Chemical tools for the study of epigenetic mechanisms

A thesis submitted to the Board of the Faculty of Physical Sciences in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (D.Phil)

at the

University of Oxford

by

Lukas Lercher

Somerville College, University of Oxford

June 2014

Abstract

The overall goal of my work was to develop and apply new chemical methods for the study of epigenetic DNA and protein modifications.

In Chapter 3 the development of Suzuki-Miyaura cross coupling (SMcc) for the post-synthetic modification of DNA is described. DNA modification by SMcc is efficient (4-6h) and proceeds under mild conditions (37°C, pH 8.5). The incorporation of various groups useful for biological investigations is demonstrated using this methodology. Using a photocrosslinker, introduced into the DNA by SMcc capture experiments are performed to identify potential binding partners of modified DNA.

In Chapter 4 a dehydroalanine (Dha) based chemical protein modification method is described that enables the introduction of posttranslational modification (PTM) mimics into histones. The PTM mimics introduced by this method are tested using western- and dot-blot and binding and enzymatic assays, confirming they function as mimics of the natural modifications.

Chapter 5 describes the use of a generated PTM mimics to elucidate the function of *O*-linked β-*N*-acetylglucosamine (GlcNAc) of histones in transcriptional regulation. It is shown that GlcNAcylation of Thr-101 on histone H2A can destabilize nucleosome by modulating the H2A/B dimer – H3/H4 tetramer interface.

N- and *C*-terminal histone tails play an important role in transcriptional regulation. In Chapter 6, nuclear magnetic resonance is used to investigate the structure of the histone H3 *N*-terminal tail in a nucleosome. The H3 tail, while intrinsically disordered, gains some α-helical character and adopts a compact conformation in a nucleosome context. This H3 tail structure is shown to be modulated by Ser-10 phosphorylation.

The effect of a new covalent DNA modification, 5-hydroxymethylcytosine (5hmC), on transcription factor binding is investigated in Chapter 7. 5hmC influences HIF1α/β, USF and MAX binding to their native recognition sequence, implying involvement of this modification in epigenetic regulation.

Acknowledgments

Firstly, I want to thank my supervisors Prof. Ben Davis and Prof. Chris Schofield for their excellent support and supervision during my almost four years of research in their labs.

Prof. Andy Baldwin I want to thank for introducing me into the wondrous world of NMR and the patients that is needed to explain math and other foreign concepts to me. For help with data acquisition, I'd also like to thank Dr. Tim Claridge and Dr. Ivan Leung.

I'd would like to thank Dr. Skirmantas Kriaucionis and Dr. Jingrong Min for hosting me in their laboratories at the Ludwig institute and the SGC Toronto respectively.

Of course, I wouldn't have been able to accomplish all of this work by myself. For fruitful collaborations and discussions I'd like to thank: Bhaskar Bushan, Justin Chalker, Dr. Yuya Lin, Dr. Mike McDonough, Dr. Joanne McGouran, Dr. Nisha Patel, Ritu Raj, Dr. Nathan Rose, Dr. Matt Schombs.

For a great working atmosphere I'd like to thank all the BGD group and LG8 especially!

I want to thank Su Pulcu, Arnold Boersma, Sarah Allison, Timmy Choi, Gabriel Villar, Massimiliano Clamer and Dr. Rodolfo Gomez from the HB group for welcoming me in LG6 at the start of my PhD.

I'd like to thank my college advisors Dr. Matt Higgins and Dr. Jonathan Burton for their advice scientific and otherwise.

Nora Bähr I'd like to thank for all of her support and understanding through these four years and for keeping me grounded.

I want to thank my parents for their continuing support throughout my studies.

Abbreviations

^{1}H NMR	proton nuclear magnetic resonance
^{13}C NMR	carbon-13 nuclear magnetic resonance
°C	degree Celsius
[M]	MS (see MS) molecular peak
Amp.	ampicillin
aq.	aqueous
Ac	acetyl
ACN	acetonitrile
Ar	argon
Boc_2O	di-*tert*-butyl dicarbonate
BSA	bolvine serum albumin
calc.	calculated
CD_3OD	deuterated methanol
$CDCl_3$	deuterated chloroform
COSY	correlated spectroscopy
D_2O	deuterated water
Da	dalton
DCC	*N, N*′-dicyclohexylcarbodiimide
DCM	dichloromethane
DMAP	4-dimethylamino pyridine
DNA	2'-deoxyribose nucleic acid
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride
eq.	equivalent
ES	electrospray
Et	ethyl
et al.	et alia
EtOAc	ethylacetate
FA	formic acid
g	gram
GlcNAc	*N*-acetylglucosamine
HMBC	hydrogen multiple bond correlation
HMQC	hydrogen multiple quantum correlated spectroscopy
hpa	3-hydroxypicolinic acid
HRMS	high resolution mass spectrometry
HSQC	heteronuclear single quantum coherence
Hz	hertz
h	hour
IPTG	isopropyl β-D-1-thiogalactopyranoside
J	NMR coupling constant (Hz)
Kan	kanamycin
kbp	kilo-base pair
kDa	kilodalton
LB	Luria Bertani
LC-MS	liquid chromatography-mass spectrometry

LRMS	low resolution mass spectrometry
m	milli
M	molarity
m/z	mass-to-charge ratio
MALDI	matrix-assisted laser desorption/ionization
MeOH	methanol
MHz	megahertz
min	minute
mL	milliliter
mol	mole
MS	mass spectrometry
NEB	New England BioLabs® Inc.
neg.	negative
nm	nanometer
NMR	nuclear magnetic resonance
NMR annotations:	
app.	apparent
br.	broad
s	singlet
d	doublet
dd	doublet doublet
ddd	doublet doublet doublet
t	triplet
q	quartet
m	multiplet
OD	optical density
ODN	oligodeoxynucleotide
ON	oligonucleotide
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pH	$-\log_{10}[H^+]$
pos.	positive
ppm	part per million
PVDF	polyvinylidene fluoride
quant.	quantitative
R	unspecific chemical group
rpm	revolution per minute
rt	room temperature
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
soln.	solution
TB	terrific broth
TEA	triethylamine
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin layer chromatography

Experiments performed in collaboration

Chapter 3

Proteomic analysis was performed by Joanne McGouran (Benedikt Kessler lab, Oxford).

Chapter 4

H3K9Me, H3K9Me$_2$, H3K9Me$_3$, H3 K9Ac, H3 S10Ph proteins were generated by Justin Chalker (University of Tulsa).

Initial HDAC experiments were performed by Justin Chalker (University of Tulsa).

Chapter 5

The H2A T101C protein in Chapter 5 was produced and purified by Joshua Price.

Native mass spectrometry in Chapter 5 was performed by Nisha A Patel (Carol V Robinson group, Oxford).

ChIP experiments in Chapter 5 were performed by Tom Milne (WIMM, Oxford).

Chapter 6

CPMG measurements were performed by Andrew Baldwin (Oxford, physical chemistry)

Table of Contents

1 Introduction

It is often stated that biological information generally flows from DNA via RNA to protein. This 'central dogma' of molecular biology is attributed to Francis Crick towards the end of the 1950s. All eukaryotic organisms contain much more DNA than is necessary to encode the sequences of all expressed proteins. Indeed, in most human cells only 3% of DNA is estimated to be directly protein coding.[1] However this value varies greatly; in a species of bladderwort 36% of DNA is reported to be protein coding.[2] Even of these genes only a small subset is being transcribed into RNA at one time in every cell. The regulation of transcription is essential for almost all biological processes, including development and responses to external cues.

One of the best studied examples of transcriptional regulation is the regulation of lactose metabolising enzymes in *Escherichia coli* (*E. coli*).[3] The genes necessary for lactose metabolism are repressed in the absence of lactose. When lactose is present and in the absence of glucose, the preferred carbon source, repression is relieved and a specific activator is recruited resulting in the up-regulation of these genes. In cases where lactose is the only energy source available the ability to metabolize lactose is essential for the survival of the bacterium. When there is no lactose present, it is proposed that expression of these proteins is be a burden for the cell, since transcription and translation, the synthesis of proteins from RNA, requires energy and the proteins would fulfil no function. Thus regulation of these genes enables optimal growth of the bacterium in all conditions. Transcriptional regulation generally is accomplished by transcription factors (TF) that can either activate or suppress transcription.[4] A combination of repressing and activating mechanisms results in tight transcriptional control of a gene.

TFs are normally modular in nature and contain one domain responsible for localisation by sequence specific DNA binding and one domain that recruits other activating or repressing factors.[4] Prokaryotes have a relatively small genome and the innate selectivity of the TF for its target DNA (6-10 bp) is proposed to be enough to ensure accurate localisation.[5] Eukaryotic cells, however, have

much larger genomes and site discrimination solely based on sequence is not sufficient to accurately target transcriptional activators to their location. For this, a second layer of regulation through organization of the DNA exists, involving the compaction of DNA into a structure called chromatin.[6]

1.1 Chromatin structure

The human genome contains over 6 billion base pairs, which would span over two metres when fully extended.[7] To contain the genome in the cell nucleus with a 6 μm diameter it must be highly compacted. This is accomplished by the condensation of DNA by association of various structural and regulatory proteins into a structure called chromatin. Generally, two types of chromatin are distinguished. Heterochromatin is tightly packaged includes telomers and centromers and is generally inactive in transcription. Euchromatin is more loosely packed leaving DNA exposed and competent for transcription.

The basic packing unit of chromatin is the nucleosome, which in higher eukaryotes contains a histone protein octamer made up of two copies each of H2A, H2B, H3 and H4. A central H3/H4 tetramer is flanked by two H2A/H2B dimers and the histone protein octamer binds around 147 bp of double stranded DNA. The first high resolution structure of this 198 kDa complex was solved by Luger *et al.*[8] more than 15 years ago. This structure revealed a flat disc shaped conformation that bound the DNA in two super-helical turns with both DNA ends extruding on the same face of the nucleosome (**Figure 1a**). Outside the central core nucleosome an additional linker histone H1 can be present,[9] which has been shown to bind near the center of the bound DNA[10] and can promote compaction.[11] Nucleosomes function to structure and compact the genomic DNA, but also fulfil an important role in regulating transcription. Indeed, these two roles are intertwined since the regulation of DNA accessibility and compaction is an important mechanism for transcriptional regulation.

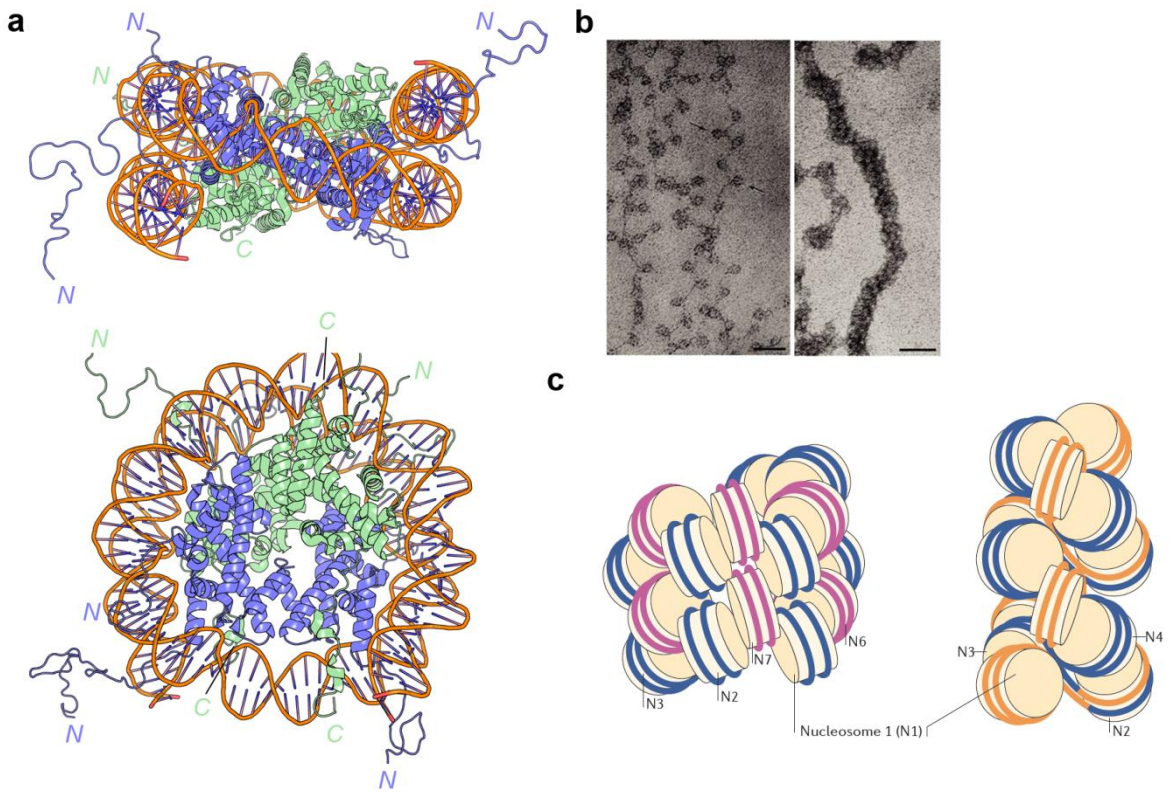


Figure 1. a View from the side and top of the nucleosome structure with the visible *N*- and *C*-terminal tails labelled (adapted from 1KX5).[12] **b** Bead on a chain (left) and compact nucleosome conformation (right; adapted from Olins *et al.)*[13] **c** Proposed modes of inter-nucleosome interactions. In the solenoid interaction model (left) interactions are formed between the n and n+1 nucleosome. In the zig-zag model (right) interactions are formed between n and n+2 nucleosomes (adapted from Luger *et al.*).[14]

When multiple histone octamers are bound to on a long DNA molecule a structure resembling beads on a string is observed (**Figure 1b**).[13,15] This extended form is further condensed in cells into a 30 nm compact chromatin fibre (similar to **Figure 1b**). The exact arrangement of this fibre is still under debate;[14] there are two competing models (**Figure 1c**). In the solenoid model, two consecutive nucleosome are stacked and wound around a helical axis with a helix repeat of around 6 nucleosomes.[16] Alternatively, a zig-zag conformation in which each nucleosome makes contact with the n+2 nucleosome is proposed. The latter model is supported by a crystal structure of a tetranucleosome,[17] which reveals that the four nucleosomes form two separate columns connected by straight linker DNA. However, the crystallised construct has a very short nucleosome repeat length (NRL) and doesn't contain H1, which is not common in packaged eukaryotic chromatin. Since most of

these structural experiments are performed *in vitro* with well-defined nucleosomal arrays containing a constant repeat length, one might argue that they are not a very accurate representation of the *in cell* structure of chromatin. Indeed, different NRL have been shown to affect fibre structure.[18] It is quite possible that *in vivo* both of these models are relevant and contribute to the overall folding of chromatin. In agreement with this, a recent electron microscopy study revealed that interactions from both models are relevant in nucleosome array folding.[19] To increase compaction further the 30 nm fibre can in turn be further packed into larger loops.

Important factors for the accurate compaction of chromatin are the histone tails. The *N*- and *C*-terminal tails of the histone proteins extrude from the structured nucleosome core. The tails are not fully resolved in the available crystal structures and are considered to be unfolded (**Figure 1a**).[14,20,21] Histone tails, like the nucleosome core, function both in structuring chromatin and in transcriptional regulation.

1.2 Histone Tails

In general, the tails of all four histones play a role in various processes relating to nucleosome and chromatin structure, including the regulation of nucleosome positioning, chromatin compaction, and DNA accessibility. They also play an important role in mediating protein-nucleosome interactions.

The presence of all four histone tails is required for efficient formation of compacted chromatin.[22-26] The H2A, H2B and H3 tails herein function in part by interacting with inter-nucleosomal DNA. The basic histone tails interact through electrostatic interactions with the DNA; shielding of the negative charge by defined interactions of the tails with DNA allows for tighter compaction. The exact sites of the interactions with the DNA of all of the tails are dependent on the precise chromatin structure, which has been found to vary with assay conditions.[20,27,28] The H4 tail is essential for chromatin condensation and can make direct contacts with the protein core of the adjacent nucleosome. The basic tail interacts with an acidic patch exposed on the surface of the H2A/B dimer. This interaction is strong enough to be resolved in crystallographic studies.[8] Residues 16-26 of the H4 *N*-terminal tail

make extensive interactions with the H2A/B acidic patch through various salt bridges and hydrogen bonds. Of these, Lys-20 has been shown to be essential for nucleosome-nucleosome interactions for crystallisation. Upon mutation of this residue to cysteine, crystallisation was not observed; this lack of crystal formation could be rescued by alkylation of the cysteine to yield a lysine mimic.[8] In further studies Lys-16 has also been shown to be important for the acid patch – H4 tail interaction.[29]

Another important role of histone tails is to recruit proteins to chromatin in a posttranslational modification (PTM) dependent manner. This function presents one of the cornerstones of epigenetic regulation.

1.3 Epigenetic mechanisms

Eukaryotic DNA is highly condensed and structured in the form of chromatin. This compaction presents a problem for all proteins requiring access to DNA, such as polymerases and TFs. When DNA dependent RNA or DNA polymerases are required to access this structure, it has to be disassembled 'in front' of the polymerase, to allow efficient passage, and reassembled in its wake to conserve the current chromatin state. TFs need accessible DNA to efficiently bind and activate transcription. By regulating accessibility of certain regions of DNA, chromatin modifying factors are able to regulate transcription. There are various mechanisms by which these factors can influence accessibility including modulation of nucleosome positioning; introduction of histone variants; modification of DNA; and addition of PTMs to proteins including histones. All of these mechanisms are proposed to contribute to the epigenetic regulation of transcription.

As per definition, epigenetics encompasses all auxiliary mechanisms that are involved in the regulation of gene expression that are not due to direct changes in the 'ATCG' of the DNA sequence. It is proposed that, for a mechanism to be classified as epigenetic it has to fulfil several criteria:[30] (i) it has to be self-perpetuating, i.e. the transcriptional state of a gene has to be stable even when the signal is removed; (ii) it has to be heritable, usually through cell division, but generational inheritance may occur; (iii) it has to be reversible, with an appropriate signal, the transcriptional state of a gene

may be reset.[30] Epigenetic regulation can happen at the DNA level, through covalent modification of DNA bases,[31,32] at the protein level, where PTM of histone proteins is the best studied mechanism to convey epigenetic information,[33] or by a combination of both.

1.3.1 Epigenetic regulation by DNA modification

Epigenetic regulation at the DNA level is conveyed by covalent C-5 cytosine modifications (**Figure 2**). Since its first description in the 70s, the C-5 methylation of cytosine to give 5-methylcytosine (5mC) has become one of the best studied mechanisms for epigenetic control.[34,35] Cytosine methylation mainly happens in a 3' cytosine – 5' guanine (CpG) context, although non-CpG methylation has been described in embryonic stem cells[36] and neuronal cells.[37] Spontaneous deamination of 5mC would yield a T:G miss match, which could be repaired to either give a T:A or a C:G basepair. This deamination results in CpGs being underrepresented in the genome.[38] Regulation of transcription by CpG methylation happens in CpG islands.[39-41] These regions are typical around 1kb in size and possess an elevated CpG content. CpG islands overlap with approximately 70% of gene promoters including almost all housekeeping genes but also various developmental genes.[40] CpG islands function as strong activators when they are unmethylated[42] and CpG methylation at CpG islands results in silencing of the promoter. This is accomplished through various mechanisms such as weakening the interaction of the DNA and transcriptional activators[43] and recruitment of 5mC binding proteins[31,44] which can promote heterochromatin formation. Cytosine methylation is performed by the DNA methyltransferases (DNMT1 and 3a/b). DNMT1 functions as the maintaining methyltransferase, copying the CpG methylation state to the newly synthesized DNA strand after replication.[45,46] DNMT3a/b are *de novo* DNA methyltransferases and are responsible for establishing DNA methylation patterns for gene silencing and imprinting. In 2009, the hydroxylation product of 5mC, 5-hydroxymethylcytosine (5hmC) was described as a novel DNA modification.[47] The modification is introduced by the ten-eleven translocation (TET) 1-3 enzymes.[48] The TET enzymes are 2-oxoglutarate (2-OG), Fe^{2+} and O_2 dependent oxygenases.[49] Products of further oxidation of 5mC by TET1-3, 5-formyl- and 5-carboxy-cytosine (5fC and 5caC) have also been described.[50-52]

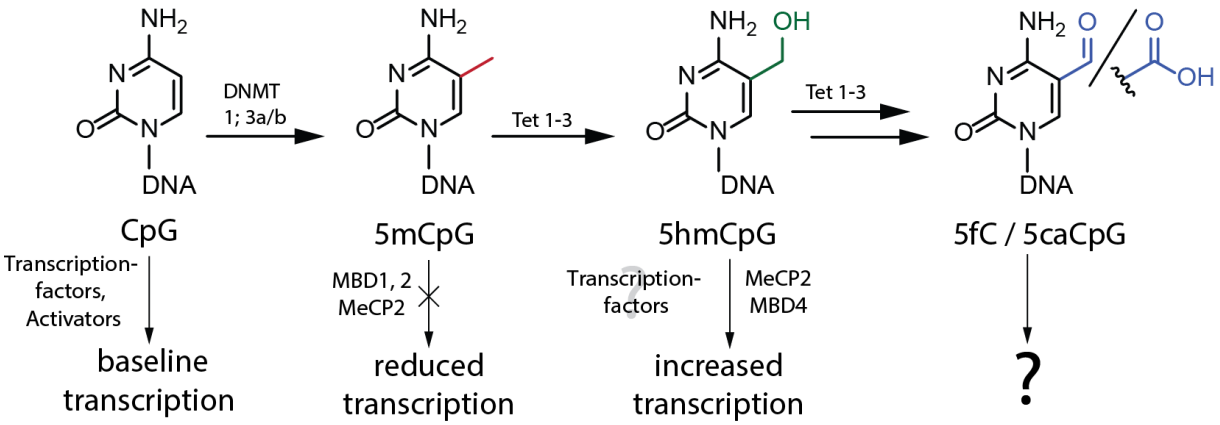


Figure 2. Regulation of transcription by cytosine C-5 modifications (adapted from Lercher *et al*).[53]

Much effort in the scientific community has recently been directed towards the elucidation of the role of these newly identified 'epigenetic' DNA modifications. A major advance towards this goal was the development of sequencing methods for 5hmC,[54-56] 5fC[57,58] and 5caC.[59] Together, results from these sequencing experiments suggest a reversal of the repressive character of 5mC by cytosine oxidation, with cytosine modifications being preferentially located at active regulatory regions. In addition to mapping, progress has been made in identifying effector proteins for these modifications.[60-62]

5hmC and further oxidation products have also been proposed as intermediates towards active and passive demethylation.[63] For active demethylation, oxidation to 5caC and decarboxylation[64] or removal by thymine DNA glycosylase (TDG)[51] have been proposed. The deamination product of 5hmC 5-hydroxymethyluracil is also a substrate for TDG and so deamination and removal by TDG is another plausible mechanism.[65] Passive demethylation would happened due to dilution of 5hmC during replication because DNMT1 does not accept hemi-5hmC containing DNA as a substrate.[66]

1.3.2 Epigenetic regulation by histone modification

The PTM of histone proteins is another cornerstone of epigenetic regulation in eukaryotic cells. Various PTMs have been reported on histone proteins including lysine acetylation, acylation with longer chain length acids,[67,68] formylation,[67,69] methylation[70,71] and ubiquitinylation[72]; arginine methylation and citrullination;[73] serine and threonine phosphorylation and GlcNAcylation;[74,75]

histidine phosphorylation;[76] glutamine methylation[77] and glutamate and arginine mono- and poly-ADP ribosylation. All of these modifications can function as an activating or repressing signal favouring eu- or heterochromatin formation respectively (**Figure 3a**). There are at least two mechanisms by which the PTMs on histone protein can affect the transcriptional state. Firstly, PTMs can directly affect the biophysical / biochemical properties of the nucleosome tails. An archetype for this mechanism is lysine acetylation, which decreases the overall charge of the tail. This can result in a decrease of the electrostatic interaction between the DNA and the tail, resulting in more open chromatin.[78,79] Phosphorylation and ADP ribosylation could have a similar effect, since negative charge is introduced, resulting in a lower overall charge of the tail. The second mechanism is the recruitment of effector proteins in a PTM dependent manner to the histone tail.[80] Many transcriptional activating or repressing complexes contain PTM 'reader' domains, which allows recruitment to sites containing activating or repressing PTMs.[81]

PTMs are transferred by 'writer' enzymes and can be removed by 'eraser' enzymes, for e.g. lysine methylation these would be lysine methyl transferases (KMT1-8) and lysine demethylases (KDM1-8). These modifying enzymes are often part of larger chromatin modifying complexes containing PTM 'readers' or contain a 'reader' domain themselves, thus interconnecting many histone marks.[82] For example PHD finger protein 8 (PHF8) contains a 'reader' plant homeodomain (PHD) finger domain which can recognize H3K4Me_3, an activating modification, and a 'eraser' jumonji domain that can demethylate H3K9Me_2, a repressing modification.[83]

While some PTMs are mainly activating, most can have different effects depending on the site of modification on the histone tails. Lysine methylation is present at multiple sites of which many are functionally well characterized. While Lys-4 methylation is found at active promoters, Lys-9 and Lys-27 are found at transcriptionally repressed promoters.[33] In addition to the site of modification, the exact methylation state of the lysine is also important. Representative mechanisms for euchromatin and heterchromatin formation as induced by histone modifications are illustrated by H3K4 and H3K9 tri-methylation.

H3K4 tri-methylation is an activating modification that is introduced by different members of the Su(var)3-9-'Enhancer of zeste'-Trithorax (SET) containing lysine methyl transferase KMT2 family (KMT2A-G). At active transcription start sites (TSS), H3K4Me$_3$ is found and can activate transcription by e.g. recruitment of transcription initiation factor TFIID subunit 3 (TAF3) which in turn can help RNA Pol II recruitment.[84] Erasure of this mark is performed by the KDM5 family.[85-90] KDM5A can also associate with the repressive polycomb group proteins (PcG) linking removal of an activating mark with transcriptional repression.[91]

H3K9 tri-methylation is a repressive modification, introduced by KTM1 family members (A, B, E, and F). This modification is important for heterochromatin organization through e.g. recruitment of the H3K9Me$_3$ reader heterochromatin protein 1 (HP1).[92-94] HP1 can bridge adjacent nucleosome containing H3K9Me$_3$[95] which results in condensation.[96] Through recruitment of KTM1 family members, HP1 also reinforces and maintains heterochromatin.

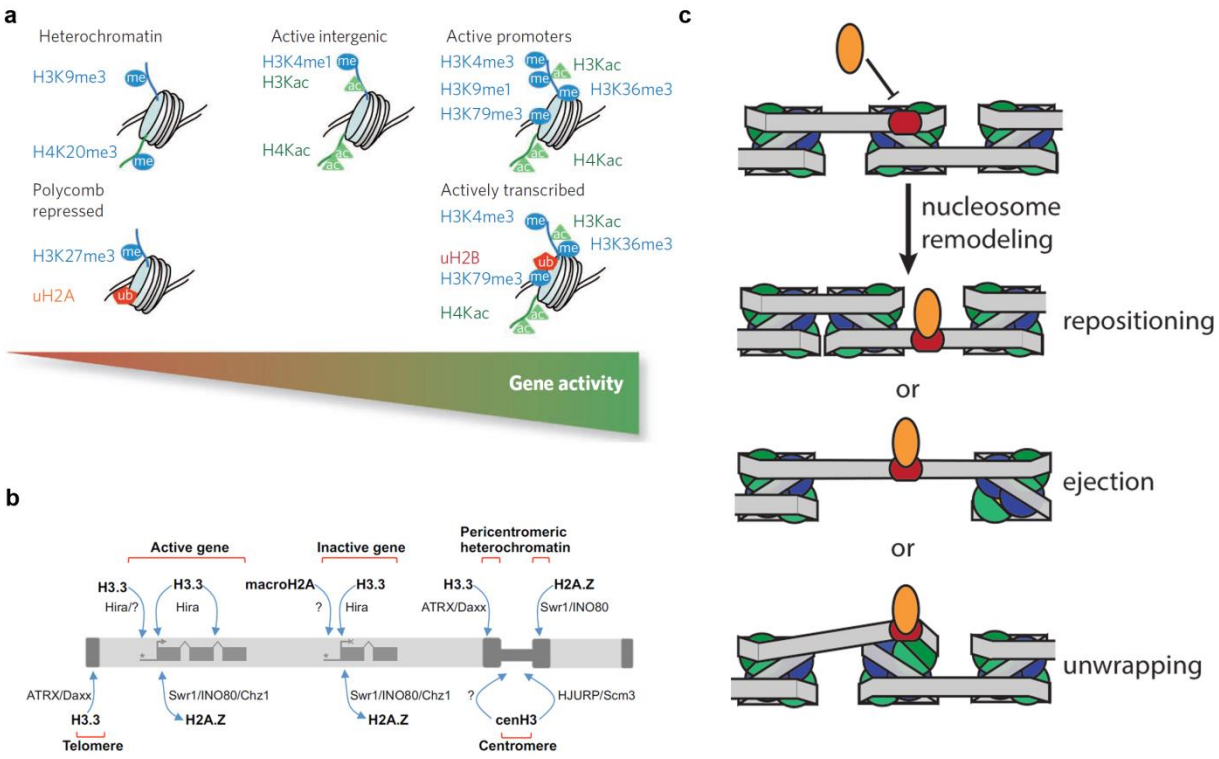


Figure 3. Epigenetic regulatory mechanisms on the nucleosome level. **a** Regulation of the transcriptional state is controlled by different histone modifications (adapted from Fierz *et al.*).[82] **b** Occupancy of different histone variants involved in

transcriptional regulation and chromatin organization (adapted from Skene *et al.*).[97] **c** Regulation of DNA accessibility by chromatin remodelers.

1.3.3 Additional mechanisms on the nucleosome level

In addition to histone PTMs, there are other mechanisms for transcriptional regulation which modulate the chromatin landscape. These include the replacement of the canonical histones with histone variants,[98-100] regulation of nucleosome positioning[101] and long non-coding RNA based mechanisms.[102-104]

1.3.3.1 Histone variants

After replication, canonical histones are rapidly deposited on both the old and newly synthetized DNA strand. In addition to these canonical histones, there are multiple variants[97] of histone H2A, H2B and H3 which are deposited in a replication independent manner and play an important role in various chromatin related processes (**Figure 3b**); e.g. transcriptional regulation[105] and DNA damage response.[106-108]

Histone H3 has two ubiquitous variants in humans, cenH3 and H3.3, and one testis specific variant, H3.4. The variant cenH3 (also called CENP-A) is found in the centromeres and is thought to be the main marker for centromeres in human cells,[109,110] since no clear DNA sequence dependency for all centromeres could be detected.[111] It is still being debated whether CENP-A containing nucleosomes contain two copies of CENP-A, identical to the normal nucleosomes, or only contain one copy.[112-114] A crystal structure of a CENP-A containing nucleosome revealed an octameric structure with two copies of CENP-A.[115] In comparison to the canonical nucleosome the CENP-A nucleosome binds a segment of DNA which is around 25 bp shorter and displays small rearrangements in the H3 fold. This is exploited by specific CENP-A binding proteins to localize to centromeres.[116]

Histone H3.3 only differs in 4-5 amino acids from the replication dependent deposited H3.1 and H3.2.[117] H3.3 is deposited in a replication independent fashion and deposition is dependent on the two chaperones HIRA[118,119] and DAXX.[120-123] The deposition of H3.3 is proposed to mainly result from

RNA polymerase action causing displacement of nucleosomes.[97] Thus, H3.3 is mainly found at sites of continued or abortive transcription such as active gene bodies and active and repressed promoters.[124] H3.3 is thought to decrease nucleosome stability resulting in an open chromatin state.[125]

Histone H2A has four main variants, H2A.Z, H2A.X, H2A.Bbd and macroH2A.[100] For H2A.Z more sub-variants have been reported.[126] H2A.Z is involved in a variety of processes such as transcriptional activation and repression, transcriptional elongation and chromosome segregation.[97] At TSS, H2A.Z is often found in the nucleosome flanking the nucleosome depleted region.[127-129] A crystal structure of an H2A.Z containing nucleosome revealed multiple divergent features, such as a different L1 loop and an extended acidic patch on the H2A.Z / H2B dimer surface.[130] The distinct acidic patch is an important recognition motif for targeting proteins to H2A.Z containing nucleosomes.[131,132] The H2A.X variant contains a *C*-terminal extension, which is important during the repair of double-strand breaks in DNA.[107,133] Upon detection of such DNA damage, a tyrosine in the *C*-terminus is phosphorylated which results in recruitment of DNA repair machinery.[106] macroH2A contains an additional *C*-terminal domain of ~25 kDa, twice the size of the H2A histone fold core. The x-ray structure of a nucleosome containing the macroH2A histone core revealed small differences in the L1 loop.[134] The macro domain is thought to function in protein recruitment.[134] MacroH2A is mainly found in heterochromatin and on the Xi inactivated X chromosome.[135] H2A.Bbd on the other hand is slightly smaller and has only 50% sequence identity with canonical H2A.[136] This variant destabilized nucleosomes[137] and is mainly found at active TSS.[138] In contrast to other variants, H2A.Bbd is only highly expressed in testis and to a lower extent in brain.[136]

1.3.3.2 Nucleosome positioning

The positioning of nucleosomes *in vivo* is regulated in part by DNA sequence preferences of nucleosomes[139] and in part by the active repositioning of nucleosome by chromatin remodelers.[140-142] Remodelers are large complexes that contain a central ATP dependent DNA helicase that is responsible for disrupting DNA-histone interactions by movement of the DNA.[140] There are four

distinct families of remodelers in eukaryotes that share conserved ATPase activity, but have very different functions.[143-146] Remodelers can govern the accessibility of DNA by repositioning or ejecting the nucleosome or by unwrapping the DNA from the nucleosome (**Figure 3c**). In a transcriptional context, exposure of DNA to a TF can result in transcriptional activation, while occupation of a TFs binding site by a nucleosome can hinder binding, resulting in transcriptional repression. In addition to this, repressing remodelers can recruit other repressing complexes such as HDACs and activating remodelers can recruit activating complexes. Also, remodelers can specifically eject canonical H2A/B dimers and replace these with H2A.Z/B containing dimers, thus depositing histone variants.[147] The inverse activity, replacement of H2AZ/B dimers with canonical dimers, has also been described.[148]

1.4 Goals of the work of this thesis

Epigenetic research has greatly benefitted from developed chemical biology methods.[82] These methods have helped to address research questions in epigenetics by providing tools to investigate protein function *in vivo*,[149] or to generation of PTM containing proteins for study *in vitro.*[82,150] The overall goal of my work was to develop and apply new chemical methods for the study of epigenetic DNA and protein modifications. Development of methodology that enables the introduction of homogeneous PTMs into proteins is important to further the role of chemical biology in this field (Chapter 4). With robust methods in hand, these were applied to learn about the effect of PTMs on larger chromatin structures such as nucleosomes (Chapter 5). In addition, recent advances in nuclear magnetic resonance (NMR) spectroscopy have made it possible to study large protein complexes up to the MDa scale.[151] These methods were applied to study the biophysical effect of histone PTMs on the structure and dynamics of the histone H3 tail in the context of a nucleosome (Chapter 6).

At the DNA level, the role of the newly discovered epigenetic DNA modification is still largely unknown. A new DNA modification methodology was developed that can be used to introduce photo-crosslinkers for the identification of DNA interacting proteins (Chapter 3). To further elucidate the role of 5hmC in transcriptional regulation, the effect of 5hmC on TF binding to DNA was investigated (Chapter 7).

1.5 Reference

(1) Consortium", T. E. P. *Nature* **2012**, *489*, 57.
(2) Ibarra-Laclette, E.; Lyons, E.; Hernandez-Guzman, G.; Perez-Torres, C. A.; Carretero-Paulet, L.; Chang, T.-H.; Lan, T.; Welch, A. J.; Juarez, M. J. A.; Simpson, J.; Fernandez-Cortes, A.; Arteaga-Vazquez, M.; Gongora-Castillo, E.; Acevedo-Hernandez, G.; Schuster, S. C.; Himmelbauer, H.; Minoche, A. E.; Xu, S.; Lynch, M.; Oropeza-Aburto, A.; Cervantes-Perez, S. A.; de Jesus Ortega-Estrada, M.; Cervantes-Luevano, J. I.; Michael, T. P.; Mockler, T.; Bryant, D.; Herrera-Estrella, A.; Albert, V. A.; Herrera-Estrella, L. *Nature* **2013**, *498*, 94.
(3) Oehler, S.; Eismann, E. R.; Kramer, H.; Muller-Hill, B. *EMBO J.* **1990**, *9*, 973.
(4) Latchman, D. S. *Int. J. Bio. Cell Bio.* **1997**, *29*, 1305.
(5) Seshasayee, A.; Sivaraman, K.; Luscombe, N. In *A Handbook of Transcription Factors*; Hughes, T. R., Ed.; Springer Netherlands: 2011; Vol. 52, p 7.
(6) Neph, S.; Vierstra, J.; Stergachis, A. B.; Reynolds, A. P.; Haugen, E.; Vernot, B.; Thurman, R. E.; John, S.; Sandstrom, R.; Johnson, A. K.; Maurano, M. T.; Humbert, R.; Rynes, E.; Wang, H.; Vong, S.; Lee, K.; Bates, D.; Diegel, M.; Roach, V.; Dunn, D.; Neri, J.; Schafer, A.; Hansen, R. S.; Kutyavin, T.; Giste, E.; Weaver, M.; Canfield, T.; Sabo, P.; Zhang, M.; Balasundaram, G.; Byron, R.; MacCoss, M. J.; Akey, J. M.; Bender, M. A.; Groudine, M.; Kaul, R.; Stamatoyannopoulos, J. A. *Nature* **2012**, *489*, 83.
(7) http://www.ensembl.org/info/about/index.html.
(8) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251.
(9) Ramakrishnan, V.; Finch, J. T.; Graziano, V.; Lee, P. L.; Sweet, R. M. *Nature* **1993**, *362*, 219.
(10) Zhou, B.-R.; Feng, H.; Kato, H.; Dai, L.; Yang, Y.; Zhou, Y.; Bai, Y. *Proc. Natl. Acad. Sci.* **2013**, *110*, 19390.
(11) Happel, N.; Doenecke, D. *Gene* **2009**, *431*, 1.
(12) Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. *J. Mol. Biol.* **2002**, *319*, 1097.
(13) Olins, D. E.; Olins, A. L. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 809.
(14) Luger, K.; Dechassa, M. L.; Tremethick, D. J. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 436.
(15) Olins, A. L.; Olins, D. E. *Science* **1974**, *183*, 330.
(16) Robinson, P. J. J.; Rhodes, D. *Curr. Opin. Struct. Biol.* **2006**, *16*, 336.
(17) Schalch, T.; Duda, S.; Sargent, D. F.; Richmond, T. J. *Nature* **2005**, *436*, 138.
(18) Routh, A.; Sandin, S.; Rhodes, D. *Proc. Natl. Acad. Sci.* **2008**, *105*, 8872.
(19) Grigoryev, S. A.; Arya, G.; Correll, S.; Woodcock, C. L.; Schlick, T. *Proc. Natl. Acad. Sci.* **2009**, *106*, 13317.
(20) Zheng, C.; Hayes, J. J. *Biopolymers* **2003**, *68*, 539.
(21) Tan, S.; Davey, C. A. *Curr. Opin. Struct. Biol.* **2011**, *21*, 128.
(22) Tse, C.; Hansen, J. C. *Biochemistry* **1997**, *36*, 11381.
(23) Hansen, J. C.; Tse, C.; Wolffe, A. P. *Biochemistry* **1998**, *37*, 17637.
(24) Fletcher, T. M.; Hansen, J. C. *J. Biol. Chem.* **1995**, *270*, 25359.
(25) Zheng, C.; Hayes, J. J. *J. Biol. Chem.* **2003**, *278*, 24217.
(26) Kan, P.-Y.; Lu, X.; Hansen, J. C.; Hayes, J. J. *Mol. Cell. Biol.* **2007**, *27*, 2084.
(27) Lee, K.-M.; Hayes, J. J. *Proc. Natl. Acad. Sci.* **1997**, *94*, 8959.
(28) Lee, K.-M.; Hayes, J. J. *Biochemistry* **1998**, *37*, 8622.
(29) Zhou, B.-R.; Feng, H.; Ghirlando, R.; Kato, H.; Gruschus, J.; Bai, Y. *J. Mol. Biol.* **2012**, *421*, 30.
(30) Bonasio, R.; Tu, S.; Reinberg, D. *Science* **2010**, *330*, 612.
(31) Klose, R. J.; Bird, A. P. *Trends Biochem. Sci.* **2006**, *31*, 89.
(32) Song, C.-X.; Yi, C.; He, C. *Nat. Biotech.* **2012**, *30*, 1107.
(33) Bannister, A. J.; Kouzarides, T. *Cell Res.* **2011**, *21*, 381.
(34) Riggs, A. D. *Cytogenet. Cell Genet.* **1975**, *14*, 9.
(35) Holliday, R.; Pugh, J. *Science* **1975**, *187*, 226.

(36) Lister, R.; Pelizzola, M.; Dowen, R. H.; Hawkins, R. D.; Hon, G.; Tonti-Filippini, J.; Nery, J. R.; Lee, L.; Ye, Z.; Ngo, Q.-M.; Edsall, L.; Antosiewicz-Bourget, J.; Stewart, R.; Ruotti, V.; Millar, A. H.; Thomson, J. A.; Ren, B.; Ecker, J. R. *Nature* **2009**, *462*, 315.
(37) Lister, R.; Mukamel, E. A.; Nery, J. R.; Urich, M.; Puddifoot, C. A.; Johnson, N. D.; Lucero, J.; Huang, Y.; Dwork, A. J.; Schultz, M. D.; Yu, M.; Tonti-Filippini, J.; Heyn, H.; Hu, S.; Wu, J. C.; Rao, A.; Esteller, M.; He, C.; Haghighi, F. G.; Sejnowski, T. J.; Behrens, M. M.; Ecker, J. R. *Science* **2013**, *341*.
(38) Consortium, I. H. G. S. *Nature* **2001**, *409*, 860.
(39) Illingworth, R. S.; Bird, A. P. *FEBS Lett.* **2009**, *583*, 1713.
(40) Deaton, A. M.; Bird, A. *Genes Dev.* **2011**, *25*, 1010.
(41) Bird, A. P. *Nature* **1986**, *321*, 209.
(42) Ramirez-Carrozzi, V. R.; Braas, D.; Bhatt, D. M.; Cheng, C. S.; Hong, C.; Doty, K. R.; Black, J. C.; Hoffmann, A.; Carey, M.; Smale, S. T. *Cell* **2009**, *138*, 114.
(43) Prendergast, G.; Ziff, E. *Science* **1991**, *251*, 186.
(44) Bogdanović, O.; Veenstra, G. *Chromosoma* **2009**, *118*, 549.
(45) Song, J.; Rechkoblit, O.; Bestor, T. H.; Patel, D. J. *Science* **2011**, *331*, 1036.
(46) Song, J.; Teplova, M.; Ishibe-Murakami, S.; Patel, D. J. *Science* **2012**, *335*, 709.
(47) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929.
(48) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
(49) Aik, W.; McDonough, M. A.; Thalhammer, A.; Chowdhury, R.; Schofield, C. J. *Curr. Opin. Struct. Biol.* **2012**, *22*, 691.
(50) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300.
(51) He, Y. F.; Li, B. Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C. X.; Zhang, K.; He, C.; Xu, G. L. *Science* **2011**, *333*, 1303.
(52) Pfaffeneder, T.; Hackner, B.; Truss, M.; Munzel, M.; Muller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. *Angew. Chem. Int. Ed.* **2011**, *50*, 7008.
(53) Lercher, L.; McDonough, M. A.; El-Sagheer, A. H.; Thalhammer, A.; Kriaucionis, S.; Brown, T.; Schofield, C. J. *Chem. Commun.* **2014**, *50*, 1794.
(54) Song, C.-X.; Szulwach, K. E.; Fu, Y.; Dai, Q.; Yi, C.; Li, X.; Li, Y.; Chen, C.-H.; Zhang, W.; Jian, X.; Wang, J.; Zhang, L.; Looney, T. J.; Zhang, B.; Godley, L. A.; Hicks, L. M.; Lahn, B. T.; Jin, P.; He, C. *Nat. Biotech.* **2011**, *29*, 68.
(55) Song, C.-X.; Clark, T. A.; Lu, X.-Y.; Kislyuk, A.; Dai, Q.; Turner, S. W.; He, C.; Korlach, J. *Nat. Meth.* **2012**, *9*, 75.
(56) Booth, M. J.; Branco, M. R.; Ficz, G.; Oxley, D.; Krueger, F.; Reik, W.; Balasubramanian, S. *Science* **2012**, *336*, 934.
(57) Song, C.-X.; Szulwach, Keith E.; Dai, Q.; Fu, Y.; Mao, S.-Q.; Lin, L.; Street, C.; Li, Y.; Poidevin, M.; Wu, H.; Gao, J.; Liu, P.; Li, L.; Xu, G.-L.; Jin, P.; He, C. *Cell* **2013**, *153*, 678.
(58) Booth, M. J.; Marsico, G.; Bachman, M.; Beraldi, D.; Balasubramanian, S. *Nat. Chem.* **2014**, *6*, 435.
(59) Lu, X.; Song, C.-X.; Szulwach, K.; Wang, Z.; Weidenbacher, P.; Jin, P.; He, C. *J. Am. Chem. Soc.* **2013**, *135*, 9315.
(60) Spruijt, Cornelia G.; Gnerlich, F.; Smits, Arne H.; Pfaffeneder, T.; Jansen, Pascal W. T. C.; Bauer, C.; Münzel, M.; Wagner, M.; Müller, M.; Khan, F.; Eberl, H. C.; Mensinga, A.; Brinkman, Arie B.; Lephikov, K.; Müller, U.; Walter, J.; Boelens, R.; van Ingen, H.; Leonhardt, H.; Carell, T.; Vermeulen, M. *Cell* **2013**, *152*, 1146.
(61) Lercher, L.; McGouran, J. F.; Kessler, B. M.; Schofield, C. J.; Davis, B. G. *Angew. Chem. Int. Ed.* **2013**, *52*, 10553.
(62) Frauer, C.; Hoffmann, T.; Bultmann, S.; Casa, V.; Cardoso, M. C.; Antes, I.; Leonhardt, H. *PLoS ONE* **2011**, *6*, e21306.
(63) Kohli, R. M.; Zhang, Y. *Nature* **2013**, *502*, 472.

(64) Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeier, C.; Truss, M.; Carell, T. *Angew. Chem. Int. Ed.* **2012**, *51*, 6516.
(65) Cortellino, S.; Xu, J.; Sannai, M.; Moore, R.; Caretti, E.; Cigliano, A.; Le Coz, M.; Devarajan, K.; Wessels, A.; Soprano, D.; Abramowitz, Lara K.; Bartolomei, Marisa S.; Rambow, F.; Bassi, Maria R.; Bruno, T.; Fanciulli, M.; Renner, C.; Klein-Szanto, Andres J.; Matsumoto, Y.; Kobi, D.; Davidson, I.; Alberti, C.; Larue, L.; Bellacosa, A. *Cell* **2011**, *146*, 67.
(66) Valinluck, V.; Sowers, L. C. *Cancer Res.* **2007**, *67*, 946.
(67) Tan, M.; Luo, H.; Lee, S.; Jin, F.; Yang, Jeong S.; Montellier, E.; Buchou, T.; Cheng, Z.; Rousseaux, S.; Rajagopal, N.; Lu, Z.; Ye, Z.; Zhu, Q.; Wysocka, J.; Ye, Y.; Khochbin, S.; Ren, B.; Zhao, Y. *Cell* **2011**, *146*, 1016.
(68) Dai, L.; Peng, C.; Montellier, E.; Lu, Z.; Chen, Y.; Ishii, H.; Debernardi, A.; Buchou, T.; Rousseaux, S.; Jin, F.; Sabari, B. R.; Deng, Z.; Allis, C. D.; Ren, B.; Khochbin, S.; Zhao, Y. *Nat. Chem. Biol.* **2014**, *10*, 365.
(69) Jiang, T.; Zhou, X.; Taghizadeh, K.; Dong, M.; Dedon, P. C. *Proc. Natl. Acad. Sci.* **2007**, *104*, 60.
(70) Martin, C.; Zhang, Y. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 838.
(71) Black, Joshua C.; Van Rechem, C.; Whetstine, Johnathan R. *Mol. Cell* **2012**, *48*, 491.
(72) Weake, V. M.; Workman, J. L. *Mol. Cell* **2008**, *29*, 653.
(73) Christophorou, M. A.; Castelo-Branco, G.; Halley-Stott, R. P.; Oliveira, C. S.; Loos, R.; Radzisheuskaya, A.; Mowen, K. A.; Bertone, P.; Silva, J. C.; Zernicka-Goetz, M.; Nielsen, M. L.; Gurdon, J. B.; Kouzarides, T. *Nature* **2014**, *507*, 104.
(74) Sakabe, K.; Wang, Z.; Hart, G. W. *Proc. Natl. Acad. Sci.* **2010**, *107*, 19915.
(75) Fong, J. J.; Nguyen, B. L.; Bridger, R.; Medrano, E. E.; Wells, L.; Pan, S.; Sifers, R. N. *J. Biol. Chem.* **2012**, *287*, 12195.
(76) Kee, J.-M.; Villani, B.; Carpenter, L. R.; Muir, T. W. *J. Am. Chem. Soc.* **2010**, *132*, 14327.
(77) Tessarz, P.; Santos-Rosa, H.; Robson, S. C.; Sylvestersen, K. B.; Nelson, C. J.; Nielsen, M. L.; Kouzarides, T. *Nature* **2014**, *505*, 564.
(78) Ausio, J.; Van Holde, K. E. *Biochemistry* **1986**, *25*, 1421.
(79) Libertini, L. J.; Ausió, J.; van Holde, K. E.; Small, E. W. *Biophys. J.* **1988**, *53*, 477.
(80) Musselman, C. A.; Lalonde, M.-E.; Cote, J.; Kutateladze, T. G. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1218.
(81) Wu, J. I.; Lessard, J.; Crabtree, G. R. *Cell* **2009**, *136*, 200.
(82) Fierz, B.; Muir, T. W. *Nat. Chem. Biol.* **2012**, 417.
(83) Horton, J. R.; Upadhyay, A. K.; Qi, H. H.; Zhang, X.; Shi, Y.; Cheng, X. *Nat. Struct. Mol. Biol.* **2010**, *17*, 38.
(84) Vermeulen, M.; Mulder, K. W.; Denissov, S.; Pijnappel, W. W. M. P.; van Schaik, F. M. A.; Varier, R. A.; Baltissen, M. P. A.; Stunnenberg, H. G.; Mann, M.; Timmers, H. T. M. *Cell* **2007**, *131*, 58.
(85) Christensen, J.; Agger, K.; Cloos, P. A. C.; Pasini, D.; Rose, S.; Sennels, L.; Rappsilber, J.; Hansen, K. H.; Salcini, A. E.; Helin, K. *Cell* **2007**, *128*, 1063.
(86) Eissenberg, J. C.; Lee, M. G.; Schneider, J.; Ilvarsonn, A.; Shiekhattar, R.; Shilatifard, A. *Nat. Struct. Mol. Biol.* **2007**, *14*, 344.
(87) Iwase, S.; Lan, F.; Bayliss, P.; de la Torre-Ubieta, L.; Huarte, M.; Qi, H. H.; Whetstine, Johnathan R.; Bonni, A.; Roberts, T. M.; Shi, Y. *Cell* **2007**, *128*, 1077.
(88) Klose, R. J.; Yan, Q.; Tothova, Z.; Yamane, K.; Erdjument-Bromage, H.; Tempst, P.; Gilliland, D. G.; Zhang, Y.; Kaelin Jr, W. G. *Cell* **2007**, *128*, 889.
(89) Lee, N.; Zhang, J.; Klose, R. J.; Erdjument-Bromage, H.; Tempst, P.; Jones, R. S.; Zhang, Y. *Nat. Struct. Mol. Biol.* **2007**, *14*, 341.
(90) Secombe, J.; Li, L.; Carlos, L.; Eisenman, R. N. *Genes Dev.* **2007**, *21*, 537.
(91) Pasini, D.; Hansen, K. H.; Christensen, J.; Agger, K.; Cloos, P. A. C.; Helin, K. *Genes Dev.* **2008**, *22*, 1345.
(92) Canzio, D.; Larson, A.; Narlikar, G. J. *Trends Cell Biol.* **2014**, *24*, 377.
(93) Jacobs, S. A.; Khorasanizadeh, S. *Science* **2002**, *295*, 2080.
(94) Lachner, M.; O'Carroll, D.; Rea, S.; Mechtler, K.; Jenuwein, T. *Nature* **2001**, *410*, 116.

(95) Canzio, D.; Chang, E. Y.; Shankar, S.; Kuchenbecker, K. M.; Simon, M. D.; Madhani, H. D.; Narlikar, G. J.; Al-Sady, B. *Mol. Cell* **2011**, *41*, 67.
(96) Canzio, D.; Liao, M.; Naber, N.; Pate, E.; Larson, A.; Wu, S.; Marina, D. B.; Garcia, J. F.; Madhani, H. D.; Cooke, R.; Schuck, P.; Cheng, Y.; Narlikar, G. J. *Nature* **2013**, *496*, 377.
(97) Skene, P. J.; Henikoff, S. *Development* **2013**, *140*, 2513.
(98) Kurumizaka, H.; Horikoshi, N.; Tachiwana, H.; Kagawa, W. *Curr. Opin. Struct. Biol.* **2013**, *23*, 109.
(99) Talbert, P. B.; Henikoff, S. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 264.
(100) Bönisch, C.; Hake, S. B. *Nucleic Acids Res.* **2012**.
(101) Dutta, A.; Workman, Jerry L. *Mol. Cell* **2012**, *48*, 1.
(102) Lee, J. T. *Science* **2012**, *338*, 1435.
(103) Bhan, A.; Mandal, S. S. *ChemMedChem* **2014**, *9*, 1932.
(104) Mercer, T. R.; Mattick, J. S. *Nat. Struct. Mol. Biol.* **2013**, *20*, 300.
(105) Bargaje, R.; Alam, M. P.; Patowary, A.; Sarkar, M.; Ali, T.; Gupta, S.; Garg, M.; Singh, M.; Purkanti, R.; Scaria, V.; Sivasubbu, S.; Brahmachari, V.; Pillai, B. *Nucleic Acids Res.* **2012**, *40*, 8965.
(106) Revet, I.; Feeney, L.; Bruguera, S.; Wilson, W.; Dong, T. K.; Oh, D. H.; Dankort, D.; Cleaver, J. E. *Proc. Natl. Acad. Sci.* **2011**, *108*, 8663.
(107) Kinner, A.; Wu, W.; Staudt, C.; Iliakis, G. *Nucleic Acids Res.* **2008**, *36*, 5678.
(108) Seo, J.; Kim, S. C.; Lee, H.-S.; Kim, J. K.; Shon, H. J.; Salleh, N. L. M.; Desai, K. V.; Lee, J. H.; Kang, E.-S.; Kim, J. S.; Choi, J. K. *Nucleic Acids Res.* **2012**, *40*, 5965.
(109) Black, B. E.; Brock, M. A.; Bédard, S.; Woods, V. L.; Cleveland, D. W. *Proc. Natl. Acad. Sci.* **2007**, *104*, 5008.
(110) Black, B. E.; Cleveland, D. W. *Cell* **2011**, *144*, 471.
(111) Warburton, P. *Chromosome Res.* **2004**, *12*, 617.
(112) Walkiewicz, M. P.; Dimitriadis, E. K.; Dalal, Y. *Nat. Struct. Mol. Biol.* **2014**, *21*, 2.
(113) Codomo, C. A.; Furuyama, T.; Henikoff, S. *Nat. Struct. Mol. Biol.* **2014**, *21*, 4.
(114) Miell, M. D. D.; Straight, A. F.; Allshire, R. C. *Nat. Struct. Mol. Biol.* **2014**, *21*, 5.
(115) Tachiwana, H.; Kagawa, W.; Shiga, T.; Osakabe, A.; Miya, Y.; Saito, K.; Hayashi-Takanaka, Y.; Oda, T.; Sato, M.; Park, S.-Y.; Kimura, H.; Kurumizaka, H. *Nature* **2011**, *476*, 232.
(116) Kato, H.; Jiang, J.; Zhou, B.-R.; Rozendaal, M.; Feng, H.; Ghirlando, R.; Xiao, T. S.; Straight, A. F.; Bai, Y. *Science* **2013**, *340*, 1110.
(117) Hamiche, A.; Shuaib, M. *BBA - Gene. Regul. Mech.* **2012**, *1819*, 230.
(118) Ray-Gallet, D.; Quivy, J.-P.; Scamps, C.; Martini, E. M. D.; Lipinski, M.; Almouzni, G. *Mol. Cell* **2002**, *9*, 1091.
(119) Tagami, H.; Ray-Gallet, D.; Almouzni, G.; Nakatani, Y. *Cell* **2004**, *116*, 51.
(120) Drané, P.; Ouararhni, K.; Depaux, A.; Shuaib, M.; Hamiche, A. *Genes Dev.* **2010**, *24*, 1253.
(121) Lewis, P. W.; Elsaesser, S. J.; Noh, K.-M.; Stadler, S. C.; Allis, C. D. *Proc. Natl. Acad. Sci.* **2010**, *107*, 14075.
(122) Liu, C.-P.; Xiong, C.; Wang, M.; Yu, Z.; Yang, N.; Chen, P.; Zhang, Z.; Li, G.; Xu, R.-M. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1287.
(123) Elsasser, S. J.; Huang, H.; Lewis, P. W.; Chin, J. W.; Allis, C. D.; Patel, D. J. *Nature* **2012**, *491*, 560.
(124) Goldberg, A. D.; Banaszynski, L. A.; Noh, K.-M.; Lewis, P. W.; Elsaesser, S. J.; Stadler, S.; Dewell, S.; Law, M.; Guo, X.; Li, X.; Wen, D.; Chapgier, A.; DeKelver, R. C.; Miller, J. C.; Lee, Y.-L.; Boydston, E. A.; Holmes, M. C.; Gregory, P. D.; Greally, J. M.; Rafii, S.; Yang, C.; Scambler, P. J.; Garrick, D.; Gibbons, R. J.; Higgs, D. R.; Cristea, I. M.; Urnov, F. D.; Zheng, D.; Allis, C. D. *Cell* **2010**, *140*, 678.
(125) Jin, C.; Felsenfeld, G. *Genes Dev.* **2007**, *21*, 1519.
(126) Bönisch, C.; Schneider, K.; Pünzeler, S.; Wiedemann, S. M.; Bielmeier, C.; Bocola, M.; Eberl, H. C.; Kuegel, W.; Neumann, J.; Kremmer, E.; Leonhardt, H.; Mann, M.; Michaelis, J.; Schermelleh, L.; Hake, S. B. *Nucleic Acids Res.* **2012**, *40*, 5951.

(127) Barski, A.; Cuddapah, S.; Cui, K.; Roh, T.-Y.; Schones, D. E.; Wang, Z.; Wei, G.; Chepelev, I.; Zhao, K. *Cell* **2007**, *129*, 823.
(128) Luk, E.; Ranjan, A.; FitzGerald, P. C.; Mizuguchi, G.; Huang, Y.; Wei, D.; Wu, C. *Cell* **2010**, *143*, 725.
(129) Ranjan, A.; Mizuguchi, G.; FitzGerald, Peter C.; Wei, D.; Wang, F.; Huang, Y.; Luk, E.; Woodcock, Christopher L.; Wu, C. *Cell* **2013**, *154*, 1232.
(130) Suto, R. K.; Clarkson, M. J.; Tremethick, D. J.; Luger, K. *Nat. Struct. Mol. Biol.* **2000**, *7*, 1121.
(131) Zlatanova, J.; Thakar, A. *Structure* **2008**, *16*, 166.
(132) Subramanian, V.; Mazumder, A.; Surface, L. E.; Butty, V. L.; Fields, P. A.; Alwan, A.; Torrey, L.; Thai, K. K.; Levine, S. S.; Bathe, M.; Boyer, L. A. *PLoS Genet* **2013**, *9*, e1003725.
(133) Fernandez-Capetillo, O.; Lee, A.; Nussenzweig, M.; Nussenzweig, A. *DNA Repair* **2004**, *3*, 959.
(134) Chakravarthy, S.; Gundimella, S. K. Y.; Caron, C.; Perche, P.-Y.; Pehrson, J. R.; Khochbin, S.; Luger, K. *Mol. Cell. Biol.* **2005**, *25*, 7616.
(135) Costanzi, C.; Pehrson, J. R. *Nature* **1998**, *393*, 599.
(136) Chadwick, B. P.; Willard, H. F. *J. Cell Biol.* **2001**, *152*, 375.
(137) Eirín-López, J. M.; Ishibashi, T.; Ausió, J. *FASEB J.* **2008**, *22*, 316.
(138) Soboleva, T. A.; Nekrasov, M.; Pahwa, A.; Williams, R.; Huttley, G. A.; Tremethick, D. J. *Nat. Struct. Mol. Biol.* **2012**, *19*, 25.
(139) Jiang, C.; Pugh, B. *Nat. Rev. Genet.* **2009**, *10*, 161
(140) Clapier, C. R.; Cairns, B. R. *Annu. Rev. Biochem.* **2009**, *78*, 273.
(141) Ho, L.; Crabtree, G. R. *Nature* **2010**, *463*, 474.
(142) Narlikar, Geeta J.; Sundaramoorthy, R.; Owen-Hughes, T. *Cell* **2013**, *154*, 490.
(143) Marfella, C. G. A.; Imbalzano, A. N. *Mutat. Res. - Fund. Mol. M.* **2007**, *618*, 30.
(144) Corona, D. F. V.; Tamkun, J. W. *BBA - Gene. Struct. Expr.* **2004**, *1677*, 113.
(145) Mohrmann, L.; Verrijzer, C. P. *BBA - Gene. Struct. Expr.* **2005**, *1681*, 59.
(146) Bao, Y.; Shen, X. *Mutat. Res. - Fund. Mol. M.* **2007**, *618*, 18.
(147) Mizuguchi, G.; Shen, X.; Landry, J.; Wu, W.-H.; Sen, S.; Wu, C. *Science* **2004**, *303*, 343.
(148) Tosi, A.; Haas, C.; Herzog, F.; Gilmozzi, A.; Berninghausen, O.; Ungewickell, C.; Gerhold, Christian B.; Lakomek, K.; Aebersold, R.; Beckmann, R.; Hopfner, K.-P. *Cell* **2013**, *154*, 1207.
(149) Frye, S. V. *Nat. Chem. Biol.* **2010**, *6*, 159.
(150) Allis, C. D.; Muir, T. W. *ChemBioChem* **2011**, *12*, 264.
(151) Mainz, A.; Religa, T. L.; Sprangers, R.; Linser , R.; Kay, L. E.; Reif, B. *Angew. Chem. Int. Ed.* **2013**, *52*, 8746.

2 Materials and methods

2.1 General considerations

2.1.1 Organic synthesis

Proton nuclear magnetic resonance (^{1}H NMR) spectra were recorded using a Bruker DPX200 (200 MHz), a Bruker AV400 (400 MHz), or a Bruker AVII500 (500 MHz) spectrometer, as indicated. Carbon nuclear magnetic resonance (^{13}C NMR) spectra were recorded using a Bruker AV400 (100 MHz) or using a Bruker AVII500 (125 MHz) spectrometer, as indicated. NMR Spectra were assigned using COSY, HMQC, and DEPT 135. All chemical shifts are quoted on the δ scale in ppm using residual solvent as the internal standard (^{1}H NMR: $CDCl_3$ = 7.26, CD_3OD = 3.31; DMSO-d6 = 2.50; CD_3CN = 1.94 and ^{13}C NMR: $CDCl_3$ = 77.0; CD_3OD = 49.0; DMSO-d6 = 39.5; CD_3CN = 1.94). Coupling constants (*J*) are reported in Hz (to the nearest 0.5 Hz) with the following splitting abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, quin = quintet, and a = apparent.

Infrared (IR) spectra were recorded using a Bruker Tensor 27 Fourier Transform spectrophotometer neat for both oils and solids. Absorption maxima (υ_{max}) are reported in wavenumbers (cm^{-1}). Low resolution mass spectra (LRMS) were recorded using a Waters Micromass LCT Premier TOF spectrometer Waters using electrospray ionization (ESI) and high resolution mass spectra (HRMS) were recorded using a Bruker MicroTOF ESI mass spectrometer. Nominal and exact m/z values are reported in Daltons.

Thin layer chromatography (TLC) was carried out using Merck aluminium backed sheets coated with 60F254 silica gel. Visualization of the silica plates was achieved using a UV lamp (λ_{max} = 254 nm), and/or ammonium molybdate (5% in 2 M H_2SO_4), and/or potassium permanganate (5% $KMnO_4$ in 1 M NaOH with 5% potassium carbonate).

Column chromatography was performed using a Biotage SP4 flash purification system using the indicated cartridges and solvent systems.

Anhydrous solvents were purchased from Fluka or Acros with the exception of dichloromethane and THF, which were dried in an alumina column under nitrogen. All other solvents were used as supplied (Analytical or HPLC grade), without prior purification. Deionized water was used for chemical reactions. Protein concentrations were determined by BCA (bicinchoninic acid) assay (Pierce) and/or OD280. Reagents were purchased from Sigma-Aldrich and used as supplied, unless otherwise indicated. All reactions using anhydrous conditions were performed using flame-dried apparatus under an atmosphere of argon or nitrogen. Brine refers to a saturated solution of sodium chloride. Anhydrous magnesium sulfate ($MgSO_4$) or sodium sulfate (Na_2SO_4) was used as drying agents after reaction workup, as indicated. Solvent mixtures are given as volume / volume (v/v).

2.1.2 Used buffers and media

Common stock solution	g reagent per L
0.5 M EDTA pH 8.0	146 g; adjust pH with HCl
1 M TRIS pH 8.0	121.1 g; adjust pH with HCl
1 M HEPES pH 7.5	238.3 g; adjust pH with HCl
5 M NaOAc pH 5.2	410 g; adjust pH with AcOH
4 M NaCl	233.8 g NaCl
2 M KCl	149.1 g KCl
5 X TBE	54 g TRIS base, 27.5 g boric acid, 20 mL of 0.5 M EDTA
10 x TAE	obtained from life technologies (15558)

Table 1. Composition of commonly used stock solutions

Commonly used media	reagent per L	
LB	10 g Tryptone, 5 g Yeast Extract, 5 g NaCl	Sigma-Aldrich (L3522)
2 x TY	16 g Tryptone, 10 g Yeast Extract, 5 g NaCl	Sigma-Aldrich (Y2627)
M9[1]	6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 0.5 g NH4Cl; adjust pH to 7.4 then autoclave; after autoclaving add 14 mg $CaCl_2{\cdot}4\ H_2O$, 120 mg $MgSO_4$, 10 mg Biotin, 10 mg Thiamine, 3 g Glucose	
TB	12 g Trypton, 24 g Yeast Extract, 9.4 g K_2HPO_4, 2.2 KH_2PO_4, 8 mL glycerol	Sigma-Aldrich (T0918)
SOC	2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose	Inivtrogen (15544) or Agilent (included with competent cells)

Table 2. Commonly used media

Common buffers

TAE	40mM Tris, 20mM AcOH, 1mM EDTA
0.5x TBE	22.5 mM Tris-borate/0.5 mM EDTA
TE 10/50	10 mM TRIS pH 7.5, 50mM EDTA
TE 10/0.1	10 mM TRIS pH 7.5, 0.1mM EDTA
PBS	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4 pH 7.4
Histone lysis and wash buffer	50 mM TRIS pH 7.5, 100mM NaCl, 1 mM EDTA, 5 mM BME
TW buffer	As wash buffer plus 1% Triton X-100
Unfolding buffer	7 M Guanidinum hydrochloride, 10 mM TRIS pH 7.5, 1 mM EDTA, 10 mM DTT, 1 mM Benzamidine
SAU-1000	7 M Urea, 20 mM NaOAc pH 5.2, 1 M NaCl, 1 mM EDTA, 5 mM BME
SAU-100	7 M Urea, 20 mM NaOAc pH 5.2, 100 mM NaCl, 1 mM EDTA, 5 mM BME
Refolding buffer	2 M NaCl, 10 mM TRIS pH 7.5, 1 mM EDTA, 5 mM BME
Reconstitution-2000	2 M KCl, 10 mM TRIS pH 7.5, 1 mM EDTA, 5 mM BME
Reconstitution-250	250 mM KCl, 10 mM TRIS pH 7.5, 1 mM EDTA, 5 mM BME
Low salt buffer	10 mM TRIS pH 7.5, 1 mM EDTA, 5 mM BME
Reaction buffer (RB)	5 M Guanidinium chloride, 50 mM TRIS pH 8.0
Cross coupling buffer	100 mM TRIS pH 8.0
TBS-T	10 mM TRIS pH 7.8, 200 mM NaCl, 0.02% (w/w) Tween 20
TBS-BT	10 mM TRIS pH 7.8, 200 mM NaCl, 0.02% (w/w) Tween 20, 3% BSA
Alkaline lysis buffer 1	50 mM glucose, 25 mM Tris pH 8.0, 10 mM EDTA
Alkaline lysis buffer 2	0.2 M NaOH, 1% (w/v) sodium dodecyl sulfate (SDS)
Alkaline lysis buffer 3	4 M sodium acetate, 2 N acetic acid

Table 3. Commonly used buffers

2.1.3 Biological Experiments

All prepared biological solutions and equipment, such as media, plastic and glassware, used in handling of the microbial cultures were autoclaved at 121 °C for 20 min in a Falcon LTE tabletop autoclave. All micro-biology work was performed in a Kendro HERAsafe KS-12 Class II Laminar Flow Cabinet. Bacterial cultures were incubated in New Brunswick Scientific, Innova 44 Incubator Shaker Series, or New Brunswick Scientific, Model G25 Incubator Shaker. Centrifugation was performed in a Beckman Coulter Avanti J-25 Centrifuge (using a JLA9.100 rotor for 700-1000mL or a F10BA-6x500y rotor for 350-450 mL or a JA-25.50 rotor for small 5-50mL volume) or Beckman Coulter Allegra X-12R Centrifuge. Microcentrifugation was performed using an Eppendorf Centrifuge 5415R, or Beckman Coulter Microfuge 18. Optical density was measured using a Molecular Devices Spectra Max Plus. Protein purifications were performed using an ATKA FPLC. Water was deionized and passed through Millipore 0.22 µm filter prior to use. Heating baths used were Grant JB Series. Polymerase chain

reaction was performed using an Applied Biosystems GeneAmp PCR System 9700. Gel electrophoretic analysis was performed using the following equipment: Apelex PS 304 Electrophoresis Power Supply (at 110 V), Bio-Rad Mini-sub Cell GT, visualized using an UVP Mini Benchtop UV Transilluminator TM-10E, and imaged using a Bio-Rad Universal Hood II (software, Quantity One 4.6.1). Determination of nucleic acid and protein concentration was performed using a NanoDrop Spectrophotometer ND-1000. Protein analysis used the following equipments: Invitrogen XCell *Sure*Lock™ and BioRad Power Pack Basic (at 200 V). Pre-cast NuPAGE Novex 4-12 % Bis-Tris protein gels (life technologies, NP0321, using MES buffer) were used for protein SDS-PAGE analysis. Protein samples were boiled for 5 min at 95% in Laemmli loading buffer (6x concentrate; 0.375M Tris pH 6.8, 12% SDS, 60% glycerol, 0.6M DTT, 0.06% bromophenol blue).

2.2 Nucleic acid analysis

DNA analysis was performed using an Agilent 1260 Infinity system using an Agilent Eclipse Plus C18 column (4.6 x 100 mm, 3.5 µm). 0.1M triethylamine acetate (TEAA) in water (solvent A) and 0.1M TEAA in 80:20 acetonitrile:water (solvent B) were used as mobile phase. The mixtures were separated by a linear gradient as stated in the example. After this the column was washed and equilibrated (5 min – 40->100% B; 10 min – 100% B; 15 min – 100->0% B; 20 min 0% B). The run was monitored by UV absorption at 260nm. Collected fractions were freeze-dried and analyzed by MALDI-TOF mass spectrometry using 3-hydroxypicolinic acid (hpa) as matrix or by ESI-MS. The hpa stock was prepared as a saturated solution in water:acetonitrile 1:1 and mixed with the sample 1:1 on the MALDI target. MALDI analysis was performed using a Waters Micromass MALDI micro MX using the manufacturer's recommended settings. For ESI analysis, samples were dissolved in water and injected in a flow of 1:1 Water:Methanol with 0.1mM NH_4OAc. The analysis was performed using a Micromass LCT (ESI-TOF-MS). The data was processed using MassLynx.

2.3 DNA Methods

2.3.1 Mutagenesis

Primers for mutagenesis were designed using the Agilent QuickChange Primer Design web tool (http://www.genomics.agilent.com/primerDesignProgram.jsp). The reactions were performed as described by the manufactures manual, either with the quick change mutagenesis XL II kit (Agilent; 200521) or with separately purchased components (PfuUltra II Hotstart, Agilent, 600630; DpnI, NEB, R0176).

2.3.2 PCR

Reactions were performed using the PfuUltraII Hotstart PCR Master Mix (Agilent, 600850) following the manufactures manual. In cases where this polymerase did not give satisfactory results the OneTaq polymerase was used (OneTaq® 2X Master Mix, M0482S, NEB) and annealing temperature was optimized. PCR product were analyzed and purified by agarose gel electrophoreses (0.7-1% agarose; 1x TEA buffer). DNA with the right molecular weight was retrieved using the gel extraction kit (Qiagen, 28704).

2.3.3 Cloning

All restriction enzymes were purchased from NEB. Digests were performed in 10-50 µL volume with 2 µL of each restriction enzyme and the recommended NEB buffer. Reactions were performed for 4 h at 37 °C. Subsequently, DNA was purified using the gel extraction kit following the manufactures instructions (Qiagen, 28704). Ligations were performed using the quick ligation kit (NEB, M2200S) following the manufactures instructions. Ligated plasmids were transformed into XL-10 gold (Agilent, 200314) following the manufactures protocol.

2.3.4 Transformations

In general, plasmid and competent cells were combined and kept on ice for 30 min. Then the cells were heat shocked at 42 °C for 40 s and put on ice for 2 min. Then 200 µL SOC medium was added and the cells were incubated at 37 °C for 1 h. Cells were subsequently plated on 1% LB/Agar plates containing the appropriate antibiotics and incubated overnight at 37 °C. For complete plasmid transformations 0.5 µL were transformed into 10 µL cells. For double transformation, mutagenesis

and ligations, 2 µL of each plasmid were transformed into 50 µL cells using 400 µL SOC during recovery.

2.3.5 Large scale DNA preparation

Large scale DNA preparation was performed essentially as described by Dyer *et al.*[2]. The 601 DNA plasmid (pUC 57) was transformed into HB101 cells, which were plated on ampicillin (100 µg / mL) containing LB-Agar plates and grown overnight. The next morning, two 15 mL pre-cultures (2xTY containing ampicillin at the same concentration) were inoculated with one colony each and grown at 37 °C until slightly turbid (4-5 h). These where then used to inoculate four flasks (700 mL TB with ampicillin at the same concentration). These were shaken at 37 °C for 22 h. The cells were then harvested (7000 rpm; 7 min; F10BA-6x500y rotor) and resuspended in a total of 120 mL alkaline lysis buffer 1 and split into two 500 mL centrifugation tubes (60 mL per tube). To each tube 120 mL Alkaline lysis buffer 2 was added and the mixture was shaken and incubated on ice for 10 – 20 min. Then 210 mL of cold Alkaline lysis buffer 3 was added and the mixture was shaken to mix and subsequently incubated on ice for 20 min. The mixture is then centrifuged (10k x g, 20 min, 4 °C) and filtered through tissue paper. Then 0.52 vol isopropanol is added the mixture is incubated at rt for 15 min and then centrifuged (10k x g, 30 min, 20 °C; in polypropylene tubes; Beckman Coulter 41121703). The supernatant was then decanted and the pellet transferred to a 50 mL Falcon tube. Then 20 mL TE 10/50 and 21 µL RNase A (100mg/mL; Qiagen; 19101) was added and the pellet was dissolved by shaking at 37 °C overnight. Then the mixture was centrifuged (JA25.50; 20k rpm; 15 min; 4 °C) to remove particulates. To the supernatant one-fifth of 4 M NaCl and two-fifth 40% Peg 6000 was added to precipitate the DNA. The mixture was shaken at 37 °C for 5 min, incubated on ice for 30 min and centrifuged at 3000g for 20 min. The supernatant was decanted and the pellet was dissolved in 15 mL TE 10/0.1 and the pellet is dissolved by shaking at 37 °C (2-3 h). The DNA concentration is then determined and 30 units of EcoRV (NEB; R0195) per nmol EcoRV site is added. The reaction is incubated at 37 °C for 16 h and completion is checked by agarose gel. If the reaction is complete, the plasmid is precipitated by addition of 0.18 vol 4 M NaCl and 0.31 vol 40% PEG 6000. The mixture is

incubated on ice for 1 h and the centrifuged (JA25.50, 27k x g, 20 min, 4 °C). The supernatant is decanted and the insert is recovered by addition of 3 vol EtOH, incubation on ice for 1 h and centrifugation (JA25.50, 27k x g, 20 min, 4 °C). Sometimes some insert precipitates in the first PEG precipitation step, in this case the PEG pellet is resuspendend and precipitation is repeated. The insert is dissolved in 3-4 mL TE 10/0.1 and purified by phenol extraction and ethanol precipitation.

2.4 Protein Expression and purification

2.4.1 Histone expression and purification

The appropriate histone encoding plasmid (used plasmids are given in each Chapter) was transformed into BL21(DE3)pLysS cells. A single colony was selected after overnight growth and used to inoculate 20-60 mL LB medium. This culture was grown at 37 °C until slightly turbid (approx. 4 h). The 10 mL culture was then added to 600 mL 2xTY medium containing the same antibiotics and grown at 37 °C until OD600 = 0.4-0.6 (approx. 4h). IPTG was added, to a final concentration of 0.5 mM and the flask was shaken at 37 °C for 2 h (3 h for Histone H4). Cells expressing histones are prone to lysis and longer expression time is detrimental on the yield. Cells were harvested by centrifugation (7000 rpm; 7 min; JLA9.100 rotor) and the pellet was resuspended in lysis buffer with protease inhibitor (one tablet per 50 mL, Roche, 11697498001, 40 ml total) and frozen and stored at -80 °C.

Frozen cells were thawed in a water bath, when the solution was viscous DNAase was added, sonicated 5 × at 60 % amplitude in 30 s bursts and then centrifuged at 20k rpm for 20 min. The supernatant was discarded and the pellet was resuspended in 40 mL TW buffer. Sonication was repeated at 60 % amplitude for 30 s, and the suspension centrifuged at 20k rpm for 10 min. The pellet was washed two more time with TW buffer and once with wash buffer. After addition of 0.2 mL DMSO the pellet was minced with a spatula and left at rt for 10 min. 7-15 mL unfolding buffer was then added, and shaken for 1 h at rt, after which it was centrifuged for 10 min at 20k rpm at rt. The supernatant was decanted and loaded onto an S200 size-exclusion column (equilibrated with 1 CV SAU-1000 buffer. The histone was eluted with SAU-1000 buffer, and analysed by SDS-PAGE. Fractions

containing purified histones were pooled, dialyzed against water containing 2 mM BME and lyophilized to avoid long storage in denaturing buffer.

The histones were further purified by ion exchange chromatography using a 1 or 5 mL SP-HP Hitrap (GE healthcare; 1mL: 17-1151-01; 5mL: 17-1152-01). For this, the lyophilised histone powder was dissolved in 1 – 4 mL (10-40 mg) SAU-100 buffer by rotation at rt for 20 min. Then particulates were removed by centrifugation (15k rpm) and the sample was loaded onto the Hitrap SP-HP equilibrated in SAU-100 buffer. The protein was eluted using a linear gradient to 70% SAU-1000 buffer over 20 CV. Pure fractions were pooled, dialyzed against water containing 2 mM BME, and lyophilized.

2.4.2 Protein Modification analysis

The chemical protein modification was monitored using liquid chromatography-mass spectrometry (LC-MS). LC-MS was performed using a Micromass LCT (ESI-TOF-MS) and a Shimadzu Prominence HPLC equipped (Chapter 5) or using a Micromass ZMD (single quadrupole) equipped with a HP 1100 HPLC (Chapter 4) with a Dionex Proswift RP-4H column (1 x 50 mm). The column was run at 0.4 mL / min. Solvents used were water with 0.1 % formic acid and 95:5 Acetonitrile: water with 0.1 % formic acid. The proteins were separated using a gradient from 5% B to 95% B in 4 min. After this the column was washed for 2 min with 95% B and then equilibrated with 5% B for 2.5 min. The electrospray source was set to 3200 V capillary voltage and 25 V cone voltage.

The data was analysed using Masslynx; the data was calibrated in Masslynx using a myoglobin standard. Deconvolution was performed using the maximum entropy algorithm.

2.4.3 Octamer, Dimer and tetramer reconstitution

Histone complex reconstitution was performed as described.[2-4] For the reconstitution, histones (powder after freeze-drying; 3-4 mg of each histone) were dissolved in 2 mL unfolding buffer (7 M Gd.HCl, 10 mM TRIS pH 7.5, 1 mM EDTA, 1 mM DTT) by rotating at 4 °C for 30 min. The concentration was then measured by nanodrop-1000. The used extinction coefficients are given in **Table 4**. The histones were then combined in a 1 / 1 / 1.1 / 1.1 ratio for H3 / H4 / H2A / H2B. The solution was

adjusted to a final concentration of 1-1.5 mg/mL and dialyzed into 3 charges of 2 L refolding buffer (2 M NaCl 10 mM TRIS pH 7.5 1 mM EDTA 5 mM BME) for at least 4 h each. After this precipitated protein was removed by centrifugation and the refolded histones were concentrated to 1-1.5 mL and purified by gel filtration (Superdex S200, 16/60, in refolding buffer). Octamers were concentrated to >3 mg/mL and either used for nucleosome reconstitution or adjusted to 50% glycerol and stored at -20 °C.

	MW	E	mg/mL
H2A	13950	4470	0.32043
H2B	13497	7450	0.551975
H3	15238	4470	0.293346
H4	11236	5960	0.530438

Table 4. Exctinction coefficients for the different histones used.

2.4.4 Nucleosome reconstitution by step-dialysis

Small scale reconstitutions by step-dialysis was used to find the optimal DNA-protein ratio for large scale reconstitutions and if only little material was needed, such as for some nMS experiments. To the desired amount of DNA (usually 0.9, 1 and 1.1 eq to protein) was added the same volume of 4 M KCl to bring the final concentration to 2 M KCl. In addition 1/100 of the volume of 1 M DTT was added to give a final concentration of 10 mM DTT. The mixture was incubated at 4 °C for 30 min before 1 eq of octamer was added (refolding buffer). The refolding reaction was then adjusted to 6 µM DNA with refolding buffer. A maximum of 100 µL was used in these reconstitutions. The mixture was then stepwise dialyzed into 2M, 850 mM, 650 mM and finally 250 mM KCl in 10 mM TRIS pH 7.5, 1 mM EDTA, 1 mM DTT. Each dialysis step was performed at least for 90 min at 4 °C.

2.4.5 Nucleosome reconstitution by salt-gradient dialysis

For large scale nucleosome reconstitutions,[2,4] salt-gradient dialysis was used. DNA and histone protein octamer were combined as described above and then put into a 6-8 kda MWCO dialysis tube. The tube was then transferred into 400 mL of refolding buffer (2 M KCl, 10 mM TRIS pH 7.5 1 mM EDTA, 1 mM DTT). Another 2 L of low-salt buffer (250 mM KCl, 10 mM TRIS pH 7.5 1 mM EDTA, 1 mM DTT) was prepared. The low salt buffer was then added at 0.8 mL / min to the refolding buffer with

the dialysis bag using a peristaltic pump. At the same time buffer was removed at the same speed using another peristaltic pump from the refolding buffer beaker, keeping the total volume in the beaker constant. This generates a slow salt gradient over 36h and results in efficient nucleosome formation. After completion of the gradient the refolded nucleosomes were dialyzed 2 times against no-salt buffer (10 mM TRIS pH 7.5, 1 mM EDTA, 1 mM DTT)

2.5 Protein Assays

2.5.1 Nucleosome analysis by native PAGE

Native PAGE analysis was performed using Novex TBE Gels (6%, 10 well, Invitrogen, EC6265BOX). The gel was prerun in 0.5x TBE for 2-3 h at 4 °C. Then 1-4 nmol of nucleosome in 5 µL, supplemented with 1 µL 30% sucrose as loading buffer, was added per well. The gel was run at 4 °C for 2-3h and subsequently stained with SYBR GOLD (Invitrogen, S-11494).

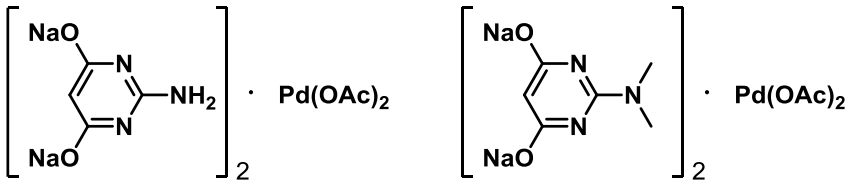
(1) Lundstrom, P.; Vallurupalli, P.; Hansen, D. F.; Kay, L. E. *Nat. Protocols* **2009**, *4*, 1641.
(2) Dyer, P. N.; Edayathumangalam, R. S.; White, C. L.; Bao, Y.; Chakravarthy, S.; Muthurajan, U. M.; Luger, K. In *Methods Enzymol.*; Allis, C. D., Carl, W., Eds.; Academic Press: 2003; Vol. Volume 375, p 23.
(3) Luger, K.; Rechsteiner, T.; Richmond, T. In *Chromatin Protocols*; Becker, P., Ed.; Humana Press: 1999; Vol. 119, p 1.
(4) Luger, K.; Rechsteiner, T. J.; Richmond, T. J. In *Methods Enzymol.*; Paul M. Wassarman, A. P. W., Ed.; Academic Press: 1999; Vol. Volume 304, p 3.

3 Nucleic acid modification by Suzuki-Miyaura cross-coupling

3.1 Introduction

Modified DNA building blocks have become a standard tool in molecular biology and biotechnology.[2,3] The accessibility of DNA with spectroscopic probes, affinity tags and drug-conjugates have greatly expanded the use of nucleic acids in these fields. DNA sequences bearing many different modifications can be prepared by solid phase synthesis (SPS) and are commercially available. For more specialized uses, the preparation of phosphoramidite building blocks is still often necessary. Post-synthetic modification of commercially available building blocks is an attractive alternative as it circumvents SPS and allows late stage diversification. Methods for this include Cu-catalyzed[2,4] or Cu-free alkyne-azide cylcoadditions,[5,6] Diels-Alder cycloaddition,[7,8] hydrazine formation,[9] reductive amination,[10] Staudinger ligation,[11] nucleophilic substitution,[12] native chemical ligation,[13] thiol-maleimide conjugation[14] and amide bond formation.[15]

Recently Pd-catalyzed Suzuki-Miyaura cross coupling (SMcc) has been reported as a method for post-synthetic DNA functionalization.[16,17] These two reports used phosphine based catalysts, which are oxidation prone, and the reaction required high temperatures, extended reaction times and inert atmosphere. In this chapter work which adapts a recently described phosphine free ligand system (**Scheme 1**) that allows rapid SMcc at 37 °C and pH 8.0 on proteins for the SMcc on oligodeoxynucleotides (ODN) is described.[18-23] It is shown that SMcc is an efficient method for the post-synthetic modification of nucleic acids and that photocrosslinker containing ODN can be generated by this method. ODN containing photocrosslinkers were used to identify potential interaction partners of a recently identified DNA modification 5-hydroxymethyl cytosine (hmC).[24-27]



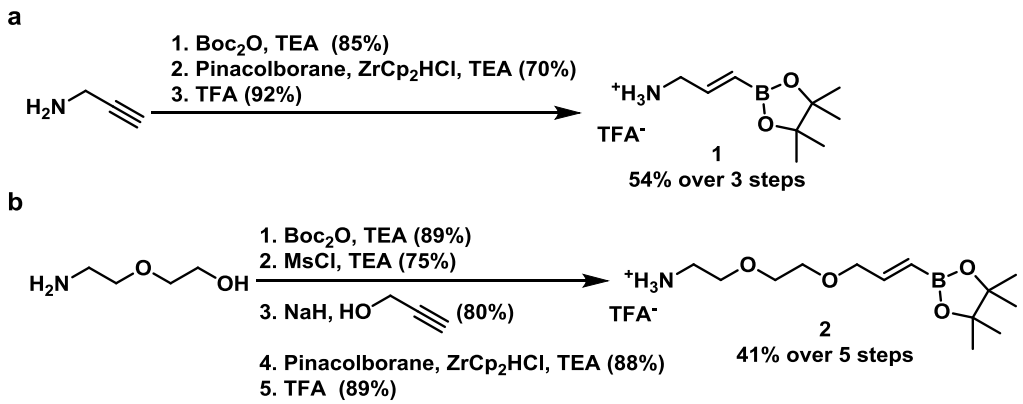
L1-Pd **L2-Pd**

Scheme 1. Ligand systems used for the SMcc described in this work. **L1**-**Pd** is derived from 2-aminopyrimidine-4,6-diol while **L2-Pd** is prepared from 2-(dimethylamino)pyrimidine-4,6-diol.

3.2 Results

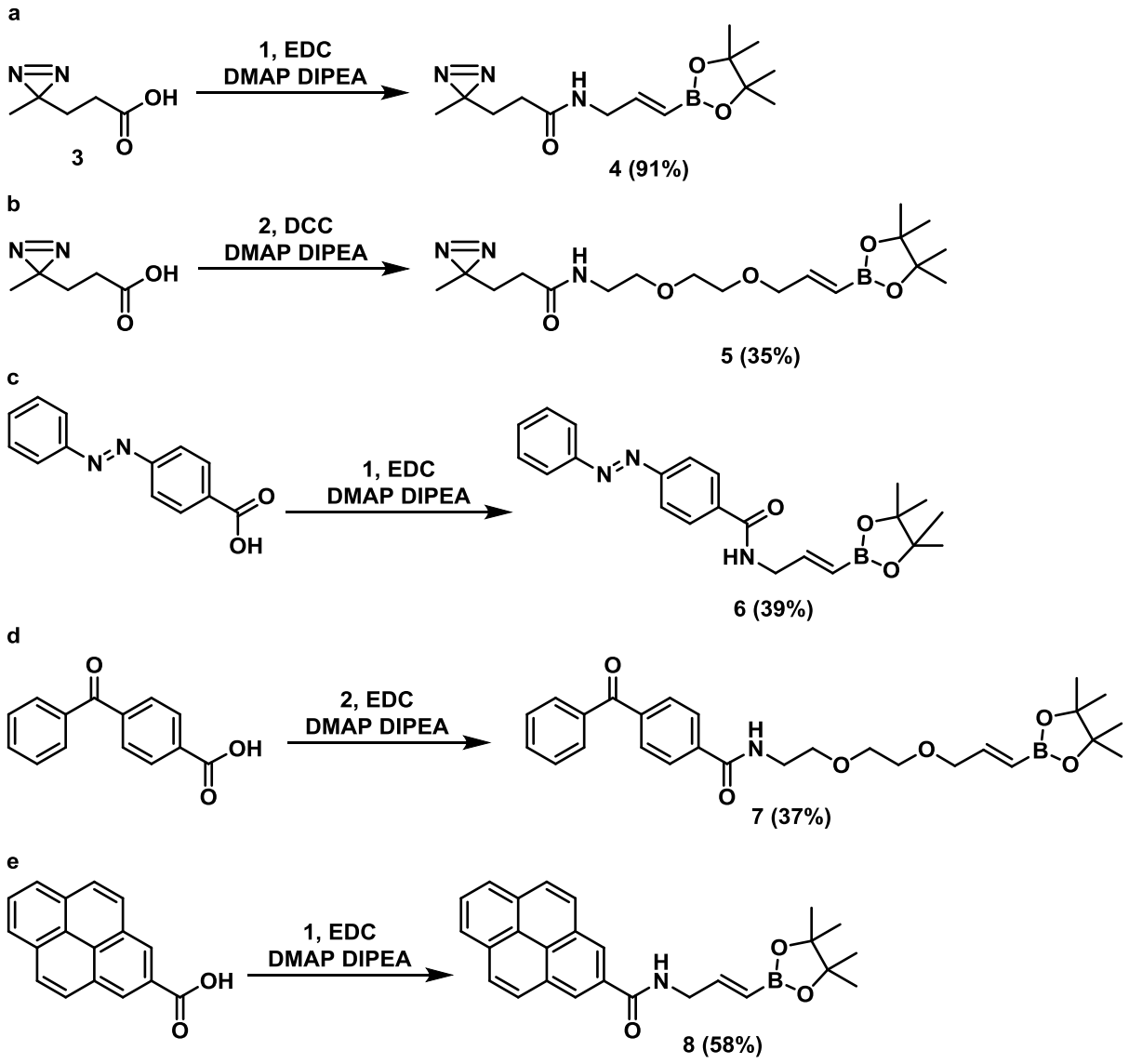
3.2.1 Synthesis of boronic acids

Boronic acid substrates for SMcc were synthesized using a modular approach. First boronic acid pinacol esters containing an amine with or without a short PEG linker were synthesized (**Scheme 2**). It was envisioned that the introduction of a hydrophilic linker would increase solubility of the boronic ester in aqueous conditions. The 3-aminoprop-1-en-1-yl boronic acid pinacol ester **1** was synthesized starting from propargylamine. Boc protection of the amine, followed by hydroboration with pinacolborane catalysed by zirconocene chloride hydride ($HZrCp_2Cl$)[28] in the presence of TEA[29] and subsequent deprotection yielded **1**. The synthesis of the longer boronic ester **2** containing a PEG linker, was started from 2-(2-aminoethoxy)ethan-1-ol. Boc protection, activation of the alcohol by mesylation and subsequent displacement with propargyl alcohol furnished the corresponding amino alkyne. Hydroboration with pinacolborane and deprotection yielded **2** in 41% over 5 steps.



Scheme 2. Synthesis of amine containing boronic acid pinacol esters with no (**a**) or a short (**b**) PEG linker.

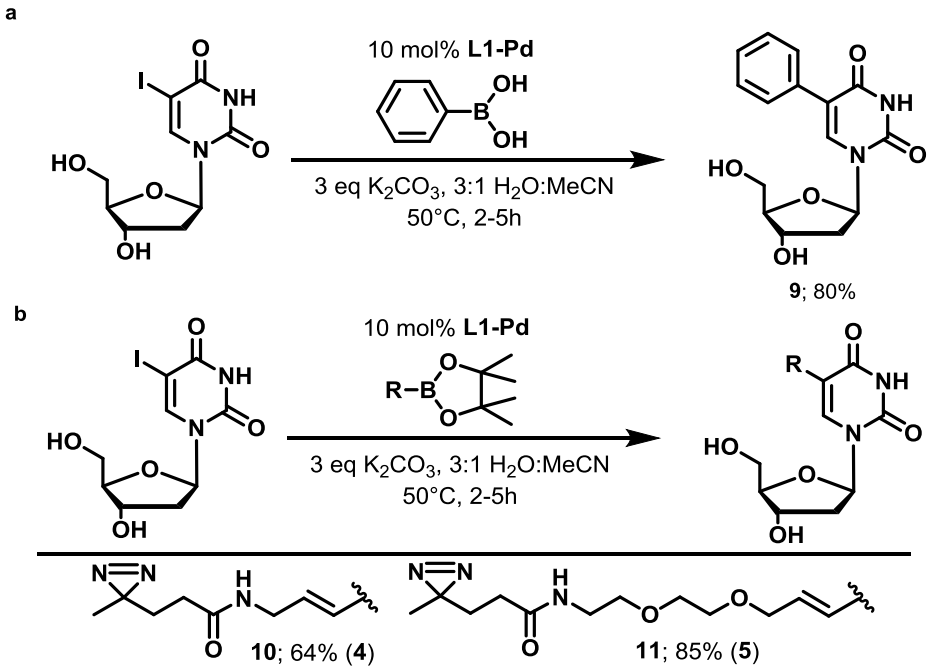
Various carboxylic acids were then coupled to the boronic esters **1** and **2** using *N,N*'-dicyclohexylcarbodiimide (DCC) or *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC) in the presence of 4-dimethylaminopyridine (DMAP) and *N,N*-diisopropylethylamine (DIPEA; **Scheme 3**). Yields were variable (37% to 91%), but sufficient material for subsequent SMcc trials were obtained in all cases. The diazirine containing acid **3** was synthesized following a one-pot literature procedure.[30-33] All the other acids that were used were commercially available.



Scheme 3. Synthesis of boronic pinacol esters bearing various functional groups used for biochemical purposes.

3.2.2 Nucleoside modification

5-Iodo-deoxyuracil (IdU) was selected as reagent for SMcc because ODN containing IdU are commercially available. To ensure that IdU is an efficient cross coupling partner, SMcc was first tested on the nucleobase. Reactions were performed in a water:acetonitrile mixture because of the low solubility of the nucleoside in water. With K_2CO_3 as base, the reaction gave good conversion after short reaction times at slightly elevated temperature (50 °C; **Scheme 4**). Reactions with phenyl boronic acid and boronic esters **4** and **5** gave good yields under these conditions. Importantly, the two diazirine moieties in boronic ester **4** and **5** were stable under these reaction conditions, allowing the introduction of photoreactive probes by SMcc. Because of various reported ligand systems that allow SMcc on the nucleobase level,[34] this reaction was not optimized.



Scheme 4. Reaction conditions for coupling of nucleobase IdU with phenylboronic acid (**a**) or different diazirine containing boronic acid pinacol esters (**b**). Boronic esters used are given in brackets.

3.2.3 DNA modification – optimization

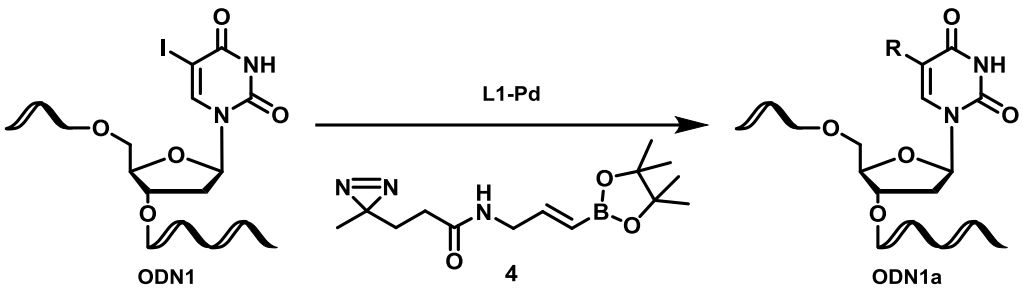
With the SMcc established as a feasible method for the modification of the IdU nucleoside, SMcc was performed on a short ODN containing all canonical bases and IdU (**Table 1**; **ODN1**) enabling identification of side reactions on canonical DNA bases.

Name	Sequence
ODN1	AT[5-IdU] GC
ODN2	CCT GAA GCA CTG TCT AGG TCT CTG
ODN3	ACC GA[5-IdU] GCA GTA G
ODN4	[Biotin] CCT GAA GCA CTG TCT AGG TCT C[5-IdU]G
ODN5	ACC G[5-IdU]T GA[5-IdU] GTG A
ODN6	AT[5-IdC] GC

Table 1. ODN used in this study. Non-canonical bases and functional groups are given in parentheses.

Firstly, reaction conditions similar to those reported for protein modification were used[18,19] (50 mM phosphate buffer pH 8.0, 10 eq **L1-Pd** and 500 eq **4**). The reactions were incubated at 37 °C for 5 h and analysed by HPLC and MALDI. Yields were determined by integration of the observed HPLC peaks. Full conversion of the starting material was observed, but in addition to 78% product the reaction yielded around 20% dehalogenated **ODN1** (**Table 2**; entry 1). To optimize the reaction the buffer conditions and pH of the reaction were varied (summarized in **Table 2**). Increasing the pH to 8.5 did not significantly improve the reaction (compare entries 1 and 2). A further increase of pH was avoided due to of the instability of nucleic acids at higher pH. Switching the buffer system to TRIS-HCl seemed to significantly improved the yield by limiting dehalogenation. Lowering the **L1-Pd** and the boronic ester concentration also resulted in higher yield. In initial trials, 0.5 eq **L1-Pd** was sufficient to allow fast reactivity. When the catalyst is stored on the bench for an extended period of time (> 3-4 weeks) some precipitation and lower reactivity is observed. Therefore, in later reactions 1 eq **L1-Pd** was used to compensate for this loss of activity. The reaction was also performed at room temperature, although this required extended reaction times. This might be beneficial if very temperature sensitive substrates are used. Further lowering the **L1-Pd** concentration down to 10 mol% was possible, but here longer reaction times were also necessary.

The need for a TRIS containing buffer to allow efficient reaction was notable. The role of TRIS in facilitating efficient SMcc could be manifold. The triol could help hydrolysis of the boronic acid ester to the boronic acid, which is more reactive for the transmetallation.[35] In addition, the presence of TRIS could result in a transesterification from the pinacol to the neopentyl glycol type boronic acid ester. These boronic acid esters have been shown by Carrow and Hartwig[35] to be significantly more reactive in transmetallation than the pinacol ester. Also, primary amines are reported as ligands for SMcc,[36] TRIS could coordinate the **L1-Pd** complexes yielding a more reactive complex.



Entry	4 [eq]	L1-Pd [eq]	Temp	Buffer	time	Substrate [%]	Dehalogenation [%]	Product [%]
1	500	10	37 °C	PO_4 8.0	5h	0	22	78[a]
2	500	5	37 °C	PO_4 8.5	5h	0	19	81[a]
3	500	5	37 °C	NH_4OAc 8.0	5h	0	16	84[a]
4	500	5	37 °C	TRIS 8.5	5h	0	5	95[a]
5	100	0.5	37 °C	,,	4h	0	0	95[a]
6	100	0.5	rt	,,	4h	51	0	49[a]
7	100	0.5	rt	,,	16h	0	0	95[a]
8	100	0.1	37 °C	,,	4h	72	2	26[a]
9	100	0.1	37 °C	,,	16h	10	3	87[a]

Table 2. Reaction optimisation for the SMcc on **ODN1**. [a] yield determined by relative integration of all observed peaks.

An example of an HPLC trace and MALDI characterization of the product for the SMcc with **ODN1** and boronic ester **5** is given in **Figure 1**.

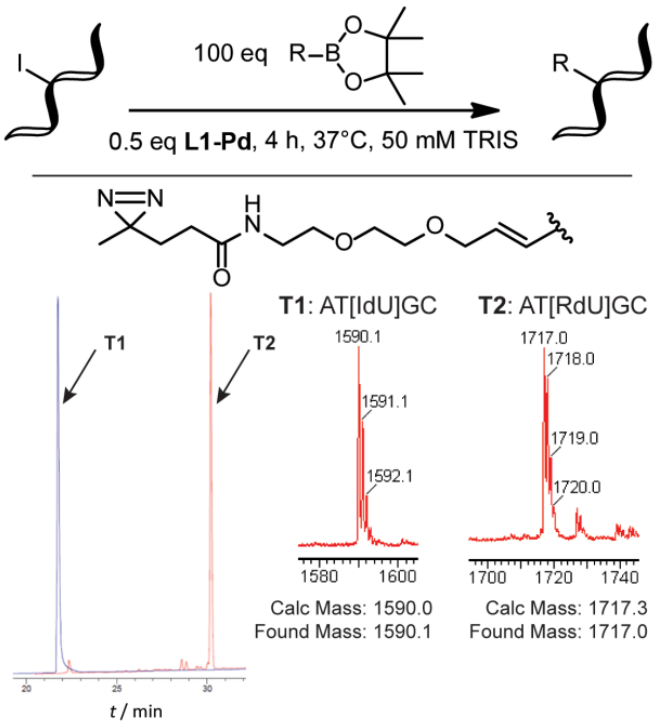


Figure 1. Example HPLC and MALDI results for the crosscoupling of **ODN1** with 100 eq **5** with catalysed by 0.5 eq **L1-Pd**. Figure adapted from Lercher *et al.*[1]

To preclude the possibility that the product arises from an undesired reaction with another nucleobase, a 24-mer (**ODN2**) that does not contain the IdU handle was subjected to the same reaction conditions. No additional peaks were observed upon HPLC analysis (**Figure 2**), confirming that the SMcc does not affect any of the canonical nucleobases.

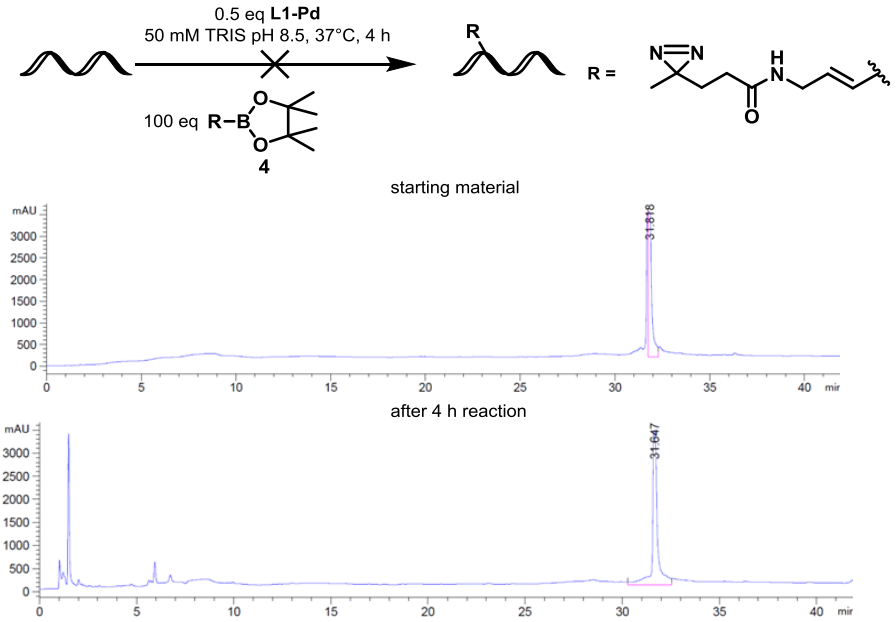


Figure 2. HPLC analysis of the control reaction of **ODN2** containing no IdU residue with 100 eq **4** in the presence of 0.5 eq **L1-Pd**.

3.2.4 DNA modification – boronic acid scope

Having established efficient reaction conditions for SMcc, the scope of boronic esters was explored using various substrates with biochemically useful moieties. Towards this goal, various vinyl boronic esters were synthesized and reacted with **ODN1** under the optimized conditions. The results are summarized in **Figure 3**. The diazirine containing boronic esters **4** and **5** were coupled in 90% and 93% yield, resulting in **ODN1a** and **ODN1b,** respectively, containing a photoreactive moiety which can be used for DNA interaction partner identification.

Azobenzenes are commonly used photoswitches in chemical biology.[37] Boronic ester **6** containing an azobenzene moiety was coupled in 71% yield to give **ODN1c**. The introduced azobenzene moiety could potentially be used to modulate protein-DNA interactions or DNA hybridization.

A third photoreactive probe, a benzophenone, was installed in 86% yield using boronic ester **7**. Pyrene moieties can be used as electron acceptor or fluorescent probes in a variety of different settings. Using boronic ester **8** it was possible to install a pyrene moiety. This reaction was slower, requiring 16 h reaction time resulting in 86% yield. Due to poor solubility of the pyrene containing boronic ester, the reaction was performed in 40% (v/v) aqueous MeCN. Because of the well-defined length and geometry of DNA, it is often used as a molecular ruler to measure distances for bivalent interactions for example lectin-carbohydrate interactions.[14,38] In addition, the conjugation of simple carbohydrates to DNA has been shown to be useful for improving cell uptake of ODN.[39] Using SMcc it was possible to generate an ODN containing a glucose moiety (**ODN1f**). Extended nucleobases containing small heterocycles at the C-5 position have fluorescent properties that are often environment sensitive[40-42] and can be used to e.g. detect abasic sites in DNA. Using 3-furanylboronic pinacol ester, it was possible to generate **ODN1g** which bears a furan moiety at the C-5 position.

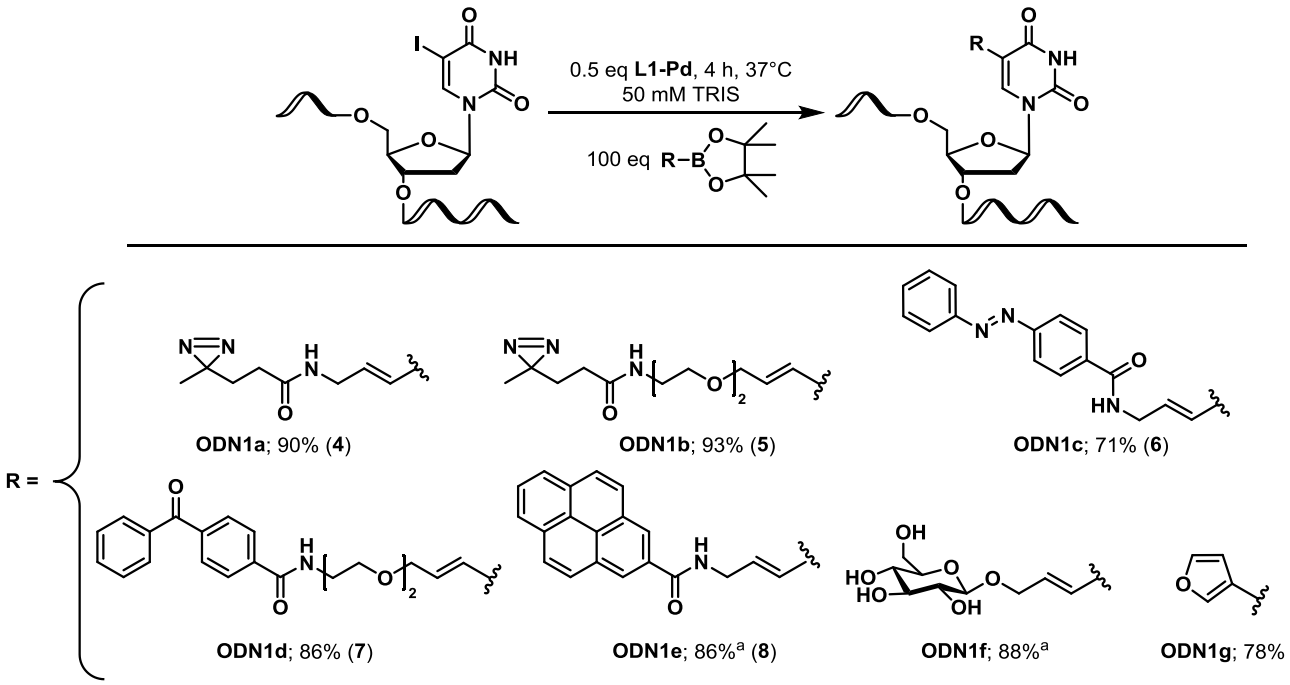


Figure 3. Scope for the SMcc of **ODN1** with different boronic esters. The boronic ester used is given in brackets for each example. [a] Reaction was shaken for 16 h at 37 °C

3.2.5 DNA modification – intermediate isolation

When performing the reaction using 3-furanylboronic pincol ester an additional peak at a retention time of 13.7 min in the HPLC chromatogram was noticed when monitoring the reaction by HPLC after 3 h (**Figure 4**).

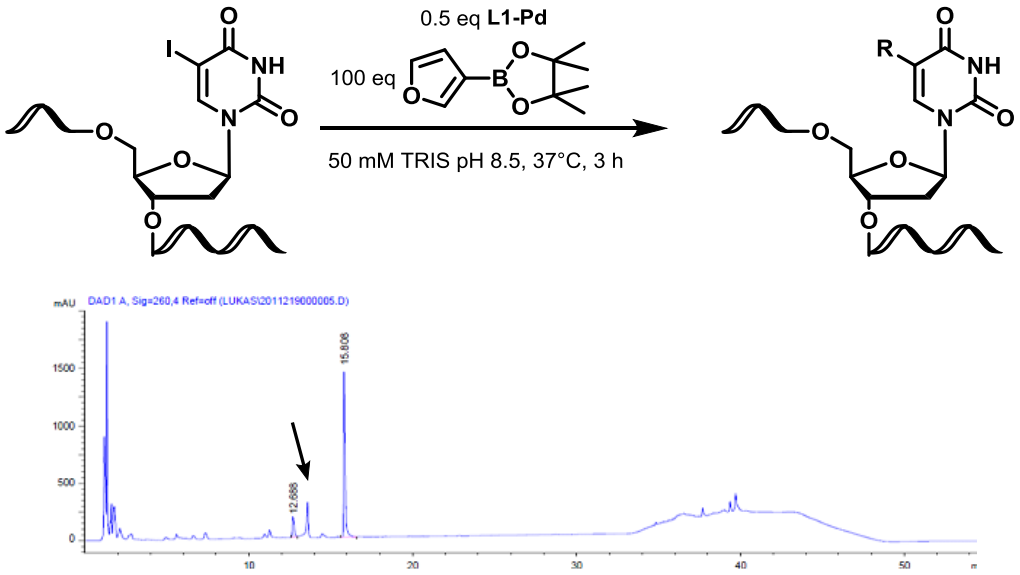


Figure 4. HPLC analysis of the reaction of **ODN1** with 100 eq 3-furanylboronic pincol ester catalysed by 0.5 eq **L1-Pd**. The starting material elutes at 12.6 min and the product at 15.8 min.

After isolation the reaction product was analysed by MALDI-MS. The mass and isotope pattern corresponds to a palladium containing species of **ODN1** after de-iodination of IdU, along with one molecule of the MALDI matrix hydroxypicolinic acid (HPA) as adduct (**Figure 5a** and **b**). Because no such species is observed upon incubation of **L1-Pd** with **ODN1**, it was reasoned that this corresponds to a palladated intermediate in the SMcc catalytic cycle. **ODN1** was purified by HPLC in a triethylamine acetate buffer prior to MALDI analysis, thus the coordinated **L1** is probably displaced by TEA during the purification. Because of the large excess of HPA during crystallization for the MALDI measurement, one molecule of HPA can coordinate to the Pd bound to the ODN (**Figure 5c**). The observed species could correspond to one of the boxed intermediates (**Figure 5d**). Similar species have been previously observed in ESI-MS monitoring of SMcc on small molecules.[43]

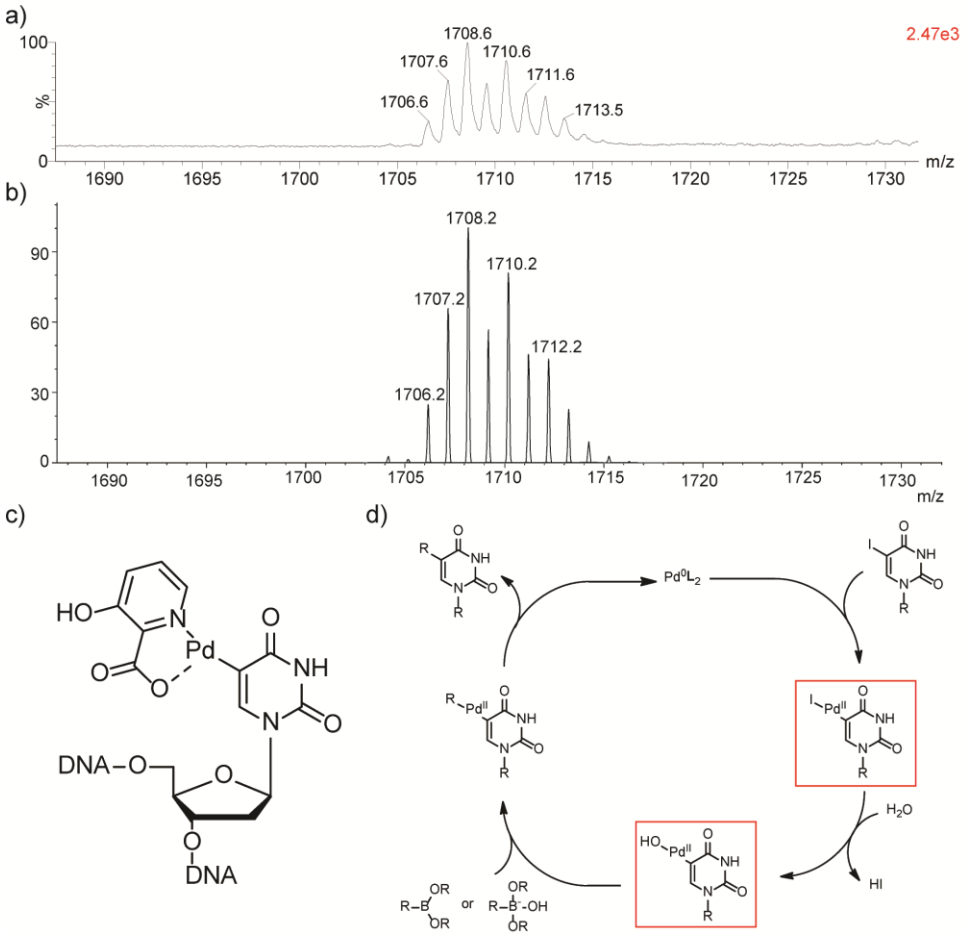
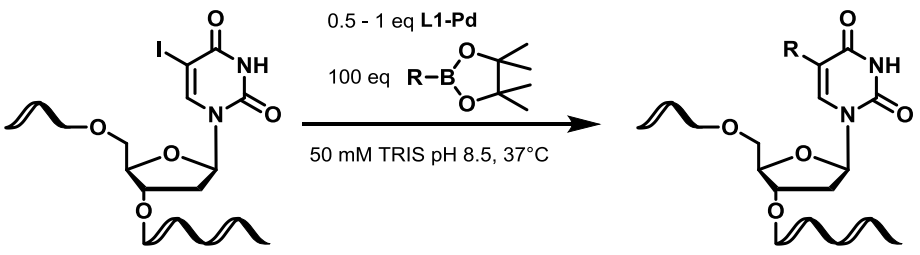


Figure 5. Analysis of the palladated intermediate from the reaction of **ODN1** with 3-furanylboronic pincol ester. **a** Observed MALDI-MS spectra. **b** Theoretical MS spectra for the possible palladated ODN. **c** Proposed structure of the intermediate. **d**

Scheme of the reaction cycle for the SMcc, the boxed intermediates could give rise to the observed species. Figure is adapted from Lercher *et al*.[1]

3.2.6 DNA modification – ODN scope

With reaction conditions established for the short 5-mer **ODN1**, it was of interest to explore the range of ODN that could serve as substrates in the SMcc (**Table 3**). Firstly, the effect of ODN length on the reaction was investigated. For this, a 13-mer (**ODN3**) and a 21-mer (**ODN4**) were used. **ODN4** also contains a biotin moiety that was later used for isolation of protein interaction partners of this ODN. For the 13-mer reactions with the short diazirine containing boronic ester **4** and the PEG chain containing diazirine boronic ester **5** proceeded with 71% and 73% yield to afford **ODN3a** and **ODN3b** respectively. The diazobenzene **6** was coupled in 59% yield to give **ODN3c**. The glucose containing boronic ester was coupled to **ODN3** to yield the glycosylated **ODN3f** in 83% yield. The slight decrease in yield when comparing the reaction of **ODN1** with **ODN3** may be attributed to increased steric restraints on the reaction by increasing DNA length. The longer **ODN4** was coupled with **4** and **5** to yield products **ODN4a** and **ODN4b** respectively, containing both a photoreactive moiety and a biotin tag for affinity purification. A 13-mer containing 2 IdU moieties gives decent yields for the crosscoupling with **4** resulting in a doubly modified product **ODN5a**. The crosscoupling of boronic ester **4** with **ODN5** gave **ODN5a** in yields slightly lower than that obtained for the 5IdU containing ODN (69% compared to 90% for **ODN1**). To see if 5-iododeoxycytidine (IdC) containing ODN are also good substrates in SMcc a 5-mer containing a central IdC (**ODN6**) was subjected to the previously optimized reaction conditions with boronic ester **4**. The SMcc product **ODN6a** was obtained in 56% yield, confirming that IdC also presents a good handle for SMcc (**Figure 6**).



Product	catalyst	time	yield

ODN3a	0.5 eq; **Pd-L2**	6	71	[a]
ODN3b	0.5 eq; **Pd-L2**	6	73	[a]
ODN3c	0.5 eq; **Pd-L2**	6	59	[a]
ODN3f	0.5 eq; **Pd-L2**	16	83	[a]
ODN4a	0.5 eq; **Pd-L1**	6	78	[b]
ODN4b	0.5 eq; **Pd-L1**	6	67	[b]
ODN5a	1 eq; **Pd-L2**	8	69	[a]

Table 3. Reaction of various ODN with different boronic acids. [a] Determined by HPLC (λ_{max} 260 nm) [b] Yield of isolated product.

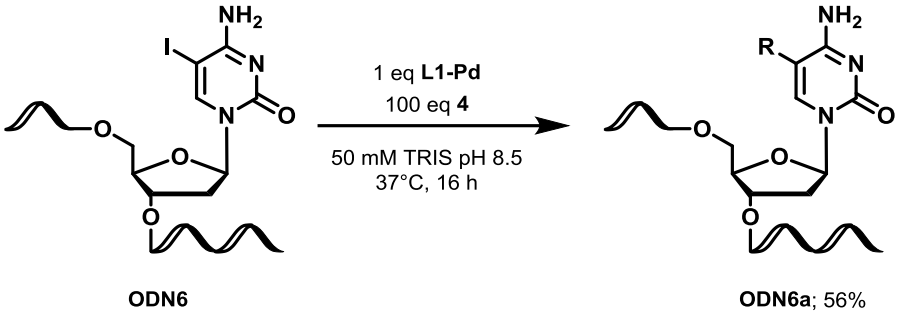


Figure 6. SMcc of the IdC containing **ODN6** with boronic ester **4**. Yield was determined by HPLC (λ_{max} 260 nm)

3.2.7 RNA modification

RNA is a more labile biological macromolecule than DNA due to rapid degradation via chemical hydrolysis[44] and by the presence of RNases.[45] Because of this instability, a modification method for RNA should be as mild and fast as possible. It is of interest to adapt SMcc to modify RNA. Initially, it was tested whether the reaction conditions caused degradation of the oligonucleotides. No degradation was observed and excellent recovery was obtained when incubating an 5-iodouridine (IU) containing 10-mer (AGU C[IU]G CAC G; **ON1**) with up to 10 eq catalyst under our standard buffer conditions (50 mM TRIS ph 8.5; Figure 7 a-c). Using the previously established reaction conditions, SMcc of **ON1** with boronic ester **5** yielded only 30% of the diazirine containing **ON1a** (Figure 7 d).

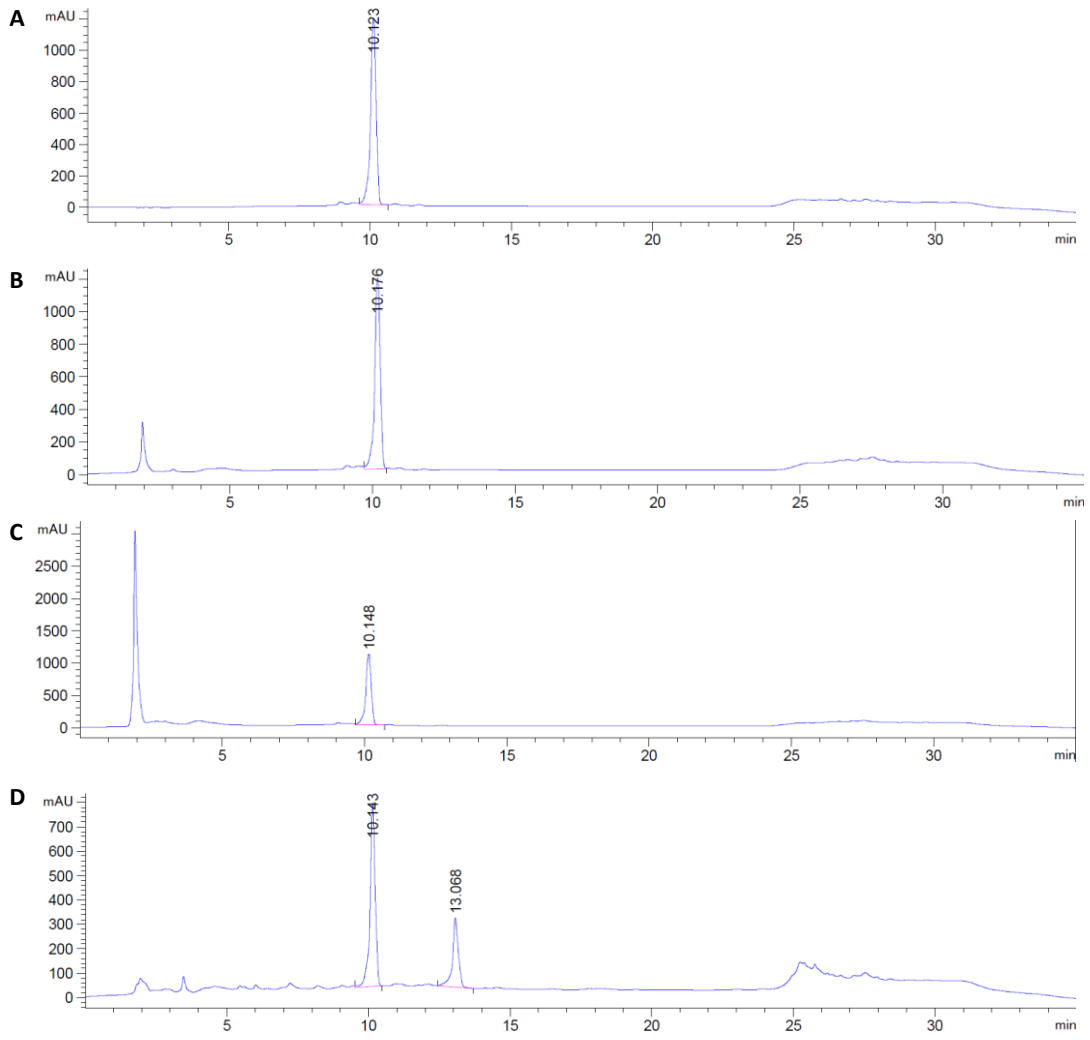


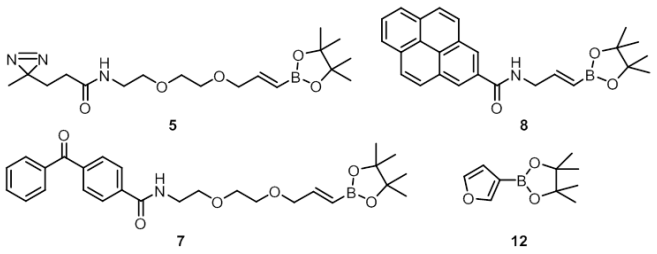
Figure 7. RNA cross coupling stability test. **a** Incubation of 10 uM RNA oligonucleotide with reaction buffer for 16h. **b** Incubation of 10 uM RNA oligo with 1 eq Pd for 16h. **c** Incubation of 10 uM RNA oligonucleotide with 10 eq Pd for 16 h. **d** Reaction with 0.5 eq Pd and 100 eq boronic acid for 4 h.

To optimise the reaction conditions, palladium and boronic ester concentrations were varied (Table 4). The highest yields, comparable to those obtained using the DNA IdU substrates, were obtained using **Pd-L2** with low ligand and high boronic ester loadings with extended reaction times (entry 8 and 9).

Entry	Ligand	Eq BOR2	Eq [Pd]	time	Start	Prod
1	L1	500	5	8h	----	51%
2	L1	100	5	8h	----	45
3	L1	500	1	8h	41%	34%
4	L1	100	1	8h	43%	21%
5	L2	500	5	6h	----	68%
6	L2	500	1	6h	----	52%
7	L2	100	1	6h	----	30%
8	L2	500	1	16h	----	82%
9	L2	100	1	16h	----	76%

Table 4. Optimization of RNA reaction between **ON1** and boronic ester **5**

Then, the scope of the SMcc with RNA was investigated. An additional photoreactive moiety, a benzophenone, was an effective substrate for SMcc and gave **ON1b** in excellent yield. Together with the previously generated **ON1a**, these three ON can be used to probe protein-RNA interactions. A pyrene moiety was introduced by SMcc to give **ON1c** in 40% yield. SMcc of the furan containing boronic ester **12** gave only 18% yield (**ON1d**), probably due to larger steric hindrance at the C-5 position in RNA.



Entry	Ligand	eq	BOR	eq	time	Product
1	L2	2	7	200	16h	**ON1b** 95%
2[a]	L2	2	8	200	16h	**ON1c** 40%
3	L2	2	12	200	16h	**ON1d** 18%

Figure 8. Boronic esters and conditions used for the SMcc on **ON1**. [a] reaction was performed with 40% (v/v) MeCN in H_2O

3.2.8 Preliminary biological application

Modification of the C-5 position of cytosine is an important epigenetic mechanism to regulate transcription. Cytosine can be methylated, hydroxymethylated, formylated or carboxylated at this position.[24,26] Two major targets towards understanding the mechanism of regulation by these epigenetic cytosine modifications are determining where in the genome they are located by sequencing[26] and determining how they can influence the epigenetic landscape.[46,47] To investigate how these modifications signal, it is essential to understand how they influence protein-DNA interactions.[48] To identify selective hmC interaction partners, it was thought to adopt a stable isotope

labelling with amino acids in cell culture (SILAC) based MS approach.[49] Additionally, it was decided to use photocrosslinking to facilitate identification and trap weak interactions.[50-52] The workflow of the experiment is detailed in **Figure 9**. Starting from **ODN4** containing a 5' terminal biotin and a 5-IdU base two positions from the 3', the photocrosslinker was first introduced by SMcc using diazirine **5** which contains a short PEG linker. The resulting **ODN4d** was then subjected to PCR with either dCTP or dhmCTP, globally replacing C with hmC. The respective probes were then incubated with differentially isotopically labeled HeLa nuclear extract. Interaction partners were crosslinked by irradiation at 365 or 311 nm. Subsequently the samples were pooled and immobilized on streptavidin beads using the biotin tag. The beads were washed extensively to minimize non-specific interaction partners. The proteins were then released from the beads, separated by SDS-PAGE, trypsinised and analysed by LC-MS.

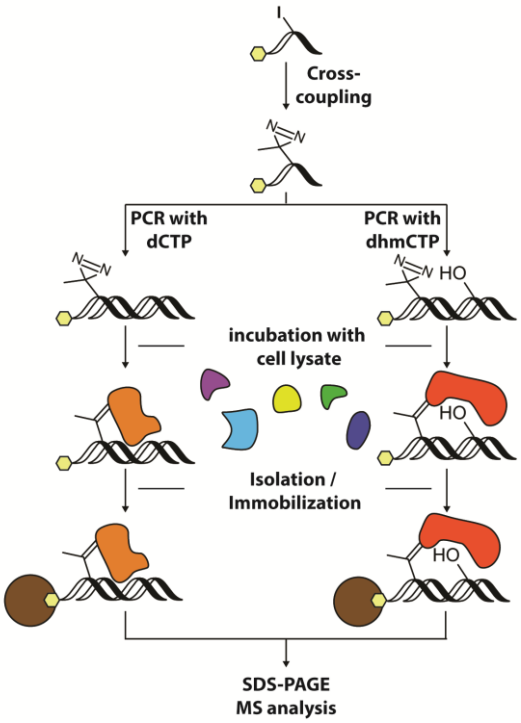


Figure 9. SILAC crosslinking workflow, adapted from Lercher *et al*.[1]

The efficiency of SMcc to introduce the diazirine based photocrosslinker has been previously established. Because in this case the differently modified nucleotides were introduced by PCR, it was investigated if the presence of the diazirine in the ODN influences the efficiency of the PCR reaction.

For this, PCRs with **ODN4** or **ODN4d** as primer were compared. The PCR was performed with a slight excess of the complementary primer to ensure that no single stranded **ODN4d** remained, as this could potentially affect the results because of binding of ssDNA proteins. As shown in **Figure 10**, both **ODN4** and **ODN4d** are amplified with similar efficiency using the Vent exo$^-$ polymerase. This polymerase was chosen because it has previously been shown to be tolerant to C-5 modifications.[6] Also, because it lacks 3'-5' exo nuclease activity there is no danger of removal of the modified base from the primer. Replacement of dCTP by dhmCTP is well established and known not to affect PCR yields[53] and the efficiency of hmC incorporation was also demonstrated in Chapter 6. Using multiple PCRs sufficient material for the photocrosslinking reactions was obtained.

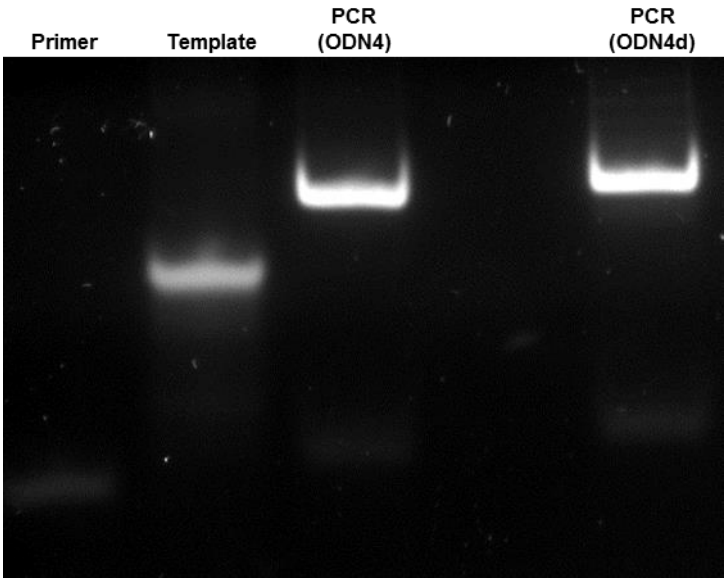


Figure 10. PAGE gel for confirmation of PCR efficiency with **ODN4a**. No free primer was observed and equal yields were obtained from the reaction with the IdU containg **ODN4** and the diazirine containing **ODN4a** (adapted from Lercher *et al*[1]).

To confirm that the diazirine photocrosslinker is stable under the PCR conditions **ODN1a** was subjected to the same PCR cycle; no degradation of the diazirine moiety was observed (**Figure 11**).

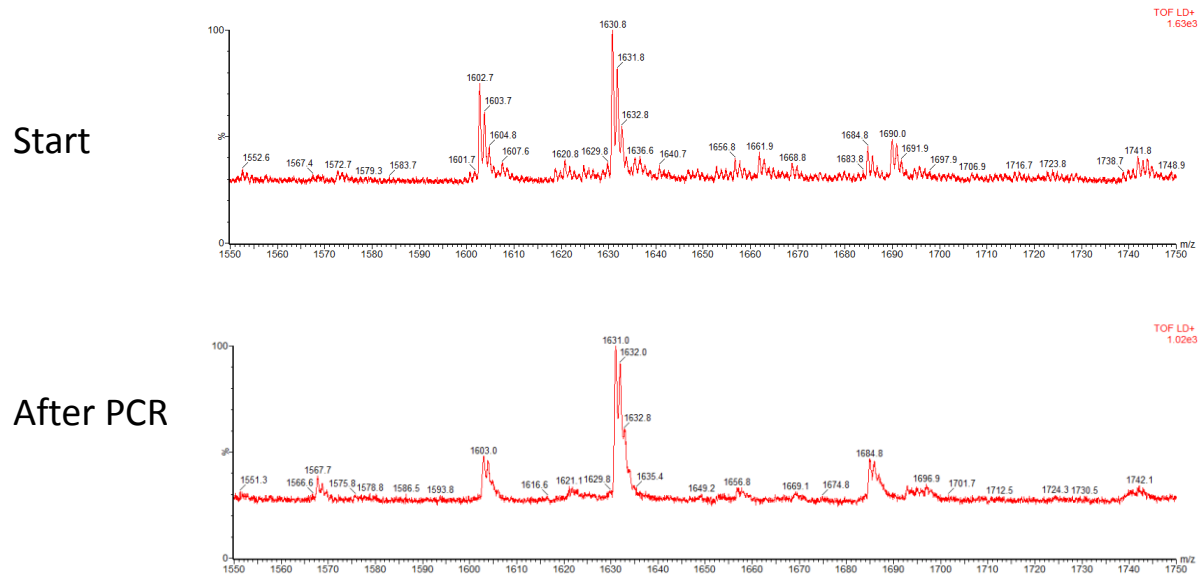


Figure 11. Thermal stability test of the incorporated diazirine moiety in **ODN1a**. No degradation of the diazirine was observed.

Next, the wavelength of the light source used for the crosslinking was optimized. For this, **ODN1a** was irradiated at different wavelengths using various UV lamps. Full decomposition of the diazirine moiety was observed with irradiation at 311 nm or 365 nm for 30 min. Using a high power Xe-lamp (>300 nm) only very poor signals could be observed in the MALDI-MS, probably due to DNA decomposition at high power and lower wavelengths. The observed products corresponded to a hydroxylated product **ODN1h**, which would result from trapping of the generated carbene by water, and one product **OND1j** corresponding to the loss of N_2 and rearrangement of the carbene to the double bond. It was decided to use 311 nm for irradiation because of complete decomposition after 30 min of irradiation and the convenience of performing the irradiation in the especially designed Caprotec Caprobox.[54,55]

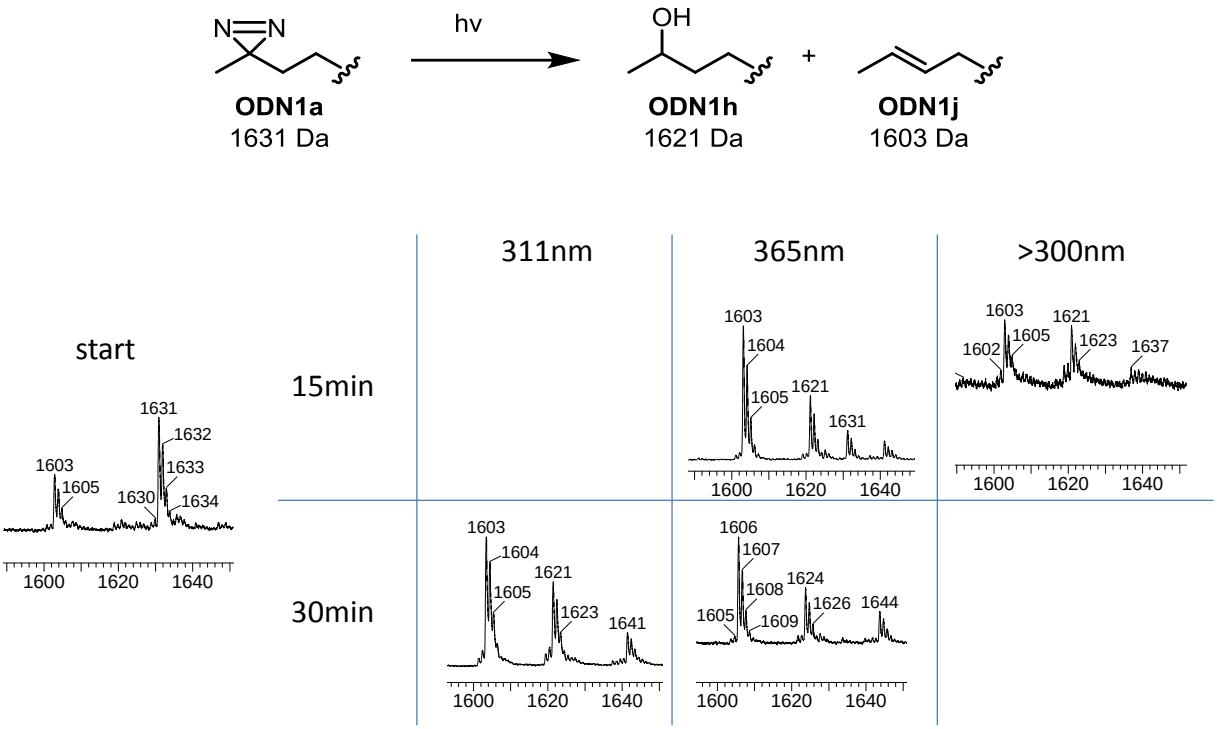


Figure 12. Photochemical decomposition of the diazirine moiety on **ODN1a**. The ODN was irradiated on ice of the indicated wavelength. Aliquots were taken at different times, desalted and analysed by MALDI-MS.

With the irradiation conditions set, initial experiments to examine crosslinking to protein targets were performed. For this, nuclear extracts were prepared from HeLa cells, mixed with the C or hmC containing probe, photocrosslinked and immobilized on streptavidin beads. The biotin-enriched sample was eluted by boiling the streptavidin beads in SDS-loading buffer and the eluted proteins were analysed by SDS-PAGE and western blot with streptavidin-alkaline phosphatase conjugate (**Figure 14**). Bands on this blot should only be observed when a covalent bond between the DNA and the protein of interest is formed, because the sample is denatured with SDS before loading on the SDS-PAGE. As expected, no proteins were detected in the input and the control without UV irradiation. Both the hmC and C probe lanes showed strong bands, corresponding to DNA-protein conjugate, confirming efficiency of this method for creating covalent DNA-protein conjugates upon UV irradiation.

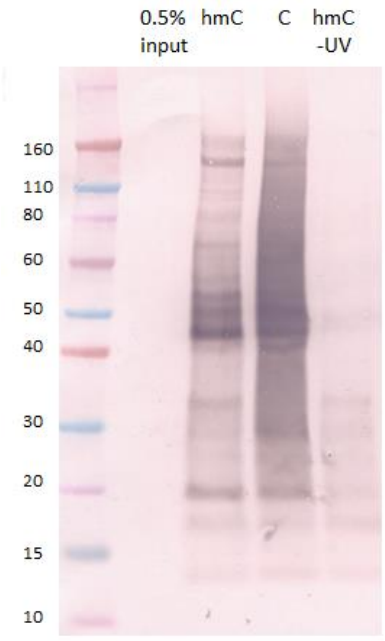


Figure 13. Western-blot analysis of the crosslinking and isolation protocol. Bands corresponding to protein-DNA cross-linked species are detected only upon irradiation (lanes hmC and C) and not when UV irradiation is omitted (lane hmc – UV)

SILAC experiments were then performed for identification of hmC interaction partners by digestion with trypsin and mass spectrometric analysis. In SILAC, isotope labelled cell extract is prepared from cells grown in medium containing ^{13}C and ^{15}N labelled "heavy" arginine and lysine. Trypsin cleaves the protein *N*-terminal to Arg or Lys, this results in formation of peptides containing one isotope labelled amino acid. To identify interaction partners, the hmC containing probe was incubated and crosslinked with normal "heavy" cell lysate and the C containing control with "light" lysate. To reduce false positive identification, experiments were performed in duplicate with heavy and light lysate switched. Interaction partners were characterised by a high H/L ratio in the first experiment and a low H/L ratio in the reverse experiment. A number of DNA binding proteins were identified (**Table 5**, full table in appendix) confirming the possibility of identifying specific DNA-protein interactions by this method.

Protein name	hmC	Ratio H/L Normalized
54 kDa nuclear RNA- and DNA binding protein	Heavy	1.6
	Light	0.7
100 kDa DNA-pairing protein	Heavy	1.8
	Light	0.3
ATP-dependent RNA helicase A	Heavy	2
	Light	0.2
ATP-dependent RNA helicase DDX3X	Heavy	1.5
	Light	0.5
DNA ligase 3	Heavy	1.3
	Light	0.6
20 kDa CGG-binding protein	Heavy	1.9
	Light	0.4

Table 5. Examples of protein interaction partners identified for hmC by the photocrosslinking SILAC method.

3.3 Conclusions

In the work in this chapter it has been shown that the SMcc is a viable method to modify both RNA and DNA. Yields are usually high, although a few features of the reaction and the oligonucleotides limit the scope. The best substrates are unhindered vinyl boronic esters. The SMcc with small heterocyclic boronic acids proceeds efficiently with DNA while the reactivity with RNA is very poor. Larger aromatic rings, or β-substituted vinyl boronic acids were not tested but it is speculated that increasing the steric demand of these system would not allow efficient SMcc.

With the increasing demand for further functionalized ODN, it is hoped that the method described herein will find use in the extended biophysical research community.

3.4 Experimental procedures

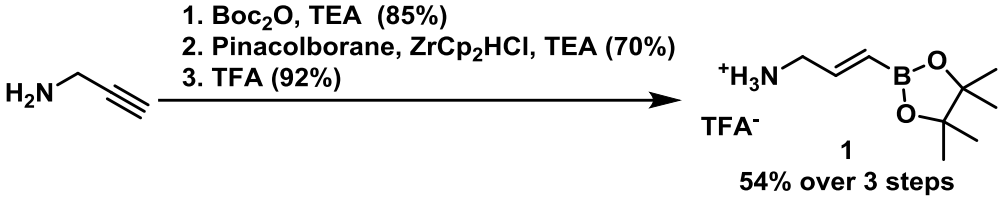
General considerations for synthesis and MS analysis of ODN and ON are given in Chapter 2.

3.4.1 Used oligonucleotides

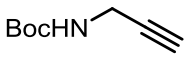
Name	Sequence (3'-5')	Obtained from
ODN1	AT[5-IdU] GC	ATDbio
ODN2	CCT GAA GCA CTG TCT AGG TCT CTG	Sigma
ODN3	ACC GA[5-IdU] GCA GTA G	ATDbio
ODN4	[Biotin] CCT GAA GCA CTG TCT AGG TCT C[5-IdU] G	ATDbio
ODN5	ACC G[5-IdU]T GA[5-IdU] GTG A	ATDbio
ODN6	AT[5-IdC] GC	ATDbio

Table 6. Used oligodexynucleotides.

3.4.2 Synthesis of the short boronic ester amines for coupling



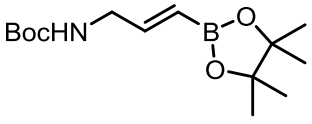
3.4.2.1 tert-*Butyl prop-2-yn-1-ylcarbamate*[56]



The literature procedure reported by Molander *et al.*[56] was followed. To a solution of di-*tert*-butyl dicarbonate (4.8 g, 22 mmol, 1.1 eq) in CH_2Cl_2 (20 mL) at 0 °C was added dropwise propargylamine (1.3 mL, 1.1 g, 20 mmol, 1 eq) and subsequently NEt_3 (5.6 mL, 4.05 g, 40 mmol, 2 eq). The resulting solution was warmed to rt and allowed to stir for 2 h. Then the reaction was quenched with sat. NH_4Cl (10 mL) and diluted with EtOAc (150 mL). The organic phase was washed with sat. NH_4Cl (100 mL), sat. $NaHCO_3$ (100 mL) and brine (100 mL), dried ($MgSO_4$), filtered and solvents were removed *in vacuo* to yield the desired compound (2.68 g, 17 mmol, 85%) as a clear oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.46 (s, 9 H, $(C\underline{H}_3)_3C$) 2.22 (t, *J*=2.5 Hz, 1 H, CC<u>H</u>) 3.92 (br s, 2 H, $C\underline{H}_2$) 4.72 (br s, 1 H, N<u>H</u>); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 28.3 ($(\underline{C}H_3)_3C$) 30.4 ($\underline{C}H_2$) 71.2 (C<u>C</u>H) 79.8 ($(CH_3)_3\underline{C}$) 80.0 (C) 155.3 (OCONH); LRMS m/z (ESI+): 178.09 $[M+Na]^+$; IR (neat): 3301, 2979, 2933, 1692, 1510, 1455 cm^{-1}.

3.4.2.2 (E)-tert-*Butyl 3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)allylcarbamate*[29,57]

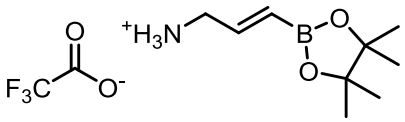


The literature procedure reported by Wang *et al.*[29] was followed. To neat *tert*-butyl prop-2-ynylcarbamate was added pinacolborane (1.4 mL, 1.23 g, 9.6 mmol, 1.5 eq), NEt_3 (100 uL, 0.64 mmol, 0.1 eq) and $ZrCp_2HCl$ (165 mg, 0.64 mmol, 0.1 eq). The solution was stirred at 65 °C for 16 h. After completion the reaction was allowed to cool to rt and was diluted with EtOAc (25 mL) and quenched

with sat. NH_4Cl. The mixture was further diluted with EtOAc (50 mL) and the organic layer was washed with sat. NH_4Cl (50 mL), sat. $NaHCO_3$ (50 mL) and brine (50 mL), dried ($MgSO_4$), filtered and solvents were removed *in vacuo*. The product was purified by automated column chromatography (Biotag KP-SIL SNAP 50g cartridge 0-40% over 10 CV EtOAc in petrol) to yield the desired compound as a yellow oil (1.27g, 4.5 mmol, 70%).

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.26 (s, 12 H, ($C\underline{H}_3$)$_2$COB) 1.44 (s, 9 H, ($C\underline{H}_3$)$_3$C) 3.84 (app s, 2 H, $C\underline{H}_2$) 4.67 (s, 1 H, N$\underline{H}$) 5.57 (app dd, *J*=18, 2 Hz, 1 H, C$\underline{H}$B) 6.58 (app d, *J*=18 Hz, 1 H, C$\underline{H}$CH$_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 24.7 ($\underline{C}H_2\underline{C}H_2$) 28.4 (($\underline{C}H_3$)$_3$C) 44.0 (CNH$\underline{C}H_2$) 79.4 ($\underline{C}$OB) 83.3 ((CH$_3$)$_3\underline{C}$) 149.4 (CHB) 155.7 (OCONH); LRMS m/z (ESI+): 306.20 [M+Na]$^+$; IR (neat): 3356, 2978, 2932, 2361, 2341, 1697, 1643, 1519 cm^{-1}.

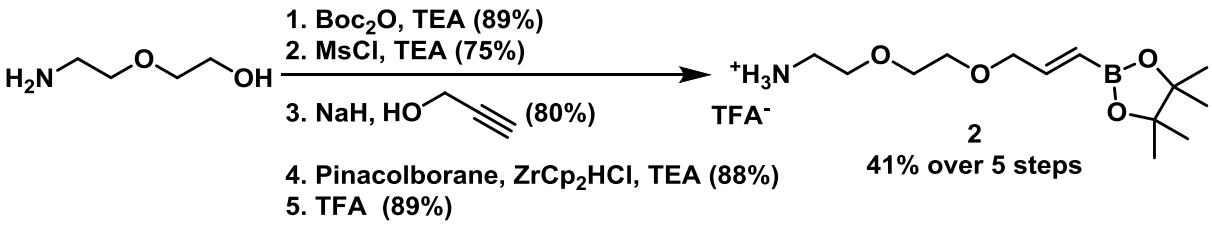
*3.4.2.3 (E)-3-(4,4,5,5-*tetra*-Methyl-1,3,2-dioxaborolan-2-yl)prop-2-en-1-amine trifluoroacetate salt (1)*



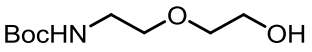
(E)-*tert*-Butyl 3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)allylcarbamate (250 mg, 0.88 mmol, 1 eq) was dissolved in 10 mL 9:1 CH_2Cl_2:CF_3COOH and stirred at rt for 1 h. Then solvents were removed *in vacuo* and remaining CF_3COOH was removed by co-evaporation (3x toluene, 3x CH_2Cl_2) to yield **1** (240 mg, 0.81 mmol, 92%) as a brown oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.23 (s, 12 H, ($C\underline{H}_3$)$_2$COB) 3.64 (br. s., 2 H, $C\underline{H}_2$) 5.73 (app d, *J*=18 Hz, 1 H, C$\underline{H}$B) 6.53 (dt, *J*=18, 6 Hz, 1 H, C$\underline{H}$CH$_2$) 8.21 (br. s., 3 H, N$\underline{H}$); ^{13}C NMR (101 MHz, CHLOROFORM-*d*) δ ppm 24.6 (($\underline{C}H_3$)$_2$COB) 42.8 ($\underline{C}H_2$) 83.8 ($\underline{C}$OB) 116.1 (*J*=293 Hz, $\underline{C}F_3$) 124.1 ($\underline{C}$HCHB) 142.1 ($\underline{C}$HB) 162.1 (*J*=36 Hz, $CF_3\underline{C}$OOH; LRMS m/z (ESI$^+$): 184.16 [M+H]$^+$ IR (neat): 2982, 1675, 1651, 1533 cm^{-1} HRMS *m/z* (ESI+): Found 184.1503 (M+H)$^+$; $C9H_{19}BNO_2$ requires 184.1503

3.4.3 Synthesis of the long-chain boronic ester amines for coupling



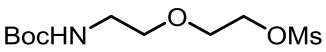
3.4.3.1 tert-*Butyl 2-(2-hydroxyethoxy)ethylcarbamate*[58,59]



The literature procedure reported by Chang et al.[58] was followed. To a solution of di-tert-butyl dicarbonate (21.83 g, 100 mmol, 1 eq) in CH_2Cl_2 (70 mL) on ice was added 2-(2-aminoethoxy)ethanol (12.0 mL, 12.6 g, 120 mmol, 1.2 eq) dropwise, subsequently NEt_3 (27.8 mL, 20.24 g, 200 mM, 2 eq). The resulting mixture was stirred at rt for 3 h. The reaction was quenched with sat. NH_4Cl, diluted with EtOAc (400 mL), washed with sat. NH_4Cl (2x200 mL) and brine (200 mL), dried ($MgSO_4$), filtered and solvents were removed *in vacuo*. The desired compound was obtained as a clear oil (18.2 g, 88.7 mmol, 89%) and was used in the next step without further purification.

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.42 (s, 9 H, Boc-H) 3.01 (s, 1 H, OH) 3.28 (d, *J*=5 Hz, 2 H, CH_2) 3.52 (dt, *J*=9.5, 5 Hz, 4 H, CH_2) 3.65 - 3.73 (m, 2 H, CH_2) 5.24 (s, 1 H, NH); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 28.4 (($CH_3)_3$)) 40.3 (CH_2) 61.5 (CH_2) 70.3 (CH_2) 72.2 (CH_2) 79.3 ($C(CH_3)_3$) 156.2 (CONH); LRMS m/z (ESI$^+$): 228.13 [M+Na]$^+$; IR (neat): 3343, 2977, 2361, 1691, 1525 cm^{-1}.

*3.4.3.2 2-(2-(*tert-*Butoxycarbonylamino)ethoxy)ethyl methanesulfonate*[60,61]

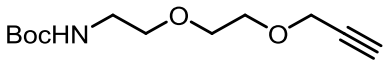


The literature procedure reported by Benanti *et al.*[60] was followed. To a solution of *tert*-butyl 2-(2-hydroxyethoxy)ethylcarbamate (17.8 g, 87 mmol, 1 eq) in 100 mL CH_2Cl_2 at 0 °C was added NEt_3 (26.7 mL, 19.4 g, 191 mmol, 2.2 eq) and then dropwise CH_3SO_2Cl (7.4 mL, 11 g, 96 mmol, 1.1 eq). The cloudy yellow solution was stirred on ice for 5 h and then quenched with sat. NH_4Cl. The reaction mixture was then diluted with EtOAc (400 mL), washed with sat. NH_4Cl (200 mL), $NaHCO_3$ (200 mL)

and brine (200 mL), dried ($MgSO_4$), filtered and solvents were removed *in vacuo*. The product was purified by column chromatography (Biotage KP-SIL SNAP 100 g cartridge 0-50% over 10CV EtOAc in petrol) to yield the desired compound (18.35 g, 65 mmol, 75%) as an orange oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.42 (s, 9 H, Boc-H) 3.04 (s, 3 H, Ms-H) 3.30 (app q, J=5 Hz, 2 H, CH_2) 3.54 (t, J=5 Hz, 2 H, CH_2) 3.68 - 3.75 (m, 2 H, CH_2) 4.30 - 4.38 (m, 2 H, CH_2) 4.92 (s, 1 H, NH); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 28.4 ($(CH_3)_3C$) 37.7 (CH_3S) 40.2, 68.6, 68.8, 70.3 (CH_2OS) 79.4 ($C(CH_3)_3$) 155.9 (OCCNH); LRMS m/z (ESI$^+$): 306.10 $[M+Na]^+$; IR (neat): 3402, 2978, 2936, 2874, 2361, 2342, 1701, 1512 cm^{-1}.

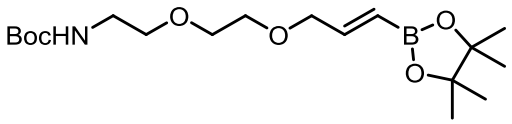
3.4.3.3 tert-*Butyl 2-(2-(prop-2-ynyloxy)ethoxy)ethylcarbamate*[58]



The literature procedure reported by Chang *et al*.[58] was followed. To a solution of propargylic alcohol (5.24 mL, 5.05 g, 90 mmol, 3 eq) in 100 mL THF on ice was added NaH (2.16 g, 90 mmol, 3 eq) portion wise. The resulting suspension was stirred on ice for 30 min then 2-(2-(*tert*-butoxycarbonylamino)ethoxy)ethyl methanesulfonate (8.5 g, 30 mmol, 1 eq) was added dropwise. The resulting suspension was stirred at rt for 19 h. The reaction was then quenched with sat. NH_4Cl (10 mL) and solvents were removed *in vacuo*. The resultant red liquid was partitioned between H_2O (50 mL) and EtOAc (350 mL). The organic phase was washed with sat. NH_4Cl (100 mL), sat. $NaHCO_3$ (100 mL) and brine (100 mL), dried ($MgSO_4$), filtered and solvents were removed *in vacuo*. The crude oil was purified by column chromatography (Biotage KP-SIL SNAP 100 g cartridge 0-50% over 10CV EtOAc in petrol) to yield the desired compound (5.81 g, 24 mmol, 80%) as a clear oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.42 (s, 9 H, Boc-H) 2.44 (t, J=2.5 Hz, 1 H, CCH) 3.30 (app q, J=5 Hz, 2 H, CH_2) 3.53 (t, J=5 Hz, 2 H, CH_2) 3.59 - 3.65 (m, 2 H, CH_2) 3.65 - 3.71 (m, 2 H, CH_2) 4.19 (d, J=2.5 Hz, 2 H, CH_2CCH) 4.99 (s, 1 H, NH); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 28.4 ($(CH_3)_3C$) 40.3 (CH_2), 58.4 (CH_2CC) 69.0, 70.1, 70.3 (CH_2) 74.7 (CCH) 79.2 (CCH) 79.5 ($C(CH_3)_3$) 156 (OCCNH); LRMS m/z (ESI$^+$): 266.14 $[M+Na]^+$; IR (neat): 3296, 2870, 1698, 1512 cm^{-1}.

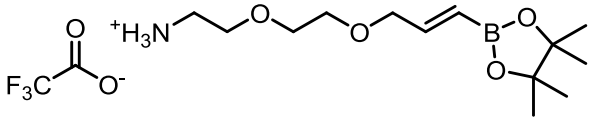
*3.4.3.4 (E)-*tert*-Butyl2-(2-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)allyloxy)ethoxy) ethylcarbamate[29]*



The literature procedure reported by Wang *et al*.[29] was followed. To neat *tert*-butyl 2-(2-(prop-2-ynyloxy)ethoxy)ethylcarbamate (4.86 g, 20 mmol, 1 eq) was added pinacolborane (3.5 mL, 3.07 g, 22 mmol, 1.1 eq), NEt_3 (270 uL, 202 mg, 2 mmol, 0.1 eq) and Cp_2ZrHCl (516 mg, 2 mmol, 0.1 eq). The solution was stirred at 65 °C for 17 h. After completion the reaction was allowed to cool to rt and was diluted with EtOAc (25 mL) and quenched with sat. NH_4Cl. The mixture was further diluted with EtOAc (150 mL) and the organic layer was washed with sat. NH_4Cl (50 mL), sat. $NaHCO_3$ (50 mL) and brine (50 mL), dried ($MgSO_4$), filtered and solvents were removed *in vacuo*. The product was purified by automated column chromatography (Biotag KP-SIL SNAP 100g cartridge 5-50% over 10CV EtOAc in petrol) to yield the desired compound as a yellow oil (6.5 g, 17.5 mmol, 88%).

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.24 (s, 12 H, pinacol-H) 1.42 (s, 9 H, Boc-H) 3.29 (s, 2 H, CH_2) 3.52 (s, 2 H, CH_2) 3.55 - 3.62 (m, 4 H, CH_2) 4.04 - 4.12 (m, 2 H, CH_2) 5.05 (s, 1 H, NH) 5.67 (d, *J*=18 Hz, 1 H, CHB) 6.60 (dd, *J*=18, 2.5 Hz, 1 H, $CHCH_2$); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 24.7 ($(CH_3)_3C$) 28.4 ($(CH_2)_2C$) 40.4, 69.7, 70.3, 72.8 (CH_2) 79.1 ($(CH_3)_3C$) 83.2 ($(CH_2)_2C$) 149.0 (CHB) 156.0 (OCONH); LRMS m/z (ESI$^+$): 394.25 $[M+Na]^+$ IR (neat): 3360, 2978, 2931, 2869, 1713, 1645, 1512 cm^{-1}; HRMS *m/z* (ESI+): Found 394.2367 $(M+Na)^+$; $C_{18}H_{34}BNO_3Na$ requires 394.2375.

3.4.3.5 (E)-2-(2-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)allyloxy)ethoxy)ethanamine (2)



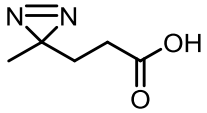
(E)-*tert*-Butyl-2-(2-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)allyloxy)ethoxy)ethylcarbamate (3.71 g, 10 mmol) was dissolved in 10% CF_3COOH in CH_2Cl_2 (50 mL total). The solution was then

stirred for 3 h. The solvents were removed *in vacuo* and CF_3COOH was subsequently co-evaporated with toluene (2x 30 mL) and CH_2Cl_2 (3x 30 mL) to yield **2** (3.41 g, 8.9 mmol, 89%) as a brown oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.26 (s, 12 H, pinacol-H) 3.22 (s, 2 H, CH_2) 3.61 (dd, *J*=6, 2.5 Hz, 2 H, CH_2) 3.67 (dd, *J*=6, 2.5 Hz, 2 H, CH_2) 3.70 - 3.79 (m, 2 H, CH_2) 4.10 (dd, *J*=5, 1.5 Hz, 2 H, CH_2) 5.61 - 5.73 (m, 1 H, CHB) 6.58 (dt, *J*=18, 5 Hz, 1 H, CHCH_2) 7.73 (s, 3 H, NH); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 24.6 ((CH_3)$_3$C) 39.9, 66.4, 69.4, 70.1, 72.6 (CH_2) 83.6 ((CH_2)$_2$C) 118.4 (CHCHB) 148.1 (CHB); LRMS m/z (ESI$^+$): 272.22 $[M+H]^+$; IR (neat): 2984, 1783, 1673, 1646, 1519 cm^{-1}; HRMS *m/z* (ESI+): Found 272.2031 $(M+H)^+$; $C_{13}H_{27}BNO_4$ requires 272.2030.

3.4.4 Amide couplings

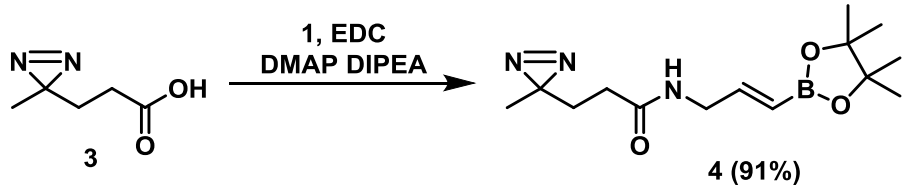
3.4.4.1 3-(3-Methyl-3H-diazirin-3-yl)propanoic acid (3)[30-33]



The literature procedure reported by Church *et al.*[30] was followed. To NH_3 (150 mL) was added levulinic acid (17.4 g, 150 mmol, 1 eq) at -78 °C. The resulting clear solution was stirred at -50 to -40 °C for 3.45 h. During this time a white precipitate was formed. The reaction mixture was then cooled to -78 °C again and hydroxylamine-O-sulfonic acid (20.36 g, 180 mmol, 1.2 eq) in MeOH (150 mL) was added dropwise over 20 min. The reaction mixture was stirred at -78 °C for 1.10 h and then warmed to rt overnight. Most of MeOH was removed *in vacuo* and the residue was taken up in MeOH (50 mL). NEt_3 was added and the resulting mixture was stirred at 0 °C for 15 min. Solvents were removed *in vacuo* and the residue was taken up in MeOH (150 mL) and cooled to 0 °C. NEt_3 (60 mL) was added and then I_2 was added in portions. The addition was continued until the colour of iodine persisted (total 33.2 g, 131 mmol). The reaction mixture was stirred at rt for 20 min. Subsequently, the solvents were removed and the residue was partitioned between Et_2O and water. The aqueous layer was acidified with 5M HCl and extracted with Et_2O (2x150 mL). The combined organic layers were washed with 20% $NaHSO_3$ and brine, dried ($MgSO_4$) filtered and solvents were removed *in vacuo* to yield sufficiently pure **3** as a yellow oil (solidifies at -20 °C; 7.82 g, 61 mmol, 40% yield).

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.02 (s, 3 H, C$\underline{H}_3$) 1.69 (t, *J*=7.5 Hz, 2 H, C$\underline{H}_2$) 2.21 (t, *J*=7.5 Hz, 2 H, C$\underline{H}_2$) 11.11 (s, 1 H, $CO_2\underline{H}$); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 19.6 (C$\underline{C}H_3$), 25.0 ($CH_3\underline{C}N_2CH_2$), 28.5, 29.3 ($\underline{C}H_2\underline{C}H_2$), 178.7 (COOH); LRMS m/z (ESI$^-$): 125.06 $[M-H]^-$; IR (neat): 2960, 1709, 1586 cm^{-1}.

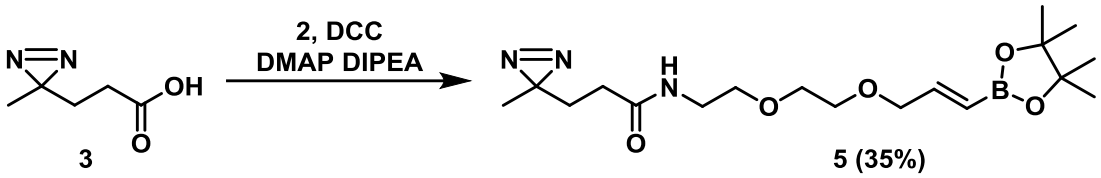
*3.4.4.2 (E)-3-(3-Methyl-3H-diazirin-3-yl)-*N*-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)allyl)propanamide (4)*



To a solution of *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC; 422 mg, 2.2 mmol, 2.2 eq) in CH_2Cl_2 (15 mL) on ice was added **3** (256 mg, 2 mmol, 2 eq). The resulting mixture was stirred on ice for 30 min and then 4-dimethylaminopyridine (DMAP; 49 mg, 0.4 mmol, 0.4 eq) and *N,N*-diisopropylethylamine (DIPEA; 435 µL, 323 mg, 2.5 mmol, 2.5 eq) and **1** (297 mg, 1 mmol, 1 eq) were added. The reaction mixture was stirred at rt overnight then quenched with NH_4Cl (sat., 20 mL) and diluted with EtOAc (80 mL). The organic phase was washed with sat. NH_4Cl (2x 30 mL), sat. $NaHCO_3$ (30 mL) and brine (30 mL), dried ($MgSO_4$), filtered and solvents were removed *in vacuo*. The residue was purified by automated column chromatography (Biotag KP-SIL SNAP 25g cartridge 0-10% over 10 CV MeOH in CH_2Cl_2) to yield **4** (266 mg, 0.91 mmol, 91%) as a brown oil.

^{1}H NMR (400 MHz, $CDCl_3$) d ppm 1.02 (s, 3 H, C$\underline{H}_3$CN$_2$) 1.25 (s, 12 H, CH_2) 1.69 - 1.82 (m, 2 H, C$\underline{H}_2$) 1.95 - 2.07 (m, 2 H, C$\underline{H}_2$) 3.87 - 4.00 (m, 2 H, C$\underline{H}_2$CH) 5.54 (dt, *J*=18, 2 Hz, 1 H, C$\underline{H}$B) 5.69 (s, 1 H, N$\underline{H}$) 6.55 (dt, *J*=18, 5 Hz, 1 H, C$\underline{H}$CHB); ^{13}C NMR (101 MHz, CDCl3) d ppm 19.9 (C$\underline{H}_3$) 24.7 (Pinacol-C$\underline{H}_3$), 25.6 (C$\underline{N}_2$) 30.5 33.9 42.9 (C$\underline{H}_2$) 83.4 ((CH$_3$)$_2$C) 148.2 (C$\underline{H}$B) 171.1 (C$\underline{O}$NH); LRMS m/z (ESI+): 316.19 $[M+Na]^+$; IR (neat): 3296, 3084, 2980, 2929, 2361, 2341, 1718, 1642, 1550 cm^{-1}; HRMS *m/z* (ESI+): Found 316.1806 $(M+Na)^+$; $C_{14}H_{24}BN_3NaO_3$ requires 316.1803.

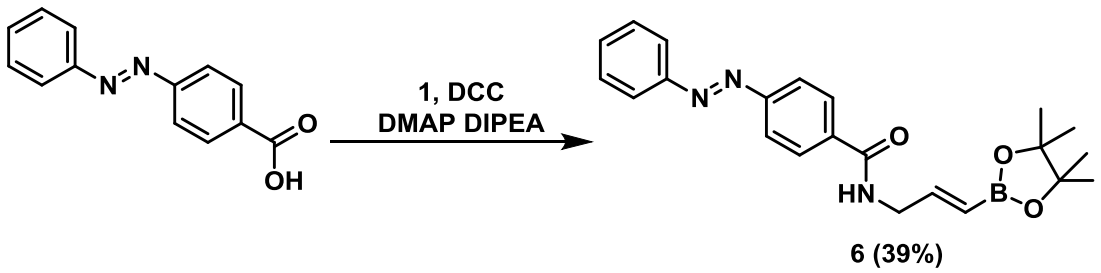
*3.4.4.3 (E)-3-(3-Methyl-3H-diazirin-3-yl)-*N*-(2-(2-(3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)allyloxy)ethoxy)ethyl)propanamide (5)*



To a solution of **3** (769 mg, 6 mmol, 1.5 eq) in CH_2Cl_2 (20 mL) on ice was added *N,N*'-dicyclohexylcarbodiimide (DCC; 1.32 g, 6.4 mmol, 1.6 eq). The resulting solution was stirred on ice for 30 min and then 4-dimethylaminopyridine (DMAP; 244 mg, 2 mmol, 0.5 eq) and *N,N*-diisopropylethylamine (DIPEA; 1.4 ml; 1.03 g, 8 mmol, 2 eq) and **2** (1.54 g, 4 mmol, 1 eq) were added. The reaction mixture was stirred at rt overnight and then filtered (wash with EtOAc or Et_2O) and diluted with EtOAc (100 mL). The organic phase was washed with sat. NH_4Cl (50 mL), sat. $NaHCO_3$ (50 mL) and brine (50 mL), dried ($MgSO_4$), filtered and solvents were removed *in vacuo*. The residue was purified by automated column chromatography (Biotag KP-SIL SNAP 25 g cartridge 0-10% over 10 CV MeOH in CH_2Cl_2) to yield **5** (738 mg, 2.1 mmol, 35%) as a brown oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.02 (s, 3 H, C$\underline{H}_3$CN$_2$) 1.26 (s, 12 H, Pinacol-C$\underline{H}_3$) 1.73 (dd, *J*=9, 7 Hz, 2 H, CN$_2$C$\underline{H}_2$C$\underline{H}_2$) 2.01 (dd, *J*=9, 7 Hz, 2 H, CN$_2$C$\underline{H}_2$C$\underline{H}_2$) 3.44 (q, *J*=5 Hz, 2 H, C$\underline{H}_2$) 3.54 (s, 2 H, C$\underline{H}_2$) 3.58 - 3.62 (m, 4 H, C$\underline{H}_2$) 4.10 (dd, *J*=5, 2 Hz, 2 H, CH$_2$) 5.71 (dt, *J*=18, 2 Hz, 1 H, C$\underline{H}$B) 6.20 (br. s., 1 H, N$\underline{H}$) 6.62 (dt, *J*=18, 5 Hz, 1 H, C$\underline{H}$CHB); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 19.8 (C$\underline{H}_3$) 24.7 ((C$\underline{H}_3$)$_2$C) 25.4 (CH$_3$C$\underline{N}$N) 30.1 (C$\underline{H}_2$) 30.6 (C$\underline{H}_2$) 39.3, 69.7, 69.8, 70.3 (C$\underline{H}_2$) 72.7 (OC$\underline{H}_2$CH), 83.3 ((CH$_3$)$_2$C) 148.8 (C$\underline{H}$B) 171.4 (CH$_2$C$\underline{O}$NH); LRMS *m/z* (ESI$^+$): 404.25 [M+Na]$^+$; IR (neat): 3308, 2978, 2928, 2869, 1726, 1646, 1551 cm^{-1}; HRMS *m/z* (ESI+): Found 404.2318 (M+Na)$^+$; $C_{18}H_{32}BN_3O_5Na$ requires 404.2330.

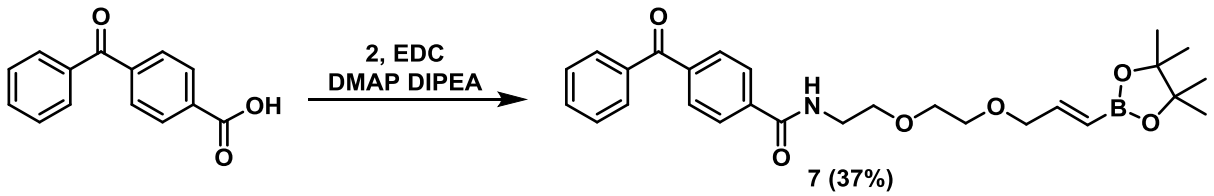
*3.4.4.4 4-((E)-Phenyldiazenyl)-*N*-((E)-3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)allyl)benzamide (6)*



To a solution of 4-(phenyldiazenyl)benzoic acid (226 mg, 1 mmol, 2 eq) in dry tetrahydrofuran (10 mL) was added *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC; 211 mg, 1.1 mmol, 2.2 eq) on ice under N_2. The resulting solution was stirred for 30 min on ice and then **1** (149 mg, 0,5 mmol, 1 eq), *N,N*-diisopropylethylamine (DIPEA; 325 uL, 1.75 mmol, 3.5 eq) and 4-dimethylaminopyridine (DMAP; 25 mg, 0.2 mmol, 0.4 eq) were added. The reaction mixture was allowed to warm to rt and stirred for 16 h. Solvents were removed *in vacuo* and the red oil was taken up in EtOAc (150 mL). The organic layer was washed with sat. NH_4Cl (2 x 50 mL), sat. $NaHCO_3$ (50 mL) and brine (50 mL); dried ($MgSO_4$), filtered and solvents were removed *in vacuo*. The red residue was purified by automated column chromatography (Biotage KP-SIL SNAP 10g; MeOH in CH_2Cl_2; 0% 5CV; 0-5% 10CV) to yield **6** a deep red solid (*trans:cis* ; 7:1; 76 mg, 0.2 mmol, 39%).

Values for the minor cis isomer are given in italic. ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.20 - 1.32 (m, (12+12x0.15)=14 H, C<u>H</u>$_3$) *4.10 - 4.17 (m, (2x0.15) H, C<u>H</u>$_2$)* 4.18 - 4.25 (m, 2 H, C<u>H</u>$_2$) *5.61 (dt, J=18, 2 Hz, 1x0.15 H, C<u>H</u>)* 5.68 (dt, *J*=18, 2 Hz, 1 H, C<u>H</u>) *6.35 (t, J=6 Hz, 1x0.15 H, N<u>H</u>)* 6.48 (t, *J*=6 Hz, 1 H, N<u>H</u>) *6.62-6.67 (m, 1x0.15 H, C<u>H</u>)* 6.65 - 6.76 (m, 1 H, C<u>H</u>) *6.81 - 6.89 (m, 4x0.15 H, Ar<u>H</u>) 7.13 - 7.19 (m, 1x0.15 H, Ar<u>H</u>) 7.22 - 7.29 (m, 2x0.15 H, Ar<u>H</u>)* 7.45 - 7.59 (m, 3 H, Ar<u>H</u>) *7.68 (d, J=9 Hz, 2x0.15 H, Ar<u>H</u>)* 7.88 - 8.00 (m, 6 H, Ar<u>H</u>); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 24.7 (<u>C</u>H$_3$) 43.4 (<u>C</u>H$_2$) 83.4 (<u>C</u>CH$_3$) 119.1 (<u>C</u>H) 120.2 (Ar-<u>C</u>H (minor)) 120.6 (Ar-<u>C</u>H (minor)) 122.9 (Ar-<u>C</u>H (major)) 123.0 (Ar-<u>C</u>H (major)) 127.7 (Ar-<u>C</u>H (minor)) 127.9 (Ar-<u>C</u>H (major)) 128.8 (Ar-<u>C</u>H (minor)) 129.1 (Ar-<u>C</u>H (major)) 131.5 (Ar-<u>C</u>H (major)) 136 (Ar-<u>C</u>) 148.1 (<u>C</u>H) 152.4 (Ar-<u>C</u>) 154.2 (Ar-<u>C</u>) 166.6 (<u>C</u>ONH); LRMS m/z (ESI$^+$): 414.18 $[M+H]^+$; HRMS *m/z* (ESI+): Found 414.1960 $(M+Na)^+$; $C_{22}H_{26}BN_3ONa$ requires 414.1963; IR (neat): 3316, 2928, 2360, 2341, 1640, 1541 cm^{-1}.

*3.4.4.5 (E)-4-Benzoyl-*N*-(2-(2-((3-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)allyl)oxy)ethoxy)ethyl)benzamide (7)*

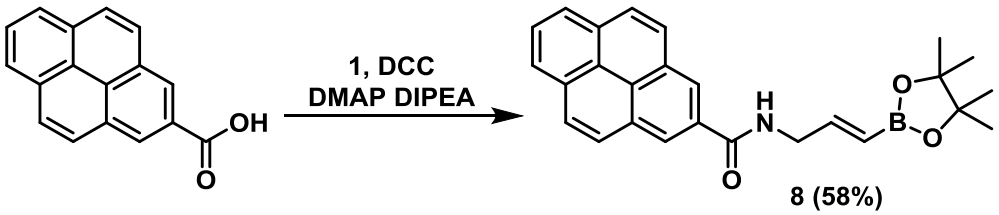


7 (37%)

To a solution of *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC; 211 mg, 1.1 mmol, 2.2 eq) in CH_2Cl_2 (10 mL) on ice was added 4-benzoylbenzoic acid (226 mg, 1 mmol, 2 eq). The resulting solution was stirred on ice for 30 min and then 4-dimethylaminopyridine (DMAP; 31 mg, 0.25 mmol, 0.5 eq) and *N,N*-diisopropylethylamine (DIPEA; 304 µL, 226 mg, 1.75 mmol, 3.5 eq) and **2** (192 mg, 0.5 mmol, 1 eq) were added. The reaction mixture was stirred at rt overnight then quenched with NH_4Cl (sat., 20 mL) and diluted with EtOAc (80 mL). The organic phase was washed with sat. NH_4Cl (2x30 mL), sat. $NaHCO_3$ (30 mL) and brine (30 mL), dried ($MgSO_4$), filtered and solvents were removed *in vacuo*. The residue was purified by automated column chromatography (Biotag KP-SIL SNAP 25g cartridge 0-5% over 10 CV MeOH in CH_2Cl_2) to yield **7** (86 mg, 0.18 mmol, 36%).

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.22 (s, 12 H, C$\underline{H}_3$) 3.58 - 3.74 (m, 8 H, C$\underline{H}_2$) 4.10 (dd, *J*=5, 2 Hz, 2 H, C$\underline{H}_2$CH) 5.72 (dt, *J*=18, 2 Hz, 1 H, C$\underline{H}$B) 6.62 (dt, *J*=18, 5 Hz, 1 H, C$\underline{H}$CHCB) 6.92 (br. s., 1 H, N$\underline{H}$) 7.45 - 7.53 (m, 2 H, C$\underline{H}$CHCCO) 7.56 - 7.64 (m, 1 H, C$\underline{H}$CHCH) 7.76 - 7.81 (m, 2 H, C$\underline{H}$CCO) 7.83 (m, *J*=8 Hz, 2 H, C$\underline{H}$CHCCONH) 7.90 (m, *J*=8 Hz, 2 H, C$\underline{H}$CCONH); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 24.7 (C$\underline{H}_3$) 39.8 (C$\underline{H}_2$CH) 69.7 (C$\underline{H}_2$) 69.7 (C$\underline{H}_2$) 70.3 (C$\underline{H}_2$) 72.7 (C$\underline{H}_2$) 83.3 (C$\underline{C}$H$_3$) 119.7 (C$\underline{H}$CHB) 127.0 (C$\underline{H}$CCONH) 128.4 (C$\underline{H}$CHCCO) 130.1 (C$\underline{H}$CHCCONH and C$\underline{H}$CCO) 132.8 (C$\underline{H}$CHCH) 137.1 (qAr-C) 137.8 (qAr-C) 139.9 (qAr-C) 148.8 (C$\underline{H}$B) 166.6 (C$\underline{O}$NH) 196.0 (C$\underline{O}$); LRMS m/z (ESI$^+$): 480.26 $[M+H]^+$; HRMS *m/z* (ESI+): Found 502.2359 $(M+Na)^+$; $C_{27}H_{34}BNO_6Na$ requires 502.2376; IR (neat): 2978, 2361, 1645, 1538 cm^{-1}.

3.4.4.6 (E)-N-*(3-(4,4,5,5*-tetra-*Methyl-1,3,2-dioxaborolan-2-yl)allyl)pyrene-2-carboxamide (8)*



8 (58%)

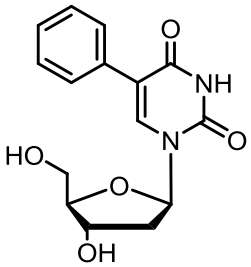
To a solution of *N*-(3-dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC, 211 mg, 1.1 mmol, 2.2 eq) in CH_2Cl_2 (10 mL) on ice was added pyrene-2-carboxylic acid (246 mg, 1 mmol, 2 eq).

The resulting solution was stirred on ice for 30 min and then 4-dimethylaminopyridine (DMAP; 25 mg, 0.2 mmol, 0.4 eq) and *N,N*-diisopropylethylamine (DIPEA; 304 µL, 226 mg, 1.75 mmol, 3.5 eq) and **1** (150 mg, 0.5 mmol, 1 eq) were added. The reaction mixture was stirred at rt overnight, then quenched with NH_4Cl (sat., 20 mL) and diluted with EtOAc (80 mL). The organic phase was washed with sat. NH_4Cl (2x30 mL), sat. $NaHCO_3$ (30 mL) and brine (30 mL), dried ($MgSO_4$), filtered and solvents were removed *in vacuo*. The residue was purified by automated column chromatography (Biotag KP-SIL SNAP 25g cartridge 0-5% over 10 CV MeOH in CH_2Cl_2) to yield **8** (120 mg, 0.29 mmol, 58%) as a yellow foam.

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.29 (s, 12 H, CH_3) 4.24 - 4.34 (m, 2 H, CH_2) 5.76 (d, *J*=18 Hz, 1 H, C<u>H</u>B) 6.53 (t, *J*=6 Hz, 1 H, N<u>H</u>) 6.78 (dt, *J*=18, 5 Hz, 1 H, CH_2C<u>H</u>) 7.82 - 8.01 (m, 6 H, Ar-<u>H</u>) 8.11 (t, *J*=7 Hz, 2 H, Ar-<u>H</u>) 8.44 (d, *J*=9 Hz, 1 H, Ar-<u>H</u>); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 24.7 (<u>C</u>H_3) 43.5 (<u>C</u>H_2) 83.3 (<u>C</u>CH_3) 119.2 (<u>C</u>HB) 124.0 (Ar-<u>C</u>) 124.1 (Ar-<u>C</u>) 124.2 (Ar-<u>C</u>) 124.4 (Ar-<u>C</u>) 125.5 (Ar-<u>C</u>) 125.6 (Ar-<u>C</u>) 126.1 (Ar-<u>C</u>) 126.9 (Ar-<u>C</u>) 128.3 (Ar-<u>C</u>) 128.4 (Ar-<u>C</u>) 130.4 (Ar-<u>C</u>) 130.5 (Ar-<u>C</u>) 130.9 (Ar-<u>C</u>) 132.3 (Ar-<u>C</u>) 148.4 (CH_2<u>C</u>H) 169.7 (<u>C</u>ONH); LRMS m/z (ESI$^+$): 434.20 [M+Na]$^+$; HRMS *m/z* (ESI+): Found 434.1888 (M+Na)$^+$; $C_{26}H_{26}BNNaO_3$ requires 434.1903; IR (neat): 2977, 1637, 1534 cm^{-1}.

3.5 Small molecule cross-couplings

3.5.1.1 5-Phenyl-2'-deoxyuridine (9)[34]

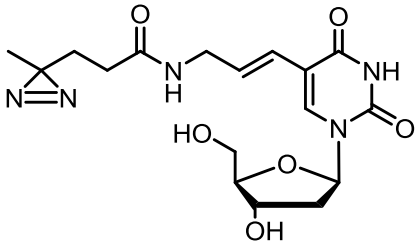


5-Iodo-2'-deoxyuridine (53 mg, 0.15 mmol, 1 eq), phenylboronic acid (28 mg, 0.23 mmol, 1.5 eq) and K_2CO_3 were assembled in a 15 mL vial. 2 mL ddH_2O, 0.5 mL MeCN and Pd(OAC)$_2$**L1$_2$** (150 µL of a 0.05 M stock, 15 µmol, 10 mol%) were added. The vial was flushed with argon, closed and heated to 50 °C for 5 h. The reaction mixture was allowed to cool down to rt, acidified with 2 M HCl (2 drops) and

extracted with CH_2Cl_2 (3x10 mL). The combined organic phases were washed with brine, dried over $MgSO_4$ and the solvent was removed *in vacuo*. The crude product was purified by column chromatography (CH_2Cl_2 to 9:1 CH_2Cl_2:MeOH) to yield **9** as colorless oil (36 mg, 0.12 mmol, 80%).

^{1}H NMR (400 MHz, CD_3OD) δ ppm 2.97 (ddd, *J*=13, 6, 4 Hz, 1 H, C$^{2'}$H) 3.06 (dt, *J*=13, 7 Hz, 1 H, C$^{2'}$H) 4.33 - 4.47 (m, 2 H, C$^{5'}$H$_2$) 4.63 (q, *J*=3 Hz, 1 H, C$^{4'}$H) 5.06 - 5.13 (m, 1 H, C$^{3'}$H) 5.94 (br. s., 1 H, 5'-OH) 6.08 (br. s., 1 H, 3'-OH) 7.05 (t, *J*=7 Hz, 1 H, C$^{1'}$H) 8.07 - 8.14 (m, 1 H, Ph-H) 8.14 - 8.21 (m, 2 H, Ph-H) 8.32 - 8.40 (m, 2 H, Ph-H) 9.01 (s, 1 H, C^{6}H); ^{13}C NMR (101 MHz, CD_3OD) δ ppm 49.7 (C$^{1'}$H) 70.6 (C$^{5'}$H) 79.9 (C$^{3'}$H) 94.1 (C$^{1'}$H) 97.2 (C$^{4'}$H) 123.1 (Ar-C) 136.8 (Ph-CH) 137.5 (Ph-CH) 137.8 (Ph-CH) 142.8 (Ar-C) 147.7 (C^{6}H) 159.6 (Ar-C) 171.7 (Ar-C) cm^{-1}; LRMS m/z (ESI+): 327.12 $[M+Na]^+$, 343.10 $[M+K]^+$; HRMS *m/z* (ESI+): Found 327.0953 $(M+Na)^+$; $C_{15}H_{16}N_2NaO_5$ requires 327.0951.

3.5.1.2 N-((E)-3-(1-((2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)allyl)-3-(3-methyl-3H-diazirin-3-yl)propanamide (10)

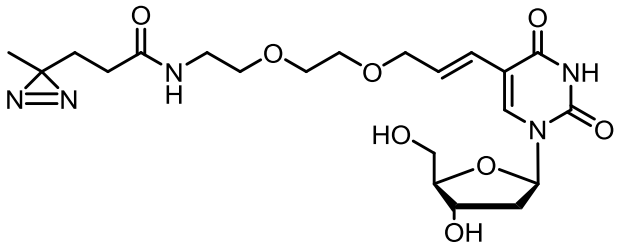


5-Iodo-2'-deoxyuridine (28.3 mg, 0.08 mmol, 1 eq) and K_2CO_3 (33.2 mg, 0.24 mmol, 3 eq) were dissolved in 4 mL 3:1 H_2O:MeCN in a 10 mL vial. To the stirring solution was added $Pd(OAC)_2$**L1$_2$** (80 µL of a 0.05 M stock, 4 µmol, 5 mol%) and subsequent **3** (35.2 mg, 0.12 mmol, 1.5 eq). The reaction mixture was stirred at 50 °C for 4 h. The solvents were removed *in vacuo* and the residue was dissolved in MeOH (30 mL), filtered and purified by automated column chromatography (Biotag SNAP KP-SIL 10 g cartridge 0-20% over 20 CV MeOH in CH_2Cl_2) to yield an inseparable mix of **5c** and starting material (30 mg, 76 µmol, 95% recovery, 60% **3** as determined by NMR) as a yellow oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 1.01 (s, 3 H, CH_3CN_2CH_2) 1.64 - 1.70 (m, 2 H, CN_2CH_2CH_2) 2.07 - 2.15 (m, 2 H, CN_2CH_2CH_2) 2.18 - 2.34 (m, 4 H, C$^{2'}$H$_2$) 3.35 (s, 1 H) 3.74 (ddd, *J*=12, 7, 3 Hz, 2 H, C$^{5'}$H$_2$) 3.79 -

3.85 (m, 2 H, C$^{5'}$H$_2$) 3.87 (d, *J*=6 Hz, 2 H, C$^{4'}$H) 3.93 (q, *J*=3 Hz, 2 H, CH$_2$CHCH) 4.38 - 4.44 (m, 2 H, C$^{3'}$H) 4.87 (s, 12 H) 6.20 - 6.31 (m, 3 H, C$^{1'}$H, CH$_2$CHCH) 6.39 - 6.52 (m, 1 H, CH$_2$CHCH) 8.15 (s, 1 H, C^{6}H) 8.53 (s, 1 H, C^{6}H); ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 19.9 (CH_3CN_2) 31.4 ($CN_2CH_2CH_2$) 31.7 ($CN_2CH_2CH_2$) 41.8 (C2') 41.9 (C2') 42.9 (CH_2CHCH) 62.6 (C5') 62.8 (C5') 72.1 (C4') 72.2 (C4') 86.8 (C1') 87.1 (C1') 89.2 (C3') 89.2 (C3') 112.8 (C5) 124.0 (CH_2CHCH) 127.8 (CH_2CHCH) 139.0 (C6) 147.4 (C6) 151.7 (CO) 174.5 (CO); LRMS m/z (ESI$^+$): 432.15 $[M+K]^+$ IR (neat): 3727, 3325, 3068, 2928, 2261, 2341, 1680, 1638, 1548 cm^{-1}; HRMS *m/z* (ESI+): Found 416.1545 $(M+Na)^+$; $C_{17}H_{23}N_5NaO_6$ requires 416.1541

3.5.1.3 N-(2-(2-((E)-3-(1-((2R,4S,5R)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)allyloxy)ethoxy)ethyl)-3-(3-methyl-3H-diazirin-3-yl)propanamide (11)

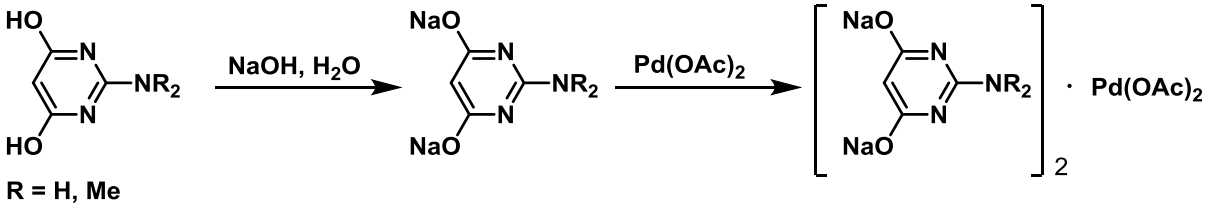


5-Iodo-2'-deoxyuridine (38 mg, 0.1 mmol, 1 eq) and K_2CO_3 (41 mg, 0.3 mmol, 3 eq) were dissolved in 4 mL 3:1 H_2O:MeCN in a 10 mL vial. To the stirred solution was added $Pd(OAC)_2$**L1$_2$** (50 µL of a 0.05 M stock, 1 µmol, 10 mol%) and subsequent **4** (57 mg, 0.15 mmol, 1.5 eq). The reaction mixture was stirred at 50 °C for 2 h after which TLC indicated complete conversion. The solvents were removed *in vacuo* and the residue was dissolved in MeOH (30 mL), filtered and purified by automated column chromatography (Biotag KP-SIL SNAP 10 g cartridge 0-20% over 20 CV MeOH in CH_2Cl_2) to yield **2** (41 mg, 85 µmol, 85%) as a yellow oil.

^{1}H NMR (400 MHz, CD_3OD) δ ppm 1.00 (s, 3 H, $CH_3CN_2CH_2$) 1.61 - 1.68 (m, 2 H, $CN_2CH_2CH_2$) 2.10 (dd, *J*=8, 7 Hz, 2 H, $CN_2CH_2CH_2$) 2.22 - 2.36 (m, 2 H, C$^{2'}$H$_2$) 3.37 (t, *J*=5 Hz, 2 H, NCH$_2$CH$_2$) 3.54 - 3.58 (m, 2 H, NCH$_2$CH$_2$) 3.63 (s, 4 H, OCH$_2$CH$_2$O) 3.77 (dd, *J*=12, 3 Hz, 1 H, C$^{5'}$H$_2$) 3.85 (dd, *J*=12, 3 Hz, 1 H, C$^{5'}$H$_2$) 3.95 (q, *J*=3 Hz, 1 H, C$^{4'}$H) 4.12 (dd, *J*=6, 1 Hz, 2 H, CH$_2$CHCH) 4.44 (dt, *J*=6, 4 Hz, 1 H, C$^{3'}$H) 6.30 (t, *J*=6 Hz, 1 H, C$^{1'}$H) 6.38 (d, *J*=16 Hz, 1 H, CH$_2$CHCH) 6.57 (dt, *J*=16, 6 Hz, 1 H, CH$_2$CH) 8.22 (s, 1 H, C^{6}H); ^{13}C NMR

(101 MHz, CD_3OD) δ ppm 19.9 ($\underline{C}H_3CN_2$) 26.5 ($\underline{C}N_2$) 31.4 ($CN_2\underline{C}H_2CH_2$) 31.7 ($CN_2\underline{C}H_2CH_2$) 40.6 ($N\underline{C}H_2CH_2O$) 41.8 (C2') 62.8 (C5') 70.7 ($NCH_2\underline{C}H_2O$) 71.4 ($O\underline{C}H_2CH_2O$) 72.1 (C4') 73.2 ($\underline{C}H_2CHCH$) 86.9 (C1') 89.2 (C3') 112.6 (C5) 125.1 ($CH_2CH\underline{C}H$) 127.9 ($CH_2\underline{C}HCH$) 139.4 (C6) 151.7 (CO) 164.5 (CO) 174.8 (CO); LRMS m/z (ESI^+): 504.24 $[M+Na]^+$ IR (neat): 3400, 3062, 2926, 2871, 2493, 2429, 2361, 2342, 1683, 1644 cm^{-1}; HRMS *m/z* (ESI+): Found 504.2067 $(M+Na)^+$; $C_{21}H_{31}N_5NaO_8$ requires 504.2065.

3.6 Catalyst preparation



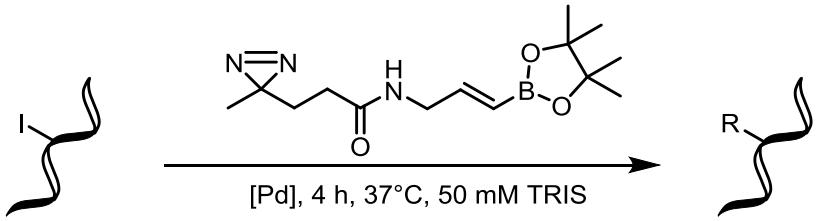
To 2-aminopyrimidine-4,6-diol (65 mg, 0.5 mmol, 2 eq, R = H) or 2-(dimethylamino)pyrimidine-4,6-diol (78 mg, 0.5 mmol, 2 eq, R = Me) in a 5 mL volumetric flask was added H_2O (3 mL) and NaOH (10 M stock, 100 µL, 1 mmol, 4 eq). The solution was stirred for ~5 min at rt until all solids were dissolved. Then $Pd(OAc)_2$ (55 mg, 0.25 mmol, 1 eq) was added and the flask was stoppered and stirred at 65 °C for 60 min. The stirring bar was removed and the solution was adjusted to 5 mL with H_2O to give a 50 mM stock solution.

3.7 Suzuki-Miyaura cross-coupling to DNA

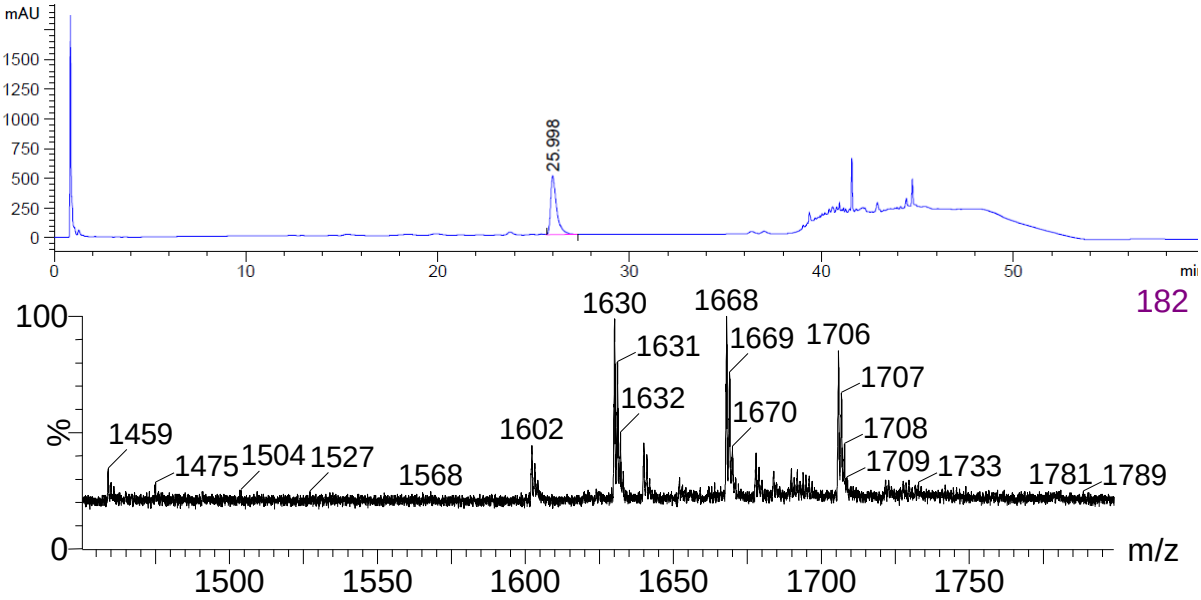
3.7.1 General reaction conditions

Unless otherwise noted, 100 µM ODN1-4, 50 mM Tris pH 8.5, 10-50 µM $Pd(OAc)_2\mathbf{LX}_2$ and 10 mM boronic ester (added as a stock solution in MeCN) were combined in an Eppendorf tube and shaken for 4h at 37 °C in an incubator. Then the reaction mixture was spun down (2 min, 16000 x g) and the supernatant was analyzed by HPLC and MS. Sometimes precipitation is observed after 24 h in the catalyst stock solution. In these cases, probably due to lower actual concentration, the reaction proceeds better with 1 eq (100 µM) $Pd(OAc)_2\mathbf{LX}_2$.

3.7.2 ODN1a

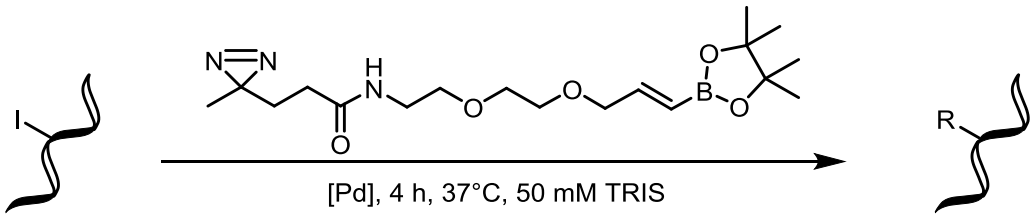


Product elutes after 26 min. Gradient is 5%-20% B in 30 min.

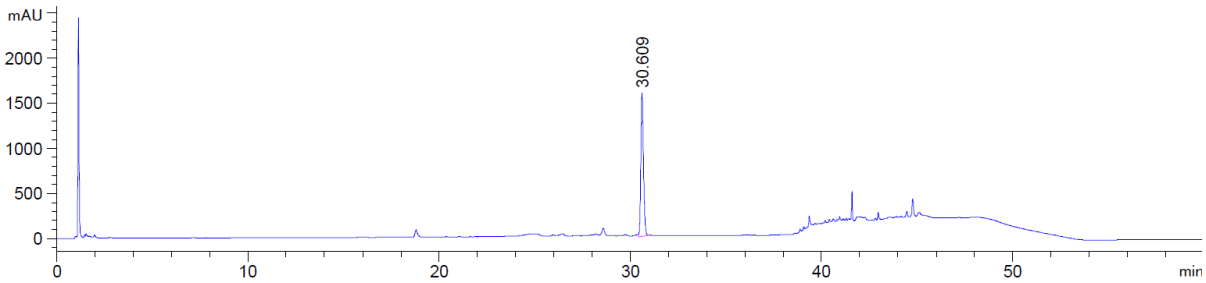


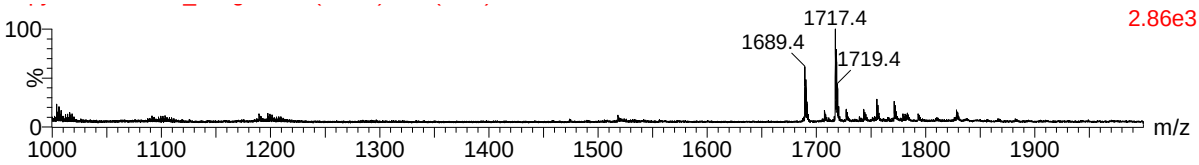
Calc 1630.2 found 1630.3

3.7.3 ODN1b



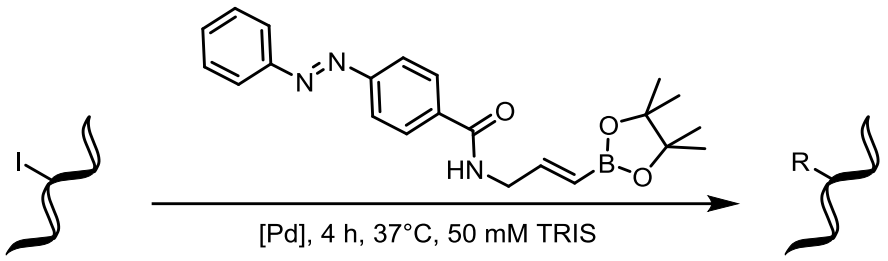
Product elutes at 30.6 min. Gradient is 5%-20% B in 35 min.



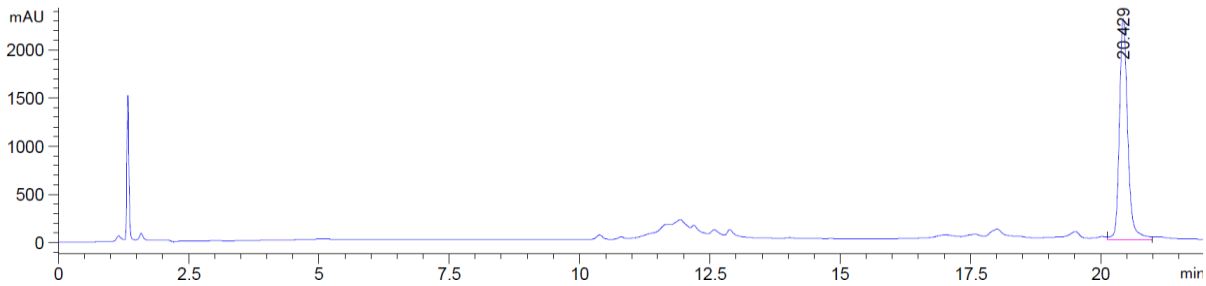


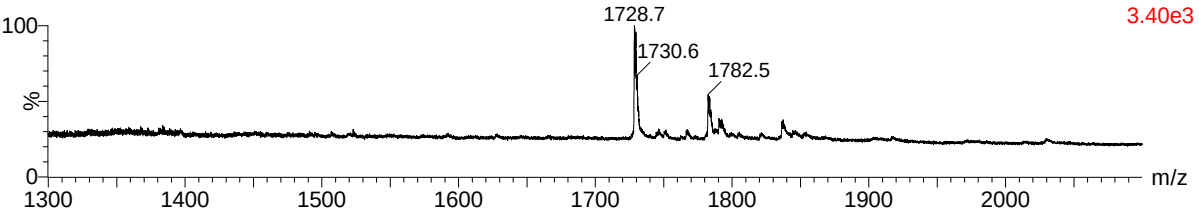
Calc 1718.2 found 1717.4

3.7.4 ODN1c



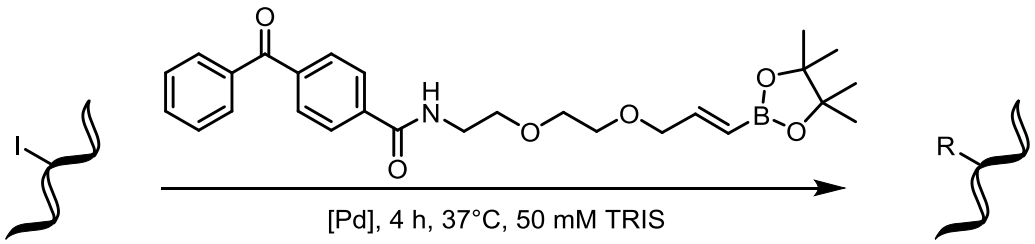
Product elutes at 20.4 min. Gradient is 5%-20% B in 20 min.



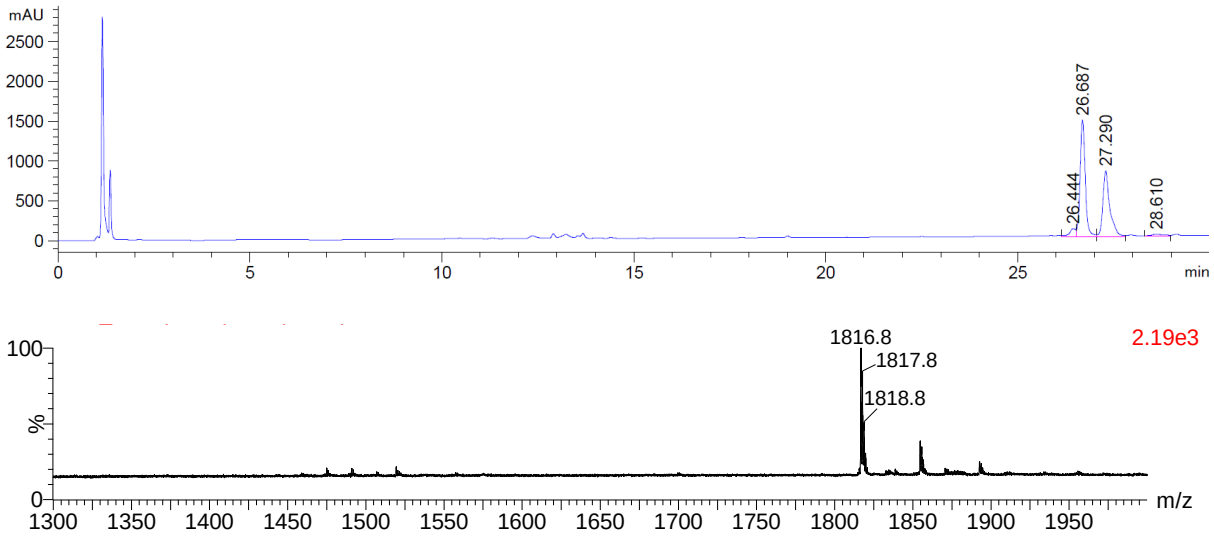


Calc 1728.2 found 1728.7

3.7.5 ODN1d

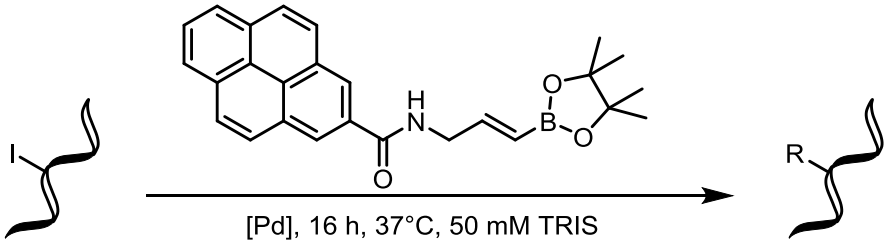


ODN1a elutes at 26.69 min. The peak at 27.29 is most likely a small molecule impurity, as no ODN was detected upon MALDI-MS analysis of this peak. Gradient is 5%-20% B in 30 min.

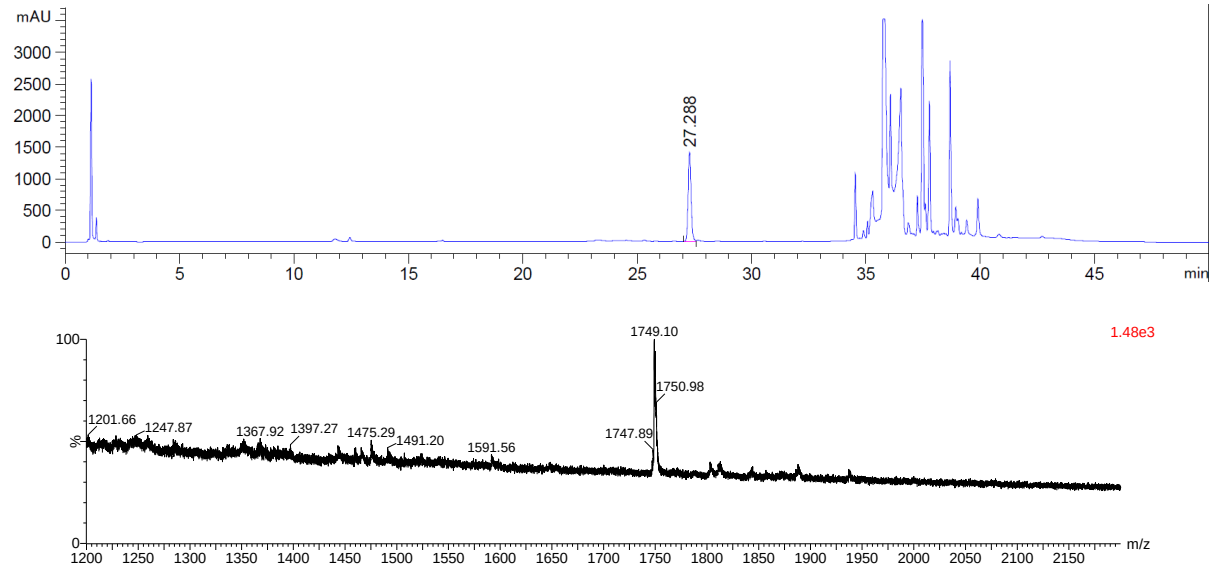


Calc 1816.3 found 1816.8

3.7.6 ODN1e

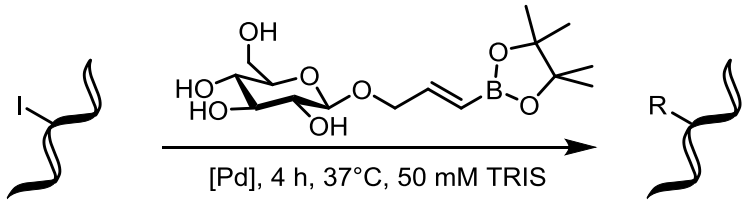


The reaction was performed in 40% MeCN. Product elutes at 27.3 min. Gradient is 5%-30% B in 30 min.

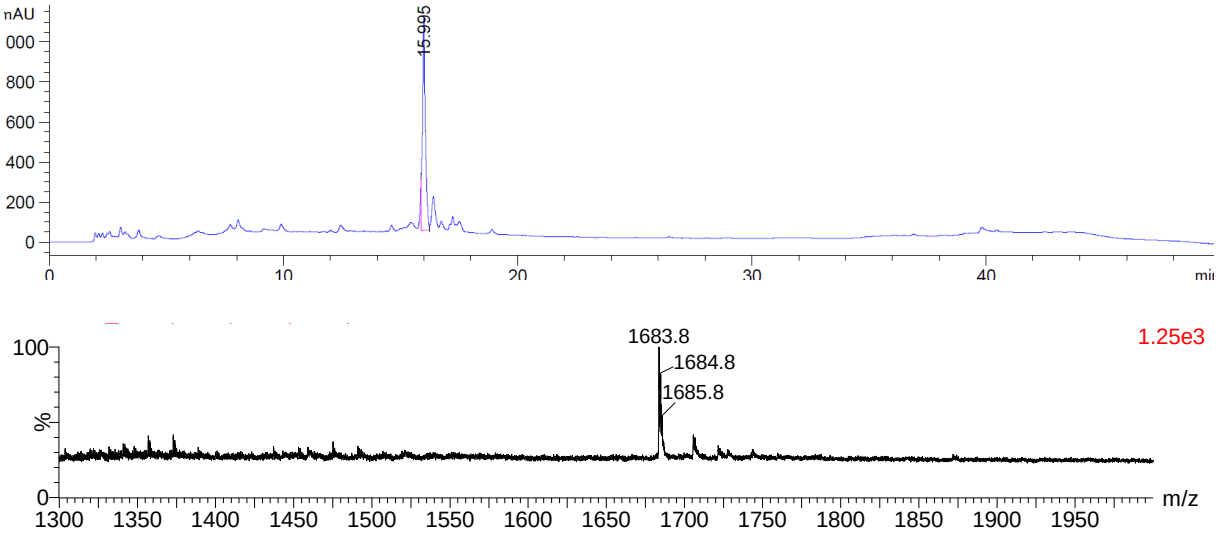


Calc 1748.2 found 1749.1

3.7.7 ODN1f

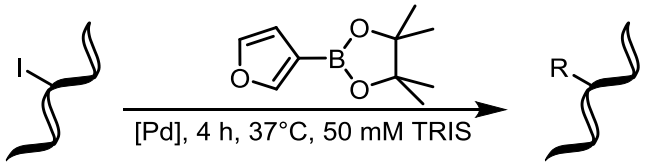


Product elutes at 16 min. Gradient is 0%-30% B in 30 min.

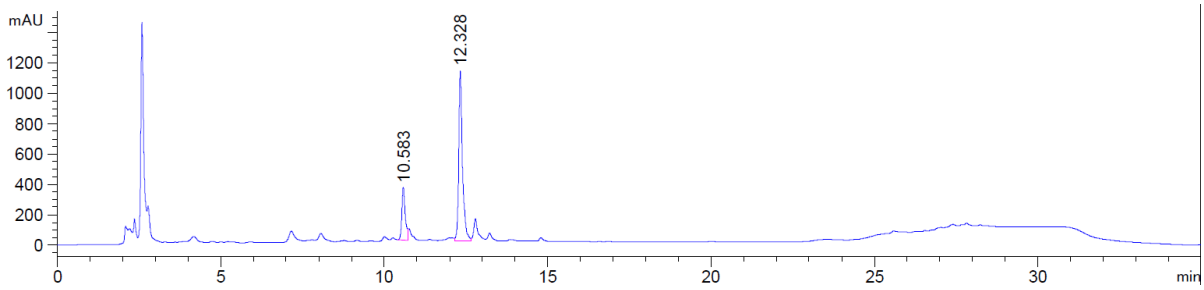


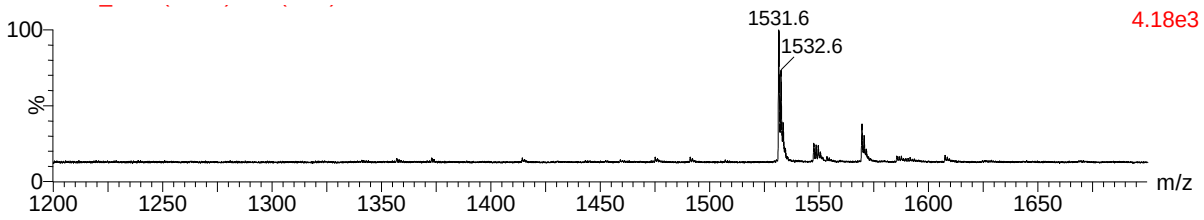
Calc 1683.2 found 1683.8

3.7.8 ODN1g



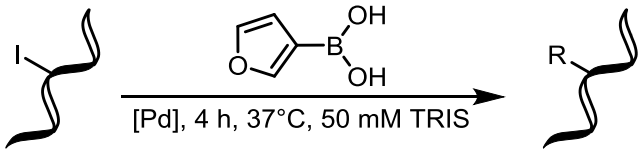
Product elutes at 12.3 min. Gradient is 5%-40% B in 20 min.



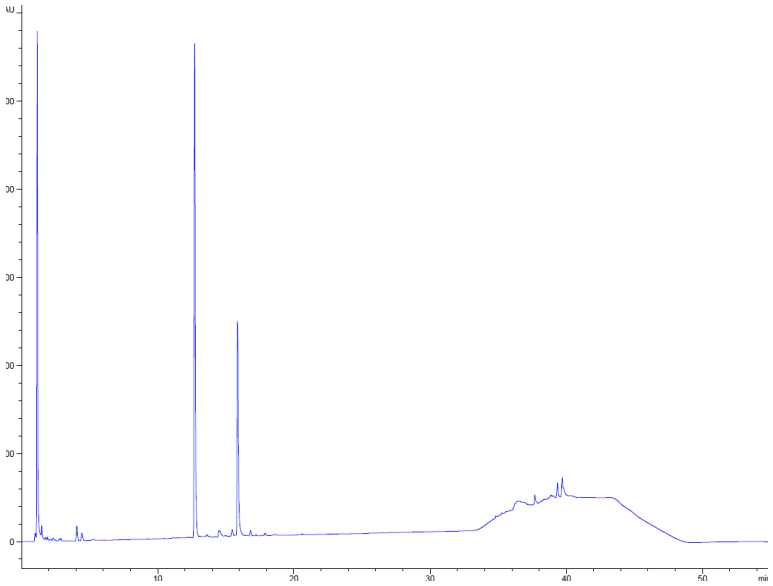


Calc. 1532.2 found 1531.6

3.7.9 ODN1g reaction with 3-furyl boronic acid

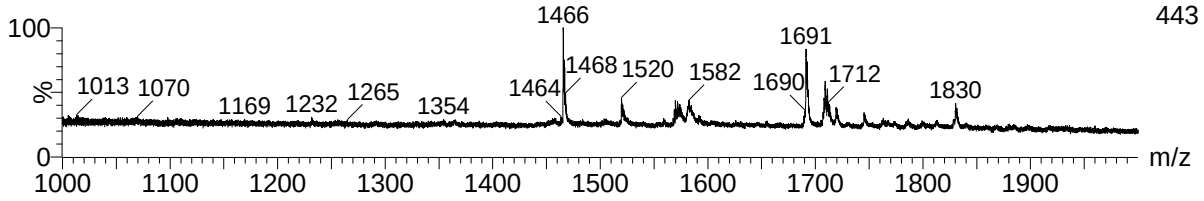


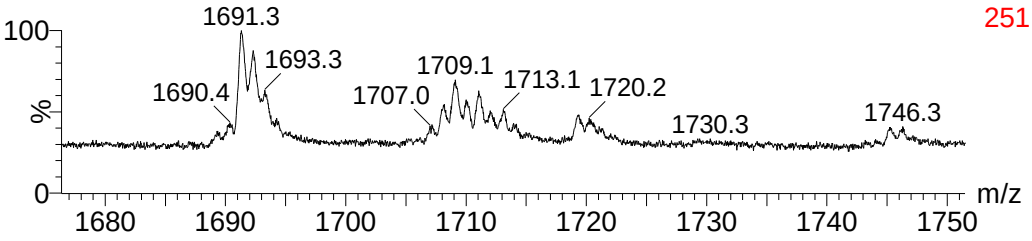
Product elutes at 15.8 min, the product of deiodination at 12.6 min. Gradient is 5%-20% B in 20 min.



3.7.10 ODN1 palladium adduct observed in PO4 buffer

When performing the cross-coupling of **4** to **ODN1** in phosphate buffer the same species can be observed in the MALDI of the crude reaction mixture.

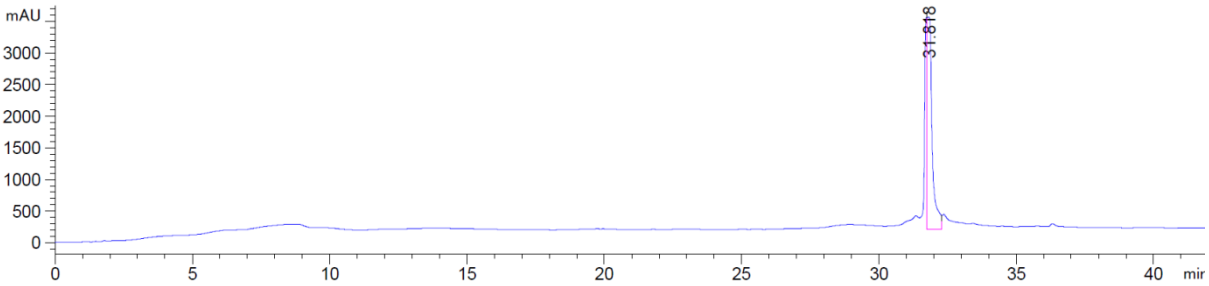




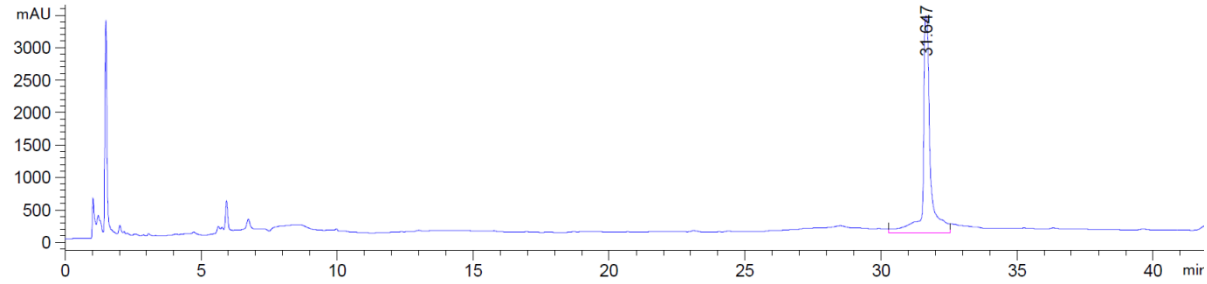
3.7.11 ODN2 (no IdU control)

ODN2 elutes at 31.8 min. Gradient is 0%-20% B in 40 min.

Before reaction:

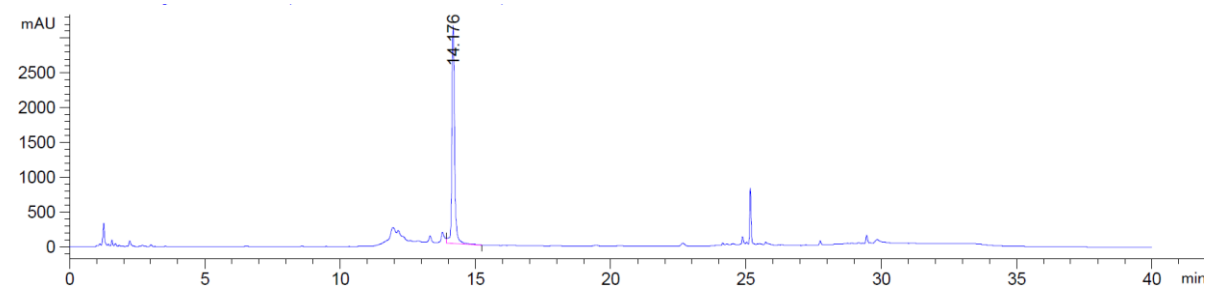


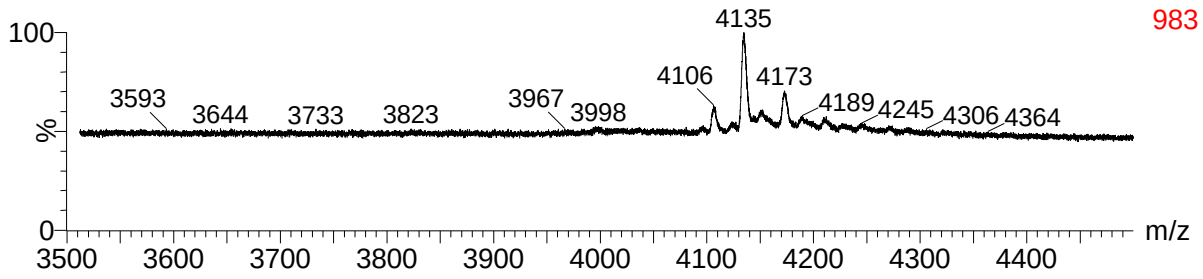
After the reaction:



3.7.12 ODN3a

The Product elutes at 14.2 min. ODN was separated on a gradient of 5%-20% B in 20 min.

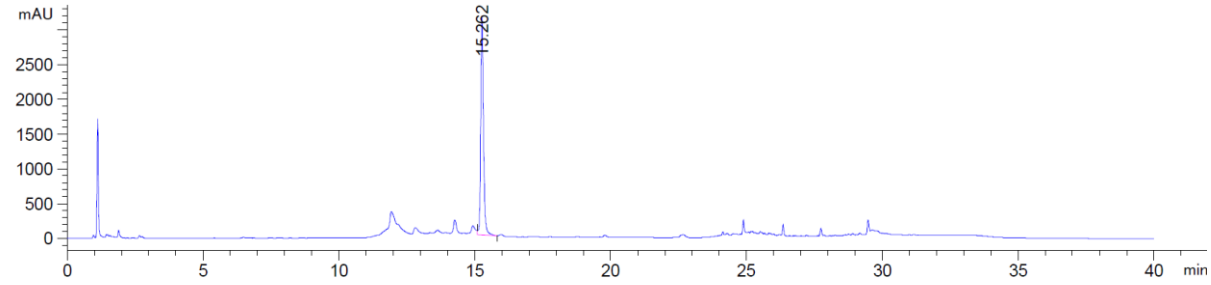


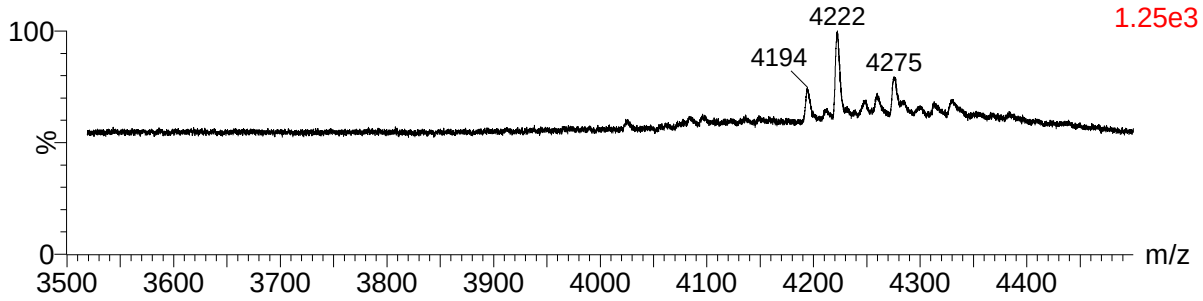


Calc. 4135 found 4135

3.7.13 ODN3b

The Product elutes at 15.3 min. ODN was separated on a gradient of 5%-20% B in 20 min.

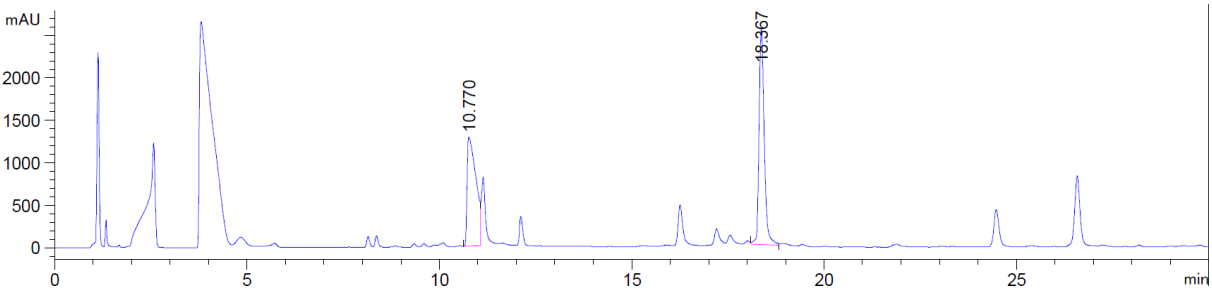


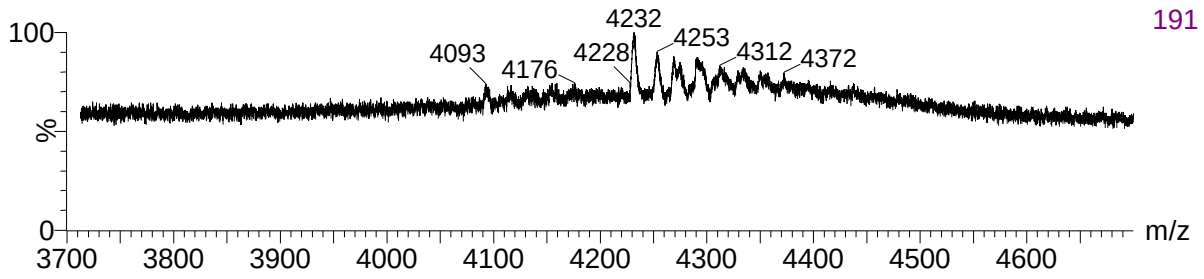


Calc. 4222 found 4222

3.7.14 ODN3c

The Product elutes at 18.4 min. ODN was separated on a gradient of 5%-20% B in 20 min.

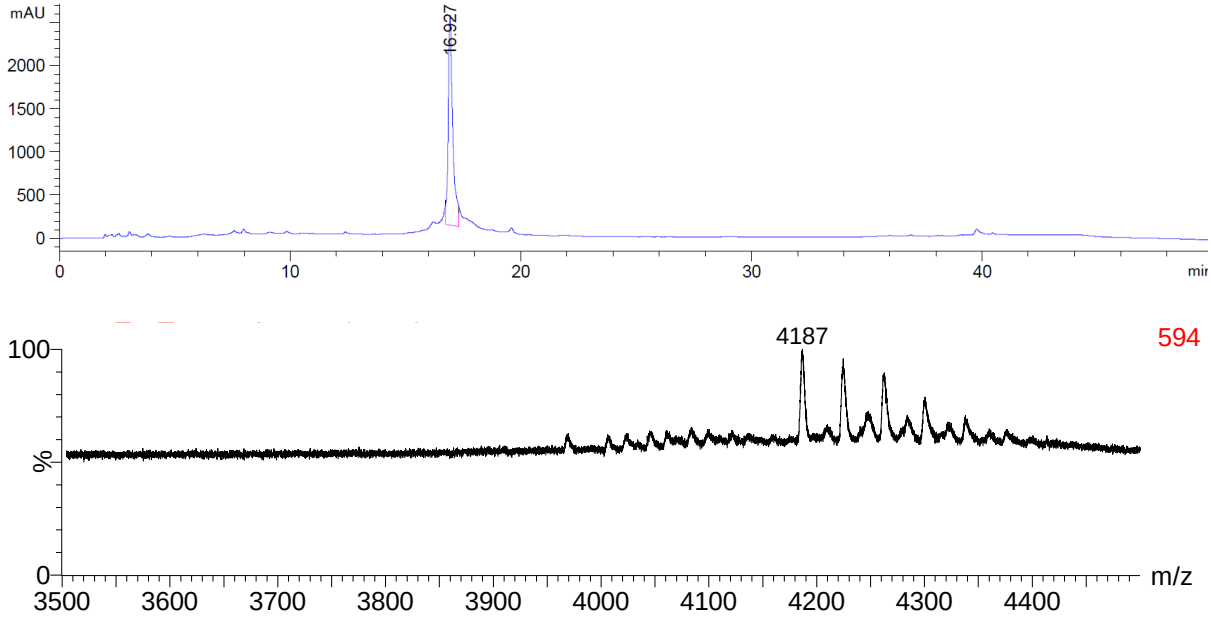




Calc. 4232 found 4232

3.7.15 ODN3f

The Product elutes at 16.9 min. ODN was separated on a gradient of 0%-40% B in 30 min.



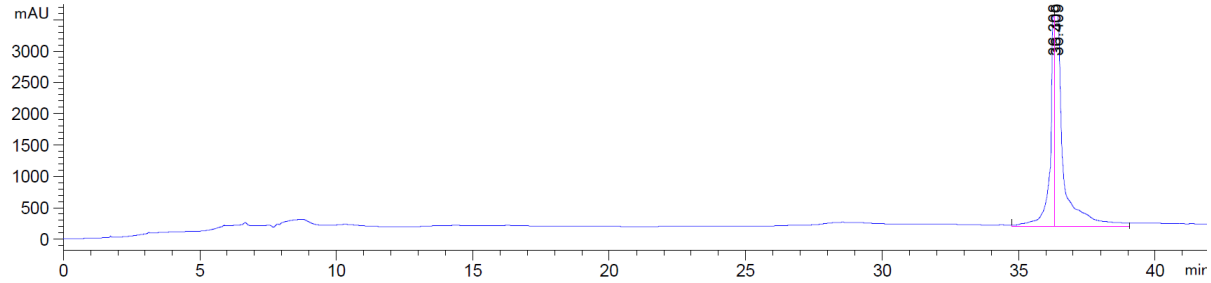
Calc. 4187 found 4187

3.7.16 ODN4 control reactions

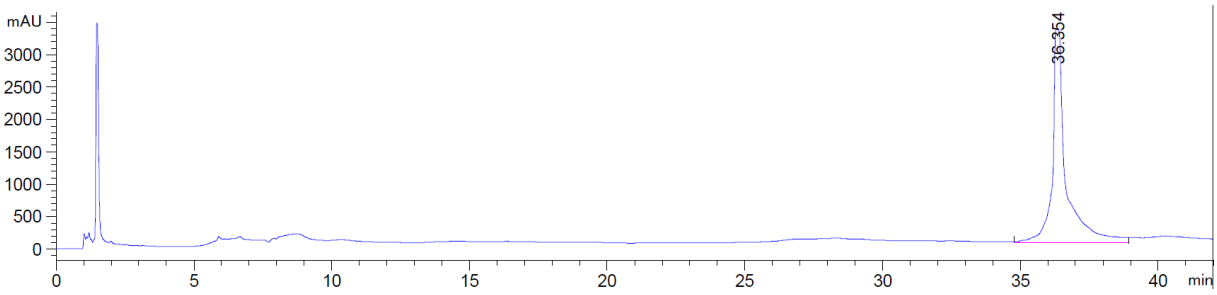
Control reactions were set up as described in the general procedure omitting one component where stated. Boronic ester **5** was used for these experiments.

ODN was separated on a gradient of 0%-20% B in 40 min. The starting material elutes at 36.4 min and the product at 40.5 min.

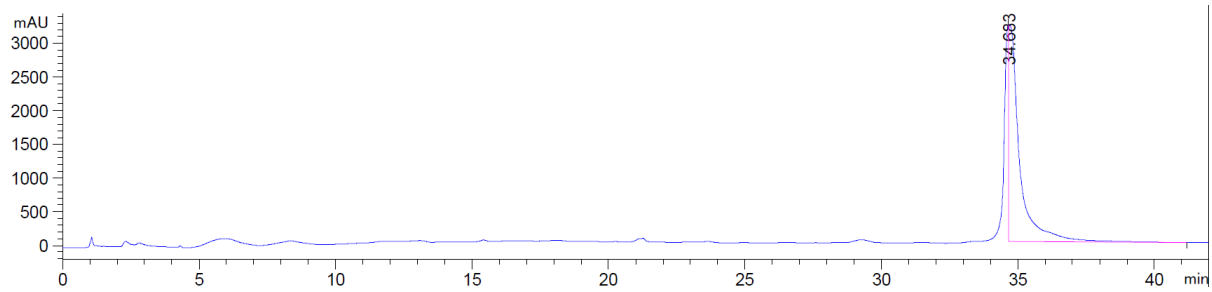
Starting material:



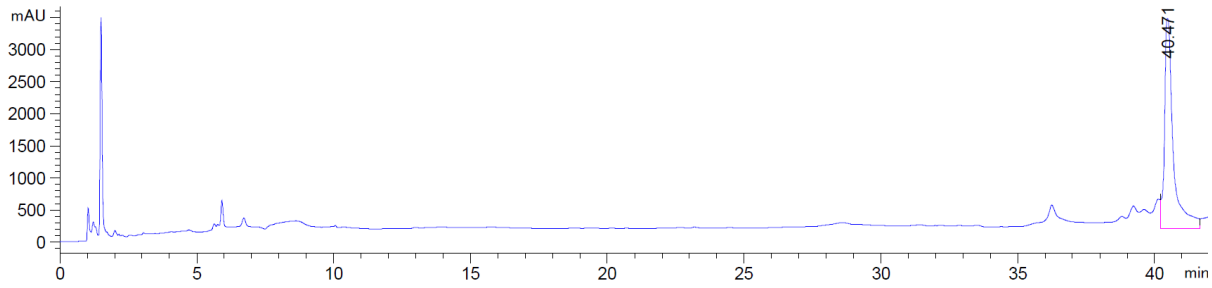
No Boronic ester:



No Palladium:



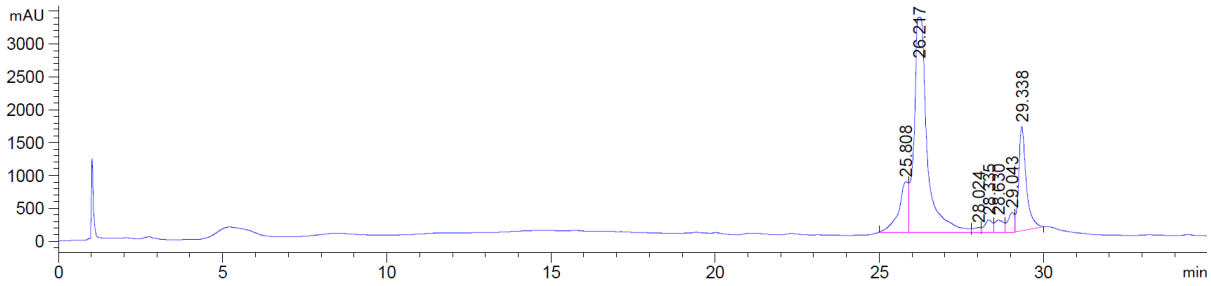
All components:



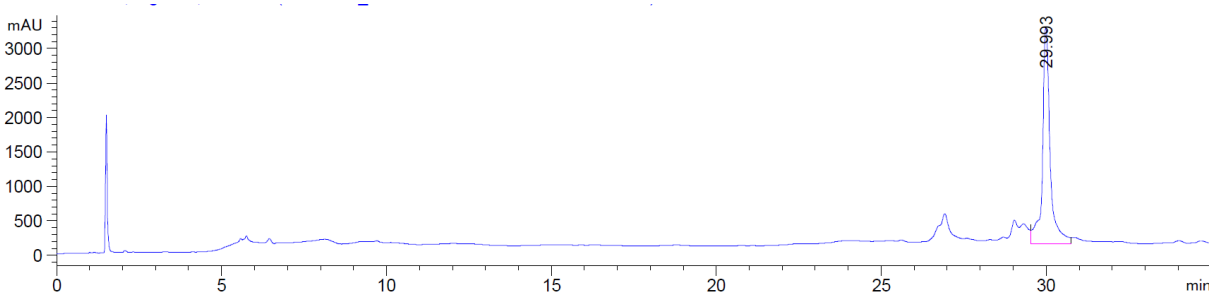
3.7.17 ODN4 – Phosphate and Tris buffer comparison

Reactions were conducted as described in the general procedure but in 100mM PO4 pH 8.5 buffer with Boronic ester **5**. The starting material elutes at 26.2 min and the product at 29.3 min. ODN was separated on a gradient of 0%-30% B in 40 min.

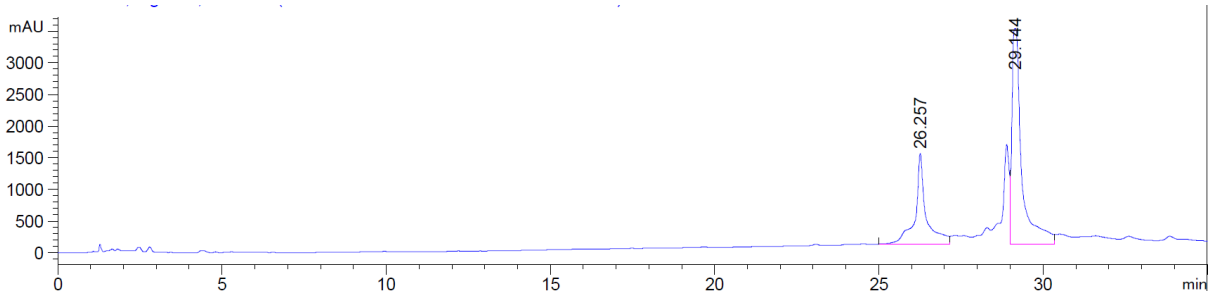
PO4 after 3h:



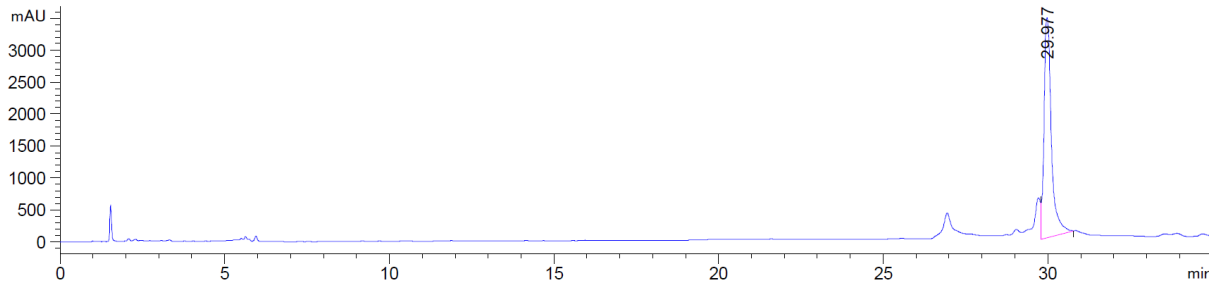
Tris after 3h:



PO4 overnight:

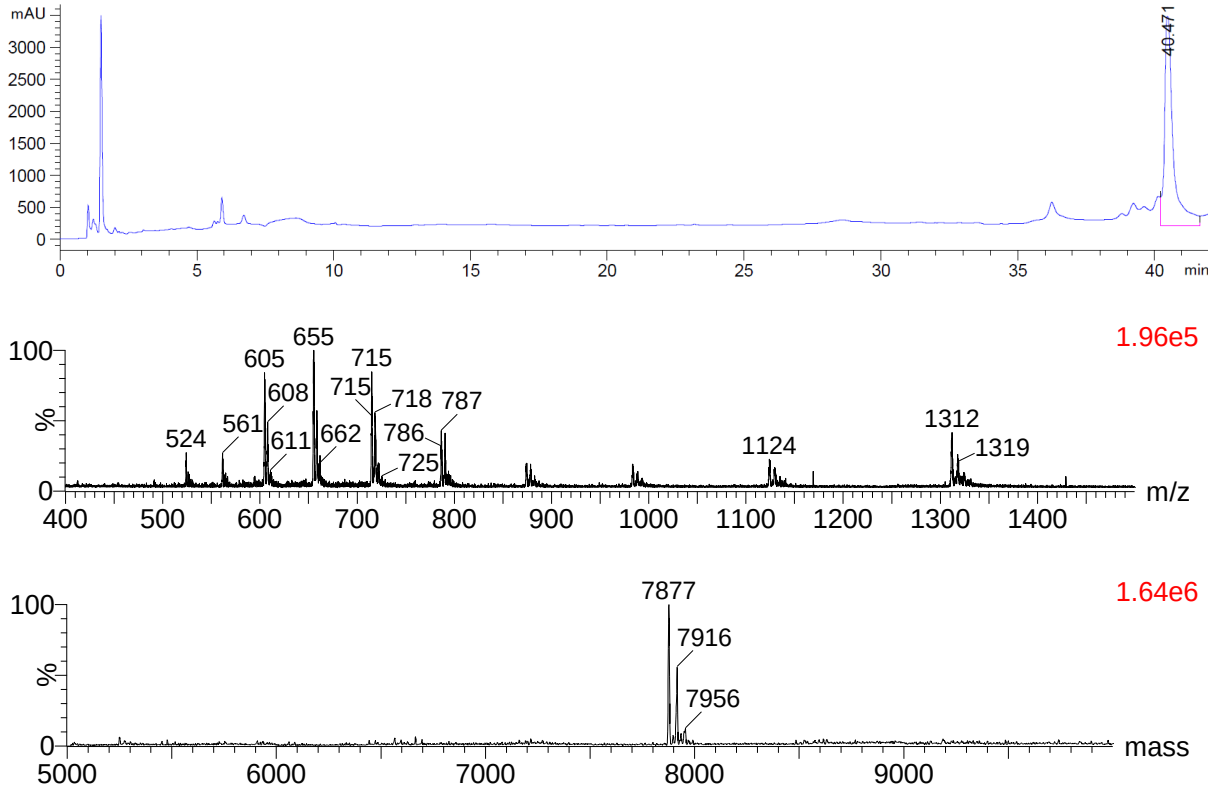


Tris overnight:



3.7.18 ODN4a

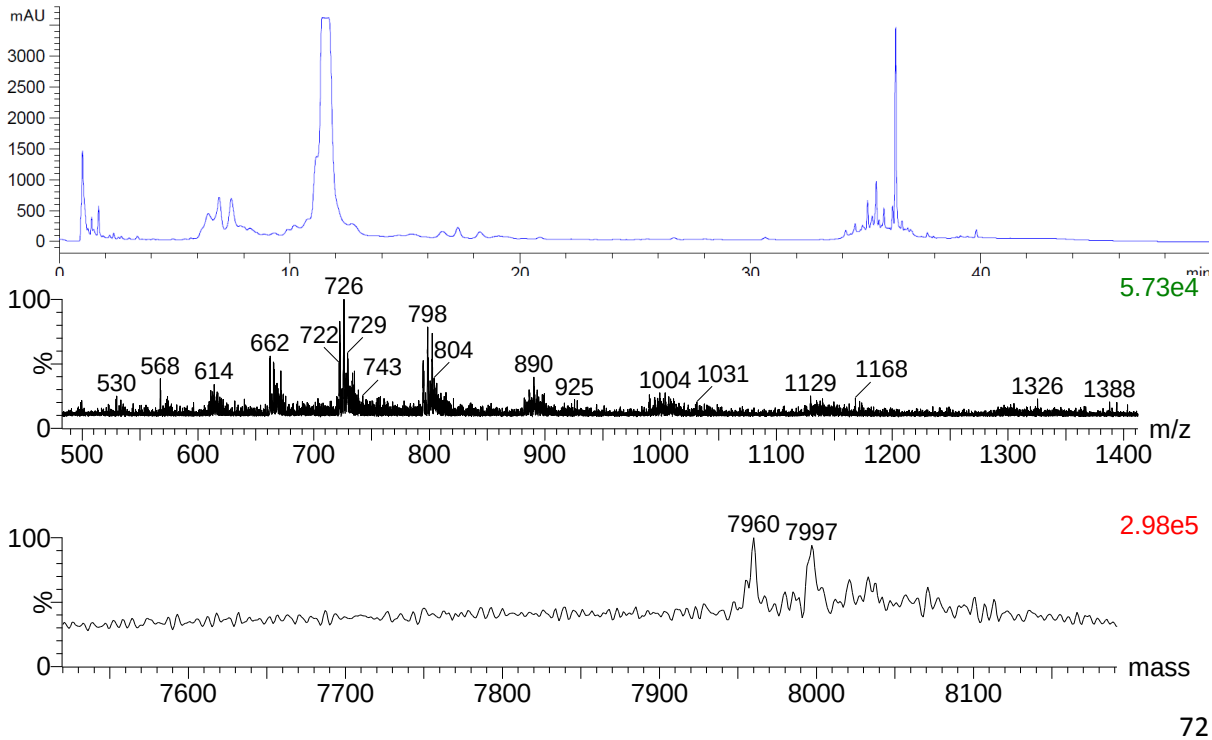
ODN was separated on a gradient of 0%-20% B in 40 min



Calc. 7875 found 7877

3.7.19 ODN4b

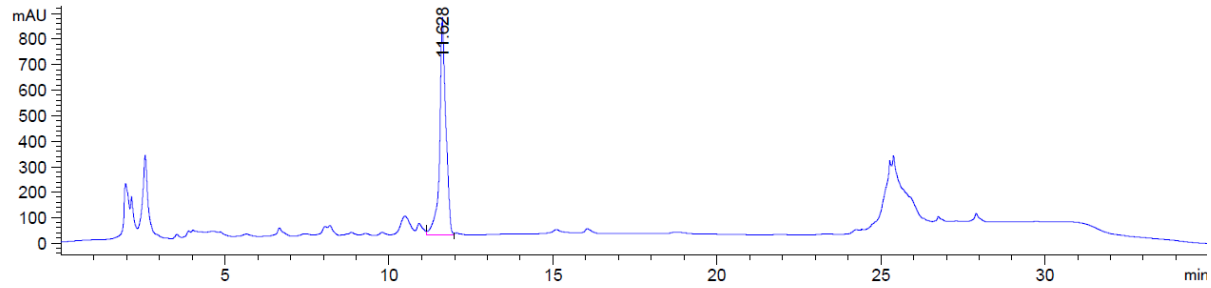
ODN was separated on a gradient of 15%-25% B in 30 min

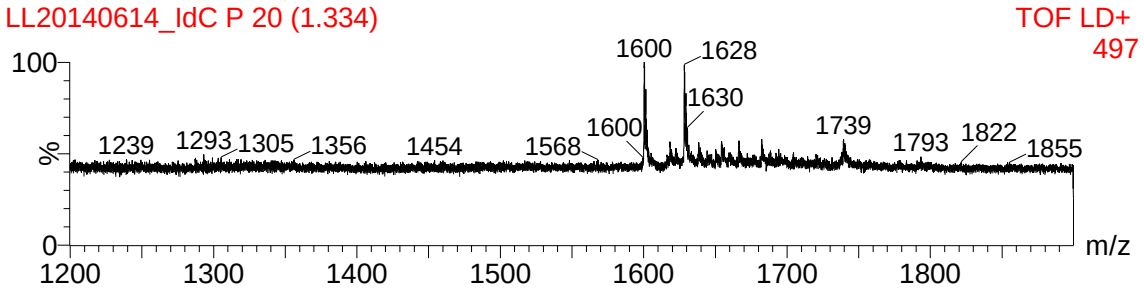


Calc. 7964 found 7960

3.7.20 ODN5a

ODN was separated on a gradient of 10%-30% B in 20 min.

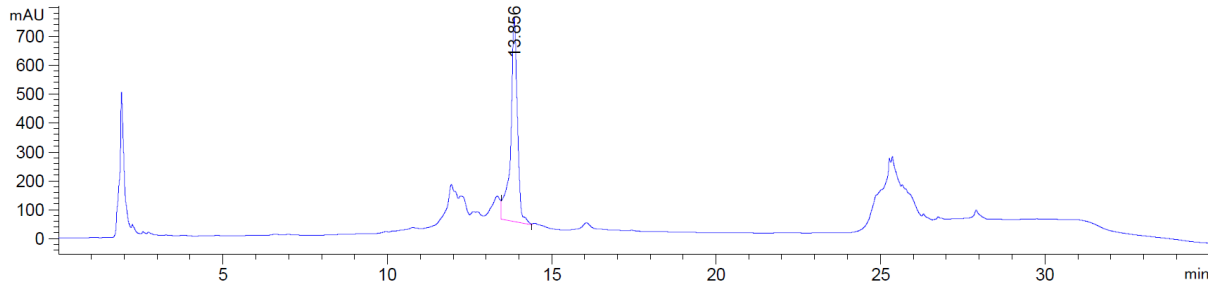


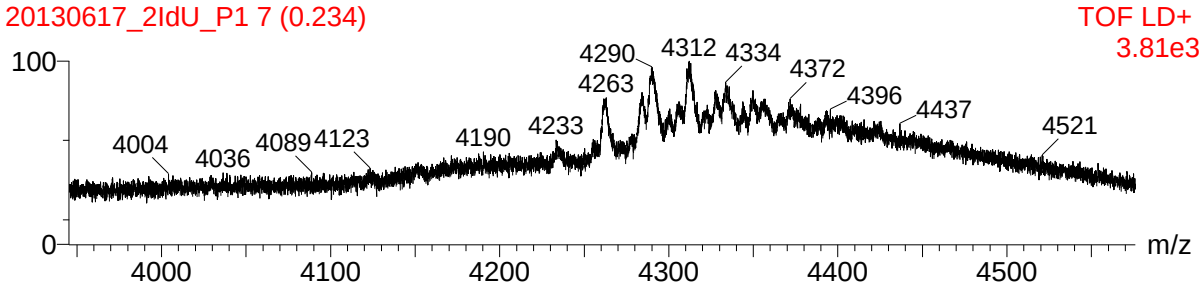


Calc. 1628 found 1628

3.7.21 ODN6a

ODN was separated on a gradient of 10%-30% B in 20 min.





Calc. 4291 found 4290

(other peaks correspond to a Na^+ adduct (4321) or loss of N_2 (4263) and the corresponding Na^+ adduct (4286))

3.7.22 dsDNA crosscoupling

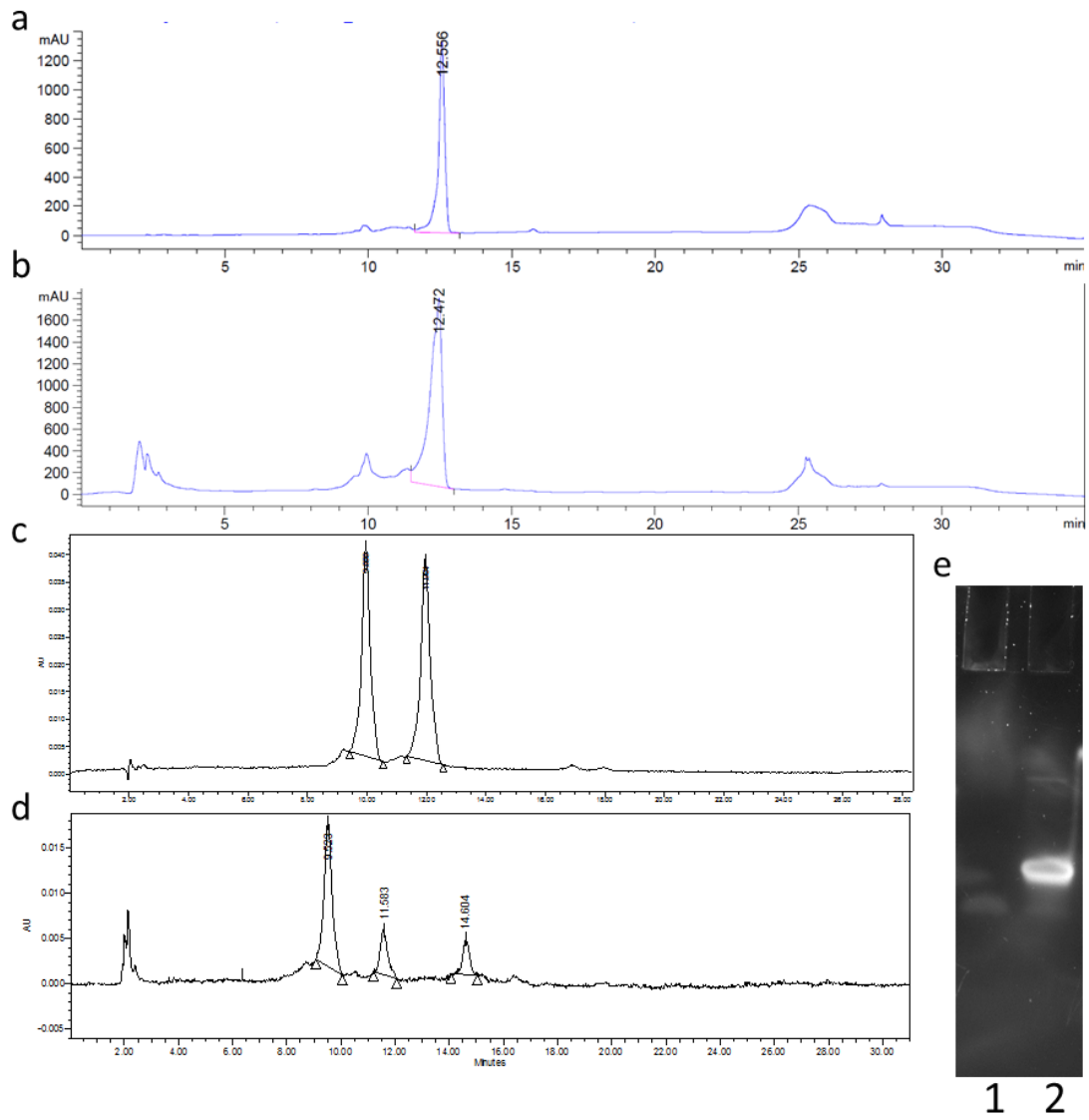


Figure 14. a HPLC chromatogram of dsODN3 after annealing (retention time: 12.6 min). **b** HPLC chromatogram of the SMcc reaction on dsODN3. Main peak retention time is 12.5 min, but the peak is broad and a smaller peak with lower retention time can be observed. **c** Denaturing HPLC chromatogram of the dsODN3. 2 Peaks with almost identical area can be observed at retention times of 10 min and 12 min, corresponding to the two ssODN. **d** Denaturing HPLC chromatogram of the SMcc reaction of dsODN3. 3 main peaks are observed at 9.5 min, 11.5 min and 14.6 min. The peak at 9.5 corresponds to the unreactive ODN, while 11.5 corresponds to the IdU containing ODN and 14.6 to the crosscoupling product. Yield of the crosscoupling product, when comparing to the unmodified ODN is 21%. **e**. PAGE gel showing the ssODN3 in lane 1 and dsODN3 after the crosscoupling reaction. The dsODN is largely still hybridized, but a faint band at the height of the ssODN can be seen, indicating some melting during the reaction.

ODN3 was annealed with the appropriate complementary strand (CTACTGCATCGGT; predicted melting temperature 43 °C, estimated using the IDT oligoanalyzer; http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/) by cooling a solution containing 150 µM of each ODN in annealing buffer (10 mM TRIS pH 7.5, 100 mM NaCl) from 90 °C to 4 °C at 1 °C per minute in a PCR block. The resulting dsODN was used as substrate for SMcc under the standard conditions (100 µM dsODN, 100 µM $PdOAc_2$**L2**$_2$**,** 10 mM **4**, 50 mM TRIS pH 8.5) and was shaken at 37 °C for 16 h. The resulting mixture was analyzed by PAGE (15%, 0.5x TBE; stained with Sybr safe), HPLC (10%-30% buffer B in 20 min) or denaturing HPLC (column was heated at 50 °C; 10%-20% B in 30 min; **Figure 15**).

3.8 Suzuki-Miyaura cross-coupling to RNA

Optimization reactions (**Table 7**) were performed with following components and shaken at 37 °C in an incubator.

	Stock [c]		Required [c]		volume		equivalents
RNA	1000	µM	100	µM	2.50	µL	
TRIS	1000	mM	100	mM	2.5	µL	
$PdOAc_2L_2$	1	mM	0.5	mM	12.5	µL	5 eq
RBPin	250	mM	50	mM	5	µL	500 eq
ddH_2O					2.50	µL	
Total					25	µL	
RNA	1000	µM	100	µM	2.50	µL	
TRIS	500	mM	50	mM	2.50	µL	
$PdOAc_2L_2$	1	mM	0.5	mM	12.50	µL	5 eq
RBPin	50	mM	10	mM	5.00	µL	100 eq
ddH_2O					2.50	µL	
Total					25	µL	

	Stock [c]		Required [c]		volume		equivalents
RNA	1000	µM	100	µM	2.50	µL	
TRIS	500	mM	50	mM	2.50	µL	
$PdOAc_2L_2$	1	mM	0.1	mM	2.50	µL	1 eq
RBPin	250	mM	50	mM	5.00	µL	500 eq
ddH_2O					12.50	µL	
Total					25	µL	
RNA	1000	µM	100	µM	2.50	µL	
TRIS	500	mM	50	mM	2.50	µL	

$PdOAc_2L_2$	1	mM	0.2	mM	5.00	µL	1 eq
RBPin	50	mM	10	mM	5.00	µL	100 eq
ddH_2O					10.00	µL	
Total					25	µL	

Table 7. Optimization reactions for the RNA SMcc.

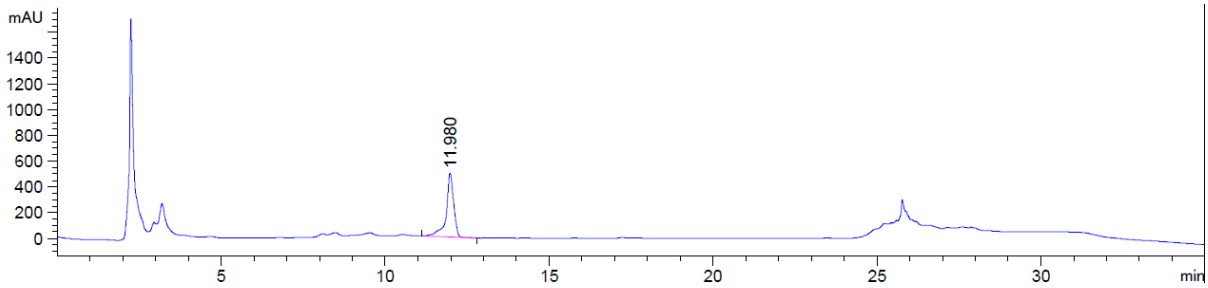
Reactions with different boronic acids in **Figure 8** were performed as described below. The reaction with the pyrene boronic acid was performed in 50% (v/v) aqueous MeCN.

	Stock [c]		Required [c]		volume		
RNA	1000	µM	100	µM	1.50	µL	
TRIS	1000	mM	100	mM	1.5	µL	
$PdOAc_2L_2$	1	mM	0.2	mM	3	µL	2 eq
RBPin	250	mM	20	mM	1.2	µL	200 eq
ddH_2O					7.80	µL	
Total					15	µL	

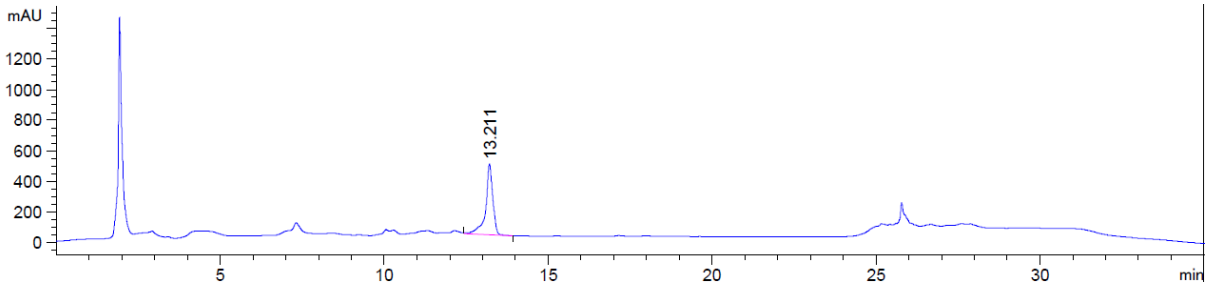
Table 8. SMcc with pyrene boronic acid and **ON1**.

HPLC traces for the analysis listed in **Table 7** are shown below.

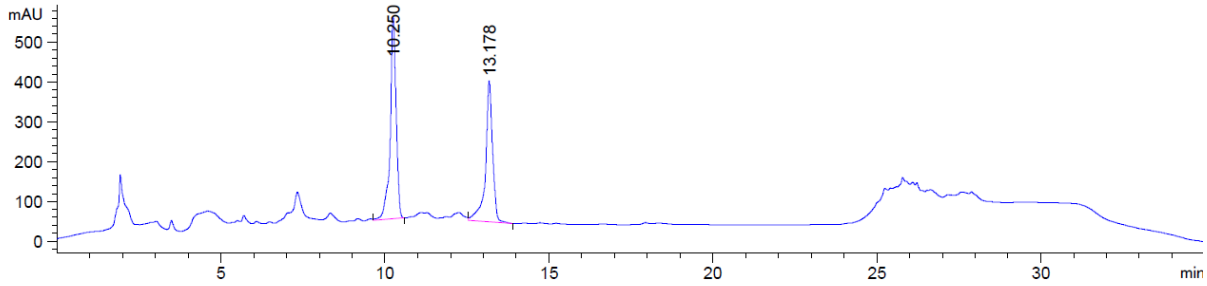
Entry 1



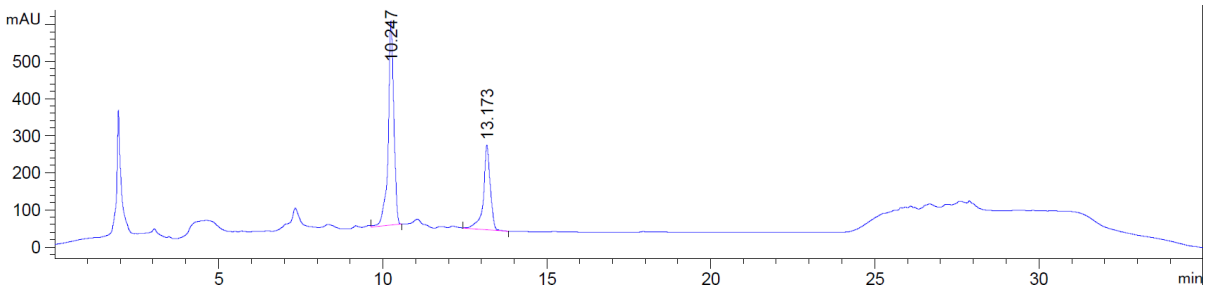
Entry 2



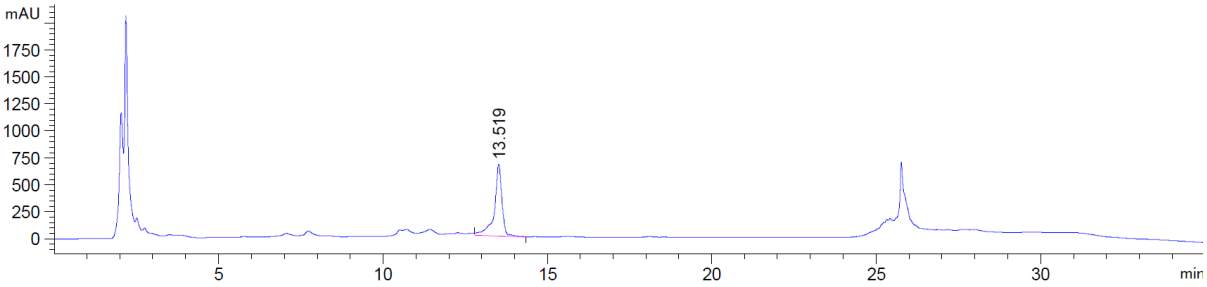
Entry 3



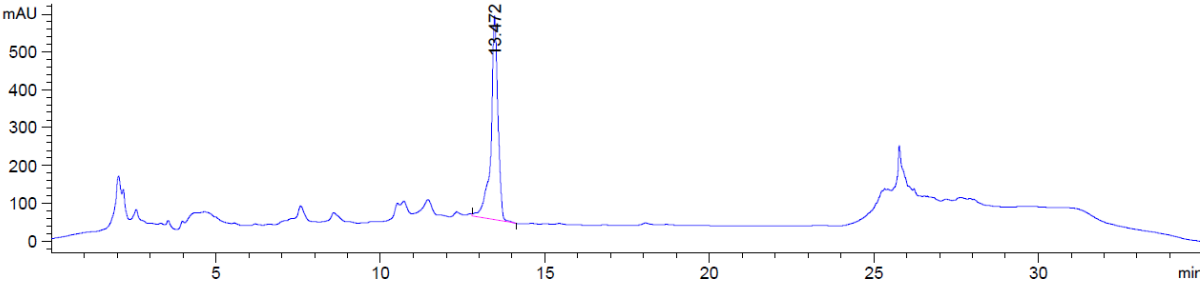
Entry 4



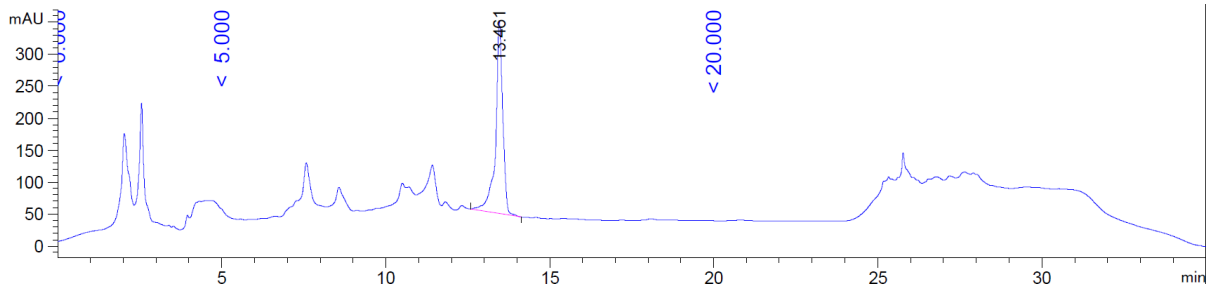
Entry 5



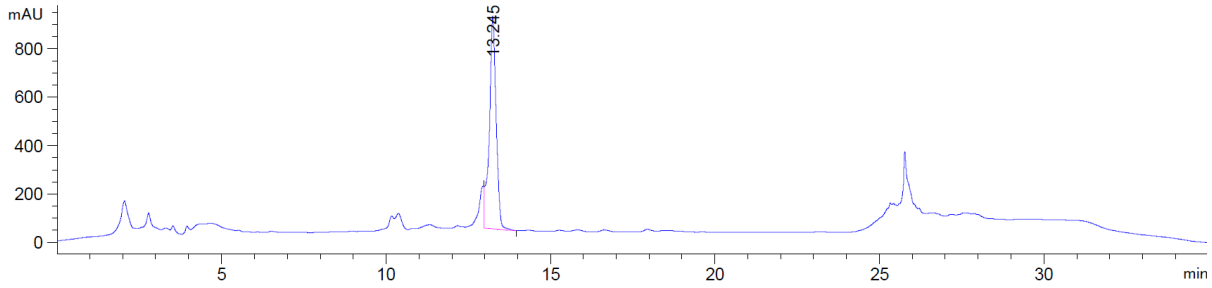
Entry 6



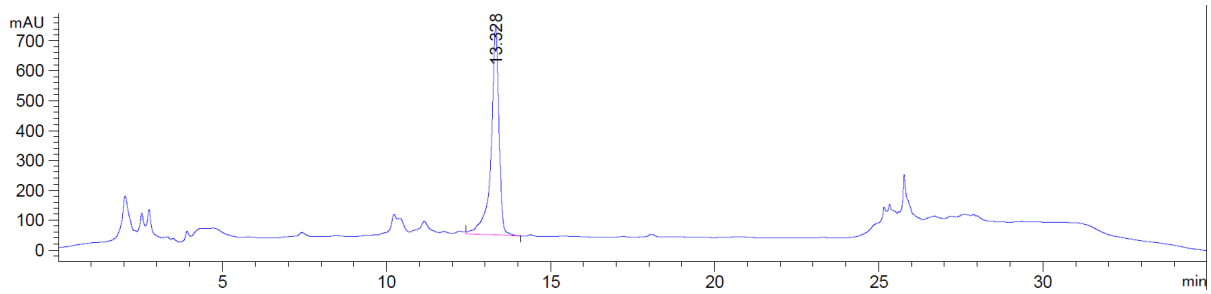
Entry 7



Entry 8

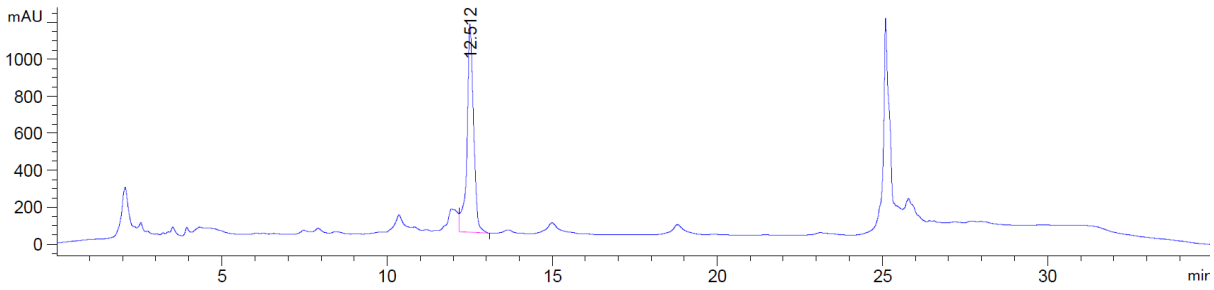


Entry 9

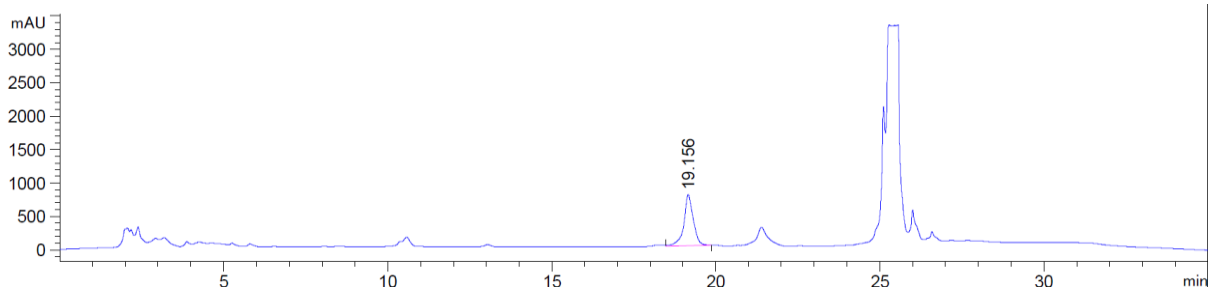


HPLC traces for the analysis listed in **Figure 8** are shown below.

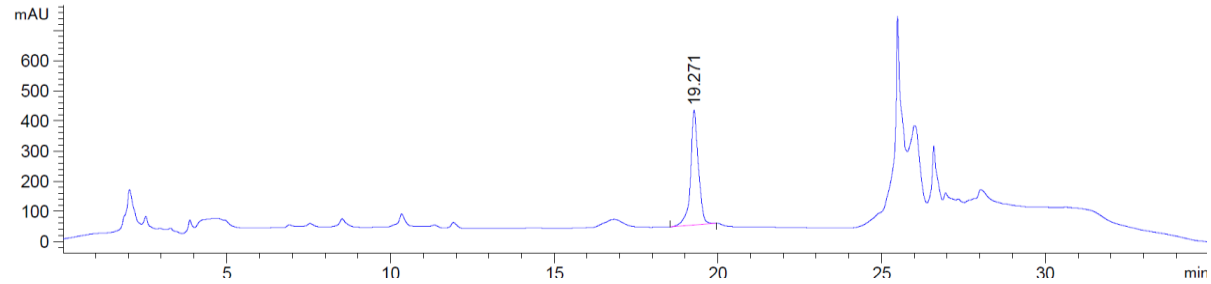
Entry 1



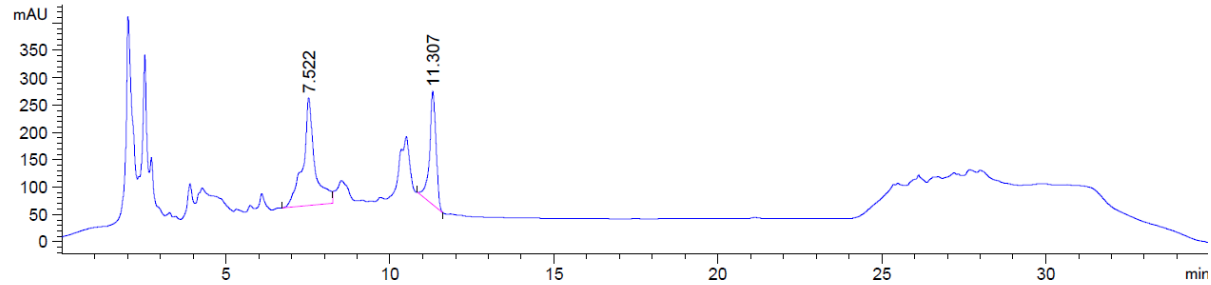
Entry 2



Entry 3

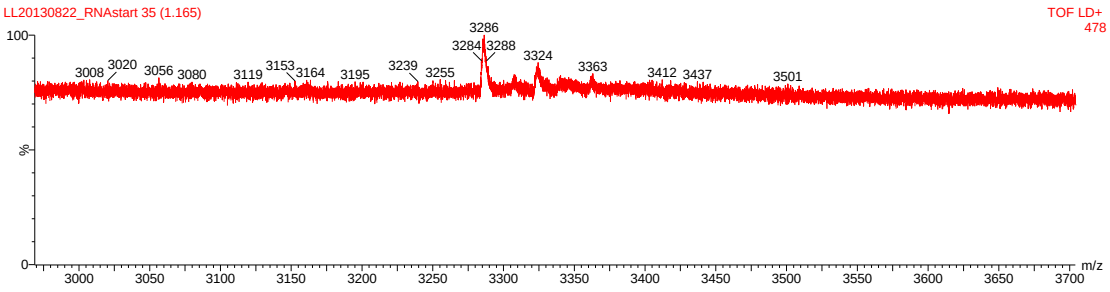


Entry 4

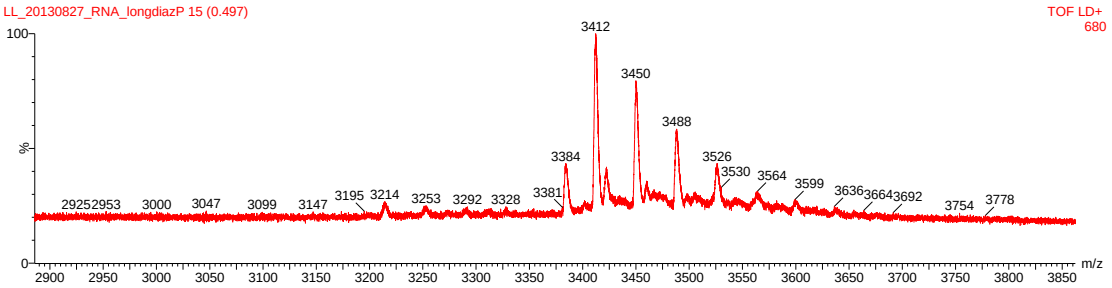


MALDI analysis of starting material and products of the SMcc on RNA.

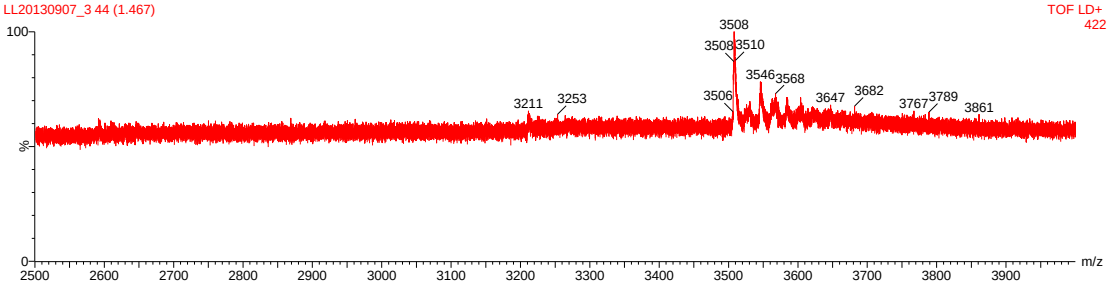
Starting material (**ON1**)



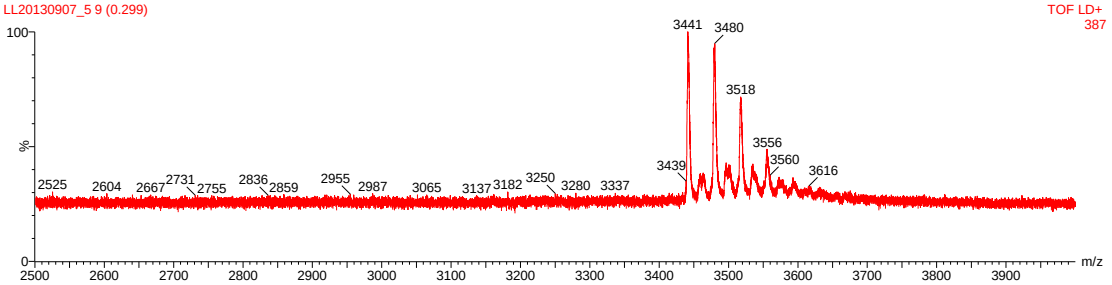
ON1a (calc 3416)



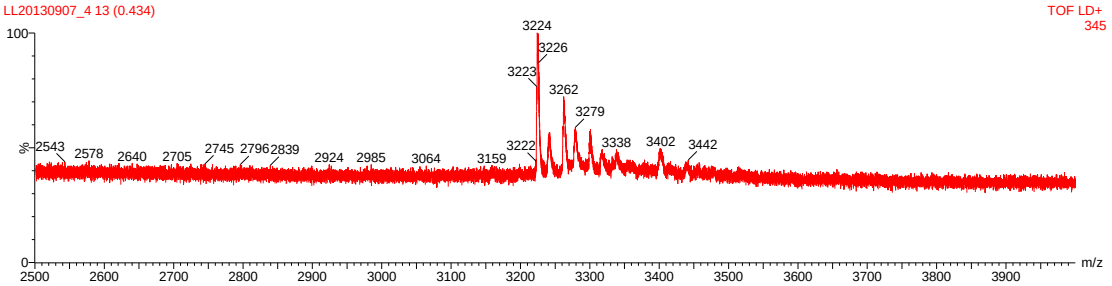
ON1b (**Figure 8** Entry 1; calc 3511)



ON1c (**Figure 8** Entry 2; calc 3443)



ON1d (**Figure 8** Entry 3; calc 3226)



3.9 DNA Photocrosslinking

Photocrosslinking probes were generated by PCR using Vent exo$^-$ (NEB, M0257S). For the hmC containing probe dCTP was replaced by dhmCTP. After PCR, the strands were analysed by PAGE and purified by phenol extraction and ethanol precipitation.

For crosslinking tests, nuclear extract was prepared from HeLa cells (2 x $1.5e^7$ cells) using the NE-PER nuclear extract preparation kit (Pierce, 78833). The obtained nuclear extract was dialyzed into binding buffer (20 mM TRIS pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM Benzamidine, 0.1 % TWEEN, 10% glycerol). To remove any biotinylated or streptavidin binding proteins, the lysate was cleared with 50 µL CaptAvidin™ Agarose (Life technologies, C-21386) by incubation at 4 °C for 1 h. Beads were removed by centrifugation giving 2 x 550 µL of 1.10 mg / mL. To the lysate, 45 µL of the hmC or C containing probe (35 µM; 2.6 µM final) and 5 µL poly dI/dC (500 ng/µL stock) was added. The mixture was incubated at 4 °C for 15 min and then divided into two 300 µL aliquots. One aliquot of each reaction was crosslinked by irradiation at 365 nm using an 8W TLC lamp for 30 min the other was incubated at 4 °C for 30 min without UV irradiation. Then the reaction was adjusted to 1 M NaCl by addition of 0.5 vol 4 M NaCl and 100 µL streptavidin beads ((Dynabeads® MyOne™ Streptavidin T1; Invitrogen; 65601) were added. The mixture was incubated at rt for 30 min, then beads were collected on a DynaMag™- Spin Magnet (Life technologie, 12320D) and washed 3 times with wash buffer (10 mM TRIS pH 7.5, 1 M NaCl, 0.1% TWEEN-20) and 3 times with denaturing buffer (10 mM TRIS pH 7.5, 1 M NaCl, 0.1% Tween-20, 6 M Urea). The proteins were then eluted by boiling the beads in 2x SDS-loading buffer for 15 min and separated by SDS-PAGE and analysed by westernblot stained with ExtrAvidine-Alkaline Phosphatase (Sigma, E2636).

For SILAC quantification, nuclear extract was prepared from $3x10^7$ “heavy” labelled and “light” unlabelled HeLa cells using the NE-PER kit. The obtained nuclear extract was dialyzed into binding buffer (20 mM TRIS pH 7.5, 150 mM NaCl, 1 mM $MgCl_2$, 1 x complete protease inhibitor, 0.1 % TWEEN, 10% glycerol). To remove any biotinylated or streptavidin binding proteins, the lysate was cleared with 50 µL CaptAvidin™ Agarose by incubation at 4 °C for 1 h. Beads were removed by centrifugation. The concentration was measured by nanodrop and the protein concentration was adjusted to 2 mg/mL for both heavy and light lysate using binding buffer (~1.30 mL). To 1.25 mL of the lysate, the hmC or C containing probe (3 nmol) and 5 µL poly dI/dC (500 ng/µL stock) was added. The mixture was incubated at 4 °C for 15 min and then divided into 100 µL aliquots in PCR tubes. The

aliquots were irradiated in the Caprotec box at 311 nm at 4 °C for 30 min. Then the reaction mixture was adjusted to 1 M NaCl by addition of 0.5 vol 4 M NaCl and 400 µL streptavidin beads were added. The mixture was incubated at rt for 1 h, then beads were collected on a DynaMag™- Spin Magnet and washed 3 times with wash buffer (10 mM TRIS pH 7.5, 1 M NaCl, 0.1% TWEEN-20) and 3 times with denaturing buffer (10 mM TRIS pH 7.5, 1 M NaCl, 0.1% Tween-20, 6 M Urea). The proteins were then eluted by boiling the beads in 2x SDS-loading buffer for 15 min and separated by SDS-PAGE. The gel was then cut in 5 slices and digested as described elsewhere (http://www.tdi.ox.ac.uk/_asset/file/in-gel.pdf). LC-MS analysis was performed by Joanna M^{c}Gouran (Benedikt Kessler lab, see appendix for full list of proteins identified).

3.10 References

(1) Lercher, L.; McGouran, J. F.; Kessler, B. M.; Schofield, C. J.; Davis, B. G. *Angew. Chem. Int. Ed.* **2013**, *52*, 10553.
(2) El-Sagheer, A. H.; Brown, T. *Chem. Soc. Rev.* **2010**, *39*, 1388.
(3) Tolle, F.; Mayer, G. *Chem. Sci.* **2013**, *4*, 60.
(4) Beyer, C.; Wagenknecht, H.-A. *Chem. Commun.* **2010**, *46*, 2230.
(5) Gutsmiedl, K.; Wirges, C. T.; Ehmke, V.; Carell, T. *Org. Lett.* **2009**, *11*, 2405.
(6) Gutsmiedl, K.; Fazio, D.; Carell, T. *Chem. Eur. J.* **2010**, *16*, 6877.
(7) Borsenberger, V.; Howorka, S. *Nucleic Acids Res.* **2009**, *37*, 1477.
(8) Schoch, J.; Wiessler, M.; Jäschke, A. *J. Am. Chem. Soc.* **2010**, *132*, 8846.
(9) Raindlová, V.; Pohl, R.; Šanda, M.; Hocek, M. *Angew. Chem. Int. Ed.* **2010**, *49*, 1064.
(10) Raindlová, V.; Pohl, R.; Hocek, M. *Chem. Eur. J.* **2012**, *18*, 4080.
(11) Wang, C. C. Y.; Seo, T. S.; Li, Z.; Ruparel, H.; Ju, J. *Bioconjugate Chem.* **2003**, *14*, 697.
(12) Shigdel, U. K.; Zhang, J.; He, C. *Angew. Chem. Int. Ed.* **2008**, *47*, 90.
(13) Rohde, H.; Schmalisch, J.; Harpaz, Z.; Diezmann, F.; Seitz, O. *ChemBioChem* **2011**, *12*, 1396.
(14) Eberhard, H.; Diezmann, F.; Seitz, O. *Angew. Chem. Int. Ed.* **2011**, *50*, 4146.
(15) Schlegel, M. K.; Hütter, J.; Eriksson, M.; Lepenies, B.; Seeberger, P. H. *ChemBioChem* **2011**, *12*, 2791.
(16) Cahová, H.; Jäschke, A. *Angew. Chem. Int. Ed.* **2013**, *52*, 3186.
(17) Omumi, A.; Beach, D. G.; Baker, M.; Gabryelski, W.; Manderville, R. A. *J. Am. Chem. Soc.* **2010**, *133*, 42.
(18) Chalker, J. M.; Wood, C. S. C.; Davis, B. G. *J. Am. Chem. Soc.* **2009**, *131*, 16346.
(19) Spicer, C. D.; Davis, B. G. *Chem. Commun.* **2011**, *47*, 1698.
(20) Spicer, C. D.; Triemer, T.; Davis, B. G. *J. Am. Chem. Soc.* **2011**, *134*, 800.
(21) Spicer, C. D.; Davis, B. G. *Chem. Commun.* **2013**, *49*, 2747.
(22) Dumas, A.; Spicer, C. D.; Gao, Z.; Takehana, T.; Lin, Y. A.; Yasukohchi, T.; Davis, B. G. *Angew. Chem. Int. Ed.* **2013**, *52*, 3916.
(23) Gao, Z.; Gouverneur, V.; Davis, B. G. *J. Am. Chem. Soc.* **2013**, *135*, 13612.
(24) Münzel, M.; Globisch, D.; Carell, T. *Angew. Chem. Int. Ed.* **2011**, *50*, 6460.
(25) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
(26) Song, C.-X.; Yi, C.; He, C. *Nat. Biotech.* **2012**, *30*, 1107.

(27) Münzel, M.; Globisch, D.; Brückl, T.; Wagner, M.; Welzmiller, V.; Michalakis, S.; Müller, M.; Biel, M.; Carell, T. *Angew. Chem. Int. Ed.* **2010**, *49*, 5375.
(28) Pereira, S.; Srebnik, M. *Organometallics* **1995**, *14*, 3127.
(29) Wang, Y. D.; Kimball, G.; Prashad, A. S.; Wang, Y. *Tetrahedron Lett.* **2005**, *46*, 8777.
(30) Church, R. F. R.; Weiss, M. J. *J. Org. Chem.* **1970**, *35*, 2465.
(31) Ikeda, Y.; Behrman, E. J. *Synth. Commun.* **2008**, *38*, 2276.
(32) Balas, L.; Durand, T.; Saha, S.; Johnson, I.; Mukhopadhyay, S. *J. Med. Chem.* **2009**, *52*, 1005.
(33) Henze, M.; Kreye, O.; Brauch, S.; Nitsche, C.; Naumann, K.; Wessjohann, L. A.; Westermann, B. *Synthesis* **2010**, *2010*, 2997.
(34) Western, E. C.; Daft, J. R.; Johnson, E. M.; Gannett, P. M.; Shaughnessy, K. H. *J. Org. Chem.* **2003**, *68*, 6767.
(35) Carrow, B. P.; Hartwig, J. F. *J. Am. Chem. Soc.* **2011**, *133*, 2116.
(36) Tao, B.; Boykin, D. W. *J. Org. Chem.* **2004**, *69*, 4330.
(37) Beharry, A. A.; Woolley, G. A. *Chem. Soc. Rev.* **2011**, *40*, 4422.
(38) Scheibe, C.; Bujotzek, A.; Dernedde, J.; Weber, M.; Seitz, O. *Chem. Sci.* **2011**, *2*, 770.
(39) Yan, H.; Tram, K. *Glycoconjugate J.* **2007**, *24*, 107.
(40) Srivatsan, S. G.; Tor, Y. *J. Am. Chem. Soc.* **2007**, *129*, 2044.
(41) Greco, N. J.; Tor, Y. *J. Am. Chem. Soc.* **2005**, *127*, 10784.
(42) Greco, N. J.; Sinkeldam, R. W.; Tor, Y. *Org. Lett.* **2009**, *11*, 1115.
(43) Aliprantis, A. O.; Canary, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 6985.
(44) Li, Y.; Breaker, R. R. *J. Am. Chem. Soc.* **1999**, *121*, 5364.
(45) Raines, R. T. *Chem. Rev.* **1998**, *98*, 1045.
(46) Lister, R.; Mukamel, E. A.; Nery, J. R.; Urich, M.; Puddifoot, C. A.; Johnson, N. D.; Lucero, J.; Huang, Y.; Dwork, A. J.; Schultz, M. D.; Yu, M.; Tonti-Filippini, J.; Heyn, H.; Hu, S.; Wu, J. C.; Rao, A.; Esteller, M.; He, C.; Haghighi, F. G.; Sejnowski, T. J.; Behrens, M. M.; Ecker, J. R. *Science* **2013**, *341*, 1237905.
(47) Wang, T.; Pan, Q.; Lin, L.; Szulwach, K. E.; Song, C. X.; He, C.; Wu, H.; Warren, S. T.; Jin, P.; Duan, R.; Li, X. *Hum Mol Genet* **2012**, *21*, 5500.
(48) Spruijt, Cornelia G.; Gnerlich, F.; Smits, Arne H.; Pfaffeneder, T.; Jansen, Pascal W. T. C.; Bauer, C.; Münzel, M.; Wagner, M.; Müller, M.; Khan, F.; Eberl, H. C.; Mensinga, A.; Brinkman, Arie B.; Lephikov, K.; Müller, U.; Walter, J.; Boelens, R.; van Ingen, H.; Leonhardt, H.; Carell, T.; Vermeulen, M. *Cell* **2013**, *152*, 1146.
(49) Mann, M. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 952.
(50) Geyer, H.; Geyer, R.; Pingoud, V. *Nucleic Acids Res.* **2004**, *32*, e132.
(51) Winnacker, M.; Welzmiller, V.; Strasser, R.; Carell, T. *ChemBioChem* **2010**, *11*, 1345.
(52) Winnacker, M.; Breeger, S.; Strasser, R.; Carell, T. *ChemBioChem* **2009**, *10*, 109.
(53) Mellén, M.; Ayata, P.; Dewell, S.; Kriaucionis, S.; Heintz, N. *Cell* **2012**, *151*, 1417.
(54) Fischer, J. J.; Michaelis, S.; Schrey, A. K.; Baessler, O. G. n.; Glinski, M.; Dreger, M.; Kroll, F.; Koester, H. *Toxicol. Sci.* **2010**, *113*, 243.
(55) Koster, H.; Little, D. P.; Luan, P.; Muller, R.; Siddiqi, S. M.; Marappan, S.; Yip, P. *Assay Drug Dev Technol* **2007**, *5*, 381.
(56) Molander, G. A.; Cadoret, F. *Tetrahedron Lett.* **2011**, *52*, 2199.
(57) Lee, Y.; Jang, H.; Hoveyda, A. H. *J. Am. Chem. Soc.* **2009**, *131*, 18234.
(58) Chang, S.-Y.; Bae, S. J.; Lee, M. Y.; Baek, S.-h.; Chang, S.; Kim, S. H. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 727.
(59) Greenwald, R. B.; Choe, Y. H.; Conover, C. D.; Shum, K.; Wu, D.; Royzen, M. *J. Med. Chem.* **2000**, *43*, 475.
(60) Benanti, T. L.; Saejueng, P.; Venkataraman, D. *Chem. Commun.* **2007**, 692.
(61) Kim, Y.-S.; Kim, K. M.; Song, R.; Jun, M. J.; Sohn, Y. S. *J. Inorg. Biochem.* **2001**, *87*, 157.

4 Chemical methods to introduce PTMs into histone proteins: multiple modifications and biological validation

To investigate the function of posttranslational modifications (PTMs), access to proteins bearing homogeneous PTMs is essential.[1,2] Such homogenous recombinant proteins can also be utilized for *in vitro* studies to investigate the biophysical effect of PTMs on protein structure and function.[3] A major target of the cell biology research community currently is to elucidate the mechanisms of epigenetic control by histone protein PTMs.[4] To allow *in vitro* characterization of functional consequences of these marks, a variety of methods to generate histones bearing homogeneous PTMs have been developed. These methods can be divided into three categories: translational incorporation, introduction via native chemical ligation and post-translational synthetic incorporation (summarised in **Figure 1**).

Site-selective translational incorporation of PTMs can be achieved by codon suppression.[5] This approach uses an aminoacyl-tRNA-synthetase (aaRS) that is orthogonal to canonical amino acids and tRNAs and specifically recognizes a PTM bearing amino acid and a tRNA selective for, most commonly, the UAG (amber) stop-codon. The orthogonallity of both aaRS and tRNA towards naturally occurring aaRS, tRNA and amino acids is essential for homogenous site-selective incorporation. By co-expressing the aaRS, the UAG-tRNA and the protein of interest bearing a TAG codon at the desired PTM site one can access a variety of PTMs directly through heterologous expression. A variety of lysine PTMs have been incorporated into histones via this method (**Figure 1a**).[6-10] While this methodology is very convenient, yields can be low because of competition between termination and incorporation, although this issue has been addressed recently.[11] For some modifications, additional steps are required after isolation of the protein[6,7] and unstable modifications cannot be accessed by this method.[12] Also, since the aaRS are often selective for only one amino acid, incorporation of a different structure requires the evolution of a new selective aaRS which can be time consuming.

Native chemical ligation (NCL)[13,14] and expressed protein ligation (EPL)[15] are powerful biochemical approaches that allow the introduction of many PTMs. In NCL two protecting-group free ('native') peptide fragments are joined by the reaction of an *N*-terminal cysteine on one fragment with a *C*-terminal activated thioester on the second fragment. These two peptide fragments can be derived from solid phase synthesis (SPS) with any number of modifications, as long as they are stable under SPS conditions. Limitations of SPS are that large peptides (>40-50 amino acids) are difficult to access and that SPS requires quite harsh conditions, which precludes access to labile modification such as phosphohistidine.[12,16] To circumvent the need to synthesise very large peptide fragments, EPL was devised. Herein, one of the fragments bearing either the *N*-terminal cysteine or the *C*-terminal thioester is expressed recombinant, while the sequence bearing the modification is synthesized by SPS. The *N*-terminal thioester is generated by use of an inteine fusion protein[15,17] and the *C*-terminal fragment bearing a cysteine at the *N*-terminus can be generated by a variety of methods.[18-21] The synthetic and the expressed protein part can then be combined, making larger proteins with modifications near the *N*- or *C*-terminal accessible.

In post-translational synthetic incorporation, a tag, usually a cysteine residue is incorporated into the protein. After production and purification of the protein, the PTM is introduced via a chemical reaction (**Figure 1b**), yielding a PTM mimic with a slight structural change. On histone proteins, this approach was pioneered by Simon *et al.*[22] who utilized different alkylation reagents to generate mimics of differentially methylated lysines. Subsequently, this alkylation approach was extended to give a non-hydrolysable acetyl-lysine mimic.[23] To generate an acetyl lysine mimic that could be processed by HDAC enzymes, Li *et al.*[24] used a thiol-ene method to react *N*-vinyl-acetamide with cysteine containing proteins. Arginine methylation was accomplished via a similar alkylation method, but required the replacement of the guanidine group with acetamidine moiety.[25]

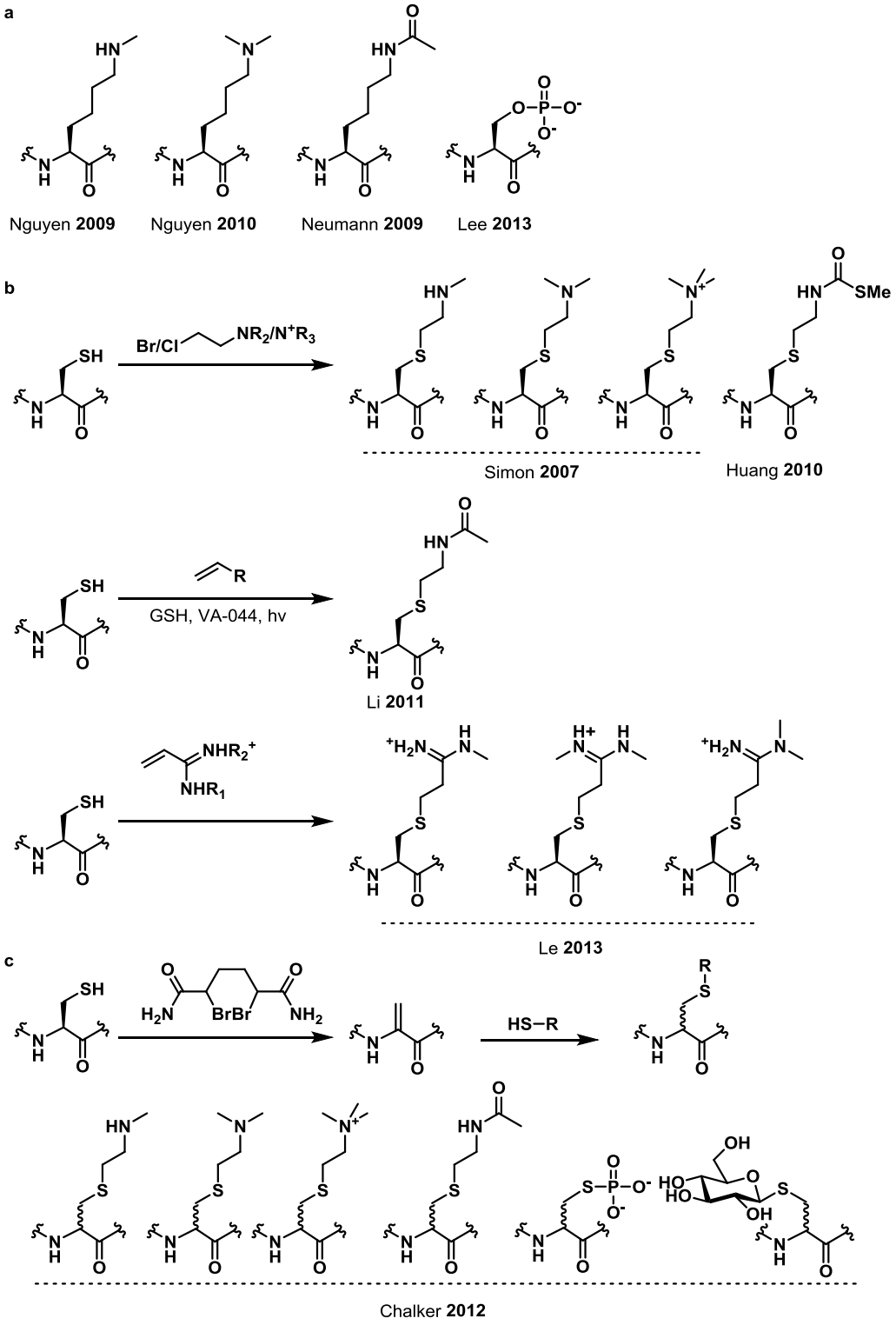


Figure 1. Summary of PTMs that can be incorporated by amber codon suppression (**a**), alkylation or thio-ene reaction (**b**), or by the Dha based strategy (**c**). Only papers reporting incorporation of the selective moieties into histones are given. References are Nguyen 2009[6], Nguyen 2010[7], Neumann 2009[8,9], Lee 2013[10,26], Simon 2007[22], Huang 2010[23], Li 2011[24], Le 2013[25], Chalker 2012[27]

In Oxford, a chemical route to dehydroalanine (Dha) via a di-alkylation / elimination approach has been developed (**Figure 2**). The addition of an appropriate thiol to the Dha Michael acceptor can yield a variety of PTM mimics (**Figure 1c**). This method allows the incorporation of additional PTMs (phosphoserine and O-GlcNAc-serine) when compared to other methodology. A major disadvantage of all cysteine based chemical methods is the replacement of the C-γ with a sulfur atom, resulting in a thioether link. This change results in reduced affinity of some proteins for the PTM mimic when compared to the natural modification.[22] Because the reaction proceeds through a planar Dha residue, the C-α stereochemistry is lost after Michael addition. This most likely yields a mixture of D- and L- stereochemistry at the site of modification. It was of interest to further explore the scope of this reaction by introducing multiple PTMs into a histone protein and validating the introduced PTMs by biological assays.

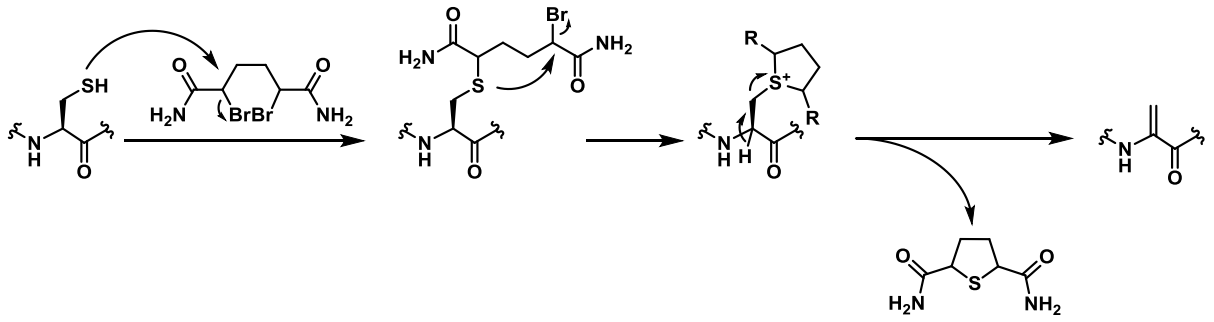


Figure 2. Proposed mechanism for the formation of Dha by reaction of cysteine with 2,5-dibromohexanediamide.

4.1 Results

4.1.1 Double modification

Histones protein tails are highly modified with 19 reported PTM sites in the first 37 residues of human histone H3. The interplay between different modifications is a major focus of current epigenetic research;[28-30] newly developed chemical methods can in theory enable the introduction of multiple PTMs. To test whether the Dha based modification methodology can be used to introduce multiple PTMs, a histone H3 mutant with lysine at position 4 and 79 replaced with cysteine was constructed. Both these lysine residues are reported to bear PTMs and are important for epigenetic

regulation.[31,32] The histone H3 variant was produced in *E. coli* and purified by SEC and ion-exchange chromatography.[33] The purity and identity of the protein was confirmed by SDS-PAGE and LC-MS (**Figure 3**). Upon addition of Ellman's reagent to the reduced double cysteine variant, a mass-shift corresponding to addition of two molecules of the reagent was observed, confirming the presence of two free cysteines in the histone (**Figure 12**).

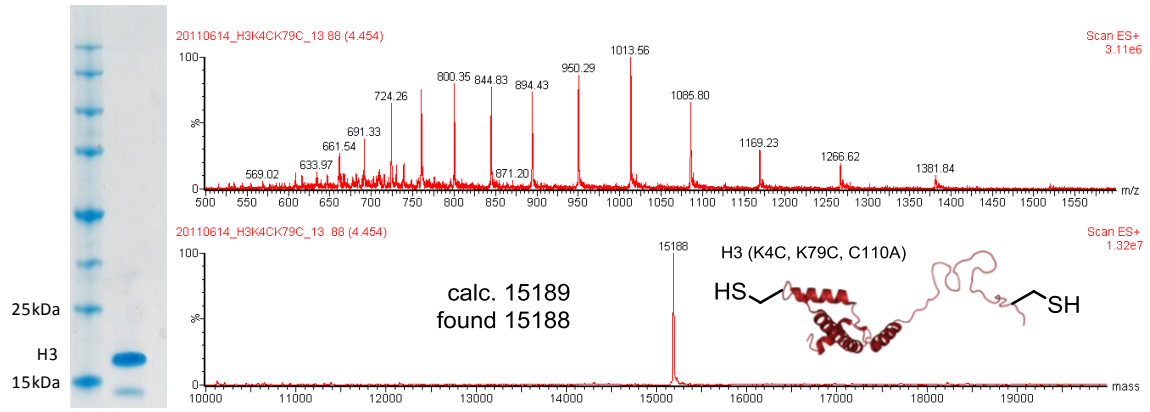


Figure 3. SDS-PAGE and LC-MS analysis of H3 (K4C K79C C110A).

The histone H3 variant (600 uM, 9 mg/mL) was then treated with the dibromide reagent **1** (50 mM, 90 eq). After incubation for 45 min at rt multiple intermediates were observed, with main products corresponding to one and two mono-alkylations (**Figure 4b**). After additional incubation at 37 °C for 2 h, complete conversion to the Histone H3 containing two Dha moieties was observed (**Figure 4c**). The conversion of both cysteines to Dha was confirmed by treatment of H3 (K4Dha K79Dha C110A) with Ellman's reagent. No addition of Ellman's reagent was observed, confirming complete reaction at both cysteine residues (**Figure 14**).

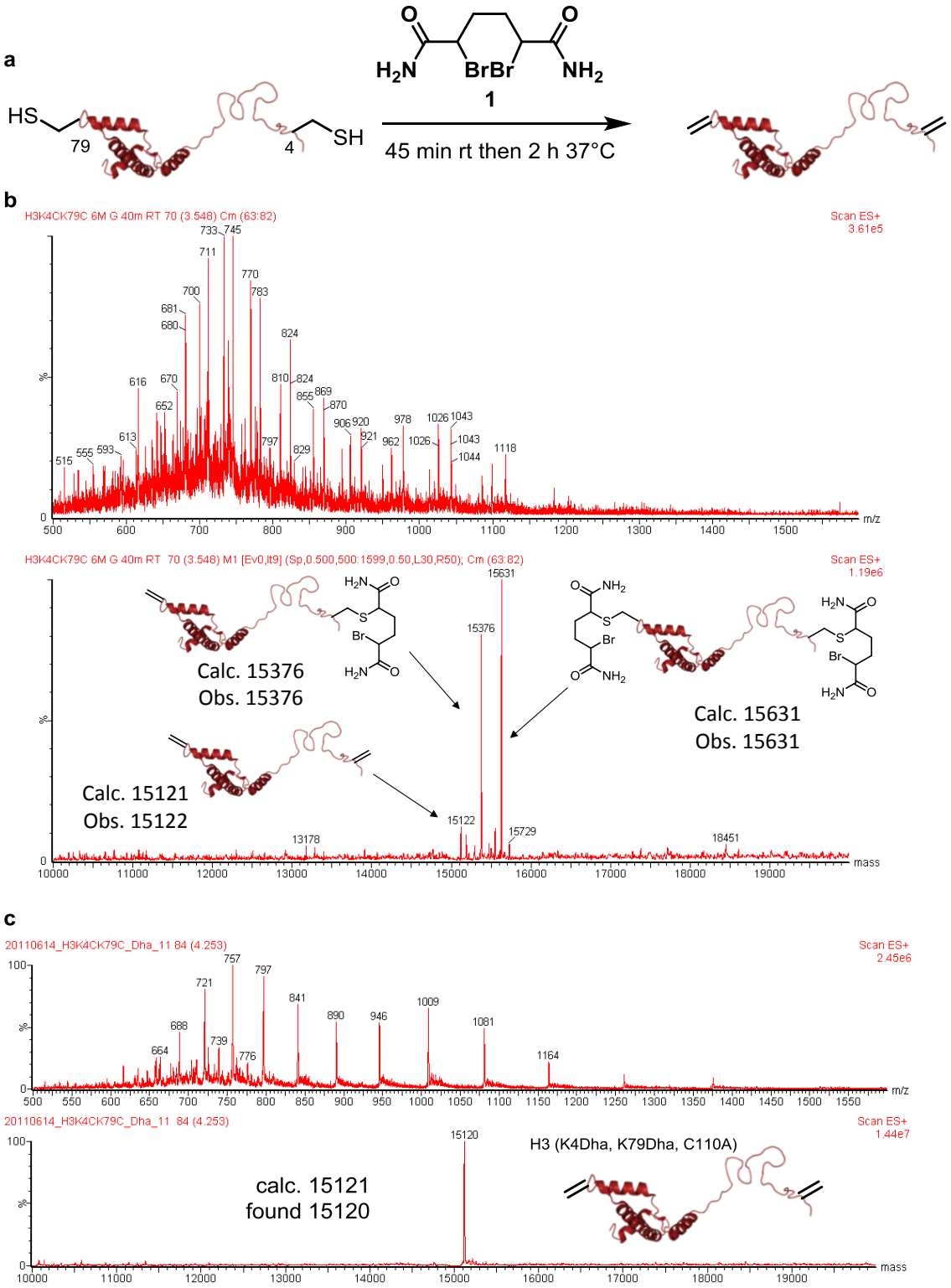


Figure 4. a Reaction scheme for conversion of the double cysteine H3 (K4C K79C C110A) to the Dha containing protein. **b** LC-MS analysis of the reaction after 45 min at room temperature. Multiple different species corresponding to intermediates of the Dha reaction were observed. **c** Final product after incubation at 37 °C for 2h.

Next, the Dha containing protein was treated with 2-(dimethylamino)ethanethiol or *N*-acetylcystamine. The reaction mixtures were incubated at rt for 1 h and subsequently analysed by LC-MS. Full conversion to the product containing two di-methyllysine mimics or two *N*-acetyllysine mimics were observed (**Figure 5**).

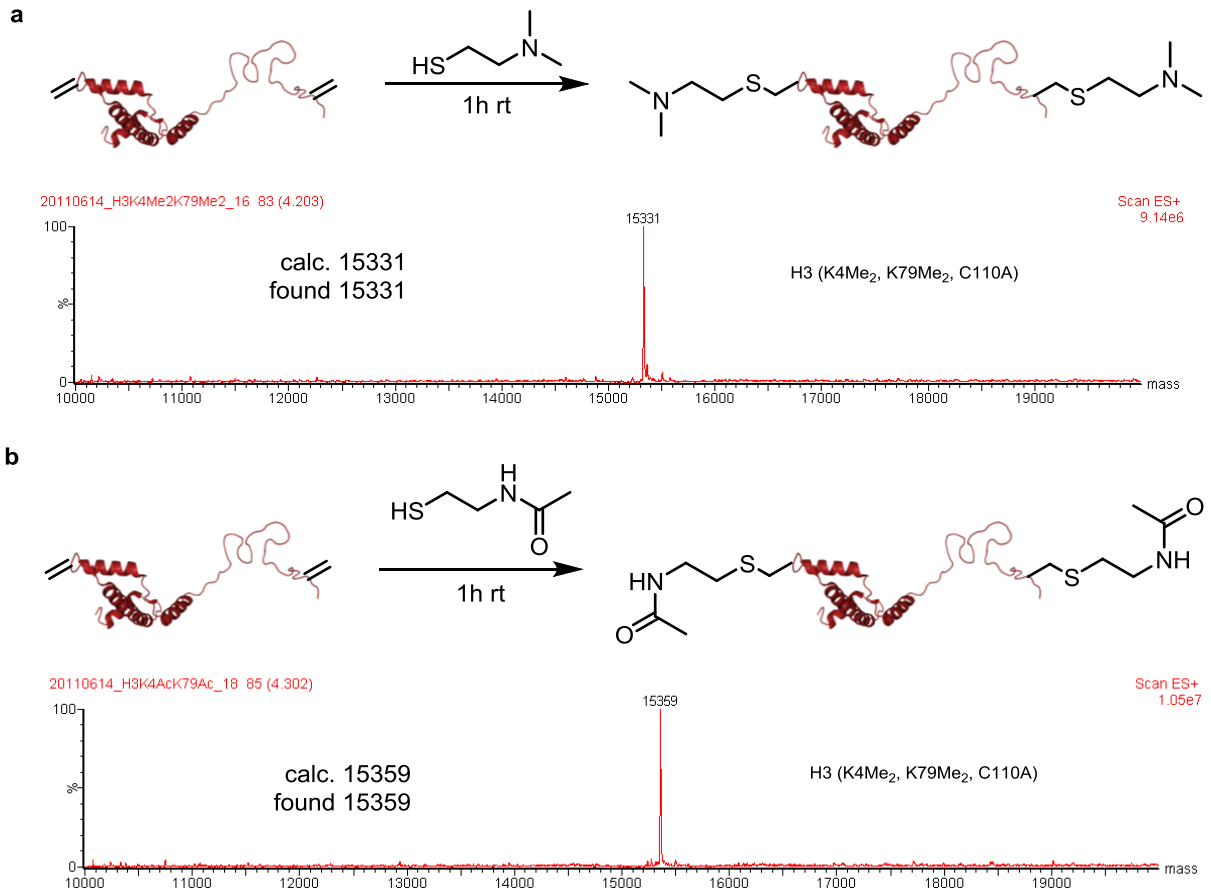


Figure 5. Generation of the double di-methyl (**a**) and double acetyl-lysine (**b**) containing histone H3 from H3 (K4Dha K79Dha C110A).

Thus, all reactions in the presented synthetic routes are selective and efficient enough to allow synthesis of homogeneously histone proteins containing two identical PTM mimics.

4.1.2 Biological confirmation

While the synthesized chemical PTMs are close to naturally occurring PTMs, the replacement of a carbon (for lysine mimics) or oxygen (for the serine mimics) by sulfur does result in a slightly different structure. Also, since the Michael addition to the planar Dha is likely to be non-stereoselective the

obtained product is likely to be a mixture of two diastereomers. To assess whether the PTM mimics behave like their natural counterparts, a variety of antibody, enzyme and 'reader' protein assays were used.

4.1.2.1 Western and dot-blotting

Most methods currently used to study the PTMs in biological context are antibody (Ab) based[34,35] (e.g. immunoprecipitation, chromatin immunoprecipitation and western blots). It was thus of interest to validate whether an Ab selective for the natural modification would retain the same specificity when the chemically generated PTMs are used. For this, histone H3 proteins containing a mono-, di- or tri-methyllysine mimic at position 9 were generated (reactions performed by Justin Chalker). The proteins were then simultaneously separated by SDS-PAGE, transferred to PVDF membranes and probed with Ab's against all three natural modifications (**Figure 6**). Specific binding of all Ab's towards their respective modification was observed with no cross-reactivity. The racemic nature of the obtained modifications does not appear to influence the specificity of the antibodies. This allows the use of these PTM mimics in e.g. antibody based assays for demethylation.

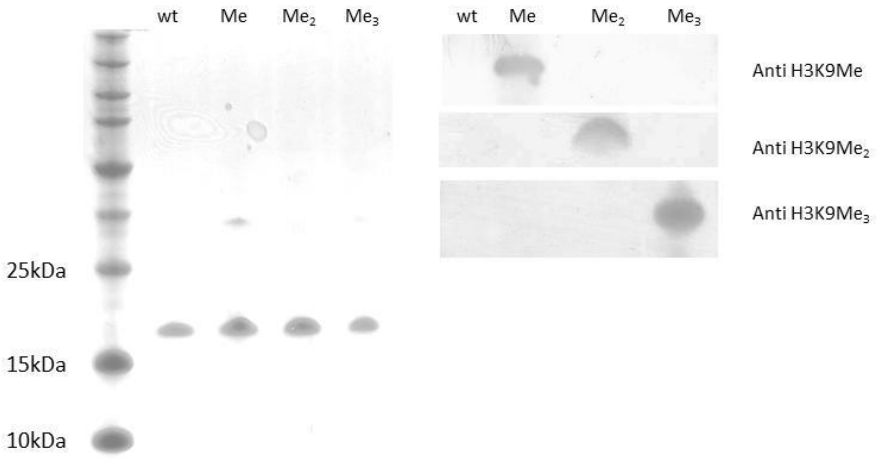


Figure 6. Comparison of antibody recognition of histone H3 K9Me$_x$ proteins. The coomassie stained loading control is shown on the left. The western blot with various antibodies raised against different natural H3 K9 methylation states is shown on the right. The methylation state of the lysine mimic is given above the lane. As control wt H3 (H3 C110A) was used.

Dot-blots are commonly used to assess Ab specificity and sensitivity.[34] While the sensitivity of Abs towards the methyl-lysine mimics have been tested by Simon *et al.*,[22] it was of interest to test the acetyllysine and phosphoserine Abs for specificity and selectivity towards our PTM mimics. As controls, commercially available peptides bearing the desired modification and wt histone H3 were used. Two different H3K9Ac antibodies were tested; one from Active Motif and one from Abcam (see methods for details). The Active Motif antibody (**Figure 7a**) showed strong binding to the chemically acetylated histone and weak binding to the wt histone H3, as expected. Weak binding was also observed for the K9Ac peptide. This H3 K9Ac peptide was obtained from Abcam and was the same as used to raise the Abcam acetyllysine Ab. Possibly, the peptide is different than the epitope used by Active Motif to raise their H3K9Ac Ab, resulting in weak binding. The Abcam Ab against H3 K9Ac (Figure 7b) showed strong binding to the K9Ac peptide control and significantly stronger binding to the synthetic H3 K9Ac than to the wt H3. From the intensities a 5-fold reduction in affinity was estimated when comparing the synthetic to the natural modification. This is similar to the 5-fold decrease in affinity reported for the methyllysine Abs.[22]

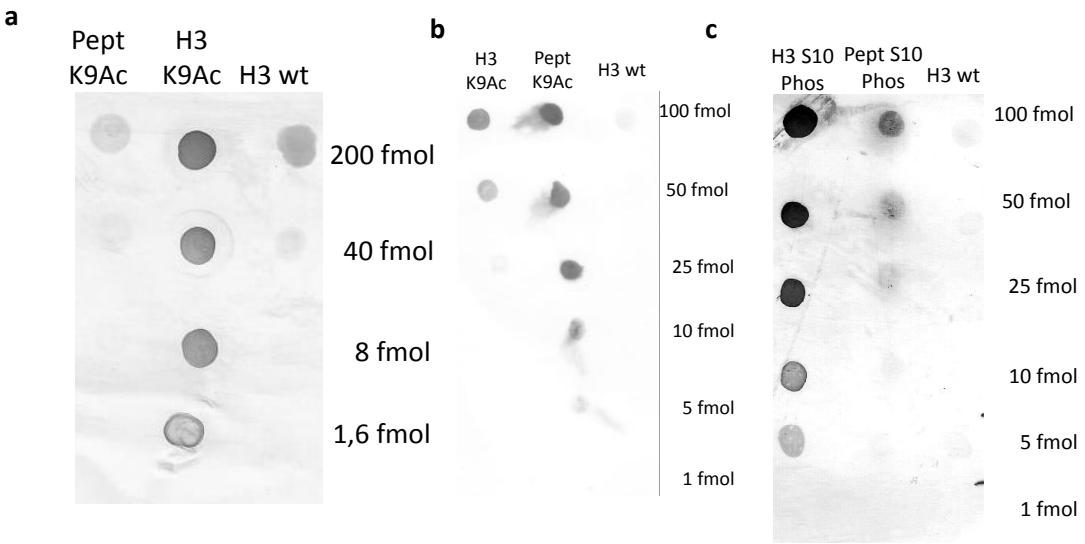


Figure 7. Dot blots of different PTM mimic containing H3 proteins with commercially available antibodies against the modification. **a** Dot blot of H3 K9Ac and the wt H3 and a K9Ac peptide control (Ab: Active-Motif 39137). **b** Dot blot of H3 K9Ac and the wt H3 and K9Ac peptide control (Ab: Abcam 39137). **c** Dot blot of H3 S10Ph and the wt H3 and H3 S10Ph control (Ab: Abcam ab5176).

To assess whether phosphocysteine is a good mimic for phosphoserine, dot-blot assays using an Ab raised to H3 S10Ph were performed. As control, an H3S10Ph peptide and the wt H3 were used. Surprisingly, the H3 S10Ph Ab showed stronger binding to the H3 S10Ph protein than the control peptide. These results confirm that the chemical PTM mimics are recognized by binding partners that interact with the natural modification.

4.1.2.2 HDAC SAHA inhibition

To further confirm functionality of the chemically introduced PTMs it was of interest to test if an enzyme that removes the natural modification would also recognize the chemical PTM. To investigate this, the ability of two histone deacetylases (HDAC1 and 2) to remove the histone H3 K9Ac mark was tested. The reactions were monitored by western blot and LC-MS. Upon incubation of H3 K9Ac with the respective HDAC a reduction in signal in the western blot was observed and LC-MS analysis showed ~50% deacetylation (initial experiments were performed by Justin Chalker).[27]

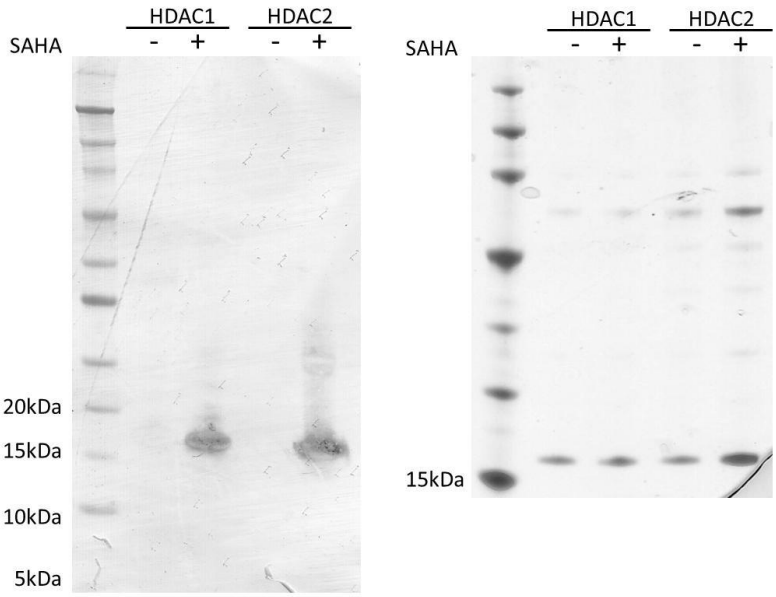


Figure 8. Western blot (left) and coomassie stained loading control (right) of the deacetylation reaction with HDAC 1 or 2 in the presence and absence of SAHA.

To confirm that the deacetylation is HDAC dependent, the H3K9Ac mimic was incubated with HDAC 1 or 2 in the presence of SAHA, a pan-HDAC inhibitor. Significantly reduced deacetylation was

observed, as indicated by an increase in the relevant signal on a western blot (Figure 8), upon addition of the known HDAC inhibitor SAHA, confirming that the observed deacetylation is HDAC activity dependent.

4.1.2.3 14-3-3ζ binding

It was also of interest to investigate the similarity between the natural and the synthetic PTMs by analysing the binding of a histone mark 'reader' module to the synthetic PTMs. For this purpose, the interaction of H3 chemically phosphorylated at serine 10 (H3 S10Ph) with 14-3-3ζ a reported 'reader' of this modification *in vitro* and *in vivo* was investigated.[36,37] An Alphascreen based competition assay was used to assess the interaction of the synthetic H3 S10Ph and 14-3-3ζ.[38] The Alphascreen assay is based on two beads (a donor and an acceptor bead) each with one interaction partner immobilized. When the two proteins interact, the beads are brought in close proximity. Upon irradiation at 680 nm the donor bead generated singlet oxygen which can diffuse only a limited distance before being quenched. When the beads are in close proximity because of an interaction between the two conjugated proteins, the singlet oxygen can diffuse to the acceptor bead, which induces emission at 615 nm. A biotinylated H3 peptide containing S10Ph (aa 1-20) was immobilised on a streptavidin donor bead and an his-tagged 14-3-3ζ on a Ni-NTA acceptor bead.

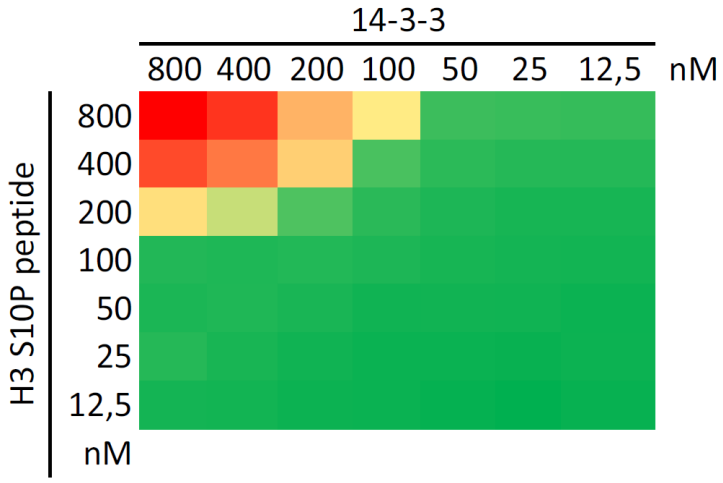


Figure 9. Heatmap of the observed fluorescence response from the alpha-screen with different concentrations of his-tagged 14-3-3ζ and biotinylated H3 S10Ph peptide.

First, titrations for both the H3S10Ph peptide and the 14-3-3ζ interaction partner were performed to find concentrations that give a robust signal (heatmap representation; **Figure 9**). A concentration between 800-400 nM of each interaction partner gave a robust response without reaching the plateau, making this range optimal for competition experiments.

For the competition experiment a concentration of 600 nM of each interaction partner was chosen. Increasing the amounts of the synthetic H3 S10Ph protein, or of H3 wt as a control, to the binding reaction were then added (Figure 10). Upon binding of the H3 S10Ph or the H3 wt control, a reduction in observed signal would be expected because some of the 14-3-3ζ is sequestered by H3.

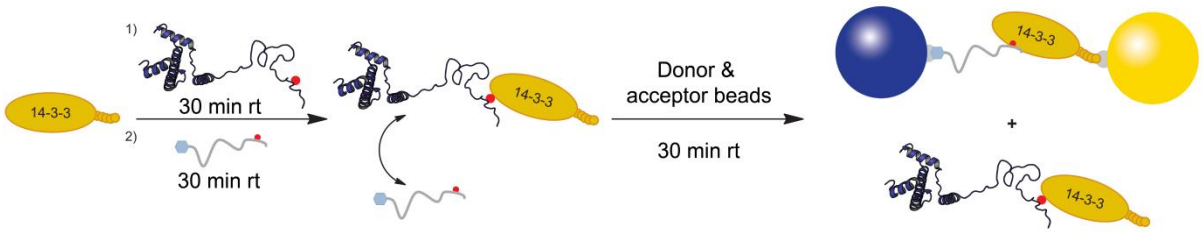


Figure 10. Schematic representation of the competition experiment. A biotinylated H3S10Ph peptide and an untagged H3 S10Ph protein are added to the his-tagged 14-3-3ζ. Upon addition of the alpha screen beads, only the peptide 14-3-3ζ complex results in signal. Increased binding of 14-3-3ζ to the untagged H3 S10Ph results in a decrease of the observed signal.

A significant decrease in signal intensity was observed upon adding two or five equivalents of the H3 S10Ph competitor (**Figure 11**). Addition of the H3 wt control results in a slight decreased signal intensity, but significantly less than the H3 S10Ph competitor. This suggests that 14-3-3ζ can specifically interact with the synthetic phosphoserine modification.

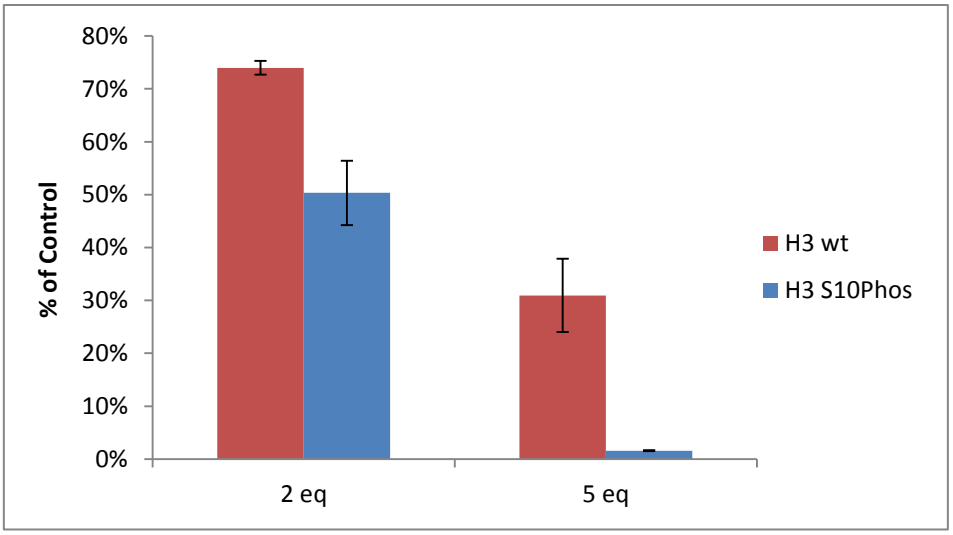


Figure 11. Signal intensity of different competitor experiments relative to the control without competitor.

4.2 Summary and outlook

It has been shown that a modification strategy utilizing the Dha intermediate is mild enough to incorporate multiple identical modifications into one protein, as exemplified by work on histone H3. PTM mimics obtained by this reaction are shown to be similar to the naturally occurring PTMs by a variety of methods. Antibodies raised to the natural PTM recognize the PTMs with high selectivity. Natural 'readers' such as the phosphoserine reader 14-3-3ζ can recognize a synthesized H3 S10Ph mimic and 'erasers' such as HDAC 1 and 2 can recognize and deacetylated a H3 K9Ac mimic.

Having shown that these mimics are functionally similar to the natural PTMs, one possible next step is the use of these modifications to investigate the molecular mechanisms of epigenetic control by PTMs that are difficult to probe by other chemical biology approaches. A strong advantage of the Dha based approach, is the ability to generate modified serine and threonine mimics and the lack of restrictions on the site of modification. This methodology was used to investigate the function of H2A T101 (Chapter 4) and H2B S121 (work carried out by Ritu Raj) GlcNAcylation.

While the presented methodology enables generation of histone H3 bearing two identical PTMs, considering the cross-talk between different histone modifications, a method that allows the

introduction of two different PTMs would be valuable. Towards this goal, orthogonal methods to introduce PTMs are currently being explored.

4.3 Experimental procedures

For mutagenesis, histone expression and LC-MS analysis of protein modification see general methods in Chapter 2. Oligonucleotides used to introduce the H3 K79C mutation are given in **Table 1**.

4.3.1 Used oligonucleotides

Name	Sequence (3'-5')	Obtained from
H3_K79C_F	GATCGCTCAGGACTTCTGCACCGACCTGCGCTTCC	Sigma
H3_K79C_R	GGAAGCGCAGGTCGGTGCAGAAGTCCTGAGCGATC	Sigma

Table 1. Used oligonucleotides

4.3.2 Histone H3 K4C K79C C110A modification

All reactions on this protein were performed in reaction buffer (RB; 3M Gd.HCl, 50 mM TRIS pH 8.0) under denaturing conditions. Previous reactions on histones were performed in phosphate buffer (100 mM sodium phosphate pH 8.0). This H3 variant proved not to be soluble in just phosphate buffer, thus buffers containing 3-4M guanidinium chloride was used in all experiments with this variant. In the original protocol by Luger *et al.*[39] histone proteins are dissolved in 7M guanidinium chloride before reconstitution of the nucleosome. Performing the reactions under denaturing conditions will not affect subsequent use of the histone and prevents precipitation of the histone during the procedure increasing the yield of modified protein.

The presence of two cysteines in the starting material was confirmed by the addition of Ellman's reagent. To H3 (10 µL, K4C, K79C, C110A) (10 mg/mL in RB) was added of an Ellman's reagent stock (2 µL, 10 mg/mL). The protein was shaken for 20 min at rt and subsequently analysed by LC-MS (**Figure 12**). Complete conversion into a product corresponding to two disulfide linked 5-thio-2-nitrobenzoic acid was observed confirming the presence of two free cysteines in the protein.

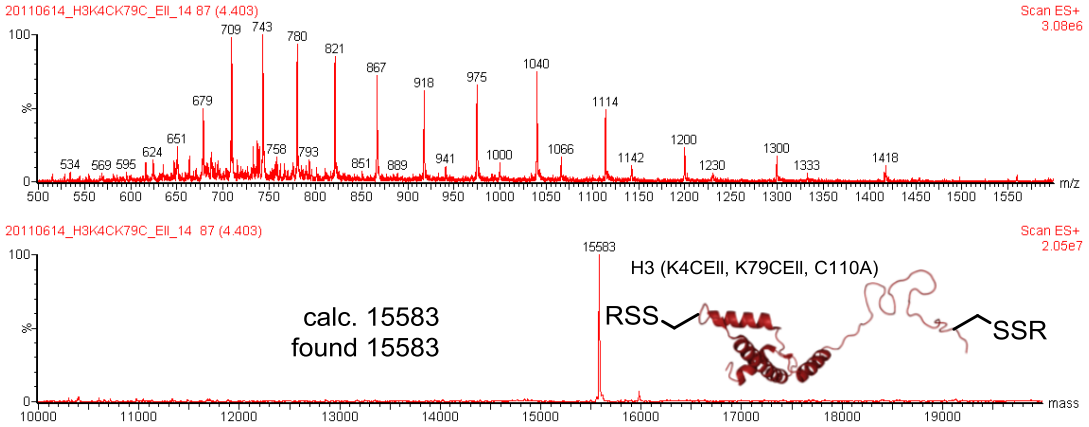


Figure 12. LC-MS analysis of the reaction of H3 (K4C K79C C110A) with Ellman's reagent.

4.3.2.1 H3 K4Dha K79Dha C110A synthesis

H3 (K4C K79C C110A; 10 mg) was dissolved in 500 µL of RB and DTT (30 mg) were added. The mixture was shaken at rt for 20 min and subsequently DTT removed by passing through a PD minitrap G-25 (equilibrated with RB), yielding a final concentration of 10 mg/mL.

To freshly reduced H3 (900 µL; K4C K79C C110A; 10 mg/mL in RB) was added di-bromide (100 µL; **1**) stock solution (0.5 M, final concentration 50 mM, 90 eq). The mixture was shaken first for 45 min at rt and then for 1.5 h at 37 °C. Then, the reaction was concentrated to 500 µL using a Vivaspin 500 concentrator (MWCO 10 kDa). Excess di-bromide was removed by PD minitrap G-25 (equilibrated and eluted with RB). The protein was analysed by LC-MS. Full conversion into the Dha containing protein H3 (K4Dha K79Dha C110A) was observed (**Figure 13**).

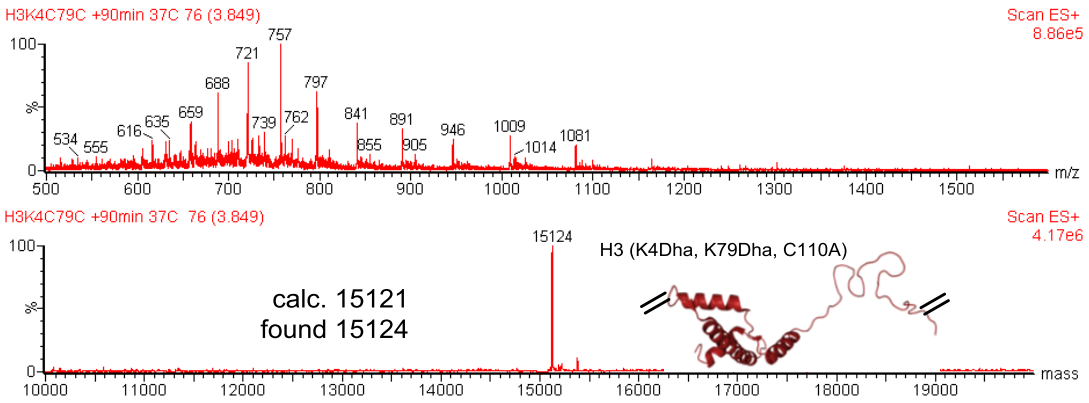


Figure 13. LC-MS analysis of H3 (K4Dha K79Dha C110A).

To confirm full consumption of the starting material, the Dha containing protein was treated with Ellman's reagent. To H3 (10 µL; K4Dha, K79Dha, C110A; 9 mg/mL in RB) was added Ellman's reagent (2 µL; 10 mg/mL stock). The protein was shaken for 20 min at rt and subsequently analysed by LC-MS (**Figure 14**). No addition of Ellman's reagent was observed confirming full consumption of the two cysteines.

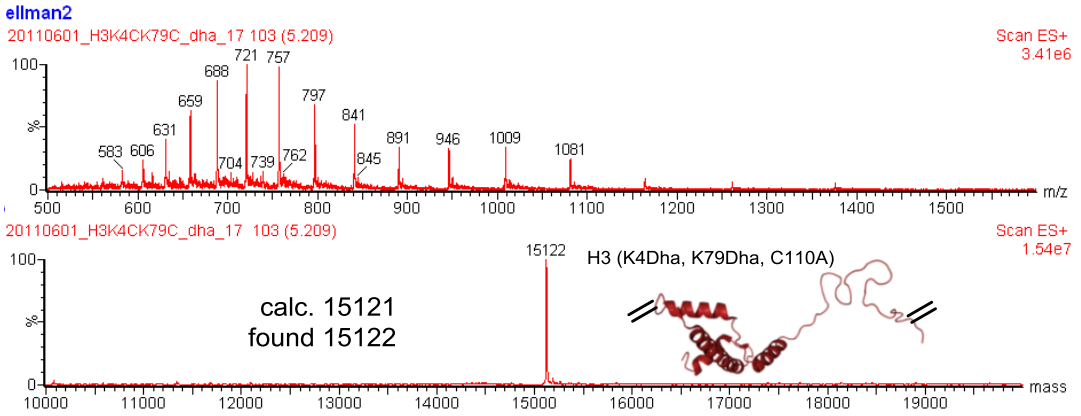


Figure 14. LC-MS analysis of the reaction of H3 (K4Dha K79Dha C110A) with Ellman's reagent.

4.3.2.2 H3 K4Me$_2$ K79Me$_2$ C110A synthesis

A stock solution of 2-(dimethylamino)ethanethiol was prepared by dissolving it in RB (30 mg in 30 µL RB), then neutralizing the solution with 22 µL 5 M NaOH. The stock solution (50 µL) was added to H3 (450 µL; K4Dha, K79Dha, C110A; 9 mg/mL in RB) and the resulting mixture was shaken at rt for 1 h. Small molecules were removed by PD minitrap G-25 (in RB). LC-MS analysis revealed full conversion to the doubly modified protein (**Figure 15**).

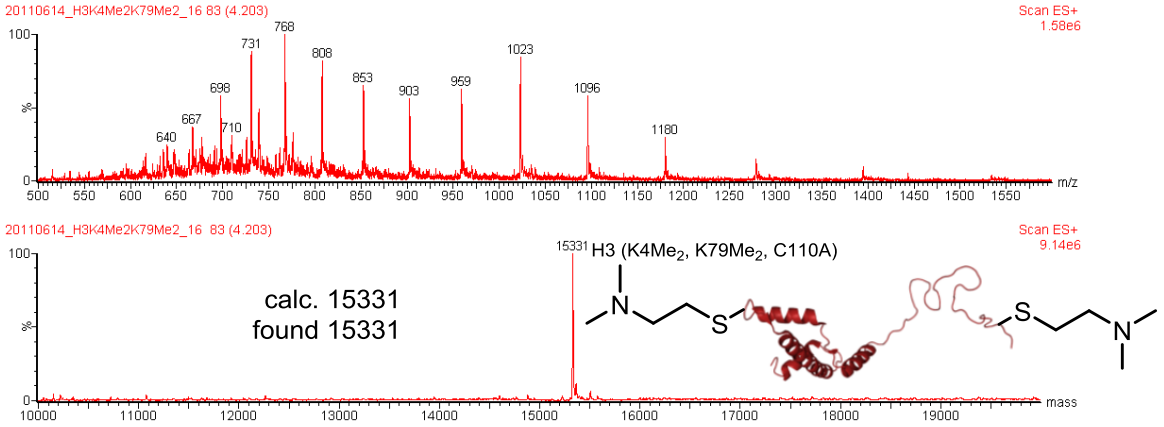


Figure 15. LC-MS analysis of H3 ($K4Me_2$ $K79Me_2$ C110A).

4.3.2.3 H3 K4Ac K79Ac C110A synthesis

To the Dha containing protein was (450 µL; 9 mg/mL in RB) was added *N*-acetylcysteamine (30 µL). The reaction was shaken at rt for 1 h. Then small molecules were removed by PD minitrap G-25 (in RB) and analysed by LC-MS, revealing full conversion to the doubly acetylated histone (**Figure 16**).

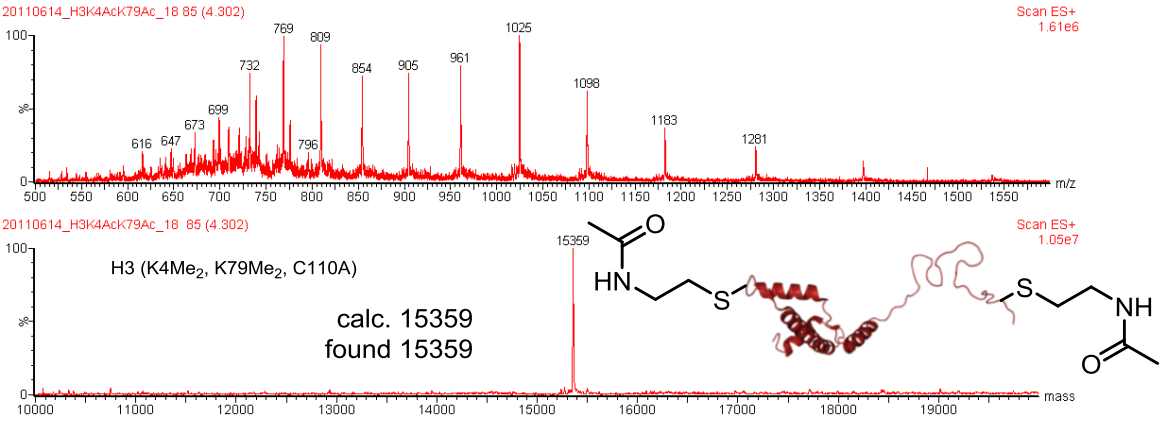


Figure 16. LC-MS analysis of H3 (K4Ac K79Ac C110A).

4.3.3 Western blots

Two aliquots of each protein were subjected to SDS-PAGE using a 4-12% NuPAGE pre-cast gel (Invitrogen) and MES or MOPS running buffer (Invitrogen). One aliquot was visualized using Instant Blue (coomassie) staining to confirm equivalent loading of all proteins. The other sample was used for western blot analysis. For the western blot, the protein was transferred to a PVDF membrane using an iBlot transfer apparatus (Invitrogen). The membrane was blocked in TBS-BT buffer (10 mL)

for 1 h at room temperature with rocking. After this, the membrane was washed three times with TBS-T buffer and then incubated with the primary Ab (dilutions are given below for the Ab's used) was added in TBS-BT. The membrane was incubated with the antibody solution for 1 hour at room temperature with rocking. After incubation, the membrane was washed with TBS-T buffer (5× 20 mL) and then incubated with the secondary Ab (see below for dilution) in TBS-BT for one hour. The blot was then washed with TBS-T (5x 20 mL) and water (3x 20 mL). The membrane was then developed using BCIP/NBT liquid substrate system (Sigma-Aldrich, B1911). The used antibodies are given in the next section.

4.3.4 Dot blot assays

A PVDF membrane (0.2 µm pore size; Invitrogen) was wetted with Methanol, washed with water (2x 10mL) and soaked in PBS. 2 µL of serial dilutions of modified H3, H3 wt and a control peptide (see below for used peptides) were spotted on the membrane and allowed to dry overnight. The dry membrane was again wetted with Methanol, washed with water (2x 10 mL) and blocked in TBS-BT for an hour at room temperature while rocking. Then the primary antibody was added and the membrane was incubated for 1 hour at room temperature while rocking. Then the membrane was washed (5x 10 mL TBS-T) and incubated with the second antibody for additional hour at room temperature while rocking. After incubation the membrane was washed with TBS-T (5x 10 mL) and water (3x 10 mL). The membrane was then developed using BCIP/NBT liquid substrate system (Sigma-Aldrich, B1911).

Primary Ab	Number	Manufacturer	Dilution used
Monoclonal Anti-polyHistidine–Alkaline Phosphatase antibody produced in mouse	A5588	Sigma	1:6000
Anti-Histone H3 (phospho S10), rabbit polyclonal	ab5176	Abcam	1:1000
Anti-Histone H3 (acetyl K9) (pAb; rabbit)	39137	Abcam	1:500
Histone H3K9ac antibody (pAb; rabbit)	39137; lot no 01008001	Active-Motif	1:500
Anti-monomethyl-Histone H3 (Lys9) (pAb; rabbit)	07-450	Millipore	1:1000
Anti-dimethyl-Histone H3 (Lys9) (pAb; rabbit)	07-441	Millipore	1:500
Anti-Histone H3 (tri methyl K9) (pAb; rabbit)	ab8898	Abcam	1:1000

Secondary Ab	Number	Manufacturer	Dilution used
anti-rabbit IgG produced in goat, conjugated with	A3687	Sigma-Aldrich	1:1000

alkaline phosphatase		
Peptide Name	**Number**	**Manufacturer**
Human Histone H3 (acetyl K9) peptide	ab16635	Abcam
Human Histone H3 (phospho S10) peptide	ab11477	Abcam

Table 2. Antibodies used in this study.

4.3.5 HDAC assay[27]

Two samples were prepared for each HDAC assay. Samples were analysed by SDS-PAGE and western blotting using the Abcam K9Ac Ab. Proteins were obtained from Active Motif (HDAC1, 31342; HDAC2, 31343). SAHA was obtained from Cayman chemicals (10009929) and was dissolved in DMSO to a final concentration of 100 mM. Reactions were performed in 25 mM Tris, pH 8.0, 138 mM NaCl using components described in **Table 3**.

HDAC1 (SAHA)			**HDAC1 (control)**		
	Vol			Vol	
H3 K9Ac mimic	5	µL	H3 K9Ac mimic	5	µL
SAHA stock	0.25	µL	DMSO	0.25	µL
recombinant HDAC1	10	µL	recombinant HDAC1	10	µL
reaction buffer	9.75	µL	reaction buffer	9.75	µL

HDAC2 (SAHA)			**HDAC2 (control)**		
	Vol			Vol	
H3 K9Ac mimic	5	µL	H3 K9Ac mimic	5	µL
SAHA stock	0.25	µL	DMSO	0.25	µL
recombinant HDAC12	10	µL	recombinant HDAC2	10	µL
reaction buffer	9.75	µL	reaction buffer	9.75	µL

Table 3. Components of the reactions for the SAHA inhibition studies of HDAC depentend deacetylation of a H3 K9Ac mimic.

All samples were mixed by pipetting and then incubated at 37 ºC for 12 hours. After this time, a 10 µL sample of each reaction was diluted with 6 x SDS-PAGE loading buffer, heated at 95 ºC for 5 minutes, and then analysed by SDS-PAGE and western blot.

4.3.6 14-3-3 alpha screen

AlphaScreen assays were performed as described in the manufacturer's protocol (AlphaScreen Histidine (Nickel chelate) Detection Kit, PerkinElmer, 6760619C) with minor modifications. All

reagents were diluted in the recommended buffer (25 mM HEPES, 100 mM NaCl, 0.1% BSA; pH = 7.4) and allowed to equilibrate to room temperature prior to addition to plates. A biotinylated 20-mer corresponding to the first 20 amino acids from H3 and a phosphoserine at position 10 was prepared using SPS (Biotin-ARTKQTARKS(p)TGGKAPRKQL). The 14-3-3ζ protein was obtained commercially as a his-tagged fusion protein (Abcam, ab87361).

For the dose response curve 4 x concentrated stocks were prepared for each peptide and 14-3-3ζ. 5 uL of each stock were mixed on a 384-well white ProxiPlate (PerkinElmer) sealed an incubated at room temperature for 1 hour. AlphaScreen streptavidin-conjugated donor and nickel chelate-conjugated acceptor beads were prepared as 2x stock (40 μg/mL) and 10 μL was added to the wells to give a final concentration of 20 μg/ml for each bead and incubated for a further 1 h in the dark at 22 °C. The plates were analysed using an Envision (PerkinElmer) plate reader. The obtained heatmap is shown below.

14-3-3 / peptide	800	400	200	100	50	25	12,5	nM
800	812583	640673	219088	35030	10021	9590	9212	
400	568400	417951	128263	11635	8001	7244	7120	
200	75750	26699	12216	7925	6335	5661	5637	
100	6942	6458	6991	6385	5664	5318	5074	
50	6147	6607	5863	4733	4855	4734	4227	
25	7281	5700	4755	4084	4024	3785	4323	
12,5	5285	4999	4332	4044	3504	2805	3659	
nM								

Figure 17. Heatmap of the obtained response from the alpha screen for the titration of biotinylated H3S10Ph peptide and his-tagged 14-3-3ζ.

For the competition assays 3 μL 14-3-3 zeta (4 μM stock; final conc. 600 nM) was mixed with the competitor (1.2 or 3 μM final concentration) and buffer was added to give 7 μL in a 384-well plate. The plate was sealed and incubated for 15 min at room temperature. Then 3 μL biotinylated H3S10Phos peptide was added (4 μM stock; final conc. 600 nM), the plate was resealed and incubated for 45 min at room temperature. AlphaScreen streptavidin-conjugated donor and nickel chelate-conjugated acceptor beads were prepared as 2x stock (50 μg/mL) and 10 μL were added to

the wells to give a final concentration of 25 μg/ml and incubated for a further 1 h in the dark at 22 °C. The plates were analysed using an Envision (PerkinElmer) plate reader. Measurements were performed in duplicates. The background was subtracted from the obtained reading and the values were normalized with respect to a control containing no inhibitor. Data for the measurements are shown below.

2 eq	H3 S10P	H3	cont	blank
	526500	874636	1337653	9578
	691425	905489	1041043	10407
-blank	516641	864777	1327794	
	681566	895630	1031184	
ratio	0,43417	0,726734		
	0,572769	0,752662		
Mean	0,5035	0,7397		
Sd	0,069299	0,012964		

5 eq	H3 S10P	H3	cont	blank
	14263	242633	647460	3948
	12940	144928	572601	5088
-blank	10315	242633	647460	
	7852	144928	572601	
ratio	0,015931	0,374746		
	0,013713	0,253105		
Mean	0,0148	0,3139		
Sd	0,001109	0,060821		

Table 4. Results for the alpha screen competition assay.

4.4 References

(1) Fierz, B.; Muir, T. W. *Nat. Chem. Biol.* **2012**, 417.
(2) Allis, C. D.; Muir, T. W. *ChemBioChem* **2011**, *12*, 264.
(3) Fierz, B.; Kilic, S.; Hieb, A. R.; Luger, K.; Muir, T. W. *J. Am. Chem. Soc.* **2012**, *134*, 19548.
(4) Bannister, A. J.; Kouzarides, T. *Cell Res.* **2011**, *21*, 381.
(5) Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem.* **2010**, *79*, 413.
(6) Nguyen, D. P.; Garcia Alai, M. M.; Kapadnis, P. B.; Neumann, H.; Chin, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 14194.
(7) Nguyen, D. P.; Alai, M. M. G.; Virdee, S.; Chin, J. W. *Chem. Biol.* **2010**, *17*, 1072.
(8) Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol.* **2008**, *4*, 232.
(9) Neumann, H.; Hancock, S. M.; Buning, R.; Routh, A.; Chapman, L.; Somers, J.; Owen-Hughes, T.; van Noort, J.; Rhodes, D.; Chin, J. W. *Mol. Cell* **2009**, *36*, 153.
(10) Lee, S.; Oh, S.; Yang, A.; Kim, J.; Söll, D.; Lee, D.; Park, H.-S. *Angew. Chem. Int. Ed.* **2013**, *52*, 5771.
(11) Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H.-R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. a.; Wang, H. H.; Carr, P. a.; Mosberg, J. a.; Rohland, N.; Schultz, P. G.; Jacobson, J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J. *Science* **2013**, *342*, 357.

(12) Kee, J.-M.; Villani, B.; Carpenter, L. R.; Muir, T. W. *J. Am. Chem. Soc.* **2010**, *132*, 14327.
(13) Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. *Science* **1994**, *266*, 776.
(14) Nilsson, B. L.; Soellner, M. B.; Raines, R. T. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 91.
(15) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci.* **1998**, *95*, 6705.
(16) Kee, J.-M.; Muir, T. W. *ACS Chem. Biol.* **2011**.
(17) Muralidharan, V.; Muir, T. W. *Nat. Meth.* **2006**, *3*, 429.
(18) Tolbert, T. J.; Wong, C.-H. *Angew. Chem. Int. Ed.* **2002**, *41*, 2171.
(19) Evans, T. C.; Benner, J.; Xu, M.-Q. *J. Biol. Chem.* **1999**, *274*, 3923.
(20) Muir, T. W. *Synlett* **2001**, *2001*, 0733.
(21) Erlanson, D. A.; Chytil, M.; Verdine, G. L. *Chem. Biol.* **1996**, *3*, 981.
(22) Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M. *Cell* **2007**, *128*, 1003.
(23) Huang, R.; Holbert, M. A.; Tarrant, M. K.; Curtet, S.; Colquhoun, D. R.; Dancy, B. M.; Dancy, B. C.; Hwang, Y.; Tang, Y.; Meeth, K.; Marmorstein, R.; Cole, R. N.; Khochbin, S.; Cole, P. A. *J. Am. Chem. Soc.* **2010**, *132*, 9986.
(24) Li, F.; Allahverdi, A.; Yang, R.; Lua, G. B. J.; Zhang, X.; Cao, Y.; Korolev, N.; Nordenskiöld, L.; Liu, C.-F. *Angew. Chem. Int. Ed.* **2011**, *50*, 9611.
(25) Le, D. D.; Cortesi, A. T.; Myers, S. A.; Burlingame, A. L.; Fujimori, D. G. *J. Am. Chem. Soc.* **2013**, *135*, 2879.
(26) Park, H.-S.; Hohn, M. J.; Umehara, T.; Guo, L.-T.; Osborne, E. M.; Benner, J.; Noren, C. J.; Rinehart, J.; Söll, D. *Science* **2011**, *333*, 1151.
(27) Chalker, J. M.; Lercher, L.; Rose, N. R.; Schofield, C. J.; Davis, B. G. *Angew. Chem. Int. Ed.* **2012**, *51*, 1835.
(28) Wang, Z.; Patel, D. J. *J. Biol. Chem.* **2011**, *286*, 18363.
(29) Lee, J.-S.; Smith, E.; Shilatifard, A. *Cell* **2010**, *142*, 682.
(30) Hart, G. W.; Slawson, C.; Ramirez-Correa, G.; Lagerlof, O. *Annu. Rev. Biochem.* **2011**, *80*, 825.
(31) Nguyen, A. T.; Zhang, Y. *Genes Dev.* **2011**, *25*, 1345.
(32) Kouzarides, T. *Cell* **2007**, *128*, 693.
(33) Dyer, P. N.; Edayathumangalam, R. S.; White, C. L.; Bao, Y.; Chakravarthy, S.; Muthurajan, U. M.; Luger, K. In *Methods Enzymol.*; Allis, C. D., Carl, W., Eds.; Academic Press: 2003; Vol. Volume 375, p 23.
(34) Egelhofer, T. A.; Minoda, A.; Klugman, S.; Lee, K.; Kolasinska-Zwierz, P.; Alekseyenko, A. A.; Cheung, M.-S.; Day, D. S.; Gadel, S.; Gorchakov, A. A.; Gu, T.; Kharchenko, P. V.; Kuan, S.; Latorre, I.; Linder-Basso, D.; Luu, Y.; Ngo, Q.; Perry, M.; Rechtsteiner, A.; Riddle, N. C.; Schwartz, Y. B.; Shanower, G. A.; Vielle, A.; Ahringer, J.; Elgin, S. C. R.; Kuroda, M. I.; Pirrotta, V.; Ren, B.; Strome, S.; Park, P. J.; Karpen, G. H.; Hawkins, R. D.; Lieb, J. D. *Nat. Struct. Mol. Biol.* **2011**, *18*, 91.
(35) Bock, I.; Dhayalan, A.; Kudithipudi, S.; Brandt, O.; Rathert, P.; Jeltsch, A. *Epigenetics* **2011**, *6*, 256.
(36) Macdonald, N.; Welburn, J. P. I.; Noble, M. E. M.; Nguyen, A.; Yaffe, M. B.; Clynes, D.; Moggs, J. G.; Orphanides, G.; Thomson, S.; Edmunds, J. W.; Clayton, A. L.; Endicott, J. A.; Mahadevan, L. C. *Mol. Cell* **2005**, *20*, 199.
(37) Winter, S.; Fischle, W.; Seiser, C. *Cell Cycle* **2008**, *7*, 1336.
(38) Du, Y.; Khuri, F. R.; Fu, H. *Current chemical genomics* **2008**, *2*, 40.
(39) Luger, K.; Rechsteiner, T. J.; Richmond, T. J. In *Methods Enzymol.*; Paul M. Wassarman, A. P. W., Ed.; Academic Press: 1999; Vol. Volume 304, p 3.

5 The effect of H2A T101 GlcNAcylation on nucleosome structure

5.1 Introduction

The addition of *O*-linked β-*N*-acetylglucosamine (GlcNAc) to serine and threonine residues,[1] as catalysed by *O*-linked *N*-acetylglucosamine transferase (OGT), is a prevalent posttranslational modification (PTM)[2] in mammalian cells. Recently, this modification has been detected on histone proteins,[3,4] implying that it may function in transcriptional regulation.[2] To date, a biological function has only been described for one of the histone GlcNAc modifications; Histone H2B S112 GlcNAcylation which has been shown to activate transcription by increased H2B ubiquitination.[5] OGT has been shown to interact with polycomb proteins[6] responsible for transcriptional repression via H3K27 methyation and the COMPASS/Set4 complex,[7] an activating complex that can catalyse H3K4 methylation. This seemingly conflicting involvement in both transcriptional repression and activation, and the observation that OGT is essential for development,[8,9] point towards an important and complicated role for OGT in transcriptional regulation.

To gain better understanding of the influence of GlcNAcylation on nucleosome structure and stability, access to homogeneously GlcNAcylated proteins at a single site is necessary. OGT is a promiscuous enzyme with many reported GlcNAcylation sites on histones.[3] In addition, not all GlcNAcylation sites on histones have been discovered yet. For Histone H2A only one GlcNAcylation site on Thr-101 has been confirmed by mass spectrometry. Upon mutation of this residue to alanine, GlcNAcylation of Histone H2A is still observed by western-blot, suggesting that there are addition GlcNAcylation sites on the protein.[3] Similarly, results were obtained for histone H2A and H4.[3] Histone H3 is also shown to be GlcNAcylated, possibly on Ser-10,[4] but no mass spectrometry evidence for the site of modification on H3 has been reported yet.

In vitro treatment of histones or nucleosomes with OGT results in multiple modifications on the proteins.[5] Chemical posttranslational modification is an attractive alternative since it allows site-

specific and complete installation of PTM mimics. In the work in this chapter, methodology developed in Oxford was used to generate homogenously GlcNAcylated histone H2A at site T101, a reported natural modification site. The product of this modification strategy is an O-GlcNAc mimic effectively replacing serine-O-GlcNAc with cysteine-S-GlcNAc. This thio-ether linked O-GlcNAc mimic has often been used in chemical biology research, as it has very similar conformational properties as the natural modification and increased stability.[10] GlcNAcyation was found to destabilize nucleosome structure, revealing a potentially new mechanism by which OGT can help to shape chromatin structure.

5.2 Results

To study the effect of GlcNAcylation on nucleosome stability, a variety of assays previously used to characterize nucleosome stability were employed. These included size exclusion chromatography (SEC),[11,12] circular dichroism (CD),[13] thermal melting monitored by CD[14] and differential scanning fluorimetry (DSF)[15] and non-denaturing mass spectrometry (nMS),[16] both of sub-complexes and of whole nucleosomes.

5.2.1 Chemical modification

In Chapter 3, the chemical addition of PTMs to histone H3 was described as an efficient method to generate site selective GlcNAcylated H3. In this chapter, the extension of this methodology to histone H2A is described using analogous reaction conditions as used for H3. First, a single-site cysteine mutant of H2A was generated, H2A T101C. The protein was produced in *E. coli* and purified from inclusion bodies by SEC and ion-exchange chromatography. Treatment of H2A T101C with 100 eq **1** gave complete conversion (>95%) to the Dha containing protein after 30 min at rt and a further 3 h at 37 °C. Subsequent addition of 1000 eq GlcNAc-SH and incubation at rt for 3.30 h yielded homogeneously modified histone H2A at site 101 as confirmed by ESI-MS.

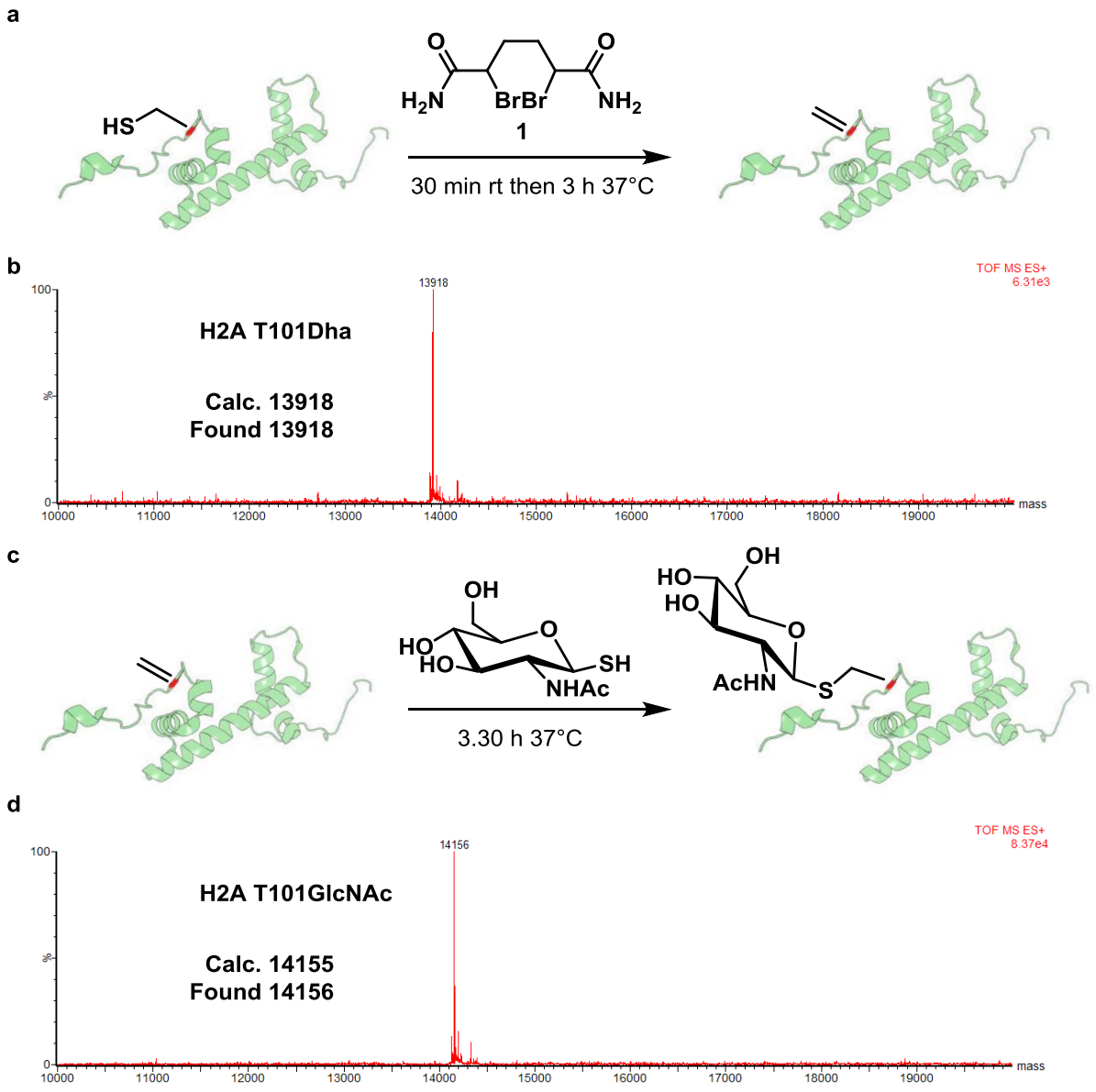


Figure 1. Generation of H2A T101 GlcNAc. **a** Reaction of H2A T101C with 100 eq **1** shows complete conversion to H2A T101Dha. **b** Deconvoluted ESI-MS spectra of the Dha containing H2A after the first reaction **c** Reaction of H2A T101Dha with GlcNAc-SH yields homogenously GlcNAcylated H2A. **d** Deconvoluted ESI-MS spectra of the GlcNAc containing H2A.

5.2.2 Dimer and octamer reconstitution

With the modified histone H2A in hand, first the effect of T101GlcNAc on H2A-H2B dimer stability was then investigated. Wt and GlcNAcylated H2A/B dimers were refolded by combining either wt or GlcNAcylated H2A with wt H2B under denaturing conditions and dialysis into 2M NaCl containing buffer. The obtained complexes were analysed and purified by SEC. The wt and GlcNAcylated H2A/H2B dimer behave identically on the size exclusion column, eluting as one peak at 81 mL (**Figure**

2a). By comparison, the larger H3/H4 tetramer elutes as one slightly asymmetric peak at 68 mL (**Figure 2b**), probably due to the dimer-tetramer equilibrium under these conditions.

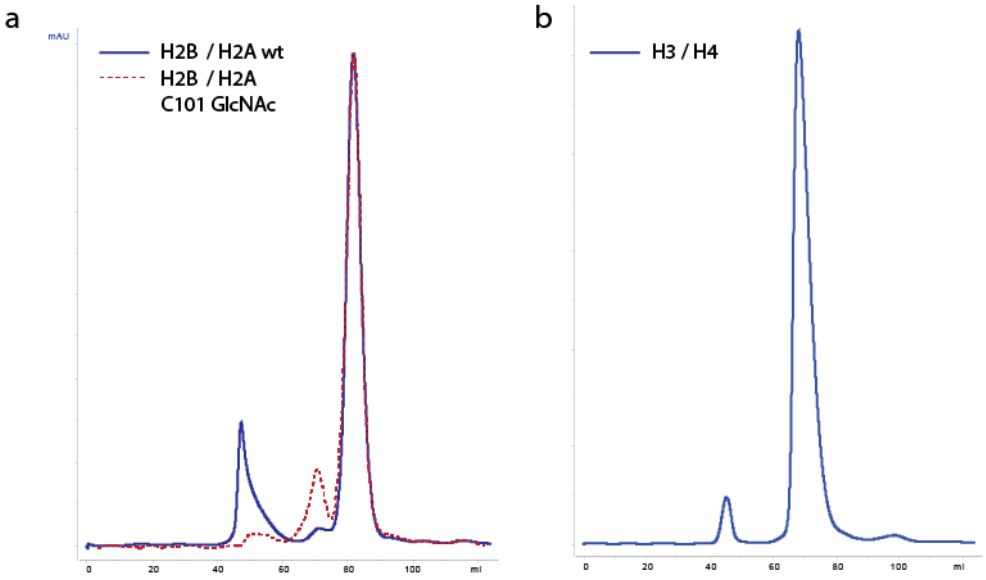


Figure 2. SEC purification of different histon multimers. **a** Comparison of the SEC UV traces for H2A/B wt and H2A/B GlcNAc. **b** SEC trace for the purification of the H3/H4 tetramer.

With the separate dimers and tetramers being reconstituted, the ability of GlcNAcylated H2A/B dimer to form histone octamers was investigated. For this, the purified dimers were combined with the H3/H4 tetramer and incubated for 1 h at 4 °C. 1.1 eq of the dimer was used, because excess dimer is easy to separate by SEC which should result in homogeneous purified GlcNAcylated octamer for further studies. When the wt dimer is combined with the H3/H4 tetramer, a larger molecular weight species eluting at 66 mL (**Figure 3a** peak **1**) is observed, corresponding to the intact octamer (Figure 3a). Upon combining GlcNAcylated H2A/B dimer with the H3/H4 tetramer, two separate peaks one eluting at 71 mL corresponding to the tetramer (**Figure 3a** peak **2**) and one eluting at 81 mL corresponding to the GlcNAcylated dimer (**Figure 3a** peak **3**) were observed. The composition of the peaks was confirmed by SDS-PAGE (**Figure 3b**). These results suggest that while GlcNacylation does not affect H2A/B dimer formation, it does inhibit octamer formation. This could be due to a destabilization at the tetramer / dimer interface.

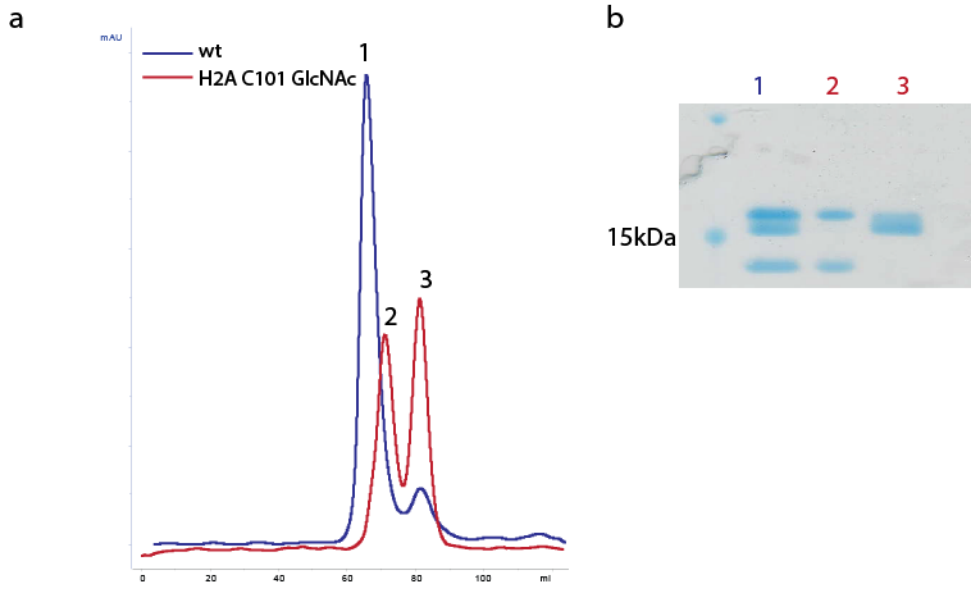


Figure 3. SEC and SDS-PAGE analysis of octamer reconstitutions with wt and GlcNAcylated H2A/B dimers and at H3/H4 tetramer. **a** SEC UV traces of the H2A/B wt (blue) and GlcNAc (red) octamer reconstitutions. **b** SDS-PAGE analysis of the indicated fractions.

5.2.3 Dimer stability

To further assess the effect of GlcNAcylation on the H2A/B dimer structure, CD spectra and melting curves of both the wt and GlcNAcylated dimer were measured. The CD spectra were identical, confirming that the GlcNAcylation does not influence dimer structure (**Figure 4a**). This is expected, as the modification site is not in the α-helical core of the protein but in a loop which folds into a short β-sheet segment upon association with the H3/H4 tetramer but is probably disordered in the free dimer. To confirm the structural integrity of the dimers, melting temperatures were measured by CD. Very similar melting temperatures (GlcNAc, 51.1 ± 0.1 °C; wt, 52.9 ± 0.1 °C; **Figure 4b** and **c**) were obtained.[13] This again confirms that the structure and stability is not significantly affected by H2A T101 GlcNAcylation.

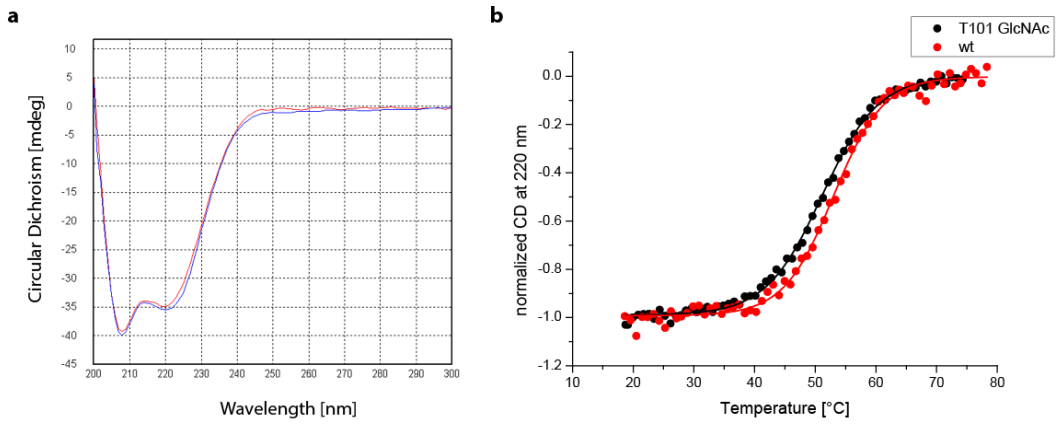


Figure 4. CD analysis of wildtype and GlcNAcylated H2A/H2B dimers. Data was processed and analysed using Pro-Data (Chirascan). **a** Overlay of the CD spectra for wildtype (red) and GlcNAcylated (blue) H2A / H2B dimers. **b** Melting profile of GlcNAcylated (black) and the wt (red) H2A / H2B dimer measured by CD.

5.2.4 Nucleosome reconstitution and stability

The isolated histone octamer is not proposed to be a relevant species *in vivo*.[17] While a decrease in stability in isolated octamer could point towards a similar effect in chromatin, the difference in the environment is substantial. Thus, it is important to investigate the effect of GlcNAcylation on a more biologically relevant species, such as nucleosomes. While the GlcNAcylated dimers do not form stable octamers, refolding of nucleosomes is also possible from separate tetramers and dimers by salt gradient dialysis.[11,12] GlcNAcylated nucleosomes were refolded from H3/H4 tetramers and GlcNacylated H2A/B dimers using a 145bp construct of the high affinity 601 DNA.[18,19] The GlcNAcylated nucleosomes yield a band migrating at the same molecular weight as the wt nucleosome in native PAGE (**Figure 5a**). The band did appear less sharp than for the wt nucleosome and more free DNA was also observed when compared to the wt reconstitution, suggesting a decrease in stability of the GlcNAcylated nucleosome. To investigate the stability of the reconstituted nucleosome, thermal melting experiments were performed and followed the unfolding by DSF or CD. In DSF a hydrophobic dye is added to the proteins, upon unfolding of the protein hydrophobic residues are exposed which can bind to the dye resulting in an increase in fluorescence.[15] Using DSF a decrease in melting temperature from 65.6 °C for the wt nucleosome to 59.6 °C for the GlcNAcylated nucleosome (**Figure 5b**) was observed. CD also showed a decrease of melting

temperature of 6.5 °C when comparing wt with the GlcNAcylated nucleosomes (64.9 °C for GlcNAc vs. 71.4 °C for wt; **Figure 5c**). Interestingly the CD traces revealed two melting transitions for both the wt and the GlcNAcylated nucleosome. This melting behaviour has been reported previously.[20] It is proposed that the first transition corresponds to the dissociation and denaturation of the H2A/B dimer, while the second corresponds to the dissociation and denaturation of the H3/H4 tetramer. This proposal is in agreement with a change only in the first transition upon GlcNAcylation, because of the destabilization of the H2A/B dimer in the nucleosome by GlcNAcylation.

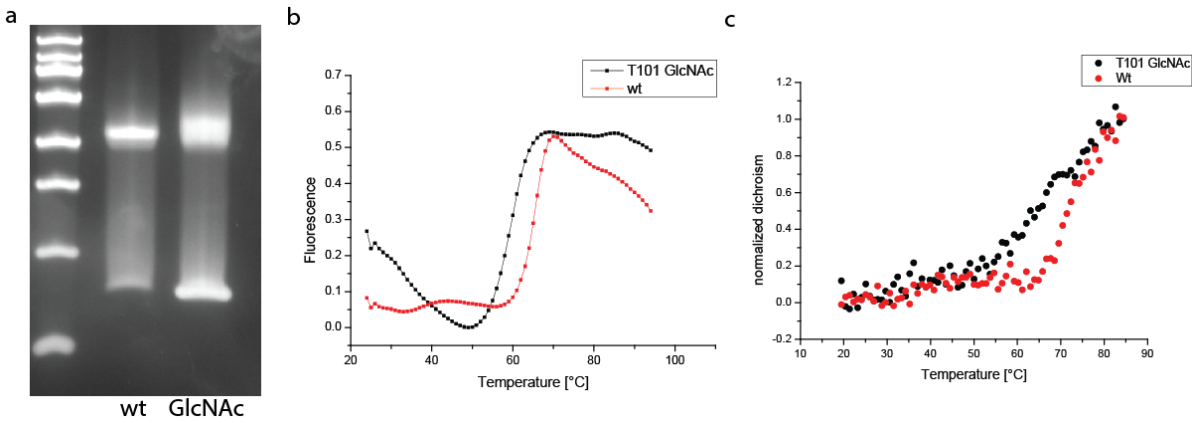


Figure 5. Comparison of stability of wt and GlcNAcylated nucleosomes. **a** Analysis of wt and GlcNAcylated nucleosomes on a 6% TBE PAGE gel. **b** Determination of melting temperature of wt (red) and GlcNAcylated (black) nucleosomes by DSF. **c** Determination of melting temperature of wt (red) and GlcNAcylated (black) nucleosomes by CD, measured at 220 nm.

5.2.5 Non-denaturing mass spectrometry

Non-denaturing mass spectrometry (nMS) is a powerful method to investigate the composition of macromolecular complexes and has been previously used to analyse nucleosome sub-complexes.[16] nMS was used to analyse the composition of nucleosomes reconstituted using the salt-gradient dialysis technique. For the wt nucleosomes two ion series corresponding to the mass of the full nucleosome (**Figure 6**) were observed, these probably correspond to differently compacted conformations. In nMS, complexes with a larger volume obtain more charge during ionization, thus the major ion series, centred around the +28 peak, would correspond to a compact conformation, while the ion series centred at +34 would correspond to an extended conformation. In the wt

sample, some hexasome in agreement with literature reports[16] and a small amount of unbound DNA was observed. The GlcNAcylated nucleosome showed a mixture of different subcomplexes. In addition to the full nucleosome, a hexasome and a nucleosome lacking one H3 was observed.

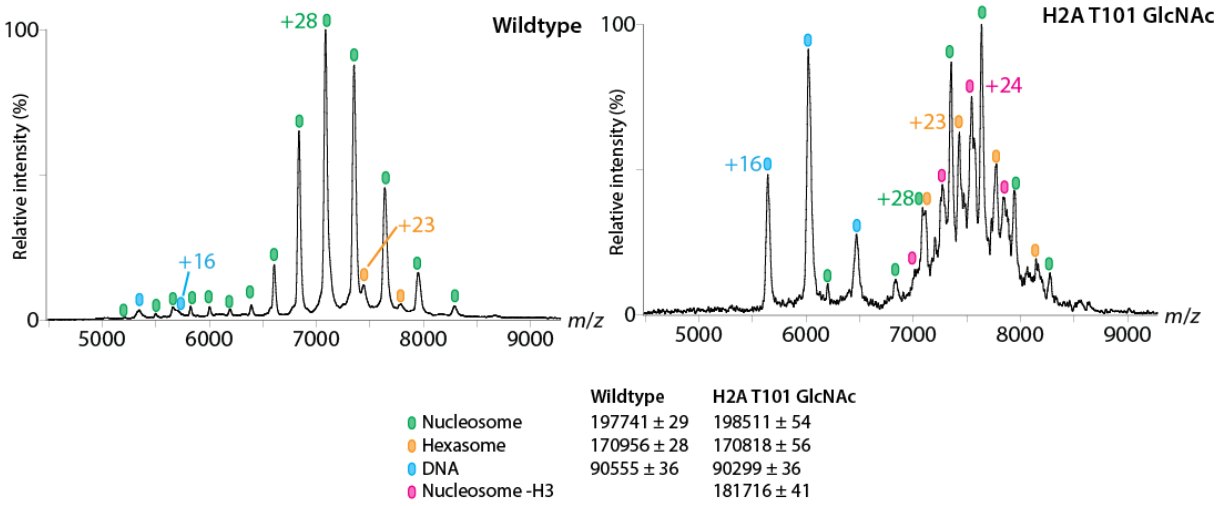


Figure 6. nMS analysis of the wt (first pannel) and GlcNAcylated (second panel) nucleosome reconstition.

Upon closer inspection of a reported nucleosome crystal structure,[21,22] it was observed that the H2A T101 alcohol is pointed towards a H3 α-helix (**Figure 7**). During salt constitutions the steric clash between this helix and the GlcNAc modification probably results in removal of one component, either the H2A/B dimer or one H3, from the nucleosome.

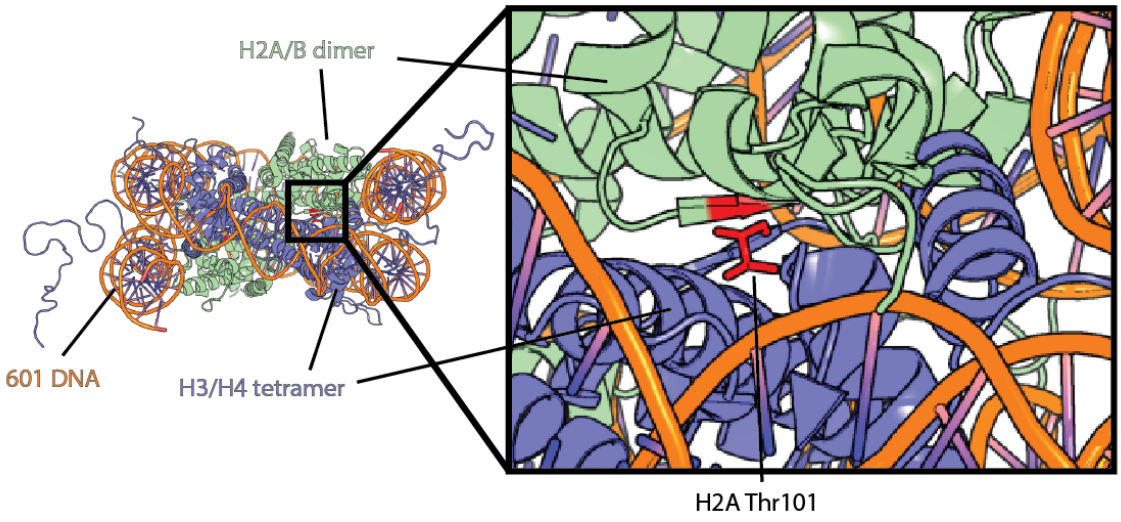


Figure 7. Nucleosome crystal structure (adapted from 1KX4)[22] with H2A Thr-101 highlighted in red.

5.2.6 Possible *In vivo* function

Having established that GlcNAcylation of H2A at Thr-101 results in destabilization of the nucleosome *in vitro*, it was of interest to consider a possible role for this modification *in vivo*. Two possible scenarios by which nucleosome destabilization could affect transcription were imagined. Firstly, destabilization would result in an open chromatin state with less nucleosome occupancy. This could enhance transcriptional activation by facilitating Pol II passage through nucleosome barriers or allowing TF binding to sites that were previously occupied by nucleosomes. Indeed, the destabilization of H2A/B dimers in nucleosomes is proposed to be an important mechanism to facilitate transcription.[23,24] .

In *in vitro* experiments, H2A T101 GlcNAcylation was only observed when H2A was used alone, not in nucleosomes or octamers. From the nucleosome structure this seems logical since this site of modification is buried in the core and thus not accessible. This suggests that T101GlcNAcylation can only happen in H2A when it is not in a nucleosome. During transcription, nucleosomes are partially disassembled when the RNA polymerase passes through and reassembled in the polymerases wake. During this disassembly, H2A/B dimers are free and could be GlcNAcylated, reinforcing transcriptional activation.

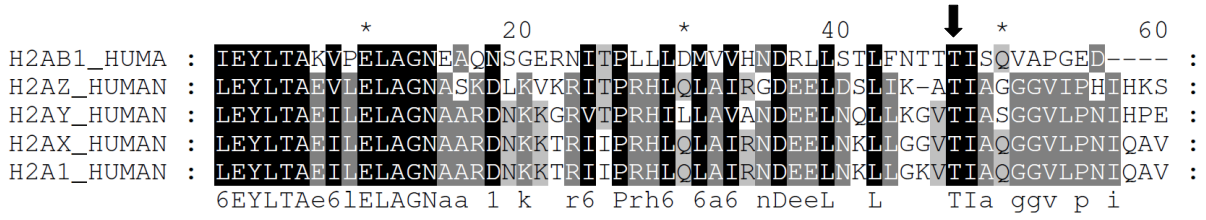


Figure 8. Alignment of different histone variants, T101 is marked by the arrow.

Interestingly, different histone variants vary somewhat in the region preceding the T101 GlcNAcylation site (**Figure 8**). H2A and different H2A variants are complexed in cells as H2A/B dimers by dedicated chaperones. Thus, it is important to consider the accessibility of the T101 residue in H2A/B chaperone complexes. There are a few chaperone-H2A/B complex structures with both H2A and H2A.Z (**Figure 9**). In the FACT complex, H2A T101 is in a flexible loop and would be accessible for

GlcNAcylation (**Figure 9c**). For H2AZ, while in the Chz1 complex T101 is accessible, in both the Anp32e and the Swr1 complex, T101 is in an α-helix stabilized by the chaperone with the group buried in the core of the protein. This suggests that the exact chaperone association of the H2A/B dimer could be an important factor in determining whether H2A T101 or the equivalent threonine in different variants is GlcNAcylated.

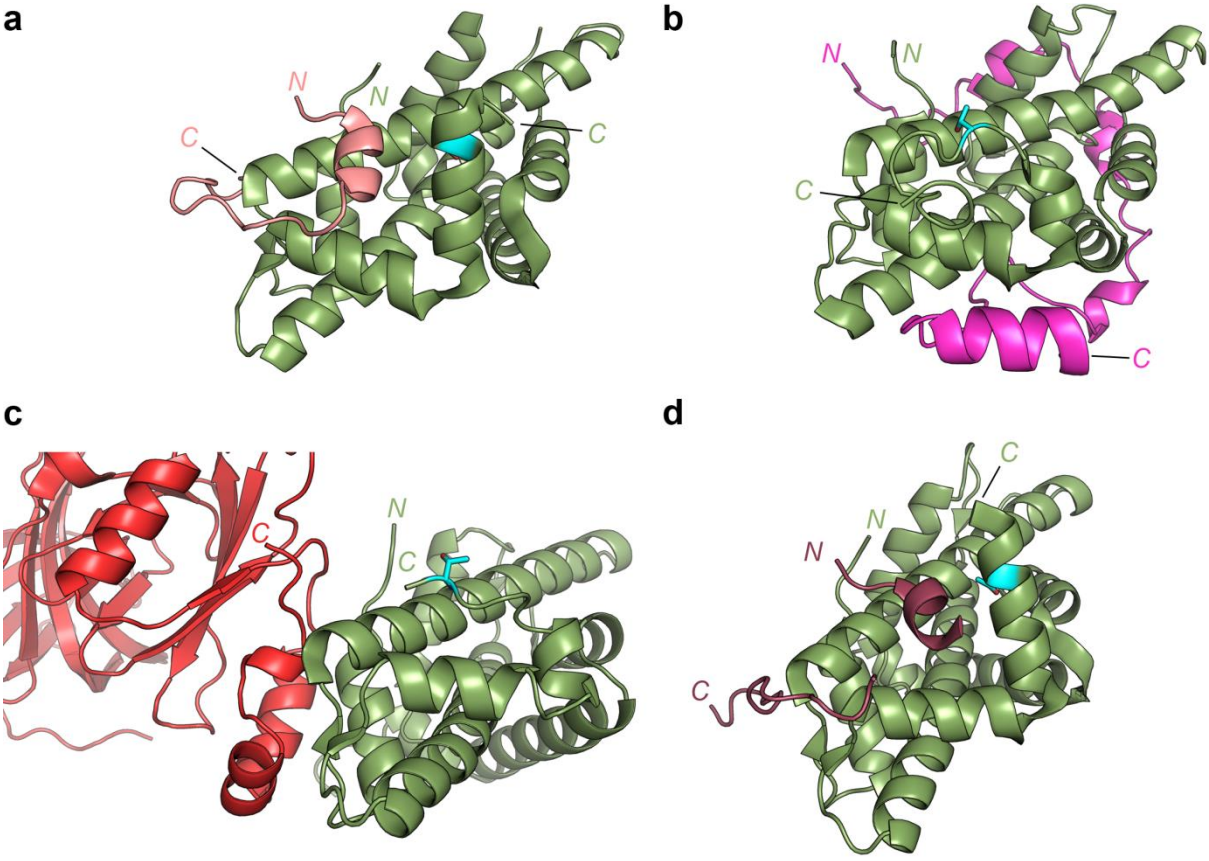


Figure 9. Structures of different histone-chaperone complexes. The H2A/B or H2AZ/B dimer is in green, the site of GlcNAcylation (T101 on H2A, T104 in H2AZ) is in light-blue and the chaperones in different red tones. **a** Anp32e-H2A/B complex adapted from pdb 4CAY.[25] **b** Chz1-H2AZ/B complex adapted from pdb 2JSS.[26] **b** FACT-H2A/B complex adapted from pdb 4KHA.[27] **d** SWR1-H2AZ/B complex adapted from pdb 4M6B[28]

By GlcNAcylation, some H2A variants could be destabilized in chromatin potentially leading to increased incorporation of non-GlcNAcylation competent variants. H2A T101 GlcNAcylation could thus also provide a targeting mechanism for the incorporation of different histone H2A variants into chromatin. Of all the histone variants, H2AZ is a promising candidate because of its importance in development[29-31] and link to PcG proteins.[29,32] Analysis of previously published OGT and H2AZ ChIP-

Seq data reveales strong co-localization of both proteins, especially at transcription start sites (**Figure 10**).

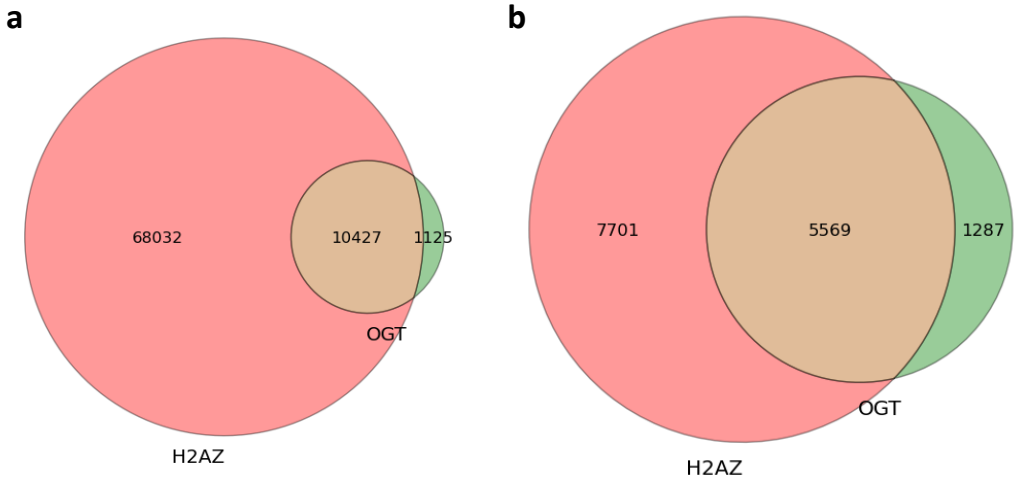


Figure 10. Venn diagram for overlap of all H2A.Z and OGT peaks (**a**) or only peaks at TSS (**b**).

It was investigated whether OGT could be involved in H2A.Z targeting *in cell*, by performing H2AZ Chromatin immunoprecipitation (ChIP) experiments in OGT siRNA knockdown cells (experiments were performed by Tom Milne (WIMM)). In ChIP, proteins that are associated with chromatin are cross-linked to the genomic DNA by e.g. formaldehyde. The cells are then lysed by sonication during which the DNA is sheared into 100-300 bp fragments. Subsequently, the protein of interest, together with the cross-linked DNA is retrieved by immunoprecipitation using the appropriate antibody. After reversal of the protein-DNA crosslink, retrieved DNA can analysed by quantitative PCR (qPCR) or sequencing. Preliminary results from H2AZ ChiP experiments, suggest that H2AZ incorporation is reduced at TSS of both repressed and active genes upon OGT KD. Further experiments will need to be performed to validate these results.

5.3 Discussion

In this chapter, I have demonstrated the utility of chemical post-translational modification as a tool to elucidate the biophysical effects and possible function of histone H2A GlcNAcylation. These provide insights into the molecular consequences of this modification from which a proposal about its *in vivo* function can be made.

Different biophysical techniques revealed that H2A T101 GlcNAcylation can destabilize the nucleosome structure, likely by modulating the dimer – tetramer association. H2A T101 is located in the *C*-terminal dimer-tetramer docking domain and this region forms a small β-sheet with a loop of the tetramer upon association, making this region perfect for regulation of dimer stability in the nucleosome.[33]

This destabilization would be in agreement with the proposed role of OGT in transcriptional activation, since nucleosome destabilization would yield an open chromatin structure which would favour transcription. OGT has been associated with both repressing and activating complexes; it is very likely that the exact complex composition can result in GlcNAcylation at different sites of the histones, explaining the complex role of OGT in transcriptional regulation. While initial OGT KD experiments were performed to see how OGT affects the chromatin landscape, it is difficult to pinpoint the exact cause of this regulation, since chromatin regulatory proteins have also been reported to be GlcNAcylated and removal of OGT could result in a change of activity in these. Additional experiments, such as localization of a H2A T101A mutant which cannot be GlcNAcylated and should be preferentially incorporated at sites of OGT recruitment, could shed more light on the exact role of modification at this site.

Sites of protein GlcNAcylation can often also be modified by phosphorylation, creating a cross-talk between these two metabolism dependent modifications.[1] Indeed, H2A T101 has also been reported as a phosphorylation site with unknown functional consequences.[34] For the histone variant H2AX, which is essential for DNA damage repair such as after exposure to ionizing radiation, T101 has been described as a phosphorylation site. T101 phosphorylation is induced upon exposure of cells to ionizing radiation and thus implicated in DNA damage response.[35] Upon Thr-101 to Ala mutation DNA the sensitivity of cells towards ionizing radiation is greatly increased, further supporting the role of this phosphorylation site in DNA damage response. It would be of interest to study the effect of T101 Phos on nucleosome structure using the presented methods.

5.4 Experimental procedures

For mutagenesis, histone expression, LC-MS analysis of protein modification, histone complex reconstitution and purification, nucleosome reconstitution and native PAGE analysis see general methods in Chapter 2. nMS analysis was performed by Nisha Patel (CVR group). H2AZ ChIP experiments were performed by Tom Milne (WIMM).

5.4.1 Histone H2A T101C modification

Histone H2A T101C was expressed and purified by Joshua Price as described in the general methods. The reaction was performed in denaturing reaction buffer (RB; 5 M Gd.HCl, 50 mM TRIS pH 8.0). 10 mg H2A T101C were weighed into an Eppendorf tube and dissolved in 500 µL RB. Then 15 mg DTT was added and the protein was shaken at rt for 30 min. Excess DTT was removed by PD minitrap (in RB) giving a final concentration of 10 mg/mL H2A T101C (**Figure 11**).

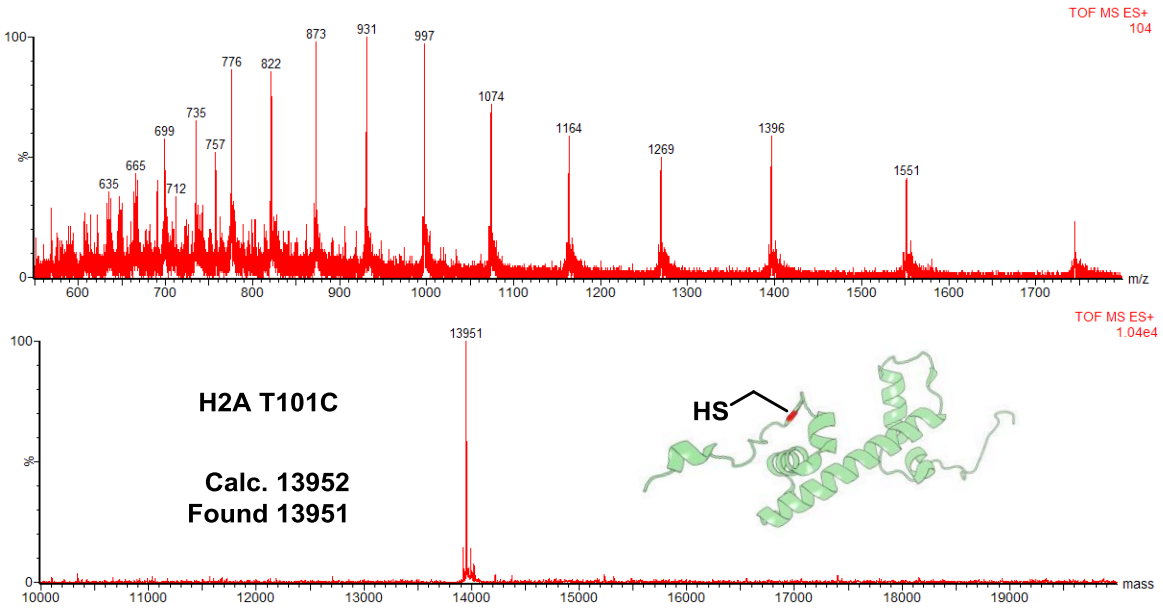


Figure 11. LC-MS analysis of the reduced H2A T101C.

5.4.1.1 H2A T101Dha Synthesis

To 900 µL of reduced H2A T101C (10 mg/mL; 0.74 mM) was added 100 µL of a DBHDA stock solution (0.5 M in DMF; 50 mM final; 90eq). The reaction was shaken at rt for 30 min and then at 37 °C for 4 h. The reaction was monitored by LC-MS and after completion; excess reagent was removed by using

a PD minitrap desalting column (in RB, each loaded with 500 µL). After 3.5 h at 37 °C full conversion of the cysteine containing starting material to Dha was observed (**Figure 12**).

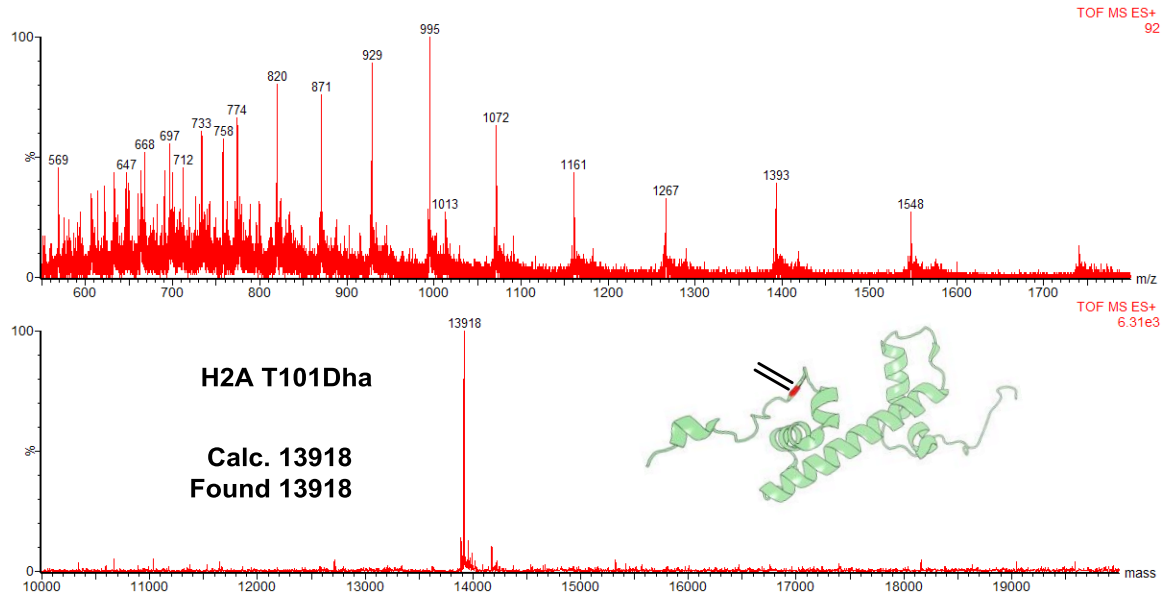


Figure 12. LC-MS analysis of H2A T101Dha.

5.4.1.2 H2A T101 GlcNAc synthesis

To 2 mL H2A T101Dha (4mg/ml, 270 µM) was added GlcNAc-SH (120 mg, 505umol, 250 mM, 925 eq) and the reaction mixture was shaken at 37 °C for 3.5 h after which full conversion to H2A T101GlcNAc was observed by LC-MS. Excess GlcNAc-SH was removed by using a PD10 desalting column (in RB) and the resultant product was stored at -20 °C for dimer reconstitution (**Figure 13**).

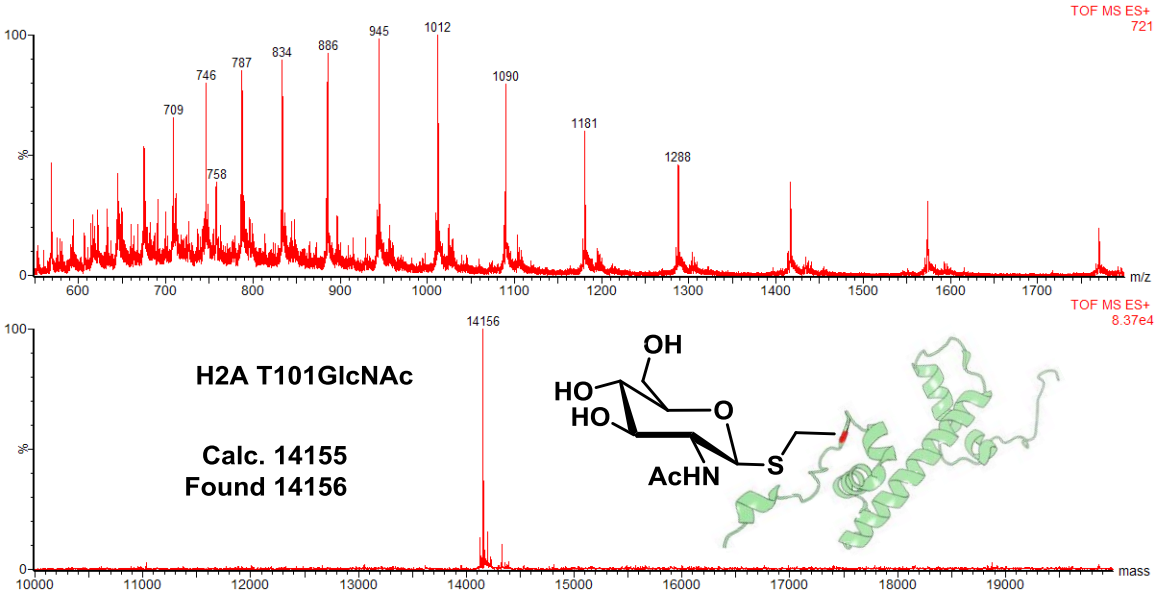


Figure 13. LC-MS analysis of H2A T101GlcNAc.

5.4.2 CD analysis of H2A/B dimers

CD spectra were recorded on a Chirascan CD-spectrophotometer (applied photophysics). H2A/B dimers were buffer exchanged into 20 mM $NaHPO_4$ pH 6.7 1mM EDTA 150 mM NaCl by viva-spin. A final concentration of 10 µM dimer in 500 µL in a micro cuvette (1 mm path length) was used for each measurement. CD spectra were recorded from 200 – 300 nm with a step size of 1 nm and 0.5 s per point. The starting spectra were recorded in triplicates. The three traces were averaged and smoothed using a window of 4 and the background was subtracted. For melting curves the temperature was increased stepwise from 25-90 °C in 1 °C steps with a hold time of 60 sec. The temperature of the sample as measured by a sample probe was recorded for each step. The raw data was exported from the Pro-data as CSV and further processed using OriginPro. The melting temperature was extracted by plotting the CD at 220 nm against temperature and fitting using a Boltzmann sigmoidal function.

5.4.3 CD analysis of nucleosomes

CD spectra were recorded on a Chirascan CD-spectrophotometer (applied photophysics). Nucleosomes (2.5 µM) were in 10 mM $NaHPO_4$ pH 7.5 1mM EDTA 150 mM NaCl. For the

measurement 300 µL in a micro cuvette (1 mm path length) was used. CD spectra were recorded from 200-300 nm with a step size of 0.5 nm and 1 sec measurement per point. Temperature was ramped stepwise from 25-95 °C in 1 °C steps with 120 sec at each temperature. Data was processed as described above for the dimer.

5.4.4 DSF analysis of nucleosomes

DSF analysis was performed on a Bio-Rad MJ Mini real time PCR machine using the Opticon Monitor 3 software. For the analysis 23 µL of 2.5 µM nucleosome in 10 mM $NaHPO_4$ pH 7.5 1mM EDTA 150 mM NaCl and 2 µL of 25 x SYPRO orange (diluted from 5,000X Concentrate in DMSO; Life technologies S-6650) were added into a real time PCR plate and sealed with transparent film. Then the sample was first kept for 1 min at 15 °C and then heated from 15-95 °C in 1 °C steps with one minute at each temperature. Fluorescence was read out for each temperature. The data was analysed using the excel sheets described by Niesen *et al*.[15] Melting temperatures were extracted by fitting using a Boltzmann sigmoidal function in OriginPro.

5.4.5 ChIP-Seq and DNase I sensitive site analysis

OGT ChIP data published by Vella *et al*[36] was used for the analysis. DNase I sensitivity data was obtained from data deposited by the ENCODE project (Digital DNaseI from ENCODE/ University of Washingtion; ES-E14; 129/Ola; E0; hotspot-broadPeak).[37,38] Data was analysed using Galaxy[39-41] and visualised using the UCSC genome browser.[42]

5.5 References

(1) Hart, G. W.; Slawson, C.; Ramirez-Correa, G.; Lagerlof, O. *Annu. Rev. Biochem.* **2011**, *80*, 825.
(2) Bannister, A. J.; Kouzarides, T. *Cell Res.* **2011**, *21*, 381.
(3) Sakabe, K.; Wang, Z.; Hart, G. W. *Proc. Natl. Acad. Sci.* **2010**, *107*, 19915.
(4) Fong, J. J.; Nguyen, B. L.; Bridger, R.; Medrano, E. E.; Wells, L.; Pan, S.; Sifers, R. N. *J. Biol. Chem.* **2012**, *287*, 12195.

(5) Fujiki, R.; Hashiba, W.; Sekine, H.; Yokoyama, A.; Chikanishi, T.; Ito, S.; Imai, Y.; Kim, J.; He, H. H.; Igarashi, K.; Kanno, J.; Ohtake, F.; Kitagawa, H.; Roeder, R. G.; Brown, M.; Kato, S. *Nature* **2011**, *480*, 557.
(6) Myers, S. A.; Panning, B.; Burlingame, A. L. *Proc. Natl. Acad. Sci.* **2011**, *108*, 9490.
(7) Deplus, R.; Delatte, B.; Schwinn, M. K.; Defrance, M.; Mendez, J.; Murphy, N.; Dawson, M. A.; Volkmar, M.; Putmans, P.; Calonne, E.; Shih, A. H.; Levine, R. L.; Bernard, O.; Mercher, T.; Solary, E.; Urh, M.; Daniels, D. L.; Fuks, F. *EMBO J.* **2013**, *32*, 645.
(8) Love, D. C.; Krause, M. W.; Hanover, J. A. *Semin. Cell Dev. Biol.* **2010**, *21*, 646.
(9) Hanover, J. A.; Krause, M. W.; Love, D. C. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 312.
(10) Aydillo, C.; Compañón, I.; Avenoza, A.; Busto, J. H.; Corzana, F.; Peregrina, J. M.; Zurbano, M. M. *J. Am. Chem. Soc.* **2013**.
(11) Bönisch, C.; Schneider, K.; Pünzeler, S.; Wiedemann, S. M.; Bielmeier, C.; Bocola, M.; Eberl, H. C.; Kuegel, W.; Neumann, J.; Kremmer, E.; Leonhardt, H.; Mann, M.; Michaelis, J.; Schermelleh, L.; Hake, S. B. *Nucleic Acids Res.* **2012**, *40*, 5951.
(12) Dyer, P. N.; Edayathumangalam, R. S.; White, C. L.; Bao, Y.; Chakravarthy, S.; Muthurajan, U. M.; Luger, K. In *Methods Enzymol.*; Allis, C. D., Carl, W., Eds.; Academic Press: 2003; Vol. Volume 375, p 23.
(13) Karantza, V.; Baxevanis, A. D.; Freire, E.; Moudrianakis, E. N. *Biochemistry* **1995**, *34*, 5988.
(14) Ausio, J.; Van Holde, K. E. *Biochemistry* **1986**, *25*, 1421.
(15) Niesen, F. H.; Berglund, H.; Vedadi, M. *Nat. Protocols* **2007**, *2*, 2212.
(16) Azegami, N.; Saikusa, K.; Todokoro, Y.; Nagadoi, A.; Kurumizaka, H.; Nishimura, Y.; Akashi, S. *Biochemistry* **2013**, *52*, 5155.
(17) Campos, E. I.; Fillingham, J.; Li, G.; Zheng, H.; Voigt, P.; Kuo, W.-H. W.; Seepany, H.; Gao, Z.; Day, L. A.; Greenblatt, J. F.; Reinberg, D. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1343.
(18) Vasudevan, D.; Chua, E. Y. D.; Davey, C. A. *J. Mol. Biol.* **2010**, *403*, 1.
(19) Lowary, P. T.; Widom, J. *J. Mol. Biol.* **1998**, *276*, 19.
(20) Mutskov, V.; Gerber, D.; Angelov, D.; Ausio, J.; Workman, J.; Dimitrov, S. *Mol. Cell. Biol.* **1998**, *18*, 6293.
(21) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251.
(22) Davey, C. A.; Sargent, D. F.; Luger, K.; Maeder, A. W.; Richmond, T. J. *J. Mol. Biol.* **2002**, *319*, 1097.
(23) Belotserkovskaya, R.; Oh, S.; Bondarenko, V. A.; Orphanides, G.; Studitsky, V. M.; Reinberg, D. *Science* **2003**, *301*, 1090.
(24) Pavri, R.; Zhu, B.; Li, G.; Trojer, P.; Mandal, S.; Shilatifard, A.; Reinberg, D. *Cell* **2006**, *125*, 703.
(25) Obri, A.; Ouararhni, K.; Papin, C.; Diebold, M. L.; Padmanabhan, K.; Marek, M.; Stoll, I.; Roy, L.; Reilly, P. T.; Mak, T. W.; Dimitrov, S.; Romier, C.; Hamiche, A. *Nature* **2014**, *505*, 648.
(26) Zhou, Z.; Feng, H.; Hansen, D. F.; Kato, H.; Luk, E.; Freedberg, D. I.; Kay, L. E.; Wu, C.; Bai, Y. *Nat. Struct. Mol. Biol.* **2008**, *15*, 868.
(27) Hondele, M.; Stuwe, T.; Hassler, M.; Halbach, F.; Bowman, A.; Zhang, E. T.; Nijmeijer, B.; Kotthoff, C.; Rybin, V.; Amlacher, S.; Hurt, E.; Ladurner, A. G. *Nature* **2013**, *499*, 111.
(28) Hong, J.; Feng, H.; Wang, F.; Ranjan, A.; Chen, J.; Jiang, J.; Ghirlando, R.; Xiao, T. S.; Wu, C.; Bai, Y. *Mol. Cell* **2014**, *53*, 498.
(29) Creyghton, M. P.; Markoulaki, S.; Levine, S. S.; Hanna, J.; Lodato, M. A.; Sha, K.; Young, R. A.; Jaenisch, R.; Boyer, L. A. *Cell* **2008**, *135*, 649.
(30) Hu, G.; Cui, K.; Northrup, D.; Liu, C.; Wang, C.; Tang, Q.; Ge, K.; Levens, D.; Crane-Robinson, C.; Zhao, K. *Cell Stem Cell* **2013**, *12*, 180.
(31) Subramanian, V.; Mazumder, A.; Surface, L. E.; Butty, V. L.; Fields, P. A.; Alwan, A.; Torrey, L.; Thai, K. K.; Levine, S. S.; Bathe, M.; Boyer, L. A. *PLoS Genet* **2013**, *9*, e1003725.
(32) Illingworth, R. S.; Botting, C. H.; Grimes, G. R.; Bickmore, W. A.; Eskeland, R. *PLoS ONE* **2012**, *7*, e34848.
(33) Bönisch, C.; Hake, S. B. *Nucleic Acids Res.* **2012**.

(34) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. *Cell* **2006**, *127*, 635.
(35) Xie, A.; Odate, S.; Chandramouly, G.; Scully, R. A. *Cell Cycle* **2010**, *9*, 3602.
(36) Vella, P.; Scelfo, A.; Jammula, S.; Chiacchiera, F.; Williams, K.; Cuomo, A.; Roberto, A.; Christensen, J.; Bonaldi, T.; Helin, K.; Pasini, D. *Mol. Cell* **2013**, *49*, 645.
(37) Consortium, M. E.; Stamatoyannopoulos, J.; Snyder, M.; Hardison, R.; Ren, B.; Gingeras, T.; Gilbert, D.; Groudine, M.; Bender, M.; Kaul, R.; Canfield, T.; Giste, E.; Johnson, A.; Zhang, M.; Balasundaram, G.; Byron, R.; Roach, V.; Sabo, P.; Sandstrom, R.; Stehling, A. S.; Thurman, R.; Weissman, S.; Cayting, P.; Hariharan, M.; Lian, J.; Cheng, Y.; Landt, S.; Ma, Z.; Wold, B.; Dekker, J. *Genome Biol.* **2012**, *13*, 418.
(38) Sabo, P. J.; Kuehn, M. S.; Thurman, R.; Johnson, B. E.; Johnson, E. M.; Cao, H.; Yu, M.; Rosenzweig, E.; Goldy, J.; Haydock, A.; Weaver, M.; Shafer, A.; Lee, K.; Neri, F.; Humbert, R.; Singer, M. A.; Richmond, T. A.; Dorschner, M. O.; McArthur, M.; Hawrylycz, M.; Green, R. D.; Navas, P. A.; Noble, W. S.; Stamatoyannopoulos, J. A. *Nat. Meth.* **2006**, *3*, 511.
(39) Goecks, J.; Nekrutenko, A.; Taylor, J.; Team, T. G. *Genome Biol.* **2010**, *11*, R86.
(40) Blankenberg, D.; Kuster, G. V.; Coraor, N.; Ananda, G.; Lazarus, R.; Mangan, M.; Nekrutenko, A.; Taylor, J. In *Current Protocols in Molecular Biology*; John Wiley & Sons, Inc.: 2001.
(41) Giardine, B.; Riemer, C.; Hardison, R. C.; Burhans, R.; Elnitski, L.; Shah, P.; Zhang, Y.; Blankenberg, D.; Albert, I.; Taylor, J.; Miller, W.; Kent, W. J.; Nekrutenko, A. *Genome Res.* **2005**, *15*, 1451.
(42) Kent, W. J.; Sugnet, C. W.; Furey, T. S.; Roskin, K. M.; Pringle, T. H.; Zahler, A. M.; Haussler; David *Genome Res.* **2002**, *12*, 996.

6 Histone tail structure and the effect of Ser-10 phosphorylation

6.1 Introduction

As discussed in Chapter 1, eukaryotic DNA is organized into a compact structure called chromatin. Nucleosomes, consisting of an octamer with each two copies of histone H2A, H2B, H3 and H4 wrapped around ~145 bp of DNA, are the basic packing unit of chromatin. All individual histones contain a folded, mainly α-helical, core and extended tails at the *C*- and/or *N*-terminus.[1,2] Histones do not only fulfil a structural role, but are important hubs for transcriptional regulation.[3] This regulation is accomplished in part by posttranslational modification (PTM) of various residues on the histone tails. While the folded nucleosome core has been extensively characterized by crystallography,[1,4] the histone tails are extended from the core and not well resolved in the reported crystal structures. From their amino acid sequence, histone tails are predicted to be intrinsically disordered.[5,6] Intrinsically disordered proteins (IDP) or proteins containing disordered domains make up a large fraction of the proteome[7] and characterizing the structure and dynamics of IDPs is a major current interest.[8,9] For structural description of these flexible domains nuclear magnetic resonance (NMR) is an important tool as it reports on the average structure adopted by the ensemble at atomic resolution.[8] By NMR many structural parameters characteristic of residual order within an IDP ensemble are measurable and structures that match the experimentally determined parameters can then be obtained by molecular dynamic approaches.[8] As most sites of epigenetic regulation are on the intrinsically disordered histone tails, elucidating the structure and dynamics of the histone tails is essential to gain better understanding of the epigenetic mechanisms that govern transcriptional regulation.

Whilst the unstructured histone tails have not been structurally described by protein crystallography, a large body of work describing the structure and function of these domains[6] exists including early 1D NMR studies,[6,10-13] small angle x-ray scattering (SAXS), UV-crosslinking, DNA foot-printing,[14]

sedimentation analysis and enzymatic assays.[15] While all these methods provide insight into the overall behaviour of the histone tails, no information of histone tail structure at atomic resolution is currently available.

SAXS studies[16,17] on nucleosomes in different salt concentrations showed an ionic strength dependent increase in size of the nucleosome particle.[18,19] This was attributed to the release of the histone tails from the nucleosome core, which was confirmed by comparing the salt dependent size increase of wt nucleosomes with nucleosomes lacking all tails,[20] or only the H3 and H4 tails.[21] When the H3/H4 tails are removed, the nucleosome structure needed to be modelled with partially unwrapped DNA ends to fit the SAXS data.[21] In contrast, in these SAXS studies wt nucleosomes are best modelled with fully-associated DNA even in the conformation with extended tails at higher ionic strength. This tentatively suggests that the tails, although partially dissociated, still retain some interactions with the DNA, stabilizing it on the nucleosome core. The exact salt concentration of dissociation increases with the length of DNA used for the nucleosome reconstitution.[20]

1D NMR studies have confirmed the presence of flexible regions in the histone proteins.[10-13] Flexible regions were distinguished from the octamer or nucleosome core by their sharp peaks and their sensitivity towards trypsin digestion.[10,13] An increase in the signal intensities of these regions from an increase in flexibility was only observed above 300mM NaCl in nucleosomes.[10] This salt dependent increase in signal is not observed in trypsinised nucleosomes.[10] From this the authors concluded that the flexible regions or tails are associated with the nucleosome core at physiological salt concentrations and only dissociate at higher than physiological salt concentrations. Because only 1D NMR was used, no tail or residue specific information about flexibility and structure was obtained.

UV-crosslinking studies have revealed that the basic histone tails can be cross-linked to nucleosomal DNA and are preferentially bound to linker DNA outside the nucleosome core.[22-25] Detailed mapping of the cross-linking sites showed that the H3 tail can make contact with nucleosomal DNA at various regions around the exit site of the H3 tail from the core.[26,27] These interactions were also detected in

the presence of linker DNA.[27] In a di-nucleosome model system the histone H3 tail was shown to make mainly intra-nucleosome contacts, while other histone tails participated in inter-nucleosome interactions.[28] Thus, a mono-nucleosome should provide a good model for the structure of the histone H3 tail. Salt titrations showed that the crosslinking of histones to DNA was reduced above 300 mM NaCl and was dependent on the presence of histone tails,[29] in agreement with a dissociation of the tail from the DNA at high ionic strength. All of these studies were performed at physiological or higher salt concentrations, suggesting that, even though the tails appear more extended at higher ionic strength, DNA-tail interactions are still present and could be an important determinant for histone tail structure.

The addition of PTMs to histones has been shown to affect the behaviour of the histone tails. The best studied modification in this respect is lysine acetylation, in part because hyper-acetylated histones and nucleosome can easily be generated *in vivo*. Lysine acetylation neutralizes the charge of the ε-amine. Thus, acetylation is expected to weaken the mainly electrostatic interaction between the positively charged histone tails and the negatively charged DNA. Indeed, hyper-acetylation results in particles with reduced thermal stability,[30] reduced sedimentation coefficient[31] and more loosely associated tails.[29] Also, an increase in α-helical content in the tails has been observed upon acetylation.[32] Other modifications have been less well studied, mainly due to challenges with sample preparation.

Histone phosphorylation is an important PTM[3] that can modulate the net charge of the tail by introducing two additional negative charges. H3 Ser-10 phosphorylation (H3 S10Ph) and the enzymes that deposit this modification are involved with various cellular processes such as regulation of transcriptional activation and mitosis.[33,34] H3 S10Ph was linked to mitosis because of the significant temporal overlap between H3 S10Ph occurrence and chromosome condensation and cell division.[35,36] Mutation of H3 S10 or deletion of the responsible mitosis regulating kinase (Aurora B in humans) results in mitotic defects in most studied organisms.[37-39] As a terminal target of multiple kinase cascades, H3 S10 phosphorylation is an important regulator of transcriptional activation.[33]

Inactivation of the responsible kinases results in transcriptional mis-regulation while mitotic phosphorylation is maintained.[40,41]

From a mechanistic perspective, histone marks can function via either recruitment of histone code 'reader' proteins or by modulating biophysical properties such as structure, dynamics and stability of histones and nucleosomes (see Chapter 1). H3 S10Ph can function by recruiting the 14-3-3ζ 'reader' protein resulting in transcriptional activation[42,43] or by blocking binding of transcriptional repressing proteins such as HP1.[44,45] Phosphorylation in other disordered proteins has also been shown to have effects on structure.[46-48] Phosphorylation in an arginine/serine rich domain can increase structuring and rigidity in this domain.[47] In the intrinsically disordered protein Sic1 phosphorylation leads to an increase in extended conformations in the structural ensemble.[48]

Recent developments in solution NMR enable studies of large complexes such as nucleosomes (198 kDa).[49-51] This is accomplished by global deuteration, which improves relaxation properties, selective isotope labelling and optimized pulse sequences. In this chapter, work that uses NMR to characterize residual structure in the histone H3 tail is described. Zhou *et al.*[52] have reported assignments of the flexible tail regions of the *Drosophila melanogaster* histone H3 in a nucleosome. Further detailed structural analysis of the H3 tail was not reported. In this chapter, it is shown that the H3 tail adopts a collapsed conformation around the exit of the DNA from the nucleosome. This structure is modulated by H3 S10 phosphorylation, providing the first molecular picture of the influence of phosphorylation on histone tail structure. Based on these insights, a model of how H3 S10 phosphorylation is able to function in many different cellular processes by reorganizing H3 tail structure is proposed.

6.2 Results

6.2.1 The H3 tail is partially structured in a nucleosome context

To investigate the structure of the histone H3 tail by NMR, a deuterated and ^{15}N labelled recombinant *Xenopus laevis* histone H3 was produced in *E. coli* transformed with the appropriate

plasmid (pET3d, H3 C110A) grown in M9 minimal media in D_2O containing deuterated glucose and $^{15}NH_4Cl$. The protein was purified by size exclusion (SEC) and ion exchange chromatography (IEXC).[53] Together with unlabelled wildtype *Xenopus laevis* H2A, H2B and H4, expressed and purified using the same methods, histone octamers containing only one labelled protein were reconstituted and purified by SEC. These were then used, together with a 145 bp construct of the high affinity '601' nucleosome positioning sequence[54] to assemble nucleosomes by salt gradient dialysis.[53] The 601 DNA sequence was selected by Lowary *et al*.[54] by a systematic evolution of ligands by exponential enrichment (SELEX) approach from a pool of random DNA sequences. This sequence has very high affinity for the histone octamer and yield highly homogeneous nucleosomes. Accurate DNA and protein concentration was found to be essential to obtain highly homogenous nucleosomes. Excess DNA results in the formation of an inhomogeneous mixture, while excess octamer results in low yield. Before large scale assembly, small scale assemblies using different concentrations were performed and analysed by native PAGE to find the correct DNA: protein ratio (**Figure 1**). The ratio that gave good nucleosome yield and no free DNA was used for large scale reconstitution of nucleosomes. After the nucleosome reconstitution, the nucleosomes were exchanged into a low-salt buffer (10 mM NaH_2PO_4 pH 6.5, 1 mM EDTA).

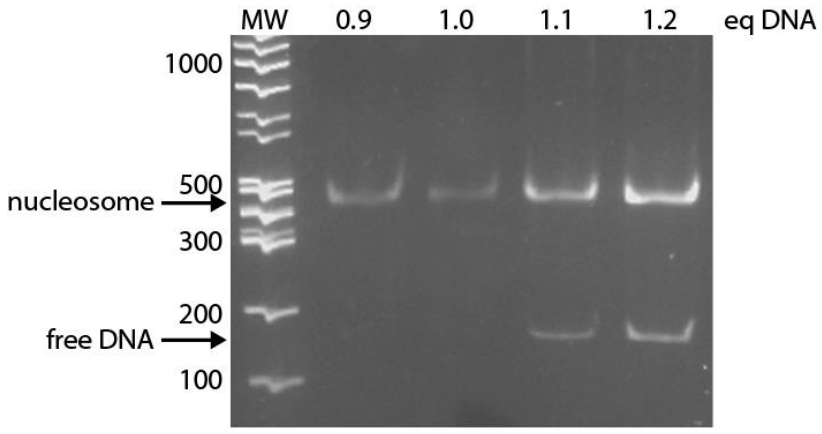


Figure 1. Native PAGE analysis of a titration of increasing amounts of DNA with constant octamer concentration.

In the ^{1}H-^{15}N HSQC, 30 resonances corresponding to the *N*-terminal histone H3 tail were observed (**Figure 2a**). The peak positions of the H3 tail matched closely the published HSQC spectra of

Drosophila melanogaster H3 in a nucleosome context, as might be expected from the complete conservation of histone H3 between the organisms (**Figure 3**).[52]

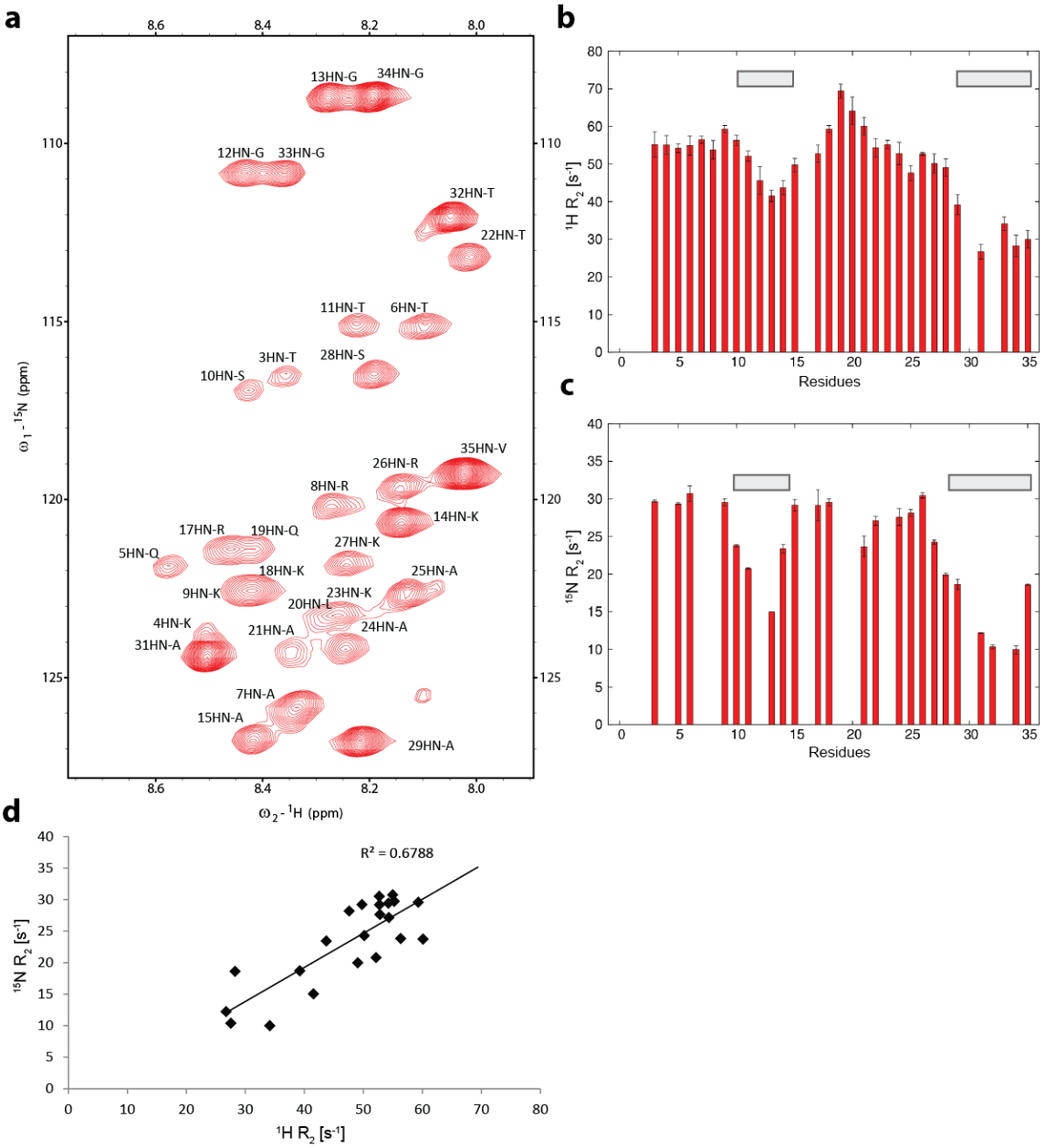


Figure 2. a ^{1}H-^{15}N HSQC spectra of the deuterated H3 tail in a nucleosome. Assigments were taken from Zhou *et al.*[52]. **b** ^{1}H R_2 relaxation rates of the histone H3 tail. Regions of low R_2 are indicated with a box above the plot. **c** ^{15}N R_2 relaxation rates of the histone H3 tail. Regions of low R_2 are indicated with a box above the plot. **d** Correlation between the measured ^{1}H and ^{15}N R_2 rates for the histone H3 tail.

Upon closer inspection of the spectra, large differences in peak intensity across different residues became obvious, with the peaks for residues closest to the nucleosome core (res 31-35) appearing

most intense. Differences in peak intensity could result from different relaxation properties in different parts of the chain but nevertheless, regions distant from the core nucleosome would be predicted to be more flexible and therefore more intense.

To investigate this possibility, ^{1}H or ^{15}N R_2 relaxation rates were measured (**Figure 2b** and **c**). All rate analyses was performed using the Function and Data Analysis (FuDa)[55] package and the quality of the peak fits were inspected manually. Only peaks with statistically sound fits were used for further analysis. A profile agreeing with the peak intensity profile with lower ^{1}H R_2 for residues 10-15 (reduction of 5-10 s^{-1}) and 29-35 (reduction of 10-20 s^{-1}) was observed. A similar profile was observed for the ^{15}N R_2 values, which ranged from 15 s^{-1} to 30 s^{-1}.

```
                  *         20         *         40         *         60
H3_DROME   : MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPF :  68
H32_XENLA  : MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPF :  68

                   *         80         *         100        *         120        *
H3_DROME   : QRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA : 136
H32_XENLA  : QRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA : 136
```

Figure 3. Sequence allignment for Histone H3 from *Drosophila melanogaster* (H3_DROME) and H3 from *Xenopus laevis* (XENLA)

It was considered that higher R_2 rates for some residues in the H3 tail (3-10 and 17-28) could indicate residual secondary structure[56] or chemical exchange between different conformations on an intermediate time scale. To test for intermediate exchange, Carr-Purcel-Meiboom-Gill (CPMG)[57,58] relaxation dispersion measurements were performed.[59,60] In this experiment, during a set delay-time a sequence of 180° refocusing pulses with a variable frequency is applied. When chemical exchange in the ms time scale is present, a low CPMG frequency will result in broadening of the peaks, because of incomplete refocusing of the magnetization, yielding a higher R_2^{eff}. At high CPMG frequencies, the magnetization is refocused more efficiently, resulting in more intense peaks and lower R_2^{eff}. From these experiments, rates of exchange and chemical shift differences between the two states can be extracted. Because of the large R_2 of the backbone amides, insufficient signal to noise (S/N) was obtained from the ^{15}N labelled sample. To increase sensitivity a deuterated, leucine and valine ^{1}H,

^{13}C-methyl labelled H3[61] was produced in *E. coli* and used to reconstitute nucleosomes. For the specific labelling *E. coli* transformed with the H3 plasmid was grown in M9 minimal medium with D_2O, deuterated ^{12}C glucose and $^{15}NH_4Cl$. For selective leucine and valine γ-$^{13}CH_3$ labelling, 2-Keto-3-methyl-^{13}C-butyric-4-^{13}C, 3-d acid was added one hour before induction with IPTG.

From the ^{1}H-^{13}C HMQC (**Figure 4**a) excellent S/N was obtained in the CPMG experiments for residues in the tail but no relaxation dispersions were observed at different temperatures and in different buffers (**Figure 4b**). As expected in the absence of exchange, a decrease in R_2^{eff} with temperature is observed for both the V35 and L20 (**Figure 4c**). Interestingly, the obtained R_2^{eff} for V35 is lower than for L20 in agreement with previous ^{1}H and ^{15}N R_2 measurements.

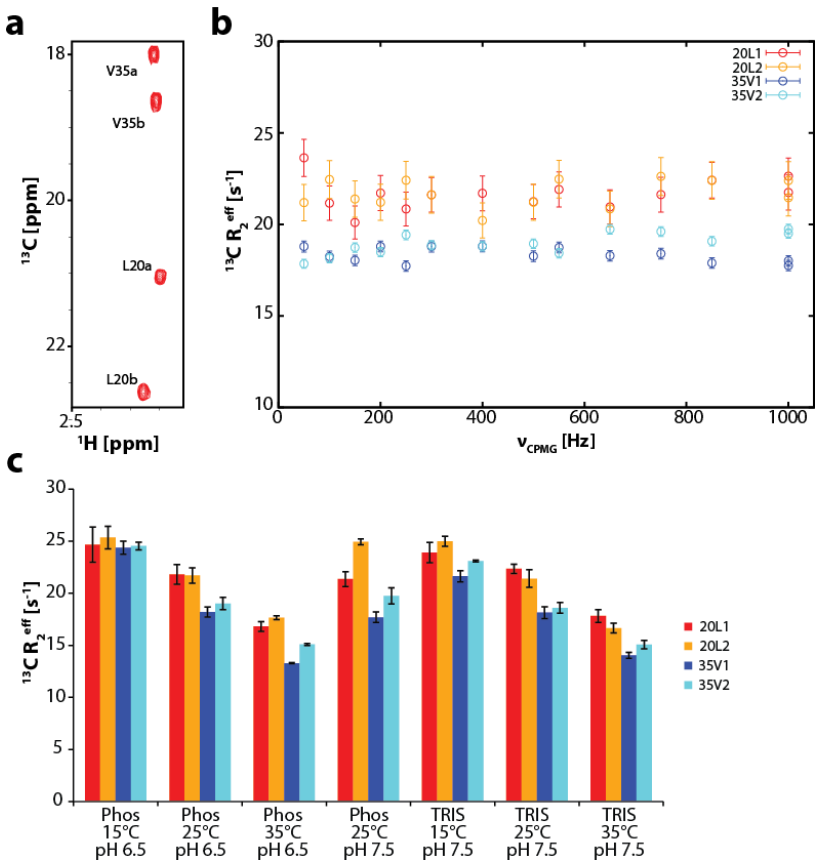


Figure 4. a ^{1}H-^{13}C HMQC spectra of nucleosomes containing the deuterated and methyl labelled H3. Because of the short acquisition time only peaks in the flexible tail region are observed. The assigment was taken from Kato *et al.*[49]. **b** Example of a flat ^{13}C R_2^{eff} profile at different CPMG rates. **c** Calculated ^{13}C R_2^{eff} rates from the CPMG experiments at different temperatures and in different buffers.

This lack of intermediate chemical exchange in the H3 tail suggests that the conformers of the tail are exchanging rapidly, on the microsecond time scale and faster. Nevertheless, the R_2 trends suggest that the tail might contain residual secondary structure. Established measures of residual structure in IDPs are the H_α, CO, C_α and C_β and to a lesser extent the HN and N chemical shifts.[8]

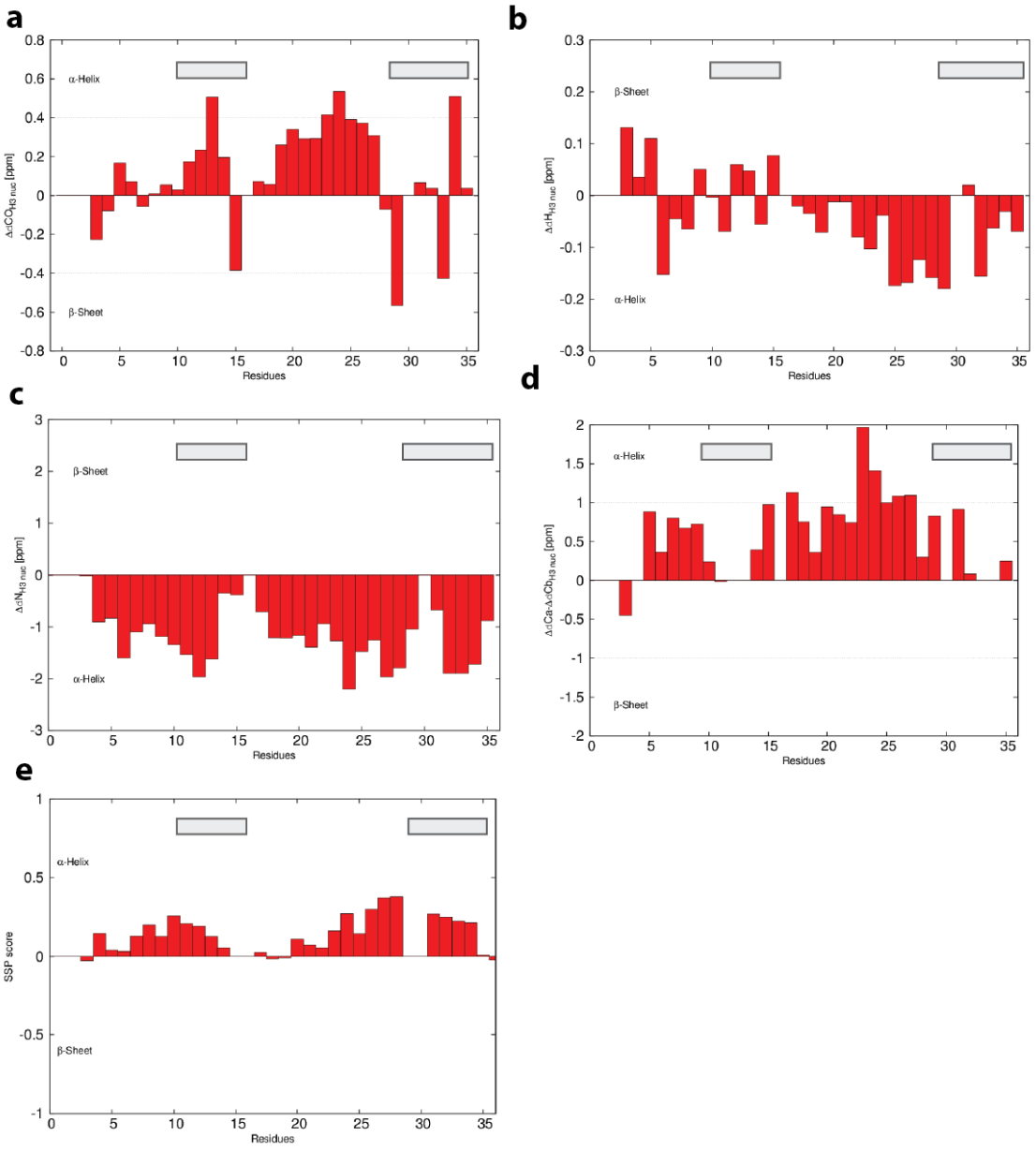


Figure 5. SCS analysis of the ΔδCO (**a**), ΔδNH (**b**), ΔδN (**c**) and $\Delta\delta C_\alpha$-$\Delta\delta C_\beta$ (**d**) chemical shifts of a H3 tail in a nucleosome. **d** SSP analysis of the combined chemical shifts. Regions of low R_2 are marked with grey bars.

CS are analysed by subtracting experimentally determined random coil values to give a secondary chemical shift (SCS).[62,63] Upfield or downfield shifts from the random coil value of different nuclei are

used to determine residual secondary structure propensity of the polypeptide chain.[64,65] To deduce residual secondary structure propensity, CO, C_α and C_β chemical shifts of a deuterated ^{15}N ^{13}C labelled H3 in the nucleosome were measured and from these, SCS were calculated (**Figure 5**).

α-Helical character, as indicated by positive $\Delta\delta C_\alpha$-$\Delta\delta C_\beta$ and ΔδCO values (**Figure 5a** and **d**) and negative ΔδN and ΔδHN values (**Figure 5b** and **c**), was observed for residues 19-27 of the H3 tail. Weak α-helical character, as indicated by small but positive $\Delta\delta C_\alpha$-$\Delta\delta C_\beta$ and negative ΔδN values, was observed for residues 5-9. Combining these chemical shifts to calculate residue-specific secondary structure propensity (SSP)[66] revealed an α-helical propensity for residue 5-12 and 20-31 with a higher score in the second stretch, similar to the analysis of individual SCS. These results are in agreement with previous reports that observed α-helical content in the histone tails by CD.[32,67]

6.2.2 Secondary structure is induced by tail-nucleosome interactions

The observed structural features could be the result of the intrinsic properties of the H3 tail polypeptide or be induced by interactions of the tail with the nucleosome core and DNA. To characterize the intrinsic properties of an isolated H3 tail a short histone H3 tail peptide (res 1-44) was prepared.

For this, a fusion protein consisting of an *N*-terminal his-tagged SUMO for solubility followed by a "Tobacco Etch Virus nuclear inclusion A endopeptidase" (TEV) recognition site and a *C*-terminal H3 tail (res 1-44) was constructed (**Figure 6**). TEV protease can cleave a variant of its optimal recognition sequence (ENLYFQ\S) in which the *C*-terminal serine is replaced with alanine (ENLYFQ\A).[68] Digestion of the construct resulted in a wt tail with an *N*-terminal alanine without any additional amino acids. The fusion protein was produced in *E. coli* with ^{15}N ^{13}C labels, purified by IMAC and SEC and digested with TEV protease. Subsequently the tag and the TEV protease were removed by another round of SEC.

Figure 6. Sequence for the SUMO-H3 tail construct.

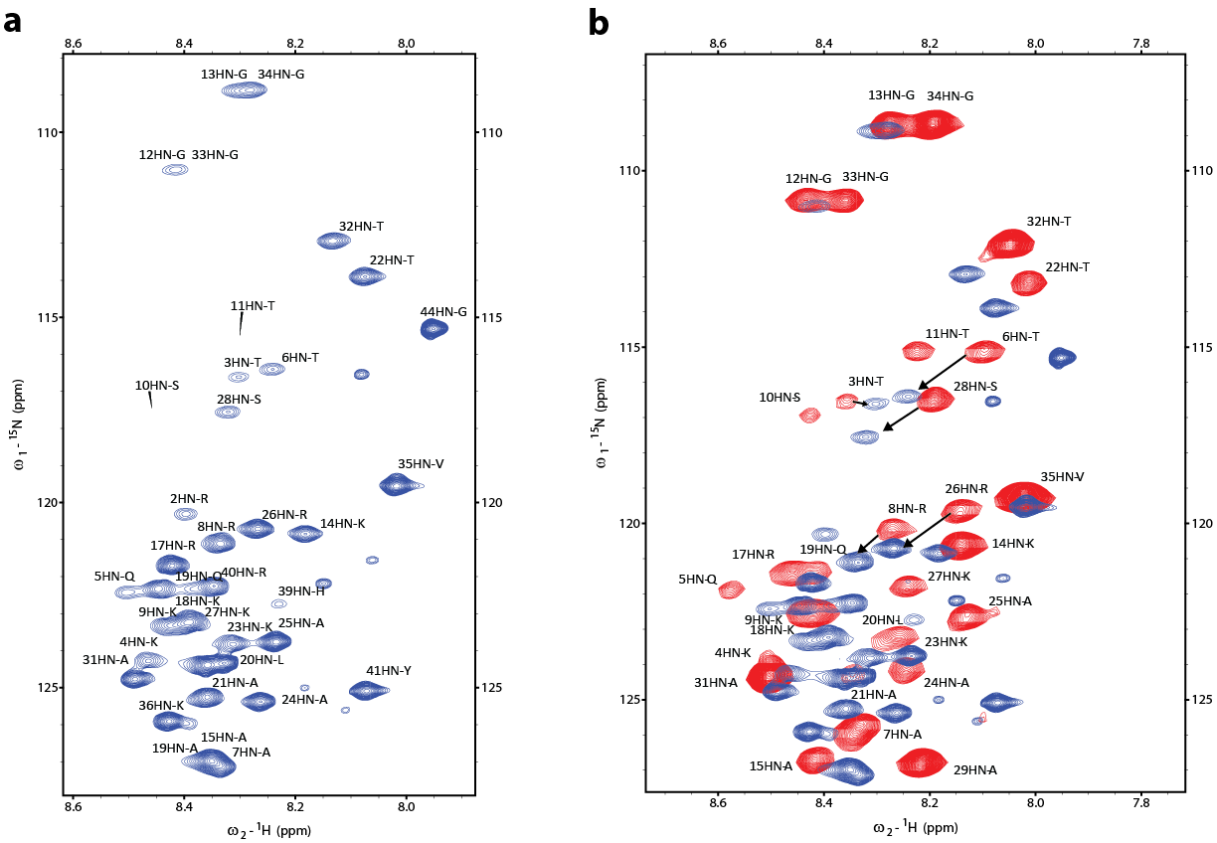


Figure 7. a ^{1}H-^{15}N HSQC of the H3 tail peptide. **b** Comparison of the ^{1}H-^{15}N HSQC of the H3 tail in a nucleosome context (red) and the H3 tail peptide (blue). Assignment for the nucleosome is shown and some large peak movements are shown by arrows, indicating clear differences in the structures adopted.

The H3 tail peptide (res 1-44) showed only 32 peaks, because of some overlap, in the ^{1}H-^{15}N HSQC which were assigned using CaCbNH and CaCbCONH spectra (**Figure 7a**). When comparing the ^{1}H-^{15}N HSQC of the H3 tail peptide with the H3 tail in the nucleosome, large discrepancies between the two became obvious (**Figure 7b**), suggesting that the nucleosomal context results in changes in the structure of the histone H3 tail.

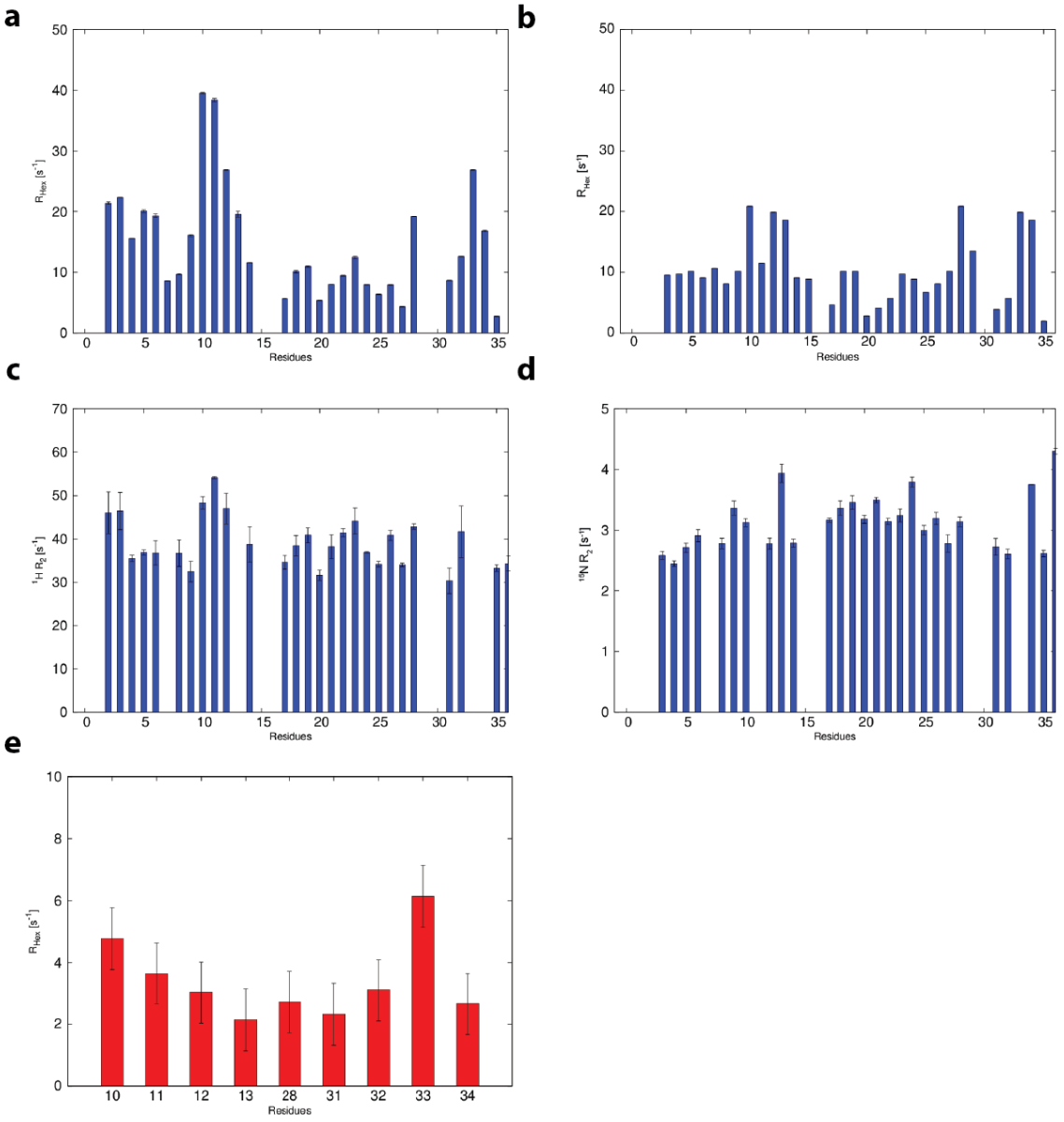


Figure 8. a R_{Hex} rates for the histone H3 tail peptide measured by the CLEANEX-PM sequence. **b** Calculated R_{Hex} rates using the Sphere web server (See experimental procedures). **c** Measured ^{1}H R_2 rates for the histone H3 tail peptide. **d** Measured ^{15}N R_2 rates for the histone H3 tail peptide. **e** R_{Hex} rates for the nucleosomal H3 tail

To compare residual secondary structure between the isolated H3 tail peptide and the H3 tail in a nucleosome, ^{1}H and ^{15}H R_2 relaxation rates were measured for a H3 tail peptide. Unstructured polypeptides can have fast proton exchange rates with water, which can affect R_2 determination.[69] The CLEANEX-PM pulse sequence was used to measure the water exchange rate (R_{Hex}) at 15 °C.[70] The R_{Hex} rates are large across the chain and range from 10-40 s^{-1} (**Figure 8a**). This is slightly higher than the calculated rates at ph 6.5, which range from 10-20 s^{-1} (**Figure 8b**). R_{Hex} corrected ^{1}H and ^{15}N R_2 relaxation rates showed a flat profile (**Figure 8c** and **d**) indicating no residual structure in the isolated H3 tail peptide.

The R_{Hex} rates for the H3 tail in the nucleosome were also measured (**Figure 8**e). For most residues this R_{Hex} is too small for accurate quantification implying significant protection in the context of the nucleosome, for these maximum R_{Hex} of 1.2 s^{-1} was estimated from the S/N. R_{Hex} rates of 2.5-5 s^{-1} were obtained for residues 10-13 and 28-34 (**Figure 8**e). These residues in the nucleosome have the lowest R_2 rates, suggesting these are in flexible unstructured regions. This is in agreement with larger R_{Hex} rates for these residues. All R_{Hex} rates are significantly lower than in the peptide, suggesting increased protection of the amides by secondary structure formation or complexation by the DNA.

Chemical shifts for N, HN, CO, C_α and C_β were also measured for the isolated H3 tail peptide. From these, SCS were calculated for the H3 tail peptide by subtracting the random coil values for each residue. For the H3 tail peptide, no clear α-helical propensity was observed (**Figure 9**). Especially the $\Delta\delta C_\alpha$-$\Delta\delta C_\beta$ values (**Figure 9c**), which are considered the best indication of secondary structure, showed only very small deviations from the random coil values. Again, this suggests that the Histone H3 tail attains structure in a nucleosomal context, while being unstructured as an isolated peptide.

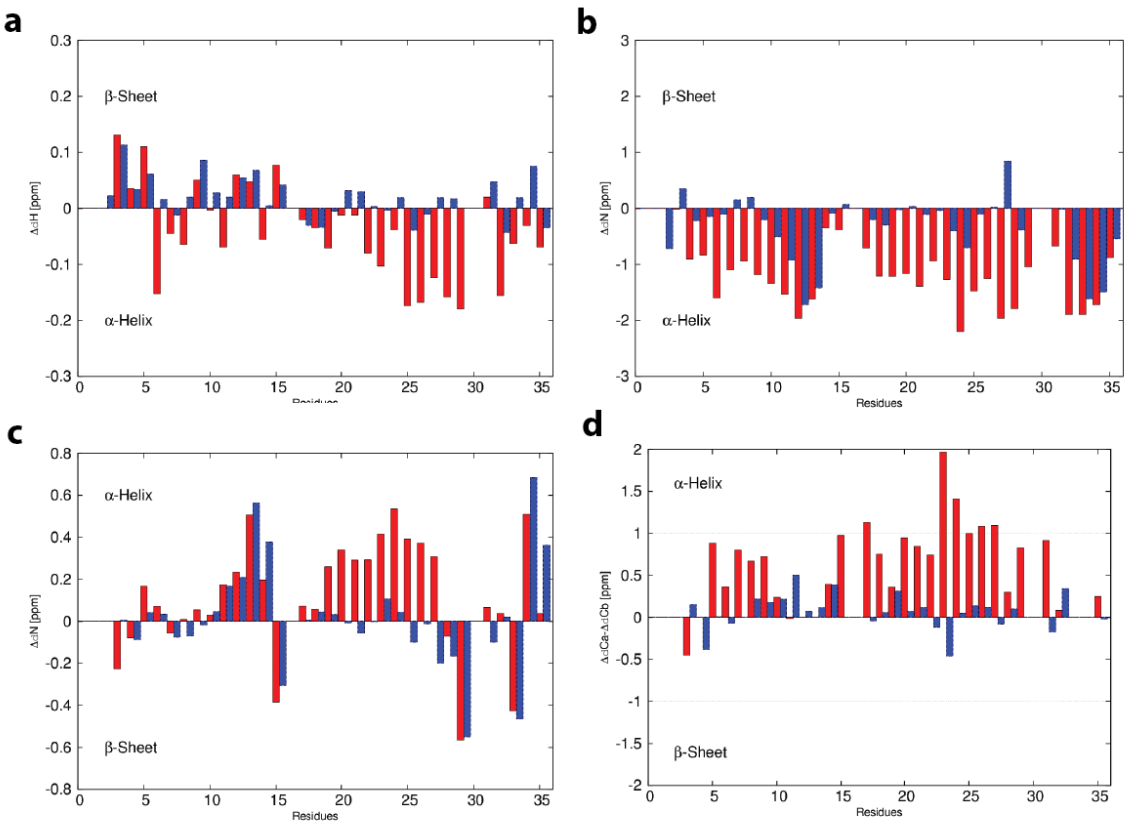


Figure 9. SCS analysis of the ΔδH (**a**), ΔδN (**b**), ΔδCO (**c**) and $\Delta\delta C_\alpha$-$\Delta\delta C_\beta$ (**d**) chemical shifts of the isolated H3 tail peptide (blue) and the H3 in a nucleosome (red).

To better visualize the effect of the nucleosome core on tail structure, the isolated H3 tail was assumed as the "random coil" value and SCS values for HN, N, CO and C_α-C_β chemical shifts were calculated by taking the difference between the chemical shifts of the H3 tail in a nucleosome context and the H3 tail peptide (**Figure 10**). All of these plots revealed increased α-helical character of the H3 tail when in a nucleosome context, especially for residues 19-27.

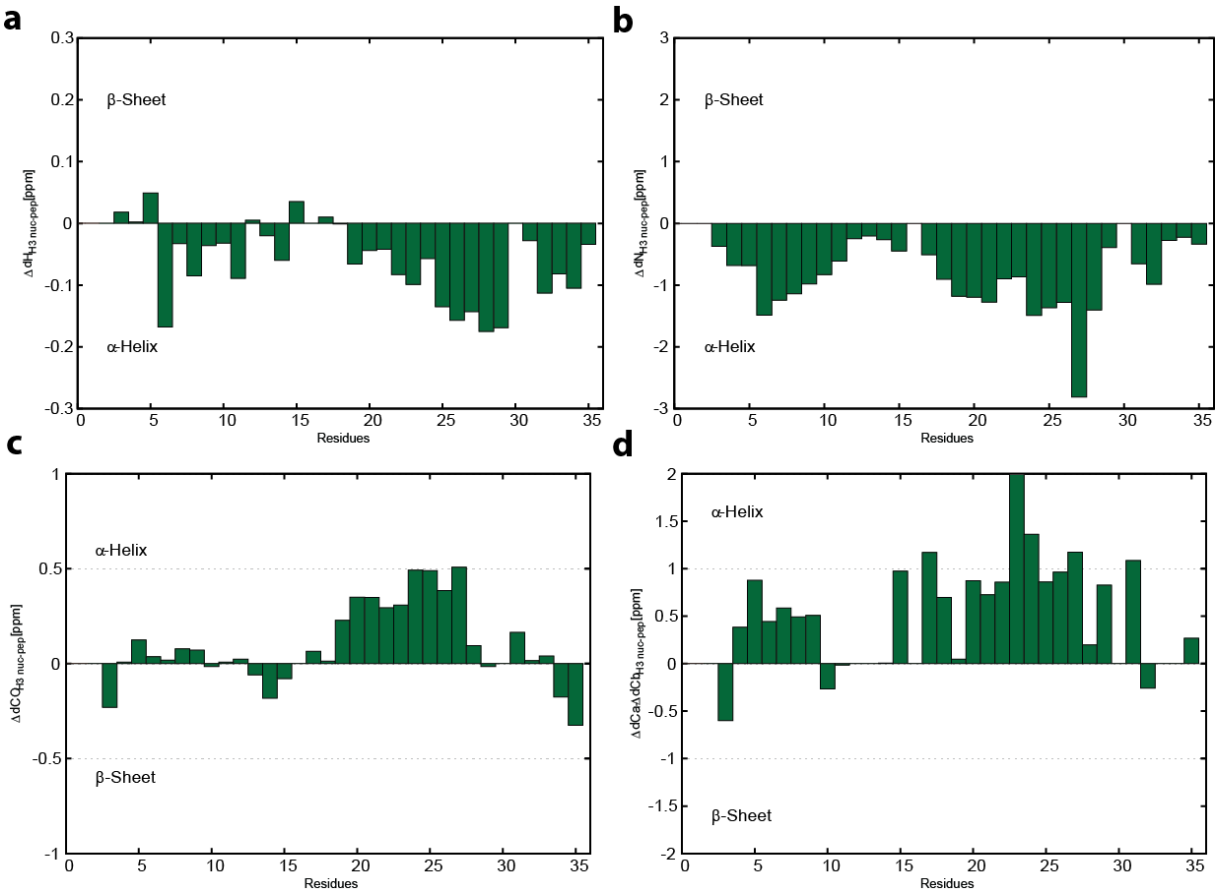


Figure 10. SCS analysis of the ΔδH (**a**), ΔδN (**b**), ΔδCO (**c**) and $\Delta\delta C_\alpha$-$\Delta\delta C_\beta$ (**d**) chemical shifts of the H3 tail in a nucleosome using the H3 tail peptide chemical shift as random coil values as given by $\Delta\delta = \delta_{nucleosome} - \delta_{peptide}$.

The observed difference between the H3 tail peptide and the H3 in a nucleosomal context could be due to interaction of the tail with the nucleosome core but could also be an artefact because only a small part of histone H3 was used in the measurements. To assess whether the H3 tail peptide is a good model for the whole H3 protein, H3/H4 tetramers with ^{15}N labelled H3 were reconstituted for comparison. If the secondary structure in the tail is induced by association with the DNA, then the H3/H4 tetramer spectra should be similar to the tail. On the other hand, if the tail is unstructured just because part of the H3 protein is removed, the H3/H4 tetramer spectrum may be expected to be similar to the nucleosome. Overlay of the ^{1}H-^{15}N HSQC of the H3 tail peptide and the H3/H4 tetramer reveals a large overlap for residues 2-36 between this complex and the H3 peptide (**Figure 11a** and **b**). When analysing the ^{1}H and ^{15}N chemical shift difference between the tetramer and the H3 tail peptide only very small and unsystematic changes in the chemical shift between the two were

observed (**Figure 11c**). These results suggest that the observed secondary structure in the H3 tail is induced only in the presence of the whole nucleosome, probably due to H3 tail – DNA interactions.

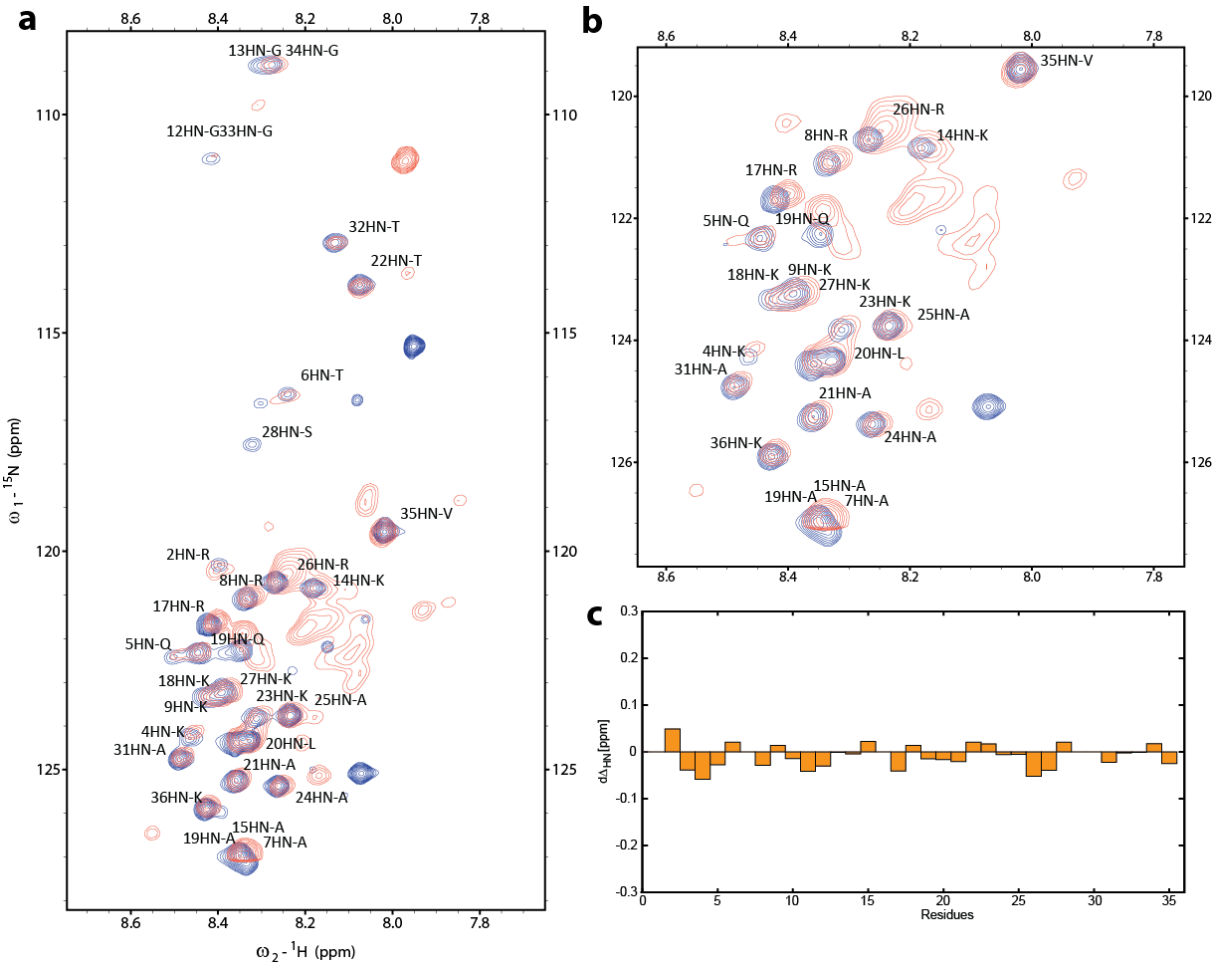


Figure 11. a / b Overlay of the ^{1}H-^{15}N HSQC of the histone H3 tail peptide (blue) and the H3/H4 tetramer (red) containing a labelled H3. **c** Difference between mean HN chemical shifts of the H3 tail peptide and the H3/H4 tetramer.

6.2.3 The histone tail adopts a compact conformation

To gain a more detailed understanding of the structure of the H3 tail in the context of the nucleosome, paramagnetic relaxation enhancement (PRE) measurements were used. For PRE measurements, a free electron moiety in the form of a stable nitroxide radical can be installed by reaction of a free cysteine in the protein with (1-Oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate via disulfide exchange (**Figure 13**). ^{1}H R_2 rates are then measured with the spin label ($R_{2\,ox}$) and after reduction of the spin label by addition of ascorbic acid ($R_{2\,red}$). PRE rates are given by $\Gamma_2 = R_{2\,ox} - R_{2\,red}$.[71]

Two independent measurements were performed, with the spin label installed close to the DNA on the H3 tail using an H3 Lys-36-Cys variant or on the core of the nucleosome using a H4 Gln-27-Cys variant (**Figure 12**).

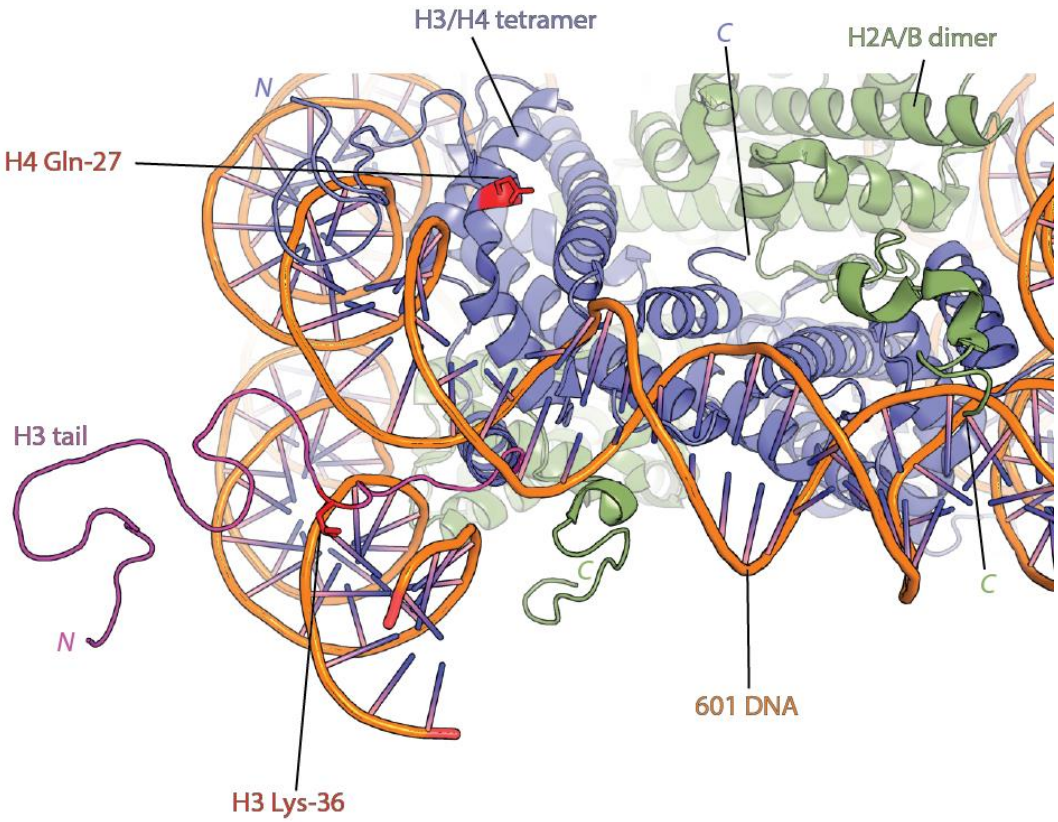


Figure 12. Sites of PRE spin labels on the nucleosome.

The H3 K36C variant was produced as deuterated and ^{15}N labelled protein in *E. coli* and purified by SEC and IEXC. The H4 Q27C variant was produced unlabelled and the nucleosomes with this variant were reconstituted with deuterated ^{15}N labelled wt H3.

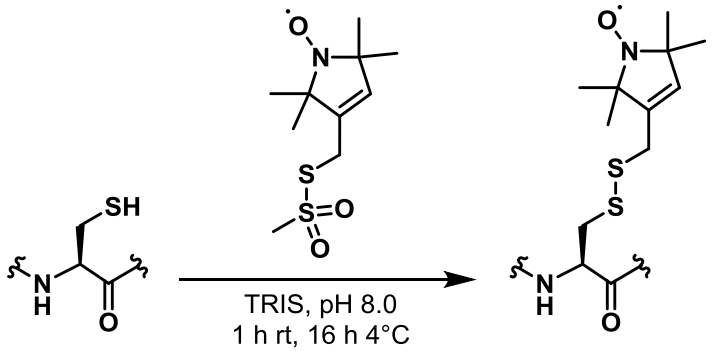


Figure 13. Modification reaction to generate the spin-labelled histone.

For an intrinsically disordered tail, the PRE effect at >20 amino acids from the spin label should be negligible.[72-74] When placing the spin label at position 36 in the H3 tail a flat profile of large PRE rates of 40-50s^{-1} for residue 3-15 was observed (**Figure 14**a), while residues 17-35 were bleached completely, suggesting even higher PRE rates. This flat profile is not in agreement with a random coil but suggests a compact structure of the histone H3 tail.[73] When placing a spin label on the surface of the nucleosome core, a PRE effect was observed for all residues with increased PRE rates at the *N*-terminus of the tail (**Figure 14b**). The distance of residue 35, close to the DNA and modelled in a nucleosome crystal structure, to the spin-label site is 34.1 Å in the model and 32.8 Å when calculated with the measured PRE values, confirming the accuracy of this method for distance determination. The observation that the PRE is larger for residues at the *N*-terminus than for residues closer to the nucleosome core strongly suggests that the tail is associating, at least transiently, with the DNA to form a compact conformation.

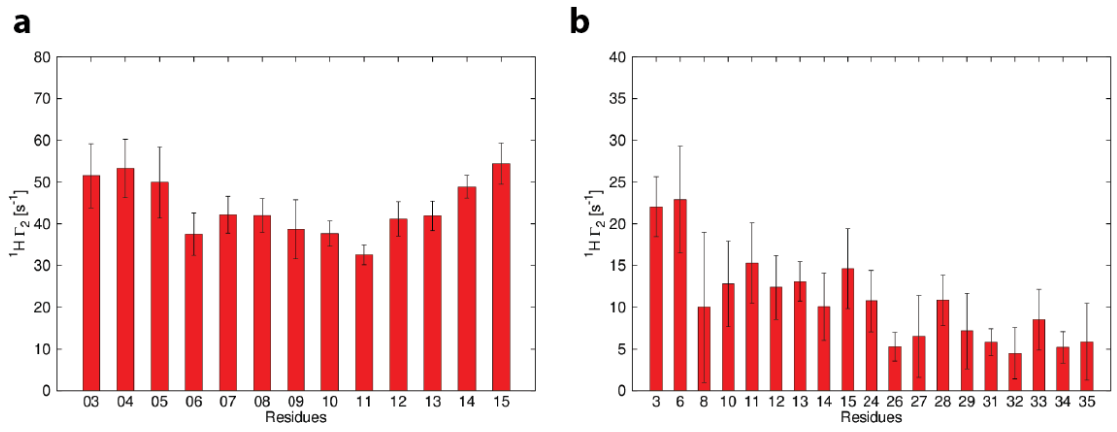


Figure 14. a Measured PRE rates for the spin label at position H3 C36. **b** Measured PRE rates for the spin label at position H4 C27.

The PRE rates were then used as distance restraints for structure calculation using XPLOR-NIH.[75,76] PRE rates were introduced as restraints using the PRE potential.[77,78] When no PRE rates were obtained because of complete bleaching of the signal, the minimum PRE was estimated from the intensity in the reduced spectra and the S/N in the oxidized spectra. Because these rates are minimum PRE values they were converted to distances which were used as maximum distance restraints using the NOE potential. The calculation script was based on example scripts from the XPLOR-NIH package and the scripts used by Religa *et al.*[79]. When using the PREs as restraints to calculate a single structure, no structures that satisfied all constraints were obtained. Due to the flexible nature of the histone H3 tail, the measured PRE values likely represent an average over multiple conformations. To take this into account, ensemble calculations with ten members were performed. Using these settings, structural ensembles that satisfied all restraints were obtained. The obtained structures revealed a compact conformation of the nucleosome tail around the entry and exit site of the DNA (**Figure 15**). Some α-helical character was observed in members of the ensemble in agreement with the chemical shift results.

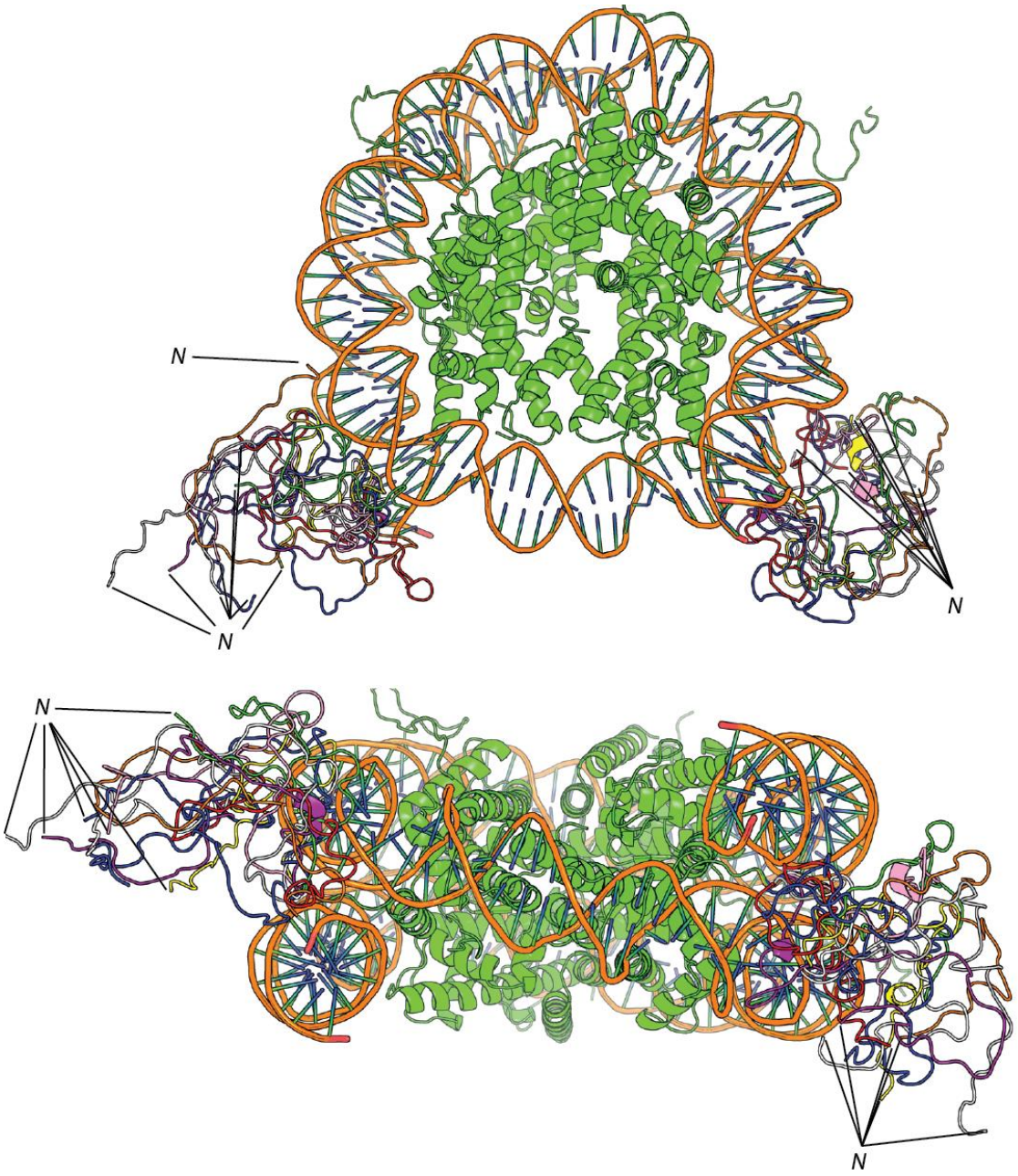


Figure 15. Top and front view of a calculated ensemble of 10 structures of the histone H3 tail using the measured PRE values as restrainted. The nucleosome core is in gree and the H3 tail of each ensemble member is in a different color.

All measurements on the nucleosome thus far described were performed at low ionic strength because these buffer conditions resulted in good quality spectra. To investigate the structure at physiological ion strength, a nucleosome sample in buffer containing 150 mM NaCl was prepared. The ^{1}H-^{15}N HSQC revealed significant peak shifts from the low-ionic strength nucleosome (**Figure 16a**). When comparing the histone peptide with the low and high ionic strength nucleosome spectra,

the high ionic strength peaks lay on a straight line between the histone peptide and the low ionic strength spectra. This suggests fast interconversion between an extended tail peptide like state and a collapsed state. From peak positions a 50 ± 6 % contribution of the compact conformation to the structural ensemble was estimated.

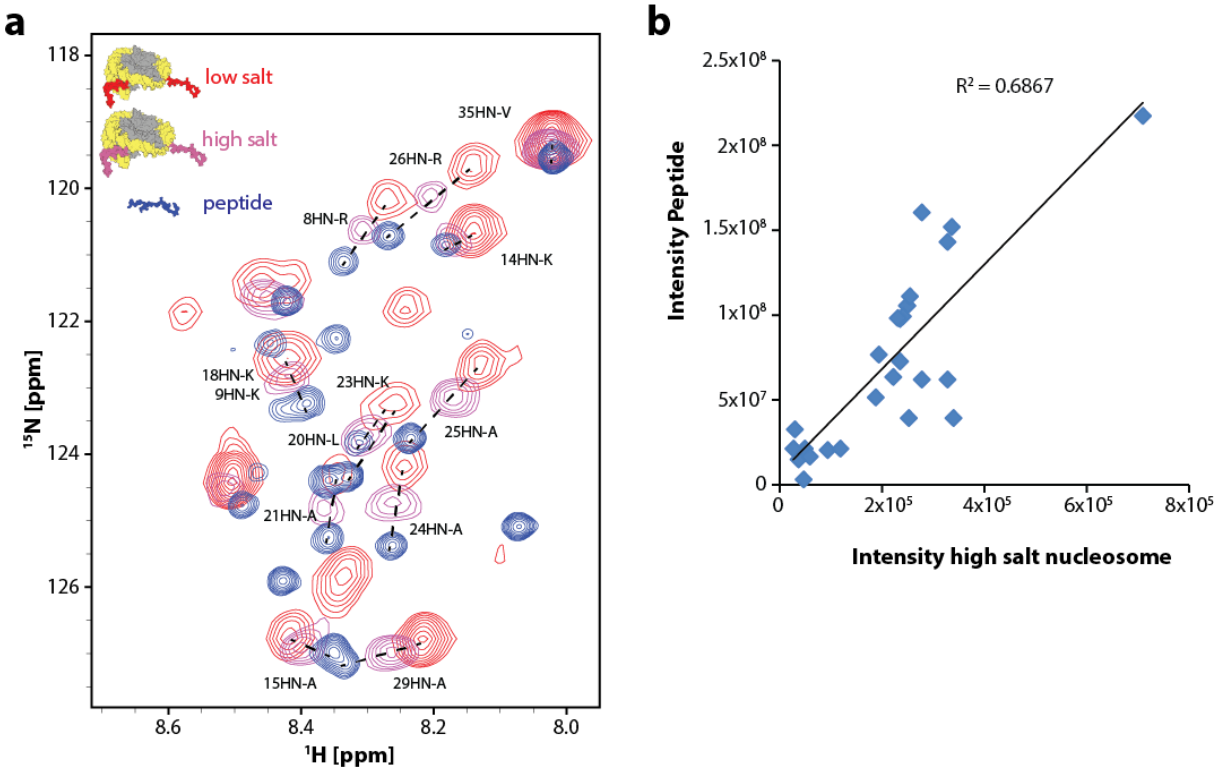


Figure 16. a Overlay of the ^{1}H-^{15}N HSQC spectra of the histone H3 tail peptide (blue), the low ionic strength nucleosome (red) and the high-ionic strength nucleosome (purple). Peak shifts are indicated by dotted lines for the labelled peaks. **b** Correlation between the peak intensity for the peptide and the nucleosome at high ionic strength.

Thus, the previously characterized compact state contributes significantly to the tail structure at physiological ionic strength. Further detailed quantification of the tail dynamics at physiological ionic strength was impossible due to very poor S/N, possibly because of increased water exchange in the extended conformation. In agreement with this, the residues with the highest water exchange rate in the peptide (res 3-6, 10-12 and 28) were not observed in the nucleosome spectra at higher ionic strength and the peak intensity of the high salt nucleosome correlates well with the peak intensity of the peptide (**Figure 16b**; R^2=0.68).

6.2.4 Phosphorylation modulates the tail-nucleosome interactions

Next, it was of interest to investigate the possibility of tail structure modulation by PTMs. Phosphorylation could influence the proposed to be mainly electrostatic interaction of the nucleosome tail with the DNA. Of the known phosphorylation sites on histone H3, Ser-10 is the best studied and can be introduced enzymatically.[80,81]

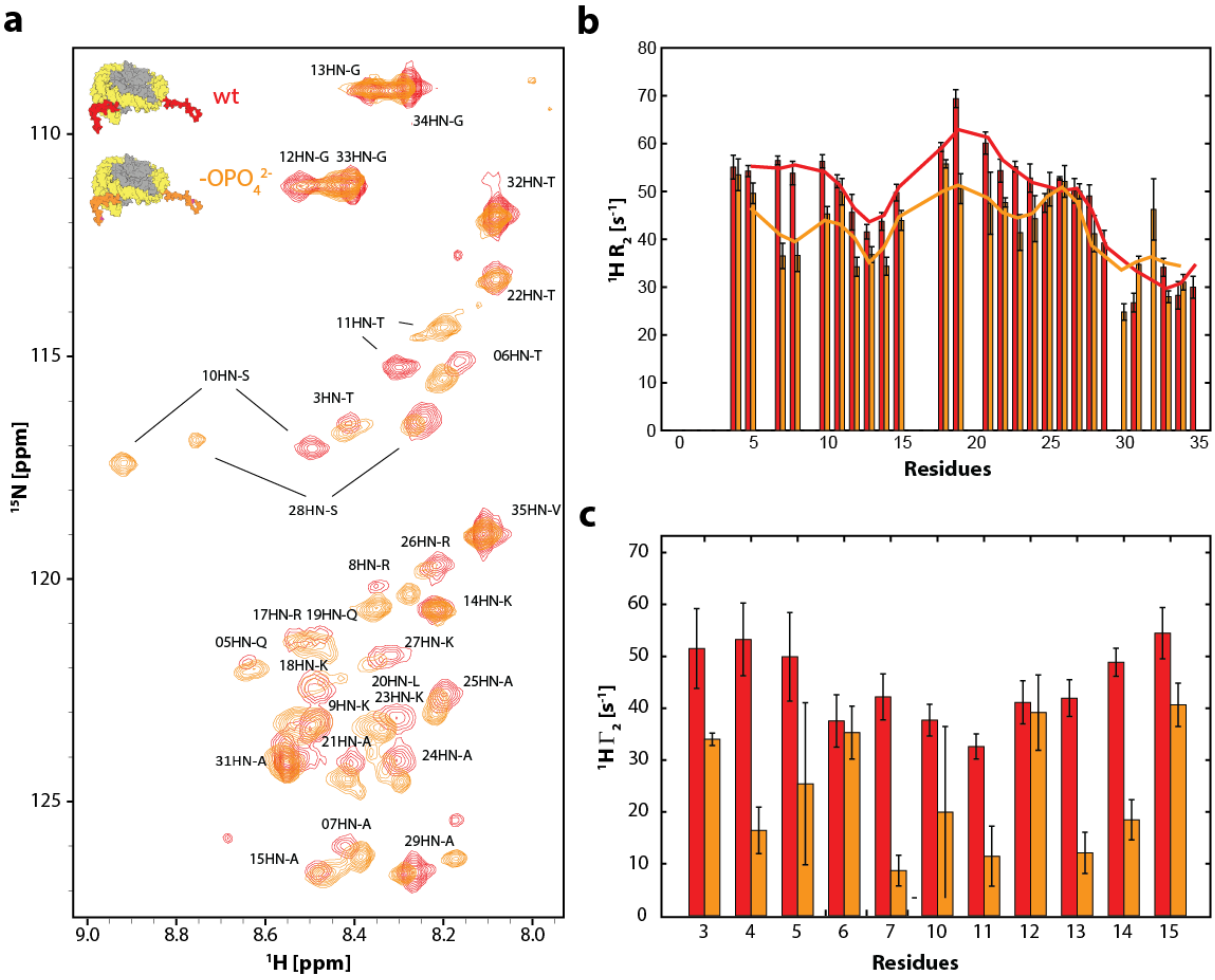


Figure 17. a Overlay of the ^{1}H-^{15}N HSQC spectra for the wt (red) and the phosphorylated (orange) H3 tail in the nucleosome context. **b** Comparison of the measured ^{1}H R_2 rates from the wt nucleosome (red) and the phosphorylated nucleosome (orange). **c** Comparison of the measured PRE rates for the wt (red) and the phosphorylated nucleosome (orange).

When treating the wt nucleosome with the kinase AuroraB,[37] phosphorylation at Ser-10 and, after prolonged incubation, at Ser-28 was observed (**Figure 17a**).[80] To only characterize the effect of Ser-10 phosphorylation, the reaction was stopped at around 50% Ser-10 phosphorylation to ensure no Ser-

28 phosphorylation is present. Then ^{1}H R_2 and PRE rates of the phosphorylated H3 tail were measured.

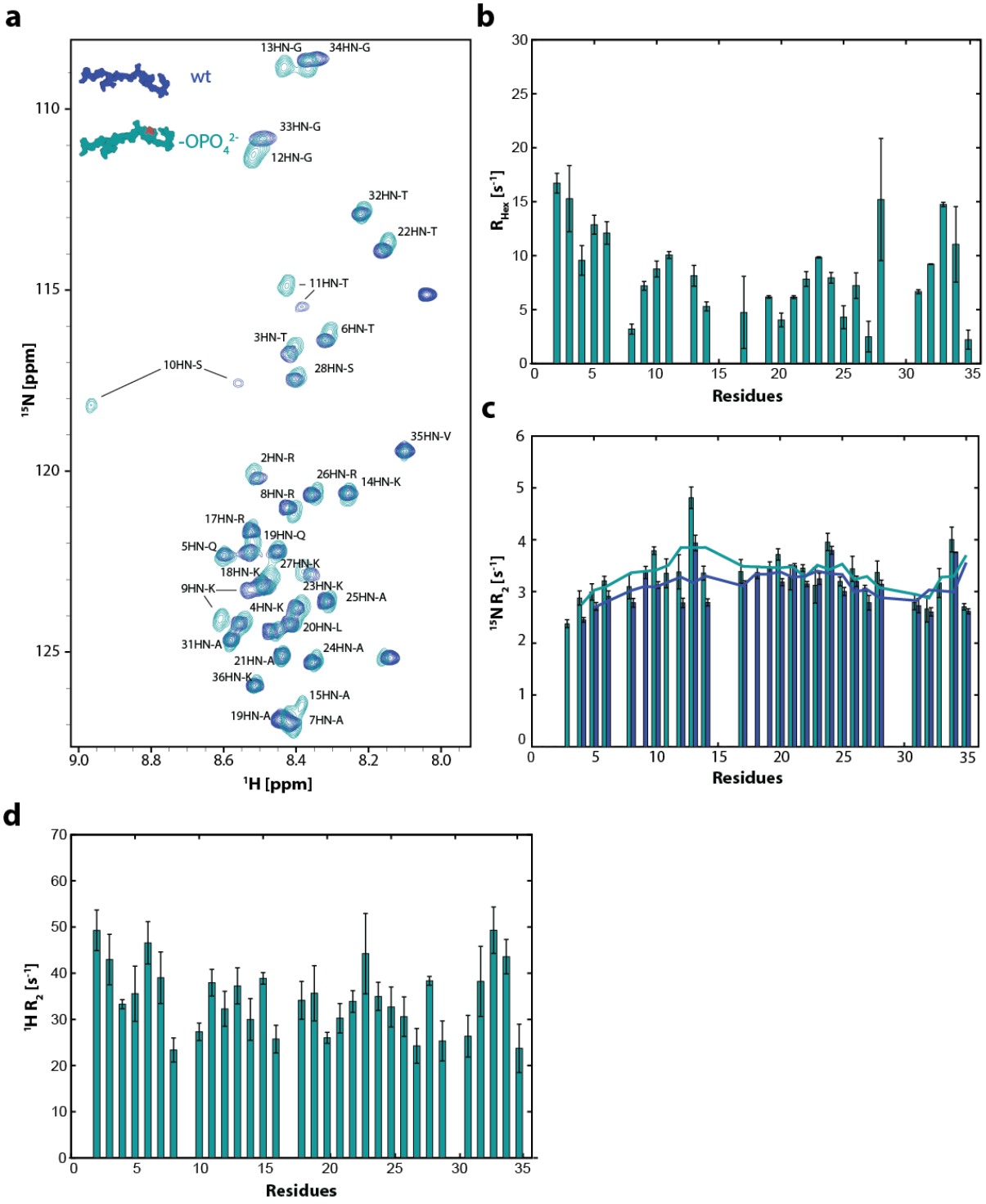


Figure 18. a Overlay of the ^{1}H-^{15}N HSQC spectra of the phosphorylated (cyan) and the wt (blue) H3 histone tail peptide. **b** Measured R_{Hex} rates for the H3 tail peptide. **c** Comparison between the ^{15}N R_2 rates for the wt H3 tail peptide (blue) and the phosphorylated H3 tail peptide (cyan). **d** Measured ^{1}H R_2 rates for the phosphorylated H3 tail peptide.

When comparing ^{1}H R_2 rates of the phosphorylated and the wt nucleosomes, a reduction in the ^{1}H R_2 for residues 3-15 and 19-24 is observed (**Figure 17b**). The measured PRE rates are also significantly lower for the most quantifiable residues (**Figure 17c**), suggesting a more extended conformation of the tail. This suggests that phosphorylation results in reduced residual structure in the H3 tail, as indicated by reduction of R_2, and in a less compact conformation, resulting in lower PRE rates.

To assess how phosphorylation changes the H3 tail – DNA interaction, the H3 tail peptide was phosphorylated using the AuroraB kinase[80] for comparison to the phosphorylated nucleosome. A fully Ser-10 phosphorylated tail peptide, with no observed Ser-28 phosphorylation was obtained after 36 h reaction with AuroraB at 30 °C (**Figure 18a**). R_{Hex} and ^{1}H and ^{15}N R_2 rates of the phosphorylated state were determined (**Figure 18b-d**). The R_{Hex} rates showed a reduction around the site of phosphorylation, especially residues 10-11, which previously had the highest R_{Hex} rates (**Figure 18b**). When analysing the ^{15}N R_2 relaxation rates, no large differences between the phosphorylated and non-phosphoylated form were observed (**Figure 18c**). In the nucleosome, large differences in the R_2 rates were observed, the lack of an effect of phosphorylation on R_2 rates in the isolated H3 tail peptide, suggests that phosphorylation of the H3 tail changes the histone tail – nucleosome interactions.

6.3 Discussion

In this chapter, detailed analysis of the H3 tail structure was performed using NMR spectroscopy. Via the use of relaxation properties, chemical shifts and PRE rates a detailed understanding of how the H3 tail behaves was obtained. Two regions (res 3-10 and 17-28) with increased R_2 rates and SCS indicating residual α-helical character were observed. These are separated from each other and from the nucleosome core by two flexible regions (res 11-15 and 29-35 respectively), with lower R_2 rates and no secondary obvious secondary structure content indicated by SCS. These features characteristic of residual secondary structure are only observed in the H3 tail in a nucleosome context, not in the isolated H3 tail peptide, indicating that the structure is induced by interactions of the H3 tail with the nucleosome. Comparing the spectra of the isolated H3 tail peptide and the H3/H4

tetramer revealed only minor chemical shift differences between the two, indicating that the H3 tail structure is induced by association of the H3 tail with the nucleosomal DNA in a nucleosome context. By measuring the nucleosome spectra at higher ionic strength it is shown that this compact conformation at low ionic strength contributes significantly to the tail structure at physiological ionic strength. Upon phosphorylation a reduction in the R_2 rates in the parts of the H3 tail with residual secondary structure and lower PRE rates are observed. This suggests that the tail is becoming more flexible and more extended from the nucleosome core.

Increased flexibility of the H3 tail by phosphorylation could have two possible functional roles. Weakening the DNA-tail interaction results both in a more flexible, and thus binding-capable tail, and in less tightly associated DNA, which is free to bind to other proteins. When chromatin is highly condensed, such as during mitosis, histone tails have been proposed to be released from nucleosomes.[35,82] Release of the H3 tail from the nucleosome by phosphorylation would favour such a conformation. This is in agreement with previous reports that show that H3 S10Ph is necessary for correct chromatin condensation during mitosis. In interphase, release of the H3 S10Ph tail would allow both H3-tail binding proteins and DNA binding proteins enhanced access to their target. Binding of DNA associating proteins[83] and polyamides[84] is reduced by the presence of the H3 and to a lesser extent the H4 tail. When a transcription factor IIIA (TFIIIA) binding site is placed at the entry/exit of the nucleosome, where the H3 tail is located in our structural ensemble, TFIIIA binding is barely detectable in the wt nucleosome.[83] Upon removal of the H3 or the H3 and H4 tail, binding is drastically enhanced, suggesting that the H3 tail binds to the DNA site competing with TFIIIA. This inhibition is only seen in the full nucleosome, not in a reconstituted H3/H4 tetramer – DNA complex, consistent with the nucleosome context being necessary for the H3 – DNA interactions.

All histone tails have been shown to have distinct features and roles in the regulation of chromatin structure. Using the described experimental methodology it should be possible to distinctly map the structure, dynamics and interactions of each tail in a nucleosome context. Also, histone tails have been shown to have high affinity for free linker DNA between nucleosomes.[23] Future studies to

distinguish intra-, inter-nucleosome and linker DNA association will shed more light on this. In addition, association of other proteins, including histone H1, with the nucleosome has been shown to reorganize tail-DNA association.

6.4 Materials and methods

For mutagenesis, general cloning procedures, histone expression, LC-MS analysis of protein modification, histone complex reconstitution and purification, nucleosome reconstitution and native PAGE analysis see the General Methods described in Chapter 2. Primers used for cloning and mutagenesis are given in **Table 1**. Plots were generated using gnuplot.[85]

Name	Sequence (3'-5')
H3_F	GCCCGTACCAAGCAGACCGCC
H3_R	AGTGCG CTCGAG CTA GCC GGG CCG GTA ACG GTG AGG
SUMO_F	ATGCTC CAT ATG GGC AGC AGC CAT CAT CAT CAT CAT C
SUMO_R	GGC GGT CTG CTT GGT ACG GGC TTG GAA GTA CAG GTT TTC CTC GAT ACC ACC
H3K36C_F	CTGCTACCGGCGGAGTCTGCAAACCTCACCGTTACCG
H3K36C_R	CGGTAACGGTGAGGTTTGCAGACTCCGCCGGTAGCAG
H4Q27C_F	GTTCTGCGTGACAACATCTGCGGTATCACCAAGCCGGCT
H4Q27C_R	AGCCGGCTTGGTGATACCGCAGATGTTGTCACGCAGAAC

Table 1. Primers used for mutagenesis and cloning.

6.4.1 Minimal media histone production

The labelled expression was performed as described by Lundstrom *et al.*[61] with minor modifications. The plasmid encoding the H3 or H3 mutants were transformed into BL21(DE3)pLysS and grown on LB-agar plates containing ampicillin (100 mg/L) and chloramphenicol (34 mg/L). The next afternoon 2x 15 mL LB in falcon tubes containing the same antibiotics were inoculated with one colony from the plate. These were grown for 2-3 h and then centrifuged (3750rpm, 10 min) and the cells were resuspended in 50 mL M9 medium (containing $^{15}NH_4Cl$, and D_2O and deuterated ^{12}C or ^{13}C glucose and the same antibiotics). These were grown ON at 30 °C and then 50 mL of pre-culture was added to 500 mL M9 medium (with the same labels and antibiotics). These cultures were grown at 37 °C until OD_{600} = 0.6-0.8 and then expression was induced by addition of 0.5 mM IPTG. Expression was

continued at 30 °C for 12-14h after which cells were harvested by centrifugation. Pellets were then processed as described in the general methods.

6.4.2 Overlap extension PCR

To ensure good expression of the H3-tail peptide a SUMO-H3 fusion gene was constructed using overlap extension PCR.[86] The histone H3 (1-44) fragment and the his-tagged SUMO fragment were amplified from their respective vectors using the PfuUltra II Hotstart PCR Master Mix (Agilent, 600850) following the manufactures specifications. PCR product were analysed and purified by agarose gel electrophoreses (0.7-1%; 1xTEA buffer). DNA with the right molecular weight was retrieved using the gel extraction kit (Qiagen, 28704) and concentration was determined by Nanodrop. A short PCR cycle was then run with 0.1 pmol of each fragment without primers using the same polymerase master mix (conditions are given in **Table 2**). Then, the primers were added (0.1 µM) and the PCR was continued for another 25 cycles using the manufactures specifications. Final products were analysed and purified by agarose gel, digested and ligated into pET22b.

	Temp	Time
	95 °C	2 min
13x	95 °C	20 s
	60 °C	1 min
	68 °C	30 s
	72 °C	3 min

Table 2. Overlap PCR conditions.

6.4.3 H3 tail peptide production

The production of the labelled H3-tail peptide was performed as described by Lundstrom *et al.*[61] with minor modifications. The plasmid encoding the H3(1-45)-SUMO construct were transformed into BL21(DE3)pLysS and grown on LB-agar plates containing ampicillin (100 mg/L) and chloramphenicol (34 mg/L). The next evening 2-15 mL LB containing the same antibiotics were inoculated with one colony from the plate. These were grown overnight and then centrifuged and the cells were resuspended in 100 mL M9 medium (containing $^{15}NH_4Cl$ and the same antibiotics). These were grown for 4 h at 37 °C and then 50 mL of pre-culture was added to 700 mL M9 medium (containing $^{15}NH_4Cl$).

These cultures were grown at 37 °C until OD_{600} = 0.6-0.8 and then expression was induced by addition of 0.5 mM IPTG. Expression was continued at 18 °C for 16h after which the cells were harvested by centrifugation. Cell pellets were stored at -80 °C until purification.

6.4.4 H3 tail peptide purification

Cells were resuspended in 35 mL lysis buffer (1-PBS, 400 mM NaCl, 20 mM Imidazole, 5 mM BME, Complete protease inhibitor) and were lysed by sonication. The lysate was then cleared by centrifugation (20 min, 20k rpm, JA25.50) and transferred to a 50 mL super-loop. The lysate was loaded onto a 1 mL HisTrap HP (GE Healthcare; 17-5247-01) equilibrated in loading buffer (25 mM TRIS pH 7.5, 500 mM NaCl, 20 mM imidazole, 5 mM BME) at 1 mL / min. The protein was eluted using a linear gradient from 0 – 100 %B (loading buffer + 400 mM imidazole) in 20 CV. The fractions containing the desired protein were then concentrated to 1.5 mL and further purified by SEC (Superdex S75 16/60 in 25 mM TRIS pH 7.5, 300 mM NaCl, 1 mM DTT, 1 mM EDTA). Fractions containing the desired protein were concentrated to 1.5-2 mL, concentration was measured using nanodrop and TEV protease (1:20 Abs) was added. The cleavage was allowed to proceed overnight (16 h) and the mixture was then purified again by size exclusion (Superdex S75 16/60, 20 mM phosphate pH 7.0, 100 mM NaCl, 1 mM EDTA, and 5 mM BME). The fractions containing the H3 fragment were then pooled, concentrated and the buffer was exchanged by viva-spin if necessary.

6.4.5 Spin label reaction

For the attachment of a spin label,[87] the cysteine mutant (H3 K36C or H4 Q27C) containing nucleosome was first dialyzed into no-salt buffer containing no DTT for 4 h. Residual DTT was removed by viva-spin. Then 10 eq (1-Oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) Methanethiosulfonate (MTSL; Toronto research chemicals O875000; dissolved in MeCN) was added and left to incubate for 1h at rt and then over night at 4 °C.[87] Subsequently excess MTSL was removed by viva-spin. After R_2 rates were recorded with the spin label, the spin label was reduced by the addition of 10 mM sodium ascorbate and incubation at 4 °C overnight.

6.4.6 General NMR acquisition and processing

Measurements were performed on a Bruker Avance 700MHz spectrometer with Topspin 3.2 or on a Varian 600MHz spectrometer. Data was processed using NMRpipe[88] and Sparky.[89] Peak and exponential fitting was performed using FuDa.[55] Samples were measured in 3 mm bruker match tubes (NORS305001, Aldrich) or Shigemi 5 mm Symmetrical NMR microtube (8 mm length, Z543349, Sigma). Unless otherwise stated, spectra were measured in 10 mM $NaHPO_4$ pH 6.5, 1 mM EDTA and 2x complete protease inhibitor. Nucleosome ^{1}H-^{15}N HSQC spectra (including ^{1}H and ^{15}N R_2 and CLEANEX-PM measurements) were recorded at 298 K with 1024 (^{1}H)-80 (^{15}N) complex points and a sweep width of 20 ppm in the indirect dimension. For H-N HSQCs the standard bruker pulse sequence "hsqcetfpf3gp" was used. A recycle delay of 1s was used. Water-suppression was accomplished by use of a flip-back pulse and use of gradients. For 32 scans in the direct dimensions the acquisition time was ~36 min.

6.4.7 ^{1}H R2 measurements

^{1}H R_2 rate measurements were performed with a custom pulse sequence based on the standard broker ^{1}H-^{15}N HSQC (hsqcetfpf3gp) with the variable delay in the initial INEPT based on a sequences reported by Mal *et al.*[90]. With 160 scans in the direct dimension the total acquisition time was 18 h. For nucleosome samples, delays of 6, 9, 12 and 15 ms were used. For the H3 peptide, delays of 8, 12, 16 and 20 ms were used. Rates were extracted using FuDa.

6.4.8 ^{15}N R2 measurements

^{15}N R_2 rate measurements were performed using the HSQCT2ETF3GPSI standard bruker pulse sequence. A recycle delay of 5 s was used. With 64 scans in the direct dimension the total acquisition time for 4 planes was 24 h. The set delay period of 16.96 ms was repeated 1, 5, 10 and 15 times. The separate 2D planes with different delays were combined and processed using nmrpipe and peak fitting was performed with FuDa.

6.4.9 PRE calculations

PRE rates are calculated from ^{1}H R_2 rates by $\Gamma_2 = R_{2\,red} - R_{2\,ox}$.

6.4.10 CLEANEX measurements

CLEANEX-PM FHSQCs with different delays were recorded using the bruker pulse sequence FHSQCCXF3GPPH. The FHSQCF3GPPH pulse sequence was used to measure the maximum intensity. Delays of 5, 10, 15 and 20 ms were used. The separate planes were combined using nmrpipe and peaks were fit using FuDa. Peak intensity at different delays (I_D) were then divided by the intensity without delay (I_0). Obtained $\frac{I_D}{I_0}$ values were fit using $\frac{I_D}{I_0} = \frac{k}{R_{1A}+k-0.6}(e^{-0.6*t} - e^{-(R_{1A}+k)*t})$.[70]

6.4.11 Theoretical R_{Hex} calculation

The R_{Hex} values were calculated using the Sphere web server (http://www.fccc.edu/research/labs/roder/sphere/sphere.html) which uses reported empirical rates and correction factors.[91,92]

6.4.12 Structure calculations

Calculations were performed using XPLOR-NIH[75,76] using the PRE potential.[77] Scripts were based on the workflow described by Religa *et al*.[79]

6.4.13 Chemical shift analysis

Random coil chemical shifts were calculated using a webserver at http://spin.niddk.nih.gov/bax/nmrserver/Poulsen_rc_CS.[93-95] Secondary chemical shifts are given by $\Delta\delta = \delta_{oberved} - \delta_{random\,coil}$.

6.4.14 Phosphorylation with AuroraB

Phosphorylation reactions were performed in 10 mM TRIS pH 7.5, 30 mM NaCl, 1 mM DTT, 3 mM $MgCl_2$, 2 mM ATP, 10 mM β-glycerolphosphate. 10-20 µM of nucleosomes were incubated with 3 µg AuroraB kinase (A31-10G, SignalChem) for 36 h at 30 °C. The reaction was then concentrated and buffer exchanged into NMR buffer (10 mM $NaHPO_4$ pH 6.5, 1 mM EDTA and 2x complete protease inhibitor) before measurements. For the H3 peptide phosphorylation, 0.5 µg AuroraB were used.

6.5 References

(1) Tan, S.; Davey, C. A. *Curr. Opin. Struct. Biol.* **2011**, *21*, 128.
(2) Andrews, A. J.; Luger, K. *Annu. Rev. Biophys.* **2011**, *40*, 99.
(3) Bannister, A. J.; Kouzarides, T. *Cell Res.* **2011**, *21*, 381.
(4) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251.
(5) Potoyan, D. A.; Papoian, G. A. *Proc. Natl. Acad. Sci.* **2012**, *109*, 17857.
(6) Zheng, C.; Hayes, J. J. *Biopolymers* **2003**, *68*, 539.
(7) Pedro Romero, Z. O., Charles R. Kissinger, J. Ernest Villafranca, Ethan Garner, Stephen Guilliot, A. Keith Dunker *{Biocomputing},* **1998**, *3*, 437.
(8) Jensen, M. R.; Zweckstetter, M.; Huang, J. R.; Blackledge, M. *Chem Rev* **2014**, *114*, 6632.
(9) Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N. *Chemical Reviews* **2014**.
(10) Walker, I. O. *Biochemistry* **1984**, *23*, 5622.
(11) Cary, P. D.; Moss, T.; Bradbury, E. M. *Eur. J. Biochem.* **1978**, *89*, 475.
(12) Hilliard, P. R.; Smith, R. M.; Rill, R. L. *J. Biol. Chem.* **1986**, *261*, 5992.
(13) Schroth, G. P.; Yau, P.; Imai, B. S.; Gatewood, J. M.; Bradbury, E. M. *FEBS Lett.* **1990**, *268*, 117.
(14) Dong, F.; Nelson, C.; Ausio, J. *Biochemistry* **1990**, *29*, 10710.
(15) Ballestar, E.; Franco, L. *Biochemistry* **1997**, *36*, 5963.
(16) Yang, C.; van der Woerd, M. J.; Muthurajan, U. M.; Hansen, J. C.; Luger, K. *Nucleic Acids Res.* **2011**, *39*, 4122.
(17) Greulich, K. O.; Ausio, J.; Eisenberg, H. *J. Mol. Biol.* **1985**, *186*, 167.
(18) Mangenot, S.; Leforestier, A.; Vachette, P.; Durand, D.; Livolant, F. *Biophys. J.* **2002**, *82*, 345.
(19) Mangenot, S.; Raspaud, E.; Tribet, C.; Belloni, L.; Livolant, F. *Eur. Phys. J. E Soft Matter* **2002**, *7*, 221.
(20) Bertin, A.; Leforestier, A.; Durand, D.; Livolant, F. *Biochemistry* **2004**, *43*, 4773.
(21) Bertin, A.; Renouard, M.; Pedersen, J. S.; Livolant, F.; Durand, D. *Biophys. J.* **2007**, *92*, 2633.
(22) Lee, K.-M.; Hayes, J. J. *Proc. Natl. Acad. Sci.* **1997**, *94*, 8959.
(23) Angelov, D.; Vitolo, J. M.; Mutskov, V.; Dimitrov, S.; Hayes, J. J. *Proc. Natl. Acad. Sci.* **2001**, *98*, 6599.
(24) Lee, K.-M.; Hayes, J. J. *Biochemistry* **1998**, *37*, 8622.
(25) Stefanovsky, V. Y.; Dimitrov, S. I.; Russanova, V. R.; Angelov, D.; Pashev, I. G. *Nucleic Acids Res.* **1989**, *17*, 10069.
(26) Usachenko, S. I.; Bavykin, S. G.; Gavin, I. M.; Bradbury, E. M. *Proc. Natl. Acad. Sci.* **1994**, *91*, 6845.
(27) Belyavsky, A. V.; Bavykin, S. G.; Goguadze, E. G.; Mirzabekov, A. D. *J. Mol. Biol.* **1980**, *139*, 519.
(28) Zheng, C.; Hayes, J. J. *J. Biol. Chem.* **2003**, *278*, 24217.
(29) Mutskov, V.; Gerber, D.; Angelov, D.; Ausio, J.; Workman, J.; Dimitrov, S. *Mol. Cell. Biol.* **1998**, *18*, 6293.
(30) Libertini, L. J.; Ausió, J.; van Holde, K. E.; Small, E. W. *Biophys. J.* **1988**, *53*, 477.
(31) Ausio, J.; Van Holde, K. E. *Biochemistry* **1986**, *25*, 1421.
(32) Wang, X.; Moore, S. C.; Laszckzak, M.; Ausió, J. *J. Biol. Chem.* **2000**, *275*, 35013.
(33) Sawicka, A.; Seiser, C. *Biochimie* **2012**, *94*, 2193.
(34) Prigent, C.; Dimitrov, S. *J. Cell Sci.* **2003**, *116*, 3677.
(35) Sauvé, D. M.; Anderson, H. J.; Ray, J. M.; James, W. M.; Roberge, M. *J. Cell Biol.* **1999**, *145*, 225.
(36) McManus, K. J.; Hendzel, M. J. *Biochem. Cell Biol.* **2006**, *84*, 640.
(37) Giet, R.; Glover, D. M. *J. Cell Biol.* **2001**, *152*, 669.
(38) Wei, Y.; Yu, L.; Bowen, J.; Gorovsky, M. A.; Allis, C. D. *Cell* **1999**, *97*, 99.
(39) Mellone, B. G.; Ball, L.; Suka, N.; Grunstein, M. R.; Partridge, J. F.; Allshire, R. C. *Curr. Biol.* **2003**, *13*, 1748.

(40) Sassone-Corsi, P.; Mizzen, C. A.; Cheung, P.; Crosio, C.; Monaco, L.; Jacquot, S.; Hanauer, A.; Allis, C. D. *Science* **1999**, *285*, 886.
(41) Wang, Y.; Zhang, W.; Jin, Y.; Johansen, J.; Johansen, K. M. *Cell* **2001**, *105*, 433.
(42) Walter, W.; Clynes, D.; Tang, Y.; Marmorstein, R.; Mellor, J.; Berger, S. L. *Mol. Cell. Biol.* **2008**, *28*, 2840.
(43) Macdonald, N.; Welburn, J. P. I.; Noble, M. E. M.; Nguyen, A.; Yaffe, M. B.; Clynes, D.; Moggs, J. G.; Orphanides, G.; Thomson, S.; Edmunds, J. W.; Clayton, A. L.; Endicott, J. A.; Mahadevan, L. C. *Mol. Cell* **2005**, *20*, 199.
(44) Hirota, T.; Lipp, J. J.; Toh, B.-H.; Peters, J.-M. *Nature* **2005**, *438*, 1176.
(45) Fischle, W.; Tseng, B. S.; Dormann, H. L.; Ueberheide, B. M.; Garcia, B. A.; Shabanowitz, J.; Hunt, D. F.; Funabiki, H.; Allis, C. D. *Nature* **2005**, *438*, 1116.
(46) Lee, G. M.; Pufall, M. A.; Meeker, C. A.; Kang, H.-S.; Graves, B. J.; McIntosh, L. P. *J. Mol. Biol.* **2008**, *382*, 1014.
(47) Xiang, S.; Gapsys, V.; Kim, H.-Y.; Bessonov, S.; Hsiao, H.-H.; Möhlmann, S.; Klaukien, V.; Ficner, R.; Becker, S.; Urlaub, H.; Lührmann, R.; de Groot, B.; Zweckstetter, M. *Structure* **2013**, *21*, 2162.
(48) Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J. D. *Structure* **2010**, *18*, 494.
(49) Kato, H.; van Ingen, H.; Zhou, B.-R.; Feng, H.; Bustin, M.; Kay, L. E.; Bai, Y. *Proc. Natl. Acad. Sci.* **2011**.
(50) Kato, H.; Jiang, J.; Zhou, B.-R.; Rozendaal, M.; Feng, H.; Ghirlando, R.; Xiao, T. S.; Straight, A. F.; Bai, Y. *Science* **2013**, *340*, 1110.
(51) Zhou, B.-R.; Feng, H.; Kato, H.; Dai, L.; Yang, Y.; Zhou, Y.; Bai, Y. *Proc. Natl. Acad. Sci.* **2013**, *110*, 19390.
(52) Zhou, B.-R.; Feng, H.; Ghirlando, R.; Kato, H.; Gruschus, J.; Bai, Y. *J. Mol. Biol.* **2012**, *421*, 30.
(53) Dyer, P. N.; Edayathumangalam, R. S.; White, C. L.; Bao, Y.; Chakravarthy, S.; Muthurajan, U. M.; Luger, K. In *Methods Enzymol.*; Allis, C. D., Carl, W., Eds.; Academic Press: 2003; Vol. Volume 375, p 23.
(54) Lowary, P. T.; Widom, J. *J. Mol. Biol.* **1998**, *276*, 19.
(55) http://www.biochem.ucl.ac.uk/hansen/fuda/.
(56) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719.
(57) Meiboom, S.; Gill, D. *Rev. Sci. Instrum.* **1958**, *29*, 688.
(58) Carr, H. Y.; Purcell, E. M. *Phys. Rev.* **1954**, *94*, 630.
(59) Baldwin, A. J.; Kay, L. E. *Nat. Chem. Biol.* **2009**, *5*, 808.
(60) Hansen, D. F.; Vallurupalli, P.; Kay, L. E. *J. Biomol. NMR* **2008**, *41*, 113.
(61) Lundstrom, P.; Vallurupalli, P.; Hansen, D. F.; Kay, L. E. *Nat. Protocols* **2009**, *4*, 1641.
(62) Wishart, D.; Sykes, B. *J. Biomol. NMR* **1994**, *4*, 171.
(63) Wishart, D.; Bigam, C.; Holm, A.; Hodges, R.; Sykes, B. *J. Biomol. NMR* **1995**, *5*, 67.
(64) Wishart, D. S.; Nip, A. M. *Biochem. Cell Biol.* **1998**, *76*, 153.
(65) Wishart, D. S. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58*, 62.
(66) Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D. *Protein Sci.* **2006**, *15*, 2795.
(67) Banères, J.-L.; Martin, A.; Parello, J. *J. Mol. Biol.* **1997**, *273*, 503.
(68) Kapust, R. B.; Tözsér, J.; Copeland, T. D.; Waugh, D. S. *Biochem. Biophys. Res. Commun.* **2002**, *294*, 949.
(69) Kim, S.; Wu, K.-P.; Baum, J. *J. Biomol. NMR* **2013**, *55*, 249.
(70) Hwang, T.-L.; van Zijl, P. M.; Mori, S. *J. Biomol. NMR* **1998**, *11*, 221.
(71) Clore, G. M.; Iwahara, J. *Chem. Rev.* **2009**, *109*, 4108.
(72) Felitsky, D. J.; Lietzow, M. A.; Dyson, H. J.; Wright, P. E. *Proc. Natl. Acad. Sci.* **2008**, *105*, 6278.
(73) Burmann, B. M.; Wang, C.; Hiller, S. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1265.
(74) Lietzow, M. A.; Jamin, M.; Jane Dyson, H.; Wright, P. E. *J. Mol. Biol.* **2002**, *322*, 655.
(75) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Marius Clore, G. *J. Magn. Reson.* **2003**, *160*, 65.

(76) Schwieters, C. D.; Kuszewski, J. J.; Marius Clore, G. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47.
(77) Iwahara, J.; Schwieters, C. D.; Clore, G. M. *J. Am. Chem. Soc.* **2004**, *126*, 5879.
(78) Braddock, D. T.; Baber, J. L.; Levens, D.; Clore, G. M. *EMBO J.* **2002**, *21*, 3476.
(79) Religa, T. L.; Sprangers, R.; Kay, L. E. *Science* **2010**, *328*, 98.
(80) Liokatis, S.; Stützer, A.; Elsässer, S. J.; Theillet, F.-X.; Klingberg, R.; van Rossum, B.; Schwarzer, D.; Allis, C. D.; Fischle, W.; Selenko, P. *Nat. Struct. Mol. Biol.* **2012**, *19*, 819.
(81) Liokatis, S.; Dose, A.; Schwarzer, D.; Selenko, P. *J. Am. Chem. Soc.* **2010**, *132*, 14704.
(82) Fletcher, T. M.; Hansen, J. C. *J. Biol. Chem.* **1995**, *270*, 25359.
(83) Vitolo, J. M.; Thiriet, C.; Hayes, J. J. *Mol. Cell. Biol.* **2000**, *20*, 2167.
(84) Gottesfeld, J. M.; Melander, C.; Suto, R. K.; Raviol, H.; Luger, K.; Dervan, P. B. *J. Mol. Biol.* **2001**, *309*, 615.
(85) Thomas Williams, C. K. *e. a. gnuplot* http://www.gnuplot.info.
(86) Nelson, M.; Fitch, D. A. In *Molecular Methods for Evolutionary Genetics*; Orgogozo, V., Rockman, M. V., Eds.; Humana Press: 2011; Vol. 772, p 459.
(87) Tzeng, S.-R.; Pai, M.-T.; Kalodimos, C. G. In *T Protein NMR Techniques* 2011; Vol. 831, p 133.
(88) Delaglio, F.; Grzesiek, S.; Vuister, G.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277.
(89) T. D. Goddard and D. G. Kneller, S., University of California, San Francisco.
(90) Mal, T. K.; Ikura, M.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 14002.
(91) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins: Struct. Funct. Bioinform.* **1993**, *17*, 75.
(92) Connelly, G. P.; Bai, Y.; Jeng, M.-F.; Englander, S. W. *Proteins: Struct. Funct. Bioinform.* **1993**, *17*, 87.
(93) Kjaergaard, M.; Poulsen, F. *J. Biomol. NMR* **2011**, *50*, 157.
(94) Kjaergaard, M.; Brander, S.; Poulsen, F. *J. Biomol. NMR* **2011**, *49*, 139.
(95) Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. *J. Am. Chem. Soc.* **2001**, *123*, 2970.

7 Effect of mC and hmC on transcription factor binding

7.1 Introduction

Binding of transcription factors (TF) to specific DNA sequences is an important mechanism in transcriptional regulation.[2] The transcriptional responses of cells to many external signals such as developmental cues, hormones and O_2 are substantially mediated by different TFs.[5] Of central importance in oxygen sensing are the hypoxia inducible factor (HIF) proteins.[6,7] The HIF proteins are α,β-hetero-dimeric basic-Helix-Loop-Helix (bHLH) TFs that specifically recognize their hypoxia response element (HRE; 5'-ACGTG-3'). In hypoxia, HIF1α and HIF2α levels rise and they translocate from the cytoplasm into the nucleus, where they can dimerise with HIF1β and bind to their HREs. At transcription start sites (TSS) HIF recruits transcriptional co-activators via the *N*- and *C*-terminal activation domain (NAD/CAD) in the HIF1α isoform.[8,9]

HIF regulation is accomplished by at least two molecular oxygen dependent mechanisms.[11] The 2-oxoglutarate (2-OG), Fe^{2+} and O_2 dependent oxygenases "prolyl hydroxylase domain-containing proteins"[12] (PHD) 1-3 can hydroxylate one or both of two proline residues in the oxygen dependent degradation domains (ODDD; Pro402 in the HIF1α NODDD; and Pro564 in the HIF1α CODDD) of HIF-α isoforms. Their hydroxylation results in association of HIF1α with an E3 ubiquitin ligase complex of which the targeting component is the von Hippel-Lindau protein[13,14] (pVHL), resulting in HIF1α ubiquitinylation and degradation by the 26S proteasome.[15,16] Another 2-OG, Fe^{2+} and O_2 dependent oxygenase factor inhibiting HIF (FIH),[17] hydroxylates Asn803 in HIF1α in its CAD, a modification that blocks recruitment of the CBP/p300 transcriptional co-activator complexes by HIF.[8,9,18]

The C-5 methylation of cytosine in a CpG context is an important mechanism for epigenetic regulation.[19] CpG methylation at promoter CpG islands results in repression by decreasing TF binding and recruitment of transcriptional repressors.[20] HREs contain a central CpG which can be subject to

cytosine methylation.[21] HIF1α/β association is substantially reduced by the presence of 5mC in the HRE.[21]

The Ten-Eleven-Translocation (TET) 1-3 enzymes are 2-OG, Fe^{2+} and O_2 dependent hydroxylases that hydroxylate 5mC to give 5-hydroxymethyl cytosine and further oxidized forms, 5-formyl- and 5-carboxy-cytosine.[22,23] It was proposed that HIF1α/β binding to the HRE could be modulated by the presence of 5hmC. In the work in this chapter the effect of cytosine hydroxylation on HIF1α/β binding to the HRE was investigated. To gain a more complete view of the effect of cytosine modification of bHLH TFs, the related bHLH transcription factors MAX[1] and Upstream stimulatory factor (USF) 1 were also investigated.[3]

The bHLH domain of MAX, USF1 and HIF1α/β is conserved (**Figure 1**) and the recognition element of these TFs is also largely conserved. MAX and USF1 recognize the E-Box sequence (3'-CACGTG-5') which is extended by one cytosine from the HRE (3'-ACGTG-5'). Dimerisation of MAX and USF is mediated by a leucine zipper (ZIP) domain and both proteins can bind to DNA as homo-dimers. MAX in addition has the ability to heterodimerise with other TFs including cMyc, vMyc, Mad and Mri. The HIF1α/β heterodimer associates via a Per-Arnt-Sim (PAS), which has been proposed to enable dimerization control by small molecules. The basic region of the bHLH domain is disordered in solution and forms an α-helix upon association with DNA which extends through the major groove (**Figure 1a**). Sequence recognition is archived by various direct and water mediated hydrogen bonds with the DNA bases. The guanine of the central CpG of the HRE is recognized by a conserved arginine present in all bHLH TFs (highlighted in **Figure 1a** and **b**). No direct interactions with the CpG cytosine are observed.

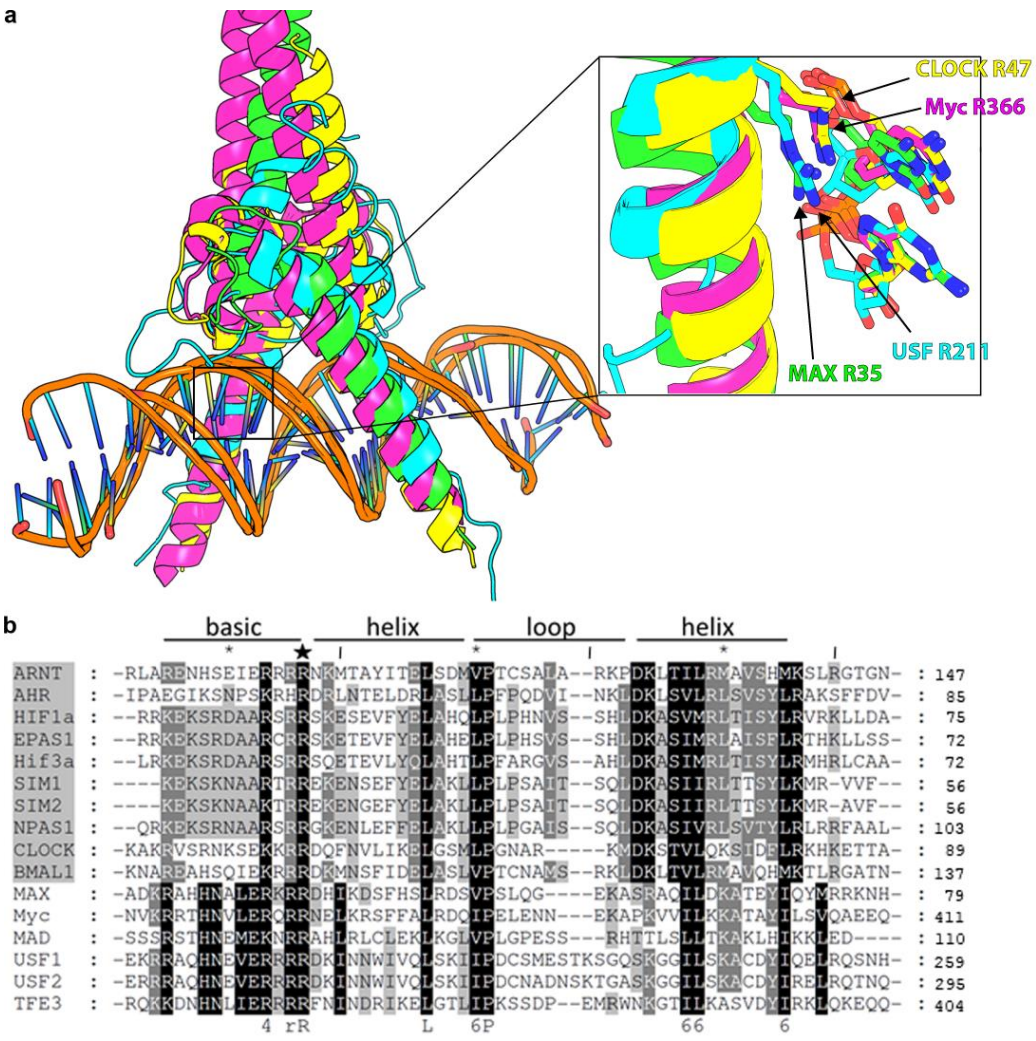


Figure 1. a Overlay of the crystal structure of MAX[1] (green), USF1[3] (cyan), MAX/MYC[24] (pink), CLOCK/BMAL1[4] (yelow) showing structural conservation. **b** Sequence alignment of the bHLH region of different bHLH-PAS (highlighted in grey) and bHLH-ZIP proteins. The conserved Arg that can form a hydrogen bond with the central G (Arg35 in MAX[25], Arg211 in USF[3], Arg47 in CLOCK[26], Arg85 in BMAL1[4]) at the CpG is highlighted by the star (adapted from Lercher *et al.*).[10]

7.2 Results

7.2.1 Influence of mC and hmC on HIF binding

One of the best studied responses to hypoxia in vertebras is the increased production of erythropoietin (EPO) to stimulate red blood cell production.[27] EPO production is regulated by HIF through binding to an HRE in the EPO promoter.[6] To test the influence of 5mC and 5hmC on HIF binding, a 59bp DNA fragment derived from the EPO promoter containing one HRE was prepared by PCR. Modifications were introduced globally by substituting dCTP with dmCTP or dhmCTP yielding

oligonucleotide probes containing all cytosines substituted by 5mC or 5hmC. The PCR efficiency was not noticeably affected by the use of differently modified triphosphates (**Figure 2a**). Plasmids encoding full length human HIF1α and HIF1β were a kind gift from Dr. Pugh and were produced by *in vitro* transcription translation (IVTT). For IVTT a plasmid with a T7 or SP6 promoter flanking the gene of interest is combined with the appropriate polymerase and dNTP to generate RNA. In the same pot all components necessary for protein production from RNA are present, derived from rabbit reticulocyte lysate. This allows production of small amounts of protein for binding assays directly from plasmid DNA. The efficiency of the IVTT reaction was first tested by performing the reaction in the presence of ^{35}S methionine and analysing the mixture by SDS-PAGE and autoradiography. Indeed, bands at slightly higher MW than expected were observed for both proteins confirming the efficiency of IVTT for producing HIF1α and HIF1β (**Figure 2b**).

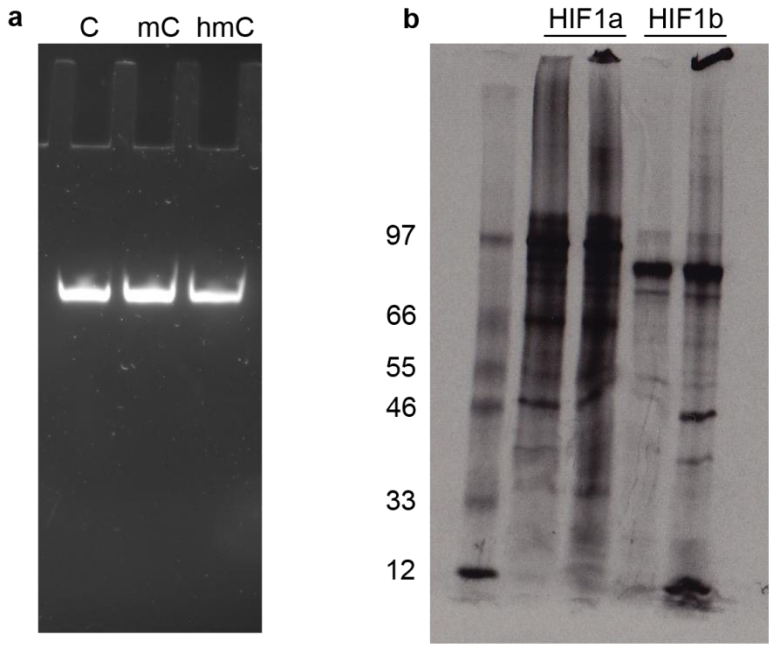


Figure 2. a Analysis of the purified PCR product for the EPO promoter fragment analysed by PAGE and stained with SYBR Safe. **b** SDS-PAGE analysis of the IVTT reactions performed with the HIF1α and HIF1β plasmids respectively. Expected MW are HIF1α: 93kDa and HIF1β: 87kDa.

To test the association between the HIF1α/β dimer and differentially modified EPO oligonucleotides, the oligonucleotides were labelled with ^{32}P-phosphate by reaction with T4 Polynucleotide Kinase and

ATP γ-^{32}P and Electrophoretic Mobility Shift Assays (EMSAs) were performed with IVTT produced HIF1α/β.

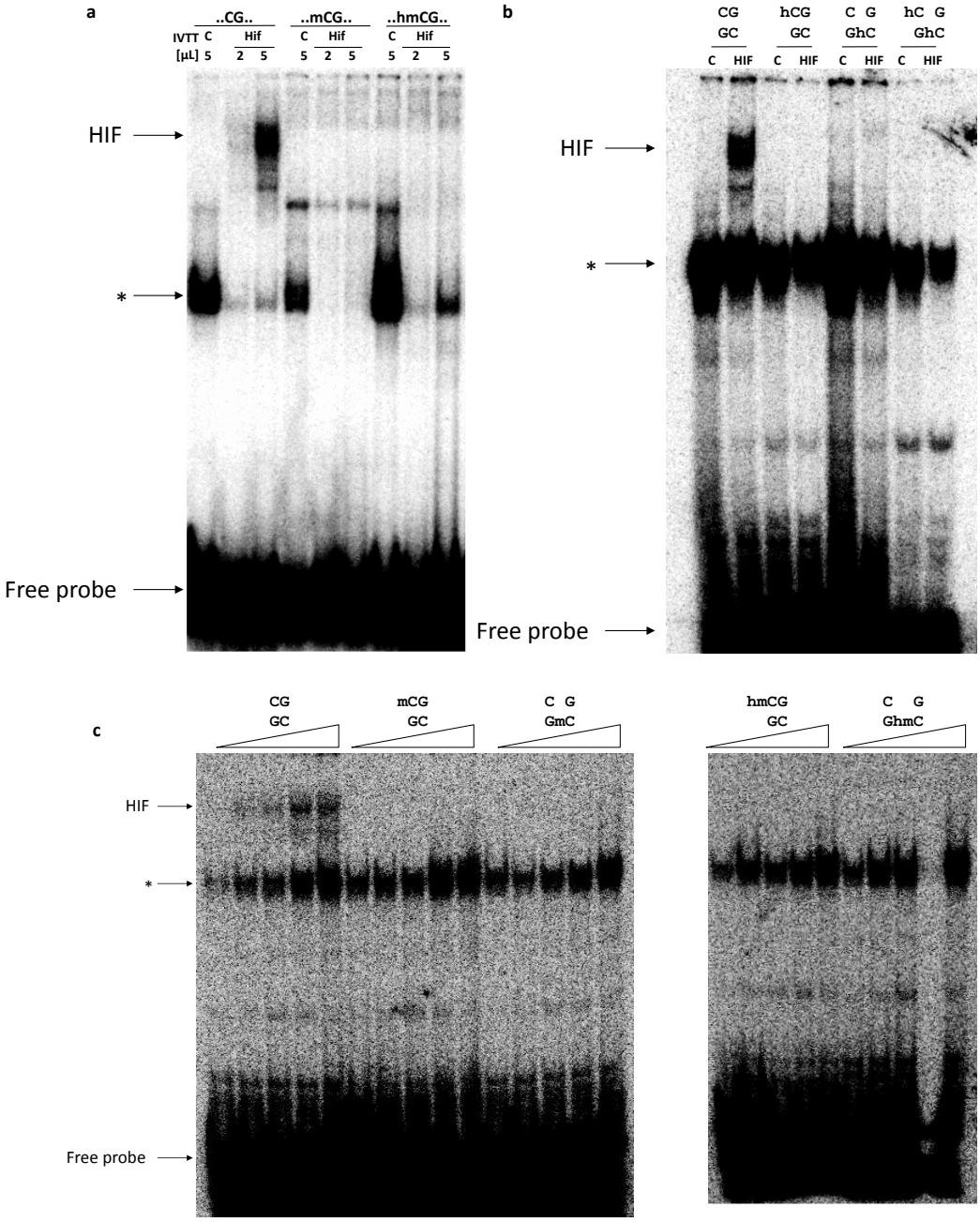


Figure 3. Modification of the central CpG in the HRE (ACGTG) reduces binding of HIF below the limit of detection. Full-length HIF1α and HIF1β (ARNT) were first produced by *in vitro* transcription translation (IVTT) and then incubated with radiolabelled DNA probes containing a HRE. The obtained complexes were separated on a 5% PAGE gel. **a** EMSA with IVTT HIF and different unmodified, fully methylated or fully hydroxymethylated probes derived from the erythropoietin (EPO) promoter.[21,28] Different modifications were introduced by substitution of dCTP with dmCTP or dhmCTP during the PCR amplification.[29] **b** EMSA with IVTT produced HIF1α/β and oligonucleotides containing 5hmC at the central CpG of the HRE. **c**

Titrations of increasing amounts of IVTT translated HIF1α/β with differently modified probes. *Non-specific binding that is present in the control reaction and not cytosine C-5 modification sensitive (adapted from Lercher *et al).*[10]

In EMSAs, the oligonucleotide probes are separated on native PAGE gels. Upon protein binding to the probe, a larger complex with a different mass to charge ratio is formed. This complex is seen as a higher molecular weight band on the native PAGE. Both HIF1α and HIF1β were produced in one IVTT reaction to ensure efficient dimer formation. As control the IVTT mixture supplemented with the manufacture supplied luciferase plasmid was used.

For the EPO fragment containing no modification, a specific band was only observed in the presence of HIF1α/β (**Figure 3a**). The global replacement of C with 5mC or 5hmC seemed to reduce binding below the limit of detection (**Figure 3a**). Because the probes had all cytosine replaced by a modified variant, it is essential to confirm that this effect is because of the modification of the central CpG in the HRE. For this, oligonucleotides encompassing part of the EPO promoter bearing only one 5hmC at the central cytosine (A**C**GTG) were prepared by solid phase synthesis.

The oligonucleotides were annealed in different combinations to give non-modified, hemi-modified or fully modified duplexes by combining equimolar concentrations of the forward and reverse single stranded oligonucleotide and cooling them slowly from 95 °C to 4 °C. After radiolabelling of the duplexes, the binding of the HIF1α/β dimer was analysed by EMSAs. Again, clear binding of the unmodified oligonucleotide by HIF1α/β was observed (**Figure 3b**). Hemi-hydroxymethylation or symmetrical hydroxymethylation both reduce binding below the limit of detection (**Figure 3b**). The binding is HIF1α/β concentration dependent and even at high HIF1α/β no binding can be detected for any modified oligonucleotides (**Figure 3c**).

7.2.2 Influence of mC and hmC on MAX and USF binding

Next, it was of interest to investigate whether the effect of C-5 modification on bHLH TF binding is consistent across this class of TFs. MAX[1] and USF1,[3] which bind as homo-dimers to the E-Box sequence (3'-CACGTG-5') were used. Reports on their sensitivity towards C-5 methylation of the

central CpG diverge. Prendergast *et al.*[30] reported that the interaction between human USF1 derived from HeLa cells and a radiolabelled unmodified E-Box oligonucleotide is efficiently outcompeted by an unlabelled oligonucleotide containing 5mC at the central E-Box CpG. This suggests USF1 can bind to a CpG methylated E-Box. In contrast, Fuji *et al.*[31] reported that mouse USF1 prepared by IVTT can bind to an unmodified E-Box containing oligonucleotide and this complex cannot be outcompeted by the addition of unlabelled oligonucleotide containing 5mC at the central E-Box CpG, suggesting CpG methylation of the E-Box inhibits binding. Different preparation methods and the use different USF proteins could be responsible for these discrepancies.

To assess the binding of USF and MAX to differently modified E-Box containing oligonucleotides, USF was prepared by IVTT (**Figure 9**). However, the yield of MAX prepared by IVTT was very low (**Figure 9**) and only very weak binding in initial EMSAs was detected. MAX was expressed as a his-tagged variant in *E.coli*, purified using IMAC and size exclusion and was obtained in high purity (**Figure 4a**).[32]

MAX binding to differently modified and radiolabelled E-Box oligonucleotides was tested by EMSAs. As expected, strong binding to the unmodified E-Box was observed (**Figure 4b**). Hemi-methylated or hemi-hydroxylated oligonucleotides led to significantly reduction in binding while symmetric modification for both 5mC and 5hmC reduced binding below the limit of detection (**Figure 4b**). To quantify the difference in affinity for the differentially modified E-Box probes, titrations with increasing concentration of MAX were performed for each probe. A sigmoidal response for the non-modified and hemi-modified oligonucleotides was observed (**Figure 4c**). The symmetric modification reduced binding affinity (**Figure 4c**) below the detection level even at the highest concentration used. The calculated Kd_{app} were similar for all hemi-modified oligonucleotides (320-370 nM; **Figure 4d**). This corresponds to a 6 fold increase in Kd_{app} when compared to the unmodified oligonucleotides.

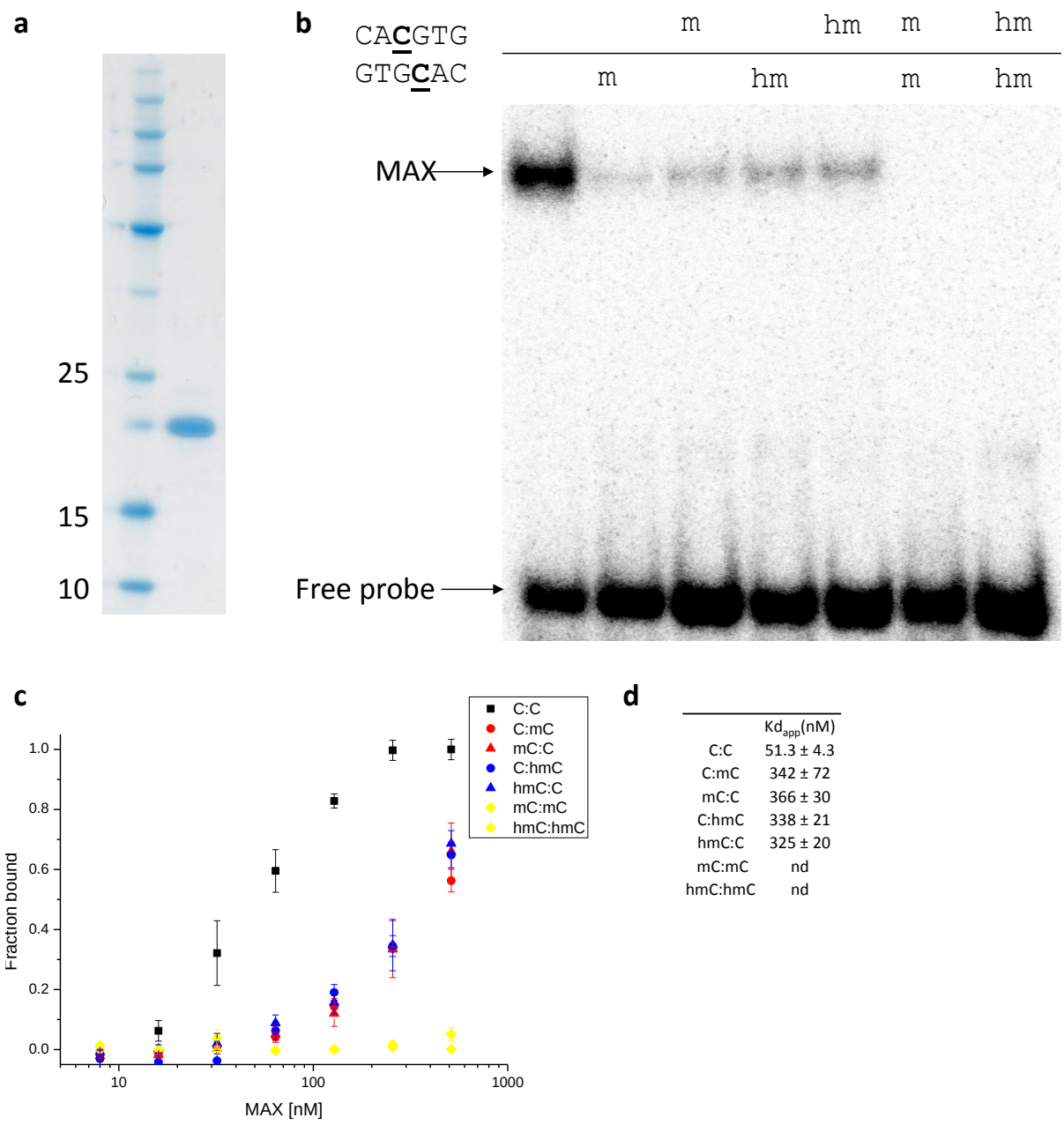


Figure 4. Analysis of MAX binding to the E-Box. **a** SDS-PAGE showing purified MAX (20.5KDa). **b** EMSA with radiolabelled dsDNA probes bearing different modifications to cytosine C-5 at the central CpG of the E-Box and MAX. The modified C is highlighted with the modification above the lane. **c** Titration curves of MAX with differently modified E-Box oligonucleotides measured by EMSA. **d** The affinity of MAX for the E-Box sequences containing different cytosine modifications as determined by EMSAs[28]. In the case of mC:mC and hmC:hmC only weak / non-specific binding was observed as evidenced by a slight reduction of free probe but no formation of a distinct complex (nd: not determined; adapted from Lercher *et al*[10]).

To investigate the effect of different CpG modifications on the binding of USF, identical EMSAs were performed using IVTT produced USF. USF efficiently binds to the non-modified E-Box probe (**Figure 5a**). Hemi-modification results in a reduction of binding affinity while symmetric modification reduces the affinity below the detection limit for both 5mC and 5hmC (**Figure 5a**). To further quantify the affinity of USF towards differently modified probes, titration curves with increasing amounts of USF containing IVTT mixture were measured. A dose dependent increase in bound complex was observed (**Figure 5b**), but the binding curves showed a plateau for all probes (at 90% bound for unmodified oligonucleotides and at 30-45% bound for hemi-modified oligonucleotides), which made the determination of the binding affinity impossible. This is probably due to the presence of large quantities of plasmid DNA and other components in the IVTT mixture that competes with the weaker binding to the hemi-modified oligonucleotides.

To assess the difference in binding affinity of USF for the differentially modified oligonucleotides, competition experiments were performed with a radiolabelled unmodified E-Box USF complex and increasing amounts of differently modified oligonucleotides (**Figure 5c**). When USF can bind to the unlabelled competitor oligonucleotide a decrease complex between USF and the radiolabelled oligonucleotide is observed. Efficient competition of the USF complex at a 100 fold excess of unlabelled probe was observed (**Figure 5c**). For both the hemi-methylated and the hemi-hydroxymethylated probe, a reduction of the labelled USF complex was only observed at 1000 fold excess (**Figure 5c**). This confirms the lower affinity of USF towards hemi-modified oligonucleotides. The symmetrically methylated or hydroxymethylated probes do not compete for USF binding (**Figure 5c**). This is in agreement with previous experiments with HIF1α/β and MAX where binding to symmetrically modified oligonucleotides was not observed.

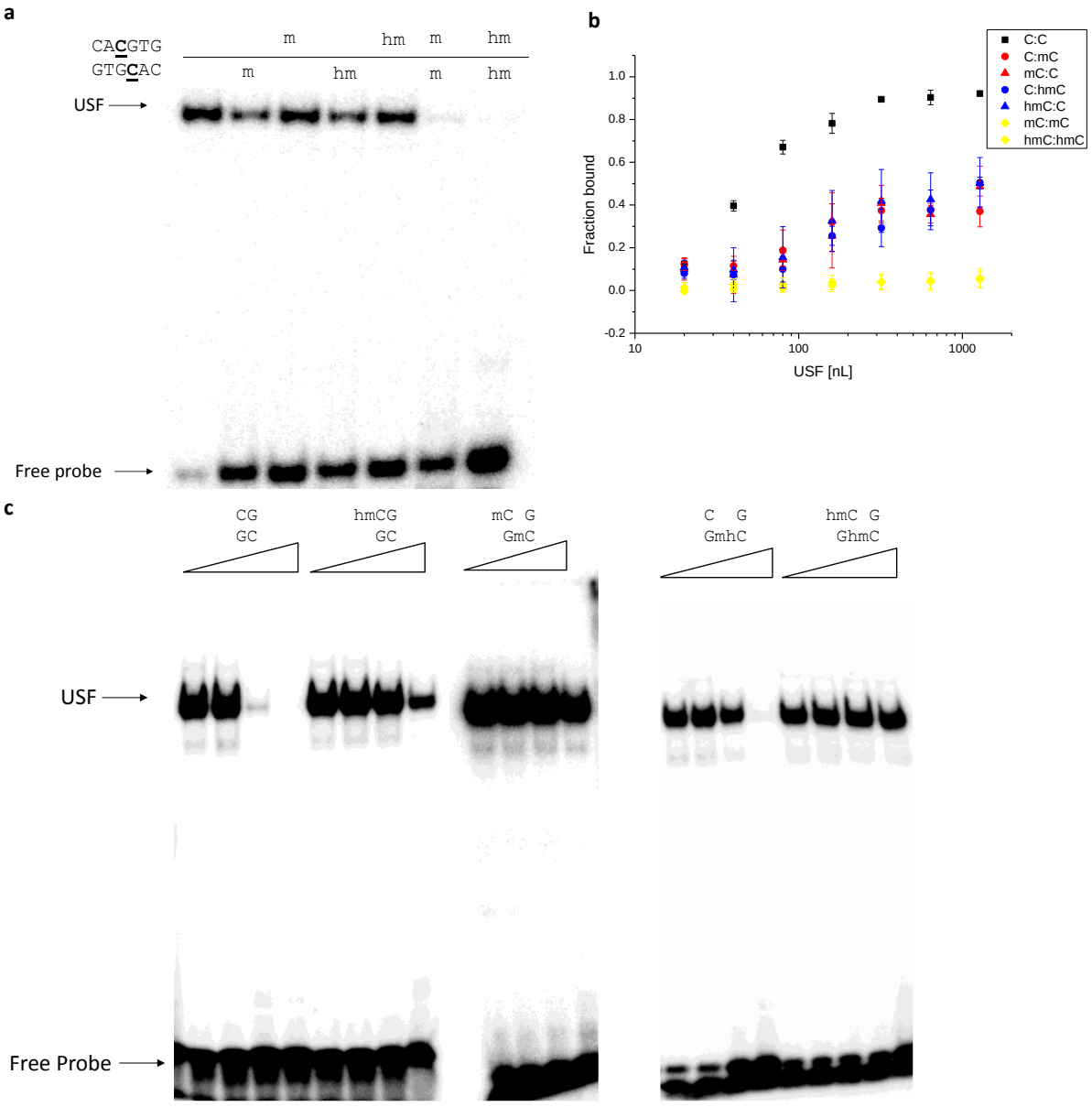


Figure 5. Analysis of USF binding to the E-Box. **a** EMSA with USF and radiolabelled dsDNA probes bearing different modifications to cytosine C-5 at the central CpG of E-Box's. The modified C is highlighted with the modification above the lane. **b** Titration curves of USF measured by EMSAs with radiolabelled dsDNA probes bearing different cytosine C-5 modifications at the central CpG of the E-Box sequence. The modification state of the cytosine in the central CpG of each series is given in the legend. Experiments were performed in triplicate. **c** Competition experiment with a radiolabelled unmodified E-Box-USF complex and unlabelled probes containing different modifications. The central CpG modification state of the unlabelled probe is given above the lanes (adapted from Lercher *et al*[10]).

7.2.3 Structural analysis

During the course of this work, structures of a dsDNA fragment containing C, mC and hmC were solved (**Figure 6**).[10,33] For this, the Dickerson-Drew dodecaeder[34] (**Figure 6b**) was used since it has excellent crystallization properties and structures down to 1.0 Å have been reported.[35]

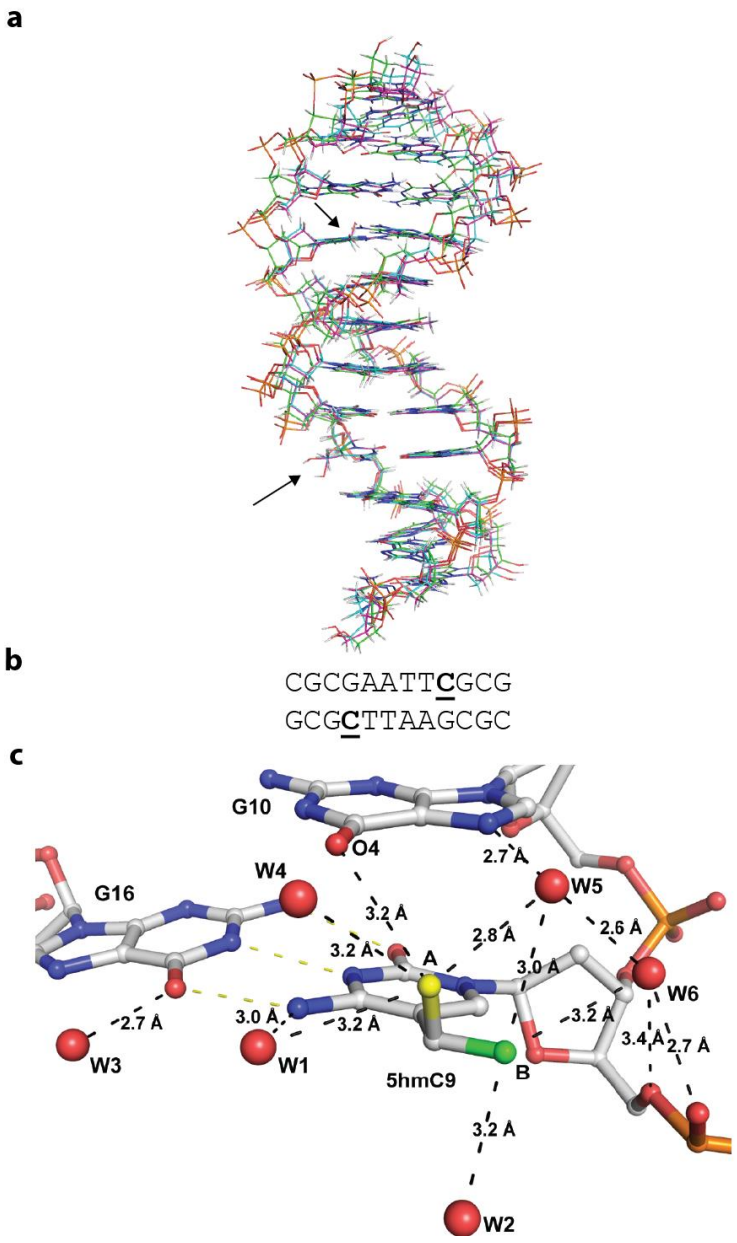


Figure 6. a Overlay of the crystal structures of Dickerson–Drew duplexes containing cytosine (green; 4C64), 5-methylcytosine (cyan; 4C63) and 5-hydroxymethylcytosine (pink; 4C5X). The sites of modification are marked with arrows. **b** Sequence of the used duplex with numbering. Modified sites are underlined. **c** Water network around 5hmC. The 5hmC hydroxyl is refined in two conformations (A, yellow, 70% occupancy; b, green, 30% occupancy). Both conformations are stabilized by extensive hydrogen bonds to surrounding waters or the O4 of the 3'-adjecent G (conformation A).

The overall structure of the B-DNA duplex is not affected by modification, as indicated by small RMSDs of all atoms between the differently modified strands (0.18 Å, C-mC; 0.27 Å, C-hmC and 0.35 Å, mC-hmC; **Figure 6a**). The 5hmC hydroxyl is observed in two distinct conformations in the crystal structure (**Figure 6c**). In the major conformation (A, 70% occupancy), the 5hmC alcohol makes a weak hydrogen bond to the O4 of the 3'-adjecent G (3.4 Å distance). The minor conformation hydroxyl (B, 30% occupancy) is oriented towards the phosphate backbone. The overall water structure around the differently modified cytosine is conserved, with minor reorientation of the waters W4 and W2 likely due to a steric clash with the 5hmC alcohol.

To gain mechanistic insight into the reduction of TF binding by C-5 modification the structures of the modified dsDNA with the published structures of bHLH TF were overlaid. When comparing the structures, a different orientation for a conserved arginine (Arg35 in MAX, Arg211 in USF) became obvious. This arginine enters the major groove of the B-DNA helix and contacts the N7 of the guanine in the central CpG (CAC**G**TG).

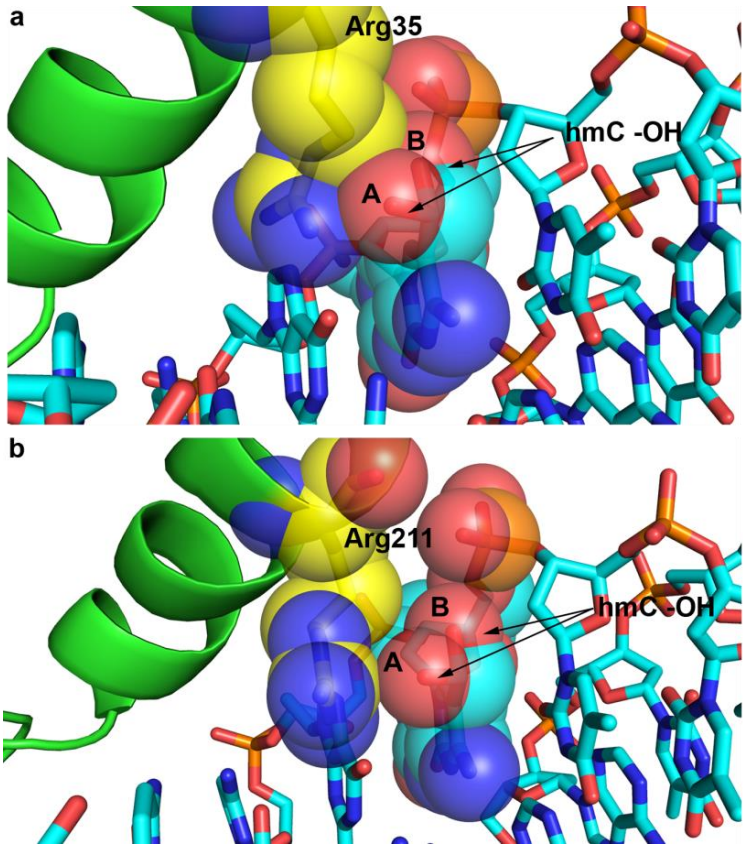


Figure 7. Structural overlay of the hmC9 containing duplex (4C5X) with the MAX-DNA complex (**a**; adapted from pdb entry 1AN2[1]) and the USF-DNA complex (**b**; adapted from pdb entry 1AN4[3]) showing the potential clash of the C5 modification with the arginine (Arg35 in MAX, Arg211 in USF). Both the major (A; 70% occupancy) and the minor (B; 30% occupancy) of the hmC9 alcohol are depicted (adapted from Lercher *et al*[10]).

The arginine in MAX adopts a conformation that results in a steric clash with the hydroxyl of hmC in the CpG (**a**). In USF, this arginine adopts a different conformation that does not clash with the hydroxyl significantly (**b**). This could explain the observed higher sensitivity of MAX towards C-5 modification than that of USF.

To explain the difference in sensitivity towards C-5 modification of HIF with MAX and USF, it was desirable to visualize the HIF-DNA interaction. Because there is no structure of the HIF1α/β bHLH domain in complex with DNA, the structure of the related heterodimeric bHLH-PAS TF BMAL1/CLOCK was used (See **Figure 1** for sequence alignment). Similar to the previously investigated TFs, BMAL1 and CLOCK contain a conserved arginine (Arg47 in BMAL1 and Arg85 in CLOCK) that contacts the N7 of the guanine (AC**G**TG). The conformation of this arginine is similar in both structures and shows a significant steric clash between the arginine side chain and the hmC hydroxyl.

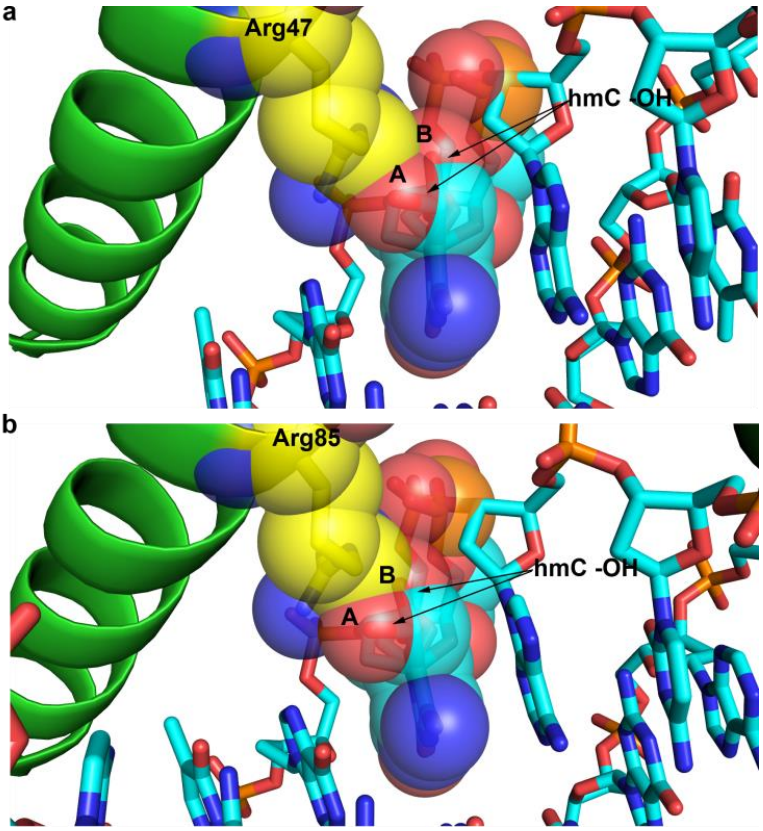


Figure 8. Structural overlay of the hmC9 containing duplex (4C5X) with the CLOCK-DNA complex (**a**) and the BMAL1-DNA complex (**b**; both adapted from pdb entry 4H10[4]) showing the potential clash of the C5 modification with the arginine (Arg47 in CLOCK, Arg85 in BMAL1). Both the major (A; 70% occupancy) and the minor (B; 30% occupancy) of the hmC9 alcohol are depicted (adapted from Lercher *et al.).*[10]

7.3 Discussion

Cytosine C-5 methylation of CpG dinucleotides in CpG islands results in transcriptional silencing by a variety of mechanism including modulation of TF binding at promoters.[19,20] While, for a variety of TFs the sensitivity for mC has been reported, only few studies have to date investigated the effect of hmC on TF binding. mC and hmC have pronounced and different effects on duplex stability of dsDNA.[36,37] mC results in an increase in melting temperature of the duplex. hmC reverses this effect and hmC containing oligonucleotides in some cases resulting in a melting temperature similar to that of unmodified DNA. This is in agreement with the observation that mC results in transcriptional repression and hmC could reverse this effect. Thus, differential effects of these two modifications on TF binding might be expected. hmC, similar to mC results in a reduction of binding affinity of HIF1α/β (**Figure 3**), MAX (**Figure 4**) and USF (**Figure 5**) for their cognate recognition sequences (**Table 1**). This effect could be caused by a steric clash between a conserved arginine residue, which is present in all TFs, and the C-5 modification (and **Figure 8**). Interestingly, while MAX and USF binding can tolerate hemi-modifications, binding of HIF1α/β to the HRE is reduced below the limit of detection by hemi-modification. The observation that HIF1α/β binding is modulated by an O_2 dependent modification could suggest TET1-3 being involved in oxygen sensing. Also, while hemi-methylated DNA is a substrate for the maintaining methyl transferase DNMT1[38], hmC is not a substrate for this enzyme[39]. Hydroxylation of an established mC would lead to a hemi-hmC containing dsDNA after replication, which can be recognized by USF or MAX. Hydroxymethylation of a symmetrically modified CpG could thus function to poise a promoter for activation, by TF recruitment to the hemi-modified CpG, after cell division.

		hemi-modification		**symmetric-modification**	
	C	mC	hmC	mC	hmC
HIF1α/β	+++	---	---	n.d.	---
MAX	+++	+	+	---	---
USF	+++	+	+	---	---

Table 1. The effect of different modifications states of the central CpG in the E-Box (for MAX and USF) or the HRE (for HIF1α/β). +++ strong binding; + weak binding; --- no binding detected n.d. not determined.

7.4 Experimental procedures

General procedure for PCR, cloning, DNA analysis, protein expression and purification can be found in Chapter 2.

7.4.1 Used oligonucleotides

Name	Sequence (3'-5')	Obtained from
MAX_FW	ATGAGCGATAACGATGACATCGAGGTGG	Sigma
MAX_RW	TTAGCTGGCCTCCATCCGGAGCTTC	Sigma
USF_FW	ATGAAGGGGCAGCAGAAAACAGCTG	Sigma
USF_RW	GGCCCAAAGCCCCTGAATCCCCA	Sigma
EPO_template	AGGGGTGGAGGGGGCTGGGCCCTACGTGCTGTCTCACACAGCCT GTCTGACCTCTCGAC	Eurofins
EPO_FW	GGTGGAGGGGGCTGGGCCCTA	Eurofins
EPO_RW	CGAGAGGTCAGACAGGCTGTGTGAGACAGC	Eurofins
EBOX_FW_C	CTCAGGCACCACGTGGTGGGGGAT	Sigma
EBOX_RW_C	ATCCCCCACCACGTGGTGCCTGAG	Sigma
EBOX_FW_mC	CTCAGGCACCAmCGTGGTGGGGGAT	Eurofins
EBOX_RW_mC	ATCCCCCACCAmCGTGGTGCCTGAG	Eurofins
EBOX_FW_hmC	CTCAGGCACCAhmCGTGGTGGGGGAT	ATD Bio
EBOX_RW_hmC	ATCCCCCACCAhmCGTGGTGCCTGAG	ATD Bio
HIF_FW_C	GCCCTACGTGCTGTCTCACACAGCCT	Sigma
HIF_RW_C	AGGCTGTGTGAGACAGCACGTAGGGC	Sigma
HIF_FW_mC	GCCCTAmCGTGCTGTCTCACACAGCCT	Eurofins
HIF_RW_mC	AGGCTGTGTGAGACAGCAmCGTAGGGC	Eurofins
HIF_FW_hmC	GCCCTAhmCGTGCTGTCTCACACAGCCT	ATD Bio
HIF_RW_hmC	AGGCTGTGTGAGACAGCAhmCGTAGGGC	ATD Bio

Table 2. Used oligonucleotides

7.4.2 MAX expression and purification

MAX was amplified from HeLa cDNA by PCR using the primers given above and cloned into pET15b. The plasmid encoding the MAX was transformed into BL21(DE3)Rosetta 2 (Merck-Millipore 71401) and grown on LB-agar plates containing ampicillin (100mg/L) and chloramphenicol (34mg/L). The next evening 100 mL LB containing the same antibiotics were inoculated with one colony from the plate. This pre-culture was grown overnight at 37 °C. The next morning 12 flasks of 800 mL 2xTY were inoculated with 8 mL of pre-culture each. These cultures were grown at 37 °C until OD_{600} = 0.5 and then expression was induced by addition of 1 mM IPTG. Expression was continued at 30 °C for 4h after which the cells were harvested by centrifugation. Cell pellets were stored at -80 °C until purification.

Cells were resuspended in 70 mL lysis buffer (20 mM TRIS pH 7.5, 300 mM NaCl, 30 mM Imidazole, 10 mM BME, Complete protease inhibitor), DNase and Lysozyme were added and cells were lysed by sonication (5 x 30 s at 60% Amp). The lysate was then cleared by centrifugation (20 min, 20k rpm, JA25.50). The lysate was then incubated with 2 mL of Ni-NTA resin (Merck Millipore; 69670; His-Bind® Resin) for 20 min at 4 °C. The resin was then washed with 25 CV lysis buffer and the protein was eluted in 5 CV lysis buffer containing 500 mM Imidazole. The eluted protein was concentrated to 2 mL and further purified by SEC (Superdex S75 16/60 in 25 mM TRIS pH 7.5, 100 mM NaCl). Fractions containing the desired protein were pooled and flash frozen and stored at -80 °C in 1.5 mL aliquots.

7.4.3 IVTT preparation of USF and HIF1α/β

USF was amplified from HeLa cDNA by PCR using the primers given above and cloned into pGEM-T Easy vectors (Promega). Plasmids encoding for hypoxic inducible factor (HIF)1-α and HIF1-β were a kind gift from Prof. Christoph W. Pugh (Welcome Trust, Oxford).

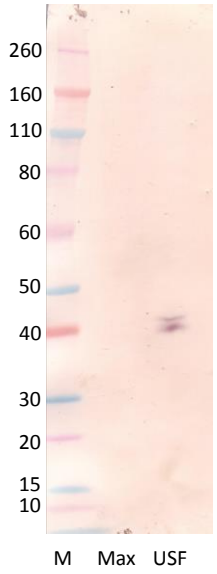


Figure 9. Biotin westerblot of an IVTT reaction with the MAX and USF encoding plasmid in the presence of Transcend™ tRNA.

USF and HIF1-α/β were produced by IVTT using the TnT® Coupled Reticulocyte Lysate System (T7 for HIF, SP6 for USF; Promega; L4601 and L4611) using the manufacturer recommended conditions. For

the control lanes, the manufacturer supplied luciferase plasmid was added to the IVTT mixture instead of the HIF1-α and -β plasmids.[40] Efficient translation of USF was confirmed by use of the Transcend™ tRNA (Promega; L5061) and biotin western blot (**Figure 9**).

7.4.4 EMSA probe preparation

For PCR generated HIF probes using Pfu Ultra (Agilent) substitution mCTP (NEB, N0356S) or hmCTP (bioline, BIO-39046) for CTP using the manufacture recommended conditions using the EPO_template as template and EPO_FW and EPO_RW as primers. The EPO template corresponds to the sequence responsible for the hypoxic upregulation of the *erythropoietin / epo* gene that contains a HRE (+97-156bp from the 3' end; the EPO_template corresponds to 38,138,574 - 38,138,632 of human chromosome 7 genomic scaffold (ref|NW_004078032.1). Oligonucleotides obtained from Sigma (unmodified), Eurofins (mC) or ATDBio (hmC) and resuspended in 10 mM TRIS pH 7.5. Concentrations were determined by measuring the A260 using the nanodrop. Theoretical absorption values and molecular weight were calculated using the web-based IDT oligo analyzer (http://eu.idtdna.com/analyzer/applications/oligoanalyzer/). The probes were annealed at an equimolar ratio (5 µM in 50 µL) in a PCR machine by cooling from 95 °C to 4 °C at 0.5 °C/min. Annealing was confirmed by PAGE gel (15 %, 0.5x TBE; data not shown). After annealing, the concentration was again measured (Nanodrop ND-1000). Probes (50 ng) were phosphorylated using polynucleotide kinase (10 u, NEB, M0201) with ATP, [γ-^{32}P] (1 µL, Perkin Elmer, NEG002A250UC) for 30 min at 37 °C. Radiolabelled probes were separated from free ATP using the QIAquick Nucleotide Removal Kit (Qiagen, 28304) as described in the manufacture manual.

7.4.5 EMSAs[28]

For assays with hypoxic inducible factor (HIF), the binding assays contained 2-5 µL of the IVTT mixture, 2 µL 10x binding buffer (final: 10 mM Tris pH 7.4, 50 mM NaCl, 50 mM KCl, 1 mM Ethylenediaminetetraacetic acid (EDTA), 5 mM DTT, 5% glycerol), 75ng dI/dC (1 µg/µL stock) diluted to 19 µL with water and incubated on ice for 10 min. Then 1 µL of radio-labelled DNA probe (0.08 ng/µL diluted from stock; see above for sequence) was added and the mixture was incubated on ice

for an additional 30 min. The complete 20 µL EMSA mixture was then loaded onto a 5% PAGE gel (0,5x TBE (45 mM Tris-borate, 1 mM EDTA, pH 8.3), 0.7 mm, cast 1 day in advance and pre-run at 240 V for 1h) and run at 240 V for 3.45-4 hours. The gel was put on Whatman paper, covered with saran wrap and dried at 80 °C for 1 h. A phosphoscreen (Biorad) was exposed to the gel for 24-72 h and imaged using a Personal Molecular Imager (PMI, Biorad). Images were processed using the Quantity one analysis software (Biorad).

The USF binding assays consisted of 1 µL binding buffer (as above), 0.1 µL dI/dC (stock 1 µg/µL; 100ng final), IVTT mixture (1 µL for fig 3 and S6; 0.02-1.28 µL for titration curves), 1 µL probe (0.05ng/µL; see above for sequence) and water to 10 µL. The reactions were incubated on ice for 20 minutes and electrophoresis was performed as described above, but at room temperature.

MAX binding assays consisted of 1 µL binding buffer (as above), 0.1 µL dI/dC (stock 1 µg/µL; 100ng final), 1 µL MAX (2.5 µM or 2x dilutions for titration curves), 1 µL probe (0.05ng/µL; see above for sequence) and water to 10 µL. Electrophoresis was performed as for the USF assays.

7.5 References

(1) Ferre-D'Amare, A. R.; Prendergast, G. C.; Ziff, E. B.; Burley, S. K. *Nature* **1993**, *363*, 38.
(2) Latchman, D. S. *Int. J. Bio. Cell Bio.* **1997**, *29*, 1305.
(3) Ferre-D'Amare, A. R.; Pognonec, P.; Roeder, R. G.; Burley, S. K. *EMBO J.* **1994**, *13*, 180.
(4) Huang, N.; Chelliah, Y.; Shan, Y.; Taylor, C. A.; Yoo, S.-H.; Partch, C.; Green, C. B.; Zhang, H.; Takahashi, J. S. *Science* **2012**, *337*, 189.
(5) Brivanlou, A. H.; Darnell, J. E. *Science* **2002**, *295*, 813.
(6) Semenza, G. L.; Wang, G. L. *Mol. Cell. Biol.* **1992**, *12*, 5447.
(7) Wang, G. L.; Semenza, G. L. *J. Biol. Chem.* **1993**, *268*, 21513.
(8) Dames, S. A.; Martinez-Yamout, M.; De Guzman, R. N.; Dyson, H. J.; Wright, P. E. *Proc. Natl. Acad. Sci.* **2002**, *99*, 5271.
(9) Freedman, S. J.; Sun, Z.-Y. J.; Poy, F.; Kung, A. L.; Livingston, D. M.; Wagner, G.; Eck, M. J. *Proc. Natl. Acad. Sci.* **2002**, *99*, 5367.
(10) Lercher, L.; McDonough, M. A.; El-Sagheer, A. H.; Thalhammer, A.; Kriaucionis, S.; Brown, T.; Schofield, C. J. *Chem. Commun.* **2014**, *50*, 1794.
(11) Chowdhury, R.; Hardy, A.; Schofield, C. J. *Chem. Soc. Rev.* **2008**, *37*, 1308.
(12) McDonough, M. A.; Li, V.; Flashman, E.; Chowdhury, R.; Mohr, C.; Liénard, B. M. R.; Zondlo, J.; Oldham, N. J.; Clifton, I. J.; Lewis, J.; McNeill, L. A.; Kurzeja, R. J. M.; Hewitson, K. S.; Yang, E.; Jordan, S.; Syed, R. S.; Schofield, C. J. *Proc. Natl. Acad. Sci.* **2006**, *103*, 9814.
(13) Min, J.-H.; Yang, H.; Ivan, M.; Gertler, F.; Kaelin, W. G.; Pavletich, N. P. *Science* **2002**, *296*, 1886.
(14) Hon, W.-C.; Wilson, M. I.; Harlos, K.; Claridge, T. D. W.; Schofield, C. J.; Pugh, C. W.; Maxwell, P. H.; Ratcliffe, P. J.; Stuart, D. I.; Jones, E. Y. *Nature* **2002**, *417*, 975.

(15) Kaelin, W. G. *Annu. Rev. Biochem.* **2005**, *74*, 115.
(16) Maxwell, P. H.; Wiesener, M. S.; Chang, G.-W.; Clifford, S. C.; Vaux, E. C.; Cockman, M. E.; Wykoff, C. C.; Pugh, C. W.; Maher, E. R.; Ratcliffe, P. J. *Nature* **1999**, *399*, 271.
(17) Elkins, J. M.; Hewitson, K. S.; McNeill, L. A.; Seibel, J. F.; Schlemminger, I.; Pugh, C. W.; Ratcliffe, P. J.; Schofield, C. J. *J. Biol. Chem.* **2003**, *278*, 1802.
(18) Lando, D.; Peet, D. J.; Whelan, D. A.; Gorman, J. J.; Whitelaw, M. L. *Science* **2002**, *295*, 858.
(19) Deaton, A. M.; Bird, A. *Genes Dev.* **2011**, *25*, 1010.
(20) Klose, R. J.; Bird, A. P. *Trends Biochem. Sci.* **2006**, *31*, 89.
(21) Wenger, R. H.; Kvietikova, I.; Rolfs, A.; Camenisch, G.; Gassmann, M. *Eur. J. Biochem.* **1998**, *253*, 771.
(22) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
(23) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929.
(24) Nair, S. K.; Burley, S. K. *Cell* **2003**, *112*, 193.
(25) Solomon, D. L. C.; Amati, B.; Land, H. *Nucleic Acids Res.* **1993**, *21*, 5372.
(26) Wang, Z.; Wu, Y.; Li, L.; Su, X.-D. *Cell Res.* **2013**, *23*, 213.
(27) Jelkmann, W. *Eur. J. Haematol.* **2007**, *78*, 183.
(28) Kvietikova, I.; Wenger, R. H.; Marti, H. H.; Gassmann, M. *Nucleic Acids Res.* **1995**, *23*, 4542.
(29) Mellén, M.; Ayata, P.; Dewell, S.; Kriaucionis, S.; Heintz, N. *Cell* **2012**, *151*, 1417.
(30) Prendergast, G.; Ziff, E. *Science* **1991**, *251*, 186.
(31) Fujii, G.; Nakamura, Y.; Tsukamoto, D.; Ito, M.; Shiba, T.; Takamatsu, N. *Biochem. J.* **2006**, *395*, 203.
(32) Farina, A.; Faiola, F.; Martinez, E. *Protein Expression Purif.* **2004**, *34*, 215.
(33) Renciuk, D.; Blacque, O.; Vorlickova, M.; Spingler, B. *Nucleic Acids Res.* **2013**, *41*, 9891.
(34) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci.* **1981**, *78*, 2179.
(35) Chiu, T. K.; Dickerson, R. E. *J. Mol. Biol.* **2000**, *301*, 915.
(36) Severin, Philip M. D.; Zou, X.; Schulten, K.; Gaub, Hermann E. *Biophys. J.* **2013**, *104*, 208.
(37) Thalhammer, A.; Hansen, A. S.; El-Sagheer, A. H.; Brown, T.; Schofield, C. J. *Chem. Commun.* **2011**, *47*, 5325.
(38) Song, J.; Teplova, M.; Ishibe-Murakami, S.; Patel, D. J. *Science* **2012**, *335*, 709.
(39) Valinluck, V.; Sowers, L. C. *Cancer Res.* **2007**, *67*, 946.
(40) Cockman, M. E.; Masson, N.; Mole, D. R.; Jaakkola, P.; Chang, G.-W.; Clifford, S. C.; Maher, E. R.; Pugh, C. W.; Ratcliffe, P. J.; Maxwell, P. H. *J. Biol. Chem.* **2000**, *275*, 25733.

8 Summary and Outlook

Chemical biology approaches have been instrumental for the detailed mechanistic investigation of biological macromolecules, this hold true in epigenetics especially, where small modifications to canonical bases and amino acids can have large functional consequences. In the work of this thesis, I have developed novel chemical biology approaches for the modification of DNA and proteins and subsequently applied these to study epigenetic mechanism.

Access to modified DNA is essential for many biochemical applications. I have developed a novel post-synthetic DNA modification approach utilizing the Suzuki-Miyaura cross coupling (SMcc). The reaction proceeds in water, at ambient temperature and does not require inert atmosphere. This is the first demonstration that SMcc can be used to modify intact DNA at 5-IdU. Using this methodology I was able to create DNA probes containing a photocrosslinker which were used to elucidate hmC binding proteins.

These and other studies have shown that the presence of mC or hmC can influence protein binding to DNA. Since hmC is an oxygen dependent modification, it was of interest to study the effect of hmC on the binding of the HIF transcription factor (TF), the master regulator of the hypoxic response. These studies revealed that HIF binding is diminished by modification, both hemi and symmetric mC and hmC, of the central CpG in its recognition sequence. This is in contrast to other basic helix-loop-helix TFs such as USF and MAX which can tolerate hemi-modification with mC or hmC.

Preliminary experiments for the modification of RNA by SMcc are reported herein. I believe SMcc is particularly well suited for the modification of RNA, since RNA is significantly less stable than RNA and the presented SMcc methodology proceeds under very mild conditions. An advantage of SMcc in comparison to other methodology is the ability to directly introduce modification on the uridine C-5. C-5 substituted uridines with fluorescent properties that depend on DNA secondary structure have been reported. Development of similar probes for RNA would be valuable since RNA can access a

plethora of secondary structures for which a selective fluorescent readout would be desirable. Using post-synthetic modification it is relatively easy to synthesise variations of fluorescent bases for screening without the need for phosphoramidite synthesis and SPS.

The posttranslational modification of proteins is an important mechanism in transcriptional regulation. For *in vitro* study of the effect of these modifications on protein structure, it is essential to have access to large quantities of proteins bearing homogeneous PTMs. I have contributed to the development of a Dha based methodology that allows the introduction of various PTMs into histone proteins. An advantage of this method is the access to phosphocysteine as a phosphoserine mimic and *S*-linked GlcNAcylation as a mimic for *O*-linked GlcNAcylation. Histone tails are densely modified with a variety of different PTMs and the crosstalk between different PTMs is of great interest. While the methodology presented in this thesis allows the introduction of multiple identical PTMs into histone H3, methods that can generate multiple different PTMs on a histone protein are desirable. Development of such methodologies is ongoing work.

The Dha based chemical PTM methodology was used to elucidate the function of *O*-GlcNAcylation on histone H2A Thr-101. Using SEC, melting profiles and non-denaturing MS it was shown that T101-GlcNAc destabilizes histone octamers and in nucleosomes, probably via a destabilization of the dimer-tetramer interface in. This leads to a decreased stability of the formed species, which *in vivo* could result in increased H2A/B exchange and consequently an open chromatin state. Overlap of OGT ChIP peaks and DNase sensitive sites is in agreement with this hypothesis. This study revealed the site-specific effects of *O*-GlcNAcylation at H2A T101. There are multiple other reported *O*-GlcNAcylation sites on histones, for which the function can be explored with the here presented methodology.

Modifications in the nucleosome core can have measurable biophysical effects on the nucleosome structure. The effect of PTMs on the structure of the histone tails is more elusive, because they are intrinsically disordered and thus difficult to characterize by conventional methods. I have used NMR

spectroscopy to describe the residual secondary structure of the H3 tail in detail. The H3 tail can adopt a compact conformation with two α-helical stretches spanning residue 3-11 and 18-28. The structure is mainly acquired through an interaction of the H3 tail with DNA. Under physiological salt concentrations the tail is in exchange between this compact and an open unstructured conformation, with about 50% contribution from both. Phosphorylation of Ser-10 shifts reduces the residual structure and the H3-tail DNA interaction, thus resulting in a more extended conformation. This study is the first description of the H3 tail structure at atomic resolution. When considering the effect of PTMs in epigenetic regulation, observed effects are often reduced to the interactions of an effector protein with the PTM. It is shown here that PTMs such as phosphorylation also effect tail structure, which could in turn affect protein binding to other sites on the H3 tail or chromatin higher order structure. All four histone proteins have *N*- or *C*-terminal tails, all of which have different functions in the organization of chromatin. The detailed structure of the remaining tails could be explored using the here presented toolbox of NMR experiments.

APPENDIX 1. Chapter 3: Proteins identified by LC-MS (forward; hmC=light; C=heavy)

Gene Names	Uniprot	PEP	Ratio H/L	Ratio H/L N	Ratio H/L	Ratio H/	log2(H/L)
ARA160;TMF1	P82094-2;P82094	0.069813	0.007701	0.0058149		1	-7.42603
	Q9P0K0	0.008838	0.006626	0.0093325	93.038	2	-6.74352
CCDC72;HSPC016;HSPC	Q9Y2S6	0.006873	0.01286	0.018093		1	-5.78842
RPL24;hCG_2023003	P83731;C9JNW5;C	2.20E-08	0.011647	0.018333	258.07	4	-5.76941
	B4DRD7	7.27E-106	0.020091	0.028473		1	-5.13426
KIAA0042;KIF14	Q15058	0.089053	0.024161	0.030588		1	-5.03089
		0.000624	0.023462	0.035915	10.311	2	-4.79927
KIAA0296;ZNF646	O15015-2;O15015	0.088792	0.028676	0.039728	43.717	2	-4.6537
GAPD2;GAPDH2;GAPDI	O14556	6.19E-05	0.031398	0.04366	5.6678	2	-4.51754
HBP;HDLBP;VGL	Q00341;B2R5V9;E	0.005329	0.033592	0.048823		1	-4.3563
CRDBP;IGF2BP1;VICKZ1	Q9NZI8;D3DTW3	0.001889	0.037391	0.050814		1	-4.29863
IBP3;IGFBP3;hCG_1735	P17936;A6XND0;A	1.28E-07	0.037068	0.052608	121.56	6	-4.24857
C6orf98;KIAA0796;KIA/	Q8NF91-1;Q8NF9	0.010116	0.034683	0.055904	18.598	2	-4.1609
HMG2;HMGB2	P26583;Q5U071;D	2.18E-14	0.0424	0.067693	192.31	4	-3.88485
HMG14;HMGN1;hCG_1	P05114;Q3B790;A	5.57E-15	0.04643	0.069726		1	-3.84216
AUF1;HNRNPD;HNRPD	Q14103-2;Q14103	5.37E-118	0.044941	0.070552		1	-3.82517
DMD;GS1-19O24.1-006	P11532-1;P11532	0.008954	0.055105	0.07077		1	-3.82072
AUF1;HNRNPD;HNRPD	Q14103-1;Q14103	7.97E-175	0.05491	0.079821	144.31	69	-3.64709
IGHM	P04220;P01871-1	1.31E-09	0.067675	0.083606		1	-3.58025
hCG_1744585;RAN;AR/	B5MDF5;P62826;A	0.009414	0.06453	0.088324		1	-3.50105
PDE5;PDE5A	O76074-1;O76074	0.004242	0.055282	0.089074		1	-3.48885
SRP9;hCG_1781062;hC	P49458;A8K0N0;C	2.29E-11	0.067365	0.093878	153.31	3	-3.41307
RPL29;RP11-632C17__/	A8K0H3;P47914;C	8.04E-05	0.069116	0.098074	58.325	2	-3.34999
DCD;AIDD;DSEP;hCG_1	A5JHP3;P81605;Q	0.000215	0.075223	0.10238		1	-3.28799
TIA1	P31483-2;P31483	9.60E-46	0.065824	0.10291	11.281	2	-3.28054
DDX49	Q9Y6V7;A8K7A1;C	0.001282	0.07964	0.10591		1	-3.23909
PC4;RPO2TC1;SUB1;hC	P53999;B7Z1Z0;Q	1.37E-37	0.073206	0.10635	129.01	3	-3.23311
EEF1A;EEF1A1;EF1A;LE	P68104;A8K9C4;A	6.02E-27	0.067829	0.10828	139.08	17	-3.20716
DNAH10;KIAA2017	Q8IVF4-1;Q8IVF4;	0.087971	0.078248	0.11224		1	-3.15534
DFS70;LEDGF;PSIP1;PSI	O75475-1;O75475	1.79E-115	0.082223	0.11709	120.76	28	-3.09431
H1F3;HIST1H1D;H1F4;H	P16402;P10412;A	3.68E-10	0.076807	0.11802	137.53	4	-3.0829
C1orf174	Q8IYL3	0.000411	0.093767	0.12966		1	-2.94719
DBP1;DDX15;DHX15	O43143;B4E0S6	0.005092	0.09722	0.13023	90.035	2	-2.94087
NPM;NPM1	P06748-1;P06748	0.012239	0.095961	0.13528		1	-2.88598
FOXA1;HNF3A;TCF3A	P55317;B4E257;B	2.51E-06	0.10571	0.13708	106.53	3	-2.86691
FAU	P35544;P62861	0.012101	0.099286	0.13928		1	-2.84394
DEK	P35659;B4DNW3;	2.16E-123	0.12934	0.14215	137.84	45	-2.81451
HNRNPA2B1;HNRPA2B	P22626-1;P22626	5.13E-193	0.10804	0.14315	89.552	72	-2.8044
KPNA2;RCH1;SRP1	P52292;A8K7D9;C	0.010612	0.11086	0.14373		1	-2.79857
G22P2;XRCC5	P13010;Q53T09;C	1.58E-63	0.099512	0.14462	118.9	30	-2.78966
REPA1;RPA1;RPA70	P27694	4.77E-125	0.10462	0.14601	124.26	33	-2.77586
HNRNPA0;HNRPA0	Q13151	7.03E-109	0.10881	0.14629	152.89	27	-2.7731
EEF1G;EF1G;PRO1608;I	B4DTG2;P26641;C	7.47E-24	0.11408	0.14683	141.27	3	-2.76778
PRSS3;PRSS4;TRY3;TRY	P35030-1;P35030	0.002293	0.10796	0.14859		1	-2.75059
LAP2;TMPO	P42166	5.73E-49	0.11296	0.1491	10.908	2	-2.74565
GSTK1;HDCMD47P;hCG	Q9Y2Q3-2;Q9Y2Q	2.28E-19	0.10584	0.15258	21.528	3	-2.71236
RPS2;RPS4;LOC392781;	P15880;A4D0Y7;C	0.000171	0.10956	0.15468	30.855	2	-2.69264
NFIC;NFI	B5MDB4;B7Z4U5;	6.83E-07	0.1059	0.16149	127.55	3	-2.63048

RBM3;RNPL	P98179	2.18E-15	0.13145	0.16281	29.412	3	-2.61874
HNRNPA1;HNRPA1;HN	P09651-1;P09651	2.01E-206	0.12628	0.16694	89.536	63	-2.5826
H1F5;HIST1H1B	P16401	1.26E-09	0.11325	0.16733	128.84	4	-2.57923
TCF13;TEAD1;TEF1	P28347;Q59EF3	5.66E-05	0.12275	0.16862		1	-2.56815
NCL;hCG_33980	P19338;B3KM80;I	2.47E-50	0.12268	0.17227	137.73	20	-2.53726
CNBP;RNF163;ZNF9	A8K7V4;Q5U0E9;I	8.10E-10	0.14176	0.17444	75.157	3	-2.5192
BAF190B;BRM;SMARC/	P51531-1;P51531	9.75E-11	0.18271	0.17456		1	-2.51821
BRUNOL2;CELF1;CUGBI	Q92879-4;Q9287!	3.19E-38	0.123	0.17905	102.05	21	-2.48157
HMGN3;PNAS-24;TRIP7	Q15651-1;Q1565:	2.95E-28	0.12038	0.18049	181.68	2	-2.47001
HNRNPL;HNRPL;P/OKcl	P14866;B2R959;C	4.81E-113	0.14368	0.18559	144.68	42	-2.42981
NS5ATP5;TFB2M	Q9H5Q4;B4DTJ4	0.000207	0.13654	0.18564		1	-2.42942
CAF;CAF1P150;CHAF1A	Q13111-1;Q1311:	2.45E-10	0.12428	0.18722		1	-2.41719
ABBP1;HNRNPAB;HNRI	Q99729-3;Q9972!	9.47E-60	0.14826	0.18958	259.7	17	-2.39912
NFYC;DKFZp667G242;R	Q13952-1;Q1395:	2.16E-11	0.13223	0.18981		1	-2.39737
LOC392748;tcag7.1276	A4D1G5;C9JLI6;P4	0.003673	0.1373	0.19091		1	-2.38904
DDX21;OK/SW-cl.65	Q9NR30-1;Q9NR3	0.002584	0.14483	0.19613	87.683	4	-2.35012
A18HNRNP;CIRBP;CIRP	Q14011;Q53XX5;I	5.97E-27	0.13937	0.19773		1	-2.3384
CGI-55;PAIRBP1;SERBP	Q8NC51-1;Q8NC5	1.28E-63	0.14067	0.19968	138.24	7	-2.32424
BBS14;CEP290;KIAA037	O15078-1;O1507!	0.13248	0.15069	0.20244		1	-2.30443
MAP4	P27816-1;P27816	7.17E-58	0.1332	0.20435	130.93	17	-2.29089
RECQ1;RECQL;RECQL1	P46063	2.58E-80	0.14322	0.20587	106.32	26	-2.28019
DDX9;DHX9;LKP;NDH2	Q08211;B3KU66;(	1.08E-11	0.15393	0.20758	238.45	8	-2.26826
TARDBP;TDP43;RP4-63	Q13148-2;Q1314!	3.85E-78	0.1746	0.2091	102.43	11	-2.25774
TOP1	P11387;B9EG90;C	5.71E-58	0.17106	0.21015	126.83	28	-2.25051
FUS;TLS	P35637-1;P35637	2.66E-33	0.189	0.21386	114.74	24	-2.22526
ZHX1	Q9UKY1;A8K0G7	6.29E-17	0.14876	0.22018		1	-2.18324
HMG1;HMGB1;RP11-5!	P09429;Q5T7C6;C	3.79E-17	0.14739	0.22414	114.91	6	-2.15753
RBM47	A0AV96-1;A0AV9(	0.000291	0.17408	0.22696	40.42	2	-2.13949
ATF1	P18846;B2R8K4;B	6.87E-13	0.16627	0.23314	146.76	8	-2.10073
VRK1	Q99986;A8K245	3.42E-05	0.1878	0.2339	50.238	2	-2.09604
SSBP;SSBP1;hCG_2014	Q04837;A4D1U3;(	1.03E-38	0.16666	0.23459	102.77	11	-2.09179
FBP3;FUBP3;hCG_3125	Q96I24-1;Q96I24;	8.31E-181	0.19578	0.24453	70.489	69	-2.03192
HNRNPA3;HNRPA3;hC(	P51991-1;P51991	2.28E-60	0.17397	0.24969	88.835	18	-2.00179
FUBP1	Q96AE4-1;Q96AE-	5.38E-133	0.19398	0.25681	91.772	39	-1.96123
HNRPQ;NSAP1;SYNCRII	O60506-3;O6050(	1.42E-127	0.21155	0.25756	124.18	30	-1.95702
FUBP2;KHSRP	Q92945-1;Q9294!	2.17E-157	0.2016	0.25896	59.322	34	-1.9492
SFRS7	Q16629-1;Q1662!	5.88E-06	0.18703	0.26497	133.6	4	-1.9161
CASP11;SFRS2IP;SIP1	Q99590-1;Q9959(	2.69E-05	0.2249	0.26548	143.66	2	-1.91332
C3orf37;DC12	Q96FZ2;B4DMK2;	1.32E-79	0.18986	0.2688	111.22	16	-1.89539
hCG_20560;PTBP1;PTB	Q9BUQ0;P26599-	6.90E-210	0.17972	0.27122	126.54	69	-1.88246
HNRNPU;HNRPU;SAFA;	Q00839-1;Q0083!	3.26E-137	0.19912	0.2717	142.8	34	-1.87991
DDX36;DHX36;KIAA148	Q9H2U1-1;Q9H2U	5.86E-118	0.19905	0.27283	61.319	14	-1.87393
HNRNPK;HNRPK;hCG_1	P61978-2;P61978	4.29E-233	0.21745	0.27505	135.77	65	-1.86223
CSDA;DBPA	P16989-1;P16989	4.16E-20	0.21785	0.27786	48.342	5	-1.84757
HSET;KIFC1;KNSL2	Q9BW19;B4DMA8	3.41E-21	0.20576	0.28124	205.21	2	-1.83013
C20orf158;DATF1;DIDC	Q9BTC0-4;Q9BTC(	5.41E-08	0.19746	0.28484	69.642	3	-1.81178
CCT3;CCTG;TRIC5;RP11	P49368;B4DUR8;(	1.52E-10	0.21061	0.28655	177.64	2	-1.80314
ADPRT;PARP1;PPOL	P09874;B2R5W3;I	0	0.23292	0.29369	150.54	176	-1.76763
MSI2	B4DHE8;Q96DH6-	2.06E-08	0.22156	0.29521	182.46	2	-1.76019
RBM4;RBM4A;RBM30;I	Q9BWF3-1;Q9BW	8.25E-128	0.22967	0.3017	105.21	42	-1.72881

C2orf12;MSSP;MSSP1;I	P29558-1;P29558	8.58E-13	0.20884	0.30363	174.97	2	-1.71961
SFRS3;SRP20;hCG_148(	P84103;B2R6F3;B	5.42E-05	0.22952	0.30679	70.222	2	-1.70468
HNRNPR;HNRPR;hCG_3	O43390-2;O4339(	1.32E-24	0.22256	0.30712	112.33	9	-1.70313
DDX48;EIF4A3;KIAA011	P38919	0.009054	0.2198	0.30755		1	-1.70111
LRP130;LRPPRC	P42704;B4DSR0;E	4.20E-48	0.24797	0.30974	96.833	15	-1.69087
OK/SW-cl.32;RPL3;hCG	P39023;Q96QL0;E	0.000142	0.22227	0.31654	102.61	3	-1.65954
DXS648E;QM;RPL10;XX	P27635;Q5HY50;C	0.014478	0.24474	0.3169	165.64	2	-1.6579
HDGF;HMG1L2;RP11-6	P51858;B2RDE8;E	1.26E-38	0.22528	0.31868	83.161	5	-1.64982
PICALM;PICALM varian	A8K5U9;A8MW24	1.99E-22	0.24046	0.31868	93.095	9	-1.64982
ELAVL1;HUR	B4DVB8;Q15717	6.07E-53	0.2493	0.32873	98.07	10	-1.60502
DRAP1	Q14919-2;Q1491(	7.08E-06	0.23868	0.32982	99.603	4	-1.60025
hCG_2017557;PCBP2	Q59HD4;Q6IPF4;C	5.33E-123	0.23347	0.33282	126.39	16	-1.58719
HNRNPH3;HNRPH3	P31942-1;P31942	1.12E-18	0.25647	0.33337	51.078	4	-1.5848
C4orf43	Q96EY4;D6RA57	7.42E-51	0.25504	0.33745	131.71	6	-1.56725
HMG2A;HMG4;HMGB3	O15347;Q16466	1.02E-27	0.22013	0.33875	103.82	8	-1.56171
PSF;SFPQ	P23246-1;P23246	4.96E-69	0.2388	0.33924	97.721	23	-1.55962
CUL1	Q13616;B3KTW0;	0.10854	0.2542	0.33962		1	-1.55801
TIAL1;hCG_40603	A8K4L9;A8K5C4;C	1.40E-51	0.25481	0.35613	9.8767	3	-1.48952
PCBP1;hCG_1776997	Q15365;Q53SS8	2.46E-111	0.2759	0.36033	111.45	25	-1.47261
SPTB2;SPTBN1;hCG_17	Q01082-1;Q0108:	0.002436	0.25301	0.36142		1	-1.46825
HNRPDL;JKTBP	O14979-1;O1497(	1.09E-10	0.29097	0.36322	135.69	3	-1.46108
TIA1;hCG_40580	P31483-1;P31483	3.34E-45	0.28083	0.37092	72.947	15	-1.43082
	B4DGK4	0.003681	0.30004	0.37271	113.13	4	-1.42387
UBC;DKFZp434K0435;h	P0CG48;A8K674;C	1.14E-25	0.26443	0.37426	198.31	11	-1.41789
ZFP29;ZSCAN2	Q7Z7L9-1;Q7Z7L9	0.32681	0.30043	0.37549		1	-1.41315
G22P1;XRCC6;CTA-216	P12956;B2RDN9;E	1.84E-94	0.33118	0.37882	123.56	28	-1.40042
RP11-375E1__A.4-005;	Q59GE4;Q5TZB9;I	6.07E-05	0.25964	0.37971		1	-1.39703
DKFZp779B0247;CSDE1	Q68DF1;O75534-1	1.75E-24	0.31478	0.38135	135.98	11	-1.39081
HNRNPM;HNRPM;NAG	P52272-1;P52272	2.14E-75	0.31263	0.38213	85.947	24	-1.38786
RPN1	P04843;Q53EP4;C	0.010782	0.28979	0.39116		1	-1.35417
DDX2B;EIF4A2;EIF4F;DI	Q14240-2;Q1424(	3.59E-06	0.29707	0.39207	155.24	2	-1.35082
SFRS2;hCG_27842;SFR	Q01130;B3KUY1;C	5.01E-25	0.31358	0.39317	29.142	6	-1.34677
BTEB4;KLF16;NSLP2	Q9BXK1;D6W5Z8	1.53E-31	0.31828	0.39555		1	-1.33807
FIR;PUF60;ROBPI;SIAHI	Q9UHX1-1;Q9UH)	3.51E-21	0.29121	0.39939	65.145	6	-1.32413
BLOCK24;HNRPLL;SRRF	Q8WVV9-1;Q8WV	0.000202	0.31742	0.40056	55.911	2	-1.31991
IQGAP1;KIAA0051	P46940;A4QPB0;E	2.41E-16	0.27648	0.40131	104.86	3	-1.31721
E1BAP5;HNRNPUL1;HN	Q9BUJ2-1;Q9BUJ2	1.70E-19	0.27648	0.40131	120.17	3	-1.31721
CGGBP;CGGBP1	Q9UFW8;C9JUJ0	1.33E-06	0.2955	0.40248	173.7	2	-1.31301
CAS;CSE1L;XPO2;CSE1L	P55060-1;P55060	0.014654	0.27639	0.4147	278.71	2	-1.26986
HCC1;RBM39;RNPC2;D	Q14498-1;Q1449{	3.73E-59	0.33697	0.41564	68.689	9	-1.26659
DAZAP1	Q96EP5-1;Q96EP5	6.31E-94	0.31466	0.42046	84.474	16	-1.24996
IGF2BP3;IMP3;KOC1;VI	O00425-1;O0042:	6.47E-11	0.33465	0.42077	87.704	7	-1.2489
NSEP1;YB1;YBX1	P67809;A0JLU4;Q	2.65E-91	0.25598	0.43396	149.46	19	-1.20437
NPTB;PTB;PTBLP;PTBP2	Q9UKA9-2;Q9UKA	7.81E-17	0.32231	0.43725	11.35	2	-1.19347
EWSR1;hCG_2010995;I	Q96MX4;Q01844-	1.28E-24	0.34603	0.4402	60.348	3	-1.18377
HSPC144;MDS012;My0	Q9P016-1;Q9P01(	6.47E-12	0.32384	0.44086	133.08	5	-1.18161
RFC1;RFC140;LLDBP	P35251-1;P35251	0.003283	0.3151	0.44467		1	-1.16919
MECP2;hCG_37678	P51608-2;P51608	1.47E-38	0.3528	0.46218	191.03	2	-1.11347
HNRNPH1;HNRPH;HNR	P31943;B4DP51;C	1.95E-79	0.37484	0.46626	72.011	15	-1.10079
DBX;DDX3;DDX3X;DBY;	O00571;B4E3E8;E	1.60E-31	0.38143	0.47004	133.36	10	-1.08914

OK/SW-cl.26;RPS3	P23396;Q53G83;(	8.67E-07	0.33187	0.47354	72.584	5	-1.07844
BAP135;GTF2I;WBSCR6	P78347-1;P78347	1.73E-17	0.32284	0.47694	98.937	7	-1.06812
H4/A;H4/B;H4/C;H4/D;	P62805;B2R4R0;C	9.67E-16	0.38387	0.47877	129.47	10	-1.0626
RPL30	P62888;E5RI99;E5	3.85E-23	0.37144	0.48455	116.23	4	-1.04528
IFI16;IFNGIP1;hCG_199	Q16666-2;Q1666(	6.66E-103	0.33879	0.48556	124.89	25	-1.04228
NMT;NMT1	P30419-1;P30419	0.000127	0.349	0.48848	6.4037	2	-1.03363
REPA2;RPA2;RPA32;RP	P15927-3;P15927	6.85E-23	0.40183	0.50021	155.12	6	-0.99939
DDX17;RP3-434P1.4-0(	Q59F66;Q92841-4	5.66E-32	0.41146	0.50126	59.592	7	-0.99637
HNRPG;RBMX;RBMXP1	P38159;B4E352;Q	1.52E-13	0.37601	0.51458	122.16	3	-0.95853
RENT3B;UPF3B;UPF3X	Q9BZI7-1;Q9BZI7;	0.000584	0.40252	0.51777	117.67	2	-0.94962
ANKT;BM-037;NUSAP1	Q9BXS6-1;Q9BXS(	6.22E-23	0.44778	0.54006	117.63	9	-0.88881
KIAA1401;TSR1	Q2NL82	5.70E-11	0.37939	0.54125	90.668	3	-0.88563
TOE1	Q96GM8;B3KSC7;	0.001022	0.40255	0.54137		1	-0.88531
CGI-25;NOSIP	Q9Y314;A8K670;E	0.006059	0.43633	0.54873		1	-0.86583
GTBP;MSH6	P52701-1;P52701	1.54E-81	0.38163	0.54939	113.91	24	-0.8641
RPL11;hCG_1739238;R	P62913-1;P62913	1.73E-33	0.41962	0.55078	131.73	10	-0.86045
RBP56;TAF15;TAF2N;h(	Q92804-1;Q9280-	8.65E-37	0.47177	0.58201	16.195	4	-0.78088
LIG3	P49916-1;P49916	1.60E-57	0.47197	0.5973	90.387	15	-0.74347
TCF6;TCF6L2;TFAM	Q00059;E5KSU5;E	5.98E-57	0.48656	0.59961	132.67	20	-0.7379
TOP2A;TOP2	C9J4C3;P11388-4;	6.39E-14	0.44586	0.60089	90.401	4	-0.73483
RPS5	P46782;Q53G25	1.69E-07	0.46927	0.60462	97.642	6	-0.7259
CGI-21;HSPC332;PRO28	Q53GS9;B7Z7L9;E	2.71E-26	0.45748	0.60779	94.209	7	-0.71836
BRD4;HUNK1;hCG_200	O60885-1;O6088!	0.008151	0.41333	0.61217		1	-0.708
AC004383.6-002;hCG_1	Q5JRC6;Q8IWS0-1	2.93E-10	0.48427	0.62059	105.87	8	-0.68829
RPS16	B4DP32;P62249;C	8.92E-07	0.47022	0.62463	98.554	5	-0.67893
MORF4L1;FWP006;HSP	A5D8W6;B3KTM8	0.00074	0.46242	0.63161	35.846	2	-0.66289
KIAA0026;MORF4L2;M	Q15014;B3KWX6;	0.008934	0.50423	0.63739		1	-0.64975
NSUN2;SAKI;TRM4;hC(	Q08J23;A8K529;B	1.88E-62	0.47584	0.64108	89.08	20	-0.64142
DDX5;G17P1;HELR;HLR	P17844;B4DN41;E	1.06E-11	0.52831	0.6607		1	-0.59793
POLRMT	O00411;Q4G0F4;(	3.47E-10	0.5319	0.67628	7.5559	3	-0.56431
NONO;NRB54	Q15233;A8K525;E	5.16E-97	0.55701	0.68159	115.12	44	-0.55302
KIF4;KIF4A	O95239-1;O9523!	9.75E-55	0.54367	0.68959	97.985	17	-0.53619
FEN1;RAD2;hCG_4084!	P39748;B4DWZ4;(	7.53E-101	0.50395	0.6927	150.98	18	-0.5297
TOP2B;topIIb;top2beta	Q02880-1;Q0288(	6.85E-11	0.51796	0.70427		1	-0.5058
MUTYH;RP4-534D1.2-0	E5KP25;E5KP30;E!	0.001055	0.55805	0.70667	51.827	2	-0.50089
KIAA0618;NUP121;POM	Q96HA1-1;Q96HA	0.006248	0.5638	0.70747		1	-0.49926
NFIA;RP5-902P15.1-00	B4DS53;E2QRI7;B	2.12E-21	0.54265	0.7153	143.27	8	-0.48338
HIP116A;HLTF;RNF80;S	Q14527-1;Q1452	4.94E-289	0.57942	0.73174	140.57	89	-0.4506
RFC2;hCG_18034	P35250-1;P35250	0.000365	0.52913	0.73468		1	-0.44481
KIAA0052;SKIV2L2	P42285;A8K6I4;B4	0.011265	0.58588	0.74895		1	-0.41706
FOXK1;MNF;KIAA0415;	P85037-1;P85037	7.14E-18	0.59275	0.75091	52.038	10	-0.41329
OCT1;OTF1;POU2F1;RP	P14859-1;P14859	5.62E-07	0.59826	0.75288		1	-0.40951
RBM42	Q9BTD8-1;Q9BTD	2.47E-19	0.5586	0.76968	12.154	2	-0.37767
ADAR2;NUP98	P52948-1;P52948	4.67E-32	0.59973	0.77259	170.25	4	-0.37223
RPL7;WUGSC:H_RG054	P18124;O95036;A	2.98E-24	0.62479	0.77447	105.9	7	-0.36872
KIAA1185;LRRC47	Q8N1G4	9.93E-30	0.63456	0.77533	17.165	5	-0.36712
HAP3;NFYB	P25208	0.012304	0.52126	0.7787		1	-0.36086
KIAA1584;SUHW4;ZNF	Q6N043-1;Q6N04	6.15E-05	0.58553	0.78153	216.72	2	-0.35563
RPS19;hCG_1995572	P39019;B0ZBD0;C	0.000332	0.63315	0.78357		1	-0.35187
NUP153;NUP153 varia	B4DIK2;Q7Z743;P	2.30E-118	0.62235	0.78413	99.288	16	-0.35084

MSN	P26038;Q6PJT4	1.32E-12	0.61129	0.78628	186.35	4	-0.34688
H2BFD;HIST1H2BN;H2E	B4DR52;Q99877;E	2.03E-32	0.64428	0.78798	148.75	9	-0.34377
DRBF;ILF3;MPHOSPH4;	Q12906-4;Q1290(	2.50E-26	0.63184	0.79158	97.566	6	-0.33719
TBX3	O15119-1;O1511!	1.21E-85	0.61289	0.79197	62.996	10	-0.33648
FOXK2;ILF;ILF1	Q01167-1;Q0116'	3.07E-14	0.65303	0.79363	11.594	2	-0.33346
C2F;EMG1	Q92979;A8K6D2;E	3.10E-06	0.62091	0.79744	18.969	3	-0.32655
DPC4;MADH4;SMAD4	Q13485;Q9BYG6;I	2.72E-08	0.6275	0.80033		1	-0.32133
LDHB;hCG_24788	P07195;Q5U077;A	8.16E-17	0.61436	0.80374	101.16	4	-0.3152
CTA-216E10.8-008;NHE	B1AHD1;P55769;(	1.93E-07	0.66665	0.82214		1	-0.28254
PES1;hCG_2010949	O00541-1;O0054'	4.90E-18	0.60845	0.82551	57.269	7	-0.27664
FLYWCH2	Q96CP2	0.016053	0.61261	0.83259		1	-0.26432
ABC50;ABCF1;DADB-12	Q8NE71-1;Q8NE7	3.30E-131	0.55943	0.83481	111.71	29	-0.26048
hCG_39182;RDX;DKFZp	A7YIJ8;A7YIK3;P3!	4.12E-13	0.63428	0.84468		1	-0.24352
EIF5B;IF2;KIAA0741;hC	O60841;A0JLR8;B	1.58E-10	0.66702	0.84802	221.19	3	-0.23783
IGKJRB;IGKJRB1;RBPJ;R	Q06330-1;Q0633(	0.001534	0.70902	0.85491		1	-0.22616
LDHA;PIG19;hCG_9667	B7Z5E3;P00338-1	4.30E-09	0.69554	0.86299		1	-0.21258
RFC3	P40938;B4DKE6;C	2.73E-11	0.69231	0.86519		1	-0.20891
ZNF787	Q6DD87	2.26E-06	0.72167	0.87537		1	-0.19204
RBM14;SIP	Q96PK6-1;Q96PK(	2.34E-113	0.72993	0.87744	35.171	39	-0.18863
LYT10;NFKB2;hCG_256	Q00653-1;Q0065'	0.005831	0.62763	0.89003		1	-0.16807
RPS9;hCG_2009111;XX	P46781;A9C4C1;B	0.002389	0.69696	0.89809		1	-0.15507
CYPB;PPIB	P23284	8.36E-28	0.70893	0.89904	131.21	20	-0.15354
CDC2;DKFZp686L20222	P06493-1;P06493	1.57E-09	0.72593	0.89991	7.1124	3	-0.15215
SMARCA5;SNF2H;WCR	O60264;B4DZC0;E	2.30E-22	0.71575	0.90112	90.789	11	-0.15021
SRBD1	Q8N5C6-1;Q8N5C	4.98E-17	0.72445	0.9095	1.6967	2	-0.13685
EZR;VIL2	P15311;B2R6J2;B'	9.21E-19	0.70123	0.91062	89.403	9	-0.13508
UHRF1;ICBP90;NP95;RI	A8K024;B2RBA9;(	8.34E-303	0.69597	0.91119	101.83	61	-0.13418
RPS20	B4DW28;P60866;	5.71E-10	0.73884	0.91348	110.32	7	-0.13055
CDC2L4;CDK9;CDC2L;C	P50750-2;P50750	3.21E-05	0.7221	0.91834		1	-0.1229
QPCTL	Q9NXS2	0.001076	0.73942	0.91973		1	-0.12072
DR1;DKFZp666G145;hC	Q01658;Q53F47;(	0.000156	0.74286	0.91993		1	-0.1204
APRIN;AS3;KIAA0979;P	Q9NTI5-1;Q9NTI5	1.23E-14	0.73323	0.92028	23.603	2	-0.11986
AROS;RPS19BP1	Q86WX3	2.51E-05	0.74651	0.92824		1	-0.10743
SRP14	P37108	1.39E-11	0.74794	0.9314	37.186	3	-0.10253
RPL13A;hCG_1998057;	P40429;Q0VGL3;(	0.087812	0.72747	0.93494	5.5781	2	-0.09705
KID;KIF22;KNSL4;hCG_2	Q14807;B7Z265;B	6.29E-127	0.78155	0.94994	117.79	29	-0.07409
KIAA1470;RCC2;TD60	Q9P258;A5PLK7	2.64E-50	0.73193	0.95118	122.23	20	-0.07221
SMARCA4;hCG_29955;	B9EGQ8;Q9HBD4;	1.26E-28	0.75978	0.95689	11.214	9	-0.06358
ESRRB;ERRB2;ESRL2;NF	A2VDJ2;Q5F0P7;C	0.007567	0.72634	0.95795	1.825	2	-0.06198
BBC1;OK/SW-cl.46;RPL	P26373;A8K4C8;B	4.13E-17	0.75668	0.97157	15.4	6	-0.04161
MSH2	P43246;B4DL39;B	7.97E-27	0.75626	0.97247	63.057	10	-0.04027
HYRC;HYRC1;PRKDC	P78527-1;P78527	4.65E-16	0.78394	0.97739	68.77	6	-0.03299
KPNB1;NTF97	Q14974;B2RBR9;E	7.12E-08	0.77766	0.98508	11.236	3	-0.02169
EAAT1;GLAST;GLAST1;S	P43003;Q8N169;(	0.00623	0.6958	0.98975		1	-0.01486
FOX2;HRNBP2;RBM9;R	O43251-8;O4325'	1.16E-16	0.79004	0.99268	50.226	3	-0.0106
RCC1;CHC1;hCG_2780!	Q16269;Q6NT97;(	2.52E-134	0.79286	0.99339	117.13	21	-0.00957
EEF2;EF2	P13639;B4DMC6;	5.88E-11	0.82347	1.0027	73.199	5	0.00389
RPS25	P62851	4.60E-07	0.77529	1.0052	85.111	5	0.007483
HUSSY-29;NSA2;TINP1;	O95478;Q5J7U2	0.000455	0.82193	1.0054		1	0.00777
BAT1;UAP56;XXbac-BC'	Q13838-2;Q1383!	7.42E-17	0.81485	1.0139	77.286	7	0.019915

CCT4;CCTD;SRB	P50991;A8K3C3;B	0.00104	0.75528	1.0177		1	0.025312
ANX2;ANX2L4;ANXA2;(	P07355-2;P07355	5.32E-45	0.75938	1.0182	68.804	5	0.026021
LRRC59;PRO1855	Q96AG4;B2RE83;I	5.58E-11	0.76328	1.0224	110.39	5	0.03196
PRO2640;RPS14	P62263;E5RH77	0.009218	0.83296	1.0373		1	0.052833
EAR2;ERBAL2;NR2F6	P10588;B4DQE4;E	3.62E-07	0.82796	1.0447	63.33	3	0.063089
APE;APEX;APEX1;APX;H	P27695;Q5TZP7	3.23E-161	0.84311	1.0449	126.46	31	0.063365
CAIN;CAN;KIAA0023;NI	P35658-3;P35658	1.03E-77	0.8177	1.0473	140.42	19	0.066675
GNB2L1;HLC7;PIG21	P63244;D6R9L0;D	3.33E-08	0.84464	1.0487	4.9983	2	0.068602
BAZ1B;WBSC10;WBSCF	Q9UIG0-1;Q9UIG(	4.04E-60	0.83297	1.0519	118.57	11	0.072998
OK/SW-cl.83;RPS8;hCG	P62241;Q5JR94;C	5.04E-21	0.82342	1.0552	6.6342	2	0.077516
ALY;BEF;THOC4	Q86V81	4.13E-11	0.75781	1.0588	9.1553	2	0.08243
TDG	Q13569;B2R848;E	5.91E-09	0.89091	1.061	120.92	5	0.085425
RPL19	P84098;Q53G49;(	1.58E-12	0.83229	1.0634	39.276	3	0.088684
HCA90;hCG_38821;RP1	Q96RR5;Q9ULW0	5.40E-09	0.844	1.0727	5.7337	2	0.101247
ADPRT2;ADPRTL2;PARF	Q9UGN5-1;Q9UG	2.56E-07	0.79715	1.0769	166.99	3	0.106884
API5;MIG8	Q9BZZ5-4;Q9BZZ5	2.05E-13	0.88943	1.086	10.225	4	0.119024
ILF2;NF45;PRO3063	Q12905;B4DY09;(	4.11E-18	0.88134	1.0885	73.433	4	0.122341
H2AFX;H2AX;H2AFR;HI	P16104;Q96QV6;I	2.68E-18	0.88644	1.089	191.73	3	0.123004
HCC1;HSPC316;SARNP;	P82979;Q567R9;A	9.26E-07	0.81997	1.0982	6.2426	2	0.135141
NIRF;RNF107;UHRF2;D	Q96PU4-1;Q96PU	9.82E-24	0.77359	1.1098	31.451	6	0.1503
SP1;TSFP1;Sp1	P08047;Q9NR52;(	9.91E-10	0.91653	1.1135		1	0.155102
GCN1L1;KIAA0219	Q92616;B4DM32;	7.27E-05	0.81923	1.1293	4.6907	2	0.175429
CHD4;CHD3	Q14839-2;Q1483!	0.036417	0.88663	1.1325	2.456	2	0.179511
PLCB3	Q01970;Q8N1A4	2.67E-37	0.85112	1.1362	11.312	6	0.184217
MAFF;CTA-447C4.3-00	Q9ULX9;B0QY71;I	0.00523	0.90053	1.1399	23.25	2	0.188907
RPL27A;L27a	P46776;Q6NZ52;(	2.02E-06	0.92662	1.1515		1	0.203514
TCOF1;DKFZp434G103!	Q13428-4;Q1342!	1.77E-09	0.93359	1.1634		1	0.218347
LAMBR;LAMR1;RPSA;L/	A6NE09;P08865;(	0.000502	0.93928	1.1664		1	0.222063
KIAA0648;PDS5;PDS5A	Q29RF7-1;Q29RF7	6.38E-26	0.94192	1.1727	47.972	5	0.229834
ANT2;SLC25A5;ANT3;C	P05141;B2RCV1;(	1.65E-10	0.92052	1.174	83.485	10	0.231432
AIM;CXXC9;DNMT;DNM	P26358-1;P26358	2.00E-48	0.94226	1.174	83.532	19	0.231432
TUBG;TUBG1;TUBG2	P23258;Q9NRH3	1.90E-28	0.91542	1.1901	77.689	10	0.251083
DDXBP1;PIAS1;hCG_23	O75925;A4ZVS7;A	0.003073	0.96349	1.1914		1	0.252658
ASNS;NARS	O43776;B4DN60;I	0.0101	0.97284	1.2044	31.015	3	0.268315
SUPV3L1;SUV3	Q8IYB8;B7Z611	2.11E-36	0.99057	1.2054	18.262	9	0.269512
LONP1;PRSS15	P36776;E5KMH8;I	6.29E-06	0.98342	1.2057	103.82	3	0.269871
SCML2;RP5-1129A6.3-(	Q9UQR0-1;Q9UQ	6.38E-08	0.93657	1.2218	24.697	5	0.289008
TRIP12;KIAA0045	D4HL82;Q14CA3;(	4.05E-05	0.90068	1.2258	8.802	2	0.293724
KIAA0170;MDC1;NFBD	Q14676-1;Q1467(	1.26E-68	0.97068	1.227	115.08	11	0.295135
TREX1;hCG_1997849	Q9NSU2-1;Q9NSU	1.71E-36	1.0108	1.2313	72.827	10	0.300182
NTH1;NTHL1;OCTS3	P78549;E5KTI5	4.69E-41	1.0132	1.263	32.497	12	0.336855
OK/SW-cl.48;PHC;SLC2	Q00325-1;Q0032!	2.76E-11	1.0475	1.2873		1	0.364348
LAP2;TMPO	P42167-1;P42167	2.80E-60	0.99137	1.2948	85.098	16	0.372729
GCS1;MOGS	Q13724;Q58F09;A	0.001144	1.0381	1.3147		1	0.394734
DUC1;DUG;MSH3	P20585;A1L480;A	8.52E-05	1.0459	1.3174	25.324	3	0.397693
CDK2;CDK2deltaT	P24941;O75100;B	0.00013	0.97649	1.3248		1	0.405775
AAG;ANPG;MID1;MPG;	P29372-2;P29372	1.90E-87	0.99798	1.3259	74.13	37	0.406972
DXS1283E;GS2;PNPLA4	P41247;C9IZF6;B4	0.000147	1.0277	1.3264	2.1196	2	0.407516
CBP1;CBP2;HSP47;PIG1	P50454;A8K259;C	0.00337	0.93081	1.34		1	0.422233
GHITM;DERP2;My021;T	B4DPG9;Q6FIA7;C	0.10256	1.0356	1.3463		1	0.429

FKSG11;ZFP91	Q96JP5-1;Q96JP5	3.80E-07	0.87031	1.3601	32.514	2	0.443713
PHB;hCG_29613;ZNF6(	P35232;A8K401;C	0.000496	1.0753	1.3953		1	0.480575
CCDC124	Q96CT7	4.86E-12	0.96295	1.4361	17.244	7	0.522156
H3.3A;H3.3B;H3F3;H3F	P84243;B2R4P9;B	8.01E-06	1.0843	1.4921	20.471	5	0.577344
DAAP-21F2.2-001;DAM	B0V043;P26640	1.56E-06	1.1979	1.5348	12.906	4	0.618051
C4orf27	Q9NWY4	2.86E-16	1.226	1.6361	142.54	11	0.710261
BPY2IP1;C19orf5;MAP1	Q66K74;A8K940;E	0.000664	0.98408	1.6426		1	0.715981
RPS7;hCG_1784266	P62081;Q57Z92;B	1.67E-05	1.2999	1.8254	84.113	2	0.868213
ACTR2;ARP2	B4DHK9;P61160;C	0.016945	1.3077	1.8801		1	0.910809
JUND	P17535	6.62E-15	1.5638	1.9391	108.23	4	0.955387
RIF1	Q5UIP0-1;Q5UIP0	0.006157	1.4995	2.1479		1	1.102927
NUP62;pp7052	P37198;Q8WYU3	0.004906	1.8031	2.58		1	1.367371
JUNB;hCG_172481	P17275;Q53HP8;C	4.29E-18	2.1192	2.6232	84.736	3	1.391328
DIMT1;DIMT1L;HUSSY-	Q9UNQ2;A8K9K8;	0.002424	2.0672	3.0832		1	1.624428
CREB1;hCG_15208	P16220-1;P16220	4.17E-14	2.7056	3.3213	19.383	3	1.731748
HMGA1;HMGIY	B4DWA0;C9J292;I	0.005836	2.7178	3.9783		1	1.992152
MN1	Q10571;A5HML1	0.12914	3.5346	4.441	289.97	2	2.150885
VDAC2;RP11-375G3.1-(	P45880-1;P45880	0.025034	4.2551	6.0233		1	2.590554

APPENDIX 2. Chapter 3: Proteins identified by LC-MS (reverse; hmC=heavy; C=light)

Gene Names	Uniprot	PEP	Ratio H/L	Ratio H/L	Ratio H/L	Ratio H/	Log2(H/L)
KIAA1187;MAP7I	Q3KQU3-1;Q3	0.005067	16.103	22.347	95.012	2	4.482
YLPM1;FLJ00353;	B4DMQ9;Q86	1.21E-07	5.1737	5.0139	30.67	3	2.3259
BCLAF1;BTF;KIAA	Q9NYF8-1;Q9	0.024252	5.0929	4.3068	40.879	2	2.1066
L18;Nbla10363;T	Q9NXF1;B4DC	2.34E-05	3.0351	3.4481	23.068	3	1.7858
KIAA0138;SAFB2	Q14151;A0PJ4	0.003036	3.4403	3.0293		1	1.599
TOP2B;topIIb;TOI	Q02880-1;Q0	1.60E-12	2.9181	2.8733	39.254	2	1.5227
CPSF100;CPSF2;K	Q9P2I0;B3KN4	6.35E-05	2.6873	2.8419		1	1.5069
AHCTF1;ELYS;MS'	Q96M44;Q8W	4.33E-08	3.6765	2.8287		1	1.5001
EC45;RPL15;TCBA	P61313;Q5U0	0.000196	3.5637	2.7977		1	1.4842
DXS648E;QM;RPI	P27635;Q5HY	0.26025	2.8193	2.7556		1	1.4624
OTT;OTT1;RBM15	Q96T37-1;Q9	0.001004	3.3142	2.7486	55.228	2	1.4587
RBP56;TAF15;TAI	Q92804-1;Q9	3.29E-35	4.2848	2.6415		1	1.4014
DDX42	Q86XP3-1;Q8	9.65E-07	2.8963	2.6245		1	1.392
ZNF326	Q5BKZ1-1;Q5	3.03E-06	2.7956	2.6177	37.191	3	1.3883
MET;SLTM	Q9NWH9;A8K	0.14794	3.1228	2.6002		1	1.3786
CLN80;COIL	P38432	6.18E-07	2.4354	2.5992	25.358	3	1.3781
HMP;IMMT;PIG4	Q16891-1;Q1	0.022615	2.3239	2.5898	51.17	2	1.3728
ACIN1;ACINUS;KI	Q9UKV3-1;Q9	5.52E-44	3.3541	2.5737	36.863	6	1.3638
EDMD;EMD;STA;I	P50402;Q6FIC	6.91E-08	3.2088	2.5342	6.3089	2	1.3415
THRAP3;TRAP15(	Q9Y2W1;Q05	2.14E-05	2.6093	2.484	33.808	2	1.3127
CPSF1;CPSF160;h	Q10570;B4DE	7.08E-08	2.6146	2.4789		1	1.3097
RPL18	Q07020;Q0QI	2.30E-21	2.5148	2.3512	6.7436	4	1.2334
HMX3;MNAR;PEL	Q8IZL8-2;Q8I	4.16E-07	2.1986	2.3059	12.623	4	1.2053
FUSIP1;FUSIP2;SF	O75494-1;O7	2.16E-18	2.6556	2.3047	9.5842	5	1.2046
NUMA1 variant p	Q4LE64;Q149	6.08E-65	2.9051	2.3027	16.086	8	1.2033
TFAP2BL1;TFAP2I	Q7Z6R9	0.004079	2.4548	2.2633		1	1.1784
VDAC2;RP11-375	P45880-1;P45	0.014099	2.8908	2.2433		1	1.1656
PRPC8;PRPF8	Q6P2Q9;B4DI	2.30E-17	3.0826	2.2168	21.213	9	1.1485
ADE2;AIRC;PAICS	P22234;Q68C	1.56E-15	2.44	2.2138	1.1524	2	1.1465
TRA2A;hCG_3797	Q13595-1;Q1	0.001749	2.4739	2.2015		1	1.1385
KIAA0365;SFRS14	A8K5G0;Q8IX	0.016807	2.2552	2.1466		1	1.1021
HSPC075;KIAA03	Q9UQ35-1;Q9	0.002231	2.498	2.1354		1	1.0945
HSPC108;SLP2;ST	Q9UJZ1;B4E1I	0.021587	2.0046	2.1276		1	1.0892
DDX6;HLR2;RCK	P26196;B2R8	0.004949	2.4675	2.1181		1	1.0828
KIAA0690;RRP12	Q5JTH9-1;Q5J	1.98E-08	2.2805	2.1031		1	1.0725
ADTAA;AP2A1;CL	O95782-1;O9	0.005984	2.2419	2.0987	1.2092	2	1.0695
KIAA0169;NUP18	Q5SRE5-1;Q5	0.001056	2.0086	2.0874		1	1.0617
HNRPG;RBMX;RB	P38159;B4E3	3.71E-20	2.4543	2.0681	57.462	18	1.0483
DXS8237E;GPATC	P98175-1;P98	1.40E-05	3.1403	2.0527	25.434	3	1.0375
RNPU1Z;RPU1;SN	P08621-1;P08	1.29E-15	2.3944	2.0519	13.374	7	1.037
DDX9;DHX9;LKP;I	Q08211;B3KU	8.68E-147	2.9925	2.0429	81.264	41	1.0306
COII;COXII;MTCO	P00403;A0S0	1.99E-05	2.5888	2.0351		1	1.0251
PLEC1;hCG_1994	Q15149-1;Q1	0.004298	1.7844	2.0309		1	1.0221
OCSP;TSPAN10	Q9H1Z9;Q6PJ	0.003766	2.8974	2.0235	11.164	2	1.0169
SNRPD1	P62314;Q7Z5	4.54E-09	1.6183	2.0203	131.03	3	1.0146
HNRNPM;HNRPM	P52272-1;P52	2.05E-121	2.741	2.0114	55.273	75	1.0082
DPM1;RP5-914P2	Q5QPK0;Q5Q	0.019022	2.4447	2.0082		1	1.0059
HPR1;THOC1	Q96FV9-1;Q9	2.16E-29	2.2217	2.0012	35.225	8	1.0009

PGDH3;PHGDH;R	O43175;B3KS	0.009453	2.0143	1.9994	31.34	4	0.9996
RPS9;hCG_20091	P46781;A9C4(	2.08E-05	2.0912	1.9956	28.729	4	0.9968
RPL27A;L27a	P46776;Q6NZ	3.15E-06	2.6582	1.9822	20.303	3	0.9871
DDX39;DKFZp547	B4DX78;O001	5.61E-26	2.1766	1.9696		1	0.9779
KIAA0663;ZC3H1	O75152;B4DL	3.11E-19	2.554	1.9681	7.9014	7	0.9768
XRCC1	P18887;B2RC'	0.005155	2.0582	1.9553		1	0.9674
CGGBP;CGGBP1	Q9UFW8;C9JU	0.007001	2.3836	1.9349		1	0.9523
RPL8	P62917;B4DV	1.65E-05	2.443	1.9303	65.765	3	0.9488
COD;SNRPB;SNRI	P14678-3;P14	0.000203	2.2389	1.9011	7.9344	7	0.9268
MIB2;SKD;ZZANK	Q96AX9-9;Q9	0.006509	2.1871	1.8857	39.563	3	0.9151
C22orf19;KIAA09	Q13769;C9JX(	0.001686	2.0989	1.8748	2.6464	2	0.9067
CPSF7	Q8N684-3;Q8	2.46E-13	2.4365	1.8709	10.151	3	0.9037
ASF;OK/SW-cl.3;S	Q07955-1;Q0'	5.27E-14	2.0715	1.8709	23.011	9	0.9037
KPNB1;NTF97	Q14974;B2RB	6.40E-12	2.1626	1.861	51.364	9	0.8961
ARS2;ASR2;SRRT	Q9BXP5-1;Q9	0.000119	2.2699	1.8588		1	0.8944
RPL19	P84098;Q53G	2.08E-36	2.3923	1.8538	12.724	3	0.8905
WDR36;DKFZp68	D6RAY5;Q8NI	0.00155	1.7963	1.853	23.72	3	0.8899
ESA1;FLOT2;M17	Q14254;Q6FG	2.21E-22	2.1963	1.8497	14.757	3	0.8873
DDX8;DHX8	Q14562;A8K6	0.40917	1.908	1.8413		1	0.8807
HNRNPF;HNRPF;(	P52597;B4DK	1.46E-97	2.3718	1.811	92.803	6	0.8568
TUBG;TUBG1;TUI	P23258;Q9NR	0.00311	1.8531	1.791		1	0.8408
KIAA0906;NUP21	Q8TEM1-1;Q8	1.03E-35	2.2599	1.79	47.416	10	0.84
PSF;SFPQ	P23246-1;P23	7.84E-170	2.5104	1.7871	112.45	116	0.8376
C16orf14;FAM19	Q9BUT9-1;Q9	0.00852	1.7775	1.787		1	0.8375
RPL6;TXREB1	B2R4K7;Q9HE	3.55E-09	1.9854	1.7846	18.906	8	0.8356
D21S2056E;NNP1	P56182;Q53F	9.88E-05	1.8708	1.7773	9.8643	2	0.8297
MYBBP1A;P160	Q9BQG0-2;Q9	0.001988	2.1051	1.7768	27.423	3	0.8293
SNRPE;SNRPEL1	P62304;Q5VY	2.81E-10	2.2616	1.7694	12.863	6	0.8233
MOSP;PLIP;PNAS	Q8WUK0-1;Q	0.000169	2.0657	1.7689		1	0.8229
DDX5;G17P1;HEL	P17844;B4DN	3.58E-49	2.5056	1.7615	22.431	10	0.8168
AQR;KIAA0560	O60306;A8K6	0.018615	1.8473	1.7584		1	0.8143
KIAA1185;LRRC47	Q8N1G4	1.68E-16	1.9279	1.7542	56.949	2	0.8108
DDX17;RP3-434P	Q59F66;Q928	1.36E-105	2.5744	1.754	84.063	42	0.8106
CHCHD3;hCG_20	C9JRZ6;Q9NX	0.003098	1.9305	1.7533		1	0.8101
MATR3;KIAA0723	A8MXP9;P432	1.44E-117	2.4205	1.7499	88.471	28	0.8073
TUBB2;TUBB2A;T	Q13885;A5D9	4.48E-21	2.1804	1.7433	42.386	11	0.8018
ACAT;ACAT1;MA'	P24752;Q96F(	0.000802	1.4699	1.7313		1	0.7919
FTP3;HNRNPH2;H	P55795;B4DF	2.21E-116	2.0971	1.7292	24.142	9	0.7901
RBM14;SIP	Q96PK6-1;Q9	4.94E-37	2.2599	1.7184	19.648	21	0.7811
NPM;NPM1	P06748-1;P06	5.67E-123	2.134	1.7157	104.89	28	0.7788
CLH17;CLTC;CLTC	Q00610-1;Q0(	7.52E-15	2.123	1.7147	79.603	6	0.778
PSEC0006;THOC6	Q86W42-1;Q8	3.21E-48	2.1631	1.7137	111.68	8	0.7771
TPR;Tpr;tpr	P12270;A0PJC	0.001806	1.7546	1.7117		1	0.7754
ASCC3L1;HELIC2;	O75643-1;O75	5.02E-14	1.8746	1.7099	31.759	7	0.7739
KIAA0105;WTAP;	Q15007-1;Q1	4.03E-30	2.22	1.7088	103.87	3	0.773
hCG_2031827;PA	Q4VC03;Q53C	1.02E-30	2.0756	1.6992		1	0.7649
C3orf26;hCG_202	Q9BQ75;B3KV	0.008748	1.9236	1.6992		1	0.7649
SFRS7	Q16629-1;Q1(	1.92E-12	1.8878	1.6891	32.111	4	0.7563
EWSR1;hCG_201(	Q96MX4;Q01	8.43E-75	2.7168	1.678	111.49	11	0.7467
SPK;SYMPK	Q92797-1;Q9	9.79E-07	1.891	1.6766		1	0.7455

SRPR	P08240;B4E0I	1.70E-05	1.8378	1.6751	30.582	2	0.7442
CYFIP1;KIAA0068	B3KWV6;Q7L5	0.000122	1.5885	1.6673	6.1114	3	0.7375
NPWBP;SIPP1;SN	Q9Y2W2;B4D	0.05776	1.6757	1.6584		1	0.7298
TUFM	P49411	7.90E-11	2.0493	1.6436	2.6195	2	0.7169
DDX46;KIAA0801	Q7L014;A8K6	0.000161	1.5349	1.6381	34.222	3	0.712
NONO;NRB54	Q15233;A8K5	1.16E-258	2.5096	1.6368	110.05	166	0.7109
OK/SW-cl.32;RPL	P39023;Q96C	3.09E-09	1.9243	1.6358	98.548	5	0.71
DDX2A;EIF4A;EIF	P60842;A8K0	2.75E-18	2.077	1.6334	29.357	6	0.7079
FLOT1;DADB-118	O75955;Q53H	1.15E-70	2.0741	1.6328	95.925	8	0.7073
CSTF1;hCG_3909	Q05048;B4DD	8.04E-15	1.857	1.6316	43.42	4	0.7063
RBM4;RBM4A;RB	Q9BWF3-1;Q9	1.08E-62	2.0421	1.6274	16.093	3	0.7026
SNRPA1	P09661;Q53G	2.89E-11	2.0909	1.6204	1.3957	2	0.6963
GHITM;DERP2;M	B4DPG9;Q6FI	0.011299	2.0623	1.6197		1	0.6957
MDU1;SLC3A2	P08195-4;P08	5.47E-24	1.8635	1.6048	74.766	8	0.6824
RPS16	B4DP32;P622	2.90E-09	2.0324	1.6011	103.89	4	0.6791
ZNF207;DKFZp76	E1P660;Q59G	2.61E-05	2.0754	1.5974	31.37	3	0.6757
NMP200;PRP19;F	Q9UMS4	1.57E-12	2.1149	1.5906	17.693	5	0.6696
KIAA0791;NUP15	O75694-1;O75	1.63E-06	1.7551	1.5899	40.916	3	0.6689
GCN1L1;KIAA021	Q92616;B4DM	0.000111	1.6082	1.5896	22.817	4	0.6687
KIAA0095;NUP93	Q8N1F7;A8K8	8.13E-09	2.0237	1.5877	50.021	2	0.6669
DIMT1;DIMT1L;H	Q9UNQ2;A8K	0.008216	1.8833	1.5653		1	0.6464
DDX48;EIF4A3;KI	P38919	6.57E-18	1.994	1.5646	117.92	5	0.6458
PAB2;PABP2;PAB	Q86U42-1;Q8	3.88E-06	1.8114	1.558	24.451	2	0.6397
LBR	Q14739	7.24E-28	1.825	1.5571	29.692	6	0.6389
OK/SW-cl.48;PHC	Q00325-1;Q0	1.21E-25	1.9453	1.554	32.606	8	0.636
DNCI2;DNCIC2;D	Q13409-1;Q1	4.21E-12	1.842	1.5489		1	0.6312
GNAI3;hCG_1997	P08754;Q5TZ	2.24E-05	1.9237	1.5386		1	0.6216
SFRS9;SRP30C	Q13242;A8K3	2.80E-08	1.9758	1.5294	9.5941	2	0.613
RPN1	P04843;Q53E	4.45E-21	1.6892	1.5281	25.3	6	0.6117
NP220;ZFML;ZNF	Q14966-1;Q14	5.37E-06	1.9673	1.5179	57.412	2	0.6021
CAS;CSE1L;XPO2;	P55060-1;P55	0.002188	1.6525	1.5123	176.82	4	0.5967
DDX47;hCG_276	Q9H0S4;Q53C	7.83E-16	1.3476	1.5063	14.714	9	0.591
VIM	P08670;B0YJC	4.57E-105	1.8797	1.5048	87.327	26	0.5896
RPL7;WUGSC:H_I	P18124;O950	1.30E-16	1.7368	1.489	112.58	6	0.5743
RP3-375P9.1-002	Q53G33;Q5TA	1.32E-11	1.9398	1.4886		1	0.574
DHC1;DNCH1;DN	Q14204;B4DS	6.90E-23	2.0465	1.487	20.512	5	0.5724
XRN2	Q9H0D6-1;Q9	3.42E-07	1.5978	1.4867	10.692	5	0.5721
EFTUD2;KIAA003	Q15029;A8KA	1.13E-19	1.8424	1.4848	62.925	8	0.5703
CD2BP2;KIAA117	O95400;Q5QT	0.006066	1.7798	1.482		1	0.5675
HNRNPH1;HNRPH	P31943;B4DP	1.85E-182	2.1446	1.4729	95.173	46	0.5587
E1BAP5;HNRNPU	Q9BUJ2-1;Q9	8.83E-51	1.8814	1.4721	95.514	13	0.5579
FUS;TLS	P35637-1;P35	2.63E-44	1.8649	1.4705	99.61	17	0.5563
DBX;DDX3;DDX3)	O00571;B4E3	4.52E-40	1.7795	1.4484	111.11	18	0.5345
RFC2;hCG_18034	P35250-1;P35	0.000599	1.6179	1.4477	17.699	2	0.5338
NEF3;NEFM;NFM	P07197;B4DG	0.014178	1.9324	1.4465		1	0.5326
MSH2	P43246;B4DL	0.001101	1.3117	1.4435	4.7158	2	0.5296
KID;KIF22;KNSL4;	Q14807;B7Z2	0.000677	1.7699	1.4421		1	0.5282
HNRNPUL2;HNRF	Q1KMD3	6.62E-14	1.5466	1.4338	49.874	3	0.5198
PAB1;PABP1;PAB	P11940-1;P11	1.66E-30	1.8059	1.4314	40.283	3	0.5174
CAIN;CAN;KIAA0(	P35658-5;P35	1.27E-06	1.6152	1.4312	68.791	3	0.5172

RPN2;hCG_3714:	P04844;B2RE4	5.10E-07	1.4623	1.4292	22.593	2	0.5152
PBSCF;SNRPF	P62306	0.002235	1.6507	1.4282		1	0.5142
HNRNPH3;HNRPH	P31942-1;P31	9.13E-29	1.8634	1.4251	18.213	12	0.5111
SCML2;RP5-1129	Q9UQR0-1;Q9	0.00445	1.6925	1.413		1	0.4988
SAP155;SF3B1;DI	O75533;A0JLT	5.01E-13	2.1988	1.4129	73.047	9	0.4987
GCS1;MOGS	Q13724;Q58F	1.29E-05	1.3395	1.3993		1	0.4847
PPP1CB	P62140;B4DJ7	7.83E-23	1.7782	1.3956		1	0.4809
PNAS-10;PNAS-1:	Q9UM00-1;Q9	1.85E-08	1.7475	1.3943		1	0.4795
LMN2;LMNB;LMI	P20700;B4DZ7	4.07E-28	1.8999	1.3931	73.648	13	0.4783
HNRNPC;HNRPC;	P07910-1;P07	8.97E-46	1.7235	1.3918	108.44	17	0.477
SFRS4;SRP75;hCG	Q08170;A8K6	0.00299	1.7655	1.3906	6.8785	3	0.4757
ATP5A;ATP5A1;A	P25705;A8K0	3.22E-20	1.6426	1.3879	21.678	5	0.4729
U2AF1;U2AF35;U	Q01081;B5BU	4.57E-35	1.5846	1.3775	24.523	6	0.4621
HADH;HADHA	P40939;B4DD	0.000669	1.5632	1.375		1	0.4594
LAP2;TMPO	P42167-1;P42	1.30E-163	1.8283	1.3748	66.687	36	0.4592
SSR4;TRAPD	P51571;A6NL	1.53E-09	1.4751	1.3705	16.733	4	0.4547
PRO2640;RPS14	P62263;E5RH	0.000283	1.762	1.3684		1	0.4525
BST2	B4E0Z6;Q105	0.000223	1.7413	1.3616	6.471	2	0.4453
BBC1;OK/SW-cl.4	P26373;A8K4	9.08E-16	1.7104	1.3508	69.138	5	0.4338
NDUFB10;hCG_4	O96000;A8K7	0.009632	1.6489	1.3471		1	0.4299
HSPA1;HSPA1A;H	P08107;A8K5	1.53E-27	1.7395	1.3371	2.452	4	0.4191
ADAR;ADAR1;DSI	P55265-4;P55	2.42E-41	2.1523	1.3361		1	0.418
BAM;BMH;CSPG6	Q9UQE7;B0A7	0.043309	1.4384	1.3349	12.142	2	0.4167
KIAA0052;SKIV2L	P42285;A8K6	3.67E-05	1.3961	1.3306	47.751	3	0.4121
BCAP31;hCG_20(	B3KQ79;C9J0	0.14934	1.5605	1.3301		1	0.4115
CYK18;KRT18;PIG	P05783;B2RA	4.95E-86	1.6096	1.3276	50.843	38	0.4088
CCG2;RPS4;RPS4)	P62701;B2R4	0.000106	1.7119	1.3214	51.277	4	0.4021
CD44;LHR;MDU2	P16070-1;P16	3.26E-14	1.6842	1.3202	14.33	3	0.4008
IGF2BP3;IMP3;KC	O00425-1;O0	5.53E-21	1.5343	1.3167	131.29	8	0.3969
KIF4B;KIF4;KIF4A	Q2VIQ3;B4DY	8.74E-06	1.3168	1.3158		1	0.3959
JM26;NPW38;PQ	O60828-1;O6	0.012481	1.5014	1.3049	42.577	2	0.3839
SMARCA5;SNF2H	O60264;B4DZ	2.99E-09	1.5516	1.3024	43.82	7	0.3812
CXorf3;THOC2;RI	Q8NI27-1;Q8I	3.10E-25	1.8292	1.2994	127.79	5	0.3778
ALDH10;ALDH3A:	P51648-2;P51	0.020632	1.4499	1.2922		1	0.3698
SPTAN1;SPTA2;D	A6NG51;Q13	2.52E-07	1.2884	1.2796	2.8414	2	0.3557
NUP107	P57740;B3KM	0.0037	1.4215	1.2763		1	0.352
ADAR2;NUP98	P52948-1;P52	3.06E-22	1.6034	1.2762	68.2	3	0.3519
BAT1;UAP56;XXb	Q13838-2;Q1	2.30E-39	1.8752	1.2752	65.671	20	0.3507
D6S218E;RPS18;I	P62269;Q5SU	2.38E-05	1.3552	1.2674	29.101	4	0.3419
DBN1;D0S117E;h	A8MV58;Q16	9.23E-06	1.5626	1.2658		1	0.34
HSP70B';HSPA6;H	P17066;B2R6	8.11E-30	1.6735	1.2656		1	0.3398
LIG3	P49916-1;P49	2.11E-08	1.2747	1.2636	53.995	3	0.3375
DRBF;ILF3;MPHO	Q12906-4;Q1	9.66E-61	1.7135	1.2549	86.728	22	0.3276
IARS;DKFZp686L1	P41252;B4DV	0.006399	1.3469	1.2529		1	0.3253
HNRNPA3;HNRPA	P51991-1;P51	7.90E-53	1.7326	1.2405	35.122	12	0.3109
TMEM109	Q9BVC6;B2R9	1.80E-11	1.642	1.233		1	0.3022
C7orf14;KIAA022	Q92621;B4DE	6.15E-06	1.2086	1.2314		1	0.3003
SFRS3;SRP20;hCG	P84103;B2R6	5.14E-17	1.582	1.2271	23.899	10	0.2953
LRRC59;PRO1855	Q96AG4;B2RE	5.74E-11	1.5766	1.2238	129.19	6	0.2914
BZW2;HSPC028;M	Q9Y6E2;B3KM	0.005815	1.3288	1.2236		1	0.2911

CBX8;PC3;RC1	Q9HC52;C9JM	0.003346	1.4639	1.2184	127.07	2	0.285
HNRNPK;HNRPK;I	P61978-2;P61	9.11E-273	1.5037	1.2071	117.96	82	0.2715
MRNP41;RAE1;h(	P78406;A8K8(	0.050575	1.479	1.2065		1	0.2708
KHDRBS1;SAM68	Q07666-1;Q0	1.26E-14	1.6143	1.2031	74.997	16	0.2668
KIAA0026;MORF4	Q15014;B3KV	0.006176	1.2321	1.1999		1	0.2629
KIAA0886;My043	Q9NQC3-1;Q9	0.007504	1.4804	1.1991		1	0.262
KIAA0656;SNAP9	O60641-1;O6(	0.012683	1.274	1.1959		1	0.2581
AFG3L2;hCG_382	Q9Y4W6;D3D	0.008159	1.3403	1.1932		1	0.2548
CAPZA2;hCG_378	P47755;A4D0	0.000515	1.5132	1.1929		1	0.2545
ANT2;SLC25A5;AI	P05141;B2RC	2.62E-20	1.4665	1.1918	98.726	11	0.2531
HCC1;HSPC316;S/	P82979;Q567	6.31E-11	1.474	1.1907	25.472	6	0.2518
VASP	P50552	9.89E-09	1.3008	1.1882	18.731	2	0.2488
PSEC0230;SRPRB	Q9Y5M8;Q54	4.16E-31	1.5373	1.1843	1.0967	2	0.244
PPP1CC;PPP1A;PI	P36873-2;P36	2.76E-33	1.4564	1.1824	85.639	4	0.2417
LAP2;TMPO	P42166	1.14E-127	1.4268	1.1778	19.957	3	0.2361
FKSG11;ZFP91	Q96JP5-1;Q96	0.001559	1.1947	1.1735	32.364	2	0.2308
KIAA0017;SAP13(	Q15393-1;Q1	8.04E-23	1.3965	1.1618	21.25	4	0.2164
CAND1;KIAA0829	Q86VP6-1;Q8	2.43E-09	1.1163	1.1538	45.859	4	0.2064
OK/SW-cl.26;RPS	P23396;Q53G	6.24E-36	1.5027	1.1529	61.225	7	0.2053
HSP90AB1;HSP9(	P08238;A8K3'	1.60E-14	1.3354	1.1508	91.013	6	0.2026
CUTL1;CUX1;Nbla	P39880-3;P39	8.59E-06	1.2275	1.1417	7.2062	2	0.1912
ACTN1;ACTN4;hC	B7TY16;A1L0\	2.84E-05	1.0655	1.1406	11.095	3	0.1898
HSC70;HSP73;HSI	P11142-1;P11	3.35E-69	1.4101	1.138	87.574	10	0.1865
ERAB;HADH2;HSI	Q99714-1;Q9'	5.26E-08	1.3928	1.1342		1	0.1817
ABCG2;ABCP;BCF	Q9UNQ0-1;Q9	2.18E-05	1.1761	1.1297		1	0.1759
EEF2;EF2	P13639;B4DM	5.53E-05	1.4082	1.1279	34.965	3	0.1736
TARDBP;TDP43;R	Q13148-2;Q1	3.76E-51	1.7138	1.1231	96.85	6	0.1675
BAP;PHB2;REA	Q99623;Q9BX	3.80E-106	1.4306	1.1217	77.711	12	0.1657
SAP145;SF3B2;DI	Q13435;Q7Z3	8.43E-06	1.7158	1.1202		1	0.1638
LAMBR;LAMR1;R	A6NE09;P088	9.56E-06	1.3649	1.12	14.415	2	0.1635
CGI-04;YARS2	Q9Y2Z4	0.000424	1.0583	1.1173	52.007	2	0.16
ANX2;ANX2L4;AN	P07355-2;P07	1.23E-56	1.3937	1.1133	55.982	14	0.1548
C22orf28;HSPC11	Q9Y3I0;B4DN	0.000632	1.4346	1.1131	101.78	3	0.1546
TOR1AIP1;RP11-5	A8MUM5;Q5.	4.20E-09	1.464	1.1113	10.508	3	0.1522
RBAP46;RBBP7;R	Q16576;Q5JN	4.37E-09	1.3947	1.1064	145.39	7	0.1459
CDC2L4;CDK9	P50750-2;P50	0.004321	1.0842	1.105		1	0.144
GRP75;HSPA9;HS	P38646;B7Z1\	2.01E-21	1.3158	1.1021	7.7323	4	0.1403
IAG2;MAGT1;PSE	B4DH58;Q9H(	0.020553	1.4037	1.1015		1	0.1395
PHB;hCG_29613	P35232;A8K4(	7.55E-81	1.3098	1.0994	92.099	14	0.1367
KIAA0648;PDS5;F	Q29RF7-1;Q2	1.26E-13	1.2454	1.0988	7.0413	3	0.1359
ADNP;ADNP1;KIA	Q9H2P0;B2RE	0.002539	1.2563	1.0894		1	0.1235
DAAP-21F2.2-001	B0V043;P266	2.93E-17	1.2447	1.0864	26.148	9	0.1196
PCBP1;hCG_1776	Q15365;Q53S	2.60E-118	1.3237	1.0845	131.96	11	0.117
CREAP1;CROP;LU	A8K3C5;O952	0.000395	1.307	1.0835		1	0.1157
KIAA0650;SMCHI	A6NHR9-1;A6	2.81E-09	1.0319	1.0819	28.659	4	0.1136
BAG2	O95816;B3KM	0.002906	1.2136	1.0797		1	0.1106
LMN1;LMNA;RP1	P02545-1;P02	8.99E-111	1.5147	1.0642	60.064	35	0.0898
CRDBP;IGF2BP1;\	Q9NZI8;D3DT	1.32E-32	1.3366	1.0632	124.32	3	0.0884
C9orf10;FAM120	Q9NZB2-6;Q9	5.95E-05	1.2388	1.0612		1	0.0857
PLCB3	Q01970;Q8N1	6.68E-05	1.0211	1.0598	4.9266	2	0.0838

HYRC;HYRC1;PRK	P78527-1;P78	5.48E-138	1.6707	1.0584	81.894	31	0.0819
POLD2;hCG_1745	P49005;A4D2	2.23E-05	1.2229	1.0565		1	0.0793
RAB39;RAB39A;R	Q14964;P203	0.005103	1.3622	1.0544		1	0.0764
INO80H;NMP238	Q9Y265-1;Q9'	4.93E-19	1.349	1.0537	90.695	7	0.0755
CDABP0035;KIAA	P37802;C9J5V	0.04266	1.1082	1.0513		1	0.0722
BAF47;INI1;SMAF	Q9H836;Q128	0.000492	1.3334	1.0496		1	0.0698
SNRPA	P09012;B2R8(	4.50E-13	1.1963	1.0496	53.413	6	0.0698
PP446;TMEM214	Q6NUQ4;B3LI	0.007652	1.1219	1.0462	2.881	2	0.0652
BSG;UNQ6505/PI	P35613-1;P35	0.22917	1.2797	1.0327		1	0.0464
API5;MIG8	Q9BZZ5-4;Q9I	1.39E-20	1.103	1.0265	5.4472	3	0.0377
ELAVL1;HUR	B4DVB8;Q157	2.02E-79	1.3061	1.0248	83.214	31	0.0353
C2F;EMG1;hCG_2	Q92979;A8K6	4.46E-08	1.3495	1.0239	19.938	3	0.0341
NXF1;TAP;DKFZp(	Q9UBU9;Q68(	0.014734	1.0364	1.022	25.024	2	0.0314
FUBP2;KHSRP	Q92945-1;Q9	7.52E-53	1.4381	1.0122	57.174	15	0.0175
KIAA1978;RAVER	Q8IY67-2;Q8I'	1.07E-05	1.2402	1.0122		1	0.0175
YAF2	C9JKP3;Q8IY5	0.00858	1.1578	1.0112	8.0244	2	0.0161
CBP1;CBP2;HSP4	P50454;A8K2	8.39E-08	1.2517	1.007		1	0.0101
DDX21;OK/SW-cl	Q9NR30-1;Q9	1.22E-47	1.5092	1.0035	110.56	27	0.005
MMAB	Q96EY8;B2R6	0.013877	1.1833	1.0034		1	0.0049
RAB5C;RABL	P51148	0.000121	1.2258	0.98917	228.33	3	-0.016
IQGAP1;KIAA005	P46940;A4QP	3.52E-23	1.3493	0.98366	62.359	7	-0.024
EIF2A;EIF2S1;hCG	P05198;Q53X	9.72E-09	1.0191	0.96381	15.243	3	-0.053
C1orf77;HT031;P	Q9Y3Y2-3;Q9'	1.27E-26	1.0754	0.95779	17.747	2	-0.062
C20orf20;MRGBP	Q9NV56;A8C4	4.42E-11	1.1775	0.95229		1	-0.071
KIAA0035;NOLC1	Q14978-2;Q1	1.35E-05	1.0122	0.94873	8.2317	4	-0.076
ARL6IP5;DERP11;	O75915;B3KC	1.23E-11	1.1645	0.94395	16.202	2	-0.083
EEF1D;EF1D;hCG_	P29692-2;P29	5.34E-31	1.2035	0.94388	96.779	13	-0.083
SMU1;hCG_3027	Q2TAY7;A0M(	0.060577	1.1395	0.93291		1	-0.1
FLN;FLN1;FLNA;F	P21333-1;P21	5.12E-201	1.4768	0.9294	61.151	36	-0.106
LAS1L;MSTP060	Q9Y4W2-1;Q9	1.08E-05	1.0906	0.92868		1	-0.107
RPL26;RPL26L1;R	A6NE05;Q9UI	0.001567	1.0578	0.92505	61.186	4	-0.112
ALY;BEF;THOC4	Q86V81	2.60E-41	1.2049	0.92329	24.071	14	-0.115
KAP1;RNF96;TIF1	Q13263-1;Q1	4.92E-05	0.99244	0.92169	40.798	6	-0.118
UBF;UBF1;UBTF;I	P17480-1;P17	1.99E-06	1.169	0.92078	31.472	3	-0.119
H1F5;HIST1H1B	P16401	1.04E-06	0.80227	0.9142	138.12	4	-0.129
HNRPQ;NSAP1;S\	O60506-3;O6(	8.97E-92	1.3108	0.91168	116.86	36	-0.133
FTE1;MFTL;RPS3/	P61247;A8K4'	0.015874	1.1417	0.9041		1	-0.145
SEC22B;SEC22L1	O75396	1.78E-57	1.1598	0.90315	100.26	5	-0.147
SRP14	P37108	1.33E-26	1.012	0.90287	17.175	6	-0.147
CTA-216E10.8-00	B1AHD1;P557	0.000344	1.1156	0.89718		1	-0.157
TIA1;hCG_40580	P31483-1;P31	3.61E-20	1.0716	0.89155	46.903	3	-0.166
DBP1;DDX15;DH)	O43143;B4E0	2.05E-16	0.82472	0.88723	151.45	8	-0.173
GATAD2B;KIAA11	Q8WXI9;B3KS	1.09E-24	1.2058	0.8777	80.84	6	-0.188
NSEP1;YB1;YBX1	P67809;A0JLL	1.81E-233	0.83556	0.87659	145.97	25	-0.19
ACTL6A;BAF53;B/	O96019-1;O9(	0.003558	1.137	0.87601		1	-0.191
KIAA0170;MDC1;	Q14676-1;Q1	6.10E-36	1.1462	0.87557	60.448	6	-0.192
CGI-21;HSPC332;	Q53GS9;B7Z7	5.09E-20	1.0744	0.8738	63.979	7	-0.195
HSPC099;KIAA04	Q5VT52-1;Q5	0.002206	0.92095	0.86523	86.272	2	-0.209
JUP	B4DE59;C9JK'	1.01E-18	1.0505	0.86156		1	-0.215
MCT1;MCTS1	B4DGY2;Q9UI	0.00137	0.88895	0.86004	12.177	2	-0.218

HCF1;HCFC1;HFC	P51610-1;P51	4.97E-08	1.0959	0.85192	33.69	2	-0.231
HSP60;HSP70L1;H	Q0VDF9;A8K8	1.62E-06	1.0753	0.84947	7.4048	2	-0.235
ZFR	Q96KR1;A8K8	2.82E-07	0.85344	0.84814	125.05	2	-0.238
SFRS2;hCG_2784	Q01130;B3KU	1.27E-23	0.87492	0.84208	16.868	6	-0.248
SF1;ZFM1;ZNF16	Q15637-5;Q1	2.28E-05	0.7631	0.84165	52.875	2	-0.249
BAZ1B;WBSC10;V	Q9UIG0-1;Q9	0.007858	1.0819	0.84063		1	-0.25
CDC2;DKFZp686L	P06493-1;P06	5.36E-05	1.0403	0.83814	43.026	3	-0.255
RPS7;hCG_17842	P62081;Q57Z	5.19E-11	0.83042	0.83291	108.34	5	-0.264
UHRF1;ICBP90;NI	A8K024;B2RB	1.48E-10	0.8691	0.8315	57.67	9	-0.266
BAF57;SMARCE1	Q969G3-1;Q9	0.000575	1.0695	0.82238		1	-0.282
MTA1L1;MTA2;PI	O94776;Q68D	2.24E-57	1.1221	0.82202	74.002	15	-0.283
HNRNPA2B1;HNF	P22626-1;P22	1.07E-155	1.0891	0.81883	91.267	74	-0.288
DB83;TMEM33;S	P57088;A6QK	0.035671	1.0143	0.81808	17.72	2	-0.29
TIAL1;hCG_4060	A8K4L9;A8K5	1.42E-53	0.97722	0.81681	95.642	14	-0.292
SNRPD3	P62318;B4DJF	1.11E-07	0.856	0.81587	104.21	2	-0.294
KIAA1470;RCC2;T	Q9P258;A5PL	4.91E-27	1.1108	0.81569	82.307	18	-0.294
TCF6;TCF6L2;TFA	Q00059;E5KS	7.05E-20	1.0276	0.81305	41.942	13	-0.299
SMARCA4;hCG_2	B9EGQ8;Q9HI	1.14E-12	1.157	0.80924	13.685	4	-0.305
hCG_2017557;PC	Q59HD4;Q6IP	1.21E-103	1.0032	0.80506	113.59	27	-0.313
POLR2;POLR2A	P24928;B4DE	2.87E-08	1.2281	0.80243		1	-0.318
TIA1	P31483-2;P31	1.66E-36	0.89041	0.79249		1	-0.336
POLD;POLD1	Q308M6;P28	4.37E-05	0.91047	0.79053	52.182	5	-0.339
FBP3;FUBP3;hCG	Q96I24-1;Q96	1.22E-45	0.98549	0.79047	62.396	13	-0.339
EEF1G;EF1G;PRO	B4DTG2;P266	9.55E-37	1.0574	0.78943	99.459	24	-0.341
MSN;hCG_39182	P26038;Q6PJ	1.67E-22	0.87406	0.78809	48.829	8	-0.344
RPS20	B4DW28;P60	0.001059	0.90418	0.7878		1	-0.344
POLR1C;POLR1E;I	O15160-1;O1	0.001602	0.8268	0.78559		1	-0.348
CHD3	Q12873-1;Q1	4.83E-14	0.83662	0.78441		1	-0.35
CORO1C;CRNN4	A7MAP0;A7M	2.58E-11	0.91023	0.78332	189.08	3	-0.352
CHD4	Q14839-2;Q1	4.30E-40	1.1288	0.77179	80.026	15	-0.374
HCC1;RBM39;RN	Q14498-1;Q1	1.88E-54	0.97677	0.77132	90.292	16	-0.375
EIF2G;EIF2S3;EIF	P41091;A8K2	1.61E-06	0.88596	0.76044	193.53	3	-0.395
TREX1;hCG_1997	Q9NSU2-1;Q9	7.79E-05	0.98496	0.75853		1	-0.399
HDGF;HMG1L2;R	B7Z525;P518	1.84E-27	0.67013	0.75538	99.394	10	-0.405
LDHA;PIG19;hCG	B7Z5E3;P003	0.011863	0.93782	0.75305		1	-0.409
TARS;TARSL2	B4DEG8;Q53C	2.51E-06	0.923	0.75192	41.139	2	-0.411
PSP1;PSPC1	Q8WXF1-1;Q8	1.14E-08	1.0421	0.75148	28.423	2	-0.412
DDOST;KIAA0115	P39656;B4DL	3.34E-22	0.94139	0.74203	156.19	4	-0.43
HIP116A;HLTF;RN	Q14527-1;Q1	5.78E-06	0.96405	0.74037	10.243	3	-0.434
FKSG13;PTRF	Q6NZI2-1;Q6N	2.14E-08	0.87654	0.73995	9.8889	2	-0.435
CSTF3;hCG_2731	Q12996;B4DF	7.30E-07	0.7574	0.73738	116.96	2	-0.44
ABC50;ABCF1;DA	Q8NE71-1;Q8	2.80E-42	0.914	0.73121	133.88	10	-0.452
RCC1;CHC1;hCG_	Q16269;Q6NT	7.29E-144	0.90155	0.72128	86.582	26	-0.471
ASE1;CAST;CD3EA	O15446-2;O1	3.19E-13	0.93814	0.72026		1	-0.473
FSCN1;FAN1;HSN	Q96IH1;Q166	0.003835	0.9269	0.71949		1	-0.475
HDAC2;HDAC1;RI	Q92769;B3KU	3.76E-06	0.91989	0.71871	99.755	3	-0.477
RBM3;RNPL	P98179	2.80E-16	0.87801	0.71852	14.762	8	-0.477
CYPB;PPIB	P23284	1.82E-55	0.91546	0.71274	108.41	29	-0.489
GNB2L1;HLC7;PIC	P63244;D6RE	2.23E-24	0.92148	0.71098	57.587	5	-0.492
GTBP;MSH6	P52701-1;P52	7.81E-32	1.0456	0.68474	95.948	13	-0.546

RPS19;hCG_1995	P39019;B0ZBI	4.09E-08	0.86388	0.68252	94.793	4	-0.551
APE;APEX;APEX1;	P27695;Q5TZ	1.47E-159	0.87753	0.68106	72.509	42	-0.554
MAP4;DKFZp547I	P27816-1;P27	9.08E-19	1.0712	0.67236	70.339	6	-0.573
PNKP	Q96T60;Q9BU	0.00301	0.80973	0.66834		1	-0.581
MRRF;RP1-57A13	Q96E11-1;Q9	0.000243	0.79096	0.66519	130.17	3	-0.588
PGAM5;PGAM5-I	Q96HS1-1;Q9	0.001244	0.68963	0.66443	111.83	2	-0.59
CCDC124	Q96CT7	4.96E-06	0.74343	0.65793	17.365	7	-0.604
DDB1;XAP1	Q16531;B7Z2	0.000246	0.84163	0.65378	6.4049	2	-0.613
RPL22	P35268;Q7Z4	0.00365	0.75948	0.65315		1	-0.615
AIM;CXXC9;DNM	P26358-1;P26	1.81E-35	0.9343	0.64864	52.558	22	-0.625
H2BFD;HIST1H2B	B4DR52;Q998	2.02E-31	0.8638	0.6484	90.738	11	-0.625
AAG;ANPG;MID1	P29372-1;P29	2.01E-52	0.8563	0.64729	24.66	25	-0.628
H2AFC;H2AFD;H2	P0C0S8;A4FT	1.03E-38	0.79852	0.64455	142.48	10	-0.634
CFIM68;CPSF6	Q16630-2;Q1	2.79E-06	0.53762	0.63989		1	-0.644
HNRPDL;JKTBP	O14979-1;O1	4.86E-10	0.79847	0.63578	55.131	4	-0.653
KIAA1649;PDIP46	B4E0L0;Q9BY	1.12E-13	0.66639	0.62456	176.38	2	-0.679
CAF1A;CAF1P60;(	Q13112;B2R7	0.007411	0.70755	0.62091		1	-0.688
BUB3	O43684-1;O4	1.16E-11	0.63775	0.61795	138.7	8	-0.694
H3.3A;H3.3B;H3F	P84243;B2R4	0.000267	0.6889	0.61715	27.25	6	-0.696
C3orf37;DC12	Q96FZ2;B4DM	3.14E-07	0.64101	0.61448	7.7988	2	-0.703
DAZAP1	Q96EP5-1;Q9	4.14E-57	0.81341	0.6076	60.756	15	-0.719
BAP135;GTF2I;W	P78347-1;P78	3.64E-61	0.78771	0.60587	73.985	21	-0.723
ACTR2;ARP2	B4DHK9;P611	1.57E-08	0.74191	0.60538	5.1545	3	-0.724
DERP9;EAP20;VP	Q9BRG1	0.015056	0.73561	0.60223		1	-0.732
TCOF1;DKFZp434	Q13428-4;Q1	3.84E-95	0.89118	0.59577	108.57	17	-0.747
hCG_20560;PTBP	Q9BUQ0;P26	8.80E-191	0.67182	0.59485	117.98	59	-0.749
WDR18	Q9BV38	0.000815	0.51722	0.59031		1	-0.76
KPNA2;RCH1;SRP	P52292;A8K7	3.71E-05	0.6219	0.58655	91.373	4	-0.77
G3BP;G3BP1;DKF	Q13283;B7Z8	0.000137	0.58954	0.58197	73.127	2	-0.781
TUBB2C;XTP3TPA	P68371;Q8IW	9.85E-18	0.62187	0.5714	187	2	-0.807
NKEFB;PRDX2;TD	P32119;B4DF	5.67E-06	0.58939	0.56646		1	-0.82
FIR;PUF60;ROBPI	Q9UHX1-2;Q9	3.80E-26	0.60914	0.5573	54.317	12	-0.843
NSUN2;SAKI;TRM	Q08J23;A8K5	7.19E-32	0.62343	0.55577	43.548	6	-0.847
PP8000;RAP1;TEI	Q9NYB0	5.86E-05	0.44967	0.54059		1	-0.887
SFRS11;hCG_228	Q05519;Q8IW	7.15E-06	0.49277	0.53997	54.5	2	-0.889
HASPP28;PDAP1	Q13442	0.002603	0.44782	0.5372		1	-0.896
ILF2;NF45;PRO30	Q12905;B4DY	3.65E-77	0.53447	0.53326	110.73	12	-0.907
U2AF2;U2AF65	P26368-1;P26	6.43E-19	0.5879	0.52334	157.64	4	-0.934
HNRNPA0;HNRPA	Q13151	2.92E-104	0.63754	0.51974	133.46	42	-0.944
HNRNPU;HNRPU;	Q00839-1;Q0	2.58E-222	0.75189	0.51852	136.5	82	-0.948
GNAS;GNAS1;GSI	Q5JWF2-1;Q5	3.06E-10	0.52874	0.50413	152.59	2	-0.988
FUBP1	Q96AE4-1;Q9	1.61E-126	0.56546	0.49544	97.354	17	-1.013
MSI2	Q96DH6-1;Q9	5.53E-12	0.42093	0.48468	42.971	2	-1.045
EZR;VIL2	P15311;B2R6	3.39E-16	0.48971	0.45547	77.121	2	-1.135
EIF5B;IF2;KIAA07	O60841;A0JLI	2.68E-23	0.66345	0.45504	100.38	7	-1.136
RPS5	P46782;Q53G	0.038659	0.55759	0.43149	76.043	3	-1.213
VAP33;VAPA;UN(	Q9P0L0-2;Q9I	3.80E-12	0.4286	0.42944	95.805	2	-1.219
SPTB2;SPTBN1;h(	Q01082-1;Q0	3.99E-12	0.47514	0.42508	73.037	2	-1.234
ACADVL;VLCAD;h	B4DEA8;Q53H	7.11E-09	0.35939	0.42141	43.061	2	-1.247
CAPRIN1;GPIAP1;	Q14444-1;Q1	1.65E-05	0.33876	0.41681		1	-1.263

HNRNPL;HNRPL;H	P14866;B2R9	4.54E-43	0.38536	0.41379	113.82	32	-1.273
RPL28;hCG_3823	B4DEP9;P467	0.010927	0.34496	0.41334		1	-1.275
FAS;FASN	P49327;Q135	4.68E-19	0.45552	0.40696	115.46	2	-1.297
HNRNPA1;HNRPA	P09651-1;P09	5.91E-166	0.51734	0.40337	101.02	64	-1.31
ABBP1;HNRNPAB	Q99729-3;Q9	3.70E-66	0.31851	0.39065	149.22	17	-1.356
C14orf92;KIAA07	O94842;B4DP	0.005121	0.30792	0.38212		1	-1.388
HNRPCL2;P542;R	Q9UKM9-1;Q	0.006728	0.31645	0.38023		1	-1.395
FEN1;RAD2;hCG_	P39748;B4DW	1.13E-197	0.35446	0.36909	134.47	46	-1.438
FACT80;SSRP1	Q08945	6.74E-05	0.3076	0.36887	20.259	2	-1.439
PAGA;PAGB;PRD	Q06830;B2R4	1.05E-07	0.46435	0.35447	162.32	2	-1.496
CYPA;PPIA	P62937;A8K2	1.29E-10	0.35069	0.3527	157.57	2	-1.503
C20orf158;DATF1	Q9BTC0-4;Q9	2.03E-08	0.29722	0.34691		1	-1.527
SSBP;SSBP1;hCG_	Q04837;A4D1	3.13E-162	0.4605	0.34642	100.52	24	-1.529
GSTK1;HDCMD47	Q9Y2Q3-2;Q9	1.07E-12	0.28612	0.34453		1	-1.537
ACS3;ACSL3;FACL	O95573;B2RB	0.001214	0.27122	0.32588	114.37	2	-1.618
BIT1;CGI-147;PTH	Q96ME4;Q9Y	0.018848	0.26459	0.31638		1	-1.66
CFIM25;CPSF25;C	O43809;B2R6	2.64E-09	0.26189	0.31496	203	3	-1.667
EIF2B;EIF2S2;hCG	P20042;B5BU	8.66E-33	0.31223	0.31474	54.43	3	-1.668
RECQ1;RECQL;RE	P46063	8.04E-104	0.27272	0.30991	119.84	38	-1.69
DXS423E;KIAA01	Q14683;A8K7	0.003161	0.2449	0.30716	77.189	3	-1.703
BAF170;SMARCC	Q8TAQ2-1;Q8	1.33E-09	0.26332	0.30397	51.398	3	-1.718
RPLP0	P05388;A8K4	3.82E-05	0.25213	0.30304	87.084	2	-1.722
ATF1;CREM;CREB	P18846;B2R8	0.002459	0.26584	0.29415	154.01	2	-1.765
CARD6	Q9BX69;Q59E	0.004841	0.38958	0.29064	250.25	3	-1.783
RPS25	P62851	2.76E-08	0.29876	0.28709	139.47	6	-1.8
BASP1;NAP22	P80723-1;P80	0.018686	0.23749	0.28703		1	-1.801
G22P2;XRCC5	P13010;Q53T	8.04E-41	0.25111	0.28687	115.83	19	-1.802
RP11-375E1__A.4	Q5TZB9;P467	0.000263	0.23271	0.2786	100.76	7	-1.844
IFI16;IFNGIP1;hC	B4DJT8;Q166	6.34E-33	0.22553	0.27685	72.553	7	-1.853
HNRNPR;HNRPR;	O43390-2;O4	6.23E-62	0.23732	0.26451	95.313	12	-1.919
BRUNOL2;CELF1;	Q92879-4;Q9	5.03E-06	0.25981	0.26393	149.61	4	-1.922
G22P1;XRCC6;CT	P12956;B2RD	8.75E-58	0.22931	0.2613	86.259	16	-1.936
NCL;hCG_33980	P19338;B3KM	1.22E-48	0.22633	0.26042	90.301	15	-1.941
DRAP1	Q14919-2;Q1	0.061773	0.20173	0.24156		1	-2.05
KIAA1401;TSR1	Q2NL82	1.63E-08	0.23616	0.23375	168.79	2	-2.097
ADPRT;PARP1;PP	P09874;B2R5	0	0.19651	0.23088	111.85	210	-2.115
MARCKSL1;MLP;I	P49006	0.02024	0.18989	0.22705		1	-2.139
CGI-55;PAIRBP1;S	Q8NC51-1;Q8	7.01E-119	0.17971	0.2172	101.77	19	-2.203
DEK	P35659;B4DN	2.16E-44	0.18542	0.20997	85.923	13	-2.252
CSDA;DBPA	P16989-1;P16	3.49E-65	0.20746	0.20964	81.159	5	-2.254
AHNAK;PM227	Q09666;B4DT	4.81E-30	0.1841	0.20913	119.93	9	-2.258
REPA1;RPA1;RPA	P27694	1.21E-76	0.1934	0.20825	112.7	11	-2.264
CGNL1;JACOP;KIA	Q0VF96-1;Q0	0.008644	0.18384	0.20494		1	-2.287
ARHGAP10;GRAF	A1A4S6;B4E2	0.27104	0.18355	0.20371		1	-2.295
VRK1	Q99986;A8K2	0.004686	0.15786	0.18877	44.964	2	-2.405
CSRP;CSRP1;CYRI	P21291;A8K2	6.17E-12	0.15531	0.18511	104.79	2	-2.434
DFS70;LEDGF;PSI	O75475-1;O7	6.34E-94	0.14444	0.18091	96.323	24	-2.467
TOP1	P11387;B9EG	1.66E-21	0.1654	0.17977	86.761	7	-2.476
ADA2B;TADA2B	Q86TJ2-1;Q8	0.021311	0.14232	0.17201		1	-2.539
SNRPB2	P08579;B5BT	9.66E-05	0.14288	0.17136		1	-2.545

H4/A;H4/B;H4/C;	P62805;B2R4	8.02E-22	0.16205	0.15736	110.36	7	-2.668
K6B;KRT6B;KRTL1	P04259	1.08E-64	0.15036	0.14475	93.795	4	-2.788
RPL24;hCG_2023	P83731;C9JN	0.000538	0.12527	0.14459	91.496	4	-2.79
REPA2;RPA2;RPA	P15927-3;P15	0.000708	0.11557	0.13309	100.75	3	-2.91
HMG2;HMGB2	P26583;Q5U0	1.17E-28	0.10965	0.12993	147.88	15	-2.944
HMG1;HMGB1;R	P09429;Q5T7	3.53E-40	0.11702	0.12095	70.795	16	-3.048
EEF1A;EEF1A1;EF	P68104;A8K9	3.43E-22	0.10652	0.11468	116.64	29	-3.124
LCN1;VEGP;LCN1	P31025;Q5VS	0.021183	0.12791	0.11386		1	-3.135
AUF1;HNRNPD;H	Q14103-1;Q1	1.23E-120	0.10669	0.11266	93.758	61	-3.15
hCG_1744585;RA	B5MDF5;P628	1.96E-05	0.089823	0.10838		1	-3.206
ACTA2;ACTSA;AC	P62736;B3KP	2.06E-111	0.099511	0.1052	93.833	4	-3.249
UBC;DKFZp434K0	P0CG48;A8K6	2.03E-18	0.087026	0.10488	126.87	3	-3.253
GAPD2;GAPDH2;	O14556	6.21E-05	0.086386	0.10397		1	-3.266
DRG1;NEDD3;DK	Q9Y295;Q9UF	0.019484	0.087837	0.10268		1	-3.284
CS	O75390;B4DJ	9.84E-11	0.11006	0.08914	175.21	5	-3.488
RPL11;hCG_1739	P62913-1;P62	3.55E-29	0.066682	0.0808	199.74	3	-3.63
H1F3;HIST1H1D;H	P16402;P104	5.79E-15	0.063144	0.08079	24.206	5	-3.63
C6orf98;KIAA079	Q8NF91-1;Q8	0.001865	0.056461	0.07181	81.492	3	-3.8
HMG2A;HMG4;H	O15347;Q164	3.68E-24	0.054903	0.07018	27.352	3	-3.833
DCD;AIDD;DSEP;I	A5JHP3;P816	0.00104	0.086486	0.06888	26.893	2	-3.86
EIF4B	P23588;B4DR	0.004125	0.098923	0.06453	51.957	2	-3.954
HMG14;HMGN1;	P05114;Q3B7	7.85E-32	0.049714	0.06009	22.375	5	-4.057
SSR1;PSEC0262;T	C9IZQ1;B2R6	1.62E-30	0.043967	0.05313	178.33	3	-4.234
		0.013944	0.038223	0.04809	30.93	3	-4.378
HMGN3;PNAS-24	Q15651-1;Q1	1.37E-33	0.039915	0.04661	287.08	4	-4.423
ACTBL2	Q562R1	2.82E-43	0.042056	0.04644	26.557	2	-4.428
KIAA0042;KIF14	Q15058	0.025655	0.028197	0.03831	55.483	4	-4.706
C21orf124;C21or	O00764-1;O0	0.015446	0.048761	0.03715		1	-4.751
KIAA0296;ZNF64	O15015-2;O1	0.080539	0.025293	0.02022		1	-5.628
SCP1;SYCP1;RP11	Q15431;B7ZL	0.021859	0.024181	0.01798		1	-5.797
	B4DNN6	0.022064	0.013346	0.01723	147.05	3	-5.859
DMD;GS1-19O24	P11532-1;P11	1.34E-05	0.020134	0.01537	49.896	5	-6.024
ARA160;TMF1	P82094-2;P82	0.020868	0.0036165	0.00389	73.411	2	-8.008
CSRP2;LMO5;SM	Q16527	1.99E-06				0	
RAB15;DKFZp686	P59190-1;P59	0.005298				0	
ACTR3;ARP3;hCG	P61158;B4DT	0.068854				0	
IBP3;IGFBP3;hCG	P17936;A6XN	0.001081				0	
LYZ;LZM;hCG_24	P61626;B2R4	5.80E-48				0	
KIAA0144;NICE4;	Q14157-2;Q1	0.00803				0	
A4;PLP2	Q04941	1.41E-08				0	
ROD1;RP11-165M	O95758-4;O9	7.33E-17				0	
A18HNRNP;CIRBI	Q14011;Q53X	3.73E-21				0	
VDAC;VDAC1	P21796;B3KT	0.013387				0	
CBX3;CCDC32;hC	Q13185;A4D1	6.01E-06				0	
EIF3B;EIF3S9;hCG	P55884-2;P55	0.010084				0	
RPL14	P50914;A8K3	6.29E-07				0	
		0.017765				0	

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Jun 16, 2014

This is a License Agreement between Lukas Lercher ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3410810804972
License date	Jun 16, 2014
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature Reviews Molecular Cell Biology
Licensed content title	Chromatin history: our view from the bridge
Licensed content author	Donald E. Olins, Ada L. Olins
Licensed content date	Oct 1, 2003
Volume number	4
Issue number	10
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 3 panel a and c
Author of this NPG article	no
Your reference number	None
Title of your thesis / dissertation	Chemical tools for the study of epigenetic mechanisms
Expected completion date	Jun 2014
Estimated size (number of pages)	180
Total	0.00 GBP
Terms and Conditions	

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use,subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   Note: For republication from the *British Journal of Cancer,* the following credit lines apply.
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   Note: For adaptation from the *British Journal of Cancer,* the following credit line applies.
   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information.Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   Note: For translation from the *British Journal of Cancer,* the following credit line applies.
   Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:

v1.1

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number 501329217. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Jun 16, 2014

This is a License Agreement between Lukas Lercher ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3410811099255
License date	Jun 16, 2014
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature Reviews Molecular Cell Biology
Licensed content title	New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?
Licensed content author	Karolin Luger, Mekonnen L. Dechassa and David J. Tremethick
Licensed content date	Jul 1, 2012
Volume number	13
Issue number	7
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 4 panel b and d
Author of this NPG article	no
Your reference number	None
Title of your thesis / dissertation	Chemical tools for the study of epigenetic mechanisms
Expected completion date	Jun 2014
Estimated size (number of pages)	180
Total	0.00 USD
Terms and Conditions	

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use,subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   Note: For republication from the *British Journal of Cancer,* the following credit lines apply.
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   Note: For adaptation from the *British Journal of Cancer,* the following credit line applies.
   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information.Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   Note: For translation from the *British Journal of Cancer,* the following credit line applies.
   Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:

v1.1

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number 501329229. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Jun 16, 2014

This is a License Agreement between Lukas Lercher ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3410811261128
License date	Jun 16, 2014
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature Chemical Biology
Licensed content title	Chromatin as an expansive canvas for chemical biology
Licensed content author	Beat Fierz, Tom W Muir
Licensed content date	Apr 17, 2012
Volume number	8
Issue number	5
Type of Use	reuse in a dissertation / thesis
Requestor type	academic/educational
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Figures	Figure 1b
Author of this NPG article	no
Your reference number	None
Title of your thesis / dissertation	Chemical tools for the study of epigenetic mechanisms
Expected completion date	Jun 2014
Estimated size (number of pages)	180
Total	0.00 USD
Terms and Conditions	

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use,subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run).NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   Note: For republication from the *British Journal of Cancer,* the following credit lines apply.
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:

   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   Note: For adaptation from the *British Journal of Cancer,* the following credit line applies.
   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information.Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:

   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication).

   Note: For translation from the *British Journal of Cancer,* the following credit line applies.
   Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:

v1.1

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number 501329231. Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

Copyright Clearance Center

1 PAYMENT 2 REVIEW 3 CONFIRMATION

Step 3: Order Confirmation

Thank you for your order! A confirmation for your order will be sent to your account email address. If you have questions about your order, you can call us at +1.855.239.3415 Toll Free, M-F between 3:00 AM and 6:00 PM (Eastern), or write to us at info@copyright.com. This is not an invoice.

Confirmation Number: 11223472
Order Date: 06/16/2014

If you paid by credit card, your order will be finalized and your card will be charged within 24 hours. If you choose to be invoiced, you can change or cancel your order until the invoice is generated.

Payment Information

Lukas Lercher
lukas.lercher@chem.ox.ac.uk
+44 7874808141
Payment Method: n/a

Order Details

Development

Order detail ID:	65181599
Order License Id:	3410820670359
ISSN:	0950-1991
Publication Type:	Journal
Volume:	
Issue:	
Start page:	
Publisher:	COMPANY OF BIOLOGISTS,
Author/Editor:	COMPANY OF BIOLOGISTS

Permission Status: Granted

Permission type: Republish or display content
Type of use: Republish in a thesis/dissertation

Requestor type	Academic institution
Format	Print, Electronic
Portion	chart/graph/table/figure
Number of charts/graphs/tables/ figures	1
Title or numeric reference of the portion(s)	Figure 2
Title of the article or chapter the portion is from	Skene, P. J.; Henikoff, S. "Histone variants in pluripotency and disease" Development 2013, 140, 2513.
Editor of portion(s)	N/A
Author of portion(s)	Skene, P. J.; Henikoff, S.
Volume of serial or monograph	140
Issue, if republishing	12

an article from a serial	
Page range of portion	2517
Publication date of portion	June 15, 2013
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	0 to 499
Made available in the following markets	all
The requesting person/organization	Lukas Lercher
Order reference number	None
Author/Editor	Lukas Lercher
The standard identifier	123
Title	Chemical tools for the study of epigenetic mechanisms
Publisher	University of Oxford
Expected publication date	Jun 2014
Estimated size (pages)	180

Note: This item will be invoiced or charged separately through CCC's **RightsLink** service. More info **$ 0.00**

This is not an invoice.

Total order items: 1 **Order Total: $ 0.00**

Confirmation Number: 11223472

Special Rightsholder Terms & Conditions

The following terms & conditions apply to the specific publication under which they are listed

Development
Permission type: Republish or display content
Type of use: Republish in a thesis/dissertation

TERMS AND CONDITIONS

The following terms are individual to this publisher:

The acknowledgement should state "Reproduced / adapted with permission" and give the source journal name. The acknowledgement should either provide full citation details or refer to the relevant citation in the article reference list. The full citation details should include authors, journal, year, volume, issue and page citation.

Where appearing online or in other electronic media, a link should be provided to the original article (e.g. via DOI):

Development: dev.biologists.org

Disease Models & Mechanisms: dmm.biologists.org

Journal of Cell Science: jcs.biologists.org

The Journal of Experimental Biology: jeb.biologists.org

Other Terms and Conditions:

None

STANDARD TERMS AND CONDITIONS

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

3.1 All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

3.2 General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, Dept 001, P.O. Box 843006, Boston, MA 02284-3006. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3 Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4 In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been

used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5 Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6 User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. **Miscellaneous.**

8.1 User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

8.2 Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here: http://www.copyright.com/content/cc3/en/tools/footer/privacypolicy.html.

8.3 The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

8.4 No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

8.5 The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such

federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to info@copyright.com.

v 1.1

Close

Confirmation Number: 11223472

Citation Information

Order Detail ID: 65181599

Development by COMPANY OF BIOLOGISTS Reproduced with permission of COMPANY OF BIOLOGISTS, in the format Republish in a thesis/dissertation via Copyright Clearance Center.

Close