

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

**Epigenetisch relevante DNA-Modifikationen:
Identifikation von Interaktionspartnern mittels Crosslinker-basierten
pull down Assays und quantitativer Massenspektrometrie**

Silvia Katharina Laube
aus
Rosenheim, Deutschland

2017

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Thomas Carell betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 29.01.2017

Silvia Laube

Dissertation eingereicht am: 06.02.2017

1. Gutachter: *Prof. Dr. Thomas Carell*

2. Gutachter: *PD Dr. Stylianos Michalakis*

Mündliche Prüfung am: 20.03.2017

für meine Familie

"If passion drives you, let reason hold the reins."

- Benjamin Franklin

Aus dieser Arbeit entstandene Publikationen:

A. Perera, D. Eisen, M. Wagner, S. K. Laube, A. F. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell, S. Michalakis; *Cell Rep* **2015**, *11*, 1-12. *TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression*.

T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, U. Müller, H. Leonhardt, P. Schär, M. Müller and T. Carell; *Nat. Chem. Bio.* **2014**, *10*, 574-581. *Tet enzymes oxidize T to hmU in mouse embryonic stem cells*.

B. Steigenberger, S. Schiesser, B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell; *Organic Letters* **2013**, *15*, 366-369. *Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and their Incorporation into Oligonucleotides by Polymerase Chain Reaction*.

Konferenzbeiträge:

S. K. Laube, D. Eisen, B. Steigenberger, T. Carell; Posterpräsentation auf der 8th European Summer School (ESS) „Proteomics“ in Brixen, Italien (**2014**). *Investigation of Epigenetic Bases and their Biological Roles Using DNA-Protein Pulldown Assays and Covalent Trapping Followed by LC-MS-Analysis*.

J. Steinbacher, S. K. Laube, B. Steigenberger, S. Schiesser, T. Pfaffeneder, B. Hackner, T. Carell; Posterpräsentation auf der IMB Konferenz „DNA-Demethylation, Repair and Beyond“ in Mainz (**2012**). *Synthesis of the Triphosphates 5-Formylcytidine and 5-Carboxylcytidine and Incorporation into DNA-Strands by PCR*.

S. K. Laube, B. Steigenberger, B. Hackner, C. Brandmayer, J. Steinbacher, T. Carell; Posterpräsentation auf dem SFB 646 Meeting in München (**2012**). *Investigation of Demethylation Pathways by Photoaffinity Labeling and Activity Studies*.

Danksagung

An dieser Stelle möchte ich mich herzlichst bei all denjenigen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

Zuallererst bedanke ich mich bei meinem Doktorvater *Prof. Dr. Thomas Carell* für die Aufnahme in seine Gruppe, die Möglichkeit an hochspannenden und vielseitigen Themen unter großartigen Arbeitsbedingungen arbeiten zu können und seine stets motivierende Unterstützung.

Herrn *PD Dr. Stylianos Michalakis* danke ich für die Übernahme des Zweitgutachtens, die gute Zusammenarbeit in den gemeinsamen Projekten und das damit verbundene Engagement und große Interesse für meine Arbeit.

Den Mitgliedern der Prüfungskommission, *Prof. Dr. Rasmus Linser*, *Prof. Dr. Lena Daumann*, *Prof. Dr. Konstantin Karaghiosoff* und *PD Dr. Dietmar Martin* danke ich für ihre Mitwirkung an der mündlichen Prüfung.

Dr. Markus Müller, *Angie Kirchner* und *Dr. David Eisen* danke ich, dass sie für das Korrekturlesen meiner Doktorarbeit ihre Freizeit geopfert haben.

Frau *Slava Gärtner* danke ich für die zuverlässige Unterstützung bei allen organisatorischen und bürokratischen Belangen. Frau *Sabine Voß*, Frau *Kerstin Kurz* und Herrn *Kristof Hufnagel* danke ich für ihre Hilfe bei allen chemischen und labortechnischen Belangen.

Dr. Arshan Perera und *Anna Geserich* danke ich für die gute und unkomplizierte Kollaboration.

Bei *Dr. Felix Gnerlich*, *Dr. Barbara Steigenberger* bedanke ich mich für die Unterstützung bei den DNA-Synthesen und beim „Clicken“, bei *Dr. Fabio Spada* und *Angie Kirchner* bedanke ich mich dafür, dass sie mir immer genügend Probenmaterial zur Verfügung gestellt haben.

Ein besonderer Dank gebührt *Dr. David Eisen*, von dem ich nicht nur sehr viel über Massenspektrometrie sondern auch über Fotografie und Freundschaft lernen konnte. Ich wünsche ihm und seiner Familie für die Zukunft alles Gute.

Dr. Markus Müller danke ich zutiefst für seine fortwährende Unterstützung, sein unerschöpfliches Wissen, sein Ideenreichtum, seine oft kritische aber immer äußerst wertvolle Meinung und seine Freundschaft. Ohne ihn hätte ich vieles nicht geschafft und es wäre nicht so schön gewesen.

Der besten Arbeitskollegin und Freundin *Dr. Jessica Steinbacher* danke ich dafür, dass sie mich immer motiviert hat, mir in so manch kritischer Situation beigestanden hat, mir manchmal den Kopf zurechtgerückt hat und für die vielen Unternehmungen in und außerhalb der Uni.

Allen Kollegen aus meinem Labor F4.017 *Dr. Maria Tomás Gamasa, Dr. Barbara Steigenberger, Thomas Wildenhof* und *Dr. Sascha Serdjukow* möchte ich für die gute Stimmung und den Zusammenhalt danken, insbesondere danke ich meiner Laborplatznachbarin *Kerstin Kurz* für's gemeinsame Lachen und manchmal auch Weinen. Außerdem danke ich meinen Praktikantinnen *Angie Kirchner* und *Katharina Strobl*, meiner Bachelorstudentin *Ann-Kathrin Schneider* und meinem Masterstudenten *Leander Runtsch* für ihre gute Arbeit und wünsche ihnen allen weiterhin viel Erfolg.

Ich danke dem gesamten AK Carell für die gute Arbeitsatmosphäre und eine überragende Zeit während der gesamten Doktorarbeit mit all den wilden Partys, tollen Mädelsabenden, gemeinsamen Kochaktionen und schönen Ausflügen. Ein großes Danke geht an *Andrea Künzel, Edris Parsa, Olga Kameroner, Dr. Ben Hackner, Dr. Karola Gasteiger, Dr. Michael Ehrlich, Dr. Toni Pfaffeneder, Dr. Ines Pfaffeneder, Korbinian Brunner, Dr. Arne Schröder, „Kinki“ (Florian Kink), Dr. Mirko Wagner, Michael Stadlmeier, Dr. Dorothea Matschkal, Dr. Sandra Koch, Michael Gattner, Dr. Andrea Kneuttinger, Dr. Johannes Harder, Dr. Stefan Schießler, Dr. Milan Vrabel, Dr. Caterina Brandmayr, Dr. Stefan Prill, Dr. Ulrike Lischke, Dr. Korbinian Heil* und *Dr. Veronika Reiter*, die durch den fachlichen Wissensaustausch und ihre Freundschaftlichkeit die Arbeit in der Gruppe einfach großartig machten.

Mein größter persönlicher Dank gilt schließlich meinen Eltern *Helga* und *Horst Laube*, meiner Schwester *Alexandra Raak*, meiner Tante *Ingrid Steinlechner* und meinen langjährigen Freundinnen *Sofie Koutsikos* und *Nina Sperber*, die mir Halt und Kraft gaben, da sie immer für mich da waren und mich bei all meinen Entscheidungen unterstützt haben. Ebenso danke ich meinem Freund *Thorsten Gröne* dafür, dass er sich Tag für Tag auf das Abenteuer mit mir einlässt, mein Leben bereichert und mich in guten und nicht so guten Lebensphasen zum Lachen bringt.

Inhaltsverzeichnis

Zusammenfassung.....	IV
Summary	VII
1. Einleitung	1
1.1 Epigenetik.....	1
1.1.1 Histone und Chromatin	1
1.1.2 DNA-Methylierung und andere Modifikationen	5
1.2 Proteomik.....	9
1.3 Technologien für die Proteomik	11
1.3.1 Trennung und Anreicherung von Proteinen	12
1.3.2 Identifikation von Proteinen via Massenspektrometrie	13
1.3.3 Quantitative MS-Ansätze.....	19
1.3.6 Crosslinking	24
1.3.4 Software und Datenbanken.....	26
2. Zielsetzung	29
3. Ergebnisse und Diskussion	30
3.1 Methodenentwicklung für quantitative DNA-Protein-Interaktionsstudien mit Crosslinker	30
3.1.1 Photoaffinitätsmarkierungs-Experimente	30
3.1.2 Validierung Isobarer Markierungsreagenzien.....	33
3.1.3 Studien zu den Inkubationsbedingungen	38
3.1.4 Einführung eines spaltbaren Crosslinkers.....	43
3.2 Anwendung der Methode	48
3.2.1 Crosslinking-Interaktionsstudie mit 5-Hydroxymethyluracil in murinen Stammzellen	48

3.2.2 Crosslinking-Interaktionsstudie mit 5-Methylcytosin und 5-Hydroxymethylcytosin in murinen Retinageweben	63
3.2.3 Studie über 5-Formylcytosin-bindende Proteine während der Entwicklung von murinen Stammzellen zu Epiblast-ähnlichen Zellen	73
3.4.1 Crosslinking-Interaktionsstudie	73
3.4.2 Vergleich mit publizierten Interaktionsstudien	87
3.4.3 Studie zu Histonmodifikationen	88
5. Material und Methoden.....	95
5.1 Chemikalien	95
5.2 Puffer und Lösungen	95
5.3 Geräte, Software und Online-Tools für Bioinformatik.....	96
5.4 Oligodesoxynukleotide	97
5.5 Oligodesoxynukleotidsynthese.....	98
5.7 Biochemische Methoden und Assays	99
5.7.1 DNA-Sonden Herstellung.....	99
5.7.2 Zell- und Gewebeproben	100
5.7.3 Zellkernlysate Präparation.....	101
5.7.4 Bradford Assay	101
5.7.5 SDS-PAGE mit Fluorescein-markierten Proben.....	102
5.7.6 <i>Pull down</i> Assays.....	102
5.7.7 Saure Extraktion von Histonen	103
5.7.8 Proteinverdau	104
5.7.9 Isobare Markierung	104
5.8 Massenspektrometrie	105
5.8.1 MALDI-TOF	105
5.8.2 Triple Quadrupol	106

Inhaltsverzeichnis

5.8.3 Orbitrap XL	106
5.8.4 Q Exactive HF.....	107
5.8.5 Proteinidentifikation und Quantifizierung.....	109
Anhang	110
Abkürzungsverzeichnis	138
Literaturverzeichnis.....	143

Zusammenfassung

Die aktive Demethylierung der DNA, die zu einer Reaktivierung stillgelegter Gene führen kann, beschäftigt Wissenschaftler nun schon seit über 30 Jahren. Als die Cytosin-Oxidationsprodukte 5-Hydroxymethylcytosin (hmC), 5-Formylcytosin (fC) und 5-Carboxylcytosin (caC) entdeckt wurden, glaubte man, damit Intermediate für diesen Prozess gefunden zu haben. Weitere Hypothesen sprechen den neu entdeckten Basen sogar darüber hinausgehende biologische Funktionen zu. Unterstützt werden diese Hypothesen durch wenige, bisher durchgeführte Studien über DNA-Protein-Wechselwirkungen und der Analyse möglicher, spezifischer Interaktoren. Die verwendeten Methoden stoßen aufgrund der komplexen Zusammenhänge immer wieder an ihre Grenzen und so ist man noch weit davon entfernt, die genauen Mechanismen und das Zusammenspiel mit anderen epigenetischen Schaltersystemen, wie beispielsweise dem Histon-Code, zu verstehen.

Um dem Ziel der Aufklärung von Wechselwirkungen zwischen epigenetisch modifizierter DNA und bestimmten Proteinen näher zu kommen, wurde in dieser Arbeit eine Methode etabliert, mit deren Hilfe komplexe Proteinproben auf DNA-Interaktoren gescreent und gleichzeitig direkte von indirekten Bindern unterschieden werden können. Dies wurde durch die Einführung eines spaltbaren, Aminogruppen-spezifischen Crosslinker an den DNA-Sonden erreicht. Durch diesen werden Proteine zunächst kovalent an die DNA-Sonden gebunden und alle übrigen Proteine mit geringer Affinität zur DNA oder zu den gebundenen Proteinen können durch mehrere Waschschrte entfernt werden. Während eines Reduktionsschrittes wird der Crosslinker gespalten, so dass die Proteine einerseits von der DNA abgetrennt und andererseits durch einen kleinen verbleibenden Teil des Crosslinkers markiert werden. Es konnte gezeigt werden, dass nach dem anschließenden Verdau die Peptide mit dem Crosslinker-Rest bei der nachfolgenden massenspektrometrischen Analyse identifizierbar sind und dem jeweiligen Protein zugeordnet werden können. Da die Struktur des Crosslinkers prinzipiell frei wählbar ist und über die Click-Reaktion an jede DNA angebracht werden kann, die ein Alkin als funktionelle Gruppe enthält, ist die Methode universell für jede DNA-Sequenz oder -Modifikation anwendbar. Durch den Einsatz isobarer Markierungsreagenzien wurde eine relative Quantifizierung der Proteine ermöglicht. Darüber hinaus wurden die biochemischen Methoden optimiert, so dass möglichst viele spezifische Proteine an der DNA angereichert und gleichzeitig unspezifisch signifikant angereicherte Proteine vermieden werden konnten. Der Amino-reaktive Crosslinker wurde in drei unabhängigen Studien eingesetzt. In allen drei Studien konnten mit den entsprechenden DNA-Sonden viele Proteine angereichert werden, die in Gene Ontology (GO)-

Analysen ein konsistentes Bild ergeben. Eine Verfeinerung der Methode und der Software-basierten Auswertung konnte noch nicht abgeschlossen werden. Deshalb wurde mit dem in diesen Studien verwendeten DNA-Protein-Crosslinker eine Eingrenzung der angereicherten Proteine auf wenige, vielversprechende Kandidaten noch nicht erreicht.

In der ersten Studie wurde versucht aufzuklären, welche Proteine in murinen Stammzellen (mESCs) mit 5-Hydroxymethyluracil (hmU) interagieren. Unsere Arbeitsgruppe konnte zeigen, dass hmU in mESCs nicht durch Desaminierung von hmC als Teil eines möglichen aktiven Demethylierungsmechanismus entsteht, sondern vielmehr durch Oxidation von Thymin (T) durch *Ten-eleven translocation* Methylcytosin Dioxygenasen (Tet-Enzyme).^[1] Für die anschließende Studie wurde hmU sowohl in Stellung gegenüber Adenosin (A) (entsteht bei der Oxidation von T) als auch gegenüber Guanosin (G) (entsteht bei der Desaminierung von hmC) untersucht. Auch hmC war Teil dieser Studie. Es stellte sich heraus, dass überwiegend Proteine angereichert werden konnten, die an der Genexpression und Chromosomenorganisation beteiligt sind. Da es im direkten Vergleich kaum Überschneidungen zwischen den gefundenen Proteinen in den drei Ansätzen gibt, kann davon ausgegangen werden, dass viele Proteine nicht nur eine DNA-Modifikation spezifisch erkennen können, sondern auch unterscheiden mit welcher Base die Modifikation gepaart ist und damit der Ursprung der Modifikation eine erhebliche Rolle spielt. Bei der Wiederholung des Experiments mit einer alternativen Basensequenz der DNA-Sonde konnte das Ergebnis auf Grund einer schlechten Zellkernlysats Präparation noch nicht eindeutig bestätigt werden.

Die zweite Studie befasste sich mit mC-, hmC- und fC-interagierenden Proteinen während der Entwicklung der Retina von Mäusen vor und nach dem Zeitpunkt, an dem die Tiere die Augen öffnen. Innerhalb dieser Zeitspanne teilen sich die neuronalen Zellen der Retina nicht mehr und es konnte beobachtet werden, dass die mC-Level fast gleich bleiben, während die hmC-Level stark ansteigen und die fC-Level stark abfallen.^[2] Außerdem ist bekannt, dass steigende hmC-Level in neuronalen Zellen mit einem Anstieg der Genexpression korrelieren. DNA-Sonden mit je einer der drei Cytosinderivate mC, hmC und fC und dem NHS-Ester-Crosslinker wurden für *pull down* Studien aus Zellkernextrakten muriner Retinagewebe zu den verschiedenen Entwicklungsstadien verwendet. Die Ergebnisse zeigen, dass die Mengen an angereicherten Proteinen an den DNA-Sonden mit den ermittelten endogenen Vorkommen der Cytosinderivate korrelieren. Unter den angereicherten hmC-Bindern wurden einige Proteine gefunden, die in einer anderen Studie über die Tet3-induzierte Genexpression als Tet3-Binder identifiziert wurden.

Bei der dritten Studie wurde untersucht, ob es unterschiedliche fC-bindende Proteine bei der Entwicklung von murinen Stammzellen zu Epiblast-ähnlichen Zellen innerhalb von 48 h gibt. Der Zeitraum wurde aufgrund der von Jessica Steinbacher durchgeführten fC-Level-Messung gewählt, die

einen deutlichen Anstieg der Modifikation zeigt. Die *pull down* Studien wurden zu den Zeitpunkten d_0 , d_1 und d_2 mit DNA-Sonden durchgeführt, die den NHS-Ester-Crosslinker und fC enthielten. Mit dem Anstieg des endogenen fC-Levels über den beobachteten Zeitraum wurde ein Sinken der Anzahl an fC-bindenden Proteinen festgestellt. Während zu den Zeitpunkten d_0 und d_1 überwiegend Proteine angereichert wurden, die an der Genexpression und Transkription beteiligt sind, wurden zum Zeitpunkt d_2 Proteine angereichert, die an der Chromatinorganisation beteiligt sind. Besonders auffällig war, dass sehr viele verschiedene Histonprotein-Varianten angereichert wurden, die v.a. zu den ersten beiden Zeitpunkten auch mit dem Crosslinker markiert wurden. Dies könnte ein Hinweis auf eine sich verändernde Chromatinstruktur sein. Auf Grund dieser Ergebnisse wurde die Studie auf die Identifikation und relative Quantifizierung von Histonmodifikationen ausgeweitet. Dadurch konnte gezeigt werden, dass der Grad an Histonmodifikationen innerhalb der ersten 24 h sinkt und danach wieder ansteigt. Möglicherweise hängt dies mit einer epigenetischen Reprogrammierung der Zellen zusammen.

Summary

For over 30 years scientists are engaged in studying the process of active demethylation of DNA, which enables the reactivation of silenced genes. When the cytosine oxidation states 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC) and 5-carboxylcytosine (caC) were discovered, they were considered as intermediates of this mechanism. Further hypotheses attribute even additional biological functions to the newly discovered bases. These hypotheses are supported by the few studies conducted till today covering DNA-protein interactions and the analysis of potential specific interactors. The applied methods reach their limit because of the complex correlations, affectively limiting the understanding of the exact mechanisms and of the interplay with other epigenetic control systems like e.g. the Histone code.

With the objective to shed light on the interaction between epigenetically modified DNA and specific proteins, this dissertation establishes a method to screen complex protein samples for DNA interactors, while at the same time distinguishing direct from indirect binders. This was achieved by the introduction of a cleavable, amino-group-specific crosslinker at the DNA probes. This crosslinker connects proteins temporarily with the DNA probes. All other proteins with lower affinity to either the DNA probe or the bound proteins can be removed by several washing steps. During protein reduction, the crosslinker will be cleaved in a way that proteins are released from the DNA probes but will be labelled with a small remaining moiety of the crosslinker. It was demonstrated, that after protein digestion the peptides, labelled with the crosslinker residue, can be identified in a subsequent mass spectrometric analysis and assigned to the respective protein. This method can be applied universally for every DNA sequence or modification, because the structure of the crosslinker is arbitrary and can be attached to any alkyne-bearing DNA by click reaction. In addition, isobaric labeling enabled a relative quantification. Furthermore, biochemical methods were optimised in order to enrich as many specific proteins as possible with the DNA, while at the same time avoiding the enrichment of unspecifically binding proteins. The amino-reactive crosslinker was applied in three independent studies. In all of the three studies, a large number of proteins were enriched with the corresponding DNA probe, so that the subsequent gene ontology analysis showed consistent results. A refinement of the method and the software based analysis is pending to date. This is why it was not possible to confine the enriched proteins to provide promising candidates with the help of the DNA-protein-crosslinker used in these studies.

The first study aimed at clarifying which proteins interact with 5-hydroxymethyluracil (hmU) in murine embryonic stem cells (mESCs). Our research group was able to demonstrate that hmU is not created by deamination of hmC as a part of a potentially active demethylation mechanism in mESCs, but indeed by the ten-eleven translocation methylcytosine dioxygenases (Tet-enzymes) mediated oxidation of thymine (T).^[1] In the subsequent study hmU was examined once paired with adenosine (A) (result of T oxidation) and once paired with guanosine (G) (result of hmC deamination). hmC was part of the study as well. It was discovered that primarily those proteins were enriched which are part of gene expression and chromosome organisation. In a direct comparison between the discovered proteins in the three approaches, barely any overlaps were found; this gives evidence that many proteins not only specifically recognize a DNA modification, but also distinguish between the base pairings. This suggests that the origin of the modification plays a significant role. The repetition of the experiment with an alternative base sequence of the DNA probe was not yet able to confirm the previous results, because of an insufficient preparation of the nuclear cell lysate.

The second study focused on proteins interacting with mC, hmC and fC during the development of the retina in mice before and after the time of eye opening. Within this period the neuronal retina cells do not longer divide and it was observed that mC levels stay constant, while hmC levels strongly increase and fC levels strongly decrease.^[2] It is also known that raising hmC levels correlate with an increased gene expression. DNA probes, each containing one of the three cytosine derivatives mC, hmC and fC, and the NHS-ester-crosslinker, were used for pull down studies from nuclear cell lysates of murine retinal tissue at the two different stages of development. The results show a correlation between the amounts of enriched proteins by the DNA probes and the measured levels of endogenous cytosine derivatives. Among the enriched hmC binders, some proteins were found which also have been identified as Tet3 binders in another study about Tet3-induced gene expression.

The third study examined if there are different fC binding proteins during the development of mESCs towards epiblast like cells (EpiLCs) within 48 hours. This time period was chosen on the basis of fC level measurements (conducted by Jessica Steinbacher), which showed a significant increase of the modification. The pull down studies were performed at the time points d_0 , d_1 and d_2 with DNA probes containing the NHS ester crosslinker and fC modification. With the rise of endogenous fC levels during the monitored period, a decrease in the number of fC binding proteins could be observed. While at the times d_0 and d_1 predominantly proteins were enriched, which are involved in gene expression and transcription, at the time d_2 proteins were enriched, which are involved in chromatin organisation. It was also particularly conspicuous that a number of different histone variants were enriched at all investigated times, and especially those proteins found at the first two time points

were labelled with the crosslinker. This could be an indication for a changing chromatin structure during development. Due to these results, the study was extended to the identification and relative quantification of histone modifications. Here it was shown, that the global levels of histone modifications drop within the first 24 hours of mESC development and then rise again within the next 24 hours. Perhaps these findings indicate a connection to epigenetic reprogramming of the cells.

1. Einleitung

1.1 Epigenetik

Am 26. Juni 2000 wurde durch Bill Clinton und die Leiter des *Human Genome Project* (HGP) die vollständige Entschlüsselung des menschlichen Erbguts verkündet und gut ein halbes Jahr später folgte die Veröffentlichung der Forschungen in *Nature*.^[3] Dies war nicht nur ein bedeutender Meilenstein in der Genomforschung, sondern ließ die ganze Welt auf einen enormen Fortschritt in der Medizin hoffen. „Die Genomwissenschaft wird die Diagnose, Prävention und die Behandlung der Meisten, wenn nicht aller menschlichen Krankheiten revolutionieren“, kommentierte der ehemalige Präsident Clinton. Allerdings wurde diese Euphorie bald durch die neue Erkenntnis eingeholt, dass man den genetischen Code nun zwar lesen konnte, allerdings noch weit davon entfernt war, ihn auch zu verstehen. Die Tatsache, dass sich die Menschen untereinander genetisch gesehen nur um 0,1 % unterscheiden, kann die vielfältigen Unterschiede in den Ausprägungen der Veranlagung eines jeden Individuums nicht erklären. Und darüber hinaus besitzen schließlich alle Zellen ein und desselben Organismus dieselbe genetische Information, können sich aber in Gestalt und Funktion grundlegend voneinander unterscheiden. Die Schlussfolgerung daraus ist, dass es Mechanismen geben muss, die dem genetischen Code übergeordnet, in Abhängigkeit von Zeit, Ort und äußeren oder umweltbedingten Faktoren, die Expression der zuständigen Gene steuern und dieses Expressionsmuster an Tochterzellen vererben. Mit der Untersuchung dieser Mechanismen befasst sich das Fachgebiet der Epigenetik.^[4] Der Hauptfokus der Forschung richtet sich dabei auf Histon-, DNA- und RNA-Modifikationen sowie RNA-vermittelte Gen-Inaktivierung.^[5-12] Die Biochemiker Stedman und Stedman beispielsweise vermuteten bereits sehr früh und noch vor der Strukturaufklärung der DNA, dass Histone an der Suppression einzelner Gene oder Gengruppen beteiligt sein könnten.^[13] Dennoch dauerte es bis etwa Mitte der 1990er Jahre, bis den Histonen mehr Funktionen als das reine Verpacken der DNA zugeschrieben wurde.^[14, 15]

1.1.1 Histone und Chromatin

Die Histone sind Proteine, die zusammen mit der DNA das Grundgerüst der Chromosomen aufbauen. Es gibt fünf Hauptfamilien der Histonproteine: H1/H5, H2A, H2B, H3 und H4. Die beiden Proteine H2A und H2B bilden zusammen ein Dimer, ebenso wie die beiden Proteine H3 und H4 (siehe Abbildung 1A). Jeweils zwei gleiche dieser Dimere lagern sich zu einem Tetramer zusammen, so dass es ein H2A-H2B-Tetramer und ein H3-H4-Tetramer gibt. Diese beiden Tetramere wiederum bilden zusammen den oktameren Nukleosomenkern, um den sich ein 146 Basenpaare (bp) langes DNA-Stück 1,65 Mal

in einer linksgängigen Windung legen kann (siehe Abbildung 1B/C). Der Anfang und das Ende des DNA-Stücks werden zusätzlich durch das Linker-Histon H1 fixiert (siehe Abbildung 1D), wodurch die Ausbildung einer Helix aus Nukleosomen ermöglicht wird (siehe Abbildung 1E, 30 nm Chromatinfaser). So entsteht eine übergeordnete Struktur, in der die DNA bis hin zum Chromosom komprimiert werden kann (siehe Abbildung 1E). Die Nukleosomenkernhistone treten mit der DNA hauptsächlich über deren Phosphodiester-Rückgrat in Kontakt, wodurch ein weitestgehend von der Sequenz unabhängiges Packen der DNA möglich wird. Allerdings besitzen die Kernhistone ein N-terminales Ende, das aus der Sphäre des Kerns ragt und somit auch zwischen und neben den DNA-Windungen hervorsteht.^[16] Die Enden der Histone spielen eine wichtige Rolle für die Stabilität des Nukleosoms bzw. Chromatins und sind häufig einer Vielzahl von unterschiedlichen post-translationalen Modifikationen (PTMs) unterworfen (vgl. Abbildung 2).^[17, 18]

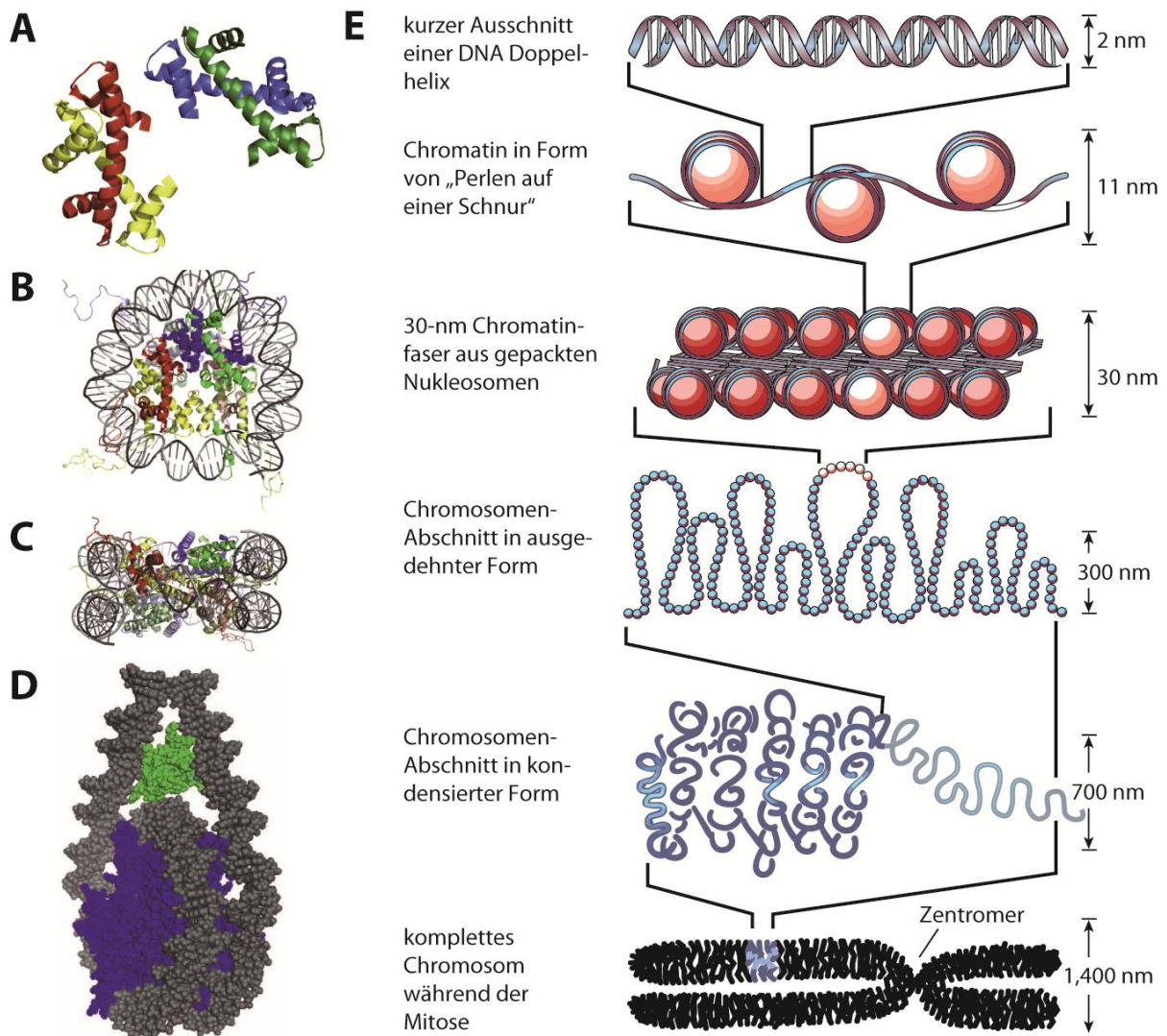


Abbildung 1: Aufbau eines Chromosoms. (A) Schematische Darstellung der H2A-H2B- (rot/gelb) und H3-H4-Dimere (grün/blau). (B) Ansicht von oben auf den Nukleosomenkern, der aus jeweils vier zusammengelagerten Dimeren besteht. Darum windet sich die DNA. Es sind deutlich die N-terminalen Enden der Histone zu erkennen, die teilweise weit zwischen den DNA-Windungen hindurch reichen. (C) Seitlicher Querschnitt durch den Nukleosomenkern. (D) Das Histon H1 (grün) fixiert Ein- und Ausgangsabschnitt der DNA (grau), die sich dadurch 1,65 Mal um den Nukleosomenkern (blau) windet. (E) Verpacken der DNA. Durch das Wickeln der DNA um das Chromatin kann die DNA bis hin zum Chromosom kondensiert werden. Die Bilder wurden aus den Referenzen ^[19, 20] adaptiert.

Berechnungen des elektrostatischen Potentials der nukleosomalen Kernpartikeloberfläche zeigen positiv geladene Regionen an den Histon-Enden und entlang der Kontaktpunkte zwischen Nukleosom und DNA, während die zentralen Bereiche des Nukleosoms eher negativ geladen sind. Diese Unterschiede im elektrostatischen Potential geben Grund zu der Annahme, dass die Kernhiston-Enden zu internukleosomalen Interaktionen führen können, weshalb sie großen Einfluss auf die Struktur des Chromatins und damit auch auf die Zugänglichkeit der DNA besitzen könnten.^[21, 22] Durch die PTMs kann das elektrostatische Potential verändert werden. So führen beispielsweise Acetylierungen der Lysine an den Histonen zur Neutralisierung der positiven Ladung des Lysins, wodurch die elektrostatische Wechselwirkung zur negativ geladenen DNA verringert wird. Die

Chromatinstruktur wird dadurch aufgelockert und der Zugang zur DNA für Proteine, wie zum Beispiel Transkriptionsfaktoren, erleichtert. Neben Histon-Acetylierungen sind viele weitere Modifikationen bekannt, die häufigsten darunter sind Methylierungen und Phosphorylierungen. Obwohl Phosphorylierungen bei weitem noch nicht so gut untersucht wurden wie Acetylierungen, dürften sie durch die Übertragung negativer Ladung ähnlichen Einfluss auf die Ladungsbilanz der Histone haben und damit ebenfalls eher zu einer Auflockerung der Chromatinstruktur führen.

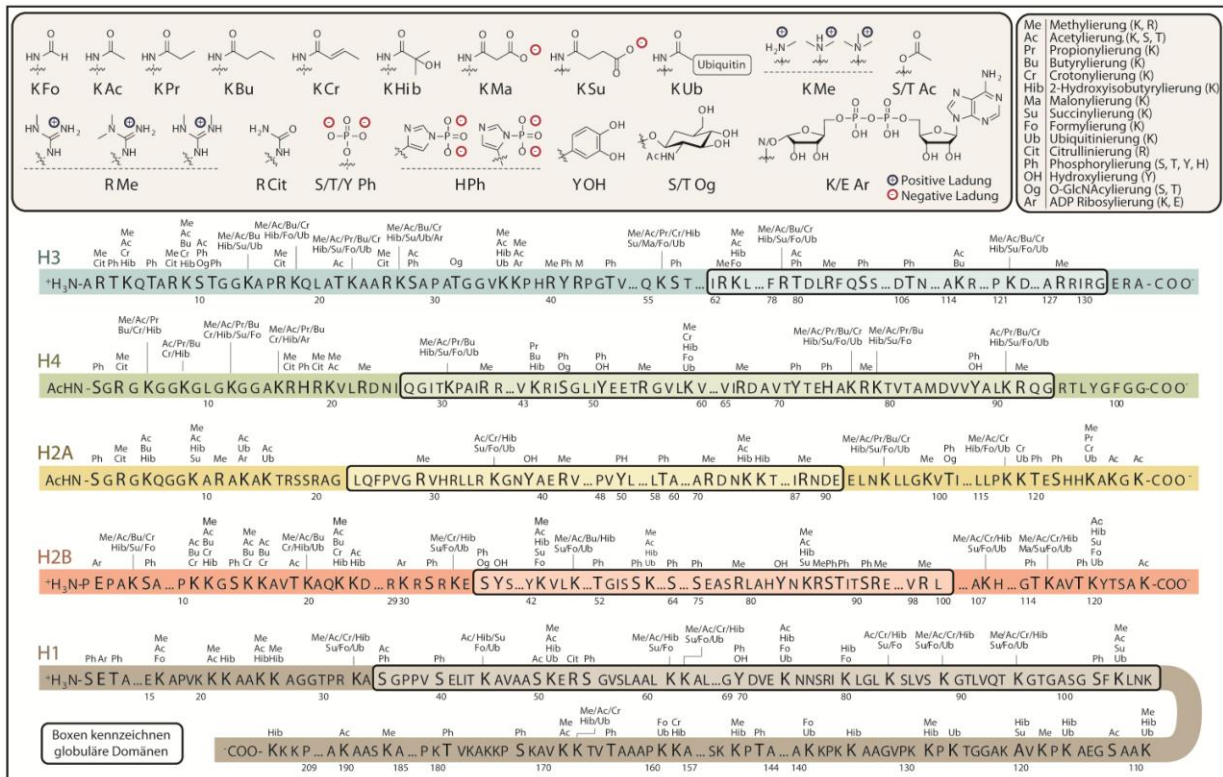


Abbildung 2: Übersicht bekannter Modifikationen der Histone H1 (grau), H2A (gelb), H2B (rot), H3 (blau) und H4 (grün). Bekannte Modifikationen sind über den jeweiligen Aminosäuren der Histone aufgeführt. Abbildung adaptiert aus Referenz [23].

Die Frage ist nun, welche Faktoren eigentlich zu derartigen Modifikationen und den damit einhergehenden Strukturveränderung des Chromatins führen können. Eine mögliche Antwort dazu liefern Histon-Methylierungen. Sie verändern zwar die Ladung der Histonproteine nicht, weisen dafür aber eine wesentlich höhere Variabilität auf. Lysine der Histone können mono-, di- oder trimethyliert und Arginine mono-, symmetrisch oder asymmetrisch dimethyliert werden. Wenn Methylierungen direkten Einfluss auf die Chromatinstruktur nehmen, dann könnte dies zum Beispiel über hydrophobe Interaktionen geschehen, welche bislang allerdings noch nicht nachgewiesen werden konnten.^[24] Viel wahrscheinlicher ist deshalb, dass die Modifikation als Signal für Nicht-Histon-Proteine dient und eine Art Plattform für enzymatische Aktivitäten darstellt, die weitere biologische Prozesse einleiten können.^[25] Beispielsweise wird das am Lysin trimethylierte Histon H3 (kurz

H3K4me3) von den Proteinen CHD1, einem ATP abhängigem Enzym, das Nukleosomen repositioniert, und JMJD2A, einer Histon-Demethylase, gebunden. Damit rekrutiert H3K4me3 also direkt Chromatin-verändernde Enzyme.^[26] Selbstverständlich ist dieser Signal-Mechanismus nicht nur auf Methylierungen beschränkt; es sind bereits viele Proteine und Proteinkomplexe bekannt, welche unterschiedliche Modifikationen erkennen und daran binden können.^[27-32] Oft besitzen solche Proteine eine oder mehrere spezifische Domänen (z.B. Chromo-, Bromodomain, Tudor Domain, PhD Finger, MBD Domain, 14-3-3), wodurch Modifikationen oder andere nukleosomale Eigenschaften simultan erkannt werden können. So werden etwa acetylierte Lysine der Histone von Bromodomains gebunden, die oft in Histon-Acetyl-Transferasen (HATs) und Chromatin-Remodeling Komplexen (z.B. SWI/SNF) gefunden werden.^[33, 34] Und das Protein HP1 bindet über seine N-terminale Chromo-Domäne an H3K9me3 und H1.4K26me, während es über seine C-terminale Chromoshadow-Domäne dimerisieren kann und somit ein bivalenter Chromatin-Binder wird. Sowohl das Protein HP1 als auch das Histon-Protein H1.4 werden vor allem mit den hoch kondensierten Bereichen des Chromatin, auch Heterochromatin genannt, in Verbindung gebracht.^[35-37]

Zusammen mit der enormen Anzahl an möglichen Modifikationen ergibt sich eine solch hohe Anzahl an Kombinationsmöglichkeiten, dass man in diesem Zusammenhang auch vom „Histon-Code“ spricht, eine Hypothese, die zur Zeit in der Literatur intensiv diskutiert wird.^[38] Demnach können verschiedene Modifikationen synergistische oder auch antagonistische Affinitäten zur Wechselwirkung mit chromatin-assoziierten Proteinen generieren, welche ihrerseits wieder dynamische Übergänge zwischen transkriptional aktiven oder inaktiven Chromatinzuständen schaffen. Damit würde das Informationspotential des genetischen Codes wesentlich erhöht und um die Möglichkeit zur direkten Antwort auf äußere Faktoren erweitert. Darüber hinaus ist auch ein Zusammenspiel von Histon-Modifikationen mit chemisch modifizierter DNA denkbar. Die Proteine UHRF1 und KDM2A beispielsweise, binden bevorzugt an Nukleosomen, die das am Lysin 9 trimethylierte Histon H3 (H3K9me3) aufweisen. Im Falle des UHRF1 wird diese Wechselwirkung signifikant verstärkt, wenn Cytosine der nukleosomalen DNA methyliert sind. KDM2A bindet im Gegensatz dazu nur dann, wenn die Cytosine nicht methyliert sind.^[39] Da DNA-Methylierungen neben den posttranslationalen Histon-Modifikationen zu den wichtigsten epigenetischen Markern in eukaryotischen Zellen zählen, soll im nächsten Kapitel näher darauf eingegangen werden.

1.1.2 DNA-Methylierung und andere Modifikationen

Obwohl bereits im Jahre 1948 zufällig von R.D. Hotchkiss bei der Auftrennung von DNA Basen aus dem Kalbsthymus mittels Papierchromatographie entdeckt, wurde erst ab 1980 die Bedeutung von 5-Methyl-Cytosin (mC) für die Regulation der Genfunktion und Differenzierung erkannt.^[40, 41] Seit dem

ist das Cytosin-Derivat Gegenstand intensiver Forschung. Es konnte bereits gezeigt werden, dass es an wichtigen epigenetischen Mechanismen zur Unterdrückung der Genexpression bei diversen biologischen Funktionen, wie der Embryonalentwicklung, dem genomischem Imprinting, der X-Chromosom-Inaktivierung und der Erhaltung der Chromosomenstabilität, beteiligt ist.^[42] In den Promotorregionen der Gene kommen nämlich besonders häufig sogenannte Cytosin-phosphatidyl-Guanin (CpG)-Inseln vor. Das sind mehrfache Wiederholungen von CpG Dinukleotiden innerhalb eines DNA-Einzelstrangs. Liegen diese CpG-Inseln im methylierten Zustand vor, wirken sie besonders attraktiv auf „*methyl-CpG-binding domain*“ (MBD)-Proteine, welche das Binden von Transkriptionsfaktoren an die DNA und damit auch ihre Transkription verhindern können.^[43, 44] Zusätzlich können MBD-Proteine Histon-remodellierende Enzyme rekrutieren und somit die Repression der Transkription weiter fördern, indem die DNA durch eine veränderte Chromatin-Struktur schlechter zugänglich wird.^[45]

Die Methylierung des Cytosins wird in Säugetieren durch drei Proteine einer Familie von Enzymen ausgeführt, die bekannt sind als DNA Methyltransferasen (DNMTs). DNMT1 wird durch seinen Reaktionspartner UHRF1 zur Instandhaltung vorhandener Methylierungsmuster während der Zellteilung mobilisiert. UHRF1 bindet dabei an hemimethylierte CpG-Orte und rekrutiert DNMT1, was zur Methylierung des gerade entstehenden DNA Strangs führt.^[46, 47] DNMT3A und DNMT3B dagegen sind an der *de novo* Methylierung der DNA beteiligt. Sie werden über DNMT3L aktiviert, das selbst keine Methyltransferase-Aktivität besitzt.^[48, 49] Die genauen Mechanismen, wie die DNMTs zu den richtigen Stellen zur *de novo* Methylierung geleitet werden, sind noch nicht bekannt, dennoch ist davon auszugehen, dass die DNA-Methylierung, ebenso wie die Modifikation der Histone, ein genau kontrollierter Prozess ist.

Da die DNA-Methylierung oft ein Teil von regulatorischen Prozessen ist, die eine langfristige Stilllegung von Genen erfordert, wie beispielsweise der X-Chromosom-Inaktivierung, dachte man im Allgemeinen, dass Methylierungen relativ stabile epigenetische Marker sind. Tatsächlich wurden in den letzten Jahren intensiver Forschung durchaus auch Kontext-spezifische Demethylierungen beobachtet. Die potentiellen Mechanismen zur Entfernung einer Methylgruppe vom Cytosin können entweder passiver oder aktiver Natur sein.^[50] Von passiver Demethylierung spricht man, wenn methylierte DNA in Abwesenheit von DNMT1 repliziert und damit sozusagen verdünnt wird. Nach der ersten Zellteilung liegt die DNA nur noch hemimethyliert in den zwei Tochterzellen vor. Und bereits nach der zweiten Zellteilung entstehen zwei Tochterzellen mit hemimethylierten sowie zwei Tochterzellen mit unmethylierten DNA-Strängen (siehe Abbildung 3).

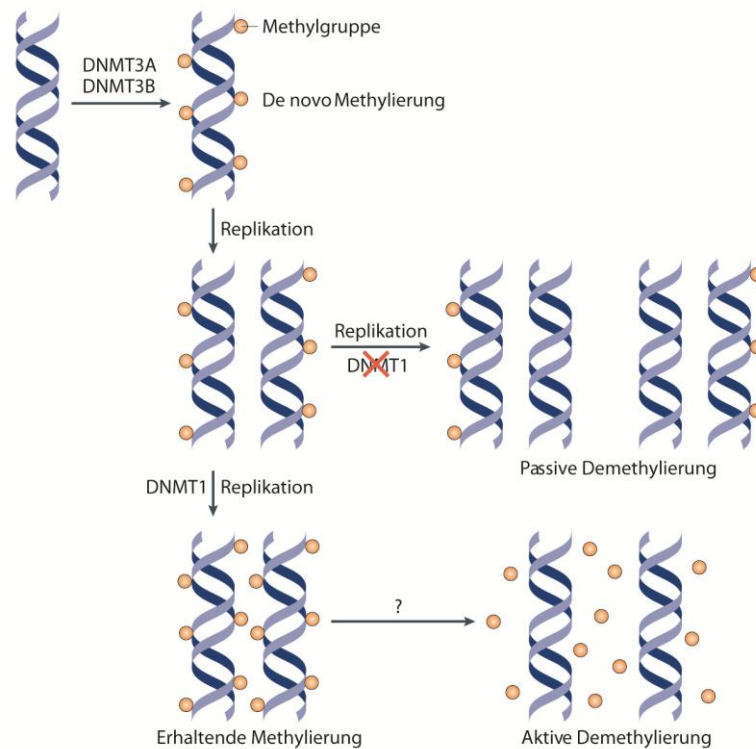


Abbildung 3: Mechanismen zur Methylierung und Demethylierung der DNA. Unmethylierte DNA wird durch Dnmt3a und Dnmt3b methyliert. Nach der Replikation liegen zwei hemimethylierte DNA-Doppelstränge vor. In Abwesenheit von Dnmt1 kommt es durch weitere Replikationszyklen zur passiven Demethylierung, da die neu entstandenen komplementären DNA-Stränge nicht methyliert werden. Der Mechanismus der aktiven Demethylierung und welche Proteine daran beteiligt sind ist noch nicht aufgeklärt. Die Abbildung wurde aus der Referenz^[50] adaptiert.

Die passive Demethylierung der DNA ist damit ein relativ träger Prozess, der zudem die Zellteilung voraussetzt. Nachdem aber auch rapide Änderungen in den DNA Methylierungsleveln beobachtet werden konnten, müssen aktive Mechanismen zur Demethylierung ebenso in Betracht gezogen werden. Genom-weite Demethylierung wurde bisher nur zu spezifischen Zeiten in der frühen embryonalen Entwicklung festgestellt, wie etwa kurz nach der Befruchtung der Eizelle. Das väterliche Genom wird hierbei innerhalb von vier bis acht Stunden, noch bevor die erste Zellteilung abgeschlossen ist, fast vollständig demethyliert.^[51, 52] Gen-spezifische Demethylierungen dagegen wurden vor allem in somatischen Zellen in Reaktion auf spezifische Signale beobachtet. Beispielsweise liegt die Promotorregion des für den Wachstumsfaktor BDNF (*brain-derived neurotrophic factor*) codierenden Gens in Neuronen methyliert vor. Solange die Zellen nicht stimuliert werden, bindet dort das repressive Methylcytosin-bindende Protein MeCP2 und unterdrückt die Transkription des Gens. Wenn nun die neuronale Zelle durch KCl depolarisiert wird, wird die Expression von BDNF hochreguliert. Damit einhergehend konnte beobachtet werden, dass MeCP2 gelöst und der Promotor demethyliert wird.^[53] Da sich die Neuronen nicht mehr teilen, muss die Demethylierung auf eine replikationsunabhängige Weise geschehen, weshalb man einen aktiven Mechanismus vermutet.

Die Vielfalt der biologischen Prozesse, in denen Demethylierungen stattfinden, legt auch die Vermutung nahe, dass sie auf unterschiedliche Art und Weise ablaufen. Daher wurden bereits mehrere Mechanismen zur aktiven Demethylierung vorgeschlagen, einschließlich der enzymatischen Entfernung der Methylgruppe von mC, der Desaminierung von mC zu Thymidin, gefolgt von Basenexzisionsreparatur (*base excision repair*, BER) des entstehenden T:G Mismatches, Nukleotidexzisionsreparatur (*nucleotide excision repair*, NER) und Radikal S-Adenosyl Methionin (SAM)-basierte Demethylierung.^[50, 54-59] Als 2009 die *ten-eleven translocation methylcytosine dioxygenase* (TET)-vermittelte Oxidation von mC zu Hydroxymethylcytosin (hmC) entdeckt wurde, folgte daraus eine weitere Hypothese. Die neue Base könnte ein Intermediat für eine oxidative Demethylierung sein und möglicherweise weitere biologische Funktionen besitzen.^[60, 61] Kurz danach konnte gezeigt werden, dass die TET-Enzyme hmC sogar noch weiter zu Formylcytosin (fC) und Carboxylcytosin (caC) oxidieren können, welche dann über die Thymin-DNA Glykosylase (TDG) ausgeschnitten und über BER in unmethyliertes Cytidin überführt werden können.^[62-65]

Auch wenn einige dieser Mechanismen bereits validiert werden konnten und der technische Fortschritt die Untersuchungen auf diesem Gebiet weiter treiben wird, sind viele Fragen noch ungeklärt: Sind diese Mechanismen global oder können sie auf spezifische Genabschnitte beschränkt sein, beispielsweise als Reaktion auf bestimmte Einflüsse? Wie kann für den letzteren Fall der Prozess auf bestimmte Gene konzentriert werden? Welche Proteine sind an den Mechanismen beteiligt und welche Prozesse finden im Anschluss statt? Wichtige Ansätze zur Beantwortung dieser Fragen liefern unter anderem Studien über Proteine mit hoher Affinität gegenüber den Modifikationen. Als mögliche Bindungspartner könnten solche Proteine die Modifikation selbst entfernen bzw. ändern oder aber als positives oder negatives Signal für weitere Proteine und Prozesse dienen. Im Falle der Cytosinderivate deuteten bereits die ersten Versuche darauf hin, dass mit ihnen einige Proteine wechselwirken, die auch an Prozessen wie BER, Transkription und Chromatinremodelling beteiligt sind.^[66, 67]

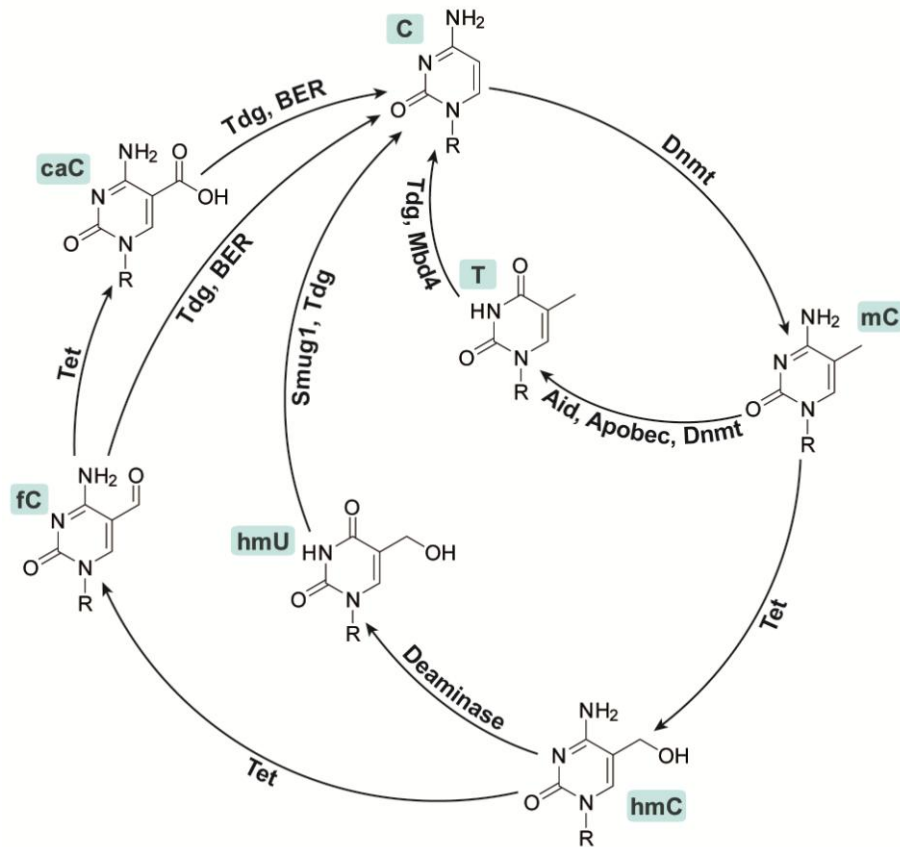


Abbildung 4: Aktuell diskutierte Möglichkeiten zur aktiven DNA-Demethylierung.

1.2 Proteomik

Der Begriff Proteomik wurde 1995 durch Marc Wilkins geprägt und definierte zunächst die groß angelegte Charakterisierung des gesamten Proteinsatzes (Proteoms) einer Zelllinie, eines Gewebes oder eines Organismus.^[68] Das Fachgebiet ist, ebenso wie die Epigenetik, ein direktes Resultat aus den neu gewonnenen Erkenntnissen, die in der Genetik erreicht wurden. Es entstand, da selbst die neuen umfassenden Informationen aus der Genetik nicht ausreichen, um den Phänotyp einer Zelle zu bestimmen. Dieser ist größtenteils auf die vorhandenen aktiven Proteine und deren Funktionen zurückzuführen. Prozesse wie alternatives Splicing oder posttranslationale Modifikationen machen es somit unmöglich, allein über den genetischen Code Mechanismen zu erklären, die zu Krankheiten führen, den Entwicklungs- oder Alterungsprozess beeinflussen oder Reaktionen auf Umwelteinflüsse darstellen. Mittlerweile ist das große Ziel der Proteomik, die Lücke zwischen genombasierter Information und der zellulären Ausprägung zu schließen. Es vereint damit viele verschiedene Forschungsgebiete unter sich (siehe Abbildung 5). Demnach sollten die Gebiete der Genom-, Transkriptom-, Metabolom- und Proteomforschungen nicht mehr separat, sondern vielmehr als ineinander greifend und sich ergänzend betrachtet werden. Nur so ist die gewaltige Aufgabe der

Beschreibung von Zusammensetzung, Dynamik und Verbindungen von Multiprotein-Modulen, die eine ganze Bandbreite von biologischen Prozessen in der Zelle katalysieren, zu bewältigen.^[69]

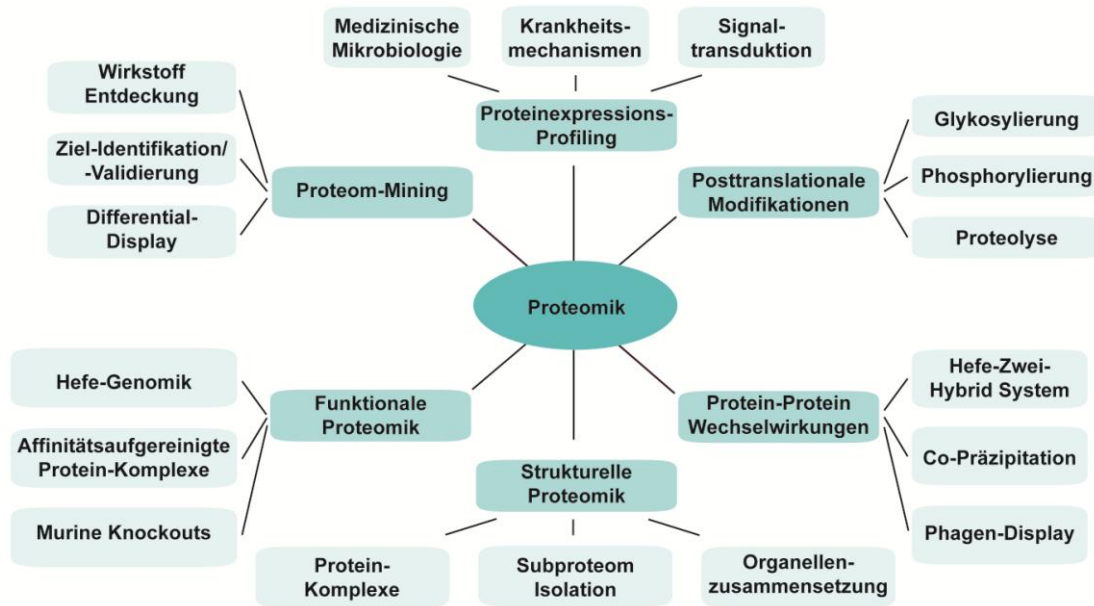


Abbildung 5: Forschungsgebiete der Proteomik.

Die größte Herausforderung der Proteomik ist, dass das Proteom hoch dynamisch ist und unmittelbar die Umgebung reflektiert, in der es untersucht wird. Interne als auch externe Faktoren können dazu führen, dass Proteine posttranslational modifiziert, verlagert, abgebaut oder erst synthetisiert werden. Deshalb ist es wichtig, sich darüber im Klaren zu sein, dass Untersuchungen eines Proteoms unter genau definierten Bedingungen und zu einem bestimmten Zeitpunkt immer nur zu Momentaufnahmen unter den angelegten Bedingungen führen. Bedenkt man alle möglichen Variablen, könnte es sogar sein, dass ein bestimmtes Genom zu unendlich vielen Proteomen führt. In vielen Fällen reicht es jedoch aus, nur eine bestimmte Gruppe von Proteinen (z.B. Histone, Enzyme, Antikörper) oder eine untergeordnete räumliche Einheit (z.B. Zellkompartimente wie Zellkerne, Mitochondrien) zu betrachten.^[70] Dann spricht man auch von einem „Sub-Proteom“. Vor allem für Mediziner sind beispielsweise solche Einblicke interessant, die Aufschluss über mögliche Wirkstoff- oder Therapieziele bieten. Die Proteomik, ebenso wie die Epigenetik, stellen neben der Genomik, Transkriptomik und Metabolomik auch wichtige Grundlagen für die systembiologische Forschung bereit. Ziel der Systembiologie ist die holistische Beschreibung von biologischen Systemen mit Hilfe mathematischer Modelle. Diese sollen es ermöglichen, Vorhersagen über biologische Prozesse zu treffen, welche dann im Labor gezielt überprüft werden können. Systembiologen verwenden die bisher gewonnen Erkenntnisse, um ihre Modelle zu erstellen und zu überprüfen. Je mehr Daten dafür zur Verfügung stehen, desto genauer werden auch die Vorhersagen.

Große Proteomprojekte wurden gestartet, die bei der Darstellung von Protein-Protein Netzwerken, Organisation von Zellorganellen oder bei der Beschreibung von Krankheitszuständen, insbesondere im Bereich der Krebsforschung, helfen sollen.^[71] Aus diesem Grund wurden Projekte, wie das „*Plasma Proteome Project*“, das „*Human Liver Proteome Project*“ oder das „*Human Brain Proteome Project*“ und viele weitere von der „*Human Proteome Organisation*“ initiiert. Nationale und internationale Organisationen (u.a. HUPO, *Human Proteome Organisation*; DGPF, Deutsche Gesellschaft für Proteomforschung) bildeten sich, um die Zusammenarbeit in der Proteomforschung zu fördern und die Entwicklung neuer Technologien und Ansätze voranzutreiben. So existieren heute bereits viele hochempfindliche und genaue Methoden für proteomische Ansätze, dennoch können manche Untersuchungen auf Grund von technologischen Limitierungen noch nicht durchgeführt werden. Eine Übersicht über die derzeit wichtigsten Techniken zur Proteomanalyse ist in Kapitel 1.3 zu finden.

1.3 Technologien für die Proteomik

Technologischer Fortschritt hatte seit jeher einen massiven Einfluss auf die Entwicklung der Proteinanalyse. Bestehende Methoden wurden immer weiter verbessert oder neuen Anforderungen angepasst. Als eines der wichtigsten Werkzeuge in der Proteomforschung gilt beispielsweise nach wie vor die zweidimensionale Gelelektrophorese, die bereits 1975 von O'Farrell und Klose entwickelt wurde und trotz einiger Limitierungen noch immer in vielen Bereichen Anwendung findet.^[72-74] Auch Immunoassays, wie ELISA (*enzyme-linked immunosorbent assay*) und Western Blots, gelten seit Jahrzehnten als bewährte Methoden zur Identifikation und Quantifizierung von Proteinen. Allerdings sind sie auf spezifische Antikörper angewiesen, die in vielen Fällen nicht zur Verfügung stehen. Insbesondere bei hoher Homologie in der Aminosäuresequenz der Zielproteine versagen diese Methoden häufig aufgrund unzureichender Spezifität und hoher Kreuzreaktivität.^[75] Für komplexe Proben, wie z.B. Blutplasma, Gewebeproben oder Zelllysate, gilt seit der Entwicklung von sanften Ionisationstechniken, wie ESI oder MALDI, die Massenspektrometrie (MS) als sehr spezifische und verlässliche Methode zur Identifikation, Quantifizierung und Charakterisierung von Proteinen. Sie ist derzeit die einzige Methode, die universell und ohne Vorwissen über die Zusammensetzung der Probe eingesetzt werden kann, um die Art und die Modifikationen eines Proteins zu bestimmen.^[76] Gleichwohl muss an bestehenden Problemen gearbeitet und an neuen technologischen Lösungen geforscht werden, um das volle Potential an Möglichkeiten auszuschöpfen, die die Massenspektrometrie für die Proteomik bereit hält. Momentan setzen sich typische MS-Experimente in der Proteomik aus drei Stufen zusammen. Diese sind: 1. Trennung (und Anreicherung) der Proteine aus Zellen, Gewebe oder eines Organismus, 2. Identifikation (und Quantifizierung) der Proteine und

3. Abgleich der erhaltenen Daten mit einer entsprechenden Datenbank. Im Folgenden soll auf diese Punkte detaillierter eingegangen werden.

1.3.1 Trennung und Anreicherung von Proteinen

Um alle Proteine einer komplexen Mischung identifizieren zu können, müssen diese zunächst aufgetrennt werden. Dafür nutzt man die unterschiedlichen Merkmale der Proteine, die durch ihre Sequenz und spezifische Struktur bedingt werden. Wegen der einfachen Anwendung, guten Reproduzierbarkeit und breiten Auflösung von Proteinen zwischen 10 – 300 kDa ist die eindimensionale **Gelelektrophorese** die am häufigsten angewendete Methode zur Auftrennung eines Proteingemisches. Die Proteine werden nach ihrer Molekülmasse aufgetrennt und die einzelnen Banden von Interesse können zur weiteren Analyse ausgeschnitten werden. Für sehr komplexe Proteingemische ist die Auflösung allerdings meistens zu gering, so dass die Banden nicht mehr klar voneinander unterscheidbar sind und innerhalb eines ausgeschnittenen Bereichs des Gels viele verschiedene Proteine zu finden sind. In diesen Fällen wird oft auf die bereits erwähnte zweidimensionale Gelelektrophorese zurückgegriffen. Bei dieser Methode werden die Proteine zunächst in einer isoelektrischen Fokussierung nach ihrer Ladung aufgetrennt und anschließend elektrophoretisch nach ihrer Masse separiert. Einer der größten Nachteile der Gel-basierten Methoden ist, dass sie langsam und schwer automatisierbar sind. Einige Proteine, wie Membranproteine, können aufgrund ihrer physikochemischen Eigenschaften auf diese Weise gar nicht aufgetrennt werden.

Flüssigchromatographische (*liquid chromatography*, LC) Verfahren sind die am häufigsten angewandten Gel-freien Methoden zur Auftrennung von Proteinen. Aufgrund der Schnelligkeit, des hohen Auflösungsvermögens, der hohen Sensitivität bei gleichzeitiger guter Reproduzierbarkeit sind LC-Verfahren essenziell für die Protein-Analytik geworden.^[77] Dies gilt insbesondere bei der nachfolgenden Analytik via Massenspektrometrie, da sich die beiden Techniken hervorragend miteinander koppeln lassen, was auch als LC-MS-Verfahren bekannt ist.^[78] Je nachdem welche Eigenschaften der Proteine für die Auftrennung genutzt werden sollen, stehen unterschiedliche Materialien und Methoden zur Verfügung.

Bei der **Umkehrphasen-Chromatographie** (*Reversed Phase Liquid Chromatography*, RPLC) werden die Analyten nach ihrer Hydrophobizität aufgetrennt. Dies wird mithilfe von Kieselgel erreicht, dessen Oberflächenpolarität mit unterschiedlich langen Kohlenstoffketten (C2-, C4-, C8- oder C18-Ketten) eingestellt werden kann. Das kontinuierliche Trennverfahren und die dafür verwendeten Puffer (meist Acetonitril und Wasser) sind besonders gut kompatibel mit kontinuierlichen Ionenquellen wie der Elektrospray Ionisation (ESI). In Experimenten mit komplexen Proben konnte gezeigt werden,

dass unter Hochdruckbedingungen (HPLC bzw. UPLC) die Verlängerung der Chromatographiesäule und die Verringerung der Partikelgröße des Säulenmaterials zu einer Verbesserung der Peakkapazität, Auflösung und Sensitivität sowie der Verringerung der Analysezeit führt.^[79-81]

Proteine können auch auf Grund ihrer ionisierbaren Amino- und Carboxylgruppen durch **Ionenaustausch-Chromatographie** (*Ion Exchange*, IEX) nach ihrer Ladung getrennt werden. Bei dieser Methode spielt vor allem die Zusammensetzung der verwendeten Puffer für die mobile Phase eine entscheidende Rolle, denn die Wechselwirkung mit der stationären Phase wird hauptsächlich durch den pH-Wert und die Salzkonzentration beeinflusst. Bei pH-Werten unter 3 werden negative Ladungen der Carboxylgruppen neutralisiert und die Aminogruppen protoniert. Damit können unter diesen Bedingungen die Proteine am Kationentauscher aufgetrennt werden. Um die Carboxylgruppen negativ zu laden und die Aminogruppen zu neutralisieren und sie damit am Anionentauscher aufzutrennen, müssten sehr hohe pH-Werte (>12) eingestellt werden. Da diese Bedingungen mit den meisten auf Silica basierten chromatographischen Säulen nicht kompatibel sind, wird in der Praxis meistens auf die Kationaustausch-Chromatographie zurückgegriffen.

Weitere wichtige Chromatographie-Verfahren sind die **hydrophile Interaktionschromatographie** (HILIC) und die **Affinitätschromatographie** (AC). Sie werden oft bei der Analyse posttranslational modifizierter Proteine angewendet. HILIC wird vor Allem zur Charakterisierung von Glycoproteinen zusammen mit oder als Alternative zur RPLC eingesetzt.^[82] Allerdings kämpft das Verfahren mit Problemen, wie der irreversiblen Adsorption von Proteinen an dem HILIC-Material sowie der möglichen Denaturierung von Proteinen durch die hohen Anteile (> 70 %) an organischen Lösemitteln der mobilen Phase. Mithilfe der AC werden oft Proteine mit Phosphorylierungen, Glycosylierungen oder oxidierten Cysteinen angereichert, wenn deren Vorkommen aufgrund niedriger bzw. dynamischer Stöchiometrie unter dem Detektionslimit des Massenspektrometers liegt.^[83]

1.3.2 Identifikation von Proteinen via Massenspektrometrie

Mit dem Aufkommen der zwei neuen Ionisierungsmethoden ESI (Elektrospray Ionisation) und MALDI (*matrix-assisted laser desorption ionization*, Matrix-unterstützte Laser-Desorption/Ionisation) Ende der 1980er Jahre, wurde die bis dahin genutzte Methode zur Proteinsequenzierung, der Edman-Abbau, nach und nach abgelöst.^[84-89] Durch die Fortschritte in der Technik der beiden MS-Methoden und der enorm schnell anwachsenden Informationen über Sequenzdaten aus den Genom-Projekten, wurde die Basis einer komplett neuen, sensitiven und vergleichsweise schnellen Proteinanalyse geschaffen. In den letzten Jahren wurden dafür die unterschiedlichsten Strategien hinsichtlich Probenvorbereitung, Fraktionierung, MS Datenerfassung, Quantifizierung und Datenanalyse

entwickelt und auf die unterschiedlichsten biologischen Fragestellungen angepasst.^[90] Dies gilt ebenso für die MS Geräte, welche durch den stetigen technischen Fortschritt eine immer bessere Massengenauigkeit, Empfindlichkeit, dynamischen Detektionsbereich und analytische Stabilität aufweisen.^[91] Ein Massenspektrometer besteht aus einer Ionenquelle, einem Analysator und einem Detektor (siehe Abbildung 6). Jedes dieser Bauteile gibt es mit verschiedenen Funktionsprinzipien, die prinzipiell frei kombinierbar sind. Aufgrund technischer Gegebenheiten und den einzelnen Vor- und Nachteilen der Komponenten existieren allerdings einige bevorzugte Kombinationen.

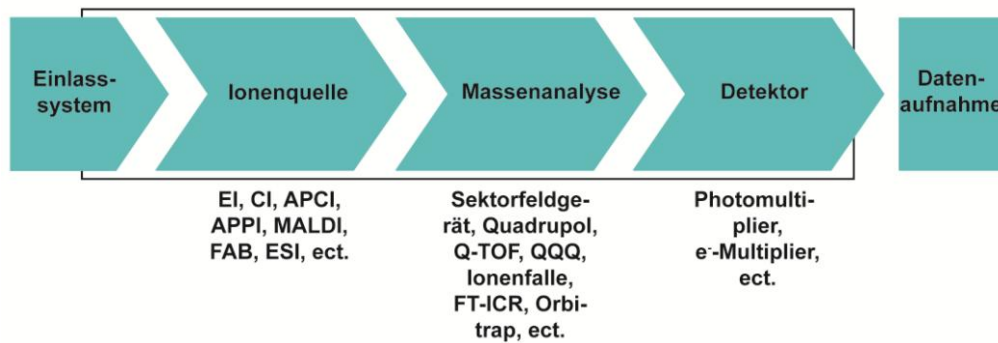


Abbildung 6: Schematischer Aufbau von Massenspektrometern.

Grundsätzlich gibt es zwei **Herangehensweisen für die MS-basierte Analytik von Proteinen**, die sogenannten „Bottom-up“- und „Top-down“-Ansätze. In Top-down-Analysen werden intakte Proteine aus der Probe isoliert und direkt massenspektrometrisch analysiert. Beispielsweise können so mit der Tandem-Massenspektrometrie Proteine *de novo* sequenziert werden. Bei der Tandem-Massenspektrometrie (auch MS/MS oder MS² genannt) erfolgt nach der ersten massenspektrometrischen Messung der durch die Ionenquelle erzeugten Ionen, deren Fragmentierung durch Zusammenstöße mit einem Inertgas, wie Stickstoff oder Argon. Die dadurch entstandenen, leichteren Fragmentionen werden erneut massenspektrometrisch gemessen. Bei der Anwendung des Verfahrens auf Proteine, erhält man durch die Fragmentierung unterschiedliche Serien an Peptidfragmentationen, anhand deren Überlappungen auf die Proteinsequenz rückgeschlossen werden kann.^[92] Die Fragmentierung mit anschließender MS-Messung kann auch mehrfach wiederholt werden, man spricht dann von MSⁿ-Experimenten. Bei der Bottom-up-Strategie werden Proteine für die Analyse zunächst enzymatisch oder chemisch zu Peptiden abgebaut, welche anschließend massenspektrometrisch nachgewiesen, identifiziert und gegebenenfalls quantifiziert werden können. Peptide sind aufgrund ihrer besseren Ionisierbarkeit im Massenspektrometer häufig einfacher zu analysieren als ganze Proteine.^[93] Eine Proteolyse kann allerdings auch nachteilig sein, wenn z.B. durch die erhöhte Komplexität einzelne Analyte aus der Probe durch Überlagerungen maskiert werden. Deshalb sind die bereits im vorigen Kapitel beschriebenen Verfahren zur

Proteinauftrennung (v.a. RPLC) auch für die Peptidauftrennung von äußerster Wichtigkeit und werden für die meisten dieser Ansätze verwendet.^[94] Die beiden Ansätze Top-down und Bottom-up können auch miteinander kombiniert werden, da sich ihre jeweiligen Vorteile gut ergänzen.^[95]

Da Proteine und Peptide polare, nicht-flüchtige und thermisch instabile Verbindungen sind, werden relativ milde **Ionisationstechniken** benötigt, um sie ohne allzu großer Zersetzung in die Gasphase zu überführen. Genau das wurde zum einen mit der Elektronenspray-Ionisation (ESI) und zum anderen mit der Matrix-unterstützten Laser-Desorption/Ionisation (MALDI) erreicht.

Bei der **MALDI**-Technik wird der Analyt aus einem kristallinen Zustand direkt in die Gasphase sublimiert.^[89, 96, 97] Dafür wird der Analyt mit einer stark UV-absorbierenden Matrix (z.B. 2,5-Dihydroxybenzoesäure (DHB), 3,5-Dimethoxy-4-hydroxy-trans-zimtsäure (Sinapinsäure) oder 3-Hydroxypicolinsäure (HPA)) entweder mit der „*dried droplet*“ oder der „*thin film*“ Methode vermischt.^[98, 99] Ein typisches Problem der MALDI-Methode ist die Inhomogenität von Proben. Dadurch können so genannte „*hot spots*“ entstehen, die starke Signale produzieren, während andere Bereiche wenig bis gar keine Signale liefern. Damit wird beispielsweise eine Quantifizierung stark erschwert.^[100] Dieses Problem kann mit neu entwickelten Ionenmatrices, wie 3,5-Dimethoxy-4-hydroxy-zimtsäure/Triethylamin oder 4-Nitroanilin/Buttersäure, teilweise durch Verbesserung der Reproduzierbarkeit der Signalintensität umgangen werden.^[100-102]

ESI ermöglicht die schonende Überführung ionisierter Moleküle aus einer flüssigen in die Gasphase mit Hilfe einer unter Spannung stehenden Kapillare. Das Ionenspray entsteht durch einen großen Spannungs- und Druckgradienten zwischen Ausgangskapillare der Ionenquelle (um die 5000 V bei Atmosphärendruck) und der Eingangskapillare des Massenspektrometers (um die 100 V bei Unterdruck auf Grund des im Analysator herrschenden Hochvakuums). Das elektrische Feld führt zu einer Ladungstrennung in der aus der Kapillare austretenden Flüssigkeit, wodurch sich deren Oberfläche zum Taylor-Konus deformiert.^[103, 104] Sobald die Oberflächenspannung durch die elektrostatischen Kräfte überwunden ist, emittiert von der Spitze des Konus ein feiner Flüssigkeitsstrahl in Richtung der Eingangskapillare.^[105] Dabei wird der größte Teil der positiven Ionen (oder im Negativ-Modus wären es negative Ionen) mit dem Strahl mitgerissen, der rasch in viele kleine Tröpfchen zerfällt.^[106, 107] Durch das allmähliche Verdampfen des Lösungsmittels kommt es zu einer Verdichtung der Oberflächenladung in den Tröpfchen bis das Rayleigh-Limit erreicht wird und sie sich durch Coulomb-Explosion in immer kleinere Tröpfchen aufteilen, die letztlich nur noch ein einziges geladenes Molekül enthalten, bevor sie die Eingangskapillare erreichen. Das Verdampfen des Lösungsmittels kann zusätzlich durch einen leichten Gasstrom (z.B. Stickstoff) gefördert werden. Wie bereits im vorigen Kapitel erwähnt, liegt der größte Vorteil der ESI-Technik darin, dass sie mit HPLC-Methoden gekoppelt werden kann. Standard-Applikationen arbeiten mit Flussraten im Bereich

von 10 – 500 $\mu\text{L}/\text{min}$. Gerade in der Proteomforschung müssen oft sehr geringe Probenmengen analysiert werden, die selbst unter diesen Maßstäben viel zu stark verdünnt werden müssten. In diesen Fällen kann auf nano-ESI-Quellen ausgewichen werden, die mit Flussraten im Bereich von 10 – 1000 nL/min arbeiten.^[105, 108] Hinzu kommt, dass nano-ESI-Quellen den Taylor-Konus schon bei etwa 2 kV erzeugen können, wodurch weniger Spannungsüberschläge auftreten, die das Spray unterbrechen.^[105] Auch die Form der Ausgangskapillare spielt für die Stabilität des Sprays eine Rolle, denn je kleiner die Spitze, desto definierter wird der Taylor-Konus erzeugt.^[109, 110] Dies ist vor allem dann wichtig, wenn eine nano-HPLC-Säule zugleich als Ausgangskapillare verwendet wird.

Das Herzstück eines jeden Massenspektrometers ist seine **Massenanalytik**-Einheit, welche die Ionen nach ihrem Verhältnis von Molekulargewicht zu Ladung (m/z) trennt. Es gibt verschiedene Analysatoren, die nach unterschiedlichen Prinzipien arbeiten und unterschiedliche Leistungsmerkmale (Massenbereich, Analysegeschwindigkeit, Auflösung, Sensitivität, Ionen­transmission, dynamischer Bereich, usw.) aufweisen. Die wichtigsten Analysatoren für Proteomanalysen sind im Folgenden aufgeführt.

Flugzeitmassenspektrometer (*time of flight*, TOF) trennen die Ionen nach ihrer Beschleunigung in einem elektrischen Feld durch die unterschiedlichen Flugzeiten der einzelnen m/z -Verhältnisse. Diese Art von Analysatoren wird meistens in Verbindung mit der MALDI-Technik verwendet. Vorteil der Methode ist der theoretisch unbegrenzte Massenbereich ($> 300 \text{ kDa}$). Nachteil ist das relativ schlechte Auflösungsvermögen R ($R = m/\Delta m$), bedingt durch die Dauer des Ionenbildungspulses, Größe des Volumens, in dem die Ionen gebildet werden und der kinetischen Energie der Ionen. Durch den Einsatz von elektrostatischen Reflektoren kann jedoch die Energiedispersion der Ionen korrigiert werden, so dass ein verbessertes Auflösungsvermögen R bis etwa 30000 erreicht werden kann.

Ein **Quadrupol** ist aus vier parallel zueinander verlaufenden Elektroden aufgebaut. An den jeweils diagonal gegenüberliegenden Elektroden werden zum einen eine konstante Gleichspannung und zum anderen eine hochfrequente Wechselspannung angelegt. Dadurch können nur Ionen mit einem bestimmten m/z -Verhältnis bis zum Detektor durch den Quadrupol wandern. Die Flugbahn anderer Ionen wird so destabilisiert, dass sie mit den Elektroden kollidieren. Durch Variation der Wechselspannung kann nach Ionen mit verschiedenen m/z -Verhältnissen gescannt werden. Vorteile von Quadrupolen liegen in ihrer einfachen Handhabung, ihrer Verlässlichkeit und den relativ niedrigen Preisen. Nachteile sind vor allem ihr begrenztes Auflösungsvermögen R (etwa bis 3000; bis 5000 nur unter starkem Empfindlichkeitsverlust möglich) und der kleine Massenbereich bis etwa m/z 4000.

Ionenfallen ermöglichen es, die Ionen durch elektromagnetische Felder in definierten Bereichen zunächst zu sammeln und dann massenabhängig zu sortieren. Die Technik eines 3D-Quadrupol-Ionenfallen-Massendetektors beispielsweise basiert auf der des Quadrupols, bietet aber zusätzlich die Möglichkeit, gefangene Ionen wiederholt zu fragmentieren sowie Massenspektren von jeder Fragmentationsstufe aufzunehmen und ermöglicht so MS^n -Analysen. Durch Anlegen verschiedener Potentiale an eine zur Mitte hin verjüngten Ringelektrode und zwei hyperbolische Endkappen wird ein dreidimensionales Wechselfeld erzeugt, das die Ionen auf eine Achter-förmige Bahn lenkt. Eine dünne Helium-Atmosphäre von bis zu einem Torr unterstützt durch die Aufnahme von kinetischer Energie die Abkühlung und Stabilisierung der Ionen oder durch hochenergetische Zusammenstöße deren Fragmentierung. Ein Nachteil besteht darin, dass die Anzahl der zu akkumulierenden Ionen begrenzt ist und durch Überschreiten der Begrenzung Raumladungseffekte auftreten können, die die gesamte Leistung der 3D-Ionenfalle beeinträchtigen. Dies wurde durch die Entwicklung der linearen Ionenfalle (LTQ, *linear trap quadrupole*) behoben.^[111, 112] Ihr Aufbau ähnelt dem eines Quadrupols, jedoch wird an die vorderen und hinteren Enden der Elektroden ein höheres Potenzial als am mittleren Teil angelegt und die Ionen bewegen sich auf axialen Bahnen im zweidimensionalen Hochfrequenzfeld. Dadurch steht den Ionen, im Vergleich zur 3D-Ionenfalle, ein größeres Volumen zur Verfügung. Raumladungseffekte können somit weitestgehend vermieden und die Empfindlichkeit gesteigert werden. Allerdings fällt die Möglichkeit zur mehrfachen Fragmentierung weg, weshalb sie oft zusammen mit anderen Techniken (Quadrupolen oder Kollisionszellen) kombiniert wird. Da der Auslass der Ionen sowohl axial als auch radial erfolgen kann, kann die LTQ sogar mit mehreren Analysesystemen gleichzeitig kombiniert werden. Vor allem wegen ihrer Schnelligkeit sind sie für Hochdurchsatzanwendungen in der Proteomik sehr beliebt.^[113, 114]

Die Technik mit dem bisher höchsten Auflösungsvermögen von R bis zu 10^8 bei einer gleichzeitig sehr guten Massengenauigkeit von weit unter 5 ppm ist die **Fourier-Transformation Ionen-Zyklotron-Resonanz (FT-ICR)**.^[115, 116] Hierbei wird ausgenutzt, dass Ionen in einem starken Magnetfeld eine kreisförmige Bewegung lotrecht zu den Magnetfeldlinien ausführen, deren Frequenz abhängig vom m/z-Verhältnis des Ions ist. Kommt es beim Einstrahlen eines elektrischen Wechselfeldes zum Resonanzfall mit der Kreisfrequenz der Ionen, vergrößert sich ihr Zyklotronradius durch Aufnahme von Energie aus dem Wechselfeld. Dieser Vorgang führt zu messbaren Signalen, die dann fouriertransformiert ausgelesen werden können. Der Vorteil gegenüber anderen Analysatoren liegt darin, dass die Ionen auf keinen Detektor treffen müssen, sondern ihre Frequenz aus kurzer Distanz gemessen werden kann. Außerdem müssen die Massen weder räumlich noch zeitlich getrennt werden; während einem Detektionsintervall werden alle Ionen gleichzeitig gemessen. Dies können bis zu 10^7 Ionen sein bevor Ladungswechselwirkungen auftreten und die Detektionsgrenze erreicht ist. Die Auflösung kann entweder durch Erhöhung der Magnetfeldstärke oder der Detektionslänge

verbessert werden. Ein Nachteil ist, dass ausreichend starke Magnetfelder (3,5 – 20 T) nur durch einen supraleitenden Magneten erzeugt werden können, wodurch Anschaffungs- und laufende Kosten einen erheblichen Aufwand darstellen.

Die **Orbitrap** wurde im Jahr 2000 von Makarov entwickelt und ist seit 2005 kommerziell erhältlich.^[117, 118] Damit ist sie die jüngste Entwicklung unter den Ionenfallen-Techniken. Das Prinzip ähnelt dem des FT-ICR, nur dass die Orbitrap mit einem elektrostatischen Feld statt einem Magnetfeld arbeitet. Sie besteht aus einer äußeren Pinolen-förmigen Elektrode mit einer koaxialen spindel-artigen Elektrode im Inneren. Aufgrund elektrostatischer Anziehung, die durch die Trägheit der Ionen ausbalanciert wird, bewegen sich die Ionen auf elliptischen Kreisbahnen um die zentrale Elektrode herum und schwingen dabei in axialer Richtung. Diese Oszillation bewirkt einen Bildstrom in der äußeren Elektrode, der aufgezeichnet wird und über Fourier-Transformation das Massenspektrum wiedergibt.^[116] Orbitrap-Massenspektrometer zeichnen sich durch eine sehr hohe Massenauflösung und –genauigkeit mit einem guten dynamischen Bereich aus.

Viele der genannten Techniken werden mit einer oder sogar mehreren weiteren Techniken kombiniert, zum einen, weil sie oft alleine für bestimmte Fragestellungen einfach nicht ausreichen, zum anderen, weil sie erst in Kombination ihr volles Potenzial entfalten können. Dabei haben sich vor allem durch technische Gegebenheiten, und anwendungsbezogene Vorlieben für bestimmte Kombinationen, **Hybridgeräte** ergeben. Tabelle 1 gibt eine grobe Übersicht über verschiedene Geräte, deren Eigenschaften und Anwendungsbereiche. Die lineare Ionenfalle wird z.B. oft mit FT-ICR- (LTQ-FT-ICR) oder Orbitrap-Analysatoren (LTQ-Orbitrap) verbunden, um als Schnittstelle die Ionen zu sammeln, zu sortieren, zu fragmentieren und nach Bedarf an andere Kammern weiterzugeben.^[91, 119-121] Somit kann die die Schnelligkeit und Sensitivität der LTQ mit der hohen Auflösung und Massengenauigkeit des FT-ICRs bzw. der Orbitrap genutzt werden.^[122-125] Als Ionenquellen dienen in dieser Konstellation dann oft ESI oder MALDI.^[126]

Tabelle 1: Übersicht über beliebte Massenspektrometer in der Proteomforschung und ihre charakteristischen Parameter.

MS-Gerät	Anwendung	Auflösungsvermögen R	Massengenauigkeit	Sensitivität	Dynamischer Bereich	Scan-Rate
LTQ	Bottom-up Proteinidentifikation, Hochdurchsatz-Analysen	2000	100 ppm	Femtomol	10 ⁴	schnell
LTQ-Orbitrap	Proteinidentifikation und quantifizierung, PTM-Identifikation	10 ⁶	2 ppm	Femtomol	10 ⁴	mittel
LTQ-FT-ICR	Bottom-up und Top-Down Proteinidentifikation und -quantifizierung, PTM-Identifikation	>10 ⁸	< 2 ppm	Femtomol	10 ⁴	langsam
Q-TOF	Bottom-up und Top-Down Proteinidentifikation, PTM Identifikation	50000	<1ppm	Attomol	10 ⁶	schnell
MALDI-TOF	Bottom-up und Top-Down Proteinidentifikation	30000	2 ppm	Attomol	10 ⁴	schnell

Heutige Elektrospray-Massenspektrometer mit einem dynamischen Detektionsbereich von 4-5 Größenordnungen, zeigen eine vom Analyten abhängige Nachweisgrenze von wenigen Attomol. Geringere Mengen eines Analyten können durch entsprechende Anreicherungsverfahren nachgewiesen werden. In proteomischen Analysen stellen oft sehr häufig vorkommende (hoch abundante) Proteine ein Problem dar, weil sie Signale von weniger stark vertretenen Proteinen maskieren und v.a. in Bottom-down Ansätzen zu einer erhöhten Komplexität der Probe führen können. Deshalb werden verschiedene affinitäts-basierte Ansätze wie Immunopräzipitations- oder *pull down* Assays angewendet, um niedrig konzentrierte Proteine anzureichern. Zum Teil beruhen diese Strategien auf denselben Methoden, die zum Nachweis von Protein-Protein-Wechselwirkungen verwendet werden. Diese werden in Kapitel 1.3.6 näher vorgestellt.

1.3.3 Quantitative MS-Ansätze

Es gibt ein breites Spektrum an Massenspektrometrie-basierten Quantifizierungsmethoden, die von markierungsfreien Methoden („*label free quantification*“, LFQ), über das Einbringen von externen isotopenmarkierten Standards bis hin zu isotopenmarkierten Analyten als interne Standards reichen. Je nach Anwendungsgebiet haben die verschiedenen Methoden ihre Vor- und Nachteile, welche in zahlreichen Übersichtsartikeln, wie z.B. von Bantscheff *et al.* zusammengefasst sind.^[127-129]

LFQ-Methoden beruhen auf LC-MS und LC-MS/MS Messungen und nutzen die Tatsache, dass die Konzentration eines Analyten mit der Ionenmenge korreliert.^[130] Dadurch ist eine Quantifizierung über den Vergleich von Signalintensitäten möglich. Eine weitere Methode ist das Zählen von identifizierten MS/MS-Spektren.^[131] Dies wird durch die lineare Korrelation zwischen Proteinabundanz und der Anzahl an generierten MS/MS-Spektren ermöglicht.^[132] LFQ-Methoden sind relativ beliebt, da keine teuren Standards, Markierungsreagenzien oder Markierungsschritte im Versuchsablauf benötigt werden. Ein weiterer Vorteil, neben den finanziellen und zeitlichen Aspekten, liegt darin, dass keine Artefakte auftreten, die bei zusätzlichen Prozessierungen der Proben eingebracht werden könnten. Des Weiteren wird die Komplexität der Spektren nicht durch das Einführen von Modifikationen erhöht, die Anzahl der Experimente ist nicht beschränkt und ein hoher dynamischer Bereich dieser Quantifizierungsmethode wurde beschrieben.^[128] Allerdings gelten LFQ-Methoden als zu vage und werden sehr kontrovers diskutiert. Robuste Ergebnisse erhält man nur durch eine Vielzahl an Spektren eines Proteins, wobei experimentelle Variationen die Quantifizierung maßgeblich beeinflussen können.^[128, 133] Nicht zuletzt erschweren die physikochemischen Eigenschaften (Masse, Ladung, Hydrophobizität, Modifikationen, u.a.) der Proteine/Peptide den Vergleich untereinander. Ihr Verhalten kann damit im LC- sowie im MS-System sehr unterschiedlich ausfallen, wodurch sowohl Signalintensitäten als auch die Anzahl der generierten Fragmentspektren betroffen sind. Diese Probleme können durch Zugabe einer externen Referenz, beispielsweise in Form von isotope-markierten synthetischen Proteinen oder Peptiden, umgangen werden. Genau dieses Prinzip wird bei der AQUA-Methode (*absolute quantification of proteins*), häufig in Kombination mit SRM/MRM (*single/multiple reaction monitoring*) Messungen, angewendet.^[134-136] Durch Zugabe schwerer isotope-markierter Standardpeptide mit bekannter Konzentration wird eine absolute Quantifizierung gemäß der Isotopenverdünnungstheorie möglich. Die Theorie besagt, dass ein Peptid, das mit stabilen Isotopen markiert wurde, chemisch identisch mit seinem entsprechenden natürlich vorkommenden Peptid ist und sich beide Peptide auch identisch während der LC-MS-Analyse verhalten.^[137] Vorausgesetzt, dass ein Massenspektrometer die Massendifferenz zwischen markiertem und unmarkiertem Peptid unterscheiden kann, wird die Quantifizierung durch den Vergleich der jeweiligen Signalintensitäten möglich. Allerdings müssen Effekte berücksichtigt werden, die vor der Zugabe der isotope-markierten Standards entstehen und das Ergebnis verfälschen können. Beispiele solcher Effekte sind ein unvollständiger Proteinverdau, unspezifische Protolysen durch endogene Proteasen, Agglutination der Analyte oder auch zelluläre Prozesse wie posttranslationale Modifikationen.

Eine dritte Möglichkeit zur Quantifizierung stellt die interne Isotopenmarkierung der Proteine selbst dar. Diese kann metabolisch, enzymatisch oder chemisch erfolgen. Metabolische Verfahren haben den Vorteil, dass die Markierung zum frühestmöglichen Zeitpunkt eingebracht werden kann, nämlich

während des Zellwachstums und der Zellteilung. Die SILAC (*stable isotope labeling by amino acids in cell culture*) Methode verwendet dafür normale Zellkulturmedien parallel zu Medien mit Aminosäuren, die schwere Isotopen wie ^{13}C oder ^{15}N enthalten.^[138, 139] So kann das Proteom von zwei unterschiedlich behandelten Zellkulturen, z.B. Knockout- gegenüber Wildtyp-Zellen, relativ zueinander verglichen werden. Da die unterschiedlich markierten Zellen vor jeglichen weiteren Arbeitsschritten zusammengegeben werden, können Fehlerquellen durch präparative Probenvorbereitung minimiert werden. Obwohl einige Studien mit Pflanzen, *C.elegans*, *Drosophila melanogaster* sowie in Maus- und Rattenmodellen zeigen konnten, dass die Methode auch auf höhere Organismen übertragbar ist, stehen die Kosten an isotoopenmarkierter Nahrung sowie die Dauer bis zur vollständigen Markierung meist in keiner Relation zu den Informationen, die aus einem solchen Experiment gewonnen werden können.^[140-142] Darüber hinaus geht der Vorteil des frühen Zusammengebens verloren, wenn die unterschiedliche Behandlung erst nach der Zellyse mit den extrahierten Proteinen stattfindet, wie z.B. im Falle von Studien zu Protein-Protein-Wechselwirkungen.

Bei der enzymatischen Markierung von Proteinen/Peptiden wird in Bottom-up Ansätzen während des enzymatischen Verdau durch die zugesetzten Proteasen ^{18}O aus schwerem Wasser eingebaut.^[143] Da jedoch die Markierung der Peptide in vielen Fällen zu unregelmäßig (unvollständig oder doppelt) verläuft, wird oft alternativ auf eine chemische Markierung der Proteine oder Peptide zurückgegriffen.^[144]

Hierbei dienen drei verschiedene funktionelle Gruppen der Peptide am häufigsten als Angriffspunkt für die Einführung des chemischen Markers: Thiole (Cysteine), primäre Amine (N-Terminus und ϵ -Aminogruppe des Lysins) und Carboxylgruppen (C-Terminus, Asparagin- und Glutaminsäuren).^[129] Die erste Methode zur Quantifizierung von Proteinen mithilfe chemischer Markierung beschreibt Gygi *et al.* 1999.^[145] Er entwickelte das ICAT (*isotope-coded affinity tag*) Reagenz, das spezifisch mit Cysteinen reagiert. Außerdem enthält es eine Polyether-Region mit entweder acht oder keinen Deuterium-Atomen sowie eine Biotingruppe, die eine Affinitätsreinigung der markierten Peptide erlaubt. Andere Thiol-spezifische Reagenzien enthalten oft Halogen-substituierte Carbonsäuren oder Amide oder verwenden die Michael-Additions-Reaktion an Carbonylgruppen für das Markieren.^[146-151] Da Cystein eine relativ seltene Aminosäure ist und eine durchaus relevante Anzahl an Proteinen existiert, die kein oder nur sehr wenige Cysteine enthalten, sind Reagenzien dieser Sorte nur begrenzt einsetzbar.

Aminogruppen-spezifische Reagenzien funktionieren meist über N-Hydroxysuccinimid- (NHS) oder andere Aktivester und Säureanhydride. Die prominentesten Beispiele sind: ICPL (*isotope-coded protein label*), iTRAQ (*isotope tags for relative and absolute quantification*), TMT (*tandem mass tags*)

und Essigsäure-/Bernsteinsäureanhydrid.^[152-159] Derzeit gehören die beiden chemischen Markierungsmethoden iTRAQ und TMT zu den populärsten, da sie nicht nur mit primären Aminogruppen reagieren, sondern auch zu den sogenannten isobaren Markern gehören. Isobare Markierungsreagenzien sind so aufgebaut, dass Proteine oder Peptide die einmal „schwer“ und einmal „leicht“ markiert wurden zunächst dieselbe Masse besitzen. Erst durch eine Fragmentierung im Massenspektrometer kommen die unterschiedlichen Markierungen zum tragen und die Analyten können unterschieden werden. Bei der Fragmentierung entstehen nämlich Reporterionen mit unterscheidbaren Massen im unteren Massenbereich eines MS/MS-Spektrums über deren Intensitäten die Quantifizierung möglich ist. Einer der größten Vorteile isobarer Quantifizierungsreagenzien ist die Möglichkeit multiplexe Analysen durchzuführen, d.h. mehrere biologische Proben auf einmal zu analysieren. Durch die unterschiedliche Kombination von schweren Kohlenstoff-, Stickstoff- und Sauerstoffisotopen innerhalb des Reagenzes (siehe Abbildung 7) können derzeit bis zu acht (iTRAQ) bzw. sogar zehn (TMT) unterschiedliche Markierungen erzeugt werden.^[160-162] Der springende Punkt dabei ist, dass anders als bei den übrigen Methoden, die Komplexität bei der LC-Trennung und im Massenspektrum der Peptide aufgrund identischer physikochemischer Eigenschaften nicht erhöht wird. Das erlaubt die simultane Bestimmung von Identität und relativer Häufigkeit der Peptide. Damit können biologische Systeme beispielsweise über mehrere Zeitpunkte oder nach unterschiedlichen Behandlungen mit nur einer massenspektrometrischen Analyse verglichen werden, was auch eine erhebliche Zeitersparnis bedeuten kann. Die Kehrseite ist, wie bei allen chemischen Markierungsverfahren, ihre Anfälligkeit gegenüber Nebenreaktionen, die zu unerwarteten Produkten führen und die Quantifizierung negativ beeinflussen könnten. Dies stellt vor allem an die Probenvorbereitung hohe Ansprüche.^[163, 164]

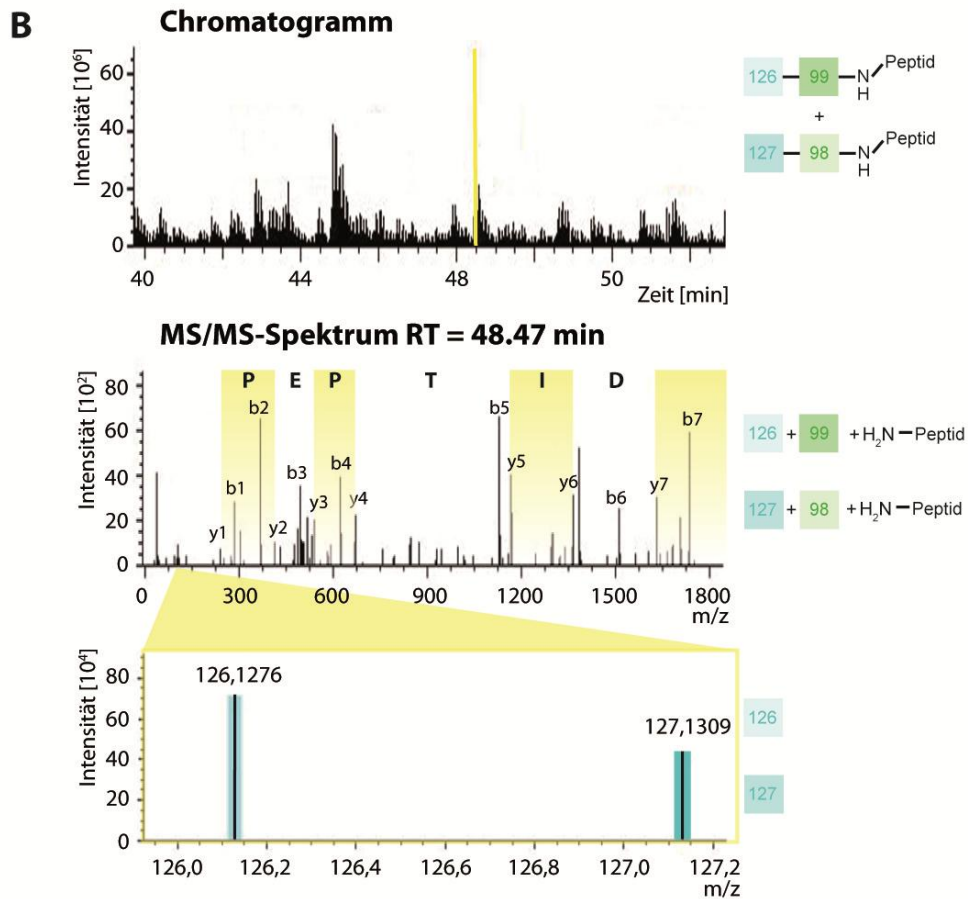
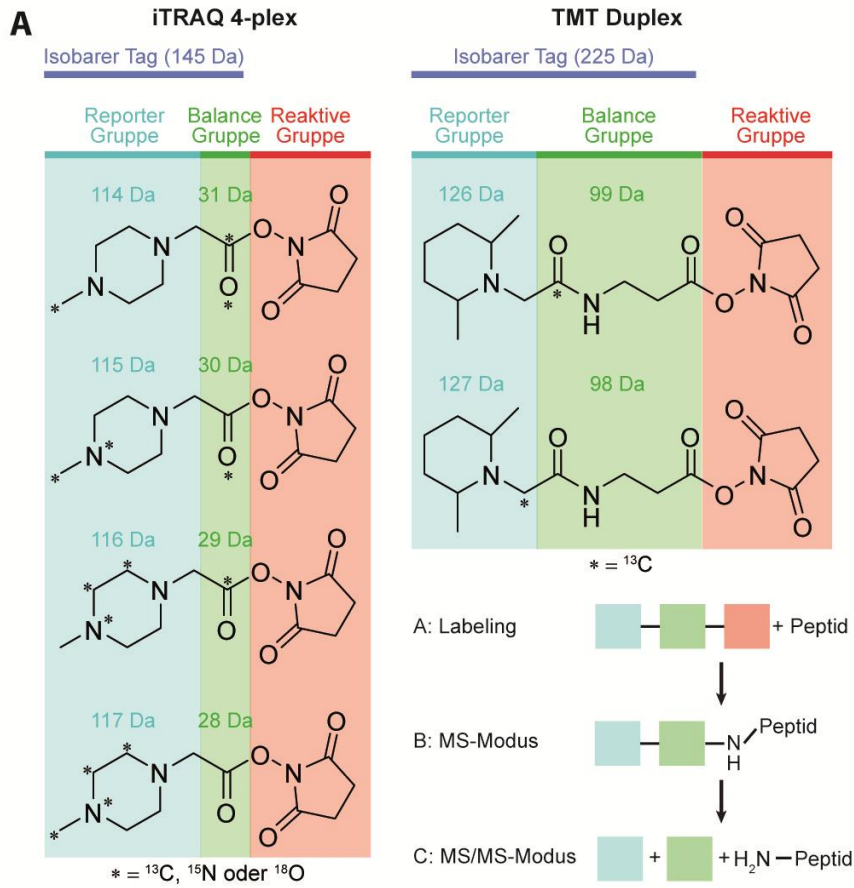


Abbildung 7: Quantifizierungsprinzip mit isobaren Markierungsreagenzien am Beispiel von iTRAQ 4-plex und TMT-Duplex Reagenzien. (A) Die Isobaren Markierungsreagenzien iTRAQ und TMT bestehen aus einer reaktiven Gruppe (rot), einer Balance Gruppe (grün) und einer Reporter Gruppe (blau). Reporter und Balance Gruppe haben zusammen eine Masse von 145 Da (iTRAQ) bzw. 225 Da (TMT). Durch unterschiedliche Isotopenverteilung besitzen die Reporter Gruppen Massen, die sich um je 1 Da voneinander unterscheiden. Dieser Unterschied wird durch die jeweilig zugehörige Balance Gruppe ausgeglichen. Durch Reaktion mit einem Peptid wird dieses am N-Terminus markiert. Im MS-Modus weisen zwei identische Peptide, die mit unterschiedlichen Reagenzien markiert wurden, dieselben Massen auf. Erst im MS/MS-Modus wird die Balance Gruppe in einer „Neutral-Verlust-Reaktion“ abgespalten und die Reporter Gruppen können unterschieden werden. (B) Schema zur Auswertung der massenspektrometrischen Messung eines mit TMT-Duplex Reagenzien markierten Peptids mit der fiktiven Aminosäuresequenz „PEPTID“. In dem theoretischen Chromatogramm (oben) wurde der Peak mit der Retentionszeit (RT) des Peptids markiert. In dem entsprechenden MS/MS-Spektrum (Mitte) werden die passenden b- und y-Ionen via Datenbankabgleich zugeordnet. Im unteren MS/MS-Spektrum ist der Massenbereich der Reporterionen vergrößert dargestellt. Über die Intensitäten der Reporterionen erfolgt dann die relative Quantifizierung des Peptids.

1.3.6 Crosslinking

Wie in Kapitel 1.1.1 bereits ausführlich am Beispiel von Histonen beschrieben, sind die meisten Proteine für die Erfüllung ihrer biologischen Funktion auf andere Proteine oder Biomoleküle angewiesen. Solche Proteine weisen dann eine spezifische Affinität zu den entsprechenden Liganden auf und es kommt zu einer Wechselwirkung bzw. Bindung. Ein zentrales Ziel der heutigen Proteomik ist die Erforschung welche Proteine auf welche Weise miteinander interagieren. Deshalb wurden einige spezielle Techniken entwickelt, die dabei helfen sollen Interaktionspartner ausfindig zu machen. Standardmethoden sind in Tabelle 2 aufgeführt.

Tabelle 2: Standardmethoden zur Untersuchung von Protein-Wechselwirkungen.

Methode	Beschreibung
Co-Immunopräzipitation (Co-IP)	Proteine mit Interaktionspartnern aus einem Proteingemisch werden über einen spezifischen Antikörper an eine Matrix gebunden und präzipitiert.
Affinitätschromatographie	Chromatographisches Trennverfahren unter Verwendung einer Protein-konjugierten stationären Phase zur Isolierung interagierender Proteine in der mobilen Phase. ^[165]
Hefe-Zwei-Hybrid-System	Verfahren unter Verwendung von Fusionsproteinen, die nach Interaktion einen funktionsfähigen Transkriptionsfaktor in der Hefezelle bilden. ^[166]
<i>Pull down Assays</i>	Proteine werden direkt (ohne Antikörper, vgl. Co-IP) an eine Matrix gebunden. Nach der Inkubation mit einem Proteingemisch können sie mitsamt Interaktionspartnern isoliert werden.
Far-Western-Blot Analyse	Proteine werden nach elektrophoretischer Trennung anstatt mit einem spezifischen Antikörper (vgl. Western-Blot) mit einem spezifischem (und markiertem) Protein inkubiert. ^[167]

Fortsetzung Tabelle 2: Standardmethoden zur Untersuchung von Protein-Wechselwirkungen.

Methode	Beschreibung
<i>Proximity ligation</i> Assay (PLA)	Verfahren unter Verwendung zweier spezifischer Antikörper, die mit einzelsträngigen Oligonukleotiden konjugiert sind und bei Interaktion der Zielproteine eine Amplifikation ligierter Oligonukleotide ermöglichen. ^[168]
<i>Label transfer</i> Protein Interaktions-Analyse	Ein markiertes Protein überträgt bei Interaktion ein Signalmolekül auf das andere unbekannte Protein.
Crosslinking Protein Interaktions-Assays	Interagierende Proteine werden kovalent aneinander gebunden.
Oberflächenplasmonenresonanzspektroskopie	Bei der Bestrahlung mit polarisiertem Licht einer Goldoberflächenmatrix, die mit Proteinen beladen ist, ändert sich im Falle einer Wechselwirkung mit Interaktionspartnern messbar der Reflexionswinkel des eingestrahnten Lichts.

Da viele Proteininteraktionen nur von sehr kurzer Dauer sind, ist bei vielen Ansätzen die Gefahr groß womöglich vorhandene Interaktionen zu „übersehen“, weil die Bindung nicht bis zur vollständigen Charakterisierung erhalten bleibt. In diesen Fällen bieten v.a. *Crosslinking Protein Interactions*-Assays eine gute Möglichkeit auch transiente Interaktionen zu untersuchen.^[169] Durch die hohen Reaktionsgeschwindigkeiten der üblichen funktionellen Gruppen (vgl. Kapitel 1.3.3 und unten), die für solche Assays verwendet werden, können selbst kurz oder schwach miteinander agierende Moleküle in einem Komplex erfasst werden, der stabil genug für die anschließende Isolation und Charakterisierung ist.

Um Proteine miteinander zu verbinden, benötigt man ein Reagenz, das mit wenigstens zwei reaktiven Gruppen ausgestattet ist. Tabelle 3 gibt eine Übersicht über die meist-verwendeten reaktiven Gruppen. Diese sind über ein Zwischenstück, dem sogenannten „*Spacer*“, verbunden. Die Struktur des Spacers kann ebenfalls den gegebenen Bedingungen oder besonderen Anforderungen angepasst werden. So gibt es sie in unterschiedlicher Länge von 0-100 Å, verzweigt, wodurch die Reagenzien mehr als zwei reaktive Gruppen besitzen können, und spaltbar, wodurch es ermöglicht wird, die Verknüpfung von Proteinen wieder aufzuheben. Desweiteren lassen sich Eigenschaften, wie die Löslichkeit des Reagenzes gut über die Struktur des Spacers beeinflussen. Natürlich können für ein Reagenz auch unterschiedliche reaktive Gruppen verwendet werden, man unterscheidet dann zwischen homo- und heterofunktionellen Reagenzien. Trägt beispielsweise ein Ende des Reagenzes einen NHS-Ester und das andere Ende ein Diaziridin, könnte zunächst ein Protein von Interesse über den NHS-Ester mit dem Reagenz verlinkt werden. Anschließend gibt man es zu einem Gemisch aus Proteinen, wartet bis sich ein Gleichgewicht eingestellt hat und aktiviert dann durch Belichtung das Diaziridin.^[170] Mittlerweile gibt es eine große Anzahl an Reagenzien mit den verschiedensten

Eigenschaften käuflich zu erwerben und die Anzahl an Synthesevorschriften ist noch viel größer. Die Auswahl eines geeigneten Crosslinker-Reagenz sollte daher vor jedem Experiment gut durchdacht sein.

Tabelle 3: Zielgruppen und häufig verwendete reaktive Gruppen bei der Biokonjugation.

Zielgruppe	Reaktivitätsklasse	Reaktive Gruppen
-NH ₂	Amin-reaktiv	NHS-Ester Imidoester Pentafluorphenylester Hydroxymethyl Phosphine
-COOH	Carbonsäure-reaktiv	Carbodiimide (z.B. EDC) als transienter Aktivator
-SH	Thiol-reaktiv	Maleimide Bromo- oder Iodoacetylene Pyridyldisulfide Thiosulfonate Vinylsulfone
-CHO	Aldehyd-/Carbonyl-reaktiv (z.B. oxidierte Zucker)	Hydrazide Alkoxyamine
Zufällig	Photoreaktiv, nicht selektiv	Diazirine Arylazide
-OH	Hydroxyl-reaktiv	Isocyanate
-N	Azid-reaktiv	Phosphine

1.3.4 Software und Datenbanken

Eine Folge der rasanten Entwicklungen in der MS-basierten Proteomforschung, sowohl was die Technik als auch die Methoden betrifft, ist die Erzeugung von riesigen Datenmengen, was eine manuelle Bearbeitung schlichtweg unmöglich macht. Dies gilt insbesondere für die weit verbreiteten Bottom-up Ansätze, bei denen komplexe Proteinproben zu Peptiden verdaut werden und anschließend via Tandem-Massenspektrometrie analysiert werden. Dabei werden je Probe mehrere tausend MS¹- und MS²-Spektren generiert. Zur Bewältigung dieser Datenflut wurden Suchmaschinen entwickelt, die anhand öffentlicher Protein-Datenbanken, wie der des „*The National Center for Biotechnology Information*“ NCBI (www.ncbi.nlm.nih.gov) und der „*universal protein database*“ UniProt (www.uniprot.org), theoretische Spektren kreieren, mit den gemessenen Spektren abgleichen und die Wahrscheinlichkeit einer Übereinstimmung von Spektren in Form eines „Score“-Wertes angeben. Dafür werden die Proteinsequenzen aus den Datenbanken einem *in silico* Verdau unterzogen und von den erhaltenen möglichen Peptidsequenzen wahrscheinliche Fragmentspektren vorausgerechnet, die sich ergeben würden, wenn diese Peptide in der Probe vorhanden wären. Die drei gängigsten Suchmaschinen derzeit sind SEQUEST, Mascot und MaxQuant/Andromeda.^[171-173] Die richtige Peptidsequenz aus tausenden von Spektren zu erkennen, stellt für die Suchalgorithmen eine

besondere Herausforderung dar, vor allem wenn die Spektren rauschen und Peaks fehlen. Deshalb müssen die Suchmaschinen einen Algorithmus aufweisen, der auch bewertet wie gut eine Sequenz dem Spektrum entspricht. Da die Wahrscheinlichkeiten der Übereinstimmungen von jedem Suchalgorithmus anders berechnet werden, kann ein und dieselbe Messung, mit unterschiedlichen Suchmaschinen ausgewertet, bisweilen recht unterschiedliche Ergebnisse liefern.^[174] Einer der Scores, den SEQUEST verwendet, misst, inwiefern experimentelle und theoretische Spektren übereinstimmen. Seine Kreuzkorrelation summiert dazu die Produkte der Intensitäten zwischen dem beobachteten Spektrum und einem vorhergesagten Spektrum der Sequenz auf. Die Robustheit der Korrelation wird zusätzlich durch einen Korrekturfaktor erhöht, der berechnet wird, indem die beiden verglichenen Spektren mehrfach relativ zueinander in m/z verschoben werden.

Mascot bewertet die Qualität der Übereinstimmungen statistisch über die Anzahl an identifizierten Peaks. Sein „Ions Score“ bezieht dabei die Anzahl an Fragmentationen ein, die im MS/MS-Spektrum gesucht werden, die Anzahl wie viele davon übereinstimmen, die Anzahl an Peaks, die innerhalb eines Spektrums oberhalb eines Intensitätsgrenzwertes beobachtet werden und die Anzahl an Peptidsequenzen im Vergleich zum Spektrum. Andromeda benutzt ein sehr ähnliches Verfahren. Die unterschiedlichen Herangehensweisen führen dazu, dass Peaks unterschiedlich beim Abgleich der Spektren einbezogen werden.

Die konträren Ansichten und Meinungen über die Art und Weise der Datenbank-Suche werden durch bestehende Softwaretools und Analysestrategien der verschiedenen Labore beeinflusst. Individuelle Datensätze, wissenschaftliche Fragestellungen, Suchmethoden und anschließende Datenprozessierung spielen alle samt eine Rolle in der Definition der optimalen Einstellungen für jedes Experiment.^[175] Nicht nur die Such-Algorithmen sind umstritten, sondern auch einige allgemeine Einstellungen, die in jedem Programm manuell angepasst werden können und die Ergebnisse maßgeblich beeinflussen können. Ein Beispiel dafür ist die Auswahl der Datenbank, die für die Suche verwendet werden soll. Üblicherweise verwendet man für den Abgleich eine Datenbank, die nur Proteinsequenzen des Organismus enthält, von der die Proteinprobe stammt. Daneben kann es sinnvoll sein Sequenzen klassischer Kontaminierungen der Datenbank hinzuzufügen. Dadurch wird sichergestellt, dass möglichst viele Peptide, die gemessen wurden, auch identifiziert werden können und somit das Risiko für falsch positive Übereinstimmungen reduziert wird. Die Verwendung von Datenbanken, die mehr als die Proteinsequenzen einer Spezies enthalten, erhöht allerdings auch die Anzahl an abzugleichenden Spektren und führt zu einer verlängerten Rechenzeit. Da falsch positive Ergebnisse auch bei der Verwendung passender Datenbanken nicht ausgeschlossen werden können, hat sich in der Praxis die Verwendung von sogenannten Köder-Datenbanken verbreitet.^[176] Dafür wird ein großes Set an Proteinsequenzen, von denen man weiß, dass sie falsch sind, wie z.B. zufällig

vertauschte Sequenzen, zur Datenbank hinzugefügt, wodurch eine Rate an falsch positiven Funden (*false discovery rate*; FDR) bestimmt werden kann. Der FDR-Wert wird dann typischerweise auf 1 – 5 % festgelegt, was bedeutet, dass 1 – 5 Proteine je 100 identifizierter Proteine falsch positiv sind. Die Verwendung und Angabe des FDRs ist in vielen Fachzeitschriften mittlerweile Voraussetzung für Publikationen.

Ein ebenfalls sehr wichtiger Parameter ist die Einstellung der Massentoleranz bei der Suchmaschine. Sie kann normalerweise separat für Vorläufer- und Fragmentationen eingestellt werden und basiert auf der Massengenauigkeit des jeweiligen zur Messung verwendeten Analysators im Massenspektrometer. Die Vorläuferionen-Massentoleranz spielt eine direkte Rolle für die Bestimmung welche Peptide mit den jeweiligen MS²-Spektren verglichen werden. Denn nur, wenn die Vorläuferionen-Masse eines Peptids innerhalb der definierten Massentoleranz liegt, wird sein MS²-Spektrum auch zum Abgleich mit dem theoretischen Fragmentspektrum des entsprechenden vorhergesagten Peptids herangezogen. Die Vorläuferionen-Massentoleranz hängt also im Allgemeinen von der Massengenauigkeit des MS¹-Spektrums ab. Bei Spektren, die mit TOF-Analysatoren aufgenommen wurden, liegt die Massengenauigkeit zum Beispiel um die 10 ppm, bei Ionenfallen um 2 ppm. Wird die Toleranzschwelle zu niedrig eingestellt, werden eventuell richtige Peptidsequenzen im Datenbankabgleich nicht berücksichtigt.^[177] Wird die Schwelle dagegen höher als nötig eingestellt, wird zum einen auf Grund der größeren Anzahl an möglichen Kandidaten die Suchzeit steigen und zum anderen die Sensitivität der Suche sinken, weil jedes vermeintlich übereinstimmende Peptid mit der größeren Anzahl an Kandidaten konkurriert, von denen jeder auch noch zufällig einen höheren Score-Wert besitzen könnte als das richtige Peptid. Selbst bei hochauflösenden Geräten kann die beobachtete Masse mehr als einige ppm von der wahren monoisotopischen Masse des Vorläuferions abweichen, wenn beispielsweise das Signal eines ¹³C Isotopes fälschlicherweise als monoisotopisches Signal durch die Peak-Erkennungssoftware des MS-Gerätes festgelegt wird. Bevor die hier angesprochenen oder die vielen weiteren Parameter und Filter für eine Datenbank-basierte Peptid Identifikation eingestellt werden, sollte bedacht werden, dass, unabhängig davon welche Suchmaschine eingesetzt wird, jedes Ergebnis eines Datenbankabgleichs im Wesentlichen auf statistischen Berechnungen beruht. Um signifikante Ergebnisse zu erhalten, spielt die Anzahl an möglichen Kandidaten und die Verteilung ihrer Score-Werte eine Schlüsselrolle. Je mehr Kandidaten es gibt und je deutlicher sich ein Kandidat mit seinem Score-Wert von den übrigen Kandidaten absetzt, desto höher ist die Wahrscheinlichkeit einer richtigen Zuordnung.

2. Zielsetzung

Das Ziel dieser Arbeit war es, Proteine zu identifizieren, die mit den vor kurzem in der DNA entdeckten Cytosin-Derivaten mC, hmC, fC oder caC wechselwirken. Hierfür sollten mit entsprechend modifizierten DNA-Oligomeren und Zellkernlysaten, auf Basis der Arbeiten von Malte Winnacker, Photoaffinitätsmarkierungs-Experimente mit anschließender massenspektrometrischer Analyse durchgeführt werden.^[178, 179] Um herauszufinden, welche DNA-bindenden Proteine spezifisch mit den Modifikationen wechselwirken, sollten diese über den Crosslinker Iodouracil kovalent an die DNA gebunden werden. Mithilfe einer Fluoreszenz-Markierung an der DNA sollten die möglichen Protein-DNA-Komplexe auf einem Gel sichtbar gemacht werden. Jedes Experiment sollte vergleichend mit unmodifizierter DNA (hinsichtlich der Cytosin-Derivate) durchgeführt werden, um unspezifische von spezifischen DNA-Bindern unterscheiden zu können.

Allerdings stellte sich bereits bei den ersten Versuchen heraus, dass die Methode von Malte Winnacker für ein Screening nach den gesuchten Proteinen nicht geeignet ist. Aufgrund des begrenzt verfügbaren Zelllysats wurden zu wenige Proteine an der DNA angereichert, um deutliche Unterschiede in den verschiedenen Ansätzen sichtbar machen zu können. Deshalb sollte ein sensitiveres Auslese-Verfahren zur Identifikation der kovalent gebundenen, spezifischen Binder entwickelt werden. Genauer gesagt, sollte eine neue Quantifizierungsmethode mit isobaren Markierungsreagenzien und ein neues Reagenz zur kovalenten, aber reversiblen, Bindung der Proteine an die DNA etabliert werden. Dadurch sollten Proteine, die direkt an die DNA gebunden werden, massenspektrometrisch von solchen Proteinen unterschieden werden können, die indirekt, beispielsweise über Proteinkomplexe, an die DNA binden. Schließlich sollte der neu entwickelte Ansatz in unterschiedlichen Zell- und Gewebeproben angewendet werden.

3. Ergebnisse und Diskussion

3.1 Methodenentwicklung für quantitative DNA-Protein-Interaktionsstudien mit Crosslinker

Lange Zeit galt die Methylierung von DNA als nicht reversibel, aber Forschungen der letzten 30 Jahre deuten stark darauf hin, dass es eine Umkehrung des Prozesses geben muss. Dennoch konnte ein Mechanismus zur aktiven Demethylierung noch nicht nachgewiesen werden. Die Entdeckung der Oxidationsstufen von Methylcytosin hmC, fC und caC führten zu der Hypothese, dass diese Zwischenstufen einer aktiven Demethylierung darstellen und weitere biologische Funktionen übernehmen könnten. Die wenigen bisherigen Studien dazu zeigten bereits, dass es spezifisch interagierende Proteine gibt, sie zeigen aber nicht wie und warum die Interaktionen zustande kommen. Ein Grund dafür ist, dass die in den Studien verwendeten Methoden eine möglicherweise entscheidende Informationsstufe auslassen. Denn sie geben keinen Aufschluss darüber, ob die identifizierten Proteine direkt oder indirekt mit der DNA wechselwirken. Wenn die Zahl möglicher Kandidaten für eine Verifizierung nicht genug eingeschränkt werden kann, wird die Erfolgsaussicht auf eine positive Verifizierung der entdeckten Interaktionen und deren weitere Untersuchung zur Art und Weise gemindert. In den nächsten Kapiteln wird die Entwicklung einer neuen Methode zum Crosslinken von Proteinen mit DNA-Oligomeren und deren Einsatz auf verschiedene Fragestellungen ausführlich dargestellt.

3.1.1 Photoaffinitätsmarkierungs-Experimente

Nachdem Malte Winnacker durch Photoaffinitätsmarkierungs-Experimente erfolgreich DNA-Reparatur-Enzyme an DNA-Sonden mit entsprechenden DNA-Schäden anreichern und massenspektrometrisch identifizieren konnte, sollte die Methode auf die neu entdeckten Cytosin-Derivate angewandt werden. Dafür wurden zunächst, wie von Malte Winnacker beschrieben, Anreicherungsexperimente von Proteinen aus Zellkernlysate an doppelsträngigen DNA-Sonden versucht, die sowohl ein Biotin-Tag als auch eine Fluorescein-Markierung trugen.^[178] Der Biotin-Tag diente dabei als Funktion zur Isolation der DNA-Sonden nach der Inkubation mit dem Zellkernlysate und damit zur Anreicherung der gebundenen Proteine. Die Fluorescein-Markierung andererseits sollte dazu dienen, die isolierten Sonden auf einem SDS-PAGE-Gel sichtbar zu machen. Desweiteren enthielt die DNA-Sonde (statt einem DNA-Schaden) mittig entweder eines der, wie in Kapitel 1.1.2 beschriebenen, modifizierten Cytosine oder unmodifiziertes Cytosin als Kontrolle. Diese Stelle wurde

im Gegenstrang, je in einem Abstand von zwei Basenpaaren, mit 5-Ioduracil (IU) flankiert, das als photoaktivierbarer Crosslinker DNA-affine Proteine kovalent binden sollte (siehe Abbildung 8). Nach 20 min Inkubation auf Eis mit Zellkernextrakt aus murinen Stammzellen (Matthias Truss) wurde belichtet und jeweils eine Probe für eine analytische SDS-PAGE genommen.

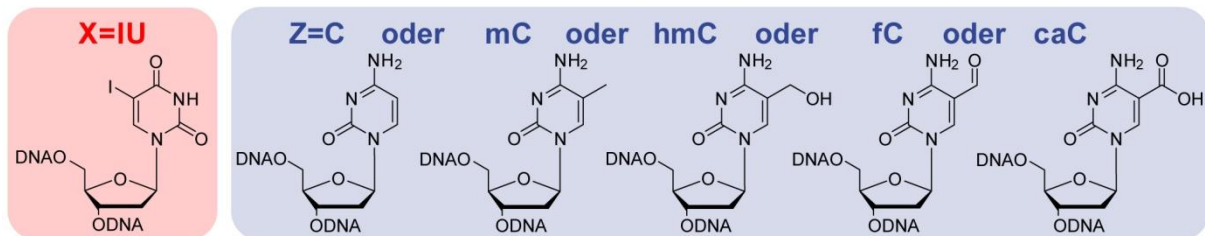


Abbildung 8: Verwendete DNA-Sonde mit Fluorescein-Tag (Flu), Biotin-Tag (Bio), 5-Ioduracil (X=IU, rot) und Cytosin (Z=C, blau) bzw. seine Oxidationsprodukte (Z=mC, hmC, fC oder caC, blau).

Wie aus Abbildung 9A hervorgeht, wurden kaum Proteine gefunden, die an diese DNA-Sonden gebunden waren und damit über die Fluorescein-Markierung hätten sichtbar sein müssen. Die wenigen, sehr schwach ausgeprägten Banden, stimmten für die verschiedenen DNA-Sonden überein. Damit konnten diese, als spezifisch an eines der modifizierten Cytosine bindenden Proteine, ausgeschlossen werden. Für diese Analysemethode wurde nur etwa je ein Sechstel der Gesamtproben verwendet. Mit den übrigen Proben sollte dann nach Anreicherung über den Biotin-Tag ein weiteres Gel präpariert werden, aus dem die Banden, die zur Kontrolle abweichen, ausgeschnitten und zur massenspektrometrischen Analyse gegeben werden sollten. Da jedoch das analytische Gel nach der Coomassie-Färbung (siehe Abbildung 9B) bereits zeigte, dass allgemein sehr wenig Proteine in den Proben vorlagen und diese aufgrund der „schmierenden“ Banden teilweise auch noch in der Probentasche des SDS-Gels ausgefallen zu sein schienen, wurde auf eine erneute gelelektrophoretische Auftrennung verzichtet und die Proben direkt nach der Anreicherung verdaut und massenspektrometrisch gemessen. Der Datenbank-Abgleich und die Auswertung erfolgte mit der Software Proteome Discoverer 1.0 (*Thermo Fisher*, Waltham, MA, USA) unter Verwendung einer *UniProt* Datenbank (*Mus Musculus* mit Kontaminanten, 2012). Wie aus Tabelle 4 hervorgeht, konnten auch hierdurch nur sehr wenige (C, mC, hmC, fC) bis gar keine (caC) spezifisch bindende Proteine identifiziert werden. Fehlende Replikate, die schlechte Sequenzabdeckung und die niedrigen Score-Werte der einzelnen identifizierten Proteine ließen eine weitere (statistische) Auswertung der Daten nicht zu.

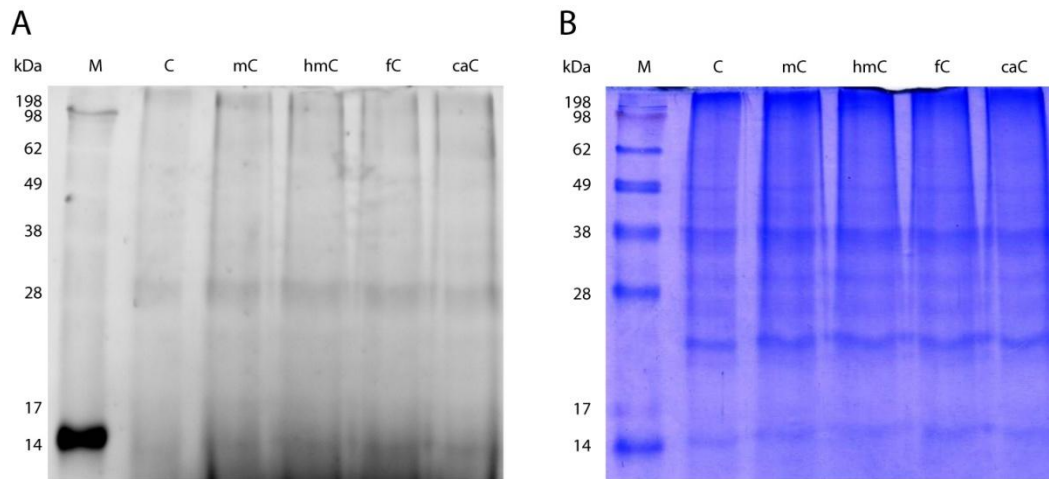


Abbildung 9: Analytische SDS-PAGE nach den *pull down* Versuchen mit Fluorescein-haltigen DNA-Sonden aus mESC-Kernextrakten. (A) Fluoreszenz-Aufnahme. (B) Coomassie Färbung. M = Marker (da kein spezieller Marker für Fluoreszenz-Aufnahmen verwendet wurde, sind unter (A) nur die Banden bei 14 und 98 kDa zu erkennen).

Tabelle 4: Ergebnisse der massenspektrometrischen Auswertung der *pull down* Versuche mit Fluorescein-haltigen DNA-Sonden. Kontaminanten wurden für die Darstellung herausgefiltert und es wurden nur Proteine verwendet, die von der Maus stammen. Wegen der fehlenden Quantifizierung sind je Ansatz nur Proteine dargestellt, die in keinem der anderen Ansätze identifiziert wurden.

UniProt Datenbanknr.	Genname	$\Sigma\#$ eindeutige Peptide	$\Sigma\#$ Peptide	Score	Abdeckung [%]	# Aminosäuren	MW [kDa]
C							
P09405	Ncl	1	1	3,35	1,98	707	76,7
Q7TPV4	Mybbp1a	1	1	2,48	0,97	1344	151,9
mC							
D6RH96	Tdeanc2	1	1	1,69	38,89	36	4,2
O64509	At2g02730/T20F6.13	1	1	0,00	4,71	276	30,0
Q8QZV0	Dedd2	1	1	1,63	3,94	330	36,8
Q8VGT1	Olf385	1	1	0,00	3,85	312	35,4
Q9CV82	2310003L06Rik	1	1	0,00	2,63	495	55,4
Q8R0Z6	Angptl6	1	1	1,68	2,41	457	51,1
Q91YR7	Prpf6	1	1	3,24	1,49	941	106,7
P63154	Crnk1	1	1	0,00	1,45	690	83,4
hmC							
P62900	Rpl31	1	1	1,84	11,20	125	14,5
Q921M7	Fam49b	1	1	1,73	4,01	324	36,8
P33610	Prim2	1	1	2,44	2,18	505	58,4
Q9D2V7	Coro7	1	1	0,00	1,74	922	100,7
fC							
G5E870	Trip12	2	2	3,71	1,04	2025	224,0
Q6P5D8	Smchd1	1	1	1,76	0,55	2007	225,5
caC							
-	-	-	-	-	-	-	-

Dieser Versuch zeigte eindeutig, dass die Methode von Malte Winnacker nicht auf die neue Fragestellung übertragbar ist. Da die interessierenden Zellstadien, wie zum Beispiel embryonale Stammzellen, nicht ohne erheblichen materiellen und finanziellen Aufwand verfügbar sind, musste die eingesetzte Zellkernlysatzmenge auf etwa ein Viertel der in den Versuchen mit DNA-Schäden eingesetzten Mengen reduziert werden. Dies führte bei dem analytischen Gel dazu, dass die Proteine wahrscheinlich in zu geringer Konzentration vorlagen, um deutliche Unterschiede in den Ansätzen erkennen zu können. Ebenfalls ist fragwürdig, inwiefern die Fluorescein-Markierung als zusätzliche Modifikation die Affinität der Proteine zu den oxidierten Cytosin-Derivaten beeinflusst. Schließlich ist Fluorescein als sperriges Addukt auch ein Substrat bei der Nukleotidexzisionsreparatur (NER).^[180] Malte Winnacker konnte zwar zeigen, dass durch die Fluorescein-Markierung weder die Struktur der DNA noch die Schadenerkennung durch Reparaturenzyme beeinflusst werden, allerdings können für die neue Fragestellung unerwünschte Wechselwirkungen oder eine sterische Hinderung der Proteine nicht ausgeschlossen werden. Deshalb wurde eine Möglichkeit gesucht, um die Fluorescein-Markierung gänzlich zu vermeiden und die Proben auf eine andere Weise vergleichbar zu machen. Da zur Identifikation der Proteine auf jeden Fall eine MS-Analyse durchgeführt werden sollte, bot es sich an, eine Quantifizierung der Proteine mit isobaren Markierungen einzuführen.

3.1.2 Validierung Isobarer Markierungsreagenzien

Die Photoaffinitätsmarkierungs-Experimente auf Basis der Fluorescein-Markierung stellten sich als ungeeignete Methode für die Untersuchung von Proteinen, die mit methylierter DNA oder den anderen Oxidationsderivaten von mC wechselwirken, heraus. Obwohl es damit in vorigen Experimenten gelang, spezifische Reparaturenzyme zu identifizieren, scheint die Methode nicht oder nur bedingt zur Identifikation von Proteinen mit niedrigen Expressionsleveln geeignet zu sein. Zudem ist die Präparation von Gelen nicht nur sehr zeitaufwendig, sondern kann auch auf Grund von Proteinen mit gleicher Mobilität den Vergleich von unterschiedlichen Proben erheblich erschweren.^[74]

Aus diesem Grund wurde nach einer anderen MS-verträglichen Möglichkeit zur Unterscheidung zweier Proben gesucht. Der Einsatz von Markierungsfreien Quantifizierungsmethoden (LFQ, *label-free quantification*), wie dem Zählen von Spektren, gilt als zu ungenau und schlecht reproduzierbar.^[127, 128] Dies betrifft vor allem Proteine, die wenig abundant sind, eine geringe Anzahl von Spektren aufweisen oder deren Konzentrationsunterschied in den zu vergleichenden Proben nur gering ausfällt.^[181] Eine bereits sehr erfolgreiche Strategie im Zusammenhang mit der Aufklärung epigenetischer Fragestellungen sind SILAC-Experimente (siehe Kapitel 1.3.3). Da diese jedoch sehr teuer und aufwendig sind, rückte ein ähnliches Verfahren in den Fokus: Isobare Markierungen. Diese

Methode nutzt dasselbe Prinzip von leicht und schwer markierten Proteinen, lässt sich aber im Gegensatz zu SILAC nachträglich über eine chemische Reaktion an den Proteinen oder Peptiden einführen und ermöglicht so eine relative Quantifizierung der Proben via Massenspektrometrie. Neben den beiden kommerziell erwerblichen Varianten iTRAQ und TMT gibt es noch einige Weitere, die sich teilweise noch in der Entwicklung befinden. Dabei weisen die verschiedenen Markierungsreagenzien unterschiedliche Eigenschaften auf, die anwendungsbezogen berücksichtigt werden müssen.

TMT Reagenzien, welche bisher als Duplex, 6plex und 10plex erhältlich sind, bieten ein ausgereiftes System zur isobaren Markierung von Proteinen oder Peptiden. Mit einem NHS-Ester als reaktive Gruppe, reagieren die Reagenzien vorzugsweise mit primären Aminen der Lysinseitenketten und N-Termini der Proteine bzw. Peptide und weisen deshalb ein hohes Identifikationspotenzial auf. Das am Caltech von Sohn *et al.* relativ neu entwickelte Reagenz CIT (*Caltech Isobaric Tag*), verspricht eine relativ günstige und einfache Synthese, bei einer gleichzeitigen theoretisch unbegrenzten Anzahl an multiplexen isobaren Tags.^[182]

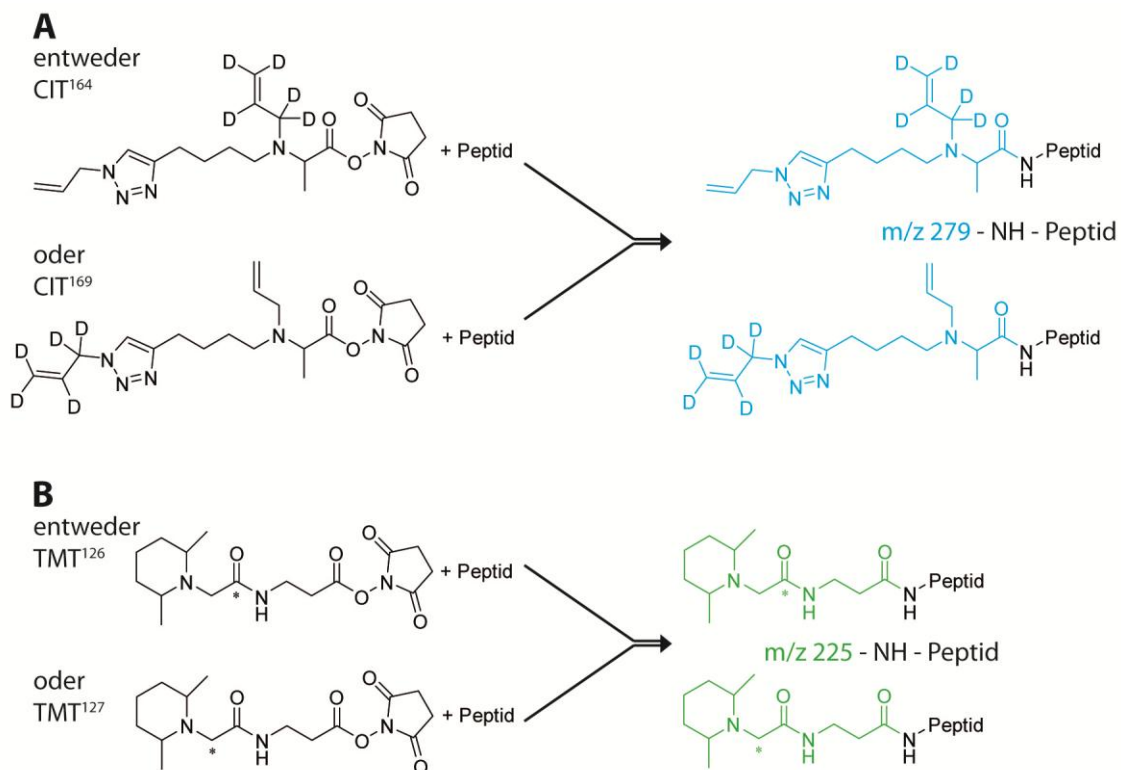


Abbildung 10: Strukturen der isobaren Duplex-Markierungsreagenzien CIT (A) und TMT (B). Peptide, die damit reagieren, weisen eine der Markierung entsprechend definierte erhöhte Masse auf. * = ¹³C

Für einen Vergleich zwischen den beiden Duplex-Reagenzien CIT (164, 169) und TMT (126, 127) wurde BSA (*bovine serum albumine*, Rinderserumalbumin) nach einem Standardprotokoll mit Trypsin verdaut und mit den jeweiligen isobaren Markierungsreagenzien nach Entwickler- bzw. Herstellerangaben markiert. Direkt im Anschluss wurden die markierten Proben und ein nicht markierter BSA-Verdau via MALDI-TOF-MS analysiert.

Die dabei aufgenommenen Spektren (siehe Abbildung 11) zeigen im direkten Vergleich eindeutig, dass in der Probe mit dem CIT-markierten BSA (siehe Abbildung 11, oben) Peptide mit einer um 279 Da erhöhten Masse, im Vergleich zu der Probe mit unmarkiertem BSA (siehe Abbildung 11, unten), gefunden werden konnten. Diese Massenverschiebung zeigt, dass Peptide mit den CIT-Reagenzien markiert wurden. Allerdings wurden gleichzeitig Peptide mit denselben Massen gefunden, wie in der Probe mit unmarkiertem BSA. D.h. die Peptide wurden nur teilweise markiert und die, von den CIT-Entwicklern beschriebene, gute Markierungseffizienz, konnte in diesem Versuch nicht bestätigt werden.

Ergebnisse und Diskussion

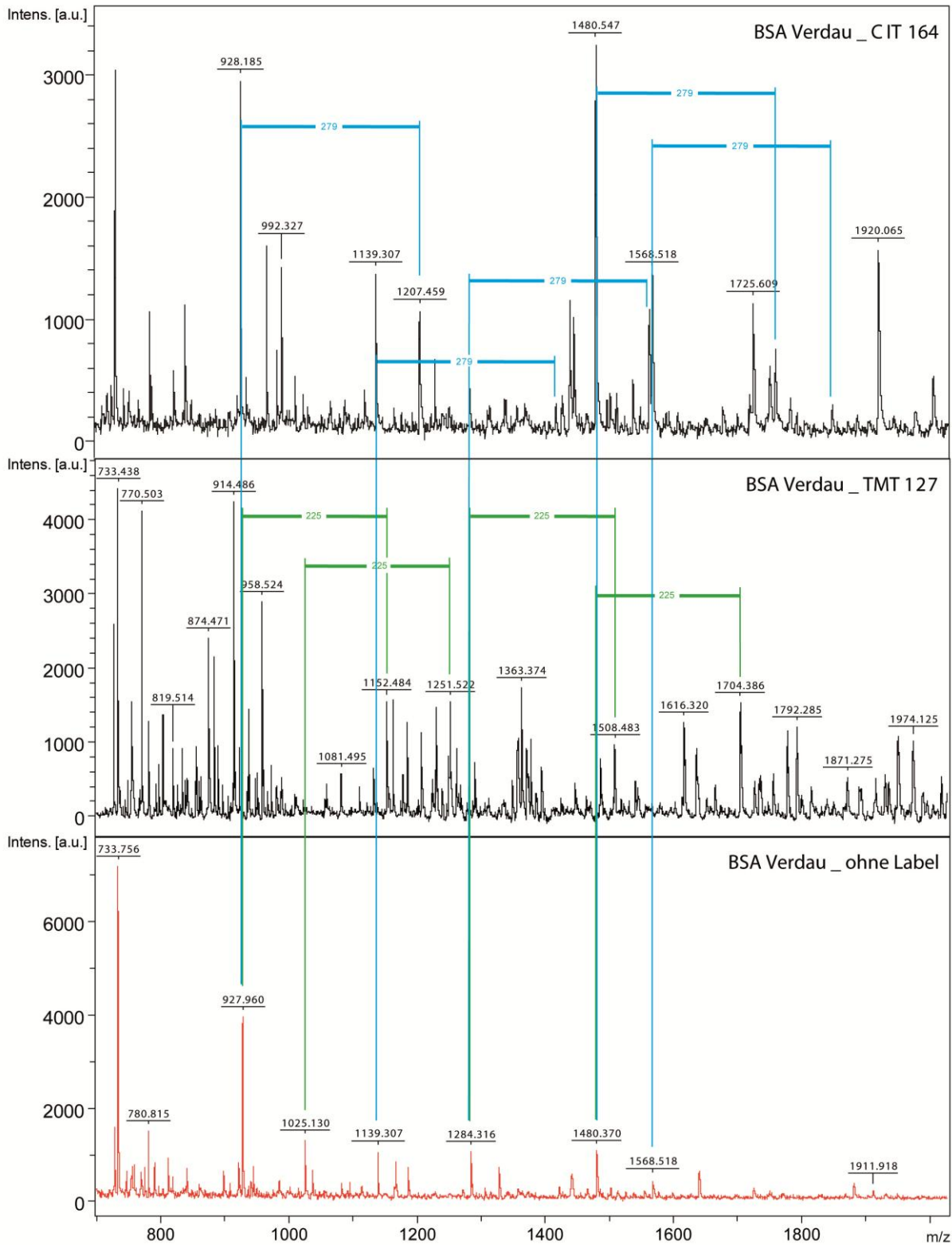


Abbildung 11: MALDI-TOF Spektren der mit CIT¹⁶⁴ (oben), TMT¹²⁷ (Mitte) markierten und nicht markierten (unten) BSA Proben nach dem proteolytischen Verdau mit Trypsin. Die jeweiligen Massenverschiebungen wurden in blau für CIT¹⁶⁴ und in grün für TMT¹²⁷ gekennzeichnet.

In der mit TMT-Reagenzien markierten Probe (siehe Abbildung 11, Mitte) dagegen konnten nahezu keine Peptide mit übereinstimmenden Massen in der unmarkierten BSA-Probe gefunden werden. Wohl aber Peptide mit den Massen um 225 Da verschoben. Hier wurden also fast alle Peptide vollständig markiert. Somit konnte gezeigt werden, dass das CIT-Reagenz nicht ohne erheblichen Aufwand zur Anpassung der optimalen Versuchsbedingungen für den Vergleich komplexer Protein Proben verwendet werden konnte. Es wurde entschieden das TMT-Reagenz für alle weiteren Versuche zu verwenden.

Ein positiver Effekt der isobaren Markierungen ist, dass gleiche Proteine, die mit unterschiedlichen Reagenzien markiert wurden, exakt dieselben physikochemischen Eigenschaften aufweisen und sich bei der Flüssigchromatographie und im MS-Modus von Massenspektrometern absolut identisch verhalten. Dadurch kann im Gegensatz zur SILAC-Methode eine Erhöhung der Spektrenkomplexität vermieden werden. Zusätzlich werden die Signale aus beiden Proben aufaddiert und führen zu einer besseren Signalintensität. Somit wird im Vergleich zu anderen Methoden, wie SILAC oder LFQ, die Detektierbarkeit verbessert und die Selektion für eine weitere Fragmentierung vereinfacht.^[183] Vor allem bei niedrig abundanten Proteinen stellt dies einen großen Vorteil dar. Den gleichen Effekt kann man auch im MS/MS-Modus beobachten, da sich auch hier die Signale gleicher Peptidfragmente aus den verschiedenen Proben aufsummieren, was zu einer verbesserten Sequenzabdeckung führt und damit auch die Identifikation des Peptids anhand eines Datenbankabgleichs der MS/MS-Signale vereinfacht. Dies wird durch die Abspaltung der Balancer Gruppe in einer „*Neutral Loss*“-Reaktion bei der Fragmentierung der Peptide ermöglicht. Parallel dazu werden auch die verschiedenen Reporterionen abgespalten, wodurch sich die Quantitäten der einzelnen Komponenten differenzieren lassen. Darüber hinaus kann die Flugeigenschaft von schlecht ionisierbaren Peptiden durch die chemische Markierung verbessert werden.^[184] Vor allem Peptide, die am C-Terminus ein Lysin aufweisen, lassen sich durch die Markierung besser detektieren.^[153]

Von den beiden getesteten isobaren Duplex-Systemen TMT^{126/127} und CIT^{164/169} konnte letzteres für die Anwendung in komplexen Peptidproben ausgeschlossen werden, da im Vergleich zum kommerziell erhältlichen TMT die Markierungseffizienz zu gering war. CIT-Reagenzien wären insofern recht interessant für bestimmte Anwendungen, da durch relativ einfache Syntheseschritte viele verschiedene Isotopenvarianten für multiplexe Anwendungen hergestellt werden könnten. Die Konstruktion isobarer Markierungsreagenzien ist nicht trivial, da neben der Austauschbarkeit der Isotopenatome in den Balance und Reporter Gruppen auch auf eine geeignete Gasphasen-Fragmentierung für die Bildung der Reporterionen im MS/MS-Modus geachtet werden muss.^[182]

Obwohl das Multiplexen von bis zu zehn unterschiedlichen Proben auch mit TMT-10plex-Reagenzien möglich gewesen wäre, wurden in dieser Arbeit aus Kostengründen die *pull down* Versuche nur mit

den TMT-Duplex-Reagenzien durchgeführt. So wurde in den Anwendungen der Methode auf die Untersuchung möglicher spezifischer Binder von DNA-Modifikationen jeweils nur eine der Modifikationen (mC, hmC, fC, caC und hmU) mit unmodifizierten Cytosin (C) bzw. Thymin (T) verglichen. Bereits fünf verschiedene isobare Reagenzien würden ausreichen, um die Cytosinderivate mC, hmC, fC und caC mit Cytosin in nur einem Ansatz zu vergleichen. Neben den bereits erwähnten technischen Vorteilen, würde enorm viel Messzeit am Massenspektrometer eingespart und zusätzlich die Vergleichbarkeit der Proben gesteigert werden, da in diesem Fall, statt der vier Proben, nur noch eine Probe zu messen wäre und systematische Fehler bei der LC/MS-Messung ausgeschlossen würden.

Ein Nachteil bei der Verwendung isobarer Markierungsreagenzien ist, dass diese, im Vergleich zur SILAC-Methode, erst relativ spät im Arbeitsablauf eingeführt werden. Während bei der SILAC-Methode die *pull down* Versuche mit bereits markierten Proteinen durchgeführt würden und die angereicherten Proteine direkt im Anschluss zusammengegeben würden, werden die isobaren Markierungen erst nach dem Verdau an den Peptiden eingeführt und diese dann im Anschluss zusammengegeben. Jeglicher systematischer Fehler in der Präparation vor dem Mischen der zu vergleichenden Proben könnte sich unterschiedlich auf die Proteinmenge der verschiedenen Proben und damit auch auf deren Verhältnis zueinander auswirken. Durch das Ausführen der Versuche mit Kreuz-Markierungsexperimenten bzw. in Replikaten wurde versucht diesen Effekt zu berücksichtigen.

3.1.3 Studien zu den Inkubationsbedingungen

Mit der neu eingeführten relativen Quantifizierungsmethode des isobaren Markierens wurde nun versucht, so viele Proteine wie möglich zu finden, die mit der modifizierten DNA spezifisch wechselwirken. Dies können Proteine sein, die aufgrund einer bestimmten Modifikation oder Sequenz entweder die DNA direkt erkennen und damit eine Art Enzym-Substrat-Komplex bilden oder als Signalmolekül eine Bindung mit der DNA eingehen und weitere Proteine rekrutieren. In beiden Fällen kann die Bindungskinetik nicht nur sehr unterschiedlich sein, sondern hängt auch von variablen Faktoren ab, vergleichbar mit der Enzymkinetik. Neben der Abhängigkeit von den Konzentrationen der Proteine und der DNA, spielen Reaktionsbedingungen, wie Temperatur, Salzkonzentration und pH-Wert eine Rolle. Die genaue Bestimmung der Zusammensetzung eines Zelllysats ist meistens aus Materialmangel nicht möglich und hängt stark davon ab, in welchem Zustand sich die Zellen zum Zeitpunkt der Lyse befinden. Jedoch ist von vielen Proteinen bzw. Proteingruppen bekannt oder es kann abgeschätzt werden, ob diese häufig oder selten auftreten. Dies kann zusätzlich durch Überexpression oder Knockdown/-out beeinflusst werden. Da sich Salzkonzentration, pH-Wert und Temperatur direkt auf die Konformation der Proteine und damit

ihrer Funktion auswirken können, bleibt man hier möglichst im physiologischen Bereich des verwendeten Zellmaterials. Allerdings kann durch Änderung der Temperatur die Reaktionsgeschwindigkeit bzw. Interaktionsdauer der Proteine im Experiment mehr oder weniger angepasst werden.^[185] Mit den folgenden Experimenten wurde versucht ein Optimum zwischen Inkubationslänge und –temperatur zu finden, um aus einer komplexen Proteinmischung möglichst viele Proteine mit einem Crosslinker an die DNA kovalent binden zu können, ohne dabei unspezifische Proteine so anzureichern, dass sie fälschlicherweise als signifikant spezifisch auftreten. Dazu wurden zunächst einige Testreihen mit Zellkernlysate von HEK-Zellen und caC-enthaltenden DNA-Sonden sowie unmodifizierten Kontroll-Sonden durchgeführt. Der Gehalt von caC in der genomischen DNA von HEK-Zellen liegt ohne Überexpression von Tet2 unterhalb der Nachweisgrenze.^[62, 63] Demnach wurde auch in dieser Testreihe zunächst keine spezifische Anreicherung für caC erwartet.

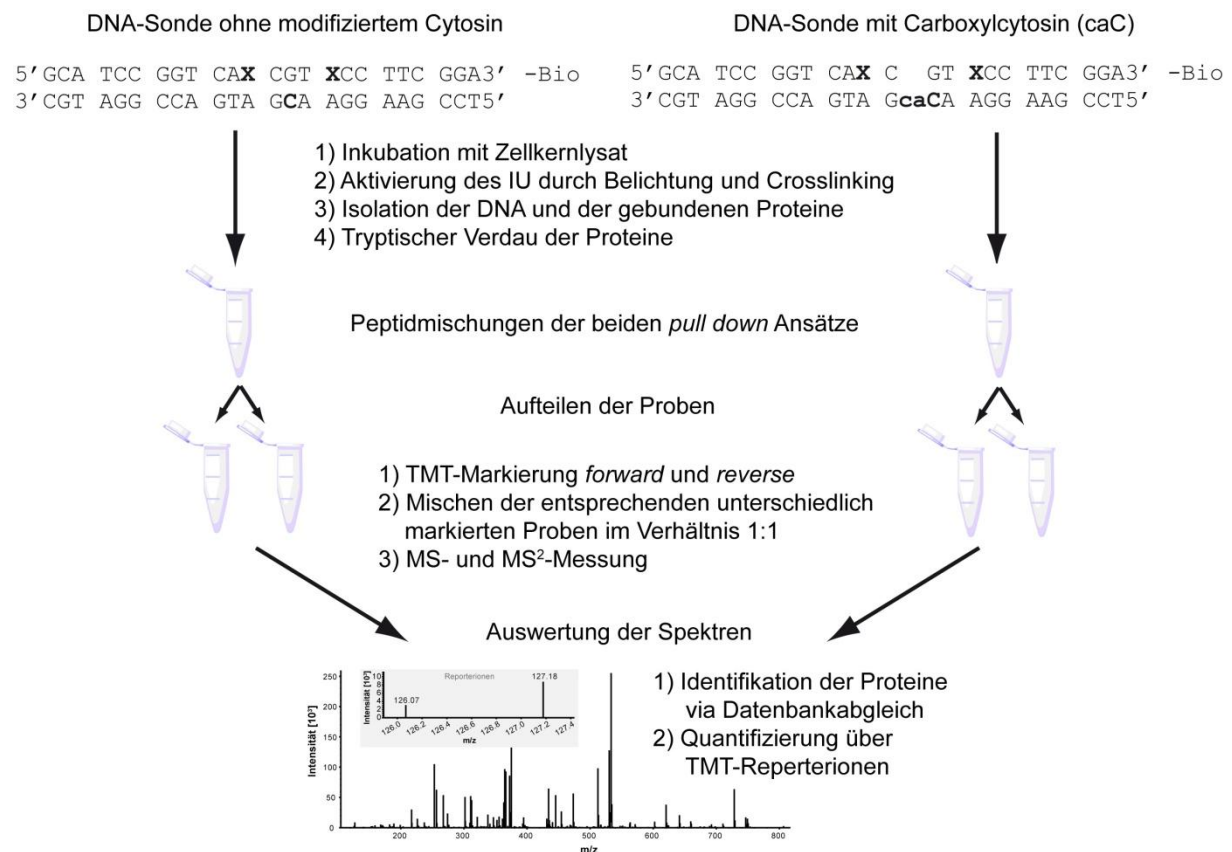


Abbildung 12: Schematische Darstellung der DNA-Sonden und des Workflows. X=5-Ioduracil (IU) und Bio=Biotin-Tag. Das genaue Schema der Probenmarkierung durch TMT-Reagenzien kann dem Methodenteil (Kapitel 5.7.9) entnommen werden.

Die DNA-Sonden (siehe Abbildung 12) wurden unterschiedlich lange (15 – 30 min) und bei unterschiedlichen Temperaturen (0 °C, RT und 37 °C) zunächst im Dunkeln mit dem Zellkernlysate inkubiert. Direkt im Anschluss daran wurden die Proben nochmals für 30 Minuten auf Eis mit Licht

der Wellenlänge 312 nm bestrahlt, um den Crosslinker 5-Ioduracil (IU) zu aktivieren und Proteine kovalent an die DNA zu binden. Nach der Isolation der DNA-Sonden mithilfe des Biotin-Tags und mit Streptavidin funktionalisierten magnetischen Partikel, der Reduktion, Alkylierung und dem tryptischem Verdau der Proteine, wurden die Proben in einem sogenannten *forward* und *reverse* Experiment mit TMT-Duplex-Reagenzien markiert. Dafür wurden die Proben aufgeteilt und je einmal mit dem TMT¹²⁶-Reagenz und dem TMT¹²⁷-Reagenz behandelt. Dann wurde die TMT¹²⁶ markierte Probe aus dem *pull down* Versuch mit caC-enthaltender DNA-Sonde mit der TMT¹²⁷ markierten Probe aus dem *pull down* Versuch mit C-enthaltender DNA-Sonde im Verhältnis 1:1 gemischt. Genauso wurde die TMT¹²⁷ markierte Probe aus dem *pull down* Versuch mit caC-enthaltender DNA-Sonde mit der TMT¹²⁶ markierten Probe aus dem *pull down* Versuch mit C-enthaltender DNA-Sonde gemischt. Durch dieses Verfahren wird einerseits eine Art technisches Replikat gemessen und zusätzlich dient es der Überprüfung, ob die zwei verschiedenen TMT-Reagenzien gleichmäßig mit den Peptiden reagiert haben. Damit sollen systematische Quantifizierungs-Fehler ausgeschlossen werden. Auf diese Weise ließen sich die Ergebnisse nach der LC-MS Messung und der weiteren Datenprozessierung auch sehr gut in einem Streudiagramm darstellen (siehe Abbildung 13).

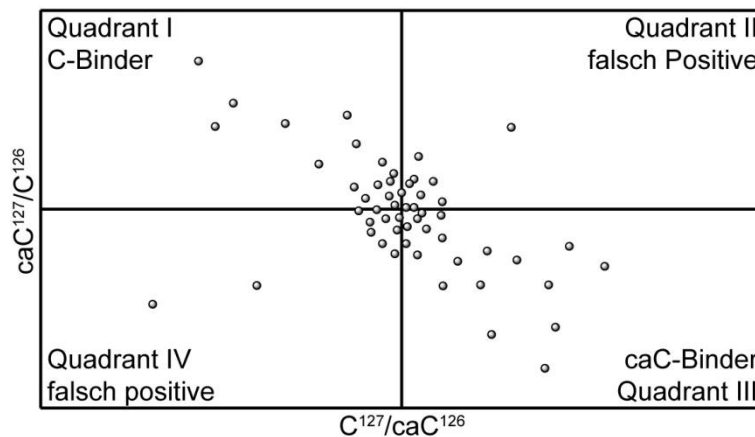


Abbildung 13: Darstellung quantitativer MS-Daten von Interaktionsstudien mit TMT-*forward/reverse* Markierung als Streudiagramm. Die meisten Proteine (graue Kreise) sollten als nicht-spezifische DNA-Binder in beiden Proben gleichermaßen angereichert sein und sich deshalb um den Ursprung des Koordinatensystems zentrieren. In Quadrant I und III befinden sich solche Proteine, die bevorzugt an den DNA-Sonden mit (QIII) oder ohne (QI) Cytosin-Modifikation binden. In den beiden Quadranten II und IV werden solche Proteine abgebildet, deren TMT-Verhältnisse sich in den *forward* und *reverse* Experiment nicht umkehren und somit falsch Positive darstellen.

Die von der Orbitrap XL erhaltenen Spektren wurden mit dem Programm Proteome Discoverer 1.4 prozessiert und die Ergebnisse in nachfolgenden Abbildungen dargestellt. Für weitere Datenanalysen wurden die Proteinlisten exportiert und mit Excel weiter bearbeitet. Bei der Analyse der Massenspektren mit dem Proteome Discoverer konnten für alle drei Versuchsbedingungen ausreichend viele Proteine für die Darstellung und Auswertung mit Hilfe der Streudiagramme

identifiziert und quantifiziert werden. Eine Übersicht dazu ist in Tabelle 5 und dem Venn Diagramm aus Abbildung 14 gegeben.

Tabelle 5: Numerische Zusammenfassung zu den Ergebnissen der *pull down* Versuche aus HEK-Zellkernlysate mit caC-haltigen DNA-Sonden unter drei verschiedenen Bedingungen.

Analyse	Bedingung 1				Bedingung 2				Bedingung 3			
	30 min auf Eis im Dunkeln				15 min bei RT im Dunkeln				15 min bei 37 °C im Dunkeln			
	Peptide	%	Proteine	%	Peptide	%	Proteine	%	Peptide	%	Proteine	%
Identifiziert gesamt	21226		190		23427		329		21732		397	
forward (caC ¹²⁶)	13409		181		15052		320		13172		371	
reverse (caC ¹²⁷)	13304		179		14820		314		13361		381	
fwd/rev Überlappung	5487		170	89	6445		305	93	4801		355	89
quantifiziert forward	8218	61	155	86	11965	80	309	97	10073	76	341	92
quantifiziert reverse	8581	64	153	85	11786	79	311	99	10564	79	358	94
quantifiziert fwd/rev	4041	73	135	71	5067	80	298	91	3628	76	316	80

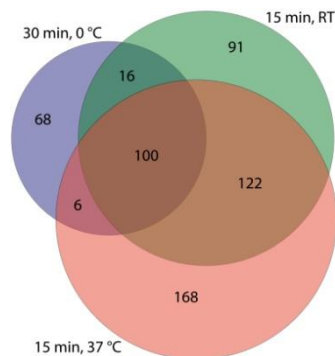


Abbildung 14: Venn Diagramm zur Veranschaulichung der unterschiedlich hohen Anzahl an identifizierten und quantifizierten Proteinen der *pull down* Versuche unter drei verschiedenen Bedingungen.

In allen Versuchsansätzen war die Markierungseffizienz ausreichend hoch, so dass 71% der unter Bedingung 1 identifizierten Proteine, 91% der unter Bedingung 2 und 80% der unter Bedingung 3 identifizierten Proteine für Streudiagramme verwendet werden konnten. Anhand dieser können deutliche Unterschiede in der Verteilung der Proteine festgestellt werden.

Ergebnisse und Diskussion

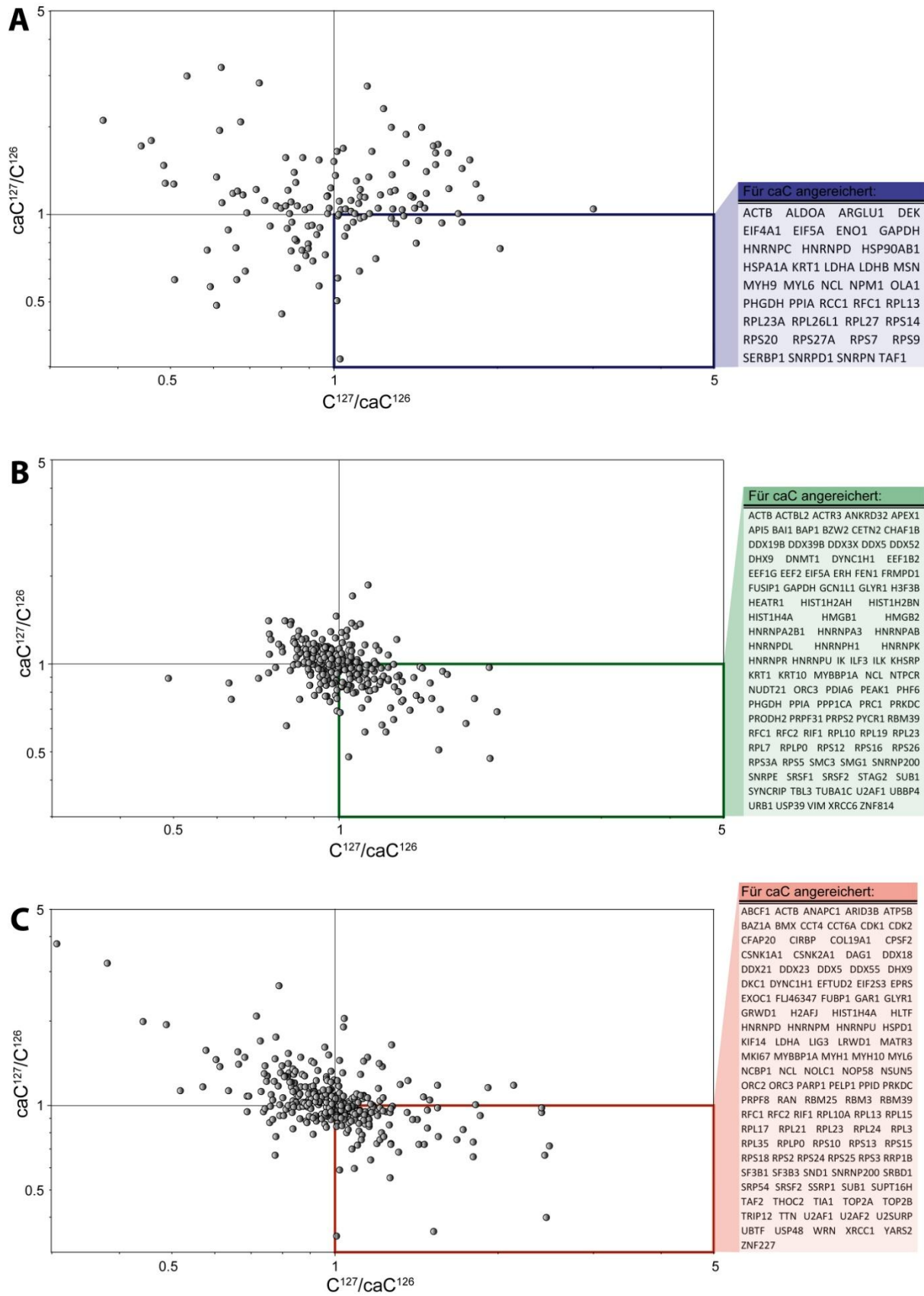


Abbildung 15: Streudiagramme zu den *pull down* Versuchen mit Inkubationszeiten im Dunkeln von (A) 30 min bei 0 °C, (B) 15 min bei RT und (C) 15 min bei 37 °C. Die jeweiligen Verhältnisse der TMT-Reporterionen Intensitäten der *forward* und *reverse* Experimente wurden auf einer logarithmischen Skala so aufgetragen, dass sich die für caC angereicherten Proteine jeweils im Quadranten unten rechts befinden.

Unter der Bedingung 1 (Abbildung 15A) wurden insgesamt nur sehr wenige Proteine an den DNA-Sonden angereichert. Dies könnte daran liegen, dass die Aktivität der Proteine bei der niedrigen Temperatur zu eingeschränkt ist, um in dem gegebenen Zeitrahmen mit der DNA in Interaktion zu treten. Eine Anhäufung von Proteinen um den Ursprung im Streudiagramm ist nicht zu erkennen, somit kann keine Aussage über signifikant angereicherte Proteine getroffen werden. Insgesamt scheint Bedingung 1 nicht geeignet für die *pull down* Studien.

Im Gegensatz dazu wurden unter den Bedingungen 2 (Abbildung 15B) und 3 (Abbildung 15C) sehr viele Proteine angereichert, die im jeweiligen Diagramm, wie erwartet, um den Ursprung liegen. Unter Bedingung 3 wurden zwar mehr Proteine identifiziert und quantifiziert als unter Bedingung 2, allerdings wurden diese mit weniger Peptiden als unter Bedingung 2 identifiziert. D.h. es wurde unter Bedingung 2 eine bessere Sequenzabdeckung und Datenqualität als unter Bedingung 3 erreicht. Dies könnte ein Hinweis darauf sein, dass bei 37 °C die Aktivität der Proteine so hoch ist, dass der Austausch von Proteinen an der DNA schneller von Statten geht als bei Raumtemperatur. Dadurch könnten Enzyme wie z.B. Dnmt1 (wurde unter Bedingung 2 für caC angereichert) sozusagen „verpasst“ werden und die Signifikanz der identifizierten Proteine würde herabgesetzt. Interessanterweise wurden unter Bedingung 2 auch mehr Histon-Proteine angereichert als unter Bedingung 3. Diese Ergebnisse führten zu der Entscheidung, dass alle weiteren Inkubationen bei *pull down* Versuchen unter Raumtemperatur durchgeführt wurden.

3.1.4 Einführung eines spaltbaren Crosslinkers

Der Einsatz von *pull down* Experimenten ist eine weit verbreitete Methode zur Untersuchung von DNA-Proteininteraktionen, da keine aufgereinigte Proteine benötigt werden, niedrig abundante Proteine selbst aus Zelllysaten und dadurch auch ganze Proteinkomplexe angereichert werden können und sie sowohl mit Immunoblotting als auch Massenspektrometrie kompatibel ist. Zu den Nachteilen zählen die Beschränkung auf *in vitro* Studien und das mögliche Auftreten signifikanter aber unspezifischer Bindungen, wodurch oft eine Validierung durch weitere Nachweismethoden, wie Chromatin Immunopräzipitations-Assays mit anschließendem Next Generation Sequencing (ChIP-seq.), Footprinting-Assays oder EMSAs, notwendig wird. Die meisten dieser Methoden sind jedoch auf hochspezifische Antikörper oder speziell aufgereinigte Proteine angewiesen und eignen sich daher nicht für Proteom-weite Untersuchungen. Eine möglichst konkrete Auswahl an in Frage kommender Validierungskandidaten wäre deshalb höchst interessant. In Affinitäts-basierten *pull down* Studien können allerdings keine direkten von indirekten spezifischen Bindern unterschieden werden, womit die Chancen auf ein erfolgreiches Validieren nur bedingt verbessert werden. Durch den Einsatz von bewährten Crosslinkern wie Ioduracil (IU) kann zwar die Auswahl für Validierungskandidaten

erheblich eingegrenzt werden, man verliert aber gleichzeitig die Information über gegebenenfalls indirekt beteiligte Proteine und Proteinkomplexe.^[186] Die photoaktivierte Reaktion von Ioduracil (IU) führt zu einer kovalenten Bindung zwischen DNA-Sonde und Proteinen.^[186] Nach dem Verdau der gebundenen Proteine liegen diese als Peptide vor, welche dann massenspektrometrisch analysiert und via Datenbankabgleich dem ursprünglichen Protein zugeordnet werden können. Dies gilt für alle Peptide bis auf jenes, welches kovalent an die DNA gebunden wurde. Da IU sehr unspezifisch mit den Peptiden reagiert und nicht klar ist, ob und welche DNA-Reste nach der Ionisierung und Fragmentierung im Massenspektrometer noch vorhanden sind, ist eine Identifikation solcher Peptide mit derzeit gegebener Soft- und Hardware nicht möglich. Eine Folge davon ist, dass die Sequenzabdeckung eines identifizierten Proteins bei der Auswertung herabgesetzt und die Datenqualität vermindert wird.

Aus diesem Grund wurde entschieden, DNA-Sonden mit einem anderen Crosslinker zu verwenden. Dieser sollte spezifisch mit einer bestimmten Aminosäure reagieren und von der DNA abspaltbar sein, damit ein Rest mit definierter Masse am Peptid zurück bleibt. Dadurch kann nicht nur die Datenqualität aufrechterhalten werden, sondern auch eine weitere aufschlussreiche Information gewonnen werden. Durch die Markierung können nämlich Proteine mit direktem Kontakt zur DNA von solchen Proteinen unterschieden werden, die indirekt, beispielsweise über Proteinkomplexe, an der DNA-Sonde angereichert wurden. Die Wahl fiel auf den Azid-PEG3-S-S-NHS Ester ($C_{18}H_{29}N_5O_8S_2$). Der NHS Ester reagiert vorzugsweise mit den Lysinen der Peptide und die Disulfidbindung ist leicht durch Reduktion zu spalten (siehe Abbildung 16).

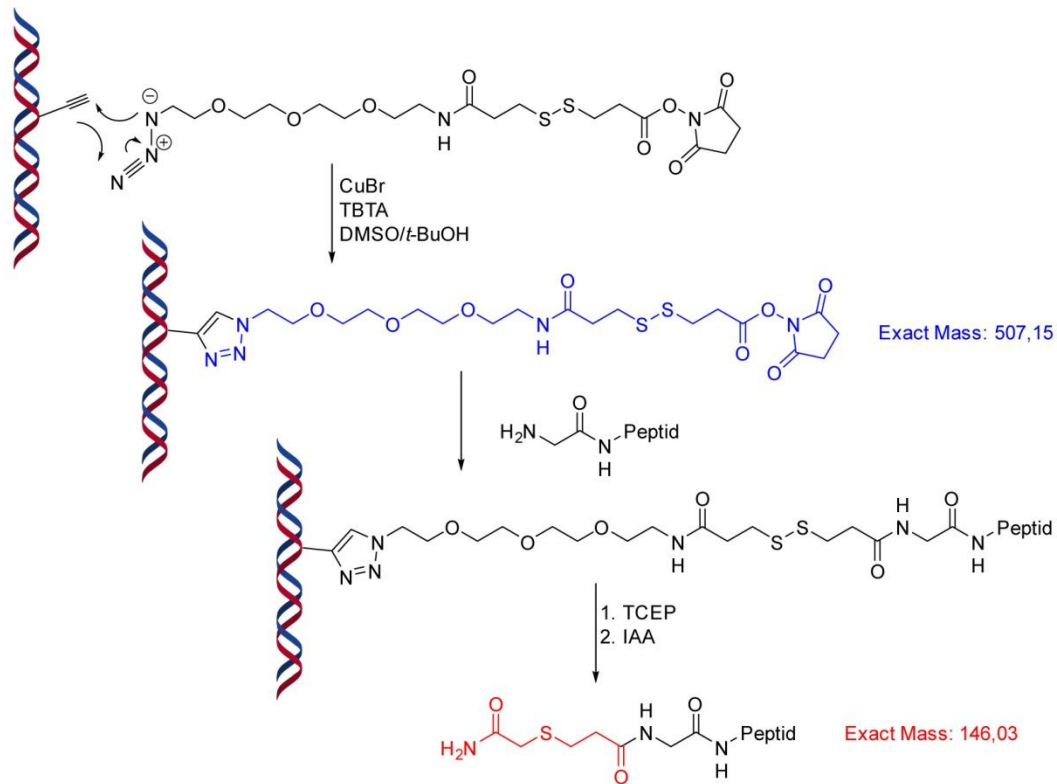


Abbildung 16: Schematische Darstellung der Click-Reaktion, des Crosslinkers und der Crosslinker-Spaltung.

Die Einführung eines solchen DNA-Protein Crosslinkers an der DNA-Sonde gelang Barbara Steigenberger mithilfe der Click-Reaktion. Hierfür wurden entsprechende DNA-Einzelstränge synthetisiert, hybridisiert und die Click-Reaktion am Doppelstrang durchgeführt. Die Einführung des Crosslinkers musste post-synthetisch am Doppelstrang durchgeführt werden, da eine HPLC-Aufreinigung wegen der Hydrolyseempfindlichkeit des NHS-Esters danach nicht mehr möglich ist. Anschließend wurde die DNA-Sonde gefällt und mit EtOH gewaschen, um unreaktierten Linker abzutrennen. Abbildung 17 zeigt die übereinander gelegten MALDI-Spektren der DNA-Sonde vor und nach der Click-Reaktion. Man kann eine deutliche Massenveränderung um das Molekulargewicht des Linkers erkennen. Die starke Verbreiterung des Signals kommt dadurch zustande, dass auf Grund der Hydrolyseempfindlichkeit des Esters auf eine Entsalzung nach der Click-Reaktion verzichtet wurde.

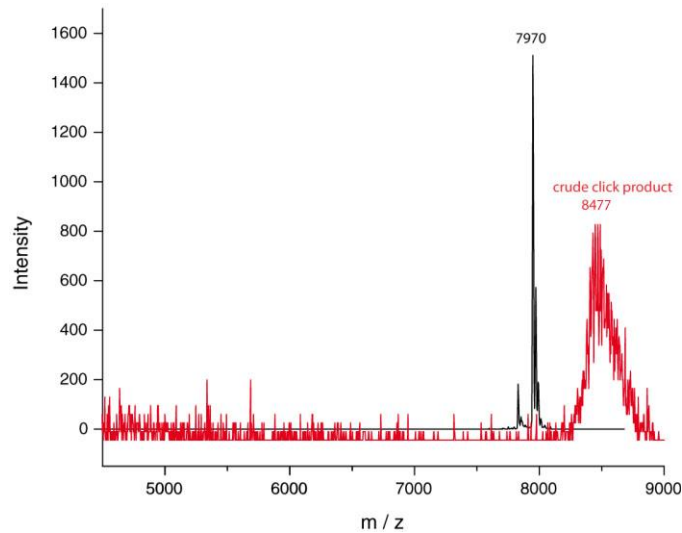


Abbildung 17: MALDI-Spektrum eines DNA-Oligomers vor (schwarz) und nach (rot) der Click-Reaktion mit dem Crosslinker. Die starke Verbreiterung des Signals nach der Click-Reaktion kommt dadurch zustande, dass auf Grund der Hydrolyseempfindlichkeit des Esters auf eine Entsalzung verzichtet wurde.

Dieser Crosslinker wurde in mehreren Experimenten zur Untersuchung verschiedener biologischer Fragestellungen eingesetzt (siehe folgende Kapitel). Dabei konnten viele Peptide identifiziert werden, die mit dem Crosslinker markiert wurden. Abbildung 18 zeigt ein exemplarisches Fragmentspektrum eines Histon H1-Peptids aus einem dieser Versuche. Darin sind die b- und y-Ionenserien des identifizierten Peptids und deren Massendifferenzen angegeben, anhand derer auf die Peptidsequenz rückgeschlossen werden kann. Das Lysin an Position 6 des Peptids SVAFKKTK weist demnach eine um die Masse des Crosslinkers erhöhte Massendifferenz auf.

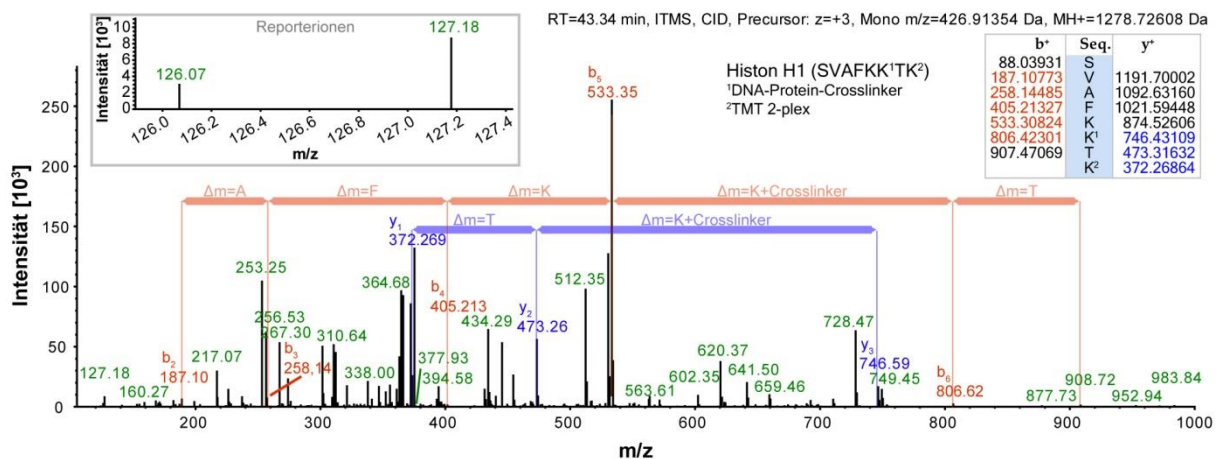


Abbildung 18: Repräsentatives Fragmentspektrum eines Peptids mit Crosslinker-Rest. Das Spektrum stammt aus der Analyse eines pull down mit hmC- und Crosslinker-haltiger DNA-Sonden aus Retinagewebe 21 Tage alter Mäuse (vgl. Kapitel 3.2.2). Angegeben sind die b- (rot) und y- (blau) Ionenserien des identifizierten Peptids und deren Massendifferenzen. Die Tabelle oben rechts gibt die berechneten (schwarz) und gefundenen (rot bzw. blau) Massendifferenzen des Histon H1-Peptids mit der Aminosäuresequenz SVAFKKTK wieder. Angereicherte Proteine wurden mit TMT quantifiziert. Der Kasten links oben zeigt das Spektrum der TMT-Reporterionen.

Ein Punkt der in dieser Arbeit noch nicht aufgeklärt werden konnte, ist inwiefern der NHS-Ester des verwendeten Crosslinkers dafür geeignet ist, spezifische Binder mit der DNA kovalent zu verlinken. Denn im Gegensatz zum photoaktivierbaren 5-Ioduracils (IU) reagiert der NHS-Ester sofort mit primären Aminen und ist zudem bei höheren pH-Werten als 8,3 hydrolyseempfindlich.^[187] Dies könnte dazu führen, dass der Ester in wässriger Lösung zu schnell abreagiert oder vermehrt auch mit nicht spezifisch wechselwirkenden Proteinen reagiert. Um dieses Problem zu umgehen, wurde von Barbara Steigenberger ein Crosslinker synthetisiert, der statt dem NHS-Ester eine photoaktivierbare Diaziridin-Gruppe besitzt. Erste Versuche mit diesem photoaktivierbaren Crosslinker (Daten nicht gezeigt), zeigten bereits, dass auch damit Proteine an die DNA gebunden werden können und diese nach Spaltung des Linkers massenspektrometrisch identifizierbar sind. Allerdings verläuft die Reaktion der Diaziridin-Gruppe über ein hochreaktives Carbeniumion, das über Additionsreaktionen mit jeder Aminosäureseitenkette oder dem Rückgrat eines Peptids innerhalb der Crosslinkerlänge reagieren kann.^[188] Ähnlich wie bei IU, erhöht sich damit erheblich die Anzahl an möglichen Addukten, die bei einem Datenbankabgleich berücksichtigt werden müssen, um solche Peptide zu identifizieren, die mit dem Crosslinker markiert wurden. Jedoch erhöht jede eingegebene Modifikation die Anzahl an möglichen theoretischen Spektren für ein Peptid. Damit steigt die Wahrscheinlichkeit, dass ein aufgenommenes Spektrum mit einem falschen theoretischen Spektrum übereinstimmt und als falsch positives Peptid identifiziert wird. Dies zeigte sich unter anderem darin, dass trotz einer anfänglich falsch berechneten Crosslinker-Masse (vgl. Abschnitt 3.2.1) relativ viele Peptide mit der Modifikation gefunden werden konnten. Ein direkter Vergleich der unterschiedlichen Crosslinker (NHS-Ester, IU und Diaziridin) in der Anwendung mit einem Gemisch aus nur wenigen Proteinen, von denen bekannt ist, inwiefern sie mit der DNA wechselwirken, könnte hier Aufschluss über die tatsächliche Auswirkung der unterschiedlichen Reaktionsmechanismen und deren jeweilige Vor- und Nachteile geben.

3.2 Anwendung der Methode

3.2.1 Crosslinking-Interaktionsstudie mit 5-Hydroxymethyluracil in murinen Stammzellen

Nachdem die Rahmenbedingungen getestet und evaluiert wurden, sollte im nächsten Schritt die Methode auf eine konkrete wissenschaftliche Frage hin angewandt werden. Eine Möglichkeit dazu bot sich, nachdem vor kurzem gezeigt werden konnte, dass Tet Enzyme nicht nur die epigenetisch wichtige DNA Base mC schrittweise zu hmC, fC und caC oxidieren können, sondern auch Thymin zu hmU. In murinen Stammzellen resultieren dadurch vergleichbar hohe Level an hmU wie an caC in der DNA. Da bereits für caC über eine mögliche Rolle bei der aktiven Demethylierung diskutiert wird, könnte auch hmU eine derartige epigenetische Funktion besitzen. Um hier einen ersten Einblick in potentielle biologische Funktionen des hmU erhalten zu können, wurden *pull down* Experimente aus Zellkernextrakten muriner Stammzellen (mESCs) durchgeführt. In vorherigen SILAC- bzw. LFQ-basierten Studien mit mC, hmC, fC und caC enthaltenden Oligonukleotiden wurde eine hohe Anzahl an spezifisch bindenden Proteine beobachtet, was darauf schließen lässt, dass die neuen Basen eine Vielzahl an unterschiedlichen Prozessen beeinflussen.^[66, 67] Mit dem in dieser Arbeit entwickeltem TMT- und Crosslinker-basierten Versuchsansatz sollten nun auch bindende Proteine identifiziert werden und direkte von indirekten bzw. sekundären Bindern unterscheidbar gemacht werden (siehe Abbildung 19).

Der Versuch wurde mit den unterschiedlichen Basenpaaren hmU:A, hmU:G und hmC:G durchgeführt. Als Referenzproben dienten DNA-Sonden mit den Basenpaaren T:A und C:G. Zusätzlich wurden die Experimente mit den hmU:A und T:A enthaltenden DNA-Sonden wiederholt, aber mit einer alternativen Basensequenz der Oligomere, um die Abhängigkeit von der Sequenz zu überprüfen. Für die relative Quantifizierung wurde TMT verwendet, wobei die Proben und Kontrollen jeweils einmal mit dem leichten und einmal mit dem schweren Tag markiert wurden, so dass jedes Experiment als *forward* und *reverse* gemessen und in einem Streudiagramm dargestellt werden konnte. Nur entsprechende Proteine, die sowohl im *forward* als auch im *reverse* Experiment relativ zum jeweiligen Kontrollstrang als angereichert quantifiziert wurden, wurden als spezifische hmU bzw. hmC Binder betrachtet. Zudem wurden alle Proteine, die mit wenigstens einem Crosslinker-Rest enthaltendem Peptid identifiziert werden konnten, markiert. Diese gelten als direkte Binder.

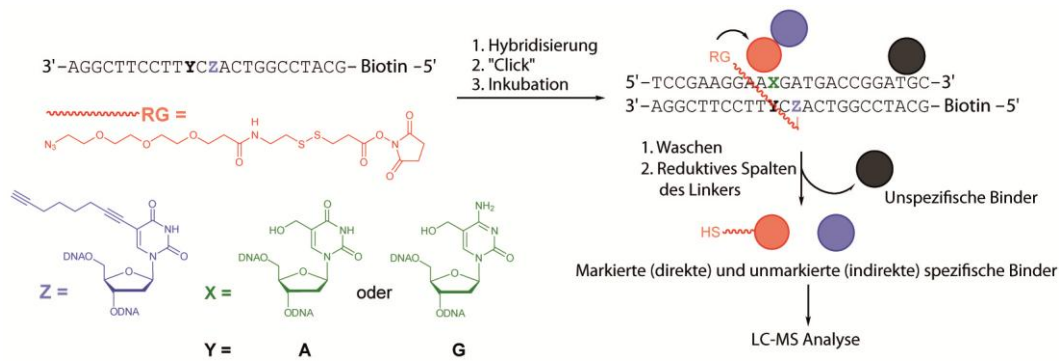


Abbildung 19: Zusammenfassung zum Versuchsansatz der *pull down* Experimente mit Crosslinker. Nach der Hybridisierung entsprechender DNA-Oligomere wurde der Crosslinker via Click-Reaktion eingeführt. Anschließend wurde die fertige DNA-Sonde mit dem Zellkernlysate aus mESCs inkubiert, wobei der NHS-Ester des Crosslinkers mit räumlich nahen Proteinen reagiert. Durch mehrmaliges Waschen wurden unspezifische Binder (schwarz) abgetrennt. Im Zuge der weiteren Probenpräparation wurde der Crosslinker reduktiv gespalten. Sowohl direkt (rot) als auch indirekt (blau) gebundene Proteine wurden isoliert und mittels LC-MS analysiert.

Die erste Auswertung der Daten wurde mit dem Programm Proteome Discoverer 1.1 durchgeführt und 2014 publiziert.^[1] Da zwischenzeitlich eine neuere Version des Programmes verfügbar war (Proteome Discoverer 1.4), die einen neuen Suchalgorithmus verwendet (SequestHT®), wurde die Auswertung der Rohdaten für diese Arbeit komplett wiederholt. Die aktuellen Ergebnisse sind nun im Folgenden dargestellt.

Die Streudiagramme (Abbildungen 20, 22, 24 und 27) wurden so erstellt, dass sich die, für das jeweilige Cytosin-Derivat (hmU oder hmC), angereicherten Proteine im Quadranten unten rechts (Q III) befinden. Der Quadrant oben links (Q I) zeigt dementsprechend Proteine, die an der jeweiligen Kontroll-DNA-Sonde (T oder C) angereichert wurden. In den Quadranten oben rechts (Q II) und unten links (Q IV) befinden sich Proteine, die nicht eindeutig quantifiziert wurden, d.h. sowohl im *forward* als auch im *reverse* Experiment ein TMT-Verhältnis größer bzw. kleiner als 1 aufweisen.

In den zugehörigen Tabellen sind alle Proteine aus den Quadranten III inklusive der jeweiligen TMT-Verhältnisse aufgelistet, getrennt danach, ob sie mit (rot) oder ohne (grau) Crosslinker identifiziert wurden. Für die GO-(*Gene Ontology*) Analysen wurden jeweils alle Proteine der Quadranten III verwendet.

Ergebnisse und Diskussion

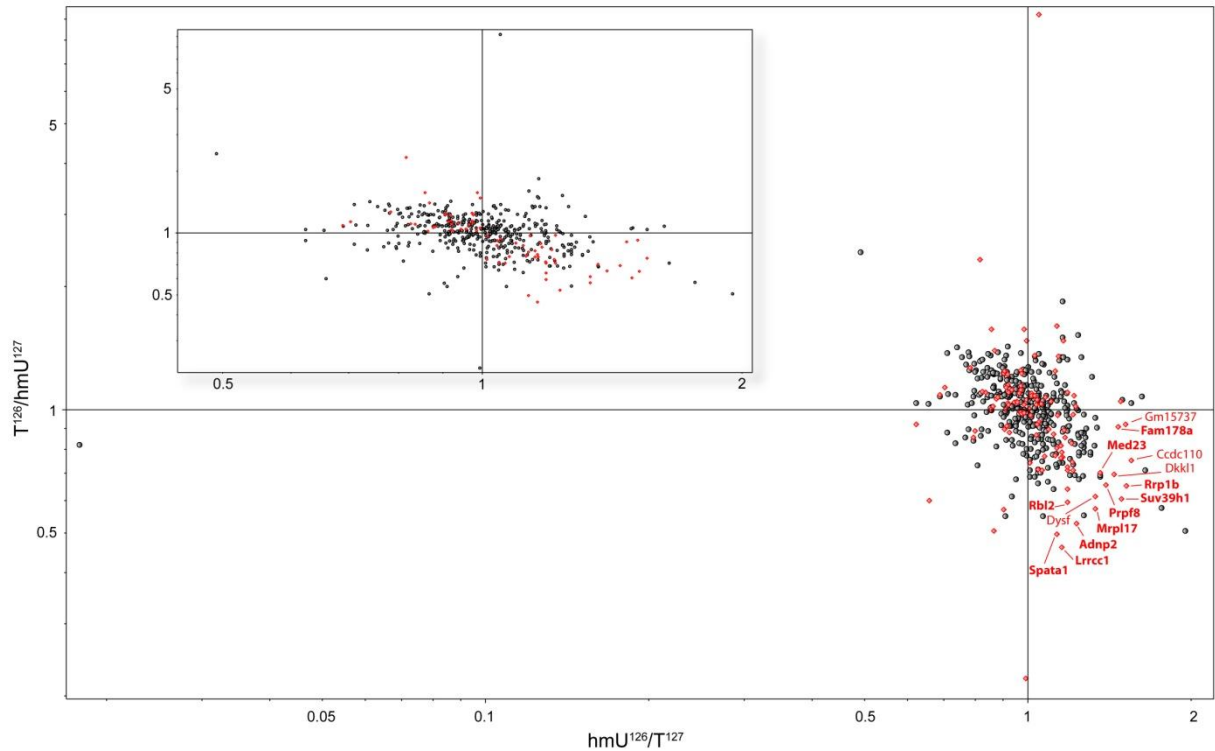


Abbildung 20: Streudiagramm zu den Ergebnissen des *pull down* Versuch mit den Basenpaaren hmU:A im Vergleich zu T:A. Alle identifizierten Proteine, die mit dem Crosslinker markiert wurden sind als rote Rauten dargestellt. Der Kasten oben links zeigt einen Ausschnitt des Streudiagramms ohne den einzelnen unspezifischen Binder ganz rechts im Diagramm. Hierin wurde auf die rote Markierung der Crosslinker-markierten unspezifischen Binder verzichtet.

Tabelle 6: Für hmU angereicherte Proteine aus dem *pull down* Versuch mit den Basenpaaren hmU:A im Vergleich zu T:A mit ihren jeweiligen TMT-Verhältnissen (Streudiagramm Abbildung 20). Proteine, die den GO-Termen aus Abbildung 21 zugeordnet wurden, sind fett hervorgehoben.

Mit Crosslinker			Ohne Crosslinker								
Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷
Adnp2	1,231	0,526	2210010C04Rik	1,002	0,906	Gltscr2	1,059	0,846	Phtf2	1,063	0,945
Ankrd17	1,170	0,889	Actrt3	1,012	0,755	Gm2663	1,083	0,916	Pik3c2g	1,145	0,742
C2cd2	1,015	0,967	Adam12	1,307	0,776	Gm3371	1,116	0,776	Pls1	1,239	0,873
Camk4	1,094	0,893	Adcy9	1,214	0,733	Gm4781	1,032	0,831	Prkg1	1,008	0,997
Ccdc110	1,553	0,752	Adh6a	1,952	0,505	Gm7964	1,021	0,978	Ptprd	1,264	0,687
Cep104	1,132	0,804	Adrbk2	1,058	0,702	Gm8369	1,345	0,904	Rad51ap1	1,069	0,798
Chd4	1,012	0,741	Aff1	1,307	0,776	Got2	1,765	0,574	Rpl19	1,364	0,684
Dkk1	1,446	0,693	Akap13	1,165	0,896	Gpr98	1,269	0,785	Scaf4	1,064	0,865
Dnah7b	1,186	0,711	Ankrd26	1,024	0,909	Grm1	1,092	0,848	Scg5	1,084	0,982
Dnaja4	1,216	0,711	Arpp19	1,017	0,858	H2afz	1,161	0,740	Scml2	1,128	0,732
Dysf	1,335	0,613	Asic1	1,135	0,722	H3f3a	1,065	0,748	Sct	1,307	0,776
Fam178a	1,471	0,905	Atp5a1	1,067	0,964	Heatr5b	1,010	0,711	Sdr16c6	1,080	0,921
Fgfbp1	1,051	0,931	Atp9b	1,258	0,816	Helz2	1,074	0,975	Sec31a	1,067	0,549
Gm15737	1,515	0,920	Atpif1	1,128	0,780	Hes2	1,096	0,988	Sept9	1,363	0,689
Hspd1	1,217	0,970	Atrx	1,181	0,985	Hhpl2	1,030	0,765	Setbp1	1,286	0,898
Lrrc8c	1,124	0,770	Bace1	1,269	0,785	Hist1h2bb	1,035	0,917	Sh3d21	1,307	0,776
Lrrcc1	1,159	0,460	Bag5	1,328	0,814	Hist1h2bm	1,028	0,917	Slc35a1	1,083	0,804
Malt1	1,076	0,767	Bmi1	1,083	0,886	Hist1h4a	1,052	0,865	Slx4ip	1,033	0,684
Med23	1,363	0,702	Brinp2	1,363	0,689	Hist2h2ac	1,269	0,827	Smpd3	1,141	0,830
Mrpl17	1,334	0,573	Cachd1	1,260	0,879	Hivep3	1,047	0,909	Smpx	1,108	0,970
Mrpl3	1,040	0,867	Cdk1	1,244	0,674	Hydin	1,010	0,685	Snrpd2	1,199	0,689
Nlrp10	1,173	0,853	Cdk3	1,017	0,956	Iqca1	1,010	0,968	Spire1	1,090	0,666
Nlrp4b	1,161	0,763	Cds2	1,014	0,845	Kif26b	1,269	0,875	Sqrdl	1,223	0,689
Nostrin	1,046	0,715	Cfap58	1,108	0,976	Krt1	1,241	0,993	Srgap2	1,089	0,718
Nr4a3	1,184	0,722	Clk1	1,078	0,959	Clk1	1,241	0,865	Srsf4	1,281	0,853
Ppl	1,137	0,968	Col19a1	1,364	0,684	Larp1	1,161	0,881	Ssh1	1,046	0,845
Proc	1,186	0,639	Col7a1	1,333	0,955	Lrrc7	1,363	0,689	Stk32a	1,010	0,930
Prpf8	1,396	0,654	Dbr1	1,244	0,787	Lrrc74b	1,194	0,912	Tenm3	1,181	0,968
Rbl2	1,186	0,593	Dgkk	1,186	0,886	Maged2	1,269	0,877	Tmpo	1,186	0,906
Rrp1b	1,522	0,650	Dmrtc2	1,202	0,835	Map3k9	1,140	0,974	Top2b	1,186	0,997
Spata1	1,132	0,496	Elmo1	1,022	0,988	Mob2	1,103	0,844	Tpm1	1,222	0,822
Suv39h1	1,492	0,603	Eme2	1,107	0,927	Mov101	1,046	0,701	Trio	1,307	0,776
Tbc1d30	1,212	0,743	Ephb3	1,269	0,807	Mylk	1,069	0,822	Ube3a	1,141	0,830
Tchp	1,152	0,761	Ern2	1,071	0,999	Myo9a	1,046	0,785	Ubn1	1,050	0,961
Ttc37	1,152	0,817	Evpl	1,193	0,951	Naca	1,052	0,901	C5orf42	1,013	0,802

Fortsetzung Tabelle 6: Für hmU angereicherte Proteine aus dem *pull down* Versuch mit den Basenpaaren hmU:A im Vergleich zu T:A mit ihren jeweiligen TMT-Verhältnissen (Streudiagramm Abbildung 20). Proteine, die den GO-Termen aus Abbildung 21 zugeordnet wurden, sind fett hervorgehoben.

Mit Crosslinker			Ohne Crosslinker								
Gen Name	hmU ¹²⁶ / _{T¹²⁷}	T ¹²⁶ / _{hmU¹²⁷}	Gen Name	hmU ¹²⁶ / _{T¹²⁷}	T ¹²⁶ / _{hmU¹²⁷}	Gen Name	hmU ¹²⁶ / _{T¹²⁷}	T ¹²⁶ / _{hmU¹²⁷}	Gen Name	hmU ¹²⁶ / _{T¹²⁷}	T ¹²⁶ / _{hmU¹²⁷}
Ttn	1,041	0,918	Fam169a	1,058	0,922	Nek4	1,118	0,916	Uso1	1,114	0,841
Ush2A	1,116	0,868	Fam227b	1,307	0,785	Nfrkb	1,244	0,771	Uvssa	1,307	0,776
Utp14b	1,205	0,827	Fam83b	1,041	0,767	Nol9	1,045	0,659	Vwa9	1,118	0,821
Zcchc6	1,159	0,786	Fgfr4	1,197	0,689	Nxpe5	1,197	0,758	Wac	1,074	0,889
Znf770	1,063	0,711	Gap43	1,245	0,853	Olfr1066	1,096	0,942	Zcchc9	1,269	0,971
			Gars	1,046	0,715	Olfr1349	1,020	0,911	Zfp689	1,307	0,776
			Gas2l2	1,269	0,550	Pfkip	1,647	0,711	Znf653	1,116	0,639
			Gdap111	1,034	0,919	Phf2011	1,038	0,992			

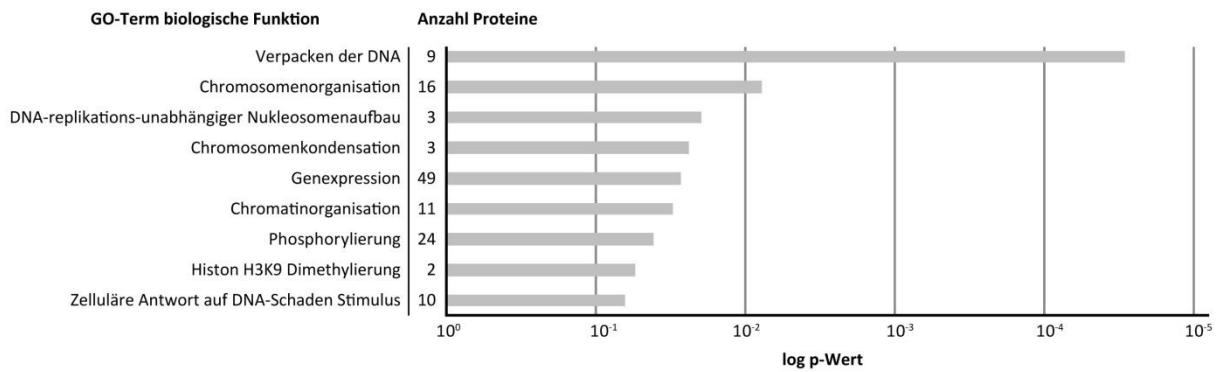


Abbildung 21: Ausschnitt einer DAVID-GO-Term Analyse der für hmU:A angereicherten Proteine. Neben den p-Werten ist auch die Anzahl der Proteine angegeben, die unter dem jeweiligen GO-Term stehen.

Insgesamt wurden 340 Proteine in den Quadranten Q-I (für T) und Q-II (für hmU) angereichert. Davon waren 72 Proteine mit dem Crosslinker markiert. Allein für hmU wurden 168 Proteine angereichert, wovon 32 Proteine mit dem Crosslinker markiert waren (vgl. Tabelle 6). Die GO-Analyse (Abbildung 21) zeigte, dass 49 der 168 spezifisch hmU-bindenden Proteine an der Genexpression und 16 Proteine an der Chromosomenorganisation beteiligt sind. Darunter befinden sich Transkriptionsfaktoren wie Dmrtc2 (*doublesex and mab-3 related transcription factor like family C2*), Phtf2 (*putative homeodomain transcription factor 2*) und Helz2 (*helicase with zinc finger 2*, transkriptioneller Coaktivator) als auch Zinkfingerproteine wie Zfp689, Zcchc6 und Zcchc9. Dmrtc2 sowie Suv39h1 (*suppressor of variegation 3-9 homolog 1*) sind Proteine, die an der Methylierung von Histon H3K9 beteiligt sind. Es wurden auch einige H2A (H2afz, Hist2h2ac), H2B (Hist1h2bb, Hist1h2bm), H3 (H3f3a) und H4 (Hist1h4a) Histonproteine unter den spezifischen Bindern entdeckt. Darüber hinaus wurde das Chromodomain Helikase DNA bindende Proteine Chd4 unter den direkten hmU Bindern gefunden, welches ein Chromatin Remodeling Faktor ist und die Polymerase I und II Transkription reguliert.

Ergebnisse und Diskussion

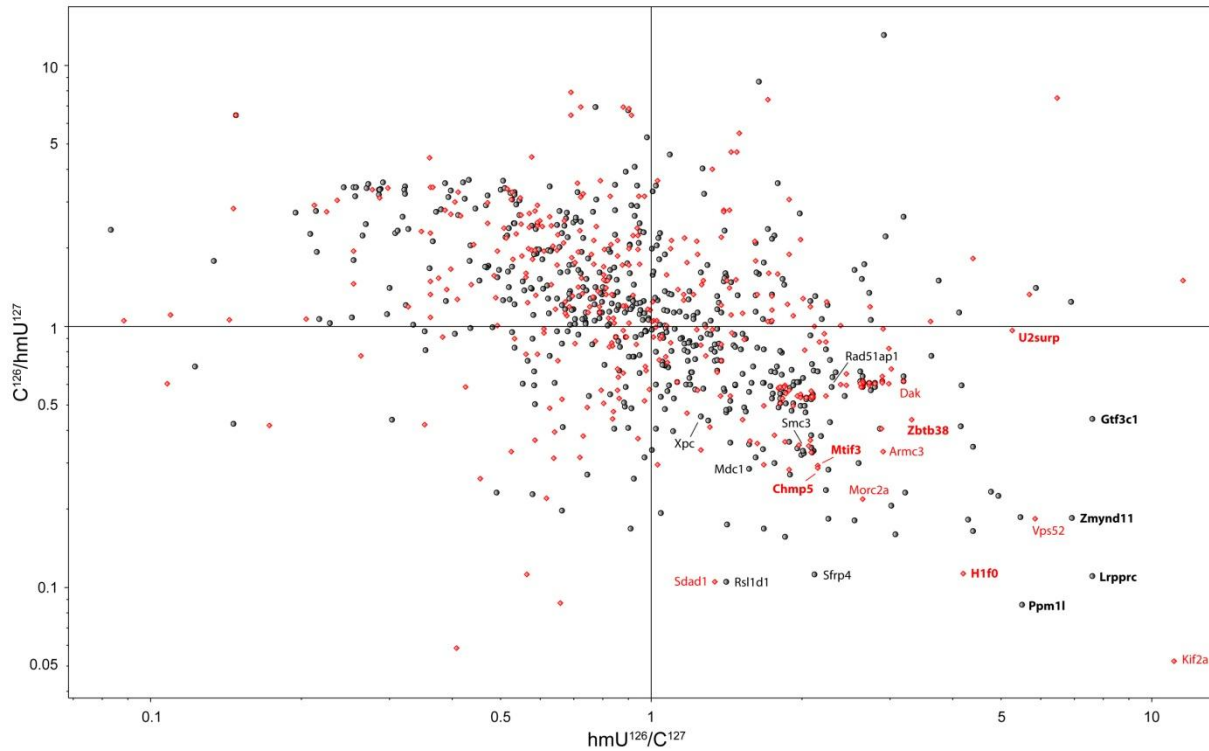


Abbildung 22: Streudiagramm zu den Ergebnissen des *pull down* Versuchs mit den Basenpaaren hmU:G im Vergleich zu C:G. Alle identifizierten Proteine, die mit dem Crosslinker markiert wurden sind als rote Rauten dargestellt.

Tabelle 7: Für hmU angereicherte Proteine aus dem *pull down* Versuchs mit den Basenpaaren hmU:G im Vergleich zu C:G mit ihren jeweiligen TMT-Verhältnissen (Streudiagramm Abbildung 22). Proteine, die den GO-Termen aus Abbildung 23 zugeordnet wurden, sind fett hervorgehoben.

Mit Crosslinker			Ohne Crosslinker								
Gen Name	hmU ¹²⁶ / C ¹²⁷	C ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / C ¹²⁷	C ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / C ¹²⁷	C ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / C ¹²⁷	C ¹²⁶ / hmU ¹²⁷
Abl2	0,541	1,962	4933411G06Rik	0,599	1,200	Kif4	0,608	2,660	Vps4b	0,769	3,630
Aff4	0,727	1,078	A430033K04Rik	0,945	1,073	Klhl38	0,533	2,094	Vps51	0,541	2,437
Ahctf1	0,533	1,991	Abcb4	0,581	1,830	Kmt2b	0,616	2,896	Wasf3	0,527	1,111
Aox1	0,812	2,033	Acad9	0,584	2,641	Lima1	0,564	1,406	Xpc	0,451	1,258
Ap1g1	0,608	2,651	Acin1	0,838	1,218	Lrpprc	0,110	7,628	Xpncp1	0,897	1,148
Apex1	0,365	2,081	Acot11	0,382	2,093	Lrriq3	0,595	1,086	Zbtb17	0,156	1,853
Apob	0,532	1,993	Acot12	0,482	1,435	Lztr1	0,616	2,896	Zc3h7b	0,616	2,651
Armc3	0,331	2,911	Adamts1	0,748	1,228	Lzts2	0,616	3,194	Zc3hav1	0,616	2,625
Atp13a1	0,641	2,896	Adck2	0,600	2,651	Map3k14	0,359	1,443	Zfyve16	0,584	2,651
BC094916	0,541	2,238	Adrbk2	0,592	2,641	Map4k2	0,525	2,043	Zmym6nb	0,928	1,053
Best1	0,293	1,032	Agap1	0,574	1,227	Map7	0,796	1,185	Zmynd11	0,184	6,931
Bhlhe40	0,988	1,287	Alpk3	0,192	1,046	Mark3	0,592	1,473	Znf516	0,682	1,912
C7	0,822	2,994	Ankdd1b	0,736	1,187	Mdc1	0,284	1,574	Znf541	0,978	1,358
Ccdc113	0,852	1,381	Ankrd63	0,592	2,651	Mfhas1	0,344	4,395			
Cfap44	0,935	2,223	Anp32c	0,380	2,189	Mgat4c	0,845	1,232			
Chd2	0,411	1,313	Apba2	0,564	1,078	Mocs1	0,533	2,094			
Chd8	0,886	1,534	Arhgap15	0,598	1,474	Mrps15	0,270	1,895			
Chmp5	0,286	2,158	Arhgap6	0,576	1,400	Mtfr2	0,600	2,726			
Clip1	0,993	1,023	Arhgef10l	0,608	2,651	Myk	0,608	2,726			
Clip2	0,871	2,239	Atic	0,351	1,572	Myo16	0,159	3,076			
Cpsf1	0,533	2,110	Atp10a	0,708	1,068	Myo18a	0,795	1,342			
Csde1	0,600	2,651	Atrip	0,359	1,888	Myo7b	0,174	1,418			
Dak	0,616	3,194	Baz1b	0,527	1,748	Mzb1	0,634	1,196			
Dclre1a	0,847	1,257	Cacna1d	0,921	2,083	Nacac	0,616	2,641			
Depdc1b	0,328	2,088	Cadps2	0,641	2,651	Nalcn	0,533	2,094			
Dnaja1	0,917	1,050	Card11	0,670	1,740	Nap11l	0,532	2,043			
Dnmt3a	0,616	2,651	Card6	0,164	4,395	Nat10	0,667	2,338			
Dscam	0,608	2,896	Cbr1	0,815	2,239	Nav1	0,495	2,026			
Dync1h1	0,686	3,026	Cc2d1a	0,578	1,899	Ndufa2	0,592	2,651			
Efcab3	0,570	1,239	Cdadc1	0,937	1,299	Nlrc5	0,669	2,625			
Efhc2	0,577	1,802	Cep104	0,698	1,784	Npffr1	0,533	2,094			
Eftud1	0,581	1,802	Cep290	0,342	2,095	Nub1	0,592	2,651			
Egfl6	0,361	1,858	Chd1	0,600	2,803	Nucls1	0,667	1,205			
Eif2s3y	0,540	1,441	Chd3	0,700	1,599	Olfr132	0,877	1,196			
Eif4g3	0,533	2,094	Chd5	0,328	2,100	Olfr1387	0,569	2,752			
Epha6	0,534	2,102	Cln5	0,592	2,651	Olfr509	0,634	2,012			
Ephb3	0,533	2,043	Cops4	0,813	1,514	Olfr514	0,715	2,100			
Gldc	0,548	1,993	Coq6	0,224	4,948	Pafah2	0,982	1,072			

Ergebnisse und Diskussion

Fortsetzung Tabelle 7: Für hmU angereicherte Proteine aus dem *pull down* Versuch mit den Basenpaaren hmU:G im Vergleich zu C:G mit ihren jeweiligen TMT-Verhältnissen (Streudiagramm Abbildung 22). Proteine, die den GO-Termen aus Abbildung 23 zugeordnet wurden, sind fett hervorgehoben.

Mit Crosslinker			Ohne Crosslinker								
Gen Name	hmU ¹²⁶ / C ¹²⁷	C ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / C ¹²⁷	C ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / C ¹²⁷	C ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / C ¹²⁷	C ¹²⁶ / hmU ¹²⁷
Gm382	0,506	1,811	Ctgf	0,605	1,681	Pald1	0,328	2,101			
Gon4l	0,405	2,894	Cttnbp2	0,532	2,110	Pan2	0,552	1,750			
Grin2b	0,862	1,097	Cxcr1	0,616	3,194	Pcdh15	0,532	2,094			
H1fo	0,113	4,215	Cyp4f40	0,395	1,108	Pex12	0,671	2,157			
Hand1	0,349	1,974	D8Erttd82e	0,856	1,169	Pgam5	0,641	3,194			
Hist1h1a	0,567	1,892	Daxx	0,488	1,535	Phf19	0,533	2,102			
Hist1h1b	0,494	1,924	Ddhd1	0,359	1,443	Pinx1	0,707	1,068			
Hspa1a	0,592	2,726	Ddx23	0,181	4,304	Pkdrej	0,655	1,527			
Hydin	0,533	1,990	Ddx4	0,588	1,730	Pla2g4c	0,594	4,178			
Ice1	0,667	1,359	Ddx46	0,470	1,551	Pnmal2	0,299	1,764			
Ikbke	0,616	2,896	Defb23	0,230	3,217	Polq	0,904	1,260			
Ints6	0,362	1,575	Dlg4	0,829	1,632	Polr3b	0,674	1,358			
Ints7	0,591	2,457	Dnah7b	0,810	1,167	Ppie	0,650	1,825			
Kiaa0922	0,496	1,682	Dppa5a	0,516	1,441	Ppm1l	0,086	5,511			
Kif2a	0,052	11,094	Dst	0,591	1,924	Ppp1r12b	0,600	2,651			
Magee1	0,996	2,125	Eif4g2	0,325	2,024	Ppp4r4	0,506	1,924			
Mecp2	0,356	1,635	Elp3	0,480	1,066	Prpf3	0,433	1,301			
Mki67	0,896	1,043	Enpp2	0,183	2,267	Prpf4b	0,734	1,381			
Morc2a	0,218	2,649	Entpd4	0,506	1,861	Prss57	0,584	2,803			
Mtif3	0,292	2,158	Epg5	0,236	2,238	Psd	0,930	1,466			
Mtmr10	0,293	1,682	Esf1	0,581	1,830	Psmc2	0,826	1,375			
Mypn	0,584	2,641	Ets1	0,168	1,684	Rabl6	0,404	2,871			
Ndst1	0,658	2,460	Fam189a2	0,605	2,806	Rad51ap1	0,599	1,175			
Nisch	0,585	1,811	Fam32a	0,605	1,589	Raggef4	0,900	1,058			
Nlrc4	0,595	1,856	Farsb	0,864	1,191	Rasip1	0,608	2,627			
Nom1	0,584	1,850	Fbxl13	0,534	2,094	Rassf1	0,832	1,472			
Obscn	0,539	1,379	Fchsd1	0,532	2,110	Rgs12	0,576	1,273			
Pbrm1	0,736	1,381	Fgfr2	0,532	2,110	Rnf152	0,930	1,317			
Pcdh20	0,532	2,094	Fgfr3	0,204	3,026	Rpl34	0,696	1,083			
Pde4dip	0,600	2,625	Fryl	0,552	1,811	Rps6	0,733	1,055			
Phf20l1	0,979	2,913	Gars	0,608	2,651	Rrbp1	0,314	1,620			
Plekha6	0,552	1,861	Ggcx	0,559	1,573	Rsl1d1	0,105	1,414			
Pnma2	0,529	1,836	Gm14151	0,742	2,280	Rufy2	0,899	1,156			
Pnpla8	0,941	1,452	Gm17673	0,180	2,549	Scn4a	0,430	2,280			
Polr1a	0,599	2,397	Gm4955	0,467	1,413	Sepsecs	0,795	1,022			
Ppwd1	0,506	1,811	Gm9837	0,848	1,341	Setdb1	0,529	1,811			
C3orf33	0,382	1,764	Gnl3	0,479	1,416	Sfrp4	0,112	2,125			
Prx	0,771	1,287	Gpr98	0,533	2,094	Sh3d21	0,532	2,094			
Psd3	0,739	1,041	Gtf2h1	0,918	1,043	Slc35a2	0,802	1,813			
Ptprh	0,909	1,014	Gtf3c1	0,440	7,628	Slc4a7	0,281	2,261			
Rims1	0,600	2,992	H2afv	0,667	1,774	Smarca2	0,404	1,023			
Rpl7	0,608	2,726	H2afx	0,612	1,975	Smc3	0,332	2,111			
Sbno1	0,584	2,630	Havcr2	0,637	2,307	Snappc2	0,466	2,039			
Sdad1	0,105	1,344	Heatr5a	0,338	1,675	Snf8	0,413	4,157			
Setbp1	0,465	1,181	Hist1h1d	0,496	1,992	Spock2	0,744	1,358			
Sh3kbp1	0,611	1,131	Hist1h2ah	0,610	1,964	Synj2	0,667	1,209			
Sh3kbp1	0,611	1,131	Hist1h2bf	0,232	4,783	Tbc1d20	0,474	1,909			
Slnf4	0,967	1,235	Hist1h2bm	0,232	4,783	Tln1	0,790	1,828			
Spink5	0,532	2,110	Hist1h3a	0,737	1,342	Tma7	0,164	4,395			
Syne1	0,437	1,526	Hist1h4a	0,814	1,683	Tmem184c	0,897	1,549			
Tbck	0,600	2,803	Hist2h2ab	0,610	1,933	Tmem194	0,676	2,115			
Tfdp2	0,766	1,070	Hnrnpc	0,802	1,101	Tmem214	0,708	1,055			
Tram1	0,349	2,068	Hspa5	0,548	2,110	Tmprss9	0,592	2,651			
Trim66	0,532	2,094	Hspa8	0,600	2,651	Tnrc18	0,333	2,015			
Ttc21a	0,870	2,175	Hspa9	0,600	2,726	Tnrc18	0,340	1,961			
U2surp	0,960	5,278	Htr7	0,966	1,003	Tpst2	0,525	2,094			
Ubr4	0,336	1,259	Ifih1	0,334	1,003	Trip11	0,975	1,118			
Vps13c	0,563	2,094	Inpp5f	0,486	1,797	Trub1	0,936	1,527			
Vps52	0,183	5,865	Ints1	0,186	5,474	Tssk2	0,846	1,466			
Wnk2	0,599	2,627	lqcd	0,534	1,990	Tubb5	0,813	1,322			
Xrra1	0,592	2,726	Jakmp2	0,322	1,999	Unbekannt	0,600	2,803			
Zbtb38	0,437	3,323	Jakmp3	0,855	1,366	Ust	0,557	1,856			
Zcchc9	0,541	2,110	Kiaa1109	0,584	2,300	Vil1	0,535	1,046			
Zfp644	0,972	1,327	Kiaa1211	0,605	1,678	Vps13b	0,299	2,602			
Zmynd8	0,282	1,888	Kif1b	0,599	1,022	Vps41	0,506	1,811			

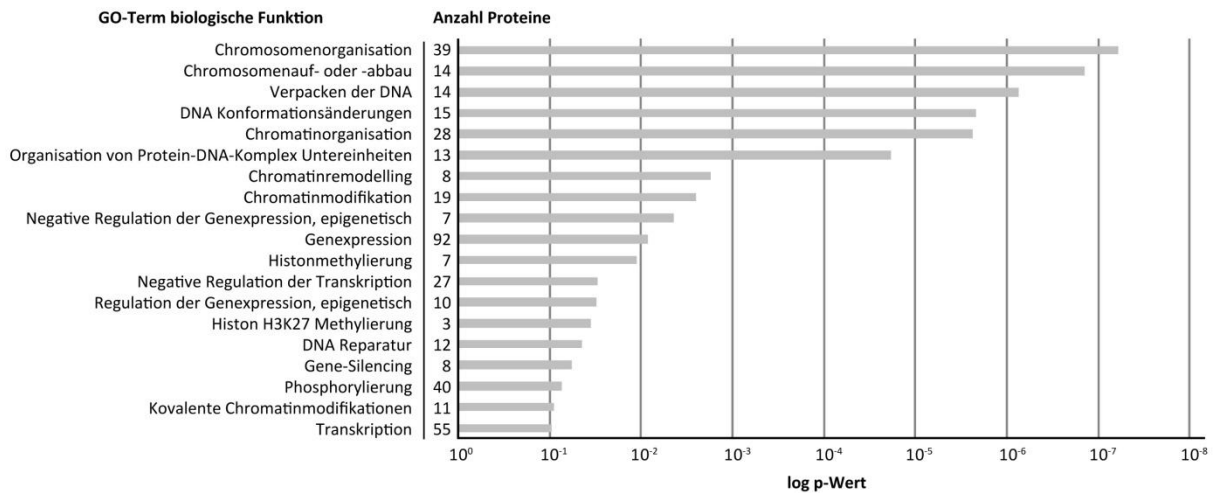


Abbildung 23: DAVID-GO-Term Analyse der für hmU:G angereicherten Proteine. Neben den p-Werten ist auch die Anzahl der Proteine angegeben, die unter dem jeweiligen GO-Term stehen.

Mit insgesamt 322 eindeutig für hmU oder 360 eindeutig für C angereicherten Proteinen konnten in diesem Ansatz sehr viel mehr Proteine als im vorigen Ansatz quantifiziert werden. Allerdings zeigt das Streudiagramm (Abbildung 22) eine breite Verteilung der Proteine, so dass sich nur wenige signifikant von der Menge absetzen. Von den 682 eindeutig quantifizierten Proteinen wurden 246 Proteine mit dem Crosslinker markiert, wovon sich 103 im Quadranten der für hmU angereicherten Proteine befinden. Eine GO-Analyse aller für hmU angereicherten Proteine ergab, dass auch in diesem Ansatz mit 92 Proteinen der Großteil an der Genexpression beteiligt ist und auch der Anteil an Proteinen, die an der Chromosomenorganisation beteiligt sind, ist mit 39 relativ hoch. Darunter befinden sich die acht Chromatinremodeller Baz1b (*bromodomain adjacent to zinc finger domain 1B*), Chd11, Chd5, Chd8 (*chromodomain helicase DNA binding protein 1-like; 5; 8*), Daxx (*fas death domain-associated protein*), Hnrnpc (*heterogeneous nuclear ribonucleoprotein C*), Smarca2 (*SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2*), Tnrc18 (*trinucleotide repeat containing 18*) und die 7 an der Methylierung von Histonen beteiligten Proteine Chd5 (*chromodomain helicase DNA binding protein 5*), Cxxc1 (*CXXC finger 1, PHD domain*), Hist1h1d (*histone cluster 1, H1d*), Kmt2b (*lysine-specific methyltransferase 2B*), Mecp2 (*methyl CpG binding protein 2*), Phf19 (*PHD finger protein 19*) und Setdb1 (*SET domain, bifurcated 1*). Desweiteren fallen unter den 55 Proteinen, die u.a. an der Transkription beteiligt sind, v.a. die DNA-Methyltransferase Dnmt3a, die Transkriptionsfaktoren Gtf2h1, Gtf3c1 und Tfdp2 und mehrere Histonproteine der Familien H1, H2A, H2B, H3 und H4 auf. Interessanterweise wurden darüber hinaus auch 12 Proteine angereichert, die an der DNA Reparatur beteiligt sind, wie zum Beispiel Mdc1 (*mediator of DNA damage checkpoint 1*), Rad51ap1 (*RAD51 associated protein 1*), Smc3 (*structural maintenance of chromosomes 3*) und Xpc (*xeroderma pigmentosum, complementation group C*).

Ergebnisse und Diskussion

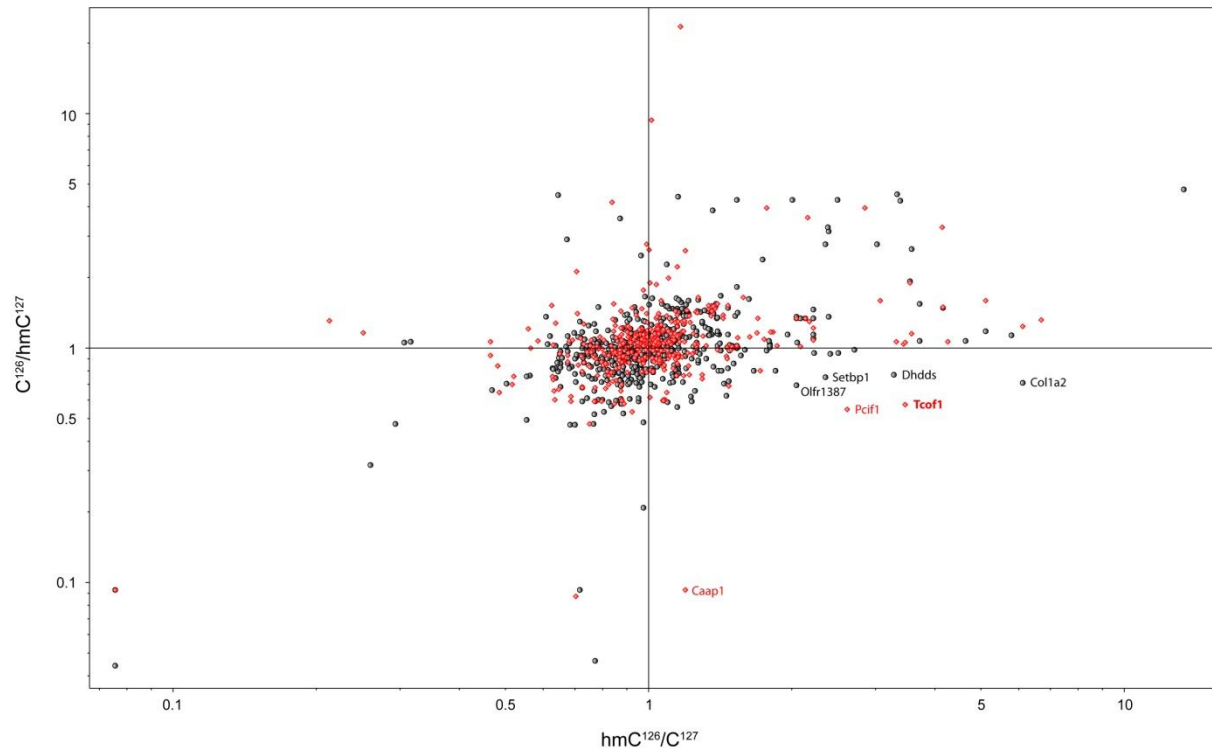


Abbildung 24: Streudiagramm zu den Ergebnissen des *pull down* Versuchs mit den Basenpaaren hmC:G im Vergleich zu C:G. Alle identifizierten Proteine, die mit dem Crosslinker markiert wurden sind als rote Rauten dargestellt.

Tabelle 8: Für hmC angereicherte Proteine aus dem *pull down* Versuchs mit den Basenpaaren hmC:G im Vergleich zu C:G mit ihren jeweiligen TMT-Verhältnissen (Streudiagramm Abbildung 24). Proteine, die den GO-Termen aus Abbildung 25 zugeordnet wurden, sind fett hervorgehoben.

Mit Crosslinker			Ohne Crosslinker					
Gen Name	C^{126}/hmC^{127}	hmC^{126}/C^{127}	Gen Name	C^{126}/hmC^{127}	hmC^{126}/C^{127}	Gen Name	C^{126}/hmC^{127}	hmC^{126}/C^{127}
2010300C02Rik	0,996	1,045	4930407110Rik	0,897	1,085	Itch	0,974	1,080
AI481877	0,975	1,496	Acacb	0,772	1,126	Kiaa1671	0,998	1,083
Alg5	0,896	1,042	Aff3	0,837	1,077	Kif3b	0,841	1,135
Ankrd12	0,914	1,178	App12	0,943	1,016	Ldhc	0,979	1,035
Arhgef17	0,919	1,460	Arcn1	0,883	1,275	Lrrk2	0,977	1,084
Bahcc1	0,980	1,401	Ash1	0,808	1,156	Lrrn2	0,972	1,128
C7	0,597	1,086	Baz1b	0,812	1,141	Lyar	0,952	2,231
Caap1	0,092	1,198	Bcl2l14	0,941	1,282	Map2	0,800	1,142
Ccdc146	0,898	1,104	Bicd2	0,915	1,157	Map3k14	0,705	1,444
Ccdc162	0,558	1,152	Bzw1	0,982	1,598	Myo16	0,762	1,148
Cenpf	0,804	1,008	Cep162	0,980	2,718	Naif1	0,619	1,232
Cep170b	0,798	1,719	Chd9	0,735	1,072	Ndst1	0,863	1,274
Cep290	0,885	1,055	Chl1	0,655	1,255	Ndufb4	0,966	1,010
Col9a2	0,762	1,188	Cilp	0,795	1,402	Nelfe	0,978	1,040
Dnah5	0,643	1,095	Col1a2	0,707	6,126	Nemf	0,587	1,098
Dnm1l	0,938	1,238	Coq6	0,756	1,130	Olfr1387	0,694	2,048
Dync2h1	0,919	1,162	Crot	0,849	1,071	P2ry12	0,594	1,078
Esf1	0,600	1,077	Ddx4	0,707	1,111	Pan2	0,678	1,032
Fer1l5	0,762	1,147	Dhdds	0,764	3,291	Pdia2	0,698	1,013
Fsip2	0,973	1,100	Dhx8	0,978	1,043	Pex12	0,971	1,419
Gm8674	0,764	1,221	Diap1	0,594	1,078	Phb	0,895	1,325
Gm906	0,839	1,216	Dnah10	0,785	1,025	Pif1	0,990	1,027
Huwe1	0,989	1,008	Dnah11	0,952	1,042	Pip5k1b	0,815	1,018
Ig hc V-III region W3082	0,942	1,051	Dnajc1	0,971	1,076	Pla2g4a	0,944	1,006
Kif3a	0,855	1,153	Dnm3	0,798	1,023	Plxnb1	0,712	1,183
Lrp2	0,884	1,063	Dnmt1	0,835	1,156	Pold3	0,588	1,236
Marco	0,904	1,019	Dot1l	0,797	1,855	Ppfia1	0,814	1,470
Mlh1	0,869	1,028	Eif3b	0,899	1,235	Ppie	0,989	1,067
Mpp7	0,891	1,142	Fam186b	0,975	1,485	Prkcd	0,921	1,196
Msh3	0,929	1,111	Fam64a	1,000	1,534	Ptpn21	0,971	1,071
Myo7a	0,947	1,245	Fancd2	0,925	1,042	Rlim	1,000	1,541
Myo9a	0,897	1,469	Fbxw19	0,959	1,113	Rps27a	0,987	1,013
Nampt	0,933	1,156	Fgfr1op	0,965	1,191	Rufy2	0,916	1,088
Ncoa7	0,987	1,045	Figl1	0,928	1,073	Serpinc3c	0,924	1,569
Olfr1454	0,990	1,090	Fmn1	0,717	1,479	Serpind1	0,950	2,510
Paip1	0,950	1,210	Fndc3c1	0,992	1,265	Setbp1	0,749	2,357
Pcif1	0,545	2,618	Galk2	0,826	1,046	Srrm2	0,914	1,149
Phf20	0,594	1,072	Galnt10	0,799	1,665	Stard9	0,594	1,078

Fortsetzung Tabelle 8: Für hmC angereicherte Proteine aus dem *pull down* Versuch mit den Basenpaaren hmC:G im Vergleich zu C:G mit ihren jeweiligen TMT-Verhältnissen (Streudiagramm Abbildung 24). Proteine, die den GO-Termen aus Abbildung 25 zugeordnet wurden, sind fett hervorgehoben.

Mit Crosslinker			Ohne Crosslinker					
Gen Name	C ¹²⁶ / hmC ¹²⁷	hmC ¹²⁶ / C ¹²⁷	Gen Name	C ¹²⁶ / hmC ¹²⁷	hmC ¹²⁶ / C ¹²⁷	Gen Name	C ¹²⁶ / hmC ¹²⁷	hmC ¹²⁶ / C ¹²⁷
Phf21a	0,943	1,136	Gbe1	0,827	1,374	Syne4	0,819	1,202
Polr3a	0,804	1,082	Gdap111	0,971	1,419	Tnrc18	0,799	1,426
C3orf33	0,864	1,059	Glud1	0,879	1,023	Trim11	0,947	1,033
Prx	0,766	1,301	Gm20489	0,895	1,235	Trim26	0,998	1,159
Rpl5	0,818	1,131	Gm6658	0,923	1,080	Trim42	0,933	1,024
Rsf1	0,594	1,078	Gm8439	0,625	1,463	Ttc39d	0,776	1,158
Rtf1	0,688	1,477	Gm973	0,936	1,044	C3orf20	0,959	1,129
S100a8	0,994	1,024	Gpr98	0,994	1,534	Unkl	0,986	1,309
Scn10a	0,971	1,024	Gstt3	0,965	1,017	Zfc3h1	0,986	1,119
Smc1b	0,612	1,009	H2afj	0,971	1,120	Znf687	0,937	1,046
Sorcs1	0,843	1,082	H2afx	0,989	1,119	Zscan18	0,938	2,416
Stag1	0,739	1,300	H2afz	0,971	1,075			
Tcof1	0,572	3,475	Hist1h2bf	0,716	1,130			
Tgm3	0,792	1,023	Hist1h2bm	0,716	1,130			
Tram1	0,827	1,074	Hist1h4a	0,954	1,267			
Ubr5	0,925	1,129	Hsp90ab1	0,856	1,512			
Wbscr16	0,968	1,063	lft122	0,996	1,068			
Wdr11	0,914	1,156	lgf1r	1,000	1,501			
Wnk2	0,928	1,105	lgfbp4	1,000	1,534			
Zdbf2	0,944	1,155	ll1rapl2	0,968	1,774			

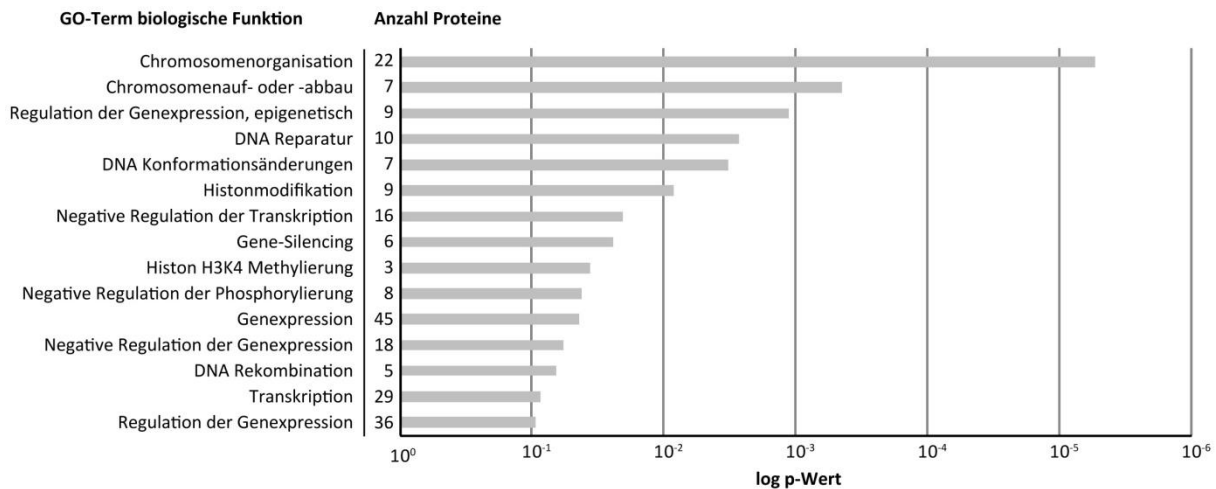


Abbildung 25: DAVID-GO-Term Analyse der für hmC:G angereicherten Proteine. Neben den p-Werten ist auch die Anzahl der Proteine angegeben, die unter dem jeweiligen GO-Term stehen.

Für hmC (165) und C (189) wurden insgesamt 354 Proteine eindeutig in den Quadranten I und III (Abbildung 24) quantifiziert. 141 Proteine wurden mit dem Crosslinker markiert, davon sind 58 Proteine für hmC angereichert. Laut der GO-Analyse (Abbildung 25) haben in der Liste der für hmC angereicherten Proteine (Tabelle 8) mit einem p-Wert von $5,33 \cdot 10^{-06}$ solche Proteine eine hohe Relevanz, die bei der Chromosomenorganisation involviert sind. Dazu gehören Proteine wie Chd9 (*chromodomain helicase DNA binding protein 9*), Dnmt1 (*DNA methyltransferase (cytosine-5) 1*), Huwe1 (*HECT, UBA and WWE domain containing 1*) und Ubr5 (*ubiquitin protein ligase E3 component n-recognin 5*). Die letzten beiden Proteine spielen auch bei der DNA Reparatur eine Rolle sowie acht weitere Proteine aus der Liste der für hmC angereicherten Proteine.

Bei dem Vergleich der drei Streudiagramme, die mit den verschiedenen DNA-Sonden erhalten wurden, fällt auf, dass bei dem Ansatz für hmC:G sich die Proteine eher auf einer Achse von Quadrant II nach IV befinden, statt auf einer Achse, die von Quadrant III nach I verläuft. Dies kann mehrere Ursachen haben. Trotz sorgfältiger Arbeitsweise, kann das Ergebnis ein Hinweis darauf sein, dass die Markierung mit den TMT-Reagenzien nicht optimal funktioniert hat und die Peptide nicht vollständig markiert wurden. Da die TMT-Reagenzien aber im Überschuss zugegeben wurden, ist es wahrscheinlicher, dass an den DNA-Sonden (hmC:G vs. C:G) schlichtweg mehr unspezifische Proteine angereichert wurden als bei den beiden Ansätzen zuvor (hmU:A vs. T:A bzw. hmU:G vs. C:G). Im weiteren Vergleich wurden die Listen der Proteine, die für das jeweilige Cytosin-Derivat angereichert wurden, abgeglichen und es wurde nach Überschneidungen gesucht. Das Ergebnis ist in Abbildung 26 als Venn-Diagramm dargestellt. Die Überschneidungen der Ansätze sind sehr gering. So wurden nur 4 Proteine gefunden, die in allen drei Ansätzen angereichert wurden: Die beiden Histonproteine Hist1h2bm (Histon H2A) und Hist1h4a (Histon H4), das DNA-bindende Protein Setbp1 (*SET-binding protein*) und das zytosolische Protein Gpr98 (*G-protein coupled receptor 98*), welches allerdings als Verunreinigung des Zellkernlysats vernachlässigt werden kann. Zwischen den Ansätzen mit hmU:A und hmC:G gibt es drei Überschneidungen: das Myosin-Protein Myo9a (*Unconventional myosin-9a*), das Protein Gdap1l1 (*Ganglioside-induced differentiation-associated protein 1-like 1*) und das Histon-Protein H2afz (Histon H2A). Bei den Ansätzen hmC:G und hmU:G gibt es 21 Überschneidungen. Davon spielen 8 Proteine (Ddx4, Esf1, H2afx, Baz1b, Cep290, Ppie, Prx und Tnrc18) eine Rolle bei der Regulation der Genexpression. Und bei den Ansätzen hmU:A und hmU:G wurden 11 sich überschneidende Proteine gefunden, wovon 5 als cytoplasmatische Proteine (Dnah7b, Hydin, Cep104, Sh3d21, Mylk) vernachlässigbar sind. Unter den 6 verbliebenen Proteinen befinden sich die PHD- und Zinkfinger Proteine Phf20l1 (*PHD finger protein 20-like protein 1*) und Zcchc9 (*Zinc finger CCHC domain-containing protein 9*), das an der DNA-Reparatur beteiligte Protein Rad51ap1 (*RAD51 associated protein 1*), das bei der Translation beteiligte Protein Gars (*Glycine-tRNA ligase*), bei der Signalübertragung beteiligte Protein Adrbk2 (*Beta-adrenergic receptor kinase 2*) und das phosphorylierende Protein Ephb3 (*Ephrin type-B receptor 3*).

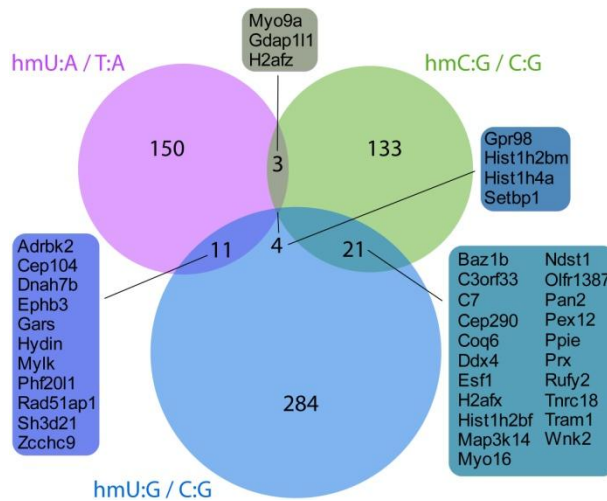


Abbildung 26: Venn-Diagramm zum Abgleich der *pull down* Ansätze mit unterschiedlichen DNA-Sonden. Die dargestellten Flächen der Kreise und deren Überschneidungen sind proportional zur dargestellten Anzahl der entsprechenden Proteine.

Obwohl Messungen von Toni Pfaffeneder zeigten, dass in mESCs die hmU-Level in etwa um den Faktor 1000 geringer sind als die hmC-Level, wurden in den hier durchgeführten *pull down* Studien vergleichbar viele Proteine für beide DNA-Modifikationen angereichert. Unsere Arbeitsgruppe konnte außerdem zeigen, dass hmU hauptsächlich durch die Oxidation von T durch Tet-Enzyme entsteht, deshalb müsste hmU in dem verwendeten Zellmaterial überwiegend gepaart mit A vorliegen. Denn eine Desaminierung von hmC durch Tdg, woraus hmU:G resultieren würde, konnte auch fast gänzlich ausgeschlossen werden. Umso überraschender ist es, dass mit 322 Proteinen für hmU:G fast doppelt so viele Proteine angereichert wurden als für hmU:A (168) oder hmC:G (165).

In allen drei Ansätzen waren laut GO-Analyse zwischen 27 und 29 % der angereicherten Proteine an der Genexpression beteiligt und zwischen 10 und 13 % an der Chromosomenorganisation. Trotzdem überschneiden sich die Proteine aus den jeweiligen Proben lediglich geringfügig, nur vier der angereicherten Proteine wurden in allen drei Ansätzen gefunden, wovon zwei Histonproteine (H2B und H4) sind und das Protein Setbp1 (vgl. Abbildung 26). Nur drei Überschneidungen gibt es zwischen den Ansätzen hmU:A und hmC:G, 11 zwischen hmU:A und hmU:G sowie 21 zwischen hmU:G und hmC:G. Dies könnte tatsächlich darauf hindeuten, dass viele Proteine nicht nur eine DNA-Modifikation spezifisch erkennen können, sondern auch unterscheiden mit welcher Base die Modifikation gepaart ist und damit der Ursprung der Modifikation eine erhebliche Rolle spielt.

In dem Ansatz hmU:A spielen laut GO-Analyse 4 % der angereicherten Proteine eine Rolle bei der DNA-Reparatur und 6 % der angereicherten Proteine des Ansatzes hmU:A bei der zellulären Antwort auf DNA Schäden. Dabei ist bekannt, dass hmU effizient von Reparaturenzymen, wie der Glykosylase Smug1, erkannt und repariert wird.^[189-191] Es ist vorstellbar, dass während der Oxidation von mC zu hmC durch Tet-Enzyme auch eine kleine Menge an hmU entsteht, die eine „*long-patch BER*“ DNA-

Reparatur induzieren und damit eine schnellere Demethylierung bedingen könnte.^[192] Andererseits wurden auch im Ansatz von hmC:G 6 % der angereicherten Proteine per GO-Analyse der DNA-Reparatur zugeschrieben, zwei Proteine (Huwe1, Ubr5) wurden sogar mit dem Crosslinker markiert. Allerdings ist hmC eigentlich nicht als Substrat für DNA-reparierende Enzyme bekannt.^[65, 193]

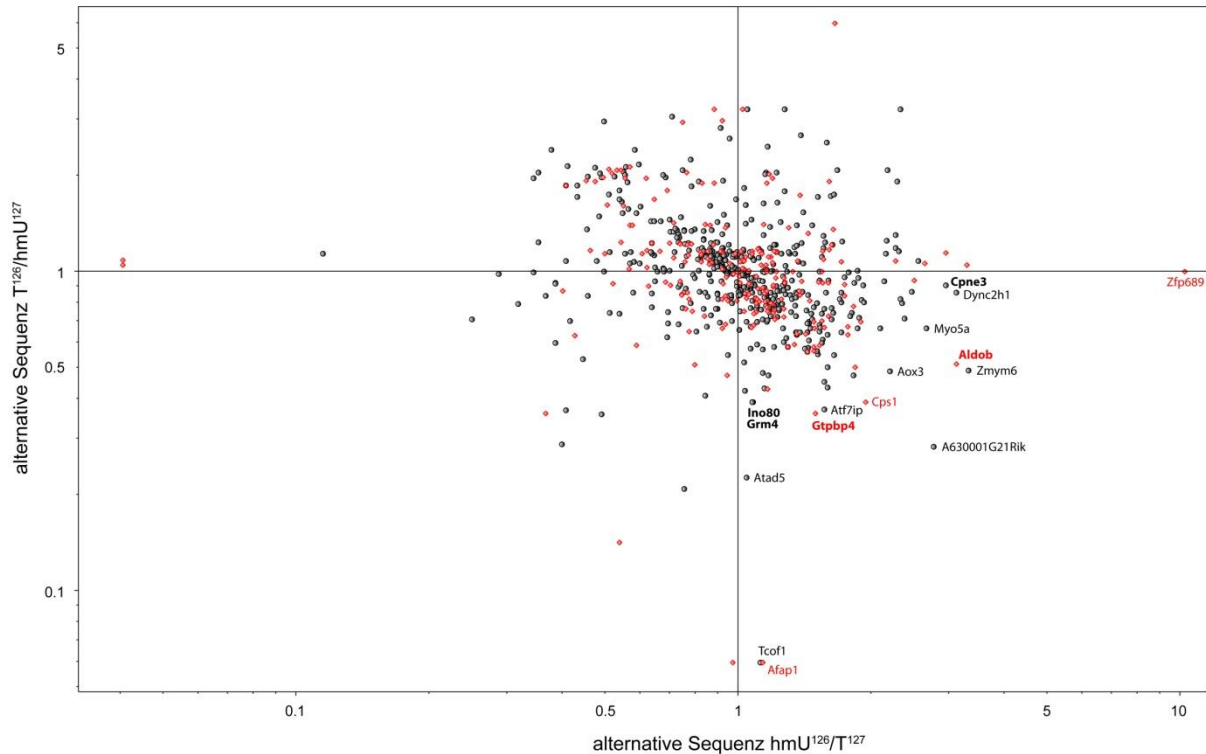


Abbildung 27: Streudiagramm zu den Ergebnissen des *pull down* Versuch mit den Basenpaaren hmU:A im Vergleich zu T:A im Kontext einer alternativen Sequenz. Alle identifizierten Proteine, die mit dem Crosslinker markiert wurden sind als rote Rauten dargestellt.

Tabelle 9: Für hmU angereicherte Proteine aus dem *pull down* Versuch mit den Basenpaaren hmU:A im Vergleich zu T:A im Kontext einer alternativen Sequenz mit ihren jeweiligen TMT-Verhältnissen (Streudiagramm Abbildung 27). Proteine, die den GO-Termen aus Abbildung 28 zugeordnet wurden, sind fett hervorgehoben.

Mit Crosslinker				Ohne Crosslinker										
Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷		Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷		Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷		Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷
Afap1	1,142	0,059		A630001G21Rik	2,783	0,282		H3f3a	1,010	0,653		Spta1	1,246	0,863
Ahnak	1,211	0,751		Abcg5	1,141	0,567		Heatr5a	1,536	0,996		Stard13	1,365	0,635
Aldob	3,123	0,511		Abl2	1,201	1,000		Hes2	1,213	0,836		Tacc2	1,751	0,955
Ampd3	1,049	0,818		Actr3	1,647	0,736		Hist1h2bm	1,094	0,902		Tcof1	1,125	0,059
Ankrd17	1,273	0,717		Adam29	1,439	0,558		Hist3h2bb	1,094	0,900		Tex11	1,049	0,692
Apeh	1,107	0,814		Afg3l2	1,069	0,569		Hivep2	1,009	0,898		Theg	1,380	0,971
Arfgap2	1,061	0,842		Akap13	1,523	0,547		Hnrnpa1	1,363	0,842		Tma7	1,317	0,838
Asic1	1,492	0,658		Akap6	1,149	0,887		Hsd11b1	1,517	0,790		Tnfrsf21	1,716	0,856
Atp9b	1,554	0,694		Aldh1a1	1,268	0,961		Htt	1,139	0,929		Tpm1	2,103	0,662
Bmp10	1,220	0,707		Alyref	1,069	0,877		Hyi	1,889	0,711		Trpc4	1,785	0,991
Bptf	1,347	0,589		Anapc5	1,636	0,705		Ikbbk	1,017	0,868		Trpm6	1,871	0,661
Ca1	1,271	0,758		Ank3	1,476	0,706		Il10ra	1,231	0,862		Ttc7	1,104	0,761
Camk4	1,088	0,733		Aox3	2,214	0,484		Il18r1	1,282	0,842		Uba6	1,266	0,899
Carf	1,455	0,806		Aqp12	1,420	0,569		Il4r	1,491	0,769		Ubn1	1,231	0,715
Ccdc136	1,224	0,995		Arhgap35	1,128	0,853		Impact	1,089	0,911		Ugt1a9	1,187	0,733
Ccdc79	1,123	0,802		Arl1	1,080	0,807		Ino80	1,080	0,388		Unc5a	2,334	0,814
Cenpc	1,149	0,849		Armxc4	1,443	0,838		lqca1p1	1,188	0,721		Upf2	1,511	0,562
Clip1	1,427	0,633		Arpp19	1,041	0,848		C1orf112	1,279	0,914		Usf1	1,263	0,788
Cnot1	1,279	0,914		Atad5	1,049	0,225		Itpr1	1,654	0,545		Ushbp1	1,450	0,690

Ergebnisse und Diskussion

Fortsetzung Tabelle 9: Für hmU angereicherte Proteine aus dem *pull down* Versuch mit den Basenpaaren hmU:A im Vergleich zu T:A im Kontext einer alternativen Sequenz mit ihren jeweiligen TMT-Verhältnissen (Streudiagramm Abbildung 27). Proteine, die den GO-Termen aus Abbildung 28 zugeordnet wurden, sind fett hervorgehoben.

Mit Crosslinker			Ohne Crosslinker								
Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷	Gen Name	hmU ¹²⁶ / T ¹²⁷	T ¹²⁶ / hmU ¹²⁷
Coq6	1,400	0,929	Atf7ip	1,570	0,367	Jph3	1,207	0,817	Vdac1	1,149	0,428
Cpa1	1,574	0,665	Atp2b4	1,132	0,753	Kdm5a	1,120	0,935	Vmn2r115	1,276	0,872
Cps1	1,950	0,388	Atp5a1	1,082	0,856	Krt17	1,406	0,857	Vmn2r22	1,226	0,782
Crocc	1,099	0,820	AW554918	1,028	0,971	Ksr2	1,181	0,769	Vps37a	1,358	0,720
Dkk1	1,487	0,692	BC048403	1,241	0,752	Lap3	1,356	0,826	Vps51	1,289	0,615
Dnah7b	1,168	0,995	Cage1	1,073	0,765	Lrrc27	1,574	0,449	Xpo4	1,543	0,664
Dock10	1,105	0,740	Ccdc38	1,596	0,431	Lrrc36	1,912	0,812	Zcwpw2	1,831	0,469
Efcab6	1,255	0,814	Cdkl3	1,324	0,742	Lrrc74b	1,006	0,833	Zfhx3	1,012	0,749
F5	1,205	0,737	Cep5711	1,078	0,949	Luc7l2	1,191	0,995	Zfp804b	1,144	0,863
Fam110b	1,751	0,908	Chd1	1,063	0,943	Mcf2	1,142	0,702	Zim1	1,356	0,643
Fb1	1,042	0,850	Chd5	1,056	0,991	Mettl16	1,707	0,817	Zmym6	3,337	0,486
Fbxo38	1,379	0,864	Chek2	1,268	0,659	Mical3	1,470	0,733	Znf526	1,073	0,863
Fgfbp1	1,200	0,787	Chml	1,299	0,887	Mpv17l	1,439	0,684			
Gcc2	1,139	0,886	Chmp7	1,098	0,819	Ms4a5	1,096	0,953			
Gdpc4	1,898	0,692	Kcap5	1,801	0,720	Mthfd1l	1,101	0,801			
Glmn	1,777	0,670	Cltc	1,166	0,692	Mylk	1,052	0,861			
Gm15800	1,525	0,921	Coq2	1,159	0,777	Myo5a	2,675	0,661			
Gm884	1,365	0,635	Cpne3	2,962	0,900	Ncam1	1,837	0,915			
Golga3	1,604	0,914	Cyp2t4	1,331	0,814	Nme4	1,061	0,898			
Golgb1	1,168	0,864	Dapk1	1,212	0,780	Oftr235	1,550	0,593			
Grm1	1,215	0,706	Dars	2,479	0,857	Osbpl9	1,036	0,971			
Gtppb4	1,504	0,358	Dbr1	1,194	0,768	Pard6b	1,384	0,642			
Haus7	1,486	0,561	Ddx17	1,159	0,848	Pbrm1	1,601	0,630			
Htatsf1	1,558	0,916	Dmtn	1,463	0,660	Pcca	1,141	0,995			
Kif6	1,535	0,585	Dnah8	1,556	0,733	Pgd	1,166	0,838			
Klk6	1,062	0,999	Dnah9	1,813	0,736	Phldb2	1,439	0,642			
Lca5l	1,157	0,801	Dnm1l	1,106	0,776	Pla2r1	1,453	0,685			
Map3k9	1,191	0,794	Dnmt3b	1,189	0,853	Plcb3	1,598	0,497			
Mmaa	1,076	0,664	Dstn	1,876	0,949	Pls1	1,037	0,517			
Myo5b	1,101	0,959	Dtwd2	1,172	0,806	Pnck	1,035	0,964			
Nnt	1,204	0,898	Dync2h1	3,123	0,853	Polq	1,110	0,586			
Nrf1	1,385	0,838	Eea1	1,532	0,677	Prg4	1,605	0,931			
Nrk	1,206	0,888	Eif2ak1	1,360	0,631	Prr12	1,051	0,961			
Oip5	1,171	0,427	Eif3d	1,930	0,796	Prr14	1,273	0,867			
Plxnc1	1,171	0,748	Esco2	1,374	0,860	Rab22a	1,434	0,752			
Prom2	1,298	0,576	F8	1,064	0,863	Rab39b	1,536	0,645			
C3orf33	1,486	0,578	Fam208b	1,106	0,692	Raf1	1,225	0,894			
Ptpn23	1,334	0,995	Fer	1,930	0,898	Rag1	1,035	0,891			
Rev3l	1,167	0,769	Flii	1,073	0,762	Ragef2	1,149	0,826			
Rnmtl1	1,088	0,677	Frmf4a	1,057	0,995	Rb1cc1	1,360	0,642			
Secisbp2	1,552	0,995	Gm20821	1,180	0,761	Rfx5	1,558	0,608			
Sgms2	1,008	0,959	Gm2178	1,589	0,853	Ric1	1,352	0,779			
Snx15	1,608	0,723	Gm21885	1,172	0,700	Rlf	1,631	0,763			
Stac3	1,222	0,949	Gm3371	1,210	0,576	Rnft2	1,070	0,670			
Stxbp3b	1,276	0,949	Gm3944	1,275	0,597	Rps16	1,142	0,477			
Tenm2	1,003	0,991	Gm7489	1,717	0,640	Rufy4	2,388	0,710			
Thbs2	1,037	0,745	Gm7964	1,449	0,606	Sacs	1,326	0,932			
Thbs3	1,829	0,775	Gm8906	1,571	0,761	Scn4a	1,014	0,838			
Trnt1	1,779	0,628	Gnl3l	1,358	0,855	Sdr16c6	1,178	0,811			
Unbekannt	1,850	0,498	Gppb1	1,239	0,761	Sepsecs	1,017	0,967			
C9orf117	1,749	0,836	Gpr110	1,302	0,576	Serpina12	1,040	0,420			
Vav1	1,113	0,881	Gpr158	1,537	0,646	Setd2	2,361	0,793			
Wdfy3	1,105	0,791	Grp2	1,537	0,935	Sh3d21	1,293	0,996			
Wdr89	2,514	0,931	Grm4	1,085	0,388	Slk	1,716	0,798			
Zbbx	1,436	0,559	Gsta1	1,151	0,989	Smarca2	1,239	0,803			
Zfp644	1,062	0,913	Gtf3c2	1,174	0,470	Smg6	1,444	0,573			
Zfp689	10,276	0,996	Gys1	2,103	0,658	Sos2	2,152	0,927			
Zmynd11	1,016	0,917	H1foo	1,078	0,906	Sp100	1,006	0,861			

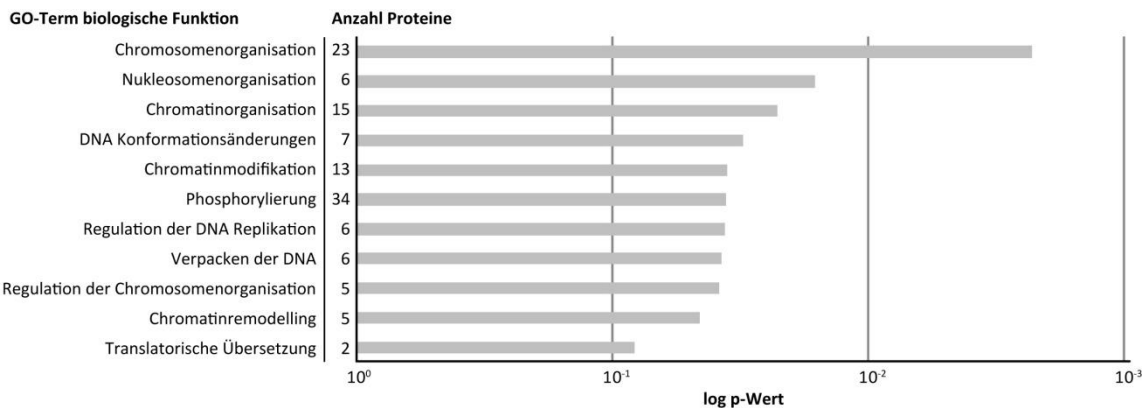


Abbildung 28: DAVID-GO-Term Analyse der für hmU:A angereicherten Proteine im Kontext einer alternativen Sequenz.

Das Streudiagramm aus Abbildung 27 zeigt die Verteilung der insgesamt 475 in den *pull down* Ansätzen mit den Basenpaaren hmU:A und T:A enthaltenden DNA-Sonden mit anderer Basensequenz eindeutig quantifizierten Proteine. Davon wurden 262 ausschließlich für hmU und 213 für T angereichert, wobei 137 Proteine (77 für hmU und 60 für T) mit dem Crosslinker markiert waren. Die GO-Analyse ergab, dass über die Hälfte der angereicherten Proteine eher in metabolischen Prozessen involviert sind (nicht gezeigt). Epigenetisch mehr relevante Prozesse wie Genexpression, Translation oder Ähnliche wurden nicht oder nur mit sehr geringen p-Werten aufgelistet (vgl. Abbildung 28). Noch einigermaßen aussagekräftig ist der GO-Term Chromosomenorganisation, unter welchem die z.T. bereits erwähnten 23 Proteine, wie z.B. Dnmt3b, H3f3a, Hist1h2bm, Chd1, Chd5, Smarca2 und Bptf, aufgelistet wurden.

Ein Vergleich zwischen den beiden hmU:A *pull down* Ansätzen mit unterschiedlichen Basensequenzen ist in Abbildung 29 visualisiert. Überschneidungen gibt es mit 23 indirekt angereicherten Proteinen und mit 5 Crosslinker-markierten Proteinen.

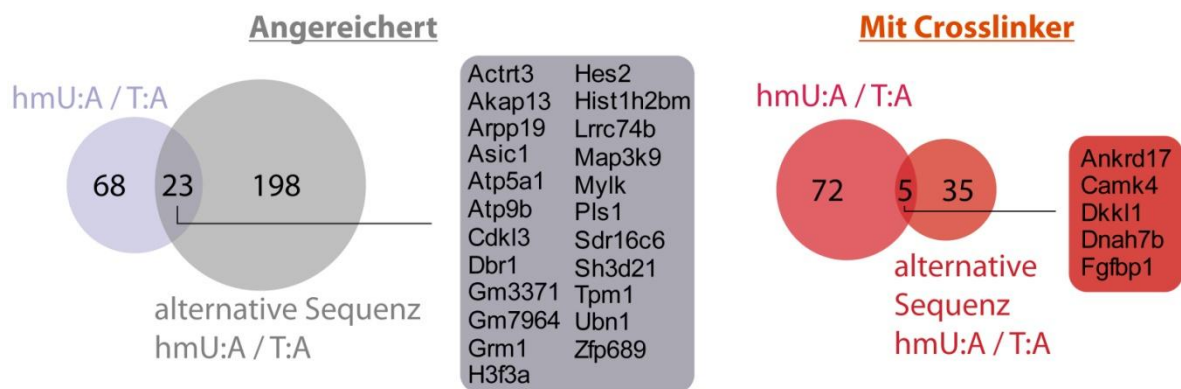


Abbildung 29: Venn-Diagramme zum Vergleich zwischen den für hmU:A angereicherten Proteinen in zwei unterschiedlichen Sequenzen.

Die Wiederholung des Experiments mit der hmU:A enthaltenden DNA-Sonde und einer zufällig gewählten anderen Basensequenz führte nicht wie erwartet zu einem ähnlichen Ergebnis. Von den 262 hierbei für hmU angereicherten Proteinen sind laut einer GO-Analyse über die Hälfte an metabolischen Funktionen und nur 8 % an der Chromosomenorganisation beteiligt. Die Ursache dafür liegt wohl zum Großteil an einer unzureichenden Zellkernlysatpräparation, denn die GO-Analyse zeigte auch, dass 55 % der für T und hmU (allein für hmU sind es 45 %) angereicherten Proteine zytosolisch sind. Aber selbst nach dem diese aus der Liste entfernt wurden, sind auch die verbleibenden nuklearen Proteine fast ausschließlich an metabolischen Prozessen beteiligt. Natürlich muss berücksichtigt werden, dass sich die Anwesenheit der zytosolischen Proteine auf den ganzen Versuch negativ auswirken. Da die eingesetzte Proteinmenge anhand der Gesamtkonzentration bestimmt wurde, wurden dem entsprechend weniger nukleare Proteine eingesetzt. Darüber hinaus

könnten die zytosolischen Proteine in Konkurrenz mit den nuklearen Proteinen stehen, auch wenn sie unter physiologischen Bedingungen nicht mit der DNA in Wechselwirkung treten können. Damit ist es nicht mehr verwunderlich, dass nur rund 17 % der Proteine aus dem ersten hmU:A Ansatz auch im zweiten Ansatz wiedergefunden werden konnten (vgl. Abbildung 29). Um eine Aussage über eine mögliche Sequenzabhängigkeit zu treffen, müsste der Versuch ein weiteres Mal wiederholt werden.

Interessant ist desweiteren, dass in allen vier *pull down* Ansätzen um die 30 % der Proteine mit dem Crosslinker markiert wurden. Mit 21 % wurden im ersten Ansatz von hmU:A am wenigsten angereicherte Proteine gefunden, die mit dem Crosslinker markiert waren, und mit 40 % im Ansatz von hmC:G am meisten, aber nur geringfügig mehr als im Ansatz mit hmU:G, worin 36 % markierte Proteine gefunden wurden. Unter den markierten Proteinen befinden sich drei Histon H1 Proteine (hmU:G), Chd4 (hmU:A), MeCP2 und Dnmt3a (hmU:G), um die interessantesten Proteine zu nennen. Sie stehen alle im Zusammenhang mit einer Einflussnahme auf die Transkription, beispielsweise können Chd4, MeCP2 die Transkription reprimieren und Dnmt3a ist an der *de novo* Methylierung der DNA beteiligt. Deshalb liegt die Vermutung nahe, dass hmU die Bindung von Transkriptionsfaktoren beeinflussen könnte, was bekanntermaßen ein Kennzeichen zur epigenetischen Reprogrammierung ist.^[194] In der Wiederholung des Ansatzes mit hmU:A und der alternativen Basensequenz wurden 29 % mit dem Crosslinker markierte Proteine identifiziert, wobei Chd4 nicht wieder gefunden werden konnte. Die ähnlich hohen Anteile an markierten Proteinen in den Ansätzen könnten ein Hinweis darauf sein, dass die Konzentration der DNA-Sonden höher als die Konzentration von direkt DNA-bindenden Proteinen war. In diesem Fall könnten auch indirekt oder nicht DNA-bindende Proteine mit dem Crosslinker reagieren bis dieser vollständig mit Proteinen gesättigt oder hydrolysiert wurde. Um dies zu vermeiden könnte man entweder weniger DNA-Sonden einsetzen oder nach der Inkubationszeit den pH-Wert erhöhen, um damit die Reaktion zu stoppen.

3.2.2 Crosslinking-Interaktionsstudie mit 5-Methylcytosin und 5-Hydroxymethylcytosin in murinen Retinageweben

Die Entdeckung, dass Tet-Enzyme mC zu hmC, fC und caC oxidieren können und die beiden letzteren Cytosin Derivate von Tdg ausgeschnitten werden können (vgl. Abbildung 4), lässt auf einen möglichen Prozess zur aktiven Demethylierung schließen.^[62-65] Die Level der Oxidationsderivate variieren signifikant zwischen unterschiedlichen Gewebe- bzw. Zelltypen.^[195] Besonders interessant ist der Verlauf in Neuronalen Zellen, da diese nicht mehr fähig zur Zellteilung sind und daher Effekte ausgeschlossen werden können, die sich aus einer erhöhten DNA-Replikation ergeben und zu einer passiven Demethylierung führen (vgl. Abbildung 3). Es konnte festgestellt werden, dass die hmC-Level in Neuronen altersabhängig ansteigen, wohingegen die Level der höheren Oxidationsstufen fC und caC unter dem Detektionslimit bleiben.^[196, 197] Desweiteren korreliert die Oxidation von 5mC zu 5hmC durch Tet Enzyme in Neuronen mit einer erhöhten Genexpression.^[196, 198-201] Dies könnte bedeuten, dass hmC mehr als die Rolle einer bloßen Übergangsstufe im Prozess einer aktiven Demethylierung spielen könnte. Vielmehr könnte hmC im Sinne einer epigenetischen Markierung zur Regulation wichtiger biologischer Prozesse beitragen. Allerdings sind weder die Mechanismen bekannt, die zu einer erhöhten 5hmC Bildung in spezifischen Genen führen, noch weiß man, wie dadurch die Transkription hochreguliert wird oder was mit der Modifikation im weiteren Verlauf geschieht.

In einer Kooperation mit Arshan Perera sollte die Rolle der genomischen mC Hydroxylierung durch Tet-Enzyme während der Reifung der Retina bei der Augenöffnung im Maus-Modell untersucht werden. Es wurde davon ausgegangen, dass bei der Entwicklung des Sehnsinns der Tiere epigenetische Änderungen stattfinden. Um herauszufinden welche Proteine bei der Regulation während der neuronalen Differenzierung eine Rolle spielen, wurden *pull down* Experimente vor und nach dem Zeitpunkt, an dem die Mäuse ihre Augen öffnen, durchgeführt. Genauer gesagt, wurden retinale Zellkernlysate von 11 und 21 Tage alten Mäusen mit doppelsträngigen DNA-Oligomeren inkubiert, die jeweils eines der Cytosin-Derivate oder Cytosin als Kontrolle enthielten. Zusätzlich enthielten die DNA-Stränge den, durch Reduktion, abspaltbaren Crosslinker Azide-PEG3-S-S-NHS Ester ($C_{18}H_{29}N_5O_8S_2$). So konnten DNA bindende Proteine angereichert, präzipitiert und tryptisch verdaut werden. Die erhaltenen Peptide wurden anschließend für ein sogenanntes „*label-swap*“-Experiment (*forward* und *reverse*) mit den TMT-Duplex Reagenzien für eine relative Quantifizierung markiert.

Proteine, die im Vergleich mit einer höheren Präferenz (mindestens 1.5-fache Anreicherung) an modifizierte als an unmodifizierte DNA sowohl im *forward* als auch im *reverse* Experiment gebunden

waren, wurden als spezifische Binder angesehen. Durch den Rest des Crosslinkers, der nach der reduktiven Spaltung am Protein gebunden bleibt, konnten zudem Proteine identifiziert werden, die direkt an die DNA gebunden wurden. Die Ergebnisse sind in den Abbildungen 30 - 32 als Streudiagramme dargestellt, die vollständigen Protein-Listen sind im Anhang zu finden (Tabellen A. 2 - Tabelle A. 13).

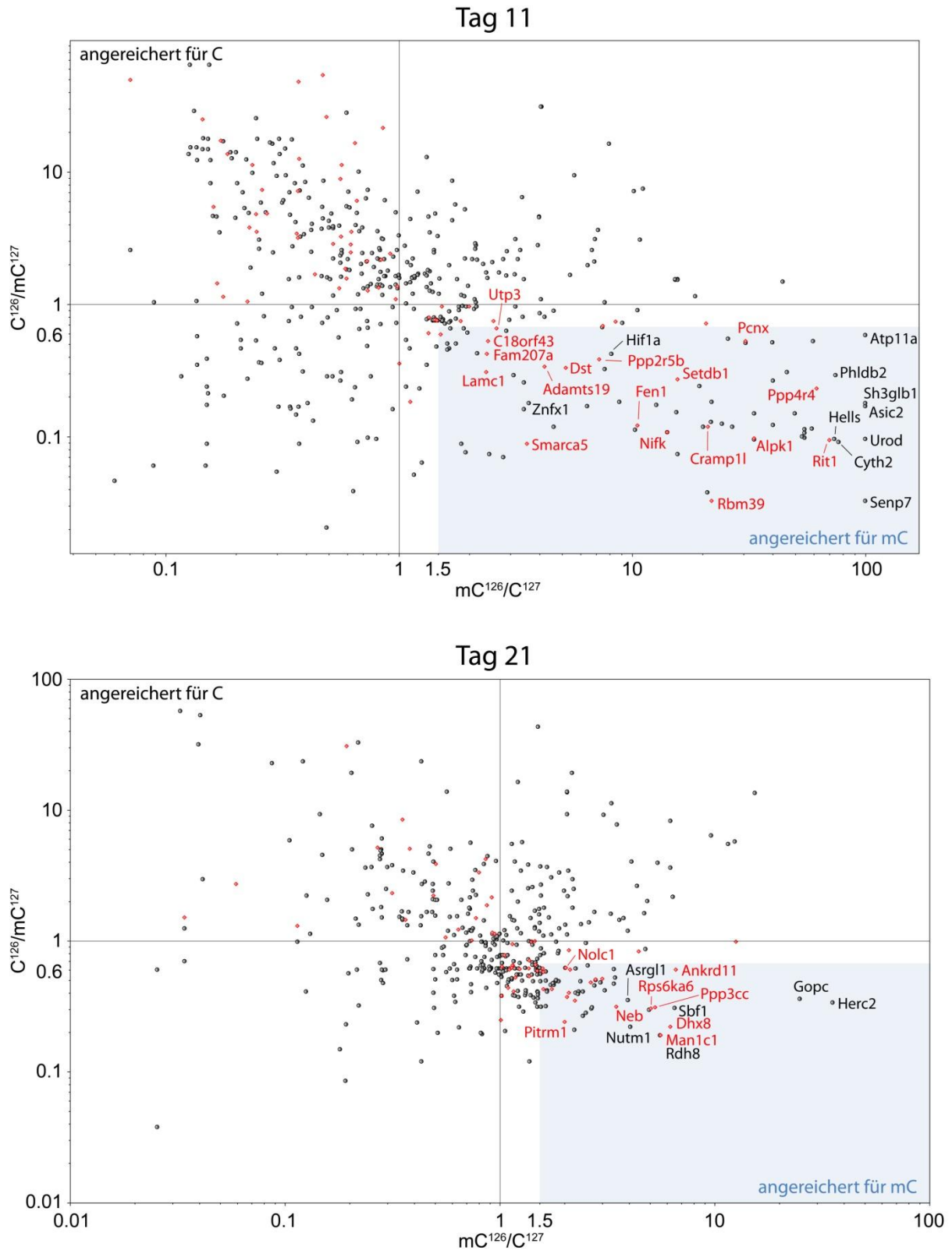


Abbildung 30: Streudiagramme zu den *pull down* Versuchen mit den mC-haltigen DNA-Sonden aus Zellkernlysaten extrahiert aus Retina-Gewebe von 11 (oben) bzw. 21 (unten) Tage alten Mäusen. Die Verhältnisse der TMT-Duplex Reporterionen Intensitäten aus den *forward* und *reverse* Experimenten wurden gegeneinander auf logarithmischen Skalen aufgetragen. Mit dem Crosslinker markierte Proteine wurden jeweils in den Quadranten I (oben links) und III (unten rechts) rot markiert.

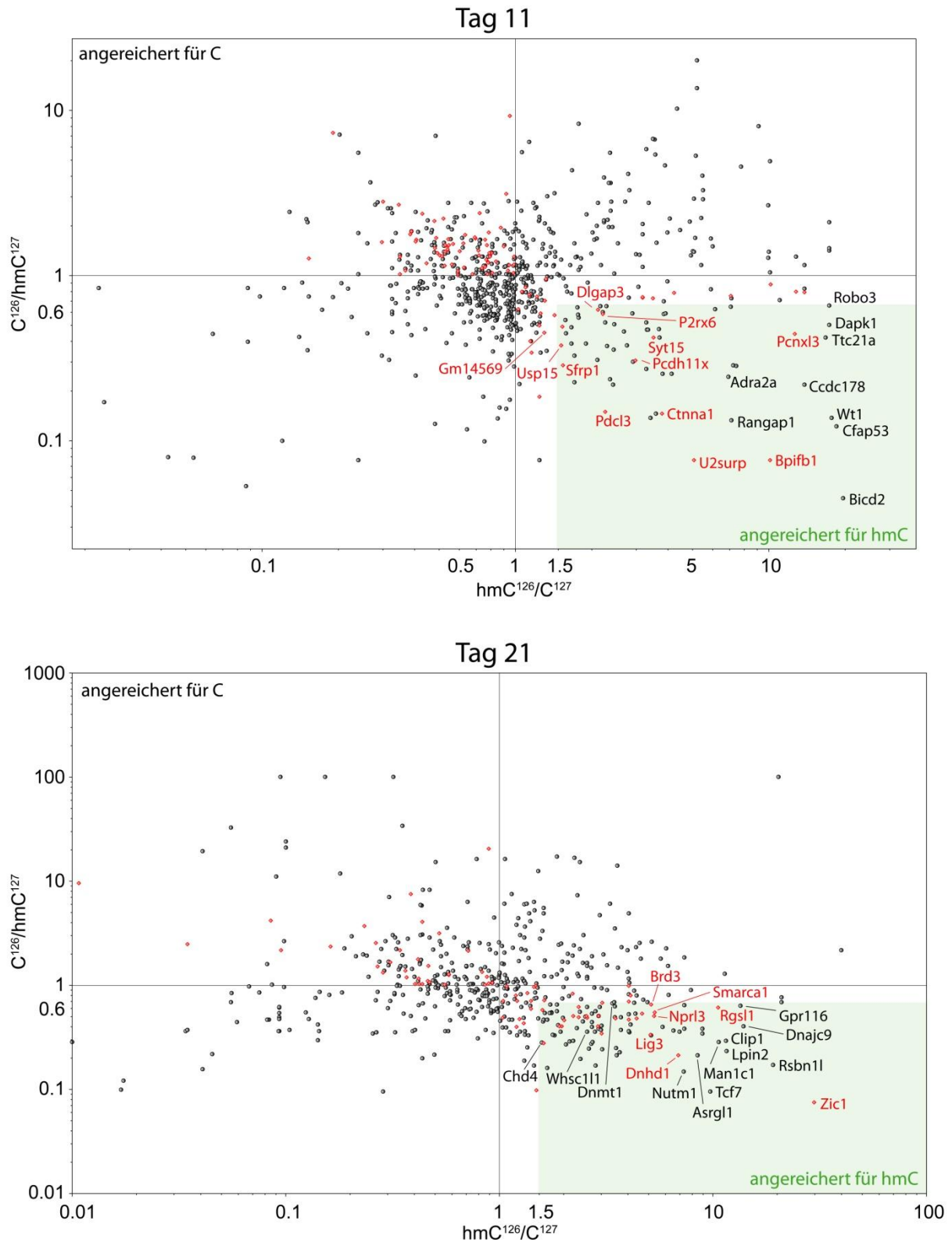


Abbildung 31: Streudiagramme zu den *pull down* Versuchen mit den hmC-haltigen DNA-Sonden aus Zellkernlysaten extrahiert aus Retina-Gewebe von 11 (oben) bzw. 21 (unten) Tage alten Mäusen. Die Verhältnisse der TMT-Duplex Reporterionen Intensitäten aus den *forward* und *reverse* Experimenten gegeneinander auf logarithmischen Skalen aufgetragen. Mit dem Crosslinker markierte Proteine wurden jeweils in den Quadranten I (oben links) und III (unten rechts) rot markiert.

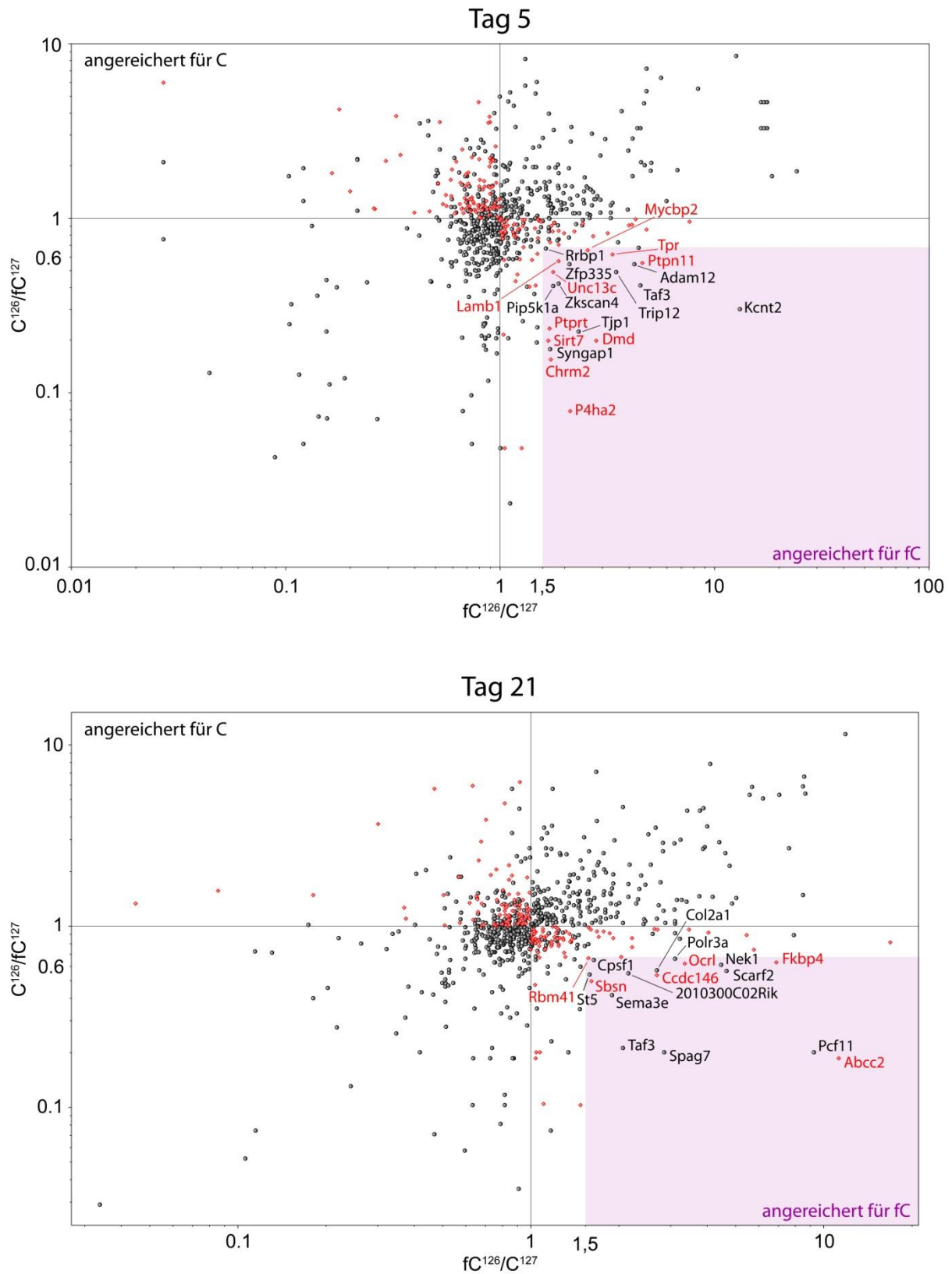


Abbildung 32: Streudiagramme zu den *pull down* Versuchen mit den fC-haltigen DNA-Sonden aus Zellkernlysaten extrahiert aus Retina-Gewebe von 5 (oben) bzw. 21 (unten) Tage alten Mäusen. Die Verhältnisse der TMT-Duplex Reporterionen Intensitäten aus den *forward* und *reverse* Experimenten wurden gegeneinander auf logarithmischen Skalen aufgetragen. Mit dem Crosslinker markierte Proteine wurden jeweils in den Quadranten I (oben links) und III (unten rechts) rot markiert.

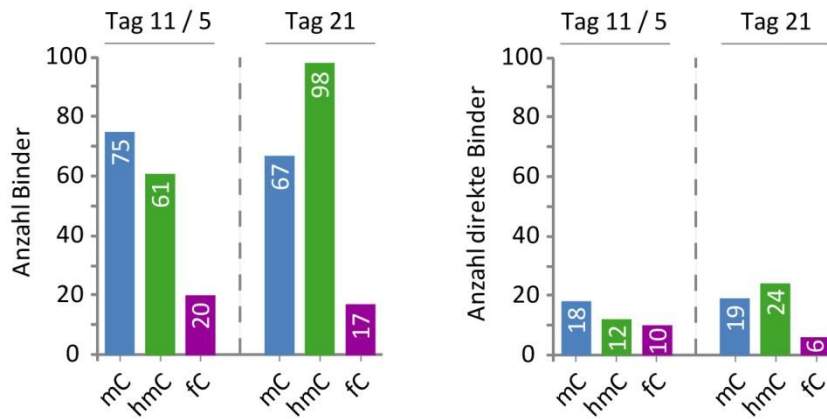


Abbildung 33: Allgemeine Ergebnisse der identifizierten mC (blau), hmC (grün) und fC (lila) Binder (links) und direkten Bindern (rechts) aus den *pull down* Versuchen mit murinen Retina Geweben.

Die Auswertung der *pull down* Versuche aus murinem Retinagewebe führte zu dem Ergebnis, dass die Anzahlen an angereicherten Proteinen für die jeweilige DNA-Modifikation mit ihrem Vorkommen in der endogenen DNA während der Retina-Entwicklung in der Maus korreliert. Während die mC-Level in den Retinageweben fünf bzw. elf und 21 Tage alter Mäuse etwa konstant bleiben, steigen die hmC-Level an und die fC-Level sinken. Damit einher geht eine gesteigerte Genexpression in den neuronalen Zellen.^[2]

Auch in den *pull down* Versuchen wurden für mC in den Retinageweben elf und 21 Tage alter Mäuse (konstante endogene mC Level) in etwa gleich viele Proteine für mC im Vergleich zu C angereichert, dabei wurden keine sich überschneidenden Proteine zwischen den beiden Altersstufen gefunden. Insgesamt konnten 142 spezifische mC-Binder in der murinen Retina identifiziert werden und davon wurden 37 als direkt an die DNA bindende Proteine über eine Anwesenheit des Crosslinkers bestätigt.

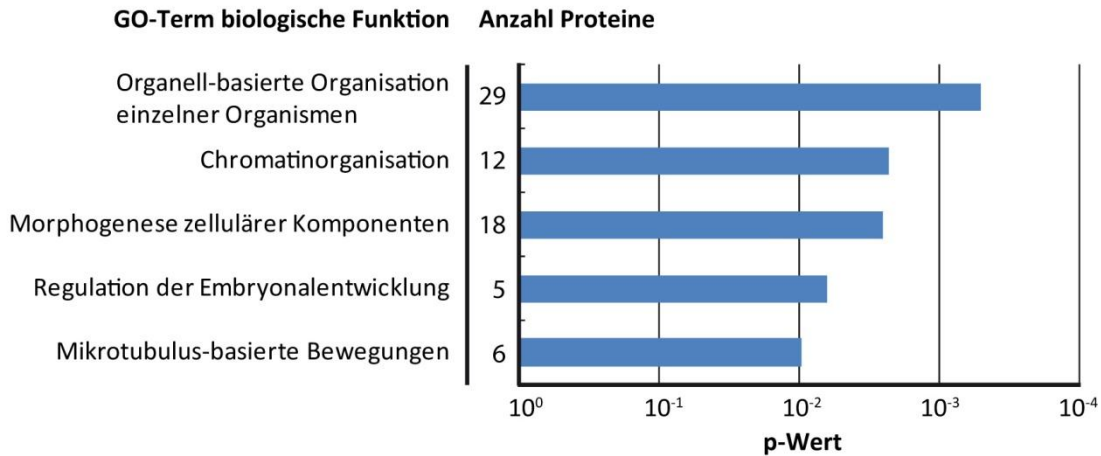
Analog dazu konnten insgesamt 159 spezifische und davon 36 direkte hmC-Binder identifiziert werden. Die Anzahl an hmC-Bindern in den Retinageweben elf und 21 Tage alter Mäuse stieg von 61 auf 98 an. Damit wurden an Tag 21 (höhere endogene hmC Level) 38 % mehr Proteine für hmC angereichert als an Tag 11 (niedrigere endogene hmC Level), dabei wurden nur drei konservierte Proteine gefunden.

Obwohl das mC-Level allgemein in der postnatalen Mäuse-Retina 15-46 fach höher als das hmC-Level ist, war die Gesamtanzahl an hmC-Bindern in etwa gleich der Anzahl an mC-Bindern (159 gegen 142) (Abbildung 33). Vor allem in der Retina elf Tage alter Mäuse wurden mehr mC-Binder (75) als hmC-Binder (61) gefunden. Umgekehrt stieg die Anzahl von hmC-Bindern am Tag 21 (98) deutlich an, während die Anzahl von mC-Bindern auf 67 gefallen war. Insgesamt gab es 24 Proteine, die sowohl

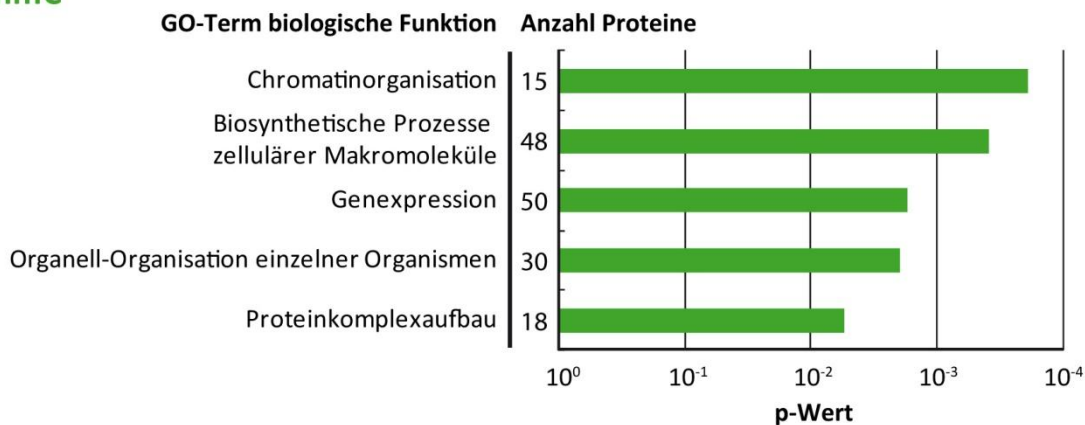
mit den mC- als auch mit den hmC-Sonden angereichert wurden. Zwei davon waren die Nukleosomen modifizierenden Proteine Nap1l1 (an Tag 11) und Smarca1 (an Tag 21).

Für fC wurden an den Tagen fünf und 21 im Allgemeinen viel weniger Proteine angereichert als für mC oder hmC und die Anzahl sinkt geringfügig über den untersuchten Zeitraum. Auch dieses Ergebnis stimmt mit den gemessenen fC-Leveln überein, die im Vergleich zu mC oder hmC um den Faktor 10^3 - 10^4 niedriger sind und im Verlauf weiter abnehmen. Möglicherweise ein Hinweis darauf, dass fC in neuronalen Zellen als kurzlebige Intermediat bei der aktiven DNA-Demethylierung fungiert.

mC



hmC



fc

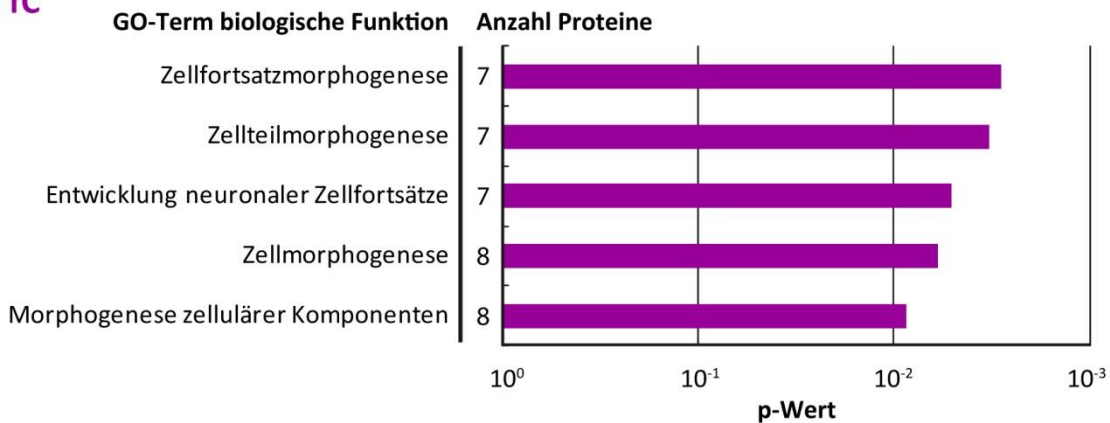


Abbildung 34: Ausschnitte aus den DAVID-GO-Analysen für mC- (blau), hmC- (grün) und fc- (lila) Binder. Für die Analysen wurden jeweils die für die entsprechenden Cytosinderivate angereicherten Proteine mit und ohne Crosslinkermodifikation aus den Geweben beider Tage zusammengefasst. Die Anzahl der Proteine gibt an, wie viele Proteine unter dem jeweiligen GO-Term stehen.

Eine Gene-Ontology Analyse (siehe Abbildung 34) zeigte, dass unter den hmC-Bindern Proteine signifikant angereichert waren, die in biologischen Funktionen, wie der Chromatinorganisation, biosynthetischen Prozessen oder der Genexpression, involviert sind und bestätigten damit den Trend

einer erhöhten Genexpression bei der Oxidation von mC zu hmC. Dahingegen wurden für mC Proteine angereichert, die in der Organellorganisation oder Zellmorphogenese involviert sind. Die meisten der Binder schienen spezifisch für eine bestimmte Modifikation zu sein, allerdings wurden 24 der identifizierten Proteine als Binder sowohl für mC als auch hmC gefunden. Darunter befanden sich die beiden Nukleosomenmodifikatoren Nap1l1 (*Nucleosome assembly protein 1-like 1*) und Smarca1 (*SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1*). Interessanterweise war Nap1l1 Binder von mC und hmC an Tag 11 und Smarca1 Binder von mC bzw. direkter Binder von hmC am Tag 21. Es gab keine Überschneidung der mC-Binder zwischen den beiden Entwicklungsstufen. Allerdings waren drei hmC-Binder konserviert zwischen den zwei Altersstufen. Darunter war der Zelldifferenzierungsregulator Sfrp1 (*Secreted frizzled-related protein 1*).

Insgesamt gibt der Datensatz starke Hinweise darauf, dass hmC eine Rolle in der epigenetischen Regulation der Genexpression spielt. Denn insgesamt 34 der identifizierten hmC-Binder sind Transkriptionsfaktoren oder –regulatoren, wie z.B. die Bromodomain-Proteine Brd2 und 3 (*Bromodomain containing 2 und 3*), der Coaktivator Casp8ap2 (*Caspase 8 associated protein 2*), die Chromatin-Remodeller Chd4, 8 und 9 (*Chromodomain helicase DNA binding protein 4, 8 und 9*), die Cytosin-5 Methyltransferase Dnmt1 (*DNA methyltransferase 1*), der Transkriptionsfaktor Sp2 (*Sp2 transcription factor*), die Topoisomerase Top2a (*Topoisomerase (DNA) II alpha*), die Peptidase Usp15 (*Ubiquitin specific peptidase 15*), die Histon-Methyltransferase Whsc1l1 und die Zinkfingerproteine Zfp668, Zfp869 und Zic1. Zwei Drittel dieser Genexpression regulierenden Proteine wurden an Tag 21 (23) identifiziert und nur 11 an Tag 11. Im Gegensatz dazu wurden nur acht mC-Binder (z.B. Smarca5 oder Zmynd11) an Tag 11 und keiner an Tag 21 gefunden, die als Regulatoren der Transkription/Genexpression klassifiziert sind. Zusätzlich wurde unter den mC-Bindern eine Gruppe von Proteinen (Cyth2, Dst, Fryl, Grin3a, Hif1a, Kndc1, Lamc1, Map1b, Myrf, Pmc1, Postn, Ppp2r5b, Sfrp1, Tead3, Tenm4) gefunden, die in der Neurogenese beteiligt sind.

Es wurde postuliert, dass die Genexpression durch das Enzym Tet3 induziert werden könnte, indem es zusätzlich zur Oxidation von mC zu hmC Transkriptionsfaktoren und Chromatinremodeller wie Nsd3 rekrutiert.^[2] Da die neuronale Isoform des Tet3-Enzyms selbst keine DNA-Bindedomäne besitzt, könnte es durch das Protein Rest rekrutiert werden. Unterstützt wird diese Hypothese durch eine Studie, die zeigt, dass in Abwesenheit von Rest in mESCs die hmC-Level Genom-weit in regulatorischen Regionen zurückgehen.^[202] Es wurden verschiedene Tet-Binder aus dieser Studie auch unter den hmC-Bindern (Nolc1, Cep290, Ctnna2, Chd4, Hist1h2bf und Nsd3 (Whsc1l1)) aus der hier vorliegenden Studie gefunden. Allerdings befanden sich weder Tet3 noch Rest darunter. Eine mögliche Ursache dafür könnte sein, dass die beiden Proteine während des Experiments noch an der

endogenen DNA gebunden sind und deshalb nicht an den DNA-Sonden binden konnten. In weiteren Experimenten könnte man versuchen vor dem *pull down* die endogene DNA aus dem Zellkernlysatz zu entfernen, allerdings würde man dadurch möglicherweise gebundene Proteine mit entfernen.

Mit den wenigen für fC angereicherten Proteinen war eine statistisch signifikante GO-Analyse schwierig, dennoch zeigten hier Proteine am meisten Relevanz, die an der Entwicklung neuronaler Zellfortsätze (Neuriten) beteiligt sind: Serin/Threonin-Proteinkinase Nek1 (*never in mitosis A-related kinase 1*), Histon-Methyltransferase Zfp335 (*zinc finger protein 335*), Semaphorin-3E (Sema3e), Inositol Polyphosphat 5-Phosphatase (Ocr1), Syngap1 (*synaptic Ras GTPase-activating protein 1*), Lamb1 (*laminin subunit beta-1*), Ptpn11 (*tyrosine-protein phosphatase non-receptor type 11*) und die E3 Ubiquitin-Protein Ligase Mycbp2. Somit ist eine mögliche biologische Funktion, die über eine reine Stellung als Intermediat bei der aktiven Demethylierung hinausgeht, nicht ganz auszuschließen. In mESCs wurden bereits fC-spezifische Binder entdeckt, allerdings sind dort die fC-Level auch höher.^[66, 67]

Im Vergleich mit Daten aus vorangegangenen Studien über hmC-Binder in mESCs, NPCs (neuronalen Progenitorzellen) und Gehirngewebe ausgewachsener Mäuse konnten keine hmC-Binder gefunden werden, die konserviert in allen Gewebe- oder Zelltypen auftraten.^[66] Zusammen genommen deuten die Befunde auf eine gewebespezifisch kontrollierte epigenetische Funktion der Cytosinderivate hin.

3.2.3 Studie über 5-Formylcytosin-bindende Proteine während der Entwicklung von murinen Stammzellen zu Epiblast-ähnlichen Zellen

3.4.1 Crosslinking-Interaktionsstudie

Zum Zeitpunkt der Nidation einer Blastozyste findet in Säugetieren eine programmierte *de novo* Methylierung spezifischer DNA-Bereiche, wie den CpG-Inseln, statt.^[203, 204] In den differenzierenden ESCs werden durch den Histon-Code Dnmts zur dauerhaften Stilllegung der Pluripotenzgene wie z.B. Oct3/4 und Nanog rekrutiert. Deshalb kann bei der Entwicklung der Blastozyste hin zum Epiblasten ein Anstieg der mC-Level beobachtet werden. Um die genauen Mechanismen, die zu diesem Prozess führen, zu untersuchen wurden von Angie Kirchner mESCs über zwei Tage zu Epiblast-ähnlichen Zellen (EpiLCs) differenziert. Messungen der Cytosin-Derivat-Level (Abbildung 35) zeigen, dass alle Level über den gesamten Zeitraum ansteigen, so auch die fC-Level. Da die Funktion des fCs bisher noch nicht aufgeklärt wurde, wurden auf Basis dieser Daten drei unterschiedliche Zeitpunkte mit möglichst großen Differenzen des fC-Gehalts ausgewählt, nämlich 0 h (d_0), 24 h (d_1) und 48 h (d_2). Die Zellen wurden zu diesen Zeitpunkten geerntet, daraus Zellkernlysate präpariert und *pull down* Studien mit fC bzw. C und Crosslinker enthaltenden DNA-Sonden durchgeführt. Die angereicherten Proteine an der fC Sonde wurden mit TMT¹²⁷ und an der C Sonde mit TMT¹²⁶ markiert und massenspektrometrisch analysiert. Um einen möglichst robusten Datensatz zu erhalten, wurden die *pull down* Versuche in drei Replikaten durchgeführt. Die Auswertung der MS-Daten und die Identifikation der Proteine erfolgte über die Software Proteome Discoverer 1.4.

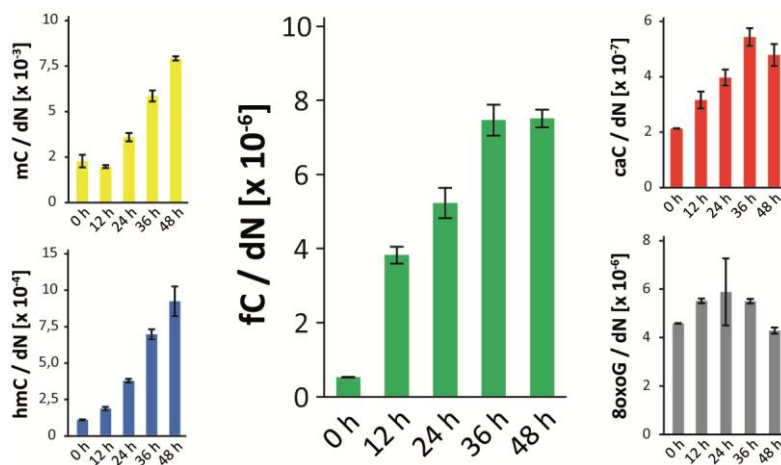


Abbildung 35: Level der DNA-Modifikationen je Nukleotid (dN) während des *Primings* von mESCs zu Epiblast-ähnlichen Zellen (EpiLCs).

Zunächst wurden alle Proteine, die mit TMT quantifiziert werden konnten, mit der Software Perseus 1.5 einem hierarchischem Clustering unterzogen. Das Ergebnis ist in Abbildung A. 1 (Anhang)

dargestellt. Die Darstellung enthält auch Informationen darüber, welche Proteine mit mindestens einem Crosslinker-markiertem Peptid identifiziert wurden. Vor allem an Tag 0 wurden sehr viele Proteine angereichert, darunter auch Solche, die unter physiologischen Bedingungen nicht mit der DNA in Wechselwirkung treten würden, da sie im Zytosol vorkommen. Offenbar konnten bei dieser Zellkernpräparation aus unbekanntem Gründen ein Großteil der zytosolischen Proteine nicht entfernt werden. Um die Datenmenge zu reduzieren, wurden für alle weiteren Analysen nur Proteine betrachtet, die im Nukleus vorkommen.

Demnach wurden zum Zeitpunkt d_0 221 nukleare Proteine für fC angereichert, zum Zeitpunkt d_1 211 Proteine und zum Zeitpunkt d_2 102 Proteine. Der Vergleich der drei Zeitpunkte im Venn-Diagramm (Abbildung 36) zeigt, dass es einige Proteine gibt, die sich überschneiden. 20 Proteine wurden zu allen drei Zeitpunkten angereichert. Darunter befinden sich die Histonproteine Hist1h1t (H1) und H2afv (H2A), 5 ribosomale Proteine (Rpl13 und 18, Rps10 und 27a, Rsl1d1) und die Transkriptionsfaktoren Mtdh (*Metadherin*), (*Heterogeneous nuclear ribonucleoprotein A/B*) und Hnrnpdl (*Heterogeneous nuclear ribonucleoprotein D-like*). Zwischen d_0 und d_1 gibt es zusätzlich 48 Proteine, die sich überschneiden, zwischen d_0 und d_2 sind es 17 Proteine und zwischen d_1 und d_2 37 Proteine. Die 48 in d_0 und d_1 konsekutiv angereicherten Proteine sind hauptsächlich an der Genexpression beteiligt, wobei sich einige Transkriptionsfaktoren, wie Tcerg1 und Gtf2f2, darunter befinden. Die Proteine, die sich in d_1 und d_2 bzw. in d_0 und d_2 überschneiden, sind an sehr unterschiedlichen biologischen Prozessen wie Transkription, DNA Verpackung und auch DNA Reparatur beteiligt. Die Ergebnisse zeigen, dass im zeitlichen Verlauf, trotz steigender fC-Level, immer weniger Proteine an den fC-Sonden angereichert wurden. Im Kontext, der ebenfalls steigenden Methylierung der DNA und der damit einher gehenden sinkenden Genexpression durch Stilllegung spezifischer Gene, scheint dies allerdings kein Widerspruch zu sein.

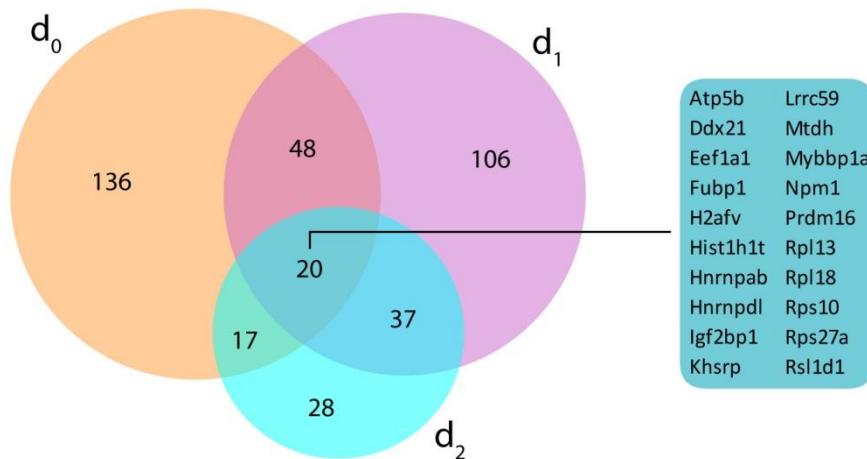


Abbildung 36: Venn-Diagramm zur Darstellung der Anzahl an Proteinen, die im Vergleich zu Cytosin an einem oder mehreren Zeitpunkten der Entwicklung von mESCs zu ELCs spezifisch an fC binden. Aufgeführt sind nur die Proteine, die zu allen drei Zeitpunkten für fC angereichert wurden. Die Listen aller übrigen Segmente des Venn-Diagramms sind im Anhang unter Abbildung A. 2 zu finden.

Zum Zeitpunkt d_0 wurden 136 exklusive Proteine identifiziert. Einer GO-Analyse (siehe Abbildung 37) nach sind über einem Drittel der angereicherten Proteine an der Genexpression und ein Viertel der Proteine an der Transkription beteiligt. Eine STRING-Analyse der 53 unter dem GO-Term „Genexpression“ aufgelisteten Proteine ist in Abbildung 38 dargestellt.^[205] Darin lassen sich die Proteine in etwa zwei Gruppen unterteilen. Unter Gruppe 1 (rote Markierung) befinden sich hauptsächlich Transkriptionsregulatoren (z.B. Phf5a), Spleißfaktoren (z.B. Sf3a2) und Ribonukleoproteine (z.B. Snrpd1) und unter Gruppe 2 (grüne Markierung) befinden sich hauptsächlich Chromatin-regulierende Proteine (z.B. Chd4). Interessanterweise wurden keine Histonproteine exklusiv zu diesem Zeitpunkt angereichert, obwohl sieben Proteine (Smarcc1, Mta2, Crebbp, Cenpv, Kdm5a, Ep400, Chd4) gefunden wurden, die bei der Chromatinmodifikation eine Rolle spielen. Weiterhin wurden auch 5 Proteine gefunden, die an der DNA-Reparatur beteiligt sind (Msh2, Mdc1, PcnA, Smc1a, Smc3).

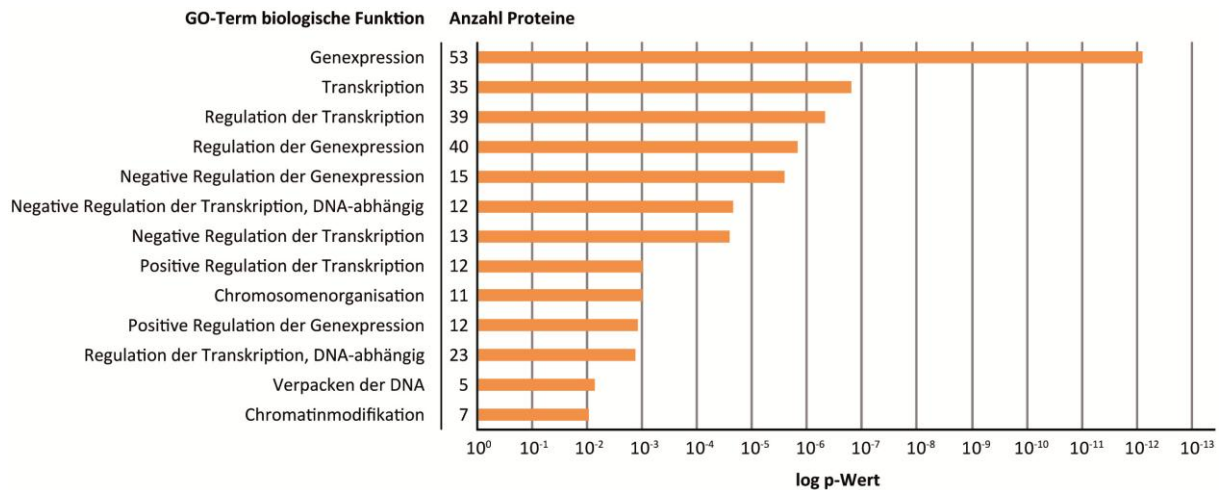


Abbildung 37: DAVID-GO-Term Analyse der 136 ausschließlich an d_0 für fC angereicherten Proteine. Die Balken geben den logarithmischen p-Wert wieder. Zusätzlich wurde die Anzahl an Proteinen angegeben, die unter den jeweiligen GO-Termen stehen.

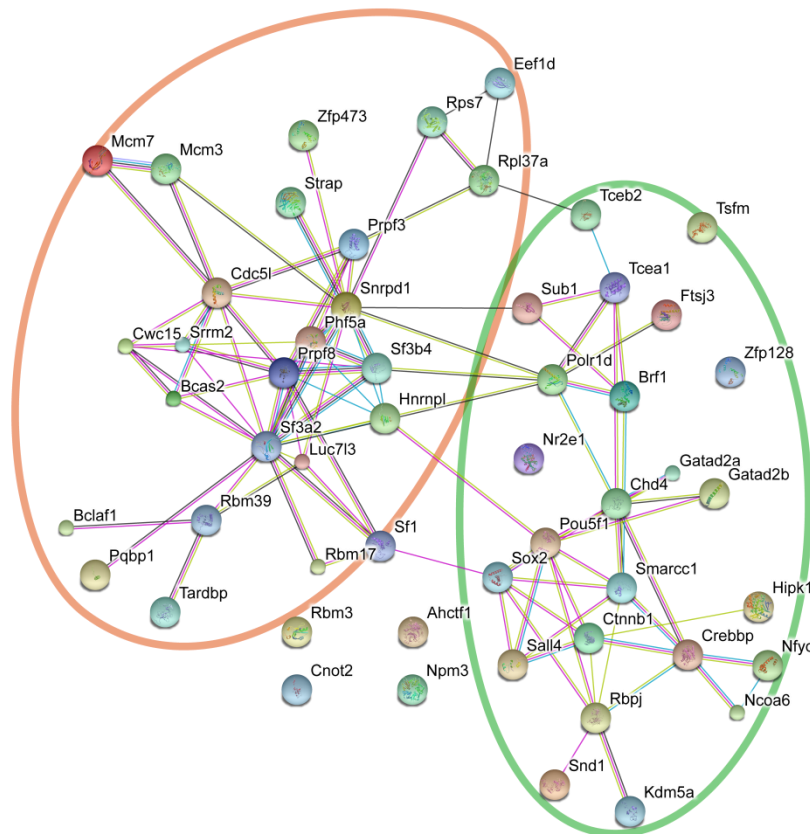


Abbildung 38: STRING-Analyse der Proteine, die unter dem GO-Term „Genexpression“ für den Zeitpunkt d_0 aufgeführt sind. Knotenpunkte repräsentieren Proteine, Verbindungslinien stellen bekannte oder vorhergesagte Protein-Protein Assoziationen dar, die nicht ausschließlich auf physischen Interaktionen beruhen müssen. Dadurch ergeben sich zwei verschiedene Gruppen: Rot eingekreist ist die Gruppe an Proteinen, die v.a. bei der Transkription involviert sind. Grün eingekreist sind Proteine, die v.a. bei der Chromatin-Regulation eine Rolle spielen.

Für den Zeitpunkt d_1 wurden 106 Proteine exklusiv angereichert. Auch zu diesem Zeitpunkt wurden am häufigsten Proteine angereichert, die im Zusammenhang von Prozessen der Genexpression bzw.

Translation gefunden werden. Auffällig ist, dass sich unter den Proteinen, die unter dem GO-Term Translation zusammengefasst sind hauptsächlich ribosomale Proteine, wie Rpl17, Rpl27, Rpl30 gefunden wurden (vgl. Abbildung 39). Zusätzlich wurde eine Gruppe von Proteinen ausgemacht, die unter dem GO-Term „Chromosomenorganisation“ mit 11 Proteinen ein Zehntel der Gesamtzahl angereicherter Proteine ausmacht. Davon sind vier Histonproteine (Hist1h1e, Hist1h1a, Hist1h3a, H2afj). Ebenfalls wurden DNA-Reparatur Proteine (Hmgn1, Nono, Morf4l2, Parp1, Pms1) gefunden.

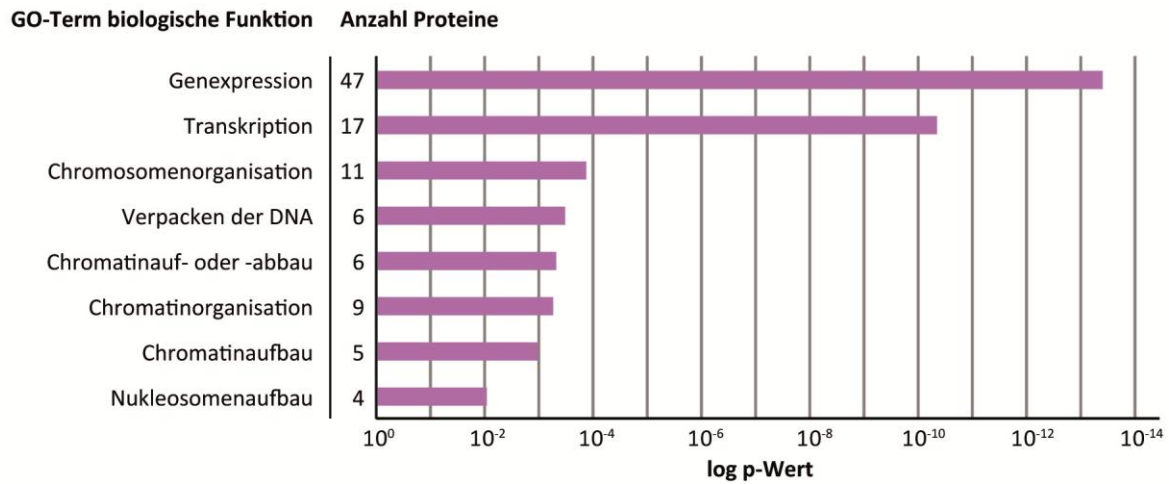


Abbildung 39: DAVID-GO-Term Analyse der 106 ausschließlich an d₁ für fC angereicherten Proteine. Die Balken geben den logarithmischen p-Wert wieder. Zusätzlich wurde die Anzahl an Proteinen angegeben, die unter den jeweiligen GO-Termen stehen.

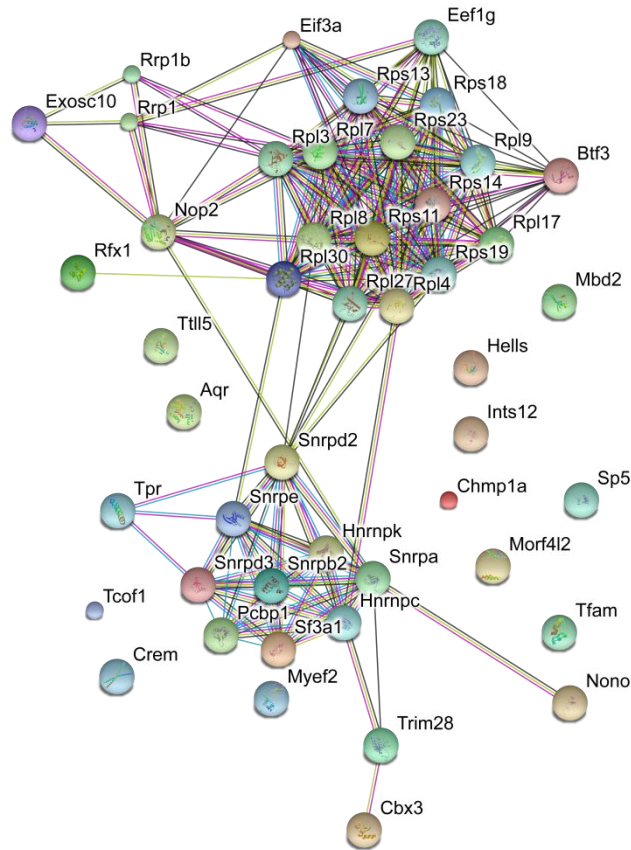


Abbildung 40: STRING-Analyse der Proteine, die unter dem GO-Term „Genexpression“ für den Zeitpunkt d_1 aufgeführt sind. Da es sich überwiegend um ribosomale Proteine handelt, ergibt sich nur eine Aufteilung in die obere Gruppe von Proteinen, die hauptsächlich an der Translation beteiligt sind und in die untere Gruppe von Proteinen, die hauptsächlich am RNA-Splicing beteiligt sind.

Die GO-Analyse mit den 28 exklusiven Proteinen, die zum Zeitpunkt d_2 angereichert wurden, ist statistisch gesehen nicht besonders aussagekräftig (vgl. p-Werte der Abbildung 41). Die höchste Relevanz besitzen die Histonproteine H1f0, Hist2h2ab, Hist1h2bm, H3f3a.

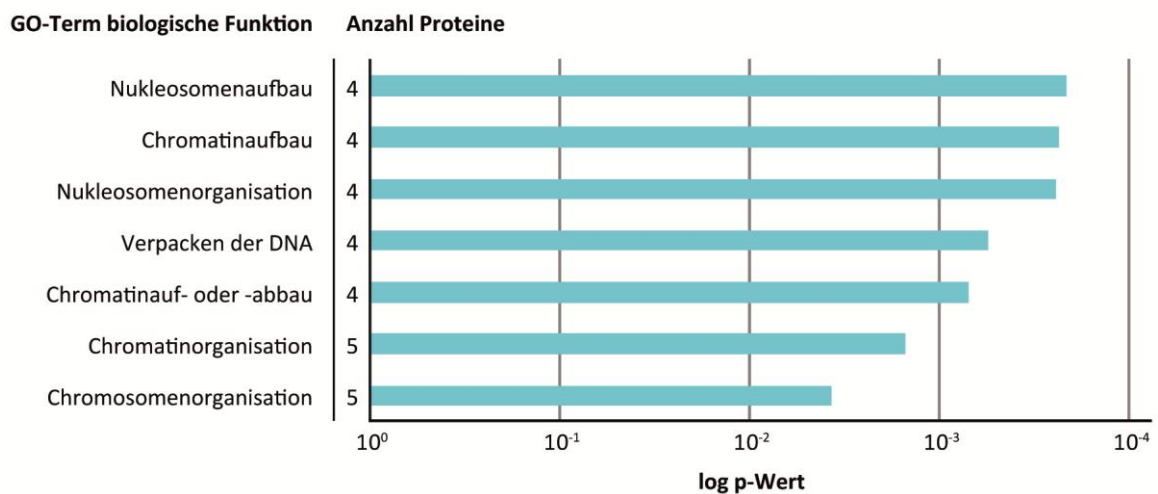


Abbildung 41: DAVID-GO-Term Analyse der 28 ausschließlich an d_2 für FC angereicherten Proteine. Die Balken geben den logarithmischen p-Wert wieder. Aufgrund der geringen Anzahl von Proteinen ist eine statistische Analyse allerdings kaum signifikant, was sich hier auch deutlich in den p-Werten niederschlägt.

In einem hierarchischen Clustering (Abbildung 42) wurden neben den Crosslinkern auch die für C angereicherten Proteine berücksichtigt. Durch das korrelationsbasierte Clustering ergeben sich verschiedene Protein-Cluster und deren hohe Dynamik wird herausgestellt. Durch GO-Analysen der einzelnen Cluster konnten die interessantesten Proteine zusammengefasst werden (siehe Tabellen 11 - 15).

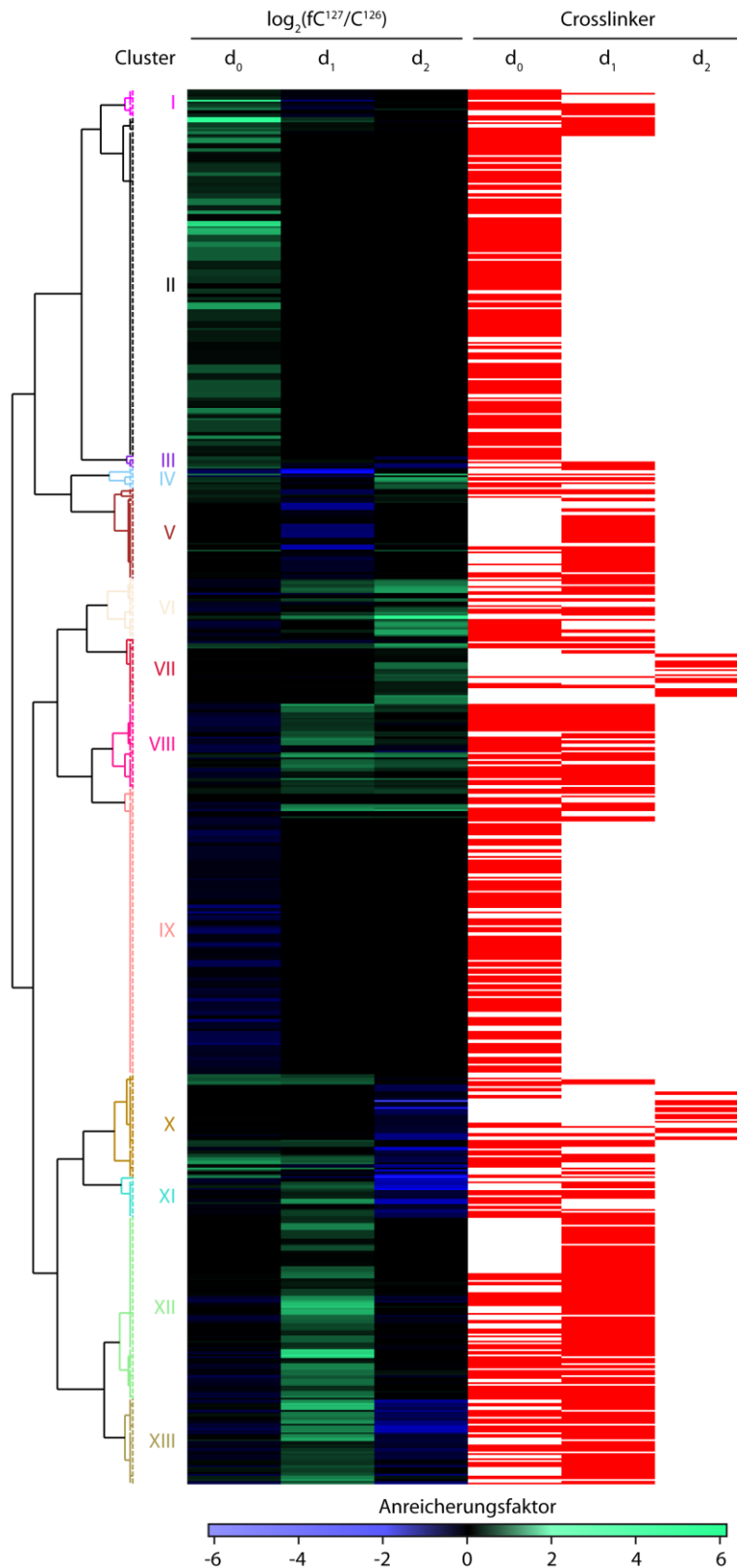


Abbildung 42: Hierarchisches Clustering dargestellt als Heatmap zur Studie fC bindender Proteine während der Entwicklung von mESCs zu mEpiLCs. Für das Clustering wurden nur im Zellkern vorkommende quantifizierte Proteine der Zeitpunkte d_0 , d_1 und d_2 verwendet. Der Anreicherungsfaktor gibt das zur Basis 2 logarithmierte Verhältnis von fC^{TMT127}/C^{TMT126} wieder, d.h. positive Werte (grün) stehen für fC angereicherte Proteine, negative Werte (blau) stehen für C angereicherte Proteine und ein Anreicherungsfaktor von 0 (schwarz) steht für Proteine, die für keine der beiden Sonden angereichert wurden bzw. gar nicht identifiziert wurden. Rot markiert sind Proteine, die mit mindestens einem crosslinker-markiertem Peptid identifiziert wurden. Die Proteine unter den einzelnen Clustern I-XIII sind in Tabelle 10 zu finden.

Tabelle 10: Liste der Proteine aus Abbildung 40 sortiert nach den gekennzeichneten Clustern. Die Spalten werden fortlaufend von oben nach unten gelesen. Die Reihenfolge entspricht der in der Heatmap.

I	II	II	II	V	VI	VIII	IX
Sall4	Fundc2	Prpf8	Anxa11	Acin1	Hmgb1	Fus	Bub3
Smarcc1	Agfg1	Zfp128	Rbm39	Nop16	Syncrip	Ncl	Rrp15
Tubb5	Rps24	Surf6	Pfn1	Srrm2		Rpl10a	Meaf6
Nr2e1	Smc3	Srsf1	Hnrnpd	Nlrc5	VII	Dlst	Poldip3
Rbm3	Dock7	Ddx47	Pqbp1	Pnpla7	Slc3a2	Hsp90b1	Ube2i
Ppp1r10	Ctnnb1	Nfrkb	Recql4	Ep300	Rps26	Rpl5	Pdhh
Scp2	Rexo1	Bdh1	Rpl19	Poldip3	H2afv	Pin4	Ruvbl1
Gatad2a	Cwc15	Brix1	Cct4	Dag1	Rsl1d1	Eef1a1	Rexo4
H1fx	Tcp1	Aifm1	Nup214	Fubp1	Rpl11	Rrp1b	Eef2
Ttn	Cast	Nop16	Npm3	Utf1	Basp1	Gps1	Brd4
Rps7	Kpnb1	Cyc1	Sptbn1	Rcc2	Top1	Atp5b	Tcf20
Pin1	Fbxo38	Cnot2	Tial1	H2afx	Ubtf	Hist1h1t	Smc2
	Zfp473	Cacybp	Tardbp	Mki67	Gar1	Igf2bp1	Sap30
II	Ep400	ldh3g	Hsp90aa1	Csrp2	Hmga1	Krt76	Ran
Srsf7	Brf1	Rpl37a		Hnrnp3	Plcb1	Anln	Ilf2
Srsf3	Ddx42	Rps10	III	Notch2	Ezh1	Nop2	Ubxn6
Nufip2	Mdc1	Ubqln1	Cox4i1	Tacc1	Mier1	Atxn2l	Tet2
Hist2h2aa1	Mrto4	Tsfm	Pgrmc1	Rps8	H3f3a	Npm1	Cbx5
Psip1	Agps	Myh9	Rbbp5	Epb41l2	Hnrnpa1	Ddx21	Aurka
Dnaja3	Msh2	Stip1	Sumo2	Arid4b	Hmgb3	Rpl6	Rsg1
Cdc5l	Dnaja1	Rbm34	Pcna	Hnrnpm	Pds5a	Rps10	Scaf11
Sf3b1	Mta2	Pcnp		Cdk1	Acin1	Khsrp	Dnttip2
Nos1ap	Rtn4	Snd1	IV	Eef1d	Dst		Utp11l
Nifk	Wdr18	Ufm1	Pdha1	Abtb1	Pex14	IX	Rpl10l
H2afy	Pds5b	Yme11	Timm13		Uqcrc2	Rps27a	Ebna1bp2
Crebbp	Wdr5	Supt16	Pabpc1	VI	Pds5b	Hnrnpab	Chchd3
Acin1	Rpa2	Heatr1	Snrnp40	Hnrnpa2b1	Oxr1	Ppan	Sdhb
Hipk1	Ftsj3	Arglu1	Hspa5	Hnrnpf	Krt13	Pbx	Eif6
Gatad2b	Taf15	Mcm7	Hist1h2ab	Snrpg	Mrpl40	Kdm5a	Srsf6
Ncoa6	Srrm2	Csnk2a1	Khdrbs1	Ppib	Hist2h2ab	Hsp90ab1	Krt73
Polr1d	Mcm3	Hnrnp1	Atp5f1	Hnrnpdl	Hist1h2bm	Tcof1	Dhx15
Bclaf1	Rif1	Phf8	Nfu1	Hnrnpu	Ddx17	Hmgb2	Trap1
Rbpj	Pou5f1	Ddx3x		Fbl		Rps3	Snrnp70
Ndufa13	Mrps31	Luc7l3	V	Rpl18	VIII	Sfpq	Tmpo
Snrpd1	Smc1a	Tceb2	Bcas2	Mybbp1a	Hmgn2	Snrpf	Emd
Tcerg1	Myh10	Dnajc8	Rbm17	Lbr	Hist1h1b	Tuba1c	Rps4x
Pabpc4	Ahctf1	Hax1	Rpl23	Sltm	Hist1h1d	Hnrnpa3	Acadm
Myof	Raly	Btaf1	Api5	Krt1	Rpl22	Sf3b2	Sav1
Ube2v2	Gon4l	Ybx3	Nudt21	H1f0	Ints12	Ikzf3	Rcc1
Ik	Cdca8	Psmc4	Nhp2l1	Rpl13	Krt5	Atn1	Dnajb6
Top2a	Rbfox2	Hist1h1t	Sf3b4	Hist1h2bb	Rbmx1l	Rbm14	Ssrp1
Cenpv	Nfyc	Ttn	Chtf8	Actb	Dek	Snrpa1	Srrt
Gmnn	Chd4	Rbm7	Creb3l2	Prdm16	Lyar	Ube2n	Dynll2
Rbpms	Gnl2	Strap	Brd4	Nop56	Atp5a1	Tmpo	Trip12
Hnrnpdl	Csrp1	Sf3a2	Ifitm3	Cct8	Hnrnpa1	Fip11l	Nup98
Phf5a	Nupl1	Ttn	Rpl21	Rpl21	Hist1h3a	Kpna2	Snap23
Stoml2	Nip7	Sf1	Rad50	Prpf6	Hist1h4a	Smarca4	Myef2
Acly	Sub1	Prpf3	Fgfr1	Fubp3	Srfbp1	Zfr	Akap3
Nup35	Dnaja2	Thyn1	Casz1	Nolc1	Hspa8	Ranbp2	Igf2bp3

Fortsetzung Tabelle 10: Liste der Proteine aus Abbildung 40 sortiert nach den gekennzeichneten Clustern. Die Spalten werden fortlaufend von oben nach unten gelesen. Die Reihenfolge entspricht der in der Heatmap.

IX	IX	XI	XII	XIII
Chchd2	Picalm	Crem	Rrp1	Nono
Prdx1		Chmp1a	Trim35	Rps19
Matr3	X	Pcbp1	Rps6	Acaca
Hint2	Hist2h2be	Ybx1	Hspd1	Ssbp1
Rbm28	Mdh2	C1qbp	Rpl4	Ssb
H2afy2	Rpsa	Sf3a1	Rps25	Serbp1
U2af2	Wbp11	Rps18	Rpl17	Ilf3
Thrap3	Rbm27	Hnrnpd	Ubap2l	Rpl9
Pak1	Myef2	Rpl12	Cpsf6	Rps14
Hnrnrm	Suc1g2	Snrpd3	Rps24	Snrpa
Ywhaz	Hnrnpc	Eif5a	Nek6	Ddx39b
Snw1	Ints5	Ppia	Snrpb	Hnrnpa0
Srsf5	Alb	Eef1g	Vdac1	Calr
Hnrnph1	Cstf2t	Rps13	Snrpd2	Vdac3
Dhx9	Snrnp35	Shmt2	Rpl7	Vdac2
Mdn1	Ercc4	Pdia3	Hspa9	Ddx5
Polr2k	Atp5a1	Hcfc1	Sp5	Morf4l2
Ndnl2	Sort1		Pum2	Ewsr1
Smarce1	Rpl13a	XII	Ccar1	Grpel1
Mcm4	Srsf7	Top2b	Krt15	Trim28
Rbbp4	Idh3b	Aspm	Tcerg1	Ddx18
Tomm40	Snrpn	Taf15	Fkbp3	Tfam
Rps27l	Slc25a4	Cirbp	Tpr	Nup62
Atp5j2	Utrn	Crem	Rpl8	Slc25a5
Chtop	Sap130	H2afy	Ttl5	Lmnb1
Tpr	Aco2	Rif1	Parp1	Rpl3
Cfl1	Trrap	Tex10	Nol11	Rplp0
Mcm2	Hspa8	Hells	Shq1	Atp5o
Rrs1	Sarnp	Myef2	Cbx3	Rpl27
Prpf19	Phb2	Mbd2	Snap23	Alyref
Hspb1	Kmt2e	Fubp1	Hmgn1	Cstf2
Nol7	Gtf2e2	Rfx1	Hnrnpk	Raly
Fytd1	Rpl13a	Rbm27	Rps23	Phb
Setd3	Hist1h2bh	Exosc10	Nup50	Lrrc59
Nup153	Vcp	Aqr	Hist1h1e	Npm1
Sf3b6	Pspc1	H2afj	Rpl30	Elavl1
Rpl9	Tcea1	Ttc37	Rpl7a	Rps11
Edf1	Ahnak	Pms1	Rybp	
Gtpbp4	Cggbp1	Erbp2ip	Btf3	
Rbm10	Aldoa	Dnajc8	Dazap1	
Gltscr2	Ptbp1	Eif3a	Snrpc	
Nup54	Gtf2f2	Rps19bp1	Hist1h1a	
Poldip3	Srsf2	Smarcal1	Snrpe	
Ddx46	Pabpn1	Ylpm1	G3bp1	
Cadps2	Pcbp2	Gapdh	Snrpb2	
Dhodh	Cenpe	Cs	Hist1h1c	
Rps9	Sox2	Sap130	Srp14	
Sf3a3	Tpt1	Mtdh	Hmga1	
Mcm5		Htt	Hnrnpc	

In Abbildung 42 ist deutlich erkennbar, dass es Cluster von Proteinen gibt, die nur zum Zeitpunkt d_0 für f_C angereichert (Cluster II, schwarz gekennzeichnet) bzw. abgestoßen (Cluster VIII und IX, hell rosa gekennzeichnet) wurden. Die GO-Term Analyse dieser Cluster zeigt, dass zum Zeitpunkt d_0 hauptsächlich Proteine angereichert wurden, die allgemeine zelluläre Prozesse steuern. Dabei

wurden für fC mehrheitlich Proteine angereichert, die mit der Transkription in Zusammenhang stehen und für C Proteine, die mit der Translation verbunden werden. In den Clustern VIII und IX sind beispielsweise viele Proteine vertreten, die vorwiegend mit Translation, RNA-Prozessierung und Chromosomenorganisation in Verbindung stehen. Sie wurden zum Zeitpunkt d_0 an C angereichert und traten an den anderen Zeitpunkten nicht mehr auf. In Cluster II sind überwiegend Proteine vertreten, die an den fC-Sonden angereichert wurden und im Zusammenhang mit Transkription und mRNA-Prozessierung stehen.

In allen anderen Clustern wurden die Proteine an den verschiedenen Zeitpunkten mehr oder weniger entweder für C oder für fC angereichert. Cluster I weist überwiegend Proteine biosynthetischer Prozesse auf, dabei wurde an d_0 zunächst für fC angereichert, an d_1 ist eine schwache Anreicherung für C erkennbar und an d_2 fand kaum eine Anreicherung statt. Da unter den Clustern III, IV, V und VI jeweils nur wenige Proteine stehen, wurden diese zusammen betrachtet. Darunter befinden sich überwiegend Proteine, die an metabolischen Prozessen, RNA-Prozessierung aber auch Nukleosomenorganisation beteiligt sind. Diese Proteine wurden jeweils an d_0 und d_1 entweder für C, fC oder gar nicht angereichert, an d_2 sind die meisten Proteine für C angereichert. Innerhalb von Cluster VII liegen einige Histon-Proteine, weswegen die GO-Analyse vor allem Prozesse wie DNA-Verpackung und Chromosomenorganisation ergibt. Auffällig ist, dass die meisten Proteine vor allem an d_2 für fC angereichert wurden. In Cluster X wurden an d_0 und d_1 die Proteine überwiegend für fC angereichert und an d_2 für C. Hierunter sind die meisten Proteine für metabolische Prozesse von Stickstoffverbindungen, wie z.B. der Nukleobasen, zuständig. Die letzten drei Cluster XI, XII und XIII wurden wieder zusammengefasst. Hier finden sich vor allem die vielen ribosomalen Proteine wieder, wodurch die GO-Analyse auf eine starke Tendenz zur Genexpression und Translation hinweist. Aber es finden sich auch viele Proteine wieder, die an metabolischen Prozessen beteiligt sind. Die Proteine wurden vor allem an d_1 für fC angereichert, an den Zeitpunkten d_0 und d_2 eher für C. Der Übersichtlichkeit halber ist in den nachfolgenden Tabellen eine Zusammenfassung der interessantesten Proteine gegeben.

Tabelle 11: Identifizierte Histone aus den *pull down* Versuchen zur Studie fC-bindender Proteine bei der Entwicklung von mESCs zu EpiLCs. Angegeben sind die Anreicherungsfaktoren der Replikate (grün: für fC angereichert, blau: für C angereichert) und ob ein Crosslinker-Rest identifiziert wurde (rot) oder nicht (weiß).

Genname	Isoform	d ₀ Log2	d ₁ Log2	d ₂ Log2	d ₀ CL	d ₁ CL	d ₂ CL
H1f0	H1.0	-0,354	-0,006	0,631			
H1fx	H1fx	0,993	-0,165	0,274			
Hist1h1a	H1.1	-0,215	0,862	-0,105			
Hist1h1b	H1.5	-0,247	0,776	0,244			
Hist1h1c	H1.2	-0,206	0,807	0,102			
Hist1h1d	H1.3	-0,106	0,846	0,415			
Hist1h1e	H1.4	-0,247	1,008	-0,073			
Hist1h1t	H1t		0,859	0,618			
Hist1h1t	H1t	0,103					
H2afj	H2afj		0,013				
H2afv	H2afv	0,416	0,429	1,167			
H2afx	H2afx	0,573	-0,006	0,559			
H2afy	H2afy	1,098					
H2afy	H2afy		0,424				
H2afy2	H2afy2	-0,060					
Hist1h2ab	H2A Typ 1	0,437		1,027			
Hist2h2aa1	H2A Typ 2-A	0,473	0,095				
Hist2h2ab	H2A Typ 2-B			0,958			
Hist1h2bb	H2B Typ 1-B		0,364	0,894			
Hist1h2bh	H2B Typ 1-H	0,376	0,394				
Hist1h2bm	H2B Typ 1-M			0,925			
Hist2h2be	H2B Typ 2-E	0,657	0,423				
H3f3a	H3f3a			0,339			
Hist1h3a	H3.1	-0,403	0,423				
Hist1h4a	H4	-0,093	0,649	0,287			

Insgesamt wurden 25 unterschiedliche Histonproteinvarianten identifiziert. Die meisten davon wurden wenigstens an einem Zeitpunkt für fC angereichert. Eine Ausnahme bilden die Histonvarianten der Familie H1 (Linkerhistone) zum Zeitpunkt d₀ (mit Ausnahme des Histons H1fx), die bevorzugt an die unmodifizierte DNA-Sonde binden oder anders gesagt, von fC abgestoßen wurden. Zum Zeitpunkt d₁ zeigen diese allerdings wieder eine deutliche Präferenz für fC-haltige DNA-Sonden. Ebenfalls besonders auffällig ist, dass die Histone zum Zeitpunkt d₂ mit Ausnahme des Histons H3f3 nicht gecrosslinkt wurden. Beides deutet stark darauf hin, dass bei der Differenzierung der Stammzellen zu EpiLCs die Chromatinstruktur einer hohen Dynamik unterliegt und dass fC-haltige DNA eine kontrollierende Funktion der stattfindenden Prozesse übernehmen könnte. Eine erhöhte Affinität der Histone zu fC-haltiger DNA könnte die Transkription negativ beeinflussen, wenn die DNA kompakter an die Histone gebunden wird und somit die Stabilität der Chromatinstruktur gefördert wird. Es wurden auch einige Repressoren gefunden, wie Rybp (d₀ und d₁), Ewsr1 (d₀ und d₁) und Crem (d₁), wobei letzteres Protein auch als Aktivator dienen kann. Weitere Aktivatoren, die ebenfalls gefunden wurden, sind: Crebbp (d₀), Psp1 (d₀ und d₁), Btf3 (d₁) und Ccar1 (d₀ und d₁).

Neben den vielen Histonproteinen wurden auch einige Proteine identifiziert, die Histone modifizieren oder diese erkennen und daran binden (Tabelle 12) sowie einige an der Transkription beteiligte Proteine (Tabelle 13), DNA-Modifikatoren (Tabelle 14) und an der DNA-Reparatur beteiligte Proteine (Tabelle 15). Insgesamt betrachtet wurden diese Proteine vor allem zum Zeitpunkt d0 identifiziert, daher finden sich auch zu diesem Zeitpunkt die meisten Crosslinks. Darüber hinaus gibt es einige Proteine, die explizit an d1 oder d2 für C oder fC angereichert wurden. Dazu gehören Ep300 (beteiligt an der Histon-Acetylierung), Ezh1 (beteiligt an der Methylierung von H3K27), Kmt2e (beteiligt an der Methylierung von H3K4) und Smarcal1 (beteiligt an der DNA-Schadensantwort). Genauso interessant sind Proteine, die an mehreren Zeitpunkten für C oder fC angereichert wurden, wie z.B. Cbx3 (wird durch H3K9 rekrutiert und ist an der epigenetischen Repression beteiligt), Gatad2a (wirkt verstärkend zur Mbd2-vermittelten Repression, effektiv nur zusammen mit Gatad2b, welches auch zum Zeitpunkt d0 für fC angereichert wurde), Nono (beteiligt an der Regulation der Transkription) und Sox2 (Transkriptionsfaktor).

Tabelle 12: Interessante Histon-Modifikatoren aus den *pull down* Versuchen zur Studie fC-bindender Proteine bei der Entwicklung von mESCs zu ELCs. Angegeben sind die Anreicherungsfaktoren der Replikate (grün: für fC angereichert, blau: für C angereichert) und ob ein Crosslinker-Rest identifiziert wurde (rot) oder nicht (weiß). Gelb markierte Funktionen sind mit der Aktivierung von Genen assoziiert, orange markierte Funktionen mit einer reprimierenden Wirkung.

Genname	d ₀ Log2	d ₁ Log2	d ₂ Log2	d ₀ CL	d ₁ CL	d ₂ CL	Funktion
Brd4	-0,137						Erkennung, Bindung und Aufrechterhaltung von acetylierten Histonen
Cbx3	-0,318	1,627					epigenetische Repression, rekrutiert durch H3K9me
Cbx5	-0,157						epigenetische Repression, rekrutiert durch H3K9me (nicht in Anwesenheit von H3Y41ph)
Chd4	0,024						Histon-Deacetylierung (NuRD-Komplex)
Ep300		-0,884					Histon Acetylierung
Ezh1			0,818				Methylierung von H3K27 (PRC/EED-EZH1-Komplex), Repression
Kdm5a							Demethylierung von H3K4me
Kmt2e			-0,979				Mono- und Dimethylierung von H3K4
Mbd2		0,059					Bindet an mC in CpG-Inseln und rekrutiert Histon-Deacetylasen
Meaf6	-0,257						Aktivierung der Transkription als Teil des NuA4-Komplex
Phf8	0,572						Histon-Demethylierung, selektiv für mono- und dimethylierte Lysine
Rbbp4	-0,277						Bindet an Kernhistone und rekrutiert Histon-Remodelierer (Deacetylasen)
Sap130	0,007		-0,266				Reprimierung der Transkription
Sap130		0,158					Reprimierung der Transkription
Sap30	-0,135						Rekrutierung des HDAC-Komplex (Histon-Deacetylierung)
Setd3	-0,546						Methylierung von H3K4 und H3K36
Smarcc1	0,173	-0,114					Chromatin-Remodeling
Smarce1	-0,295						Chromatin-Remodeling
Smc1a	0,370						Aufrechterhaltung der Chromosomenstruktur
Smc3	0,121						Aufrechterhaltung der Chromosomenstruktur
Wdr18	0,696						Methylierung von H3K4
Wdr5	0,715						Methylierung von H3K4

Tabelle 13: Weitere Interessante Proteine aus den *pull down* Versuchen zur Studie fC-bindender Proteine bei der Entwicklung von mESCs zu ELCs. Angegeben sind die Anreicherungsfaktoren der Replikate (grün: für fC angereichert, blau: für C angereichert) und ob ein Crosslinker-Rest identifiziert wurde (rot) oder nicht (weiß).

Genname	d ₀ Log2	d ₁ Log2	d ₂ Log2	d ₀ CL	d ₁ CL	d ₂ CL	Funktion
Cnot2	0,184						Interagiert mit Histon-Deacetylasen, Repressor der Polymerase II Transkription
Gatad2a	0,720	-0,546					Verstärkung der Mbd2-vermittelten Repression, nur zusammen mit Gatad2b
Gatad2b	0,801						Verstärkung der Mbd2-vermittelten Repression, nur zusammen mit Gatad2a
Hipk1	0,803						Inaktivierung von MYB-Transkriptionsfaktor Aktivität
Nono	-0,404	1,091	-1,154				Regulation der Transkription
Phf5a	0,264						Regulation der Transkription
Polr1d	0,071						Transkription
Polr2k	-0,221						Transkription
Pou5f1	0,422						Transkriptionsfaktor
Prdm16	0,051	0,189	0,476				Regulation der Transkription
Snw1	-0,114						Regulation der Transkription
Sox2	0,714		-2,384				Transkriptionsfaktor

Tabelle 14: Interessante DNA-Modifikatoren aus den *pull down* Versuchen zur Studie fC-bindender Proteine bei der Entwicklung von mESCs zu ELCs. Angegeben sind die Anreicherungsfaktoren der Replikate (grün: für fC angereichert, blau: für C angereichert) und ob ein Crosslinker-Rest identifiziert wurde (rot) oder nicht (weiß).

Genname	d ₀ Log2	d ₁ Log2	d ₂ Log2	d ₀ CL	d ₁ CL	d ₂ CL	Funktion
Mcm2	-0,232						Einleitung der DNA-Replikation und -Elongation
Mcm3	0,166						Einleitung der DNA-Replikation und -Elongation
Mcm4	-0,295						Einleitung der DNA-Replikation und -Elongation
Mcm5	-0,176						Einleitung der DNA-Replikation und -Elongation
Mcm7	0,159						Einleitung der DNA-Replikation und -Elongation
Tet2	-0,161						Oxidation von mC zu hmC
Top1			0,108				Topoisomerase
Top2a	0,250						Topoisomerase
Top2b		0,299					Topoisomerase

Tabelle 15: Proteine aus den *pull down* Versuchen zur Studie fC-bindender Proteine bei der Entwicklung von mESCs zu ELCs, die an der DNA-Reparatur beteiligt sind. Angegeben sind die Anreicherungsfaktoren der Replikate (grün: für fC angereichert, blau: für C angereichert) und ob ein Crosslinker-Rest identifiziert wurde (rot) oder nicht (weiß).

Genname	d ₀ Log2	d ₁ Log2	d ₂ Log2	d ₀ CL	d ₁ CL	d ₂ CL
Ercc4			-0,206			
Msh2	0,977					
Parp1	-0,018	0,343	-0,055			
Rad50		-0,089				
Smarcal1		0,779				
Trip12	-0,350					
Mdc1	0,521					

Allerdings lässt sich aus diesen Daten keine klare Präferenz von aktivierenden oder reprimierenden Faktoren für oder gegen fC erkennen. Daraus lässt sich schließen, dass diese Prozesse stark genspezifisch ablaufen und vermutlich von der umgebenden Histonstruktur und/oder der DNA-Sequenz abhängig sind. Hier könnten beispielsweise durch das ChIP-Seq-Verfahren weitere Informationen zur genomischen Lokalisation der bindenden Proteine erhalten werden.^[206]

3.4.2 Vergleich mit publizierten Interaktionsstudien

Da es aus dem Jahr 2013 bereits zwei sehr ähnliche Studien gibt, die sich mit dem Thema einer möglichen biologischen Funktion des fCs beschäftigten, wurden die Daten aus diesen Studien für einen Vergleich herangezogen. Iurlaro *et al.* hat dazu zwei unterschiedliche DNA-Sonden für *pull down* Studien verwendet.^[67] Kurz gesagt, wurden via PCR entsprechende DNA-Oligonukleotide mit den Sequenzen aus den Promotorregionen der Gene Pax6 (280 bp mit 14 CpG-Inseln) oder Fgf15 (248 bp mit 8 CpG-Inseln) erhalten. Diese wurden dann mit Zellkernextrakten muriner Stammzellen (mESCs) inkubiert und Affinitäts-gebundene Proteine wurden mittels LFQ-Methode massenspektrometrisch analysiert. Spruijt *et al.* verwendete synthetisch hergestellte DNA-Oligonukleotide (27 bp mit vier auf einander folgenden CpG-Dinukleotiden) für SILAC-basierte *pull down* Studien mit fC aus mESC-Zellkernextrakten.^[66]

Für den Vergleich zwischen den Daten von Iurlaro *et al.*, Spruijt *et al.* und den eigenen, wurden Letztere zu einer Gesamtliste, aller für fC angereicherten Proteine der verschiedenen Zeitpunkte d_0 , d_1 und d_2 , zusammengefasst. Ebenso wurden die beiden Listen der angereicherten Proteine aus den *pull down* Versuchen mit Pax6- und Fgf15-Oligonukleotiden von Iurlaro *et al.* zusammengefasst. Aus dem Datensatz von Spruijt *et al.* wurden alle sowohl im *forward* als auch im *reverse*-Experiment für fC angereicherten Proteine verwendet.

Obwohl die Bedingungen hinsichtlich Zellstadien, biochemischer Methoden, Oligomerlänge und –sequenz recht unterschiedlich in den drei verschiedenen Versuchsansätzen sind, gibt es entsprechende Übereinstimmungen in den Ergebnissen. Das Venn-Diagramm aus Abbildung 43 zeigt wie viele Proteine der Datensätze sich überschneiden. Es wurden 11 Proteine in allen drei Datensätzen gefunden. Darunter sind die Proteine Cbx3, Parp1 und Wdr18, die wichtige biologische Prozesse steuern. Zusätzlich wurden weitere 37 Proteine gefunden, die sich mit den Daten von Iurlaro *et al.* überschneiden und weitere 49 Proteine überschneiden sich mit den Daten von Spruijt *et al.* Unter den 37 Proteinen befinden sich Chd4, Top1, Top2b, Top2a, Gatad2b, Gatad2a und Wdr5; und unter den 49 Proteinen befinden sich Msh2, Nono, Sap130 und Sox2 (siehe Abbildung A. 3).

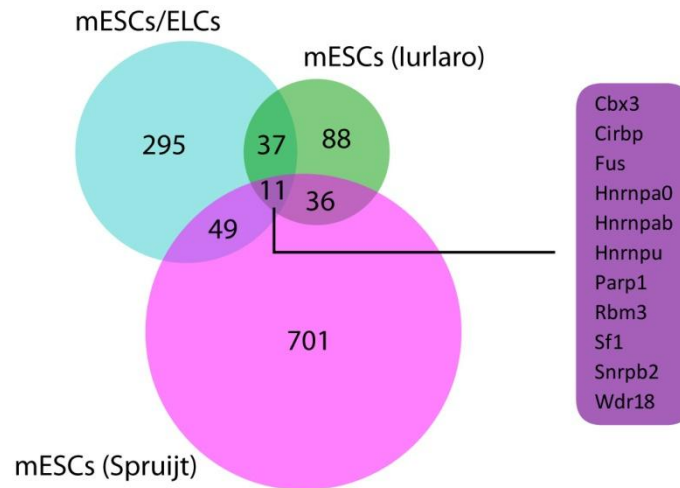


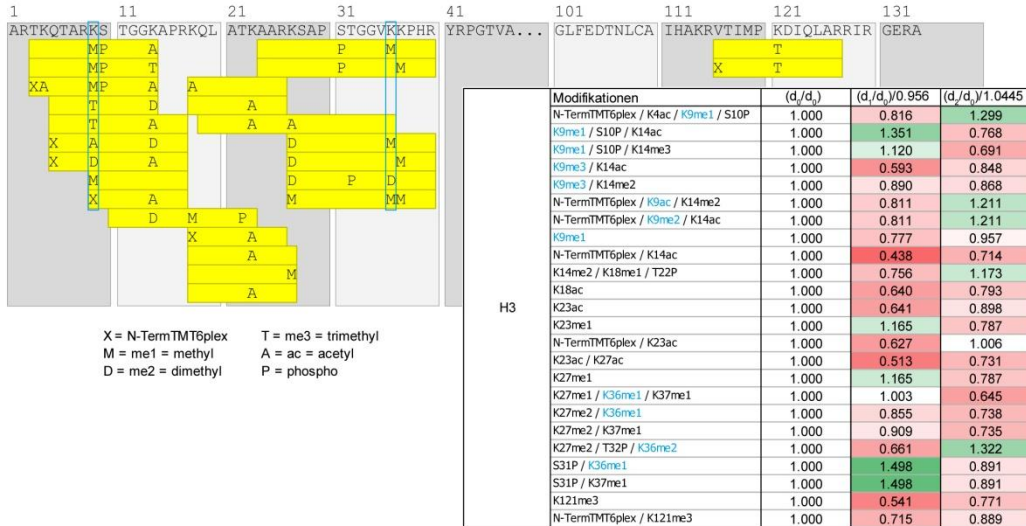
Abbildung 43: Venn-Diagramm zum Vergleich von drei Datensätzen zur Studie fC-bindender Proteine unter verschiedenen Versuchsbedingungen. Blau: alle fC-bindenden Proteine aus den Zeitpunkten d_0 , d_1 und d_2 bei der Studie zur Entwicklung von ELCs. Grün: alle fC-bindenden Proteine, die mit zwei sequenziell unterschiedlichen DNA-Oligomeren aus PCR-Produkten angereichert werden konnten.^[67] Magenta: alle für fC angereicherten Proteine aus der SILAC-basierten *pull down* Studie aus mESC-Zellkernextrakten.^[66] Die Proteinlisten der beiden Segmente „37“ und „49“ sind in Abbildung A. 3 zu finden.

3.4.3 Studie zu Histonmodifikationen

Da die Histone und ihre Modifikationen offensichtlich eine große Rolle bei der Stammzellentwicklung spielen, wurden die Histonproteine aus den Zellkernlysaten der drei Zeitpunkte d_0 , d_1 und d_2 in Triplikaten isoliert, verdaut und mit drei verschiedenen TMT-6plex Reagenzien markiert. Histonpeptide aus der Probe von d_0 wurden mit TMT¹²⁹ markiert, aus der Probe von d_1 mit TMT¹³⁰ und aus Probe d_2 mit TMT¹³¹. Anschließend wurden die Proben gemischt und massenspektrometrisch gemessen. Dadurch konnten die Häufigkeiten der Histon-Modifikationen relativ zum Zeitpunkt d_0 verglichen werden. Insgesamt wurden 17 Histone identifiziert (siehe Tabelle A. 15). Mit dem Programm Proteome Discoverer wurde nach Acetylierungsmustern an Tyrosinen (Y) und Lysin (K) sowie nach Methylierungsmustern an Lysin gesucht, da von diesen bekannt ist, dass sie die Chromatinstruktur und damit die Zugänglichkeit zur DNA beeinflussen können. Desweiteren wurde nach Phosphorylierungen an Serinen (S), Threoninen (T) und Tyrosinen gesucht. Oxidationen an Methioninen (M) und Tryptophanen (W) und selbstverständlich die TMT-Modifikationen an N-Termini und Lysin waren ebenfalls als Standardmodifikationen für die Suche eingestellt. Für die Quantifizierung der Modifikationen wurden in der Auswertung nur solche Peptide berücksichtigt, die eindeutig zu allen drei Zeitpunkten und in allen drei Replikaten durch die Anwesenheit des jeweiligen TMT-Reporters im MS/MS-Spektrum quantifiziert werden konnten. Die Intensitäten der Reporterionen wurden auf das Gesamtverhältnis eines ganzen Proteins mit einer Sequenzabdeckung von 100% (H4) normalisiert und in Relation zum Signal des TMT¹²⁹-Reporters aus der Probe zum Zeitpunkt d_0 gesetzt. So konnten Acetylierungen und Methylierungen für alle fünf bekannten

Histonklassen H1, H2A, H2B, H3 und H4 gefunden und relativ zum Zeitpunkt d_0 quantifiziert werden. Die Auswertung wurde für jedes Histon bzw. jede Histonvariante einzeln durchgeführt (vgl. Abbildung 44 und Tabellen A. 16 - A. 23). Wegen der uneinheitlichen Nomenklatur der Histonvarianten wurden zusätzlich zu den Bezeichnungen die entsprechenden Gennamen in Klammern angegeben, bzw. sind auch in Tabelle A. 15 übersichtlich dargestellt.

Histon H3 (H3f3a)



Histon H4 (Hist1h4a)

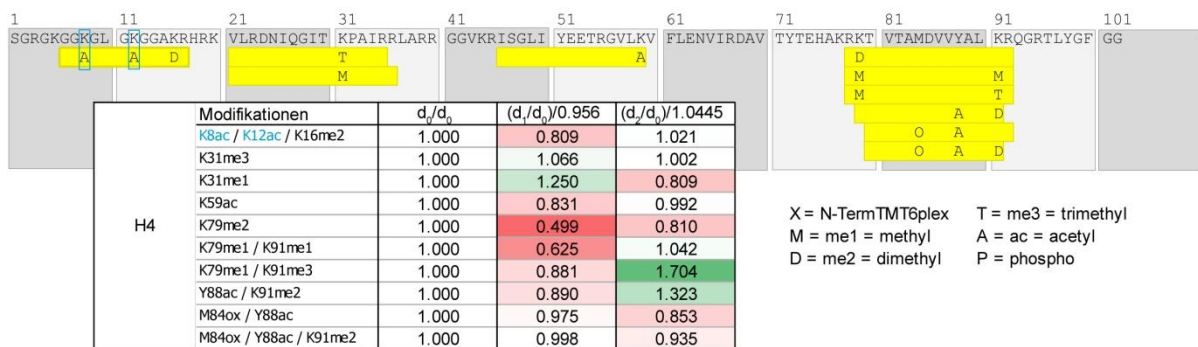


Abbildung 44: Auswertung der Modifikationen der Histone H3 (H3f3a) und H4 (Hist1h4a). Gezeigt sind die vollständigen Aminosäuresequenzen der Proteine. Die gelben Balken zeigen die massenspektrometrisch identifizierten Peptide mit den entsprechenden Modifikationen an. Diese sind im Kasten auch tabellarisch zusammengefasst. Die Werte geben zusätzlich den Faktor an, um den sich das normalisierte TMT-Verhältnis des Peptids von d_1 bzw. d_2 im Vergleich zu d_0 ändert. Ein Anstieg im Level der Modifikationen ist grün gekennzeichnet, ein Sinken ist rot gekennzeichnet. Je intensiver die Farbe, desto höher ist der betragsmäßige Unterschied. Weiß kennzeichnet ein Verhältnis von 1 und bedeutet, dass es keine Veränderung gibt.

Global scheint die Anzahl von acetylierten, methylierten oder phosphorylierten Histonen nach 24 h (d_1) zunächst abzunehmen, nach weiteren 24 h (d_2) nähern sie sich in Summe fast wieder dem Ausgangswert (d_0). Dies stimmt mit früheren Studien zur embryonalen Stammzellentwicklung von Lee *et al.* überein, durch welche sie zu dem Schluss gelangten, dass dies eine Art der Reprogrammierung des epigenetischen Codes darstellen könnte.^[207]

Acetylierungen der Lysine eines Histons führen im Allgemeinen zum Ausgleich seiner positiven Ladung und damit zu einer verminderten Wechselwirkung mit den negativen Ladungen an der DNA. Die Chromatin-Struktur wird dadurch aufgelockert, der Zugang zur DNA erleichtert und somit eine Transkription begünstigt. In Abbildung 45 wurden die Werte aller Acetylierungen der quantifizierten Histone zum Zeitpunkt d_1 bzw. d_2 gemittelt und in Relation zum Mittelwert am Zeitpunkt d_0 aufgetragen.

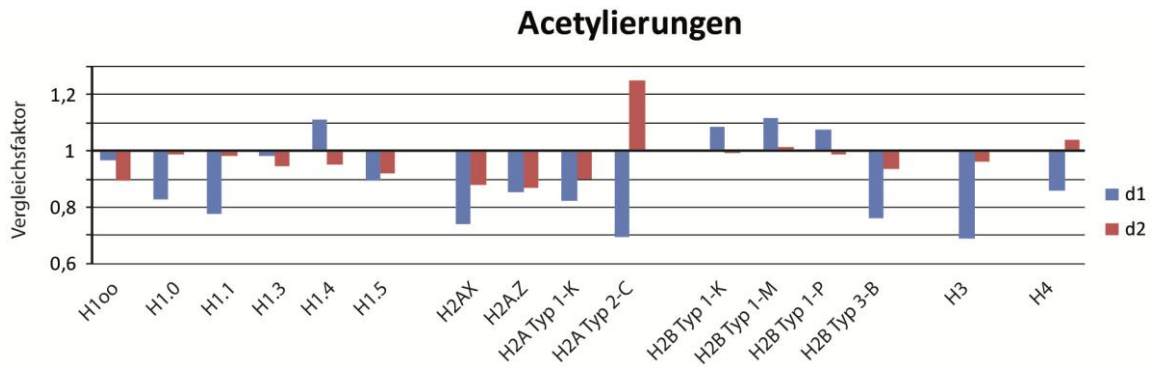


Abbildung 45: Diagramm zum Verlauf der Histon-Acetylierungen zu den Zeitpunkten d_1 und d_2 im Vergleich zu d_0 .

Die meisten Histone weisen demnach unterschiedliche Acetylierungslevel an den verschiedenen Zeitpunkten der Stammzellentwicklung auf. Während zum Zeitpunkt d_1 die Histone tendenziell weniger acetyliert sind als zum Zeitpunkt d_0 , wurden zum Zeitpunkt d_2 ähnlich viele Acetylierungen wie am Ausgangspunkt d_0 gefunden. Am meisten Varianz besitzt das Histon H2A type 2-C (Hist2h2ac). Wesentlich häufiger als Acetylierungen wurden Methylierungen an den isolierten Histonen gefunden. Sie tritt an den Lysinen und Argininen auf, welche mit bis zu zwei (an R) bzw. mit bis zu drei (an K) Methylgruppen modifiziert sein können. Methylierungen an den Histonen können sowohl an der Aktivierung als auch Stilllegung spezifischer Gene beteiligt sein (siehe dazu auch Kapitel 1.1.1). Eine vergleichbare Auswertung der Methylierungen wie schon für die Acetylierungen, lässt eine hohe Dynamik im zeitlichen Verlauf für die unterschiedlichen Histone und die verschiedenen Methylierungsstufen erkennen (siehe Abbildung 46).

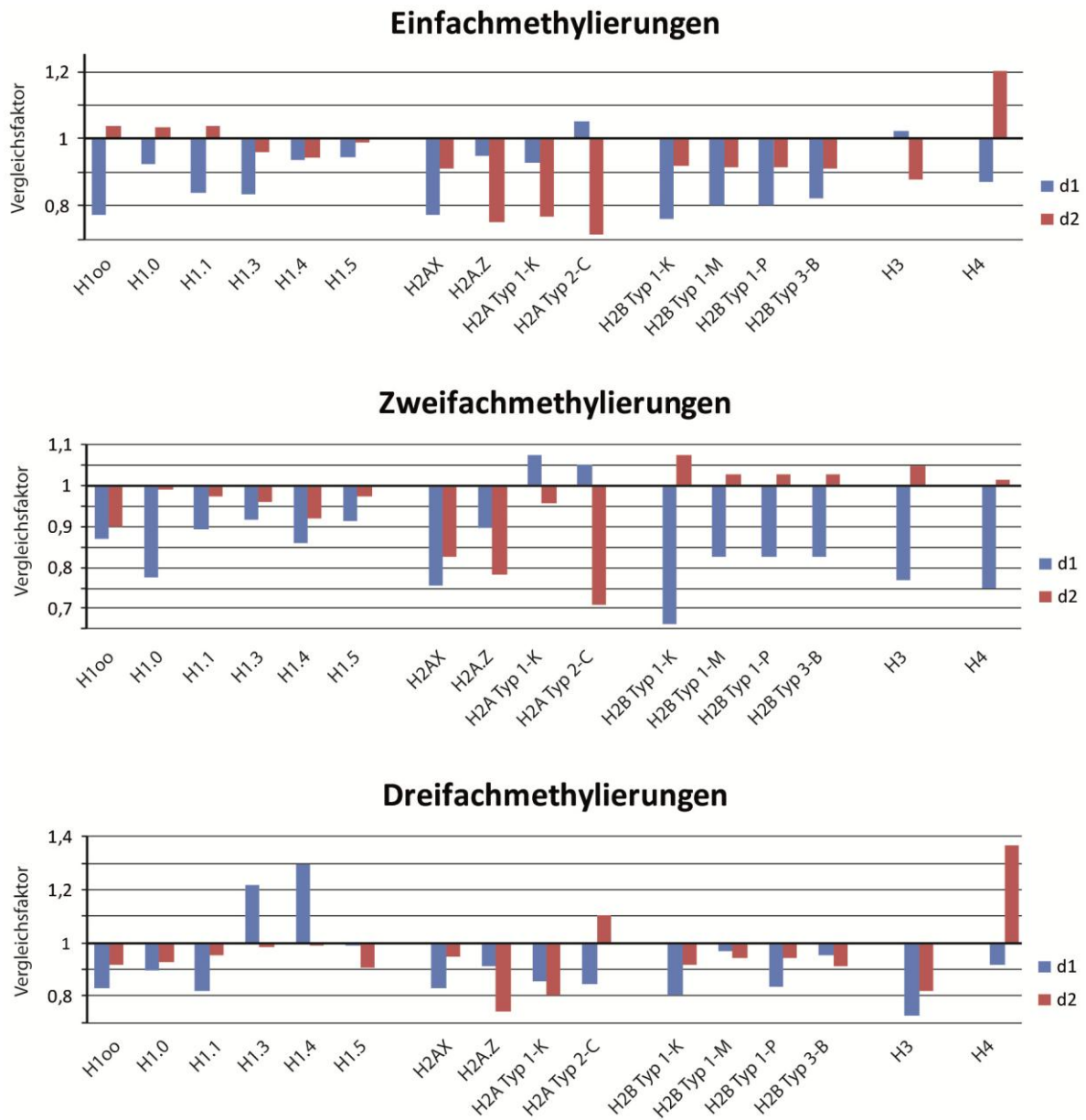


Abbildung 46: Diagramm zum zeitlichen Verlauf von Methylierungen der Histonvarianten.

Eine Analyse der Phosphorylierungen für die Varianten der H1 Proteine (Abbildung 47), zeigt eine hohe Dynamik für H1oo (H1foo), H1.0 (H1f0) und H1.1 (Hist1h1a), welche an d₁ zunächst weniger Phosphorylierungen als an d₀ aufweisen, sich aber an d₂ wieder den Werten von d₀ annähern. Für die Varianten H1.3 (Hist1h1d), H1.4 (Hist1h1e) und H1.5 (Hist1h1b) wurden sowohl nach 24 h als auch nach 48 h im Vergleich zu den Ausgangswerten beinahe gleich bleibend erniedrigte Werte für Phosphorylierungen gefunden.

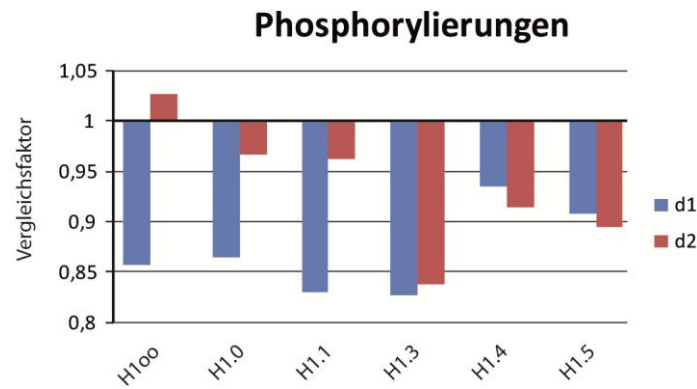


Abbildung 47: Diagramm zum zeitlichen Verlauf von Phosphorylierungen der Histonvarianten von H1 Proteinen.

Histone der Klasse H1 wurden im Vergleich zu den anderen Histonklassen kaum untersucht.^[208] Bis zum heutigen Kenntnisstand spielen die H1 Linker-Histone eine wichtige Rolle bei der Aufrechterhaltung der übergeordneten Chromatinstruktur, können aber auch auf die Regulation der Genexpression Einfluss nehmen.^[209] Obwohl die Biologie der H1 Histone äußerst komplex ist, da es viele verschiedene Isoformen und noch viel mehr posttranslationale Modifikationen gibt, konnte beispielsweise bereits gezeigt werden, dass H1 Histone sowohl die DNA Methylierung als auch die H3 Histon Methylierung beeinflussen können.^[10, 210, 211] Vor allem für die Histone H1.1 (Hist1h1a), H1.5 (Hist1h1b), H1.3 (Hist1h1d) und H1.4 (Hist1h1e) konnte eine Affinität zu DNMTs und eine Silencing-Aktivität nachgewiesen werden.^[210] Diese Histone wurden auch in den *pull down* Studien identifiziert. Interessanterweise wurden die Histone an d_0 schwach von fC abgestoßen, wohingegen sie an d_1 für fC angereichert wurden. An d_2 scheint die Affinität für fC wieder etwas nachzulassen. DNMTs wurden in keiner der *pull down* Proben angereichert. Ob und inwiefern hier ein Zusammenhang besteht und die unterschiedlichen posttranslationalen Modifikationen eine Rolle für die Funktionalität der H1 Histone spielen, müsste noch untersucht werden. Immerhin wurden in der Literatur bereits einige wichtige Zusammenhänge zwischen Chromatin Remodeling und Phosphorylierten H1 Histonen hergestellt. Durch die Anwesenheit von Linker-Histonen werden Chromatin-Remodeling Komplexe, wie der SWI/SNF-Komplex, inhibiert. Durch Phosphorylierung der Histone kann dieser Effekt wieder aufgehoben werden.^[212] In den hier untersuchten Linker-Histonen sinken die Phosphorylierungslevel in den ersten 24 h und steigen danach wieder leicht an. Das würde bedeuten, dass es bei der Differenzierung von Stammzellen anfänglich zu einer Erhöhten Umstrukturierung des Chromatins kommen kann. Interessanterweise wurde das Histon H1.2 (Hist1h1c) nur bei den *pull down* Versuchen gefunden, nicht jedoch bei der Histonfällung. Der Sequenz-Vergleich dieser H1 Varianten beim Menschen ergab eine hohe Homologie.^[208] Möglicherweise fehlte bei der hier vorliegenden Analyse der Histone ein Varianten-spezifisches Peptid des Histons H1.2, wodurch die eindeutige Identifikation nicht möglich war.

Das Histon H1oo (H1foo) besteht aus 304 Aminosäuren und wurde mit einer Sequenzabdeckung von 94% identifiziert. Es spielt eine wichtige Rolle bei der Genexpression während der Oogenese und der frühen Embryogenese, wahrscheinlich durch Destabilisierung der Chromatinstruktur.^[213] Aus der relativen Quantifizierung geht hervor, dass Acetylierungen im Vergleich zum Ausgangszeitpunkt d_0 leicht abnehmen, während Phosphorylierungen und einfache Methylierungen zunächst stärker abnehmen (d_1) und dann wieder leicht ansteigen bis sie um den Ausgangswert liegen (d_2). Zweifach und dreifach Methylierungen scheinen insgesamt weniger zu werden.

Das Histon H1.0 (H1fv) besteht aus 194 Aminosäuren und wurde mit einer Sequenzabdeckung von 100% identifiziert. Es häuft sich vor allem in Zellen an, die in der Endstufe der Differenzierung sind oder niedrige Zellteilungsraten aufweisen. Da H1.0-Knockout-Mäuse sich normal entwickeln und fortpflanzen können, wird bei dem Histon von einer redundanten Funktionalität ausgegangen.^[214] Die relative Quantifizierung der Histon H1.0 Modifikationen zeigt, dass Acetylierungen, zweifach Methylierungen und Phosphorylierungen zum Zeitpunkt d_1 stark absinken und zum Zeitpunkt d_2 wieder annähernd den Ausgangswert erreichen. An den Werten der einfach und dreifach Methylierungen gibt es fast keine Änderungen.

Die Kernhistone H2A, H2B, H3 und H4 stehen mit der DNA in direktem Kontakt und haben deshalb direkten Einfluss auf Mechanismen wie Transkription, DNA-Reparatur, DNA-Replikation und chromosomale Stabilität. Die erhaltenen Daten zeigen, dass ihre Modifikationen betragsmäßig deutlicheren Schwankungen unterliegen als die der Linkerhistone H1.

Acetylierungen der Histon Proteine H2A scheinen im Wesentlichen abzunehmen, mit der Tendenz an d_2 wieder leicht anzusteigen, mit der Ausnahme von H2A Typ 2-C (Hist2h2ac). Hier steigt der Acetylierungsgrad an d_2 sogar über den Ausgangswert bei d_0 . Die Methylierungen folgen keiner klaren Tendenz, sie scheinen viel mehr vom Histon-Typ und vom Grad der Methylierung abhängig zu sein.

Auch die Histonmodifikationen der Familie H2B unterliegen einer klaren Dynamik. Acetylierungen steigen zunächst an, um dann wieder zu sinken. Eine Ausnahme bildet die Histonvariante H2B des Typs 3-B (Hist3h2bb), dessen Acetylierungsgrad zunächst sinkt und dann wieder leicht ansteigt, jedoch nicht über den Ausgangswert von d_0 . Methylierungen nehmen alle tendenziell zunächst eher ab und steigen bei d_2 wieder mehr oder weniger an.

Bei dem Histon H3 (H3f3a) konnten zum Zeitpunkt d_1 weniger Acetylierungen festgestellt werden als zu den Zeitpunkten d_0 und d_2 . Die Einfachmethylierungen bleiben zu den Zeitpunkten d_0 und d_1 fast konstant und gehen zum Zeitpunkt d_2 leicht zurück. Zweifach methylierte Peptide treten an d_1 weniger oft auf als an d_0 , allerdings wurden an d_2 sogar leicht mehr als an d_0 gefunden. Und die

dreifach Methylierungen sinken zunächst an d_1 fast genauso stark ab wie die Acetylierungen, an d_2 werden aber nur geringfügig mehr der dreifach Methylierungen gegenüber d_1 gefunden.

Bei Histon H4 (Hist1h4a) bleiben über den betrachteten Zeitraum die Acetylierungen relativ konstant um den Ausgangswert erhalten. Einfach- und Dreifachmethylierungen sind an d_2 höher als an d_0 und d_1 . Zweifachmethylierungen sinken von d_0 nach d_1 und steigen an d_2 wieder auf das Ausgangsniveau.

Acetylierungen der Histone führen im Allgemeinen zu einer Aktivierung der Transkription. Die hier beobachtete globale Abnahme an Acetylierten Histonen steht im Einklang mit den steigenden mC-Leveln in der DNA und der damit verbundenen Stilllegung spezifischer Gene. Im Gegensatz zu den Acetylierungen können Methylierungen sowohl einen negativen als auch positiven Effekt auf die Transkription haben. Nach dem bisherigen Kenntnisstand (vgl. Kapitel 1.1.1) hängt es von mehreren Faktoren ab, ob Gene durch methylierte Histone aktiviert oder deaktiviert werden. D.h. nicht nur die Position der Methylierung innerhalb der Peptidsequenz der Histonproteine, sondern auch der Grad der Methylierung und die Kombination mit anderen Histon-Modifikationen sowie DNA-Modifikationen, können eine Rolle spielen. Dies spiegelt sich auch in den hier erhaltenen Daten wieder. Gerade bei den Methylierungen konnte keine klare Tendenz einer Zu- oder Abnahme erkannt werden. Wegen der hohen Komplexität, die dieses Muster annehmen kann, ist es schwierig direkte Zusammenhänge mit Vorgängen während der Differenzierung, geschweige denn mit der Genregulation herzustellen. Dennoch könnte der hier vorliegende Datensatz als wertvolle Grundlage für weitere Untersuchungen zu diesem Thema dienen. Man könnte beispielsweise in Expressionsdatenbanken suchen, aus denen hervorgeht, welche Gene ein- bzw. ausgeschaltet werden.

5. Material und Methoden

5.1 Chemikalien

Chemikalien wurden von den Firmen *Sigma/Aldrich/Fluka* (Deisenhofen), *Merck* (Darmstadt), *AppliChem* (Darmstadt), *Jena Bioscience* (Jena) und *Roth* (Karlsruhe) in den üblichen Qualitätsstufen *puriss, p.a., purum* und *molecular biology grade* bezogen, wenn nicht anders angegeben. Für Puffer, Lösungen und Medien wurde bidestilliertes Wasser (*ddH₂O*) verwendet, das durch eine Milli-Q Plus Anlage der Firma *Merck Millipore* (Darmstadt) erhalten wurde. Puffer und Lösungen für LC-MS-Messungen wurden mit H₂O (LC-MS Qualität) der Firma *Fluka* (St. Louis, MO, USA) und Acetonitril (ACN, LC-MS Qualität) der Firma *Carl Roth* (Karlsruhe) hergestellt.

5.2 Puffer und Lösungen

Tabelle 16: Puffer und Lösungen.

Puffer oder Lösung	Komponenten	Konzentration / Menge
ABC-Puffer (pH 8.0)	NH ₄ HCO ₃	50 mM
1 x B&W-Puffer	Tris-HCL (pH 7.5)	5 mM
	EDTA	0.5 mM
	NaCl	1 mM
CLB-Puffer	HEPES	10 mM
	NaCl	10 mM
	KH ₂ PO ₄	1 mM
	NaHCO ₃	5 mM
	EDTA	5 mM
	CaCl ₂	1 mM
	MgCl ₂	0.5 mM
Coomassie-Färbelösung	Coomassie Brilliant Blue R 250	0.25 % (w/v)
	Ethanol	20 % (v/v)
	Essigsäure	10 % (v/v)
DTT-Lösung	DTT	10 mM
	NH ₄ HCO ₃	25 mM
Entfärber-Lösung	Ethanol	20 % (v/v)
	Essigsäure	10 % (v/v)
5% Hydroxylamin-Lösung	HONH ₃ HCl	5 % (w/v)
Hybridisierungspuffer	NaCl	150 mM
	Tris-HCl (pH 7.4)	10 mM
IAA-Lösung	Iodacetamid	1 M
10x Laufpuffer	Tris-HCl	0.25 M
	Glycin	1.29 M
	SDS	1 % (w/v)

Fortsetzung Tabelle 16: Puffer und Lösungen.

Puffer oder Lösung	Komponenten	Konzentration / Menge
Lyse-Puffer A	HEPES (pH 7.9)	10 mM
	MgCl ₂	1.5 mM
	NaCl	10 mM
	Tablette <i>compl. mini protease inhibitor (Roche)</i>	1
Lyse-Puffer C	HEPES (pH 7.9)	20 mM
	Glycerin	20 % (v/v)
	NaCl	0.42 M
	MgCl ₂	1.5 mM
	EDTA	0.2 mM
	Tablette <i>compl. mini protease inhibitor (Roche)</i>	1
PMSF-Lösung	PMSF	200 mM
5 x reduzierender Auftragspuffer	Tris-HCl (pH 6.8)	0.225 M
	Glycerin	50 % (v/v)
	SDS	5 % (w/v)
	Bromphenolblau	0.05 % (w/v)
	DTT	0.25 M
TCEP-Lösung	TCEP	1 M
TEAB-Puffer	TEAB	10 mM
	MgCl ₂	1 mM
TSE-Puffer	Tris-HCl (pH 7.5)	10 mM
	Saccharose	300 mM
	EDTA	1 mM

5.3 Geräte, Software und Online-Tools für Bioinformatik

Tabelle 17: Geräte und Software.

Gerät	Software	Firma
Eppendorf Zentrifugen 5810R und 5424R		Eppendorf (Hamburg)
Eppendorf Thermomixer Comfort		Eppendorf (Hamburg)
Eppendorf BioPhotometer Plus		Eppendorf (Hamburg)
NanoDrop® ND-1000“-Spektrophotometer	NanoDrop 1000 3.8.1	PEQLAB Biotechnologie (Erlangen)
Mini-PROTEAN® Tetra“ Elektrophoresekammer		Bio-Rad (München)
Ultimate 3000 HPLC	Dionex Chromatography MS Link 2.14 Build 3818 (basierend auf Chromeleon 6.80 SR13 Build 3818) und Thermo Xcalibur 3.0.63	Dionex/Thermo Fisher Scientific (Waltham, MA, USA)

Fortsetzung Tabelle 17: Geräte und Software.

Gerät	Software	Firma
Q Exactive HF Spektrometer	Thermo Xcalibur 3.0.63	Thermo Fisher Scientific (Waltham, MA, USA)
LTQ Orbitrap XL Spektrometer	Thermo Xcalibur 3.0.63	Thermo Fisher Scientific (Waltham, MA, USA)
	Proteome Discoverer 1.4.1.14	Thermo Fisher Scientific (Waltham, MA, USA)
	Perseus 1.5.0.9	Max-Planck-Institut für Biochemie (Martinsried)

GO-Analysen wurden mit den DAVID Bioinformatik-Datenbanken (<https://david.ncifcrf.gov/>) und STRING-Analysen wurden mithilfe der STRING Bioinformatik-Datenbank (<http://string-db.org/>) durchgeführt.^[205, 215-217]

5.4 Oligodesoxynukleotide

Die für diese Arbeit verwendeten Oligodesoxynukleotide wurden entweder selbst synthetisiert, freundlicherweise von Barbara Steigenberger synthetisiert und zur Verfügung gestellt oder durch die Firmen *Metabion GmbH* (Martinsried) bzw. *Ella Biotech GmbH* (Martinsried) in gereinigter HPLC-Qualität und lyophilisiert bezogen. Tabelle 18 gibt einen Überblick über die verwendeten Oligodesoxynukleotide und deren Bezugsquellen.

Tabelle 18: Verwendete Oligodesoxynukleotide und deren Bezugsquellen. Sequenzen sind in 5'- nach 3'-Ende angegeben. A=Adenosin, C=Cytosin, G=Guanin, T=Tyrosin, caC=5-Carboxycytosin, X=5-Ioduracil (IU), Y=5-Octadiinyluracil, mC=5-Methylcytosin, hmC=5-Hydroxymethylcytosin, fC=5-Formylcytosin, hmU=5-Hydroxymethyluracil, Flu=Fluorescein, Bio=Biotin. Bezugsquellen: a) Selbst synthetisiert, b) synthetisiert und bereitgestellt durch Barbara Steigenberger, c) *Metabion GmbH*, d) *Ella Biotech GmbH*.

Oligodesoxy-nukleotid-Nr.	Sequenzen der Oligodesoxynukleotide (5' → 3')	Bezugsquellen
S1	TCC GAA GGA AcaCG ATG ACC GGA TGC	a)
S2	Flu- GCA TCC GGT CAX CGT XCC TTC GGA -Bio	c)
S3	TCC GAA GGA ACG ATG ACC GGA TGC	a), c)
S4	GCA TCC GGT CAX CGT XCC TTC GGA -Bio	c)
S5	Bio- GCA TCC GGT CAY CGT TCC TTC GGA	b)
S6	TCC GAA GGA AmCG ATG ACC GGA TGC	c)
S7	TCC GAA GGA AhmCG ATG ACC GGA TGC	c)
S8	TCC GAA GGA AfCG ATG ACC GGA TGC	a)
S9	Bio- GCA TCC GGT CAY CAT TCC TTC GGA	b)
S10	TCC GAA GGA ATG ATG ACC GGA TGC	c)

Fortsetzung Tabelle 18: Verwendete Oligodesoxynukleotide und deren Bezugsquellen. Sequenzen sind in 5'- nach 3'-Ende angegeben. A=Adenosin, C=Cytosin, G=Guanin, T=Tyrosin, caC=5-Carboxycytosin, X=5-Ioduracil (IU), Y=5-Octadiinyluracil, mC=5-Methylcytosin, hmC=5-Hydroxymethylcytosin, fC=5-Formylcytosin, hmU=5-Hydroxymethyluracil, Flu=Fluorescein, Bio=Biotin. Bezugsquellen: a) Selbst synthetisiert, b) synthetisiert und bereitgestellt durch Barbara Steigenberger, c) *Metabion GmbH*, d) *Ella Biotech GmbH*.

Oligodesoxy-nukleotid-Nr.	Sequenzen der Oligodesoxynukleotide (5' → 3')	Bezugsquellen
S11	TCC GAA GGA AhmUG ATG ACC GGA TGC	d)
S12	Bio- GCT CAC GCT AGY CAA CTC CGT GCA	b)
S13	TGC ACG GAG TTG ACT AGC GTG AGC	c)
S14	TGC ACG GAG ThmUG ACT AGC GTG AGC	d)

5.5 Oligodesoxynukleotidsynthese

Die Festphasen-DNA-Synthese erfolgte nach den Standardprotokollen des Geräteherstellers mittels CE-Phosphoramidit-Chemie an einem *ABI 394 DNA synthesizer* (*Applied Biosystems*, Foster City, CA, USA). Als Festphasenmaterial wurde ein CPG-Träger von (500 Å, *Glen Research*, Sterling, VA, USA) verwendet. Die Standardphosphoramidite von dA (Bz-dA), dC (Bz-dC oder Ac-dC), dG (*i*Bu-dG) und dT wurden von *Link Technologies* (Bellshill, NL, GB) und trockenes Acetonitril (*Rotidry*, ≥ 99.9%, < 10 ppm Wasser) wurde von *Roth* bezogen. Die Sonderphosphoramidite fdC und cadC wurden freundlicherweise von Felix Gnerlich und Barbara Steigenberger zur Verfügung gestellt. Die für Kupplung, *Deblocking*, *Capping* und Oxidation verwendeten Reagenzien und Lösungen wurden verwendet, wie im Folgenden aufgelistet:

- **Kupplung:** 0.25 M BTT in MeCN (≤ 10 ppm H₂O), hergestellt nach Welz *et al.*^[218]
- **Deblocking:** 3% Dichloressigsäure in DCM
- **Capping A:** Ac₂O (20% Vol.), 2,6-Lutidin (30% Vol.) in MeCN (≤ 10 ppm H₂O)
- **Capping B:** *N*-Methylimidazol (16% Vol.) in MeCN (≤ 10 ppm H₂O)
- **Oxidation:** 25 mM Iod in 2,6-Lutidin/H₂O/MeCN (6:65:30)

Standardphosphoramidite und Sonderphosphoramidite wurden als 0.1 M Lösungen in MeCN (≤ 10 ppm H₂O) eingesetzt. Während Standardphosphoramidite für 45 s gekuppelt wurden, wurden die Kupplungszeiten bei Sonderphosphoramiditen auf 165 s verlängert. Bei den 24 bp langen fdC-Oligonukleotiden wurde eine „DMT-off“ Synthese (Abspaltung der 5'-DMT-Schutzgruppe am Träger) und für cadC-Oligonukleotide wurde eine „DMT-on“ Synthese (keine Abspaltung der 5'-DMT-Schutzgruppe am Träger) durchgeführt. Die synthetisierten Stränge wurden mit 0.4 M NaOH in MeOH/H₂O (4/1 v/v) für 18 h bei RT vom Träger abgespalten. Die Lösung wurde filtriert, und der

Träger mit 600 μL 1 M Triethylammoniumacetat-Lösung gewaschen und ebenfalls filtriert. Die vereinigten Filtrate wurden in einer *SpeedVac* auf ca. 30 % des Ausgangsvolumens reduziert und im Anschluss analysiert. *DMT-on* Ansätze wurden für die Abspaltung der terminalen DMT-Gruppe bis zur Trockene eingedampft und in 100 μL 80 % Essigsäure aufgenommen. Der Ansatz wurde für 20 min bei RT geschüttelt (1200 rpm) und die DNA durch Zugabe von 60 μL 3 M NaOAc sowie 1600 μL EtOH gefällt (mindestens 60 min bei $-20\text{ }^{\circ}\text{C}$). Die DNA wurde anschließend durch Zentrifugation vom Überstand abgetrennt.

Die Identität der synthetisierten Oligodesoxynukleotide wurde via MALDI-TOF-MS überprüft. Analyse und Aufreinigung der Stränge erfolgte mittels RP-HPLC. Verwendet wurde hierfür ein Waters HPLC-System (Waters Alliance 2695 mit PDA 2996, Präparative HPLC: 1525EF mit 2482 UV Detektor, Waters, Eschborn) und VP 250/10 Nucleosil 100-7 C 18 Säulen des Herstellers *Macherey Nagel* (Düren). Dabei wurde ein Gradient aus 0.1 M Triethylammoniumacetat jeweils in Wasser und 80 % MeCN eingesetzt. Die präparativen Trennungen wurden auf folgenden RP-Säulen durchgeführt: CC 250/10 Nucleosil 100-7 C 18 oder CC 250/10 Nucleodur 100-7 C 18 (*Macherey-Nagel*, Düren). Für die analytischen HPLC Trennungen wurden die Säulen CC 250/4 Nucleosil 4 120-3 C 18 und CC 250/4 Nucleodur 4 120-3 C 18 eingesetzt.

Die Konzentration der synthetisierten Oligodesoxynukleotide wurde mittels UV/Vis-Spektroskopie bei 260 nm (Nanodrop ND-1000 von *PreQlab*) bestimmt und mit den Extinktionsfaktoren der Basen korrigiert (dA: 15 L/(mmol-cm), dC: 7 L/(mmol-cm), dG: 12 L/(mmol-cm) und dT: 8 L/(mmol-cm), fdC: 11 L/(mmol-cm), cadC: 7 L/(mmol-cm)). Nach erfolgter präparativer Reinigung und analytischer Charakterisierung wurde das Wasser an einer *SpeedVac* oder an einer Lyophylle entfernt und mit $d\text{dH}_2\text{O}$ auf eine Konzentration von 100 pmol/ μL eingestellt. Die Oligodesoxynukleotide wurden bei $-20\text{ }^{\circ}\text{C}$ gelagert.

5.7 Biochemische Methoden und Assays

5.7.1 DNA-Sonden Herstellung

Zur Durchführung der *pull down* Assays wurden zunächst doppelsträngige DNA-Sonden hergestellt. Hierfür wurden äquimolare Mengen der unter Tabelle 19 aufgeführten DNA-Stränge zusammengegeben, mit Puffer auf eine Endkonzentration von 25 pmol/ μL gebracht und in einem Heizblock für 2 min auf $85\text{ }^{\circ}\text{C}$ erhitzt und über 4 h auf Raumtemperatur abgekühlt.

Ioduracil enthaltende DNA-Sonden wurden in Hybridisierungspuffer hybridisiert und bei jedem Schritt darauf geachtet, dass möglichst unter Lichtausschluss gearbeitet wurde. Nach dem Abkühlen auf Raumtemperatur wurden diese DNA-Sonden bei $-20\text{ }^{\circ}\text{C}$ bis zur weiteren Verwendung gelagert.

Die **Alkin** (5-Octadiinyluracil) enthaltenden DNA-Sonden wurden in 0.25 x PBS (*Sigma-Aldrich*, St. Louis, MO, USA) hybridisiert und nach dem Abkühlen auf Raumtemperatur lyophilisiert, um im nächsten Schritt via Click-Reaktion den Crosslinker einzuführen. Hierzu wurden zu 10 nmol der doppelsträngigen DNA 0,5 µL einer 200 mM Lösung des Azid-PEG₃-S-S-NHS-Esters (C₁₈H₂₉N₅O₈S₂, *Jena Bioscience*, Jena, Deutschland oder freundlicher Weise von Kristof Hufnagel synthetisiert und zur Verfügung gestellt) in DMSO gegeben. In einem separaten Eppendorfggefäß wurde so viel CuBr in einer TBTA Lösung (DMSO/tBuOH 3:1, 100 mM) gelöst, um ein 1:1 Verhältnis von Cu(I):TBTA zu erhalten. Diese Lösung wurde sofort zu der DNA/Azid Mischung gegeben, mit 45 µL DMSO/tBuOH (3:1) weiter verdünnt und bei 37 °C für 3 h geschüttelt (950 rpm). Das resultierende Produkt wurde nach einer Ethanol-Fällung erhalten. Um die Hydrolyse des NHS-Esters zu verhindern, wurden die so erhaltenen DNA-Sonden über Nacht lyophilisiert und anschließend für *pull down* Studien verwendet oder bei -80 °C bis zur weiteren Verwendung gelagert.

Tabelle 19: Hybridisierungsschemata zu den verwendeten DNA-Sonden aller durchgeführten *pull down* Studien.

Versuch	enthaltende Modifikation	Strang-Nr.	Gegenstrang-Nr.	entsprechende Kontrollbase	Kontrollstrang-Nr.	Kontrollgegenstrang-Nr.
Kapitel 3.1.1	mC	S6	S2	C	S3	S2
	hmC	S7				
	fC	S8				
	caC	S1				
Kapitel 3.1.3	caC	S1	S4	C	S3	S4
Kapitel 3.2.1	hmU:A	S11	S9	T:A	S10	S9
	hmU:G		S5	C:G	S3	S5
	hmC:G	S7	S12	T:A (alt.)	S13	S12
	hmU:A (alt.)	S14				
Kapitel 3.2.2	mC	S6	S5	C	S3	S5
	hmC	S7				
	fC	S8				
Kapitel 3.4.1	fC	S8	S5	C	S3	S5

5.7.2 Zell- und Gewebeproben

Murine embryonale Stammzellen (mESCs) wurden freundlicherweise zunächst von *Matthias Truss* (Charité, Berlin) und später von Fabio Spada kultiviert und zur Verfügung gestellt. Die Kultur und Entwicklung der mESCs zu **epiblast-ähnlichen Zellen (epiblast-like cells, EpiLCs)** wurde freundlicherweise von Angie Kirchner durchgeführt.

Human Embryonic Kidney 293T (HEK-293T) Zellen wurden in *Dulbecco's Modified Eagle Medium* (DMEM, *Gibco, LifeTechnologies*, Carlsbad, CA, USA) mit 10% fötalem Kälberserum (FBS, *Gibco, LifeTechnologies*) und 1x *Antibiotic-Antimycotic* (100 U/mL Penicillin, 100 µg/mL Streptomycin und 25 µg/mL Fungizon, *Gibco, LifeTechnologies*) in T75 Zellkulturflaschen (*Cellstar, Greiner*, Frickenhausen) in einem wassergesättigtem 5% CO₂/95% Luftgemisch bei 37 °C kultiviert. Bei Konfluenz der Zellen wurde das Medium abgenommen, die Zellen mit PBS Puffer gewaschen und

nach Zugabe von 2 mL Trypsin-Lösung (0.25% Trypsin, EDTA, Phenolrot, *Gibco, LifeTechnologies*) für 5 min bei 37 °C inkubiert. Danach wurden 13 mL Medium zugegeben und die Zellen resuspendiert. 1 mL der Zellsuspension wurde dann in eine neue Zellkulturflasche mit 19 mL frischem Medium überführt (1:15 Splitt), die restlichen 14 mL Zellsuspension wurden in einem 15 mL Falcon pelletiert, mit PBS gewaschen und standen anschließend für die Zellkernextraktion zur Verfügung.

Die **Gewebeproben der Retinae** von Mäusen wurden freundlicherweise von Arshan Perera (LMU, München) zur Verfügung gestellt.

5.7.3 Zellkernlysate Präparation

Für die Präparation von **Nuklearextrakten aus mESCs und HEK-293T-Zellen** wurde nach einem angepassten Protokoll von Dignam *et al.* verfahren.^[219] Kurz gefasst, wurden die Zellen nach der Ernte mit PBS gewaschen und in fünffachem Pelletvolumen Lyse-Puffer A resuspendiert. Die Zellen wurden 10 min auf Eis inkubiert und danach bei 4 °C und 2000 rpm 10 min zentrifugiert. Das Zellpellet wurde in seinem zweifachem Volumen Lyse-Puffer A resuspendiert und die Zellen durch 20 Stöße eines *WheatonDounce*-Homogenisators lysiert. Die Zellkerne wurden durch 10 minütiges Zentrifugieren bei 4 °C und 2000 rpm pelletiert und der Überstand vorsichtig entfernt. Es wurden das zweieinhalbfache Volumen Lyse-Puffer C zu dem Pellet gegeben, durch auf und ab Pipettieren homogenisiert und 30 min auf Eis inkubiert. Nach Zentrifugation bei 4 °C und 14500 rpm für 20 min wurde der Überstand abgenommen, aliquotiert und bis zur weiteren Verwendung bei -80 °C gelagert.

Die Zelllyse und Präparation der **Zellkernlysate aus murinen Retinae** wurde wie in Guillemin *et al.* beschrieben durchgeführt.^[220] Es wurden jeweils 1-2 Retinae lysiert. Dazu wurden 100 µL CLB-Puffer gegeben und mit 4 Stößen eines *WheatonDounce*-Homogenisators homogenisiert. Nach 10 min Inkubation auf Eis wurden noch einmal 6 Stöße ausgeführt. Es wurden 10 µL 2.5 M Saccharose zugegeben und 10 min bei 4 °C und 6300 x g zentrifugiert. Das Pellet wurde in 100 µL TSE-Puffer resuspendiert und mit 30 Stößen homogenisiert. Danach wurde das Zellkernlysate aliquotiert und bei -80 °C bis zur weiteren Verwendung gelagert.

5.7.4 Bradford Assay

Proteinkonzentrationen der Zellkernlysate wurden mit Hilfe des Bradford-Assays bestimmt. Dafür wurden 1-2 µL der Proteinprobe mit *ddH₂O* auf ein Endvolumen von 50 µL verdünnt und mit 950 µL einer 1:5 verdünnten Lösung des Bradford-Reagenz (*Bio Rad, Hercules, CA, USA*) versetzt. Dieses Reagenz enthält den Farbstoff Coomassie-Brilliantblau-G250, welcher mit basischen und aromatischen Seitenketten der Aminosäuren interagiert und dadurch sein Absorptionsmaximum von 465 nm auf 595 nm wechselt. Nach einer Inkubationszeit von 5 min wurden die Proben an einem

vorkalibrierten BioPhotometer gemessen, womit der Proteingehalt in μg erhalten wurde. Kalibriert wurde das Photometer mit Hilfe von BSA-Standardlösungen bekannter Konzentrationen.

5.7.5 SDS-PAGE mit Fluorescein-markierten Proben

Die diskontinuierliche SDS-Polyacrylamid-Gelelektrophorese nach Lämmli wurde zur Analyse der *pull down* Experimente mit Fluorescein-markierten DNA-Sonden verwendet. Hierfür wurden 80 μL der *pull down* Probe entnommen und zunächst mit einem Zentrifugalfilter (MWCO = 10 kDa, Merck Millipore, Darmstadt) auf 20 μL eingengt. Die Probe wurde dann zusammen mit reduzierendem Auftragspuffer 5 min auf 95 °C erhitzt und auf das Gel aufgetragen. Die Zusammensetzung der gegossenen SDS-Gele ist aus Tabelle 20 ersichtlich. Desweiteren wurde als Proteingrößenstandard *SeeBlue™ Plus2 pre-stained Protein Standard* (Invitrogen, Oslo, Norwegen) auf das Gel aufgetragen. Die Elektrophorese erfolgte dann bei konstanten 130 V für 80 min im 1x Laufpuffer.

Fluorescein-markierte Proteinbanden wurden dann unter einem *LAS3000-Imager* (Fujifilm, Tokio, Japan) Geldokumentationssystem ($\lambda_{\text{Anregung}} = 460 \text{ nm}$, $\lambda_{\text{Emission}} = 515 \text{ nm}$) sichtbar gemacht und fotografiert. Zusätzlich wurde das Gel mit Coomassie-Färbelösung gefärbt und überschüssiger Farbstoff wurde durch Waschen mit Entfärber-Lösung entfernt.

Tabelle 20: Zusammensetzung von Trenngel und Sammelgel der SDS-PAGE.

Komponente	Trenngel 12%	Sammelgel 5%
ddH ₂ O	1.32 mL	850 μL
Rotiphorese® Gel 30	1.6 mL	207.5 μL
Tris-HCl pH 8.8 (1.5 M)	1 mL	-
Tris-HCl pH 6.8 (0.5 M)	-	157.5 μL
10% (w/v) SDS	40 μL	12.5 μL
10% (w/v) APS	40 μL	12.5 μL
TEMED	4 μL	1.25 μL

5.7.6 Pull down Assays

Zu 1 nm der jeweiligen **IU enthaltenden DNA-Sonden** wurden 220 μg Zellkernlysate gegeben und mit PBS auf ein Gesamtvolumen von 500 μL gebracht. Je nach Experiment wurde 15-30 min auf Eis oder bei Raumtemperatur im Dunkeln inkubiert. Danach wurden die Proben in geöffneten Reaktionsgefäßen auf Eis 30 min unter einer UV-Licht-Lampe mit 312 nm bestrahlt. Nachdem 80 μL der Probe für die Analyse mittels SDS-PAGE entnommen wurde, wurde die restliche Probe über Nacht bei -20 °C gelagert. Um die an der biotinylierten DNA gebundenen Proteine anzureichern, wurden mit Streptavidin beschichtete, magnetische Partikel verwendet. 1 mg *Dynabeads®* (Invitrogen, Oslo, Norwegen) wurden nach Herstellerangaben in 1x *B&W* Puffer gewaschen und in

500 μL 2x B&W Puffer resuspendiert, bevor die 420 μL des Belichtungsansatzes zugegeben und 2-3 h bei leichtem Rotieren inkubiert wurden. Die magnetischen Partikel wurden dann mithilfe eines Magneten fixiert, der Überstand wurde abgenommen und verworfen, und es wurde drei Mal mit 1x B&W Puffer gewaschen. Eluiert wurde mit 20 μL 10% (w/v) SDS und 30 μL ddH₂O bei 95 °C für 5 min. Die Proteine wurden mit Aceton gefällt, 10 min bei 13000 rpm zentrifugiert und in 100 μL DTT-Lösung via Ultraschall resuspendiert. Nach 30 min Inkubation bei 56 °C wurden 100 μL IAA-Lösung zugegeben und weitere 30 min bei Raumtemperatur inkubiert. Nach Zugabe von 1 μL β -Mercaptoethanol wurde über Nacht bei -20 °C in Aceton gefällt, 10 min bei 13000 rpm zentrifugiert und das Pellet in 200 μL Natriumhydrogencarbonat-Lösung für den anschließenden tryptischen Verdau resuspendiert.

Das Protokoll für den Protein *pull down* wurde für Experimente mit dem neuen Crosslinker angepasst, da das Belichten der Proben durch den reaktiven NHS-Ester hinfällig war. Es wurden 1-2 nmol der entsprechenden lyophilisierten **DNA-Sonde mit NHS-Ester-Crosslinker** in 10-20 μL trockenem DMSO gelöst und zu 80-125 μg Zellkernlysate je nmol DNA-Sonde gegeben. Zellkernlysate wurden zuvor mit TEAB-Puffer auf eine Konzentration von etwa 0.5 $\mu\text{g}/\mu\text{L}$ verdünnt. Nach 20 min Inkubation bei Raumtemperatur wurden 100-200 μg nach Herstellerangaben vorbereitete und in TEAB-Puffer resuspendierte *Streptavidin Magnetic Particles* (Roche, Mannheim) zugegeben und weitere 3 h unter leichtem Rotieren inkubiert. Die Partikel wurden magnetisch fixiert, drei Mal mit 500 μL TEAB-Puffer gewaschen und in 100 μL TEAB-Puffer resuspendiert. Reduziert wurde mit 5 μL TCEP-Lösung für 60 min bei 60 °C und 750 rpm im Thermomixer (Eppendorf) und alkyliert wurde mit 10 μL Iodacetamid (IAA)-Lösung für 40 min bei Raumtemperatur im Dunkeln. Durch Zugabe von 10 μL 10% (w/v) SDS und Erhitzen auf 95 °C für 5 min wurde eluiert. Die Partikel wurden mit 20 μL ddH₂O nachgespült, dann wurden die Proteine mit Aceton bei -80 °C eine Stunde gefällt. Das Präzipitat wurde bei 14500 rpm abzentrifugiert, der Überstand verworfen und das Pellet in 50-80 μL TEAB-Puffer für den anschließenden Verdau resuspendiert.

5.7.7 Saure Extraktion von Histonen

Histone wurden aus den Zellkernextrakten von mESCs bzw. EpiLCs nach Shechter *et al.* isoliert.^[221] Es wurden je Zeitpunkt ($d_0=0$ h, $d_1=24$ h und $d_2=48$ h) 100 μg Zellkernextrakt in 200 μL Lyse-Puffer C für die saure Präzipitation verwendet. Es wurden 200 μL einer 0.8 N H₂SO₄ Lösung zugegeben, gemischt und für 30 min bei 750 rpm inkubiert. Danach wurde für 30 min bei 13200 rpm zentrifugiert, um Zellkernfragmente zu entfernen. Der Überstand wurde in ein neues Gefäß überführt und 132 μL TCA wurde tropfenweise zugegeben. Während einer Inkubationszeit von 50 min auf Eis wurde das geschlossene Gefäß mehrmals geschwenkt. Danach folgte ein weiterer Zentrifugationsschritt für

15 min bei 13200 rpm. Der Überstand wurde verworfen und das Histon-Pellet wurde zwei Mal mit eiskaltem Aceton gewaschen, bevor es in ABC-Puffer (pH 8.0) für den anschließenden tryptischen Verdau und TMT-Markierung aufgenommen wurde.

5.7.8 Proteinverdau

50 µg BSA wurden in 50 µL TEAB-Puffer gelöst, mit 5 µL 1 M TCEP-Lösung 60 min bei 60 °C und 750 rpm reduziert und mit 10 µL 1 M IAA-Lösung für 40 min bei Raumtemperatur im Dunkeln alkyliert. Alle reduzierten und alkylierten Proteinproben wurden mit *MS-Grade* Trypsin (*Thermo Fisher Scientific*, Waltham, MA, USA) nach Herstellerangaben über Nacht bei 650 rpm und 37 °C verdaut. Zum Abstoppen des Verdau wurden die Proben für 5 min auf 99 °C erhitzt, auf Raumtemperatur abgekühlt und es wurde jeweils 1 µL PMSF-Lösung zugegeben.

5.7.9 Isobare Markierung

Die **CIT-Markierungsreagenzien** wurden von Barbara Steigenberger nach Sohn *et al.* synthetisiert und gelöst in DMSO mit einer Konzentration von 50 µg/µL bei -80 °C gelagert.^[182] Nach Entwicklerangaben wurden je 50 µg BSA-Verdau mit je 8 µL der CIT¹⁶⁴- oder CIT¹⁶⁹-Reagenzlösung für 2 h bei Raumtemperatur inkubiert. Anschließend wurden die Proben mittels MALDI MS analysiert.

Die **TMT-Markierungsreagenzien** wurden entweder käuflich erworben (*Thermo Fisher Scientific*, Waltham, MA, USA) oder *in house* von Gengo Kashiwazaki synthetisiert. Die Markierung der tryptisch verdauten Proteine wurde nach Herstellerangaben durchgeführt. 50 µg BSA-Verdau wurden mit 400 µg, *pull down* Proben und extrahierte Histone wurden mit 200 - 400 µg TMT Reagenz markiert. Es wurden die TMT-6plex Reagenzien TMT¹²⁹, TMT¹³⁰ und TMT¹³¹ für die extrahierten Histone verwendet, sonst die TMT-Duplex Reagenzien TMT¹²⁶ und TMT¹²⁷. Wurde in *forward/reverse*-Experimenten eine Probe mit TMT¹²⁶ markiert, dann wurde die entsprechende Kontrollprobe mit dem TMT¹²⁷ Reagenz markiert und *vice versa*. In einigen Experimenten wurde jede Probe und jede Kontrollprobe aufgeteilt, um sowohl mit TMT¹²⁶ als auch mit TMT¹²⁷ markiert werden zu können. Diese Vorgehensweise ist an SILAC-Experimente angelehnt und in Abbildung 48 veranschaulicht. Überschüssige TMT Reagenzien wurden nach 2 h durch die Zugabe von 8 µL 5% Hydroxylamin-Lösung gequencht und die unterschiedlich markierten Proben im Verhältnis 1:1 gemischt.

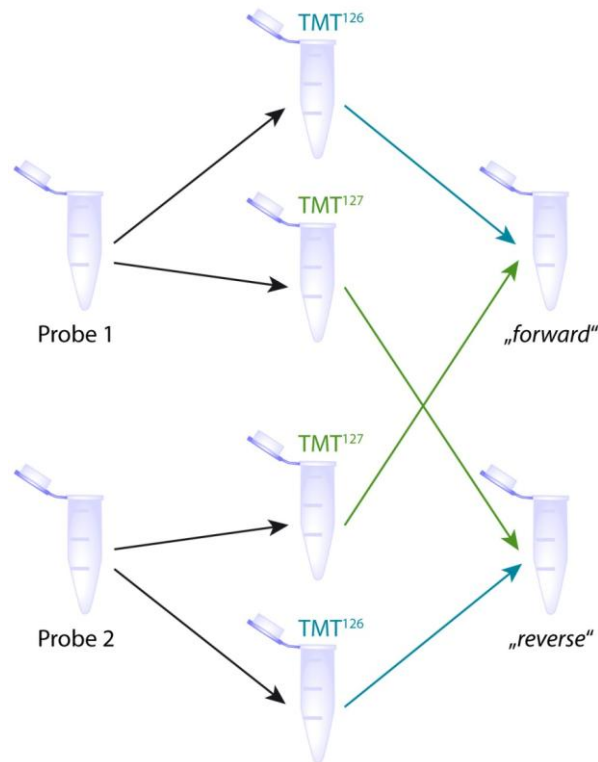


Abbildung 48: Schema zur TMT-Duplex Markierung bei *forward/reverse* Experimenten.

5.8 Massenspektrometrie

Wegen ihrer unterschiedlichen Vorteile, wurde je nach Anwendung auf unterschiedliche massenspektrometrische Methoden zurückgegriffen. Ein MALDI-TOF Massenspektrometer wurde für verschiedene Qualitätskontrollen (Oligodesoxynukleotide) und einfache Analysen (BSA-Verdau mit isobarer Markierung) verwendet. Insbesondere die Messung von komplexen Proben (Protein-Verdau) und deren Quantifizierungen (TMT) wurden an einer Orbitrap durchgeführt. Desweiteren wurden die Detektion und Quantifizierung von modifizierten Basen der DNA mit Hilfe eines Triple-Quadrupol-Massenspektrometers durchgeführt.

5.8.1 MALDI-TOF

Alle MALDI-TOF-Spektren wurden mit einem *Bruker*-Autoflex II-Massenspektrometer und einem AnchorChipTM-Target (*Bruker MTP var/384*) und der Messsoftware FlexControl 3.0 generiert. Die Standard-Geräteparameter der jeweiligen verwendeten Messmethode können der Tabelle 21 entnommen werden.

Synthetisierte Oligodesoxynukleotide wurden über 30 min auf einer 0.025 µm-Membran (VSWP, *Merck Millipore*, Darmstadt) gegen *ddH₂O* dialysiert. Die entsalzten Proben wurden dann auf einer HPA-Matrix (3-Hydroxypicolin-Säure (50 mg), 15-Krone-5 (10 µL), Ammonium-Hydrogencitrat (10 mg)

in 500 μL H_2O und 500 μL MeCN) co-kristallisiert und im Negativ-Modus gemessen. Bei der Analyse von hybridisierten Doppelsträngen nach der Click-Reaktion mit dem Crosslinker wurde wegen der Hydrolyseempfindlichkeit des NHS-Esters die vorige Entsalzung auf 10 min gekürzt bzw. ganz darauf verzichtet.

Verdaute Proteinproben wurden vor der Messung nach Herstellerangaben mit *ZipTips* (C18, *Merck Millipore*, Darmstadt) entsalzt und dann auf einer HCCA-Matrix (15 mg α -Cyano-4-Hydroxymethyl-Säure in 1 mL 30 % (v/v) Acetonitril und 0,1 % (v/v) Trifluoressigsäure) co-kristallisiert. Die Messungen wurden im Positiv-/Reflektor-Modus ausgeführt. Dabei wurden je 1000 Schuss mit einer Frequenz von 20 Hz aufgenommen und zu einem Gesamtspektrum aufsummiert.

Tabelle 21: Standard-Geräteparameter des Bruker-Autoflex II.

Probenart	Proteinverdau/Peptide	Einzelsträngige DNA	Doppelsträngige DNA (mit CL)
Messbereich m/z	700-4000	1000-10000	5000-9000
Signalunterdrückung	< 700 Da	< 2000 Da	< 1800 Da
Ionenquelle I	19 kV	19 kV	19 kV
Ionenquelle II	16.8 kV	15.6 kV	15.6 kV
Linsenspannung	8.8 kV	8.85 kV	8.85 kV
Reflektorspannung I	20 kV	20 kV	20 kV
Reflektorspannung II	0	0	0
Laser-Abschwächung	52%	30%	44%

5.8.2 Triple Quadrupol

Quantitative LC/UV-ESI-MS/MS-Analysen verdauter DNA-Proben wurden von Toni Pfaffeneder oder Jessica Steinbacher mit einem *Agilent* (Santa Clara, CA, USA) 1290 UHPLC-System, das mit einem UV-Detektor ausgerüstet war, und mit einem *Agilent* 6490 Triple Quadrupole Massenspektrometer durchgeführt. Dabei wurde eine verbesserte Methode der stabilen Isotopenverdünnungsmethode verwendet, wodurch alle Nukleoside zeitgleich analysiert werden konnten.^[1] Die Chromatographie wurde mit einer Poroshell 120 SB-C8-Säule (*Agilent*, 2.7 μm , 2.1 mm \times 150 mm) durchgeführt.

5.8.3 Orbitrap XL

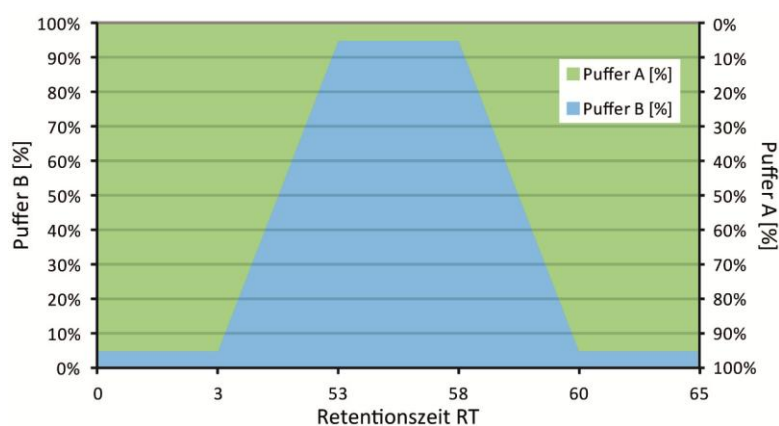
Die Proben der Versuche aus den Kapiteln 3.1, 3.2.1 sowie 3.2.2 wurden vor der LC-MS/MS-Messung an einer *SpeedVac* komplett eingengt und in 30-50 μL 2% (v/v) Acetonitril und 0.1% (v/v) Ameisensäure aufgenommen und für die Messung in Microvials überführt.

Die chromatographische Auftrennung der Peptide gemäß ihrer physikochemischen Eigenschaften vor der massenspektrometrischen Analyse wurde an einem *Dionex/Thermo Fisher Scientific* (Waltham, MA, USA) Ultimate 3000 Nano-LC-System durchgeführt. 5 – 15 μL der Proben wurden auf einer μ -

Vorsäule (Acclaim® PepMap100, C18, 5 µM, 100 Å, 300 µm i.D. x 5 mm) zunächst entsalzt und aufkonzentriert. Anschließend wurden die Proben auf eine analytische nano-Säule (Reprosil-Pur®, C18, 3 µM, 120 Å, in 75 µm i.D. x 150 mm und 8 µm Picotip®-Emitter) gespült und mit einer Flussrate von 0.3 µL/min eines mehrstufigen Gradienten in 65 min aufgetrennt (siehe Tabelle 22). Das Eluat wurde mittels ESI ionisiert und in der LTQ Orbitrap XL der Firma *Thermo Fisher Scientific* (Waltham, MA, USA) analysiert.

Tabelle 22: Gradient der chromatographischen Auftrennung vor der MS-Analyse an der LTQ Orbitrap XL mit Puffer A (0.1 % Ameisensäure in H₂O) und Puffer B (0.1 % Ameisensäure in ACN).

Zeit [min]	Puffer A [%]	Puffer B [%]
0 – 3	95	5
3 – 53	95 – 5	5 – 95
53 – 58	5	95
58 – 60	5 – 95	95 – 5
60 - 65	95	5



Die massenspektrometrischen Analysen starteten mit einem *Full-Mass-Scan* im Massenbereich zwischen m/z 300 und 1650 mit einer Auflösung von 60000. Danach folgten drei MS/MS-Scans mit dem FTMS-Analysator bei einer Auflösung von 7500 und einer hohen Kollisionsenergie von 70 eV in der HCD-Zelle sowie drei zusätzlichen MS/MS-Scans mit einer normalisierten Kollisionsenergie von 35 % in einer Ionenfalle der drei intensivsten Vorläuferionen mit Ladungen höher als Eins in einem Isolationsfenster von m/z 2.0.

5.8.4 Q Exactive HF

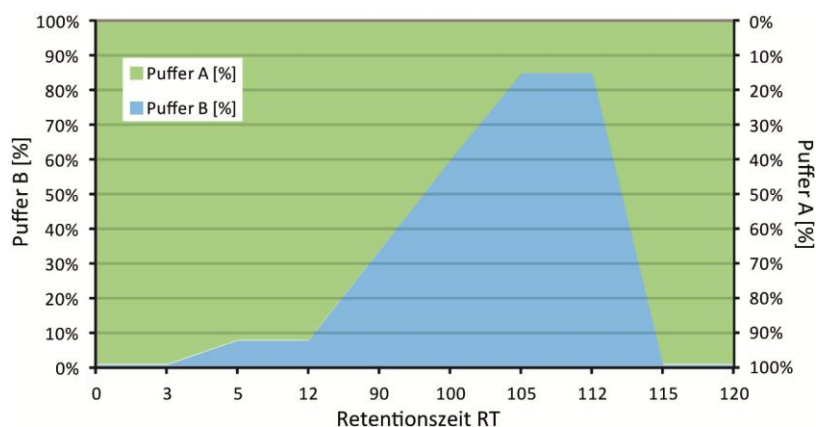
Vor den Messungen an der Q Exactive HF der Firma *Thermo Fisher Scientific* (Waltham, MA, USA) wurden die Proben aus Kapitel 3.2.3 nach der Markierung mit TMT zunächst über *StageTips* nach dem Protokoll von Rappsilber *et al.* von Verunreinigungen wie Detergenzien oder Reaktionsnebenprodukte befreit.^[222] Dafür wurden die Peptidproben zunächst an der *SpeedVac* eingengt und mit 50-80 µL 1 % TFA in H₂O angesäuert. Die *StageTips* wurden mit 200 µL 1 % TFA in H₂O equilibriert und danach wurden die Proben aufgetragen. Es wurde mit je 200 µL 0.2 % TFA in H₂O, Isopropanol und ein weiteres Mal 0.2 % TFA in H₂O gewaschen, bevor mit 150 µL 80 % MeCN und 5 % NH₃ in H₂O eluiert wurde. Nach dem Entfernen des Lösemittels an der *SpeedVac*, wurden die

Proben in 30-50 µL 2% (v/v) Acetonitril und 0.1% (v/v) Ameisensäure aufgenommen und für die massenspektrometrische Messung in Microvials überführt.

Die Proben wurden dann mittels einem UltiMate 3000 Nano-LC-System (*Dionex, Thermo Fisher Scientific, Waltham, MA, USA*) gekoppelt an ein Q Exactive HF Massenspektrometer (*Thermo Fisher Scientific*) analysiert. Dafür wurden 5 µL der Proben auf eine µ-Vorsäule (*Acclaim C18 PepMap100, 5 µm, 100 µm i.D. x 5 mm, Dionex, Thermo Fisher Scientific*) aufgetragen und anschließend auf eine selbst gepackte analytische nano-Säule (*SilicaTip Emitter, 75 µm i.D. 8 µm Spitze, 15 cm, New Objective, Woburn, MA, USA; gepackt mit ReproSil®-Pur C18 AQ 1.9 µm, Dr. Maisch GmbH, Ammerbruch-Entringen, Deutschland*) gespült und mit einer Flussrate von 0.3 µL/min eines mehrstufigen Gradienten in 120 min aufgetrennt. Der Gradient wurde zusammen mit Michael Stadlmeier so gestaltet, dass Signale von Reaktionsnebenprodukten der TMT Reagenzien minimiert werden (Tabelle 23). Die Säulentemperatur wurde mit einem Säulenofen (*Sonation GmbH, Biberach, Deutschland*) konstant auf 30 °C gehalten.

Tabelle 23: Gradient der chromatographischen Auftrennung vor der MS-Analyse an der Q Exactive HF mit Puffer A (0.1 % Ameisensäure in H₂O) und Puffer B (0.1 % Ameisensäure in ACN).

Zeit [min]	Puffer A [%]	Puffer B [%]
0 – 3	99	1
3 – 5	99 – 92	1 – 8
5 – 12	92	8
12 - 90	92 - 66	8 – 34
90 - 100	66 - 40	34 – 60
100 - 105	40 - 15	60 – 85
105 - 112	15	85
112 - 115	15 - 99	85 - 1
115 - 120	99	1



Auch die Methode der massenspektrometrischen Analyse wurde optimiert, um die Messung von Ionen der TMT Nebenprodukte zu vermeiden. Dafür wurden in den ersten 45 min der Messung zwei verschiedene Scan-Events eingeführt. Das erste Scan-Event startet mit einem *Full-Mass-Scan* im Massenbereich zwischen m/z 334.9 und 385.0 mit einer Auflösung von 60000, *AGC target* (Ladungen) von $3e^6$ Ionen und einer maximalen Injektionszeit von 100 ms. Die fünf häufigsten Vorläuferionen mit Ladungen höher als eins wurden in einem Isolationsfenster von m/z 1.5 für die Fragmentierung in der HCD-Zelle bei stufenweise steigenden normalisierten Kollisionsenergien von 24, 30 und 36

ausgewählt. Die MS/MS-Scans wurden bei einer Auflösung von 30000, AGC *target* von $1e^5$ Ionen und einer maximalen Injektionszeit von 100 ms aufgenommen. Um mehrfach-Fragmentierungen gleicher Vorläuferionen zu vermindern, wurde die dynamische Exklusion auf 10 s gesetzt. Um die MS/MS-Signalintensität zu erhöhen, wurde der Apex-Trigger so eingestellt, dass die Peakspitzen zwischen 10 und 20 s nach der ersten Detektion des Vorläuferions gesucht werden.

Nach diesem ersten Scan-Event folgte ein zweites mit denselben Einstellungen, bis auf Folgende: Der *Full-Mass-Scan* wurde auf einen Massenbereich zwischen m/z 390 und 1750 mit einer Auflösung von 120000 eingestellt und es wurden die zehn intensivsten Vorläuferionen für die anschließenden MS/MS-Scans ausgewählt.

Für die Messung nach den ersten 45 min wurde der *Full-Mass-Scan* auf einen Massenbereich von m/z 335 und 1750 mit einer Auflösung von 120000 eingestellt und es wurden die 15 intensivsten Vorläuferionen der zuvor beschriebenen Fragmentierung unterworfen.

5.8.5 Proteinidentifikation und Quantifizierung

Der Datenbank-Abgleich der aus den Messungen an den Orbitrap-Massenspektrometern erhaltenen Daten erfolgte mit dem Proteome Discoverer 1.4 (*Thermo Fisher Scientific*, Waltham, MA, USA) in Verbindung mit dem SEQUEST HT-Algorithmus und einer UniProt-Datenbank (*Mus Musculus*, November 2014). Als Detektionslimit wurde das Signal-Rausch-Verhältnis (S/N) auf 3 (LTQ Orbitrap XL) oder 1.5 (Q Exactive HF) gesetzt. Als Protease wurde Trypsin mit maximal zwei übersprungenen Spaltstellen gewählt. Die maximale Massentoleranz der Vorläufer-Ionen wurde mit 10 ppm und die der Fragment-Ionen mit 0.8 Da (Orbitrap XL) oder 0.02 Da (Q Exactive HF) festgelegt. Als statische Modifikation wurde Carbamidomethylcystein eingestellt. Dynamische Modifikationen waren: Na-Kation (D, E) und Oxidation (M) sowie der DNA-Protein Crosslinker-Rest (+145.020 Da; K, Y) und TMT-Duplex bzw. -6plex (N-Terminal, K) soweit verwendet.

Für die relative Quantifizierung wurden mit TMT markierte, nicht-redundante Peptide verwendet. Das Toleranzfenster für die Integration betrug 20 ppm und die Integrationsmethode wurde auf „*most confident centroid*“ gesetzt. Die Signale der TMT-Duplex Reporterionen 126 und 127 wurden auf den Protein-Median normalisiert und daraus wurden Verhältnisse berechnet, die als Anreicherungsfaktoren die Präferenz eines Proteins angeben an eine der zu vergleichenden DNA-Sonden zu binden. In den *forward/reverse*-Experimenten wurden nur solche Proteine berücksichtigt, für die sowohl im *forward* als auch im *reverse* Experiment ein Anreicherungsfaktor berechnet werden konnte. Für Triplikate wurden die Anreicherungsfaktoren gemittelt.

Anhang

Tabelle A. 1: Liste der für C angereicherten Proteine aus den pull down Versuchen mit caC-DNA-Sonden und HEK-Zellkernlysate unter drei verschiedenen Bedingungen (Kapitel 3.1.3).

für C angereicherte Proteine						
unter Bedingung 1	unter Bedingung 2			unter Bedingung 3		
Actb	Abcf1	Msh6	Rps15	Acin1	Hmgn2	Rpl11
Aldoa	Acad	Myh10	Rps15	Akap8	Hnrnpa0	Rpl14
Arglu1	Acin1	Nat10	Rps18	Alb	Hnrnpa1	Rpl29
Dek	Alyref	Ncapp2	Rps19	Aldh18a1	Hnrnpa2b1	Rps5
Eif4a1	Atp13a3	Nhlrc2	Rps23	Alyref	Hnrnpa3	Ryr1
Eif5a	C14orf166	Nolc1	Rps24	Ankrd53	Hnrnpab	Sf1
Eno1	C1orf167	Nono	Rps25	Apc	Hnrnpc	Sfpq
Gapdh	Chd4	Npm1	Rps3	Api5	Hnrnp1	Smarca2
Hnrnpc	Chk4	Oma1	Rps6	Apob	Hnrnpk	Smc1a
Hnrnpd	Cpd	Parp1	Rps7	Bzw1	Hnrnp1	Snrpd1
Hsp90ab1	Csnk2a1	Pdcd11	Rsl1d1	Bzw2	Hnrnp1	Snrpd2
Hspa1a	Ddx1	Ppib	Sfpq	Ccnb3	Hsd12	Snrpd3
Krt1	Ddx21	Ptbp1	Slc39a4	Cdk5rap2	Ilf2	Snrpe
Ldha	Dek	Ran	Smu1	Chd4	Ilf3	Snrpg
Ldhb	Dhx15	Rcc1	Snrpd3	Col13a1	Khdrbs1	Snrpn
Msn	Eftud2	Rnf44	Snrpn	Cpsf6	Khsrp	Srp72
Myh9	Eif2s3	Rpl10a	Srp9	Ddx17	Kpna5	Srsf1
Myl6	Fau	Rpl13a	Ssrp1	Ddx39b	Krt1	Srsf3
Ncl	Fbl	Rpl14	Supt16h	Diaph2	Krt10	Srsf4
Npm1	Fus	Rpl18a	Tcof1	Eef1a1	Larp6	Stag2
Ola1	Gon4l	Rpl21	Top1	Eef1g	Maoa	Stap2
Phgdh	Grp	Rpl22	Top2b	Eef2	Maz	Syncrip
Ppia	Grwd1	Rpl23a	U2af2	Eif2ak2	Mdn1	Tex10
Rcc1	H1foo	Rpl24	Vrk1	Eif4a2	Nin	Thoc6
Rfc1	H1fx	Rpl26		Eif4a3	Nono	Xrcc5
Rpl13	Hist1h1c	Rpl27		Eif5a	P2rx5	Xrcc6
Rpl23a	Hist1h1d	Rpl28		Eif5b	Pbrm1	Ywhaz
Rpl26l1	Hist1h1e	Rpl29		Erh	Pcbp2	
Rpl27	Hmgb3	Rpl30		Exoc7	Pdcd11	
Rps14	Hnrnpa0	Rpl31		Ezr	Plrg1	
Rps20	Hnrnpa1	Rpl34		Fau	Poldip3	
Rps27a	Hnrnp1	Rpl35		Fus	Ppia	
Rps7	Igf2bp1	Rpl35a		Gcn111	Ppib	
Rps9	Kif2a	Rpl39		Heat1	Prc1	
Serbp1	Matr3	Rpl6		Hist1h1e	Ptbp1	
Snrpd1	Maz	Rpl7a		Hmgb1	Ptbp3	
Snrpn	Mcts1	Rpl9		Hmgb2	Rcc1	
Taf1	Msh2	Rps11		Hmgb3	Rcc2	

Tabelle A. 2: Angereicherte Proteine für fC an Tag 5 (Kapitel 3.2.2). Es wurden nur Proteine mit einem Quotienten größer als 3/2 bzw. kleiner als 2/3 berücksichtigt.

mit Crosslinker			ohne Crosslinker		
Genname	fC ¹²⁶ /C ¹²⁷	C ¹²⁶ /fC ¹²⁷	Genname	fC ¹²⁶ /C ¹²⁷	C ¹²⁶ /fC ¹²⁷
Chrm2	1,727	0,154	Adam12	4,245	0,544
Dmd	2,819	0,198	Kcnt2	13,161	0,300
Lamb1	1,875	0,564	Pip5k1a	1,774	0,407
Mycbp2	2,573	0,651	Rrbp1	1,642	0,667
P4ha2	2,133	0,078	Syngap1	1,721	0,176
Ptpn11	4,615	0,550	Taf3	4,510	0,410
Ptprt	1,707	0,232	Tjp1	2,332	0,222
Sirt7	1,681	0,199	Trip12	3,489	0,488
Tpr	3,353	0,617	Zfp335	2,116	0,544
Unc13c	1,767	0,488	Zkscan4	1,885	0,421

Anhang

Tabelle A. 3: Angereicherte Proteine für C im Vergleich zu fC an Tag 5 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	fC ¹²⁶ /C ¹²⁷	C ¹²⁶ /fC ¹²⁷	Genname	fC ¹²⁶ /C ¹²⁷	C ¹²⁶ /fC ¹²⁷
Aurkaip1	0,027	5,971	Actbl2	0,516	1,809
Cdc27	0,294	2,116	Adal	0,027	2,086
Gck	0,526	3,541	Atp13a3	0,508	1,877
Kif17	0,653	1,587	Cdc14a	0,465	3,599
Msh2	0,590	1,656	Chmp2a	0,500	1,741
Mtcl1	0,165	1,802	Clspn	0,513	1,574
Ncor1	0,329	3,827	Ddhd1	0,465	2,978
Olfr695	0,178	4,190	Dis3	0,121	1,927
Rpgrip1	0,596	1,902	Eml1	0,525	2,460
Tcof1	0,344	2,290	Fam160b1	0,423	3,482
Trim30c	0,628	2,099	Fhod1	0,104	1,736
Wdfy4	0,513	1,574	Ifih1	0,216	2,156
			Krt17	0,646	1,826
			Plcl2	0,612	2,209
			Ppp1r12a	0,217	2,166

Tabelle A. 4: Angereicherte Proteine für mC an Tag 11 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷	Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷
Adamts19	4,216	0,336	Abcb10	20,941	0,038
Alpk1	33,387	0,097	Asic2	100,000	0,170
Cramp1l	21,091	0,119	Atp11a	100,000	0,587
Dst	5,184	0,331	Casp8ap2	19,383	0,242
Fam107a	2,384	0,422	Ccdc73	10,285	0,112
Fam205a	1,508	0,591	Ccdc88b	1,659	0,657
Fen1	10,545	0,121	Cd84	8,816	0,183
Lamc1	2,372	0,308	Cep350	1,609	0,506
Nifk	14,115	0,108	Cfap54	54,713	0,107
Pcnx	30,636	0,526	Clhc1	1,640	0,464
Ppp2r5b	7,199	0,384	Col6a6	4,615	0,119
Ppp4r4	61,918	0,230	Cyb5r2	14,115	0,108
Rbm39	21,832	0,033	Cyp4f40	33,387	0,149
Rit1	70,193	0,094	Cyth2	76,710	0,092
Setdb1	15,660	0,271	Fam227b	39,848	0,516
Smarca5	3,536	0,088	Fam71e2	2,878	0,629
C19orf43	2,406	0,525	Fmnl1	1,962	0,608
Utp3	2,615	0,659	Fyb	45,900	0,308
			Gm608	6,399	0,171
			Gm6578	1,890	0,656
			Grin3a	2,157	0,425
			Hells	73,314	0,096
			Hif1a	8,143	0,420
			Hsp90b1	26,837	0,119
			C10orf11	7,633	0,325
			Map1b	12,674	0,174
			Mb21d2	59,574	0,525
			Mtcl1	1,853	0,088
			Muc5ac	2,805	0,070
			Myrf	20,076	0,119
			Nap1l1	21,807	0,129
			Olfr68	3,430	0,256
			Pcm1	25,681	0,548
			Pcnxl3	54,770	0,114
			Pex16	15,437	0,153
			Phldb2	74,288	0,291
			Phlpp2	1,609	0,456
			Phrf1	1,752	0,505
			Postn	1,932	0,076
			Rimbp3	3,104	0,292
			Samd9l	30,636	0,509

Anhang

Fortsetzung Tabelle A. 4: Angereicherte Proteine für mC an Tag 11 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷	Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷
			Senp7	100,000	0,033
			Sh3glb1	100,000	0,179
			She	21,900	0,184
			Smarcal1	58,794	0,115
			Srek1	49,669	0,149
			Tenm4	2,436	0,074
			Trappc10	33,387	0,095
			Ttc21a	39,998	0,122
			Urod	100,000	0,097
			Vwa3a	3,425	0,161
			Wt1	54,713	0,098
			Xpnpep2	24,220	0,126
			Yeats2	53,497	0,100
			Zmynd11	39,998	0,265
			Znfx1	3,590	0,179
			Zw10	15,660	0,074

Tabelle A. 5: Angereicherte Proteine für C im Vergleich zu mC an Tag 11 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷	Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷
7530416G11Rik	0,184	13,633	1700039E15Rik	0,213	7,016
Arfgap2	0,228	3,813	4930480E11Rik	0,152	17,859
Arhgap10	0,071	49,416	5730559C18Rik	0,379	3,240
Ash1l	0,619	2,820	Akap11	0,410	6,380
Asxl3	0,160	5,465	Akap6	0,200	14,082
Atmin	0,620	2,463	Anxa4	0,243	25,526
Bpifb6	0,436	1,686	Asrg1	0,176	16,972
Cacnb4	0,660	6,041	BC094916	0,422	2,496
Car4	0,647	16,549	C130060K24Rik	0,320	6,132
Chd4	0,246	3,521	Cdh6	0,388	11,141
Chd6	0,564	3,239	Cdk1	0,280	16,633
Dnah5	0,369	3,167	Cdk11b	0,521	3,864
Fan1	0,370	47,893	Chaf1a	0,360	4,177
Herc2	0,473	54,048	Chd9	0,227	4,872
Man1c1	0,172	17,183	Cmtr2	0,152	17,859
Myh7b	0,259	7,306	Col25a1	0,349	2,725
Mylk	0,522	2,863	Cpsf1	0,513	4,845
Ncapg2	0,365	3,419	Csf2	0,190	14,082
Pck1	0,489	26,122	Dmxl2	0,476	8,586
Pinx1	0,569	11,263	Dyrk2	0,309	13,692
Psmc2	0,561	8,890	Elovl2	0,521	3,769
Ptchd4	0,625	3,518	Enpp6	0,127	15,301
Rpl4	0,235	11,263	Eprs	0,465	1,637
Rps6ka4	0,144	24,845	Fam178b	0,155	12,227
Syt3	0,272	4,852	Fbln2	0,505	1,547
Tpp2	0,373	12,566	Flna	0,521	3,769
Tsc1	0,244	4,811	Flvcr2	0,420	3,641
Ttn	0,588	1,861	Gabrg2	0,390	3,373
Ttn	0,596	1,561	Gm10112	0,348	17,637
C3orf38	0,368	7,130	Gm27179	0,199	4,521
			Gm933	0,316	5,877
			Gucy1b2	0,522	2,208
			Haa0	0,375	7,252
			Hist1h2ah	0,251	2,625
			Hist1h2al	0,251	3,070
			Hist1h4a	0,203	8,189
			Hmg20b	0,606	7,105
			Hnrnpul2	0,218	5,330
			ldh3a	0,170	3,487
			Ifih1	0,306	17,684

Anhang

Fortsetzung Tabelle A. 5: Angereicherte Proteine für C im Vergleich zu mC an Tag 11 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	mC^{126}/C^{127}	C^{126}/mC^{127}	Genname	mC^{126}/C^{127}	C^{126}/mC^{127}
			Igfbp1	0,268	4,959
			Kansl1	0,469	5,416
			Kcnj11	0,423	4,047
			Kiaa1211	0,595	27,873
			Kif23	0,165	4,596
			Kntc1	0,144	14,814
			Loxl2	0,160	4,653
			Lrsam1	0,327	9,383
			Mapre3	0,659	4,872
			Mcmbp	0,359	8,261
			Mmp24	0,289	11,691
			Mob1b	0,152	14,814
			Myo15a	0,192	12,734
			Nrbf2	0,132	28,938
			Nutm1	0,136	15,301
			Obox1	0,326	14,982
			Onecut3	0,341	4,736
			Osgepl1	0,232	1,885
			Pbrm1	0,338	6,982
			Pck2	0,247	17,859
			Pcsk2	0,556	3,905
			Pla2g4a	0,286	16,286
			Plekhs1	0,568	2,440
			Pmp2	0,501	3,813
			Ppfia2	0,563	1,607
			Prep	0,561	2,058
			Prg4	0,471	4,448
			Prkcd	0,288	2,456
			Prrc2b	0,127	64,364
			Ptprq	0,144	17,943
			Pxn	0,492	1,599
			Qars	0,595	1,828
			Rabggta	0,612	1,818
			Rbbp6	0,071	2,565
			Rgs8	0,615	5,409
			Rpf2	0,154	64,364
			Rpl8	0,338	1,540
			Rps6	0,561	2,329
			Rsph9	0,268	4,959
			Scaf4	0,373	1,596
			Shroom2	0,501	4,279
			Slc26a9	0,255	5,877
			Slc44a2	0,595	5,477
			Svep1	0,298	9,292
			Timm17a	0,602	2,184
			Tmem214	0,235	9,855
			Tram1	0,568	2,566
			Trim66	0,396	8,403
			Ube2s	0,125	13,633
			Vil1	0,650	3,972
			Vmn2r55	0,155	8,189
			Wrn	0,221	12,345
			Zbtb7c	0,136	12,288
			Zfp668	0,467	3,729

Anhang

Tabelle A. 6: Angereicherte Proteine für hmC an Tag 11 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	hmC ¹²⁶ /C ¹²⁷	C ¹²⁶ /hmC ¹²⁷	Genname	hmC ¹²⁶ /C ¹²⁷	C ¹²⁶ /hmC ¹²⁷
Bpifb1	10,154	0,076	Adra2a	6,932	0,243
Ctnna1	3,804	0,145	Akap17b	3,588	0,146
Dlgap3	2,122	0,617	Bicd2	19,643	0,045
Gm14569	1,536	0,487	Casp8ap2	2,917	0,299
P2rx6	2,212	0,583	Ccdc178	13,859	0,218
Pcdh11x	2,982	0,304	Ccm2	1,758	0,639
Pcxl3	12,665	0,442	Ccna2	1,905	0,387
Pdcl3	2,263	0,149	Cep120	3,933	0,589
Sfrp1	1,542	0,286	Cep131	1,666	0,387
Syt15	3,515	0,421	Cfap53	18,556	0,121
U2surp	5,069	0,076	Cfap54	3,413	0,137
Usp15	1,516	0,376	Chek2	3,319	0,467
			Cpd	2,295	0,652
			Cpsf1	3,299	0,27
			Cyth2	2,106	0,308
			D2hgdh	1,576	0,536
			Dapk1	17,336	0,501
			Fam193a	7,111	0,645
			Fam227b	7,462	0,281
			Fen1	2,434	0,217
			Fer14	3,345	0,467
			Gm340	7,242	0,285
			Gm6406	3,301	0,519
			Gm6578	2,685	0,458
			Golph3l	2,307	0,375
			Hapln2	1,718	0,225
			Igf2bp2	4,143	0,252
			Kif2a	1,714	0,327
			March8	1,767	0,579
			Myh10	3,705	0,42
			Myo1f	2,168	0,648
			Nap1l1	2,265	0,519
			Nit1	3,821	0,252
			Nrip1	3,694	0,473
			Osbp18	2,909	0,589
			Pcdhb7	6,14	0,624
			Pkd2l2	1,589	0,446
			Polr2a	2,212	0,6
			Pqbp1	2,446	0,355
			Rangap1	7,118	0,132
			Robo3	17,336	0,653
			Rpl5	2,765	0,376
			Shroom2	1,786	0,554
			Trhde	3,865	0,583
			Ttc21a	16,793	0,418
			Wt1	17,695	0,137
			Zfp428	2,456	0,506
			Zfp668	1,667	0,486
			Zw10	2,37	0,236

Anhang

Tabelle A. 7: Angereicherte Proteine für C im Vergleich zu hmC an Tag 11 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	hmC ¹²⁶ /C ¹²⁷	C ¹²⁶ /hmC ¹²⁷	Genname	hmC ¹²⁶ /C ¹²⁷	C ¹²⁶ /hmC ¹²⁷
Abcb1b	0,634	1,756	4932429P05Rik	0,425	1,623
Armc10	0,624	1,562	AF366264	0,485	7,012
Cmya5	0,408	1,638	Arap2	0,280	2,685
Copa	0,481	2,133	Baiap2l1	0,261	1,567
D430041D05Rik	0,547	1,524	Ccdc108	0,339	1,799
Dmrct2	0,405	1,603	Cdk1	0,523	1,579
Duox1	0,569	1,545	Cep112	0,545	1,955
Fer1l5	0,431	2,353	Chrm2	0,469	1,528
Gcc2	0,302	2,801	Clspn	0,326	1,601
Gne	0,530	1,525	Csf2	0,448	1,929
Gvin1	0,545	1,659	Cyp1a2	0,324	2,544
Krit1	0,297	1,591	Dpysl5	0,128	2,421
Lct	0,522	2,209	Ebf2	0,578	2,234
Med13	0,389	1,841	Enpp2	0,404	2,421
Myo15b	0,389	1,845	Fam47c	0,452	2,150
Ncoa6	0,489	1,680	Fam47e	0,413	1,837
Spaca5	0,191	7,303	Fbxw7	0,151	2,096
Stxbp2	0,347	2,685	Flnc	0,562	1,892
Uggt1	0,385	1,765	Gm10610	0,468	1,813
Wdfy4	0,532	1,506	Gmpr	0,546	2,751
Xirp2	0,545	1,514	Hfm1	0,524	1,514
			Il31ra	0,483	1,512
			Irf2	0,150	2,182
			Kntc1	0,389	1,850
			Mapk15	0,287	2,767
			Myef2	0,481	1,585
			Myo3a	0,326	2,382
			Nrbf2	0,315	2,544
			Ophn1	0,575	2,093
			Pmp2	0,556	1,659
			Ppfia4	0,619	1,680
			Ppp1r12a	0,240	1,799
			Prkce	0,240	5,506
			Prpf4b	0,389	1,841
			Sympk	0,626	2,289
			Tgfbr2	0,268	3,661
			Trim66	0,469	1,885
			Ufd1l	0,203	7,115
			Zscan29	0,589	2,240

Tabelle A. 8: Angereicherte Proteine für mC an Tag 21 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷	Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷
1110002E22Rik	1,592	0,577	Acox1	2,225	0,209
Ankrd11	6,580	0,600	Adarb1	1,503	0,365
C030017K20Rik	2,657	0,480	Ak9	2,250	0,617
Dhx8	6,228	0,222	Akr1c6	2,311	0,393
Fam179a	2,795	0,508	Ankrd36	1,642	0,318
Fryl	2,096	0,399	Asrgl1	3,941	0,351
Hist1h1b	1,589	0,614	Atp5o	3,401	0,472
Krr1	2,235	0,349	Axdnd1	2,564	0,647
Man1c1	5,614	0,190	Cd300lf	2,762	0,455
Neb	3,481	0,313	Cltc	1,816	0,491
Nolc1	2,031	0,619	Cnot4	3,075	0,417
Pitrm1	2,008	0,241	Cypt3	1,542	0,556
Ppp3cc	5,304	0,310	Dcdc2	2,665	0,303
Rps6ka6	5,037	0,303	Dennd4c	1,639	0,585
Svep1	2,055	0,373	Fahd1	1,724	0,460
Sycp2l	1,600	0,429	Flvcr2	1,825	0,502
Tnrc18	1,746	0,427	Gm8251	2,792	0,502

Anhang

Fortsetzung Tabelle A. 8: Angereicherte Proteine für mC an Tag 21 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷	Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷
Vps16	3,011	0,514	Gopc	24,930	0,359
Wasf2	2,128	0,603	Gorasp2	2,947	0,482
			H2afy	1,553	0,589
			Herc2	35,414	0,337
			Huwe1	1,658	0,409
			Kndc1	2,251	0,399
			Lonrf1	1,560	0,314
			Lrif1	4,951	0,298
			Mamstr	1,787	0,462
			Mettl14	1,548	0,493
			Mob1b	2,432	0,409
			Nutm1	4,060	0,221
			Pcsk2	1,879	0,490
			Plk4	3,389	0,610
			Prg4	1,509	0,390
			Rdh8	5,571	0,190
			Rpl13	3,432	0,525
			Rps6	1,656	0,417
			Sbf1	6,498	0,306
			Sfrp1	2,333	0,409
			Sh2d4a	1,503	0,406
			Smarca1	1,634	0,499
			Spryd4	2,021	0,621
			Srbd1	1,537	0,621
			Stk4	2,024	0,621
			Svs1	1,769	0,475
			Tead3	2,251	0,399
			Tmem214	2,443	0,269
			Txnl4b	2,513	0,283
			Usp48	2,692	0,310
			Znf687	2,372	0,490

Tabelle A. 9: Angereicherte Proteine für C im Vergleich zu mC an Tag 21 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷	Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷
Hapln2	0,317	2,305	2010300C02Rik	0,489	2,425
Hipk2	0,034	1,507	Abcg1	0,146	9,255
Ireb2	0,506	3,872	Amfr	0,316	3,798
Nap111	0,490	2,209	Amot	0,396	3,313
Phactr4	0,269	5,167	Arfgef3	0,257	1,754
Pi4k2b	0,194	30,550	Baz1a	0,342	1,524
Ppl	0,382	5,048	Cald1	0,465	4,645
Rims1	0,352	8,422	Casz1	0,490	4,044
Ttn	0,157	2,057	Ccdc158	0,356	3,445
Uhrf1bp1l	0,059	2,721	Clybl	0,255	7,537
			Cmya5	0,472	5,271
			Col11a1	0,033	57,251
			Dhx15	0,042	2,948
			Dr1	0,486	1,661
			Drosha	0,502	1,556
			Eml5	0,277	4,229
			Fen1	0,352	1,680
			Fhod3	0,591	5,062
			Fyb	0,578	2,734
			Gm8973	0,105	5,845
			H1fx	0,280	1,812
			Hells	0,377	1,666
			Hivep2	0,220	32,835
			Ift140	0,511	2,057
			Il1rapl2	0,281	5,005

Anhang

Fortsetzung Tabelle A. 9: Angereicherte Proteine für C im Vergleich zu mC an Tag 21 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷	Genname	mC ¹²⁶ /C ¹²⁷	C ¹²⁶ /mC ¹²⁷
			L3mbtl2	0,282	4,620
			Larp1	0,488	2,080
			Magi3	0,041	52,757
			Mcm10	0,634	1,538
			Moap1	0,300	3,669
			Nlrp4c	0,087	22,817
			Nsrp1	0,205	5,012
			Nxf7	0,661	2,054
			Pde1b	0,433	1,541
			Pde3a	0,490	2,722
			Plin2	0,277	4,481
			Postn	0,587	1,524
			Prdm11	0,283	1,516
			Rars2	0,459	1,666
			Rbm25	0,283	6,049
			Rev3l	0,434	2,708
			Rgl1	0,204	19,289
			Rtca	0,253	3,669
			Shkbp1	0,502	2,057
			Shroom3	0,431	23,532
			Sipa1l1	0,126	2,212
			Slc28a3	0,568	13,830
			Srek1	0,449	2,819
			Tcf21	0,149	4,525
			Trappc10	0,237	3,615
			Trp53rk	0,040	31,849
			Trpv3	0,244	2,257
			Ttc21a	0,279	4,880
			Vil1	0,121	23,532
			Znf608	0,087	22,817
			Zw10	0,297	2,484

Tabelle A. 10: Angereicherte Proteine für hmC an Tag 21 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	hmC ¹²⁶ /C ¹²⁷	C ¹²⁶ /hmC ¹²⁷	Genname	hmC ¹²⁶ /C ¹²⁷	C ¹²⁶ /hmC ¹²⁷
5430421N21Rik	3,022	0,343	1700039E15Rik	2,736	0,284
Brd3	5,161	0,648	4931408C20Rik	1,521	0,479
Ccdc141	2,222	0,504	5730559C18Rik	2,140	0,288
Cep290	1,600	0,578	Aga	5,845	0,397
D430041D05Rik	4,671	0,529	Alpk2	3,534	0,249
Dnhd1	6,925	0,210	Armc8	1,937	0,355
Duox1	2,356	0,616	Arntl2	1,741	0,647
Hist1h2bh	2,919	0,503	Asrgl1	8,507	0,212
Lig3	5,116	0,334	Atp10a	2,540	0,498
Lrrc63	2,015	0,459	Brd2	2,219	0,293
Nolc1	3,503	0,479	Cand1	7,370	0,643
Npm2	1,971	0,399	Cd70	3,037	0,243
Nprl3	5,322	0,503	Chd4	1,586	0,280
Pde10a	1,613	0,276	Chd8	2,582	0,614
Ppp3cc	4,415	0,477	Chd9	5,149	0,331
Rgs1	10,606	0,607	Chrm2	3,720	0,372
Rpl4	2,576	0,505	Clip1	11,533	0,293
Setdb1	2,903	0,402	Ctnna2	8,939	0,343
Sfrp1	4,055	0,463	Ctnna3	2,716	0,397
Smarca1	5,370	0,547	Dennd1b	1,764	0,535
Sp2	1,922	0,399	Dmd	1,861	0,653
C3orf38	2,589	0,482	Dnajc9	13,925	0,403
Zfml	2,366	0,490	Dnmt1	3,485	0,623
Zic1	29,930	0,075	Dtl	3,556	0,212
			Flvcr2	1,615	0,467

Anhang

Fortsetzung Tabelle A. 10: Angereicherte Proteine für hmC an Tag 21 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	hmC ¹²⁶ /C ¹²⁷	C ¹²⁶ /hmC ¹²⁷	Genname	hmC ¹²⁶ /C ¹²⁷	C ¹²⁶ /hmC ¹²⁷
			Gapdh	7,370	0,380
			Gm20821	2,919	0,380
			Gpr108	1,870	0,558
			Gpr116	13,464	0,631
			Hist1h2bf	2,919	0,503
			Hspa4	2,582	0,616
			Igfbp1	2,816	0,355
			Inpp5b	3,005	0,373
			Larp6	4,045	0,526
			Larp7	2,060	0,267
			Lcorl	2,402	0,195
			Lpin2	11,611	0,233
			Man1c1	10,689	0,284
			Myl3	1,609	0,445
			Myo15a	2,707	0,271
			Nlrc4	1,687	0,467
			Nutm1	7,325	0,147
			Osbpl11	2,629	0,245
			Osbpl8	4,362	0,535
			Pdzd2	6,576	0,364
			Plekha6	1,687	0,160
			Ppfia2	3,407	0,668
			Prrc2b	8,939	0,380
			Rrp15	4,141	0,417
			Rsbn1l	19,231	0,170
			Serpinf1	1,823	0,331
			Shank1	1,639	0,535
			Shroom2	2,731	0,443
			Slc22a23	3,025	0,406
			Smpd3	1,994	0,485
			Sms	2,960	0,616
			Snupn	4,066	0,352
			Svep1	1,785	0,334
			Svs1	1,761	0,543
			Tada2b	7,001	0,358
			Tcf7	9,737	0,095
			Timm17a	1,888	0,411
			Tmem214	3,484	0,493
			Tnks2	2,390	0,553
			Top2a	4,045	0,360
			Tram1	1,726	0,539
			Trim66	3,720	0,492
			Usp48	2,830	0,169
			Whsc1l1	2,589	0,355
			Xrn1	2,308	0,388
			Xylt2	2,716	0,533
			Zfp869	2,825	0,561
			Znf185	2,466	0,447
			Zyg11a	3,664	0,226

Anhang

Tabelle A. 11: Angereicherte Proteine für C im Vergleich zu hmC an Tag 21 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	hmC ¹²⁶ /C ¹²⁷	C ¹²⁶ /hmC ¹²⁷	Genname	hmC ¹²⁶ /C ¹²⁷	C ¹²⁶ /hmC ¹²⁷
Abca5	0,234	3,693	Akap2	0,425	1,595
Adamts19	0,163	2,348	Amfr	0,1	20,883
Arrdc3	0,27	1,51	Atp1a1	0,354	1,935
Casp8ap2	0,524	3,142	Cald1	0,536	1,889
Dnah7a	0,466	1,533	Cbx8	0,435	5,838
Grm1	0,011	9,504	Cytip	0,214	1,895
Ogdhl	0,095	2,147	Eif2a	0,305	1,673
Opa1	0,387	7,434	Fbxl17	0,345	1,526
Phf3	0,035	2,481	Fbxw10	0,472	2,859
Plcb3	0,085	4,134	Git1	0,529	2,649
Rcor3	0,344	2,147	Gm1110	0,617	2,848
Setd2	0,265	2,552	Gramd1b	0,18	11,839
Skiv2l	0,309	1,65	Hbb-bs	0,281	2,228
Wdr72	0,438	4,05	Hecw2	0,231	1,958
Zfyve20	0,418	1,758	Helz	0,24	1,911
			Hhla1	0,1	23,956
			Irf2	0,493	2,196
			Jph2	0,472	8,203
			Kiaa1109	0,603	1,636
			Kif2c	0,375	2,789
			Kiz	0,503	15,087
			Klf6	0,204	2,944
			L3mbtl2	0,535	2,439
			Lmo7	0,364	1,5
			Lrch2	0,153	100
			Nrk	0,082	1,588
			OTTMUSG00000017677	0,352	33,699
			Pcnx13	0,319	100
			Plcb4	0,43	3,009
			Ppp4r4	0,433	1,756
			Rit1	0,041	19,235
			Rpl3	0,439	8,203
			Samd9l	0,567	1,992
			Srek1	0,056	32,317
			Stat4	0,438	3,065
			Sub1	0,095	100
			Synj2	0,428	5,873
			Tceal3	0,331	2,228
			Tmod3	0,091	10,979
			Tsen54	0,319	3,008
			Ttc21a	0,326	2,863
			Ttc28	0,304	1,551
			Ubr5	0,276	1,676
			Upf3a	0,427	1,638
			Vwa3a	0,305	6,97
			Wt1	0,337	2,558
			Zfp707	0,573	3,55
			Zfp831	0,099	2,649
			Znf687	0,189	2,254
			Zw10	0,294	2,648

Anhang

Tabelle A. 12: Angereicherte Proteine für fC an Tag 21 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	fC126/C127	C126/fC127	Genname	fC126/C127	C126/fC127
Abcc2	11,249	0,186	2010300C02Rik	2,162	0,545
Ccdc146	2,700	0,535	Col2a1	2,700	0,568
Fkbp4	6,901	0,628	Cpsf1	1,649	0,647
Ocrl	3,360	0,619	Nek1	4,477	0,608
Rbm41	1,573	0,663	Pcf11	9,273	0,201
Sbsn	1,617	0,493	Polr3a	3,109	0,656
			Scarf2	4,673	0,564
			Sema3e	1,899	0,415
			Spag7	2,862	0,201
			St5	1,592	0,540
			Taf3	2,065	0,212

Tabelle A. 13: Angereicherte Proteine für C im Vergleich zu fC an Tag 21 (Kapitel 3.2.2).

mit Crosslinker			ohne Crosslinker		
Genname	fC ¹²⁶ /C ¹²⁷	C ¹²⁶ /fC ¹²⁷	Genname	fC ¹²⁶ /C ¹²⁷	C ¹²⁶ /fC ¹²⁷
Abcc6	0,568	1,860	Fgl2	0,406	1,941
Cand1	0,470	5,721	Spag16	0,568	1,860
Dennd4c	0,634	5,937	Stk4	0,662	1,520
Sybu	0,086	1,558	Tyk2	0,522	1,503
Synj2	0,302	3,657	Zfp13	0,532	2,391
Vps11	0,664	2,297	Zfp606	0,580	1,860
			Zfp697	0,440	2,026

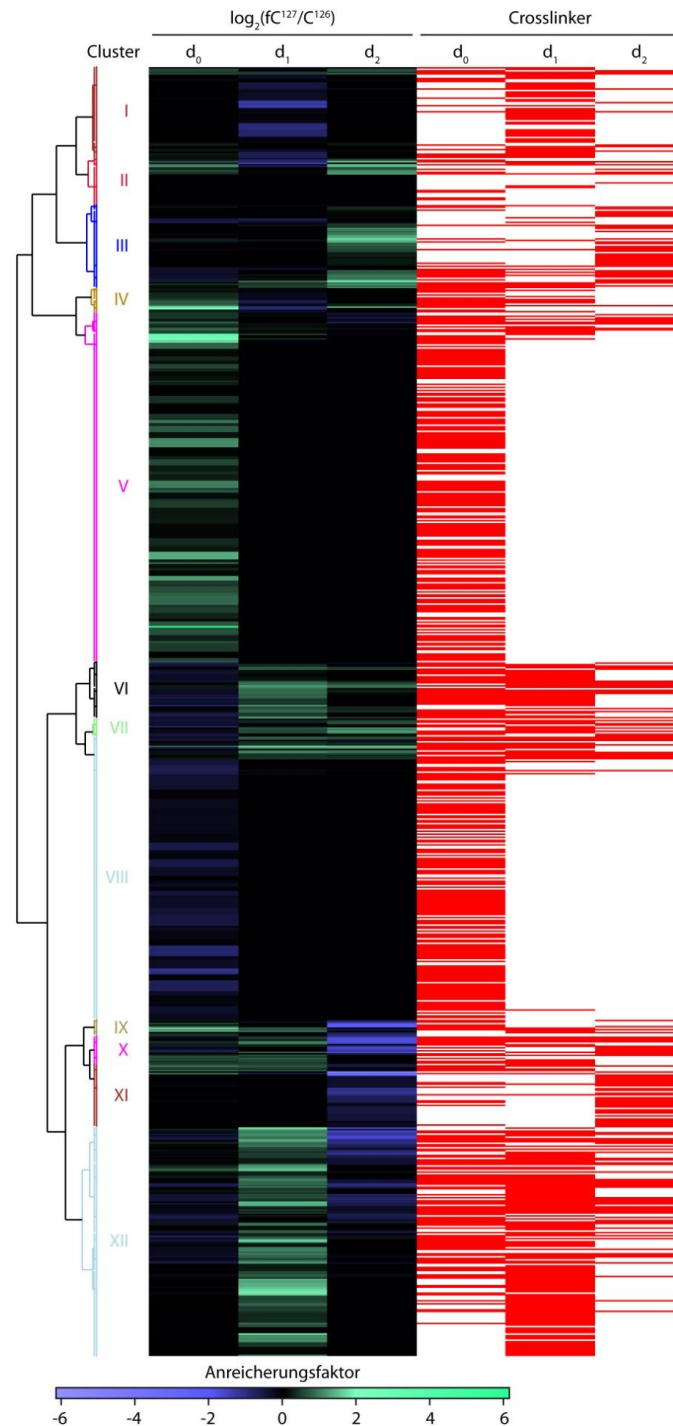


Abbildung A. 1: Hierarchisches Clustering als Heatmap zur Studie fC bindender Proteine in *epiblast like cells* (Kapitel 3.4.1). Für das Clustering wurden alle quantifizierten Proteine der Zeitpunkte d_0 , d_1 und d_2 verwendet. Der Anreicherungsfaktor gibt den zur Basis 2 logarithmierten Verhältnis von fC^{TMT127}/C^{TMT126} wieder, d.h. positive Werte (grün) stehen für fC angereicherte Proteine, negative Werte (blau) stehen für C angereicherte Proteine und ein Anreicherungsfaktor 0 (schwarz) steht für Proteine, die für keine der beiden Sonden angereichert wurden bzw. gar nicht identifiziert wurden. Rot markiert sind Proteine, die mit mindestens einem crosslinker-markiertem Peptid identifiziert wurden. Die Proteine unter den einzelnen Clustern I-XII sind in Tabelle A. 14 zu finden.

Anhang

Tabelle A. 14: Liste der Proteine des hierarchischen Clustering aus Abbildung A. 1

I	I	II	III	V	V	V	V	V	V	VI
4930451G09Rik	Rpl21	Pdha1	Pde8b	Actg2	Cwc15	Hsd17b12	Nup214	Sec61g	Uqcrc1	Rpl26
4933436I01Rik	Rpl23	Sod1	Pds5a	Agfg1	Cyb5a	Hsd17b4	Nup35	Selh	Usp32	Rpl29
Abtb1	Rps17	Ss18	Pds5b	Agps	Cyc1	Hsp90aa1	Nupl1	Sf1	Vapa	Rpl35a
Acin1	Rps8	Ssr1	Pex14	Ahctf1	Cyp2d11	Idh2	Nwd2	Sf3a2	Vapb	Rpl36
Adck1	Rsl1d1	Tas2r115	Plcb1	Ahnak	Cyp51a1	Idh3g	Pabpc4	Sf3b1	Wdr18	Rps16
Api5	Sdk2	Tdgf1	Ppp1r12a	Aifm1	Dctn2	Ik	Pam16	Slc17a8	Wdr5	Rps21
Arid4b	Sf3b4	Timm13	Prdm16	Ak2	Ddx3x	Kif19	Pcna	Slc26a3	Ybx3	Rrp1b
Bcas2	Slc3a2	Vamp3	Prpf6	Ak4	Ddx42	Kpnb1	Pcnp	Slc2a3	Yme11	Srfbp1
Brd4	Slit2		Rpl11	Akap12	Ddx47	Ktn1	Pdhx	Slc38a10	Ythdf3	Srp14
Casz1	Spata1	III	Rpl13	Ank2	Dnah2	Larp4	Pds5b	Smc1a	Ywhae	
Cdk1	Srp54c	Acin1	Rpl18	Anxa11	Dnaja1	Lin7c	Pfn1	Smc3	Ywhaq	VII
Chtf8	Srrm2	Actb	Rpl21	Apool	Dnaja2	Lman1	Pgrmc1	Snd1	Zfp128	Bsg
Cox5b	Tacc1	Afg3l2	Rpl36a	Arglu1	Dnaja3	Lman2	Phf5a	Snrpd1	Zfp473	Cyb5b
Creb3l2	Tapbp	Asap2	Rpn1	Armc1	Dnajc15	Lrpap1	Phf8	Snx12	Zfp873	Fbl
Csrp2	Tbcd	Atp5i	Rps15	Armc2	Dnajc8	Luc7l3	Pkdcc	Sod2		Gm6576
Dag1	Tma7	Rps26	Arpc4	Dock7	Map4	Map4	Polr1d	Spcs3	VI	Hnrrpa2b1
Dhtkd1	Tomm5	Ccdc65	Snrnp40	Arpc5	Echs1	Mapre1	Pou5f1	Sptbn1	Alpl	Hnrrpdl
Diaph2	Ttn	Cct8	Srp54	Bclaf1	Eif3c	Mast4	Pqbp1	Srrm2	Anln	Hnrrpfl
Eef1d	Utf1	Cntnap5b	Sucla2	Bdh1	Eif4a1	Mcm3	Prpf3	Srsf1	Atp5a1	Hnrrpnu
Elmo1	Vrtn	Col24a1	Syncrip	Bola3	Ep400	Mcm7	Prpf8	Srsf3	Atp5b	Krt1
Ep300		Ddx17	Top1	Brf1	Erp29	Mdc1	Psip1	Srsf7	Atxn2l	Krt2
Epb41l2	II	Dnah7b	Trpm3	Brix1	Fam207a	Mrpl11	Psmc4	Stip1	Dazap1	Ppib
Fam184b	Sept11	Dst	Ubtf	Btaf1	Fam227b	Mrpl27	Ptprb	Stoml2	Dek	Rpl18a
Fgfr1	2210010C04Rik	Ezh1	Uqcrc2	Cacybp	Fbxo38	Mrpl43	Ptprk	Strap	Dlst	Sltm
Flnb	Acpp	Fubp3		Casc4	Fh	Mrps11	Qil1	Sub1	Eef1a1	Snrgg
Fubp1	Aox4	Gar1	IV	Cast	Fkbp2	Mrps31	Qki	Supt16	Fau	
Gab1	Atp5f1	Gm10250	Cobl	Cbln4	Fkbp9	Mrps36	Rab1	Surf6	Fus	VIII
Glud1	Atp5l	Gm9493	Gatad2a	Ccdc12	Ftsj3	Mrps6	Rab11a	Taf15	Gm10036	Sept2
Gm8674	Bbx	Gnb2l1	Gm10094	Ccdc39	Fundc2	Mrto4	Rab7a	Taf4a	Gps1	Sept9
Gprin3	C8g	H1f0	H1fx	Ccdc66	Gatad2b	Msh2	Raly	Tardbp	Hist1h1b	Sept11
H2afx	Cdv3	H2afv	Itga6	Ccdc88a	Gfm1	Mta2	Rap1a	Tceb2	Hist1h1c	Acadm
Hnrrnp3	Cirbp	H3f3a	Mrrf	Cct4	Gm10032	Mtco2	Rbbp5	Tcerg1	Hist1h1d	Acsbg1
Hnrrnpm	Cox5a	Hist1h2bb	Nr2e1	Cct7	Gm10073	Mtx2	Rbfox2	Tcp1	Hist1h1t	Akap3
Ifitm3	Csrp1	Hist1h2bm	Pdia4	Cdc5l	Gm10093	Myh1	Rbm34	Tdh	Hist1h3a	Aldh18a1
Krt79	Dnajc19	Hist2h2ab	Pin1	Cdca8	Gm17093	Myh10	Rbm39	Tdpoz1	Hist1h4a	Aldh2
Ktn1	Eif3f	Hmga1	Ppp1r10	Cenpv	Gm17669	Myh9	Rbm7	Thyn1	Hmgn2	Aldh8a1
Mia2	Galnt13	Hmgb1	Rbm3	Cep250	Gm20390	Myof	Rbpj	Tial1	Hnrrpa1	Ap2b1
Mki67	Gm21698	Hmgb3	Rps7	Chd4	Gm4951	Nars	Rbpms	Tmed4	Hspa8	Ap2m1
Mrgprx2	Gm4861	Hnrrpa1	Sall4	Chdh	Gmnn	Ncoa6	Rbpms2	Tomm6	Igf2bp1	Asic5
Mrpl14	Golph3	Il18rap	Scp2	Cisd1	Gnl2	Ndufa13	Recql4	Tomm7	Ints12	Asph
Nbeal2	Hist1h2ab	Krt13	Smarcc1	Cisd2	Gon4l	Ndufa3	Rexo1	Tomm70a	Krt10	Atn1
Nlrc5	Hsp90ab1	Krt17	Snx3	Ckap2	Gosr1	Ndufa6	Rgs5	Top2a	Krt5	Atp2a2
Nop16	Hspa5	Krt75	Try10	Cnn2	H2afy	Ndufab1	Rif1	Trim71	Krt76	Atp4a
Notch2	Kdm5a	Lamp1	Tubb5	Cnot2	Hax1	Ndufs4	Rpa2	Trpc4	Mrpl12	Atp5g1
Nudt21	Khdrbs1	Lbr	C19orf43	Colgalt1	Heatr1	Ndufs6	Rpl19	Tsfm	Mthfd1l	Atp5j2
Paf	Mettl13	Mier1	C7orf50	Cox4i1	Hipk1	Nfrkb	Rpl37a	Tspan13	Ncl	Aurka
Pik3cd	Msx3	Mrpl30	Vasp	Cox7a2	Hist1h1t	Nfyc	Rpl38	Ttn	Nop2	Bag3
Pnpla7	Myh11	Mrpl40		Crebbp	Hist2h2aa1	Nifk	Rps10	Ttn	Npm1	Brd4
Poldip3	Nfu1	Mthfd2	V	Crip2	Hist3h2bb	Nip7	Rps12	Ttr	Pin4	Bub3
Rab11b	Nhp2	Mybbp1a	9530053A07Rik	Csnk2a1	Hk2	Nop16	Rps24	Txn	Rbmxl1	Cadps2
Rabep2	Nhp2l1	Nolc1	Acat1	Ctnna1	Hnrrpd	Nos1ap	Rps5	Ube2v2	Rpl10a	Canx
Rad50	Occ1	Nop56	Acin1	Ctnnb1	Hnrrpdl	Npm3	Rsu1	Ubqln1	Rpl22	Capzb
Rbm17	Pabpc1	Oxr1	Acly	Ctsd	Hnrrpl	Nras	Rtna	Ufm1	Rpl23a	Cbx5
Rcc2	Pbdc1	Pclo	Acot10	Ctsm	Hsd17b10	Nufip2	Sdf211	C8orf59	Rpl24	Ccbl2

Anhang

Fortsetzung Tabelle A. 14: Liste der Proteine des hierarchischen Clustering aus Abbildung A. 1

VIII	VIII	VIII	VIII	IX	XI	XII	XII	XII	XII
Ccpg1	Hint2	Pdhh	Smarca4	Pcbp2	Hist1h2bh	Ccdc116	Lmnb1	Rplp0	Vdac1
Cct2	Hmgb2	Pdia6	Smarcae1	Ptbp1	Hnrnpc	Cd63	Lrrc59	Rps11	Vdac2
Cct3	Hnrnpa3	Picalm	Smc2	Rpl22l1	Hspa8	Cd9	Lyar	Rps12	Vdac3
Cfl1	Hnrnpab	Plekha6	Sms	Sox2	Idh3b	Cirbp	Lzts1	Rps13	Vim
Chchd2	Hnrnph1	Poldip3	Snap23	Srsf2	Immt	Ckap4	M6pr	Rps14	Ylpm1
Chchd3	Hnrnpm	Poldip3	Snrnp70	Sumo2	Ints5	Cpsf6	Map4	Rps15	Ywhaq
Cherp	Hsp90b1	Polr2k	Snrpa1	Tcea1	Klrc2	Crem	Mbd2	Rps18	Znf207
Chtop	Hspb1	Ppan	Snrpf		Kmt2e	Cs	Morf4l2	Rps19	Znf706
Clec3a	Ifitm2	Ppp1r18	Snw1	X	Ktn1	Cstf2	Mst1r	Rps19bp1	
Coro1c	Igf2bp3	Prdx1	Spata5	C1qbp	Mdh2	D8ErtD738e	Mtstp8	Rps20	
Cpt1a	Ikzf3	Prkcs	Sptan1	Cald1	Myef2	Ddx18	Mtdh	Rps23	
Ctrc	Ilf2	Prpf19	Srrt	Caprin1	Myo3b	Ddx21	Myef2	Rps24	
Cyb5r3	Ins1	Psrc1	Srsf5	Chmp1a	Phb2	Ddx26b	Myl12b	Rps25	
Ddost	Kalrn	Rac3	Srsf6	Crem	Ppia	Ddx39b	Myl6	Rps28	
Ddx46	Khsrp	Ran	Ssrp1	Ddx39a	Rbm27	Ddx5	Nav1	Rps3a	
Dhodh	Kif1b	Ranbp2	Suclg1	Eif5a	Rpl13a	Dnajc19	Ndufv2	Rps6	
Dhx15	Kpna2	Rbbp4	Tcf20	Gm10108	Samd14	Dnajc8	Nek6	Rrbp1	
Dhx9	Krt10	Rbm10	Tcof1	Gm10134	Sap130	Eef1b	Nol11	Rrp1	
Dlat	Krt42	Rbm14	Tet2	Golm1	Sarnp	Eif2a	Nono	Rybp	
Dnajb6	Krt73	Rbm28	Tgoln1	Gtf2e2	Slc25a4	Eif3a	Npm1	Sap130	
Dnttip2	Lamc1	Rcc1	Thrap3	Hist2h2be	Snrnp35	Elavl1	Nup50	Serbp1	
Dynll2	Lamp2	Rexo4	Tiam2	Hyou1	Snrpn	Erb2ip	Nup62	Serpinh1	
Ebna1bp2	Lasp1	Rpl10l	Timm23	Immt	Sort1	Ewsr1	Parp1	Sf3a1	
Eci1	Lima1	Rpl28	Tmcc2	Pspc1	Srsf7	Exosc10	Pcbp1	Shmt2	
Eef1	Marcks11	Rpl32	Tmpo	Rpl13a	Suclg2	Fam110d	Pdcd6ip	Shq1	
Eef2	Matr3	Rpl6	Tmpo	Rpl27a	Tomm5	Fam136a	Pdia3	Skida1	
Eif2s2	Mcm2	Rpl9	Tomm40	Rplp2	Tpt1	Fkbp3	Phb	Slc25a5	
Eif6	Mcm4	Rps10	Tpm2	Rpsa	Trrap	Fubp1	Pitpnm3	Slc38a10	
Emd	Mcm5	Rps15a	Tpr	Ssb	Ush2A	G3bp1	Pitrm1	Smarcal1	
Eno3	Mdn1	Rps27a	Trap1	Tbc1d8b	Utrn	Gapdh	Pms1	Snap23	
Etfa	Meaf6	Rps27l	Trip12	Vcp	Wbp11	Gm10126	Prom2	Snrpa	
Etfb	Mesdc2	Rps29	Tuba1c	Ybx1	Ywhaq	Gm15013	Prrc2c	Snrpb	
Fam114a1	Mfap1	Rps3	U2af2			Grpel1	Pum2	Snrpb2	
Fip1l1	Mp68	Rps4x	Ube2i	XI	XII	H2afj	Raly	Snrpc	
Fig2	Mrpl3	Rps9	Ube2n	Aco2	1810009A15Rik	H2afy	Rbm27	Snrpd2	
Fscn2	Mrpl41	Rrp15	Ubxn6	Alb	1810022K09Rik	Haus4	Rbmx12	Snrpd3	
Ftl1	Mrps22	Rrs1	Unc13c	Aldoa	Acaca	Hcfc1	Rfx1	Snrpe	
Fyttl1	Myef2	Rsg1	C1orf174	Atp5a1	Adgb	Hells	Rif1	Sp5	
G3bp2	Ndn12	Rttn	Uqcrf51	Atp5c1	Akap12	Hist1h1a	Rpl12	Ssbp1	
Galnt2	Ndufa2	Ruvbl1	Utp11l	Cggbp1	Alyref	Hist1h1e	Rpl14	Taf15	
Ganab	Ndufa7	Ryr1	Wdr91	Cstf2t	Aqr	Hmga1	Rpl17	Tcerg1	
Gja1	Ndufb10	Sap30	Ywhaz	D1Pas1	Armcx4	Hmgn1	Rpl27	Tex10	
Gltscr2	Nedd4	Sav1	Zfr	Dock6	Aspm	Hnrnpa0	Rpl3	Tex261	
Gm10320	Nipsnap1	Scaf11	Znf593	Eef1g	Atp5d	Hnrnpc	Rpl30	Tfam	
Gm11360	Nol7	Sco1	Znf638	Eif4a2	Atp5e	Hnrnpd	Rpl31	Tmed9	
Gm6104	Nova2	Sdha		Emc7	Atp5j	Hnrnpk	Rpl34	Top2b	
Gm9755	Nt5c1a	Sdhb	IX	Ercc4	Atp5o	Hspa9	Rpl35	Tpr	
Golga5	Nup153	Setd3	Cenpe	Fau	Atpif1	Hspd1	Rpl37	Trim28	
Gosr2	Nup54	Sf3a3	Erh	Gltscr1	Bcap31	Hspe1	Rpl4	Trim35	
Got2	Nup98	Sf3b2	Gm10020	Gm3839	Bsg	Htt	Rpl5	Ttc37	
Gpr89a	Osgepl1	Sf3b6	Gtf2f2	Gm6970	Btf3	Ilf3	Rpl7	Ttl5	
Gtpbp4	P4hb	Sfpq	Ndufa4	Gm8730	Calr	Itpr2	Rpl7a	Ubap2	
H2afy2	Pak1	Slc12a5	Nmt1	Grin3b	Cbx3	Krt15	Rpl8	Ubap2l	
Hars	Pde8b	Slc25a1	Pabpn1	Hao2	Ccar1	Krt78	Rpl9	Uqcrh	

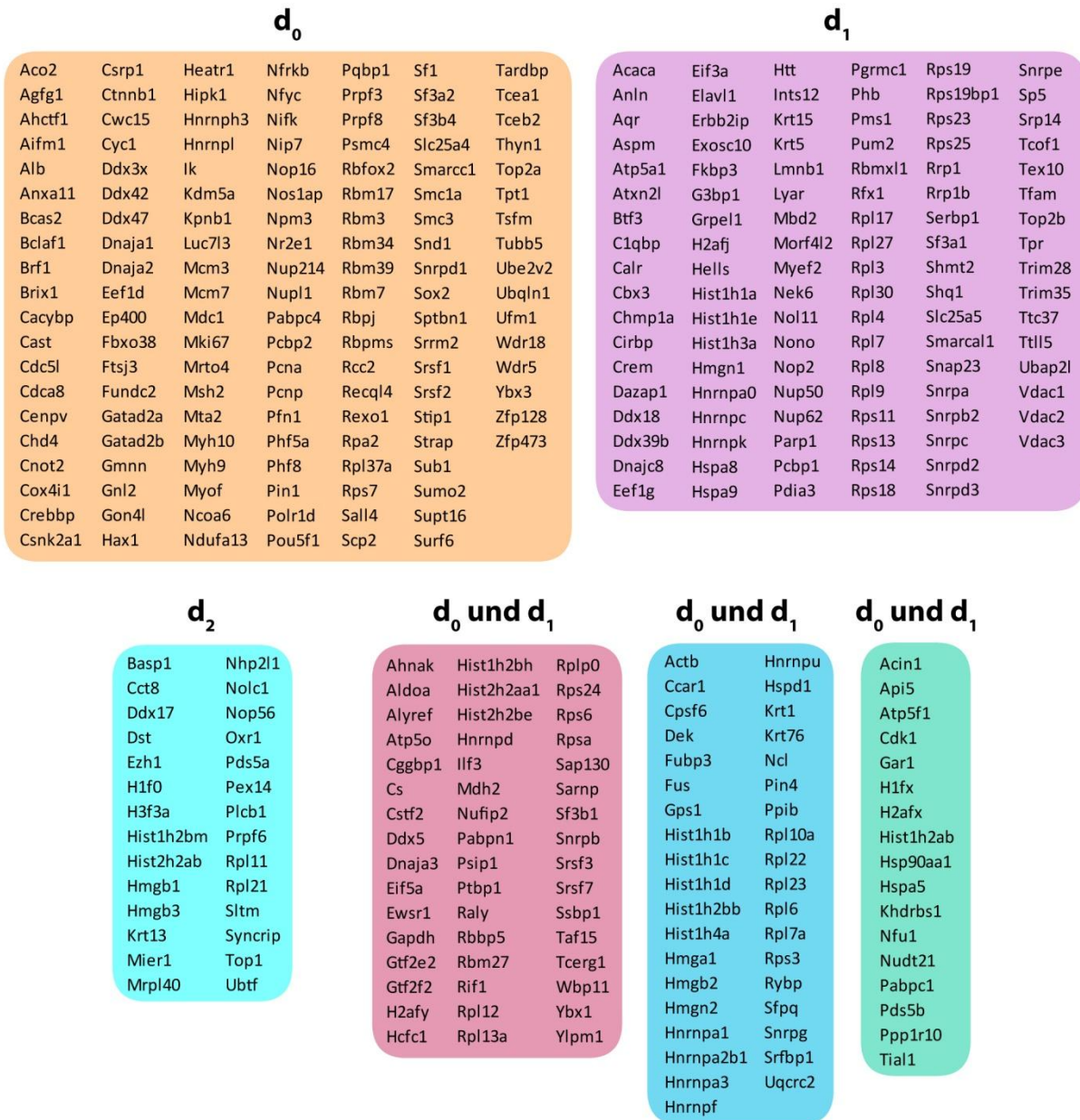


Abbildung A. 2: Ergänzung zum Venn-Diagramm (Abbildung 33) aus Kapitel 3.4.1 (fc-Crosslinking-Interaktionsstudie).

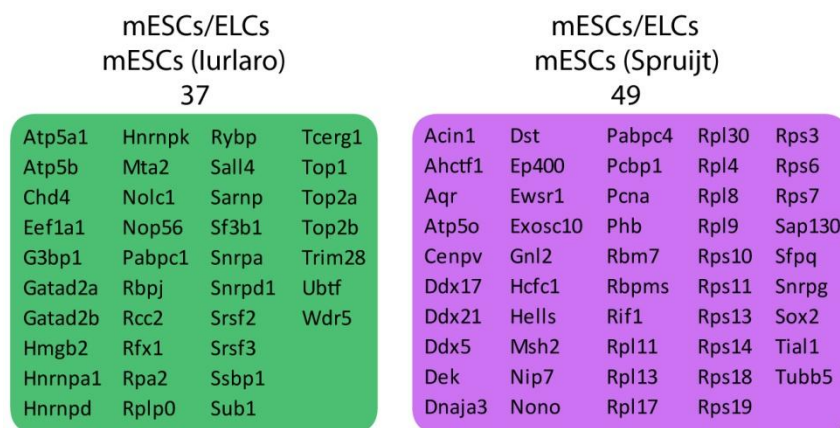


Abbildung A. 3: Ergänzung zum Venn-Diagramm (Abbildung 40) aus Kapitel 3.4.2 (Vergleich mit publizierten fc Daten)

Accession	Genname	Beschreibung	# AAs	MW [kDa]	ΣCoverage	Σ# Unique Peptides	Σ# Peptides	Σ# PSMs	A5: 129/129 Count	A5: 129/129 Variability [%]	A5: 130/129 Count	A5: 130/129 Variability [%]	A5: 131/129 Count	A5: 131/129 Variability [%]	
P62806	Hist1h4a	Histone H4	103	11,4	100,00	12	49	638	1,000	0,0	1,005	33,4	1,030	42	17,2
P43274	Hist1h1e	Histone H1.4	219	22,0	100,00	1	120	3935	1,000	0,0	0,872	42,6	0,919	123	26,0
POC056	H2afz	Histone H2A.Z	128	13,5	100,00	1	40	527	1,000	0,0	0,849	21,7	0,938	49	29,6
P10922	H1f0	Histone H1.0	194	20,8	100,00	1	102	2495	1,000	0,0	0,796	23,6	0,944	79	23,1
P43276	Hist1h1b	Histone H1.5	223	22,6	100,00	4	122	3823	1,000	0,0	0,862	26,2	0,906	126	23,3
P43275	Hist1h1a	Histone H1.1	213	21,8	99,53	4	115	3241	1,000	0,0	0,763	36,9	0,951	148	20,6
F8W135	H3f3a	Histone H3	135	15,2	99,26	13	53	711	1,000	0,0	0,781	19,4	1,063	41	38,8
P43277	Hist1h1d	Histone H1.3	221	22,1	99,10	3	114	3044	1,000	0,0	0,812	32,9	0,905	139	20,2
Q8CGP7	Hist1h2ak	Histone H2A type 1-K	130	14,1	97,69	3	38	271	1,000	0,0	0,843	47,4	1,133	23	47,0
Q64523	Hist2h2ac	Histone H2A type 2-C	129	14,0	97,67	2	38	613	1,000	0,0	0,818	26,9	1,157	19	33,0
Q8CGP0	Hist3h2bb	Histone H2B type 3-B	126	13,9	97,62	2	47	1339	1,000	0,0	0,812	41	0,966	41	23,2
Q8CGP2-2	Hist1h2bp	Isoform 2 of Histone H2B type 1-P	138	15,6	97,10	2	56	1655	1,000	0,0	0,838	45	1,000	45	23,9
P10854	Hist1h2bm	Histone H2B type 1-M	126	13,9	96,83	2	53	1474	1,000	0,0	0,838	45	0,974	45	22,2
Q8CGP1	Hist1h2bk	Histone H2B type 1-K	126	13,9	96,83	3	49	1138	1,000	0,0	0,838	45	0,966	45	21,7
P27661	H2afx	Histone H2AX	143	15,1	95,80	4	39	388	1,000	0,0	0,835	38	1,086	38	34,5
Q9DSZ7	Subh2bv	Histone H2B subacrosomal variant	123	14,2	95,12	1	23	123	1,000	1	0,739	1	0,695	1	
Q8VK3	H1f0o	Histone H1.0o	304	32,2	94,41	3	116	2559	1,000	0,0	0,838	66	0,919	66	22,9

Tabelle A. 15: Identifizierte und quantifizierte Histone aller Zeitpunkte aus der Studie zu Histonmodifikationen (Kapitel 3.4.3).

Tabelle A. 16: Ergänzung zu den Daten aus Kapitel 3.4.3: Quantifizierte Modifikationen der Histon 1 Isoform H1oo.

Modifikationen	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
K42me3 / K49me3 / K50ac	1,000	0,708	1,110
K47me2 / K48me1 / S49P	1,000	0,562	1,428
K47me3 / K48me3 / K54me3	1,000	1,226	1,226
K47me3 / K48me1 / K54ac / K55ac	1,000	1,086	1,056
K47me3 / K48me3 / K54ac	1,000	1,036	0,690
N-TermTMT6plex / K47me2 / K48me2 / K55me1	1,000	0,529	0,909
N-TermTMT6plex / K47me3 / K48me1 / K54me1	1,000	0,529	0,909
N-TermTMT6plex / K59me3 / T61P / K64me1	1,000	1,072	0,795
K59me3 / M62O / K64me3 / K67ac	1,000	0,994	0,717
N-TermTMT6plex / K59me1 / K64me1 / K67me2	1,000	0,941	0,705
N-TermTMT6plex / K59me1 / K64me3	1,000	0,882	0,737
N-TermTMT6plex / K71me1	1,000	0,986	1,387
K91me1	1,000	0,664	0,804
N-TermTMT6plex / K110me3	1,000	0,793	0,617
K112me2 / S117P / K119me2 / K123me1	1,000	1,051	1,071
K112me2 / S117P / K119me1	1,000	0,597	0,796
N-TermTMT6plex / S117P / K125ac	1,000	0,696	0,886
K123ac / K125ac / T126P / K127me2	1,000	2,147	0,686
K123ac / K125ac / T126P / K128me2	1,000	2,147	0,686
K123me3 / K125me1 / T126P / K127me2	1,000	0,922	1,417
K127ac / C130C	1,000	1,149	0,744
K128me3 / C130C / K133ac	1,000	0,772	0,952
K128ac / C130C / K133me1	1,000	1,318	1,016
K128me1 / C130C / K133ac	1,000	1,318	1,016
K128me2 / C130C	1,000	0,904	0,779
K128me1 / C130C / K133me1	1,000	1,023	0,782
N-TermTMT6plex / C130C / K142me3	1,000	0,836	0,949
N-TermTMT6plex / K142me3 / K147ac	1,000	0,840	0,788
N-TermTMT6plex / K142me2 / K147ac / K148ac	1,000	0,799	0,706
S149P / K154ac / M162O	1,000	1,529	0,923
K153me1 / K154me1 / K159me1 / K164me2	1,000	0,688	0,551
K154ac / K159ac / K164ac	1,000	0,851	0,676
K154me1 / K159me1 / K164me3	1,000	0,326	0,637
K154me2 / K159me2 / K164ac / K167ac	1,000	1,092	0,632
K154me2 / K159me3 / K164me3 / K167me1	1,000	0,953	0,914
K154ac / K167me3	1,000	0,888	0,771
K159me3 / K164me2	1,000	0,698	0,604
N-TermTMT6plex / M162O / K164me1	1,000	0,874	1,656
N-TermTMT6plex / M162O / K167me1	1,000	0,874	1,656
K164me3 / K167ac	1,000	0,868	0,770
Y171ac / C173C / K174me2 / M179O / K182me2	1,000	0,922	0,607
K173me3 / K174me2 / K176me1	1,000	0,886	1,002
K190me1 / K194me1 / K198ac	1,000	0,459	0,518
K204ac / K206me3	1,000	1,276	0,907
K229me1 / K231me3	1,000	0,845	0,868
K231me1 / K234me3 / K239ac	1,000	0,910	1,582
K234ac / K239ac / K244ac	1,000	0,924	0,696
K234me2 / S235P / K236ac / K241me1	1,000	0,621	1,212
K234me1 / K239ac / K241me3	1,000	0,812	1,148
K234me3 / S235P / K236me3 / K241me1	1,000	0,628	1,058
S235P / K241ac / K251me1	1,000	0,689	1,264
K239ac / K241ac	1,000	0,757	0,683
K241me3 / K251me1	1,000	1,019	0,734

Anhang

Tabelle A. 17: Ergänzung zu den Daten aus Kapitel 3.4.3: Quantifizierte Modifikationen der Histon H1 Isoform H1.0.

Modifikationen	d_0/d_0	$(d_1/d_0)/0,956$	$(d_2/d_0)/1,0445$
S18P / K19me2 / S21P / K26me3	1,000	1,041	0,807
Y27ac / S28P / K39ac	1,000	1,113	0,806
N-TermTMT6plex / K51me2 / K54me3	1,000	1,110	0,833
T77P / K79me3 / K81ac / K83ac	1,000	0,547	0,999
K79ac / K81ac / S84P / K87me2	1,000	0,501	0,559
K79ac / K81ac / S85P / K87me2	1,000	0,501	0,559
K79me1 / S84P / K87ac	1,000	1,859	1,107
N-TermTMT6plex / K81me2 / K87ac	1,000	0,873	1,128
K84ac / K96ac	1,000	1,038	0,656
K84me1 / S89P	1,000	1,051	0,917
N-TermTMT6plex / K84ac / K96ac	1,000	1,002	1,019
N-TermTMT6plex / K84me1	1,000	1,002	0,894
T83P / K84me1	1,000	1,051	0,917
K96me1	1,000	0,661	0,871
S103P / K108me2 / T109P	1,000	0,881	0,740
K108me1 / T109P	1,000	0,709	0,865
K108me2 / T109P / K110me3	1,000	0,927	0,719
T109P / K110me1	1,000	0,709	0,865
K111me2 / K115me3 / T118P / K136me3	1,000	0,812	0,823
K114me2 / K115me3 / T118P / K136me3	1,000	0,865	0,952
N-TermTMT6plex / K114me1 / T118P / K136ac	1,000	0,832	1,014
K114ac / K120me1	1,000	0,801	0,664
K114me2 / K115me3 / T118P / K120me3	1,000	0,925	1,033
K114me3 / K115me3 / T118P / K120me2	1,000	0,892	1,090
N-TermTMT6plex / K114me2 / K115ac / T118P	1,000	0,293	1,249
N-TermTMT6plex / K115me1 / T118P / K136ac	1,000	0,832	1,014
N-TermTMT6plex / K115ac / K120ac / K121me3	1,000	0,912	1,041
K120ac / K121ac / K124ac / K126me3	1,000	0,716	0,755
K120me3 / K121ac / K124ac / K126ac	1,000	0,649	0,756
N-TermTMT6plex / K131ac	1,000	0,839	0,757
K127me2 / K131me2 / K135ac / K136ac	1,000	1,148	0,788
K127me1 / K131me3 / K135ac / K136ac	1,000	1,358	0,972
N-TermTMT6plex / S130P / K131me2 / K136ac	1,000	0,873	0,833
K131ac / K138me1 / K143me1	1,000	0,799	0,948
K131ac / S134P / K135ac / K143me3	1,000	0,864	1,449
K131me3 / K135me2 / K138ac / K143ac	1,000	0,821	0,469
S134P / K135me3 / K138ac / K143ac	1,000	0,927	1,373
S134P / K135ac / K136me2 / T140P	1,000	0,639	1,352
S134P / K135me2 / K136ac / T140P	1,000	0,639	1,352
K135ac / K136ac / K138me1 / T140P	1,000	1,232	0,857
K135ac / K136ac / K138me1	1,000	0,800	0,987
K135me2 / K136ac / K138ac / K143me1	1,000	0,673	2,272
K135me1 / K138ac / T140P / K143ac	1,000	1,321	0,804
K136ac / K138ac / K143ac	1,000	0,725	1,295
K138me2 / T140P / K143me3	1,000	1,000	0,859
K147me2 / K148ac / K154ac	1,000	0,930	0,775
K147ac	1,000	0,875	0,703
K147ac / K153ac / K154me1	1,000	0,630	1,155
K147me1 / K153ac / K154ac	1,000	0,632	1,057
N-TermTMT6plex / T151P / K153me1 / K156me3	1,000	0,949	0,980
K157ac	1,000	0,947	0,750
N-TermTMT6plex / K179ac / K181ac / K187me3	1,000	1,138	0,832
S179P / K180me3 / K182me3 / K183me3	1,000	0,954	0,850
K183me3 / K185me3 / T186P / K1188me3	1,000	1,052	0,972
K183me1 / K187me3 / K191me3	1,000	1,552	1,181

Tabelle A. 18: Ergänzung zu den Daten aus Kapitel 3.4.3: Quantifizierte Modifikationen der Histon H1 Isoform H1.1.

Modifikationen	d ₀ /d ₀	(d ₁ /d ₀)/0.956	(d ₂ /d ₀)/1.0445
K22ac / K24ac / K25me3 / K28me3	1,000	0,763	1,022
K22me3 / K24me3 / K25ac / K28ac	1,000	0,893	0,805
N-TermTMT6plex / S59P / K64me1 / K65me1	1,000	1,082	0,781
K65me3 / S66P / Y72ac / K76me3	1,000	0,964	0,638
K65me1 / Y72ac / K76me1	1,000	1,336	0,680
S79P / K82ac / K86me2 / K91me2	1,000	0,618	0,748
S79P / K82ac / K86me3 / K91me1	1,000	0,657	0,705
N-TermTMT6plex / K86me2 / S87P / K91me2	1,000	0,952	0,597
N-TermTMT6plex / K110me2 / K111me1 / K120me3	1,000	0,717	0,847
N-TermTMT6plex / K110me1 / K111me2 / K120me3	1,000	0,717	0,847
K111ac / K115me3 / K120me3	1,000	0,992	1,016
K111me2 / K115me1	1,000	1,264	1,701
K111me2 / K115me3 / K120ac	1,000	0,761	1,109
K111me1 / T119P / K120ac	1,000	0,297	0,681
N-TermTMT6plex / S114P / K115ac / K120me1	1,000	0,773	0,807
N-TermTMT6plex / S114P / K115me1 / K120ac	1,000	0,536	0,663
S114P / K120me3	1,000	1,015	1,314
K115me2 / K120me2 / S122P / K124me2	1,000	0,980	0,794
K115me1 / T119P / K120ac	1,000	0,392	0,738
K115me3	1,000	1,658	1,683
N-TermTMT6plex / K115ac / T119P / K120me1	1,000	0,598	0,802
N-TermTMT6plex / K115me2 / K120me1	1,000	2,372	0,866
N-TermTMT6plex / K115me1 / T118P / K120ac	1,000	0,606	0,737
N-TermTMT6plex / K115me1 / T119P / K120ac	1,000	0,486	0,678
K115me2 / K120me3	1,000	1,174	1,063
N-TermTMT6plex / K115ac	1,000	0,698	1,111
K115me2 / K120ac / K124me1 / K126me1	1,000	0,772	0,921
T117P / K123ac / K125ac	1,000	0,728	1,377
T117P / K125me1 / K131me2	1,000	0,925	1,422
T117P / K119ac / K131me2	1,000	0,725	1,269
T117P / T118P / K131me2	1,000	0,724	0,712
T118P / K125me1 / K131me2	1,000	0,925	1,422
T118P / K119ac / K131me2	1,000	0,725	1,269
T118P / K119me3 / K123me3 / K131me2	1,000	0,682	0,899
T118P / K120me3	1,000	1,015	1,314
T118P / K123ac / K125ac	1,000	0,728	1,377
K119me3 / S121P / K123me3 / K131me2	1,000	0,799	0,944
K120me3	1,000	1,046	1,053
K120me3 / K124me1	1,000	0,911	1,212
N-TermTMT6plex / K120me2 / K124me1	1,000	2,114	1,028
N-TermTMT6plex / K120me1 / K124me2	1,000	1,822	1,064
N-TermTMT6plex / K120me3	1,000	2,004	1,036
K120me3 / S122P / K124ac / K126me2	1,000	0,815	0,871
S122P / K124ac / K126ac / K132ac	1,000	0,408	1,534
N-TermTMT6plex / S122P / K132me1	1,000	0,503	1,127
N-TermTMT6plex / S122P / K133me1	1,000	0,457	1,025
K123ac / K125ac / K131me1	1,000	0,621	1,149
K123me1 / K125ac / K131ac	1,000	0,596	1,040
K123me3 / K125me3 / S127P / K131me2	1,000	0,772	1,184
K124me2 / K126ac / S128P / K132me1	1,000	1,024	0,821
K124me2 / K126ac / S128P / K133me1	1,000	1,098	0,840
K124me2 / S128P / K132ac / K133me1	1,000	1,002	0,962
K124me1 / K126ac / S128P / K133me2	1,000	0,962	1,335
S128P / K132me1	1,000	0,363	0,576
S128P / K133me1 / K135ac	1,000	0,856	1,006
K132me2	1,000	0,761	0,811
K132ac / K135me3 / K136me2 / T137P	1,000	0,780	0,892
K132me1 / K133me1 / T137P	1,000	0,755	1,137
K133ac / K135me3 / K136me2 / T137P	1,000	0,780	0,892
T137P / K143me2 / K144me3	1,000	1,051	0,719
K143me2	1,000	1,078	1,051
N-TermTMT6plex / K143ac / K144me2	1,000	0,871	0,806
N-TermTMT6plex / K143me2 / K144ac	1,000	0,871	0,806
K143me1 / K147me2	1,000	0,611	1,508
N-TermTMT6plex / K143me1 / K144ac / K147me2	1,000	0,671	0,759
N-TermTMT6plex / K143me2 / T145P / K147me3	1,000	1,167	1,001
N-TermTMT6plex / K143me3 / K147ac / T148P	1,000	0,551	0,736

Anhang

Fortsetzung Tabelle A. 18: Ergänzung zu den Daten aus Kapitel 3.4.3: Quan. Modifikationen der Histon H1 Isoform H1.1.

Modifikationen	d ₀ /d ₀	(d ₁ /d ₀)/0.956	(d ₂ /d ₀)/1.0445
K144ac / K147me1 / T148P	1,000	0,667	0,608
N-TermTMT6plex / K144me3 / K147ac / T148P	1,000	0,458	0,809
N-TermTMT6plex / K144me3 / T145P / K147me1	1,000	0,711	1,205
N-TermTMT6plex / K144me3 / T145P / K150me1	1,000	0,711	1,205
K144me1 / K147me1	1,000	0,863	1,552
N-TermTMT6plex / T145P / K147me3 / T148P	1,000	0,897	0,795
K147me3 / T148P / K153ac	1,000	0,876	0,460
K147me3 / T148P / K151ac	1,000	0,876	0,460
T148P / K150me2 / K151me2	1,000	1,213	0,714
T148P / K150me3 / K151me1	1,000	1,213	0,714
K150ac / K151ac / K153ac	1,000	0,524	0,873
K150me2 / K153ac / K154ac / K159me1	1,000	0,600	0,750
K153ac / K160me3 / S162P / K163me3	1,000	0,664	1,027
K153me1 / S158P / K160ac / S162P	1,000	0,756	1,217
K153me1 / S158P / S162P / K163ac	1,000	0,756	1,217
K154ac / K160me3 / S162P / K163me3	1,000	0,664	1,027
K154me1 / S158P / K160ac / S162P	1,000	0,756	1,217
K154me1 / S158P / S162P / K163ac	1,000	0,756	1,217
K154me1 / S158P / K159ac	1,000	0,997	1,012
K154me1 / S158P / K160ac	1,000	0,997	1,012
S158P / K159me3 / K160me3 / K163ac	1,000	0,632	1,097
N-TermTMT6plex / S158P / K163ac / K166me3	1,000	0,729	1,442
N-TermTMT6plex / S158P / K160ac / K166me3	1,000	0,703	1,320
K159me2 / T161P / S162P / K163me1	1,000	0,903	0,998
K159me1 / K160me1 / T161P	1,000	1,923	1,004
K159ac	1,000	2,051	0,773
N-TermTMT6plex / K159ac / K160ac / T161P	1,000	1,350	1,344
N-TermTMT6plex / S162P / K163me2	1,000	0,746	0,743
N-TermTMT6plex / K163ac	1,000	0,890	1,236
K163me1 / K166ac	1,000	1,471	1,151
N-TermTMT6plex / K163me2 / K166me3	1,000	0,751	0,898
N-TermTMT6plex / K163me1 / K166me2	1,000	0,647	0,934
N-TermTMT6plex / K163me3	1,000	0,682	1,030
N-TermTMT6plex / K163me3 / K166me2	1,000	0,851	0,999
K163ac / K166ac / K169me1	1,000	0,511	0,867
K163ac / K166ac / K167me2	1,000	1,095	1,719
K163ac / K166ac / K167me1	1,000	0,511	0,867
K163me1 / K167ac / K169ac	1,000	0,674	0,909
K166me2 / K168ac / K171me3	1,000	0,627	0,792
K166me2 / K168ac / K173me3	1,000	0,580	0,825
N-TermTMT6plex / K172me1	1,000	0,832	0,707
N-TermTMT6plex / K175ac / K178ac	1,000	0,630	0,515
K178me3 / S179P / K182ac / K184ac	1,000	1,026	0,786
K182ac / K184ac / S191P	1,000	0,687	1,032
K184me1 / K187ac / K189ac / K192ac	1,000	0,914	1,957
K184ac / K187ac / K189ac / K192me2	1,000	1,452	0,954
N-TermTMT6plex / K184ac / K187ac / S191P	1,000	0,906	0,755
K187ac / K189ac / K192ac / K194me1	1,000	0,745	1,021
K187ac / K189ac / K192me1 / K194me1	1,000	0,897	0,851
K187ac / K189ac / K194me2	1,000	0,897	0,851
K187ac / K189me2 / S191P / K192me3	1,000	0,474	0,686
K187ac / K189me3 / S191P / K192me2	1,000	0,474	0,686
K194ac / K197ac / K199ac	1,000	1,068	0,895
N-TermTMT6plex / T196P / K200ac / K204ac	1,000	1,369	0,742
T196P / K197me1 / K199ac / T200P	1,000	0,891	1,137
T196P / K197me1 / T200P / K205me2	1,000	0,829	1,049
T196P / K197me1 / T200P / K203me2	1,000	1,246	1,019
T196P / K199me1 / T200P / K205me2	1,000	1,072	0,980
T196P / K199me1 / T200P / K203me2	1,000	1,246	1,019
K197ac / K199ac / K203ac / K205me1	1,000	0,972	1,962
K197ac / K199ac / K203ac / K206me1	1,000	0,958	1,333
K199me2 / T200P / K203ac / K205me3	1,000	1,189	0,747
K199me2 / T200P / K203me3 / K205ac	1,000	1,189	0,747
K203ac / K205ac / K206me2	1,000	0,652	1,102
K203ac / K205ac / K206me3	1,000	0,861	0,965
K203me3 / K205ac / K206ac	1,000	0,861	0,965
K203me3 / K205me3 / K206ac / K210ac	1,000	1,064	0,848
K203me1 / K206ac / K210ac / K211ac	1,000	0,863	0,885

Tabelle A. 19: Ergänzung zu den Daten aus Kapitel 3.4.3: Quantifizierte Modifikationen der Histon H1 Isoform H1.3.

Modifikationen	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
K24me2 / K25me3 / K32me3	1,000	0,635	0,741
K25ac / T26P	1,000	0,658	0,479
K25me1 / K32ac	1,000	0,777	0,957
N-TermTMT6plex / T45P / K46me2	1,000	0,938	0,751
K52ac / K63me1	1,000	0,844	0,572
K52me1 / K63ac	1,000	0,844	0,572
N-TermTMT6plex / S58P / K63me1 / K64me1	1,000	1,082	0,781
S78P / K81me3 / K85ac / K90me1	1,000	1,199	1,431
K85ac / S89P / K90me1	1,000	0,868	0,886
K85ac / K90me3 / T92P	1,000	0,767	1,250
K85me2 / K90me1 / T92P / K97ac	1,000	0,557	1,020
K85me2 / S86P / S89P / K90me1	1,000	1,163	0,843
K85me2 / S89P / K90me1 / K97ac	1,000	0,671	1,039
K85me1 / S86P / K90me3	1,000	1,006	0,801
N-TermTMT6plex / T99P / K106me3 / K109ac	1,000	0,764	0,678
S104P / K106me2 / K109me1 / K110ac	1,000	0,923	0,800
S104P / K106me3 / K110ac	1,000	0,858	0,799
K106me1	1,000	0,624	0,849
N-TermTMT6plex / K109ac / K110me2	1,000	0,541	0,823
N-TermTMT6plex / K109me2 / K110ac	1,000	0,601	0,887
K109ac / K110me2 / K117me1	1,000	0,661	1,098
K109ac / K110me2 / K119me1	1,000	0,636	1,074
K109me2 / K110ac / K117me1	1,000	0,661	1,098
K109me2 / K110ac / K119me1	1,000	0,636	1,074
K110me3 / K117me2 / K119me2	1,000	0,890	1,444
S113P / K121me3 / K127ac	1,000	0,766	0,818
K119ac / K121ac / K122me3 / K127me3	1,000	0,630	0,773
N-TermTMT6plex / K121me1	1,000	1,287	1,319
N-TermTMT6plex / K121me2 / K127me3	1,000	0,798	0,890
K122ac / K127ac / K129me1	1,000	1,011	0,857
K127me2 / K129me2 / K136ac	1,000	0,899	1,984
K127me2 / K129me2 / K130ac	1,000	0,842	1,478
K127me3 / K129me3 / K137ac / K139ac	1,000	4,273	1,043
K129ac / K130me1 / K136me3	1,000	0,749	1,495
K129me1 / K130ac / K136me3	1,000	0,592	1,431
K129me3 / K130me3 / K136ac / K137ac	1,000	2,555	0,987
K129me3 / K130me3 / K137ac / K139ac	1,000	3,425	1,036
K130me3 / K136me3 / K137ac / K139ac	1,000	2,696	0,990
K130ac / K136ac / K140me1	1,000	0,457	0,896
K136ac / K137me1	1,000	0,638	0,892
K136me1 / K137ac	1,000	0,638	0,892
K139ac / K140ac / T142P / K148me2	1,000	0,930	0,912
K139me2 / K140ac / T142P / K148ac	1,000	1,013	0,727
N-TermTMT6plex / K140me1 / T142P / K148me3	1,000	0,664	0,931
K140me1 / T146P / K149ac / K152ac	1,000	1,005	0,860
N-TermTMT6plex / T146P / K149ac / K152me1	1,000	0,850	0,765
T146P / K148ac / K149ac / K152me1	1,000	0,828	0,817
T146P / K148ac / K152me2	1,000	0,940	0,897
T146P / K148me1 / K149ac / K152ac	1,000	0,864	0,870
K148me2 / K149ac / T150P / K152me3	1,000	0,501	0,559
N-TermTMT6plex / K148ac / K149ac / K152ac	1,000	0,712	1,053
N-TermTMT6plex / K148me1 / K149ac / K152me2	1,000	1,104	0,807
N-TermTMT6plex / K148me3 / K149ac	1,000	1,104	0,807
N-TermTMT6plex / K148me1	1,000	0,622	0,632
K152me3 / K153ac / K156ac	1,000	0,980	0,741
N-TermTMT6plex / K152me3 / K153ac	1,000	1,093	0,953
T154P / K160me1 / K168ac	1,000	0,843	0,655
N-TermTMT6plex / K157me1	1,000	0,503	0,504
K158ac / K159ac / K167ac	1,000	0,872	0,585
K158ac / K159ac / K167ac	1,000	1,032	0,996
K158ac / K159ac / K166me2 / K169me1	1,000	1,074	0,767
K158ac / K159ac / K166me1 / K168me2	1,000	1,319	0,843
K158ac / K159ac / K166me1 / K169me2	1,000	1,319	0,843
K158me2 / K159ac / K165ac	1,000	0,720	0,801
K159me1 / K165me1 / K168ac / K169ac	1,000	0,698	0,734
K160me1 / K168me2 / K169ac / S171P	1,000	0,846	0,821
K169ac / S171P / K172ac / K175ac	1,000	0,887	0,589

Fortsetzung Tabelle A. 19: Ergänzung zu den Daten aus Kapitel 3.4.3: Quan. Modifikationen der Histon H1 Isoform H1.3.

Modifikationen	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
K172me1 / K175me3	1,000	0,947	0,813
K175me1	1,000	0,635	0,741
K184me3 / K187ac / K191ac	1,000	1,736	1,142
K184ac / K191ac / K193me3	1,000	0,772	0,683
K184me3 / K191ac / K193ac	1,000	0,796	0,733
K184me2 / K191me1 / K193ac	1,000	0,729	1,509
K184me2 / K191me1 / K195ac	1,000	0,729	1,509
N-TermTMT6plex / K187ac / K191ac	1,000	0,767	0,749
N-TermTMT6plex / K187ac / S188P / K196me1	1,000	0,428	0,593
N-TermTMT6plex / K187ac / S188P / K193me1	1,000	0,388	0,528
K191me1 / K193me2 / K195ac / K196me3	1,000	0,815	1,268
K191me1 / K193me2 / K195me3 / K196ac	1,000	0,815	1,268
K191me1 / K193me3 / K195ac / K196ac	1,000	1,052	0,834
K191ac / K193me2 / K196me3	1,000	1,635	0,757
K191ac / K193me3 / K196me2	1,000	1,635	0,757
K191me2 / K193ac / K196me3	1,000	1,635	0,757
K191me2 / K193me3 / K196ac	1,000	1,635	0,757
K191me3 / K193ac / K196me2	1,000	1,635	0,757
K191me3 / K193me2 / K196ac	1,000	1,635	0,757
K193me3 / K195me3 / K196ac	1,000	0,858	1,466
N-TermTMT6plex / K198ac / K201me2 / K203me2	1,000	0,896	0,800
K198me1 / K201me3 / K206ac / K208ac	1,000	1,883	1,502
N-TermTMT6plex / K198ac / S200P / K201me2	1,000	0,909	0,796
N-TermTMT6plex / K198me2 / K203ac / K206ac	1,000	1,057	0,833
S200P / K201me1 / K203me1	1,000	1,213	0,856
K201me3 / K203me1 / K206ac / K208ac	1,000	2,090	1,367
K201ac / K203ac / K208me1	1,000	0,908	1,003
K201ac / K203ac / K206me1	1,000	0,908	1,003
K201me2 / S205P / T210P	1,000	1,358	0,731
N-TermTMT6plex / K201ac / K203ac / K206me2	1,000	0,961	0,760
K203ac / K206ac / K208me2 / K211me3	1,000	0,698	1,211
N-TermTMT6plex / S205P / K208me1 / T210P	1,000	0,998	1,142
K206ac / K208ac	1,000	0,603	1,147
K208ac / K211ac / K213me3	1,000	1,094	0,782
K208me3 / K211ac / K213ac	1,000	1,094	0,782
K208me2 / T210P	1,000	1,029	0,693
K208me1 / T210P / K211me1	1,000	1,029	0,693

Tabelle A. 20: Ergänzung zu den Daten aus Kapitel 3.4.3: Quantifizierte Modifikationen der Histon H1 Isoform H1.4.

Modifikationen	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
K33ac / T44P / K45me2	1,000	1,107	0,824
K33me2 / K45me3	1,000	0,517	0,719
K45me2	1,000	0,973	0,677
N-TermTMT6plex / K45me2	1,000	0,977	1,028
K51ac / K62me1	1,000	0,844	0,572
K51me1 / K62ac	1,000	0,844	0,572
N-TermTMT6plex / S57P / K62me1 / K63me1	1,000	1,082	0,781
S77P / K80me3 / K84ac / K89me1	1,000	1,199	1,431
K84ac / S88P / K89me1	1,000	0,868	0,886
K84ac / K89me3 / T91P	1,000	0,767	1,250
K84me2 / K89me1 / T91P / K96me1	1,000	0,680	1,224
K84me2 / S85P / S88P / K89me1	1,000	1,163	0,843
K84me2 / S88P / K89me1 / K96ac	1,000	0,671	1,039
K84me1 / S85P / K89me3	1,000	1,006	0,801
N-TermTMT6plex / T98P / K106me3 / K109ac	1,000	1,158	0,849
S103P / K105me2 / K108me1 / K109ac	1,000	0,923	0,800
S103P / K105me3 / K109ac	1,000	0,858	0,799
K105me1	1,000	0,624	0,849
N-TermTMT6plex / K108ac / K109me2	1,000	0,541	0,823
N-TermTMT6plex / K108me2 / K109ac	1,000	0,601	0,887
K108ac / K109me2 / K116me1	1,000	0,661	1,098
K108ac / K109me2 / K118me1	1,000	0,636	1,074
K108me2 / K109ac / K116me1	1,000	0,661	1,098

Fortsetzung Tabelle A. 20: Ergänzung zu den Daten aus Kapitel 3.4.3: Quan. Modifikationen der Histon H1 Isoform H1.4.

Modifikationen	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
K108me2 / K109ac / K118me1	1,000	0,636	1,074
K109me3 / K116me2 / K118me2	1,000	0,890	1,444
K109ac / K116ac / K118me3	1,000	0,772	0,683
K109me3 / K116ac / K118ac	1,000	0,796	0,733
N-TermTMT6plex / K120me1	1,000	1,287	1,319
K126me3 / K128me3 / K136ac / K138ac	1,000	4,273	1,043
K126me2 / K128me2 / K135ac	1,000	0,899	1,984
K126me2 / K128me2 / K129ac	1,000	0,842	1,478
K128ac / K129me1 / K135me3	1,000	0,749	1,495
K128me1 / K129ac / K135me3	1,000	0,592	1,431
K128ac / K129ac / K136me3 / K138me3	1,000	2,923	1,384
K128ac / K129ac / K136me2 / K139me1	1,000	1,074	0,767
K128ac / K129ac / K136me1 / K138me2	1,000	1,319	0,843
K128ac / K129ac / K136me1 / K139me2	1,000	1,319	0,843
N-TermTMT6plex / K129ac / K135ac	1,000	0,469	0,725
K129me1 / K135me1 / K138ac / K139ac	1,000	0,698	0,734
K129ac / K135ac / K139me1	1,000	0,457	0,896
K135ac / K136me1	1,000	0,638	0,892
K135me1 / K136ac	1,000	0,638	0,892
K135me1 / K138ac	1,000	1,217	0,731
K135me1 / K139ac	1,000	1,317	0,715
N-TermTMT6plex / K138ac / K139ac / K147me3	1,000	0,855	0,843
K139me2 / K147me1 / K148me1 / S149P	1,000	1,107	0,883
K139me2 / K147me1 / K148me1 / T150P	1,000	1,107	0,883
T143P / K48me2	1,000	0,655	0,971
T143P / K147me1 / K148me1	1,000	0,655	0,971
N-TermTMT6plex / T143P / K147me2	1,000	0,743	0,700
N-TermTMT6plex / T145P / K147me2	1,000	0,743	0,700
K147ac / K148me3 / K151me1	1,000	0,985	1,252
K147me3 / K148ac / K151me1	1,000	0,985	1,252
K147me2 / T150P	1,000	0,939	0,699
K147me1 / K148ac / S149P / K151ac	1,000	2,619	1,360
K147me1 / K148ac / T150P / K151ac	1,000	2,619	1,360
K147me1 / K148me1 / S149P	1,000	0,939	0,699
K147me1 / K148me1 / T150P	1,000	0,939	0,699
N-TermTMT6plex / K147me2 / T150P	1,000	0,859	0,987
N-TermTMT6plex / K147me1 / K148me1 / S149P	1,000	0,859	0,987
K147me1 / K150me2 / K151me3 / K154me3	1,000	0,657	0,576
K148me2 / T150P	1,000	0,939	0,699
K150ac / K151ac / K154ac	1,000	0,968	0,647
N-TermTMT6plex / K151me3	1,000	0,771	0,743
K152ac	1,000	0,993	0,938
K152ac	1,000	0,854	0,820
N-TermTMT6plex / K152ac	1,000	0,724	0,090
T153P / K159me1 / K167ac	1,000	0,843	0,655
N-TermTMT6plex / K156me1	1,000	0,503	0,504
K158ac / K159ac / K167ac	1,000	0,872	0,585
N-TermTMT6plex / K158ac / K159me2 / K167me1	1,000	1,172	0,706
N-TermTMT6plex / K158me2 / K159ac / K167me1	1,000	1,175	0,725
N-TermTMT6plex / K158me2 / K159ac / K170me1	1,000	1,222	0,713
K167ac / K168ac	1,000	0,957	0,774
K168me2 / S171P	1,000	1,050	0,735
K168me1 / K170me1 / S171P	1,000	1,081	0,708
K170me2 / S171P	1,000	1,081	0,708
K174ac / K179me3	1,000	1,153	0,960
K176ac / K179me3	1,000	1,085	0,984
N-TermTMT6plex / K176ac	1,000	0,841	1,603
K182ac / K185ac / K189me2	1,000	0,894	0,885
K185ac / K189me3	1,000	0,817	0,809
K185me3 / K189ac	1,000	0,775	0,762
K185me1 / S186P / T192P	1,000	0,818	0,898
K185me3 / S186P / K189me3 / K191me3	1,000	0,962	1,220
K189me3 / K191me3 / T192P / K194ac	1,000	1,142	0,981
K189me2 / K191me3 / T192P / K194ac	1,000	1,253	0,965
K191ac / K194ac / K196ac / K199me3	1,000	4,783	0,865
K191ac / K194ac / K196ac / K201me3	1,000	3,824	0,915
T192P / K201ac	1,000	1,055	0,654

Anhang

Fortsetzung Tabelle A. 20: Ergänzung zu den Daten aus Kapitel 3.4.3: Quan. Modifikationen der Histon H1 Isoform H1.4.

Modifikationen	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
K194ac / K196ac / K199ac / K201me3	1,000	3,071	0,856
K194ac / K196ac / K199ac / K200me3	1,000	0,744	0,787
K194me3 / K196me1 / K199me1	1,000	1,952	1,423
K194ac / K196ac	1,000	1,066	0,855
K194ac / K196ac / K199ac / S203P	1,000	0,763	1,507
K194ac / K196ac / K199ac / T202P	1,000	1,770	1,150
K194ac / K199me3 / K201me3	1,000	1,260	0,771
K194me1 / K196ac / K199me2 / K201me3	1,000	1,197	0,807
K196ac / K199me3 / K201me3	1,000	1,213	0,824
S203P / K204me1 / K206me3 / K209ac	1,000	1,026	0,786
K212me2 / T213P / K216me1	1,000	0,749	0,728
T213P / K216me1 / K217ac / K218me3	1,000	0,772	1,004
N-TermTMT6plex / K216ac / K217me1	1,000	0,547	1,064

Tabelle A. 21: Ergänzung zu den Daten aus Kapitel 3.4.3: Quantifizierte Modifikationen der Histon H1 Isoform H1.5.

Modifikationen	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
K22ac / K25ac / K26ac / K31me3	1,000	2,806	0,711
K25me1	1,000	0,767	1,737
N-TermTMT6plex / K25ac / K26ac / K31me3	1,000	0,894	0,897
N-TermTMT6plex / K25me1	1,000	0,831	2,050
K26me1	1,000	0,767	1,737
K26me1 / K31me2	1,000	2,875	0,686
K51ac	1,000	0,687	1,083
S57P / K62me2	1,000	0,829	0,882
S57P / K62me1 / K63me1	1,000	0,829	0,882
K72me3 / K75ac / K77ac / K79ac	1,000	1,595	0,728
N-TermTMT6plex / K74me1	1,000	0,787	1,227
S77P / K80me3 / K84ac / K89me1	1,000	1,199	1,431
K84ac / S88P / K89me1	1,000	0,868	0,886
K84ac / K89me3 / T91P	1,000	0,767	1,250
K84me2 / K89me1 / T91P / K96me1	1,000	0,680	1,224
K84me2 / S85P / S88P / K89me1	1,000	1,163	0,843
K84me2 / S88P / K89me1 / K96ac	1,000	0,671	1,039
K84me1 / S85P / K89me3	1,000	1,006	0,801
N-TermTMT6plex / T98P / K105me3 / K108ac	1,000	1,158	0,849
S103P / K105me2 / K108me1 / K109ac	1,000	0,923	0,800
S103P / K105me3 / K109ac	1,000	0,858	0,799
K105me1	1,000	0,624	0,849
N-TermTMT6plex / K108ac / K109me2	1,000	0,541	0,823
N-TermTMT6plex / K108me2 / K109ac	1,000	0,601	0,887
K108ac / K109me2 / K116me1	1,000	0,661	1,098
K108ac / K109me2 / K118me1	1,000	0,636	1,074
K108me2 / K109ac / K116me1	1,000	0,661	1,098
K108me2 / K109ac / K118me1	1,000	0,636	1,074
K109me3 / K116me2 / K118me2	1,000	0,890	1,444
K109ac / K116ac / K118me3	1,000	0,772	0,683
K109me3 / K116ac / K118ac	1,000	0,796	0,733
K109me2 / K116me1 / K118ac	1,000	0,729	1,509
K109me2 / K116me1 / K120ac	1,000	0,729	1,509
K116me1 / K118me2 / K120ac / K121me3	1,000	0,815	1,268
K116me1 / K118me2 / K120me3 / K121ac	1,000	0,815	1,268
K116me1 / K118me3 / K120ac / K121ac	1,000	1,052	0,834
K118me3 / K120me3 / K121ac	1,000	0,858	1,466
N-TermTMT6plex / K120me1	1,000	1,287	1,319
N-TermTMT6plex / K121ac / K126ac / K128ac	1,000	0,901	0,721
K121ac / K128me2 / K129me2 / T134P	1,000	0,893	0,947
K121ac / K128me1 / K129me3 / T134P	1,000	0,816	0,838
K121ac / K128me3 / K129me1 / T134P	1,000	0,893	0,947
K121me2 / K128ac / K129ac / K136me3	1,000	0,860	1,031
T122P / K126ac / K128me1 / K136me3	1,000	0,609	0,835
K126ac	1,000	0,824	1,137
K126ac / K128ac / K129me2 / K136me3	1,000	0,877	1,005
K126ac / K128ac / K129me3	1,000	1,205	0,902

Anhang

Fortsetzung Tabelle A. 21: Ergänzung zu den Daten aus Kapitel 3.4.3: Quan. Modifikationen der Histone H1 Isoform H1.5.

Modifikationen	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
K126ac / K128me1 / T134P / K136me3	1,000	1,146	0,651
K126ac / K129me1 / T134P / K136me3	1,000	1,146	0,651
K126me2 / K128ac / K129ac / K136me3	1,000	1,242	1,043
K126me2 / K128me2 / K129ac / T134P	1,000	1,135	0,977
K126me1 / K128ac / T134P / K136me3	1,000	1,146	0,651
K126me1 / K129ac / T134P / K136me3	1,000	1,146	0,651
K126me1 / K129me3 / T134P / K136ac	1,000	0,944	0,812
K126me3 / K128ac / K129ac	1,000	1,170	0,851
K126me3 / K128ac / K129ac / K136me2	1,000	0,890	0,981
K126me3 / K128me1 / T134P / K136ac	1,000	1,079	0,924
K126me3 / K129ac / K136ac	1,000	1,150	0,914
K126me3 / K129me1 / T134P / K136ac	1,000	1,079	0,924
N-TermTMT6plex / K126me3	1,000	0,572	0,792
N-TermTMT6plex / K126me3	1,000	0,762	1,163
K127ac / K128ac / T133P / K135me1	1,000	0,786	0,751
K127ac / K128ac / T133P / K136me1	1,000	0,786	0,751
K127ac / K128me1 / T133P / K138me3	1,000	0,716	1,289
K127ac / K128me3 / K135me3 / K136me3	1,000	0,609	0,997
K127me3 / K128me3 / K135me3 / K136ac	1,000	0,581	0,874
K128ac / K129me1 / T134P / K136me3	1,000	1,146	0,651
K128me2 / K129me2 / T134P / K136ac	1,000	0,944	0,812
K128me1 / K129ac / T134P / K136me3	1,000	1,146	0,651
K128me1 / K129me3 / T134P / K136ac	1,000	0,944	0,812
K128me3 / K129me1 / T134P / K136ac	1,000	1,052	0,950
K129ac / T134P / K136ac / K139me2	1,000	0,702	0,816
K129me1 / T134P / K139ac / K140ac	1,000	0,730	1,121
K129me1 / T134P / K140ac / K145ac	1,000	0,769	0,982
T134P / K136me1 / K139ac / K140ac	1,000	0,736	1,072
T134P / K137me1 / K139ac / K140ac	1,000	0,764	1,089
K136ac / K137ac / K139ac / K140me3	1,000	0,879	0,812
K136me3 / K137ac / K139ac / K140ac	1,000	0,938	0,789
N-TermTMT6plex / K145me2 / K146ac	1,000	1,419	0,943
K145ac / K146me3 / T147P / K149me2	1,000	0,422	0,380
N-TermTMT6plex / K146me2 / K149ac	1,000	0,908	0,766
K149me1 / K150ac / K153ac	1,000	0,890	0,741
N-TermTMT6plex / K149ac / K150ac / T151P	1,000	1,065	1,133
K149me1 / K153ac / K154me3	1,000	0,519	0,968
K149me1 / K153me3 / K154ac	1,000	0,519	0,968
T151P / K164ac	1,000	0,994	0,938
T151P / K154me3 / K156me3 / K164ac	1,000	0,922	0,973
T151P / K157ac	1,000	0,965	0,904
N-TermTMT6plex / K154me1	1,000	0,503	0,504
K154ac / K156ac / K157ac / K164me1	1,000	1,401	0,848
K154me2 / K156ac / K157ac / K164ac	1,000	0,840	0,708
N-TermTMT6plex / K156me2 / K157ac / K114ac	1,000	0,893	0,635
N-TermTMT6plex / K157ac / K114ac / K115me2	1,000	0,893	0,635
K165ac / K168me2	1,000	0,822	0,735
S169P / K172ac / K174me3	1,000	0,924	0,541
N-TermTMT6plex / K171me3 / K172ac / K174ac	1,000	0,973	0,664
K178me3 / K180ac / K188me2 / K190me1	1,000	3,073	1,252
K180me1 / K181me1 / S185P	1,000	1,182	0,932
K184ac / K188ac / K190me3	1,000	1,066	0,855
K184ac / S185P / K188me1 / K190me1	1,000	0,583	0,466
N-TermTMT6plex / K184me1 / S185P	1,000	0,118	0,826
K184ac / S185P / K188me2 / K190me1	1,000	1,089	0,900
K184me3 / S185P / K188me2 / K190me1	1,000	0,643	0,751
K184ac / S185P / K188me1 / S194P	1,000	1,198	0,887
K184ac / S185P / K190me1 / S194P	1,000	1,198	0,887
K184me3 / K188me3 / S194P / K195me1	1,000	1,992	0,968
K184me3 / S185P / K188me3 / K195me1	1,000	1,892	0,979
S185P / K188ac / K190ac / K193me1	1,000	1,093	0,668
S185P / K188me1 / K190ac / K193ac	1,000	1,093	0,668
S185P / K188me1 / K190me1 / S194P	1,000	0,990	1,351
S185P / K190ac	1,000	1,216	0,774
K188me1 / K193ac / S194P / K195me2	1,000	0,939	0,879
K188me1 / K193me2 / S194P / K195ac	1,000	0,939	0,879
K188me1 / K190ac / K193ac / K195ac	1,000	0,816	1,366

Anhang

Fortsetzung Tabelle A. 21: Ergänzung zu den Daten aus Kapitel 3.4.3: Quan. Modifikationen der Histon H1 Isoform H1.5.

Modifikationen	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
K188me3 / K190me3 / S194P / K195me1	1,000	1,704	0,961
K188ac / K190ac	1,000	0,629	0,885
K188ac / K190ac / K193ac	1,000	0,827	1,434
K188ac / S194P	1,000	0,981	0,735
K190ac / K193ac	1,000	0,766	1,088
K190ac / K193ac / K195ac	1,000	0,827	1,434
K193ac / K195ac	1,000	0,572	0,954
K193ac / S194P / K195ac	1,000	0,624	0,887
S194P / K195ac	1,000	1,177	0,762
S194P / K198ac	1,000	1,114	0,680
S194P / K195ac / K198ac / K200ac	1,000	0,821	1,267
N-TermTMT6plex / K195me3 / S197P	1,000	0,646	0,850
K198me1 / K200ac / K203ac / K205ac	1,000	1,068	0,895
K198ac / K200ac / S202P / K203ac	1,000	0,915	0,896
K200ac / K202ac / K204ac / T206P	1,000	0,894	0,741
K200me3 / S202P	1,000	1,108	0,767
T202P / K203me1	1,000	1,092	0,594
K202me1 / K204me1 / T206P	1,000	0,903	1,049
S202P / K203me3 / K208ac / K210ac	1,000	0,639	0,571
T206P / K208me2 / K210me2	1,000	1,351	0,979
T206P / K208me1 / K210me3	1,000	1,351	0,979
T206P / K208me3 / K210me1	1,000	1,351	0,979
N-TermTMT6plex / K220me1 / K221ac	1,000	0,936	0,723

Tabelle A. 22: Ergänzung zu den Daten aus Kapitel 3.4.3: Quantifizierte Modifikationen der H2A-Histone.

Isoform	Modifikation	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
	N-TermTMT6plex / K5me2 / K9me1	1,000	1,107	0,704
	N-TermTMT6plex / K9me3	1,000	1,107	0,704
	N-TermTMT6plex / K9ac	1,000	0,775	1,506
	K13ac / K15ac	1,000	0,727	0,799
	N-TermTMT6plex / K13me3	1,000	0,632	1,068
1-K	N-TermTMT6plex / K15me2 / T16P	1,000	1,158	1,194
	K119me3 / K125me1 / K129ac	1,000	0,823	0,772
	K119me3 / K125me1 / K127ac	1,000	0,823	0,772
	T122P / K127me1 / K129me1	1,000	1,080	0,833
	K125me3 / K127ac / K129me1	1,000	1,023	0,732
	K125me3 / K127me1 / K129ac	1,000	1,023	0,732
	K4me2 / K7ac / S9P / K11me3	1,000	0,665	0,808
	N-TermTMT6plex / K4me2 / K7me2 / S9P	1,000	1,032	0,758
	K7me1 / K11me2 / K13me3	1,000	1,105	0,955
	N-TermTMT6plex / K7me2 / K15me1	1,000	1,059	0,720
	N-TermTMT6plex / K7me2 / K13me1	1,000	1,115	0,791
	N-TermTMT6plex / K11me2 / K13me3 / K15me3	1,000	0,784	0,655
H2A.Z	N-TermTMT6plex / K11me1 / K13me3 / K15me3	1,000	1,026	0,612
	N-TermTMT6plex / K11me3 / K13me2 / K15me2	1,000	1,066	0,594
	N-TermTMT6plex / K11me3 / K13me3 / K15me2	1,000	0,896	0,782
	N-TermTMT6plex / K11me3 / K13me3 / K15me1	1,000	1,066	0,594
	N-TermTMT6plex / K11me1 / T14P	1,000	0,639	0,773
	N-TermTMT6plex / K13me3 / K15ac	1,000	1,129	0,905
	N-TermTMT6plex / K13me2 / T14P	1,000	0,805	0,913
	N-TermTMT6plex / K5me2 / K9me1	1,000	1,107	0,704
	N-TermTMT6plex / K9me3	1,000	1,107	0,704
	N-TermTMT6plex / K9ac	1,000	0,775	1,506
H2A2-C	N-TermTMT6plex / K13me3	1,000	0,632	1,068
	K13ac / K15ac	1,000	0,727	0,799
	N-TermTMT6plex / K15me3 / S16P	1,000	1,158	1,194
	K119me3 / K124ac	1,000	0,682	1,403

Fortsetzung Tabelle A. 22: Ergänzung zu den Daten aus Kapitel 3.4.3: Quan. Modifikationen der H2A-Histone.

Isoform	Modifikation	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
H2AX	N-TermTMT6plex / K13me3	1,000	0,632	1,068
	N-TermTMT6plex / K15me3 / S16P	1,000	1,158	1,194
	K119me1 / K127me3	1,000	0,901	0,903
	K119me2 / S120P / K127ac / K133me1	1,000	0,681	0,879
	K119me1 / S121P	1,000	0,916	0,841
	K119me1 / S121P / K127ac / K133me3	1,000	0,862	0,908
	K119me1 / T123P	1,000	0,889	0,890
	N-TermTMT6plex / K127ac / K133ac / K134me1	1,000	0,640	0,974
	T123P / K127me3 / K133ac / K134me2	1,000	0,859	0,791
	T123P / K133me3 / K133me2 / K134ac	1,000	0,859	0,791
	N-TermTMT6plex / S121P / T123P / K133ac	1,000	1,016	1,057
	N-TermTMT6plex / S121P / T123P / K134ac	1,000	1,016	1,057

Tabelle A. 23: Ergänzung zu den Daten aus Kapitel 3.4.3: Quantifizierte Modifikationen der H2B-Histone.

Isoform	Modifikation	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
H2B3-B	N-TermTMT6plex / K11ac / K12ac / K15me3	1,000	0,822	0,886
	N-TermTMT6plex / S14P / K16me1 / K20me1	1,000	0,960	0,961
	K15me2 / T19P	1,000	0,721	0,903
	K15me1 / K16me1 / T19P	1,000	0,778	0,820
	K15me3 / K16me1 / T19P	1,000	0,845	0,713
	N-TermTMT6plex / K15me1 / K20me2 / K23me2	1,000	0,680	1,224
	N-TermTMT6plex / K16me1 / T19P / K20me1	1,000	1,320	0,803
	K16me3 / K20ac / K23ac / K24ac	1,000	0,449	0,688
	K20ac / K24me1	1,000	0,659	0,693
	K24me1	1,000	0,696	0,964
	K27me1	1,000	0,670	1,015
	Y83ac / K85me3	1,000	1,027	1,584
	K108me3	1,000	1,107	0,660
	K116me1 / K130me1 / K135me2	1,000	1,214	0,920
	K116me1 / K130me3	1,000	0,858	0,853
	N-TermTMT6plex / T119P / K120ac	1,000	0,730	0,769
	K130ac	1,000	0,936	0,905
	K130me3	1,000	1,490	0,932
	Y131ac	1,000	0,988	0,938
	K135me3	1,000	1,490	0,932
	K15me2 / T19P	1,000	0,721	0,903
	K15me1 / K16me1 / T19P	1,000	0,778	0,820
	K15me3 / K16me1 / T19P	1,000	0,845	0,713
	N-TermTMT6plex / K15me1 / K20me2 / K23me2	1,000	0,680	1,224
	N-TermTMT6plex / S15P / K17me1 / K20me1	1,000	0,960	0,961
	N-TermTMT6plex / K16me1 / T19P / K20me1	1,000	1,320	0,803
	K16me3 / K20ac / K23ac / K24ac	1,000	0,449	0,688
	K20ac / K24me1	1,000	0,659	0,693
	K24me1	1,000	0,696	0,964
	K27me1	1,000	0,670	1,015
	N-TermTMT6plex / Y42P / K43ac / K46ac	1,000	3,136	1,340
	Y42P / K43ac / K46ac	1,000	1,117	0,883
	K43me1	1,000	0,604	0,941
Y83ac / K85me3	1,000	1,027	1,584	
K108me3	1,000	1,107	0,660	
K116me1 / K120me1 / K125me2	1,000	1,214	0,920	
K116me1 / K120me3	1,000	0,858	0,853	
N-TermTMT6plex / T119P / K120ac	1,000	0,730	0,769	
K120ac	1,000	0,936	0,905	
K120me3	1,000	1,490	0,932	
K120me3 / T122P / K125me3 / K130me3	1,000	0,358	1,031	
K120me3 / Y121P / K125me3 / K130me3	1,000	0,358	1,031	
Y121ac	1,000	0,988	0,938	
K125me3	1,000	1,490	0,932	

Anhang

Fortsetzung Tabelle A. 23: Ergänzung zu den Daten aus Kapitel 3.4.3: Quan. Modifikationen der H2B-Histone.

Isoform	Modifikation	d_0/d_0	$(d_1/d_0)/0.956$	$(d_2/d_0)/1.0445$
	N-TermTMT6plex / K11ac / K12ac / K15me3	1,000	0,822	0,886
	N-TermTMT6plex / S14P / K16me1 / K20me1	1,000	0,960	0,961
	K15me2 / T19P	1,000	0,721	0,903
	K15me1 / K16me1 / T19P	1,000	0,778	0,820
	K15me3 / K16me1 / T19P	1,000	0,845	0,713
	N-TermTMT6plex / K15me1 / K20me2 / K23me2	1,000	0,680	1,224
	N-TermTMT6plex / K16me1 / T19P / K20me1	1,000	1,320	0,803
	K20ac / K24me1	1,000	0,659	0,693
	K24me1	1,000	0,696	0,964
	K27me1	1,000	0,670	1,015
H2B1-M	N-TermTMT6plex / Y42P / K43ac / K46ac	1,000	3,136	1,340
	Y42P / K43ac / K46ac	1,000	1,117	0,883
	K43me1	1,000	0,604	0,941
	Y83ac / K85me3	1,000	1,027	1,584
	K108me3	1,000	1,107	0,660
	K116me1 / K120me1 / K125me2	1,000	1,214	0,920
	K116me1 / K120me3	1,000	0,858	0,853
	N-TermTMT6plex / T119P / K120ac	1,000	0,730	0,769
	K120ac	1,000	0,936	0,905
	Y121ac	1,000	0,988	0,938
	K125me3	1,000	1,490	0,932
	N-TermTMT6plex / K11ac / K12ac / K15me3	1,000	0,822	0,886
	N-TermTMT6plex / S14P / K16me1 / K20me1	1,000	0,960	0,961
	K15me2 / T19P	1,000	0,721	0,903
	K15me1 / K16me1 / T19P	1,000	0,778	0,820
	K15me3 / K16me1 / T19P	1,000	0,845	0,713
	N-TermTMT6plex / K15me1 / K20me2 / K23me2	1,000	0,680	1,224
	N-TermTMT6plex / K16me1 / T19P / K20me1	1,000	1,320	0,803
	K20ac / K24me1	1,000	0,659	0,693
H2B1-K	K24me1	1,000	0,696	0,964
	K27me1	1,000	0,670	1,015
	N-TermTMT6plex / Y42P / K43ac / K46ac	1,000	3,136	1,340
	Y42P / K43ac / K46ac	1,000	1,117	0,883
	K43me1	1,000	0,604	0,941
	Y83ac / K85me3	1,000	1,027	1,584
	K108me3	1,000	1,107	0,660
	K116me3 / K120ac / K123ac / K124ac	1,000	0,521	0,735
	N-TermTMT6plex / T119P / K120ac	1,000	0,730	0,769
	Y121P / K125me3	1,000	0,792	0,887

Abkürzungsverzeichnis

8oxoG	8-Oxoguanin
A	Adenin
Å	Angström
Abb.	Abbildung
ABC-Puffer	Ammonium-Bicarbonat Puffer
Ac	Acetyl
ACN	Acetonitril
Aid	<i>activation-induced deaminase</i>
APCI	<i>atmospheric pressure chemical ionization</i>
Apobec	<i>apolipoprotein B pre-mRNA editing enzyme, catalytic polypeptide</i>
APPI	<i>atmospheric pressure photo ionization</i>
AQUA	<i>absolute quantification of proteins</i>
ATP	Adenosintriphosphat
BDNF, Bdnf	<i>brain-derived neurotrophic factor</i>
BER	Basenexzisionsreparatur, <i>base excision repair</i>
bp	Basenpaar(e)
BSA	Rinderserumalbumin, <i>bovine serum albumin</i>
Bsp.	Beispiel
bzw.	beziehungsweise
C	Cytosin
C.elegans	<i>caenorhabditis elegans</i>
caC	Carboxylcytosin
CI	chemische Ionisation

Abkürzungsverzeichnis

CID	<i>collision induced dissociation</i>
Co-IP	Co-Immunopräzipitation
CpG	Cytosin-phosphatidyl-Guanin-Insel
Da	Dalton
DMSO	Dimethylsulfoxid
dN	2'-Desoxynukleosid
DNA	Desoxyribonukleinsäure, <i>deoxyribonucleic acid</i>
DNMT, Dnmt	DNA-Methyltransferase
DTT	Dithiothreitol
etc.	<i>et cetera</i>
EDTA	Ethylendiamintetraessigsäure, <i>ethylenediaminetetraacetic acid</i>
EI	Elektronenstoßionisation
ELISA	<i>enzyme linked immunosorbent assay</i>
EMV	<i>electron multiplier voltage</i>
ESI	Elektrospray-Ionisierung
Et	Ethyl
et al.	<i>et alii</i>
EtOAc	Ethylacetat
EtOH	Ethanol
FA	Ameisensäure, <i>formic acid</i>
FAB	<i>fast atom bombardment</i>
fC	Formylcytosin
FDR	<i>false discovery rate</i>
FT-ICR	Fourier-Transformation Ionen-Zyklotron-Resonanz
G	Guanin
g	Gramm
h	Stunde(n)

Abkürzungsverzeichnis

HATs	Histon-Acetyl-Transferasen
HDAC	Histon-Deacetylase
HEK	<i>human embryonic kidney</i>
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethansulfonsäure
HGP	Humanes Genom Projekt
hmC	Hydroxymethylcytosin
HP1	Heterochromatin-Protein 1
HPLC	Hochdruckflüssigkeitschromatographie
IAA	Iodacetamid
ICAT	<i>isotope-coded affinity tag</i>
ICPL	<i>isotope-coded protein label</i>
iTRAQ	<i>isotope tags for relative and absolute quantification</i>
IU	Ioduracil
IUPAC	<i>International Union of Pure and Applied Chemistry</i>
KCl	Kaliumchlorid
L	Liter
LC-MS	<i>liquid chromatography-mass spectrometry</i>
LFQ	markierungsfreie Quantifizierungsmethode (<i>label free quantification</i>)
M	mol/L
MALDI	<i>matrix assisted laser desorption ionization</i>
MBD, Mbd	<i>methyl-CpG-binding domain</i>
mC	Methylcytosin
me	Methyl
MeCP2	<i>methyl-CpG binding protein 2</i>
mESC	embryonale Stammzelle der Maus
min	Minute(n)
MRM	<i>multiple reaction monitoring</i>

Abkürzungsverzeichnis

MS	Massenspektrometrie, Massenspektrometer
MS/MS, MS ²	Tandemmassenspektrometrie
NCBI	„ <i>The National Center for Biotechnology Information</i> “
NER	Nukleotidexzisionsreparatur, <i>nucleotide excision repair</i>
NHS	N-Hydroxysuccinimid
Oct4	Octamer-bindender Transkriptionsfaktor 4
p	Signifikanzwert berechnet durch einen t-Test
PAGE	Polyacrylamidgelelektrophorese
PBS	<i>phosphate buffered saline</i>
PEG	Polyethylenglycol
PLA	<i>proximity ligation assay</i>
PMSF	Phenylmethylsulfonylfluorid
Q	Quadrupol
QQQ	Triple-Quadrupol
RNA	Ribonukleinsäure, <i>ribonucleic acid</i>
ROS	<i>reactive oxygen species</i>
rpm	<i>rounds per minute</i>
RT	Retentionszeit oder Raumtemperatur
SAM	S-Adenosylmethionin
SDS	Natriumdodecylsulfat, <i>sodium dodecyl sulphate</i>
SILAC	<i>stable isotope labeling by amino acids in cell culture</i>
Smug1	<i>single-stranded-selective monofunctional uracil-DNA glycosylase 1</i>
SRM/MRM	<i>single/multiple reaction monitoring</i>
SWI/SNF	<i>switch/sucrose non-fermentable</i>
T	Tesla (Einheit) oder Thymin (Nukleinbase)
TBTA	Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amin)
TCA	Trichloressigsäure, <i>trichloroacetic acid</i>

Abkürzungsverzeichnis

TCEP	Tris-(2-Carboxyethyl)Phosphat Hydrochlorid
TDG, Tdg	Thymin-DNA Glykosylase
TEMED	N,N,N',N',-Tetramethyldiamin
TET, Tet	<i>ten-eleven translocation methylcytosine dioxygenase</i>
TFA	Trifluoressigsäure, <i>trifluoroacetic acid</i>
TMT	<i>tandem mass tag</i>
TOF	<i>time of flight</i>
Tris-HCl	Tris-(hydroxymethyl)-aminoethanhydrochlorid
u.a.	unter anderem
Uhrf	<i>ubiquitin-like, containing PHD and RING finger domains</i>
UniProt	<i>„universal protein database“</i>
UV	ultraviolettes Licht
v/v	<i>volume per volume</i>
w/v	<i>weight per volume</i>
z.B.	zum Beispiel

Aminosäuren im IUPAC Einbuchstaben-Code

Literaturverzeichnis

- [1] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, and T. Carell, *Nat. Chem. Biol.* **2014**. *10*, 574-581. *Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA.*
- [2] A. Perera, D. Eisen, M. Wagner, S. K. Laube, A. F. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell, and S. Michalakis, *Cell Rep.* **2015**. *11*, 283-94. *Tet3 is recruited by rest for context-specific hydroxymethylation and induction of gene expression.*
- [3] E. S. Lander, L. M. Linton; Birren, B., *et al.*, *Nature* **2001**. *409*, 860-921. *Initial sequencing and analysis of the human genome.*
- [4] R. Holliday, *Biol. Rev. Camb. Philos. Soc.* **1990**. *65*, 431-71. *Mechanisms for the control of gene activity during development.*
- [5] T. L. Sheppard, *Nat. Chem. Biol.* **2015**. *11*, 175-175. *Rna epigenetics: M6a partial differential.*
- [6] C. He, *Nat. Chem. Biol.* **2010**. *6*, 863-865. *Grand challenge commentary: Rna epigenetics?*
- [7] A. J. Bannister and T. Kouzarides, *Cell Res.* **2011**. *21*, 381-95. *Regulation of chromatin by histone modifications.*
- [8] S. R. Bhaumik, E. Smith, and A. Shilatifard, *Nat. Struct. Mol. Biol.* **2007**. *14*, 1008-16. *Covalent modifications of histones during development and disease pathogenesis.*
- [9] I. Cantone and A. G. Fisher, *Nat. Struct. Mol. Biol.* **2013**. *20*, 282-9. *Epigenetic programming and reprogramming during development.*
- [10] H. Cedar and Y. Bergman, *Nat. Rev. Genet.* **2009**. *10*, 295-304. *Linking DNA methylation and histone modification: Patterns and paradigms.*
- [11] Y. Fu and C. He, *Curr. Opin. Chem. Biol.* **2012**. *16*, 516-24. *Nucleic acid modifications with epigenetic significance.*
- [12] S. I. S. Grewal and D. Moazed, *Science* **2003**. *301*, 798-802. *Heterochromatin and epigenetic control of gene expression.*
- [13] E. Stedman and E. Stedman, *Nature* **1950**. *166*, 780-781. *Cell specificity of histones.*

- [14] V. Tordera, R. Sendra, and J. E. Pérez-Ortín, *Experientia* **1993**. 49, 780-788. *The role of histones and their modifications in the informative content of chromatin.*
- [15] V. Ramakrishnan, *Annu. Rev. Biophys. Biomol. Struct.* **1997**. 26, 83-112. *Histone structure and the organization of the nucleosome.*
- [16] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, *J. Mol. Biol.* **2002**. 319, 1097-1113. *Solvent mediated interactions in the structure of the nucleosome core particle at 1.9Å resolution.*
- [17] C. L. Peterson and M. A. Laniel, *Curr. Biol.* **2004**. 14, R546-51. *Histones and histone modifications.*
- [18] J. Ausio, F. Dong, and K. E. Van Holde, *J. Mol. Biol.* **1989**. 206, 451-463. *Use of selectively trypsinized nucleosome core particles to analyze the role of the histone "tails" in the stabilization of the nucleosome.*
- [19] G. Felsenfeld and M. Groudine, *Nature* **2003**. 421, 448-453. *Controlling the double helix.*
- [20] A. R. Cutter and J. J. Hayes, *FEBS Letters* **2015**. 589, 2914-2922. *A brief review of nucleosome structure.*
- [21] L. Marino-Ramirez, M. G. Kann, B. A. Shoemaker, and D. Landsman, *Expert Rev. Proteomics* **2005**. 2, 719-29. *Histone structure and nucleosome stability.*
- [22] D. J. Clark and T. Kimura, *J. Mol. Biol.* **1990**. 211, 883-96. *Electrostatic mechanism of chromatin folding.*
- [23] H. Huang, B. R. Sabari, B. A. Garcia, C. D. Allis, and Y. Zhao, *Cell* **2014**. 159, 458-458 e1. *Snapshot: Histone modifications.*
- [24] C. Martin and Y. Zhang, *Nat. Rev. Mol. Cell. Biol.* **2005**. 6, 838-849. *The diverse functions of histone lysine methylation.*
- [25] T. Kouzarides, *Cell* **2007**. 128, 693-705. *Chromatin modifications and their function.*
- [26] R. J. Sims, 3rd, C. F. Chen, H. Santos-Rosa, T. Kouzarides, S. S. Patel, and D. Reinberg, *J. Biol. Chem.* **2005**. 280, 41789-92. *Human but not yeast chd1 binds directly and selectively to histone h3 methylated at lysine 4 via its tandem chromodomains.*
- [27] D. E. Sterner and S. L. Berger, *Microbiol. Mol. Biol. Rev.* **2000**. 64, 435-59. *Acetylation of histones and transcription-related factors.*
- [28] Y. Zhang and D. Reinberg, *Genes Dev.* **2001**. 15, 2343-60. *Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails.*

- [29] S. J. Nowak and V. G. Corces, *Trends Genet.* **2004.** *20*, 214-20. *Phosphorylation of histone h3: A balancing act between chromosome condensation and transcriptional activation.*
- [30] A. Shilatifard, *Annu. Rev. Biochem.* **2006.** *75*, 243-69. *Chromatin modifications by methylation and ubiquitination: Implications in the regulation of gene expression.*
- [31] D. Nathan, K. Ingvarsdottir, D. E. Sterner, G. R. Bylebyl, M. Dokmanovic, J. A. Dorsey, K. A. Whelan, M. Krsmanovic, W. S. Lane, P. B. Meluh, E. S. Johnson, and S. L. Berger, *Genes Dev.* **2006.** *20*, 966-76. *Histone sumoylation is a negative regulator in *saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications.*
- [32] P. O. Hassa, S. S. Haenni, M. Elser, and M. O. Hottiger, *Microbiol. Mol. Biol. Rev.* **2006.** *70*, 789-829. *Nuclear adp-ribosylation reactions in mammalian cells: Where are we today and where are we going?*
- [33] S. Mujtaba, L. Zeng, and M. M. Zhou, *Oncogene* **2007.** *26*, 5521-5527. *Structure and acetyl-lysine recognition of the bromodomain.*
- [34] A. H. Hassan, P. Prochasson, K. E. Neely, S. C. Galasinski, M. Chandy, M. J. Carrozza, and J. L. Workman, *Cell* **2002.** *111*, 369-79. *Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes.*
- [35] A. J. Bannister, P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire, and T. Kouzarides, *Nature* **2001.** *410*, 120-124. *Selective recognition of methylated lysine 9 on histone h3 by the hp1 chromo domain.*
- [36] M. Lachner, D. O'carroll, S. Rea, K. Mechtler, and T. Jenuwein, *Nature* **2001.** *410*, 116-120. *Methylation of histone h3 lysine 9 creates a binding site for hp1 proteins.*
- [37] S. Daujat, U. Zeissler, T. Waldmann, N. Happel, and R. Schneider, *J. Biol. Chem.* **2005.** *280*, 38090-5. *Hp1 binds specifically to lys26-methylated histone h1.4, whereas simultaneous ser27 phosphorylation blocks hp1 binding.*
- [38] T. Jenuwein and C. D. Allis, *Science* **2001.** *293*, 1074-1080. *Translating the histone code.*
- [39] T. Bartke, M. Vermeulen, B. Xhemalce, S. C. Robson, M. Mann, and T. Kouzarides, *Cell* **2010.** *143*, 470-484. *Nucleosome-interacting proteins regulated by DNA and histone methylation.*
- [40] R. D. Hotchkiss, *J. Biol. Chem.* **1948.** *175*, 315-32. *The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography.*
- [41] A. Razin and A. Riggs, *Science* **1980.** *210*, 604-610. *DNA methylation and gene function.*
- [42] Z. D. Smith and A. Meissner, *Nat. Rev. Genet.* **2013.** *14*, 204-20. *DNA methylation: Roles in mammalian development.*

- [43] J. Boyes and A. Bird, *Cell* **1991**. *64*, 1123-34. *DNA methylation inhibits transcription indirectly via a methyl-cpg binding protein.*
- [44] P. Gu, X. Xu, D. Le Menuet, A. C. Chung, and A. J. Cooney, *Stem Cells* **2011**. *29*, 1041-51. *Differential recruitment of methyl cpg-binding domain factors and DNA methyltransferases by the orphan receptor germ cell nuclear factor initiates the repression and silencing of oct4.*
- [45] C. P. Tan and S. Nakielny, *Mol. Cell. Biol.* **2006**. *26*, 7224-35. *Control of the DNA methylation system component mbd2 by protein arginine methylation.*
- [46] M. Bostick, J. K. Kim, P.-O. Estève, A. Clark, S. Pradhan, and S. E. Jacobsen, *Science* **2007**. *317*, 1760-1764. *Uhrf1 plays a role in maintaining DNA methylation in mammalian cells.*
- [47] X. Liu, Q. Gao, P. Li, Q. Zhao, J. Zhang, J. Li, H. Koseki, and J. Wong, *Nat. Commun.* **2013**. *4*, 1563. *Uhrf1 targets dnmt1 for DNA methylation through cooperative binding of hemimethylated DNA and methylated h3k9.*
- [48] H. Gowher, K. Liebert, A. Hermann, G. Xu, and A. Jeltsch, *J. Biol. Chem.* **2005**. *280*, 13341-8. *Mechanism of stimulation of catalytic activity of dnmt3a and dnmt3b DNA-(cytosine-c5)-methyltransferases by dnmt3l.*
- [49] D. Jia, R. Z. Jurkowska, X. Zhang, A. Jeltsch, and X. Cheng, *Nature* **2007**. *449*, 248-51. *Structure of dnmt3a bound to dnmt3l suggests a model for de novo DNA methylation.*
- [50] S. C. Wu and Y. Zhang, *Nat. Rev. Mol. Cell. Biol.* **2010**. *11*, 607-20. *Active DNA demethylation: Many roads lead to rome.*
- [51] W. Mayer, A. Niveleau, J. Walter, R. Fundele, and T. Haaf, *Nature* **2000**. *403*, 501-502. *Embryogenesis: Demethylation of the zygotic paternal genome.*
- [52] J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, and J. Walter, *Curr. Biol.* **2000**. *10*, 475-8. *Active demethylation of the paternal genome in the mouse zygote.*
- [53] K. Martinowich, D. Hattori, H. Wu, S. Fouse, F. He, Y. Hu, G. Fan, and Y. E. Sun, *Science* **2003**. *302*, 890-3. *DNA methylation-related chromatin remodeling in activity-dependent bdnf gene regulation.*
- [54] S. K. Bhattacharya, S. Ramchandani, N. Cervoni, and M. Szyf, *Nature* **1999**. *397*, 579-83. *A mammalian protein with specific demethylase activity for mcpg DNA.*
- [55] A. Sancar, L. A. Lindsey-Boltz, K. Unsal-Kacmaz, and S. Linn, *Annu. Rev. Biochem.* **2004**. *73*, 39-85. *Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints.*

- [56] H. D. Morgan, W. Dean, H. A. Coker, W. Reik, and S. K. Petersen-Mahrt, *J. Biol. Chem.* **2004**. 279, 52353-60. *Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: Implications for epigenetic reprogramming.*
- [57] C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, and W. Reik, *Nature* **2010**. 463, 1101-5. *Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by aid deficiency.*
- [58] K. Rai, I. J. Huggins, S. R. James, A. R. Karpf, D. A. Jones, and B. R. Cairns, *Cell* **2008**. 135, 1201-12. *DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45.*
- [59] G. Barreto, A. Schafer, J. Marhold, D. Stach, S. K. Swaminathan, V. Handa, G. Doderlein, N. Maltry, W. Wu, F. Lyko, and C. Niehrs, *Nature* **2007**. 445, 671-5. *Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation.*
- [60] S. Kriaucionis and N. Heintz, *Science* **2009**. 324, 929-930. *The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain.*
- [61] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao, *Science* **2009**. 324, 930-935. *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by mll partner tet1.*
- [62] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, and G. L. Xu, *Science* **2011**. 333, 1303-7. *Tet-mediated formation of 5-carboxylcytosine and its excision by tdg in mammalian DNA.*
- [63] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang, *Science* **2011**. 333, 1300-1303. *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.*
- [64] T. Pfaffeneder, B. Hackner, M. Truss, M. Munzel, M. Müller, C. A. Deiml, C. Hagemeyer, and T. Carell, *Angew. Chem. Int. Ed. Engl.* **2011**. 50, 7008-12. *The discovery of 5-formylcytosine in embryonic stem cell DNA.*
- [65] A. Maiti and A. C. Drohat, *J. Biol. Chem.* **2011**. 286, 35334-8. *Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: Potential implications for active demethylation of cpG sites.*
- [66] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. Jansen, C. Bauer, M. Munzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. Van Ingen, H. Leonhardt, T. Carell, and M. Vermeulen, *Cell* **2013**. 152, 1146-59. *Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives.*

- [67] M. Iurlaro, G. Ficiz, D. Oxley, E. A. Raiber, M. Bachman, M. J. Booth, S. Andrews, S. Balasubramanian, and W. Reik, *Genome Biol.* **2013.** *14,* R119. *A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation.*
- [68] M. R. Wilkins, C. Pasquali, R. D. Appel, K. Ou, O. Golaz, J. C. Sanchez, J. X. Yan, A. A. Gooley, G. Hughes, I. Humphery-Smith, K. L. Williams, and D. F. Hochstrasser, *Biotechnology (N Y)* **1996.** *14,* 61-5. *From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis.*
- [69] H. J. Hartwell Lh, Leibler S, Murray Aw, *Nature* **1999.** *402,* C47-52. *From molecular to modular cell biology.*
- [70] J. E. Celis and J. M. Moreira, *Mol. Cell. Proteomics* **2008.** *7,* 1779. *Clinical proteomics.*
- [71] M. Tyers and M. Mann, *Nature* **2003.** *422,* 193-197. *From genomics to proteomics.*
- [72] J. Klose, *Humangenetik* **1975.** *26,* 231-43. *Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals.*
- [73] P. H. O'farrell, *J. Biol. Chem.* **1975.** *250,* 4007-21. *High resolution two-dimensional electrophoresis of proteins.*
- [74] S. P. Gygi, G. L. Corthals, Y. Zhang, Y. Rochon, and R. Aebersold, *Proc. Natl. Acad. Sci. USA* **2000.** *97,* 9390-5. *Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology.*
- [75] T. M. a. C. Hilgendorf, *Eur. Pharm. Rev.* **2012.** *17,* 9-12. *Quantification of membrane drug transporters and application in drug discovery and development.*
- [76] D. Nedelkov, U. A. Kiernan, E. E. Niederkofler, K. A. Tubbs, and R. W. Nelson, *Mol. Cell. Proteomics* **2006.** *5,* 1811-8. *Population proteomics: The concept, attributes, and potential for cancer biomarker research.*
- [77] A. Oliva, J. B. Farina, and M. Llabres, *Curr. Pharm. Anal.* **2007.** *3,* 230-248. *New trends in analysis of biopharmaceutical products.*
- [78] M. R. Emmett and R. M. Caprioli, *J. Am. Soc. Mass Spectrom.* **1994.** *5,* 605-13. *Micro-electrospray mass spectrometry: Ultra-high-sensitivity analysis of peptides and proteins.*
- [79] J. A. Anspach, T. D. Maloney, R. W. Brice, and L. A. Colon, *Anal. Chem.* **2005.** *77,* 7489-94. *Injection valve for ultrahigh-pressure liquid chromatography.*

- [80] J. A. Anspach, T. D. Maloney, and L. A. Colon, *J. Sep. Sci.* **2007**. *30*, 1207-13. *Ultrahigh-pressure liquid chromatography using a 1-mm id column packed with 1.5-microm porous particles.*
- [81] R. G. Kay, B. Gregory, P. B. Grace, and S. Pleasance, *Rapid Commun. Mass Spectrom.* **2007**. *21*, 2585-93. *The application of ultra-performance liquid chromatography/tandem mass spectrometry to the detection and quantitation of apolipoproteins in human serum.*
- [82] M. P. Lam, S. O. Siu, E. Lau, X. Mao, H. Z. Sun, P. C. Chiu, W. S. Yeung, D. M. Cox, and I. K. Chu, *Anal. Bioanal. Chem.* **2010**. *398*, 791-804. *Online coupling of reverse-phase and hydrophilic interaction liquid chromatography for protein and glycoprotein characterization.*
- [83] G. Han, M. Ye, and H. Zou, *Analyst* **2008**. *133*, 1128-38. *Development of phosphopeptide enrichment techniques for phosphoproteome analysis.*
- [84] K. Biemann, *Annu. Rev. Biochem.* **1992**. *61*, 977-1010. *Mass spectrometry of peptides and proteins.*
- [85] K. Biemann and S. A. Martin, *Mass Spectrom. Rev.* **1987**. *6*, 1-75. *Mass spectrometric determination of the amino acid sequence of peptides and proteins.*
- [86] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, *Science* **1989**. *246*, 64-71. *Electrospray ionization for mass spectrometry of large biomolecules.*
- [87] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, *Mass Spectrom. Rev.* **1990**. *9*, 37-70. *Electrospray ionization-principles and practice.*
- [88] M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, and M. Mann, *Nature* **1996**. *379*, 466-9. *Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry.*
- [89] M. Karas and F. Hillenkamp, *Anal. Chem.* **1988**. *60*, 2299-301. *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.*
- [90] B. Domon and R. Aebersold, *Nat. Biotechnol.* **2010**. *28*, 710-21. *Options and considerations when selecting a quantitative proteomics strategy.*
- [91] J. R. Yates, C. I. Ruse, and A. Nakorchevsky, *Annu. Rev. Biomed. Eng.* **2009**. *11*, 49-79. *Proteomics by mass spectrometry: Approaches, advances, and applications.*
- [92] K. G. Standing, *Curr. Opin. Struct. Biol.* **2003**. *13*, 595-601. *Peptide and protein de novo sequencing by mass spectrometry.*
- [93] K. A. Resing and N. G. Ahn, *FEBS Letters* **2005**. *579*, 885-889. *Proteomics strategies for protein identification.*

- [94] A. Bensimon, A. J. Heck, and R. Aebersold, *Annu. Rev. Biochem.* **2012**. *81*, 379-405. *Mass spectrometry-based proteomics and network biology.*
- [95] C. H. Borchers, R. Thapar, E. V. Petrotchenko, M. P. Torres, J. P. Speir, M. Easterling, Z. Dominski, and W. F. Marzluff, *Proc. Natl. Acad. Sci. USA* **2006**. *103*, 3094-9. *Combined top-down and bottom-up proteomics identifies a phosphorylation site in stem-loop-binding proteins that contributes to high-affinity rna binding.*
- [96] M. Karas, D. Bachmann, U. Bahr, and F. Hillenkamp, *Int. J. Mass Spectrom. Ion Processes* **1987**. *78*, 53-68. *Matrix-assisted ultraviolet laser desorption of non-volatile compounds.*
- [97] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, and T. Matsuo, *Rapid Commun. Mass Spectrom.* **1988**. *2*, 151-153. *Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry.*
- [98] F. Xiang and R. C. Beavis, *Org. Mass Spectrom.* **1993**. *28*, 1424-1429. *Growing protein-doped synaptic acid crystals for laser desorption: An alternative preparation method for difficult samples.*
- [99] F. Xiang, R. C. Beavis, and W. Ens, *Rapid Commun. Mass Spectrom.* **1994**. *8*, 199-204. *A method to increase contaminant tolerance in protein matrix-assisted laser desorption/ionization by the fabrication of thin protein-doped polycrystalline films.*
- [100] B. M. Ham, J. T. Jacob, and R. B. Cole, *Anal. Chem.* **2005**. *77*, 4439-47. *Maldi-tof ms of phosphorylated lipids in biological fluids using immobilized metal affinity chromatography and a solid ionic crystal matrix.*
- [101] D. W. Armstrong, L.-K. Zhang, L. He, and M. L. Gross, *Anal. Chem.* **2001**. *73*, 3679-3686. *Ionic liquids as matrixes for matrix-assisted laser desorption/ionization mass spectrometry.*
- [102] R. Lemaire, M. Wisztorski, A. Desmons, J. C. Tabet, R. Day, M. Salzet, and I. Fournier, *Anal. Chem.* **2006**. *78*, 7145-53. *Maldi-ms direct tissue analysis of proteins: Improving signal sensitivity using organic treatments.*
- [103] G. Taylor, *Proc. R. Soc. London, Ser. A* **1964**. *280*, 383-397. *Disintegration of water drops in an electric field.*
- [104] J. Zeleny, *Physical Review* **1917**. *10*, 1-6. *Instability of electrified liquid surfaces.*
- [105] M. S. Wilm and M. Mann, *Int. J. Mass Spectrom. Ion Processes* **1994**. *136*, 167-180. *Electrospray and taylor-cone theory, dole's beam of macromolecules at last?*
- [106] I. Marginean, P. Nemes, L. Parvin, and A. Vertes, *Appl. Phys. Lett.* **2006**. *89*, 064104. *How much charge is there on a pulsating taylor cone?*

- [107] P. Nemes, I. Marginean, and A. Vertes, *Anal. Chem.* **2007**. *79*, 3105-16. *Spraying mode effect on droplet formation and ion chemistry in electrosprays.*
- [108] M. Wilm and M. Mann, *Anal. Chem.* **1996**. *68*, 1-8. *Analytical properties of the nanoelectrospray ion source.*
- [109] Y. Li and R. B. Cole, *Anal. Chem.* **2003**. *75*, 5739-46. *Shifts in peptide and protein charge state distributions with varying spray tip orifice diameter in nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry.*
- [110] A. El-Faramawy, K. W. Siu, and B. A. Thomson, *J. Am. Soc. Mass Spectrom.* **2005**. *16*, 1702-7. *Efficiency of nano-electrospray ionization.*
- [111] J. W. Hager and J. C. Le Blanc, *J. Chromatogr. A* **2003**. *1020*, 3-9. *High-performance liquid chromatography-tandem mass spectrometry with a new quadrupole/linear ion trap instrument.*
- [112] J. C. Schwartz, M. W. Senko, and J. E. Syka, *J. Am. Soc. Mass Spectrom.* **2002**. *13*, 659-69. *A two-dimensional quadrupole ion trap mass spectrometer.*
- [113] D. J. Douglas, A. J. Frank, and D. Mao, *Mass Spectrom. Rev.* **2005**. *24*, 1-29. *Linear ion traps in mass spectrometry.*
- [114] J. R. Yates, D. Cociorva, L. Liao, and V. Zabrouskov, *Anal. Chem.* **2006**. *78*, 493-500. *Performance of a linear ion trap-orbitrap hybrid for peptide analysis.*
- [115] M. B. Comisarow and A. G. Marshall, *J. Mass Spectrom.* **1996**. *31*, 581-5. *The early development of fourier transform ion cyclotron resonance (ft-icr) spectroscopy.*
- [116] M. W. Senko, J. D. Canterbury, S. Guan, and A. G. Marshall, *Rapid Commun. Mass Spectrom.* **1996**. *10*, 1839-44. *A high-performance modular data system for fourier transform ion cyclotron resonance mass spectrometry.*
- [117] Q. Hu, R. J. Noll, H. Li, A. Makarov, M. Hardman, and R. Graham Cooks, *J. Mass Spectrom.* **2005**. *40*, 430-43. *The orbitrap: A new mass spectrometer.*
- [118] A. Makarov, *Anal. Chem.* **2000**. *72*, 1156-62. *Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis.*
- [119] X. Han, A. Aslanian, and J. R. Yates, 3rd, *Curr. Opin. Chem. Biol.* **2008**. *12*, 483-90. *Mass spectrometry for proteomics.*
- [120] R. H. Perry, R. G. Cooks, and R. J. Noll, *Mass Spectrom. Rev.* **2008**. *27*, 661-699. *Orbitrap mass spectrometry: Instrumentation, ion motion and applications.*

- [121] J. V. Olsen, L. M. De Godoy, G. Li, B. Macek, P. Mortensen, R. Pesch, A. Makarov, O. Lange, S. Horning, and M. Mann, *Mol. Cell. Proteomics* **2005**. *4*, 2010-21. *Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a c-trap.*
- [122] A. Makarov, E. Denisov, O. Lange, and S. Horning, *J. Am. Soc. Mass Spectrom.* **2006**. *17*, 977-82. *Dynamic range of mass accuracy in Itq orbitrap hybrid mass spectrometer.*
- [123] J. E. P. Syka, J. A. Marto, D. L. Bai, S. Horning, M. W. Senko, J. C. Schwartz, B. Ueberheide, B. Garcia, S. Busby, T. Muratore, J. Shabanowitz, and D. F. Hunt, *J. Proteome Res.* **2004**. *3*, 621-626. *Novel linear quadrupole ion trap/ft mass spectrometer: Performance characterization and use in the comparative analysis of histone h3 post-translational modifications.*
- [124] K. Breuker, M. Jin, X. Han, H. Jiang, and F. W. McLafferty, *J. Am. Soc. Mass Spectrom.* **2008**. *19*, 1045-53. *Top-down identification and characterization of biomolecules by mass spectrometry.*
- [125] M. W. Senko, C. L. Hendrickson, M. R. Emmett, S. D. H. Shi, and A. G. Marshall, *J. Am. Soc. Mass Spectrom.* **1997**. *8*, 970-976. *External accumulation of ions for enhanced electrospray ionization fourier transform ion cyclotron resonance mass spectrometry.*
- [126] M. Hardman and A. A. Makarov, *Anal. Chem.* **2003**. *75*, 1699-705. *Interfacing the orbitrap mass analyzer to an electrospray ion source.*
- [127] M. Bantscheff, S. Lemeer, M. M. Savitski, and B. Kuster, *Anal. Bioanal. Chem.* **2012**. *404*, 939-65. *Quantitative mass spectrometry in proteomics: Critical review update from 2007 to the present.*
- [128] M. Bantscheff, M. Schirle, G. Sweetman, J. Rick, and B. Kuster, *Anal. Bioanal. Chem.* **2007**. *389*, 1017-31. *Quantitative mass spectrometry in proteomics: A critical review.*
- [129] S.-E. Ong and M. Mann, *Nat. Chem. Biol.* **2005**. *1*, 252-262. *Mass spectrometry-based proteomics turns quantitative.*
- [130] R. D. Voyksner and H. Lee, *Rapid Commun. Mass Spectrom.* **1999**. *13*, 1427-37. *Investigating the use of an octupole ion guide for ion storage and high-pass mass filtering to improve the quantitative performance of electrospray ion trap mass spectrometry.*
- [131] W. Zhu, J. W. Smith, and C. M. Huang, *J. Biomed. Biotechnol.* **2010**. *2010*, 840518. *Mass spectrometry-based label-free quantitative proteomics.*
- [132] H. Liu, R. G. Sadygov, and J. R. Yates, 3rd, *Anal. Chem.* **2004**. *76*, 4193-201. *A model for random sampling and estimation of relative protein abundance in shotgun proteomics.*
- [133] W. M. Old, K. Meyer-Arendt, L. Aveline-Wolf, K. G. Pierce, A. Mendoza, J. R. Sevensky, K. A. Resing, and N. G. Ahn, *Mol. Cell. Proteomics* **2005**. *4*, 1487-502. *Comparison of label-free methods for quantifying human proteins by shotgun proteomics.*

- [134] S. A. Gerber, J. Rush, O. Stemman, M. W. Kirschner, and S. P. Gygi, *Proc. Natl. Acad. Sci. USA* **2003**. *100*, 6940-5. *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem ms.*
- [135] L. Anderson and C. L. Hunter, *Mol. Cell. Proteomics* **2006**. *5*, 573-88. *Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins.*
- [136] H. Keshishian, T. Addona, M. Burgess, E. Kuhn, and S. A. Carr, *Mol. Cell. Proteomics* **2007**. *6*, 2212-29. *Quantitative, multiplexed assays for low abundance proteins in plasma by targeted mass spectrometry and stable isotope dilution.*
- [137] J. F. Pickup and K. Mcpherson, *Anal. Chem.* **1976**. *48*, 1885-1890. *Theoretical considerations in stable isotope dilution mass spectrometry for organic analysis.*
- [138] S. E. Ong, *Mol. Cell. Proteomics* **2002**. *1*, 376-386. *Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics.*
- [139] Y. Oda, K. Huang, F. R. Cross, D. Cowburn, and B. T. Chait, *Proc. Natl. Acad. Sci. USA* **1999**. *96*, 6591-6. *Accurate quantitation of protein expression and site-specific phosphorylation.*
- [140] C. C. Wu, M. J. Maccoss, K. E. Howell, D. E. Matthews, and J. R. Yates, 3rd, *Anal. Chem.* **2004**. *76*, 4951-9. *Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis.*
- [141] W. Schutz, N. Hausmann, K. Krug, R. Hampp, and B. Macek, *Plant Cell* **2011**. *23*, 1701-5. *Extending silac to proteomics of plant cell lines.*
- [142] J. Krijgsveld, R. F. Ketting, T. Mahmoudi, J. Johansen, M. Artal-Sanz, C. P. Verrijzer, R. H. A. Plasterk, and A. J. R. Heck, *Nat. Biotech.* **2003**. *21*, 927-931. *Metabolic labeling of c. Elegans and d. Melanogaster for quantitative proteomics.*
- [143] K. J. Reynolds, C. Fenselau, *Curr. Protoc. Protein Sci.* **2004**. 23.4.1-23.4.8. *Quantitative protein analysis using proteolytic [¹⁸O]water labeling.*
- [144] C. J. Mason, T. M. Therneau, J. E. Eckel-Passow, K. L. Johnson, A. L. Oberg, J. E. Olson, K. S. Nair, D. C. Muddiman, and H. R. Bergen, 3rd, *Mol. Cell. Proteomics* **2007**. *6*, 305-18. *A method for automatically interpreting mass spectra of 18o-labeled isotopic clusters.*
- [145] S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold, *Nat. Biotech.* **1999**. *17*, 994-999. *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.*
- [146] M. Shen, L. Guo, A. Wallace, J. Fitzner, J. Eisenman, E. Jacobson, and R. S. Johnson, *Mol. Cell. Proteomics* **2003**. *2*, 315-24. *Isolation and isotope labeling of cysteine- and methionine-containing tryptic peptides: Application to the study of cell surface proteolysis.*

- [147] C. Pasquarello, J. C. Sanchez, D. F. Hochstrasser, and G. L. Corthals, *Rapid Commun. Mass Spectrom.* **2004**. *18*, 117-27. *N-t-butyliodoacetamide and iodoacetanilide: Two new cysteine alkylating reagents for relative quantitation of proteins.*
- [148] Y. Shi, R. Xiang, J. K. Crawford, C. M. Colangelo, C. Horvath, and J. A. Wilkins, *J. Proteome Res.* **2004**. *3*, 104-11. *A simple solid phase mass tagging approach for quantitative proteomics.*
- [149] Y. Shi, R. Xiang, C. Horvath, and J. A. Wilkins, *J. Proteome Res.* **2005**. *4*, 1427-33. *Quantitative analysis of membrane proteins from breast cancer cell lines bt474 and mcf7 using multistep solid phase mass tagging and 2d lc/ms.*
- [150] Y. Qiu, E. A. Sousa, R. M. Hewick, and J. H. Wang, *Anal. Chem.* **2002**. *74*, 4969-79. *Acid-labile isotope-coded extractants: A class of reagents for quantitative mass spectrometric analysis of complex protein mixtures.*
- [151] R. Sebastiano, A. Citterio, M. Lapadula, and P. G. Righetti, *Rapid Commun. Mass Spectrom.* **2003**. *17*, 2380-6. *A new deuterated alkylating agent for quantitative proteomics.*
- [152] A. Schmidt, J. Kellermann, and F. Lottspeich, *Proteomics* **2005**. *5*, 4-15. *A novel strategy for quantitative proteomics using isotope-coded protein labels.*
- [153] P. L. Ross, Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, and D. J. Pappin, *Mol. Cell. Proteomics* **2004**. *3*, 1154-69. *Multiplexed protein quantitation in *saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents.*
- [154] A. Thompson, J. Schafer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, R. Johnstone, A. K. Mohammed, and C. Hamon, *Anal. Chem.* **2003**. *75*, 1895-904. *Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by ms/ms.*
- [155] J. Ji, A. Chakraborty, M. Geng, X. Zhang, A. Amini, M. Bina, and F. Regnier, *J. Chromatogr. B Biomed. Sci. Appl.* **2000**. *745*, 197-210. *Strategy for qualitative and quantitative analysis in proteomics based on signature peptides.*
- [156] F. Y. Che and L. D. Fricker, *Anal. Chem.* **2002**. *74*, 3190-8. *Quantitation of neuropeptides in *cpe(fat)/cpe(fat)* mice using differential isotopic tags and mass spectrometry.*
- [157] X. Zhang, Q. K. Jin, S. A. Carr, and R. S. Annan, *Rapid Commun. Mass Spectrom.* **2002**. *16*, 2325-32. *N-terminal peptide labeling strategy for incorporation of isotopic tags: A method for the determination of site-specific absolute phosphorylation stoichiometry.*
- [158] M. O. Glocker, C. Borchers, W. Fiedler, D. Suckau, and M. Przybylski, *Bioconjug. Chem.* **1994**. *5*, 583-90. *Molecular characterization of surface topology in protein tertiary structures by amino-acylation and mass spectrometric peptide mapping.*

- [159] S. Wiese, K. A. Reidegeld, H. E. Meyer, and B. Warscheid, *Proteomics* **2007**. 7, 340-50. *Protein labeling by itraq: A new tool for quantitative mass spectrometry in proteome research.*
- [160] L. Choe, M. D'ascenzo, N. R. Relkin, D. Pappin, P. Ross, B. Williamson, S. Guertin, P. Pribil, and K. H. Lee, *Proteomics* **2007**. 7, 3651-60. *8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for alzheimer's disease.*
- [161] A. Pierce, R. D. Unwin, C. A. Evans, S. Griffiths, L. Carney, L. Zhang, E. Jaworska, C. F. Lee, D. Blinco, M. J. Okoniewski, C. J. Miller, D. A. Bitton, E. Spooncer, and A. D. Whetton, *Mol. Cell. Proteomics* **2008**. 7, 853-63. *Eight-channel itraq enables comparison of the activity of six leukemogenic tyrosine kinases.*
- [162] G. C. Mcalister, E. L. Huttlin, W. Haas, L. Ting, M. P. Jedrychowski, J. C. Rogers, K. Kuhn, I. Pike, R. A. Grothe, J. D. Blethrow, and S. P. Gygi, *Anal. Chem.* **2012**. 84, 7469-78. *Increasing the multiplexing capacity of tmts using reporter ion isotopologues with isobaric masses.*
- [163] P. Feist and A. Hummon, *Int. J. Mass Spectrom. Ion Processes* **2015**. 16, 3537. *Proteomic challenges: Sample preparation techniques for microgram-quantity protein analysis from biological samples.*
- [164] R. L. Gundry, M. Y. White, C. I. Murray, L. A. Kane, Q. Fu, B. A. Stanley, and J. E. Van Eyk, *Curr. Protoc. Mol. Biol.* **2009**. Chapter 10, Unit10 25. *Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow.*
- [165] D. Ratner, *J. Mol. Biol.* **1974**. 88, 373-83. *The interaction bacterial and phage proteins with immobilized escherichia coli rna polymerase.*
- [166] S. Fields and O.-K. Song, *Nature* **1989**. 340, 245-246. *A novel genetic system to detect protein-protein interactions.*
- [167] J. R. Glenney, Jr. and K. Weber, *Methods Enzymol.* **1983**. 102, 204-10. *Detection of calmodulin-binding polypeptides separated in sds-polyacrylamide gels by a sensitive [¹²⁵i]calmodulin gel overlay assay.*
- [168] O. Soderberg, K. J. Leuchowius, M. Gullberg, M. Jarvius, I. Weibrecht, L. G. Larsson, and U. Landegren, *Methods* **2008**. 45, 227-32. *Characterizing proteins and their interactions in cells and tissues using the in situ proximity ligation assay.*
- [169] J. De Gunzburg, R. Riehl, and R. A. Weinberg, *Proc. Natl. Acad. Sci. USA* **1989**. 86, 4007-11. *Identification of a protein associated with p21ras by chemical crosslinking.*
- [170] D. W. Thomas, L. Eades, C. Wilson, and M. J. Solvay, *Cell Immunol.* **1985**. 96, 351-62. *Modulation of ia and photoreactive antigen on antigen-presenting cells: Fun with a photoprobe.*

- [171] J. K. Eng, A. L. McCormack, and J. R. Yates, *J. Am. Soc. Mass Spectrom.* **1994**. *5*, 976-89. *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.*
- [172] D. N. Perkins, D. J. Pappin, D. M. Creasy, and J. S. Cottrell, *Electrophoresis* **1999**. *20*, 3551-67. *Probability-based protein identification by searching sequence databases using mass spectrometry data.*
- [173] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann, *J. Proteome Res.* **2011**. *10*, 1794-805. *Andromeda: A peptide search engine integrated into the maxquant environment.*
- [174] J. A. Paulo, *WebmedCentral* **2013**. *4*, WMCPLS0052. *Practical and efficient searching in proteomics: A cross engine comparison.*
- [175] J. K. Eng, B. C. Searle, K. R. Clauser, and D. L. Tabb, *Mol. Cell. Proteomics* **2011**. *10*, R111.009522. *A face in the crowd: Recognizing peptides through database search.*
- [176] J. E. Elias and S. P. Gygi, *Nat. Methods* **2007**. *4*, 207-14. *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.*
- [177] B. Cooper, *J. Proteome Res.* **2011**. *10*, 1432-5. *The problem with peptide presumption and low mascot scoring.*
- [178] M. Winnacker, S. Breeger, R. Strasser, and T. Carell, *Chem. Bio. Chem.* **2009**. *10*, 109-18. *Novel diazirine-containing DNA photoaffinity probes for the investigation of DNA-protein-interactions.*
- [179] M. Winnacker, V. Welzmler, R. Strasser, and T. Carell, *Chem. Bio. Chem.* **2010**. *11*, 1345-9. *Development of a DNA photoaffinity probe for the analysis of 8-oxodg-binding proteins in a human proteome.*
- [180] C. N. Buechner, K. Heil, G. Michels, T. Carell, C. Kisker, and I. Tessmer, *J. Biol. Chem.* **2014**. *289*, 3613-24. *Strand-specific recognition of DNA damages by xpd provides insights into nucleotide excision repair substrate versatility.*
- [181] E. L. Hendrickson, Q. Xia, T. Wang, J. A. Leigh, and M. Hackett, *Analyst* **2006**. *131*, 1335-41. *Comparison of spectral counting and metabolic stable isotope labeling for use with quantitative microbial proteomics.*
- [182] C. H. Sohn, J. E. Lee, M. J. Sweredoski, R. L. Graham, G. T. Smith, S. Hess, G. Czerwieniec, J. A. Loo, R. J. Deshaies, and J. L. Beauchamp, *J. Am. Chem. Soc.* **2012**. *134*, 2672-80. *Click chemistry facilitates formation of reporter ions and simplified synthesis of amine-reactive multiplexed isobaric tags for protein quantification.*

- [183] L. Dayon, N. Turck, S. Kienle, P. Schulz-Knappe, D. F. Hochstrasser, A. Scherl, and J. C. Sanchez, *Anal. Chem.* **2010**. *82*, 848-58. *Isobaric tagging-based selection and quantitation of cerebrospinal fluid tryptic peptides with reporter calibration curves.*
- [184] E. Ernoult, E. Gamelin, and C. Guette, *Proteome Sci.* **2008**. *6*, 27. *Improved proteome coverage by using itraq labelling and peptide offgel fractionation.*
- [185] R. M. Daniel, M. J. Danson, and R. Eisenthal, *Trends Biochem. Sci.* **2001**. *26*, 223-5. *The temperature optima of enzymes: A new perspective on an old phenomenon.*
- [186] M. C. Willis, B. J. Hicke, O. C. Uhlenbeck, T. R. Cech, and T. H. Koch, *Science* **1993**. *262*, 1255-7. *Photocrosslinking of 5-iodouracil-substituted rna and DNA to proteins.*
- [187] S. Madler, C. Bich, D. Touboul, and R. Zenobi, *J. Mass Spectrom.* **2009**. *44*, 694-706. *Chemical cross-linking with nhs esters: A systematic study on amino acid reactivities.*
- [188] U. K. Shigdel, J. Zhang, and C. He, *Angew. Chem. Int. Ed. Engl.* **2008**. *47*, 90-3. *Diazirine-based DNA photo-cross-linking probes for the study of protein-DNA interactions.*
- [189] H. L. Lewis, D. R. Muhleman, and J. F. Ward, *Radiat. Res.* **1978**. *75*, 305-16. *Serologic assay of DNA base damage. I. 5-hydroxymethyldeoxyuridine, a radiation product of thymidine.*
- [190] R. J. Boorstein, A. Cummings, Jr., D. R. Marenstein, M. K. Chan, Y. Ma, T. A. Neubert, S. M. Brown, and G. W. Teebor, *J. Biol. Chem.* **2001**. *276*, 41991-7. *Definitive identification of mammalian 5-hydroxymethyluracil DNA n-glycosylase activity as smug1.*
- [191] B. Kavli, M. Otterlei, G. Slupphaug, and H. E. Krokan, *DNA Repair (Amst.)* **2007**. *6*, 505-16. *Uracil in DNA--general mutagen, but normal intermediate in acquired immunity.*
- [192] F. Santos, J. Peat, H. Burgess, C. Rada, W. Reik, and W. Dean, *Epigenetics Chromatin* **2013**. *6*, 39. *Active demethylation in mouse zygotes involves cytosine deamination and base excision repair.*
- [193] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, and A. Bellacosa, *Cell* **2011**. *146*, 67-79. *Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair.*
- [194] J. Silva and A. Smith, *Cell* **2008**. *132*, 532-6. *Capturing pluripotency.*
- [195] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, and T. Carell, *PLoS ONE* **2010**. *5*, e15367. *Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates.*

- [196] M. A. Hahn, R. Qiu, X. Wu, A. X. Li, H. Zhang, J. Wang, J. Jui, S. G. Jin, Y. Jiang, G. P. Pfeifer, and Q. Lu, *Cell Rep.* **2013**. 3, 291-300. *Dynamics of 5-hydroxymethylcytosine and chromatin marks in mammalian neurogenesis.*
- [197] M. Munzel, D. Globisch, T. Bruckl, M. Wagner, V. Welzmler, S. Michalakis, M. Müller, M. Biel, and T. Carell, *Angew. Chem. Int. Ed. Engl.* **2010**. 49, 5375-7. *Quantification of the sixth DNA base hydroxymethylcytosine in the brain.*
- [198] B. M. Colquitt, W. E. Allen, G. Barnea, and S. Lomvardas, *Proc. Natl. Acad. Sci. USA* **2013**. 110, 14682-7. *Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity.*
- [199] M. Mellen, P. Ayata, S. Dewell, S. Kriaucionis, and N. Heintz, *Cell* **2012**. 151, 1417-30. *Mecp2 binds to 5hmc enriched within active genes and accessible chromatin in the nervous system.*
- [200] C. X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C. H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, and C. He, *Nat. Biotechnol.* **2011**. 29, 68-72. *Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine.*
- [201] K. E. Szulwach, X. Li, Y. Li, C. X. Song, H. Wu, Q. Dai, H. Irier, A. K. Upadhyay, M. Gearing, A. I. Levey, A. Vasanthakumar, L. A. Godley, Q. Chang, X. Cheng, C. He, and P. Jin, *Nat. Neurosci.* **2011**. 14, 1607-16. *5-hmc-mediated epigenetic dynamics during postnatal neurodevelopment and aging.*
- [202] A. Feldmann, R. Ivanek, R. Murr, D. Gaidatzis, L. Burger, and D. Schubeler, *PLoS Genet.* **2013**. 9, e1003994. *Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions.*
- [203] F. Santos, B. Hendrich, W. Reik, and W. Dean, *Dev. Biol.* **2002**. 241, 172-82. *Dynamic reprogramming of DNA methylation in the early mouse embryo.*
- [204] M. A. Surani, *Nature* **2001**. 414, 122-128. *Reprogramming of genome function through epigenetic inheritance.*
- [205] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen, and C. Von Mering, *Nucleic Acids Res.* **2015**. 43, D447-52. *String v10: Protein-protein interaction networks, integrated over the tree of life.*
- [206] T. S. Furey, *Nat. Rev. Genet.* **2012**. 13, 840-52. *Chip-seq and beyond: New and improved methodologies to detect and characterize protein-DNA interactions.*
- [207] J. H. Lee, S. R. Hart, and D. G. Skalnik, *Genesis* **2004**. 38, 32-8. *Histone deacetylase activity is required for embryonic stem cell differentiation.*

- [208] S. W. Harshman, N. L. Young, M. R. Parthun, and M. A. Freitas, *Nucleic Acids Res.* **2013**. *41*, 9593-609. *H1 histones: Current perspectives and challenges.*
- [209] Y. Fan, T. Nikitina, E. M. Morin-Kensicki, J. Zhao, T. R. Magnuson, C. L. Woodcock, and A. I. Skoultchi, *Mol. Cell. Biol.* **2003**. *23*, 4559-72. *H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo.*
- [210] B. J. K. Seung-Min Yang, Laura Norwood Toro, and Arthur I. Skoultchi, *PNAS* **2013**. *110*, 1708-1713. *H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone h3 methylation.*
- [211] A. Z. Jacek R. Wisniewski, Sonja Krüger, and Matthias Mann, *Mol. Cell. Proteomics* **2007**. *6*, 72-87. *Mass spectrometric mapping of linker histone h1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue.*
- [212] P. J. Horn, L. M. Carruthers, C. Logie, D. A. Hill, M. J. Solomon, P. A. Wade, A. N. Imbalzano, J. C. Hansen, and C. L. Peterson, *Nat. Struct. Biol.* **2002**. *9*, 263-7. *Phosphorylation of linker histones regulates atp-dependent chromatin remodeling enzymes.*
- [213] M. T. Masataka Furuya, Takahide Teranishi,, Y. H. Kazuya Matsumoto, Kazuhiro Saeki3, Hitoshi Ishimoto,, and A. I. a. Y. Y. Kazuhiro Minegishi, *J. Reprod. Dev.* **2007**. *53*, 895-902. *H1foo is indispensable for meiotic maturation of the mouse oocyte.*
- [214] W. E. Allen M. Sirotkin, Genhong Cheng, Andres Klein-Szanto, Raju Kucherlapat, and A. a. I. Skoultchi, *Proc. Natl. Acad. Sci. USA* **1995**. *92*, 6434-6438. *Mice develop normally without the hi0 linker histone.*
- [215] W. Huang Da, B. T. Sherman, and R. A. Lempicki, *Nat. Protoc.* **2009**. *4*, 44-57. *Systematic and integrative analysis of large gene lists using david bioinformatics resources.*
- [216] W. Huang Da, B. T. Sherman, and R. A. Lempicki, *Nucleic Acids Res.* **2009**. *37*, 1-13. *Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists.*
- [217] C. Von Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork, and B. Snel, *Nucleic Acids Res.* **2003**. *31*, 258-61. *String: A database of predicted functional associations between proteins.*
- [218] R. Welz and S. Müller, *Tetrahedron Lett.* **2002**. *43*, 795-797. *5-(benzylmercapto)-1h-tetrazole as activator for 2'-o-tbdms phosphoramidite building blocks in rna synthesis.*
- [219] J. D. Dignam, R. M. Lebovitz, and R. G. Roeder, *Nucleic Acids. Res.* **1983**. *11*, 1475-89. *Accurate transcription initiation by rna polymerase ii in a soluble extract from isolated mammalian nuclei.*

- [220] I. Guillemin, M. Becker, K. Ociepka, E. Friauf, and H. G. Nothwang, *Proteomics* **2005**. 5, 35-45. *A subcellular prefractionation protocol for minute amounts of mammalian cell cultures and tissue.*
- [221] D. Shechter, H. L. Dormann, C. D. Allis, and S. B. Hake, *Nat. Protoc.* **2007**. 2, 1445-57. *Extraction, purification and analysis of histones.*
- [222] J. Rappsilber, M. Mann, and Y. Ishihama, *Nat. Protoc.* **2007**. 2, 1896-906. *Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using stagetips.*