N., Manjappara, U. V., and Chaiken, I. M. (2002). Peptide-small molecule hybrids via orthogonal deprotection-chemoselective conjugation to cysteine-anchored scaffolds. A model study. Org Lett *4*, 4041-4044.

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Sun, Z. W., and Allis, C. D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. Nature *418*, 104-108.

Suto, R. K., Clarkson, M. J., Tremethick, D. J., and Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat Struct Biol *7*, 1121-1124.

Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., and Luger, K. (2003). Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. J Mol Biol *326*, 371-380.

Tanny, J. C., Erdjument-Bromage, H., Tempst, P., and Allis, C. D. (2007). Ubiquitylation of histone H2B controls RNA polymerase II transcription elongation independently of histone H3 methylation. Genes Dev *21*, 835-847.

Taunton, J., Hassig, C. A., and Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science *272*, 408-411.

Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol *14*, 1025-1040.

Thambirajah, A. A., Li, A., Ishibashi, T., and Ausio, J. (2009). New developments in post-translational modifications and functions of histone H2A variants. Biochem Cell Biol *87*, 7-17.

Thorne, A. W., Sautiere, P., Briand, G., and Crane-Robinson, C. (1987). The structure of ubiquitinated histone H2B. Embo J *6*, 1005-1010.

Tirat, A., Schilb, A., Riou, V., Leder, L., Gerhartz, B., Zimmermann, J., Worpenberg, S., Eidhoff, U., Freuler, F., Stettler, T.*, et al.* (2005). Synthesis and characterization of fluorescent ubiquitin derivatives as highly sensitive substrates for the deubiquitinating enzymes UCH-L3 and USP-2. Anal Biochem *343*, 244-255.

Tolbert, T., and Wong, C.-H. (2000). Intein-Mediated Synthesis of Proteins Containing Carbohydrates and Other Molecular Probes. J Am Chem Soc *122*, 542105428.

Trievel, R. C., Beach, B. M., Dirk, L. M., Houtz, R. L., and Hurley, J. H. (2002). Structure and catalytic mechanism of a SET domain protein methyltransferase. Cell *111*, 91-103.

Tsunaka, Y., Kajimura, N., Tate, S., and Morikawa, K. (2005). Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. Nucleic Acids Res *33*, 3424-3434.

van der Knaap, J. A., Kumar, B. R., Moshkin, Y. M., Langenberg, K., Krijgsveld, J., Heck, A. J., Karch, F., and Verrijzer, C. P. (2005). GMP synthetase stimulates histone H2B deubiquitylation by the epigenetic silencer USP7. Mol Cell *17*, 695-707.

van Leeuwen, F., Gafken, P. R., and Gottschling, D. E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell *109*, 745-756.

Varshavsky, A. (2005). Ubiquitin fusion technique and related methods. Methods Enzymol *399*, 777-799.

Vijay-Kumar, S., Bugg, C. E., and Cook, W. J. (1987). Structure of ubiquitin refined at 1.8 A resolution. J Mol Biol *194*, 531-544.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J. M.*, et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.

Vitaliano-Prunier, A., Menant, A., Hobeika, M., Geli, V., Gwizdek, C., and Dargemont, C. (2008). Ubiquitylation of the COMPASS component Swd2 links H2B ubiquitylation to H3K4 trimethylation. Nat Cell Biol *10*, 1365-1371.

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I., and Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Wahl, G. M., and Carr, A. M. (2001). The evolution of diverse biological responses to DNA damage: insights from yeast and p53. Nat Cell Biol *3*, E277-286.

Walsh, C. T. (2005). Posttranslational Modification of Proteins: Expanding Nature's Inventory (Portland, OR: Roberts & Company Publishers).

Wan, Q., and Danishefsky, S. J. (2007). Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. Angew Chem Int Ed Engl *46*, 9248-9252.

Wang, L., Xie, J., and Schultz, P. G. (2006). Expanding the genetic code. Annu Rev Biophys Biomol Struct *35*, 225-249.

Wang, S. S., Gisin, B. F., Winter, D. P., Makofske, R., Kulesha, I. D., Tzougraki, C., and Meienhofer, J. (1977). Facile synthesis of amino acid and peptide esters under mild conditions via cesium salts. J Org Chem *42*, 1286-1290.

Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Peng, W., Zhang, M. Q., and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897-903.

Watt, F., and Molloy, P. L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev *2*, 1136-1143.

Weake, V. M., and Workman, J. L. (2008). Histone ubiquitination: triggering gene activity. Mol Cell *29*, 653-663.

West, M. H., and Bonner, W. M. (1980). Histone 2B can be modified by the attachment of ubiquitin. Nucleic Acids Res *8*, 4671-4680.

White, C. L., Suto, R. K., and Luger, K. (2001). Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. Embo J *20*, 5207-5218.

Wolffe, A. P., and Hayes, J. J. (1999). Chromatin disruption and modification. Nucleic Acids Res *27*, 711-720.

Wood, A., Krogan, N. J., Dover, J., Schneider, J., Heidt, J., Boateng, M. A., Dean, K., Golshani, A., Zhang, Y., Greenblatt, J. F*., et al.* (2003a). Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Mol Cell *11*, 267-274.

Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2003b). The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. J Biol Chem *278*, 34739-34742.

Wood, A., Schneider, J., Dover, J., Johnston, M., and Shilatifard, A. (2005). The Bur1/Bur2 complex is required for histone H2B monoubiquitination by Rad6/Bre1 and histone methylation by COMPASS. Mol Cell *20*, 589-599.

Wu, B., Droge, P., and Davey, C. A. (2008). Site selectivity of platinum anticancer therapeutics. Nat Chem Biol *4*, 110-112.

Wu, J. W., Hu, M., Chai, J., Seoane, J., Huse, M., Li, C., Rigotti, D. J., Kyin, S., Muir, T. W., Fairman, R*., et al.* (2001). Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling. Mol Cell *8*, 1277-1289.

Wyce, A., Xiao, T., Whelan, K. A., Kosman, C., Walter, W., Eick, D., Hughes, T. R., Krogan, N. J., Strahl, B. D., and Berger, S. L. (2007). H2B ubiquitylation acts as a barrier to Ctk1

nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related complex. Mol Cell *27*, 275-288.

Wysocki, R., Javaheri, A., Allard, S., Sha, F., Cote, J., and Kron, S. J. (2005). Role of Dot1-dependent histone H3 methylation in G1 and S phase DNA damage checkpoint functions of Rad9. Mol Cell Biol *25*, 8430-8443.

Xia, Y., Pao, G. M., Chen, H. W., Verma, I. M., and Hunter, T. (2003). Enhancement of BRCA1 E3 ubiquitin ligase activity through direct interaction with the BARD1 protein. J Biol Chem *278*, 5255-5263.

Xiao, T., Kao, C. F., Krogan, N. J., Sun, Z. W., Greenblatt, J. F., Osley, M. A., and Strahl, B. D. (2005). Histone H2B ubiquitylation is associated with elongating RNA polymerase II. Mol Cell Biol *25*, 637-651.

Xu, P., Duong, D. M., Seyfried, N. T., Cheng, D., Xie, Y., Robert, J., Rush, J., Hochstrasser, M., Finley, D., and Peng, J. (2009). Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. Cell *137*, 133-145.

Yan, L. Z., and Dawson, P. E. (2001). Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. J Am Chem Soc *123*, 526-533.

Yang, B., Britton, J., and Kirchmaier, A. L. (2008). Insights into the impact of histone acetylation and methylation on Sir protein recruitment, spreading, and silencing in Saccharomyces cerevisiae. J Mol Biol *381*, 826-844.

Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A., and Korsmeyer, S. J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. Nature *378*, 505-508.

Zeng, L., and Zhou, M. M. (2002). Bromodomain: an acetyl-lysine binding domain. FEBS Lett *513*, 124-128.

Zhang, W., Xia, X., Reisenauer, M. R., Hemenway, C. S., and Kone, B. C. (2006). Dot1a-AF9 complex mediates histone H3 Lys-79 hypermethylation and repression of ENaCalpha in an aldosterone-sensitive manner. J Biol Chem *281*, 18059-18068.

Zhang, X. Y., Varthi, M., Sykes, S. M., Phillips, C., Warzecha, C., Zhu, W., Wyce, A., Thorne, A. W., Berger, S. L., and McMahon, S. B. (2008). The putative cancer stem cell marker USP22 is a subunit of the human SAGA complex required for activated transcription and cell-cycle progression. Mol Cell *29*, 102-111.

Zhao, Y., Lang, G., Ito, S., Bonnet, J., Metzger, E., Sawatsubashi, S., Suzuki, E., Le Guezennec, X., Stunnenberg, H. G., Krasnov, A*., et al.* (2008). A TFTC/STAGA module mediates histone H2A and H2B deubiquitination, coactivates nuclear receptors, and counteracts heterochromatin silencing. Mol Cell *29*, 92-101.

Zheng, C., and Hayes, J. J. (2003). Intra- and inter-nucleosomal protein-DNA interactions of the core histone tail domains in a model system. J Biol Chem *278*, 24217-24224.

Zhu, B., Zheng, Y., Pham, A. D., Mandal, S. S., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2005). Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. Mol Cell *20*, 601-611.