

# **A Method for the Quantitative Analysis of Protein-Protein Interactions *In Vivo***



**Dissertation**

for the award of the degree

“Doctor rerum naturalium”

of the Georg-August-Universität Göttingen

within the doctoral program *Biomolecules: Structure-Function-Dynamics*  
of the Georg-August University School of Science (GAUSS)

submitted by

**Nils Arne Rall**

from Buchholz i. d. N.

**Göttingen, 2016**

### **Thesis Committee**

Prof. Dr. Heinz Neumann, Applied Synthetic Biology Group, University of Göttingen

Prof. Dr. Henning Urlaub, Max Planck Research Group Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry

Prof. Dr. Matthias Dobbelstein, Molecular Oncology, University of Göttingen

### **Members of the Examination Board**

Referee:

Prof. Dr. Heinz Neumann,  
Applied Synthetic Biology Group, University of Göttingen

2<sup>nd</sup> Referee:

Prof. Dr. Henning Urlaub,  
Max Planck Research Group Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry

Prof. Dr. Matthias Dobbelstein,  
Molecular Oncology, University of Göttingen

### **Members of the Extended Examination Board**

Prof. Dr. Blanche Schwappach, Membrane Protein Biogenesis, University of Göttingen

Prof. Dr. Steven Johnsen, Translational Cancer Research, University of Göttingen

Prof. Dr. Andre Fischer, DZNE Research Group Epigenetic Mechanism in Dementia, Deutsches Zentrum für Neurodegenerative Erkrankungen

**Date of oral examination:** 22.03.2016

# Table of Contents

List of Figures.....	III
List of Tables .....	V
Abbreviations .....	VI
1 Introduction.....	1
1.1 Chromatin Structure and Function .....	1
1.2 Histones and Nucleosomes .....	2
1.3 Post-Translational Modifications of Histones .....	4
1.4 Nucleosome Dynamics .....	6
1.5 Higher-Order Chromatin Structure.....	7
1.6 Chromatin Condensation .....	10
1.7 Expansion of the Genetic Code .....	13
1.8 Mass Spectrometry-Based Proteomics .....	19
1.9 Mass Spectrometry-Based Quantification .....	21
2 Research Objective .....	25
3 Materials and Methods.....	27
3.1 Materials.....	27
3.1.1 Devices and Instruments.....	27
3.1.2 Chemicals .....	28
3.1.3 Consumables and Other Materials.....	28
3.1.4 Protein Size Standards .....	29
3.1.5 Culture Media and Agar Plates.....	29
3.1.6 Antibiotics .....	30
3.1.7 Unnatural Amino Acids .....	30
3.1.8 Isotopic Labeled Amino Acids .....	30
3.1.9 Enzymes .....	30
3.1.10 Cell Lines.....	31
3.1.11 Plasmids.....	31
3.1.12 Buffers and Solutions .....	33
3.1.13 DNA Kit Systems .....	35
3.1.14 Antibodies.....	35
3.1.15 Conjugated Beads .....	35
3.2 Methods .....	36
3.2.1 Microbiological Methods .....	36
3.2.2 Protein Biochemical Methods .....	36
3.2.3 Nucleic Acid Biochemical Methods .....	44
4 Results.....	45
4.1 <i>In Vivo</i> Crosslinking of Histones in <i>S. cerevisiae</i> .....	45
4.1.1 Scanning of Histone Surfaces with Genetically Encoded UV-Inducible Crosslinker .....	45
4.1.2 Spatio-Temporal Scanning of Binding Interfaces of N-Terminal H4 Domain .....	46

4.2	Isolation of Histone pBPA Mutants and Crosslink Products .....	48
4.2.1	Immunoprecipitation of pBPA Crosslink Products.....	48
4.2.2	Optimization of Immunoprecipitation Procedure for SILAC Experiments.....	50
4.3	Identification of Crosslink Products by MS-Based SILAC Approach .....	60
4.3.1	SILAC Labeling Efficient in BY4741 Deletion Strain.....	61
4.3.2	Optimization of SILAC Procedure Successfully Produces Peptide Pairs .....	62
4.3.3	Low Identification Yield Independent of Histone Variant or Crosslinker Positioning .....	64
4.3.4	Verification of Histone Interaction Partners by Molecular Shift Assay.....	68
4.3.5	Bidirectionally Labeled Samples for Verification of Histone Interaction Partners .....	69
4.3.6	Quantity of Protein Identifications Independent of MaxQuant Setup .....	75
4.3.7	Enrichment of Crosslink Products by Acid Extraction or Hydrazide Reactivity.....	77
4.3.8	Improving Result Evaluation by Introduction of Internal Quality Control .....	80
4.3.9	Enhanced Immunoprecipitation Improves Overall Quality of SILAC Analysis .....	81
4.3.10	Improved Overall Quality Allows Reproduction of Results.....	95
5	Discussion.....	107
5.1	<i>In Vivo</i> Crosslinking of Histones in <i>S. cerevisiae</i> .....	107
5.2	Isolation of Histone pBPA Mutants and Crosslink Products .....	108
5.3	Identification of Crosslink Products by MS-based SILAC Approach.....	113
5.4	Concluding Remarks and Outlook .....	117
6	Acknowledgements.....	120
7	Appendix.....	121
8	References.....	122
9	Curriculum Vitae.....	150

## List of Figures

Figure 1.1: The various levels of chromatin organization .....	1
Figure 1.2: Structures of the core histones and the nucleosome core particle.....	2
Figure 1.3: Schematic illustration of interconnected nucleosome structures. ....	3
Figure 1.4: Schematic overview of histone post-translational modifications.....	4
Figure 1.5: Formation of higher-order chromatin structures. ....	8
Figure 1.6: Mitotic chromatin condensation driven by a cascade of histone modifications.....	13
Figure 1.7: Expansion of the genetic code .....	15
Figure 1.8: Genetically encoded crosslinker amino acids for <i>in vivo</i> crosslinking. ....	16
Figure 1.9: <i>In vivo</i> crosslinking by genetically encoded crosslinker amino acid pBPA. ....	17
Figure 1.10: Dependency of crosslink formation on pBPA incorporation and UV-irradiation.....	18
Figure 1.11: Subcellular localization of produced H2A amber mutants and crosslink products.....	19
Figure 1.12: Main principles of mass spectrometry modes. ....	21
Figure 1.13: Basic principle of Stable Isotope Labeling with Amino Acids in Cell Culture.....	23
Figure 4.1: Overview of crosslink patterns across the surfaces of histone H3 and H4. ....	45
Figure 4.2: Overview of crosslink patterns on N-terminal sites of histone H4 during mitosis. ....	47
Figure 4.3: Immunoprecipitation of pBPA crosslink products. ....	49
Figure 4.4: Immunoprecipitation of pBPA crosslink products. ....	50
Figure 4.5: Growth conditions and characterization of YPH499 $\Delta$ Arg4 strain. ....	52
Figure 4.6: Immunoprecipitation of pBPA crosslink products under SILAC conditions. ....	53
Figure 4.7: Increasing pBPA concentration improves full-length protein yield. ....	54
Figure 4.8: Increasing tRNA <sup>CUA</sup> levels increases pBPA incorporation efficiency.....	56
Figure 4.9: SILAC immunoprecipitation of H4 Q27pBPA crosslink products for MS analysis. ....	57
Figure 4.10: Improved tRNA <sup>CUA</sup> levels cause amber suppression of $\Delta$ LYS2 in YPH499 strain.....	58
Figure 4.11: Double deletion strain of BY4741 qualified for SILAC. ....	60
Figure 4.12: SILAC labeling efficient in BY4741 deletion strain.....	61
Figure 4.13: Optimization of SILAC procedure produces peptide pairs.....	62
Figure 4.14: Identification of proteins from H3 T6pBPA SILAC analysis .....	63
Figure 4.15: SILAC immunoprecipitation of H3 A21pBPA and H4 R17pBPA crosslink products. ....	65
Figure 4.16: Identification of proteins from H3 A21pBPA SILAC analysis. ....	66
Figure 4.17: Identification of proteins from H4 R17pBPA SILAC analysis. ....	67
Figure 4.18: Identification of H2A-H4 crosslink product by genomic tagging.....	69
Figure 4.19: SILAC immunoprecipitation of H3 T6pBPA crosslink products.....	70
Figure 4.20: Identification of proteins from H3 T6pBPA SILAC analysis. ....	71
Figure 4.21: Enrichment of crosslink products by acid extraction of free histone proteins. ....	78
Figure 4.22: Enrichment of crosslink products by hydrazide chemistry. ....	79
Figure 4.23: Verification of STH1 as interaction partner of histone H2A at position A61.....	80
Figure 4.24: Comparative SILAC immunoprecipitation of H2A A61pBPA crosslink products. ....	82

## List of Figures

---

Figure 4.25: Identification of proteins from H2A A61pBPA SILAC analysis with agarose beads. ....	83
Figure 4.26: Identification of proteins from H2A A61pBPA SILAC analysis with magnetic beads.....	84
Figure 4.27: Identification of chromatin-related proteins from H2A A61pBPA SILAC analysis. ....	90
Figure 4.28: SILAC immunoprecipitation of H2A A61pBPA crosslink products.....	95
Figure 4.29: Identification of proteins from H2A A61pBPA SILAC analysis. ....	96
Figure 4.30: Proteins with significance in both H2A A61pBPA data sets. ....	98
Figure 4.31: Identification of chromatin-related proteins from H2A A61pBPA SILAC analysis. ....	102
Figure 4.32: Chromatin-related proteins from H2A A61pBPA SILAC analysis .....	103
Figure 5.1: Hydrazide to benzophenone coupling reaction. ....	112
Figure 7.1: Immunoprecipitation of pBPA crosslink products under SILAC conditions. ....	121

## List of Tables

Table 3.1: Amino acid mixes for cell culturing .....	29
Table 3.2: Ingredients for LB, YPDA and SC media.....	29
Table 3.3: Overview of antibiotics used in growth media and agar plates .....	30
Table 3.4: Overview of unnatural amino acids used for genetic code expansion.....	30
Table 3.5: Overview of isotopic labeled amino acids used for SILAC .....	30
Table 3.6: Overview of cell lines used.....	31
Table 3.7: Overview of plasmids used for genetic code expansion.....	31
Table 3.8: Overview of commonly used buffers and solutions .....	33
Table 3.9: Overview of antibodies used for Western blotting .....	35
Table 3.10: Overview of antibody conjugated beads used for immunoprecipitation .....	35
Table 3.11: Composition of polyacrylamide gels for SDS-PAGE.....	38
Table 4.1: Potential crosslink products of H3 T6 with a log <sub>2</sub> SILAC ratio greater than 1.5 .....	64
Table 4.2: Potential crosslink products of H3 A21 with a log <sub>2</sub> SILAC ratio greater than 1.5 .....	68
Table 4.3: Potential crosslink products of H4 R17 with a log <sub>2</sub> SILAC ratio greater than 1.5 .....	68
Table 4.4: Potential crosslink products in forward and reverse sample of H3 T6pBPA.....	72
Table 4.5: Potential crosslink products in forward sample of H3 T6pBPA.....	72
Table 4.6: Potential crosslink products in reverse sample of H3 T6pBPA.....	73
Table 4.7: Proteins abundant in samples of cells grown in light media.....	73
Table 4.8: Proteins abundant in samples of cells grown in heavy media. ....	74
Table 4.9: Overview of applied changes to MaxQuant control parameters.....	75
Table 4.10: Colour code for Table 4.11 .....	76
Table 4.11: Overview of changes in protein identification produced by variation of MaxQuant parameters. ....	76
Table 4.12: Intersecting hits with high significance in the magnetic bead samples of H2A A61pBPA. ....	85
Table 4.13: Intersecting hits with high significance in the agarose bead samples of H2A A61pBPA. ....	85
Table 4.14: Intersecting hits with high significance in both data sets of H2A A61pBPA. ....	86
Table 4.15: Significant hits solely detected in the magnetic bead sample of H2A A61pBPA.....	86
Table 4.16: Significant hits solely detected in the agarose bead sample of H2A A61pBPA. ....	87
Table 4.17: Comparison of exclusive protein hits with chromatin biological background in both samples ...	87
Table 4.18: Intersecting hits with high significance in both data sets of H2A A61pBPA SILAC analysis .....	97
Table 4.19: Intersecting hits with chromatin biological background in both data sets of H2A A61pBPA SILAC analysis.....	100

## Abbreviations

µg	microgram
µL	microliter
µm	micrometer
AA	amino acid
aaRS/RS	aminoacyl-tRNA-synthetase
ABK	3'-azibutyl-N-carbamoyl-lysine
ACN	Acetonitrile
AG	<i>Arbeitsgruppe</i>
Amp	ampicillin
bp	base pair(s)
BPARS	BPA aminoacyl-tRNA-synthetase
BSA	bovine serum albumin
Cm	Chloramphenicol
DiZPK	3-(3-methyl-3 <i>H</i> -diazirine-3-yl)-propaminocarbonyl-Nε-L-lysine
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
e.g.	exempli gratia
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
et al.	et alii/aliae/alia
EtOH	ethanol
FDR	False discovery rate
FRET	Förster/fluorescence resonance energy transfer
g	gram
GFP	green fluorescent protein
His	histidine
HeLa	Henrietta Lacks human cell line
HRP	horseradish peroxidase
iTRAQ	Isobaric Tags for Relative and Absolute Quantitation
IP	Immunoprecipitation
Kan	kanamycin
L	liter
leu	leucine
LB	lysogeny broth
M	mole
VI	

---

MeOH	methanol
mg	milligram
min	minute
mL	milliliter
mM	millimole
mRNA	messenger RNA
ms	millisecond
MS	mass spectrometry
MS1	Mass spectrometric survey scan
MS2	tandem MS/MS analysis
Myc	myelocytomatosis
Ng	nanogram
nm	nanometer
nM	nanomole
OD <sub>600</sub>	optical density at 600 nm wavelength
PAGE	polyacrylamide gel electrophoresis
<i>Pfu</i>	<i>Pyrococcus furiosus</i>
PIC	protease inhibitor cocktail
pAzF	<i>p</i> -4-azido-L-phenylalanine
pBPA	<i>p</i> -benzoyl-L-phenylalanine
PMSF	phenylmethanesulfonyl fluoride
PVDF	polyvinylidene difluoride
RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SAXS	Small-angle X-ray scattering
SC	Synthetic complete
SDS	sodium dodecyl sulfate
SILAC	Stable Isotope Labeling with Amino Acid in Cell Culture
<i>Taq</i>	<i>Thermus aquaticus</i>
TBS	Tris buffered saline
TCA	trichloroacetic acid
TEMED	N,N,N',N'-tetramethylethylenediamine
tmdF	<i>p</i> -trifluoromethyl-diazirinyL-L-phenylalanine
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
UAA	unnatural amino acid

## Abbreviations

---

ura	uracil
UV	ultraviolet
v	volume
w	weight
WB	western blot
WT	wild-type
YPDA	Yeast peptone dextrose adenine

# 1 Introduction

## 1.1 Chromatin Structure and Function

The cell nucleus is an organelle of immense organization for storing and translating genetic information. Storing vast amounts of genetic information into the restricted dimensions of the nucleus requires that DNA molecules be highly compacted. The cell must possess mechanisms that allow for this packaging, yet still provide access to distinct regions of DNA that are essential for cellular processes. The opposing roles of compact versus accessible DNA impose a great challenge to the cell. Although the nucleus is not separated into substructures by membranes, it is compartmentalized for different functions in nucleic acid metabolism. Compartmentalization is the joined result of highly malleable properties of chromatin structure and chromatin remodeling proteins. The nuclear DNA is organized together with structural proteins into dynamic higher-order chromatin structures which reflect and control gene expression during the cell cycle and cellular differentiation. Chromatin remodeling proteins aid to orchestrate these dynamic changes by modifying chromatin's basic structure or by recruiting other necessary chromatin factors.

Chromatin structure rearrangement is a key regulator of DNA-dependent processes such as transcription, replication and repair. Chromatin assembly exhibits a continuously increasing degree of condensation from its basic building blocks, the nucleosome core particles, lining up like beads on a string, allowing the systematic build-up of the characteristic hierarchical structure of chromatin up to the most condensed form, the mitotic metaphase chromosome (Figure 1.1).



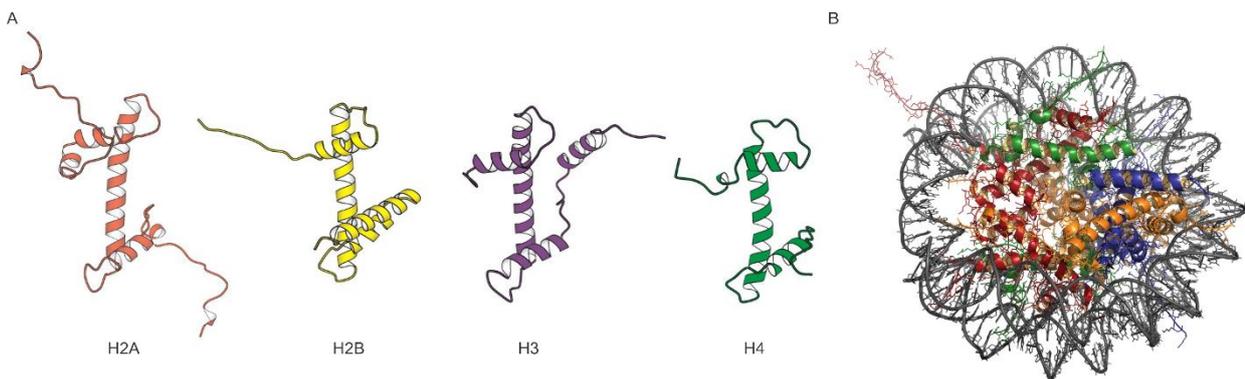
**Figure 1.1:** The various levels of chromatin organization

Artistic illustration of the hierarchical DNA compaction in the process of chromatin organization.

Image: Water colour by Nicolas Bouvier for G. Almouzni

## 1.2 Histones and Nucleosomes

The core histones H2A, H2B, H3 and H4 are abundant, small and highly basic proteins. They are amongst the most conserved proteins in living organisms, also present in archaeobacteria (Starich et al., 1996), indicating that sequence, structure and function are highly interdependent in these proteins (Sullivan and Landsman, 2003). All four core histones exhibit a common structural motif, the histone fold (Arents et al., 1991) (Figure 1.2A). It is characterized by a long central  $\alpha$ -helix, flanked on either side by a loop and a short  $\alpha$ -helix. The helices and loops are denoted, from their N-terminus to their C-terminus, as  $\alpha$ 1, L1,  $\alpha$ 2, L2, and  $\alpha$ 3. Apart from the central histone-fold region, core histones H2A and H2B contain an additional  $\alpha$ -helix in the N-terminal region, termed  $\alpha$ N, whereas core histone H2B contains an additional  $\alpha$ -helix in the C-terminal region, termed  $\alpha$ C (Ramakrishnan, 1997). Besides these structural features, all core histone proteins possess a lysine-rich N-terminal tail. Only histone H2B contains an additional C-terminal tail (Hansen, 2002).



**Figure 1.2:** Structures of the core histones and the nucleosome core particle.

A) Structures of the histone folds of the four core histones H2A, H2B, H3 and H4. Modified from *Biochemistry*, 7<sup>th</sup> Ed., 2012. B) Structure of the nucleosome core particle consisting of histone octamer and wrapped-around, left-handed DNA superhelix. (pdb: 3AV1)

Notably, there are non-allelic variants of histone H2A and H3 which can replace canonical histones at specific time points during the cell cycle and differ in structure (primary structure, N- or C-terminal extensions or truncations). While there are numerous other non-canonical histone variants with important chromatin functions, the work in this thesis focuses on experiments pertaining to the canonical histones. Therefore, an in-depth review of histone variants will not be presented here. For a detailed review, see Weber and Henikoff, 2014.

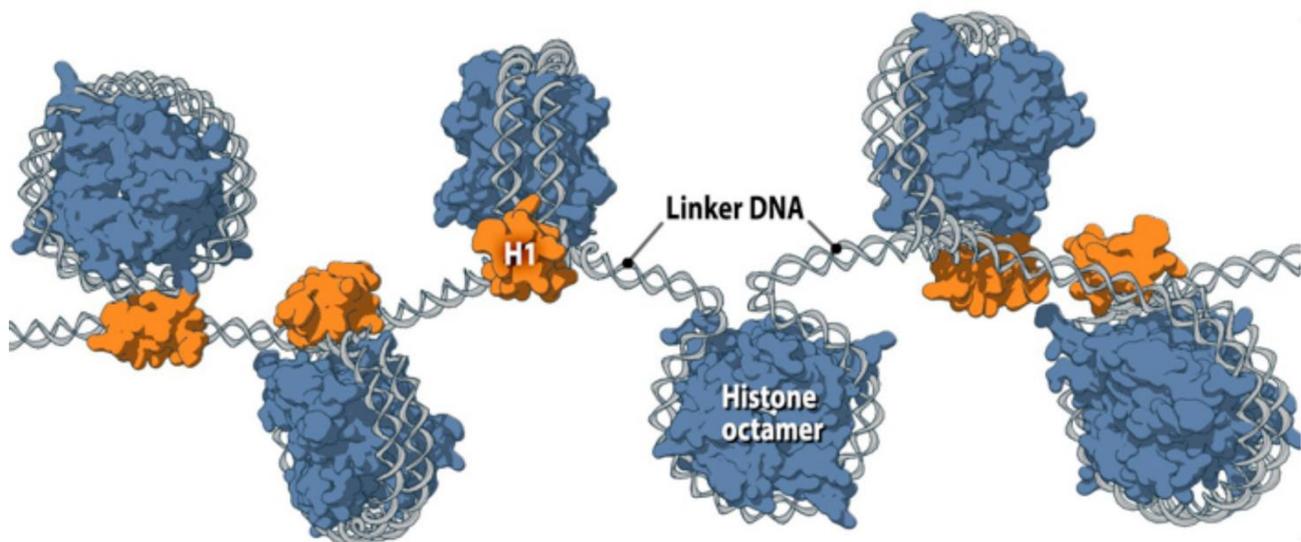
In higher eukaryotes, the core histones are encoded from clustered, multicopy genes whose expression is tightly regulated by the cell cycle and coupled to DNA replication (Hentschel and Birnstiel, 1981). In contrast to higher eukaryotes, the genome of *S. cerevisiae* possesses only four loci encoding the four core histones. Each locus consists of a set of two genes that are divergently transcribed from a central promoter (Osley, 1991).

Two of these loci, HTA1-HTB1 and HTA2-HTB2, carry genes for histones H2A and H2B (Smith and Andresson, 1983). The others, HHT1-HHF1 and HHT2-HHF2, carry genes for histones H3 and H4 (Hereford et al., 1979). The two gene pairs HHF-HHT encode identical proteins for histones H3 and H4, respectively

(Cherry et al., 1997). In contrast, the two HTA-HTB loci encode for slightly different histone H2A and H2B isoforms (Cherry et al., 1997). Although the genes encoding the major histones are essential for growth, deletion strains with either one gene set are still viable (Clark-Adams et al., 1988; Cross and Smith, 1988; Kayne et al., 1988; Norris and Osley, 1987; Smith and Stirling, 1988). Deletion of H2A.Z (HTZ1), the only H2A variant in yeast, allows viability, but produces a slow-growth phenotype (Adam et al., 2001; Jackson et al., 1996; Jackson and Gorovsky, 2000). H2A.Z and major H2A cannot substitute for each other, indicating that these histones have distinct important functions (Jackson and Gorovsky, 2000).

The nucleosome core particle comprises a 147 bp-long DNA strand wrapped 1.7 times around an octameric complex of histone proteins. This octamer consists of two copies of each of the four core histones H2A, H2B, H3 and H4 in basic nucleosomes (Luger et al., 1997) (Figure 1.2B).

Following DNA replication and passing of the replication fork, nucleosome assembly occurs during the S-phase of the cell cycle (Lucchini and Sogo, 1995). Consequently, parental histones are deposited into nucleosomes on both emerging branches of the replication fork (Sogo et al., 1986). *De novo* assembly of nucleosomes occurs in an ordered and step-wise mechanism in which two H3-H4 dimers, either sequentially or as a (H3-H4)<sub>2</sub> tetramer, are deposited on the DNA and subsequently joined by two H2A-H2B dimers. Assembly is completed with the addition of the linker histone H1 which acts as a mediator between the nucleosome core and the linker DNA (Worcel et al., 1978) (Figure 1.3). The length of linker DNA can vary from 8 to 114 bp between different species, cell types or chromosomal regions (van Holde, 1988).



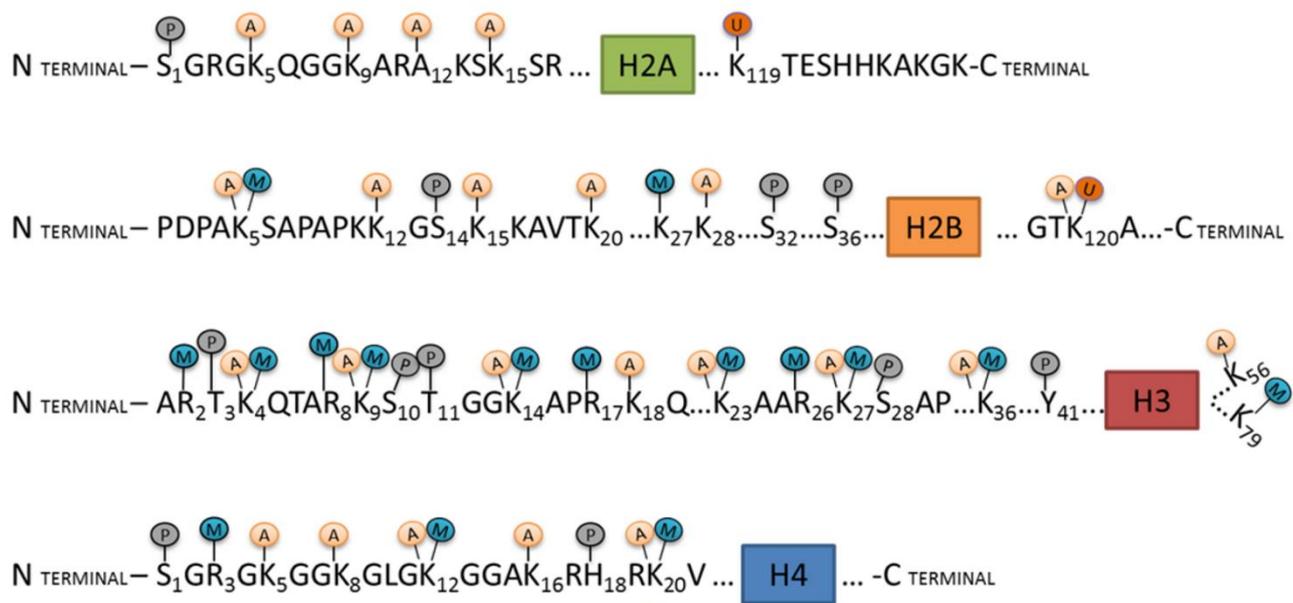
**Figure 1.3:** Schematic illustration of interconnected nucleosome structures.

Each nucleosome is assembled from the four core histones H2A, H2B, H3 and H4 (or other histone variants), forming a histone octamer, and a segment of DNA which is wrapped around the histone octamer. Adjacent nucleosomes are interconnected by the linker histone H1 and linker DNA. Modified from Mechanobiology Institute, Singapore.

*In vivo* nucleosome assembly is aided and regulated by histone chaperones, a class of chromatin-associated proteins. They assist nucleosome assembly by facilitating the interaction between histones while remaining absent from the final product (Loyola and Almouzni, 2004). This aid is believed to be accomplished by charge neutralization mediated by highly acidic residues on the surfaces of histone chaperones. This reduces charge repulsion and prevents formation of non-nucleosomal interactions between positively charged histones and negatively charged DNA (Verreault, 2000). Classification of histone chaperones is based upon their associated histone pair, which can also include histone variants. The binding specificity towards particular histone variants implicates the histone chaperone's preference towards specific cellular processes.

### 1.3 Post-Translational Modifications of Histones

The flexible N-terminal tail domains as well as the structured globular domains of histones are object of a wide range of small chemical changes at the amino acid level, termed post-translational modifications (PTMs). Amino acids are subject to chemical modifications in a specific manner, such as acetylation, ubiquitylation and SUMOylation of lysines, methylation of lysines and arginines, phosphorylation of serines and threonines and the ADP-ribosylation of glutamic acids. Moreover, arginines can either be mono- or dimethylated, and lysine residues can even be mono-, di or trimethylated (Margueron et al., 2005; Peterson and Laniel, 2004) (Figure 1.4).



**Figure 1.4:** Schematic overview of histone post-translational modifications.

The core histones H2A, H2B, H3 and H4 of the nucleosome core particle are subjected to numerous post-translational modifications at their N- and C-terminal tail domains which protrude from the nucleosome. Modified from Xu et al., 2013.

For a long time, the primary method for discovering histone modifications was Edman degradation, which favoured the analysis of the first 20-30 amino acids (Mersfelder and Parthun, 2006). More recently, many histone modifications have been discovered by the advancement of mass spectrometric techniques (Cocklin and Wang, 2003; Freitas et al., 2004; Zhang et al., 2002a; Zhang et al., 2003). These advancements have led to the identification of many new sites and modifications (Arnaudo and Garcia, 2013; Chen et al., 2009), including the identification of serine and threonine as possible acetylation targets (Tweedie-Cullen et al., 2012). They also led to the discovery of a novel class of acyl-based modifications targeting lysine residues (Jiang et al., 2007; Zhang et al., 2009; Chen et al., 2007; Tan et al., 2011; Peng et al., 2011). This class includes formylation, propionylation, butyrylation, crotonylation, succinylation and malonylation.

The long-lasting prevailing view was that modifications of the histone N-terminal tail regulate the interaction between the highly basic tail domains and nucleosomal or linker DNA (Angelov et al., 2001). Although some studies support this view, it appears that the primary mechanism of histone tail modifications is to alter the site-specific binding behaviour of molecular interaction partners according to the different states of modification (Grewal and Moazed, 2003; Iizuka and Smith, 2003; Jenuwein and Allis, 2001; Martin and Zhang, 2005). A steadily growing number of biological processes reported to be dictated by specific and unique patterns of PTMs on histone tails led to the proposal of the “histone code” whose outcome depends on the kind and number of histone modifications (Hansen et al., 1998; Jenuwein and Allis, 2001; Strahl and Allis, 2000). The establishment, removal and recognition of histone modifications is context-dependent and performed by specific proteins (writers, erasers, and readers, respectively), some of which can be grouped into distinct families (Bottomley, 2004; Kouzarides, 2007; Marmorstein, 2001). Some modifying enzymes and proteins recognizing histone modifications occur in multi-subunit protein complexes, linked to chromatin remodeling, transcription, silencing, DNA repair and chromosome segregation (Bottomley, 2004; Khorasanizadeh, 2004; van Attikum and Gasser, 2005). Prominent families of chromatin remodelers in eukaryotes are ISWI, INO80, NuA, RSC, and SWF/SNF (for a review, see Clapier and Cairns, 2009).

Histone modifications can be established in an autonomous as well as an interdependent manner (Fischle et al., 2003). This interrelationship, termed histone modification cross-talk, can occur between modifications on the same or on distinct histone tails. It operates either by producing a complex, distinct modification pattern that modulates the binding of histone-modifying enzymes, or by sequential modification patterns in which defined modifications solely recruit non-histone regulatory proteins (Fischle et al., 2003; Turner, 2002).

A recently discovered cross-talk involves the acetylation of lysine 16 of histone H4 (H4 K16) and the phosphorylation of serine 10 of histone H3 (H3 S10) (Wilkins et al., 2014). During chromatin compaction, the H2A-H2B acidic patch plays a central role in the formation of higher-order chromatin structures (Fan et al., 2004). Original histone octamer crystal structures indicated strong inter- and intranucleosomal interactions of the H4 tail with the acidic patch (Luger et al., 1997; Schalch et al., 2005). These interactions were inhibited by the acetylation of H4 K16, resulting in the reduced formation of higher-order chromatin structures (Shogren-Knaak et al., 2006). Since H4 K16 is the most abundant target for acetylation in yeast and most of the species' genome is laid out in an open conformation state, it comes to mind that H4 K16 acetylation could be a key regulator for chromatin condensation *in vivo* (Lohr et al., 1977). *In vivo* experiments using a genetically

encoded crosslinker in yeast confirmed the role of H4 K16 acetylation as a driving force in chromatin compaction and discovered its cross-talk with H3 S10 phosphorylation, a hallmark of mitosis (Wilkins et al., 2014). They were able to show that H3 S10 phosphorylation draws in a deacetylase to remove the H4 K16 acetylation to allow interaction with the acidic patch.

The importance of PTM on histone proteins cannot be overstated. The precise combination of their chemical signals regulates chromatin dynamics and function. The loss of a single PTM can lead to vast changes across the chromatin's architectural landscape.

### 1.4 Nucleosome Dynamics

Nucleosomes are not static molecules, but rather intrinsically dynamic (Luger, 2006). DNA wrapped around nucleosomes can spontaneously unwrap and rewrap from the histone octamer in solution. This occurrence is termed DNA breathing (Hansen, 2002; Luger, 2003). The histone octamer can translationally reposition along the DNA, which is termed nucleosome sliding (Becker 2002). The H2A-H2B dimers are less stably associated in the nucleosome than the (H3-H4)<sub>2</sub> tetramer, which is reflected by a sequential release of first H2A-H2B dimers and then the (H3-H4)<sub>2</sub> tetramer from the DNA in increasing salt concentration (Hansen, 2002).

Several studies confirmed the rapid unwrapping and rewrapping of nucleosomal DNA from the histone octamer (within 50-250 ms) over a stretch of 10 to 20 bp by using reconstituted mononucleosomes and stopped-flow FRET (Böhm et al., 2011; Li et al., 2005; Li and Widom, 2004; Tomschik et al., 2005). This transient unwrapping creates a window of opportunity for proteins to bind or translocate along the DNA (Li et al., 2005). It can be regulated by histone variants and PTMs such as H3 K56 acetylation. The latter can influence histone-DNA interactions by contacting the phosphate DNA backbone. Through reconstitution of nucleosomal arrays with recombinant histones carrying H3 K56 acetylation, single-molecule FRET experiments have shown that this particular H3 acetylation increases DNA breathing by a factor of 7 (Neumann et al., 2009).

Compared to the rapid process of unwrapping and rewrapping of nucleosomal DNA, spontaneous translational repositioning of nucleosomes is a very slow process under physiological conditions. Chromatin remodelling enzymes can enhance this mobility by lowering the energy barrier of nucleosome repositioning by coupling the disruption of histone-DNA contacts to ATP-hydrolysis (Becker and Horz, 2002; Johnson et al., 2005; Saha et al., 2006; Tsukiyama, 2002). DNA accessibility, in whose regulation ATP-dependent chromatin remodeling complexes are greatly involved, is a key requirement for processes such as transcription, replication and DNA repair. All remodelers are multi-subunit complexes, highly conserved across species, which share a common ATPase domain belonging to the SWI2/SNF2-family ATPase subunit (Becker and Hörz, 2002). Apart from aiding in the mobilization of nucleosomes, they are capable of histone dimer eviction and histone variant exchange. Mechanisms for the catalyzed nucleosome translocation are still under debate with several different models being suggested (Johnson et al., 2005; Langst and Becker, 2004; Saha et al., 2006). All models are based on energy-dependent introduction of disturbances and relaxation of histone-DNA contacts, proposing mechanisms which include looping of DNA (bulge diffusion) (Saha et al., 2005; Zofall et al., 2006), twisting (twist diffusion) (Gavin et al., 2006) or disassembly and reassembly of the histone octamer (histone eviction) (Gutierrez et al., 2007; Lorch et al., 1999; Phelan et al., 2000).

Besides ATP-dependent chromatin remodeling complexes, nucleosome mobility is aided by histone chaperones and assembly factors (Luger and Hansen, 2005). They affect deposition, eviction and nuclear import of histones and maintain a storage pool of histones for nucleosome assembly (Burgess and Zhang, 2013). Thereby, histone chaperones assist DNA-dependent processes such as transcription, replication and DNA repair.

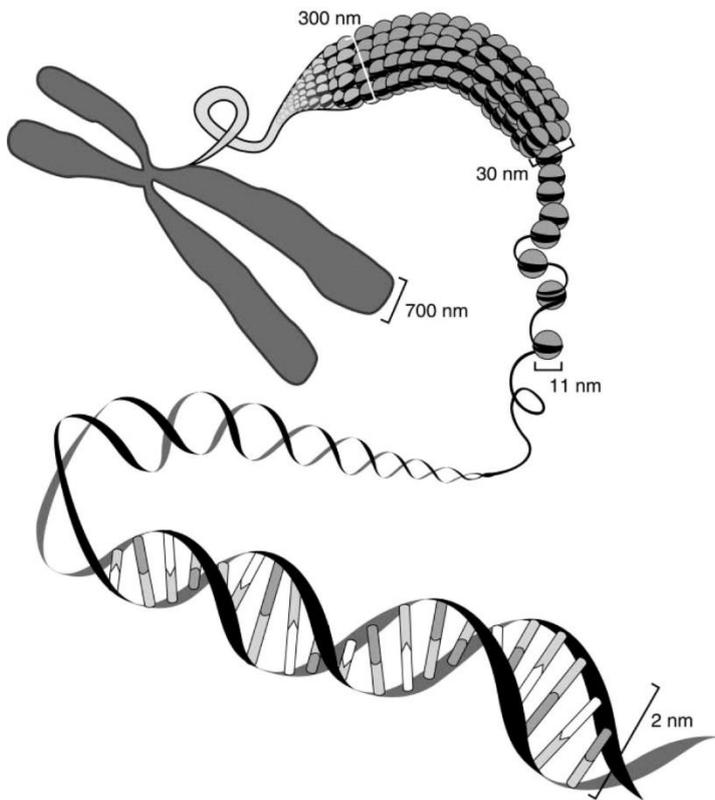
Numerous chaperones capable of interacting with histones and facilitating their deposition onto DNA *in vitro* have been identified (for examples, see Earnshaw et al., 1980; Ishimi et al., 1984; Munakata et al., 2000; Smith and Stillman, 1989). *In vivo*, some of these chaperones may have specialized functions in histone folding, storage, delivery to the nucleus, or modification (Loyola and Almouzni, 2004) without participating in nucleosome deposition per se. The precise *in vivo* function and specificity of most histone chaperones remains to be determined.

### 1.5 Higher-Order Chromatin Structure

Levels of chromatin folding beyond the linear array of nucleosomes are defined as higher-order chromatin structures (Woodcock, 2006). A hierarchical system has been proposed for classifying chromatin structure. The term higher-order chromatin structure encompasses a wide range of hierarchical levels of chromatin folding from positioned nucleosomes up to the large-scale organization of interphase chromosomes (Woodcock and Dimitrov, 2001).

The formation of higher-order chromatin structures is based upon the systematic build-up of the characteristic hierarchical structure of chromatin starting from the primary 10 nm chromatin fibre, a nucleosomal array. The secondary structure, termed 30 nm chromatin fibre, is formed by a folding of the primary structure and involves internucleosomal contacts and interactions with linker histones as well as non-histone chromosomal proteins. The tertiary structure of chromatin that can reach diameters of 300-400 nm and ultimately concludes the chromosome assembly is formed by interactions between secondary structure elements and sustained by intranucleosomal contacts or looped chromatin domains (Woodcock, 2006; Woodcock and Dimitrov, 2001) (Figure 1.5).

Reconstitution of nucleosomal arrays using recombinant histones revealed that chromatin fibres can still fully condense when any of the histone tails is deleted, with the exception of histone H4 (Dorigo et al., 2003). Still, chromatin structure is dynamic and regulated by a plethora of factors, including PTMs, histone variants, nucleosome repeat length, and the presence of linker histones and non-histone chromosomal proteins. Thereby, a high degree of heterogeneity and complexity in structure is being created within all levels of chromatin folding.



**Figure 1.5:** Formation of higher-order chromatin structures.

Schematic illustration of the packaging of DNA into higher-order chromatin structures. Taken from Weier, 2001.

Investigation of dynamic properties of chromatin organization became possible by advancements in imaging technology, fluorescent protein assays and monitoring of the mobility of many chromosomal proteins including histones in living cells of higher eukaryotes (Gasser, 2002; Kimura, 2005; Phair et al., 2004). Fluorescence Recovery after Photobleaching (FRAP) experiments using GFP-tagged histones in HeLa cells allowed monitoring of histone turnover, revealing different turnover rates for H2A-H2B and (H3-H4)<sub>2</sub> tetramers (Kimura and Cook, 2001). Other photobleaching studies reported rapid turnover of proteins from chromatin involved in transcription and DNA repair (Bustin et al., 2005; Hager et al., 2004; Mone et al., 2004; Phair et al., 2004).

Still, all these approaches face limitations in regard to studying chromatin structure *in vivo* (Horowitz-Scherer and Woodcock, 2006). Although they all create a highly detailed description of the local and global primary structure of chromatin, the results remain essentially one-dimensional (Horowitz-Scherer and Woodcock, 2006; Woodcock, 2006).

Therefore, experimental studies of chromatin secondary structure were conducted using structural biological methods. Built upon observations from transmission microscopy, electron cryo-microscopy, atomic force microscopy and X-ray diffraction of chromatin in various ionic strength environments, distinct structural models for the formation of higher-order chromatin structures have been proposed (McBryant et al., 2006; Robinson and Rhodes, 2006; Woodcock, 2006).

Most models have an open zig-zag arrangement of nucleosome arrays in common that results in a helical structure, 30 nm in diameter. These models propose either a one-start or a two-start helix. In a one-start helix, consecutive nucleosomes of an array coil up and follow each other immediately along the same helical path (Kruithof et al., 2009; Robinson et al., 2006; Robinson and Rhodes, 2006). The linker DNA between adjacent nucleosomes continues the curvature established in the nucleosome and thus is bent. In a two-start helix, consecutive nucleosomes of an array arrange in a zig-zag course into a helical structure, which is interconnected by straight linker DNA segments (Dorigo et al., 2004). A pronounced example for a one-start helix is the solenoid model in which nucleosome arrays coil around a central cylinder with linker DNA in the interior of the fibre and with six to eight nucleosomes per turn (Finch and Klug, 1976; McGhee et al., 1983; Thoma et al., 1979).

Until recently, the one-start solenoid model and the two-start zig-zag model were dominating the organization of this secondary chromatin structure *in vitro* (Chen and Li, 2010; Robinson and Rhodes, 2006; Schalch et al., 2005). However, inclusion of the existence of one structure must not exclude the other. It was proposed that each structure can form depending on the linker DNA length (Routh et al., 2008). This would lead to the assumption that both structures can co-exist within a 30 nm chromatin fibre under certain conditions (Grigoryev et al., 2009). Therefore, the controversy of how nucleosomes are organized in condensed 30 nm chromatin fibres continues.

Different models have been described for the formation of large-scaled higher-order chromatin structures. The radial loop model suggests an arrangement of loops made from 30 nm chromatin fibres, forming smaller, thicker loops, which are distributed radially around the axis of the chromatid. Stabilization of the structure is supposed to be maintained by non-histone structural proteins at the base of each loop (Paulson and Laemmli, 1977; Laemmli et al., 1978). The hierarchical helical model proposes a progressive folding of a 30 nm chromatin fibre into larger fibres, including ~100 nm and then ~200 nm fibres, up to a final ~400 nm chromatin fibre (Sedat and Manuelidis, 1978; Belmont et al., 1989; Horn and Peterson, 2002).

The common dogma of all models is the actual existence of the elusive 30 nm fibre *in vivo*. In 1986, pioneering cryo-EM work allowed the first imaging of native mammalian chromosomes (Dubochet et al., 1986). The observed mitotic chromosomes revealed a homogeneous and grainy texture with ~11 nm spacing instead of higher-order chromatin structures including 30 nm chromatin fibres (Dubochet et al., 1986; Dubochet et al., 1988). This observation aligns with a number of publications that have questioned whether mammalian chromosomes contain regular 30 nm chromatin fibres (Dekker, 2008; Fussner et al., 2011a, b; Fussner et al., 2012; Maeshima et al., 2010; Nishino et al., 2012; Woodcock, 1994).

Cryo-EM studies by Woodcock et al. (1994) were able to show that 30 nm chromatin fibres are only present in a minor group of cells defined by highly condensed, transcriptionally inactive chromatin and rather long (~210 bp) nucleosomal repeat lengths. In yeast, investigation of transcriptionally active loci using Chromosome-conformation-capture (3C) exposed a loose arrangement of 10 nm chromatin fibres instead of compact 30 nm fibres (Dekker, 2008). Further visual proof was produced by electron spectroscopic imaging (ESI) that allows assessing the folding of genomic DNA by mapping its phosphorus and nitrogen. It revealed

a highly diffuse mesh of 10 nm fibres in pluripotent mammalian cells (Fussner et al., 2011a, b) and condensed chromatin structures of 10 nm fibres in differentiated mammalian cells (Fussner et al., 2012).

A combinational effort in investigating the structure of mitotic chromosomes in HeLa cells by cryo-EM, SAXS and ultra-SAXS revealed an arrangement of nucleosome fibres in an irregular manner lacking any kind of higher-order chromatin structures above a diameter of 11 nm (Nishino et al., 2012). Maeshima et al. postulate that the folding of nucleosome arrays into irregularly arranged fibres is regulated by nucleosome array concentrations and exhibits a “polymer melt” behavior (2010). Their model states that low nucleosome array concentrations allow the formation of 30 nm chromatin fibres by selective intra-array interactions. An increase in nucleosome array concentrations to a level resembling *in vivo* conditions would lead to a disruption of these interactions and formation of inter-array interactions. Therefore, cellular intermolecular crowding of nucleosome arrays would interfere with the formation of a 30 nm chromatin fibre *in vivo* (Maeshima et al., 2010).

Conclusively, these observations challenge the traditional view of chromatin from possessing static regular to dynamic irregular structural properties.

### 1.6 Chromatin Condensation

Although chromatin has been shown to condense in the presence of divalent or multivalent cations *in vitro*, its condensation is far from being an undirected and random process *in vivo* (De Frutos et al., 2001). Cellular processes depend on the maintenance of histone modification patterns across transcriptionally active and inactive loci. Therefore, chromatin condensation must be highly orchestrated to fulfil this requirement throughout the cell cycle (Valls et al., 2005; Kouskouti and Talianidis, 2005).

Regulating the degree of compaction is a key element in partitioning the eukaryotic genome into functionally distinct chromosomal domains. The underlying mechanisms that direct segments of the genome either into a compacted heterochromatic or an open euchromatic state are manifold and remain to be investigated.

By these means, histones represent integral and dynamic components of the transcription regulating machinery. The densest condensation of chromatin occurs during the transition from interphase to mitosis with the formation of the pronounced x-shaped metaphase chromosome.

Initially, DNA topoisomerase II (Topo II) and condensin were proposed to be the major constituents driving condensation in mitosis (Earnshaw et al., 1985; Gasser et al., 1986; Saitoh et al., 1994; Strunnikov et al., 1995). Condensins represent ring-shaped pentameric complexes composed of two proteins of the structural maintenance of chromosomes (SMC) family and three non-SMC subunits, namely a kleisin and two HEAT-repeat proteins. They are conserved across all eukaryotic species and key factors in mitotic chromosome condensation (Hirano, 2012; Hudson et al., 2009). The association of condensin with chromatin occurs at the N-terminal tail of histone H2A and is regulated by phosphorylation by the chromosomal passenger complex (CBC) kinase Aurora B, which also phosphorylates H3 S10 (Tada et al., 2011).

The exact mechanism by which condensins promote condensation is still under investigation. It has been proposed that positive supercoiling DNA is a net product of reactions involving oligomerized condensin complexes which eventually leads to the folding of a chromatin fibre during chromosome condensation

(Swedlow and Hirano, 2003). Inactivation or depletion of condensin subunits has been shown to abolish the resolution of chromosome bridges in anaphase, resulting in severe defects in chromosome segregation (Hirano, 2012; Hudson et al., 2009). Formation of condensed chromosomes was still observed, however, with strongly reduced structural integrity (Gerlich et al., 2006; Hudson et al., 2003; Vagnarelli et al., 2006). Therefore, there is probably not a single, but rather multiple interlinked mechanisms driving condensation.

With cohesion and KIF4, other factors contributing to chromatin condensation have been identified. Cohesion is mediated by the cohesin complex, which is composed of two proteins from the SMC family, two proteins from the kleisin family and an accessory subunit. It is involved in resolving sister chromatids and alignment of chromosomes on the spindle at metaphase (Gimenez-Abian, et al. 2004; Lavoie et al., 2002; Nakajima et al., 2007; Shintomi and Hirano, 2009).

KIF4 is a DNA-binding kinesin motor protein which interacts with condensin and localizes to the arms of mitotic chromosomes. KIF4 and the condensin subunit SMC2 rely on each other for their localization on chromosome arms. Mitotic chromosomes in KIF4-deficient cells revealed a fat and short structure, a phenotype similarly seen in condensin-deficient cells, although their structural integrity is even more compromised (Samejima et al., 2012). A double mutant exhibited a total loss of structural integrity. This phenotype was partly rescued upon depletion of Topo II which suggests an opposing pathway for the enzyme. Condensin promotes lateral chromosome condensation by forming supercoiled loops of chromatin. Assumably, KIF4 clusters these loops together or, in combination with other proteins, forms supercoiled loops of its own to compact chromosomes further. Topo II could try to oppose this by untangling these loops in order to keep chromosome arms from becoming too long as they compact laterally (Samejima et al., 2012).

Modifying histones also represent a key element in partitioning the chromatin landscape into functionally distinct domains. Histone modifications function either by disrupting intra- and internucleosomal contacts or by affecting the interaction of chromatin with non-histone proteins. By these means, histones serve as integral and dynamic surfaces for a plethora of different PTMs, creating highly detailed, cell cycle-dependent binding interfaces for a multitude of histone modifiers and binders.

Depending on the type of modification, localized changes in charge can occur. Thus, acetylation and phosphorylation of histone tails can potentially alter histone-DNA or internucleosomal contacts, influencing chromatin structure by electrostatic mechanisms (Mersfelder and Parthun, 2006).

Histone acetylation is a crucial regulator of the degree of chromatin folding and condensation by promoting the formation of transcriptionally active euchromatin. Acetylation levels of lysine residues have been shown to increase globally upon entry into interphase and decrease in mitosis (Masumoto et al., 2005; Patzlaff et al., 2010; Vaquero et al., 2006; Wako et al., 2002; Wako et al., 2005). Hyperacetylation of histone tails was reported to disrupt the formation of higher-order chromatin structures, but also diminish histone-DNA interactions (Tse et al., 1998; Annunziato et al., 1988).

The most prominent candidate for a modification directly modulating chromatin compaction is acetylation of lysine 16 of histone H4 (Shia et al., 2006b; Shogren-Knaak et al., 2006). With 80% of all H4 molecules being acetylated at lysine 16, it is the most abundant acetylation site in *S. cerevisiae* (Clarke et al., 1993; Smith et al., 2003), while most of the yeast genome exists in a decondensed state (Lohr et al., 1977; Smith et al., 2003).

Native chemical ligation has been used to produce nucleosome arrays harboring histone H4 acetylated at residue K16. These arrays subsequently exhibited unfolding of chromatin on a global scale, equivalent to a deletion of the entire N-terminal domain of histone H4 (Shogren-Knaak et al., 2006). This behaviour was also observed in nucleosome arrays containing histones enzymatically acetylated at H4 K16 (Robinson et al., 2008). The histone H4 tail forms an  $\alpha$ -helix over a short stretch of residues, centered around lysine 16, while the rest of the tail remains unstructured (Yang and Arya, 2011). In its unmodified state, the basic residues lysine 16, arginine 19, lysine 20 and arginine 23 of the  $\alpha$ -helix face in one common direction. Their side chains are believed to allow a strong binding of the tail to the H2A/H2B acidic patch of an adjacent nucleosome (Luger and Richmond, 1998). Upon acetylation of K16, the orientation of lysine 16 is changed and the  $\alpha$ -helix becomes destabilized. This results in a reduced interaction between the H4 tail and the acidic patch, which eventually fails to stabilize the packaging of chromatin and thus leads to chromatin decondensation (Yang and Arya, 2011).

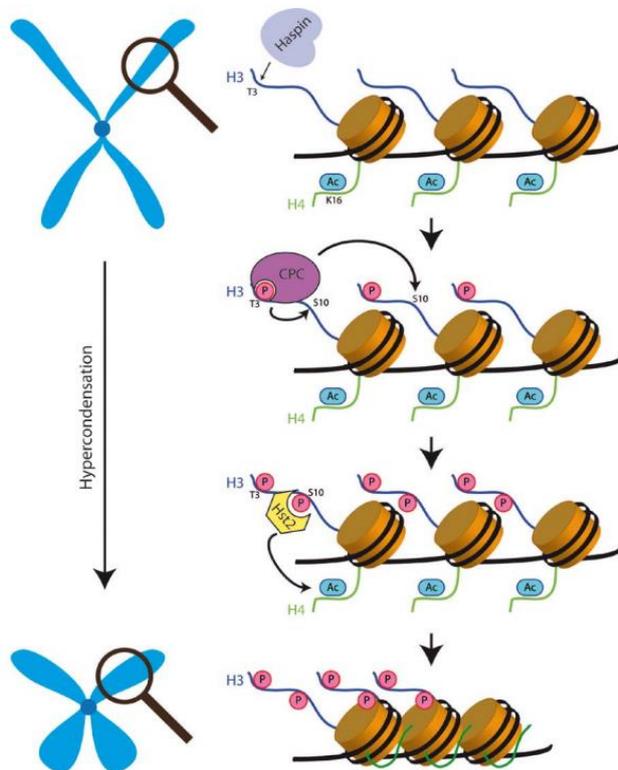
Histone phosphorylation has been associated with a multitude of cellular processes, including transcriptional regulation, apoptosis, cell cycle progression, DNA repair, chromosome condensation, and developmental gene regulation (Cheung et al., 2000; Cruickshank et al., 2010; Houben et al., 2007; Johansen and Johansen, 2006; Kouzarides, 2007; Loomis et al., 2009).

Global phosphorylation levels exhibit an opposite behaviour to acetylation levels by decreasing in interphase and rising upon entry into mitosis (Gurley et al., 1978; Perez-Cadahia et al., 2009; Sawicka and Seiser, 2014). This anticorrelating interrelationship between these two types of modifications throughout the cell cycle aroused suspicion about a cross-talk. Predominantly, phosphorylation of serine and threonine residues on particular histone tails was reported to be involved in chromatin condensation during mitosis and meiosis. Apart from phosphorylation of threonine 119 of histone H2A being linked to regulation of chromatin structure and function in mitosis (Aihara, 2004), the most extensively studied phosphorylation site influencing chromatin compaction during mitosis is H3 S10. It was identified as the major phosphorylation site on histone H3 (Paulson and Taylor, 1982) and as conserved hallmark of mitotic chromosomal condensation across different mammalian cell lines (Hendzel et al., 1997).

Immunofluorescence studies clearly demonstrate the temporal and spatial relationship between chromosome condensation and H3 S10 phosphorylation. Dephosphorylation at this residue sets on during anaphase and is completed within telophase, even before detectable traces of chromosome decondensation can be found (Hendzel et al., 1997; Maile et al., 2004). These observations encourage the suggestion that histone H3 S10 phosphorylation is important for chromosome condensation and segregation. However, it is necessary to point out that there is evidence that H3 S10 phosphorylation is only necessary for the initiation of condensation rather than for maintaining it (van Hooser et al., 1998).

As the very first, an evident cross-talk between H3 S10 phosphorylation and H4 K16 acetylation driving chromatin condensation *in vivo* was observed by deploying a genomically encoded crosslinker into the natural chromatin landscape of *S. cerevisiae* (Wilkins et al., 2014). A model based on site-specific *in vivo* crosslinking data proposes the phosphorylation of H3 T3, a well-established hallmark of mitosis, by Haspin kinase as the initial step in a sequence of PTM events temporally occurring upon entry of mitosis, eventually leading to

mitotic chromatin condensation. This phosphorylation recruits the chromosome passenger complex, whose kinase Aurora B (Ipl1 in yeast) then establishes the phosphorylation of H3 S10. In the following, this modification draws in the deacetylase Hst2 of the sirtuin histone deacetylase (SIR) family, which removes the acetylation of H4 K16 (Wilkins et al., 2014) (Figure 1.6).



**Figure 1.6:** Mitotic chromatin condensation driven by a cascade of histone modifications.

Schematic overview of cross-talk between H3 S10 phosphorylation and H4 K16 acetylation which drives chromatin condensation in *S. cerevisiae*. Taken from Wilkins et al., 2014.

Thereby, the well-documented interaction between the H4 tail and the acidic patch can occur, which drives compaction of mitotic chromosomes (Luger et al., 1977; Schalch et al., 2005; Shia et al., 2006b; Shogren-Knaak et al., 2006; Wilkins et al., 2014).

### 1.7 Expansion of the Genetic Code

Evolution gave rise to an immensely large variety of life-forms, ranging from single- to multi-celled organisms. Despite this diversity, all living organisms are built upon the same range of organic compounds, which include nucleotides and amino acids. They represent the essential building blocks for larger cellular macromolecules such as DNA and RNA as well as proteins, which are indispensable for cellular life.

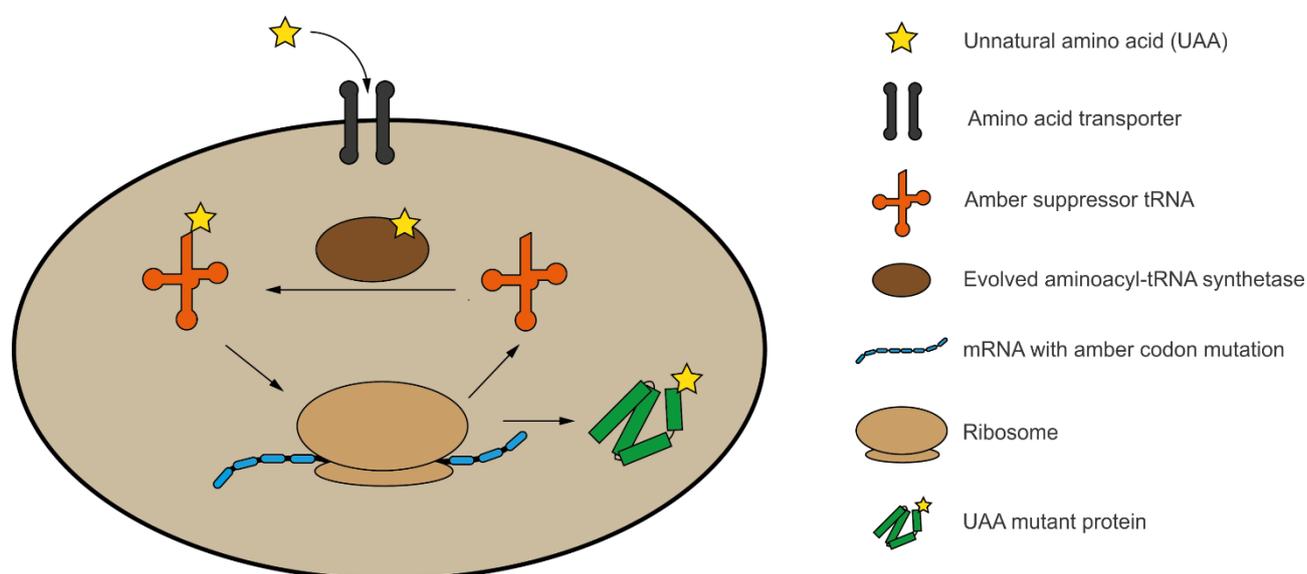
Genetic information is stored as DNA, transcribed into RNA and ultimately translated with the help of amino acids and ribosomes into proteins. This flow of information from genes into proteins is described as the central dogma of molecular biology.

All biological processes are built upon the presence and participation of proteins. Structure and function of these proteins are determined by a common repertoire of 20 canonical amino acids in all organisms. Although this limitation of variety in building blocks perfectly ensures the fidelity of protein synthesis, it greatly reduces diversity of protein structure and function. Naturally, amino acids contain only a limited set of functional groups, such as alcohols, amides, carboxylic acids, nitrogen bases and thiols. To overcome this restriction, some archaea and eubacteria encode non-canonical amino acids such as selenocysteine or pyrrolysine. Selenocysteine, originating from an enzymatically modified serine and then charged to selenocysteinyl-tRNA, was reported to be incorporated into proteins in response to opal (UGA) stop codons, and was regarded as an expansion of the genetic code as the 21<sup>st</sup> amino acid (Böck et al., 1991a; Böck et al., 1991b). Pyrrolysine was termed the 22<sup>nd</sup> amino acid while being directly charged onto pyrrolysyl-tRNA by its cognate aminoacyl-tRNA synthetase PylS (Srinivasan et al., 2002; Atkins and Gesteland, 2002; Krzycki, 2005; Blight et al., 2004; Polycarpo et al., 2004).

Enabling an organism to access an expanded pool of amino acids would allow to overcome the constitutional limitations in protein structure and function. To provide an organism with the ability to genetically incorporate unnatural amino acids with novel functions into proteins would give total control over protein function and allow the introduction of versatile investigation tools for improved analysis of cellular processes. The genetic incorporation of amino acids from an expanded repertoire requires reprogramming of the genetic code in order to accommodate additional amino acids. The unique arrangement of amino acids within a protein is enforced by the nucleotide sequence of the coding gene, encoded on the messenger RNA (mRNA) in the form of base triplets. These base triplets are termed codons and are composed of a combination of the four bases adenine (A), cytosine (C), guanine (G) and thymine (T). The resulting genetic code possesses 64 possible codons of which only 61 encode the 20 canonical amino acids and the remaining three encode termination signals that stop protein biosynthesis.

The process in which genetic information is decoded from mRNA into proteins is called translation and is performed by ribosomes. Protein biosynthesis requires the aminoacylation of a transfer RNA (tRNA) with its correct amino acid by a specific aminoacyl-tRNA-synthetase (aaRS). The aminoacyl-tRNA contains an amino acid-dependent anti-codon which allows complementary binding to the codon on the mRNA and thus ensures site-specific addition of the amino acid to the very C-terminal residue of the growing peptide chain during translation.

By exploiting the degeneracy of the genetic code and mimicking the encoding of the canonical amino acids with an evolved aminoacyl-tRNA-synthetase/tRNA pair, Chin and colleagues successfully developed a method for the genetic incorporation of unnatural amino acids (UAAs) by suppression of an amber (TAG) stop codon (2003) (Figure 1.7). To ensure that the introduced components worked completely orthogonal to endogenous translation components, both tRNA and aaRS were evolved. The tRNA was evolved to ensure that it cannot be charged with canonical amino acids by endogenous aaRSs, and the amino acid specificity of the aaRS was evolved to only charge the evolved tRNA with the desired non-canonical amino acid.



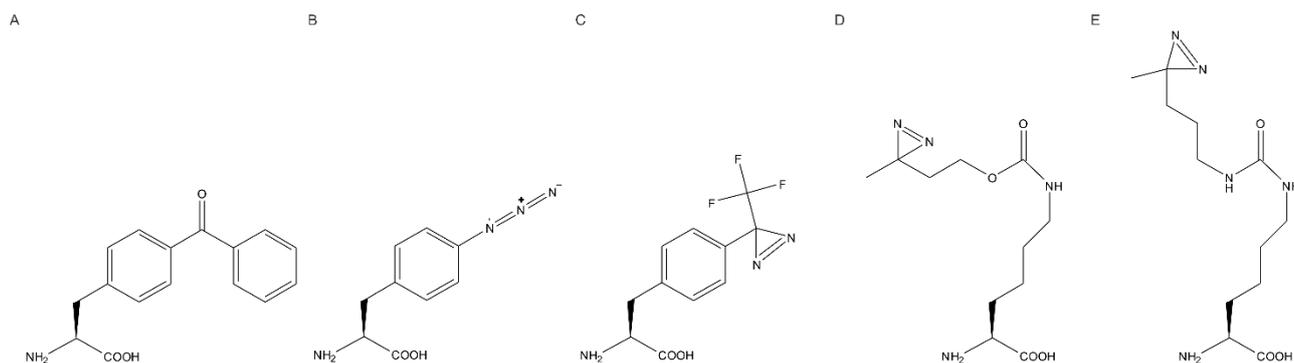
**Figure 1.7:** Expansion of the genetic code

*In vivo* incorporation of an unnatural amino acid (UAA) at a designated position on a protein in response to an amber stop codon by introduction of an evolved aminoacyl-tRNA-synthetase/tRNA<sub>CUA</sub> pair.

The utilization of genetic code expansion in combination with diverse UAAs in a steadily growing number of publications speaks for the versatile applicability of this method. To name a few examples of its application, it was successfully used for the introduction of post-translational modifications into histones and non-histone proteins (Neumann et al., 2008; Neumann et al., 2009; Park et al., 2011). Furthermore, it was employed to introduce chemical handles into proteins which can be modified *in vivo* by exogenous probes for i.e. labeling (Chin et al., 2002; Deiters and Schultz, 2005; Lang et al., 2012a, b; Nguyen et al., 2009; Plass et al., 2012). Other studies used the system to incorporate photolabile UAAs which allow activation of UAA-specific functions, such as crosslinking or uncaging of functional groups, by light (Chin et al., 2002a; Deiters et al., 2006; Gautier et al., 2010; Lemke et al., 2007).

Understanding biological processes requires the identification of participating proteins and their interaction partners. It is often difficult to characterize protein-protein interactions due to the transient or weak nature of some interactions. Site-directed incorporation of genetically encoded photo-crosslinkers into proteins *in vivo* would allow the induction of cross-linking in living cells to trap such interactions for subsequent identification. The common functional groups of photo-crosslinkers are aryl azide, aryl diazine, benzophenone and trifluoromethylphenyl, exhibiting differences in properties such as activation wavelength, crosslinking efficiency, linker lengths and chemical characteristics such as charge and hydrophobicity (Chin et al., 2002; Chin et al., 2003; Ai et al., 2011; Chou et al., 2011).

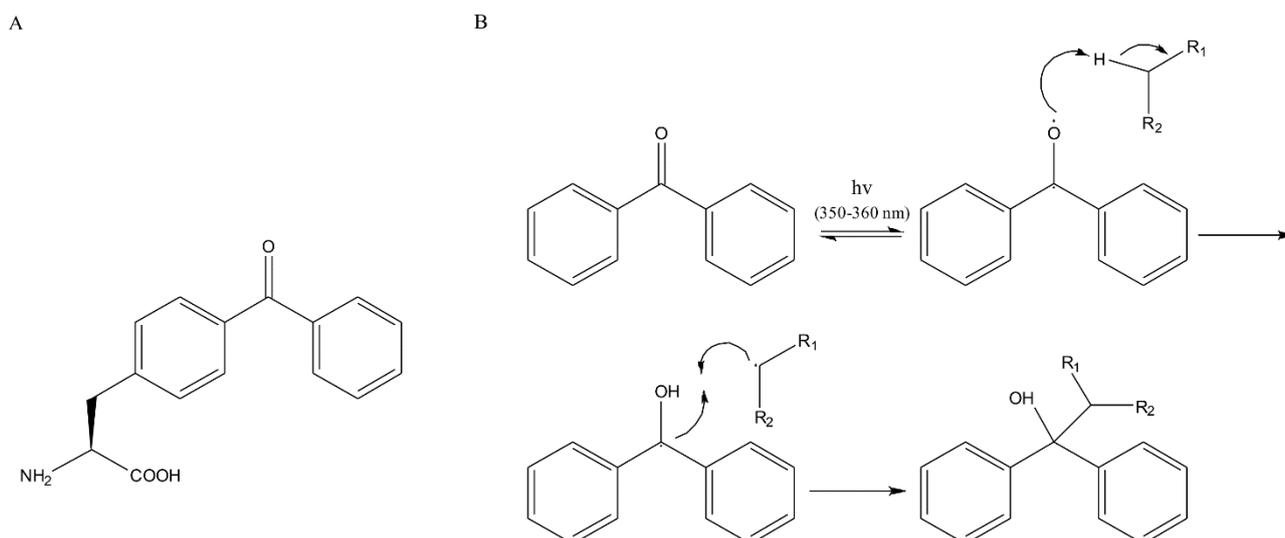
There is a number of available crosslinker amino acids for expression in eukaryotic systems (Figure 1.8).



**Figure 1.8:** Genetically encoded crosslinker amino acids for *in vivo* crosslinking.

Chemical structures of genetically encoded crosslinker amino acids. A) *p*-benzoyl-L-phenylalanine (pBPA) B) *p*-azido-L-phenylalanine (pAzF) C) *p*-trifluoromethyl-diazirinyl-L-phenylalanine (tmdF) D) 3'-azibutyl-N-carbamoyl-lysine (ABK) E) 3-(3-methyl-3*H*-diazirine-3-yl)-propaminocarbonyl-N $\epsilon$ -L-lysine (DiZPK).

Benzophenones generally have a higher crosslinking yield than acryl azide based crosslinkers (Dormán and Prestwich, 1994). They possess distinct advantages as they are activated with low energy ultraviolet light at a wavelength of 360-365 nm that minimizes damage to proteins or DNA, emphasizing its strength for *in vivo* studies. Upon irradiation with the relevant wavelength, benzophenones generate a triplet ketyl biradical that can react with protein functional groups by a sequential abstraction-recombination mechanism. The photochemical basis for the formation of the biradical is absorption of a quantum ( $\lambda \sim 350$  nm) by the benzophenone chromophore, which promotes a non-bonding electron on the carbonyl oxygen into the carbonyl  $\pi^*$  orbital. The resulting triplet excited state can abstract a hydrogen within a distance of 3.1 Å due to the electron deficiency on the ketyl oxygen. Excitation of benzophenones is a reversible process and, in the event that a suitable proton donor is not present during the lifetime of the excited state, benzophenones return to their ground state. The ground state is then available for re-excitation, which is a great advantage compared to other crosslinker chemistries (Preston and Wilson, 2013) (Figure 1.9B). Still, benzophenones have been shown to be not absolutely unreactive in the dark. Their carbonyl moiety can also undergo imine formation with adjacent amines to form specific covalent attachments, which adds to the chance of abnormal crosslink product formation (Weber and Beck-Sickinger, 1997).



**Figure 1.9:** *In vivo* crosslinking by genetically encoded crosslinker amino acid pBPA.

A) Chemical structure of the crosslinker amino acid *p*-benzoyl-L-phenylalanine (pBPA). B) Mechanism of photo-induced crosslinking by benzophenones exemplified by its reaction with an adjacent methylene group.

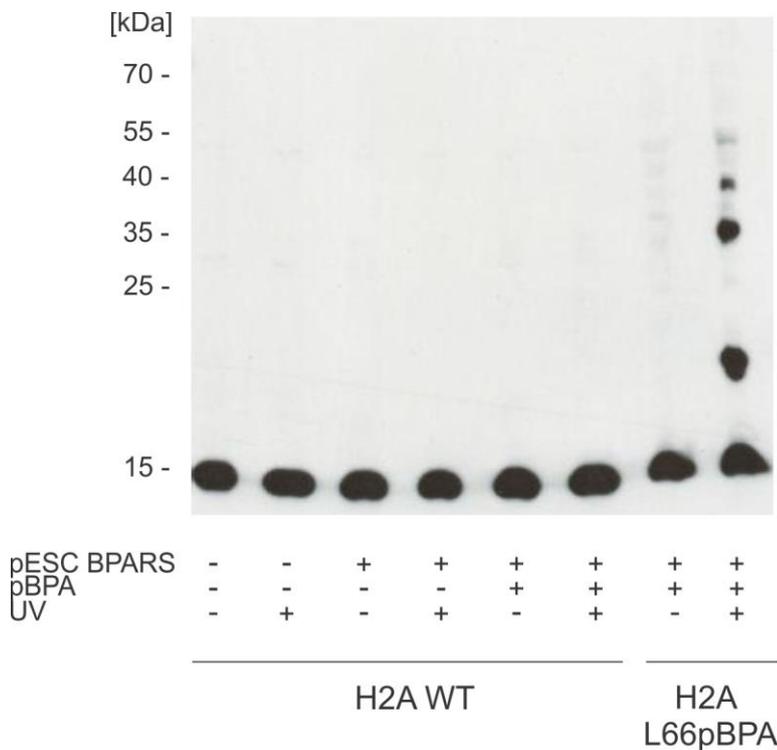
A number of studies helped to establish *p*-benzoyl-L-phenylalanine (pBPA) as a well-suited chemical crosslinker for visualizing interactions in living cells (Chin et al., 2002; Chin et al., 2003; Das and Oliver, 2011; Mori and Ito, 2006; Schlieker et al., 2004; Tagami et al., 2010; Wilkins et al., 2014) (Figure 1.9A).

Combining genetic code expansion and pBPA has been successfully applied by several studies to reveal structural and mechanistic details in *S. cerevisiae* from analyzing single interaction sites to mapping the interactome of whole protein surfaces. Carvalho et al. employed it to gain insight into the mechanistic details of the ERAD pathway (Carvalho et al., 2010; Stanley et al., 2011). Several groups used the system to investigate mitochondrial protein import (Mohibullah and Hahn, 2008; Schulz et al., 2011; Tamura et al., 2009, Wilkins et al., 2014). Moreover, the system was proven capable of mapping whole interaction networks by work of Mohibullah and Hahn (2008) and, more recently, by Hoffmann and Neumann (2015). Mohibullah and Hahn scanned 61 different surface-exposed positions of the TATA-binding protein (TBP) by site-specific incorporation of pBPA (2008). By using a yeast shuffle strain, they completely exchanged the endogenous TBP with pBPA-containing mutants which led to the identification and mapping of several interaction partners, including the Spt-Ada-Gcn5 acetyltransferase (SAGA) complex (Mohibullah and Hahn, 2008). Hoffmann performed crosslinking experiments to map the interactome of the conserved histone chaperone FACT (facilitates chromatin transcription) complex. In a tedious approach, he introduced pBPA into 119 different sites of Spt16 and 67 different sites of Pob3, both major subunits of FACT. Thereby, a distinctive binding interface for histones H2A and H2B was discovered which interfered with binding of Pob3 to Importin- $\alpha$ , suggesting a possible regulatory role in FACT recruitment to chromatin (Hoffmann and Neumann, 2015). Lastly, by substituting single amino acids with pBPA on the N-terminal domain of histone H2A, the system was successfully applied to elucidate the downstream events of H3 S10 phosphorylation driving chromatin condensation in *S. cerevisiae* (Wilkins et al., 2014).

The orthogonal translation components and pBPA have shown to produce absolute amounts of histone pBPA mutants accounting for approximately 10% of the endogenous histone level (Rall, 2012; Wilkins et al., 2014). Also, the system apparently does not alter the viability of the cells, which grew at a comparable rate to WT cells, and with no observable changes in phenotype, in the presence of the BPARS/tRNA<sub>CUA</sub> pair and pBPA (Rall, 2012).

Although amber is the least abundant stop codon in *S. cerevisiae*, its abundance is still high enough to raise concerns about pBPA incorporation into genetic amber codons. Other studies have identified potential open reading frames (ORFs) that are at risk for natural extensions of their native protein products (Lajoie et al., 2013; Mukai et al., 2010; Isaacs et al., 2011).

By comparing plasmid-borne wild-type H2A and a plasmid-borne H2A pBPA mutant, it was shown that crosslink formation depends on the incorporation of pBPA in response to an amber stop codon and UV-irradiation (Figure 1.10) (Rall, 2012; Wilkins et al., 2014).



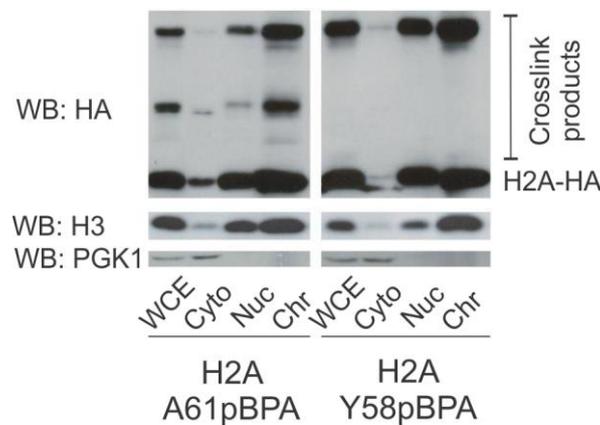
**Figure 1.10:** Dependency of crosslink formation on pBPA incorporation and UV-irradiation.

Western blot of whole-cell lysates from yeast cells transformed with/without BPARS/tRNA<sub>CUA</sub> pair-containing plasmid (pESC BPARS) and either an H2A wild-type- or an H2A L66pBPA mutant-encoding plasmid, and grown in the absence or presence of pBPA. Crosslink product formation could only be observed upon availability of an amber codon in the recombinantly expressed H2A pBPA mutant and UV-irradiation. Modified from Rall, 2012 and Wilkins et al., 2014.

It cannot be ruled out that pBPA is incorporated into amber stop codons within the genomic background of the cells, leading to extended translation of affected genes and production of crosslink products upon UV-irradiation. However, UV-irradiation of cells grown in the presence of pBPA and expressing plasmid-borne wild-type H2A with no amber mutation failed to produce any detectable crosslink products (Figure 1.10). This indicated that although crosslink formation is likely to occur in the genomic background of yeast, no crosslinks are formed with the investigated histones. Therefore, all crosslink products which are formed upon UV-treatment of a histone pBPA mutant are genuine and represent trapped histone-protein interactions.

Subcellular fractionation showed that all histone crosslink products were exclusively formed in the chromatin fraction of a yeast cell (Figure 1.11) (Rall, 2012; Wilkins et al., 2014). This proved that all plasmid-borne

histone mutants are correctly incorporated into the chromatin landscape of *S. cerevisiae*. This is of critical importance for the directed and site-specific investigation of histone interactions, since aggregation of expressed histone pBPA mutants in other cellular compartments would lead to capturing of false interactions. Considering the authenticity and specificity of formed crosslink products (Figure 1.10), the localization of crosslink product formation in the chromatin fraction of yeast suggests the exclusive capturing of chromatin-related interactions.



**Figure 1.11:** Subcellular localization of produced H2A amber mutants and crosslink products.

Western blot of subcellular fractions of yeast cells expressing distinct histone pBPA mutants showed that UV-induced crosslink products were localized in the chromatin fraction. *Controls:* anti-H3 (nuclear); anti-PGK1 (cytoplasmic). *Legend:* WCE: Whole-cell extract; Cyto: Cytoplasmic fraction; Nuc: Nuclear fraction; Chr: Chromatin fraction. Modified from Rall, 2012 and Wilkins et al., 2014.

Consequently, this approach has proven to be a powerful and reliable tool for capturing histone-protein interactions in living yeast cells and elucidating structural and mechanistic details of biological processes.

## 1.8 Mass Spectrometry-Based Proteomics

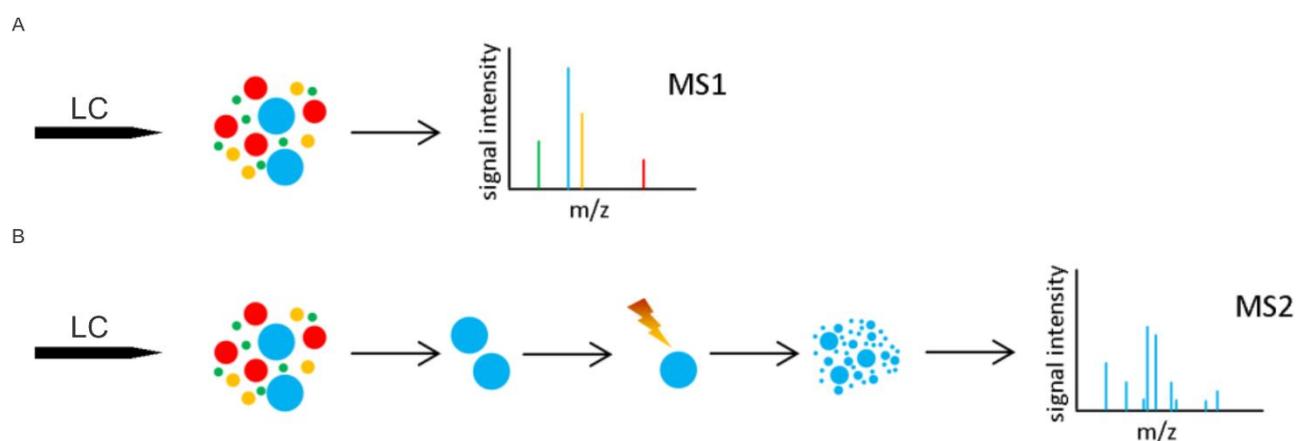
In the past decades, mass spectrometry (MS) has proven to be a robust and sensitive technology for protein analysis, becoming an invaluable tool for proteomics. MS is a fast and sensitive technology that allows for high throughput analyses of a multitude of complex protein or peptide mixtures. Its spectrum of application reaches from sequencing and quantifying proteins or peptides, over detection of posttranslational modifications, to providing structural information about folding and topology of proteins (Aebersold and Mann, 2003). The two main approaches used to study proteins are termed “top-down” and “bottom-up” proteomics. While the first is applied to the analysis of whole proteins by MS, the second aims at analyzing peptides derived from protein digestion and using the peptide measurements to identify the corresponding proteins. Even though the “top-down” approach preserves the protein configuration and allows the precise mass measurement of a protein, the mass limitation of measurable proteins by MS and the ease in handling peptides instead makes the “bottom-up” approach the most widely used to uniquely identify a protein in a complex biological sample.

Proteins extracted from tissues, cell lysates, or subcellular compartments (plasma membrane, cytoplasm, organelles, nucleus, etc.) are first digested into peptides by a protease and then typically separated, based on their hydrophobicity, on a reversed-phase liquid chromatography (LC) column, which is coupled to the mass spectrometer instrument itself. Peptides eluted from the LC column need to be ionized and volatilized to be capable of being analyzed by MS. There are different techniques for this process, the most common are Matrix Assisted Laser Desorption Ionization (MALDI) and Electro Spray Ionization (ESI) (Lottspeicher and Zorbas, 1998). Ionized peptides are sorted in the ion separator, which separates the positively charged ions not directly by their molecular mass, but their mass-to-charge ( $m/z$ ) ratio. Technically, the separation is performed by a combination of magnetic and electric fields or so-called time-of-flight (TOF) components (Domon and Aebersold, 2006). The peptide ions are then recorded by an ion detector which monitors the mass-to-charge ( $m/z$ ) ratio of the analyzed peptides. The signals are typically digitized to produce a mass spectrum of the intensity in function of the  $m/z$  (Figure 1.12A).

There is a wide abundance of mass analyzers available for proteomics measurements, each exploiting different physical principles to perform the separation, resulting in specific strengths and weaknesses. Therefore, the choice of the mass analyzer and of the acquisition method is a decisive factor for the type and quality of the measurements.

Throughout the recent years, the basic principle of how mass spectrometry is set up has been mainly conserved. Still, newer generation mass spectrometers have greatly improved in performance, providing higher mass resolution and allowing very precise mass measurements of analyzed peptides. The higher mass resolution is a key factor in providing a means to better distinguish between acquired signals, especially considering the vast amount of peptides per analyzed biological sample in a single MS scan. More importantly, the improved sensitivity of the new generation of mass spectrometers enables detection of peptides present at very low concentrations, which is vital for the analysis of rare biological elements such as peptides with PTMs or low-abundant crosslink products.

A prominent approach is called shotgun which is generally used for discovery projects where no prior knowledge on the composition of the sample is available, to identify a large number of peptides (McDonald and Yates, 2003). The first stage is the measurement of all  $m/z$  ratios of all peptides at a given time, resulting in the MS1 spectrum. Then, the most intense ions are semi-stochastically selected and transferred into the collision cell where they collide with gas atoms (argon, helium or nitrogen) to break apart at their amino acid bonds into smaller peptide fragments which represent substructures of the original peptide sequence. This fragmentation process is called collision-induced decay (CID). The MS2 spectrum is then generated from the mass and intensity of the generated peptide fragments (Figure 1.12B).



**Figure 1.12:** Main principles of mass spectrometry modes.

A) An MS1 scan of all precursor ions, which were eluted from an LC column and then ionized, is acquired by the mass spectrometer. B) Specific precursor ions are selected (typically those with the highest intensity) in the MS1 scan and fragmented in a collision cell by collision-induced decay (CID) to subsequently generate an MS2 scan by a second mass spectrometer.

These are then searched and matched against theoretical spectra generated from a specific protein database. The most widely used search engines are Sequest, Mascot and Andromeda. Although the number of identified peptides is very high, there is a large number of peptides being measured in an MS1 scan which are not being accounted for in the MS2 spectra by stochastic selection. Therefore, this technique is not capable of analyzing all peptides of a complex peptide sample, limiting the reproducibility of results from replicate analysis.

Multiple techniques address this undersampling problem by reducing the complexity of the peptide elements prior to MS analysis by prefractionation or selective enrichment. Prefractionation is accomplished by separating peptides according to their specific properties (molecular mass, isoelectric point, hydrophobicity, etc.) and thus splitting the original complex sample into multiple fractions of lower complexity. Selective enrichment focuses on the analysis of a particular subset of peptides, e.g. those harboring a certain PTM such as phosphorylation or glycosylation (Witze et al., 2007). However, due to the automated selection process the approach remains biased to high-abundant peptides, rendering the identification of lower-concentrated peptides difficult.

### 1.9 Mass Spectrometry-Based Quantification

Apart from identification of peptides within a sample, mass spectrometry (MS) can be utilized to quantify peptides. MS-based quantification is mainly performed by two main strategies: unlabeled and labeled. Since MS intensities of peptide ions are proportional to their amount in a sample (within the linear dynamic range of the used mass spectrometer), mass spectra from an MS survey (MS1) scan can be used to extract quantitative information about each detected peptide within a sample. Peptides can be distinct from each other in terms of sequence, PTMs or charge state, which results in different ionization efficiencies. Therefore, each peptide is detected by MS with a different efficiency, termed the response factor. MS can accordingly only produce

relative quantities, which only allows the comparison of identical peptides in different samples. Notably, the incorporation of isotope-labeled synthetic peptides of known concentration and with the same sequence and PTMs of the peptides of interest allows measurement of the response factor, and therefore, absolute quantification (Schmidt et al., 2011; Ludwig et al., 2012).

Spectral counting is an unlabeled quantification MS strategy based on the shotgun approach, which allows semi-quantitative analysis (Liu et al., 2004). The quantification strategy is based on the hypothesis that the number of times a peptide precursor ion is selected for fragmentation in a large data set is directly related to the abundance of a peptide represented by its precursor ion in the sample. Due to the low accuracy of this technique, it has been frequently used for the analysis of samples with low to moderate mass resolution.

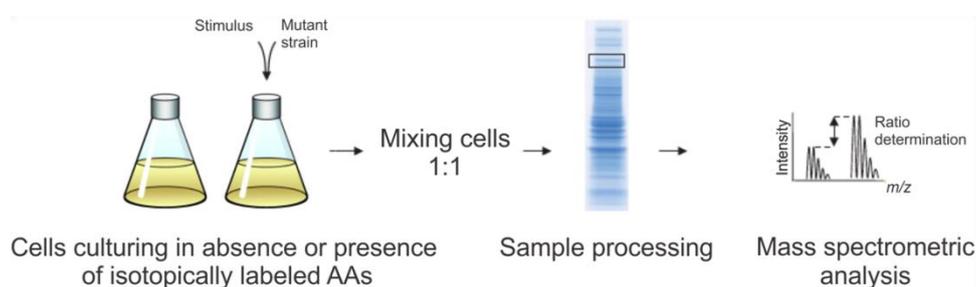
The increased mass resolution of the new generation of mass spectrometers greatly improved the performance of label-free quantification strategies towards quantitative analysis of biological samples. In a typical label-free approach, peptides are detected at MS1 level and then integrated across the retention time dimension ( $m/z$ -by-RT) and used to reconstruct a chromatographic elution profile of the monoisotopic peptide mass. This allows the computational quantification of the number of detections per peptide, which can be used as a quantitative measurement of the original peptide concentration. Thereby, every peptide signal within the sensitivity range of the mass spectrometer can principally be extracted and incorporated into the quantification process independent of MS2 spectra (Listgarten and Emili, 2005). Contrary to the previously described shotgun approach, the common undersampling problem is greatly reduced by this dynamic range of peptide detection. Differential stable isotope labeling has proven to be a very precise and robust method for quantitative proteomics. It allows for absolute and relative quantification of peptide concentrations, either by comparing relevant peptides of a sample against their spiked-in isotopic labeled synthetic analogue, or by comparing peptide intensity values between multiple biological samples within a single MS1 measurement (Aebersold and Mann, 2003; Kusmierz et al., 1990; Lill, 2003; Yan and Chen, 2005). Labeling can be performed by metabolic, chemical or enzymatic approaches (Conrads et al., 2002; Gygi et al., 1999; Zhou et al., 2002; Yao et al., 2001). Labeling strategies can be divided into two main approaches, each exploiting a different principle: The first introduces known mass shifts to the MS1 spectrum of different samples by using stable isotope labels of different mass. Examples for this approach are Isotope Coded Affinity Tag (ICAT) (Gygi et al., 1999), Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC) (Ong et al., 2002), Isotope-Code Protein Label (ICPL) (Schmidt et al., 2005), and  $^{15}\text{N}$  or  $^{18}\text{O}$  labeling. The second approach employs isobaric labels which do not produce mass shifts in MS1 spectra, but generate reporter ions of different masses after fragmentation in the MS2 spectrum. Examples for use of this principle are Isobaric Tags for Relative and Absolute Quantification (iTRAQ) (Ross et al., 2004) and Tandem Mass Tag (TMT) (Thompson et al., 2003).

Label quantification approaches have two main advantages over label-free quantification approaches: Firstly, multiple samples can be prepared and measured simultaneously in one MS run, because samples can be distinguished by different labeling, which reduces measurement variabilities. Secondly, the labels can be directly used for quantification of peptides. However, labeling processes introduce additional steps to the sample preparation, thereby possibly impairing the quantification of low-abundant peptides.

In recent years, SILAC has become increasingly popular as an excellent approach for high-accuracy quantitative proteomics (Ong et al., 2002; Ong and Mann, 2006) (Figure 1.13). It was originally described for mammalian cell culture (Ong et al., 2003) but has been quickly adapted for labeling in *S. cerevisiae* (Jiang and English, 2002; Gruhler et al., 2005), *E. coli* (Kerner et al., 2005), *Plasmodium falciparum* (Nirmalan et al., 2004), and *Arabidopsis thaliana* (Gruhler et al., 2005). Its application spectrum ranges from detection of specific biological changes in functional proteomics assays (Schulze and Mann, 2004; de Hoog et al., 2004; Blagoev et al., 2003) to acquiring temporal profiles of protein abundances (Andersen et al., 2005) and changes in modification states of proteins (Blagoev et al., 2004). SILAC can be adapted to practically every cell culture system that uses defined sources of amino acids. It involves culturing of cells in media containing either light (e.g.  $^1\text{H}$ ,  $^{12}\text{C}$  and  $^{14}\text{N}$ ) or heavy (e.g.  $^2\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$ ) labeled amino acids to incorporate the isotopically labeled amino acids into proteins through the metabolic cycle. Trypsin is the most common protease in proteomic workflows, because it possesses a very efficient, yet specific cleaving activity, targeting the carboxyl-termini of lysine and arginine residues (Olsen et al., 2004). Therefore, heavy labeled L-arginine and L-lysine in combination with trypsin digest is an excellent combination for SILAC experiments, since all peptides except the very C-terminal peptide of a protein are principally quantifiable (Ibarrola et al., 2003).

Most media used for cell culturing provide an overabundance of amino acids and other nutrients. This can lead to metabolic conversion of arginine to proline and the formation of additional heavy labeled proline peptide satellites (Ong et al., 2003; Van Hoof, 2007). This conversion usually occurs when an excess of arginine is provided to cells in growth medium. This can be addressed by reducing the concentrations of stable isotope labeled amino acids or using adapted arginine-to-proline converting cell types.

Although *S. cerevisiae* can normally synthesize all amino acids, SILAC labeling can be performed by using deletion strains in which the biosynthesis pathways of the specific amino acids used for labeling are disrupted (Gruhler et al., 2005).



**Figure 1.13:** Basic principle of Stable Isotope Labeling with Amino Acids in Cell Culture

Cell cultures are grown in media containing either light or heavy labeled arginine and lysine to incorporate the modified amino acids into proteins through the metabolic cycle. Quantification of differences in protein abundance caused by e.g. stimuli is accomplished by mixing equal amounts of cells, isolation of proteins and extraction of peptides after trypsin digestion, MS/MS analysis and determination of intensity ratios between specific light and heavy peptide pairs.

When the mixed light and heavy isotope labeled proteins or peptides are analyzed by MS, they are separated by a residue-specific mass difference corresponding to the number of stable isotope labels as well as the

number of labeled amino acid residues in the analyte. Because quantitative information is encoded by the used isotope labeled amino acid, its choice is important and should be targeted specifically to the goals of the experiment. The relative abundance of the proteins is based on the intensities of the light and heavy peptides. Since differential labeling and combining of the individual samples occurs early on during sample preparation, this approach reduces the introduction of additional error sources through extra experimental sample processing steps (Ong and Mann, 2006). Additionally, modification of proteins or peptides does not require chemical reactions and the level of incorporation is high. New generation mass spectrometers display rapid data acquisition rates which allow the identification and quantification of large amounts of proteins from a single SILAC experiment. Due to the increased mass resolution of these instruments, proteins can be easily identified and quantified by as few as two peptides observed from each protein (Steen and Mann, 2004; Ong et al., 2004).

## 2 Research Objective

Protein-protein interactions are beyond dispute the pacemakers of all processes of life. Little happens in cells without the physical contact of one protein with another. Whether a cell divides, secretes a protein or triggers its own death, protein-protein interactions are the driving force. Understanding these interactions is as important as ever since it helps to gain information about the functions of the involved proteins and their role in biological processes.

The expansion of the genetic code has proven to be a powerful tool for elucidating mechanistic and molecular details of biological processes across the bacterial and eukaryotic kingdom. The genetic encoding and incorporation of the unnatural crosslinker amino acid pBPA into histone proteins allows in a unique and effective manner the site-specific investigation of protein-protein interactions within the chromatin landscape of living yeast.

Although this approach can be applied to visualize protein interactions of spatio-temporal resolution with great success, it fails to provide identification of the trapped interaction partners. A combination of *in vivo* crosslinking, immunoprecipitation and a quantitative mass spectrometry-based approach would have a broad applicability for the objective and quantitative analysis of protein-protein interactions by using stable isotope ratios. Therefore, this PhD project was aimed towards the development of a method which combines pBPA-mediated *in vivo* crosslinking, SILAC immunoprecipitation and mass spectrometry for the quantitative identification of histone-protein interactions in living yeast cells.



### 3 Materials and Methods

The following paragraphs serve to give a detailed account of all materials and methods used during this project for verification of the described findings and to effectively enable replication of the study.

#### 3.1 Materials

All equipment, consumable supplies, chemicals and cell lines used are listed in the following section.

##### 3.1.1 Devices and Instruments

Autoclave HST 4-5-8	Zirbus, Bad Grund
Accumax Pipet Help	Accumax, India
Biophotometer	Eppendorf, Hamburg
Bunsen burner Fuego basis	WLD-Tec, Göttingen
Celvin S Imager	Biostep, Meinsberg
Centrifuge Allegra 2IR Beckman	Coulter, Krefeld
Centrifuge Avanti J-20 XPIJA-20	Beckman Coulter, Krefeld
Centrifuge HERAEUS Pico 17	Thermo Scientific, Langenselbold
Centrifuge 5415R	Eppendorf, Hamburg
Gel Shaker Duomax 1030	Heidolph, Schwabach
Hamilton Syringe 50 µL	Hamilton, USA
Incubator IPP 400	Memmert, Schwabach
Incubator Mytron WB 60 k	Schütt, Göttingen
Incubator Unitron and Multitron	Infors HT, Einsbach
Magnetic stirrer MR Hei-Standard	Heidolph, Kelheim
OptiMax X-ray Processor 1170-1000	Protec, Austria
Orbitrap Q-Fusion	Thermo Scientific, Langenselbold
pH meter	Beckman Coulter, Krefeld
Pipet Research Plus (10, 100, 1000 µL)	Eppendorf, Hamburg
Power Supply 300V	VWR International, Darmstadt
Power Supply EV231	Consort, Belgium
Power Supply MP-250V	Major Science, USA
Rotor JA-20	Beckman Coulter, Krefeld
Rotor JA-30.50	Beckman Coulter, Krefeld
Rotor JA-8.100	Beckman Coulter, Krefeld
Rotator PTR-30	Grant-Bio, UK
Scanner CanoScan 5600F	Canon Deutschland, Krefeld
Tankblotter Criterion 1.3 L	BioRad, Munich
Tankblotter Mini Trans-Blot 0.45 L	BioRad, Munich
Thermoshaker comfort 1.5 mL	Eppendorf, Hamburg

Table Top Shaker 3017	GFL GmbH, Burgswedel
Ultra-centrifugal rotor mill ZM200	Retsch, Haan
Unitron Incubators	Infors HT, Einsbach
UV-lamp 230V, 50/60 Hz	Vilber Luormat, France
Vertical Double Gel Systems	PeqLab, Erlangen
Vortex generator VV3	VWR International, Darmstadt
X-Ray Cassette 18 x 24 cm	Rego X-Ray GmbH, Augsburg

#### 3.1.2 Chemicals

All chemicals and organic substances were obtained from the companies listed below, unless stated otherwise, and fulfilled the purity grade “pro analysis”.

AppliChem, Darmstadt

BioRad, Munich

Fluka, Buchs

Merck, Darmstadt

Roth, Karlsruhe

Sigma-Aldrich, Steinheim

VWR International, Darmstadt

#### 3.1.3 Consumables and Other Materials

Amersham ECL Prime Western blot Detection Reagent	GE Healthcare, Munich
Amersham ECL Select Western blot Detection Reagent	GE Healthcare, Munich
Amersham Hyperfilm ECL	GE Healthcare, Munich
Amicon Ultra-15 Centrifugal Filter Unit (10k MWCO)	Merck, Darmstadt
Eppendorf Tubes (1.5 mL, 2.0 mL)	Eppendorf, Hamburg
Falcon Tubes (15 mL, 50 mL)	Sarstedt, Nümbrecht
Immobilion-P PVDF transfer membrane	Merck, Darmstadt
Instant Blue	Biozol, Eching
NuPAGE Antioxidant Reagent	Thermo Scientific, Langenselbold
PCR Soft Tubes	Biozym, Austria
peqGOLD Universal Agarose	Peqlab, Erlangen
Petri Dishes 92 x 16 mm	Sarstedt, Nümbrecht
Pipet Tips	Sarstedt, Nümbrecht
Spectra/Por Dialysis Membrane, 50 mm	Spectrum Laboratories, Netherlands
Syringe Filters (0.2 and 0.45 µm)	Sartorius, Göttingen
UV Cuvettes	Eppendorf, Hamburg
Whatman Filter Paper	Whatman, Dassel

### 3.1.4 Protein Size Standards

GeneRuler 1 kb DNA ladder

Thermo Scientific, Langenselbold

PageRuler Prestained Protein Ladder

Thermo Scientific, Langenselbold

Prestained Protein Molecular Weight Marker

Fermentas, St. Leon-Rot

### 3.1.5 Culture Media and Agar Plates

**Table 3.1:** Amino acid mixes for cell culturing

3-letter-code	Name	Regular Amino Acid Dropout	SILAC Amino Acid Dropout
Ade	adenine	2 g	2 g
Ala	alanine	2 g	2 g
Arg	arginine	2 g	-
Asn	asparagine	2 g	2 g
Asp	aspartic acid	2 g	2 g
Cys	cysteine	2 g	2 g
Gln	glutamine	2 g	2 g
Glu	glutamic acid	2 g	2 g
Gly	glycine	2 g	2 g
His	histidine	-	-
Ile	isoleucine	2 g	2 g
Leu	leucine	-	-
Lys	lysine	2 g	-
Met	methionine	2 g	2 g
Paba	4-aminobenzoic-acid	0.2 g	0.2 g
Phe	phenylalanine	2 g	2 g
Pro	proline	2 g	2 g
Ser	serine	2 g	2 g
Thr	threonine	2 g	2 g
Trp	tryptophan	-	-
Tyr	tyrosine	2 g	2 g
Ura	uracil	-	-
Val	valine	2 g	2 g

**Table 3.2:** Ingredients for LB, YPDA and SC media

LB medium	YPDA medium	SC medium (-ura/-leu)	SC medium (-ura/-leu/-arg/-lys)
10 g tryptone/peptone	20 g tryptone/peptone	1.7 g yeast nitrogen base	1.7 g yeast nitrogen base
5 g yeast extract	10 g yeast extract	5 g ammonium sulfate	5 g ammonium sulfate
5 g NaCl	40 mg adenine	2 g regular amino acid dropout	2 g SILAC amino acid dropout
Add H <sub>2</sub> O to 1 L	Add H <sub>2</sub> O to 900 mL	3.5 mM NaOH	3.5 mM NaOH
		80 mg histidine	80 mg histidine
		80 mg tryptophan	80 mg tryptophan
		Add H <sub>2</sub> O to 900 mL	Add H <sub>2</sub> O to 900 mL

For agar plates, the corresponding medium was supplemented with 1.5% (w/v) agar-agar. All media were sterilized by autoclaving at 130 °C for 20 min. Before use, each yeast medium was supplemented with 2% glucose to a final volume of 1 L.

### 3.1.6 Antibiotics

**Table 3.3:** Overview of antibiotics used in growth media and agar plates

Antibiotic	Work concentration [ $\mu\text{g/mL}$ ]	Company
Ampicillin (Amp)	100	AppliChem, Darmstadt
Chloramphenicol (Cm)	50	AppliChem, Darmstadt
Kanamycin (Kan)	50	AppliChem, Darmstadt

### 3.1.7 Unnatural Amino Acids

**Table 3.4:** Overview of unnatural amino acids used for genetic code expansion

Amino acid	Stock solution [M]	Working concentration [mM]	Company
<i>p</i> -benzoyl-L-phenylalanine (pBPA)	0.1 (in 0.1 M NaOH)	1	Chem-Impex international, USA

### 3.1.8 Isotopic Labeled Amino Acids

**Table 3.5:** Overview of isotopic labeled amino acids used for SILAC

Name	Stock solution [mg/mL]	Working concentration [mg/L]	Company
$^3\text{C}_6$ -L-arginine:HCl	125 (in dH <sub>2</sub> O)	30	Cambridge Isotope Laboratories, USA
$^3\text{C}_6$ - $^{15}\text{N}_4$ -L-arginine:HCl	125 (in dH <sub>2</sub> O)	30	Cambridge Isotope Laboratories, USA
$^{13}\text{C}_6$ -L-lysine:2HCl	250 (in dH <sub>2</sub> O)	60	Cambridge Isotope Laboratories, USA

### 3.1.9 Enzymes

*Taq* DNA Polymerase

Thermo Scientific, Langenselbold

Trypsin

Serva, Heidelberg

*Pfu* Turbo DNA Polymerase

Agilent, Böblingen

Phusion DNA Polymerase

Thermo Scientific, Langenselbold

Proteinase K

Thermo Scientific, Langenselbold

All enzymes were used according to the manufacturers' recommendations.

## 3.1.10 Cell Lines

Table 3.6: Overview of cell lines used

Cell line	Genotype	Source
<i>E. coli</i> DH10B	F-mcrA $\Delta$ (mrr-hsdRMS-mcrBC) $\Phi$ 80lacZ $\Delta$ M15 $\Delta$ lacX74 recA1 endA1 araD139 $\Delta$ (ara, leu)7697 galU galK $\lambda$ -rpsL nupG	Invitrogen, Darmstadt
<i>S. cerevisiae</i> BY4741 WT	MATa his3 $\Delta$ 1 leu2 $\Delta$ 0 met15 $\Delta$ 0 ura3 $\Delta$ 0	AG Neumann
<i>S. cerevisiae</i> BY4741 $\Delta$ Arg4 $\Delta$ Lys2	MATa his3 $\Delta$ 1 leu2 $\Delta$ 0 met15 $\Delta$ 0 ura3 $\Delta$ 0 arg4::KanMx lys2::his3	AG Neumann
<i>S. cerevisiae</i> BY4741 HTA1-FLAG	MATa his3 $\Delta$ 1 leu2 $\Delta$ 0 met15 $\Delta$ 0 ura3 $\Delta$ 0 HTA1-1Flag::His3MX6	AG Neumann
<i>S. cerevisiae</i> BY4741 STH1-3Myc	MATa his3 $\Delta$ 1 leu2 $\Delta$ 0 met15 $\Delta$ 0 ura3 $\Delta$ 0 arg4::KanMx lys2::his3 STH1-3Myc::His3MX6	AG Neumann
<i>S. cerevisiae</i> BY4741 STH1-GFP	MATa his3 $\Delta$ 1 leu2 $\Delta$ 0 met15 $\Delta$ 0 ura3 $\Delta$ 0 arg4::KanMx lys2::his3 STH1-GFP::His3MX6	AG Neumann
<i>S. cerevisiae</i> ATCC 201388 BMS1-GFP	MATa leu2D0 met15D0 ura3D0 BMS1-GFP::His3MX6	Thermo Scientific, Langenselbold
<i>S. cerevisiae</i> ATCC 201388 CHD1-GFP	MATa leu2D0 met15D0 ura3D0 CHD1-GFP::His3MX6	Thermo Scientific, Langenselbold
<i>S. cerevisiae</i> ATCC 201388 NOB1-GFP	MATa leu2D0 met15D0 ura3D0 NOB1-GFP::His3MX6	Thermo Scientific, Langenselbold
<i>S. cerevisiae</i> ATCC 201388 NOP12-GFP	MATa leu2D0 met15D0 ura3D0 NOP12-GFP::His3MX6	Thermo Scientific, Langenselbold
<i>S. cerevisiae</i> ATCC 201388 REG2-GFP	MATa leu2D0 met15D0 ura3D0 REG2-GFP::His3MX6	Thermo Scientific, Langenselbold
<i>S. cerevisiae</i> YPH499 WT	MATa ura3 $\Delta$ 52 lys2 $\Delta$ 801-amber ade2 $\Delta$ 101-ochre trp1- $\Delta$ 63 his3- $\Delta$ 200 leu2 $\Delta$ 1	AG Kehlenbach
<i>S. cerevisiae</i> YPH499 $\Delta$ Arg4	MATa ura3 $\Delta$ 52 lys2 $\Delta$ 801-amber ade2 $\Delta$ 101-ochre trp1- $\Delta$ 63 his3- $\Delta$ 200 leu2 $\Delta$ 1 arg4 $\Delta$ 0	AG Kehlenbach

## 3.1.11 Plasmids

Table 3.7: Overview of plasmids used for genetic code expansion

Vektor	Insert	Auxotrophic marker/Antibiotic resistance	Source
pESC	BPARS-tRNA <sub>CUA</sub>	Leu/Amp	Christian Hoffmann, Göttingen
pESC	BPARS-1xSUP4 tRNA <sub>CUA</sub>	Leu/Amp	Bryan Wilkins, Göttingen
pESC	BPARS-2xSUP4 tRNA <sub>CUA</sub>	Leu/Amp	Bryan Wilkins, Göttingen
pESC	BPARS-3xSUP4 tRNA <sub>CUA</sub>	Leu/Amp	Bryan Wilkins, Göttingen
pRS	H2A-FLAG-L66TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-WT	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-E57TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-Y58TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-L59TAG	Ura/Amp	Marco Winkler, Göttingen

### 3 Materials and Methods

pRS	H2A-HA-A60TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-A61TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-E62TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-I63TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-L64TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-E65TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-L66TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-A67TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-D90TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H2A-HA-D91TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H3-HA-T6 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H3-HA-A15 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H3-HA-A21 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H3-HA-A29 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H3-HA-P38 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H3-HA-R52 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H3-HA-K56 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H3-HA-R69 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H3-HA-E73 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H3-HA-Q76 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H3-HA-L82 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-G9 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-R17 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-R19 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-K20 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-R23 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-Q27 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-R45 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-A56 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-S60 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-R67 TAG	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-E74	Ura/Amp	Marco Winkler, Göttingen
pRS	H4-HA-K77	Ura/Amp	Marco Winkler, Göttingen

### 3.1.12 Buffers and Solutions

All buffers were prepared with dH<sub>2</sub>O.

**Table 3.8:** Overview of commonly used buffers and solutions

Buffer/Solution	Ingredients
PBS (1x)	137 mM NaCl 2.7 mM KCl 10 mM Na <sub>2</sub> HPO <sub>4</sub> 1.8 mM KH <sub>2</sub> PO <sub>4</sub> pH 7.4
TBS (1x)	50 mM Tris-HCl, pH 7.5 150 mM NaCl
SDS Sample Buffer (1x)	2.5% glycerol 12.5 mM Tris-HCl, pH 6.8 25 mM DTT 0.5% SDS (w/v) 0.025% bromophenol blue (w/v)
SDS Running buffer (1x)	25 mM Tris base 192 mM glycine 0.1% SDS (w/v)
MOPS Running buffer (1x)	50 mM MOPS 50 mM Tris base 1 mM EDTA 0.1% SDS (w/v) pH 7.7
WB Transfer Buffer (1x)	25 mM Tris base 192 mM glycine 0.1% SDS (w/v) 10-20% MeOH (v/v)
PonceauS Solution (0.5%)	0.5% PonceauS (w/v) 5% TCA (w/v)
TE Buffer (10x)	100 mM Tris-HCl, pH 8.0 10 mM EDTA, pH 8.0
PEG Solution (50%)	50% PEG 4000 (w/v)
Single-stranded DNA	10 mg/mL salmon testes DNA (w/v)
Resuspension Buffer	1x TE Buffer 0.1 M LiOAc
PEG Buffer	1x TE Buffer 0.1 M LiOAc 50% PEG Solution (w/v)
PIC (1x)	75 μM pefabloc SC 150 nM leupeptin 37.5 μM O-phenanthroline 500 nM pepstatin A Sterilized using 0.2 μm syringe filter

### 3 Materials and Methods

Lysis Buffer	1x PBS 2x PIC 2 mM PMSF
Guanidinium Chloride Extraction buffer	7 M Guanidinium-HCl 20 mM Tris-HCl, pH 7.5 10 mM DTT
Urea Dialysis Buffer	7 M Urea 10 mM Tris-HCl, pH 8.0 1 mM EDTA, pH 8.0 100 mM NaCl 5 mM $\beta$ -ME 1 mM PMSF
Urea Loading Buffer	7 M Urea 10 mM Tris-HCl, pH 8.0 1mM EDTA, pH 8.0 1x PIC 1 mM PMSF 1 mM DTT
Urea Elution buffer	7 M Urea 10 mM Tris-HCl, pH 8.0 1mM EDTA, pH 8.0 1x PIC 1 mM PMSF 1 mM DTT 1 M NaCl
IP buffer (1.2x)	168 mM NaCl 12 mM Tris-HCl, pH 8.0 1.2 mM EDTA, pH 8.0 0.12% SDS (w/v) 0.12% Na-deoxycholate (w/v) 1.2% Triton X-100 (w/v)
DTT Buffer	10 mM DTT 100 mM $\text{NH}_4\text{HCO}_3$
IAA Buffer	55 mM iodoacetamide 100 mM $\text{NH}_4\text{HCO}_3$
Trypsin Buffer	1.5 $\mu\text{g}$ Trypsin 42 mM $\text{NH}_4\text{HCO}_3$ 5 mM $\text{CaCl}_2$
Digestion Buffer	42 mM $\text{NH}_4\text{HCO}_3$ 5 mM $\text{CaCl}_2$
MS Buffer A	0.1% FA
MS Buffer B	80% ACN, 0.08% FA

### 3.1.13 DNA Kit Systems

All plasmid purification kits were used according to the manufacturer's protocol.

peqGOLD Plasmid Miniprep Kit

Peqlab, Erlangen

QIAGEN Plasmid Mini Kit

QIAGEN, Hilden

### 3.1.14 Antibodies

**Table 3.9:** Overview of antibodies used for Western blotting

Antigen	Host	Conjugate	Diluent (w/v)	Product code	Company
HA	rabbit	-	3% BSA 1x TBS	ab9110	Abcam,UK
Myc	mouse	-	3% Milk 1x TBS	sc-40	Santa Cruz, Heidelberg
FLAG	rabbit	-	3% Milk 1x TBS	F7425	Sigma-Aldrich, Steinheim
PGK	mouse	-	3% Milk 1x TBS	22C5D8	Thermo Scientific, Langenselbold
H2A	rabbit	-	3% Milk 1x TBS	ab13923	Abcam,UK
H3	rabbit	-	3% Milk 1x TBS	ab1791	Abcam,UK
H4	rabbit	-	3% Milk 1x TBS	ab7311	Abcam,UK
rabbit	goat	HRP	5% Milk 1x TBS	A6154	Sigma-Aldrich, Steinheim
mouse	goat	HRP	5% Milk 1x TBS	A4416	Sigma-Aldrich, Steinheim

### 3.1.15 Conjugated Beads

**Table 3.10:** Overview of antibody conjugated beads used for immunoprecipitation

Conjugate	Host	Bead material	Concentration	Diluent	Product code	Company
$\alpha$ -HA IgG <sub>1</sub>	mouse	Agarose	3.5 mg/mL	1x IP Buffer	26182	Thermo Scientific, Langenselbold
$\alpha$ -HA IgG <sub>1</sub>	mouse	1 $\mu$ m magnetic particles	10 mg/mL	1x IP Buffer	88836	Thermo Scientific, Langenselbold
Hydrazide group	-	1 $\mu$ m magnetic particles	30 mg/mL	0.5 M EDTA, pH 8.0	FD-101	Bioclone Inc., USA

### 3.2 Methods

#### 3.2.1 Microbiological Methods

##### 3.2.1.1 Transformation of *S. cerevisiae*

The transformation of *S. cerevisiae* was performed by a standard lithium acetate/heat shock approach. Prior to the transformation, a 5 mL overnight culture of the desired cell type (Table 3.6) was prepared from a single colony. The next day, the overnight culture was used to inoculate a 50 mL main culture to a final OD<sub>600</sub> of 0.5. The cells were allowed to grow for one cell cycle and then collected by centrifugation at 4800 rpm for 1 min at RT. Subsequently, the cells were washed once with dH<sub>2</sub>O. The resulting pellet was resuspended in 2 mL of Resuspension Buffer (Table 3.8). From the cell resuspension, 100 µL was mixed with 1 µg plasmid DNA, 10 µg of single-stranded DNA and 700 µL PEG Solution (Table 3.8) were incubated for 30 min at 30 °C. The heat shock was then performed at 42 °C for 7 min. Following heat shock, the cells were spun down at 13.000 rpm for 1 min and the supernatant was discarded. The cell pellet was washed with 1 mL dH<sub>2</sub>O and then resuspended in 100 µL dH<sub>2</sub>O. Finally, the resuspended cells were plated on selective synthetic complete (SC) agar plates (Table 3.2) depending on the plasmid's auxotrophic marker and then incubated for 2 days at 30 °C.

##### 3.2.1.2 Transformation of *E. coli*

Chemically competent *E. coli* DH10B cells (Table 3.6) were transformed with plasmids using the heat shock method. A mixture of 50 µL chemical competent cells and 150-500 ng plasmid DNA were incubated on ice for 10 min. After heat shocking for 90 sec at 42 °C, the cells were incubated for further 3 min on ice. For recovery, the cells were resuspended in 1 mL LB medium (Table 3.2), followed by incubation for 45 min at 37 °C and 750 rpm in a Thermoshaker. Transformants were plated on agar media containing the appropriate antibiotics (Table 3.3) or used for the inoculation of an overnight culture.

#### 3.2.2 Protein Biochemical Methods

##### 3.2.2.1 Recombinant Protein Expression

All plasmids used in this study were kindly provided by Dr. Bryan J. Wilkins and Marco Winkler from the Research group of Applied Synthetic Biology at the Institute of Microbiology and Genetics of the University of Göttingen.

The plasmid pRS426 was used as the backbone to carry the genomic coding region of a single *S. cerevisiae* histone variant, namely H2A, H3 or H4, its native promoter region and a coding sequence for a C-terminal HA-epitope. To study potential interaction sites by genomic code expansion, site-specific amber codon mutations were introduced using standard QuikChange PCR protocol.

Specific incorporation of pBPA at amber codon mutated sites required the co-transformation of a plasmid which contains the DNA gene sequence for an evolved *E. coli* amber suppressor tyrosyl-tRNA-synthetase/tRNA<sub>CUA</sub> pair under control of constitutive promoters. Initial experiments were performed with the original plasmid developed by Chin et al., 2003. During the process of systemic optimization, an enhanced version of the original plasmid was assembled. It encodes the evolved *E. coli* tRNA<sub>CUA</sub> gene surrounded by

the ~55 bp upstream (5') and ~30 bp downstream (3') flanking regions of the yeast suppressor tRNA gene *SUP4* as well as the evolved *E. coli* Tyrosyl-aaRS (BPARS) under control of a constitutive promoter.

### 3.2.2.2 Crosslinking in *S. cerevisiae* cells

Transformed yeast cells (Table 3.6) carrying plasmids for the expression of the evolved *E. coli* amber suppressor tyrosyl-tRNA-synthetase/tRNA<sub>CUA</sub> pair and an amber codon mutated histone variant (Table 3.7) were cultured in the appropriate standard SC dropout medium (Table 3.2) supplemented with 2% glucose. Recombinant protein expression of full-length mutant protein was enabled by supplementation of 1 mM pBPA final concentration. Cells were grown at 30 °C with shaking at 210 rpm.

Prior to crosslinking, a 5 mL overnight culture was prepared from a single colony of transformed cells. The next day, the overnight culture was used to inoculate a 20 mL main culture to a final OD<sub>600</sub> of 0.5. Cells were harvested in late exponential phase.

Samples were normalized to 12 ODs and resuspended in 100 µL SC medium at 4°C for UV treatment. Live cells were irradiated with 365 nm UV-light from a distance of ~5 cm for 7 min at 4 °C (Vilber Lourmat lamp, 2x 8 W, 365 nm tubes, 32 W, 230 V #VL-208.BL).

Whole cell lysate samples were prepared as described in 3.2.2.4. As preparation for Western blot analysis (3.2.2.7), samples were centrifuged and stored at -20 °C.

### 3.2.2.3 Cell Cycle Synchronization

Cell cycle synchronization in *S. cerevisiae* was performed by treating transformed BY4741 cells (Table 3.1) with the antimetabolic agent nocodazole (Sigma-Aldrich, Steinheim).

Cell culturing was identical to the crosslink approach (3.2.2.2). Transformed cells were cultured in the appropriate standard SC dropout medium (Table 3.2) supplemented with 2% glucose. Recombinant protein expression of full-length pBPA mutants was enabled by supplementation of 1 mM pBPA final concentration. Cells were grown at 30 °C with shaking at 210 rpm.

A single colony of transformed BY4741 yeast cells carrying plasmids for the expression of the evolved *E. coli* amber suppressor tyrosyl-tRNA-synthetase/tRNA<sub>CUA</sub> pair and an amber codon mutated histone variant (Table 3.7) was used to inoculate a 5 mL overnight culture. The following day, the overnight culture was used to inoculate a 50 mL main culture to a final OD<sub>600</sub> of 0.2, followed by incubation for one cell cycle at 30 °C. For cell arrest, nocodazole was added to 15 µg/ml final concentration from 100X stock solution (1.5 mg/mL in DMSO) and the culture was incubated for 1.5 h at 30°C before harvest. The synchronization process was monitored by microscopy until >95% of the cells were arrested in the separation of daughter cells. Samples were taken before and after nocodazole treatment, normalized to 12 ODs and resuspended in 100 µL SC medium at 4°C for UV treatment. Live cells were irradiated with 365 nm UV-light from a distance of ~5 cm for 10 min at 4 °C (Vilber Lourmat lamp, 2x 8 W, 365 nm tubes, 32 W, 230 V #VL-208.BL). Whole cell lysate samples were prepared as described in 3.2.2.4. As preparation for SDS-PAGE (3.2.2.5) and further analysis, samples were centrifuged and stored at -20 °C.

### 3.2.2.4 Cell Lysis

Several methods for lysis of yeast cells were performed. For whole cell lysate, cell pellets of 12 ODs were resuspended in 200  $\mu$ L 100 mM NaOH and incubated for 10 min at room temperature. Subsequently, cells were collected by centrifugation and then resuspended in 100  $\mu$ L preheated 1x SDS sample buffer at 95  $^{\circ}$ C, and boiled for 10 min. In preparation for immunoprecipitation, washed cell pellets from 1 L cell cultures were resuspended in 5 mL lysis buffer (v/v, Table 3.8) and lysed by flash freezing in liquid nitrogen and disruption by ultra-centrifugal rotor mill (Retsch ZM200). The adherent cell lysate was scraped off the rotor mill and gently transferred into a centrifugation tube. Triton-X 100 was added to the lysate to a final concentration of 1% and then clarified by centrifugation with rotor JA-20 (Beckman Coulter, Krefeld) at 20.000 rpm for 30 min at 4 $^{\circ}$ C. Afterwards, the pellet could be used for protein extraction.

### 3.2.2.5 SDS-PAGE for Crosslinking Samples

**Table 3.11:** Composition of polyacrylamide gels for SDS-PAGE

8% Resolving gel	15% Resolving gel	4% Stacking gel
375 mM Tris-HCl, pH 8.8	375 mM Tris-HCl, pH 8.8	125 mM Tris-HCl, pH 6.8
0.1% SDS (w/v)	0.1% SDS (w/v)	0.1% SDS (w/v)
8% acrylamide	15% acrylamide	4% acrylamide
0.22% bisacrylamide	0.4% bisacrylamide	0.11% bisacrylamide
0.1% APS (w/v)	0.1% APS (w/v)	0.05% APS (w/v)
0.04% TEMED (v/v)	0.04% TEMED (v/v)	0.1% TEMED (v/v)

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed to separate proteins according to their electrophoretic mobility. SDS is a strong anionic detergent which denatures secondary and non-disulfide-linked tertiary structures and, when bound to a protein, confers a negative charge which is proportional to the molecular weight. SDS-PAGE was performed in a discontinuous buffer system. The stacking gel possesses a lower pH than the SDS-running buffer and the resolving gel has larger pores due to a lower amount of acrylamide/bisacrylamide. The glycine of the SDS-running buffer has a neutral charge and therefore a slower mobility than the negatively charged chloride anions. The proteins are stacked between these two fronts. At transition into the resolving gel, the pH increases. The charge of the glycine becomes negative, which leads to a higher mobility. The protein band is sharpened at the border to the resolving gel where pore size decreases. Due to the narrow sieve-like properties of polyacrylamide resolving gel the proteins are separated by size. To estimate the molecular weight of protein species, a molecular weight standard is co-electrophoresed (3.1.4).

The polymerization of the polyacrylamide is initiated by the free radical of ammonium persulfate (APS) and the cross-linking agent TEMED (N,N,N',N'-tetramethylethylenediamine).

During this study, 12 and 15% gels were used. Samples of *S. cerevisiae* expression cultures were prepared as described in chapters 3.2.2.2 and 3.2.2.4.

Electrophoresis was performed in 1x SDS running buffer (Table 3.8) at 130 V for three hours. After electrophoresis, the proteins were either detected by Instant Blue staining or transferred onto a polyvinylidene difluoride (PVDF) membrane by Western blot (3.2.2.7).

### 3.2.2.6 Bis-Tris SDS-PAGE

The traditional SDS-PAGE system relies on denaturation and surrounding of proteins by an SDS negatively charged detergent micelle (Laemmli, 1970). This eliminates most of the charge and idiosyncratic solubility differences from one protein to another and gives a reasonable separation based on the negative charge which is proportional to the molecular weight. The highly alkaline operating pH of this system may cause chemical modifications of the sample proteins, e.g. deamination and alkylation, reoxidation of reduced disulfides from cysteine residues, as well as protein instability, which results in poor band resolution.

Bis-Tris systems have quite different charge-shielding characteristics. Bis has a tertiary amine with a pKa of 6.46 and a pKb of 7.54, rendering the operating pH more neutral. This allows the maintenance of reduced state of proteins and improved protein stability, resulting in higher band resolution.

Bis-Tris SDS-PAGE was performed by using premade NuPAGE Novex Bis-Tris gels with a gradient of 4-12%. The gels were placed in an Xcell SureLock Mini-Cell Electrophoresis System (Thermo Scientific, Langenselbold), filled with 1x MOPS-running buffer (Table 3.8) and supplemented with 500  $\mu$ L NuPAGE Antioxidant reagent (Thermo Scientific, Langenselbold). IP samples were prepared as described in chapters 3.2.2.10 and 3.2.2.11. Electrophoresis was performed at 130 V for 45 min. After electrophoresis, the proteins were detected by Instant Blue staining and the gel stored for further analysis at 4 °C.

### 3.2.2.7 Western Blot

To verify the expression of a protein, immuno blotting was performed to detect those proteins by specific antibodies against the protein-tag. In general, a protein that has been immobilized on a polyvinylidene difluoride (PVDF) membrane is specifically recognized and bound by a primary antibody. This primary antibody is then in turn bound by a secondary antibody that is conjugated to a horseradish peroxidase (HRP). Prepared whole cell lysates of crosslinked cells (3.2.2.2) were subjected to SDS-PAGE and then transferred to a PVDF membrane using the wet blot method. The SDS-gel was placed onto the PVDF membrane and sandwiched by 2 layers of 3 mm Whatman paper on each side, then enclosed in a blotting cassette. The cassette was clamped into the blotting tank filled with 1x transfer buffer (Table 3.8) and placed at 4 °C. Transfer of proteins from 15% SDS-gels was performed for 30 min at 100 V and 500 mA, and from 8% SDS-gels for 4 h at 50 V and 250 mA. The transfer efficiency was determined by Instant Blue staining of the gel and 0.5% PonceauS staining (Table 3.8) of the membrane. Non-specific binding of antibodies was blocked by incubation of the membrane with 3% BSA or non-fat dried milk in PBS or TBS (pH 7.5), respectively, depending on antibodies used. All blocking procedures were performed for 1 h. Primary and secondary antibodies were chosen and prepared according to Table 3.8.

All blocked membranes were incubated in primary antibody overnight at 4 °C. Before application of secondary antibody, all membranes were washed three times with 1x TBS for 20 min.

Immunocomplexes of primary antibody-bound proteins were detected using HRP-conjugated secondary antibodies (Table 3.9). Membranes were incubated for 2 h at room temperature and washed with 1x TBS and 0.5% Tween 20 for 10 min, followed by thorough washing with 1x TBS.

For chemiluminescent detection, all blots were incubated with substrate, provided by Pierce ECL Select (Thermo Scientific, Langensfeld), for 5 minutes. Excess substrate solution was removed and the membranes were placed either in an x-ray cassette or on the Celvin S Imager (Biostep, Meinsberg). The film development was performed by using an x-ray film processor of the AG Brauns. Chemiluminescent detection was performed over a range of exposure times. These times were dependent upon the intensity of the signal.

#### **3.2.2.8 Bradford Protein Assay**

Protein concentrations can be determined by performing the Bradford assay (Bradford, 1976). The Bradford reagent consists of the dye Coomassie Brilliant Blue G250 in a phosphoric acid solution. The assay is based on an absorbance shift of Coomassie Brilliant Blue G250 from 465 nm in the cationic, unbound red form to 595 nm in the anionic, blue form when it is bound to protein. This absorbance at 595 is proportional to the protein concentration in a range of 0.1-0.9. The spectrophotometer was calibrated to such a degree that an absorbance of 0.1 equates to a protein concentration of 0.1 mg/mL. For spectrophotometrical determination of the protein concentration, 10  $\mu$ L protein solution was added to 490  $\mu$ L Bradford reagent in a 1 mL cuvette (1 cm optical path length) and the absorbance was measured against a blank after an incubation time of 5 min.

#### **3.2.2.9 Concentration of Protein Solutions**

For rapid concentration of protein solutions by centrifugation, the protein solution was transferred into an Amicon centrifugal concentrator (Merck, Darmstadt). These concentrators consist of two chambers which are separated by a membrane with a specific molecular weight cutoff. Proteins significantly larger than this molecular weight cutoff are retained in the top chamber during centrifugation, whereas the buffer and smaller proteins are collected into the bottom chamber. Thus, the volume of the protein solution decreases and the protein concentration increases. Samples were centrifuged at 4000 rpm at 4 °C until the required protein concentration was achieved. The protein concentration was determined using a standard Bradford protein assay (3.2.2.8).

#### **3.2.2.10 Immunoprecipitation under Normal Growth Conditions**

Antibody-driven immunoprecipitation allows the targeted isolation and purification of a protein of interest from whole cell lysate. A specific, bead-coupled antibody (Table 3.10) is incubated with cell extract, leading to the formation of an antibody/antigen complex with the target protein. Subsequent removal of the antibody-coupled beads from the cell extract physically isolates the protein of interest from the rest of the sample. The sample can then be separated by SDS-PAGE for Western blot analysis or mass spectrometric analysis.

Transformed yeast cells (Table 3.6) carrying plasmids for the expression of the evolved *E. coli* amber suppressor BPA-tRNA-synthetase/tRNA<sub>CUA</sub> pair and an amber codon-mutated histone variant (Table 3.7) were cultured in the appropriate standard SC dropout medium (Table 3.2) supplemented with 2% glucose.

Recombinant protein expression of full-length pBPA mutants was enabled by supplementation of 1 mM pBPA final concentration. Cells were grown at 30 °C with shaking at 210 rpm.

Two days before harvest, a single colony of transformed yeast cells was used to inoculate a 5 mL overnight culture. The next morning, the entire overnight culture was used to inoculate a 100 mL culture (supplemented with 1mM pBPA) and cells were grown over four cell cycles at 30 °C. Then, the whole 100 mL cell culture was used to inoculate a 900 mL main cell culture (+1 mM pBPA), which was grown overnight at 30 °C. The next morning, the main culture was harvested by centrifugation with rotor JA-8.1000 (Beckman Coulter, Krefeld) at 4.800 rpm and 4 °C for 15 min, washed with dH<sub>2</sub>O and resuspended in 20 mL SC media. Live cells were irradiated with 365 nm UV-light from a distance of ~5 cm at 4 °C in 3 intervals of 5 min irradiation and short mixing by slow agitation (Vilber Lourmat lamp, 2x 8 W, 365 nm tubes, 32 W, 230 V #VL-208.BL).

Preparation of cell lysis was performed by resuspending the cells in 1:1 v/v lysis buffer (Table 3.8). The cells were lysed by flash freezing in liquid nitrogen and disruption by ultra-centrifugal rotor mill (Retsch ZM200). The adherent cell lysate was scraped off the rotor mill and gently transferred into a centrifugation tube. Triton-X 100 was added to the lysate to a final concentration of 1% and then clarified by centrifugation with rotor JA-20 (Beckman Coulter, Krefeld) at 20.000 rpm for 30 min at 4 °C. The pellet was washed twice in 1x PBS and collected by centrifugation. The insoluble nuclear pellet fraction was resuspended in 1 mL DMSO and incubated at 37 °C for 1 h.

To extract the target proteins from the insoluble fraction, 25 mL extraction buffer (Table 3.8) was added to the sample, followed by vortexing and incubation for 1 h at room temperature on a tumbler. The protein extract was clarified by centrifugation with rotor JA-20 (Beckman Coulter, Krefeld) at 20.000 rpm for 30 min at 4 °C, and the resulting supernatant was collected. The remaining pellet was subjected again to 25 mL extraction buffer for 1 h at room temperature on a tumbler and centrifugation. The resulting supernatant was pooled with the previously collected supernatant and dialyzed in dialysis buffer (Table 3.8) overnight at 4 °C.

Isolation of histone proteins was performed by affinity chromatography with a 5 mL ion exchange HiTrap SP FF column (GE Healthcare), equilibrated in urea loading buffer (Table 3.8). The sodium chloride concentration for the equilibration and washing of the column depends on the histone variant: histones H2A and H2B require 100 mM NaCl, respectively 10% urea elution buffer (Table 3.8), and histones H3 and H4 require 200 mM NaCl, respectively 20% urea elution buffer (Luger et al., 1999). Elution of histone proteins was performed by washing the column with 100% urea elution buffer, containing 1 M NaCl.

The resulting peak fractions were pooled and concentrated with a 10k MWCO Amicon ultra-centrifugal concentrator (Merck, Darmstadt) (3.2.2.9) to a final concentration of 0.7-1.0 mg/mL. The concentration was determined by standard Bradford test (3.2.2.8).

The concentrated sample was subsequently mixed with a 5-fold excess of 1.2x IP-buffer (Table 3.8), resulting in a final concentration of 1x IP-buffer in the sample, and incubated overnight at 4 °C on a tumbler. The next day, the sample was centrifuged with rotor JA-30.50 at 30.000 rpm for 1 h at 4 °C. In preparation for the IP, either anti-HA agarose or magnetic beads (Table 3.10) were washed once in 1x TBS, once in 1x IP-buffer, followed by centrifugation at 2.000 rpm for 1 min, or by magnetic Eppendorf tube rack. IP reaction was performed for 3 h at 4 °C on a tumbler. Beads were washed once with 1x IP-buffer, 1x TBS (150 mM NaCl +

0.05% Tween 20), 1x TBS 300 (300 mM NaCl, 0.05% Tween 20), 1x TBS (150 mM NaCl), and 1x TBS 300 (300 mM NaCl). Proteins were eluted by resuspending the beads in 50  $\mu$ L preheated 2x SDS loading buffer at 95 °C, and boiling for 20 min. As preparation for Bis-Tris SDS-PAGE (3.2.2.6) and further analysis, samples were centrifuged and stored at -20 °C.

#### 3.2.2.11 Immunoprecipitation under SILAC Growth Conditions

The immunoprecipitation protocol in preparation for a SILAC experiment differs from the regular immunoprecipitation protocol (3.2.2.10). SILAC takes advantage of metabolic incorporation of non-radioactive isotopic labeled amino acids substituted with stable isotopic nuclei (e.g.  $^2\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$ ) into proteins for mass spectrometry-based quantitative proteome analysis.

Two cell cultures of transformed yeast cells lacking the genes *LYS2* and *ARG4* (Table 3.6) and carrying plasmids for the expression of the evolved *E. coli* amber suppressor tyrosyl-tRNA-synthetase/tRNA<sub>CUA</sub> pair and an amber codon-mutated histone variant (Table 3.7) were cultured in the appropriate SILAC SC dropout medium (Table 3.2) supplemented with 2 mM pBPA and either unlabeled (“light”) or isotopically labeled (“heavy”) amino acids (e.g.  $^{13}\text{C}_6$ - $^{15}\text{N}_4$ -L-arginine and  $^{13}\text{C}_6$ -L-lysine) (Table 3.5) with final concentrations of 30 mg/L L-arginine and 60 mg/L L-lysine, respectively. Cells were grown at 30 °C with shaking at 210 rpm. For a forward sample, the heavy isotope labeled cell culture was then treated with UV-light for crosslinking, mixed with the non-crosslinked, light isotope labeled cell culture. Opposite treatment resulted in the assembly of a reverse sample. After mixing, the samples were processed (3.2.2.12.1, 3.2.2.12.2) and eventually analyzed by mass spectrometry (3.2.2.12.3). The incorporation of heavy isotopic labeled amino acids results in a mass difference of labeled peptides, which renders them easily distinguishable from unlabeled peptides and allows very accurate quantification. Therefore, this approach allows the quantitative identification of crosslink products.

Two days before harvest, two single colonies of transformed yeast cells were used to inoculate a 5 mL overnight culture each, either supplemented with light or heavy isotope labeled amino acids. This specific individual supplementation was kept throughout the following culturing steps until cell harvest. The next morning, both overnight cultures were used to inoculate each a 100 mL cell culture, which were grown for four cell cycles at 30 °C. Then, both 100 mL cell cultures were used to inoculate each a 900 mL main cell culture, which were grown at 30 °C overnight. The next morning, both main cultures were harvested by centrifugation with rotor JA-8.1000 (Beckman Coulter, Krefeld) at 4.800 rpm and 4 °C for 15 min, washed with dH<sub>2</sub>O and resuspended in 20 mL SC media. Heavy isotope labeled cells were irradiated with 365 nm UV-light from a distance of ~5 cm at 4 °C in 3 intervals of 5 min irradiation and short mixing by slow agitation (Vilber Lourmat lamp, 2x 8 W, 365 nm tubes, 32 W, 230 V #VL-208.BL).

Preparation of cell lysis was performed by resuspending the cells in 1:1 v/v lysis buffer (Table 3.8) and mixing non-crosslinked cells with crosslinked cells in a 1:1 ratio based on OD. Cells were lysed and subjected to immunoprecipitation as described in 3.2.2.10. Proteins were eluted by resuspending the beads in 50  $\mu$ L preheated 2x SDS loading buffer at 95 °C, and boiling for 20 min. As preparation for Bis-Tris SDS-PAGE (3.2.2.6) and further analysis, samples were centrifuged and stored at -20 °C.

### 3.2.2.12 Mass Spectrometric Analysis of Crosslink Products by SILAC

In SILAC experiments, proteins are metabolically labeled during protein turnover by culturing cells in media containing light and heavy isotopic labeled amino acids. Different treatment of cell cultures, e.g. formation of crosslink products by UV-radiation, results in distinct peptide peaks in mass spectra which can be accurately quantified relative to each other to identify proteins.

#### 3.2.2.12.1 In-Gel Trypsin Digest of SILAC samples

Prior to in-gel trypsin digest, the samples acquired by immunoprecipitation (3.2.2.11) were subjected to Bis-Tris SDS-PAGE (3.2.2.6).

The most commonly used protease in proteomics is trypsin (3.1.9), which specifically cleaves at L-arginine or L-lysine. Taking advantage of this specificity, both amino acids are the most commonly used heavy isotopic labeled amino acids (e.g.  $^{13}\text{C}_6$ - $^{15}\text{N}_4$ -L-arginine and  $^{13}\text{C}_6$ -L-lysine, Table 3.5) in SILAC. A trypsin digest results in all peptides (except the very C-terminal of a protein) containing single heavy isotope labeling amino acids (more if trypsin misses one or more cleavages) which renders them quantifiable by mass spectrometry.

Before digestion, each gel lane was horizontally cut into 23 gel bands with a metal cutter made in-house. Each gel band was cut into small pieces (~1 mm<sup>3</sup> cubes) with a surgeon knife and transferred with a spatula into mini Eppendorf tubes containing 150  $\mu\text{L}$  dH<sub>2</sub>O. The tubes were incubated for 5 min at 26 °C and 1050 rpm in a Thermoshaker, followed by centrifugation at 13.300 rpm for 10 sec. The supernatant was removed, the gel pieces were resuspended in 150  $\mu\text{L}$  acetonitrile (ACN) and incubated for 15 min at 26 °C and 1050 rpm. The tubes were centrifuged at 13.300 rpm for 10 sec and the supernatant removed. The gel pieces were dried in a SpeedVac (Eppendorf) for 5 min and resuspended in 150  $\mu\text{L}$  DTT buffer (Table 3.8) to reduce disulfide bridges, followed by incubation for 50 min at 56 °C and 1050 rpm and centrifugation at 13.300 rpm for 10 sec. The supernatant was removed and the gel pieces resuspended in 150  $\mu\text{L}$  ACN. The tubes were incubated for 15 min at 26 °C and 1050 rpm, centrifuged at 13.300 rpm for 10 sec and the supernatant removed. The gel pieces were resuspended in 100  $\mu\text{L}$  IAA buffer (Table 3.8) to alkylate reduced cysteine residues and incubated for 20 min at 26 °C and 1050 rpm. The tubes were centrifuged at 13.300 rpm for 10 sec and the supernatant removed. The gel pieces were resuspended in 150  $\mu\text{L}$  100 mM ammonium bicarbonate (NH<sub>4</sub>HCO<sub>3</sub>) and incubated for 15 min at 26 °C and 1050 rpm. 150  $\mu\text{L}$  ACN was added and the tubes were further incubated for 15 min at 26 °C and 1050 rpm. The tubes were centrifuged at 13.300 rpm for 10 sec and the supernatant removed. The gel pieces were dried in a SpeedVac (Eppendorf) for 10 min and resuspended in 20  $\mu\text{L}$  trypsin buffer (Table 3.8), followed by an incubation for 15 min on ice. Subsequently, 10-20  $\mu\text{L}$  trypsin buffer were added to submerge the gel pieces completely, followed by 20  $\mu\text{L}$  digestion buffer (Table 3.8) to prevent the gels from drying during digestion. In the following, the samples were incubated at 37 °C overnight.

#### 3.2.2.12.2 Extraction of Peptide Digestion Products

Gel pieces previously treated overnight with trypsin (3.2.2.12.1) were covered with 20  $\mu\text{L}$  dH<sub>2</sub>O and incubated for 15 min at 37 °C and 850 rpm in a Thermoshaker. Then, 80  $\mu\text{L}$  ACN was added and the sample tubes were further incubated for 15 min at 37 °C and 850 rpm. The tubes were centrifuged at 13.300 rpm for 10 sec and

the supernatants individually collected. The gel pieces were resuspended in 65  $\mu\text{L}$  5% formic acid and incubated for 15 min at 37 °C and 850 rpm. Subsequently, 50  $\mu\text{L}$  ACN were added and the gel pieces were incubated for 15 min at 37 °C and 850 rpm, followed by centrifugation at 13.300 rpm for 10 sec. The supernatants were collected and pooled with the priorly collected extracts. Finally, the extracts were dried in a SpeedVac (Eppendorf) for 4 h. For further mass spectrometric analysis, the dried extracts were resuspended in 10-20  $\mu\text{L}$  0.1% (v/v) trifluoroacetic acid by vortexing and incubation for 5 min in a sonication bath, followed by centrifugation at 10.000 rpm for 15 min. Unless the samples were directly analysed, they were stored at -20 °C.

#### **3.2.2.12.3 Mass Spectrometric Measurement**

All samples were analyzed on an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) coupled with an Ultimate 3000 RSLC system (Dionex, USA) in the research group “Bioanalytical Mass Spectrometry” of Prof. Dr. Henning Urlaub at the MPI for Biophysical Chemistry. Peptides were separated on an in-house-made capillary column (ReproSil-Pur 120 C18-AQ, 1.9  $\mu\text{m}$ , Dr. Maisch GmbH, 300 x 0.075 mm) with a linear gradient of increasing MS buffer B (Table 3.8) and decreasing buffer A (Table 3.8) at a constant flow rate of 300  $\text{nL min}^{-1}$ .

#### **3.2.2.12.4 Data Processing**

Peak list files containing peptide masses and their corresponding fragment ion masses and intensities were extracted from MS raw files by MaxQuant (version 1.5.2.8, Cox and Mann, 2008) and MS/MS spectra were searched against Uniprot yeast database (Uniprot ID UP000002311) via Andromeda search engine (Cox et al., 2011). The SILAC ratio of a protein is calculated by comparing the summed MS intensities of matching light and heavy isotopic labeled peptides.

### **3.2.3 Nucleic Acid Biochemical Methods**

#### **3.2.3.1 Preparation of Plasmid DNA**

Isolation and purification of plasmid DNA from transformed *E. coli* cells (3.2.1.2) was performed with the help of kit systems (3.1.13) according to manufacturer’s manual. In general, chemically competent *E. coli* DH10B cells (see Table 3.6) were transformed with the desired plasmid (Table 3.7) and used to inoculate an overnight culture (5 mL LB, 37 °C) containing a selective antibiotic (Table 3.3). The next day, cells were harvested by centrifugation at 13.300 rpm for 10 min and disrupted by alkaline lysis using the kit buffers. The plasmid DNA purification was performed over silica columns.

#### **3.2.3.2 DNA Quantification**

The DNA concentration was determined spectrophotometrically at a wavelength of 260 nm ( $A_{260}$ ) with pure water as a reference. The purity of the sample was calculated by the quotient of  $A_{260}$  and  $A_{280}$ .

## 4 Results

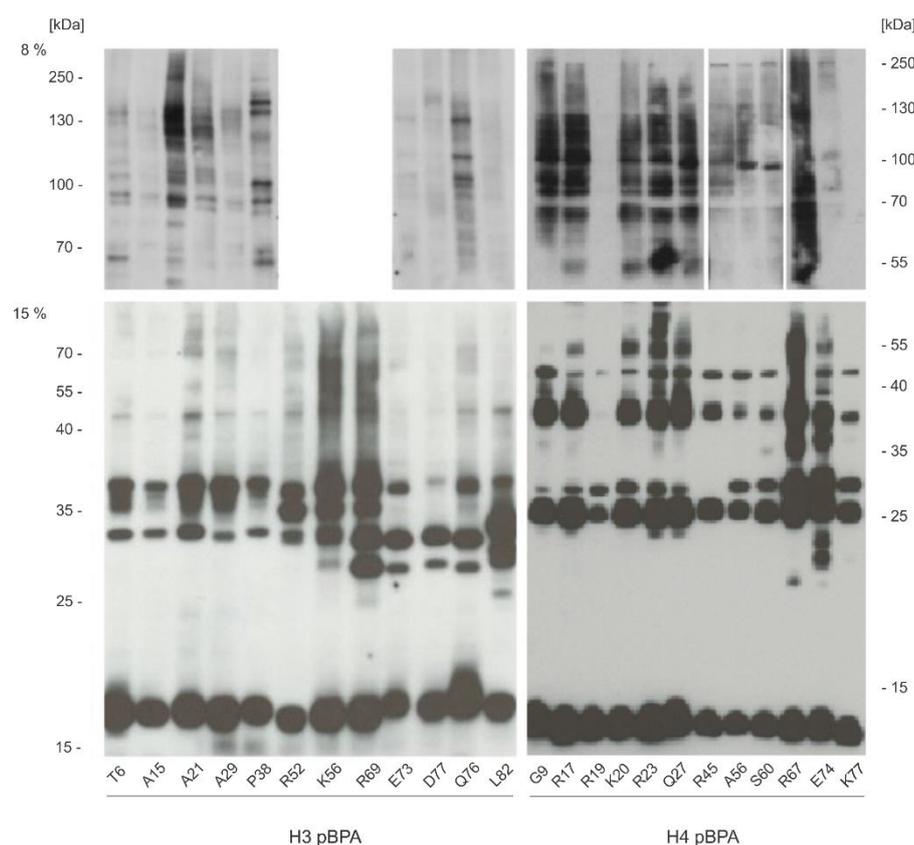
### 4.1 *In Vivo* Crosslinking of Histones in *S. cerevisiae*

In this project, the system for genetic code expansion developed by Chin and colleagues (2003) was used for the incorporation of the UV-inducible crosslinker *p*-benzoyl-L-phenylalanine (pBPA) on permissive sites on histones. Previous research has depicted the pBPA-driven *in vivo* crosslinking approach as an intriguing and reliable technique for the analysis of inter- and intranucleosomal interactions for the purpose of elucidating chromatin dynamics in a site-directed and UV-dependent manner (Rall, 2012; Wilkins et al. 2014).

#### 4.1.1 Scanning of Histone Surfaces with Genetically Encoded UV-Inducible Crosslinker

A comprehensive crosslink screening of histone pBPA mutants was performed to investigate the dynamics of histone-protein interactions in living yeast cells. Repositioning of the genetically encoded crosslinker amino acid pBPA across the surface of HA-epitope-tagged histones allows investigation of the interaction topology of the chromatin landscape in *S. cerevisiae* at single amino acid resolution.

The scanning was performed on 12 representative positions across the histone surfaces of H3 and H4 from the N-terminal domain across the central globular domain to the C-terminal domain. Positions for incorporation of the crosslinker were selected based on structural data (pdb: 1ID3). All experiments were performed in BY4741 cells expressing the pESC BPARS plasmid for the incorporation of pBPA in response to the amber suppression codon as well as distinct histone amber mutants.



**Figure 4.1:** Overview of crosslink patterns across the surfaces of histone H3 and H4.

Western blot analysis of whole-cell lysates from BY4741 cells expressing H3 or H4 amber variants with pBPA introduced at positions covering the N-terminal, globular and C-terminal domains. Western blot analysis was performed using anti-HA antibody. Upon UV treatment, every histone pBPA mutant produced a characteristic pattern of crosslink products.

Western blot analysis of whole-cell lysates from cells expressing amber mutants of H3 and H4 in the presence of pBPA revealed comparable amounts of full-length H3-HA and H4-HA protein as well as a high variety of crosslink product patterns depending on the site of amber codon positioning (Figure 4.1). Every crosslink pattern, indicated by a higher weight than the free histone proteins, was characterized by a specific banding pattern varying in number and molecular weight of bands and individual densities. All samples showed continuously distinct bands at approximately 27 kDa which might represent a crosslinked dimeric H3/H4 complex. Adjacent positions frequently exhibited similar banding patterns, which suggests the presence of potential interaction surfaces with identical proteins (e.g. H3: A21 and A29, H4: R23 and Q27).

Still, the variety of distinctly formed crosslink patterns depending on the position of the crosslinker greatly reflects the dynamic nature as well as specificity of interactions across the histone surface. The abundance of formed crosslink products decreased while moving away from the N-terminal tail towards the C-terminal domain, emphasizing the role of the N-terminal domain as a prominent binding surface. Notably, the majority of crosslink products of both histones were above 70 kDa. This suggests the assumption that many of the substituted sites on both histones interact with high molecular weight protein complexes, possibly involved in remodeling or regulation of chromatin structure.

However, both histones exhibited a variety of crosslink product formation with lower molecular weight, while such interactions were more abundant on histone H4. Some crosslink products were frequently observed and appear almost on every investigated position. They may represent trapped interactions with neighbouring histones or proteins involved in general cellular processes.

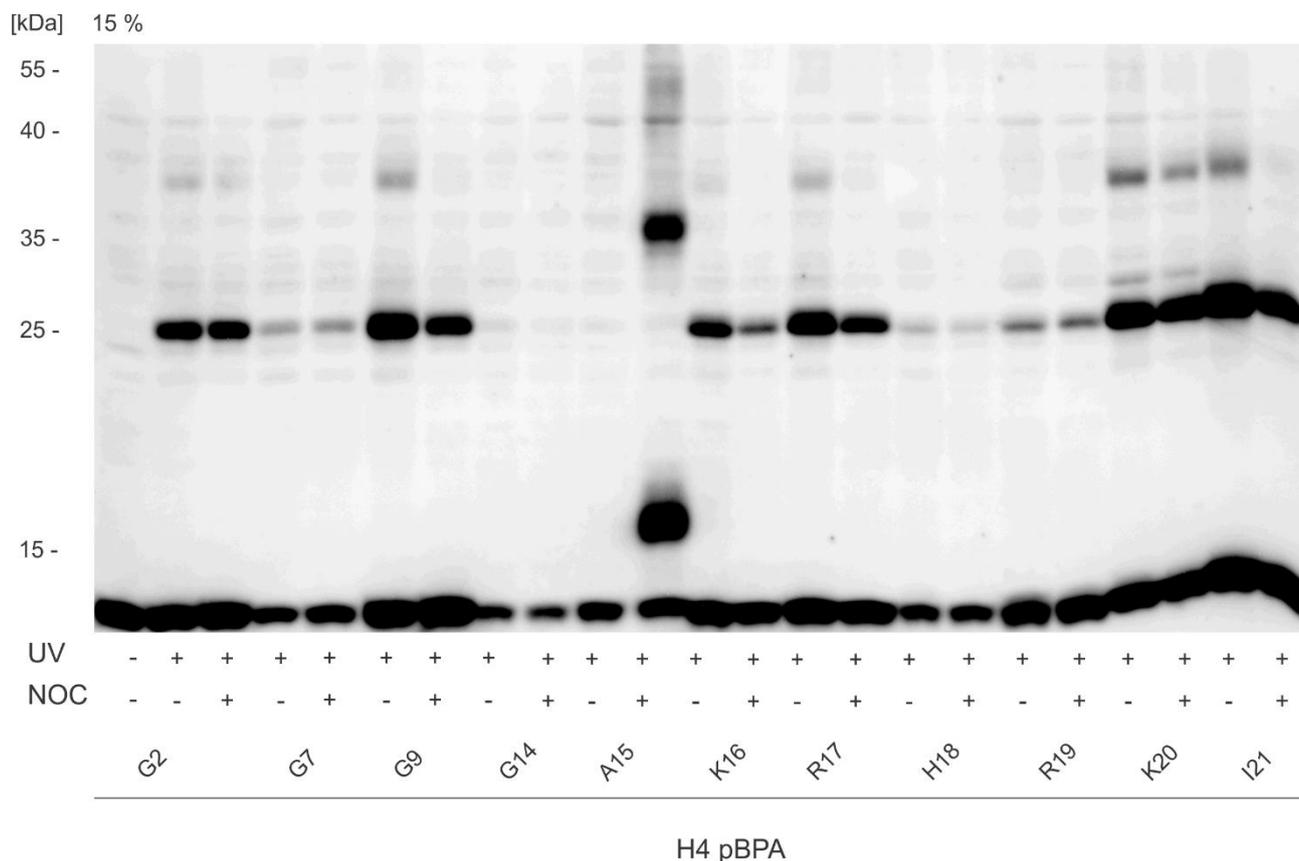
Overall, scanning the surfaces of histones H3 and H4 with the genetically encoded crosslinker pBPA revealed a highly diverse interaction interface of the chromatin landscape in *S. cerevisiae*. It has also proven that *in vivo* crosslinking is a powerful tool for the investigation of protein interactomes.

### **4.1.2 Spatio-Temporal Scanning of Binding Interfaces of N-Terminal H4 Domain**

The previous chapter has shown the diverse nature of histone-protein interactions. Still, it has to be considered that they can be transient and dependent on a particular stage of the cell cycle. Although the prepared samples, while being unsynchronized in their cellular phase, represent specific overviews of trapped interactions on single residues throughout the cell cycle, some interactions might remain difficult to be trapped. Hence, the impact of chromatin condensation on the intensity of crosslink products was investigated by the use of yeast cells synchronized with respect to their cell cycle.

Synchronizing the cells allows monitoring the interaction in space and time and analysis of specific crosslink behaviour over the course of mitosis. Synchronization of the cell cycle was accomplished by using the antimetabolic agent nocodazole which causes cell arrest and chromatin hypercondensation in metaphase. Since hypercondensation is thought to enhance nucleosomal contacts, cell arrest by nocodazole in metaphase was expected to result in the formation of increased amounts of crosslink product. All experiments were performed in BY4741 cells expressing the pESC BPARS plasmid for the incorporation of pBPA in response to the amber suppression codon as well as distinct histone amber mutants. Combined synchronization and crosslinking experiments were performed at positions covering the N-terminal domain of histone H4, considering its

importance for higher-order chromatin folding (Dorigo et al., 2003, Krajewski and Ausio, 1996). The synchronization was performed for 1.5 h.



**Figure 4.2:** Overview of crosslink patterns on N-terminal sites of histone H4 during mitosis.

Western blot analysis of whole-cell lysates from BY4741 cells expressing H4 amber variants with pBPA introduced at positions covering the N-terminal domains before and after treatment with nocodazole. Western blot analysis was performed using anti-HA antibody. Upon nocodazole treatment, a global reduction of crosslink product formation with different magnitudes was observed in all samples across the histone tail. *Legend:* Noc: Nocodazole.

Comparative Western blot analysis of unsynchronized and synchronized cells showed global changes in crosslink pattern across the H4 N-terminal domain upon arrest in metaphase (Figure 4.2). All sample pairs exhibited comparable amounts of full-length H4-HA protein. Nocodazole treatment led to a severe reduction of crosslink products in all samples across both histone tails, but with different degrees. All samples, except A15, displayed continuously distinct bands at approximately 27 kDa. These 27 kDa bands might represent a crosslinked dimeric H3/H4 complex. Notably, the intensity of these bands decreased upon metaphase arrest. This was to be expected since the reduction of space within nucleosome arrays during chromatin condensation should prohibit extensive inter- and intranucleosomal contacts. Still, the extent of reduction in crosslink product formation of residues on the N-terminal domain of H4 displays strongly the occurrence of an interaction of the H4 tail and the acidic patch upon chromatin condensation and thus a shielding of the tail from extensive protein interactions.

Another interesting band occurred at a molecular weight of approximately 37 kDa at positions G2, G9, K16, R17, K20, and I21. For G9, K16, R17 and I21, it only appeared in the unsynchronized samples, but vanished in the synchronized samples. For G2 and K20, it was detectable in both, but less dense in the synchronized sample. Also, K20 displayed an additional band at approximately 30 kDa which was detectable in both lanes, but less dense in the synchronized sample.

Opposing the disappearing of crosslinks upon synchronization, position A15 appeared to be an exception to this observation by showing no crosslinks in the unsynchronized sample, but two very prominent crosslinks after nocodazole-mediated mitotic arrest at ~15 kDa and ~35 kDa.

Summing up, synchronization of cells and *in vivo* crosslinking has proven to be a robust technique for resolving temporal correlations between histone-protein interactions and cell cycle phases. It can also be used to gain insight into structural and mechanistic details of chromosome dynamics.

### 4.2 Isolation of Histone pBPA Mutants and Crosslink Products

The expression of histones harboring the site-specifically encoded crosslinker amino acid pBPA allows the mapping of interaction networks with single amino acid resolution across the chromatin landscape in living yeast cells. Every crosslink experiment produces a specific banding pattern varying in number and molecular weight of bands and individual densities. Thereby, each site produces a vast amount of structural information details of which need to be revealed.

Identification of interaction partners of a protein of interest is an established way to get an insight into its function. This can either be done by a molecular weight shift approach or by isolating the interaction partners and subjecting them to mass spectrometry. The mass shift assay is a cost- and time-effective method for the identification of crosslink products. It is performed by the candidate approach in which a putative interaction partner is genomically tagged in a duplicated crosslink experiment. Actual crosslinking to the putative interaction partner would hence result in an additional molecular weight shift, detectable by Western blot. This approach is ideal for the identification of known binding partners to a certain residue or region. Still, the throughput of this technique is very low.

To identify all interactions trapped in a single crosslinking experiment, a more efficient and high-throughput method needs to be employed. The combination of affinity purification of a protein of interest followed by quantitative mass spectrometry using SILAC presents a powerful means for the discovery and identification of novel protein-protein interactions. The following paragraphs display the results achieved in the efforts to optimize enrichment of crosslink products in yeast cells for subsequent SILAC analysis.

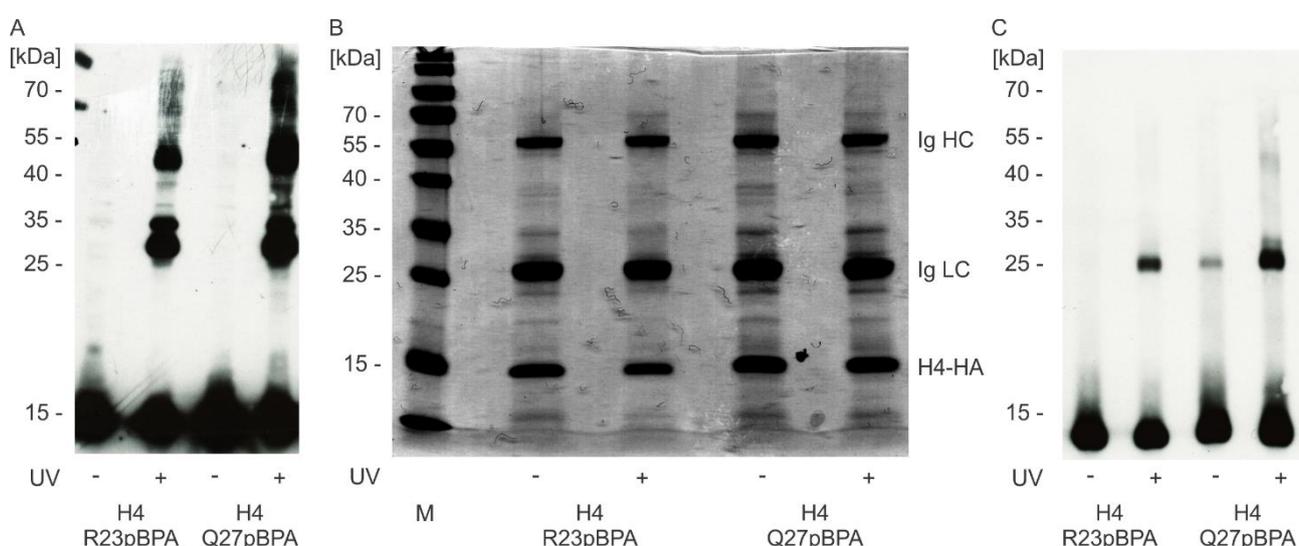
#### 4.2.1 Immunoprecipitation of pBPA Crosslink Products

Because of their low abundance, analysis of crosslink products has been difficult. Direct MS analysis of these complexes has proven challenging because of stoichiometric underrepresentation of crosslink products compared to free protein.

In the following, a protocol for the targeted enrichment of crosslinks was established. H4 is reported to be very unstable and has proven to be hard to isolate in the past. Therefore, initial trials were performed on H4 amber

mutants in order to find optimal conditions for crosslink product isolation in general. The immunoprecipitation (IP) procedure includes mechanical cell disruption, protein extraction with chaotropic agents, affinity purification and antibody-mediated protein isolation. In the process of establishing and optimizing the immunoprecipitation protocol, a multitude of parameters were tested for beneficial effects on quality and quantity of isolated crosslink products. Every step was determined by trial and error testing of various purification methods and buffer compositions. The obtained protein solutions are highly concentrated in a small volume which allows maximum saturation of the beads during immunoprecipitation while preventing protein precipitation.

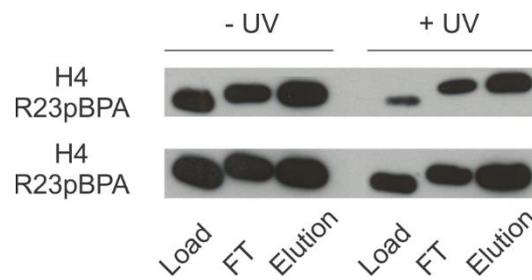
Large scale cultures of BY4741 cells expressing H4-HA with amber mutations at position R23 or Q27 in the presence of 1 mM pBPA were treated with UV-irradiation for crosslinking. H4-HA full-length protein and its respective crosslink products were immunoprecipitated using anti-HA antibodies conjugated to agarose beads (Figure 4.3).



**Figure 4.3:** Immunoprecipitation of pBPA crosslink products.

A) Western blot of the whole-cell lysates of yeast strain BY4741 expressing either H4 R23pBPA or H4 Q27pBPA in the presence of 1 mM pBPA. UV-irradiation caused formation of distinct crosslink patterns in both positions. B) Coomassie-stained gel of IP elution fractions of H4 R23pBPA or H4 Q27pBPA mutants. Full-length H4-HA (~13 kDa) was enriched in all samples. C) Western blot analysis of IP elution samples of H4 R23pBPA and H4 Q27pBPA showed enrichment of crosslink products in the UV-treated samples (+ UV). *Legend:* M: Marker; Ig LC: Immunoglobulin light chain; Ig HC: Immunoglobulin heavy chain.

Western blot analysis of the whole-cell lysate control samples exhibited a UV-dependent crosslink product formation on both positions, verifying the good initial quality of the analyzed IP samples (Figure 4.3A). Samples of the non-bound material fraction, the flow-through, showed no significantly decreased amount of H4-HA full-length protein in comparison to the load sample, which indicates an overabundance of available isolated protein in the sample (Figure 4.4).



**Figure 4.4:** Immunoprecipitation of pBPA crosslink products.

Western blot analysis of the IP procedure of H4 R23pBPA and H4 Q27pBPA showed no significant reduction in H4-HA full-length protein concentration in the non-bound material fraction (FT). This indicates an overabundance of available isolated protein in the sample. *Legend:* Load: IP load; FT: IP flow-through fraction; Elution: IP elution fraction.

While the control elution samples, which were not subjected to UV-irradiation, contained only the enriched H4-HA protein, the elution of the UV-treated samples revealed the isolation of H4-HA crosslink products (Figure 4.3C). To further display the enrichment of the H4-HA protein, all elution fractions were loaded on a 10% Bis-Tris gel and stained with Instant Blue (Figure 4.3B). All samples showed dense bands at the expected size of H4-HA (~13 kDa), which verifies the enrichment of the target protein. Also, all lanes showed two prominent bands of high density at ~25 kDa and ~55 kDa. Both can be assigned to the anti-HA-antibody, the band at ~25 kDa representing the light chain (Ig LC) and the band at ~55 kDa representing the heavy chain (Ig HC) of the antibody.

Furthermore, all samples showed bands with a molecular weight of 10 kDa and sample dependent variations in density, likely representing truncation products of the H4-HA protein. Moreover, it could not clearly be visually differentiated between non-UV-treated and UV-treated samples by Coomassie staining (Figure 4.3B). Although all samples showed a variety of additional bands of distinct molecular weight, they were present across sample pairs. The lack of more prominent crosslink product bands might be the outcome of a general low abundance of crosslink products and a competitive binding of free H4-HA protein and crosslink products to the beads in favor of the H4-HA protein. However, these results prove that the developed IP protocol for HA-epitope-tagged histones with anti-HA-conjugated agarose beads can be successfully applied to isolate and enrich crosslink products.

#### 4.2.2 Optimization of Immunoprecipitation Procedure for SILAC Experiments

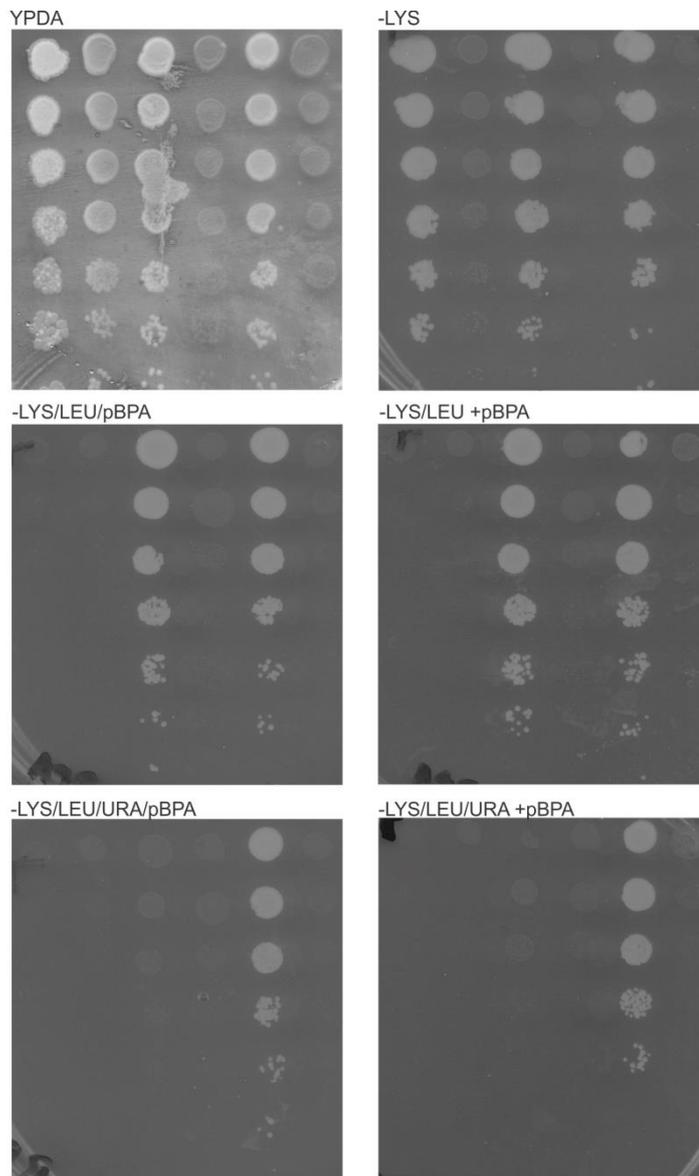
SILAC experiments involve culturing of cells in media containing either light (e.g.  $^{12}\text{C}$  and  $^{14}\text{N}$ ) or heavy (e.g.  $^{13}\text{C}$  and  $^{15}\text{N}$ ) labeled amino acids, mostly L-arginine and L-lysine, to incorporate the isotopically labeled amino acids into proteins through the metabolic cycle. Although *S. cerevisiae* can normally synthesize all amino acids, SILAC labeling can be performed by using deletion strains in which the genes *LYS2* and *ARG4* for the biosynthesis of these specific amino acids are disrupted.

#### 4.2.2.1 Growth Conditions and Characterization of SILAC Strain

The usage of stable isotopically labeled amino acids can be costly, depending on the number of experiments and required amounts of cells. Therefore, it is important to screen for optimal but cost-efficient growth conditions. SILAC-specific medium is deprived of L-lysine and L-arginine, and later supplemented with light or heavy forms of both amino acids.

To determine optimal growth conditions, YPH499 cells lacking the genes *LYS2* and *ARG4* and expressing H2A-HA with an amber mutation at position Y58 were grown in the presence of 1 mM pBPA under different concentrations of light L-arginine and L-lysine. The concentrations of L-arginine and L-lysine were supplemented at 15/15, 15/30, 30/30 and 30/60 (mg/L). Cells grown in media not restricted in both amino acids reached optical densities (OD<sub>600</sub>) of 5.0 and higher. Limiting the concentrations of both amino acids to 15 mg/L resulted in stalling of growth at 2.0 OD<sub>600</sub>. An increase in L-lysine to a final concentration of 30 mg/L led to an extended growth up to 3.0 OD<sub>600</sub>. Further increasing of L-arginine to 30 mg/L resulted in only a minor increment until growth stalled at 3.5 OD<sub>600</sub>. When the cells were grown in the presence of 30 mg/L L-arginine and 60 mg/L L-lysine, cells were able to reach 5.0 OD<sub>600</sub> before their growth stalled. Therefore, for further experiments under SILAC conditions, concentrations of 30 mg/L L-arginine and 60 mg/L L-lysine were chosen to enable optimal, but cost-efficient growth conditions. The original YPH499 wild-type strain possesses a disrupted *LYS2* gene which is accomplished by an amber stop codon at position 801. For SILAC applicability, an additional *ARG4* deletion was introduced to allow selection for L-arginine and L-lysine.

Since the crosslinking approach relies on suppression of amber stop codons, the cells could be susceptible to an amber suppression-mediated read-through of the *LYS2* gene, which would compromise the planned incorporation of heavy L-lysine. Also, it was observed that distinct natural amino acids can compete for insertion at amber codons. The reason for background suppression *in vivo* is mis-acylation of the suppressor tRNA molecules by the evolved synthetase with an endogenous amino acid, such as tyrosine or phenylalanine, in the rich media (Johnson et al., 2011; Chen et al., 2007; O'Donoghue et al., 2012; Heinemann et al., 2012; Odoi et al., 2013). Therefore, BY4741 and YPH499  $\Delta$ Arg4 cells, either wild-type or expressing the pESC plasmid (leu) and/or H4 R23 amber mutant (ura), were grown on different restrictive agar plates in the absence or presence of 1 mM pBPA.



YPH499 $\Delta$ Arg4	-	+	-	+	-	+	-	+	-	+	-	+
BY4741 WT	+	-	+	-	+	-	+	-	+	-	+	-
pESC BPARS	-	-	+	+	+	+	-	-	+	+	+	+
pRS H4 R23pBPA	-	-	-	-	+	+	-	-	-	-	+	+

**Figure 4.5:** Growth conditions and characterization of YPH499  $\Delta$ Arg4 strain.

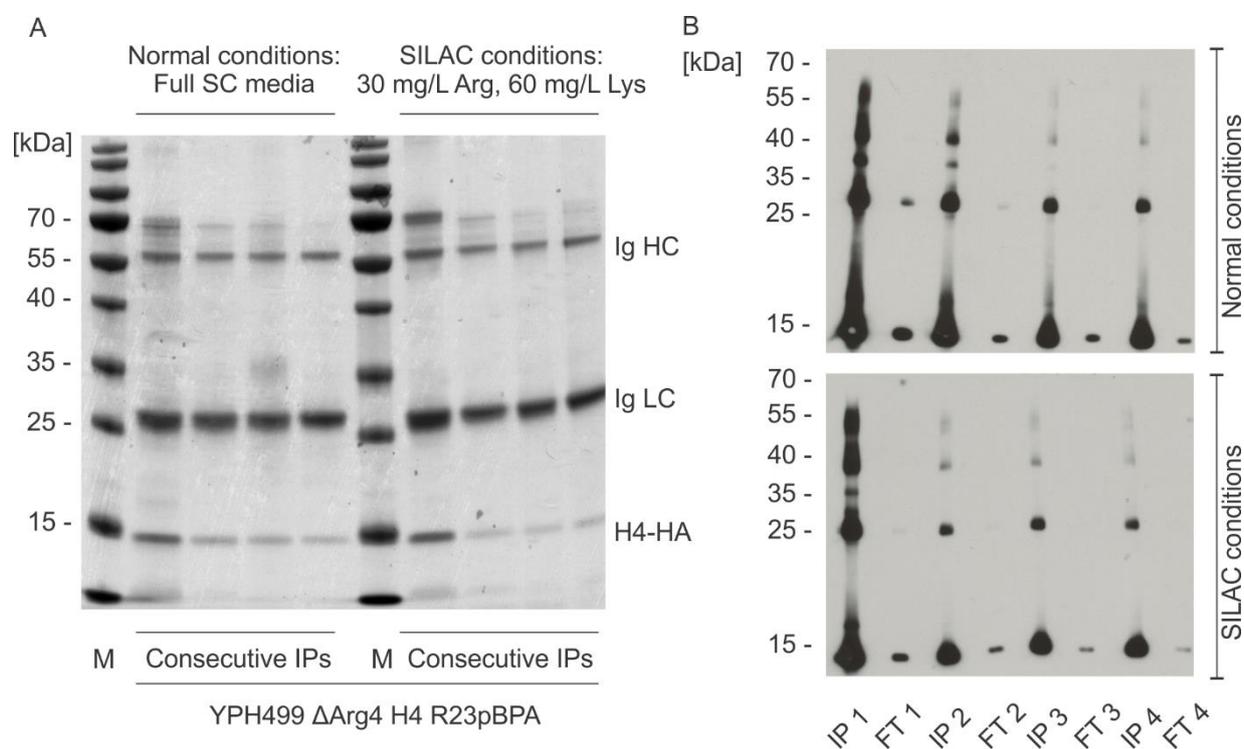
Growth assay for evaluation of susceptibility of *LYS2* amber mutation in YPH499  $\Delta$ Arg4 to amber suppression.

BY4741 and YPH499  $\Delta$ Arg4 cells, either wild-type or expressing the pESC BPARS plasmid (*leu*) and/or H4 R23 amber mutant (*ura*), were grown on different restrictive agar plates in the absence or presence of 1 mM pBPA. YPH499  $\Delta$ Arg4 were incapable of survival without L-lysine in the presence of the amber suppression system.

The depicted growth assay in Figure 4.5 revealed that only the BY4741 strain with intact *LYS2* gene was able to grow in the absence of L-lysine. YPH499  $\Delta$ Arg4 cells expressing the pESC plasmid and/or H4 R23 amber mutant were incapable of survival without L-lysine, independent of the absence or presence of pBPA. Additionally, no growth caused by insertion of natural amino acids in response to the amber codon in the disrupted *LYS2* gene was observed in the absence of pBPA. In conclusion, YPH499  $\Delta$ Arg4 cells can be used in combination with the amber suppression system for SILAC experiments.

#### 4.2.2.2 Immunoprecipitation of Crosslink Products under SILAC Conditions

To quantify the effect of SILAC growth conditions on the yield of crosslink products, immunoprecipitation was performed on YPH499  $\Delta$ Arg4 expressing H4-HA with an amber mutation at position R23 either under normal or restrictive SILAC conditions. Samples were equilibrated by OD and analyzed by Western blot and Coomassie gel staining.



**Figure 4.6:** Immunoprecipitation of pBPA crosslink products under SILAC conditions.

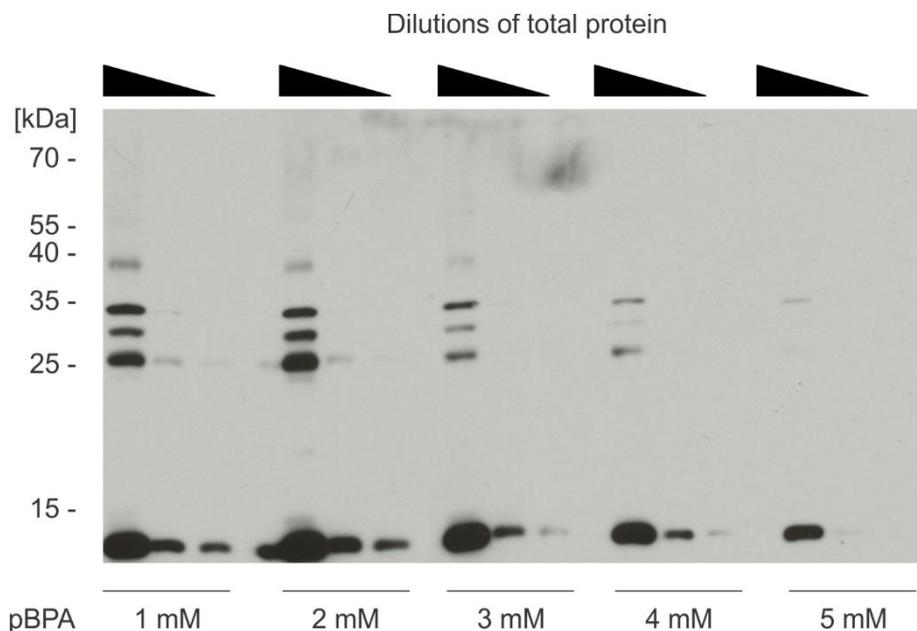
A) Coomassie-stained gel of consecutive IP elution fractions of H4 R23pBPA isolated from YPH499  $\Delta$ Arg4 cells under normal or restricted SILAC growth conditions. Quantification of the stained H4-HA proteins indicated a minor reduction of total protein upon growth under SILAC conditions. B) Western blot analysis of consecutive IP elution samples of H4 R23pBPA isolated from YPH499  $\Delta$ Arg4 cells under normal or restricted SILAC growth conditions. SILAC growth conditions reduced the overall amount of full-length H4-HA protein and the formation of crosslink products. *Legend:* M: Marker; Ig LC: Immunoglobulin light chain; Ig HC: Immunoglobulin heavy chain.

Consecutive IPs showed that the total amount of H4-HA protein and crosslink products slightly decreased under SILAC growth conditions (Figure 4.6B). In both cases the first performed IP resulted in a saturation of the beads with proteins of the sample. However, the amount of crosslink products, indicated by higher molecular weight than the H4-HA full-length protein, was significantly reduced in cells grown with restricted concentrations of L-arginine and L-lysine. Repeated consecutive rounds of IPs against the flow-through from samples revealed remaining crosslink products in the normally grown samples, but almost no remaining crosslink products in the restrictively grown samples. The residual eluates were loaded on a 10% Bis-Tris gel and stained with Instant Blue (Figure 4.6A). Quantification of the stained H4-HA proteins indicated a minor

reduction of total protein under SILAC conditions. This was supported by purification data indicating a reduction of signal of the same volume of elution fractions, proportional to the load peaks, from ~1200 mAu under normal conditions to ~650 mAu under SILAC conditions (See Figure 7.1 in appendix). These findings emphasize the influence of restrictive growth conditions on the formation of crosslink products. Although the principle of immunoprecipitation can be successfully applied to H4-HA crosslink samples derived from SILAC growth conditions, the protocol required further optimization of crosslink-products enrichment for enhanced identification by MS.

#### 4.2.2.3 Increased pBPA Concentration Improves Full-Length Protein Yield

Earlier experiments have shown a contribution of 10% of plasmid-borne histone amber mutants to the endogenous histone pool under normal growth conditions (Rall, 2012; Wilkins et al., 2014). To investigate whether the decrease of crosslink product formation under nutrition-restricted SILAC growth conditions could be compensated by increased pBPA concentrations, YPH499  $\Delta$ Arg4 cells expressing H4 R23pBPA were grown in increasing amounts of pBPA, final concentrations ranging from 1 mM to 5 mM. All samples were treated with UV-light and analyzed by Western blot.



**Figure 4.7:** Increasing pBPA concentration improves full-length protein yield.

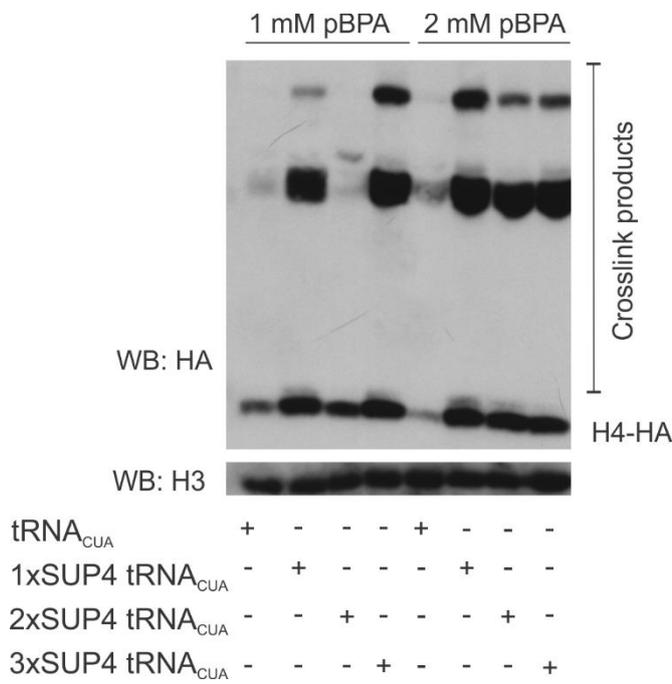
Western blot analysis of dilution series of whole-cell lysate (from left to right: 1:1, 1:5, 1:10) obtained from H4 R23pBPA-expressing BY4741 cells grown in SC media supplemented with different concentrations (1, 2, 3, 4, 5 mM) of pBPA. The formation of full-length H4-HA protein and its crosslink products increased at a concentration of 2 mM pBPA, but decreased exponentially with further increase.

As expected from cells grown in the presence of pBPA, Western blot analysis revealed a band at the molecular size of full-length H4-HA protein which was observable in all samples (Figure 4.7). However, the amount of produced protein differed between samples, depending on the pBPA concentration. A doubling of pBPA concentration to 2 mM resulted in ~30% increase in full-length protein formation. However, further increasing of pBPA did not result in a linear rise of total protein amounts, but a decrease. Titration of total protein amounts showed a decrease of protein expression at 3 mM. This negative effect continued with further increasing amounts of pBPA, peaking in a minimal expression of full-length H4-HA protein in the presence of 5 mM pBPA.

All samples showed formation of crosslink products after UV-irradiation. However, the amounts of produced crosslinks correlated with the amounts of available full-length H4-HA protein. Therefore, a slight increase in crosslink product formation was observed in the presence of 2 mM pBPA. Still, the amount of formed crosslinks decreased rapidly with further increasing pBPA concentrations, until there was hardly any formation detectable in the presence of 5 mM pBPA. While culturing the cells, increasing degrees of non-definable amino acid precipitation were observable in the media in correlation to rising amounts of pBPA. Therefore, it appears possible that the decrease in total protein yield is a result of a lack of nutrients caused by pH-related amino acid precipitation as well as poor solubility of pBPA concentrations above 3 mM. For further experiments, a final concentration of 2 mM pBPA was determined to be optimal for increased amber mutant yields.

#### 4.2.2.4 Increased tRNA<sub>CUA</sub> Levels Improve pBPA Incorporation Efficiency

The original *E. coli* BPARS/tRNA<sub>CUA</sub> pair for genetic code expansion, developed by Chin and colleagues (2003), is encoded on a pRS vector with a high-copy 2 $\mu$  yeast replicative origin. However, they state that the expression of the amber suppressor tRNA is probably driven by a cryptic promoter on the plasmid or in the gene, but fail to provide further explanation (Chin et al., 2003; Chen et al., 2007). Thereby, basal expression levels of the tRNA<sub>CUA</sub> allow for maintaining cell growth and selection, but prohibit high yields of amber mutants. To optimize the system for higher protein yields, advantage was taken of an improved genetic code expansion system which was developed by Chen (2007). It works by inserting the *E. coli* tRNA<sub>CUA</sub> gene, which contains an intact B box, within the flanking regions of the yeast suppressor tRNA gene *SUP4*, which contain internal A and B boxes. The produced chimaera gene cassette was subsequently substituted in copy numbers of up to 3 for the original *E. coli* tRNA<sub>CUA</sub> gene in the original pRS plasmid. The optimized plasmids were transformed separately into YPH499  $\Delta$ Arg4 cells expressing H4-HA with an amber mutation at position R23 and grown under SILAC conditions in the presence of 1-2 mM pBPA.



**Figure 4.8:** Increasing tRNA<sub>CUA</sub> levels increases pBPA incorporation efficiency.

Western blot analysis of whole-cell lysate obtained from YPH499  $\Delta$ Arg4 cells expressing H4 R23pBPA by amber suppression mediated by pESC BPARS-tRNA<sub>CUA</sub>, pESC BPARS-1xSUP4 tRNA<sub>CUA</sub>, pESC BPARS-2xSUP4 tRNA<sub>CUA</sub> or pESC BPARS-3xSUP4 tRNA<sub>CUA</sub> in SILAC SC media supplemented with 1 or 2 mM pBPA. Insertion of a single or multiple SUP4-tRNA<sub>CUA</sub> cassettes significantly increased the overall yield in full-length H4-HA protein and crosslink products. A combination of pESC BPARS-1xSUP4 tRNA<sub>CUA</sub> and a supplementation of 2 mM pBPA proved to produce optimal yield and allow optimal growth.

*Legend:* tRNA<sub>CUA</sub>: pESC BPARS-tRNA<sub>CUA</sub>;

1xSUP4 tRNA<sub>CUA</sub>: pESC BPARS-1xSUP4 tRNA<sub>CUA</sub>;

2xSUP4 tRNA<sub>CUA</sub>: pESC BPARS-2xSUP4 tRNA<sub>CUA</sub>;

3xSUP4 tRNA<sub>CUA</sub>: pESC BPARS-3xSUP4 tRNA<sub>CUA</sub>.

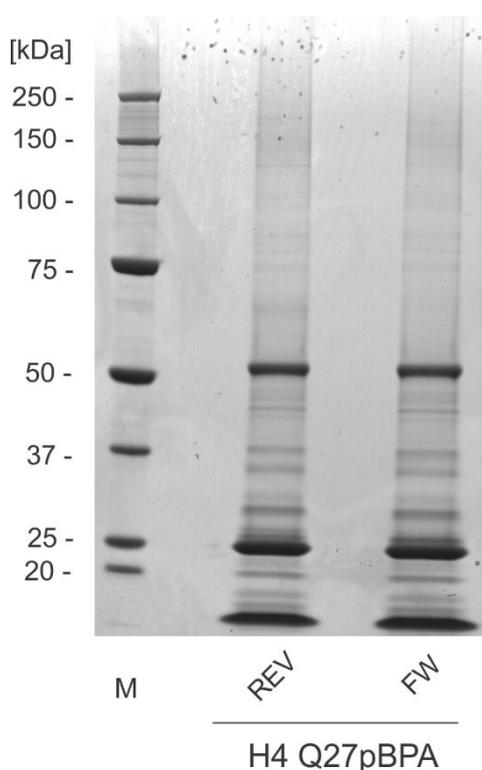
Western blot analysis showed an overall increase in full-length H4-HA protein expression, up to 5-fold, in all samples compared to the original pESC BPARS plasmid (Figure 4.8). In the presence of 1 mM pBPA, the amount of crosslink product formation after UV-irradiation increased linearly with the number of SUP4-tRNA<sub>CUA</sub> cassettes, rendering the 3xSUP4-tRNA<sub>CUA</sub> construct the most efficient. Doubling of the pBPA concentration to 2 mM and usage of the 1xSUP4-tRNA<sub>CUA</sub> construct resulted in similar amounts of full-length protein and crosslink products as the last noted 3xSUP4-tRNA<sub>CUA</sub> in the presence of 1 mM pBPA.

However, increasing the number of SUP4-tRNA<sub>CUA</sub> cassettes did not lead to higher crosslink product formation, but a decline. The doubling times of cells which either express a 2xSUP4-tRNA<sub>CUA</sub> or a 3xSUP4-tRNA<sub>CUA</sub> construct increased 1.5- to 2-fold, independent of used pBPA concentration. Usage of a 1xSUP4-tRNA<sub>CUA</sub> construct showed no observable effect on growth rate. Therefore, both multicopy constructs appeared to represent a metabolic burden for the cells, which may result in impaired protein biosynthesis. Still, by using the SUP4-tRNA<sub>CUA</sub> chimaera gene for expression of the *E. coli* tRNA<sub>CUA</sub>, the amounts of expressed full-length histone-HA protein and its crosslink products were successfully increased by a factor of 5 under SILAC growth conditions. Since the cells are already experiencing metabolic stress in terms of nutrient restriction under SILAC growth conditions, the combination of the 1xSUP4-tRNA<sub>CUA</sub> construct and a supplementation of 2 mM pBPA to the growth media will be used for further cell culturing under SILAC conditions.

#### 4.2.2.5 YPH499 Strain Compromises Peptide Pair Formation in SILAC Experiments

Based on previously described crosslinking experiments (Figure 4.1), position Q27 on histone H4 was chosen for initial SILAC analysis.

Two times two cell cultures of transformed YPH499  $\Delta$ Arg4 cells carrying plasmids for the expression of the BPARS/1xSUP4-tRNA<sub>CUA</sub> pair and an amber codon-mutated histone H4 (Q27) were cultured in appropriate SILAC SC dropout media supplemented with 2 mM pBPA and either light or heavy 30 mg/L L-arginine and 60 mg/L L-lysine. Equal parts of heavy (here: <sup>13</sup>C<sub>6</sub>-L-arginine and <sup>13</sup>C<sub>6</sub>-L-lysine), UV-treated cells were mixed with light, non-UV-treated cells, resulting in the forward sample. Mixing of light, UV-treated cells with heavy, non-UV-treated cells resulted in assembly of the reverse sample. Both samples were concentrated to final concentrations of ~0.71 mg/mL and subjected to immunoprecipitation of H4-HA and its crosslink products. The acquired elution samples were loaded on a 4-12% Bis-Tris gel and stained with Instant Blue. Gel analysis shows an even saturation of the beads with H4-HA protein as well as banding patterns in both samples (Figure 4.9). Subsequently, the samples were digested and analyzed by MS. Quantification of SILAC pairs was expected to result in the formation of peptide pairs with a mass shift of 6 or 12 Da. However, no peptide pair formation was observed in either sample.



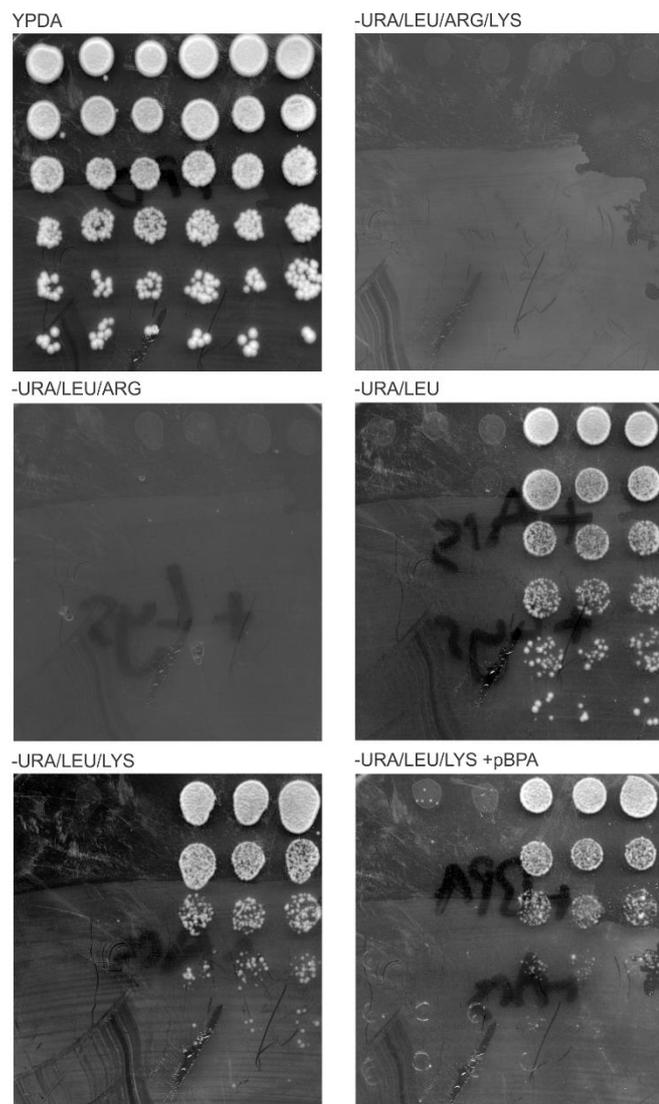
**Figure 4.9:** SILAC immunoprecipitation of H4 Q27pBPA crosslink products for MS analysis.

Coomassie-stained gel of IP elution fractions from forward and reverse samples of H4 R23pBPA mutants expressed in YPH499  $\Delta$ Arg4 cells. Full-length H4-HA (~13 kDa) was enriched in both samples. *Legend:* M: Marker; REV: Reverse sample; FW: Forward sample.

#### 4.2.2.6 Improved tRNA<sub>CUA</sub> levels Cause Amber Suppression of $\Delta$ LYS2 in YPH499 Strain

The YPH499 yeast strain contains a disrupted *LYS2* gene, which is accomplished by an insertion of an amber stop codon at position 801. Suppression of this amber codon may compromise the incorporation of supplemented light or heavy L-lysine for SILAC. As previously shown, the original pESC BPARS plasmid has no detectable effect on mediating read-through of the *LYS2* gene as well as suppressing the amber codon with natural amino acids in the absence of pBPA.

While improving the tRNA<sub>CUA</sub> levels resulted in a significant increase of full-length amber mutant protein and crosslink products, it remained to be determined if the improved system had an effect on the susceptibility of the cells for amber suppression-mediated read-through. Therefore, YPH499  $\Delta$ Arg4 cells, either wild-type or expressing the pESC BPARS-1xSUP4 tRNA<sub>CUA</sub> plasmid (leu) and H4 R23 amber mutant (ura), were grown on different restrictive agar plates in the absence or presence of 1 mM pBPA (Figure 4.10).



**Figure 4.10:** Improved tRNA<sub>CUA</sub> levels cause amber suppression of  $\Delta$ LYS2 in YPH499 strain.

Growth assay for evaluation of susceptibility of *LYS2* amber mutation in YPH499  $\Delta$ Arg4 to optimized amber suppression.

YPH499  $\Delta$ Arg4 cells, either wild-type or expressing the pESC BPARS-1xSUP4 tRNA<sub>CUA</sub> plasmid (leu) and H4 R23pBPA mutant (ura), were grown on different restrictive agar plates in the absence or presence of 1 mM pBPA. Independent of pBPA, YPH499  $\Delta$ Arg4 cells were capable of survival without L-lysine by amber suppression of the amber stop codon mutation in the supposedly knocked-out *LYS2* gene.

1x SUP4 tRNA <sub>CUA</sub>	-	-	-	+	+	+	-	-	-	+	+	+
H4 R23pBPA	-	-	-	+	+	+	-	-	-	+	+	+

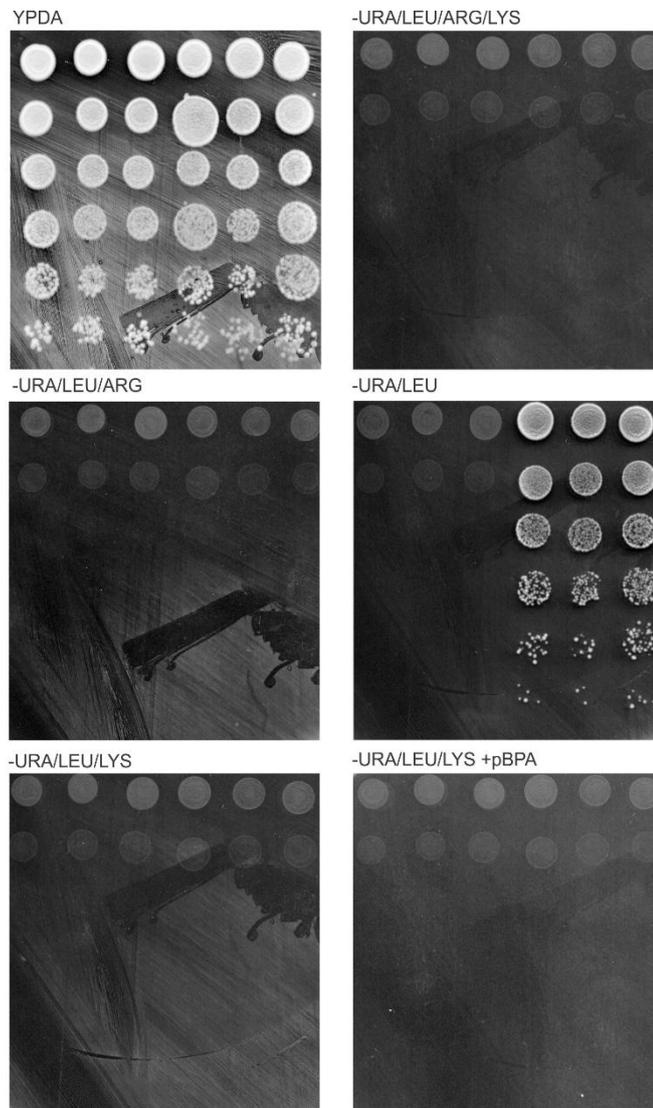
As expected, the transformed cells were unable to grow in the absence of L-arginine and L-lysine combined as well as L-arginine alone. However, supplementation of only L-arginine allowed the cells harboring the pESC BPARS-1xSUP4 tRNA<sub>CUA</sub> plasmid to grow in the absence of L-lysine. Notably, this effect was observed independent of the absence or presence of pBPA. Cell growth in the absence of pBPA was evidently the result of mis-acylated natural amino acids in response to the amber codon.

Still, the cells growing in the absence of L-lysine and presence of pBPA showed slightly reduced growth, which may be caused by augmented occupation of the amber suppression system for the biosynthesis of the H4 amber mutant with the unnatural amino acid.

Taken together, these findings confirm that the improved genetic code expansion system rendered the YPH499  $\Delta$ Arg4 no longer suitable for SILAC experiments. Therefore, an alternative strain had to be chosen.

#### **4.2.2.7 Double Deletion Strain of BY4741 Qualified for SILAC**

Since the YPH499 strain revealed to be susceptible to the improved genetic code expansion system, an alternative  $\Delta$ Arg4/ $\Delta$ Lys2 strain with a different cellular backbone was selected. In a BY4741 wild-type strain, both genes were disrupted by introduction of an auxotrophic marker cassette: *ARG4* was replaced with a *KanMX* cassette and *LYS2* was knocked out with a *HIS3* cassette. Therefore, the strain was expected to be eligible for SILAC experiments. To test its suitability, BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells, either wild-type or expressing the pESC BPARS-1xSUP4 tRNA<sub>CUA</sub> plasmid (leu) and H4 R23 amber mutant (ura), were grown on different restrictive agar plates in the absence or presence of 1 mM pBPA (Figure 4.11).



1x SUP4 tRNA <sub>CUA</sub>	-	-	-	+	+	+	-	-	-	+	+	+
H4 R23pBPA	-	-	-	+	+	+	-	-	-	+	+	+

**Figure 4.11:** Double deletion strain of BY4741 qualified for SILAC.

Growth assay for evaluation of suitability of BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells for SILAC experiments. BY4741  $\Delta$ Arg4/ $\Delta$ Lys2, either wild-type or expressing the pESC BPARS-1xSUP4 tRNA<sub>CUA</sub> plasmid (leu) and H4 R23pBPA mutant (ura), were grown on different restrictive agar plates in the absence or presence of 1 mM pBPA. BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells were incapable of survival without L-lysine in the absence or presence of the amber suppression system.

In an opposing way from the previously made observations in YPH499 cells, the transformed BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells were completely incapable of growth in the absence of L-arginine and/or L-lysine.

Their growth was fully dependent on the supplementation of both amino acids to the growth media. Additionally, usage of the pESC BPARS-1xSUP4 tRNA<sub>CUA</sub> plasmid did not promote any noticeable changes to the phenotype of the cells nor their growth behaviour.

Collectively, this growth assay proved the dependency of the BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 strain on the supplementation of L-arginine and L-lysine, rendering it suitable for SILAC experiments.

### 4.3 Identification of Crosslink Products by MS-Based SILAC Approach

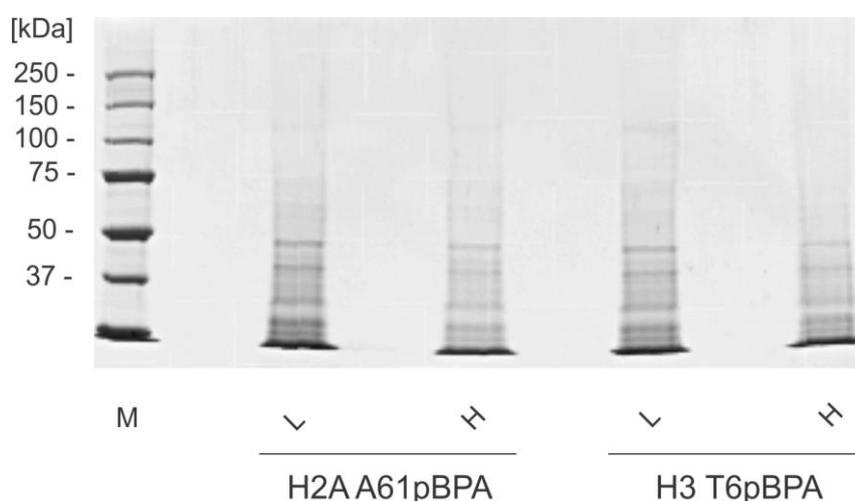
Quantitative proteomics combined with SILAC immunoprecipitation represent a powerful means for the identification of protein interaction partners. The following paragraphs will depict the results of the effort to combine *in vivo* crosslinking with a genetically encoded crosslinker amino acid in *S. cerevisiae* with a SILAC-based MS approach for quantitative identification of crosslink products.

### 4.3.1 SILAC Labeling Efficient in BY4741 Deletion Strain

While the YPH499  $\Delta$ Arg4 strain has proven to be an interference factor for SILAC experiments, the BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 strain showed to be applicable. The mass difference between  $^{13}\text{C}_6$  and  $^{12}\text{C}_6$  is 6 Da. Combining  $^{13}\text{C}_6$ -L-arginine and  $^{13}\text{C}_6$ -L-lysine can thereby only produce mass shifts of 6 or 12 Da. Depending on the peptide sequence, the identical mass of both labeled amino acids can be of hindrance for the identification of these peptides. Therefore,  $^{13}\text{C}_6$ -L-arginine was replaced by  $^{13}\text{C}_6$ - $^{15}\text{N}_4$ -L-arginine in all further SILAC experiments, which produces an increased mass shift of 10 Da through the isotopically labeled  $^{15}\text{N}_4$ . Thereby, peptide pairs with mass shifts of 6, 10, 12, 16 or 20 Da can be produced, depending on the peptide sequence, which further increases the identification rate of the approach.

To ensure complete metabolic labeling of protein, cells should undergo at least 5-10 rounds of doubling in media supplemented with the isotopically labeled amino acids. Thereby, protein labeling is achieved by: (I) dilution of non-labeled protein with *de novo* synthesized labeled protein and (II) degradation of non-labeled protein.

For incorporation testing, BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells co-expressing the pESC BPARS/1xSUP4-tRNA<sub>CUA</sub> plasmid and a pRS plasmid encoding either an H2A or H3 amber mutant were grown in appropriate SILAC SC media supplemented with 2 mM pBPA and either light or heavy 30 mg/L L-arginine and 60 mg/L L-lysine. Cells from each culture were lysed and digested individually. A mixture of peptides from a gel band is analyzed by MS and the ratio of heavy labeled peptides (H) to the remaining light non-labeled peptides (L) is determined by comparative analysis of the raw files by MaxQuant. The incorporation efficiency can be calculated as  $(1 - 1/\text{Ratio (H/L)})$  on the peptide level. For proper SILAC experiments, it is required to obtain a labeling efficiency of more than 95%.



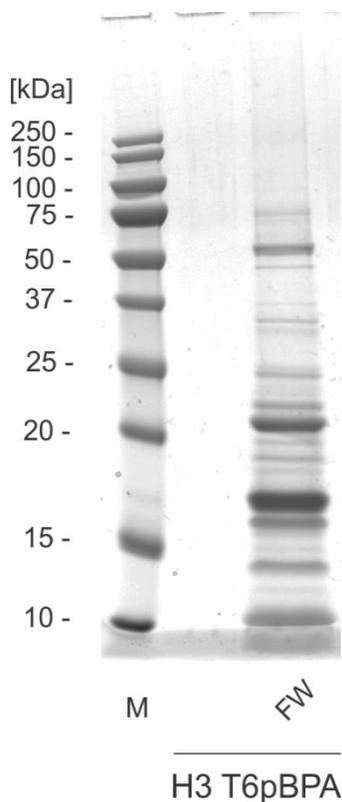
**Figure 4.12:** SILAC labeling efficient in BY4741 deletion strain.

Coomassie-stained gel of whole-cell lysate samples of H2A A61pBPA and H3 T6pBPA mutants expressed in BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells in SILAC SC media either supplemented with light (L) or heavy (H) L-arginine and L-lysine. All samples were subjected to MS analysis for evaluation of incorporation efficiency of labeled amino acids. *Legend:* L: light, unlabeled L-arginine and L-lysine; H: heavy, labeled  $^{13}\text{C}_6$ - $^{15}\text{N}_4$ -L-arginine and  $^{13}\text{C}_6$ -L-lysine. M: Marker.

All peptides identified from the whole-cell lysates (Figure 4.12) demonstrated a label incorporation of 97%, independent of the expressed histone amber mutant. These results indicate that proteins from *S. cerevisiae* BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells can be successfully labeled by heavy amino acids in culture.

### 4.3.2 Optimization of SILAC Procedure Successfully Produces Peptide Pairs

To test this approach, two cell cultures of transformed BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells co-expressing the pESC BPARS/1xSUP4-tRNA<sub>CUA</sub> plasmid and a pRS plasmid encoding an amber codon-mutated histone H3 (T6) were cultured in appropriate SILAC SC dropout media supplemented with 2 mM pBPA and either light or heavy 30 mg/L L-arginine and 60 mg/L L-lysine. Equal parts of heavy (here:  $^{13}\text{C}_6$ - $^{15}\text{N}_4$ -L-arginine and  $^{13}\text{C}_6$ -L-lysine), UV-treated cells were mixed with light, non-UV-treated cells based on OD, resulting in a forward sample. After affinity purification, the sample was concentrated to a final concentration of  $\sim$ 0.89 mg/mL and subjected to immunoprecipitation of H3-HA and its crosslink products. The acquired elution sample was loaded on a 10% Bis-Tris gel and stained with Instant Blue.



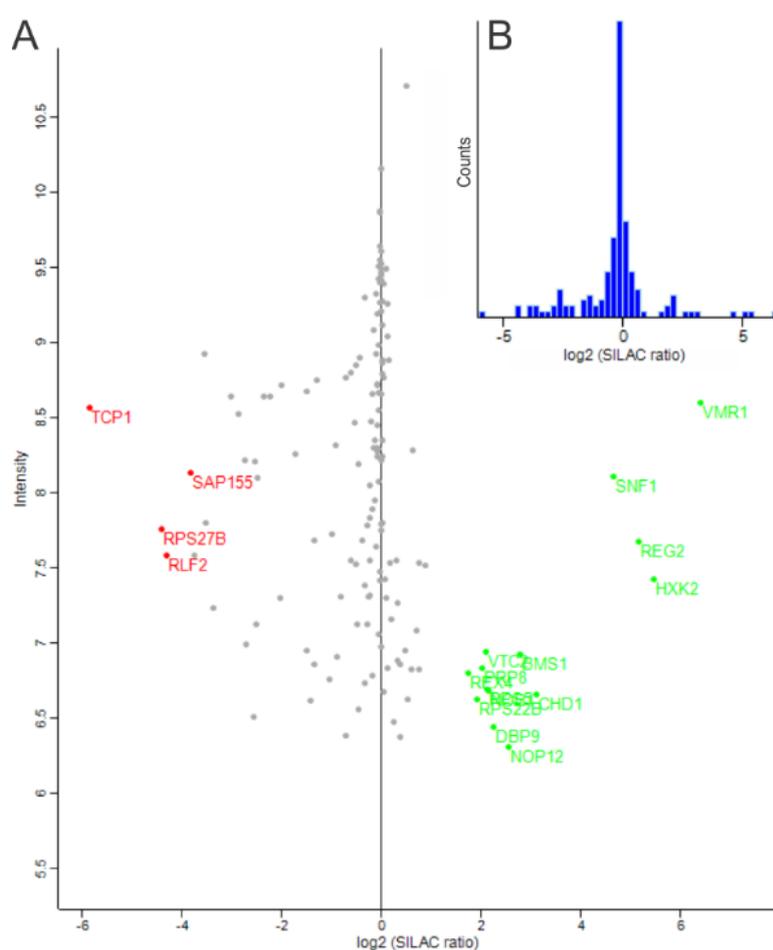
**Figure 4.13:** Optimization of SILAC procedure produces peptide pairs.

Coomassie-stained gel of IP elution fractions from forward samples of H3 T6pBPA mutants expressed in BY4741  $\Delta$ Arg4/ $\Delta$ Lys2. Full-length H3-HA ( $\sim$ 15 kDa) was enriched in the forward sample. *Legend:* M: Marker; FW: Forward sample.

The stained gel shows a successful saturation of the beads with full-length H3-HA protein of a molecular weight of approximately 15 kDa, and its crosslink products (Figure 4.13). Subsequently, the sample was digested and analyzed by MS. Quantification of SILAC pairs was expected to result in the formation of peptide pairs with mass shifts of 6, 10, 12, 16 or 20 Da.

MaxQuant was used to extract peak list files containing peptide masses and their corresponding ion masses and intensities from raw MS files and match these to sequences obtained from protein database Uniprot,

containing the proteome of *S. cerevisiae* wild-type strain 288C, the originating strain of the used BY4741 cells. The overall percentage of identified MS2 spectra was 6.55% with a 9.05% share of identified MS2 spectra of SILAC clusters. Matching of the peak files to protein sequences resulted in the identification of 250 proteins from 2026 peptides. Normalized SILAC ratios of all identified proteins were plotted against summed peptide intensities in Figure 4.14A, showing the heavily enriched proteins with gene names. The histogram of the  $\log_2$  SILAC ratios shows a normal distribution centred at zero, supporting similar overall protein loads of the light and heavy sample fractions (Figure 4.14B). Most of the proteins in both samples were relatively within a one-to-one ratio (H/L), which indicates a successful SILAC partner identification. By using a significance test with a Benjamini-Hochberg FDR of <5%, 12 proteins were found to be significantly upregulated, all possessing a relative quantity of more than factor 1.5 (H/L) (Table 4.1).



**Figure 4.14:** Identification of proteins from H3 T6pBPA SILAC analysis

A) The  $\log_2$  SILAC (i.e., heavy/light arginine and/or lysine) ratios calculated for each protein identified in the H3 T6pBPA immunoprecipitation experiment is plotted versus total peptide intensity. B) Distribution of normalized  $\log_2$  SILAC ratios.

*Legend:* in green: Proteins enriched in the heavy sample fraction; in red: Proteins enriched in light sample fraction.

Screening of the identified proteins revealed that a majority was involved in ribosomal biogenesis (BMS1, NOP12, DBP9, NOB1, RPS5, RPS22A/B, and REX4), metabolism (VMR1, HXK2) or spliceosomal assembly

(PRP8). Only two (REG2 and CHD1) were categorized into a chromatin biology-related background. REG2 has been identified as the regulatory subunit of protein phosphatase 1 (PP1), which is conserved across eukaryotic organisms. PP1 targets phosphorylation sites, including H3 S10, and thereby counteracts Aurora B kinase, signalling mitotic exit (Pinsky et al., 2009). CHD1 is an ATP-dependent chromatin remodelling factor which is involved in the recognition of H3 K4 methylation as substrate for the transcription-regulatory histone acetylation (HAT) complexes SAGA and SLIK. Both proteins, REG2 and CHD1, represent potentially trapped histone-protein interactions of great interest.

This SILAC experiment was performed in anticipation of identifying an admissible number of crosslinks. This expectation was not met, however, the high ratio between generally identified proteins and significantly enriched proteins indicates an increased level of background noise in the sample. Reducing the overall background noise may lead to identification of crosslinks which are treated as background signal under the current experimental setup. Also, since the anti-HA agarose beads were fully saturated with protein, the quantitative outcome of this SILAC experiment may either be caused by a general low abundance of crosslink products or by competitive displacement of the crosslink products from binding to the beads by an overabundance of full-length H4-HA protein. Therefore, procedures for depletion of free full-length histone-HA proteins should be investigated to increase the overall crosslink product concentration in the sample. Also, the used MaxQuant settings may have to be evaluated to prevent a cut-off too strict for acceptance.

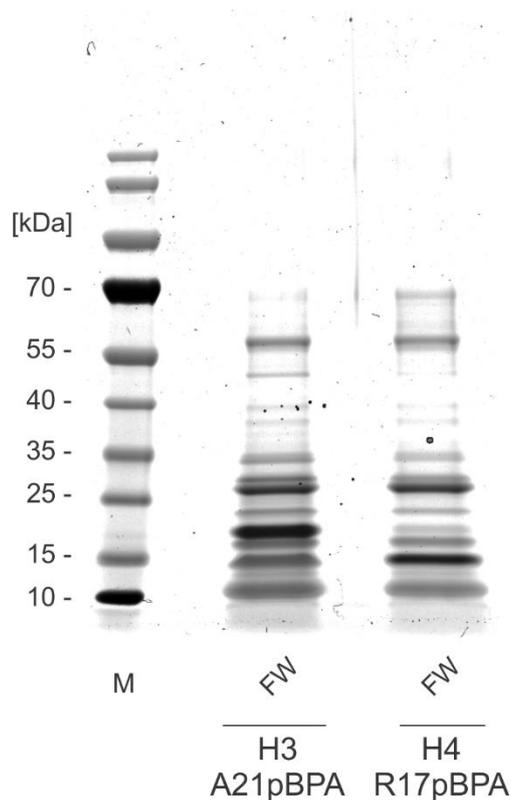
**Table 4.1:** Potential crosslink products of H3 T6 with a  $\log_2$  SILAC ratio greater than 1.5

Protein Names	Gene Names	Unique Peptides	SILAC ratio, normalized	Significance B (corr)
ABC transporter ATP-binding protein VMR1	VMR1	1	6.398	1.84 E-36
Hexokinase-2	HXK2	3	5.475	1.58 E-23
Glc7p type-1 protein phosphatase subunit REG2	REG2	1	5.159	1.49 E-19
Chromo domain-containing protein 1	CHD1	8	3.113	2.19 E+01
Ribosome biogenesis protein BMS1	BMS1	7	2.787	4.63 E+03
Nucleolar protein 12	NOPI2	3	2.557	1.44 E+05
ATP-dependent RNA helicase DBP9	DBP9	2	2.251	9.21 E+06
20S-pre-rRNA D-site endonuclease NOB1	NOB1	3	2.158	2.95 E+07
40S ribosomal protein S5	RPS5	4	2.120	4.69 E+07
Pre-mRNA-splicing factor 8	PRP8	1	2.013	1.65 E+08
40S ribosomal protein S22-B	RPS22B	3	1.910	5.30 E+08
RNA exonuclease 4	REX4	2	1.737	3.28 E+09

### 4.3.3 Low Identification Yield Independent of Histone Variant or Crosslinker Positioning

In order to determine if histone choice or pBPA positioning has an influence on the number of identified crosslink products, SILAC experiments were conducted with cells either expressing a histone H3 A21pBPA mutant or a histone H4 R17pBPA mutant. Both samples were prepared as forward samples, meaning only the heavy labeled cell portion of each sample was treated with UV-light, the light portion remained untreated. Before immunoprecipitation of the HA-epitope-tagged histones and their crosslink products, both samples were concentrated to final concentrations of 0.84 mg/mL (H3 A21) and 0.71 mg/mL (H4 R17). The acquired elution samples were loaded on a 10% Bis-Tris gel and stained with Instant Blue. The stained gel showed a

successful saturation of the beads in both lanes with full-length H3-HA protein or H4-HA protein, respectively, and their specific crosslink products (Figure 4.15).



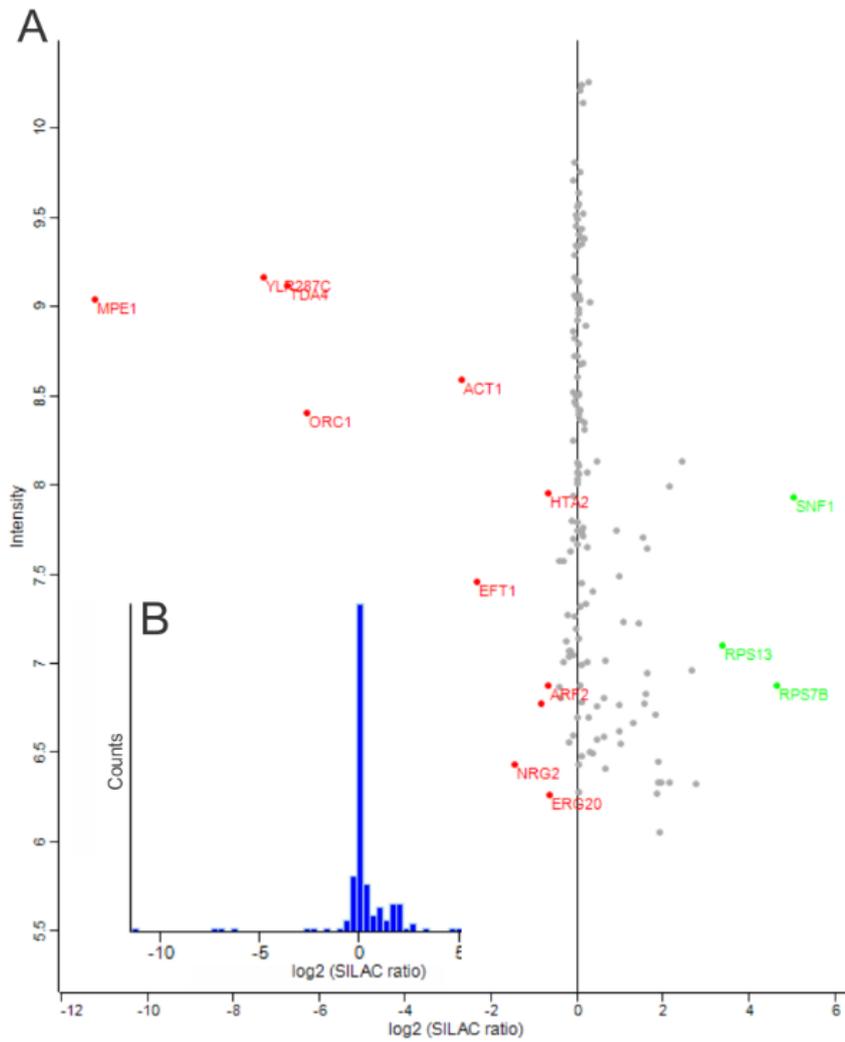
**Figure 4.15:** SILAC immunoprecipitation of H3 A21pBPA and H4 R17pBPA crosslink products.

Coomassie-stained gel of IP elution fractions from forward samples of H3 A21pBPA and H4 R17pBPA mutants expressed in BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells. Full-length H3-HA (~15 kDa) and H4-HA (~13 kDa) were enriched in the respective samples. *Legend:* M: Marker; FW: Forward sample.

The overall percentage of identified MS2 spectra for H3 A21 was 5.82% with a 9.95% share of identified MS2 spectra of SILAC clusters. For H4 R17, the overall percentage of identified MS2 spectra was 6.47% with an 8.15% share of identified MS2 spectra of SILAC clusters.

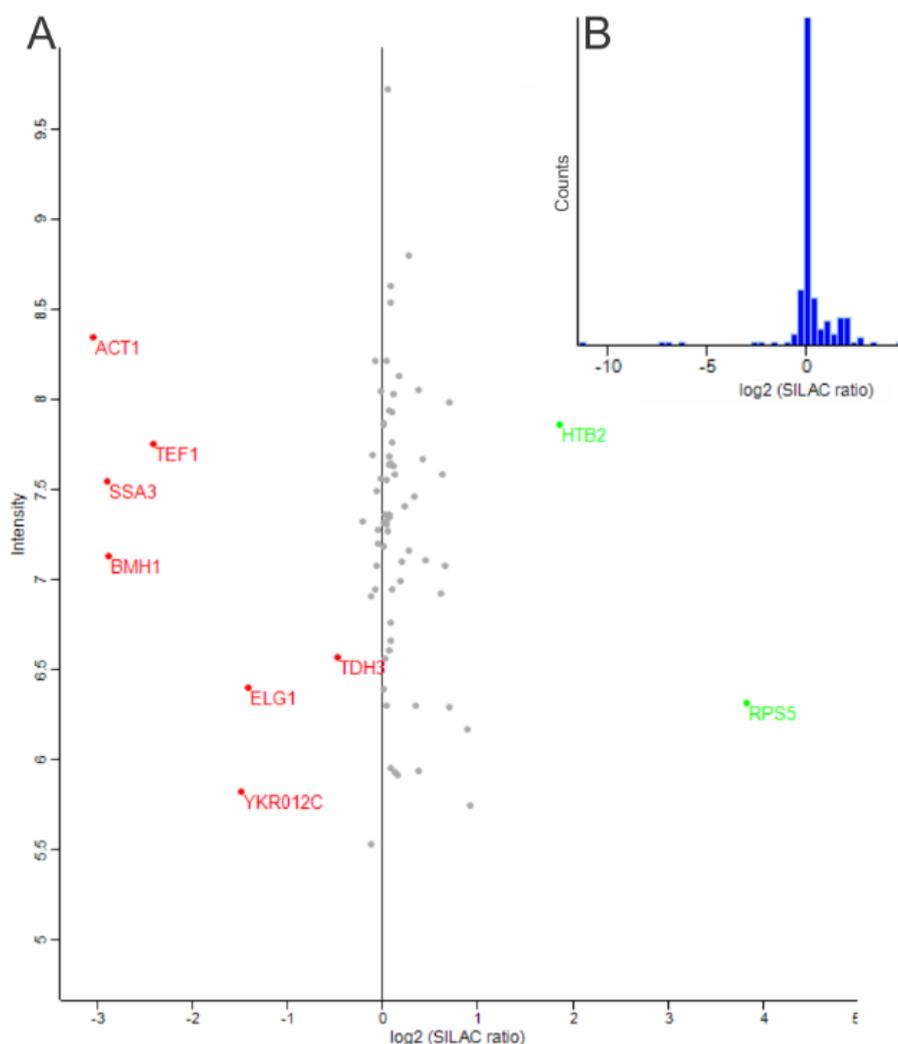
Analysis of the raw MS files revealed the detection of 2191 peptides in the H3 A21 sample which allowed the identification of 223 proteins, 3 of which were significantly enriched in the heavy sample (Figure 4.16A). The H4 R17 sample produced 1432 detectable peptides, leading to the identification of 114 proteins (Figure 4.17A). Only 2 proteins were categorized as significantly elevated in the heavy portion of the H4 R17 sample.

The histograms of the  $\log_2$  SILAC ratios of both analyses showed a normal distribution centred at zero, supporting similar overall protein loads of the light and heavy sample fractions (Figure 4.16B; Figure 4.17B).



**Figure 4.16:** Identification of proteins from H3 A21pBPA SILAC analysis.

A) The log<sub>2</sub> SILAC (i.e., heavy/light arginine and/or lysine) ratios calculated for each protein identified in the H3 A21pBPA immunoprecipitation experiment are plotted versus total peptide intensity. B) Distribution of normalized log<sub>2</sub> SILAC ratios. *Legend:* in green: Proteins enriched in the heavy sample fraction; in red: Proteins enriched in light sample fraction.



**Figure 4.17:** Identification of proteins from H4 R17pBPA SILAC analysis.

A) The log<sub>2</sub> SILAC (i.e., heavy/light arginine and/or lysine) ratios calculated for each protein identified in the H4 R17pBPA immunoprecipitation experiment are plotted versus total peptide intensity. B) Distribution of normalized log<sub>2</sub> SILAC ratios. *Legend:* in green: Proteins enriched in the heavy sample fraction; in red: Proteins enriched in light sample fraction.

As already observed in the prior SILAC experiment, most of the proteins in both samples were relatively within a one-to-one ratio (H/L), which indicates a successful SILAC partner identification. Still, the overall number of identified proteins in both samples as well as the ratio of significantly enriched proteins to overall identified proteins remained to be lower than expected, independent of the histone variant or crosslinker positioning.

Under the assumption that the identical treatment and preparation of the cells produced the same amount of background noise, the greater overall yield of identified proteins in this and the earlier analyzed histone H3 sample compared to the histone H4 sample allows the assumption that there are histone variant-dependent differences in occupancy with binding proteins. However, the number of proteins marked as significantly increased remained lower than anticipated in both samples. From 3 proteins in the H3 A21 sample, the majority of 2 was categorized into ribosome biogenesis (RPS7B and RPS13) (Table 4.2). In the H4 R17 sample, half of

the proteins were involved in ribosome biogenesis (RPS5) (Table 4.3). In eukaryotic cells, ribosome biogenesis takes place in the cytoplasm and in the nucleolus. The packaging of ribosomal DNA (rDNA) in the nucleolus is highly regulated by chromatin structure. While subcellular fractionation could show the exclusive formation of crosslink products within the chromatin fraction of the cells (Figure 1.11), it remains to be deduced if the overabundance of this class of proteins is random by nucleolar overcrowding, unspecific binding to the beads, or represents specific trapped interactions between chromatin components and potential ribosomal substructures in the nucleolus. The remaining identified protein supposedly interacting with H3 A21 was the protein kinase SNF1 that is involved in responses to environmental stress (Hedbacker and Carlson, 2008). More interestingly, analysis of the MS data from H4 R17 revealed histone H2B as significantly enriched protein in the heavy, UV-treated portion of the sample. The crystal structure of the nucleosome core particle shows an interaction between residues 16 to 25 of the highly basic histone H4 tail and acidic patch on the exposed surface of the H2A-H2B dimer (Luger et al., 1997). Therefore, this finding may constitute the first trapping and quantitative detection of known interactions between histones by combination of *in vivo* crosslinking and SILAC in yeast. Nevertheless, the overall number of significant hits remains to be improved and novel protein interaction partners need to be verified.

**Table 4.2:** Potential crosslink products of H3 A21 with a  $\log_2$  SILAC ratio greater than 1.5

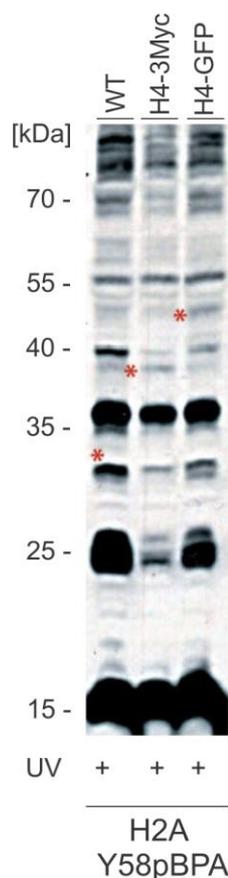
Protein Names	Gene Names	Unique Peptides	SILAC ratio, normalized	Significance B (corr)
Carbon catabolite-derepressing protein kinase	SNF1	1	5.043	1.10 E-07
40S ribosomal protein S7-B	RPS7B	3	4.641	1.06 E-06
40S ribosomal protein S13	RPS13	4	3.383	0.00039726

**Table 4.3:** Potential crosslink products of H4 R17 with a  $\log_2$  SILAC ratio greater than 1.5

Protein Names	Gene Names	Unique Peptides	SILAC ratio, normalized	Significance B (corr)
40S ribosomal protein S5	RPS5	5	3.819	5.79 E-35
Histone H2B	HTB1	2	1.860	3.75 E-09

#### 4.3.4 Verification of Histone Interaction Partners by Molecular Shift Assay

The verification of histone interaction partners which were identified by SILAC experiments can be performed by a molecular weight shift approach. It works by duplication of the crosslink experiment in a yeast strain where the presumed interaction partner is genomically tagged, which results in a detectable shift of bands in a western blot if the interaction can be trapped. As an example, Figure 4.18 shows the successful identification of an H2A-H4 crosslink by this approach (Rall, 2012; Wilkins et al, 2014).



**Figure 4.18:** Identification of H2A-H4 crosslink product by genomic tagging.

Western blot analysis of whole-cell lysates from cells with a genomically tagged copy of histone H4 and expressing H2A Y58pBPA. Both tagged strains showed an additional crosslink band of higher molecular weight and a reduced H2A-H4 crosslink band. Modified from Rall, 2012 and Wilkins et al., 2014.

*Legend:* \*: Crosslink product of corresponding size to H2A-H4 interaction.

By taking advantage of a yeast GFP library (Thermo Fisher), verification of potential interaction partners obtained from the previously performed SILAC experiment on histone H3 with an amber mutation on position T6 was performed (4.3.2). Cells harbouring genomically tagged copies of BMS1, CHD1, NOB1, NOP12 and REG2 were transformed with plasmids for expression of a histone H3 T6pBPA mutant, crosslinked and analyzed against a UV-treated BY4741 H3 T6pBPA control by Western blot. Dependent on the used GFP-tagged strain, band shifts were expected at 135 kDa for BMS1, at 170 kDa for CHD1, at 52 kDa for NOB1 and NOP12 and at 39 kDa for REG2.

No mass shifts at the expected heights of putative interaction partners were observed for any of the five tested GFP-tagged yeast strains. Additionally, no crosslink products of these specific molecular weights were observed in the BY4741 reference sample. The reader is referred to the bachelor's thesis of Tim Garbers for a more detailed analysis (2016).

Therefore, it remained questionable if the identified proteins represent actual trapped interaction partners of histone H3 at position T6.

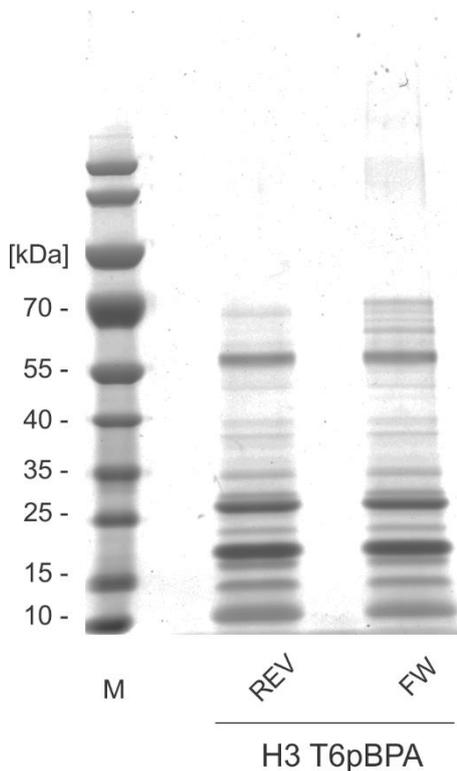
#### 4.3.5 Bidirectionally Labeled Samples for Verification of Histone Interaction Partners

Bidirectional labeling of biological replicates is a popular approach for the verification of potential hits in MS-based SILAC experiments. While the experimental setup of the preceding SILAC experiments followed a forward approach by labeling the UV-treated portion of the sample for screening, a reverse sample is obtained from crosslinking the unlabeled portion of a sample and mixing it with the non-UV-treated labeled portion. A

genuine crosslink product is expected to be detected in the UV-treated portions of both forward and reverse samples.

Analysis of global expression levels and localization of proteins in *S. cerevisiae* together were able to detect more than 4500 proteins, indicating that at least 80% of the yeast genome is expressed in logarithmically growing cells. Expression levels follow a bell-shaped curve with an average expression of about 3000 proteins (Ghaemmaghami et al., 2003; Huh et al., 2003). The resulting dynamic range of protein expression may cause differences in the formation of crosslinks between samples, which can be limited to one of two samples. These crosslink products can be detected by a combined forward and reverse labeling approach. Additionally, effects caused by differences in culturing can be determined, since cells grown in the presence of heavy amino acids have shown an increased stalling of growth at lower optical densities.

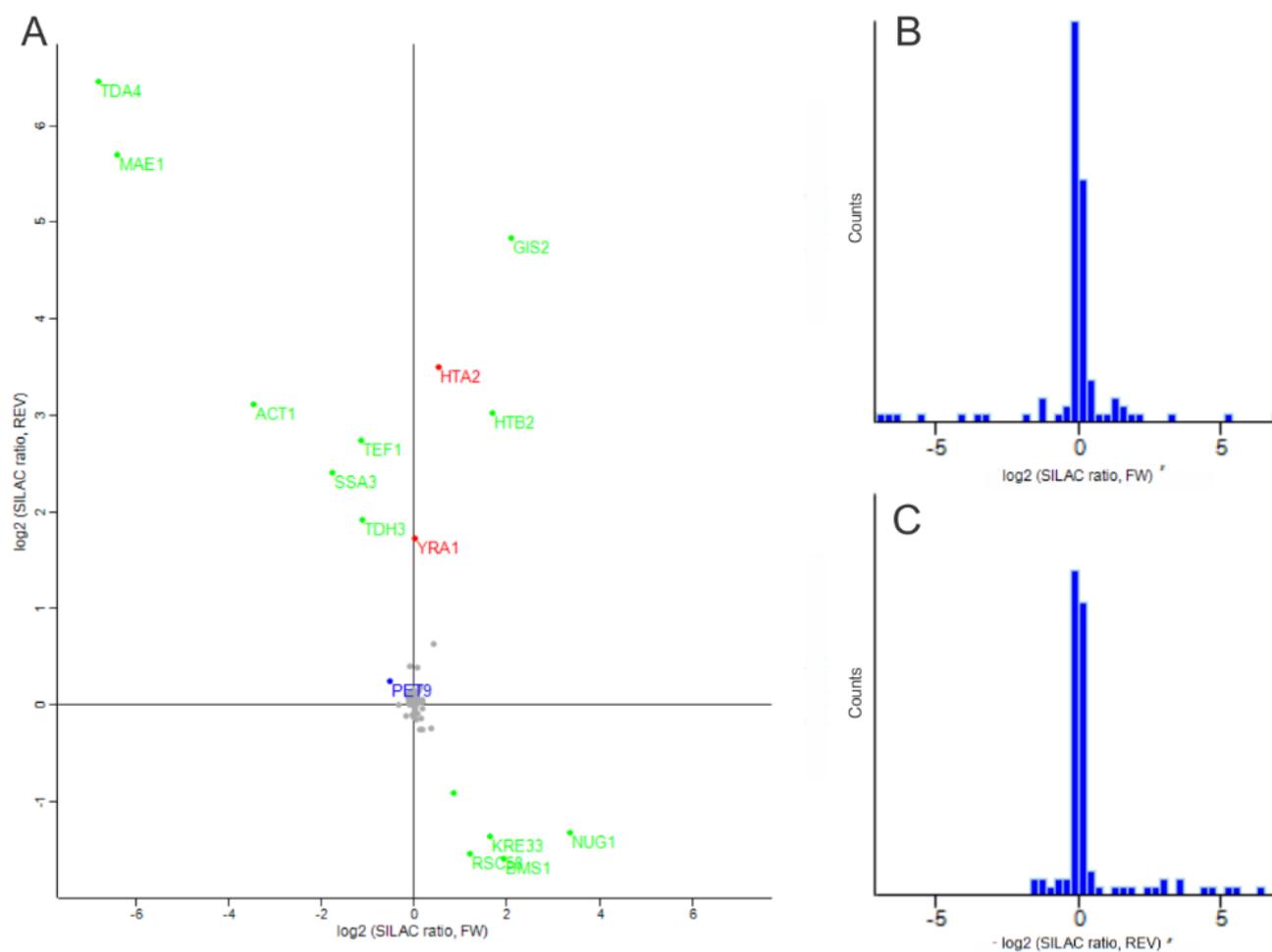
In order to investigate the above-mentioned points, a forward and reverse SILAC experiment of H3 T6pBPA was performed. Two times two cell cultures of transformed BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells carrying plasmids for the expression of the BPARS/1xSUP4-tRNA<sub>CUA</sub> pair and an amber codon-mutated histone H3 (T6) were cultured in appropriate SILAC SC dropout media supplemented with 2 mM pBPA and either light or heavy 30 mg/L L-arginine and 60 mg/L L-lysine. Before immunoprecipitation of H3-HA and its crosslink products, the forward sample was concentrated to a final concentration of 0.81 mg/mL and the reverse sample to a final concentration of 0.85 mg/mL. The acquired elution samples were loaded on a 4-12% Bis-Tris gel and stained with Instant Blue. Gel analysis showed an even saturation of the beads with H3-HA protein as well as banding patterns in both samples (Figure 4.19). Subsequently, the samples were digested and analyzed by mass spectrometry.



**Figure 4.19:** SILAC immunoprecipitation of H3 T6pBPA crosslink products.

Coomassie-stained gel of IP elution fractions from forward and reverse samples of H3 T6pBPA mutants expressed in BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells. Full-length H3-HA (~15 kDa) was enriched in both samples.

*Legend:* M: Marker; REV: Reverse sample; FW: Forward sample.



**Figure 4.20:** Identification of proteins from H3 T6pBPA SILAC analysis.

A) The log<sub>2</sub> SILAC (i.e., heavy/light L-arginine and/or L-lysine) ratios calculated for each protein identified in the forward and reverse experiments of H3 T6pBPA immunoprecipitation are plotted against each other. B) Distribution of normalized log<sub>2</sub> SILAC ratios from forward sample. C) Distribution of normalized log<sub>2</sub> SILAC ratios from reverse sample. *Legend:* in green: Proteins enriched in both samples; in blue: Proteins enriched in forward sample; in red: Proteins enriched in reverse sample.

The overall percentage of identified MS2 spectra was 5.53% with a 6.93% share of identified MS2 spectra of SILAC cluster. Matching of the peak files to proteins resulted in the identification of 173 proteins from 1870 peptides. Normalized SILAC ratios of all identified proteins from the forward sample are plotted against normalized SILAC ratios of all identified proteins from the reverse sample in Figure 4.20A, showing the heavily enriched proteins with gene names. The histograms of the log<sub>2</sub> SILAC ratios show a normal distribution centred at zero for both samples, supporting similar overall protein loads of the light and heavy sample fractions (Figure 4.20B, C). Only two proteins are significantly increased in the UV-treated portion of both samples: histone H2B and the Zinc finger protein GIS2 (Table 4.4).

**Table 4.4:** Potential crosslink products in forward and reverse sample of H3 T6pBPA.

Protein Names	Gene Names	Unique Peptides FW/REV		SILAC ratio, normalized FW/REV		Significance B (corr) FW/REV	
Histone H2B	HTB2; HTB1	4	3	1.702	3.026	4.56 E-09	2.31 E-17
Zinc finger protein GIS2	GIS2	1	6	2.105	4.840	3.77 E-13	2.93 E-42

Notably, comparison of the SILAC ratios between the forward and reverse samples revealed a doubling of the  $\log_2$  values for both proteins, which represents a 4-fold increase in abundance. Trapping of the interaction with histone H2B may show the successful incorporation of the histone H3 amber mutant into the chromatin landscape. GIS2 is involved in environmental stress response and acts as a translational activator for mRNAs with internal ribosome entry sites.

The performed significance B test also revealed some identified proteins to be restricted exclusively to the UV-treated portion of one sample. The proteins NOP4, TY4B, SDA1 and NOP12 could only be identified in the UV-treated portion of the forward sample (Table 4.5).

**Table 4.5:** Potential crosslink products in forward sample of H3 T6pBPA.

Protein Names	Gene Names	Unique Peptides, FW/REV		SILAC ratio, normalized, FW/REV		Significance B (corr), FW/REV	
Nucleolar protein 4	NOP4	3	-	1.102	-	0.000160155	1
Reverse transcriptase/ribonuclease H	TY4B-H	2	1	5.443	-	1.81 E-79	1
Protein SDA1	SDA1	4	-	1.387	-	1.87 E-06	1
Nucleolar complex protein 14	NOP14	2	-	1.365	-	2.75 E-06	1

Opposing these observations, the proteins ADH1, histone H2A, RPP0, ORC1 and YRA1 were exclusively present in the crosslinked fraction of the reverse sample (Table 4.6). The proteins NOP4 and NOP12 are both nucleolar proteins involved in the biogenesis of ribosomal subunits. TY4B-H is a reverse transcriptase/ribonuclease and is involved in the transposition of native Ty4 transposons in *S. cerevisiae*. SDA1 is a nucleolar protein required for passage through G1 (Zimmerman and Kellogg, 2001). Found in the UV-treated portion of the reverse sample, histone H2A could indicate the successful incorporation of the expressed amber mutant into nucleosomes. With ORC1, the large subunit of the origin recognition complex (ORC) which is involved in chromosomal replication and transcription silencing, a potentially interesting interaction partner of histone H3 at position T6 was detected. The remaining proteins are involved in cell metabolism (ADH1), translation (RPP0) and nuclear export of mRNAs (YRA1).

**Table 4.6:** Potential crosslink products in reverse sample of H3 T6pBPA.

Protein Names	Gene Names	Unique Peptides,		SILAC ratio, normalized,		Significance B (corr),	
		FW/REV		FW/REV		FW/REV	
Alcohol dehydrogenase 1	ADH1	-	6	-	4.519	1	5.14 E-37
Histone H2A	HTA1	2	2	0.534	3.503	0.072	8.25 E-23
60S acidic ribosomal protein P0	RPP0	-	3	-	3.497	1	9.81 E-23
Origin recognition complex subunit 1	ORC1	1	1	-	5.285	1	4.16 E-50
RNA annealing protein YRA1	YRA1	2	3	-	1.725	0.974	1.69 E-06

The obtained data can also be used to identify media-specific background noise which can only be detected in either the unlabeled or labeled portion of the samples, independent of UV-treatment. Screening of the identified proteins revealed the proteins TDA4, MAE1, ACT1, SSA3, TEF1 and TDH3 to be exclusively present and enriched in the lysate of cells grown in the absence of heavy amino acids (Table 4.7). MAE1 and TDH3 are involved in cell metabolism, TEF1 in translation, ACT1 is responsible for shaping of the cell and SSA3 belongs to the heat shock protein 70 (HSP70) family, involved in protein folding and stress response.

**Table 4.7:** Proteins abundant in samples of cells grown in light media.

Protein Names	Gene Names	Unique Peptides,		SILAC ratio, normalized,		Significance B (corr),	
		FW/REV		FW/REV		FW/REV	
Topoisomerase I damage-affected protein 4	TDA4	1	1	-6.792	6.455	-	4.96 E-74
NAD-dependent malic enzyme	MAE1	1	1	-6.401	5.694	-	6.18 E-58
Actin	ACT1	9	6	-3.455	3.113	3.10 E-108	2.69 E-18
Heat shock protein SSA3	SSA3	1	-	-1.750	2.407	2.19 E-29	1.78 E-11
Elongation factor 1-alpha	TEF1	4	5	-1.130	2.739	2.74 E-13	1.82 E-14
Glyceraldehyde-3-phosphate dehydrogenase 3	TDH3	2	5	-1.121	1.913	4.11 E-13	1.03 E-07

In the presence of heavy amino acids, cells showed a restrictive enrichment of the proteins YCR087C-A, RSC58, KRE33, BMS1 and NUG1 (Table 4.8). The most interesting candidate is RSC58 which is part of the RSC chromatin remodelling complex. Its exclusive abundance in cells grown in the presence of heavy amino acids raises concern about a possible effect of labeling on chromatin structure. The remaining proteins KRE33, BMS1 and NUG1 are involved in ribosome biogenesis. YCR087C-A lacks detailed functional information, but is known to be located in the nucleolus.

**Table 4.8:** Proteins abundant in samples of cells grown in heavy media.

Protein Names	Gene Names	Unique Peptides,		SILAC ratio, normalized,		Significance B (corr),	
		FW/REV		FW/REV		FW/REV	
UPF0743 protein YCR087C-A	YCR087C-A	5	8	0.865	-0.916	0.003173368	0.34 E-11
Chromatin structure-remodeling complex protein RSC58	RSC58	5	4	1.206	-1.544	3.52 E-05	1.35 E-27
UPF0202 protein KRE33	KRE33	5	3	1.646	-1.364	1.45 E-08	5.16 E-22
Ribosome biogenesis protein BMS1	BMS1	3	2	1.938	-1.589	2.37 E-11	4.27 E-29
Nuclear GTP-binding protein NUG1	NUG1	2	2	3.376	-1.318	1.51 E-31	1.09 E-20

Besides the rediscovery of BMS1, no other protein of significance from the previous SILAC experiment of H3 T6 was detected (4.3.2). Given the above-mentioned observations, the detection of BMS1 in the first experiment may also have been caused by metabolic labeling of the cells with heavy amino acids. This and the non-detection of the other proteins may be evidence for the falseness of these hits and would explain the failed efforts of verification by mass shift assay (4.3.4). It remains to be elucidated if the failed detection of these proteins is the result of dynamic expression levels or sample preparation and acquisition.

Although the experiment failed to reproduce the outcome of the previously performed SILAC experiment on H3 T6, it helped to provide knowledge about the interpretation of potential hits and sort them into different categories as well as to reveal an accumulation of specific proteins as a possible effect of labeling.

### 4.3.6 Quantity of Protein Identifications Independent of MaxQuant Setup

MaxQuant is a quantitative proteomics software which provides an algorithmic pipeline for the analysis of large MS data sets. It provides a means for feature detection and peptide quantification, MS/MS ion search based on the protein database search engine Mascot, identification and validation of hits, and visualization of the output.

The preset parameter values have been determined by the developers Cox and Mann (2008) by analysis of a plethora of data sets of very different types and monitoring of the performance of the algorithms by a multitude of users. The algorithms and basic principles performed by this software should be applicable to any kind of high-mass-resolution data.

In the following, it was investigated if modification MaxQuant control parameters would result in altered outcomes of identification from the same SILAC sample. The priorly analyzed H3 T6 forward and reverse samples will serve as exemplary samples. The changes to the parameters can be seen in Table 4.9. All chosen parameters were tested for their ability to improve peptide identification rates in order to determine optimal parameter values for crosslink product detection with the described experimental setup.

**Table 4.9:** Overview of applied changes to MaxQuant control parameters.

Parameter	Function	Presetting	Setup A	Setup B	Setup C
<b>Max. labeled AAs</b>	Maximum number of labeled amino acids per peptide	3		5	8
<b>Max. missed cleavages</b>	Maximum number of tolerated missed cleavages	2		2	3
<b>First search peptide tolerance</b>	Mass tolerance for peptides in first search round used for mass- and time-dependent recalibration of peptide masses	20		30	30
<b>Main search peptide tolerance</b>	Peptide mass tolerance during main search	4.5		9	9
<b>Max. peptide mass [Da]</b>	Mass threshold for peptides	4600	9000	9000	9000
<b>Max. peptide length for unspecific search</b>	Maximum peptide length for unspecific search	25	9		
<b>Min. peptide length for unspecific search</b>	Minimum peptide length for unspecific search	8		7	7
<b>PSM FDR</b>	False discovery rate control of identifications on peptide level	0.01	0.02		
<b>Protein FDR</b>	False discovery rate control of identifications on protein level	0.01	0.02		
<b>Site decoy fraction</b>	False discovery rate for assessment of the likelihood of false positive identifications	0.01	0.02		

**Table 4.10:** Color code for Table 4.11

Log <sub>2</sub> (x) ratio H/L normalized		
Forward sample	Reverse sample	Prediction
≥1	≤-1	True positive*
≥1	NaN	May be positive
NaN	≤-1	May be positive
≥1	≥0	False positive*
≤0	≤-1	False negative*
NaN	≥1	May be negative
≤-1	NaN	May be negative
≤-1	≥1	True negative*
=	=	Background

**Table 4.11:** Overview of changes in protein identification produced by variation of MaxQuant parameters.

Changed parameters are described in Table 4.9. Color coding is described in Table 4.10.

Control	Setup A	Setup B	Setup C
RSC58 (1.21/-1.54)	RSC58 (1.21/-1.54)	KRE33 (1.65/-1.76)	RSC58 (1.1/-0.52)
KRE33 (1.65/-1.36)	KRE33 (1.65/-1.36)	BMS1 (1.95/-1.61)	KRE33 (1.94/-1.68)
BMS1 (1.94/-1.59)	BMS1 (1.94/-1.59)	<u>NOP4 (1.09/NaN)</u>	BMS1 (1.82/-1.61)
NUG1 (3.38/-1.32)	NUG1 (3.38/-1.32)	<u>NOP14 (1.29/NaN)</u>	NUG1 (3.31/-1.28)
<u>NOP4 (1.1/NaN)</u>	<u>NOP4 (2.06/NaN)</u>	<u>SDA1 (1.31/NaN)</u>	<u>NOP14 (1.18/NaN)</u>
<u>NOP14 (1.37/NaN)</u>	<u>NOP14 (1.37/NaN)</u>	HTA2; HTA1 (1.92/4.85)	<u>SDA1 (1.2/NaN)</u>
<u>SDA1 (1.39/NaN)</u>	<u>SDA1 (1.39/NaN)</u>	HTB2; HTB1 (1.71/1.82)	<u>CNM67 (1.18/NaN)</u>
<u>TY4B-J; TY4B-H (5.44/NaN)</u>	<u>CUE2 (4.6/NaN)</u>	TH2 (-1.18/-0.32)	HTA2; HTA1 (0.23/4.14)
YRA1 (0.03/1.73)	<u>TY4B-H; TY4B-J (7.23/NaN)</u>	<u>GIS2 (NaN/4.87)</u>	HTB2; HTB1 (0.47/1.81)
HTA2; HTA1 (0.53/3.5)	HTA2; HTA1 (0.53/3.5)	<u>CDC19 (NaN/1.36)</u>	YRA1 (0.09/1.67)
HTB2; HTB1 (1.7/3.03)	HTB2; HTB1 (0.47/3.03)	<u>RPP0 (NaN/3.48)</u>	<u>GIS2 (NaN/4.8)</u>
GIS2 (2.11/4.84)	GIS2 (2.11/4.84)	<u>YLR287C (NaN/3.21)</u>	<u>CDC19 (NaN/1.36)</u>
<u>CDC19 (NaN/1.34)</u>	<u>YRA1 (NaN/1.73)</u>	<u>ADH1 (NaN/4.56)</u>	<u>RPP0 (NaN/3.47)</u>
<u>RPP0 (NaN/3.5)</u>	<u>CDC19 (NaN/1.34)</u>	<u>BMH1; BMH2 (-4.1/NaN)</u>	<u>ADH1 (NaN/4.56)</u>
<u>ADH1; ADH2 (NaN/4.52)</u>	<u>RPP0 (NaN/3.5)</u>	<u>ELG1 (-1.39/NaN)</u>	<u>SSA4 (-4.14/NaN)</u>
<u>ORC1 (NaN/5.29)</u>	<u>ADH1; ADH2 (NaN/4.52)</u>	<u>SSA3 (-1.5/NaN)</u>	<u>ELG1 (-1.36/NaN)</u>
<u>YLR287C (-6.96/NaN)</u>	<u>ORC1 (NaN/5.29)</u>	<u>MAE1 (-6.39/5.39)</u>	<u>MAE1 (-6.38/NaN)</u>
<u>TCP1 (-5.58/NaN)</u>	<u>YLR287C (-6.96/NaN)</u>	<u>ACT1 (-3.45/2.79)</u>	<u>BMH1; BMH2 (-4.07/2.16)</u>
<u>BMH1; BMH2 (-4.08/NaN)</u>	<u>TCP1 (-5.58/NaN)</u>	<u>TEF1 (-1.27/2.75)</u>	<u>ACT1 (-3.88/3.01)</u>
<u>SSA4 (-3.22/NaN)</u>	<u>BMH1; BMH2 (-4.08/NaN)</u>	<u>TDH3 (-1.13/1.92)</u>	<u>SSA3 (-1.68/2.84)</u>
<u>ELG1 (-1.19/NaN)</u>	<u>SSA4 (-3.22/NaN)</u>	RSC58 (/)	<u>TEF1 (-1.47/2.68)</u>
<u>TDA4 (-6.79/6.46)</u>	<u>ELG1 (-1.19/NaN)</u>	NUG1 (/)	<u>TDH3 (-1.12/1.94)</u>
<u>MAE1 (-6.4/5.69)</u>	<u>RSM22 (-1.09/NaN)</u>	CUE2 (/)	<u>RPS0B; RPS0A (-0.16/1.35)</u>
<u>ACT1 (-3.46/3.11)</u>	<u>TDA4 (-7.33/6.46)</u>	TY4B-H; TY4B-J (/)	NOP4 (/)
<u>SSA3 (-1.75/2.41)</u>	<u>MAE1 (-6.4/5.69)</u>	YRA1 (/)	CUE2 (/)
<u>TEF1 (-1.13/2.74)</u>	<u>ACT1 (-3.46/3.11)</u>	ADH1; ADH2 (/)	TY4B-H; TY4B-J (/)
<u>TDH3 (-1.12/1.91)</u>	<u>SSA3 (-1.75/2.84)</u>	ORC1 (/)	ADH1; ADH2 (/)
CUE2 (/)	<u>TEF1 (-1.13/2.74)</u>	TCP1 (/)	ORC1 (/)
TY4B-H; TY4B-J (/)	<u>TDH3 (-1.12/1.91)</u>	SSA4 (/)	YLR287C (/)
RSM22 (/)	TY4B-J; TY4B-H (/)	RSM22 (/)	TCP1 (/)
THI2 (/)	THI2 (/)	TDA4 (/)	RSM22 (/)
ADH1 (/)	ADH1 (/)	TY4B-J; TY4B-H (/)	TDA4 (/)
CNM67 (/)	CNM67 (/)	CNM67 (/)	TY4B-J; TY4B-H (/)
RPS0B; RPS0A (/)	RPS0B; RPS0A (/)	RPS0B; RPS0A (/)	THI2 (/)

The effects of modifying the run parameters of MaxQuant are shown in Table 4.11. Statistical evaluation showed only minor differences in the overall outcome between the different setups. There were setting-dependent deviations observable in the detection of single proteins, but the group of highly significant proteins changed only marginally. Although there were some variations (e.g., YRA1, GIS2, MAE1, CNM67, NUG1), the majority of proteins exhibited no changes in probability across the tested setups (e.g., KR33, BMS1, NOP14, CDC19, ACT1, ADH1). It needs to be pointed out that application of setup B led to the marking of RSC58 and NUG1, both true positives in the control and setup A and C, as insignificant. This might be evidence for this setup being too strict in cut-off.

Overall comparison of the detected proteins showed variation of different magnitudes in their normalized H/L ratios across setups. For example, histone H2B can be detected in all setups. In the control setup, it had a ratio of 1.7 in the forward sample and 3.03 in the reverse sample. Setup A resulted in a decrease of the forward ratio to 0.47, but remained steady for the reverse sample. Compared to the control, setup B showed only a decrease by approximately half for the ratio of the reverse sample. Setup C caused a severe decrease of the forward ratio to 0.47 as well as to 1.81 for the reverse sample. The range of these values shows that the SILAC ratios obtained from the same data set may vary by a factor of up to 4 depending on the chosen parameters.

Therefore, the adjustment of the parameter values may have a decisive effect on the statistical significance of proteins with borderline values. Nevertheless, the performed modification of the run parameters of MaxQuant did not result in the identification of new crosslink products with the present experimental setup.

Consequently, for the following experiments, it was refrained from changing the preset parameters of MaxQuant.

#### **4.3.7 Enrichment of Crosslink Products by Acid Extraction or Hydrazide Reactivity**

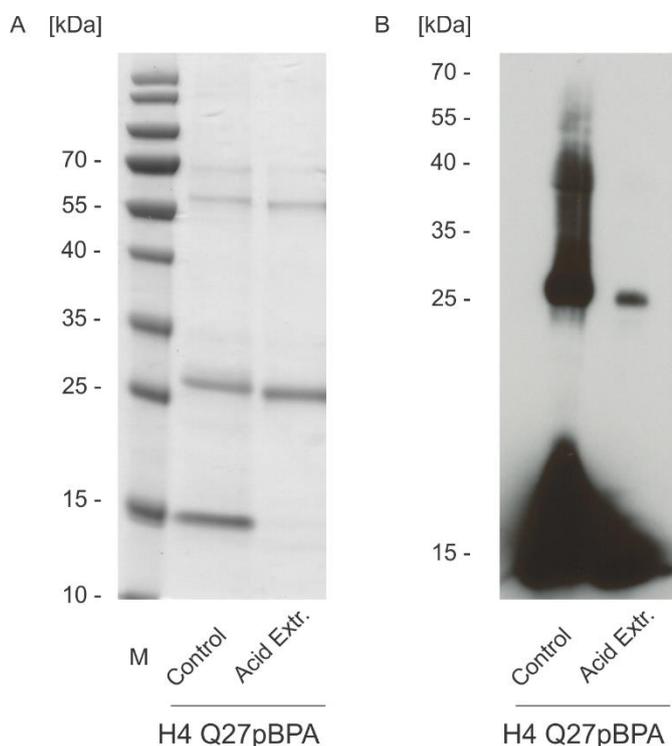
The abundance of crosslink products has asserted itself to be a possible limiting factor in the acquisition of reliable MS data. Therefore, different approaches have been pursued in order to produce concentrated crosslink samples.

The general obstacle these approaches aimed to overcome was the displacement of HA-tagged crosslink products in the course of immunoprecipitation by an overabundance of free, non-crosslinked HA-tagged histones. Therefore, a reduction of free histone-HA proteins should result in an improved ratio towards crosslink products.

Acid extraction is known to be a robust technique for the isolation and study of histone proteins from a multitude of organisms. Most of the histone extraction work has been based on solubility of histones in dilute acid conditions which favour the precipitation of most other nuclear proteins and nucleic acids (Shechter et al., 2007). Therefore, crosslink products were expected to precipitate under acidic conditions due to their crosslinked non-histone protein attachment, allowing the removal of free histone-HA protein in the acid fraction and isolation of crosslink products from the pellet.

To test this approach, an acid extraction step with 0.4 N H<sub>2</sub>SO<sub>4</sub> was added to the standard immunoprecipitation protocol (3.2.2.11). After mechanic disruption of cells expressing an H4 Q27pBPA mutant and washing of the cell lysate with Triton X-100, the histone fraction was extracted by resuspending and incubating the pellet in

H<sub>2</sub>SO<sub>4</sub> overnight. The next day, the pellet was collected and further processed according to the protocol. A fraction of the acquired elution samples was loaded on a 12% Bis-Tris gel and stained with Instant Blue (Figure 4.21A). A second fraction was used for Western blot analysis (Figure 4.21B).



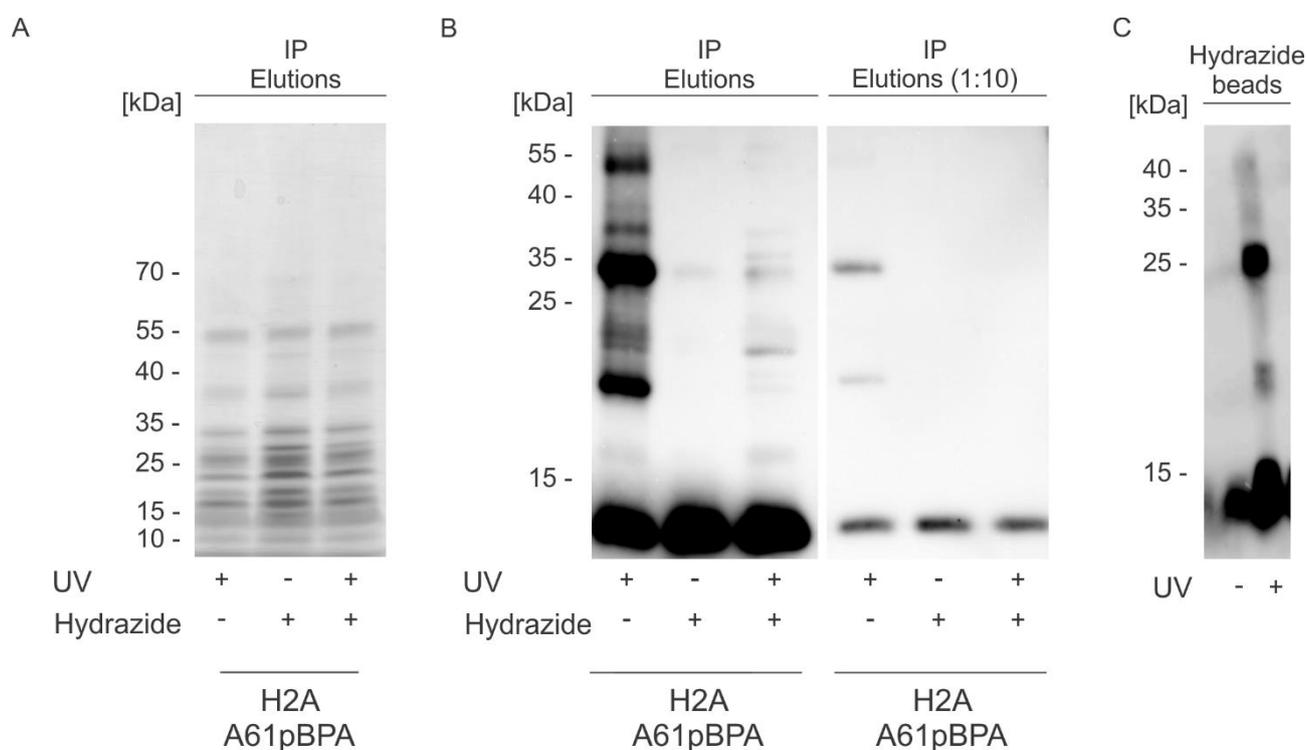
**Figure 4.21:** Enrichment of crosslink products by acid extraction of free histone proteins.

A) Coomassie gel staining of IP elution samples obtained from H4 Q27pBPA mutants by immunoprecipitation with or without prior acid extraction of free H4-HA protein. Full-length H4-HA (~13 kDa) was undetectable after acid extraction. B) Western blot analysis of IP elution fractions showed detectable crosslinks of H4 R27pBPA only in the control sample. No crosslink products were detected after acid extraction. *Legend:* M: Marker.

The stained gel showed an absolute removal of free H4-HA protein from the acid-treated sample in comparison to an untreated sample. However, this comparison also revealed a reduction in overall background intensity and decreased intensity of bands with higher molecular weight (Figure 4.21A). This suggests an overall reduction in protein concentration after acid extraction. Western blot analysis confirmed this observation by uncovering an absolute loss of crosslink products in the acid-treated IP sample (Figure 4.21B). The non-histone protein portion of the crosslink products was incapable of diminishing the solubility of the histone crosslink products in acid. Therefore, acid extraction proves not applicable for the enrichment of histone crosslink products.

Ketones and aldehydes can react with hydrazides under mild acidic conditions by nucleophilic addition-elimination to form a hydrazone. In its unreacted, ground state, the benzophenone moiety of pBPA possesses a free keto group which could be targeted by hydrazides. Coupling of hydrazides to beads would allow the targeted depletion of unreacted pBPA mutants from a sample in order to concentrate crosslink products for a higher yield during immunoprecipitation.

Therefore, cells harboring a histone H2A A61pBPA mutant were grown and UV-treated and non-UV-treated samples were prepared for immunoprecipitation. Before immunoprecipitation, all samples were adjusted to a mild acidic pH of 6 to facilitate the hydrazone formation and incubated with hydrazide-terminated beads. Immunoprecipitation of H2A-HA and its crosslink products was performed with anti-HA antibody-coupled magnetic beads.



**Figure 4.22:** Enrichment of crosslink products by hydrazide chemistry.

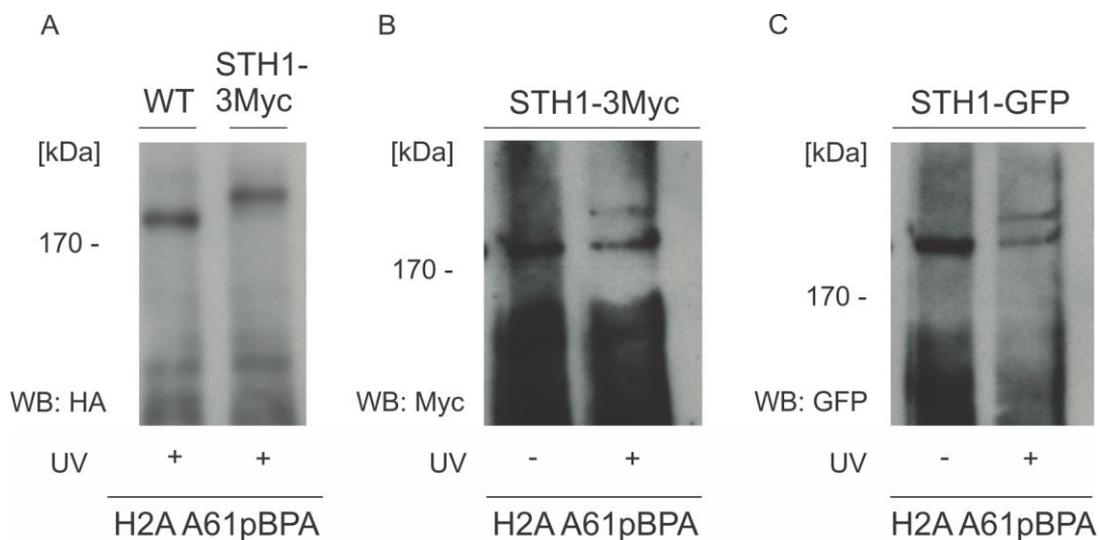
A) Coomassie gel staining of IP elution samples obtained from H2A A21pBPA mutants by immunoprecipitation with or without prior hydrazide treatment. Full-length H2A-HA (~15 kDa) was detectable in all samples. B) Western blot analysis of IP elution fractions (undiluted/1:10) showed a reduction of overall signal in the samples treated with hydrazide. C) Western blot analysis of boiled hydrazide beads showed signals for full-length H2A-HA (~15 kDa) in both lanes. Crosslink products were detected in the UV-treated IP elution sample of H2A A61pBPA obtained from the boiled hydrazide beads.

A fraction of the acquired elution samples was loaded on a 12% Bis-Tris gel and stained with Instant Blue (Figure 4.22A). A second fraction was used, either undiluted or 1:10 diluted, for Western blot analysis (Figure 4.22B). After incubation of the samples with hydrazide beads, a reduction of full-length H2A protein was expected. Analysis of the elution fractions by gel staining showed no changes of banding patterns after treatment with hydrazides (Figure 4.22A). Still, Western blot analysis of the IP elution samples showed a reduction of crosslink products after hydrazide treatment. After extensive washing of the hydrazide beads, boiling them revealed the reaction of a fraction of full-length H2A-HA protein as well as a crosslink products fraction with the hydrazide group (Figure 4.22C). Therefore, the hydrazone formation may not be restricted to unreacted pBPA mutants, but include their crosslink products. This rendered this approach unsuitable for the enrichment of crosslink products from pBPA mutants.

### 4.3.8 Improving Result Evaluation by Introduction of Internal Quality Control

A hindrance for the assessment of the effectiveness and efficiency of the aforementioned SILAC experiments was the lack of an internal control. A site-specific and confirmed interaction partner would act as a quality control for the capability of the procedure to identify true crosslink products at a permissive site. The verification of potential crosslink candidates was previously described in chapter 4.3.4.

Before the beginning of this project, Dr. Bryan Wilkins was able to identify the ATPase STH1 of the RSC chromatin remodeling complex as potential interaction partner of histone H2A at position A61 by combining *in vivo* crosslinking, immunoprecipitation and mass spectrometry. Parallel to this project and with the invaluable help of Corinna Krüger, he was able to produce further evidence for the authenticity of this interaction. Verification was performed by crosslinking of H2A-HA with an A61pBPA substitution in a cellular background in which STH1 was marked genomically with a 3Myc-tag at the C-terminal domain to produce a mass shift. An expected crosslink product of STH1 and H2A-HA is supposed to possess a molecular weight of ~170 kDa. Addition of a 3Myc-tag should result in a visually detectable mass shift of 3.6 kDa.



**Figure 4.23:** Verification of STH1 as interaction partner of histone H2A at position A61.

A) Western blot analysis of whole-cell lysate samples obtained from H2A A61pBPA mutants expressed in wild-type or STH1-3Myc cells and treated with UV-light. A band with the expected size of an H2A:STH1 crosslink product was detected in wild-type samples at ~170 kDa. In the STH1-3Myc cellular background, the band shifted for the expected size of the added 3Myc-tag, 3.6 kDa. B) Western blot analysis of IP elution fractions of H2A A61pBPA expressed in STH1-3Myc cellular background. A band with the expected size of the H2A:STH1-3Myc crosslink product (~173.6 kDa) appeared upon UV-irradiation. C) Western blot analysis of IP elution fractions of H2A A61pBPA expressed in STH1-GFP cellular background. A band with the expected size of the H2A:STH1-GFP crosslink product (~198 kDa) appeared upon UV-treatment.

Comparative Western Blot analysis between crosslinking of H2A A61pBPA in wild-type and in STH1-3Myc yeast cells revealed a defined band at a molecular weight of approximately 170 kDa in the crosslinked wild-type cells which reappeared with increased size in the STH1-3Myc cellular background (Figure 4.23A). Without dispute, this observed mass shift confirmed the identification of STH1 as interaction partner of H2A at position A61 by *in vivo* crosslinking experiment with the genetically encoded crosslinker amino acid pBPA in yeast.

To further strengthen this result, anti-HA immunoprecipitation of crosslink products was performed from STH1-3Myc or STH1-GFP cellular background and analyzed for a crosslink product of expected size by blotting against the respective 3Myc- (Figure 4.23B) or GFP-tag (Figure 4.23C). STH1 has a regular size of 156 kDa, a crosslink product with H2A is expected to have a molecular size of 171 kDa. Addition of the respective tags should yield in an increase in size of 3.6 kDa for STH1-3Myc:H2A to a final molecular weight of 174.6 kDa, whereas STH1-GFP:H2A is supposed to weigh 198 kDa after the addition of the 27 kDa GFP-tag.

Analysis of the non-UV-treated control sample of STH1-3Myc showed a defined band of approximately 170 kDa. The non-crosslinked control sample of STH1-GFP displayed a defined band at marginally higher molecular weight. Since both samples were not subjected to UV-irradiation, both bands were considered as antibody-specific background signal and dismissed for redundancy.

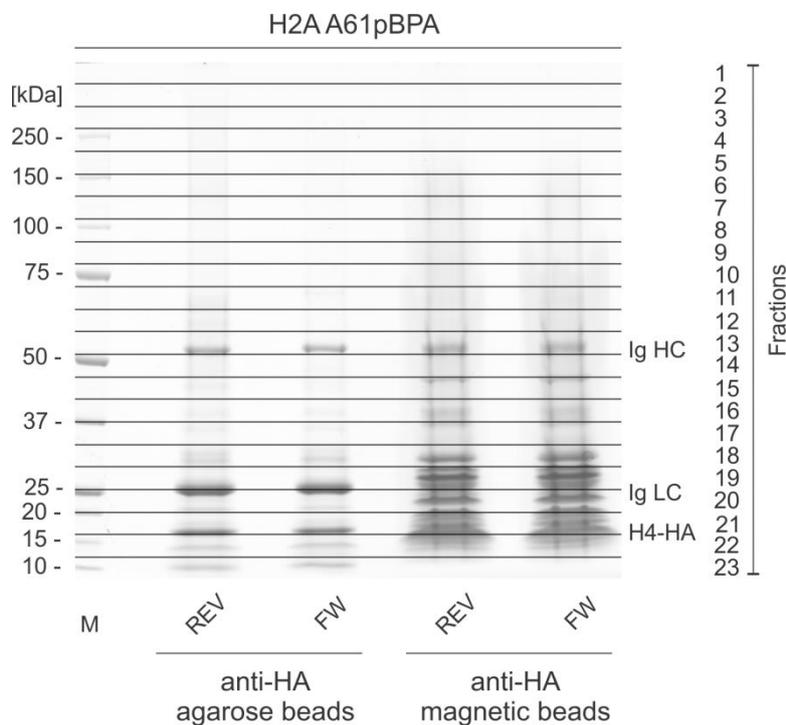
Upon UV-irradiation, both strains showed the appearance of a single additional band of higher molecular weight (Figure 4.23B, C). In the STH1-3Myc background, the detected band ran at an expected lower height than in the STH1-GFP background. This difference in mass was expected due to the difference in size of the added tags. Since the formation of bands was UV-dependent in both strains and fulfilled the expectations of estimated sizes for an STH1:H2A crosslink product in the respective cell line, these immunoprecipitations provided further evidence of a successful detection of an interaction partner of histone H2A. Therefore, STH1 was used as internal quality control for further SILAC experiments, which were performed on position H2A A61, to better assess the effectiveness and efficiency of the developed protocol.

#### **4.3.9 Enhanced Immunoprecipitation Improves Overall Quality of SILAC Analysis**

Agarose beads possess a porous surface, which gives them high surface area for interaction with proteins and allows them to maintain hydration for protein stability. This renders them perfect for the application which they were originally designed for: bulk protein purification in columns. However, employing them for the isolation of specific proteins from a solution using bead-bound antibodies creates some disadvantages which became apparent in the course of this project. The porous surface structure of the beads results in trapping of the antibodies which hinders the washing process, ultimately increasing background signals. The washing procedure was performed in prep columns which increases the chance of accidentally losing sample in the transfer process. Furthermore, the diffusion rate is slow, thereby increasing the time needed for each incubation step. Long incubation periods and extensive washing may cause mechanical and proteolytic damage to the proteins and therefore interfere with the reproducibility of the results.

The usage of magnetic beads rather than agarose beads could potentially solve all of the listed problems. Their non-porous structure offers no hidden surface for trapping of antibodies which can reduce the background and the number of required washing steps. Incubation times can be reduced due to their increased diffusion rate. Less washing steps and faster incubation times mean less chance of proteolytic damage to the sample, ultimately increasing the reproducibility of the experiment. Also, separation of the beads from the supernatant by a magnet reduces the chance of sample loss and renders the whole process faster and more user-friendly. Using the SILAC protocol and STH1 as internal quality control, a comparison was made of overall H2A A61pBPA protein and crosslink product binding between magnetic beads and agarose beads, both conjugated with anti-HA-antibodies. For this, two times two forward and reverse samples of H2A A61pBPA mutant cells were prepared.

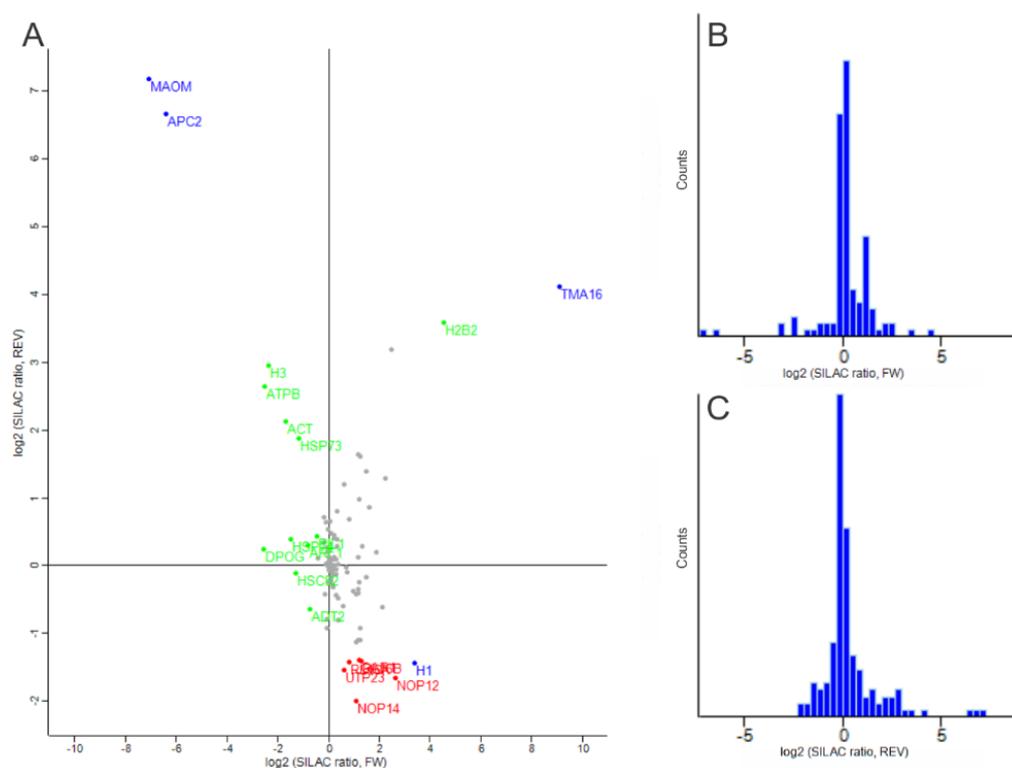
Before immunoprecipitation of H2A-HA and its crosslink products, the forward samples were concentrated to a final concentration of 0.82 mg/mL and the reverse samples to a final concentration of 0.75 mg/ml. Immunoprecipitations were performed with identical concentrations of 0.35 mg anti-HA IgG<sub>1</sub>, conjugated to either magnetic or agarose beads. The acquired elution samples were loaded on a 4-12% Bis-Tris gel and stained with Instant Blue (Figure 4.24).



**Figure 4.24:** Comparative SILAC immunoprecipitation of H2A A61pBPA crosslink products.

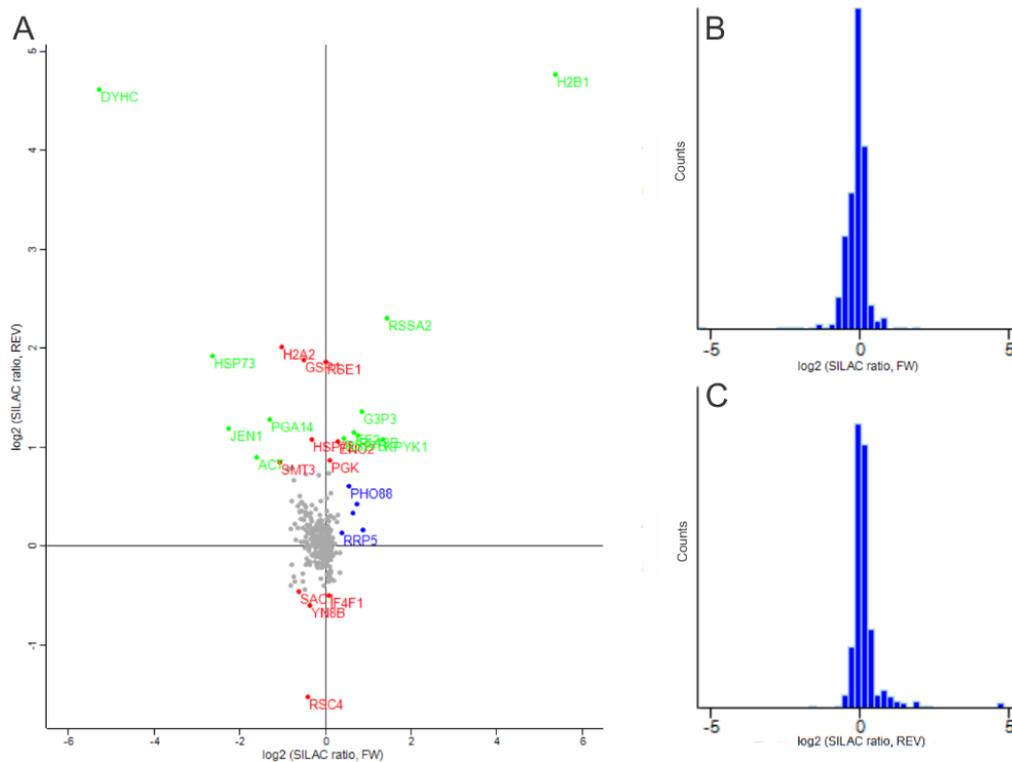
Coomassie-stained gel of IP elution fractions from forward and reverse samples of H2A A61pBPA mutants expressed in BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells obtained with anti-HA antibody-conjugated either agarose or magnetic beads. Full-length H2A-HA (~15 kDa) was enriched in both samples. The usage of magnetic beads resulted in more pronounced overall pulldown of proteins. For improved evaluation of MS/MS results, the cutting fractions of the gel are implemented. *Legend:* M: Marker; REV: Reverse sample; FW: Forward sample; Ig LC: Immunoglobulin light chain; Ig HC: Immunoglobulin heavy chain.

Gel analysis showed noticeable differences between the samples, depending on the kind of beads used. Although all beads were washed extensively after immunoprecipitation, usage of magnetic beads resulted in a much more pronounced pulldown of proteins compared to the agarose beads. Overall, both treatments shared identical bands, but their densities were much higher in the magnetic beads samples. Also, the latter showed a variety of additional bands, mostly of lower molecular weight, which were absent in the agarose beads samples. Additionally, the signal for H2A-HA, expected to be at approximately 15 kDa, showed significantly greater saturation in the magnetic beads sample. First evaluation of the comparison suggests a higher pulldown efficiency of anti-HA antibody-conjugated magnetic beads compared to their agarose-based analogue. Subsequently, the samples were digested and analyzed by MS.



**Figure 4.25:** Identification of proteins from H2A A61pBPA SILAC analysis with agarose beads.

A) The log<sub>2</sub> SILAC (i.e., heavy/light L-arginine and/or L-lysine) ratios calculated for each protein identified in the forward and reverse experiments of H2A A61pBPA immunoprecipitation with anti-HA agarose beads are plotted against each other. B) Distribution of normalized log<sub>2</sub> SILAC ratios from forward sample. C) Distribution of normalized log<sub>2</sub> SILAC ratios from reverse sample. *Legend:* in green: Proteins enriched in both samples; in blue: Proteins enriched in forward sample; in red: Proteins enriched in reverse sample.



**Figure 4.26:** Identification of proteins from H2A A61pBPA SILAC analysis with magnetic beads.

A) The log<sub>2</sub> SILAC (i.e., heavy/light arginine and/or lysine) ratios calculated for each protein identified in the forward and reverse experiments of H2A A61pBPA immunoprecipitation with anti-HA magnetic beads are plotted against each other. B) Distribution of normalized log<sub>2</sub> SILAC ratios from forward sample. C) Distribution of normalized log<sub>2</sub> SILAC ratios from reverse sample. *Legend:* in green: Proteins enriched in both samples; in blue: Proteins enriched in forward sample; in red: Proteins enriched in reverse sample.

The usage of agarose beads produced an overall percentage of identified MS2 spectra of only 1.02% with a 5.14% share of identified MS2 spectra of SILAC clusters. Employing magnetic beads resulted in an overall percentage of identified MS2 spectra of 2.78 with a 15.06% share of identified MS2 spectra of SILAC clusters. Matching of the peak files to proteins resulted in the identification of 232 proteins from 1790 peptides after usage of antibody-conjugated agarose beads (Figure 4.25A). Usage of their magnetic analogue led to an enhanced detection of 4885 peptides, resulting in the identification of 480 proteins (Figure 4.26A).

The histograms of the log<sub>2</sub> SILAC ratios of all samples showed a normal distribution centred at zero for both samples, supporting similar overall protein loads of the light and heavy sample fractions (Figure 4.25B, C; Figure 4.26B, C)

Statistical evaluation produced an intersecting set of 150 proteins for both samples. Quantification of peptide levels showed recruitment of 1817 unique peptides for these intersecting proteins in the magnetic bead samples with a mean value of 12 unique peptides per identified protein. Immunoprecipitation with agarose beads led to the identification of these proteins from 735 unique peptides with a mean of 5 unique peptides per protein. Pulldown with agarose beads resulted in isolation of 82 proteins exclusive to these samples. In contrast, magnetic beads produced a sole selection of 329 proteins.

Potential crosslink products were characterized by a normalized SILAC ratio of at least 1.5 in both forward and reverse samples. From 150 intersecting proteins, 28 were categorized as significant, 9 of which were enriched in the magnetic beads samples (1 potential crosslink included, Table 4.12), 15 in the agarose bead samples (1 potential crosslink included, Table 4.13), and 5 in both, one of which was categorized as crosslink product in both (Table 4.14).

**Table 4.12:** Intersecting hits with high significance in the magnetic bead samples of H2A A61pBPA.

Protein Names	Gene Names	Unique Peptides	SILAC ratio, normalized, FW/REV		Significance B (corr), FW/REV	
Pyruvate kinase 1	CDC19	16	1.344	1.073	3.01 E-22	8.62 E-05
60S ribosomal protein L19-B	RPL19B	6	0.757	1.116	1.79 E-08	4.30 E-05
Pyruvate decarboxylase isozyme 1	PDC1	8	0.741	0.422	3.32 E-08	0.151
Inorganic phosphate transport protein PHO88	PHO88	4	0.544	0.608	3.26 E-05	0.031
Enolase 2	ENO2	12	0.282	1.053	0.019	0.0001
Phosphoglycerate kinase 1	PGK1	8	0.111	0.863	0.247	0.001
Pre-mRNA-splicing factor RSE1	RSE1	2	0.008	1.864	0.657	3.55 E-12
Histone H2A	HTA2	4	-1.007	2.009	0.009	5.83 E-14
Hydrophilin YNL190W	YNL190W	5	-1.292	1.274	8.049 E-04	2.65 E-06

**Table 4.13:** Intersecting hits with high significance in the agarose bead samples of H2A A61pBPA.

Protein Names	Gene Names	Unique Peptides	SILAC ratio, normalized, FW/REV		Significance B (corr), FW/REV	
Translation machinery-associated protein 16	TMA16	8	9.098	4.122	3.37 E-05	0.004
ADP-ribosylation factor 1	ARF1	2	-0.805	0.307	7.73 E-07	0.815
Histone H3	HHT3	1	-2.371	2.94	6.90 E-42	0.019
NAD-dependent malic enzyme	MAE1	1	-7.064	7.180	0	1.19 E-08
Histone H1	HHO1	2	3.393	-1.440	0.002	8.78 E-04
60S ribosomal protein L16-A	RPL16A	3	0.800	-1.419	0.510	0.001
rRNA-processing protein UTP23	UTP23	10	0.604	-1.547	0.633	3.54 E-04
60S ribosomal protein L16-B	RPL16B	6	1.284	-1.402	0.268	0.001
Nucleolar complex protein 14	NOP14	3	1.075	-2.005	0.361	3.88 E-06
Myosin-3	MYO3	3	-	-1.946	1	7.3 E-06
H/ACA complex subunit 1	GAR1	2	1.219	-1.401	0.294	0.001
Nucleolar protein 12	NOP12	6	2.630	-1.656	0.018	1.33 E-04
ADP, ATP carrier protein 2	PET9	6	-0.735	-0.639	5.26 E-06	0.134
Sphingolipid protein PIL1	PIL1	3	-0.445	0.430	0.003	0.740
ATP-dependent chaperone HSC82	HSC82	2	-1.288	-0.118	2.9 E-14	0.760

## 4 Results

**Table 4.14:** Intersecting hits with high significance in both data sets of H2A A61pBPA.

Protein Names	Gene Names	Unique Peptides (magn./aga. beads)		SILAC ratio, normalized, FW/REV, (magn. beads)		SILAC ratio, normalized, FW/REV, (aga. beads)		Significance B (corr), FW/REV, (magn. beads)		Significance B (corr), FW/REV, (aga. beads)	
Histone H2B	HTB1	4	8	5.371	4.765	4.556	3.583	0.00	6.46 E-73	3.37 E-05	0.004
Actin	ACT1	9	4	-1.603	0.897	-1.686	2.121	2.75 E-05	0.001	1.26 E-22	0.0937
TATA-binding protein factor MOT1	MOT1	2	1	-2.10855	-	1	8.43 E-70	2.69 E-08	2.69 E-08	8.43 E-70	1
Heat shock protein SSA3	SSA3	-	-	-2.623	1.924	-3.115	-	3.51 E-12	6.76 E-13	7.51 E-12	0.138
ATP synthase subunit beta	ATP2	1	2	-	-0.709	-2.515	2.637	1	5.31 E-06	1.00 E-46	0.036

Comparison of the sole protein populations of both bead types showed 22 proteins to be significant in the magnetic bead samples (Table 4.15) versus 6 proteins in the agarose bead samples (Table 4.16).

**Table 4.15:** Significant hits solely detected in the magnetic bead sample of H2A A61pBPA.

Protein Names	Gene Names	Unique Peptides	SILAC ratio, normalized, FW/REV		Significance B (corr), FW/REV	
40S ribosomal protein S0-B	RPS0B	2	1.442	2.305	2.93 E-25	5.44 E-18
Glyceraldehyde-3-phosphate dehydrogenase 3	TDH3	5	0.844	1.358	4.45 E-10	5.20 E-07
Elongation factor 2	EFT1	8	0.669	1.146	5.15 E-07	2.58 E-05
40S ribosomal protein S27-B	RPS27B	2	0.430	1.088	7.54 E-04	6.78 E-05
Carboxylic acid transporter protein homolog	JEN1	2	-2.252	1.183	2.69 E-0	1.37 E-05
Dynein heavy chain	DYN1	1	-5.291	4.614	1.24 E-45	1.88 E-68
Mitochondrial protein import protein MAS5	YDJ1	4	1.809	-	3.11 E-38	1
Alcohol dehydrogenase 1	ADH1	7	0.880	0.166	8.63 E-11	0.648
Mitochondrial pyruvate carrier 2	MPC2	1	0.843	-	4.76 E-10	1
Heat shock protein SSB2	SSB2	6	0.658	-	7.50 E-07	1
Fructose-bisphosphate aldolase	FBA1	5	0.648/0.337		1.07 E-06	0.268
rRNA biogenesis protein RRP5	RRP5	6	0.397/0.128		0.001	0.757
Protein MSS51	MSS51	2	-1.482/-		1.11 E-04/1	
Altered inheritance of mitochondria protein 18	AIM18	1	-2.438/-		1.09 E-10/1	
Ketol-acid reductoisomerase	ILV5	3	-1.385		1/3.00 E-07	
Eukaryotic initiation factor 4F subunit p150	TIF4631	17	0.079/-0.505		0.351/8.77 E-04	
Heat shock protein SSA2	SSA2	5	-0.318/1.075		0.475/8.32 E-05	
UPF0674 endoplasmic reticulum membrane protein YNR021W	YNR021W	4	-0.366/-0.598		0.399/1.035 E-04	

Chromatin structure-remodeling complex subunit RSC4	RSC4	3	-0.421/-1.523		0.321/3.52 E-21	
GTP-binding nuclear protein GSP1/CNR1	GSP1	5	-0.514	1.880	0.213	2.28 E-12
Phosphoinositide phosphatase SAC1	SAC1	2	-0.629	-0.463	0.119	0.002
Ubiquitin-like protein SMT3	SMT3	2	-1.067	0.847	0.006	0.002

**Table 4.16:** Significant hits solely detected in the agarose bead sample of H2A A61pBPA.

Protein Names	Gene Names	Unique Peptides	SILAC ratio, normalized, FW/REV		Significance B (corr), FW/REV	
Anaphase-promoting complex subunit 2	APC2	1	-6.421	6.357	1.34 E-281	1.24 E-07
Heat shock protein SSA4	SSA4	0	-1.471	0.395	7.52 E-18	0.762
DNA polymerase gamma	MIP1	2	-2.576	0.243	7.48 E-49	0.855
Protein BMH1	BMH1	1	-3.169	-	4.32 E-72	1
DNA-directed RNA polymerase III subunit RPC1	RPO31	1	-	6.494	1	2.54 E-07
Phosphate metabolism protein 7	PHM7	2	-	-1.971	1	5.62 E-06

Usage of magnetic beads identified 2 proteins as potential crosslink products, enriched in forward and reverse, as well as 2 proteins either enriched in forward or reverse sample. Analysis of the agarose bead samples revealed only one protein to be enriched in the forward sample.

Altogether, application of magnetic beads for immunoprecipitation resulted in the identification of 480 proteins from 4885 peptides of which 329 were identified exclusively, and a total of 31 proteins classed as significantly enriched with 6 potential crosslink candidates. On the contrary, usage of agarose beads led to the identification of 232 proteins from 1790 peptides with a sole selection of 82 proteins, and a total of 22 significant protein hits with 3 potential crosslink products. However, in neither data set the known control protein STH1 was identified as a significantly increased hit.

The significant increase in protein identifications by using magnetic beads raised the question if potential crosslink candidates, including the internal quality control STH1, were detected but discarded as background noise due to low peptide counts and therefore poor ion statistics. Therefore, the exclusively identified protein groups of both pulldown approaches were screened for hits with roles in chromatin remodeling and regulation.

**Table 4.17:** Comparison of exclusive protein hits with chromatin biological background in both samples

Protein Names	Gene Names	Unique Peptides	Fraction (total peptides)	Excl. to aga. beads	Excl. to magn. beads	SILAC ratio, normalized, FW/REV		Significance B (corr), FW/REV	
Chromatin modification-related protein EAF6	EAF6	1	4 (1)	+		-	-	1	1
Protein BMH1	BMH1	1	5 (1), 6 (1), 15 (1)	+		-3.169	-	4.32 E-72	1
Anaphase-promoting complex subunit 2	APC2	1	2 (1), 17 (1), 18 (4), 19 (1),	+		-6.421	6.357	1.34 E-281	1.24 E-07
GLC7-interacting protein 4	GIP4	1	8 (4)		+	-	-	1	1
Histone deacetylase HDA1	HDA1	2	18 (1), 20 (1)		+	-	-	1	1

## 4 Results

Histone deacetylase HOS3	HOS3	1	2 (1), 19 (1), 20 (2), 21 (1), 22 (2), 23 (1)		+	-	-	1	1
ISWI one complex protein 2	IOC2	2	7 (3)		+	-	0.364	1	0.224
Nuclear protein STH1/NPS1	STH1	4	5 (7)		+	0.231	0.017	0.046	0.854
Chromatin structure- remodeling complex subunit RSC9	RSC9	6	11 (8)		+	0.132	0.165	0.193	0.651
ISWI one complex protein 3	IOC3	10	7 (21), 8 (4)		+	0.052	0.079	0.457	0.902
Chromo domain- containing protein 1	CHD1	25	4 (16), 5 (39), 6 (14), 7 (10), 8 (2), 11 (2)		+	0.015	0.066	0.623	0.942
Chromatin structure- remodeling complex protein RSC58	RSC58	23	11 (56), 14 (2)		+	-0.002	-0.058	0.714	0.522
Origin recognition complex subunit 2	ORC2	8	1( 1), 9 (12)		+	-0.007	-0.012	0.743	0.715
DNA topoisomerase 2- associated protein PAT1	PAT1	5	8 (10)		+	-0.029	-0.402	0.861	0.006
Histone-lysine N- methyltransferase, H3 lysine-79 specific	DOT1	7	10 (14), 11 (9)		+	-0.053	0.212	0.992	0.528
Origin recognition complex subunit 1	ORC1	4	7 (4), 9 (3), 10 (1)		+	-0.155	0.719	0.786	0.010
Putative DNA helicase INO80	INO80	14	4 (14), 5 (27), 6 (3), 9 (2),		+	-0.201	0.085	0.690	0.884
Chromatin structure- remodeling complex protein RSC3	RSC3	9	7 (14)		+	-0.236	-0.068	0.623	0.486
Bromodomain- containing factor 1	BDF1	13	7 (1), 8 (22), 9 (7), 10 (1), 11 (5), 16 (2)		+	-0.304	0.080	0.499	0.898
Chromatin structure- remodeling complex subunit RSC4	RSC4	3	9 (5)		+	-0.421	-1.523	0.321	3.52 E-21
Chromatin structure- remodeling complex protein RSC14	LDB7	2	18 (4)		+	-0.428	0.209	0.311	0.536
Chromatin modification- related protein EAF7	EAF7	9	10 (19), 11 (1)		+	-0.492	0.273	0.236	0.387
ISWI one complex protein 4	IOC4	3	10 (6)		+	-0.533	-0.362	0.195	0.013
Ino80 subunit 1	IES1	2	8 (3), 9 (1)		+	-0.549	-	0.180	1
Chromatin structure- remodeling complex protein RSC30	RSC30	3	7 (5), 8 (1)		+	-0.574	0.043	0.159	0.978

Remarkably, the screening revealed STH1 to be present in the data set acquired with magnetic beads, but not in the one obtained from agarose beads (Table 4.17). Although its forward and reverse SILAC ratios were too low for unambiguous quantification (0.231/0.017), the identification of STH1 provided first proof of the effectiveness of a combination of *in vivo* crosslinking, immunoprecipitation, and SILAC for the capturing and identification of histone crosslink products.

In total, the screening resulted in 3 hits (see Table 4.17: APC2, BMH1, EAF6) for the agarose bead samples and 22 hits (see Table 4.17: BDF1, CHD1, DOT1, EAF7, GIP4, HOS3, HDA1, IES1, INO80, IOC2, IOC3, IOC4, LDB7, ORC1, ORC2, PAT1, RSC3, RSC4, RSC9, RSC30, RSC58, STH1) for the magnetic bead samples of H2A amber mutant A61. The hits exclusive to the agarose bead samples were the APC2 subunit of the Anaphase-Promoting Complex (APC), the transcription regulator protein BMH1 and the EAF6 subunit of the acetyltransferase complexes NuA3 and NuA4, which acetylate histones H3 and H2A/H4, respectively. Notably, APC2 and BMH1 were already described as significantly enriched before, but none of the proteins could potentially be categorized as a potential crosslink product based on their SILAC ratios.

The magnetic beads produced a larger yield of chromatin-related hits with a generally higher complexity. BDF1 is a subunit of the SWR1 chromatin remodeling complex that exchanges H2A for its variant H2A.Z. CHD1 is an ATP-dependent chromatin remodeling factor which is involved in the recognition of H3 K4 methylation as substrate for the transcription-regulatory histone acetylation (HAT) complexes SAGA and SLIK. DOT1 is a histone methylase which targets lysine 79 of histone H3 and is involved in telomeric silencing, meiotic checkpoint control and DNA damage response. EAF7 is part of the NuA4 histone acetyltransferase complex which targets the N-terminal domains of H2A and H4. The regulator protein GIP4 is involved in control of protein phosphatase 1 (GLC7/PP1) which targets e.g. H3 S10 phosphorylation. HOS3 and HDA1 are both histone deacetylases. HOS3 predominantly targets all four core histones, whereas HAD1 targets all core histones except H2A. The ATPase INO80 and the subunit protein IES1 are both part of the INO80 chromatin remodeling complex which promotes nucleosome shifts. IOC2 is a subunit of the ISWIa complex, in contrast IOC3 and IOC4 are both part of the ISWIb complex. Both complexes are responsible for the ATP-dependent repositioning of nucleosomes. The ATPase ORC1 and the subunit protein ORC2 are both part of the Origin Recognition Complex (ORC) which aids in the assembly of transcriptionally silent chromatin at the mating-type loci Hidden MAT Left (HML) and Hidden MAT Right (HMR) by recruitment of deacetylases from the Silent Information Regulator (SIR) family. PAT1 is a Topoisomerase II-associated protein which is necessary for structure integrity of centromeric chromatin and chromosome transmission during meiosis and mitosis. Strikingly, LDB7, RSC3, RSC4, RSC9, RSC30, RSC58 and the ATPase STH1 are all components of the RSC chromatin remodeling complex.

The multitude of related proteins from different chromatin remodeling complexes as well as single histone modifiers encourages the idea of successful capturing of histone-protein interactions within a nucleosomal landscape. Notably, the ratios of light and heavy peptide pairs were very low for these proteins compared to the earlier discussed most abundant proteins in the sample (Figure 4.27).



hypotheses testing which describes a simultaneous test for each protein on the null hypothesis of no change in protein measure between the two samples. This includes computing a test statistic and applying a multiple-testing procedure to determine which hypothesis to reject while controlling a defined false-positive error rate (Dudoit et al., 2004).

Significance of protein ratios can be determined by obtaining a robust and asymmetrical estimate of the standard deviation of the main distribution while taking the dependence of the distribution on the summed protein intensity into account. The accuracy of a protein ratio is evaluated by calculating the coefficient of variability over all redundant quantifiable peptides. By statistical measure, valid hits should not exceed a p-value of 0.05. It needs to be pointed out that the p-value of 0.05 as the cut-off between significance and insignificance is used by convention, not on principle.

Review of the chromatin-related protein hits from the forward and reverse SILAC experiment of H2A A61pBPA revealed the majority of p-values to be significantly higher than the named threshold (Table 4.17). Notably, out of all proteins of both analyses, only one, APC2, fulfilled the threshold criterion in both forward ( $p_{fw} = 1.34 \text{ E-}72$ ) and reverse ( $p_{rev} = 1.24 \text{ E-}07$ ) samples. Other proteins could only satisfy the criterion in one sample, while displaying a significant deviation from this value in the other, e.g., BMH1 ( $p_{fw} = 4.32 \text{ E-}72$ ,  $p_{rev} = 1$ ), IES1 ( $p_{fw} = 0.861$ ,  $p_{rev} = 0.006$ ), IOC4 ( $p_{fw} = 0.786$ ,  $p_{rev} = 0.010$ ), PAT1 ( $p_{fw} = 0.195$ ,  $p_{rev} = 0.013$ ), RSC4 ( $p_{fw} = 0.321$ ,  $p_{rev} = 3.52 \text{ E-}21$ ) and STH1 ( $p_{fw} = 0.046$ ,  $p_{rev} = 0.854$ ). Considering a limitation of statistical power by variability of biological samples, regardless of whether the returning p-value is low or high, a repetition of the same experiment would likely result in a substantially different level of evidence against the null hypothesis. The computed p-values seem to not produce reliable information about probable results of low abundant proteins in a replicated experiment. The deviation of p-values between biological replicates cast doubts on their applicability for testing the significance of crosslink products.

The MS raw files also contain information about the gel fractions in which peptides of an identified protein were measured. It can be assumed that trapping of an interaction by *in vivo* crosslinking of amber mutants results in the formation of a unique crosslink product with a defined molecular weight. By calculating the estimated molecular weight of a crosslink product and validating this against the local information of the detected peptides within the gel, this should shed light on the probability of an identified trapped interaction. A unique crosslink product with a defined molecular weight should only be detectable in a single fraction, with a minor exception to adjacent fractions due to cutting artefacts. Detection of peptides in multiple, nonadjacent fractions could indicate an unspecific binding of the protein to the beads or degradation. Therefore, such a particular hit may be a false positive.

In the following, the detected proteins related to chromatin structure and function were evaluated based on their estimated molecular weight in complex with histone H2A and their fractions of detection (Table 4.17). Complexes with high molecular weight were expected to be located within the upper fractions of the analyzed gel, whereas light complexes were supposed to be detected in the lower fractions according to their molecular weight (Figure 4.24).

Firstly, the hits acquired from the agarose bead samples were evaluated. A complex of the 12.9 kDa heavy EAF6 and histone H2A-HA is expected to be 28 kDa. It was identified from 1 peptide in fraction 4, which

speaks against acceptance. A BMH1 crosslink product is expected to be 45 kDa. Detection of BMH1 occurred in fractions 5, 6 and 15 with 1 peptide each. Fraction 15 equates to a molecular weight between 37 and 50 kDa, which fits the expected complex weight. However, fractions 5 and 6 cover molecular weights between 100 and 150 kDa which would correlate to a 2- to 3-fold weight of the complex. Also, BMH1 was only detected in the forward sample with a large negative  $\log_2$  value, which argues against it being a crosslink product. APC2 was detected in fraction 1 with 1 peptide and fractions 17 to 19 with 1, 4 and 1 peptides, respectively. Its crosslink product with H2A is estimated to be 115 kDa. Although it was detectable in the top fraction of the gel, it was mainly detected in the lower fractions of the gel which makes it highly unlikely to be a crosslink product. This is supported by its divergent SILAC ratios in the forward and reverse sample, which indicated a label-dependent background signal. To sum up, none of the potential crosslink products acquired from the agarose bead samples was categorized as valid.

Screening of the MS raw data files from the magnetic beads samples revealed a significantly increased number of chromatin remodeling proteins, histone modifiers and regulator proteins. GIP4 was identified from 4 peptides in fraction 8 which corresponds to a molecular weight of 100 kDa. This correlates with the crosslink product's estimated molecular weight of 100 kDa. However, no SILAC ratio was detected for this protein by which it could have been validated further. Therefore, this candidate remains unclear. A potential HDA1-H2A-HA complex would have an estimated size of 95 kDa and would be expected to be detected in fractions 8 to 9. Detection of single peptides in the not directly adjacent fractions 18 and 20 in combination with missing SILAC ratios suggest a true negative. The same line of argumentation holds true for HOS3. While having an expected molecular weight of 95 kDa, it was detected in fractions 2, 19, 20, 21, 22 and 23. Therefore, HOS3 can be rejected as a candidate. IOC2 in complex with H2A has an approximate weight of 110 kDa which correlates roughly to fractions 7 and 8. Indeed, it was identified by 3 peptides in fraction 7, which renders it potentially positive. However, although 2 of 3 peptides were unique, it was only detected in the reverse sample with a small, but positive  $\log_2$  value (0.364). For that reason, IOC2 could not be fully categorized as potentially true yet.

STH1 was detected solely in fraction 5 from 7 peptides, of which 4 are unique, and produced SILAC pairs in both forward and reverse samples with ratios of 0.231 and 0.0172, respectively. The estimated weight of 170 kDa was expected to be found between fractions 4 to 6, which concurs with the local information from the MS raw file. Also, both SILAC ratios, regardless of their small values, were positive, which allows the assumption of both to be part of the crosslinked sample fraction. Under consideration of STH1's earlier proved authenticity as interaction partner (0), these findings provided more evidence for the effectiveness of this approach.

The detection of RSC9 occurred in fraction 11 from 8 peptides of which 6 were of unique character. The including fraction correlated with an expected size of 80 kDa. SILAC ratios of 0.132 in the forward sample and 0.165 in the reverse sample suggest an allocation to the crosslinked fractions, although the values remain comparably small, as already observed at STH1. Nevertheless, RSC9 might be also a valid crosslink product. IOC3 was detected mainly in fraction 7 with 21 peptides and in the adjacent fraction 8 with 4 peptides. Fractions 7 and 8 correlated with an estimated crosslink product size of 105 kDa. Since the majority of peptides

was found in fraction 7, the detection of the remaining peptides in the adjacent fraction is probably a cutting artefact. The calculated SILAC ratios were small, in detail 0.052 in the forward sample and 0.079 in the reverse sample, but still positive, therefore, the protein is possibly affiliated to the crosslink portion of the samples.

The complex formed by CHD1 and H2A is supposedly 185 kDa in size. Analysis of the fractions revealed peptides to be detected in fractions 4 through 8 and 11. Although the majority of peptides was measured in fractions 4 (16 peptides) and 5 (39 peptides), which was in line with the molecular weight of the complex, the detection of further peptides in other fractions of lower molecular weight raised suspicion about the validity of this hit as potential crosslink product. It remains to be determined if the flux of peptides across the fractions is the result of protein degradation due to the large size of the complex or evidence for false background noise. The SILAC ratios in the forward and reverse sample showed positive, but very small  $\log_2$  values (0.015 and 0.066). Collectively, CHD1 cannot be ruled out completely as potential candidate for a trapped interaction with histone H2A.

The remaining candidate proteins (RSC58, ORC2, PAT1, DOT1, ORC1, INO80, RSC3, BDF1, RSC4, LDB7, EAF7, IOC4, IES1, and RSC30) could not be properly evaluated based on their SILAC ratios, since they all exhibited partially or completely negative  $\log_2$  values in variable magnitudes.

RSC58 and H2A-HA were expected to form a 75 kDa crosslink product which was accordingly detected in fraction 11 from a majority of 56 peptides. However, 2 peptides were also detected in fraction 14, which would correlate to unbound RSC58. The crosslink product of ORC2 weighs approximately 85 kDa, which correlated positively with the majority of its detected peptides being in fraction 9. But a single peptide was also detected in fraction 1, although this may be an artefact of the gel electrophoresis procedure.

Solely fraction 8 contained 10 peptides of PAT1 which fulfilled the expectation of a 105 kDa crosslink product with H2A-HA. DOT1 was identified from 14 peptides in fraction 10 and 9 peptides from the neighbouring fraction 11. This is approximately in line with an 80 kDa crosslink product.

The ATPase ORC1 was detected from multiple fractions, namely 7 (4 peptides), 9 (3 peptides) and 10 (1 peptide). Detection in fraction 7 coincided with a 120 kDa crosslink product, but alternative detections produced doubts about the significance. Another ATPase, INO80, was also detected in multiple fractions: The majority of peptides was detected in fraction 4 (14 peptides) and 5 (27 peptides), which is in line with a 185 kDa crosslink product. However, detection of 3 peptides in fraction 6 and 2 peptides in fraction 9 raised suspicion about protein degradation or background noise.

RSC3 was identified from 14 peptides in fraction 7, of which 9 were classified as unique. With a molecular weight of 115 kDa, the potential crosslink product was identified in the correct segment of the gel.

BDF1 with an estimated complex size of 90 kDa was identified by a majority of 22 peptides in fraction 8, which equated to the correct molecular weight, but peptides were also detected in the adjacent fractions 7 (1 peptide) and 9 (7 peptides), as well as fractions 10 (1 peptide), 11 (5 peptides) and 16 (2 peptides). Therefore, it remains to be determined if this comprehensive detection is the result of degradation or a pointer towards background noise.

Appropriate to its estimated size of 87 kDa, a potential crosslink product of RSC4 was detected in fraction 9 from 5 peptides, including 3 unique ones. The identification of a crosslinked complex from LDB7 and H2A

was done from 4 peptides in fraction 18, correctly located to the expected size of 35 kDa. EAF7 was detected accordingly to an expected molecular weight of 65 kDa from 19 peptides in fraction 10 and 1 peptide in the neighbouring fraction 11. The subunit protein IOC4 was singly detected from 6 peptides in fraction 10 which would be in line with a 70 kDa crosslink product. Peptides of IES1 were detected in fractions 8 (3 peptides) and 9 (1 peptide), which correlated with a crosslink product of 95 kDa. RSC30 was identified from 5 peptides in fraction 7 and 1 peptide in fraction 8. This supports the detection of trapped interaction between RSC30 and H2A-HA with an estimated size of 115 kDa.

In summary, the assessment of validity for potential crosslink candidates from the magnetic bead SILAC sample of H2A A61pBPA produced a number of potentially valid hits based on their fraction of discovery and to some extent on their SILAC ratios. As already stated, the latter can only be drawn on with reservation.

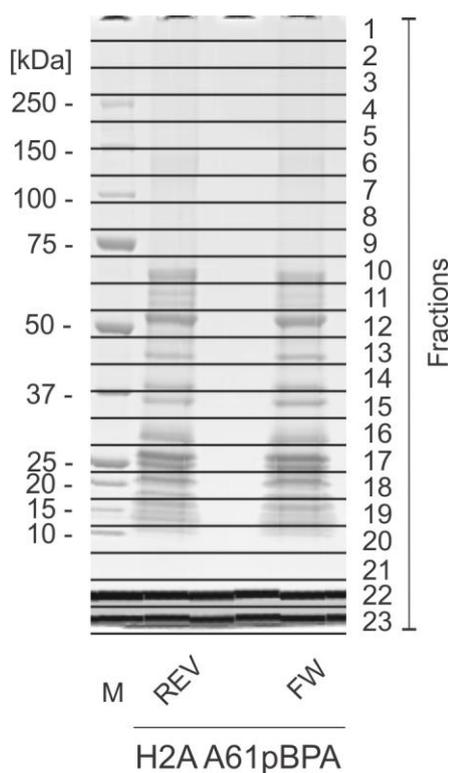
Based on the named criteria, the proteins IOC2, IOC3, STH1, RSC9 and conditionally CHD1 were classified as potential crosslink products. While their SILAC ratios were ambiguous, only the correct assignment of DOT1, EAF7, IES1, IOC4, ORC2, PAT1 RSC3, RSC4, and RSC30 to their expected fractions provided potential evidence for crosslink character. BDF1, INO80, ORC1 and RSC58 remained inconclusive. GIP4, HDA1 and HOS3 were excluded from analysis by reason of missing SILAC ratios.

In conclusion, the application of anti-HA antibody-conjugated magnetic beads to immunoprecipitation in preparation for SILAC-based MS analysis of histone pBPA mutants enhanced the pulldown efficiency and thus the overall yield in peptides and identified proteins significantly. It yielded the identification of a substantially increased number of proteins, including a significant number related to chromatin structure and function as well as the known interaction partner STH1 (0), which indicates an improved efficiency in the capturing of crosslink products. Although the majority of these proteins can be correctly assigned to their expected fractions of discovery, all proteins are defined by small SILAC ratios, assumably because of poor ion statistics based on low intensity spectra. Additionally, a high degree of deviation in significance (*p*) values of individual proteins was observed between forward and reverse samples, which creates doubts on their applicability for testing the significance of crosslink products. Thus, the parameters for testing the probability of crosslink products may need to be refined.

### 4.3.10 Improved Overall Quality Allows Reproduction of Results

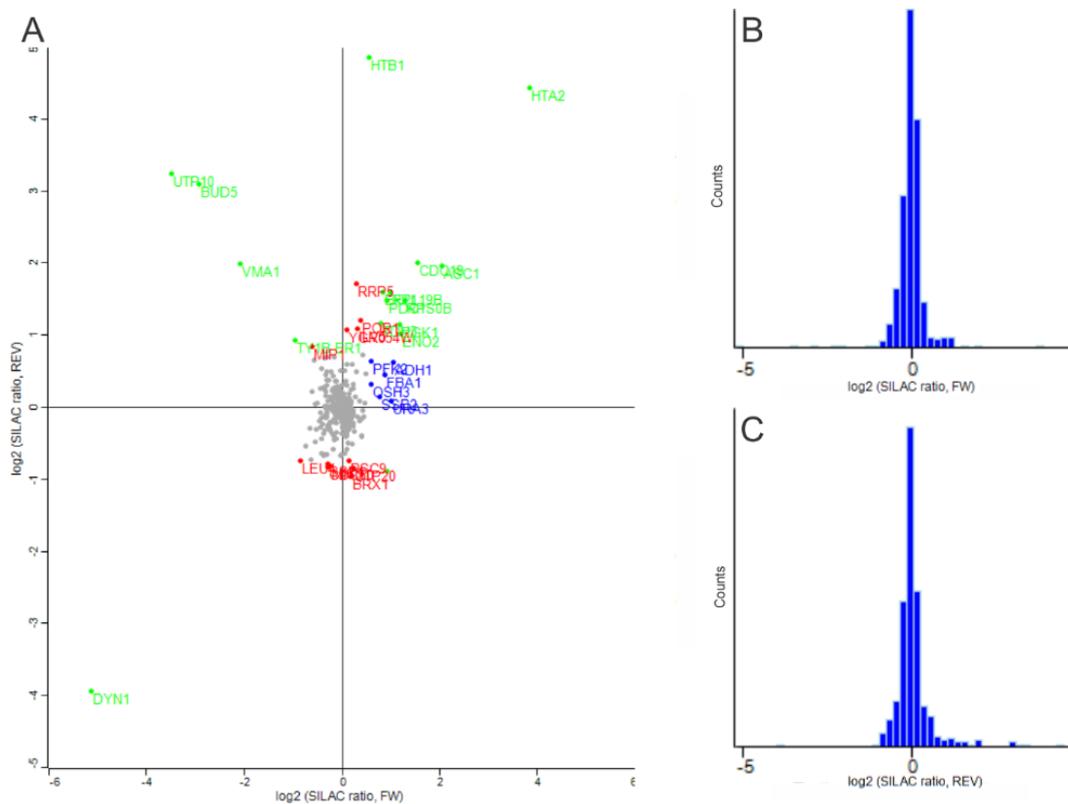
The preceding experiment promises compelling improvements in regard to pulldown efficiency and crosslink identification. However, it remains to be verified that the results are reproducible and that the observed low SILAC ratios and fluctuating significance values of potential crosslinks are a feature of the current experimental setup. Therefore, the previous SILAC experiment of a histone H2A A61 amber mutant was repeated and analyzed for conformity of overall quality and identified proteins.

The forward sample was concentrated to a final concentration of 0.90 mg/mL and the reverse sample to a final concentration of 0.85 mg/mL. Immunoprecipitations were performed with identical concentrations of 0.35 mg anti-HA IgG<sub>1</sub> conjugated to magnetic beads.



**Figure 4.28:** SILAC immunoprecipitation of H2A A61pBPA crosslink products.

Coomassie-stained gel of IP elution fractions from forward and reverse samples of H2A A61pBPA mutants expressed in BY4741  $\Delta$ Arg4/ $\Delta$ Lys2 cells. Full-length H2A-HA (~15 kDa) was enriched in both samples. For improved evaluation of MS/MS results, the cutting fractions of the gel are implemented.  
*Legend:* M: Marker; REV: Reverse sample; FW: Forward sample.



**Figure 4.29:** Identification of proteins from H2A A61pBPA SILAC analysis.

A) The log<sub>2</sub> SILAC (i.e., heavy/light L-arginine and/or L-lysine) ratios calculated for each protein identified in the forward and reverse experiments of H2A A61pBPA immunoprecipitation with anti-HA magnetic beads are plotted against each other. B) Distribution of normalized log<sub>2</sub> SILAC ratios from forward sample. C) Distribution of normalized log<sub>2</sub> SILAC ratios from reverse sample. *Legend:* in green: Proteins enriched in both samples; in blue: Proteins enriched in forward sample; in red: Proteins enriched in reverse sample.

The acquired elution samples were loaded on a 4-12% Bis-Tris gel and stained with Instant Blue. Usage of magnetic beads resulted again in a much more pronounced pulldown of proteins compared to the previously used agarose beads (Figure 4.28). Full-length H2A-HA and potential crosslink products were clearly observable. Therefore, the samples were digested and analyzed by MS.

The resulting overall percentage of identified MS2 spectra was 2.91% with a 14.89% share of identified MS2 spectra of SILAC clusters. Analysis of the MS raw files by MaxQuant showed the identification of 535 proteins from 5541 peptides (Figure 4.29). This is slightly higher than in the last experiment in which 4885 peptides led to the identification of 480 proteins. Although in both cases the magnetic beads were fully saturated, minor differences in the final concentrations of the samples may be responsible for an altered overall yield.

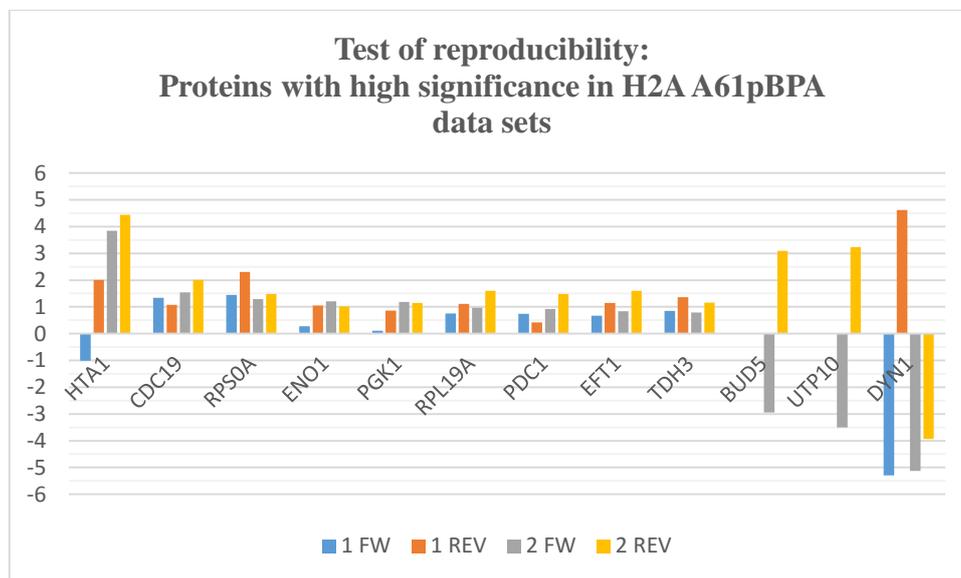
Comparison revealed a major overlap of 369 proteins between the preceding and the present experiment. The identification of the overlapping 369 proteins was accomplished with 3558 unique peptides with a mean of 9.642 unique peptides per protein. This matched approximately the number of 3386 unique peptides and a mean of 8.957 of the previous analysis. Again, the minor deviation might be caused by concentration differences. In order to quantify the degree of reproducibility and to establish a standard deviation for evaluation of potential crosslink products, a comparison between the SILAC ratios of all 369 overlapping

proteins of both experiments was performed. For the forward samples, a standard deviation of 0.229 was calculated, whereas a value of 0.274 was calculated for the reverse samples. This deviation has to be taken into account when analyzing the significance of reappearing candidates for crosslink products. Notably, further experiments have to be conducted in order to establish a measurement range for improved evaluation quality. To further study the reproducibility of the experiments, the current data set was screened for candidates which exhibit high significance values ( $p \leq 0.01$ ) in both forward and reverse samples and were also detectable in the prior data set. This group of highly abundant proteins should be expected to have similar H/L ratios if they display constant factors in the protein profile of the investigated site. Table 4.18 shows the result of this screening.

**Table 4.18:** Intersecting hits with high significance in both data sets of H2A A61pBPA SILAC analysis

Protein Name	Gene Name	SILAC ratio, normalized, FW/REV (4.3.9)		SILAC ratio, normalized, FW/REV (4.3.10)		$\Delta$ SILAC ratio FW/REV		Significance B (corr), FW/REV (4.3.9)		Significance B (corr), FW/REV (4.3.10)	
Histone H2A	HTA1	-1.007	2.009	3.844	4.435	4.851	2.42	0.009	5.83 E-14	3.44 E-102	4.29 E-63
Pyruvate kinase 1	CDC19	1.344	1.073	1.539	2.003	0.195	0.930	3.01 E-22	8.62 E-5	9.47 E-18	3.15 E-14
40S ribosomal protein S0-A	RPS0A	1.442	2.305	1.289	1.485	0.152	0.819	2.93 E-25	5.44 E-18	6.66 E-13	1.72 E-08
Enolase 1	ENO1	0.282	1.053	1.207	1.013	0.925	0.039	0.019	1.17 E-4	1.71 E-11	1.14 E-04
Phosphoglycerate kinase 1	PGK1	0.111	0.863	1.178	1.144	1.067	0.281	0.247	1.79 E-03	5.13 E-11	1.36 E-05
60S ribosomal protein L19-A	RPL19A	0.757	1.116	0.973	1.604	0.216	0.488	1.79 E-08	4.3 E-05	5.91 E-08	1.16 E-09
Pyruvate decarboxylase isozyme 1	PDC1	0.741	0.422	0.927	1.476	0.185	1.054	3.32 E-08	0.152	2.47 E-07	2.12 E-08
Elongation factor 2	EFT1	0.669	1.146	0.835	1.601	0.166	0.455	5.15 E-07	2.58 E-05	3.33 E-06	1.23 E-09
Glyceraldehyde-3-phosphate dehydrogenase 3	TDH3	0.844	1.358	0.786	1.164	0.057	0.194	4.45 E-10	5.20 E-07	1.22 E-05	9.66 E-06
Bud site selection protein 5	BUD5	-	-	-2.938	3.088	-	3.089	1	1	5.71 E-23	1.51 E-31
U3 small nucleolar RNA-associated protein 10	UTP10	-	-	-3.501	3.235	-	3.236	1	1	6.66 E-32	1.91 E-34
Dynein heavy chain	DYN1	-5.291	4.614	-5.126	-3.937	0.164	8.552	1.24 E-45	1.88 E-68	2.41 E-66	3.14 E-57

With the exceptions of BUD5 and UTP10 of which no SILAC peptide pairs were detected in the previous analysis, all proteins possess high significance in both experiments. However, although the majority displays similar ratios in both experiments, a number of outliers with significant changes in parameter values were observed (Figure 4.30). Notably, in these cases it has to be distinguished between changes in magnitude or valence. Furthermore, dynamic variation of experimental parameters (mixing, isotope incorporation and isotope impurity) and inter- and intra-individual variability of the cells have to be considered as potential causes for the observed changes.



**Figure 4.30:** Proteins with significance in both H2A A61pBPA data sets.

Overview of intersecting proteins with chromatin biological background in both data sets of H2A A61pBPA SILAC analysis. Normalized  $\log_2$  SILAC ratios from forward and reverse samples of both H2A A61pBPA SILAC analyses are plotted for each candidate. *Legend:* 1 FW/REV: Forward and reverse samples of first H2A A61pBPA SILAC analysis (4.3.9); 2 FW/REV: Forward and reverse samples of second H2A A61pBPA SILAC analysis (4.3.10).

The majority of the analyzed group (CDC19, RSP0A, ENO1, RPL19A, EFT1, THD3, PDC1, and PGK) showed steady tendencies and valences of their SILAC ratios, but with different magnitudes of change. A small difference was detected for THD3 with a change of 0.057 between the forward samples and 0.194 between the reverse samples. It was followed by EFT1 which differed by 0.166 in the forward samples and by 0.455 in the reverse samples. RPL19A exhibited a variation of 0.216 amongst the forwards samples and 0.488 between the reverse samples.

The remaining three of this subgroup showed a significant difference between two similar sample fractions. ENO1 revealed a difference of 0.925 between the forward samples, but only of 0.039 between the reverse samples. Also, RSP0A showed only a minor difference of 0.152 among the forward samples, but a major difference of 0.819 between the reverse samples. CDC19 displayed a difference of 0.195 between the forward samples, but differed by 0.930 between the reverse samples. A value of 0.185 indicated only a minor difference

---

between the forward samples for PDC1, opposed to a major deviation of 1.054 between the reverse samples. Lastly, PGK1 differed in a factor of 0.281 in the reverse samples, but 1.067 in the forward sample.

The remaining two of this group of overlapping proteins, DYN1 and histone H2A, showed the largest differences between both data sets. While DYN1 differed only by a factor of 0.164 among its forward samples, the difference between the reverse samples amounted to 8.552 which resulted in a change in H/L ratio of 4.614 to -3.937. Histone H2A exhibited the most dramatic change with a factor of 4.851 for the forward samples and 2.42 for the reverse samples. Notably, H2A also showed a change of valence between its forward samples.

The magnitude of these changes has to be evaluated by taking the calculated standard deviation into account as well as the dynamic range of gene expression in yeast. Yet, since most proteins were able to retain their tendencies and valence of their SILAC ratios, it can be assumed that reproducibility of results can be achieved for high abundant proteins. Under the assumption that crosslink products are generally low abundant, the degree of reproducibility needs to be determined for proteins with low intensity spectra and therefore poor ion statistics.

In order to do so and further compare the outcomes of both experiments, both data sets were screened for overlapping proteins with functions in chromatin structure and regulation which possibly represent trapped interactions of histone H2A and proteins of the chromatin landscape at position A61 (Table 4.19).

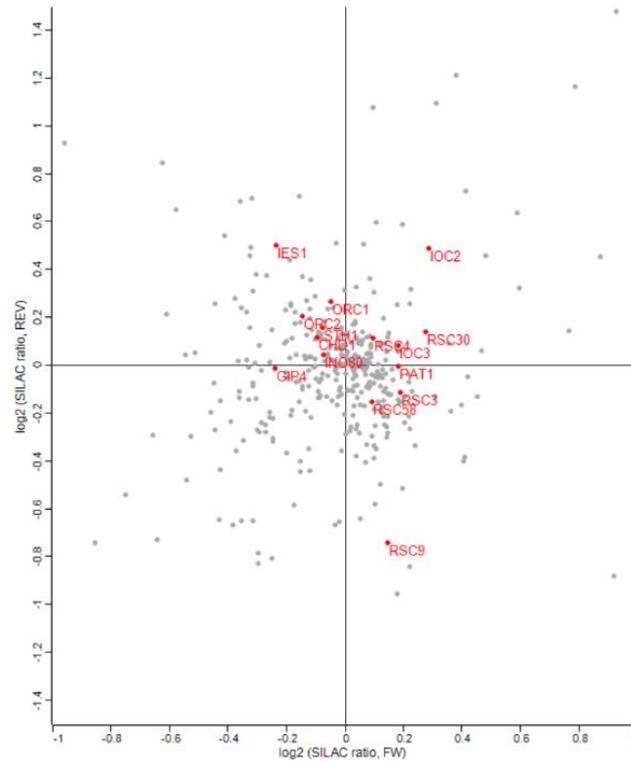
## 4 Results

**Table 4.19:** Intersecting hits with chromatin biological background in both data sets of H2A A61pBPA SILAC analysis

Protein Name	Gene Name	Fraction (total peptides), (4.3.9/4.3.10)		SILAC ratio, normalized, FW/REV (4.3.9/4.3.10)				Significance B (corr), FW/REV (4.3.9/4.3.10)			
Chromo domain-containing protein 1	CHD1	4 (16), 5 (39), 6 (14), 7 (10), 8 (2), 11 (2)	4 (13), 5 (102), 6 (41), 7 (41), 8 (25), 9 (9), 11 (7), 12 (1), 13 (2), 14 (1), 15 (1), 16(1)	0.015	0.066	-0.094	0.119	0.623	0.942	0.742	0.628
Histone-lysine N-methyltransferase, H3 lysine-79 specific	DOT1	10 (14), 11 (9)	10 (2), 11 (2)	-0.053	0.212	-0.103	-	0.992	0.528	0.719	1
Chromatin modification-related protein EAF7	EAF7	10 (19), 11 (1)	10 (3)	-0.492	0.273	-0.413	-	0.236	0.387	0.161	1
GLC7-interacting protein 4	GIP4	8 (4)	8 (6), 10 (1)	-	-	-0.238	-0.012	1	1	0.416	0.989
Histone deacetylase HOS3	HOS3	2 (1), 19 (1), 20 (2), 21 (1), 22 (2), 23 (1)	9 (1), 15 (1)	-	-	-	-	1	1	1	1
INO80 subunit 1	IES1	8 (3), 9 (1)	8 (6)	-0.549	-	-0.236	0.500	0.180	1	0.420	0.054
Putative DNA helicase INO80	INO80	4 (14), 5 (27), 6 (3), 9 (2)	4 (15), 5 (56), 6 (20), 7 (11), 8 (10), 9 (6)	-0.201	0.085	-0.074	0.041	0.690	0.884	0.792	0.849
ISWI one complex protein 2	IOC2	7 (3)	7 (10)	-	0.364	0.285	0.487	1	0.224	0.114	0.061
ISWI one complex protein 3	IOC3	7 (21), 8 (4)	7 (50), 8 (23), 9 (3)	0.052	0.079	0.182	0.083	0.457	0.902	0.319	0.728
ISWI one complex protein 4	IOC4	10 (6)	1 (1), 10 (5)	-0.533	-0.362	0.363	-	0.195	0.013	0.044	1
Chromatin structure-remodeling complex protein RSC14	LDB7	18 (4)	18 (4)	-0.428	0.209	-0.310	-	0.311	0.536	0.292	1
Origin recognition complex subunit 1	ORC1	7 (4), 9 (3), 10 (1)	7 (12), 8 (1), 9 (5)	-0.155	0.719	-0.050	0.264	0.786	0.010	0.855	0.302
Origin recognition complex subunit 2	ORC2	1 (1), 9 (12)	5 (1), 8 (1), 9 (16)	-0.007	-0.012	-0.146	0.202	0.743	0.715	0.614	0.424

DNA topoisomerase 2-associated protein PAT1	PAT1	8 (10)	8 (8), 10 (1)	-0.029	-0.402	0.180	-0.003	0.861	0.006	0.322	0.982
Chromatin structure-remodeling complex protein RSC3	RSC3	7 (14)	7 (39), 8 (8)	-0.236	-0.068	0.189/ -0.111		0.623	0.486	0.299	0.676
Chromatin structure-remodeling complex protein RSC30	RSC30	7 (5), 8 (1)	7 (4), 8 (2)	-0.574	0.043	0.276/ 0.140		0.159	0.978	0.127	0.573
Chromatin structure-remodeling complex subunit RSC4	RSC4	9 (5)	9 (11)	-0.421	-1.523	0.0951/ 0.113		0.321	3.52 E-21	0.609	0.645
Chromatin structure-remodeling complex protein RSC58	RSC58	11 (56), 14 (2)	11 (78), 13 (1), 14 (7), 16 (4)	-0.002	-0.058	0.091/ -0.154		0.714	0.52	0.622	0.554
Chromatin structure-remodeling complex subunit RSC9	RSC9	11 (8)	10 (9), 11 (2)	0.132	0.165	0.147/ -0.741		0.193	0.651	0.421	0.002
Nuclear protein STH1/NPS1	STH1	5 (7)	5 (7)	0.231	0.017	-0.078/ 0.155		0.046	0.854	0.782	0.535

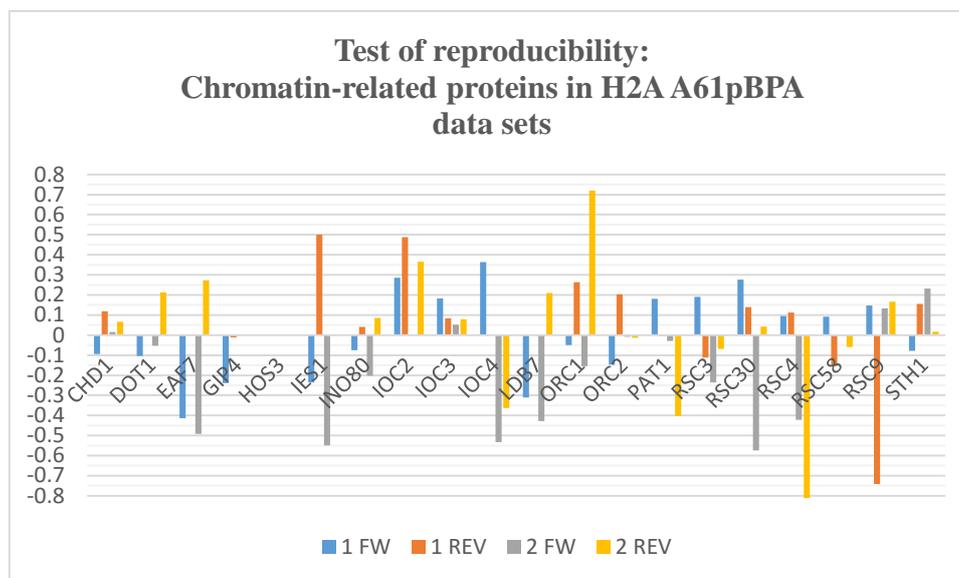
Strikingly, from the previously identified 22 chromatin-related proteins, the screening revealed 20 to be detectable in the present sample, including the known interaction partner STH1. Surveying the SILAC ratios, it became evident that all proteins exhibited, as observed in the previous analysis, very low ratios of light and heavy peptide pairs insufficient for quantification (Figure 4.31).



**Figure 4.31:** Identification of chromatin-related proteins from H2A A61pBPA SILAC analysis.

The  $\log_2$  SILAC (i.e., heavy/light L-arginine and/or L-lysine) ratios [E-1] calculated for chromatin-related proteins identified in the forward and reverse experiments of H2A A61pBPA immunoprecipitation with anti-HA magnetic beads are plotted against each other.

The tendency of SILAC ratio stayed constant for 4 proteins (HOS3, INO80, IOC3 and ORC1), but changed for 16 (CHD1, DOT1, EAF7, GIP4, IES1, IOC2, IOC4, LDB7, ORC2, PAT1, RSC3, RSC4, RSC9, RSC30, RSC58 and STH1) (Figure 4.32).



**Figure 4.32:** Chromatin-related proteins from H2A A61pBPA SILAC analysis

Overview of intersecting Chromatin-related hits in both data sets of H2A A61pBPA SILAC analysis. Normalized log<sub>2</sub> SILAC ratios from forward and reverse samples of both H2A A61pBPA SILAC analyses are plotted for each candidate.

*Legend:* 1 FW/REV: Forward and reverse samples of first H2A A61pBPA SILAC analysis. 2 FW/REV: Forward and reverse samples of second H2A A61pBPA SILAC analysis.

Notably, this included proteins which displayed a partial or complete change of quality, or changes in absolute detection but not necessarily tendency. Looking at proteins with no tendency change, HOS3 showed again no detection of peptide pairs, but also different fraction patterns. Therefore, HOS3 was no longer considered as potential crosslink product.

The H/L ratio of INO80 remained to be of divergent log<sub>2</sub> value, but with a 2-fold decrease in both forward and reverse samples. The tendency of detection in the unlabeled portion of the sample, indicated by change of valence, raised concerns about the probability of this crosslink product. Also, the degree of fractionation, although reproducible across both analyses, supported the doubt of proteinaceous background signal.

IOC3 displayed low, but positive SILAC ratios in both analyses. Still, the ratios of the forward samples showed a 3-fold increase in the present data. The fractions of discovery also coincided with the forerunner analysis: Although the degree of fractionation was slightly higher, the majority of peptides was similarly located. The sum of these observations speaks for IOC3 as potential crosslink.

ORC1 was earlier identified from multiple fractions which partially correlated with a 120 kDa crosslink product, but alternative detections produced doubts about the significance. This tendency was also observed

here. Also, the trend of detection in the unlabeled portion of the sample raised questions about the probability of this crosslink product.

Consequently, only IOC3 could potentially be a crosslink product out of the group of proteins with steady SILAC ratio tendencies.

Because of their ambiguous SILAC ratios, the remaining 16 proteins were only analyzed under reservations. Coming from different initial bases, IOC2, RSC4 and RSC30 exhibited positive  $\log_2$  ratios in both forward and reverse samples in the present experiment, which rendered them as potential crosslink products. However, IOC2 was only detectable in the forward sample of the prior analysis, whereas RSC4 was detectable in both, but with completely negative  $\log_2$  ratios. RSC30 displayed diverging ratios in both samples. All three proteins were detected in the same fractions in both experiments, respectively.

CHD1, RSC9 and STH1 were all candidates for crosslink products in the last analysis. They all displayed a negative  $\log_2$  in either the forward or the reverse sample in the present experiment. CHD1 showed a change of valence in the forward sample, but remained constant in the reverse sample while exhibiting a doubling of value. Analysis of the peptide distribution across fractions revealed a similar, but more wide-spread pattern. It remains to be clarified if this is the result of protein degradation due to the large size of the complex or evidence for false background noise. Taken together, a final statement about the probability of a CHD1 crosslink product cannot be made yet.

RSC9 showed comparable fraction patterns between both experiments, but a change of prefix in the reverse sample from a positive (0.165) to a negative (-0.741) SILAC ratio. The magnitude of this change was significant, whereas the H/L ratios in the forward sample remained stable (0.147 and 0.193). Therefore, the cause and its effect on the significance of this candidate need to be determined.

Known interaction partner of histone H2A at position 61, STH1 was again detected from peptides in the correct fraction according to the size of an estimated crosslink product with H2A. In the current analysis, its divergent SILAC ratios (-0.078/0.155) rather suggested this protein to be a negative hit. However, comparison with the prior SILAC ratios (0.231/0.017) showed a definite independence of protein abundance from labeling status in the forward sample, while the current ratios could suggest otherwise.

Contrarily, in the earlier experiment, IOC4, PAT1, RSC3 and RSC58 were categorized as unlikely to be crosslink products because of their negative SILAC ratios in both forward and reverse samples. However, in the current analysis, all four proteins exhibited positive  $\log_2$  ratios in the forward sample. PAT1, RSC3 and RSC58 still exhibited negative  $\log_2$  values in the reverse sample, while IOC4 lacked any detectable peptide pairs in the same. The main distribution of PAT1 peptides remained similar across experiments, but with a single outlier. Ratios changed from -0.029 in the forward sample and -0.402 in the reverse sample to 0.180 and -0.003, respectively. Both changes were significant, but more pronounced in the reverse sample. The forward ratio of RSC3 changed from -0.236 to 0.189, while the reverse ratio remained negative, though changing from -0.068 to -0.111. Fractionation remained comparable. The SILAC ratios of RSC58 exhibited change from -0.002 to 0.091 in the forward sample and from -0.058 to -0.154 in the reverse sample, while the present analysis revealed more peptide outliers, but a similar overall pattern.

Both IES1 and ORC possessed negative forward ratios and positive reverse ratios. While IES1 was detectable with a negative  $\log_2$  value only in the forward sample, ORC2 was identified in both forward and reverse samples with a negative ratio. The fractions of discovery remained the same for IES1, but changed slightly for ORC2 in terms of outliers. The current outcome suggests both proteins to be background noise.

DOT1, EAF7 and LDB7 displayed divergent ratios in forward and reverse samples in the previous analysis of H2A A61pBPA and therefore were categorized as potential background noise. In the present data set, all lacked detectable peptide pairs in the reverse sample. Although all three were identified from the same fractions as before, their steady negative SILAC ratios in the forward sample as well as their prior positive ratios in the reverse samples render them doubtful.

GIP4 showed no detectable SILAC pairs in the prior analysis, but appeared in the present. However, both ratios turned out to be of negative value, which renders it most likely background noise.

Collectively, comparison of the previous with the current data set revealed an overlap of 20 chromatin-related proteins, including the internal quality control STH1. All proteins exhibited very low peptide ion statistics which can be assumed to be caused by a general low abundance of peptides from crosslink products. While considering a general degree of value deviation by experimental parameters (variability of cell proteome, mixing, isotope incorporation), two proteins (IOC2 and IOC3) of the imitation-switch (ISWI) class of chromatin remodeling complexes and two proteins (RSC30 and STH1) of the Chromatin Structure Remodeling (RSC) complex were identified as potential crosslink products of histone H2A at position A61. Under the assumption of a mostly random sequencing of peptides which does not identify every available peptide in every experiment, low abundant peptides are statistically at a disadvantage causing a large quantity of values to be missing from an experiment, which may cause fluctuations in ion statistics. It remains to be determined in which magnitude these fluctuations can occur. This knowledge should allow consideration of IOC4, PAT1, RSC3, RSC4 and RSC58 as cases for crosslink candidates.

Overall, examination of two independent SILAC experiments of position A61 on histone H2A under identical experimental parameters revealed an intersection of 369 proteins between both data sets from comparable quantities of peptides. Taking into account a certain degree of setup-caused fluctuation in ion statistics, proteins of high abundance and significance showed a high degree of conservation and reproducibility.

The evaluation of potential crosslink products proved to be more complicated. In consequence of low abundant peptides, all potential crosslink products exhibited poor ion statistics and therefore low SILAC ratios, which made them harder to distinguish from the overall background noise. Alignment of both data sets produced a group of 20 coinciding proteins with functions in chromatin structure and regulation, of which a total of 9 may be crosslink candidates based on differences in magnitude or valence of their SILAC ratios. Notably, the majority of these proteins belonged to two distinct chromatin remodeling classes, ISWI and RSC, which supports the suggestion of successful capturing of binding partners of histone H2A within the chromatin structure of *S. cerevisiae*.

On the whole, these findings indicate that quantitative proteomics combined with *in vivo* crosslinking and SILAC immunoprecipitation possesses the effectiveness to identify site-specific interaction partners of histone pBPA mutants. However, the low abundance of the crosslink products which results in poor ion statistics and

therefore low SILAC ratios remains the bottleneck of this method. As long as this problem remains, a non-ambiguous statement about authenticity of crosslink products cannot be made. Therefore, future efforts should be directed towards increasing the crosslink product portion of the samples in order to produce indisputable data.

## 5 Discussion

This PhD project was aimed towards the development of a method to identify histone interaction partners in living yeast cells. This method was set to combine the genetic encoding of *p*-benzoyl-L-phenylalanine (pBPA) for *in vivo* crosslinking, SILAC immunoprecipitation and mass spectrometry for the objective and quantitative identification of protein-protein interactions in the chromatin landscape of *S. cerevisiae*.

Firstly, the genetic code expansion system was tested for effectiveness to analyze histone-protein interactions *in vivo*. Secondly, an immunoprecipitation protocol for the pulldown of HA-tagged histone pBPA mutants and their UV-induced crosslink products was established. Thirdly, the influence of SILAC conditions on crosslink formation was investigated and the IP protocol was accordingly adapted and optimized for optimal yield of crosslink products. Fourthly, SILAC analyses of captured histone-protein interactions on different histone positions were performed and evaluated for quality and credibility.

### 5.1 *In Vivo* Crosslinking of Histones in *S. cerevisiae*

The genetic code expansion system allows for a selective production of histone pBPA mutants when expressed in the presence of the unnatural amino acid pBPA and an evolved *E. coli* BPARS/tRNA<sub>CUA</sub> pair. This pair acts as orthogonal cellular component for the suppression of an amber stop codon. The amino acid specificity of the aaRS is altered so that it aminoacylates the evolved tRNA with a non-canonical amino acid only.

Notably, a series of studies have shown that natural amino acids can compete for the evolved synthetase that suppresses the amber codon when using the pBPA system. This leads to the mis-acylation of the suppressor tRNA and the production of protein with endogenous amino acids (Johnson et al., 2011; Chen et al., 2007; O'Donoghue et al., 2012; Heinemann et al., 2012; Odoi et al., 2013). Multiple reaction monitoring (MRM) by Chen and colleagues revealed a preference of the evolved *E. coli* BPARS to perform a mis-acylation with tryptophan and leucine at rates ranging from 0.6 to 4% compared to pBPA. Small amounts of uncharged and/or hydrophobic residues were also occasionally detected, but no charged amino acid (Chen et al., 2007). Chen and colleagues postulated that the increased misincorporation of tryptophan and leucine is the result of increased intracellular concentrations of both amino acids due to their usage as plasmid-borne auxotrophic markers for the expression of the BPARS/tRNA<sub>CUA</sub> plasmid (tryptophan) and the amber mutant plasmid (leu). Intracellular amino acid concentrations in yeast undergo cell cycle-dependent oscillations, and mean concentrations can range from 0.1 mM to 90 mM (Hans et al., 2002). Intracellular leucine concentrations change back and forth between 0.2 and 0.8 mM, whereas intracellular tryptophan concentrations oscillate between 0.1 and 0.4 mM.

Since the auxotrophic markers in the conducted experiments of Chen et al. (2007) and in this study are expressed from plasmids with high-copy 2 $\mu$  yeast replicative origins, it is highly likely that their concentrations will exceed natural intracellular levels. Therefore, single or combined amino acid concentrations will most probably be higher than intracellular pBPA levels of 1-2 mM. Because of this, in the current experimental setup a misincorporation of leucine caused by intracellular availability of pBPA is likely to occur. This seems to be a detrimental factor for the specificity of the BPARS and its selection. However, a misincorporation of leucine in context of *in vivo* crosslinking would merely result in the expression of a full-length histone incapable of

crosslinking, but detectable by Western blot. Due to the described low misincorporation rate of natural amino acids (Chen et al., 2007), a negative influence on the significance of the observed results by decreased crosslinking efficiency can be dismissed.

Previous research produced evidence that all UV-induced histone-protein crosslink products are genuine and represent trapped histone-protein interactions exclusively captured within the chromatin landscape of *S. cerevisiae* (Rall, 2012; Wilkins et al., 2014).

A steadily growing number of groups has successfully used this *in vivo* crosslinking system to study protein-protein interactions (Carvalho et al., 2010; Stanley et al., 2011; Mohibullah and Hahn, 2008; Schultz et al., 2011; Wilkins et al., 2014; Hoffmann and Neumann, 2014). To demonstrate the applicability of this approach for the analysis of nucleosomal interactions *in vivo*, pBPA was introduced into various sites spanning the surface of histones H3 and H4 (4.1.1). This revealed a site-specific and reproducible unique nature of crosslink patterns, which supports the specificity of this approach for the targeted investigation of intra- and internucleosomal interactions. Since the cells were not subjected to cell-cycle synchronization and hence the majority will statistically most likely occupy a rather open chromosome state, each sample represents a specific overview of trapped interactions on a particular residue throughout the cell cycle.

By using the antimetabolic agent nocodazole, cells were arrested in metaphase, which resulted in a time-resolved overview of mitotic interactions on the N-terminus of histone H4 (4.1.2). Wilkins et al. had tremendous success in elucidating the dynamics of mitotic chromosome condensation with a combination of *in vivo* crosslinking and nocodazole-mediated cell arrest (2014).

Consequently, the performed experiments established the pairing of genetic code expansion in yeast and the unnatural crosslinker amino acid pBPA as a highly specific and reliable tool for the targeted investigation of histone-protein interactions in the chromatin landscape of *S. cerevisiae*.

### **5.2 Isolation of Histone pBPA Mutants and Crosslink Products**

The purification and enrichment of crosslinks was a major concern of this PhD project. In the following, the procedure will be discussed and analyzed for further optimization.

Cell lysis was performed by fine rotor grinding of the cells in a liquid-nitrogen environment. In order to maximize the contact area of the cells with the high-speed rotor for optimal disruption, the cell pellets were resuspended in a minimum amount of lysis buffer. The soluble cytoplasmic fraction was subsequently removed from the insoluble nuclear fraction by centrifugation.

Assumably the majority of histone studies carried out so far made use of the solubility of histone proteins in dilute acidic conditions for their isolation. A serious objection against the use of acid for the extraction of histone crosslink products is the denaturing and degrading effect of acid to non-histone proteins (Spelsberg et al., 1973b). On another note, a C-C crosslink product which is formed by recombination of a diradical with a close substrate is unlikely to be destroyed during subsequent analysis. However, the presence of a hydroxyl group which is formed when the ketyl oxygen of pBPA abstracts a hydrogen renders it susceptible to acid-catalyzed dehydration (Egnaczyk et al., 2001). This made it seem appropriate to dissociate the nucleoprotein complex in a solution of high ionic strength, such as guanidine-hydrochloride or NaCl plus urea, to secure the

quality of the samples and help to solubilize the histones and their crosslink products by reducing aggregation (Arnold and Young, 1972; Gilmour and Paul, 1969; van den Broek et al., 1973). In the development phase of this purification protocol, only one guanidine-HCl extraction step was performed, which resulted in approximately a third of the total amount of histone amber mutants to remain in the insoluble pellet fraction. Addition of a second step and use of detergent significantly increased the overall preparation of histone pBPA protein and crosslinks.

Although the overall extraction yield of this approach may be lower compared to acid extraction, it was chosen for quality instead of quantity to ensure the purification and isolation of intact crosslink products for proteomic analysis.

To determine the yields of isolated histone pBPA mutants and crosslink products, pooled elution fractions were measured by Bradford test after affinity purification. The protein purification of forward and reverse samples, which consist each of 1 L light cell culture and 1 L heavy cell culture, yielded on average 2.1-3.5 mg of histone protein solution for H2A and H3 pBPA mutants, whereas H4 pBPA mutants produced only up to approximately 2 mg. Considering a 10% contribution of the genetic code expansion system using the pESC BPARS plasmid to the histone pool of the cells (Rall, 2012; Wilkins et al., 2014), a protein yield of around 210-350  $\mu$ g of histone pBPA protein and crosslink products can be calculated for histones H2A and H3, as well as 200  $\mu$ g for histone H4.

In general, overall yields of recombinant proteins with specific incorporated unnatural amino acids by an orthogonal aaRS/tRNA<sub>CUA</sub> pair have been reported to account for up to 50% of endogenous protein levels in optimized systems (Smolskaya et al., 2013). The pESC plasmid used in this study for the expression of the *E. coli* BPARS/tRNA<sub>CUA</sub> pair contains the coding sequences of the original plasmid developed by Chin et al. for the incorporation of pBPA (2003). However, this system can only provide basal expression levels of tRNA<sub>CUA</sub> since expression is dependent on a cryptic promoter on the plasmid or in the gene which is not further described (Chin et al., 2003; Chen et al., 2007).

Consequently, a combinatorial effect of aforementioned decreased pBPA availability and low expression levels of tRNA<sub>CUA</sub> limits the expression levels of histone pBPA mutants. Therefore, although the system can sufficiently provide cell growth and selection as well as produce enough protein for Western blot analysis, it prevents high yields of mutant protein.

Since higher protein yields were required for further analysis, several efforts were geared towards the production of greater yields of pBPA mutant protein. Increasing the availability of pBPA resulted only in a marginal increase in total yield, but also in accumulation of growth and expression defects at higher concentrations (4.2.2.3). An investigation of the incorporation efficiencies of various unnatural amino acids has shown that pBPA produces substantially smaller amounts of pBPA mutants, assumably because of reduced intracellular availability due to low solubility (Chen et al., 2007). This coincides with the observation of precipitating pBPA at higher concentrations (4.2.2.3).

Since low expression levels of amber suppressor tRNA were determined as cause for low protein yields and attributed to an inefficient promoter (Chin et al., 2003; Chen et al., 2007), the genetic code expansion system was optimized based on the work of Chen et al. (2007). In this, the originally used evolved *E. coli* tRNA<sub>CUA</sub>

gene, which contains an intact B box, was inserted within the flanking regions of the yeast suppressor tRNA gene *SUP4*. These flanking regions contain internal A and B boxes to allow higher tRNA<sub>CUA</sub> expression levels by providing an intact promoter.

Testing an array of vectors encoding multiple copies of this chimaera gene cassette revealed that one gene cassette copy in combination with 2 mM pBPA provided optimal yield of mutant protein while maintaining optimal growth (4.2.2.4). The amount of produced histone pBPA mutant protein from this combination was estimated to exceed the levels produced by the former system 5-fold. Considering the previously reported contribution of 10% of the former system to the endogenous histone pool (Rall, 2012; Wilkins et al., 2014), a 5-fold increase could be equated to a 50% contribution to endogenous levels. This level would be in line with other optimized genetic code expansion systems for improved protein yields (Smolskaya et al., 2013). Higher numbers of gene cassettes and higher concentrations of pBPA impaired cell growth and crosslink formation (Figure 4.8), most likely due to supraphysiological concentrations of expressed tRNA<sub>CUA</sub> compared to endogenous tRNAs.

Considering the average protein yields of histones H2A, H3 and H4 and the 5-fold increase in protein levels by the improved system, employing the optimized pESC BPARS-1xSUP4 tRNA<sub>CUA</sub> plasmid would allow estimation of an overall yield in mutant protein and crosslink products of approximately 1.05-1.75 mg for pBPA mutants of histones H2A and H3, and 1 mg for histone H4 from one 2 L SILAC sample.

Prediction of the crosslink efficiency cannot be performed in a straightforward manner since it strongly depends on the positioning of the crosslinker amino acids (Pavlova et al., 2013; Majmudar et al., 2009a; Majmudar et al., 2009b; own observations), adjacent amino acids which can influence the directionality of the crosslink formation (Wittelsberger et al., 2006), as well as strength of the UV-light source, distance between the lamp and the sample, and duration of irradiation. These variables lead to a variance of reported crosslink efficiencies for pBPA ranging from 50-100% (Chin et al., 2002; Dorman and Prestwich, 1994; Kauer et al., 1986; Hui et al., 2015; Shoelson et al., 1993). Under the assumption of an average crosslink efficiency of 50-80 % and a minimal total protein yield of 1 mg, crosslink products would account for 500-800 µg per SILAC experiment. Although these estimated numbers most likely have to be adjusted downwards based on an observed overabundance of free, unreacted histone pBPA mutants, they suggest that the purification protocol efficiently allows the isolation and enrichment of significant amounts of crosslink products.

Reasonable up-scaling of the experimental setup to maximize crosslink product yield would be by a factor of 2-2.5, therefore, up to 2.5 L of cell culture for both light and heavy sample fractions. However, it appears more practical to increase the amount of anti-HA magnetic beads used during immunoprecipitation in order to fully deplete the available histone pool for enhanced enrichment of crosslink products. Although both approaches would decrease the overall cost-efficiency per experiment, the latter could greatly improve the ion statistics for the identification of crosslink products.

An alternative purification procedure could also be considered, where the isolation of nuclei from cells following lyticase digestion would produce spheroblasts. These modified cells have weaker outer membranes and would allow for efficient lysis via mechanical disruption. However, considering 2 L of cell culture for a single SILAC experiment, large quantities of lyticase and a large scale homogenizer would be required to

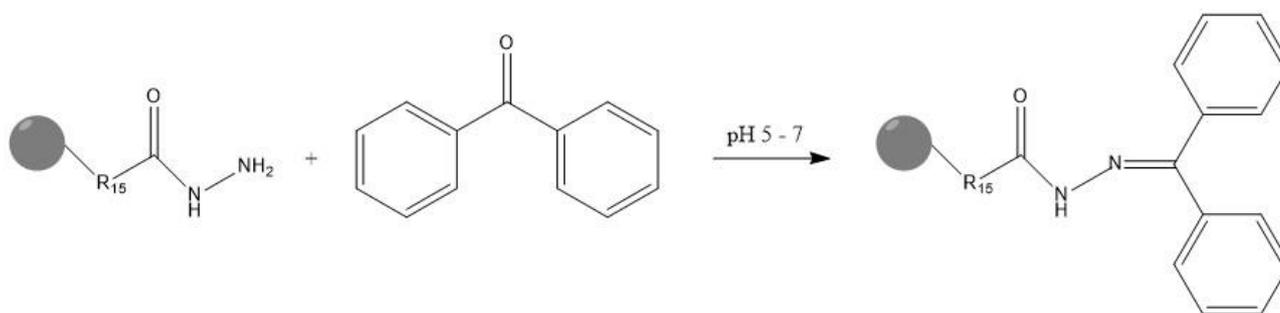
isolate nuclei for further immunoprecipitation. Still, this approach would greatly reduce the number of work steps.

However, a general obstacle for the successful enrichment of histone-HA crosslink products during immunoprecipitation was their displacement by an overabundance of free, non-crosslinked HA-tagged histones. The crosslink products represent only a small fraction of the total HA-tagged protein in the sample and therefore are outcompeted for binding by the large concentration of the crosslink-free histone. In an attempt to overcome this issue, two approaches were tested during this project for their effectiveness in separating the crosslink products from the crosslink-free histone fraction to improve the targeted isolation of crosslink products.

Histone acid extraction is based on the increased solubility of histones in dilute acid conditions which favour the precipitation of most other nuclear proteins and nucleic acids (Shechter et al., 2007). Based on this, it was expected that crosslinking of a histone to another protein with a different charge and size would reduce or diminish the solubility in acidic conditions and lead to precipitation of the crosslink products. This would then allow the removal of free histone-HA protein in the acid fraction and isolation and enrichment of the crosslink products from the pellet.

As described in chapter 4.3.7, enrichment of crosslink products by acid extraction was unsuccessful due to loss of crosslink products co-migrating with the unbound-histone fraction (Figure 4.21). Under acidic conditions, the crosslinked proteins did not prevent the histone-protein complexes from going into solution. Also upon success, it would have remained to be determined if the crosslink products would still be intact for MS analysis, due to the already mentioned potential degrading effect of acid to non-histone proteins (Spelsberg et al., 1973b).

Another approach to remove the overabundance of non-crosslinked histones from the samples was to utilize the chemistry of hydrazides. These chemical groups react with carbonyl groups (aldehydes or ketones) at pH 5-7 to form stable hydrazone linkages (Figure 5.1). Aldehydes (and ketones) do not readily exist in protein. However, under oxidative stress by reactive oxygen species (ROS), addition of carbonyl groups to protein side chains (especially Arg, Lys, Pro and Thr) can occur (Fedorova et al., 2014). Since our system introduces a unique keto group into our protein of interest, this chemistry can be exploited for the enrichment of crosslink products. When pBPA is crosslinked, its keto group is absent due to the radical recombination chemistry that occurs when it finds a reactive partner. In contrast, when the protein is not crosslinked to another protein, the keto moiety remains. It was assumed that since all proteins show a natural absence of ketones, hydrazide would react only with the free keto group of unreacted pBPA mutants. Using hydrazide-coupled isolation beads, the goal was to react all of the unreacted histones to the beads for an initial purification of these proteins from the solution before performing immunoprecipitation. Once this was achieved, all protein tagged with HA in the remaining solution should be crosslinked and therefore have little competition for the HA beads during the IP procedure. It was postulated that this would allow to increase the concentration of crosslink products for SILAC.



**Figure 5.1:** Hydrazide to benzophenone coupling reaction.

Reaction scheme for chemical conjugation of benzophenone group of pBPA to hydrazide-terminated magnetic beads.

The results indicated that the incubation of a crosslinked immunoprecipitation sample with hydrazide-terminated beads under acidic conditions (pH 6) led to coupling of crosslink products to the hydrazide group (4.3.7, Figure 4.22). One possible explanation for this could be the fact that besides aldehydes and ketones, carboxylic acids can also react to form an amide bond between the primary amine of the hydrazide group and the carboxyl group.

The direct coupling of carboxylic acid and an amine is highly favored due to the ready availability of these compounds (Mitchell and Reid, 1931). However, this reaction requires very high temperatures or activation of the acid by a coupling agent to overcome the thermodynamic barrier of salt formation. In this case, amide bond formation is highly unlikely to occur within the described experimental setup. However, there is a multitude of enzymes devoted to form amide bonds which reportedly also have been used for the synthesis of bioactive peptides and drug molecules (Wang, 2005). Enzymatic amide bond formation between carboxylic acids and amines can be performed by activation of the carboxyl group at the expense of ATP, leading to the formation of an acyl-adenylate or acylphosphate intermediate, before final amide formation. While aminoacyl-tRNA synthetases and non-ribosomal protein synthetases (NRPSs) hydrolyze ATP to AMP, members of the ATP-grasp family hydrolyze ATP to ADP (Goswani and van Lanen, 2015). While it can be speculated if the undesired pulldown of crosslink products is the result of enzymatic amide formation or of another nature, hydrazides remain to be disqualified for crosslink enrichment, although more effort should be invested in optimizing this promising approach. It appears worthwhile to repeat the experiment at different pH levels.

An alternative approach with similar bioconjugation characteristics for the enrichment of crosslink products would be the usage of hydroxylamine derivatives. These compounds can form stable oxime bonds with carbonyl groups under similar acidic conditions as hydrazone formation, but are hydrolytically more stable (Kalia and Raines, 2008).

Summing up the previous paragraphs, a protocol for the purification and isolation of intact histone pBPA mutants and their corresponding crosslink products from *S. cerevisiae* cells was successfully established to allow SILAC-based MS analysis for protein identification and quantification. However, the efficient enrichment of crosslink products still poses an obstacle which is of critical importance for the effective

improvement of the conducted SILAC analyses. Attempts to address this issue were without success, but require more refinement.

### 5.3 Identification of Crosslink Products by MS-based SILAC Approach

The SILAC experiments performed in this study were aimed towards the quantitative identification of specific interaction partners of histones captured by *in vivo* crosslinking in yeast with the genetically encoded crosslinker amino acid pBPA.

The basic principle of protein interaction partner identification by a SILAC-based approach is the distinction of specific from non-specific binding proteins based on the unbiased and definite ratio of heavy to light isotopes measured for each peptide.

MS analysis revealed a sufficient metabolic incorporation of isotope-labeled amino acids into proteins with an efficiency of 97% (4.3.1) and production of valid peptide pairs (4.3.2). Factors such as imperfections in mixing, isotope impurities or incomplete isotope incorporation can introduce systemic error and lead to a limitation of the dynamic range of detectable differences.

In the conducted experiments, the majority of proteins showed a low scatter of SILAC ratios around a factor of one which is the result of correct pairing of heavy and light peptides from evenly abundant proteins in both sample fractions. Also, a successful peptide pairing indicates the absence of arginine-to-proline conversion. Therefore, it can be deduced that neither mixing errors nor insufficient isotope incorporation or arginine-to-proline conversion have influenced the overall outcome of the conducted experiments.

In this project, mixing of light and heavy cell populations was performed based on OD<sub>600</sub> before cell lysis. However, mixing could also be performed at later points in time, which could reduce possible effects of inefficient cell lysis. One popular possibility is mixing after cell lysis based on protein concentrations of the individual lysates (Wang and Huang, 2008). An alternative approach is mixing of the protein solution after cell lysis and affinity purification (Kaake et al., 2010). Still, it should be noted that the later the mixing occurs and the more steps are performed beforehand, the higher the risk of introducing experimental variability which can be mistaken as biological variation.

Based on the performed crosslink overview across the surface of histones H3 and H4 (4.1.1), the SILAC analyses were expected to result in the identification of a multitude of genuine interaction partners. Disproving this expectation, the overall number of identified proteins per run was unexpectedly low, independent of histone variant or crosslinker position (4.3.3). Although chromatin-binding proteins like CHD1 and REG2 were identified as significantly enriched (4.3.2), they could neither be confirmed in mass shift assays (4.3.4), nor reproduced in an additional SILAC experiment (4.3.5).

Exchanging anti-HA agarose beads for anti-HA magnetic beads in immunoprecipitation resulted in a significantly increased number of overall identified proteins and significantly enriched proteins per SILAC experiment (4.3.9), as well as a high reproducibility of protein identification (4.3.10).

All conducted SILAC experiments exhibited an abundance of seemingly unrelated hits of protein classes compared to the investigated histone context, such as cytoskeletal and structural proteins, heat shock proteins, as well as glycolytic and ribosomal proteins. Although the majority of these hits exhibited SILAC ratios either

at or near the threshold, a reoccurring number showed significantly higher ratios, such as ENO1, SSA3, and THD3. However, either their detection was not in accordance with their actual localization (e.g., ENO1, cytoplasmic) or they could not be confirmed (BMS1, 4.3.4). Therefore, these proteins represent a population which binds non-specifically to the affinity matrix used.

More stringent washing regimes can only slightly improve the pulldown quality of the beads since it will also increase the risk of removing binding partners of low abundance or low affinity. Therefore, an optimal preservation of all specific interactions will inevitably lead to the co-purification of non-specific binding proteins.

Identification and removal of these proteins can be performed by a negative control database which provides information about non-specific binding of proteins to the affinity matrix used. With a quantitative MS approach, Trinkle-Mulcahy et al. identified populations of proteins binding non-specifically to affinity matrices such as agarose, sepharose or magnetic beads (2008). As already suspected during analysis of the SILAC samples, their data supports the classification of single proteins (e.g., ACT1, ENO1, CHD1, CDC19, DYN1, and THD3) or classes (e.g., nucleoporins, ribosomal proteins (RPL, RPS), heat shock proteins) as non-specific background. Notably, the cited dataset also names histones as non-specific background. Although this suggests the rejection of histones as positive controls for crosslinking within the chromatin landscape, it concurs with the observed strong deviation of SILAC ratios of histones.

A similar database derived from a negative control would allow the distinction between non-specific and specific binding proteins and would greatly facilitate the evaluation of future SILAC analyses. Therefore, a SILAC experiment should be performed with an HA-tagged wild-type histone. Although the majority of these proteins possess a SILAC ratio of 1, some exhibit significantly higher ratios. It is important to highlight these proteins as false positives, so that they can be considered as low priority for future analyses.

Notably, although their datasets reveal DNA topoisomerase I as well as eukaryotic initiation and translation elongation factor as non-specifically binding proteins, they do not contain homolog proteins of any kind of chromatin complex detected in the course of this project (RSC, ISWI, and INO80). Also, the authors state that based on their data magnetic beads produced generally lower background in nuclear extracts than agarose beads, which would support the capturing of more valid hits by using magnetic beads as observed during this project.

Validity of identified proteins is a major concern of this project since it aims to identify specific interaction partners of histones. The usage of magnetic beads resulted in the identification of a multitude of chromatin-binding proteins (4.3.9) which also were detected in large parts in a replicate experiment (4.3.10). However, all proteins exhibited insufficient SILAC ratios, which prevents unambiguous distinction of their specificity. It has been stated that the random detection and underrepresentation of low abundant peptides impairs the reproducibility of these hits between experiments (Kuster et al., 2005; Li et al., 2005). However, the repeated detection of the majority of these proteins indicates that they represent an inherent part of the sample. Although this also holds true for non-specific binding proteins, the confirmation of STH1 as valid interaction partner of H2A at position A61 (0) proves that such can be captured and identified by the described procedure. Still, for already discussed reasons, the quality of the obtained data alone is not sufficient to identify crosslink products.

To further strengthen the tendency and credibility of the obtained results, more identified proteins of the IOC or RSC family should be subjected to verification attempts.

The percentages of identified peptides in the MS2 spectra are comprehensively low which displays a general loss of information for each performed analysis. Generally, information is discarded in the course of an MS/MS analysis where peptides from the MS1 survey scan are not selected for fragmentation and MS2 analysis. Also, any MS2 spectra which are not confidently identified will be lost. One factor facilitating loss of information is a low signal-to-noise ratio which can be caused e.g. by low concentrations of peptides, low ionization efficiency or apparatus-related background noise. A low signal-to-noise ratio results in a hindered distinction of low abundant peptides from background noise, which eventually leads to a non-selection for further MS2 analysis (Käll and Vitek, 2011; Bantscheff et al., 2012). The quantification of low abundance proteins represents a challenge since the selection of precursor masses for MS/MS analysis is biased towards peptides of high abundance. Consequently, this results in a random detection and underrepresentation of low abundant peptides, which impairs the reproducibility of these hits between experiments and thus undermines the reliability of SILAC ratios (Kuster et al., 2005; Li et al., 2005). Therefore, the SILAC ratios of low abundant proteins, such as crosslink products with very low protein counts per sample, are not sufficient to unambiguously distinguish their specificity.

Notably, since none of the samples were subjected to cell cycle synchronization, they represent specific overviews of trapped interactions on a particular residue over the course of the whole cell cycle. Therefore, it can be assumed that the already low protein counts of trapped interactions underlie a sample-dependent fluctuation which could impair proper quantification even more. Treatment of cells with nocodazole would result in a comprehensive arrest of all cells in mitosis which would lead to an increased accumulation of mitosis-related crosslink products and higher protein counts per captured interaction. Therefore, this approach could aid in improved SILAC ratios for a number of chromatin-binding products which are, for instance, involved in chromatin condensation (Wilkins et al., 2014). Still, positions on histones for this approach must be chosen with care, since nocodazole arrest was observed to cause a massive loss of crosslink products due to increased chromatin condensation and therefore shielding of the histone tails from interactions (Figure 4.2). Selection of precursor ions from an MS1 survey scan can be performed according to predetermined rules of the mass spectrometer's operation "data-dependent acquisition" (DDA) which triggers acquisition of product ion spectra above a certain threshold based on the intensity of precursor ions detected in scan data (Stahl et al., 1996). Lowering the threshold would allow acquisition of lower-quality peptide spectra, which could lead to an increased identification of peptides and proteins. However, this could also increase the level of noise which could affect the detection of real peptides. Investigating the effect of different threshold values, Wong et al. observed an increased number of spectra acquired upon extremely low threshold values, without a corresponding increase of identifications. They suggest a combination of a mass spectrometer with high mass accuracy (e.g. Orbitrap) and an acquisition threshold set at or very closely below the noise level for an optimal gain of identifications while maintaining overall quality (Wong et al., 2009).

The newer generation of mass spectrometers allows the acquisition of highly increased amounts of MS2 spectra by increased scan rates and resolution. Still, a large portion of these spectra cannot be assigned to peptide

sequences and therefore remains unidentified for reasons such as presence of unknown protein sequences, possibility of chemical or post-translational modification, insufficient spectral quality for high-confidence matching, or a low signal-to-noise ratio arising from chemical or electrical noise (Elias et al., 2005; Wong et al., 2007). Kelstrup et al. assessed the influence of scan speed on the identification rate of proteins from a yeast proteome SILAC experiment (2012). Performing the MS/MS analysis on an Orbitrap tandem mass spectrometer with prolonged scan times, they were able to show a significantly improved sensitivity, mass accuracy and spectral quality, and based thereupon a higher peptide and protein identification rate from samples of low concentration.

Consequently, these studies have shown that the overall quality of the acquired data and based thereupon the number of identified proteins could be improved by fine tuning of the mass spectrometric analysis.

It is important to note that other factors such as sample loading amount or concentration may also affect the optimal instrument threshold, since the signal-to-noise ratio would depend on the sample concentration.

For better evaluation of the total yield of peptides from a single SILAC experiment for further optimization of the procedure, the peptide concentrations after trypsin digest should be measured. Classically, peptide concentrations can be estimated based on the presence of aromatic amino acids like tryptophan and tyrosine, and measurement of their fluorescence at 280 nm against a peptide sample of known concentration. Although all core histones lack tryptophan, they still contain tyrosine as prominent target of phosphorylation. Additionally, it is highly likely that caught interaction partners contain tryptophan. Therefore, this approach would represent a non-destructive method to minimize sample loss and to determine the amount of peptides after digestion. Alternatively, peptide concentrations can also be measured at 205 nm by which the absorbance arises primarily from the peptide backbone.

Evaluation of the identification process revealed that the majority of proteins was generally identified based on low numbers of unique peptides. It is generally recommended to require  $\geq 2$  unique peptides to be identified within a single protein for its positive identification (Carr et al., 2004). In a typical MS/MS analysis, a large population of proteins is identified only by a single peptide match. This may be caused by low concentration, few tryptic peptides in small proteins or missed capturing of reporter ions. Still, Washburn et al. reported the correct identification of abundant proteins with multiple peptides as well as low abundant proteins by one or two in yeast (2001). Therefore, omitting these hits will result in loss of information.

Quantification of peptides can also be conducted with other approaches. Label-free quantification employs spectral counting by measuring the frequency with which the peptide of interest has been sequenced by the mass spectrometer. It exploits the fact that the number of spectra for each peptide is proportional to the amount of protein in the sample. Application of this method for the determination of changes in protein abundance in complex biological samples has certain limitations. For instance, the measurement of small changes in the quantity of low abundant proteins can be difficult since it is often masked by sampling error.

Isotopic label approaches offer higher reliability and sensitivity than label-free quantification. An often discussed alternative to SILAC is Isobaric Tag for Relative and Absolute Quantification (iTRAQ) which uses isobaric tags that react with free primary amino groups at N-termini and  $\epsilon$ -amino groups of lysine side chains of peptides. Each label has an overall mass of 145 Da and is composed of a unique charged reporter group, a

neutral balance group and a reactive peptide group. When these peptides are fragmented during an MS2 scan, the reporter group breaks off and produces distinct ions whose  $m/z$  values depend on the label used. The isobaric nature of the tags also allows the comparison of the same peptides from each sample as a single peak in the mass spectrum, which reduces the complexity of the data as to light and heavy peptides in a SILAC experiment. The relative intensities of these reporter ions are directly proportional to the relative abundance of each peptide in the sample, which allows quantification (Ross et al., 2004).

Opposing the metabolic labeling of SILAC, iTRAQ relies on chemical labeling after trypsin digestion. The complexity of the samples is highly increased after digestion, which renders the late labeling step of this technique even more prone to variability and error. More importantly, iTRAQ faces the same problem of poor ion statistics and reduced reliability in regard to low abundant peptides as SILAC (Karp et al., 2010). Also, quantification of peptides can be impaired in the course of an iTRAQ analysis when two peptides have similar  $m/z$  values and cannot be properly resolved during precursor ion selection. The resulting MS2 spectrum will contain fragment ions and reporter ions from both peptides which may lead to the identification of one peptide, but leaving its iTRAQ ratio impaired by the second peptide's ratio (Ow et al., 2011).

Consequently, the increased chance of sample variability caused by late labeling as well as no expectable improvement towards detection of low abundant peptides renders iTRAQ incapable of further enhancing crosslink detection. Therefore, SILAC remains the superior approach for providing an unbiased and sensitive readout for peptide quantities as required for the successful identification of histone crosslink products.

Overall, the identification of *in vivo* crosslink products by SILAC immunoprecipitation is still in an early stage. The analysis and goal-directed optimization steps performed during this project succeeded in allowing acquisition of samples of steady quality, which permits reproduction of results. By repeatedly detecting a confirmed interaction partner, the developed approach proved its potential effectiveness for the detection and identification of crosslink products. However, poor SILAC ratios prevent the unambiguous distinction of crosslink specificities and therefore impair the credibility of potential interaction partners so far. In order to be fully capable of using the potential of this approach for the identification of histone-protein interaction partners, further work has been conducted to effectively increase the amounts of identified and quantified peptides per sample.

#### **5.4 Concluding Remarks and Outlook**

The function and activity of proteins is often modulated by other proteins they interact with. To understand cellular behaviour at the system level, a complete description of interactions among different proteins and the dynamic nature of these interactions is required.

Identification and quantification of proteins based on measured peptide intensities represents the highest standard for an unbiased and definite determination of protein interaction partners.

Inspired by the display of effectiveness of pBPA *in vivo* crosslinking in elucidating the mechanistic details of chromatin condensation in *S. cerevisiae* (Wilkins et al., 2014), the research objective of this PhD thesis was the development of a method for the quantitative identification of histone interaction partners for further decoding of chromatin structure dynamics.

First, the effectiveness of pBPA *in vivo* crosslinking for capturing histone-protein interactions in living yeast cells was successfully confirmed. Secondly, an immunoprecipitation protocol for pulldown of HA-tagged histone pBPA mutants and their UV-induced crosslink products was established and adapted to SILAC conditions. The targeted enrichment of crosslink products presented an obstacle during this project which could not satisfactorily be solved. Thirdly, initial SILAC analyses were performed and optimized until reproducible results of increased quality were obtained. Optimization led to the identification of numerous chromatin-related proteins, including a confirmed interaction partner. However, all candidate proteins suffered from poor ratios of light and heavy peptides, preventing distinction of their specificities.

During this project it became apparent that the concentration of crosslink products might represent the main bottleneck of this method, preventing sufficient quantification. Low protein counts per crosslink product result in low quantities of respective peptides, which increases the risk of them not being selected for MS/MS analysis.

Therefore, it is of considerable importance to increase the concentration of crosslink products per sample for an effective improvement of ion statistics.

First level of possible optimization could be the *in vivo* crosslinking approach. For *in vivo* interaction studies, pBPA is the most frequently used UV-inducible crosslinker amino acid. Although it inserts into C-H bonds within a 3.1 Å reactive radius (De Ruijter et al., 2003), it was reported that pBPA preferentially reacts with methionine's thioether sidechain even over distances beyond 3.1 Å (Wittelsberger et al., 2006). Consequently, the crosslinking efficiency of pBPA can be altered severely when being introduced in close proximity to a methionine, potentially allowing capturing of abnormal crosslinks. Also, pBPA was shown to produce substantially smaller amounts of mutant protein in comparison to other unnatural amino acids, assumably because of reduced intracellular availability due to low solubility (Chen et al., 2007).

In order to remove these potential interference factors, other crosslinker amino acids (Figure 1.8) with higher solubility and crosslink efficiency should be tested for increased overall yields of mutant protein and crosslink products. The aryl azide *p*-azido-L-phenylalanine (pAzF) has been reported to have a high incorporation efficiency, assumably by being less hydrophobic than pBPA, and produces significantly less steric hindrance due to smaller size (Chen et al., 2007; Chou et al., 2011). *p*-Trifluoromethyl-diaziranyl-L-phenylalanine (tmdF) possesses a higher degree of incorporation efficiency, too, as well as a higher crosslinking efficiency than pBPA at 365 nm (Hino et al., 2011). Both 3'-azibutyl-N-carbamoyl-lysine (ABK) and the pyrrolysine-derived 3-(3-methyl-3*H*-diazirine-3-yl)-propaminocarbonyl-Nε-L-lysine (DiZPK) have displayed higher incorporation efficiency, flexibility and crosslinking efficiency than pBPA (Ai et al., 2011; Chou et al., 2011; Hino et al., 2011; Zhang et al., 2011).

A major advantage of pBPA is its return to ground state after excitement with 365 nm light and lack of a reaction partner, as well as its capability of re-excitement. On the contrary, the photo-activation of ABK, pAzF and DiZPK is irreversible. Also, the applicability of pAzF in living cells can potentially be diminished by reduction of the azide to an amine in cellular environment (Staros et al., 1978). While each crosslinker amino acid has its inherent benefits and caveats, all can be considered as potent alternatives to pBPA in enhancing the total amount of crosslink products without increasing the metabolic burden of the cells.

Second level of optimization could be the purification process. As described as in chapter 5.2, an alternative lysis procedure could help to increase the overall yield of protein. More attention should be paid to selective purification strategies such as hydrazide or hydroxylamine chemistry for the enrichment of crosslink products to effectively improve IP pulldown efficiency. Also, scaling up of cell culture quantities or higher amounts of beads for immunoprecipitation should be tested.

A third level of optimization could target the MS/MS analysis. The MS/MS setup is biased to high abundant peptides and tends to neglect their low abundant counterparts. As described as in chapter 5.3, optimization should be performed on behalf of improved peptide detection. Here, prolonged scan intervals in the MS1 scan and an optimized acquisition threshold for MS2 spectra could yield improved peptide detection and better ion statistics for quantification.

Once the SILAC ratios are unambiguous, this method will open enormous possibilities for the investigation of protein-protein interactions *in vivo*. Combining the demonstrated power of *in vivo* crosslinking for large-scale capturing of interaction partners (4.1.1) with SILAC-based MS will allow the acquisition of vast amounts of interactome data with single amino acid resolution to effectively help to elucidate biological processes.

In regard to chromatin structure, highly detailed interactomes of histone surfaces could be generated and investigated for changes upon mutation or drug administration. Binding patterns of specific proteins could be mapped and aligned with other candidates to investigate signaling cascades. Also, the addition of directed cell phase arrest would give this approach a spatio-temporal resolution and would allow the investigation of binding characteristics over the course of the cell cycle.

Histone modifications have been described to have cell phase-dependent gradient-like features, e.g., the phosphorylation of H3 S10 is first observable in the pericentromeric heterochromatin during late G2 phase. It subsequently spreads over the chromosomal arms, is completed by prophase, and remains visible during metaphase (Hendzel et al., 1997). Combining cell phase arrest at distinct phases and the absence or presence of a non-phosphorylatable H3 S10A mutation could yield the identification of numerous H3 S10 phosphorylation-dependent interaction partners and allow the development of a conclusive regulation model. This approach would also be suited to probe other non-histone proteins (e.g., HST2), but also proteins out of chromatin context (e.g., importins, exportins).

## 6 Acknowledgements

First, I would like to thank Prof. Dr. Heinz Neumann for giving me the opportunity to accomplish this thesis in his workgroup and for his continuous support and supervision during my PhD studies.

I am grateful to Prof. Dr. Henning Urlaub and Prof. Dr. Matthias Dobbstein for support and agreeing to be my co-examiners. Furthermore, I want to thank Prof. Dr. Blanche Schwappach, Prof. Dr. Steven Johnsen and Prof. Dr. Andre Fischer for participating in my extended thesis committee.

A special thanks goes to Dr. Bryan Wilkins for countless brainstorming sessions, support in writing this thesis and being an awesome person. The contribution from Corinna Krüger is acknowledged, whose help was crucial for the verification of STH1.

I would like to acknowledge all my colleagues of the Applied Synthetic Biology workgroup for the wonderful time both in and out of the lab. Your company resulted in a truly unique and pleasurable working experience.

I am grateful to the staff of the Molecular Structural Biology department of Prof. Dr. Ralf Ficner for providing a well-functioning work environment. A special thanks goes to Marita Kalck for her administrative support.

I would like to acknowledge Dr. Kuan-Ting Pan, Dr. Samir Karaca and Annika Kühn of the Bioanalytical Mass Spectrometry department of Prof. Dr. Henning Urlaub for their help with processing and analyzing my SILAC samples.

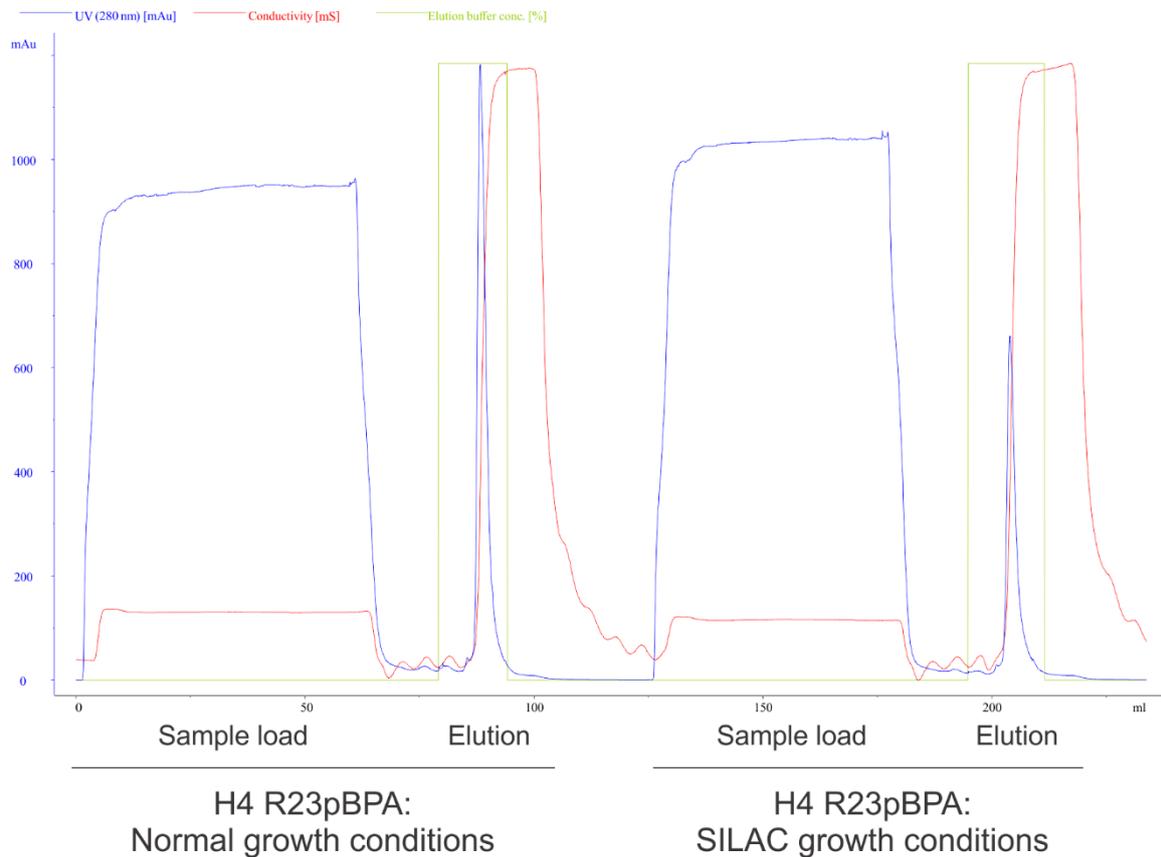
Further acknowledgements go to all my friends in Göttingen and all over the world for support and lending an ear.

I would like to thank my girlfriend Freya Gehrke for her support and invaluable help while writing this thesis.

Finally, I would like to thank my parents, Prof. Dr. Klaus Rall and Irene Rall, for their endless faith in me and their support beyond all description. Without you, I wouldn't be the person I am today.

## 7 Appendix

For better clarity, but without compromising the overall significance of the made observations, the results section of this thesis contains only a selected excerpt from the extracted MS protein identification files. Data files containing the complete lists of identified and quantified proteins were handed to the GGNB on a data carrier.



**Figure 7.1:** Immunoprecipitation of pBPA crosslink products under SILAC conditions.

The chromatogram shows the ion exchange affinity chromatography of H4 R23pBPA mutant protein and its crosslink products obtained from same amounts of YPH499  $\Delta$ Arg4 yeast cells grown either under normal, or restrictive SILAC growth conditions. Cell growth under SILAC growth conditions resulted in a significant decrease in total protein yields compared to normal growth conditions.

## 8 References

- Adam, M, Robert, F, Larochele, M and Gaudreau, L (2001) H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. *Mol Cell Biol.* 21, 6270–6279.
- Aebersold, R and Mann, M (2003) Mass spectrometry-based proteomics. *Nature.* 422(6928):198–207.
- Ai, H; Shen, W; Sagi, A, Chen, PR and Schultz, PG (2011) Probing protein-protein interactions with a genetically encoded photo-crosslinking amino acid. *Chembiochem.* 12, 1854–7.
- Ai, HW, Shen, W, Sagi, A, Chen, PR, Schultz, PG (2011) Probing protein-protein interactions with a genetically encoded photo-crosslinking amino acid. *Chembiochem.* 12, 1854–1857
- Aihara, H, Nakagawa, T, Yasui, K, Ohta, T, Hirose, S, Dhomae, N, Takio, K, Kaneko, M, Takeshima, Y, Muramatsu, M, Ito, T (2004) Nucleosomal histone kinase-1 phosphorylates H2A Thr 119 during mitosis in the early *Drosophila* embryo. *Genes Dev.* 18, 877–888.
- Andersen, JS, Lam, YW, Leung, AK, Ong, SE, Lyon, CE, Lamond, AI, Mann, M (2005) Nucleolar proteome dynamics. *Nature.* 433, 77–83.
- Angelov, D, Vitolo, JM, Mutskov, V, Dimitrov, S and Hayes, JJ (2001) Preferential interaction of the core histone tail domains with linker DNA. *Proc Natl Acad Sci U S A.* 98, 6599–6604.
- Arents, G, Burlingame, RW, Wang, BC, Love, WE. and Moudrianakis, EN (1991) The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci U S A.* 88, 10148–10152.
- Arnaudo, AM, and Garcia, BA (2013) Proteomic characterization of novel histone post-translational modifications. *Epigenetics Chromatin.* 6, 24.
- Arnold, EA and Young, KE (1972) Low molecular weight chromatin-associated RNA from rat liver. *Biochim Biophys Acta.* 269(2):252–8.
- Atkins, JF and Gesteland, R (2002) The 22nd amino acid. *Science.* 296, 1409–10.
- Balciunas, D, and Ronne, H (1999) Yeast genes GIS1-4: multicopy suppressors of the Gal<sup>-</sup> phenotype of *snf1 mig1 srb8/10/11* cells. *Mol. Gen. Genet.* 262:589–599.

- Bantscheff, M, Lemeer, S, Savitski, MM, Kuster, B (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem.* 404: 939–965.
- Bradford, M (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal Biochem.* 72:248–54.
- Baxter, J, Sen, N, Martínez, VL, De Carandini, ME, Schwartzman, JB, Diffley, JF, Aragón, L (2011) Positive supercoiling of mitotic DNA drives decatenation by topoisomerase II in eukaryotes. *Science.* 331, 1328–1332.
- Becker, PB (2002) Nucleosome sliding: facts and fiction. *EMBO J.* 21:4749–4753.
- Becker, PB and Horz, W (2002) ATP-dependent nucleosome remodeling. *Annu Rev Biochem.* 71, 247–273.
- Belmont, AS, Braunfeld, MB, Sedat, JW, Agard, DA (1989) Large-scale chromatin structural domains within mitotic and interphase chromosomes in vivo and in vitro. *Chromosoma.* 98:129–143,
- Blagoev, B, Kratchmarova, I, Ong, SE, Nielsen, M, Foster, LJ, Mann, M (2003) A proteomics strategy to elucidate functional protein–protein interactions applied to EGF signaling. *Nat. Biotechnol.* 21, 315–318.
- Blagoev, B., Ong, SE, Kratchmarova, I. and Mann, M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.* 22, 1139–1145.
- Blight, SK, Larue, RC, Mahapatra, A, Longstaff, DG, Chang, E, Zhao, G, Kang, PT, Green-Church, KB, Chan, MK, Krzycki, JA (2004) Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature.* 431, 333–5.
- Böck, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B., Zinoni, F (1991a) Selenocysteine: the 21st amino acid. *Mol. Microbiol.* 5, 515–20.
- Böck, A.; Forchhammer, K.; Heider, J. & Baron, C (1991b) Selenoprotein synthesis: an expansion of the genetic code. *Trends Biochem. Sci.* 16, 463–7.
- Böhm, V, Hieb, AR, Andrews, AJ, Gansen, A, Rocker, A, Tóth, K, Luger, K, and Langowski, J (2011) Nucleosome accessibility governed by the dimer/tetramer interface. *Nucleic Acids Research.* 39(8):3093–3102.
- Bottomley, MJ (2004) Structures of protein domains that create or recognize histone modifications. *EMBO Rep.* 5, 464–469.

- Buchenau, P, Saumweber, H, Arndt-Jovin, DJ (1993) Consequences of topoisomerase II inhibition in early embryogenesis of *Drosophila* revealed by in vivo confocal laser scanning microscopy. *J. Cell Sci.* 104, 1175–1185.
- Burgess, RJ and Zhang, Z (2013) Histone chaperones in nucleosome assembly and human disease. *Nature Structural & Molecular Biology.* 20(1):14–22.
- Bustin, M, Catez, F and Lim, JH (2005) The dynamics of histone H1 function in chromatin. *Mol Cell.* 17, 617–620.
- Carr, S, Aebersold, R, Baldwin, M, Burlingame, A, Clauser, K, Nesvizhskii, A (2004) The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol. Cell Proteom.* 3:531–533.
- Carvalho, P, Stanley, AM, and Rapoport, TA (2010) Retrotranslocation of a misfolded luminal ER protein by the ubiquitin-ligase Hrd1p. *Cell.* 143(4):579–591.
- Chen, P., Li, G (2010) Dynamics of the higher-order structure of chromatin. *Protein & Cell.* 967–971.
- Chen, S, Schultz, PG, Brock, A (2007) An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in *Saccharomyces cerevisiae*. *J Mol Biol.* 371:112–122.
- Chen, Y, Chen, W, Cobb, M.H, and Zhao, Y (2009) PTMap – a sequence alignment software for unrestricted, accurate, and full-spectrum identification of post-translational modification sites. *Proc. Natl. Acad. Sci. U. S. A.* 106, 761–6.
- Chen, Y, Sprung, R, Tang, Y, Ball, H, Sangras, B, Kim, SC, Falck, JR, Peng, J, Gu, W, and Zhao, Y (2007) Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol. Cell. Proteomics.* 6, 812–9.
- Cherry, JM, Ball, C, Weng, S, Juvik, G, Schmidt, R, Adler, C, Dunn, B, Dwight, S, Riles, L, Mortimer, RK and Botstein, D (1997) Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature.* 387, 67–73.
- Cheung, P, Allis, CD and Sassone-Corsi, P (2000) Signaling to chromatin through histone modifications. *Cell.* 103:263–271.
- Chin JW, Cropp TA, Anderson, JC, Mukherji, M, Zhang, Z, Schultz, PG (2003) An expanded eukaryotic genetic code. *Science.* 301:964–967.

- Chin, JW, Martin, AB, King, DS, Wang, L, Schultz, PG (2002) Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc.* 124, 9026–7.
- Chin, JW, Martin, AB, King, DS, Wang, L, Schultz, PG (2002a) Addition of a photocrosslinking amino acid to the genetic code of *Escherichiacoli*. *Proc Natl Acad Sci U S A.* 99:11020–11024.
- Chin, JW, Schultz, PG (2002b) In vivo photocrosslinking with unnatural amino acid mutagenesis. *Chembiochem.* 3(11):1135–7.
- Chou, C; Uprety, R; Davis, L; Chin, JW and Deiters, A (2011) Genetically encoding an aliphatic diazirine for protein photocrosslinking *Chem. Sci.* 2, 480.
- Christensen, MO, Larsen, MK, Barthelmes, HU, Hock, R, Andersen, CL, Kjeldsen, E, Knudsen, BR, Westergaard, O, Boege, F, Mielke, C (2002) Dynamics of human DNA topoisomerases IIalpha and IIbeta in living cells. *J. Cell Biol.* 157, 31–44.
- Clapier, CR, Cairns, BR (2009) The Biology of Chromatin Remodeling Complexes. *Annu Rev Biochem.* 78(1):273–304.
- Clark-Adams, CD, Norris, D, Osley, MA, Fassler, JS and Winston, F (1988) Changes in histone gene dosage alter transcription in yeast. *Genes Dev.* 2, 150–159.
- Clarke, DJ, O'Neill, LP and Turner, BM (1993) Selective use of H4 acetylation sites in the yeast *Saccharomyces cerevisiae*. *Biochem J.* 294 (Pt 2), 557–561.
- Cocklin, RR and Wang, M (2003) Identification of methylation and acetylation sites on mouse histone H3 using matrix-assisted laser desorption/ionization time-of-flight and nanoelectrospray ionization tandem mass spectrometry. *J Protein Chem.* 22, 327–334.
- Conrads, TP, Isaaq, HJ, and Veenstra, TD (2002) New tools for quantitative phosphoproteome analysis. *Biochem Biophys Res Commun.* 290(3): p. 885–90.
- Cox, J, Neuhauser, N, Michalski, A, Scheltema, RA, Olsen, JV, Mann, M (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res.* 10: 1794–805.
- Cox, J. and Mann, M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.* 26, 1367–72.

- Cross, SL and Smith, MM (1988) Comparison of the structure and cell cycle expression of mRNAs encoded by two histone H3-H4 loci in *Saccharomyces cerevisiae*. *Mol Cell Biol.* 8, 945–954.
- Cruickshank, MN, Besant, P, and Ulgiati, D (2010) The impact of histone posttranslational modifications on developmental gene regulation. *Amino Acids.* 39:1087–1105.
- Das, S and Oliver, DB (2011) Mapping of the SecA·SecY and SecA·SecE interfaces by site-directed in vivo photocross-linking. *J. Biol. Chem.* 286, 12371–80.
- De Frutos, M, Raspaud, E, Leforestier, A, Livolant, F (2001) Aggregation of nucleosomes by divalent cations. *Biophys. J.* 81:1127–1132.
- De Hoog, CL, Foster, LJ and Mann, M (2004) RNA and RNA binding proteins participate in early stages of cell spreading through spreading initiation centers. *Cell.* 117, 649–662.
- De Ruijter, AJM, Van Gennip, AH, Caron, HN, Kemp, S, Van Kuilenburg, ABP (2003) Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochemical Journal.* 370(3):737–749.
- Deiters, A, Groff, D, Ryu, Y, Xie, J, Schultz, PG (2006) A genetically encoded photocaged tyrosine. *Angew Chem Int Ed Engl.* 45: 2728–2731.
- Deiters, A, Schultz, PG (2005) In vivo incorporation of an alkyne into proteins in *Escherichia coli*. *Bioorg Med Chem Lett.* 15:1521–1524.
- Dekker, J (2008) Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. *J Biol Chem.* 283:34532–34540.
- Domon, B and Aebersold, R (2006) Mass spectrometry and protein analysis. *Science.* 312(5771):212–217.
- Dorigo, B, Schalch, T, Bystricky, K and Richmond, TJ (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, RR and Richmond, TJ (2004) Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science.* 306, 1571–1573.
- Dormán, G and Prestwich, PD (1994) Benzophenone photophores in biochemistry. *Biochemistry.* 33(19):5661–5673.

- Dorman, G and Prestwich, GD (1994) Benzophenone photophores in biochemistry. *Biochemistry*. 33,5661.
- Dubochet, J, Adrian, M, Chang, JJ, Homo, JC, Lepault, J, McDowell, AW, Schultz, P (1988) Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys*. 21:129–228.
- Dubochet, J, Adrian, M, Schultz, P, Oudet, P (1986) Cryo-electron microscopy of vitrified SV40 minichromosomes: the liquid drop model. *EMBO J*. 5:519–528.
- Dudoit S, van der Laan, MJ, Pollard KS (2004) Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Stat Appl Genet Mol Biol*. 3:Article13.
- Earnshaw, WC, Heck, MM (1985) Localization of topoisomerase II in mitotic chromosomes. *J Cell Biol* 100: 1716–1725. Earnshaw WC, Heck MM (1985) Localization of topoisomerase II in mitotic chromosomes. *J Cell Biol*. 100: 1716–1725.
- Earnshaw, WC, Honda, BM, Laskey, RA, and Thomas, JO (1980) Assembly of nucleosomes: the reaction involving *X. laevis* nucleoplasmin. *Cell*. 21373–383.
- Egnaczyk, GF, Greis, KD, Stimson, ER and Maggio, JE (2001) Photoaffinity cross-linking of Alzheimer's disease amyloid fibrils reveals interstrand contact regions between assembled beta-amyloid peptide subunits. *Biochemistry*. 40, 11706–11714.
- Fan, JY, Rangasamy, D, Luger, K, Tremethick, DJ (2004) H2A.Z alters the nucleosome surface to promote HP1 alpha-mediated chromatin fiber folding. *Mol. Cell*. 16:655–661.
- Finch, JT and Klug, A (1976) Solenoidal model for superstructure in chromatin. *Proc Natl Acad Sci U S A*. 73, 1897–1901.
- Fischle, W, Wang, Y and Allis, CD (2003) Histone and chromatin cross-talk. *Curr Opin Cell Biol*. 15, 172–183.
- Franklin, SG, and Zweidler, A (1977) Non-allelic variants of histones 2a, 2b and 3 in mammals. *Nature*. 266, 273–275.
- Freitas, MA, Sklenar, AR and Parthun, MR (2004) Application of mass spectrometry to the identification and quantification of histone post-translational modifications. *J Cell Biochem*. 92, 691–700.

- Fussner, E, Ching, RW, Bazett-Jones, DP (2011a) Living without 30 nm chromatin fibers. *Trends Biochem Sci.* 36:1–6.
- Fussner, E, Djuric, U, Strauss, M, Hotta, A, Perez-Iratxeta, C, Lanner, F, Dilworth, FJ, Ellis, J, Bazett-Jones, DP (2011b) Constitutive heterochromatin reorganization during somatic cell reprogramming. *EMBO J.* 30:1778–1789.
- Fussner, E, Strauss, M, Djuric, U, Li, R, Ahmed, K, Hart, M, Ellis, J, Bazett-Jones, DP (2012) Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. *EMBO Rep.* 13:992–996.
- Garbers, T (2016) Untersuchung von Interaktionen zwischen Histonen und Proteinen mittels genetisch kodierten Quervernetzern in vivo. Bachelor's thesis. Georg-August-University Göttingen, Germany.
- Gasser, SM (2002) Visualizing chromatin dynamics in interphase nuclei. *Science.* 296, 1412–1416.
- Gasser, SM, Laroche, T, Falquet, J, Boy de la Tour, E, Laemmli, UK (1986) Metaphase chromosome structure. Involvement of topoisomerase II. *J Mol Biol.* 188: 613–629.
- Gautier, A, Nguyen, DP, Lusic, H, An, W, Deiters, A, and Chin, JW (2010) Genetically encoded photocontrol of protein localization in mammalian cells. *J Am Chem Soc.* 132:4086–4088.
- Gavin, I, Horn, PJ and Peterson, CL (2001) SWI/SNF chromatin remodeling requires changes in DNA topology. *Mol Cell.* 7, 97–104.
- Gerlich, D, Hirota, T, Koch, B, Peters, JM, Ellenberg, J (2006) Condensin I stabilizes chromosomes mechanically through a dynamic interaction in live cells. *Curr Biol.* 16: 333–344.
- Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. *Nature.* 425:737–741.
- Gilmour, RS and Paul, J (1969) RNA transcribed from reconstituted nucleoprotein is similar to natural RNA. *J Mol Biol.* 40(1):137–9.
- Giménez-Abián, JF, Clarke, DJ, Devlin, J, Giménez-Abián, MI, De La Torre, C, Johnson, RT, Mullinger, AM and Downes, CS (2000) Premitotic chromosome individualization in mammalian cells depends on topoisomerase II activity. *Chromosoma.* 109, 235–244.

- Glover, DM, Leibowitz, MH, McLean, DA and Parry, H (1995) Mutations in aurora prevent centrosome separation leading to the formation of monopolar spindles. *Cell*. 81:95–105.
- Goswami, A and van Lanen, SG (2015) Enzymatic strategies and biocatalysts for amide bond formation: tricks of the trade outside of the ribosome. *Mol Biosyst*. 11(2):338–53.
- Grewal, SI and Moazed, D (2003) Heterochromatin and epigenetic control of gene expression. *Science*. 301, 798–802.
- Grigoryev, SA, Arya, G, Correll, S, Woodcock, CL, and Schlick, T (2009) Evidence for Heteromorphic Chromatin Fibers from Analysis of Nucleosome Interactions. *Proc. Natl. Acad. Sci. USA*. 106, 13317–13322.
- Gruhler, A, Olsen, JV, Mohammed, S, Mortensen, P, Faergeman, NJ, Mann M, et al. (2005) Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics*. 4, 310–327.
- Guillemette, B, Gaudreau, L (2006) Reuniting the contrasting functions of H2A.Z. *Biochem Cell Biol*. 84:528–535.
- Gurley, LR, D'Anna, JA, Barham, SS, Deaven, LL, Tobey, RA (1978) Histone phosphorylation and chromatin structure during mitosis in Chinese hamster cells. *Eur J Biochem*. 84: 1–15.
- Gutierrez, JL, Chandy, M, Carrozza, MJ and Workman, JL (2007) Activation domains drive nucleosome eviction by SWI/SNF. *Embo J*. 26, 730–740.
- Gygi, SP, Rist, B, Gerber, SA, Turecek, F, Gelb, MH, Aebersold, R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol*. 17(10): 994–9.
- Hager, GL, Nagaich, AK, Johnson, TA, Walker, DA and John, S (2004) Dynamics of nuclear receptor movement and transcription. *Biochim Biophys Acta*. 1677, 46–51.
- Hansen, JC (2002) Conformational dynamics of the chromatin fiber in solution: Determinants, mechanisms, and functions. *Annu Rev Biophys Biomol Struct*. 31: 361Y392.
- Hansen, JC, Tse, C and Wolffe, AP (1998) Structure and function of the core histone N-termini: more than meets the eye. *Biochemistry*. 37, 17637–17641.
- Hardy, CD, Crisona, NJ, Stone, MD, Cozzarelli, NR (2004) Disentangling DNA during replication: a tale of two strands. *Philos. Trans. R. Soc. Lond. B: Biol. Sci*. 359, 39–47.

- Hedbacker K, Carlson M (2008) SNF1/AMPK pathways in yeast. *Front Biosci.* 13: 2408–2420.
- Heinemann I.U., Rovner A.J., Aerni H.R., Rogulina S., Cheng L., Olds W., Fischer J.T., Soll D., Isaacs F.J., Rinehart J. (2012) Enhanced phosphoserine insertion during *Escherichia coli* protein synthesis via partial UAG codon reassignment and release factor 1 deletion. *FEBS Letter.* 586: 3716–3722.
- Hendzel, MJ, Wie, Y, Mancini, MA, Van Hooser, A, Ranalli, T, Brinkley, BR, et al. (1997) Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma.* 106: 348–360.
- Henikoff, S, Ahmad, K (2005) Assembly of variant histones into chromatin. *Annu Rev Cell Dev Biol.* 21: 133–53.
- Hentschel, CC, Birnstiel, ML (1981) The organization and expression of histone gene families. *Cell* 25, 301–313.
- Hereford, L, Fahrner, K, Woolford, J, Rosbash, M and Kaback, DB (1979) Isolation of yeast histone genes H2A and H2B. *Cell.* 18, 1261–1271.
- Hino, N, Oyama, M, Sato, A, Mukai, T, Iraha, F, Hayashi, A, Kozuka-Hata, H, Yamamoyo, T, Yokoyama, S, Sakamoto, K (2011) Genetic incorporation of a photo-crosslinkable amino acid reveals novel protein complexes with GRB2 in mammalian cells *Journal of Molecular Biology.* 406(2): 343–353.
- Hirano, T and Mitchison, TJ (1993) Topoisomerase II does not play a scaffolding role in the organization of mitotic chromosomes assembled in *Xenopus* egg extracts. *J. Cell Biol.* 120, 601–612.
- Hirano, T (2012) Condensins: universal organizers of chromosomes with diverse functions. *Genes Dev.* 26:1659–1678.
- Hoffmann, C and Neumann, H (2015) In vivo mapping of FACT-histone interactions identifies a role of Pcb3 C-terminus in H2A-H2B binding. *ACS Chem. Biol.*
- Holm, C, Goto, T, Wang, JC, Botstein, D (1985) DNA topoisomerase II is required at the time of mitosis in yeast. *Cell.* 41, 553–563.
- Horn, PJ, Peterson, CL (2002) Molecular biology. Chromatin higher order folding—wrapping up transcription. *Science.* 297:1824–1827.

- Horowitz-Scherer, RA and Woodcock, CL (2006) Organization of interphase chromatin. *Chromosoma*. 115, 1–14.
- Houben, A, Demidov, D, Caperta, AD, Karimi, R, Agueci, F and Vlasenko, L (2007) Phosphorylation of histone H3 in plants – a dynamic affair. *Biochim Biophys Acta*. 1769:308–315.
- Hudson, DF, Marshall, KM, and Earnshaw, WC (2009) Condensin: Architect of mitotic chromosomes. *Chromosome Res*. 17:131–144.
- Hudson, DF, Vagnarelli, P, Gassmann, R, Earnshaw, WC (2003) Condensin is required for nonhistone protein assembly and structural integrity of vertebrate mitotic chromosomes. *Dev Cell*. 5: 323–336.
- Huh, WK, Falvo, JV, Gerke, LC, Carroll, AS, Howson, RW, Weissman, JS, O'Shea, EK (2003) Global analysis of protein localization in budding yeast. *Nature*. 425:686–691.
- Hui, JZ, Al Zaki, A, Cheng, Z, Popik, V, Zhang, H, LuningPrak, ET, Tsourkas, A (2014) Facile Method for the Site-Specific, Covalent Attachment of Full-Length IgG onto Nanoparticles. *Small*, 10(16): 3354–3363.
- Ibarrola, N, Kalume, DE, Gronborg, M, Iwahori, A and Pandey, A (2003) A proteomic approach for quantitation of phosphorylation using stable isotope labeling in cell culture. *Anal. Chem*. 75, 6043–6049.
- Iizuka, M and Smith, MM (2003) Functional consequences of histone modifications. *Curr Opin Genet Dev*. 13, 154–160.
- Isaacs F.J., Carr P.A., Wang H.H., Lajoie M.J., Sterling B., Kraal L., Tolonen A.C., Gianoulis T.A., Goodman D.B., Reppas N.B., et al. (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science*. 333:348–353.
- Ishimi, Y, Hirosumi, J, Sato, W, Sugasawa, K, Yokota, S, Hanaoka, F and Yamada, M (1984) Purification and initial characterization of a protein which facilitates assembly of nucleosome-like structure from mammalian cells. *Eur. J. Biochem*. 142:431–439.
- Jackson, JD, Falciano, VT and Gorovsky, MA (1996) A likely histone H2A.F/Z variant in *Saccharomyces cerevisiae*. *Trends Biochem Sci*. 21, 466–467.
- Jackson, JD, Gorovsky, MA (2000). Histone H2A.Z has a conserved function that is distinct from that of the major H2A sequence variants. *Nucleic Acids Res*. 28:3811–3816.

- Jenuwein, T, Allis CD (2001) Translating the histone code. *Science*. 293:1074–1080.
- Jiang, H and English, AM (2002) Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine. *J. Proteome Res.* 1, 345–350.
- Jiang, T, Zhou, X, Taghizadeh, K, Dong, M, and Dedon, PC (2007) N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* 104, 60–5.
- Johansen, KM, and J. Johansen (2006) Regulation of chromatin structure by histone H3S10 phosphorylation. *Chromosome Res.* 14:393–404.
- Johnson, DB, Xu, J, Shen, Z, Takimoto, JK, Schultz, MD, Schmitz, RJ, Xiang, Z, Ecker, JR, Briggs, SP, Wang, L (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* 7:779–786.
- Johnson, CN, Adkins, NL and Georgel, P (2005) Chromatin remodeling complexes: ATP-dependent machines in action. *Biochem Cell Biol.* 83, 405–417.
- Kaake RM, Wang X, Huang L. (2010) Profiling of protein interaction networks of protein complexes using affinity purification and quantitative mass spectrometry. *Mol Cell Proteomics.* 9(8):1650–65.
- Kalia, J and Raines, RT (2008) Hydrolytic stability of hydrazones and oximes. *Angew Chem Int Ed Engl.* 47, 7523–6.
- Käll, L. and Vitek, O (2011) Computational mass spectrometry-based proteomics. *PLoS Comput. Biol.* 7, e1002277.
- Karp, NA, Huber, W, Sadowski, PG, Charles, PD, Hester, SV and Lilley, KS (2010) Addressing accuracy and precision issues in iTRAQ quantitation. *Mol Cell Proteomics.* Vol.9, No.9, pp.1885–97.
- Kauer, JC, Erickson-Viitanen, S., Wolfe, HR. and DeGrado, WF. (1986) p-Benzoyl-L-phenylalanine, a new photoreactive amino acid. Photolabeling of calmodulin with a synthetic calmodulin-binding peptide. *J. Biol. Chem.* 261, 10695.
- Kayne, PS, Kim, UJ, Han, M, Mullen, JR, Yoshizaki, F and Grunstein, M (1988) Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. *Cell.* 55, 27–39.

- Kelstrup, CD, Young, C, Lavalley, R, Nielsen, ML and Olsen, JV (2012) Optimized fast and sensitive acquisition methods for shotgun proteomics on a quadrupole orbitrap mass spectrometer. *J. Proteome Res.* 11, 3487–3497.
- Kerner, MJ, Naylor, DJ, Ishihama, Y, Maier, T, Chang, HC, et al. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell.* 122, 209–220.
- Khorasanizadeh, S (2004) The nucleosome: from genomic organization to genomic regulation. *Cell.* 116, 259–272.
- Kimura, H (2005) Histone dynamics in living cells revealed by photobleaching. *DNA Repair. Amst.* 4, 939–950.
- Kimura, H and Cook, PR (2001) Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *The Journal of Cell Biology.* 153(7):1341–1353.
- Kouskouti, A, Talianidis, I (2005) Histone modifications defining active genes persist after transcriptional and mitotic inactivation. *EMBO J.* 24:347–357.
- Kouzarides, T (2007) Chromatin modifications and their function. *Cell.* 128, 693–705.
- Kruithof, M, Chien, FT, Routh, A, Logie, C, Rhodes, D, van Noort, J (2009) Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber. *Nat Struct Mol Biol.* 16: 534–540.
- Krzycki, JA (2005) The direct genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* 8, 706–12.
- Kusmierz JJ, Sumrada R, and Desiderio DM. (1990) Fast atom bombardment mass spectrometric quantitative analysis of methionine-enkephalin in human pituitary tissues. *Anal Chem.* 62(21):2395–2400.
- Kuster, B, Schirle, M, Mallick, P, and Aebersold, R (2005) Scoring proteomes with proteotypic peptide probes. *Nat Rev Mol Cell Biol.* 6(7):577–583.
- Laemmli, UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* 227, 680–685.
- Laemmli, UK, Cheng, SM, Adolph, KW, Paulson, JR, Brown, JA, Baumbach, WR (1978) Metaphase chromosome structure: the role of nonhistone proteins. *Cold Spring Harb Symp Quant Biol.* 42:351–360.

- Lajoie M.J., Rovner A.J., Goodman D.B., Aerni H.R., Haimovich A.D., Kuznetsov G., Mercer J.A., Wang H.H., Carr P.A., Mosberg J.A., et al. (2013) Genomically recoded organisms expand biological functions. *Science*. 342:357–360.
- Lang, K, Davis, L, Torres-Kolbus, J, Chou, C, Deiters, A, Chin, JW (2012a) Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem*. 4: 298–304.
- Lang, K, Davis, L, Wallace, S, Mahesh, M, Cox, DJ, Blackman, ML, Fox, JM, Chin, JW (2012b) Genetic Encoding of Bicyclononynes and trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels-Alder Reactions. *J Am Chem Soc*. 134:10317–10320.
- Langst, G and Becker, PB (2004) Nucleosome remodeling: one mechanism, many phenomena? *Biochim Biophys Acta*. 1677, 58–63.
- Lavoie, BD, Hogan, E, Koshland, D (2002) In vivo dissection of the chromosome condensation machinery: reversibility of condensation distinguishes contributions of condensin and cohesin. *J Cell Biol*. 156: 805–815.
- Lemke, EA, Summerer, D, Geierstanger, BH, Brittain, SM, Schultz, PG (2007) Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat Chem Biol*. 3:769–772.
- Li, B, Pattenden, SG, Lee, D, Gutierrez, J, Chen, J, Seidel, C, Gerton, J, and Workman, JL (2005) Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling. *Proc. Natl. Acad. Sci. USA*. 102, 18385–18390.
- Li, G and Widom, J (2004) Nucleosomes facilitate their own invasion. *Nat Struct Mol Biol*. 11: 763–769.
- Li, X, Yi, EC, Kemp, CJ, Zhang, H, and Aebersold, R (2005) A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry. *Mol Cell Proteomics*. 4(9):1328–1340.
- Lill, J (2003) Proteomic tools for quantitation by mass spectrometry. *Mass Spectrom Rev*. 22(3):182–194.
- Listgarten, J and Emili, A (2005) Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. *Mol Cell Proteomics*. 4(4):419–434.
- Liu, CC and Schultz, PG (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem*. 79, 413–44.

- Liu, H, Sadygov, RG and Yates, JR (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem.* 76(14): p. 4193–201.
- Lo, WS, Duggan, L, Emre, NC, Belotserkovskya, R, Lane, WS, Shiekhattar, R, and Berger, SL (2001) Snf1 - a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription. *Science.* 293:1142–1146.
- Lohr, D, Kovacic, RT, Van Holde, KE (1977) Quantitative analysis of the digestion of yeast chromatin by staphylococcal nuclease. *Biochemistry.* 16(3):463–471.
- Loomis, RJ, Naoe, Y, Parker, JB, Savic, V, Bozovsky, MR, Macfarlan, T, Manley, JL and Chakravarti, D (2009) Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. *Mol Cell.* 33:450–461.
- Lorch, Y, Zhang, M and Kornberg, RD (1999) Histone octamer transfer by a chromatin-remodeling complex. *Cell.* 96, 389–392.
- Lottspeicher, F and Zorbas, H. (1998) *Bioanalytik.* Spectrum Akademischer Verlag, Heidelberg.
- Loyola, A., Almouzni, G (2004) Histone chaperones, a supporting role in the limelight. *Biochim Biophys Acta.* 1677:3–11.
- Lucchini, R and Sogo, JM (1995) Replication of transcriptionally active chromatin. *Nature.* 374:276–280.
- Ludwig, C, Claassen, M, Schmidt, A, Aebersold, R (2012) Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. *Mol Cell Proteomics.* 11(3): p. M111 013987.
- Luger, K (2003) Structure and dynamic behavior of nucleosomes. *Curr. Opin. Genet. Dev.* 13:127–135.
- Luger, K (2006) Dynamic nucleosomes. *Chromosome Res.* 14, 5–16.
- Luger, K and Hansen, JC (2005) Nucleosome and chromatin fiber dynamics. *Curr Opin Struct Biol.* 15, 188–196.
- Luger, K., Mader, AW, Richmond, RK, Sargent, DF, Richmond, TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature.* 389:251–260.

- Luger, K., Richmond, TJ (1998) The histone tails of the nucleosome. *Curr Opin Genet. Dev.* 8:140–6.
- Maeshima, K, Hihara, S, Eltsov, M (2010a) Chromatin structure: does the 30-nm fibre exist in vivo? *Curr Opin Cell Biol.* 22:291–297.
- Maile, T, Kwoczynski, S, Katzenberger, RJ, Wassarman, DA, and Sauer, F (2004) TAF1 activates transcription by phosphorylation of serine 33 in histone H2B. *Science.* 304:1010–1014.
- Majmudar, CY, Lee, LW, Lancia, JK, Nwokoye, A, Wang, Q, Wands, AM, Wang, L, Mapp, AK (2009) Impact of nonnatural amino acid mutagenesis on the in vivo function and binding modes of a transcriptional activator. *J Am Chem Soc.* 131:14240–14242.
- Majmudar, CY, Wang, B, Lum, JK, Hakansson, K, Mapp, AK. (2009) A high resolution interaction map of three transcriptional activation domains with a key coactivator from photo-cross-linking and multiplexed mass spectrometry. *Angew Chem Int Ed Engl.* 48:7021–7024.
- Margueron, R, Trojer, P and Reinberg, D (2005) The key to development: interpreting the histone code? *Curr Opin Genet Dev.* 15, 163–176.
- Marmorstein, R (2001) Protein modules that manipulate histone tails for chromatin regulation. *Nat Rev Mol Cell Biol.* 2, 422–432.
- Martin, C and Zhang, Y (2005) The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol.* 6, 838–849.
- Masumoto, H, Hawke, D, Kobayashi, R, Verreault, A (2005) A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature.* 436: 294–298.
- McBryant, SJ, Adams, VH and Hansen, JC (2006) Chromatin architectural proteins. *Chromosome Res.* 14, 39–51.
- McDonald, WH and Yates, JR (2003) Shotgun proteomics: integrating technologies to answer biological questions. *Curr Opin Mol Ther.* 5(3): 302–9.
- McGhee, JD, Nickol, JM, Felsenfeld, G and Rau, DC (1983) Higher order structure of chromatin: orientation of nucleosomes within the 30 nm chromatin solenoid is independent of species and spacer length. *Cell.* 33, 831–841.

- Mersfelder, EL, Parthun, MR (2006) The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. *Nucleic Acids Res.* 34: 2653–62.
- Mersfelder, EL and Parthun, MR (2006) The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. *Nucleic Acids Res.* 34, 2653–2662.
- Mizuguchi, G, Shen, X, Landry, J, Wu, WH, Sen, S (2004) ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science.* 303:343–348.
- Mohibullah, N and Hahn, S (2008) Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3. *Genes & Development.* 22(21):2994–3006.
- Mone, MJ, Bernas, T, Dinant, C, Goedvree, FA, Manders, EM, Volker, M, Houtsmuller, AB, Hoeijmakers, JH, Vermeulen, W and van Driel, R (2004) In vivo dynamics of chromatin-associated complex formation in mammalian nucleotide excision repair. *Proc Natl Acad Sci U S A.* 101, 15933–15937.
- Mori, H and Ito, K (2006) Different modes of SecY-SecA interactions revealed by sitedirected in vivo photo-cross-linking. *Proceedings of the National Academy of Sciences.* 103(44):16159–16164.
- Mukai, T, Hayashi, A, Iraha, F, Sato, A, Ohtake, K, Yokoyama, S, Sakamoto, K (2010) Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res.* 38:8188–8195.
- Munakata, T, Adachi, N, Yokoyama, N, Kuzuhara, T, and Horikoshi, M (2000) A human homologue of yeast anti-silencing factor has histone chaperone activity. *Genes Cells.* 5:221–233.
- Nakajima, M, Kumada, K, Hatakeyama, K, Noda, T, Peters, JM, Hirota, T (2007) The complete removal of cohesin from chromosome arms depends on separase. *J Cell Sci.* 120: 4188–4196.
- Neumann, H, Peak-Chew, SY, Chin, JW (2008) Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol.* 4:232–234
- Neumann, H, Hancock, SM, Buning, R, Routh, A, Chapman, L, Somers, J, Hughes, TO, van Noort, J, Rhodes, D, and Chin, JW (2009) A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Molecular Cell.* 36(1):153–163.
- Neumann, H. (2012) Rewiring translation – Genetic code expansion and its applications. *FEBS Letters.* 586(15):2057–2064.

- Nguyen, DP, Lusic, H, Neumann, H, Kapadnis, PB, Deiters, A, Chin, JW (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry. *J Am Chem Soc.* 131:8720–8721.
- Nirmalan, N, Sims, PF and Hyde, JE (2004) Quantitative proteomics of the human malaria parasite *Plasmodium falciparum* and its application to studies of development and inhibition. *Mol. Microbiol.* 52, 1187–1199.
- Nishino, Y, Eltsov, M, Joti, Y, Ito, K, Takata, H, Takahashi, Y, Hihara, S, Frangakis, AS, Imamoto, N, Ishikawa, T, Maeshima, K (2012) Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J.* 31:1644–1653.
- Norris, D and Osley, MA (1987) The two gene pairs encoding H2A and H2B play different roles in the *Saccharomyces cerevisiae* life cycle. *Mol Cell Biol.* 7, 3473–3481.
- Odoi, KA, Huang, Y, Rezenom, YH, Liu, WR (2013) Nonsense and sense suppression abilities of original and derivative *Methanosarcina mazei* pyrrolysyl-tRNA synthetase-tRNA(Pyl) pairs in the *Escherichia coli* BL21(DE3) cell strain. *PLoS One.* 8:e57035.
- O'Donoghue, P, Prat, L, Heinemann, IU, Ling, J, Odoi, K, Liu, WR, Söll, D (2012) Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNA<sup>Pyl</sup> for genetic code expansion. *FEBS Letters.* 586:3931–3937.
- Olsen, JV, Ong, SE and Mann, M (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics.* 3, 608–614.
- Ong, SE, Blagoev, B, Kratchmarova, I, Kristensen, DB, Steen, H, Pandey, A, Mann, M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics.* 1(5): p. 376–386.
- Ong, SE, Kratchmarova, I and Mann, M (2003) Properties of <sup>13</sup>C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.* 2, 173–181.
- Ong, SE, Mann, M (2006) A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat Protoc.* 1(6):2650–60.
- Ong, SE, Mittler, G and Mann, M (2004) Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods.* 1, 119–126.

- Osley, MA (1991) The regulation of histone synthesis in the cell cycle. *Annu Rev Biochem.* 60, 827–861.
- Ow, SY, Salim, M, Noirel, J, Evans, C and Wright, PC (2011) Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and highresolution HILIC fractionation. *Proteomics.* Vol.11, No.11, pp. 2341–6.
- Park, HS, Hohn, MJ, Umehara, T, Guo, LT, Osborne, EM, Benner, J, Noren, CJ, Rinehart, J, Soll, D (2011) Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science.* 333:1151–1154.
- Patzlaff, JS, Terrenoire, E, Turner, BM, Earnshaw, WC, Paulson, JR (2010) Acetylation of core histones in response to HDAC inhibitors is diminished in mitotic HeLa cells. *Experimental Cell Research.* 316: 2123–2135.
- Paulson, JR, Laemmli, UK (1977) The structure of histone-depleted metaphase chromosomes. *Cell.* 12:817–828.
- Paulson, JR. and Taylor, SS (1982) Phosphorylation of histones 1 and 3 and nonhistone high mobility group 14 by an endogenous kinase in HeLa metaphase chromosomes. *J. Biol. Chem.* 257, 6064–6072.
- Pavlova, O, Ieva, R, Bernstein, HD (2013) Monitoring the assembly of a secreted bacterial virulence factor using site-specific crosslinking. *JVisExp.*
- Peng, C, Lu, Z, Xie, Z, Cheng, Z, Chen, Y, Tan, M, Luo, H, Zhang, Y, He, W, Yang, K, Zwaans, BMM, Tishkoff, D, Ho, L, Lombard, D, He, TC, Dai, J, Verdin, E, Ye, Y, and Zhao, Y (2011) The first identification of lysine malonylation substrates and its regulatory enzyme. *Mol. Cell. Proteomics.* 10, M111.012658.
- Perez-Cadahia, B, Drobic, B, Davie, JR (2009) H3 phosphorylation: dual role in mitosis and interphase. *Biochem Cell Biol.* 87: 695–709.
- Peterson, CL and Laniel, MA (2004) Histones and histone modifications. *Curr Biol.* 14, R546–551.
- Phair, RD, Scaffidi, P, Elbi, C, Vecerova, J, Dey, A, Ozato, K, Brown, DT, 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.
- Phelan, ML, Schnitzler, GR and Kingston, RE (2000) Octamer transfer and creation of stably remodeled nucleosomes by human SWI-SNF and its isolated ATPases. *Mol Cell Biol.* 20, 6380–6389.

- Pinsky, BA, Nelson, CR, Biggins, S (2009) Protein phosphatase 1 regulates exit from the spindle checkpoint in budding yeast. *Curr Biol.* 19: 1182–1187.
- Plass, T, Milles, S, Koehler, C, Szymanski, J, Mueller, R, Wiessler, M, Schultz, C, Lemke, EA (2012) Amino acids for Diels-Alder reactions in living cells. *Angew Chem Int Ed Engl.* 51:4166–4170.
- Polycarpo, C, Ambrogelly, A, Berube, A, Winbush, SM, McCloskey, JA, Crain, PF, Wood, JL, Söll, D (2004) An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. *Proc. Natl. Acad. Sci. U. S. A.* 101, 12450–4.
- Preston, GW; Wilson, AJ (2013) Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chem. Soc. Rev.* 42, 3289–3301.
- Prigent, C, and Dimitrov, S (2003) Phosphorylation of serine 10 in histone H3, what for? *J Cell Sci.* 116:3677–3685.
- Ramakrishnan, V (1997) Histone structure and the organization of the nucleosome. *Annu Rev Biophys Biomol Struct.* 26, 83–112.
- Rall, NA (2012) In vivo investigation of the histone H4 tail and the H2A acidic patch interaction with genetically encoded crosslinkers. Master's thesis. Georg-August-University Göttingen, Germany.
- Ray-Gallet, D, Quivy, JP, Scamps, C, Martini, EMD, Lipinski, M, and Almouzni, G (2002) HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. *Molecular Cell.* 9(5):1091–1100, 2002.
- Robinson, PJ and Rhodes, D (2006) Structure of the '30nm' chromatin fibre: A key role for the linker histone. *Curr Opin Struct Biol.* 16, 336–343.
- Robinson, PJ, Fairall, L, Huynh, VA 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.
- Ross, PL, Huang, YN, Marchese, JN, Williamson, B, Parker, K, Hattan, S, Khainovski, N, Pillai, S, Dey, S, Daniels, S, Purkayastha, S, Juhasz, P, Martin, S, Bartlet-Jones, M, He, F, Jacobson, A, Pappin, DJ (2004) Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Molecular & Cellular Proteomics.* 3:1154–1169.

- Ross, PL, Huang, YN, Marchese, JN, Williamson, B, Parker, K, Hattan, S, et al. (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using aminereactive isobaric tagging reagents. *Molecular & Cellular Proteomics*. 3(12): p. 1154–1169.
- Routh, A, Sandin, S, Rhodes, D (2008) Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. *Proc Natl Acad Sci*. 105: 8872–8877.
- Saha, A, Wittmeyer, J and Cairns, BR (2005) Chromatin remodeling through directional DNA translocation from an internal nucleosomal site. *Nat Struct Mol Biol*. 12, 747–755.
- Saha, A, Wittmeyer, J and Cairns, BR (2006) Chromatin remodelling: the industrial revolution of DNA around histones. *Nat Rev Mol Cell Biol*. 7, 437–447.
- Saitoh, N, Goldberg, IG, Wood, ER, Earnshaw, WC (1994) ScII: an abundant chromosome scaffold protein is a member of a family of putative ATPases with an unusual predicted tertiary structure. *J Cell Biol*. 127: 303–318.
- Samejima, K, Samejima, I, Vagnarelli, P, Ogawa, H, Vargiu, G, Kelly, DA, de Lima Alves, F, Kerr, A, Green, LC, Hudson, DF, Ohta, S, Cooke, CA, Farr, CJ, Rappsilber, J, Earnshaw, WC (2012) Mitotic chromosomes are compacted laterally by KIF4 and condensin and axially by topoisomerase IIalpha. *J Cell Biol*. 199: 755–770.
- Sassone-Corsi, P, Mizzen, CA, Cheung, P, Crosio, C, Monaco, L, Jacquot, S, Hanauer, A, and Allis, CD (1999) Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3. *Science*. 285:886–891.
- Sawicka, A, Seiser, C (2014) Sensing core histone phosphorylation - a matter of perfect timing. *Biochim Biophys Acta*. 1839: 711–718.
- Schalch, T, Duda, S, Sargent, DF and Richmond, TJ (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*. 436, 138–141.
- Schalch, T, Duda, S., Sargent, DF, Richmond, TJ (2005) X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*. 436:138.
- Schlieker, C, Weibezahn, J, Patzelt, H, Tessarz, P, Strub, C, Zeth, K, Erbse, A, Schneider-Mergener, J, Chin, JW, Schultz, PG, Bukau, B, and Mogk, A (2004) Substrate recognition by the AAA+ chaperone ClpB. *Nature Structural & Molecular Biology*. 11(7):607–615.

- Schmidt, A, Beck, M, Malmström, J, Lam, H, Claassen, M, Campbell, D, Aebersold, R (2011) Absolute quantification of microbial proteomes at different states by directed mass spectrometry. *Mol Syst Biol.* 7: p. 510.
- Schmidt, A, Kellermann, J and Lottspeich, F (2005) A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics.* 5(1): 4–15.
- Schulz, C, Lytovchenko, O, Melin, J, Chacinska, A, Guiard, B, Neumann, P, Ficner, R, Jahn, O, Schmidt, B and Rehling, P (2011) Tim50's presequence receptor domain is essential for signal driven transport across the TIM23 complex. *The Journal of Cell Biology.* 195(4):643–656.
- Schulze, WX and Mann, M (2004) A novel proteomic screen for peptide–protein interactions. *J. Biol. Chem.* 279, 10756–10764.
- Sedat, J, Manuelidis, L (1978) A direct approach to the structure of eukaryotic chromosomes. *Cold Spring Harb Symp Quant Biol.* 42(Pt 1):331–350.
- Shamu, CE and Murray, AW (1992) Sister chromatid separation in frog egg extracts requires DNA topoisomerase II activity during anaphase. *J. Cell Biol.* 117, 921–934.
- Shechter, D, Dormann, HL, Allis, CD, Hake, SB (2007) Extraction, purification and analysis of histones. *Nat Protoc.* 2:1445–1457.
- Sherman, F (1998) An introduction to the genetics and molecular biology of the yeast *Saccharomyces cerevisiae*. *The Encyclopedia of Molecular Biology and Molecular Medicine.* 6:302–325.
- Shia, WJ, Pattenden, SG and Workman, JL (2006) Histone H4 lysine 16 acetylation breaks the genome's silence. *Genome Biol.* 7, 217.
- Shintomi, K and Hirano, T (2009) Releasing cohesin from chromosome arms in early mitosis: opposing actions of Wapl-Pds5 and Sgo1. *Genes Dev* 23: 2224–2236.
- Shoelson, SE, Sivaraja, M, Williams, KP, Hu, P, Schlessinger, J, and Weiss, MA (1993) Specific phosphopeptide binding regulates a conformational change in the PI 3-kinase SH2 domain associated with enzyme activation. *EMBO J.* 795–802.
- Shogren-Knaak, MA, Fry, CJ, and Peterson, CL (2003) A native peptide ligation strategy for deciphering nucleosomal histone modifications. *J Biol Chem.* 278:15744–15748.

- Shogren-Knaak, MA, Ishii, H., Sun, JM, Pazin, MJ, Davie, JR, Peterson, CL (2006) Histone H4- K16 acetylation controls chromatin structure and protein interactions. *Science*. 311:844–847.
- Smith, CM, Gafken, PR, Zhang, Z, Gottschling, DE, Smith, JB and Smith, DL (2003) Mass spectrometric quantification of acetylation at specific lysines within the amino-terminal tail of histone H4. *Anal Biochem*. 316, 23–33.
- Smith, MM and Andresson, OS (1983) DNA sequences of yeast H3 and H4 histone genes from two non-allelic gene sets encode identical H3 and H4 proteins. *J Mol Biol*. 169, 663–690.
- Smith, MM and Stirling, VB (1988) Histone H3 and H4 gene deletions in *Saccharomyces cerevisiae*. *J Cell Biol*. 106, 557–566.
- Smith, S and Stillman, B (2002) Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. *Cell*. 58(1):15–25.
- Smith, S, and B. Stillman (1989) Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. *Cell*. 58:15–25.
- Smolskaya, S, Zhang, ZJ, Alfonta, L (2013) Enhanced Yield of Recombinant Proteins with Site-Specifically Incorporated Unnatural Amino Acids Using a Cell-Free Expression System. *PLoS ONE*. 8(7): e68363.
- Sogo, JM, Stahl, H, Koller, Th, Knippers, R (1986) Structure of replicating simian virus 40 minichromosomes. *J. Mol. Biol*. 189: 189–204.
- Spelsberg, TC, Mitchell, WM, Chytil, F (1973) Structural alterations of acidic proteins by acid treatment of chromatin. *Molecular and Cellular Biochemistry*. Vol. 1, No. 2, 09.1973, p. 243–246.
- Srinivasan, G; James, CM and Krzycki, J (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science*. 296, 1459–62.
- Stahl, DC, Swiderek, KM, Davis, MT, Lee, TD (1996) Data-Controlled Automation of Liquid Chromatography/Tandem Mass Spectrometry Analysis of Peptide Mixtures. *J Am Soc Mass Spectrom*. 7:523–40.
- Stanley, AM, Carvalho, P, and Rapoport, T (2011) Recognition of an ERAD-L substrate analyzed by site-specific in vivo photocrosslinking. *FEBS Letters*. 585(9):1281–1286.

- Starich, MR, Sandman, K, Reeve, JN and Summers, MF (1996) NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. *J Mol Biol.* 255, 187–203.
- Staros, JV, Bayley, H, Standring, DN, and Knowles, JR (1978) Reduction of aryl azides by thiols: Implications for the use of photoaffinity reagents. *Biochemical and Biophysical Research Communications.* 80(3):568–572.
- Steen, H and Mann, M (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* 5, 699–711.
- Strahl, BD and Allis, CD (2000) The language of covalent histone modifications. *Nature*, 403, 41–45.
- Strunnikov, AV, Hogan, E, Koshland, D (1995) SMC2, a *Saccharomyces cerevisiae* gene essential for chromosome segregation and condensation, defines a subgroup within the SMC family. *Genes Dev.* 9: 587–599.
- Sullivan, SA and Landsman, D (2003) Characterization of sequence variability in nucleosome core histone folds. *Proteins.* 52, 454–465.
- Swedlow, JR and Hirano, T. (2003) The making of the mitotic chromosome: Modern insights into classical questions. *Mol Cell.* 11: 557–569.
- Swedlow, JR, Sedat, JW, Agard, DA (1993) Multiple chromosomal populations of topoisomerase II detected in vivo by time-lapse, three-dimensional wide-field microscopy. *Cell.* 73, 97–108.
- Tada, K, Susumu, H, Sakuno, T, Watanabe, Y (2011) Condensin association with histone H2A shapes mitotic chromosomes. *Nature.* 474(7352):477–483.
- Tagami, H, Ray-Gallet, D, Almouzni, G, and Nakatani, Y (2004) Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell.* 116(1):51–61, 2004.
- Tagami, S, Sekine, S, Kumarevel, T, Hino, N, Murayama, Y, Kamegamori, S, Yamamoto, M, Sakamoto, K, Yokoyama, S (2010) Crystal structure of bacterial RNA polymerase bound with a transcription inhibitor protein. *Nature.* 468, 978–82.
- Tamura, Y, Harada, Y, Shiota, T, Yamano, K, Watanabe, K, Yokota, M, Yamamoto, H, Sesaki, H, and Endo, T. (2009) Tim23–Tim50 pair coordinates functions of translocators and motor proteins in mitochondrial protein import. *The Journal of Cell Biology.* 184(1):129–141.

- Tan, M, Luo, H, Lee, S, Jin, F, Yang, JS, Montellier, E, Buchou, T, Cheng, Z, Rousseaux, S, Rajagopal, N, Lu, Z, Ye, Z, Zhu, Q, Wysocka, J, Ye, Y, Khochbin, S, Ren, B, and Zhao, Y (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*. 146, 1016–28.
- Tavormina, PA, Come, MG, Hudson, JR, Mo, YY, Beck, WT, et al. (2002) Rapid exchange of mammalian topoisomerase II alpha at kinetochores and chromosome arms in mitosis. *J. Cell Biol.* 158, 23–29.
- Teperek-Tkacz, M, Meglicki, M, Pasternak, M, Kubiak, JZ, and Borsuk, E (2010) Phosphorylation of histone H3 serine 10 in early mouse embryos: active phosphorylation at late S phase and differential effects of ZM447439 on first two embryonic mitoses. *Cell Cycle*. 9:4674–4687.
- Thoma, F, Koller, T and Klug, A (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol.* 83, 403–427.
- Thompson, A, Schäfer, J, Kuhn, K, Kienle, S, Schwarz, J, Schmidt, G, Neumann, T, Johnstone, R, Mohammed, AK, Hamon, C (2003) Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical Chemistry*. 75(8): p. 1895–1904.
- Tomschik, M., Zhengm H, van Holdem K, Zlatanovam J, Leubam SH (2015) Fast, long-range, reversible conformational fluctuations in nucleosomes revealed by single-pair fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U.S.A.* 102: 3278–3283.
- Trinkle-Mulcahy, L, Boulon, S, Lam, YW, Urcia, R, Boisvert, FM, Vandermoere, F, Morrice, NA, Swift, S, Rothbauer, U, Leonhardt, H, Lamond, A (2008) Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *The Journal of cell biology*. 183(2):223–39.
- Tsukiyama, T (2002) The in vivo functions of ATP-dependent chromatin-remodelling factors. *Nat Rev Mol Cell Biol.* 3, 422–429.
- Turner, BM (2002) Cellular memory and the histone code. *Cell*. 111, 285–291.
- Tweedie-Cullen, RY, Brunner, AM, Grossmann, J, Mohanna, S, Sichau, D, Nanni, P, Panse, C, and Mansuy, IM (2012) Identification of combinatorial patterns of post-translational modifications on individual histones in the mouse brain. *PLoS One*. 7, e36980.
- Uemura, T and Yanagida, M (1984) Isolation of type I and II DNA topoisomerase mutants from fission yeast: single and double mutants show different phenotypes in cell growth and chromatin organization. *EMBO. J.* 3, 1737–1744.

- Uemura, T, Ohkura, H, Adachi, Y, Morino, K, et al. (1987) DNA topoisomerase II is required for condensation and separation of mitotic chromosomes in *S. pombe*. *Cell*. 50, 917–925.
- Vagnarelli, P, Hudson, DF, Ribeiro, SA, Trinkle-Mulcahy, L, Spence, JM, Lai, F, Farr, CJ, Lamond, AI, Earnshaw, WC (2006) Condensin and Repo-Man-PP1 co-operate in the regulation of chromosome architecture during mitosis. *Nat Cell Biol*. 8: 1133–1142.
- Valls, E., Sanchez-Molinam, S, Martinez-Balbas, MA (2005) Role of histone modifications in marking and activating genes through mitosis. *J. Biol. Chem*. 280:42592–42600.
- Van Attikum, H and Gasser, SM (2005) The histone code at DNA breaks: a guide to repair? *Nat Rev Mol Cell Biol*. 6, 757–765.
- Van den Broek, HWJ, Noorden, LD, Sevall, JS, Bonner, J (1973) Isolation, purification and fractionation of nonhistone chromosomal proteins. *Biochemistry*. 12: 229–236.
- Van Holde, KE (1998) *Chromatin*. Springer-Verlag, New York.
- Van Hoof, D, Pinkse, MWH, Oostwaard, DWV, Mummery CJ, Heck, AJR, and Krijgsveld, J (2007) An experimental correction for arginine-to-proline conversion artifacts in silac-based quantitative proteomics. *Nat Methods*. 4(9):677–678.
- Van Hooser, A, Goodrich, DW, Allis, CD, Brinkley, BR, and Mancini, MA (1998) Histone H3 phosphorylation is required for the initiation, but not maintenance, of mammalian chromosome condensation. *J Cell Sci*. 111 (Pt 23):3497–3506.
- Vaquero, A, Scher, MB, Lee, DH, Sutton, A, Cheng, HL, Alt, FW, Serrano, L, Sternglanz, R, Reinberg, D (2006) SirT2 is a histone deacetylase with preference for histone H4 Lys 16 during mitosis. *Genes Dev*. 20: 1256-1261.
- Verreault, A (2000) De novo nucleosome assembly: new pieces in an old puzzle. *Genes Dev*. 14:1430–1438.
- Wako, T, Fukuda, M, Furushima-Shimogawara, R, Belyaev, ND, Fukui, K (2002) Cell cycle-dependent and lysine residue-specific dynamic changes of histone H4 acetylation in barley. *Plant Mol Biol*. 49: 645–653.
- Wako, T, Murakami, Y, Fukui, K (2005) Comprehensive analysis of dynamics of histone H4 acetylation in mitotic barley cells. *Genes Genet Syst* 80: 269–276.

- Wang X., Huang L (2008) Identifying dynamic interactors of protein complexes by quantitative mass spectrometry. *Mol. Cell. Proteomics* 7, 46–57.
- Wang, JC (2002) Cellular roles of DNA topoisomerases: a molecular perspective. *Nat. Rev. Mol. Cell Biol.* 3, 430–440.
- Wang, MX (2005) Enantioselective biotransformations of nitriles in organic synthesis. *Top. Catal.* 35: 117–130.
- Washburn, MP, Wolters, D, Yates, JR (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *3rd Nat. Biotechnol.* 19, 242–247.
- Weber, CM, Henikoff S (2014) Histone variants: Dynamic punctuation in transcription. *Genes Dev.* 28:672–682.
- Weier, HUG (2001) DNA fiber mapping techniques for the assembly of high-resolution physical maps. *J Histochem Cytochem* 49:939–948.
- Wilkins, BJ, Rall, NA, Ostwal, Y, Kruitwagen, T, Hirasami-Hamada, K, Winkler, M, Barral, Y, Fischle, W, Neumann, H (2014) A cascade of histone modifications induces chromatin condensation in mitosis. *Science.* 343:77–80.
- Williams, SP, Athey, BD, Muglia, LJ, Schappe, RS, Gough, AH and Langmore, JP (1986) Chromatin fibers are left-handed double helices with diameter and mass per unit length that depend on linker length. *Biophys J.* 49, 233–248.
- Wittelsberger, A, Thomas, BE, Mierke, DF, and Rosenblatt, M (2006) Methionine acts as a “magnet” in photoaffinity crosslinking experiments. *FEBS Letters.* 580(7):1872–1876.
- Witze, ES, Old, WM, Resing, KA, and Ahn, NG (2007) Mapping protein posttranslational modifications with mass spectrometry. *Nat Methods.* 4(10):798–806.
- Wong, CCL, Cociorva, D, Venable, JD, Xu, T, Yates, JR (2009) Comparison of different signal thresholds on data dependent sampling in orbitrap and LTQ mass spectrometry for the identification of peptides and proteins in complex mixtures. *J Am Soc Mass Spectrom.* 20:1405–1414.
- Woodcock, CL (1994) Chromatin fibers observed in situ in frozen hydrated sections. Native fiber diameter is not correlated with nucleosome repeat length. *J Cell Biol* 125:11–19.

- Woodcock, CL and Dimitrov, S (2001) Higher-order structure of chromatin and chromosomes. *Curr Opin Genet Dev*, 11, 130–135.
- Woodcock, CL, Frado, LL and Rattner, JB (1984) The higher-order structure of chromatin: evidence for a helical ribbon arrangement. *J Cell Biol*, 99, 42–52.
- Woodcock, CL. (2006) Chromatin architecture. *Curr Opin Struct Biol*. 16: 213–220.
- Worcel, A, Han, S, Wong, L (1978) Assembly of newly replicated chromatin. *Cell*. 15: 969–977.
- Worcel, A, Strogatz, S and Riley, D (1981) Structure of chromatin and the linking number of DNA. *Proc Natl Acad Sci U S A*. 78, 1461–1465.
- Xu, YZ, Kanagaratham, C and Radzioch, D (2013). Chromatin Remodelling During Host-Bacterial Pathogen Interaction, *Chromatin Remodelling*, Dr. Danuta Radzioch (Ed.), ISBN: 978-953-51-1087-3.
- Yamamoto, Y, Verma, UN, Prajapati, S, Kwak, YT, and Gaynor, RB (2003) Histone H3 phosphorylation by IKK-alpha is critical for cytokine-induced gene expression. *Nature*. 423:655–659.
- Yan, W and Chen, SS (2005) Mass spectrometry-based quantitative proteomic profiling. *Brief Funct Genomic Proteomic*, 4(1):27–38.
- Yang, D, Arya, G (2011) Structure and binding of the H4 histone tail and the effects of lysine 16 acetylation. *Phys. Chem. Chem. Phys*. 13:2911.
- Yao, X, Freas, A, Ramirez, J, Demirev, PA, Fenselau, C (2001) Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem*. 73(13): p. 2836–42.
- Zhang, K, Chen, Y, Zhang, Z, and Zhao, Y (2009) Identification and verification of lysine propionylation and butyrylation in yeast core histones using PTMap software. *J. Proteome Res*. 8, 900–6.
- Zhang, K, Tang, H, Huang, L, Blankenship, JW, Jones, PR, Xiang, F, Yau, PM and Burlingame, AL (2002a) Identification of acetylation and methylation sites of histone H3 from chicken erythrocytes by highaccuracy matrix-assisted laser desorption ionization-time-of-flight, matrix-assisted laser desorption ionization-postsource decay, and nanoelectrospray ionization tandem mass spectrometry. *Anal Biochem*, 306, 259–269.
- Zhang, L, Eugeni, EE, Parthun, MR and Freitas, MA (2003) Identification of novel histone post-translational modifications by peptide mass fingerprinting. *Chromosoma*, 112, 77–86.

Zhang, Lin, S, Song, X, Liu, J, Fu, Y, Ge, X, Fu, X, Chang, Z, Chen, PR (2011) A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nat. Chem. Biol.* 7: 671–677.

Zhou, H, Ranish, JA, Watts, JD, Aebersold, R (2002) Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat Biotechnol.* 20(5): p. 512–5.

Zimmerman ZA and Kellogg DR (2001) The Sda1 protein is required for passage through start. *Mol Biol Cell.* 12:201–219.

Zlatanova, J, Thakar, A (2008) H2A.Z: view from the top. *Structure.* 16:166–179.

Zofall, M, Persinger, J, Kassabov, SR and Bartholomew, B (2006) Chromatin remodeling by ISW2 and SWI/SNF requires DNA translocation inside the nucleosome. *Nat Struct Mol Biol,* 13, 339–346.

## 9 Curriculum Vitae

### Personal

Name: Nils Arne Rall  
Date and place of birth: 03.09.1986, Buchholz in der Nordheide, Germany  
Nationality: German  
E-Mail: nrall@gwdg.de

### Education

02/2013-03/2016      Doctoral thesis entitled „A Method for the Quantitative Analysis of Protein-Protein Interactions *In Vivo*”  
Supervisor Prof. Dr. H. Neumann, University of Göttingen

2010-2012      M.Sc. Microbiology and Biochemistry, University of Göttingen

2007-2010      B.Sc. Biology, University of Göttingen

06/2006      Allgemeine Hochschulreife (Abitur)

### Publication record

Wilkins, BJ, Rall, NA, Ostwal, Y, Kruitwagen, T, Hiragami-Hamada, K, Winkler, M, Barral, Y, Fischle, W, Neumann, H (2014) A cascade of histone modifications induces chromatin condensation in mitosis. *Science*. 343:77–80.