Synthetic post-translational protein modification as a switch for activity



VOLUME I

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford.

Sarah Claire Faulkner

University College

Trinity Term 2019

AUTHOR'S DECLARATION

The work presented in this thesis was conducted at the Chemistry Research Laboratory at the University of Oxford under the supervision of Professor Ben Davis, and at the Sir William Dunn School of Pathology under the supervision of Professor Fumiko Esashi. All the work is my own, except where otherwise stated, and has not been submitted for any other degree at this or any other university.

Sarah Claire Faulkner

October 2019.

ABSTRACT

The ‘post-translational mutagenesis’ chemistry, utilising a bio-orthogonal radical reaction with Dha on proteins, has opened up a vast range of novel possibilities in generating post-translational modification (PTM) mimics. Two case studies are described in this thesis, around the theme of PTMs as switches for protein-protein interactions.

The first case study recapitulates the natural PTMs H3K36me2 and -me3. Functional evidence from in vivo observations suggests that the chromodomain reader protein MRG15 should preferentially bind H3K36me3 over its dimethyl counterpart. However, in vitro studies using peptides or isolated H3 as a substrate have failed to recapitulate this observation. Here, using synthetic, ‘post-translational mutagenesis’, fully-intact model nucleosomes containing both H3K36me2 and H3K36me3 were generated. Binding of MRG15 to these intact nucleosomes revealed a marked preference for H3K36me3 containing nucleosomes over those containing H3K36me2, recapitulating physiologically relevant observations. A model was suggested in which the H3K36 side chain is normally sequestered in intact nucleosomes in a hydrogen bonding interaction with the phosphate of the DNA backbone, and the interaction entirely abrogated only when the third methyl group is added to form H3K36me3, providing a ‘methyl-switch’ mechanism for context-dependent reader selectivity.

The second case study was centred around installing a natural PTM in an unnatural location to produce a novel sensor. Dha chemistry was once again employed to install *O*- and *N*-linked GlcNAc mimics in the complement determining region (CDR) of a nanobody cAbLys3, with the aim of reversibly blocking its antigen binding capacity. Enzymatic extension of cAbLys3-*O*-GlcNAc to produce an Fc decasaccharide was also investigated as an alternative means of tuning a nanobody’s effects for potential therapeutic applications.

Overall, the aim of this thesis was to demonstrate the use of the post-translational mutagenesis method to install mimics of labile PTMs; both to study their function in a natural setting, and to create novel biosensors.

ACKNOWLEDGEMENTS

Thank you to my supervisors, Ben and Fumiko. Studying at Oxford has been a dream of mine since I was a child growing up in New Zealand, and I appreciate being given the opportunity to make it a reality. To Ben, thanks for the insightful, novel, and sometimes out-there project ideas; to Fumiko, thanks for the thoughtful suggestions and knowledgeable guidance, particularly as I navigated the field of cell biology and DNA damage for the first time.

Thank you to the members of both labs who have helped me grow as a scientist, and a special thanks to those of you patient enough to proofread parts of this thesis; Abirami, Aziz, and Tom. There are too many people who have helped me become a better scientist in the lab to name, but I'd like to especially thank Mylene, Jitka, Nan, Matt B., Seb, Marjorie, Kim, Yiqun, and the nucleosome crew Ritu, Abirami and Simon. Jessi and Tom, thank you for cheering me on and listening to my constant moaning while I write up. Thanks to Brian J. for the histone help. Thanks also to the NMR staff, particularly Nader; and the mass spec staff, particularly James Wickens for patiently answering my phone calls whenever the open access was down.

I'd also like to acknowledge the people who made my time in the lab bearable when things were tough and better when they were going well; many of them I've named previously, but I'd also like to acknowledge Jessi, Alex, Tobi, Yue Jin, Brian L., and the various members of LG9 and LG10 who have put up with me taking over the lab speakers to play my aggressively terrible pop music. Thank you to Hayley and Chris B., who were there to support me during the most difficult times. Thanks also to the Esashi lab for welcoming me and teaching a chemical biologist how to run Western blots; Marjorie, Xanita, Lucia, Davide, Brian, and Isobel.

Thanks to UCBC. You taught me how to row as a novice in my first year with the club and how to be a more confident person as vice-captain in my second. Thanks to my friends from back home who haven't lost touch; (the better) Sarah, Richard, Sophie, Anya, Alice, Merve, Andrew, Hamish, and everyone else. It isn't easy to stay in contact with someone when you're separated by half a globe and 12 hours' time difference, but I appreciate that you guys made the time to talk to me and meet up when we were in the same country.

Finally, thanks to my parents for your constant support even while I'm on the other side of the world. I'm sorry I didn't get home for Christmas as often as you'd like, but thank you for taking me to the beach to get my yearly dose of vitamin D when I did. And Mum, thanks for putting Phoebe on video chat every Sunday.

PUBLICATIONS

T.H. Wright, B.J. Bower, J.M. Chalker, G.J.L. Bernardes, R. Wiewiora, W-L. Ng, R. Raj, S. Faulkner, M.R.J. Vallée, A. Phanumartwiwath, O.D. Coleman, M-L. Thézénas, M. Khan, S.R.G. Galan, L. Lercher, M.W. Schombs, S. Gerstberger, M.E. Palm-Espling, A.J. Baldwin, B.M. Kessler, T.D.W. Claridge, S. Mohammed, B.G. Davis. Posttranslational Mutagenesis: a Chemical Strategy for Exploring Protein Side-chain Diversity. *Science* **2016**, *354*, 597 and aag1465

Sarah Faulkner, Antony M. Couturier, Brian Josephson, Tom Watts, Benjamin G. Davis, and Fumiko Esashi. Intact nucleosomal context enables chromodomain reader MRG15 to distinguish H3K36me3 from -me2. **2020**, *Manuscript in submission.*

TABLE OF CONTENTS

AUTHOR'S DECLARATION I

ABSTRACT II

ACKNOWLEDGEMENTS III

PUBLICATIONS V

1 INTRODUCTION 1

TABLE OF CONTENTS FOR CHAPTER 1 2

LIST OF ABBREVIATIONS FOR CHAPTER 1 3

1.1. POST-TRANSLATIONAL MODIFICATIONS IN NATURE 6

1.2 METHODS FOR ENGINEERING PTMS 14

1.3 AIMS 26

1.4. REFERENCES 27

2. PROBING H3K36 METHYLATION WITH SYNTHETICALLY INSTALLED PTMS

PART 1: NUCLEOSOME CONSTRUCTION 44

TABLE OF CONTENTS FOR CHAPTER 2 45

LIST OF ABBREVIATIONS FOR CHAPTER 2 47

2.1 INTRODUCTION 50

2.2 RESULTS AND DISCUSSION 59

2.3 CONCLUSIONS 74

2.4 EXPERIMENTAL 76

2.5 REFERENCES 137

2.6 APPENDIX 144

3. PROBING H3K36 METHYLATION WITH SYNTHETICALLY INSTALLED PTMS

PART 2: INTERACTIONS WITH MRG15 184

TABLE OF CONTENTS FOR CHAPTER 3 185

LIST OF ABBREVIATIONS FOR CHAPTER 3 187

3.1 INTRODUCTION 191

3.2 Results and discussion 199
3.3 Conclusions 219
3.4 Experimental 222
3.5 References 260
3.6 Appendix 267

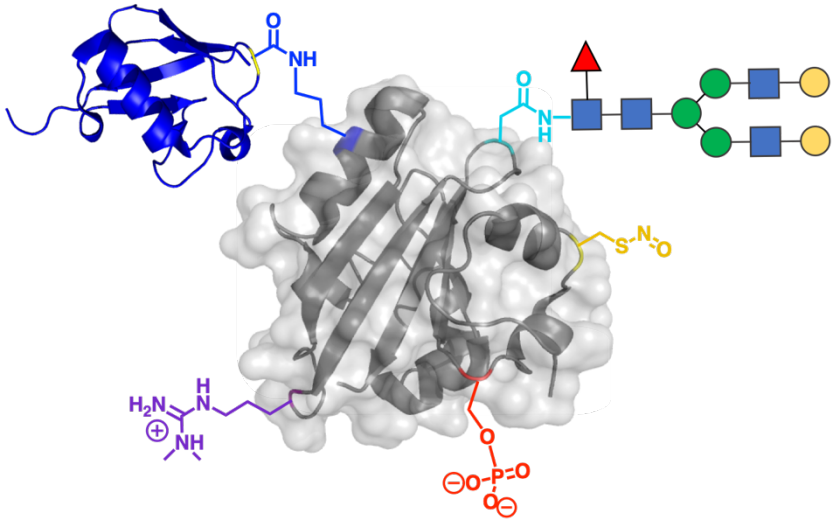
4. CONSTRUCTING LOGIC-GATED ANTIBODIES THROUGH SITE-SELECTIVE CHEMICAL GLYCOSYLATION 279

table of contents for chapter 4 280
list of abbreviations for chapter 4 282
4.1 Introduction 284
4.2 Results and discussion 295
4.3 Conclusions 311
4.4 Experimental 314
4.5 References 378
4.6 Appendix 390

5. CONCLUDING REMARKS 423

list of abbreviations for chapter 3 424
5. concluding remarks 425

1



Chapter 1: Introduction

1. Introduction

TABLE OF CONTENTS FOR CHAPTER 1

LIST OF ABBREVIATIONS FOR CHAPTER 1 3

1.1 POST-TRANSLATIONAL MODIFICATIONS IN NATURE 6

1.1.1 POST-TRANSLATIONAL MODIFICATIONS FINE-TUNE PROTEIN PROPERTIES 6

1.1.2 REGULATING PROTEIN ACTIVITY AT THE GENE EXPRESSION LEVEL IS SIMPLER IN PROKARYOTES 8

1.1.3 PTMS PROVIDE A DYNAMIC MECHANISM FOR RAPID RESPONSES 13

1.2 METHODS FOR ENGINEERING PTMS 14

1.2.1 SYNTHETIC, SITE-SELECTIVE PTM OF PROTEINS PRESENTS CHALLENGES 14

1.2.2. GENETIC CODE EXPANSION FOR SITE-SELECTIVE PTM INCORPORATION 16

1.2.3. SOLID PHASE PROTEIN SYNTHESIS-BASED METHODS FOR INSTALLING PTMS 19

1.2.4. CYSTEINE AS A REACTIVE HANDLE FOR SITE-SELECTIVE PTM 20

1.2.5. BIOORTHOGONAL REACTIONS AT DHA 22

1.3 AIMS 26

1.4 REFERENCES 27

LIST OF ABBREVIATIONS FOR CHAPTER 1

Abbreviation	Meaning
Ac	Acetylation
Aha	Azidohomoalanine
bp	Base pair
CDR	Complement determining region
C. elegans	*Caenorhabditis elegans*
CF_2pSerine	γ-difluorophosphoserine
Cit	Citrullination
Cr	Crotonylation
DBHDA	2,5-dibromohexane diamide
Dha	Dehydroalanine
DNA	Deoxyribonucleic Acid
E1cB	Elimination unimolecular conjugate base
E. coli	*Escherichia coli*
EF-Tu	Elongation Factor Thermo Unstable
Fuc	Fucose
Gal	Galactose
GFP	Green Fluorescent Protein
Gl	Glutarylation
Glc	Glucosylation
GlcNAc	N-acetylglucosamine
GOPAL	Genetically Encoded Orthogonal Protection and Activated Ligation
H2A	Histone 2A

H2B	Histone 2B
H3	Histone H3
H3K36me2	Histone H3 dimethylated at lysine at position 36
H3K36me3	Histone H3 trimethylated at lysine at position 36
H4	Histone H4
$LiAlH_4$	Lithium Aluminium Hydride
Ma	Malonylation
Man	Mannose
MDBP	2,5-methyldibromopentonate
me	monomethylation
me2	Dimethylation
me3	Trimethylation
MORF	Mortality-Factor
MRG15	MORF-related gene on chromosome 15
mRNA	Messenger ribonucleic acid
$NaBH_4$	Sodium Borohydride
NCL	Native Chemical Ligation
NO	Nitrosylation
NO_2	Nitration
NP-β	*Nostoc Punctiforme* β
O-GlcNAc	*O*-linked N-Acetylglucosamine
OH	Hydroxylation
p	Phosphorylation
Palm	Palmitoylation
PDB	Protein Data Bank

PIC	Pre-Initiation Complex
p-IF	*para*-Iodophenylalanine
Pol II	Polymerase II
PTM	Post-translational modification
RNA	Ribonucleic acid
RNAP	Ribonucleic acid polymerase
RNAPII	Ribonucleic acid polymerase II
RS	Aminoacyl tRNA Synthetase
SETD2	SET-Domain Containing 2
SSG	Glutathionylation
Succ	Succinylation
Sulf	Sulfation
SUMO	Small Ubiquitin-like Modifier
TFIIB	Transcription Factor II B
tRNA	Transfer ribonucleic acid
UAA	Unnatural Amino Acid
Ub	Ubiquitination
Ubl	Ubiquitin-like protein
UBQ	Ubiquitin

1. Introduction

1.1 Post-translational modifications in nature

1.1.1 Post-translational modifications fine-tune protein properties

In 1958, Francis Crick outlined what was to become the central dogma of molecular biology.[1] Here, he proposed that the information an organism needed to synthesise its proteins was encoded within DNA; indeed, we now know that triplet codons within DNA correspond to each of the twenty amino acids. By 1963, this hypothesis had expanded to include mRNA as the intermediate linking the reading of the DNA code, transcription, to its use as a template to build a primary amino acid sequence, translation.[2]

The twenty canonical amino acids provide a range of chemical functionalities, including charged, polar, and apolar side chains.[3] However, we now know that these can be further diversified by the addition of various chemical groups to the side chains after the central process of transcription and translation has taken place. This occurs in a manner not explicitly coded for within DNA, and is known as post-translational modification (PTM). Over 200 PTMs are now known and these can increase the structural diversity of a proteome by up to two orders of magnitude;[4,5,6] in the human proteome, which is expected to contain over 1 million proteins, this provides enormous scope for expanding their functionality.[7]

There are several kinds of post-translational processing which proteins can undergo. Not all PTMs involve addition of functional groups; proteolytic cleavage can convert the pro- form of a protein to its mature state.[8,9] Reactive prosthetic groups such as the heme of cytochrome P450 can also be conjugated to the protein to facilitate catalysis.[10,11] Moreover, appended groups can also serve to localise the protein to a particular cellular destination, such as lipids anchoring proteins to the cellular membrane, or to add stability and facilitate interactions with recognition partners, as glycans do.[12,13]

However, for this thesis there is one class of PTMs which is particularly relevant; labile, small molecule PTMs which can be reversibly added to a protein. The addition and removal of these PTMs provides a mechanism for switching a certain function of the protein on and off.

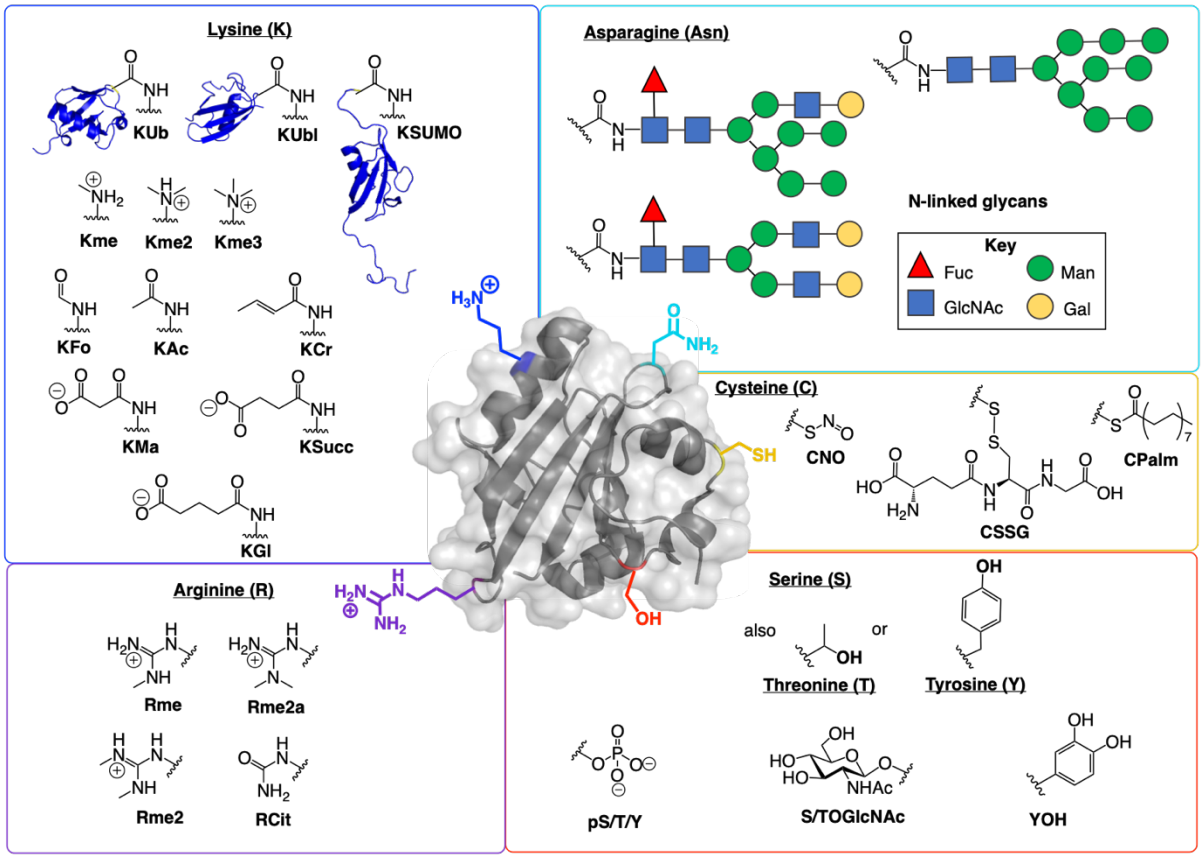


Fig. 1.1.1.1 *Selected examples of common PTMs. Biologically relevant PTMs commonly found on proteins were identified from the dbPTM database.*[14] *Abbreviations are as follows:*

Lysine*: Ub= ubiquitination (pictured: PDB ID 2MJB), Ubl= ubiquitin-like protein (pictured: PDB ID 1P0R), SUMO= small ubiquitin-like modifier (pictured: PDB ID 1A5R), me= monomethylation, me2= dimethylation, me3= trimethylation, Fo= formylation, Ac= acetylation, Cr= crotonylation, Ma= malonylation, Succ= succinylation, Gl= glutarylation.*

Arginine: *me= monomethylation, me2a, me2= demethylation, Cit= citrullination*

Serine/ Threonine/ Tyrosine: *p= phosphorylation, OGlcNAc= O-GlcNAcylation, OH= hydroxylation.*

Cysteine: *NO= nitrosylation, SSG= glutathionylation, Palm= palmitoylation.*

Asparagine*: N-linked glycans pictured: complex (top left), high mannose (top right)*[15] *and the Herceptin G2F glycan*[16]*. Fuc= fucose, GlcNAc= N-Acetylglucosamine, Man= mannose, Gal= galactose.*

Central protein pictured taken from PDB ID 6P4I.

1.1.2 Regulating protein activity at the gene expression level is simpler in prokaryotes

Whilst PTMs are known to occur in both prokaryotic and eukaryotic organisms, the nature of the modifications varies greatly between the two, and in general eukaryotic PTMs are much more diverse and widespread.[17] This is reflective in general of the greater complexity of eukaryotes compared to prokaryotes,[18] which can be understood on examining the mechanisms by which each type of cell responds to an external stimulus.

A classic example in prokaryotes is the *lac* operon system (see fig. 1.1.2.1A).[19] When the cell is growing on glucose, the constitutively expressed *lac* repressor binds to the *lac* operon for genes *lacZ, lacY* and *lacA* encoding β-Galactosidase, β-Galactoside permease and β-Galactoside transacetylase, respectively. However, when lactose enters the cell and isomerizes to allolactose which subsequently binds the *lac* repressor, these genes are expressed, allowing the organism to catabolise lactose and making it more permeable to further lactose uptake. Here, the organism is able to effectively up-regulate gene expression as a response to an external stimulus.

Importantly, the success of genetic regulation systems such as the *lac* operon is dependent on the coupling of transcription and translation in prokaryotes. This is not the case in eukaryotes,

however, where the two processes are uncoupled; as a result, the fine-tuning provided by PTMs is needed to produce a response with similar efficacy.

The uncoupling of transcription and translation in eukaryotes arises from several factors. First, many eukaryotic organisms are multicellular, and not every cell has an identical function; humans have over 200 different cell types, each expressing a very different profile of genes to generate a different cellular proteome.[20] However, the genetic information carried by each cell is identical. Broadly speaking, this means that every cell carries the information needed to express every protein in the body, but much of this information will not be used. Not only is the amount of genetic material present in eukaryotic cells therefore greater, but much of it needs to be stored in such a way that it cannot be inappropriately transcribed. This need for heterogeneous and multi-layered chromatin packing in turn curtails the usefulness of a simple repressor system for regulation.

To achieve this, the DNA is stored in the form of chromatin wrapped tightly around histone protein octamers (see fig. 1.1.2.1B) . An important consequence of this is that the accessibility of the chromatin provides a major pathway for regulating transcription.[21] In order for transcription to occur, a pre-initiation complex (PIC) consisting of RNA Polymerase II (RNAPII) and many transcription factors must bind to the gene's promoter, necessitating physical contact between the large multiprotein complex and DNA; this cannot occur if the DNA is sequestered as heterochromatin or physically blocked by nucleosomes. Indeed, nucleosomes are not uniformly distributed through the genome, but are more dense in regions of heterochromatin, and depleted in regulatory loci, where transcription factors are often bound.[2223] Moreover, chromatin state is not constant, but changes over time as the expression profile needed for a particular cell changes; for example, some genes are switched on very

early in development and subsequently turned off, or vice versa.[24,25] Therefore, the mechanisms regulating chromatin accessibility must also be dynamic.

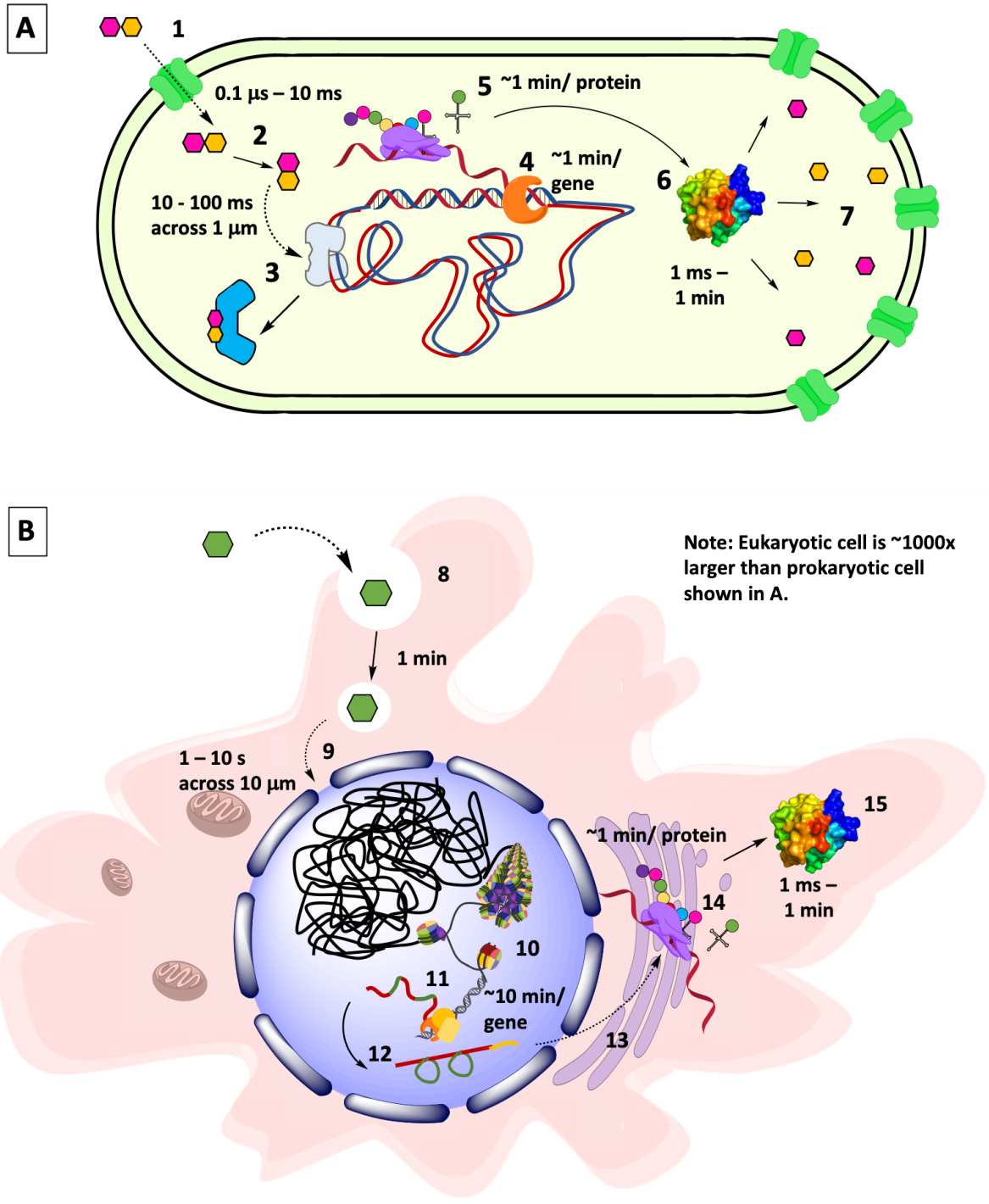


Fig. 1.1.2.1 *Comparison of gene upregulation in response to an external stimulus in prokaryotes (lac operon example) and eukaryotes. All timescales are taken from ref* [26]. ***A.*** *Schematic showing key points of the lac operon system in prokaryotes.* ***1.*** *Lactose is taken up into the cell.* ***2.*** *Lactose is isomerized to allolactose.* ***3****. Allolactose binds the lac repressor; the*

*repressor is then released from the lac operon, allowing RNAP to bind and begin transcription (**4**). **5**. Ribosomes can bind directly to mRNA as it is being transcribed and being translation. In the lac operon, three proteins are transcribed; β-Galactosidase, β-Galactoside permease and β-Galactoside transacetylase. **6**. Protein folding gives functional structures. **7.** The upregulated proteins increase processing and cellular uptake of lactose.*

***B**. Schematic showing key points of gene upregulation in a eukaryotic cell in response to a generic external stimulus. **8.** Clathrin-mediated endocytosis of the stimulus into cytosol. **9**. Diffusion of stimulus into nucleus. **10**. Chromatin is organised into fibres by wrapping around histone octamers to form nucleosomes, which can then condense into higher-order bundles. **11**. Gene transcription by RNAPII. Transcription occurs at the same rate as in prokaryotes (10- 100 nt/ s) but genes are ~10x longer. **12**. mRNA undergoes processing, including polyadenylation and intron excision (introns shown here in green, exons in red, polyA tail in yellow). **13**. mRNA is exported to the cytosol. **14.** Translation of mRNA by the ribosome; this begins in the cytosol but often moves to the rough endoplasmic reticulum as shown here. **15**. Protein folding; this is often mediated by chaperones (not shown).*

A key mechanism for achieving dynamic regulation of chromatin structure is the use of PTMs on nucleosome histones. With many modifiable residues, including a large number of lysines, histone proteins have at least 2^{28} PTM possibilities; moreover, many of these are known to be on the extended tail regions which protrude from the nucleosome core and are accessible to external proteins.[6] So far, over 20,000 interactions between the methylation and acetylation PTMs alone and various writers, readers and erasers have been identified, with more constantly being discovered.[27] These reader proteins are able to distinguish subtle differences between PTMs with remarkable accuracy; this is exemplified by chromodomain readers, which interact with methylated lysine.[28] Chromodomain readers are able to discriminate between the multiple

lysine residues on the H3 tail, and in some cases even between mono-, di- and tri-methylated states.[29,30] While the mechanism for achieving such exquisite discrimination has been determined for some chromodomain proteins, for others, such as MRG15 (MORF-related gene on Chromosome 15), it remains a mystery.

Furthermore, the mechanism by which the mRNA is transcribed causes further uncoupling of transcription and translation. Once the PIC has bound, there is no guarantee that a transcription attempt will be successful. 99% of transcripts produced are short polynucleotides of 10 nucleotides or less, indicating that most attempts are abortive.[3132] This is due to competition between the growing mRNA strand, beginning when it reaches 5-6 nucleotides in length and Transcription Factor II B (TFIIB) on the PolII saddle. In most cases, TFIIB wins out and initiation is abortive, but in some cases the growing mRNA wins and RNA Pol II 'breaks free' of the promoter to begin elongation. [33–37]

The mRNA produced then undergoes processing before it is translated. This includes 5' capping, 3' cleavage and, most significantly, splicing to remove introns.[38] The use of introns and exons in eukaryotic protein coding provides an efficient method for introducing more genetic diversity without bloating the overall DNA length. However, this means the mRNA initially transcribed is not the same as what is ultimately translated and cannot be used immediately in protein synthesis.

Finally, and perhaps most significantly, uncoupling is caused by the physical separation of transcription and translation with the compartmentalization of eukaryotic cells. Transcription occurs in the nucleus, where DNA is stored. The mRNA must then be exported to the cytoplasm for translation by the ribosome, unlike in prokaryotic cells where the mRNA can be translated

directly as it is coming off the RNA polymerase. Much like the case for transcription, most attempts at protein synthesis are also abortive, generating only short fragments.[39]

In comparison to prokaryotes, then, the additional complexity of eukaryotes impedes their ability to respond to stimuli through regulation of gene expression. This provides a particular niche for PTMs, not only in modulating their more complex genome, but in providing the means to tune protein response to stimuli in a precise manner.

1.1.3 PTMs provide a dynamic mechanism for rapid responses

Once a protein is expressed, it may also require fine control on a timescale that is not feasible from simply up- or down-regulating its expression. For example, glycolysis is one of the major metabolic pathways in cells, and metabolite turnover occurs on the order of one minute.[26] In prokaryotes, modulation of mRNA levels occurs on a similar timescale. Conversely, in eukaryotes it takes approximately 15 minutes from receiving an external signal to produce a folded protein (see fig. 1.1.2.1B), and the lag between an increase in mRNA levels and increased in protein levels is hours to days, depending greatly on the half-lives of the mRNA and protein, and the protein's expression level.[40] Moreover, once produced in eukaryotes, degradation of mRNA takes approximately 10 hours, and reduction of protein levels on the order of days.[26] In short, not only is there a significant lag time between genetic upregulation and increased protein concentration, but once this increase is attained it takes a long time to return to baseline.

PTMs can again provide a solution, rapidly and reversibly altering the properties of proteins as necessary. The most well-known example of this is phosphorylation, an anionic modification estimated to occur on 13,000 proteins in humans.[41]

Phosphorylation can alter the conformation of a protein via interaction between the negative phosphate group and positive charges such as basic side chains or the δ+ dipole of an α-helix. It can also facilitate new interactions between proteins. These properties enable phosphorylation signal cascades, where an external stimulus is rapidly amplified within a cell by a series of phosphorylation events to elicit a response.

Phosphorylation is not the only PTM involved in dynamic signalling events; *O*-GlcNAcylation has recently been found to occur in a similar fashion, switching protein activity on or off as needed.[42] Indeed, switching protein activity by the addition or removal of a PTM is common to many small-molecule PTMs.[43]

1.2 Methods for engineering PTMs

1.2.1 Synthetic, site-selective PTM of proteins presents challenges

As PTMs play such an integral role in the eukaryotic proteome, isolating modified proteins is a significant goal in molecular biology; however, it also presents significant challenges. Traditionally, proteins are isolated by overexpression in an appropriate system followed by purification. Bacterial expression systems are usually used where possible due to their high yields and ease of manipulation. However, this is not usually feasible for isolating proteins with PTMs of interest from eukaryotic systems, as prokaryotes utilise a different and more limited set of PTMs. Expressing PTM-bearing proteins in mammalian systems is also challenging. Many PTMs are transient and/ or heterogeneous, making isolation of a single

variant for studying challenging. A particularly significant example is the isolation of glycosylated antibodies for therapeutic applications; commercially available antibodies are often present as a heterogeneous mixture of glycoforms.[44]

In vivo, there are some examples of PTMs being installed on proteins via purely chemical means; for example, glycation and formylation of nucleosomes are non-enzymatic processes that are driven by oxidative stress.[45,46] However, most PTMs installed on proteins are there to fulfil a specific purpose, and therefore must generally be added in a precise location.

In vivo, this is achieved using appropriate enzymes. Here, protein-protein interactions are used to recognise a specific site where the PTM is to be added; the surface area needed for recognition is approximately 800 $Å^2$.[47,48] Some of the enzymes are highly promiscuous, capable of adding PTMs to a variety of substrates; for example, casein kinase has over 300 known substrates.[49] Others are specific to a single protein substrate. However, even in promiscuous enzymes a specific amino acid sequence motif, present in a range of substrates, is generally recognised.

In contrast, small molecule reagents for organic synthesis have a much smaller recognition area, and consequently will tend to react with an appropriate amino acid side chain anywhere on the protein, resulting in multiply labelled and often heterogeneous product. This is the case for labelling of lysine residues through reactions with hard electrophiles such as activated esters or isothiocyanates, or reductive alkylation with aldehydes.[50–52] Although such chemistries are commonly employed due to ease of use, they give variable product mixtures as a result and as such are limited to applications where specificity is not essential, such as conjugating antigens for potential vaccines.[53]

In order to generate a protein bearing a given PTM at a specific residue, then, the modified residue must either be included at the time of synthesis, or the protein substrate must be tuned

such that the position of reactivity is a unique or uncommon one which can preferentially react with the reagent to give a homogeneous, specifically labelled product- a 'tag and modify' approach.

1.2.2. Genetic code expansion for site-selective PTM incorporation

The first is the use of genetic code expansion to install an unnatural amino acid (UAA). Here, there are two strategies commonly employed to generate modified proteins; selective pressure incorporation, and the more ubiquitous stop codon suppression.

In selective pressure incorporation, limiting conditions are used and the growth media is supplemented with a UAA such that every instance of an amino acid is replaced by an unnatural counterpart. One such example is auxotrophic methionine suppression, where every Met is replaced by a given UAA, often with an engineered metRS capable of aminoacylating the Met AUG tRNA.[54] This can work well in proteins where the native amino acid replaced occurs only once in the primary sequence, but otherwise provides limited specificity. Moreover, as every instance of the amino acid is replaced everywhere in the proteome, this method may prove lethal to the organism.

In order to use these stop codon suppression techniques, it is crucial that the tRNA/ RS pair used is orthogonal to the organism used for protein expression to avoid cross-reactivity with native amino acids.[55–58] An efficient system has been elucidated over many years of research, and amber stop codon suppression for UAA incorporation has been implemented in a range of organisms including bacteria, yeast, mammalian cell cultures, *C. elegans*, and even in complex living mammals such as mice.[59–64]

In some cases, this method can be used to incorporate residues bearing PTMs or PTM mimics directly. The most successful example of this is acetyllysine, which has been used successfully to study histone acetylation, transcription factor acetylation, and enzymatic activity of a prolyl isomerase.[65–69]

It is also possible to encode phosphoserine (pSer), although this also requires reengineering of the elongation factor EF-Tu to accommodate the UAA, and efficiency is low.[70,71]

Where direct incorporation of the native PTM is not possible, it is sometimes feasible to substitute this with a close mimic. Examples include sulfated tyrosine, carboxylic acid derivatives of tyrosine, or nitrotyrosine as mimics of phosphotyrosine.[72–74]

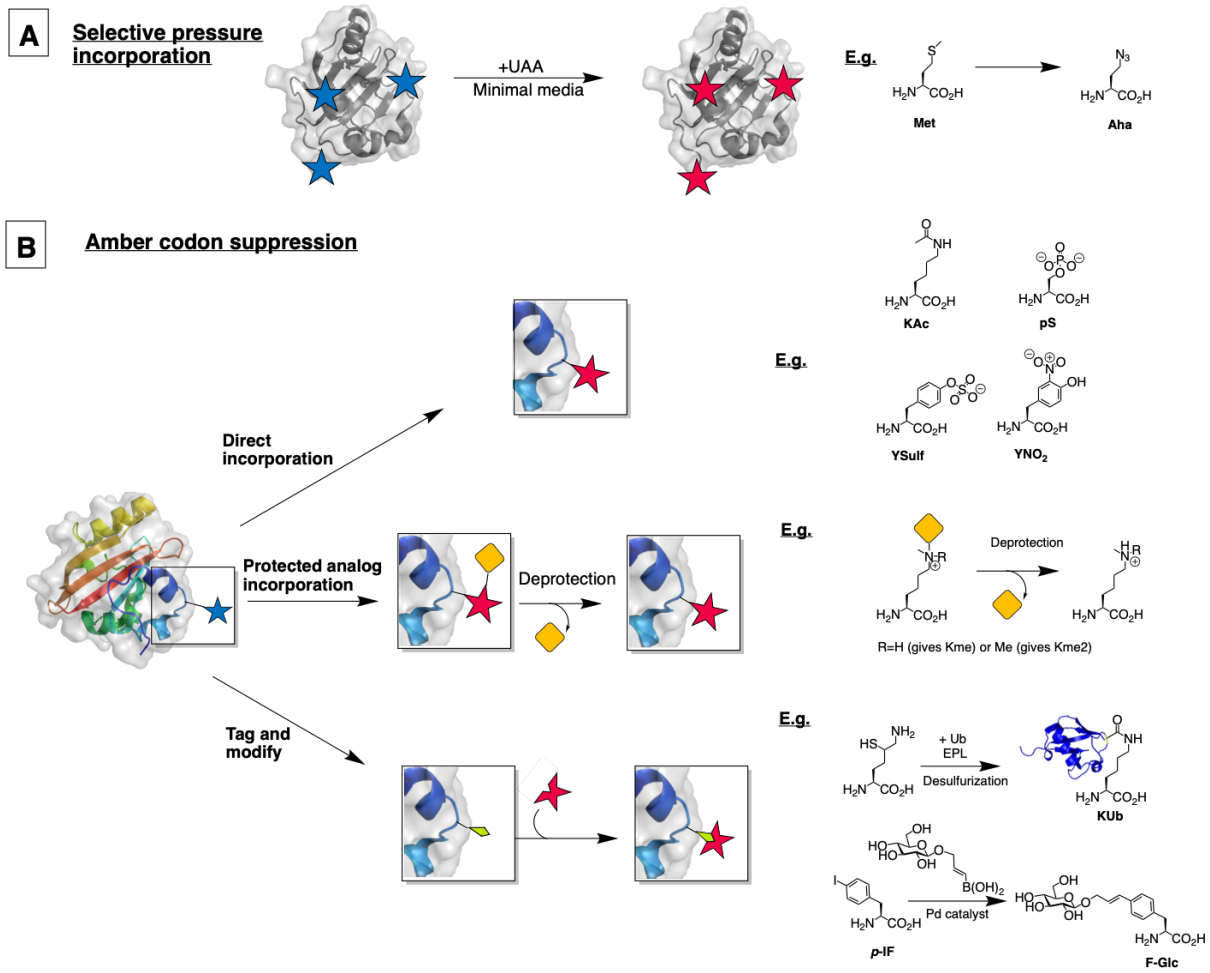


Fig. 1.2.2.1*. **A.** Schematic showing selective pressure incorporation, where every instance of a natural amino acid (blue star) in a protein is replaced with a UAA (red star). **Met**= Methionine, **Aha**= Azidohomoalanine. **B**. Schematic showing three strategies for*

incorporating PTMs or analogues using amber codon suppression. ***KAc****= Acetyllysine,* ***pS****= Phosphoserine,* ***YSulf****= Sulfated tyrosine,* ***YNO₂****= nitrated tyrosine,* ***KUb****= Ubiquitinated lysine,* **p-*IF*=** para-*Iodophenylalanine,* ***F-Glc****= Glucosylated phenylalanine (here an unnatural analog).Example protein= 6P4I.*

In addition to the need for orthogonality between organisms, the UAA incorporated must differ significantly enough from native amino acids in order to be efficiently discriminated. This means that small modifications, such as lysine methylation, are difficult to successfully encode. To overcome this, protected mono-and di-methyllysine variants have been genetically incorporated, then deprotected in the expressed protein by light or metal catalysis to give the modified amino acids.[75–79] As this method requires that at least one bond between protecting group and the lysine N be hydrolysed to an -N-H bond, it has not been possible to extend to trimethyllysine.

Finally, UAA incorporation has also been used to install a unique reactive handle from which a PTM can be appended. This strategy has been used extensively for installation of ubiquitin, a protein PTM too large to be accommodated by the translation machinery in a UAA.

The first method developed, Genetically Encoded Orthogonal Protection and Activated Ligation (GOPAL) involves installing a caged lysine UAA at the site ubiquitin is to be installed, followed by deprotection.[80,81] However, yields are generally poor using this method so a second involves incorporating a δ-thiol-Lys to make an isothiopeptide bond to UBQ, which can then be desulfurized to give the native linkage.[82]

This method has also been used successfully for the incorporation of a reactive azide handle, which has been used for a variety of applications including synthesis of a phosphoramidate mimic of phosphoserine, and remodelling of cell surface glycans.[83,84]

Other reactive handles include p-iodophenylalanine, which has found particular utility in the synthesis of glycoproteins.[85–88]

1.2.3. Solid phase protein synthesis-based methods for installing PTMs

Whilst powerful, UAA incorporation does not allow for installation of all PTMs, and in some cases the unnatural 'scar' left by attachment to a reactive handle affects the behaviour of a PTM mimic in a way that differs from the wild-type. Another strategy for specificity is to chemically synthesise part or all of the protein via solid-phase peptide synthesis, allowing residues bearing PTMs to be built in at precise sites. Total chemical synthesis of post-translationally modified proteins has been successfully implemented in some cases where other methods are currently unfeasible; for example, in generating a single glycoform of the multiply glycosylated protein erythropoietin, or in the synthesis of a fully D-amino acid chaperone to study the handedness of folding.[89,90] However, this method necessitates synthesis of a denatured peptide chain followed by refolding, which some proteins cannot survive. Moreover, synthesis of peptide chains is technically challenging and cumulative side reactions from small impurities and diminishing returns on yield mean that chain length is limited to approximately 50 amino acids.[91] In order to synthesise an entire protein, which is typically several hundred amino acids, it is necessary to synthesise several chains and stitch these together by native chemical ligation (NCL); as well as being synthetically challenging, this may also introduce chemical mutations at the sites of ligation which can affect activity.

One solution to this has been found in expressed-protein ligation based methods. Here, only the part of the protein bearing the PTM is chemically synthesised, while the rest of the truncated protein is expressed recombinantly; the two are then ligated.[92] Originally, this was done with a cysteine introduced at the terminus of the expressed protein, and was used successfully to synthesise such proteins as ribonuclease C , a glycoprotein, and lipidated Ras, a membrane protein.[93,94]

More recently, a trans-splicing intein method has also been used in which the peptide and expressed protein fragment bear two halves of an intein, which combine to guide splicing of the two protein fragments.[95] These methods have been successfully implemented to study histone modifications in cellulo.[96]

However, these methods work best for modifications at either end of the protein; as the length of the synthetic peptide needed increases, the synthetic challenge also increases and the stability of the remaining expressed fragment may increasingly become an issue. Moreover, in the case of split intein chemistry careful optimization of the exact intein sequence used is needed.

1.2.4. Cysteine as a reactive handle for site-selective PTM

The third general strategy involves targeting naturally occurring amino acids with low abundance to install chemical modifications. Cysteine is commonly chosen, as it has the lowest abundance of any of the 20 canonical amino acids with a frequency in primary amino acid sequences of approximately 2%.[97] Moreover, with careful selection of an appropriate soft electrophile, cysteine can be targeted over other nucleophilic side chains.[98]

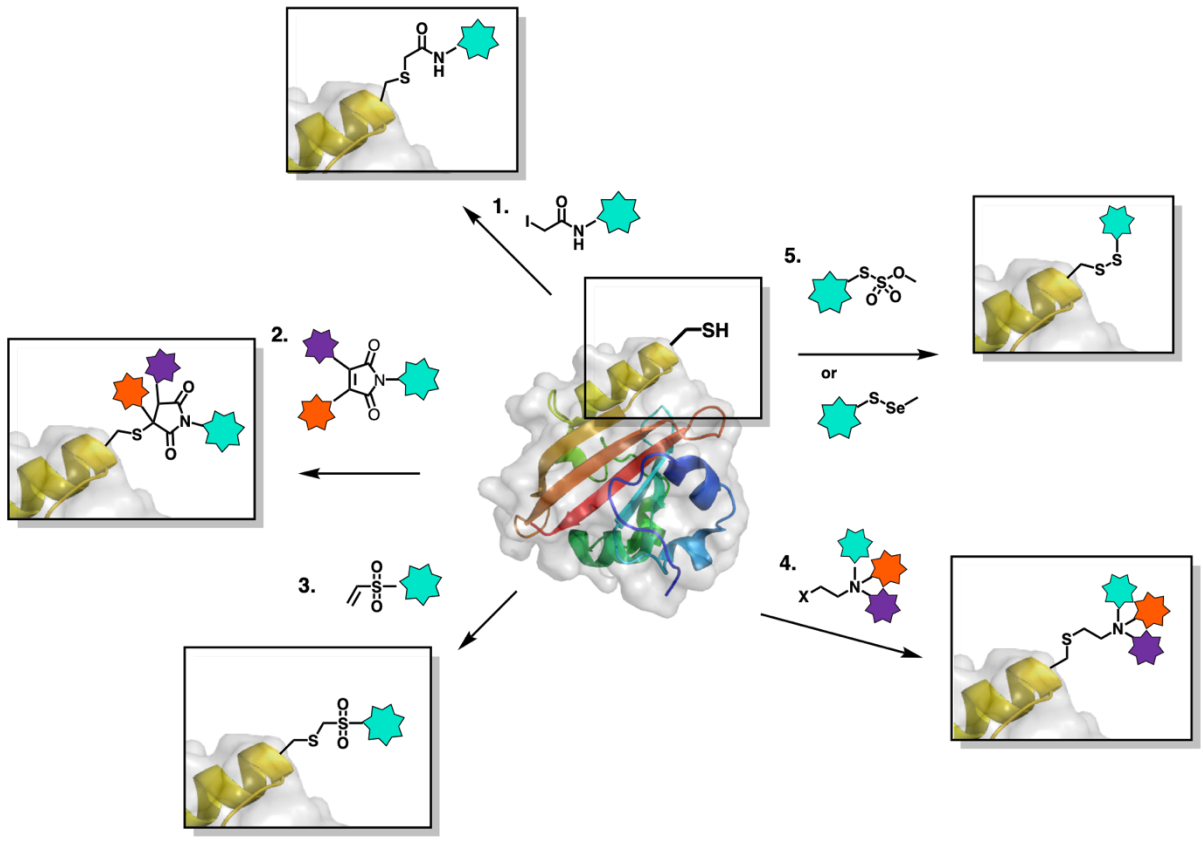


Fig. 1.2.4.1 *Reactions commonly used for selective cysteine labelling.* ***1.*** *Iodoacetamide* ***2.*** *Maleimide* ***3.*** *Vinyl sulfone* ***4.*** *Alkyl halide (X= Br or I)* ***5****. Disulfide exchange. Example protein= 6P4I.*

Nucleophilic substitution with iodoacetamide is typically used in mass spectrometry to cap cysteine residues peptides for analysis, but the iodoacetamide group has also been used as a linker to functionalise cysteines in intact proteins with a range of larger PTMs such as glycans.[99–102] Reaction with more minimal halide substrates has also enabled the installation of thio-methyllysine mimics.[103–106]

The use of Michael addition reactions has also proved a common approach, owing to their commercial availability and ease of use. The most ubiquitous of these is reaction with maleimide-based reagents, but vinyl sulfones have also been employed. [107–110]

Finally, thiol exchange can be used to form a new disulfide bond to cysteine, a strategy which has been employed to install PTM mimics such as prenylation and glycosylation. [111,112]

One major issue with these methods, however, is the possibility for side reactions with other nucleophilic side chains such as lysine. Conversely, the 20 canonical amino acids do not typically act as electrophiles; therefore, if an electrophilic residue could be installed on a protein, it would serve as a reactive handle which could selectively be targeted by nucleophilic reagents. This was realised with the utilisation of the Michael acceptor dehydroalanine (Dha).

1.2.5. Bioorthogonal reactions at Dha

Dha is found in nature as a product of serine dehydration or phosphoserine elimination in both peptides and proteins;[113,114] in some cases, it is a byproduct of protein ageing, whilst in others it facilitates catalysis.[115–117]

Initially, it was generated synthetically using elimination of an incorporated phosphoserine or chemically sulfonated serine, but harsh conditions were necessary and the scope for reaction was limited.[118,119] A method to generate Dha from phenylselenocysteine, introduced by Amber stop codon suppression, was also developed in the lab of Peter Schultz.[120,121]

However, the Davis group developed a more accessible method for converting cysteine to dehydroalanine.[122,123] The residue of interest is first mutated to cysteine using standard site-directed mutagenesis protocols, avoiding the need for harsh chemical conditions or UAA to install the reactive handle, which may lower protein yields. Initially, O-mesitylenesulfonylhydroxylamine was used; however, as this resulted in some unwanted

oxidation of methionine, lysine and histidine residues, other reagents were explored. The most successful of these proved to be 2,5-dibromohexane diamide (DBHDA) and 2,5-methyldibromopentonate (MDBP), the latter of which has been found particularly useful for sensitive proteins.[124,125] The mechanism of Dha formation has been studied in detail, and found to occur via an alkylation- E1cB elimination mechanism.[126]

In cases with more than one cysteine residues present in the target protein, selectivity can still be achieved where the target cysteine is more accessible or more reactive; indeed, if folded correctly most natural cysteine residues are locked in disulfide linkages and thus unaffected.[127–129] If these fail, selenocysteine can also be incorporated via amber codon suppression to generate selectivity; as it has greater acidity, DBHDA will react preferentially with it over cysteine. [130,131]

Dha can also be formed on a range of protein structural motifs. It has been formed on unfolded proteins, notably ubiquitin[132–137] and histones H2A, H2B, and H3, [138,139] in regions which would go on to successfully refold into C-termini, a-helices and linker regions. It has also been successfully formed on native proteins without disrupting their structure, such as the β-barrel of Green Fluorescent Protein (GFP)[140], β-sheets of NP-β[141], a monoclonal antibody (mAb) light chain in an intact antibody[123,125,142], α-helices and loops of phosphatases.[127,143] Moreover, it has been used to modify the active sites of enzymes such as Aurora A kinase [128], *N*-Acetylneuraminidate lyase[144], and Pantothenate Synthase. [124]

1. Introduction

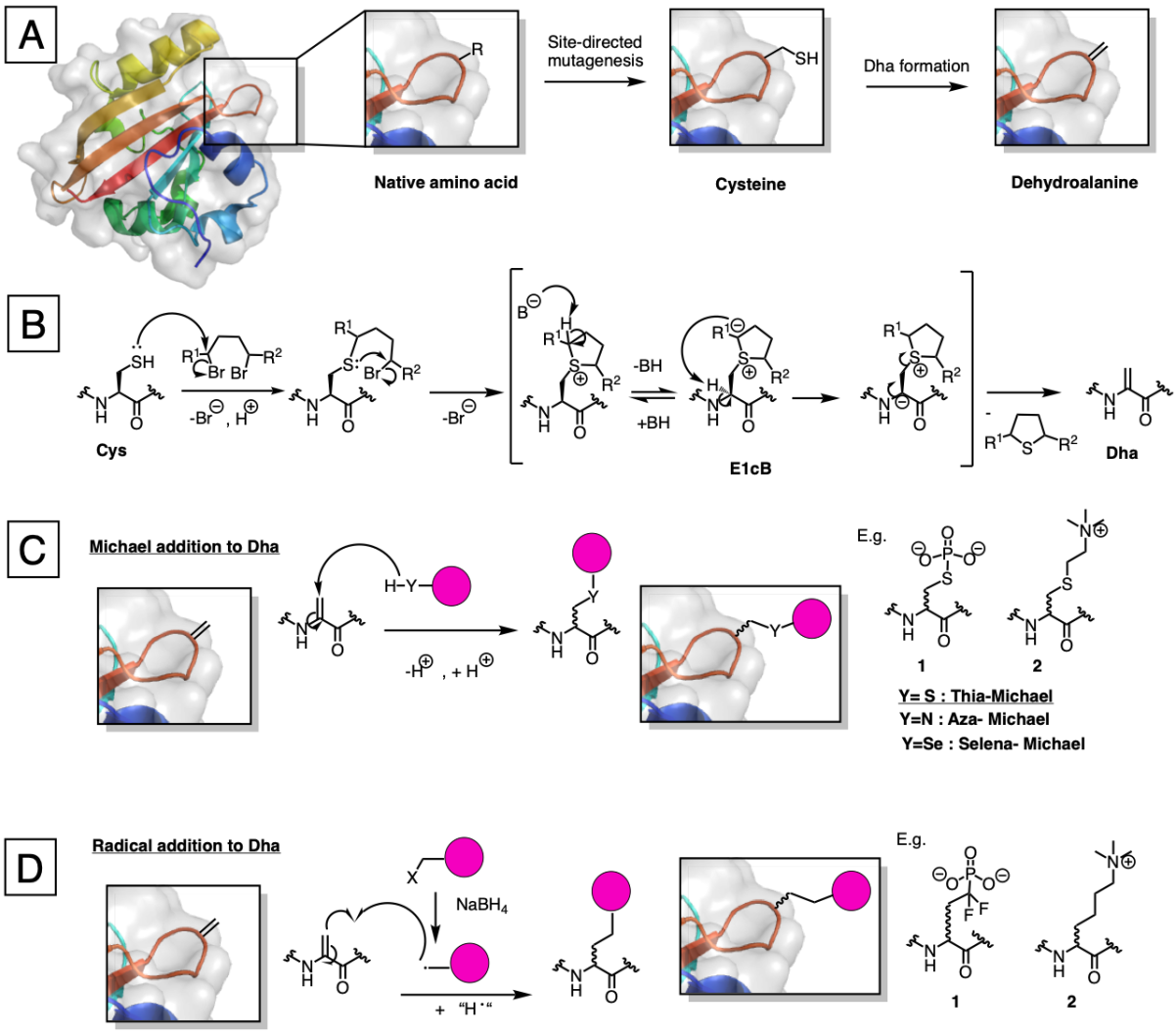


Fig. 1.2.5.1. A*. General scheme for conversion of a native amino acid to dehydroalanine (Dha).* ***R****= any canonical amino acid side chain.*

B*. Mechanism for the formation of Dha from cysteine with DBHDA:* R^1=R^2= $CONH_2$. *The mechanism for Dha formation with MDBP has additional steps, discussed more in Chapter 2.*

C*. General mechanism for Michael addition to Dha, where the pink circle represents any functional group designed to mimic a PTM.* ***1****= Phosphocysteine (mimics phosphoserine),* ***2****= Trimethyl-thialysine (mimics trimethyllysine).*

D. *General mechanism for radical addition to Dha, where the pink circle represents any functional group designed to mimic a PTM.* ***1****=* CF_2*pSerine (mimics phosphoserine)* ***2****= Trimethyllysine. Example protein= 6P4I.*

1. Introduction

Dha is an α-β unsaturated carbonyl, and as such can act as a Michael acceptor in reactions with nucleophiles.[145] This has been demonstrated with several variants, including thiamichael,[146,147] azamichael,[124,125] and selenamichael[148] additions. The thiamichael addition in particular has been used to install a range of PTM mimics, such as thiaphosphate and thia-GlcNAc.[138,149,150] It has also been used to install methylated thialysine mimics to study this PTM, in particular the trimethyl variant which, as noted before, was not accessible via amber codon suppression.[120] However, some literature suggests that the presence of the unnatural thia-linker causes the PTM mimic to behave in an unnatural way.[151,152] Indeed, this reflects the general issue that a chemical 'scar' is often left in the modified residue formed by reaction of nucleophiles with Dha, prohibiting the formation of truly biomimetic PTMs.

A breakthrough was made when the Davis group developed a method for radical C-C bond formation to Dha.[153] As most residues are inert to reaction with the C radical, it provided a powerful new method for selective installation of PTM mimics similar or identical to those found in nature.[154]

Initially, radicals were formed using $NaBH_4$ and a metal, Zn or In, as a redox catalyst, although this leads to an unwanted oxidative cleavage side product.[119] However, control reactions unexpectedly revealed that the reaction still proceeded with $NaBH_4$ alone, with cleaner results.[153,155,156] Whilst rare, examples of radical formation from halides by $LiAlH_4$ or $NaBH_4$ are known in the literature.[157–161]

The caveat to using this method is the unavoidable generation of a 50:50 mix of D and L epimers at the Cα centre. As the Dha is planar, attack can occur from either face, and studies on amino acids, peptides, and modified proteins suggest this does indeed occur.[162,163] Specifically, it has been noted that some of the native interaction partners such as antibodies against a certain PTM are capable of recognising only the natural L 50% of the total modified

protein, whilst others seem to process both epimers indiscriminately.[153] The effect of the two different epimers must therefore be evaluated on a case-by-case basis.

1.3 Aims

Post-translational modifications are able to tune the properties of proteins. This allows certain functions, notably interactions with other biomolecules, to be switched on or off. In eukaryotic cells, there are two key biological functions where the switchable nature of PTMs is important; in controlling the interactions of chromatin in the nucleus, and in rapid cell signalling in the cytoplasm. Recreating these PTMs in-vitro, however, is challenging, and several different chemical biology strategies have been developed to address the problem.

Recently, a method of selectively converting cysteine to a variety of PTM mimics via a radical reaction has been developed in the Davis group, allowing the exploration of previously inaccessible PTMs. This thesis describes two case studies using this method.

In the first case study, the histone PTM H3K36me3 was investigated This PTM is thought to be vital for normal development; a methyltransferase, SETD2 (SET-Domain Containing 2), exists solely to catalyse the conversion of H3K36me2 to H3K36me3, and malfunction of this methyltransferase causes cancers or embryonic lethality.[164–168] It is evident, then, that the switch from H3K36me2 to H3K36me3 must also be important; however, the mechanism effecting change is unclear given the relatively subtle chemical change to lysine on going from the me2 to me3 state.

H3K36 trimethylation has previously been challenging to study, as there is no way of genetically encoding the modification, its distance from the *N*-terminus decreases the efficacy of NCL-based methods, and the trimethyl-thialysine analog has several chemical differences that are thought to impact its function. However, Dha chemistry provides a method for installing the native trimethyllysine, albeit racemised at the Cα centre.

Chapter 2 details the construction of precisely methylated H3K36me2 and me3 containing nucleosomes. In Chapter 3, the interactions of these synthetic nucleosomes with chromodomain protein MRG15 (MORF-related gene on chromosome 15) are studied, and a mechanism for discrimination between the two states elucidated.

The second case study was centred around installing a natural PTM in an unnatural location to produce a novel sensor. Dha chemistry was once again employed to install *O*- and *N*-linked GlcNAc mimics in the complement determining region (CDR) of a nanobody, with the aim of reversibly blocking its antigen binding capacity. As up-regulated *O*-GlcNAcases are a known cancer biomarker, it was anticipated that this could form the basis for further medicinal applications.[169,170]

Overall, this thesis aims to demonstrate the use of radical-Dha chemistry to install mimics of labile PTMs, both to study their function in a natural setting, and to create novel biosensors.

1.4 References

1. CRICK, F. H. On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–63 (1958).
2. Watson, J. D. Involvement of RNA in the Synthesis of Proteins: The ordered interaction of three classes of RNA controls the assembly of amino acids into proteins.

Science (80-.). **140**, 17–26 (1963).

3. Bischoff, R. & Schlüter, H. Amino acids: Chemistry, functionality and selected non-enzymatic post-translational modifications. *J. Proteomics* **75**, 2275–2296 (2012).
4. Deribe, Y. L., Pawson, T. & Dikic, I. Post-translational modifications in signal integration. *Nat. Struct. Mol. Biol.* **17**, 666–672 (2010).
5. Duan, G. & Walther, D. The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *PLOS Comput. Biol.* **11**, e1004049 (2015).
6. Walsh, C. *Posttranslational modification of proteins: expanding nature's inventory*. (Roberts and Company Publishers, 2006).
7. Nørregaard Jensen, O. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* **8**, 33–41 (2004).
8. Rogers, L. D. & Overall, C. M. Proteolytic post-translational modification of proteins: proteomic tools and methodology. *Mol. Cell. Proteomics* **12**, 3532–42 (2013).
9. Kolattukudy, P. E., Sirakova, T. D., Barret, A. J., Rawlings, N. D. & Woessner, J. F. Handbook of Proteolytic Enzymes. (1998).
10. Gilardi, G. & Di Nardo, G. Heme iron centers in cytochrome P450: structure and catalytic activity. *Rend. Lincei* **28**, 159–167 (2017).
11. SMITH, M. H. Relative Distribution of Enzymes and Prosthetic Groups. *Nature* **213**, 627–628 (1967).
12. Jiang, H. *et al.* Protein Lipidation: Occurrence, Mechanisms, Biological Functions, and Enabling Technologies. *Chem. Rev.* **118**, 919–988 (2018).
13. Varki, A. & Gagneux, P. *Biological Functions of Glycans*. *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2015). doi:10.1101/GLYCOBIOLOGY.3E.007
14. Lu, C.-T. *et al.* dbPTM 3.0: an informative resource for investigating substrate site

specificity and functional association of protein post-translational modifications. *Nucleic Acids Res.* **41**, D295–D305 (2013).

15. Preston, R. J. S. *et al.* Elucidating the role of carbohydrate determinants in regulating hemostasis: insights and opportunities. *Blood* **121**, 3801–10 (2013).
16. Parsons, T. B. *et al.* Optimal Synthetic Glycosylation of a Therapeutic Antibody. *Angew. Chemie Int. Ed.* **55**, 2361–2367 (2016).
17. Gnad, F. *et al.* Evolutionary constraints of phosphorylation in eukaryotes, prokaryotes, and mitochondria. *Mol. Cell. Proteomics* **9**, 2642–53 (2010).
18. Szathmáry, E., Jordán, F. & Pál, C. Can Genes Explain Biological Complexity? *Science (80-.).* **292**, 1315 LP – 1316 (2001).
19. Zubay, G. The Mechanism of Activation of the Lac Operon. in *Metabolic Interconversion of Enzymes* 149–156 (Springer Berlin Heidelberg, 1972). doi:10.1007/978-3-662-37966-0_14
20. Alberts, B. *et al. Molecular biology of the cell*. (Garland Science, 2002).
21. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
22. Lee, C.-K., Shibata, Y., Rao, B., Strahl, B. D. & Lieb, J. D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36**, 900–905 (2004).
23. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
24. Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* **463**, 474–484 (2010).
25. Laan, S. Van der & Maiorano, D. Post-translational modifications in embryonic cell cycle. *Cell Cycle* **13**, 1364 (2014).
26. Shamir, M., Bar-On, Y., Phillips, R. & Milo, R. SnapShot: Timescales in Cell Biology.

Cell **164**, 1302-1302.e1 (2016).

27. Xu, Y. *et al.* WERAM: a database of writers, erasers and readers of histone acetylation and methylation in eukaryotes. *Nucleic Acids Res.* **45**, D264–D270 (2016).
28. Cavalli, G. & Paro, R. Chromo-domain proteins: linking chromatin structure to epigenetic regulation. *Curr. Opin. Cell Biol.* **10**, 354–360 (1998).
29. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Exp. &Amp; Mol. Med.* **49**, e324 (2017).
30. Sharifi-Zarchi, A. *et al.* DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *BMC Genomics* **18**, 964 (2017).
31. Goldman, S. R., Ebright, R. H. & Nickels, B. E. Direct detection of abortive RNA transcripts in vivo. *Science* **324**, 927–8 (2009).
32. Holstege, F. C., Fiedler, U. & Timmers, H. T. Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J.* **16**, 7468–80 (1997).
33. Cramer, P. *et al.* Structure of Eukaryotic RNA Polymerases. *Annu. Rev. Biophys.* **37**, 337–352 (2008).
34. Lee, T. I. & Young, R. A. Transcription of Eukaryotic Protein-Coding Genes. *Annu. Rev. Genet.* **34**, 77–137 (2000).
35. Kornberg, R. D. The molecular basis of eukaryotic transcription. *Proc. Natl. Acad. Sci.* **104**, 12955 LP – 12961 (2007).
36. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. & Darst, S. A. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285–90 (2002).
37. Murakami, K. S., Masuda, S. & Darst, S. A. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science* **296**, 1280–4 (2002).

38. Lodish, H. F. *Molecular cell biology*. (W.H. Freeman, 2000).

39. Baboo, S. *et al.* Most Human Proteins Made in Both Nucleus and Cytoplasm Turn Over within Minutes. *PLoS One* **9**, e99346 (2014).

40. Hargrove, J. L., Hulsey, M. G. & Beale, E. G. The kinetics of mammalian gene expression. *BioEssays* **13**, 667–674 (1991).

41. Vlastaridis, P. *et al.* Estimating the total number of phosphoproteins and phosphorylation sites in eukaryotic proteomes. *Gigascience* **6**, 1–11 (2017).

42. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu. Rev. Biochem.* **80**, 825–858 (2011).

43. DENHARDT, D. T. Signal-transducing protein phosphorylation cascades mediated by Ras/Rho proteins in the mammalian cell: the potential for multiplex signalling. *Biochem. J.* **318**, 729–747 (1996).

44. Wohlschlager, T. *et al.* Native mass spectrometry combined with enzymatic dissection unravels glycoform heterogeneity of biopharmaceuticals. *Nat. Commun.* **9**, 1713 (2018).

45. Zheng, Q. *et al.* Reversible histone glycation is associated with disease-related changes in chromatin architecture. *Nat. Commun.* **10**, 1289 (2019).

46. Jiang, T., Zhou, X., Taghizadeh, K., Dong, M. & Dedon, P. C. *N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage*. (2006).

47. Edwards, A. M. *et al.* Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends Genet.* **18**, 529–36 (2002).

48. Boeckmann, B. *et al.* Protein variety and functional diversity: Swiss-Prot annotation in its biological context. *C. R. Biol.* **328**, 882–899 (2005).

49. MEGGIO, F. & PINNA, L. A. One-thousand-and-one substrates of protein kinase CK2? *FASEB J.* **17**, 349–368 (2003).

50. Kalkhof, S. & Sinz, A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal. Bioanal. Chem.* **392**, 305–12 (2008).

51. Nakamura, T., Kawai, Y., Kitamoto, N., Osawa, T. & Kato, Y. Covalent Modification of Lysine Residues by Allyl Isothiocyanate in Physiological Conditions: Plausible Transformation of Isothiocyanate from Thiol to Amine. *Chem. Res. Toxicol.* **22**, 536–542 (2009).

52. Jentoft, N. & Dearborn, D. G. Labeling of proteins by reductive methylation using sodium cyanoborohydride. *J. Biol. Chem.* **254**, 4359–65 (1979).

53. Smith, M. L. *et al.* Modified Tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. *Virology* **348**, 475–488 (2006).

54. Johnson, J. A., Lu, Y. Y., Van Deventer, J. A. & Tirrell, D. A. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. *Curr. Opin. Chem. Biol.* **14**, 774–780 (2010).

55. Xie, J. & Schultz, P. G. An expanding genetic code. *Methods* **36**, 227–238 (2005).

56. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci.* **99**, 11020–11024 (2002).

57. Greiss, S. & Chin, J. W. Expanding the Genetic Code of an Animal. *J. Am. Chem. Soc.* **133**, 14196–14199 (2011).

58. Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A Genetically Encoded Photocaged Amino Acid. *J. Am. Chem. Soc.* **126**, 14306–14307 (2004).

59. Xie, J. & Schultz, P. G. A chemical toolkit for proteins—an expanded genetic code. *Nat. Rev. Mol. Cell Biol.* **7**, 775 (2006).

60. Chin, J. W. *et al.* An Expanded Eukaryotic Genetic Code. *Science (80-.).* **301**, 964–967 (2003).

61. Hancock, S. M., Uprety, R., Deiters, A. & Chin, J. W. Expanding the Genetic Code of Yeast for Incorporation of Diverse Unnatural Amino Acids via a Pyrrolysyl-tRNA Synthetase/tRNA Pair. *J. Am. Chem. Soc.* **132**, 14819–14824 (2010).

62. Mukai, T. *et al.* Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* **371**, 818–822 (2008).

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

64. Ernst, R. J. *et al.* Genetic code expansion in the mouse brain. *Nat. Chem. Biol.* **12**, 776–778 (2016).

65. Neumann, H. *et al.* A Method for Genetically Installing Site-Specific Acetylation in Recombinant Histones Defines the Effects of H3 K56 Acetylation. *Mol. Cell* **36**, 153–163 (2009).

66. Oppikofer, M. *et al.* A dual role of H4K16 acetylation in the establishment of yeast silent chromatin. *EMBO J.* **30**, 2610–2621 (2011).

67. Arbely, E. *et al.* Acetylation of lysine 120 of p53 endows DNA-binding specificity at effective physiological salt concentration. *Proc. Natl. Acad. Sci.* **108**, 8251–8256 (2011).

68. Thao, S., Chen, C.-S., Zhu, H. & Escalante-Semerena, J. C. Nε−Lysine Acetylation of a Bacterial Transcription Factor Inhibits Its DNA-Binding Activity. *PLoS One* **5**, e15123 (2010).

69. Lammers, M., Neumann, H., Chin, J. W. & James, L. C. Acetylation regulates Cyclophilin A catalysis, immunosuppression and HIV isomerization. *Nat. Chem. Biol.*

6, 331–337 (2010).

70. Park, H.-S. *et al.* Expanding the Genetic Code of Escherichia coli with Phosphoserine. *Science (80-.).* **333**, 1151–1154 (2011).

71. Lee, S. *et al.* A Facile Strategy for Selective Incorporation of Phosphoserine into Histones. *Angew. Chemie Int. Ed.* **52**, 5771–5775 (2013).

72. Liu, C. C. & Schultz, P. G. Recombinant expression of selectively sulfated proteins in Escherichia coli. *Nat. Biotechnol.* **24**, 1436 (2006).

73. Xie, J., Supekova, L. & Schultz, P. G. A Genetically Encoded Metabolically Stable Analogue of Phosphotyrosine in Escherichia coli. *ACS Chem. Biol.* **2**, 474–478 (2007).

74. Heinz Neumann, †, Jennifer L. Hazen, ‡, John Weinstein, ‡, Ryan A. Mehl, *,‡ and & Jason W. Chin*, †. Genetically Encoding Protein Oxidative Damage. (2008). doi:10.1021/JA710100D

75. Nguyen, D. P., Garcia Alai, M. M., Kapadnis, P. B., Neumann, H. & Chin, J. W. Genetically Encoding N^{ϵ} -Methyl- l -lysine in Recombinant Histones. *J. Am. Chem. Soc.* **131**, 14194–14195 (2009).

76. Nguyen, D. P., Alai, M. M. G., Virdee, S. & Chin, J. W. Genetically Directing ε-N, N-Dimethyl-l-Lysine in Recombinant Histones. *Chem. Biol.* **17**, 1072–1076 (2010).

77. Wang, Y.-S. *et al.* A genetically encoded photocaged Nε-methyl-l-lysine. *Mol. Biosyst.* **6**, 1557–1560 (2010).

78. Groff , D., Chen , P. R., Peters , F. B. & Schultz, P. G. A Genetically Encoded ε-N-Methyl Lysine in Mammalian Cells. *ChemBioChem* **11**, 1066–1068 (2010).

79. Ai, H., Lee, J. W. & Schultz, P. G. A method to site-specifically introduce methyllysine into proteins in E. coli. *Chem. Commun.* **46**, 5506–5508 (2010).

80. Virdee, S., Ye, Y., Nguyen, D. P., Komander, D. & Chin, J. W. Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nat. Chem.*

Biol. **6**, 750–757 (2010).

81. Castañeda, C. *et al.* Nonenzymatic Assembly of Natural Polyubiquitin Chains of Any Linkage Composition and Isotopic Labeling Scheme. *J. Am. Chem. Soc.* **133**, 17855–17868 (2011).

82. Virdee, S. *et al.* Traceless and Site-Specific Ubiquitination of Recombinant Proteins. *J. Am. Chem. Soc.* **133**, 10708–10711 (2011).

83. Serwa, R. *et al.* Chemoselective Staudinger-Phosphite Reaction of Azides for the Phosphorylation of Proteins. *Angew. Chemie Int. Ed.* **48**, 8234–8239 (2009).

84. Saxon, E. & Bertozzi, C. R. Cell Surface Engineering by a Modified Staudinger Reaction. *Science (80-.).* **287**, 2007 LP – 2010 (2000).

85. Spicer, C. D. & Davis, B. G. Rewriting the bacterial glycocalyx via Suzuki–Miyaura cross-coupling. *Chem. Commun.* **49**, 2747 (2013).

86. Spicer, C. D., Triemer, T. & Davis, B. G. Palladium-Mediated Cell-Surface Labeling. *J. Am. Chem. Soc.* **134**, 800–803 (2012).

87. Spicer, C. D. & Davis, B. G. Palladium-mediated site-selective Suzuki–Miyaura protein modification at genetically encoded aryl halides. *Chem. Commun.* **47**, 1698 (2011).

88. Bilyard, M. K. *et al.* Palladium-mediated enzyme activation suggests multiphase initiation of glycogenesis. *Nature* **563**, 235–240 (2018).

89. Wang, P. *et al.* Erythropoietin derived by chemical synthesis. *Science* **342**, 1357–1360 (2013).

90. Weinstock, M. T., Jacobsen, M. T. & Kay, M. S. Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity. *Proc. Natl. Acad. Sci.* **111**, 11679 LP – 11684 (2014).

91. Kent, S. B. H. Total chemical synthesis of proteins. *Chem. Soc. Rev.* **38**, 338–351

(2009).

92. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci.* **95**, 6705–6710 (1998).

93. Piontek, C. *et al.* Semisynthesis of a Homogeneous Glycoprotein Enzyme: Ribonuclease C: Part 2. *Angew. Chemie Int. Ed.* **48**, 1941–1945 (2009).

94. Bader, B. *et al.* Bioorganic synthesis of lipid-modified proteins for the study of signal transduction. *Nature* **403**, 223–226 (2000).

95. Shah, N. H., Dann, G. P., Vila-Perelló, M., Liu, Z. & Muir, T. W. Ultrafast Protein Splicing is Common among Cyanobacterial Split Inteins: Implications for Protein Engineering. *J. Am. Chem. Soc.* **134**, 11338–11341 (2012).

96. David, Y., Vila-Perelló, M., Verma, S. & Muir, T. W. Chemical tagging and customizing of cellular chromatin states using ultrafast trans-splicing inteins. *Nat. Chem.* **7**, 394–402 (2015).

97. Fodje, M. N. & Al-Karadaghi, S. Occurrence, conformational features and amino acid propensities for the π-helix. *Protein Eng. Des. Sel.* **15**, 353–358 (2002).

98. Weerapana, E., Simon, G. M. & Cravatt, B. F. Disparate proteome reactivity profiles of carbon electrophiles. *Nat. Chem. Biol.* **4**, 405–407 (2008).

99. Abo, M., Li, C. & Weerapana, E. Isotopically-Labeled Iodoacetamide-Alkyne Probes for Quantitative Cysteine-Reactivity Profiling. *Mol. Pharm.* **15**, 743–749 (2018).

100. Davis, N. J. & Flitsch, S. L. A novel method for the specific glycosylation of proteins. *Tetrahedron Lett.* **32**, 6793–6796 (1991).

101. Macmillan, D., Bill, R. M., Sage, K. A., Fern, D. & Flitsch, S. L. Selective in vitro glycosylation of recombinant proteins: semi-synthesis of novel homogeneous glycoforms of human erythropoietin. *Chem. Biol.* **8**, 133–45 (2001).

102. Swanwick, R. S., Daines, A. M., Tey, L.-H., Flitsch, S. L. & Allemann, R. K.

Increased Thermal Stability of Site-Selectively Glycosylated Dihydrofolate Reductase. *ChemBioChem* **6**, 1338–1340 (2005).

103. Simon, M. D. *et al.* The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **128**, 1003–1012 (2007).
104. LINDLEY, H. A New Synthetic Substrate for Trypsin and its Application to the Determination of the Amino-acid Sequence of Proteins. *Nature* **178**, 647–648 (1956).
105. Smith, H. B. & Hartman, F. C. Restoration of activity to catalytically deficient mutants of ribulosebisphosphate carboxylase/oxygenase by aminoethylation. *J. Biol. Chem.* **263**, 4921–5 (1988).
106. Chalker, J. M., Bernardes, G. J. L., Lin, Y. A. & Davis, B. G. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. - An Asian J.* **4**, 630–640 (2009).
107. Crankshaw, M. W. & Grant, G. A. Modification of cysteine. *Curr. Protoc. protein Sci.* **3**, 11–15 (1996).
108. Zhang, Y., Bhatt, V. S., Sun, G., Wang, P. G. & Palmer, A. F. Site-Selective Glycosylation of Hemoglobin on Cys β93. *Bioconjug. Chem.* **19**, 2221–2230 (2008).
109. Masri, M. S. & Friedman, M. Protein reactions with methyl and ethyl vinyl sulfones. *J. Protein Chem.* **7**, 49–54 (1988).
110. Hemelaar, J. *et al.* Specific and covalent targeting of conjugating and deconjugating enzymes of ubiquitin-like proteins. *Mol. Cell. Biol.* **24**, 84–95 (2004).
111. Gamblin, D. P. *et al.* Chemical site-selective prenylation of proteins. *Mol. Biosyst.* **4**, 558 (2008).
112. Bernardes, G. J. L., Gamblin, D. P. & Davis, B. G. The Direct Formation of Glycosyl Thiols from Reducing Sugars Allows One-Pot Protein Glycoconjugation. *Angew. Chemie* **118**, 4111–4115 (2006).

113. Ortega, M. A. & van der Donk, W. A. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem. Biol.* **23**, 31–44 (2016).

114. Repka, L. M., Chekan, J. R., Nair, S. K. & van der Donk, W. A. Mechanistic Understanding of Lanthipeptide Biosynthetic Enzymes. *Chem. Rev.* **117**, 5457–5520 (2017).

115. Li, H. *et al.* The Phosphothreonine Lyase Activity of a Bacterial Type III Effector Family. *Science (80-.).* **315**, 1000 LP – 1003 (2007).

116. Masters, P. M. In vivo decomposition of phosphoserine and serine in noncollagenous protein from human dentin. *Calcif. Tissue Int.* **37**, 236–241 (1985).

117. Wang, Z., Lyons, B., Truscott, R. J. W. & Schey, K. L. Human protein aging: modification and crosslinking through dehydroalanine and dehydrobutyrine intermediates. *Aging Cell* **13**, 226–234 (2014).

118. Strumeyer, D. H., White, W. N. & Koshland Jr, D. E. Role of serine in chymotrypsin action. Conversion of the active serine to dehydroalanine. *Proc. Natl. Acad. Sci. U. S. A.* **50**, 931 (1963).

119. Yang, A. *et al.* A chemical biology route to site-specific authentic protein modifications. *Science (80-.).* **354**, 623–626 (2016).

120. Guo, J., Wang, J., Lee, J. S. & Schultz, P. G. Site-Specific Incorporation of Methyl- and Acetyl-Lysine Analogues into Recombinant Proteins. *Angew. Chemie* **120**, 6499–6501 (2008).

121. Wang, J., Schiller, S. M. & Schultz, P. G. A Biosynthetic Route to Dehydroalanine-Containing Proteins. *Angew. Chemie* **119**, 6973–6975 (2007).

122. Bernardes, G. J. L., Chalker, J. M., Errey, J. C. & Davis, B. G. Facile conversion of cysteine and alkyl cysteines to dehydroalanine on protein surfaces: versatile and

switchable access to functionalized proteins. *J. Am. Chem. Soc.* **130**, 5052–5053 (2008).

123. Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2**, 1666–1676 (2011).

124. Dadová, J. *et al.* Precise probing of residue roles by post-translational β, γ-C, N aza-Michael mutagenesis in enzyme active sites. *ACS Cent. Sci.* **3**, 1168–1173 (2017).

125. Freedy, A. M. *et al.* Chemoselective installation of amine bonds on proteins through Aza-Michael ligation. *J. Am. Chem. Soc.* **139**, 18365–18375 (2017).

126. Galan, S. R. G. *et al.* Post-translational site-selective protein backbone α-deuteration. *Nat. Chem. Biol.* **14**, 955–963 (2018).

127. Rowan, F. C. *et al.* Insights into Aurora-A kinase activation using unnatural amino acids incorporated by chemical modification. *ACS Chem. Biol.* **8**, 2184–2191 (2013).

128. Bertoldo, J. B. *et al.* A water-bridged cysteine-cysteine redox regulation mechanism in bacterial protein tyrosine phosphatases. *Chem* **3**, 665–677 (2017).

129. Nathani, R. I. *et al.* A novel approach to the site-selective dual labelling of a protein via chemoselective cysteine modification. *Chem. Sci.* **4**, 3455–3458 (2013).

130. Whedon, S. D. *et al.* Selenocysteine as a latent bioorthogonal electrophilic probe for deubiquitylating enzymes. *J. Am. Chem. Soc.* **138**, 13774–13777 (2016).

131. Liu, J., Chen, Q. & Rozovsky, S. Utilizing selenocysteine for expressed protein ligation and bioconjugations. *J. Am. Chem. Soc.* **139**, 3430–3437 (2017).

132. Mulder, M. P. C. *et al.* A cascading activity-based probe sequentially targets E1–E2–E3 ubiquitin enzymes. *Nat. Chem. Biol.* **12**, 523 (2016).

133. Weber, A. *et al.* A linear diubiquitin-based probe for efficient and selective detection of the deubiquitinating enzyme OTULIN. *Cell Chem. Biol.* **24**, 1299–1313 (2017).

134. Tan, X.-D. *et al.* A diubiquitin-based photoaffinity probe for profiling K27-linkage

targeting deubiquitinases. *Chem. Commun.* **53**, 10208–10211 (2017).

135. An, H. & Statsyuk, A. V. Facile synthesis of covalent probes to capture enzymatic intermediates during E1 enzyme catalysis. *Chem. Commun.* **52**, 2477–2480 (2016).
136. Haj-Yahya, N. *et al.* Dehydroalanine-based diubiquitin activity probes. *Org. Lett.* **16**, 540–543 (2013).
137. Meledin, R., Mali, S. M., Singh, S. K. & Brik, A. Protein ubiquitination via dehydroalanine: development and insights into the diastereoselective 1, 4-addition step. *Org. Biomol. Chem.* **14**, 4817–4823 (2016).
138. Raj, R., Lercher, L., Mohammed, S. & Davis, B. G. Synthetic Nucleosomes Reveal that GlcNAcylation Modulates Direct Interaction with the FACT Complex. *Angew. Chemie Int. Ed.* **55**, 8918–8922 (2016).
139. Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J. & Davis, B. G. Conversion of Cysteine into Dehydroalanine Enables Access to Synthetic Histones Bearing Diverse Post-Translational Modifications. *Angew. Chemie Int. Ed.* **51**, 1835–1839 (2012).
140. Nathani, R., Moody, P., Smith, M. E. B., Fitzmaurice, R. J. & Caddick, S. Bioconjugation of green fluorescent protein via an unexpectedly stable cyclic sulfonium intermediate. *ChemBioChem* **13**, 1283–1285 (2012).
141. Vallée, M. R. J., Schombs, M. W., Balaban, Z. J., Colyer, J. & Davis, B. G. Ready display of antigenic peptides in a protein 'mimogen'. *Chem. Commun.* **52**, 3014–3017 (2016).
142. Gunnoo, S. B. *et al.* Creation of a gated antibody as a conditionally functional synthetic protein. *Nat. Commun.* **5**, 4388 (2014).
143. Chooi, K. P. *et al.* Synthetic phosphorylation of p38α recapitulates protein kinase activity. *J. Am. Chem. Soc.* **136**, 1698–1701 (2014).
144. Timms, N. *et al.* Structural Insights into the Recovery of Aldolase Activity in N-

Acetylneuraminic Acid Lyase by Replacement of the Catalytically Active Lysine with γ-Thialysine by Using a Chemical Mutagenesis Strategy. *ChemBioChem* **14**, 474–481 (2013).

145. Dadová, J., Galan, S. R. & Davis, B. G. Synthesis of modified proteins via functionalization of dehydroalanine. *Curr. Opin. Chem. Biol.* **46**, 71–81 (2018).
146. Windle, C. L. *et al.* Extending enzyme molecular recognition with an expanded amino acid alphabet. *Proc. Natl. Acad. Sci.* **114**, 2610–2615 (2017).
147. Rowan, F., Richards, M., Widya, M., Bayliss, R. & Blagg, J. Diverse Functionalization of Aurora-A Kinase at Specified Surface and Buried Sites by Native Chemical Modification. *PLoS One* **9**, e103935 (2014).
148. Lin, Y. A. *et al.* Rapid cross-metathesis for reversible protein modifications via chemical access to Se-allyl-selenocysteine in proteins. *J. Am. Chem. Soc.* **135**, 12156–12159 (2013).
149. Chen, Z. & Cole, P. A. Synthetic approaches to protein phosphorylation. *Curr. Opin. Chem. Biol.* **28**, 115–122 (2015).
150. Lercher, L. *et al.* Generation of a synthetic GlcNAcylated nucleosome reveals regulation of stability by H2A-Thr101 GlcNAcylation. *Nat. Commun.* **6**, 7978 (2015).
151. Seeliger, D. *et al.* Quantitative assessment of protein interaction with methyl-lysine analogues by hybrid computational and experimental approaches. *ACS Chem. Biol.* **7**, 150–154 (2011).
152. Chen, Z., Notti, R. Q., Ueberheide, B. & Ruthenburg, A. J. Quantitative and structural assessment of histone methyllysine analogue engagement by cognate binding proteins reveals affinity decrements relative to those of native counterparts. *Biochemistry* **57**, 300–304 (2017).
153. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring

protein side-chain diversity. *Science (80-.).* **354**, (2016).

154. Imiołek, M. *et al.* Selective radical trifluoromethylation of native residues in proteins. *J. Am. Chem. Soc.* **140**, 1568–1571 (2018).

155. Wright, T. H. & Davis, B. G. Post-translational mutagenesis for installation of natural and unnatural amino acid side chains into recombinant proteins. *Nat. Protoc.* **12**, 2243 (2017).

156. Wright, T. H. Post-translational mutagenesis: radical methods for protein modification. (University of Oxford, 2017).

157. Barltrop, J. A. & Bradbury, D. Chain photoreaction of sodium borohydride with halogenated aromatic hydrocarbons. Initiation by aryl radicals. *J. Am. Chem. Soc.* **95**, 5085–5086 (1973).

158. Ashby, E. C., DePriest, R. N., Goel, A. B., Wenderoth, B. & Pham, T. N. Occurrence of electron transfer in the reduction of organic halides by lithium aluminum hydride and aluminum hydride. *J. Org. Chem.* **49**, 3545–3556 (1984).

159. Stork, G. & Baine, N. H. Cyclization of vinyl radicals: a versatile method for the construction of five- and six-membered rings. *J. Am. Chem. Soc.* **104**, 2321–2323 (1982).

160. Corey, E. J. & Suggs, J. W. Method for catalytic dehalogenations via trialkyltin hydrides. *J. Org. Chem.* **40**, 2554–2555 (1975).

161. Groves, J. T. & Ma, K. W. Dehalogenations with sodium borohydride. Evidence for a free radical reaction. *J. Am. Chem. Soc.* **96**, 6527–6529 (1974).

162. Gutiérrez-Jiménez, M. I. *et al.* Bifunctional chiral dehydroalanines for peptide coupling and stereoselective S-Michael addition. *Org. Lett.* **18**, 2796–2799 (2016).

163. Key, H. M. & Miller, S. J. Site-and stereoselective chemical editing of thiostrepton by Rh-catalyzed conjugate arylation: New analogues and collateral enantioselective

synthesis of amino acids. *J. Am. Chem. Soc.* **139**, 15460–15466 (2017).

164. Nimura, K. *et al.* A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf–Hirschhorn syndrome. *Nature* **460**, 287–291 (2009).

165. Ho, T. H. *et al.* High-resolution profiling of histone h3 lysine 36 trimethylation in metastatic renal cell carcinoma. *Oncogene* **35**, 1565–1574 (2016).

166. Zhu, X. *et al.* Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nat. Genet.* **46**, 287 (2014).

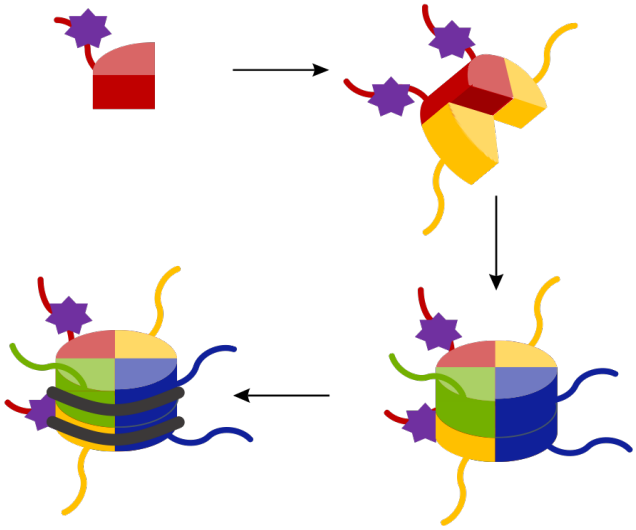
167. Newbold, R. F. & Mokbel, K. Evidence for a tumour suppressor function of SETD2 in human breast cancer: a new hypothesis. *Anticancer Res.* **30**, 3309–11 (2010).

168. Hu, M. *et al.* Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *Proc. Natl. Acad. Sci.* **107**, 2956–2961 (2010).

169. Yi, W. *et al.* Phosphofructokinase 1 Glycosylation Regulates Cell Growth and Metabolism. *Science (80-.).* **337**, 975 LP – 980 (2012).

170. Fardini, Y., Dehennaut, V., Lefebvre, T. & Issad, T. O-GlcNAcylation: A New Cancer Hallmark? *Front. Endocrinol. (Lausanne).* **4**, 99 (2013).

2



Chapter 2: Probing H3K36 methylation with synthetically installed PTMs

PART 1: NUCLEOSOME CONSTRUCTION

TABLE OF CONTENTS FOR CHAPTER 2

LIST OF ABBREVIATIONS FOR CHAPTER 2 47

2.1 INTRODUCTION 50

2.1.1 CHROMATIN IS PACKAGED AROUND HISTONES 50

2.1.2 HISTONES ARE MARKED WITH PTMS 53

3.1.3 HISTONE METHYLATION PLAYS VERSATILE ROLES IN MODULATING TRANSCRIPTION 54

3.1.4 METHYLATION AT H3K36 56

3.1.5 AIMS 58

2.2 RESULTS AND DISCUSSION 59

2.2.1 SITE- DIRECTED MUTAGENESIS, EXPRESSION, AND PURIFICATION OF H3K36C 59

2.2.2 IODIDE SYNTHESIS 62

2.2.3 OCTAMER ASSEMBLY 65

2.2.4 NUCLEOSOME ASSEMBLY WITH 145 BP DNA 69

2.2.5 NUCLEOSOME ASSEMBLY WITH 186 BP DNA 70

2.3 CONCLUSIONS 74

2.4 EXPERIMENTAL 76

2.4.1 GENERAL EXPERIMENTAL CONSIDERATIONS 76

2.4.2 SITE-DIRECTED MUTAGENESIS TO GENERATE H3K36C AND H4 78

2.4.3 HISTONE EXPRESSION 81

2.4.4 HISTONE PURIFICATION 82

2.4.5 OPTIMIZATION OF THE DHA FORMATION REACTION ON H3K36C 101

2.4.6 POST-TRANSLATIONAL MUTAGENESIS 106

2.4.7 OCTAMER REFOLDING 112

2.4.8 NUCLEOSOME REFOLDING WITH 145 BP DNA 116

2.4.9 186 BP DNA PREPARATION 120

2.4.10 NUCLEOSOME ASSEMBLY WITH 186 BP DNA 129

2.4.11 Reagent synthesis 135

2.5 REFERENCES 137

2.6 APPENDIX 144

2.6.1 Sequencing 144

2.6.2 LCMS-ESI+ spectra for Dha formation timecourse 162

2.6.3 Post-translational mutagenesis optimization 168

2.6.4 Single histone refolding and CD analysis 174

2.6.5 Octamer refolding 177

2.6.6 Nucleosome reconstitution 179

LIST OF ABBREVIATIONS FOR CHAPTER 2

Abbreviation	**Meaning**
2x YT	2x Yeast Tryptone
ADP	Adenosine diphosphate
Aly	Aly/REF export factor/ THO complex subunit 4
β-ME	β- mercaptoethanol
bp	Base pair
COSY	Correlation spectroscopy
DBHDA	2,5-dibromohexane diamide
DEPTQ	Distortionless enhancements by polarisation transfer with retention of quaternaries
Dha	Dehydroalanine
DNA	Deoxyribonucleic Acid
Dpn I	*Diplococcus pneumoniae* nuclease I
dUTP	Deoxyuridine triphosphate
E. coli	*Escherichia coli*
EcoRV	*Escherichia coli* restriction endonuclease V
EcoRI	*Escherichia coli* restriction endonuclease I
EDTA	Ethylenediaminetetraacetic acid
EHMT2	Euchromatic histone-lysine N-methyltransferase 2
ESI	Electrospray ionization
FPLC	Fast protein liquid chromatography
H1	Histone H1
H2A	Histone 2A
H2A.Z	Histone 2A (Z variant)

H2B	Histone 2B
H3	Histone H3
H4	Histone H4
HMT	Histone Methyltransferase
HPLC	High Performance Liquid Chromatography/ High Pressure Liquid Chromatography
HSQC	heteronuclear single quantum correlation
IPTG	Isopropyl-β-D-thiogalactoside
Iws1	Interacts with Spt6 1
LBA	Luria-Bertani media with Ampicillin
LB	Luria-Bertani
LCMS-ESI+	Liquid chromatography mass spectrometry positive electrospray ionization
LSD2	Lysine-specific histone demethylase 2
MDBP	2,5-methyldibromopentonate
mRNA	Messenger ribonucleic acid
MWCO	Molecular weight cut-off
NSD	Nuclear SET domain protein
OD_{600}	Optical density at 600 nm
O-GlcNAc	*O*-linked N-Acetylglucosamine
PCR	Polymerase chain reaction
PTM	Post-translational modification
RNA	Ribonucleic acid
RNAPII	Ribonucleic acid polymerase II
SDS-PAGE	Sodium dodecyl sulfate Polyacrylamide gel electrophoresis

Set2	SET domain containing protein 2 (yeast)
SETD2	SET domain containing protein 2 (human)
SOC	Super Optimal broth with Catabolite repression
Spt6	Transcription elongation factor Spt6 (SUPT6H)
SUMO	Small Ubiquitin-like Modifier
TLC	Thin-layer chromatography
TOF	Time of Flight
UV	Ultraviolet
Wnt	Wingless-related integration site
WT	Wild type
X. laevis	*Xenopus laevis*

2.1 Introduction

2.1.1 Chromatin is packaged around histones

Within each eukaryotic cell, the genetic information necessary to produce every protein in the organism is encoded in approximately 2 m of DNA and stored within a nucleus just 10 μM wide.[1] It is clear then that an efficient system for packaging and compacting the DNA is essential. Moreover, transcription of this genetic information into mRNA for downstream translation into proteins needs to be co-ordinated such that the correct proteins are transcribed at the correct time; this system for compaction must therefore have a level of organization associated with it to facilitate this.

The cell achieves both of these functions by wrapping DNA into protein- bound structures called nucleosomes. Each nucleosome contains a core octamer comprising two copies of each of the highly basic histone proteins H2A, H2B, H3 and H4, around which 147 bp of DNA is wrapped 1.65 times.[2] This structure is nearly invariant across eukaryotes and is optimised to facilitate translation; 147 bp is the average length of an exon within a genic region of DNA, and indeed nucleosomes are more prevalent in exons compared to other regions.[3,4] Between each nucleosome is 20-80 bp of linker DNA, depending on cell type.[5]

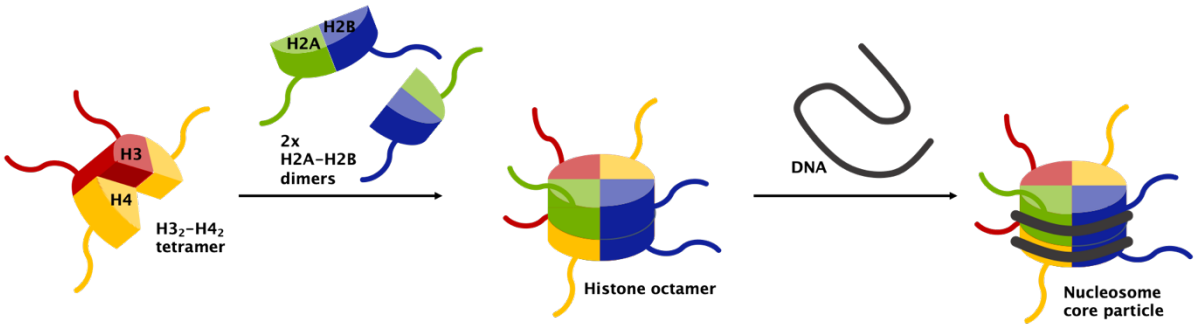


Fig. 2.1.1.1 *Histone assembly. Two copies each of H3 and H4 first form an $H3_2$-$H4_2$ tetramer, which then associate with two H2A-H2B dimers to form an octamer. DNA can then wrap around the histone octamer, forming the nucleosome core particle.*

Regions where there is no higher level of organization beyond the nucleosomes are known as euchromatin.[6] This form of chromatin is relatively open and therefore accessible to proteins such as transcription factors and RNA polymerases; it is a hallmark of active genes.[7] An additional histone, H1, can further organise chromatin into chromatosomes by positioning between the nucleosome and linker region of DNA.[8] This facilitates compaction of the chromatin into a fibre 30 nm wide.[9] The exact structure of this 30 nm fibre is debated, but it involves stacking interactions between the basic tail of H4 and the acidic surface of the H2A-H2B dimer.[10,11] This fibre then coils into a 300 nm filament which can wind on itself further to tightly compact the DNA.[12] The compacted form of chromatin is inaccessible and therefore transcriptionally silent; it is known as heterochromatin.[13]

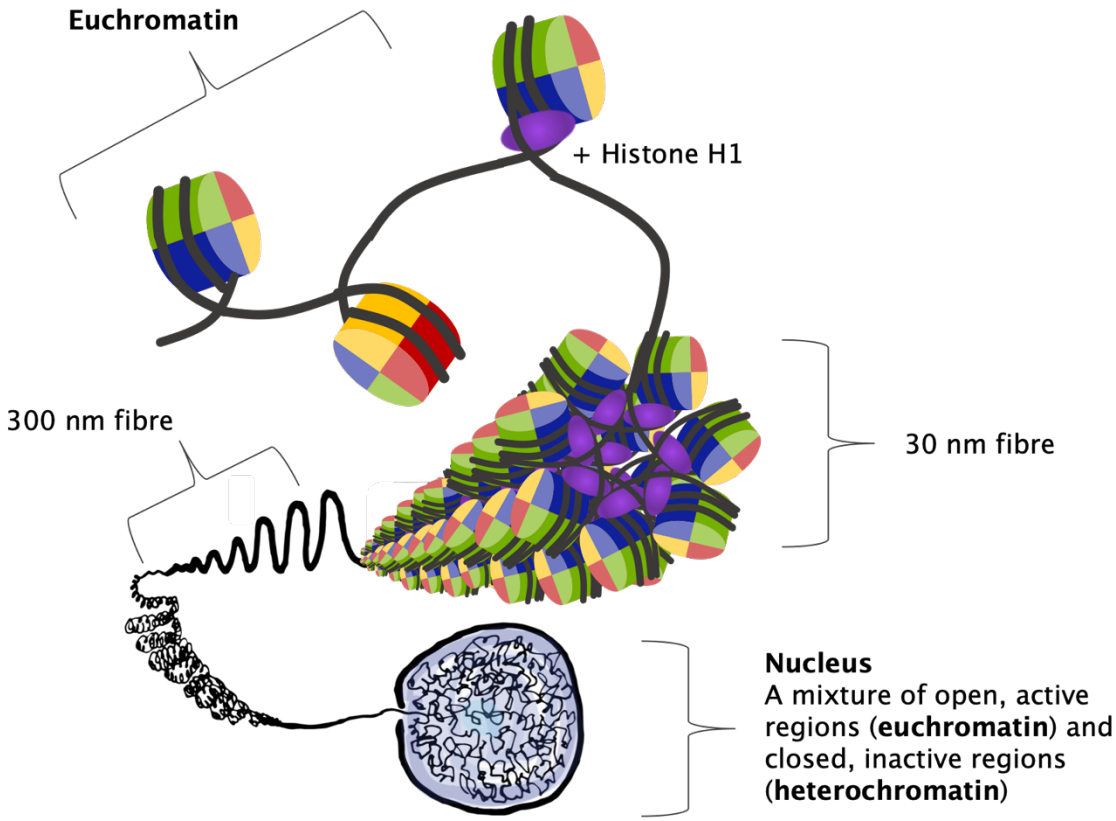


Fig.2.1.1.2 *Schematic overview of chromatin packing.*

Histones are highly stable and resilient proteins, inherited through cell division between generations of cells.[14] In addition to the structured core, each histone has a flexible N-terminal tail which protrudes from the nucleosome.[15,16] These tails allow for dynamic interactions with DNA and adjacent nucleosomes to facilitate transitions in the chromatin structure, as well as interaction with a range of partners within the nucleus.

As well as being inherited from parent to daughter cells, histones are also highly evolutionarily conserved across species.[17,18] Many of the seminal studies on nucleosomes have been carried out using histones isolated from the African clawed frog *X. laevis* and recombinantly expressed in *E. coli.*[19,20] However, conservation between species is so great that it is possible to mix histones from different organisms to reconstitute the nucleosome core particle. For example, nucleosomes with a mouse H2A.Z incorporated with *X.laevis* H2B, H3 and H4 have

successfully been generated.[21] This conservation and cross-species compatibility strongly suggests that conclusions drawn from work on *X. laevis* are applicable to a range of other eukaryotes, including humans.

2.1.2 Histones are marked with PTMs

Condensing the chromatin with the aid of nucleosomes provides a solution for DNA storage, but the question of organization remains as yet unanswered. Specifically, if all DNA was wrapped around identical histone octamers, how would the various transcription, replication and repair machineries access the DNA, and how would they distinguish which parts were to be accessed at which point?

The answer lies in modifications which can be found on the DNA or, of particular interest to us, post-translational modifications (PTMs) on the histone octamer. A vast range of histone PTMs are possible; known modifications include acetylation, ADP-ribosylation, biotinylation, butyrylation, citrullination, crotonylation, formylation, glutathionylation, hydroxyisobutyrylation, proline isomerization, malonylation, methylation, *O*-GlcNAcylation, hydroxylation, oxidation, phosphorylation, SUMOylation and ubiquitination.[22–26] Some of these PTMs can arise non-enzymatically through chemical reactions within the nucleus as, for example, formylation of histones occurs as a by-product of oxidative DNA damage.[27] However, most are introduced by 'writer' enzymes, recognised and bound by 'reader' proteins, and removed by 'eraser' enzymes, which suggests these modifications are purposely introduced by the cell to fulfil a specific function.[23] Moreover, multiple PTMs are often present together in varying permutations.

These observations led Strahl and Allis to propose the histone code hypothesis in 2000; "multiple histone modifications, acting in a combinatorial or sequential fashion in one or multiple histone tails, specify unique downstream functions."[28] In essence, there is a histone code arising from the combinations of these PTMs which are instrumental in regulating chromatin structure and access to DNA for the plethora of chromatin binding partners.

How then do these histone PTMs modulate chromatin accessibility? Some are able to act directly on chromatin structure, the best known example in this category being acetylation of basic residues.[29] The nucleosome is held together largely by coulombic interactions between the negatively charged DNA and the positively charged histones; mitigating the histone's charge, as occurs on acetylation of arginine or lysine to neutralise their positive charge, decreases this stabilizing coulombic interaction and loosens the DNA from the octamer. The nucleosome can also be pulled apart by the addition of steric bulk to a histone; for example, H2A T101 GlcNAcylation destabilises the H2A-H2B dimer by physically hindering the interface between the two.[30]

However, many other PTMs do not provide an immediately apparent means to directly alter chromatin structure. Of this second class of PTMs, the most well-known example is methylation.

3.1.3 Histone methylation plays versatile roles in modulating transcription

Histone methylation involves the addition of methyl groups to arginine, lysine, or more rarely, histidine, and is generally known to be involved in transcriptional regulation.[31–34] Of these modifications, lysine methylation is most relevant to this work and will be discussed further.

Lysine can bear one, two, or three methyl groups.[32,35,36] Each methyl group is relatively small (15 Da) and does not negate lysine's usual +1 net charge. Although methylation can alter pKa and basicity of the lysine, this does not provide a readily apparent direct mechanism for effect on chromatin organization. One mode of action is thought to arise from the linked methylation of histones and DNA acting in concert.[37] However, the majority of methyllysine's effects appear to arise from interactions with a plethora of partners within the cell.[38–40]

As might be expected from highly basic proteins, histones, and in particular H3, contain a larger than usual number of modifiable lysine residues. The most extensively studied of these are mostly found on the H3 tail; H3K4, H3K9, H3K27, H3K36, along with H3K79 and H4K20.[41]

Methylation at each of these residues serves to modulate transcription. Very broadly speaking, methylated H3K4, H3K36 and H3K79 are associated with active transcription, whereas H3K9, H3K27 and H4K20 methylation is associated with transcriptional repression.[34,42] However, the function of this marker is highly dependent on several factors, namely modification site, number of methyl groups, and cellular context.

The site of modification is crucial for determining the effect of methylation; for example, when trimethylation is found at H3K4 it is associated with active transcription, whereas at H3K27 it is linked to transcriptional repression.[43–45]

The methylation state is also key; H3K4me1, for example, is known for its enhancer function, whereas H3K4me3 is involved in promotor function.[46,47] Even a single methyl group can produce such marked differences, as in the case of H3K79me2 and -me3; the former is thought to play a role in cell cycle, the latter in the Wnt signalling pathway.[48,49]

Finally, the same modification can have different functions depending on context. For example, di- and tri- methylation and H3K4 is generally known to be activating but has been found to play a repressive role in some cases during development.[50,51]

3.1.4 Methylation at H3K36

H3K36 methylation is a marker for active transcription. It is installed by histone methyltransferase enzymes (HMTs) which follow RNA polymerase II (RNAP II).
In yeast, a single methyltransferase Set2 is capable of generating all three of H3K36's mono-, di- and tri- methylation states. It interacts with the hyperphosphorylated C-terminal domain of RNAP II to follow the transcription machinery and generate H3K36's various methylation states, including me3.[52]

In humans, this process is more complex. Once again, H3K36 methylation follows RNAPII, but in this instance several different enzymes are required to reach the trimethylated state.

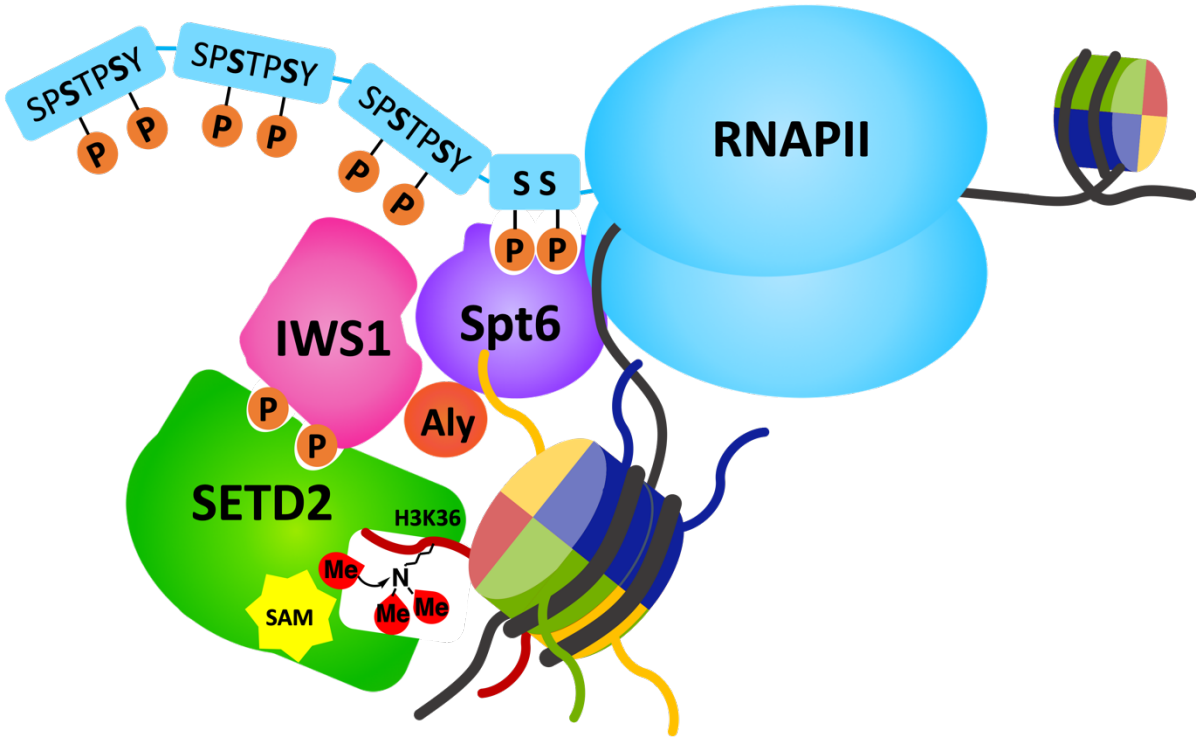


Fig. 3.1.4.1 *Schematic of H3K36me2 methylation by SETD2 to form H3K36me3. The RNAPII tail is hyperphosphorylated, allowing Spt6 to bind and recruit IWS1 and Aly. IWS1 is phosphorylated by AKT3/ AKT1 and can then recruit SETD2.*[53] *SETD2 then uses cofactor S-adenosylmethionine (SAM) to methylate H3K36me2. Note that the nucleosome is depicted as fully assembled during methylation for convenience, but this may not be the case* in-vivo.

First, NSD, LSD2 and EHMT2 form a complex which associates with transcription elongation factors and the phosphorylated C-terminal domain of RNAP II.[54] Together, they are able to catalyse the addition of up to two methyl groups.

SETD2 is separately recruited to the C-terminal domain of RNAP II by factors IWS1 and Spt6, and adds the final methyl group to H3K36me2 to generate H3K36me3.[55]

As SETD2 is the only enzyme capable of generating H3K36me3, we know that it, and by extension H3K36me3, must make a crucial contribution to normal cell functioning from the pathologies that arise when SETD2 malfunctions.[56] SETD2 plays a key role in genome

stability; functionally inactive mutants have been found responsible for clear cell carcinoma and some leukaemias, and it has additionally been implicated in suppressing breast cancer.[57–60] In a laboratory setting, we know that homozygous disruption of SETD2 is embryonic lethal.[61]

3.1.5 Aims

Although it is known that specifically attaining a trimethylated state at H3K36 is important for normal cell function, this modification's mechanism of action is unclear. Indeed, previous studies with known binding partners on K36me2 and -me3 containing peptides or whole H3 species have failed to discriminate between the two.[62] Considering this, it is hypothesized then that its function is due to an ensemble effect on the entire nucleosome rather than H3 as a whole.

The aim of the work described in this chapter was to generate synthetic nucleosomes containing the H3K36me2 and -me3 histones in order to elucidate a model for H3K36me3's observed biological action. Precisely methylated H3 variants were first made using semi-synthetic methodology, and then used to assemble nucleosomes for further study. Characterization data from the assembly revealed a difference in conformation that may begin to explain how biological systems discriminate between H3K36me2 and -me3 -containing nucleosomes.

2.2 Results and discussion

2.2.1 Site- directed mutagenesis, expression, and purification of H3K36C

In order to generate the desired methyllysine variants, a 'tag and modify' approach was chosen. The position of interest in H3 was first mutated to a cysteine, then transformed to dehydroalanine (Dha), which could undergo a 'post-translational mutagenesis' reaction with an appropriate alkyl iodide.

This strategy was particularly appealing for use on H3 for several reasons. First, the disordered tail region of H3 in which K36 is found is particularly robust to the chemical treatments required to generate Dha and append the side chain; indeed, the reaction was initially developed in the Davis group on other residues within the H3 tail.[63] We began with the fully functional H3C110A mutant (denoted here as 'H3 WT') commonly used in biophysical studies.[64] As cysteine is not present anywhere else in this H3 variant, it provided us with a reactive handle to target for further transformation.

Second, the reaction is also known to be amenable to the concentrated guanidinium buffer used to keep histones in an unfolded state. This is ideal because it allowed the attachment of the side chain whilst H3 is denatured, after which it could be lyophilised and then refolded with other histones as needed in octamer preparation.

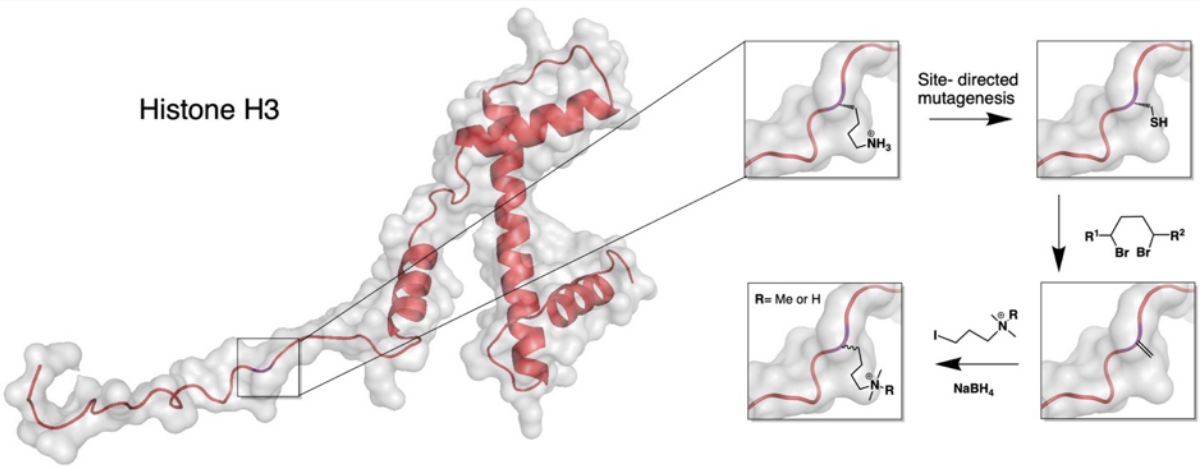


Fig. 2.2.1.1. *Strategy for generating precisely methylated H3K36me2 and me3. H3K36 is first mutated to cysteine (H3K36C) using site-directed mutagenesis, then converted to Dha using an alkylating agent (DBHDA: $R^1 = R^2 = CONH_2$ or MDBP: $R^1 = CO_2Me$, $R^2 = H$). Dha can then be converted to di- or trimethyllysine using the appropriate iodide and $NaBH_4$.*

This strategy involves first mutating the position of interest to cysteine for further reaction to Dha. Plasmids encoding the wild-type histones in the pET3d vector were kindly donated to the Davis group by Dr. Rob Klose, University of Oxford. A plasmid for expression of H3K36C in *E. coli* was generated via site-directed mutagenesis. The protein was then expressed in a BL21-DE3 (pLysS) expression system and purified by the standard method of sequential inclusion body isolation, size- exclusion and cation exchange chromatography.

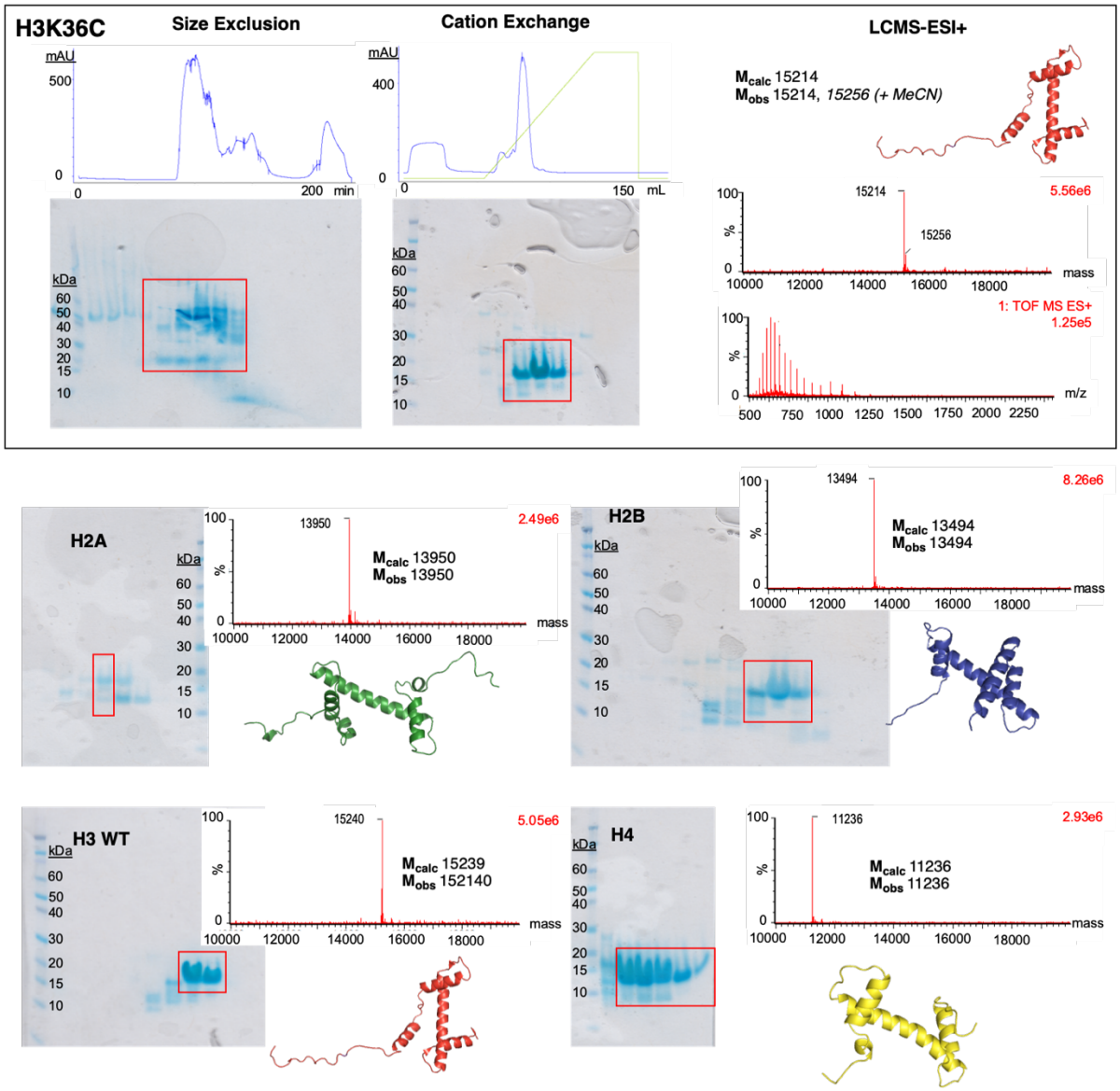


Fig. 2.2.1.2. *Expression and purification of the five histone variants. Full characterization data i.e. size exclusion and cation exchange FPLC traces, SDS-PAGE analysis, and LCMS-ESI+ ion series and deconvoluted spectrum are shown as an example for H3K36C. Final SDS-PAGE from cation exchange step and deconvoluted LCMS-ESI+ spectra are shown for all other variants; for full characterization, please see experimental section.*

Although the mutant H3 gave a noticeably lower yield of 27.5 mg/ mL (66 mg total) protein compared to 42 mg/ L (101 mg total) for the wild-type, a substantial amount was still obtained for further experiments.

2.2.2 Iodide synthesis

(3-iodopropyl)-trimethylammonium iodide required for the synthesis of H3K36me3 was synthesised using a protocol previously established within the Davis group:[63]

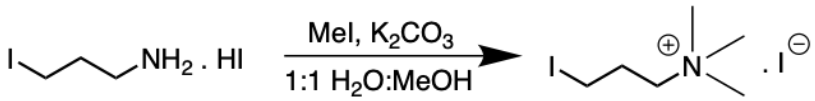


Fig. 2.2.2.1 *Synthetic route for generating (3-iodopropyl)-trimethylammonium iodide.*

It was obtained in good yield and purity and used, along with (3-iodopropyl)dimethylamine (synthesised by T. Wright) to access the methylated H3 variants.

The first step in transforming the reactive cysteine handle to methylated lysine is the conversion of cysteine to dehydroalanine (Dha). Two reagents have previously been favoured for Dha formation on histones within the Davis group; 2,5-methyldibromopentonate (MDBP) and 2,5-dibromohexane diamide (DBHDA). As the former was commercially available and had recently been used with success on H3 and H2B, it was trialled first.[30]

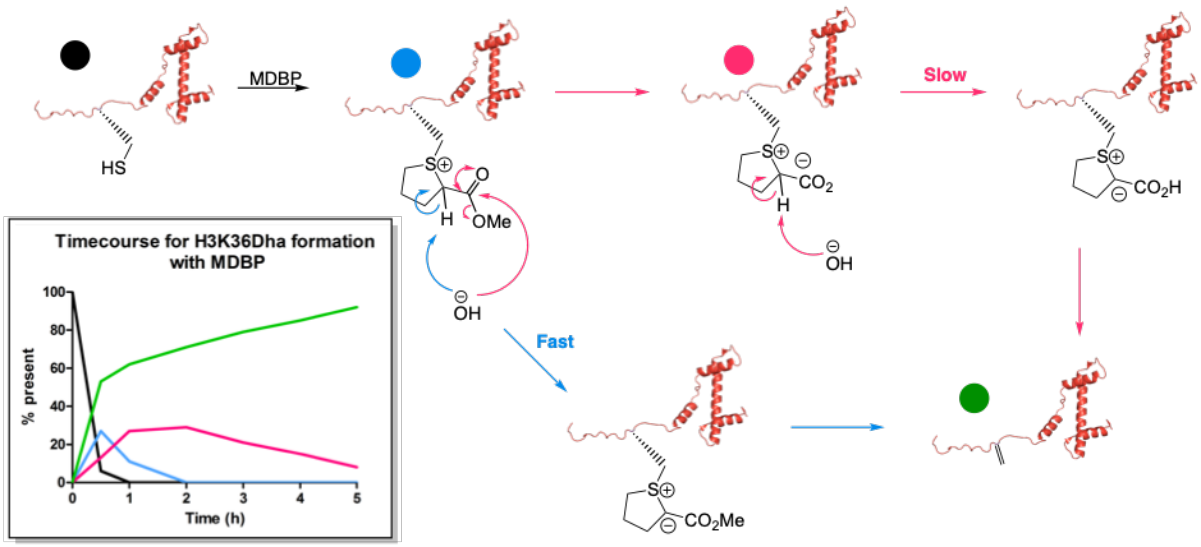


Fig. 2.2.2.2. *Dha synthesis with MDBP showing starting material H3K36C (black), methyl ester (blue) and carboxylic acid (pink) intermediates, ylid intermediates, and H3K36Dha product (green). The percentage of each species present over time is shown.*

In the reactions with MDBP, two reaction pathways were observed. The major, desired pathway was alkylation followed by rapid sulfonium ylide formation and elimination to form Dha within 1 hour. However, a second pathway was also present whereby the methyl ester hydrolysed to form a carboxylic acid, which hindered subsequent ylide, and therefore Dha, formation. Nonetheless, after 5 hours, 92% conversion to H3K36Dha was achieved. At this point the reaction is typically passed through a PD-10 desalting column once it is at or near completion to remove excess small molecule contaminants, after which it is incubated overnight under a nitrogen atmosphere to equilibrate for the anaerobic radical reaction. It was reasoned that the reaction might continue towards completion during this time.

With a protocol for Dha formation in hand, the next step was to optimise the radical reaction. Initial studies showed that using a higher concentration of $NaBH_4$ and (3-iodopropyl)-trimethylammonium iodide yielded better conversion to the desired H3K36me3 product. However, a di-addition side product was also observed. Subsequently, it was found that a large

amount of iodide (20 mg) combined with a lower amount of $NaBH_4$ (2 mg) was ideal for maximising conversion while minimising di-addition (work carried out by B.Josephson).

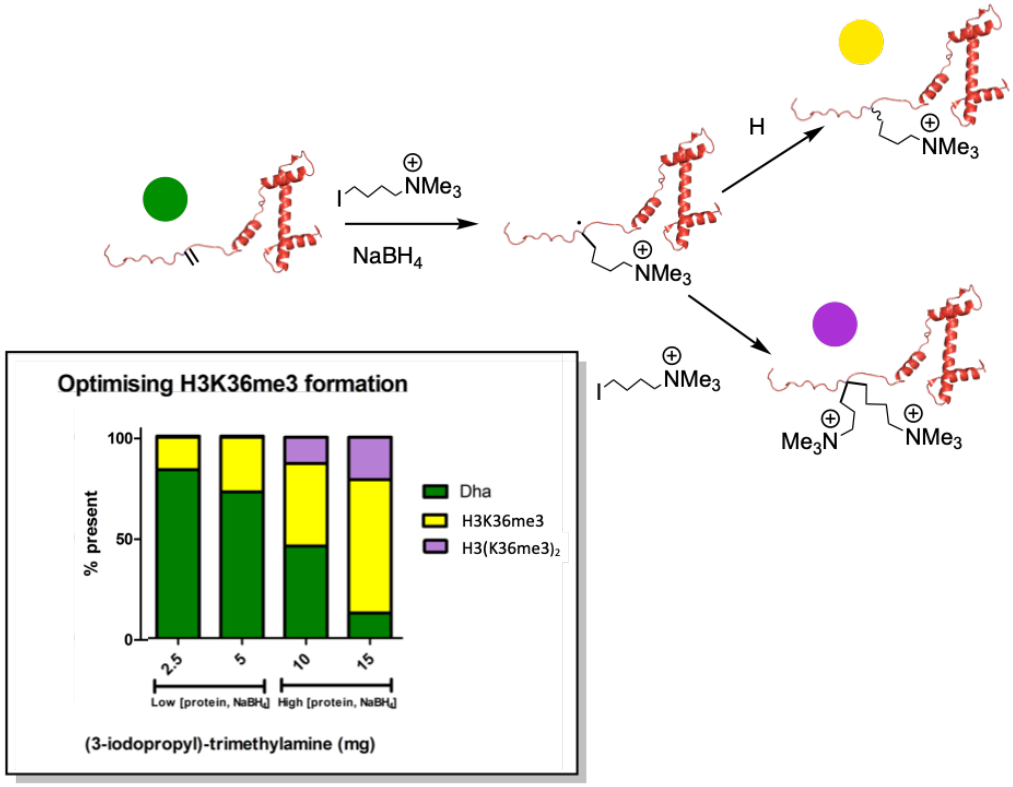


Fig. 2.2.2.3. *Optimization of the radical reaction with Dha.*

The final synthesis of the two methylated H3K36 variants, gave good yield (synthesised by B. Josephson).

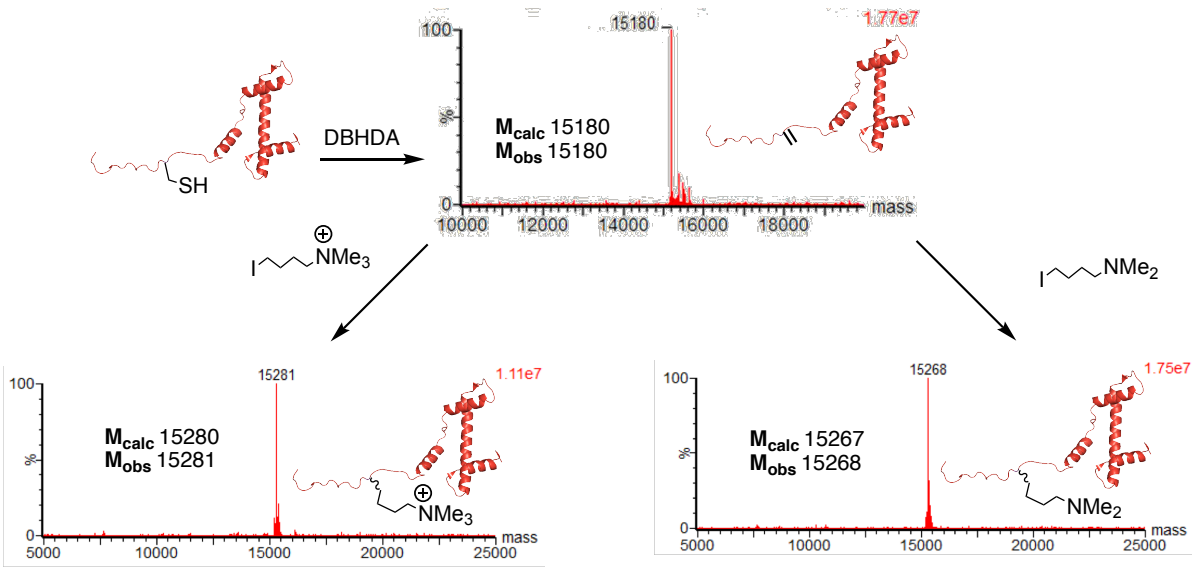


Fig. 2.2.2.4 *Final synthesis of di- and tri-methyl H3.* M_{calc}*= mass calculated from expected product structure,* M_{obs}*= mass found in deconvoluted LCMS-ESI+ spectrum.*

2.2.3 Octamer assembly

The *in-vitro* assembly of nucleosomes is a relatively well-established process, with standard literature protocols available.[65] Broadly speaking, there are two strategies for assembly. The first involves assembling the entire histone octamer separately, then introducing the DNA to form the nucleosome. In the second strategy, the $H3_2$-$H4_2$ tetramer and H2A-H2B dimers are first assembled separately, then combined with the DNA to form the DNA-bound octamer in one step.

Both of these methods have been used previously within the Davis group without significant difference in outcome.[30,66] For simplicity and minimal loss of the limiting component, the methylated H3 variants, the former strategy was chosen. Briefly, lyophilised histones were reconstituted in unfolding buffer (6 M Gdn. HCl, 20 mM Tris. HCl pH 7.5, 5 mM DTT) and mixed in an equimolar ratio. The histone mixture was then dialysed three times against

refolding buffer (2M NaCl, 10 mM Tris. HCl pH 7.5, 1 mM EDTA and 5 mM β-ME). The histones consequently refolded together to form the octamer.

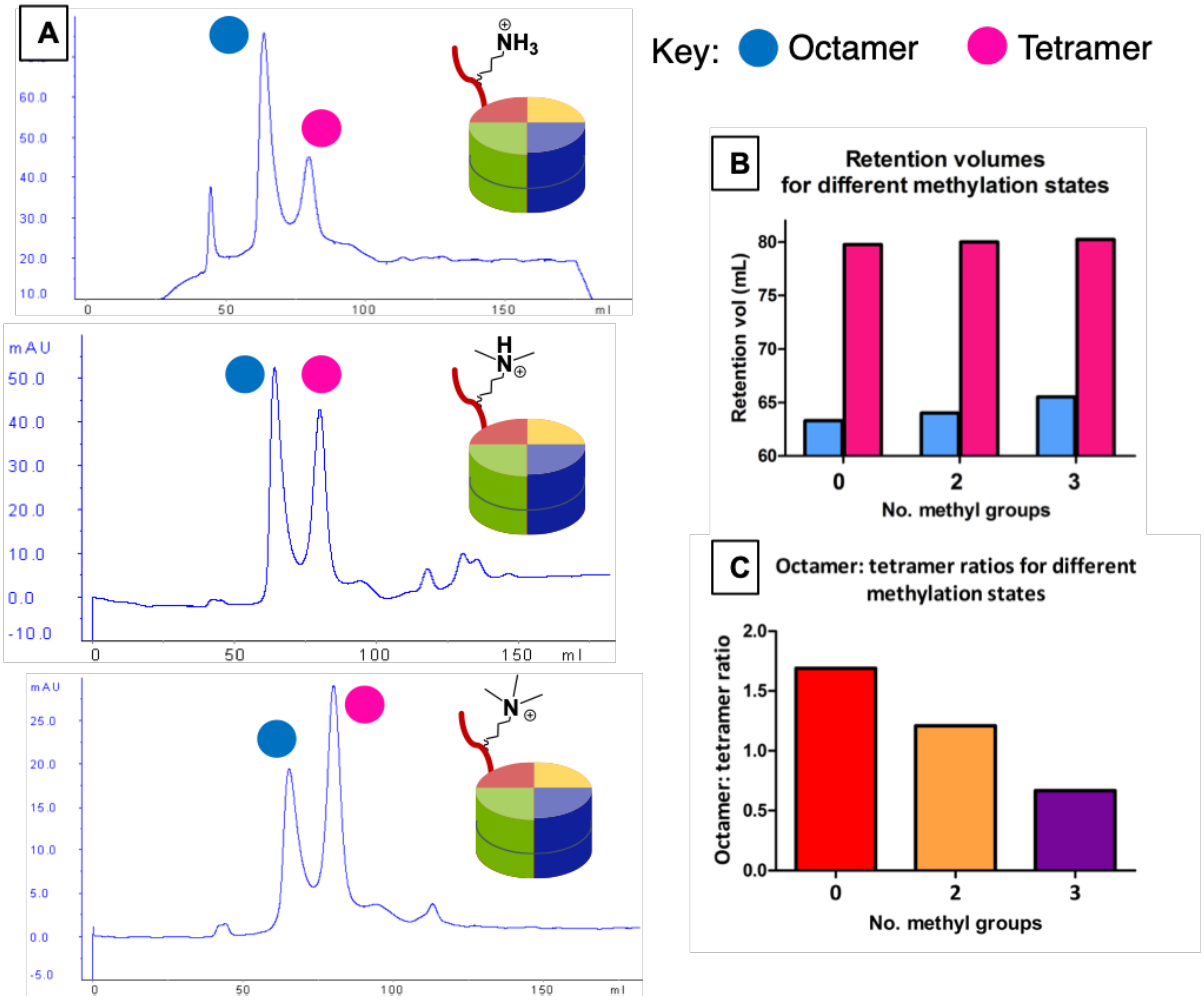


Fig. 2.2.3.1 (A) *Size exclusion traces for octamer assembly with WT H3, H3K36me2 and H3K36me3. Note that while only one H3 tail is pictured in each diagram, there are two present per octamer.* ***(B).*** *Retention volumes octamer and tetramer for different methylation states.* ***(C).*** *Ratios between octamer and tetramer peaks for each methylation state, as determined by comparing relative heights of peaks in SEC trace. .*

Intriguingly, octamer assembly revealed an apparent trend through the methylation states. As the number of methyl groups increased from 0 to 2 to 3, the ratio between the octamer and tetramer peak shifted to favour the tetramer, and the retention volume also increased.

Each methyl group added replaces a hydrogen on the K36 side chain, with -me3 representing a state where all three hydrogens have been replaced. One would also expect the addition of each methyl group to correspond to an increase in hydrophobicity and reduction of the hydration sphere around K36, as with each hydrogen lost a potential hydrogen-bonding partner to the surrounding water is lost also. As the interactions between histones are largely hydrophobic, it is proposed that increasing the hydrophobicity of the tail causes it to interfere with the interface between the $H3_2$-$H4_2$ tetramer and H2A-H2B dimers.[67] Moreover, it is hypothesized that the increase in retention time is due to the decrease in hydration around K36, and corresponding decrease in hydrodynamic radius.

An important consequence of our semi-synthesis of H3 to be considered here is the unavoidable generation of two epimers. This arises from the fact that Dha is planar, and radical attack is therefore in theory possible from either face. Indeed, previous small-molecule studies on the reaction have suggested a 50:50 mixture of the epimers results. It may be possible to obtain an enantiomeric excess if the local environment provided by the surrounding residues is chiral. However, in previous examples where the radical post-translational mutagenesis reaction with Dha has been employed on H3, this 50:50 ratio has been inferred from subsequent reactivity with antibodies and demethylase enzymes.[63]

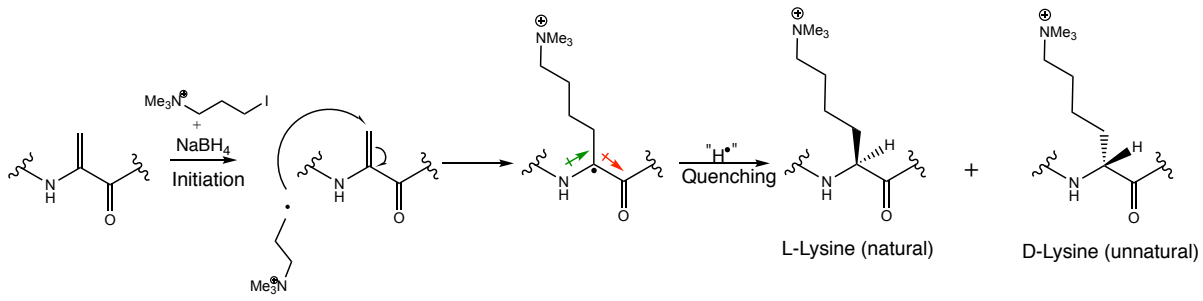


Fig. 2.2.3.2 *Scheme showing the post-translational mutagenesis radical reaction resulting in the generation of two epimers. The radical first attacks the β-carbon to generate a*

captodatively stabilized intermediate, which is then quenched by hydrogen. The reaction is not stereoselective.

It cannot therefore be conclusively determined whether the differences in octamer assembly between the WT, containing recombinantly expressed H3 with 100% of K36 present as the natural L epimer, and the synthetically generated H3K36me2 and me3 are down to differences in methylation state at K36 or the effect of the different epimers.

However, several inferences can be made to inform the interpretation of the data. First, as K36 is present on the tail of H3, it would not be expected to significantly impact folding of the core octamer. It is also worth noting that in previous studies where H3K9me3 synthesised via the same Dha-radical reaction was used for octamer assembly, the ratio of octamer: tetramer was much greater and resembled that for the WT in these studies.[63] Moreover, in studies with H2A-Thr101GlcNAc synthesised via Dha, using different ratios of pre-assembled H2A-H2B dimer and $H3_2$-$H4_2$ tetramer to compensate for the effect of the different epimers did not appear to have a significant effect on the outcome.[30]

In this study, if the unnatural epimer at position 36 was unable to be incorporated, the octamer: tetramer ratio might be predicted to be approximately half of that for the WT. However, this is not the case; the octamer: tetramer ratio with H3K36me2 is approximately 75% of that with WT H3. This suggests that at least some of the unnatural epimer is being incorporated. In contrast, both H3K36me2 and H3K36me3 are expected to be present in an approximately 50:50 ratio between epimers. This means that differences between the two are likely to arise from the differences in methylation state.

2.2.4 Nucleosome assembly with 145 bp DNA

With the octamer in hand, nucleosomes were first reconstituted using a 145 bp dsDNA fragment containing the strong positioning '601' sequence (prepared by S. Nadal, Ben Davis group)[68,69]. Nucleosome reconstitution with 145 bp DNA is again a well-established protocol; briefly, octamer and DNA were mixed and dialysed sequentially into buffers with increasingly low concentrations of KCl (Tris. HCl 10 mM pH 7.5, ETDA 1 mM, DTT 1 mM and [KCl]= 2 M, 0.85 M, 0.65 M and 0.25 M). In initial tests with the WT octamer, different DNA: octamer ratios were screened to determine optimal conditions to use for the methylated variants. Similarly, the inclusion of a heat shifting step after assembly was also trialled at 37°C and 55°C. Heating the nucleosomes allowed the DNA to reposition on the histone octamer, so that those trapped in a locally stable conformation could assume the globally most stable conformation. This also ensured a more uniform nucleosome sample. The best conditions overall were found to be 1:1 ratio of DNA: octamer with a 55°C heat shifting step.

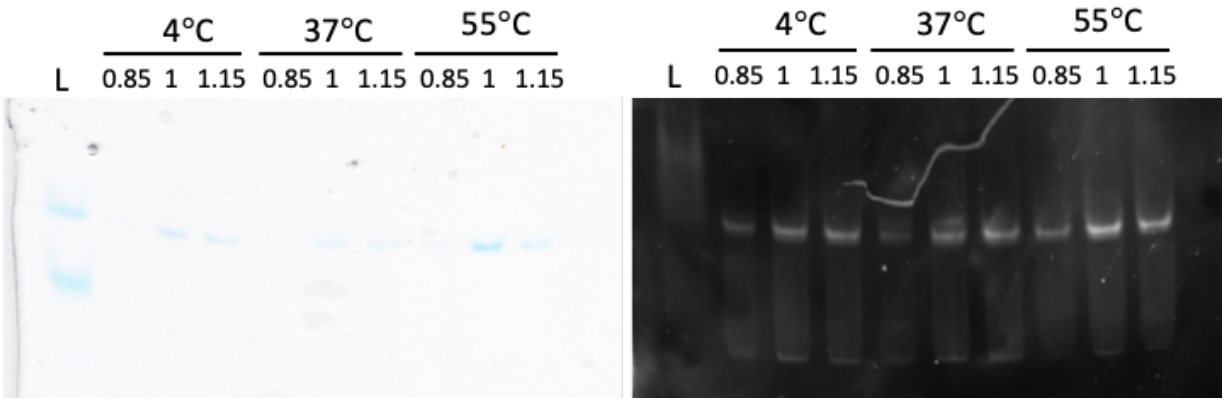


Fig. 2.2.4.1 *6% TBE gel of WT 145 bp nucleosome assembled under different conditions stained with Coomassie (left) and SYBR gold (right). Heat shift temperature and DNA: octamer ratios for each sample are labelled. L= NativeMark unstained protein standard (Thermo Fisher Scientific). See SI for uncropped scans.*

These conditions were then successfully used to repeat reconstitution of the WT nucleosome, as well as to generate H3K36me2 and -me3 containing nucleosome for further experiments. In general, the bands for methylated octamers appeared more diffuse.

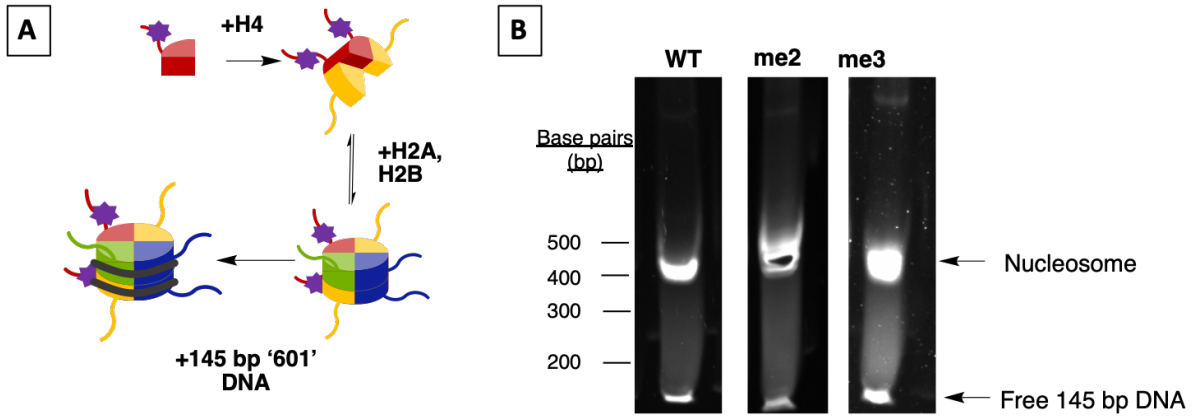


Fig. 2.2.4.2 *(A) Schematic for nucleosome assembly. H3K36 is indicated with a purple star. (B) Assembly of WT, me2 and me3 nucleosomes as shown by 6% TBE gel stained with SYBR gold.*

2.2.5 Nucleosome assembly with 186 bp DNA

186 bp DNA bearing two biotin tags at the 5' end was synthesised according to protocols previously used within the Davis group. A plasmid bearing 16 repeats of a 186 bp fragment containing the strong positioning 601 sequence and flanked by EcoRV and EcoRI sequences (kindly donated to the Ben Davis group by Prof. Till Bartke, Imperial College, London) was expressed on a large scale. The plasmid was then purified and the 186 bp fragments excised using the appropriate restriction enzymes, after which the fragment was isolated from the vector by careful ethanol precipitation. Importantly, the use of EcoRI left a 5' overhang which could then be recognised by the Klenow 3' to 5' exo fragment to append two molecules of biotin-11-dUTP. The biotinylated DNA was then purified by chloroform-phenol extraction to remove protein. The presence of the biotin tag was confirmed by Southern blot.

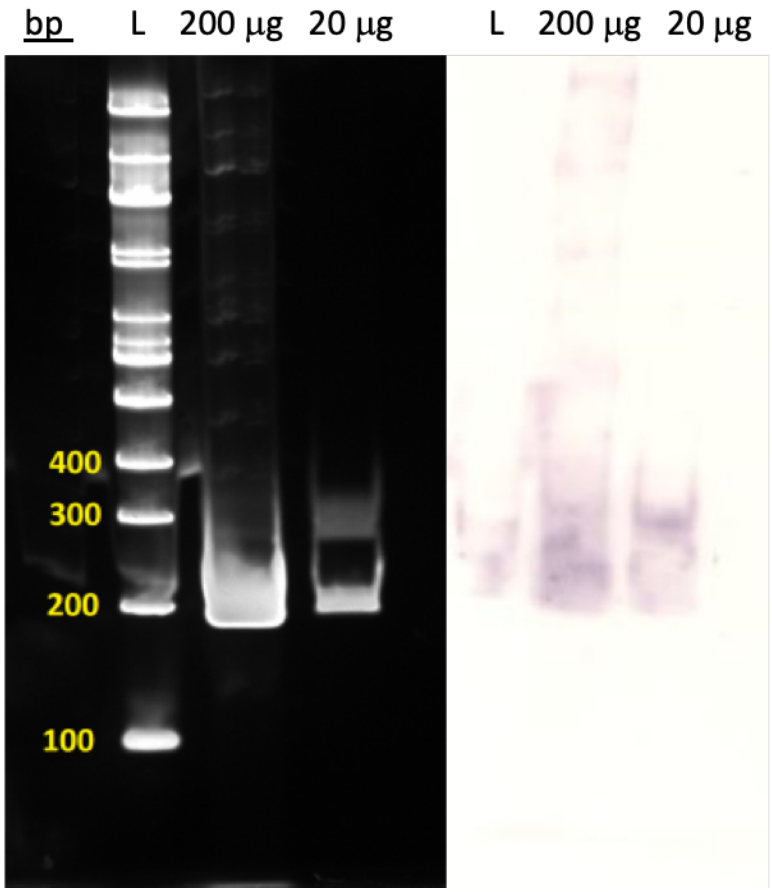


Fig. 2.2.5.1 *(Left) 6% TBE gel of 186 bp DNA stained with SYBR gold, with NEB 100 bp ladder. (Right) Southern blot detection of biotinylated bands. Labels denote amount of DNA loaded. Single-stranded 186 bp DNA can be visualised just below 200 bp, while double-stranded 186 bp DNA can be visualised between 300 and 400 bp.*

Significantly for future experiments, the 186 bp biotinylated DNA contained some degradation product, as shown by a diffuse lower molecular-weight band on the gel. An attempt was made to remove the contaminants by repeating the final purification step, but this exacerbated the degradation, most likely due to the new exposure to environmental DNAses caused by handling.

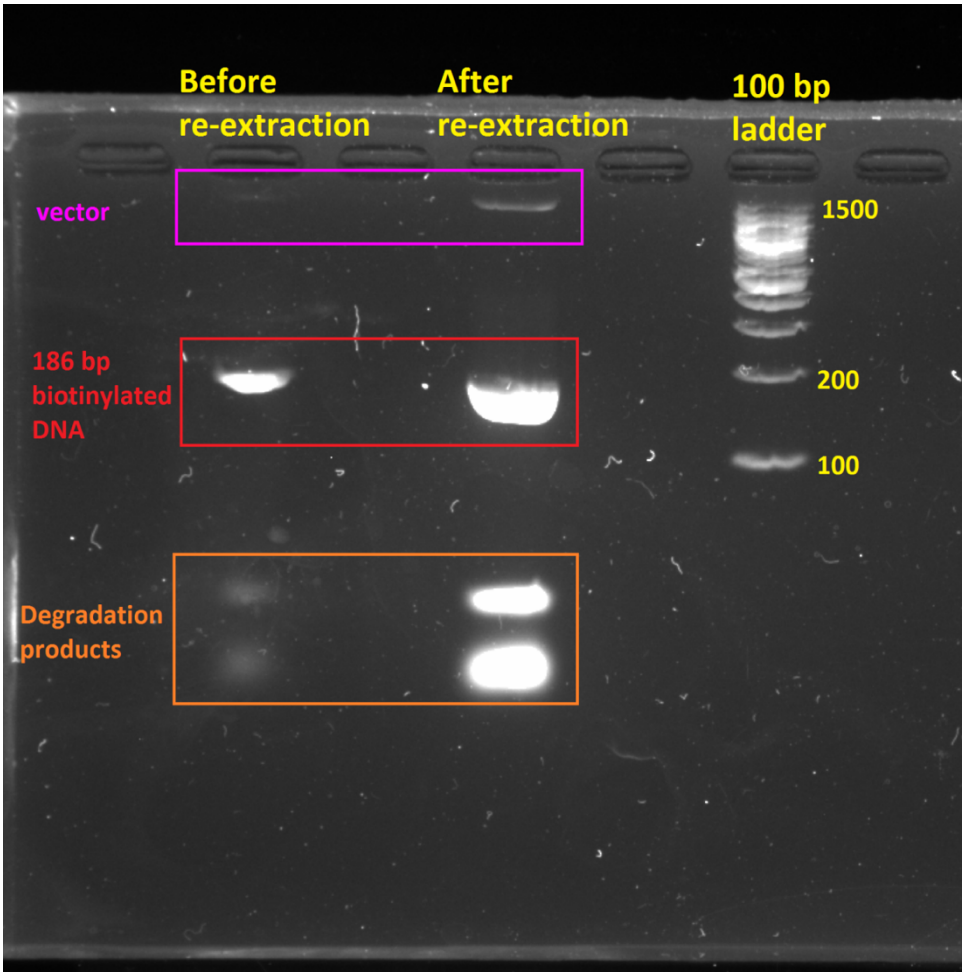


Fig. 2.2.5.3. *Agarose gel electrophoresis analysis of 186 bp biotinylated DNA before and after re-extraction.*

With the difficulty of isolating the DNA in a DNAse-free environment in mind, nucleosome assembly was carried forward under the assumption that the degraded DNA fragments would not interfere. As the DNA concentration used to determine the volume needed for a given DNA: octamer ratio was measured by nanodrop, which measures the absorbance of purine and pyrimidine rings and hence does not discriminate between free nucleotide contaminants and our 186 bp fragment, the measured 186 bp DNA concentration was taken to be higher than the actual concentration. To compensate, increasing ratios of DNA: octamer for reconstitution with WT octamer were screened. Ratios between 2:1 and 3:1 were found to be successful; the

intensity of the resulting nucleosome band increased with the ratio, with the caveat that the intensity of the free DNA band also increased.

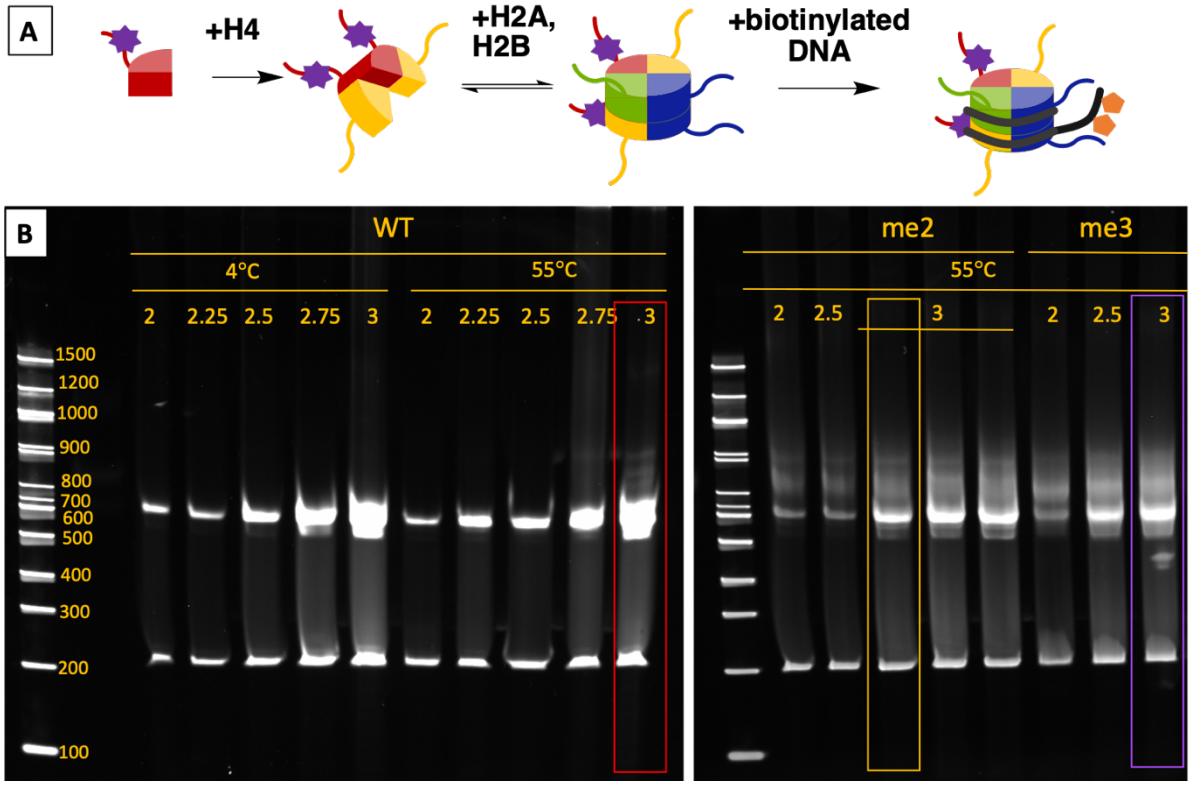


Fig. 2.2.5.4 *(A) Overview of nucleosome formation with 186 bp DNA. Histones are coloured red (H3), yellow (H4), green (H2A) and blue (H2B), modification at K36 is indicated by a purple star and biotin is indicated by orange pentagons. (B) 186 bp DNA nucleosomes visualised on 6% TBE gel, SYBR gold stain. Temperatures indicate 2h incubation at 55 ℃ (heat shift) or 4 ℃ (no heat shift) after assembly. Numbers represent ratio of DNA: octamer used in assembly (where octamer=1). Samples used in subsequent pulldown experiments are highlighted in red (WT), yellow (H3K36me2) and purple (H3K36me3).*

During the 186 bp octamer reconstitution it was also noted that, unlike with the more robust 145 bp nucleosome assembly, starting with an octamer concentration over 1 mg/ mL was essential.

With conditions for WT octamer reconstitution in hand, assembly of the H3K36me2 and -me3 octamers was then carried out. In general, bands for H3K36me2 and -me3 -containing nucleosomes appeared more diffuse, suggesting less uniform formation. This might be expected given the trend observed across octamers in Fig. 2.2.3.1, where the stability of the octamer relative to tetramer decreased with methylation state. However, analysis of the components by denaturing SDS-PAGE revealed all contained the expected histones (Fig. 2.2.5.5.)



Fig. 2.2.5.5. *Denaturing SDS-PAGE analysis of nucleosomes. L= GeneFlow BluEye pre-stained protein standard.*

2.3 Conclusions

A common mode for regulating the accessibility of chromatin is by histone PTMs. For some PTMs such as acetylation, the chromatin structure is directly modulated by altering the strength of interaction between DNA and histone octamer. However, other PTMs effect change in the nucleus by recruiting specific reader proteins which can then form multimeric complexes to carry out a range of functions. Perhaps the most well characterised PTM in this latter category is lysine methylation.

Each lysine can bear up to three methyl groups, and in some cases each methylation state of a single residue has a unique function. A striking example is H3K36, which has a dedicated methyltransferase enzyme SETD2 solely responsible for converting H3K36me2 to H3K36me3. Adding a single group to H3K36me2 does not significantly change its charge or add much steric bulk, but hampering this transition causes disfunction in a range of core biological processes. The mechanism by which the two methylation states are differentiated is currently unknown, and was investigated in this work.

As previous studies on peptides or free H3 variants have failed to recapitulate any such discrimination by readers, it was hypothesised that the difference might arise on the scale of the entire nucleosome. To this end, synthetic H3K36me2 and H3K36me3 variants were constructed using the 'post-translational mutagenesis' technique, and used to first assemble octamers, then nucleosomes. In the octamer assembly, a striking difference in the ratio between octamer and tetramer peaks in SEC purification was observed, suggesting a destabilization of the interaction between $H3_2$-$H4_2$ tetramer and H2A-H2B dimers.

This was further reflected in 186 bp nucleosome assembly, where the methylated nucleosomes gave streaked bands in the 6% TBE gel analysis, suggesting a less uniform assembly.

These observations suggest that methylation at H3K36 affects the properties of the nucleosome as a whole. To gain more insight into a potential mechanism and investigate whether this could form the basis for selectivity of reader proteins, binding of these synthetic

nucleosomes with the reader MRG15 was subsequently studied, and will be described in Chapter 3.

2.4 Experimental

2.4.1 General experimental considerations

Chemicals and solvents were purchased from Acros UK, Alfa Aesar UK, Carbosynth, Fischer UK or Sigma-Aldrich UK and used as delivered unless stated otherwise. Thin layer chromatography (TLC) was carried out using plastic 0.20 mm Polygram® Sil g/UV254 plates that were dried under a stream of nitrogen and visualized under UV (λ_{max} 254 nm), or by use of anisaldhyde, anisidinium phthalate, sulfuric acid or ninhydrin dip. Melting points were recorded using a Leica hotstage microscope and are uncorrected. Flash column chromatography was performed using Geduran® Si 60.8.2 or a Biotage Flash Purification System with a Kinesis Telos or Biotage Snap columns.

NMR Spectroscopy General

Deuterated solvents were used as the lock, and the residual protonated solvent as the internal reference peak. Spectra were analysed using MestReNova. Correlation spectroscopy (COSY) and heteronuclear single quantum correlation (HSQC) were used when necessary to assign peaks. Samples were prepared by dissolving approx. 20 mg in 0.5 mL of solvent. The following numbering system was used in assignments. Other positions in the chain (e.g. α and β) not explicitly indicated here are labelled as standard.

^{1}H NMR Spectroscopy

Chemical shift information (δH) is given in parts per million (ppm) relative to the centre of the solvent signal ($CDCl_3$ 7.26 ppm, D_2O 4.79 ppm DMSO-d6 2.50 ppm, or MeOH-d4 3.31 ppm). Spectra were recorded on a Bruker AVIIIHD 400 nanobays (400.1 MHz) or AVII 500 (500.1 MHz). The number of protons for a resonance signal is given by nH based on integral values. Multiplicity is denoted by the bracketed symbol. Coupling constants (J) are reported to the nearest 0.1 Hz and are determined using MestReNova.

^{13}C NMR Spectroscopy

Chemical shift information (δC) is given in parts per million (ppm) relative to the centre of the solvent signal ($CDCl_3$ 77.16 ppm, DMSO-d6 39.52 ppm, MeOH-d4 49.00 ppm). Spectra were recorded on a Bruker AVIIIHD 400 nanobays (100.6 MHz) or AVII 500 (125.7 MHz). Unless otherwise stated, signals are assumed to be singlets. Distortionless enhancements by polarisation transfer with retention of quaternaries (DEPTQ) pulse sequence with broadband proton decoupling were used to record spectra.

Small Molecule Mass Spectrometry

Low-resolution mass spectra were recorded on an Agilent 6120 Quadrupole spectrometer using electrospray ionization (ESI). Positive and/ or negative mode results are given as appropriate. High-resolution mass spectra were recorded on a Bruker MicroTOF ESI mass spectrometer.

Peptide mass spectrometry

Liquid-chromatography-mass spectrometry (LC-MS) was performed using:

1. a) Waters LCT Classic coupled to a Shimadzu 20 Series HPLC using a Thermo Proswift (250 x 4.6 mm x 5 μm) column. The gradient was programmed as follows: 95% A (5 min isocratic) to 100% B after 15 min then isocratic for 3 min.

   b) Waters LCT Premier XE coupled to a Waters 1525 Micro HPLC using a Merck Chromolith C18 guard cartridge. Water (solvent A) and acetonitrile (solvent B), both containing 0.1% formic acid, were used as mobile phase at a variable flow rate. The gradient program was: 95% A (1 min isocratic, 0.4 ml/min) to 100% B after 5 min (0.4 ml/min) then isocratic for 3 min (flow rate increasing to 1.0 ml/min). The spectra were calibrated automatically. The electrospray source parameters were: capillary voltage: 3000 V, cone voltage: 160 V. Nitrogen was used as the desolvation gas and nebuliser at a total flow of 600 l/hr in both spectrometers.

2.4.2 Site-directed mutagenesis to generate H3K36C and H4

The H4 WT plasmid was unavailable, but an H4 R17C mutant was taken from Davis group plasmid bank; mutagenesis was carried out to revert this mutation back to WT.
All manipulations were carried out in a laminar flow hood. Mutagenesis was carried out using a QuikChange II XL Site-Directed Mutagenesis kit following the manufacturer's protocol with modifications.

Mutant		**Sequence**
H3K36C	F	5'-ctgctaccggcggagtctgcaaacctcaccgttaccg-3'
	R	5'-cggtaacggtgaggtttgcagactccgccggtagcag-3'
H4 C17R	F	5'-tctgggtaaaggtggtgctaaacgccaccgtaaag-3'
	R	5'-ctttacggtggcgtttagcaccacctttacccaga-3'

Table 2.4.2.1. *Primer sequences*

Primers were designed using the online Agilent Quikchange primer design tool.

For PCR to generate the mutant, the following reagents were combined on ice:

Reagent	**Quantity**
10x reaction buffer	5µL
Quick Solution	3µL
Plasmid	5 µL of 2 ng/µL stock
Forward primer	5 µL of 25 ng/µL stock
Reverse primer	5 µL of 25 ng/µL stock
dNTP mix	1µL
MilliQ H_2O	25 µL
PfuUltra HF DNA polymerase	1µL of 2.5 U/µL

Table 2.4.2.2 *Composition of Quikchange PCR reactions*

PCR was then carried out using the following cycling parameters:

Segment	**Cycles**	**Temperature (°C)**	**Time**
1	1	95	1 min
2	20	95	50 s
		60	50 s
		68	7 min
3	1	68	7 min

Table 2.4.2.3*. PCR cycling parameters for Quikchange PCR*

For the digest of parent DNA, *Dpn* I restriction enzyme (1 µL) was added directly to each reaction; the reactions were mixed by vortex, and incubated at 37°C for 2 hours.

For transformation into competent cells, XL-10 gold cells were thawed on ice and aliquoted (50 µL per reaction) into 14 mL Falcon polypropylene round-bottom tubes containing β-ME

mix (2 µL). The cells were incubated on ice for 10 minutes, before the *Dpn* I-treated PCR product (2 µL) was added and the cells were allowed to incubate for a further 30 mins. The cells were then heat-shocked at 42°C for 42 seconds before incubating for 2 minutes on ice to allow recovery, followed by addition of SOC media (500 µL) and incubation at 37°C for 1 hour. The cells were then spread on LBA plates (150 µL on LB agar supplemented with 100 µg/ mL ampicillin) and incubated at 37°C for approximately 20 hours.

For each of the two mutants, four colonies were cultured for miniprep and sequencing. LB broth (10 mL, agar supplemented with 100 µg/ mL ampicillin and 0.8% glucose) was inoculated with a single colony from the previous transformation. Cultures were incubated for 16 hours before pellets were harvested by centrifugation (8000 rpm for 10 minutes). DNA was then extracted using a QIAprep Spin Miniprep kit according to manufacturer's protocol with some modifications; the final elution step was carried out using a reduced volume of MilliQ H_2O (40µL) with incubation for an extended period of time (10 minutes) before centrifugation to increase final concentration and yield. Plasmid concentration was determined by nanodrop. Samples of each (8 µL of approx. 100 ng/µL) were submitted for Sanger Sequencing using T7F primers. DNA sequence alignments and translation were carried out using SerialCloner; protein sequence alignments were carried out using the MUSCLE server. For each mutant, a plasmid with the desired sequence was successfully identified; see appendix. **2.6.1.3** for full sequencing data.

H3K36C translated protein sequence
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVCKPHRYRPGTVALREIRRY
QKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAI
HAKRVTIMPKDIQLARRIRGERA*

H4 WT (H4C17R) translated protein sequence

MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRG
VLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG*

2.4.3 Histone expression

2.4.3.1 Transformation into BL21 (DE3) pLysS cells

Plasmid (0.5 µL) and BL21 (DE3) pLysS cells (10 µL) were incubated together on ice for 30 mins. The cells were then heat-shocked at 42°C for 42 seconds before incubating for 2 minutes on ice to allow recovery, followed by addition of SOC media (300 µL) and incubation at 37°C for 1 hour. The cells were then spread on LBA plates (75 µL on LB agar supplemented with 100 µg/ mL ampicillin) and incubated at 37°C for approximately 20 hours.

2.4.3.2 Small-scale cultures

H3 WT, H2A and H2B

For each histone, 4x 10 mL aliquots of 2 XYT media containing 100 µg/ mL ampicillin and 34 µg/ mL chloramphenicol were inoculated with colonies from the transformation. These were incubated at 37 °C with shaking overnight. Note that only three starter cultures are needed for each histone; the fourth is added as a backup in case one fails to grow.

H4 WT and H3K36C

Protocol was as above, but the 10 mL cultures were inoculated in the morning and incubated at 37 °C for 6 hours before adding to 800 mL media for large-scale culture. Note that for most of the starter cultures no turbidity was observed after 6 hours, but the large-scale cultures grew to OD more rapidly than when overnight starter cultures were used (2-3 hours instead of 4-5 hours). This is attributed to each stage of inoculation (plate to starter culture, then starter culture to large scale) occurring when more cells are in the log phase of growth.

2.4.3.3 Large-scale cultures

For each histone, 3 x 800 mL 2 XYT media containing 100 µg/ mL ampicillin and 34 µg/ mL chloramphenicol were inoculated with 10 mL starter culture. These were grown to OD_{600}= 0.4 to 0.6 at 37 °C with shaking, at which point protein production was induced with addition of 0.5 mM IPTG (400 µL of 1M stock). They were then incubated for a further 3 hours (4 hours for H4) before cells were split into 2 x 400 mL in 500 mL centrifuge bottles and pelleted by centrifugation (F500 rotor, 8000 rpm, 10 minutes at 4 °C). Each pellet was resuspended in 17.5 mL of lysis buffer (50 mM Tris. HCl pH 7.5, 100 mM NaCl, 1 mM EDTA and 5 mM β-ME) supplemented with c0mplete protease inhibitor tablets (EDTA free, one per 50 mL buffer). Note: 1 L buffer was made up and β-ME and protease inhibitor tablets were added to 250 mL just prior to initial use. Remaining buffer was stored at 4 °C. Cell suspensions were then transferred to 3x 50 mL falcon tubes, frozen on dry ice, and stored at -80 °C.

2.4.4 Histone purification

2,4,4,1 Inclusion body purification

Note that several histones were taken through this protocol in pairs simultaneously; H3K36C with H2B, and H2A with H4. H3 WT was taken through individually as an initial test.

Cell suspension was thawed in Buchi water bath at either room temperature or 37 °C. Approx. 5 mg of DNAse I was then added, and the cell suspension incubated at 4 °C for 1 hour with gentle rocking. The three cell suspensions were then combined in a glass beaker and lysed by sonication (0.5 s on at 60% amplitude, 0.5 s off for a period of 30s followed by 1 min off; this 1:30 cycle was then repeated 4x for a total of 5 sonication periods). This was then split between 4 x 40 mL centrifuge tubes (30 mL per tube) and centrifuged to pellet solids (20,000 rpm, 4 °C, 20 mins, JA 25.50 rotor). The supernatant was then discarded; the four pellets were

suspended in TW buffer (lysis buffer with 1% Triton X-1000) by pipette (10 mL serological) and combined in a 100 mL beaker. The suspensions were then subjected to the same sonication procedure as previous, divided in 2 x 20 mL between two centrifuge tubes, and solids pelleted by centrifugation (20,000 rpm, 4 °C, 10 mins, JA 25.50 rotor). The pellet was once again resuspended in TW buffer (20 mL per tube), centrifuged (20,000 rpm, 4 °C, 10 mins). This procedure was repeated once more with TW buffer (total of 3 washes) and once more with lysis buffer. 0.2 μL DMSO was then added to each pellet, which was minced into a thick paste with a spatula and left for 10 minutes. 7.5 mL unfolding buffer was then added to each tube, and incubated with rotation at room temperature for 1 hour. Solids were then pelleted by centrifugation (20,000 rpm; 10 mins 20 °C or 15 mins 4 °C). The supernatant was then decanted, filtered (0.45 μM), and either stored overnight at 4 °C or loaded directly onto size-exclusion column for purification.

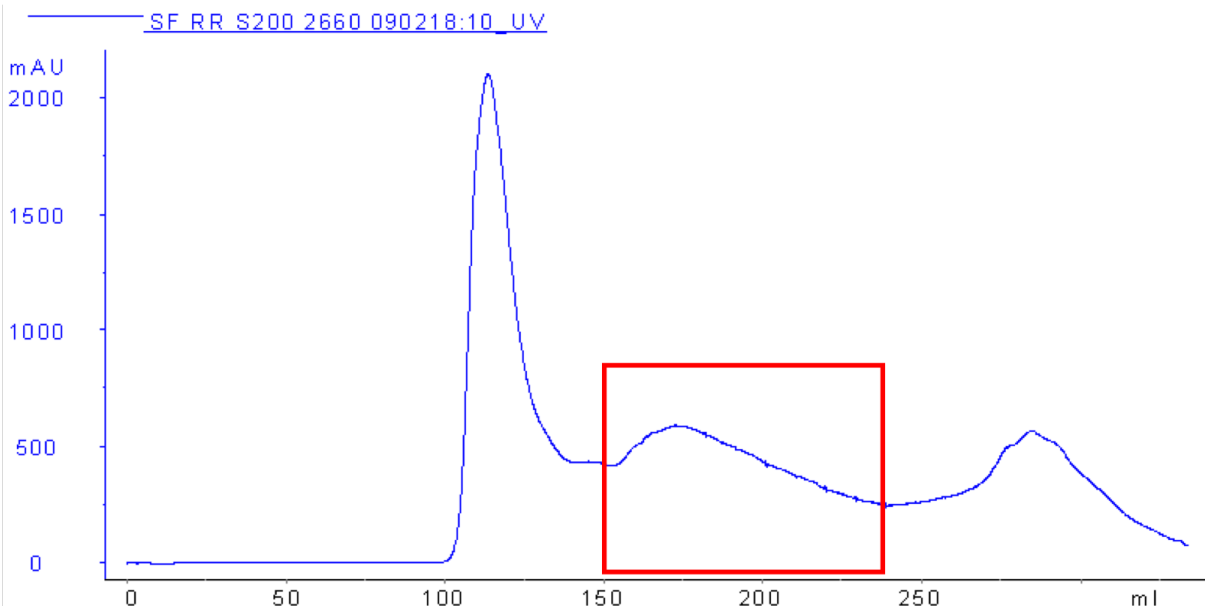
2.4.4.2 Size exclusion purification

A size-exclusion column (HiLoad 26/60 Superdex 200, GE Healthcare Life Sciences) was equilibrated with SAU-1000 buffer (7M Urea, 20 mM NaOAc pH 5.2, 1M NaCl, 1 mM EDTA, 5 mM β-ME) using equilibration protocol RR S2660 wash (buffer B= MilliQ, Buffer A= SAU-1000 when starting from column stored in ethanol). The solubilized inclusion body proteins from 2.4.4.1 were then loaded using a SuperLoop (50 mL, GE Healthcare Life Sciences) and eluted in SAU-1000 buffer. Initially the elution program RR 2660 (0.35 C.V. elute before fractionation begins) was used for H3WT; this was altered to SF RR 2660 (0.28 C.V. elute before fractionation begins) as proteins began to elute earlier on the newer column used. When purifying two histones in parallel, the first was typically left to run overnight and the second was loaded directly on the next morning i.e. no further equilibration step was needed.

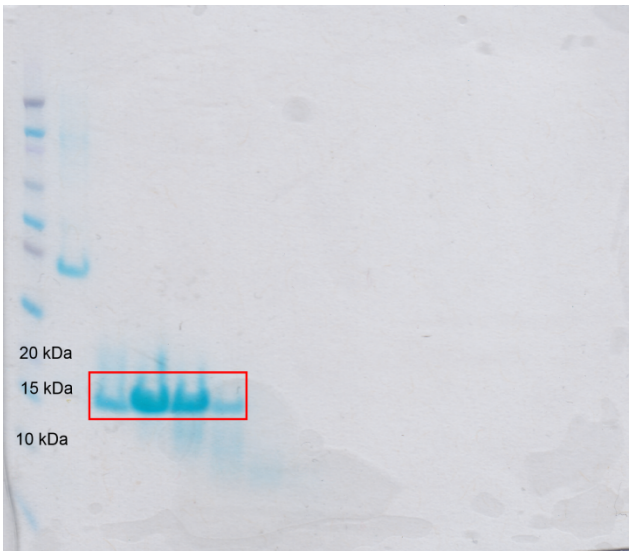
Fractions corresponding to peaks in the UV trace were analysed by SDS-PAGE and all histone-containing fractions were combined and dialysed against IEX binding buffer (7M Urea, 10 mM Tris.HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 5 mM β-ME) using 3.5 kDa dialysis casettes.

H3 WT

UV trace from SEC



SDS-PAGE

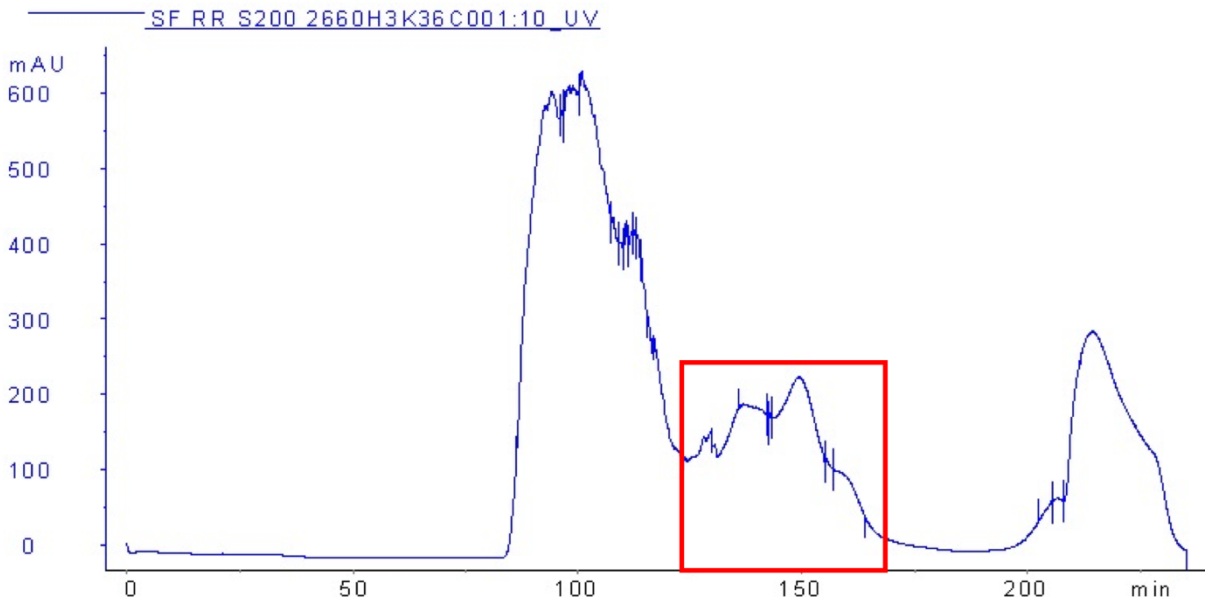


Fractions highlighted in red correspond to those containing H3 WT

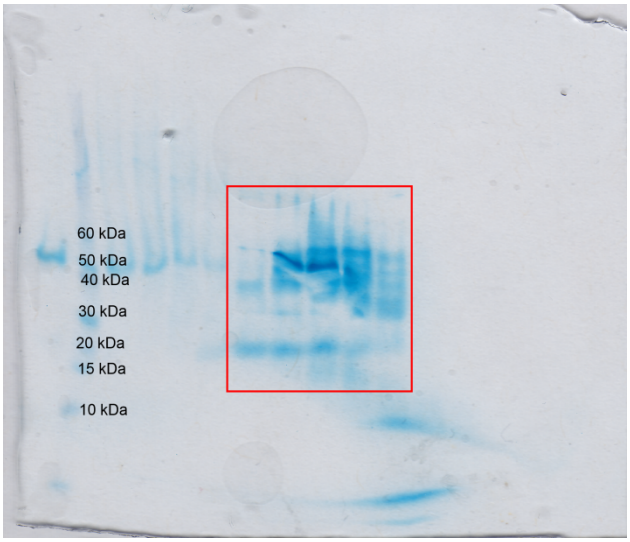
Fig. 2.4.4.2.1 *Characterization of H3 WT after SEC purification*

H3K36C

UV trace from SEC



SDS-PAGE

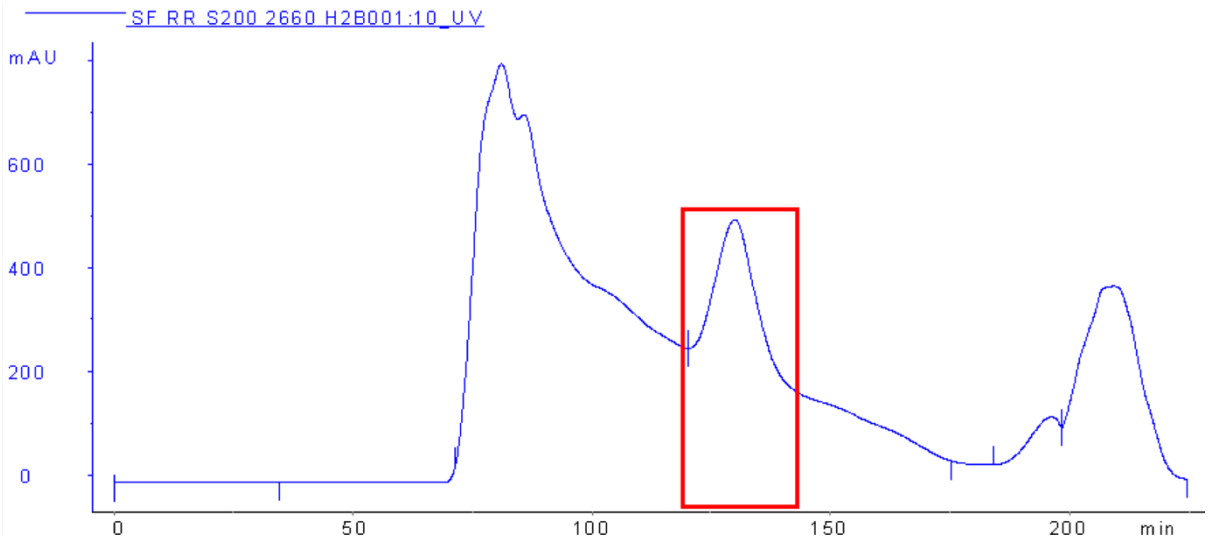


Fractions highlighted in red correspond to those containing H3K36C

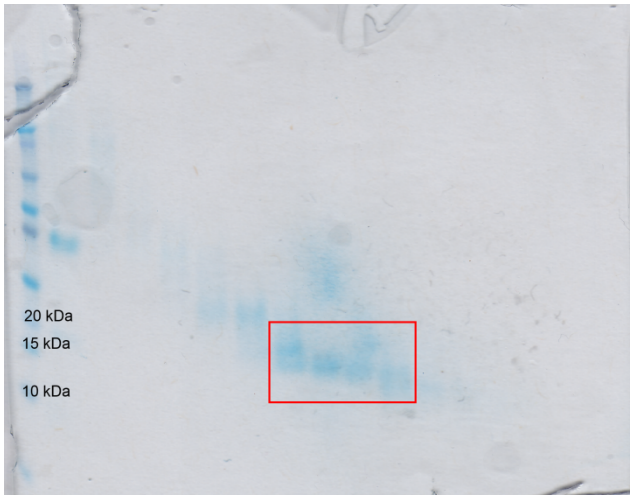
Fig. 2.4.4.2.2 *Characterization of H3K36C after SEC purification. From fig. 2.2.1.2.*

H2B

UV trace from SEC



SDS-PAGE

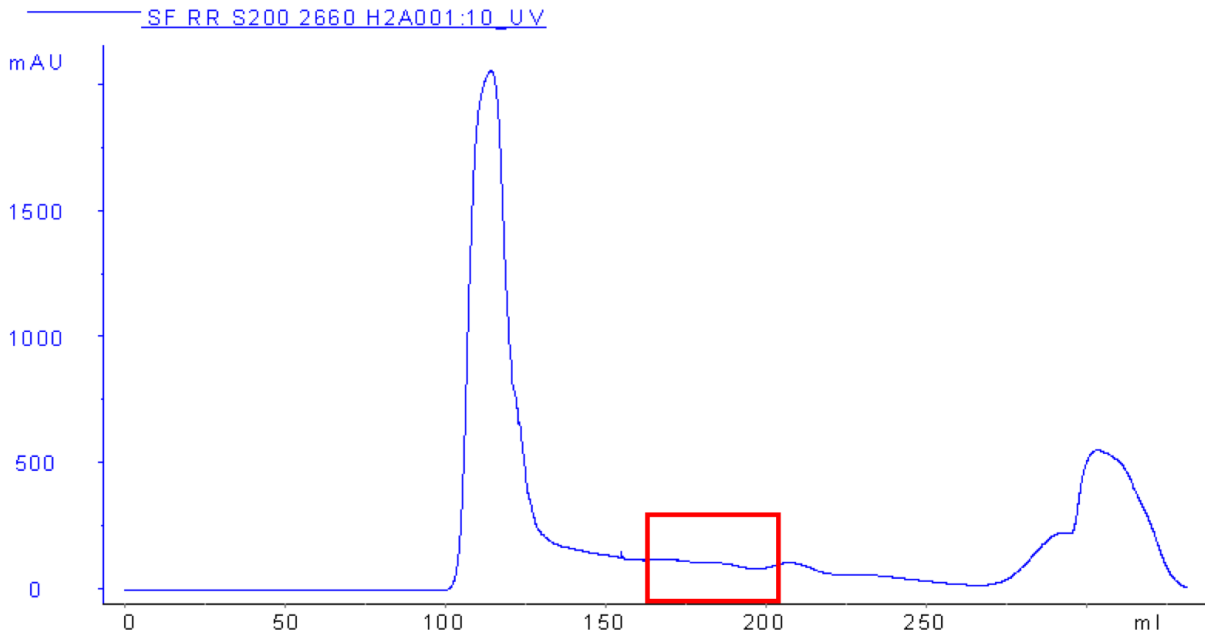


Fractions highlighted in red correspond to those containing H2B

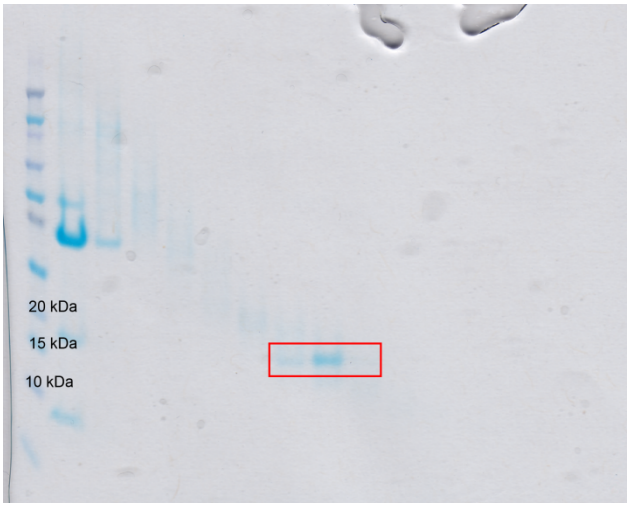
Fig. 2.4.4.2.3 *Characterization of H2B after SEC purification*

H2A

UV trace from SEC



SDS-PAGE



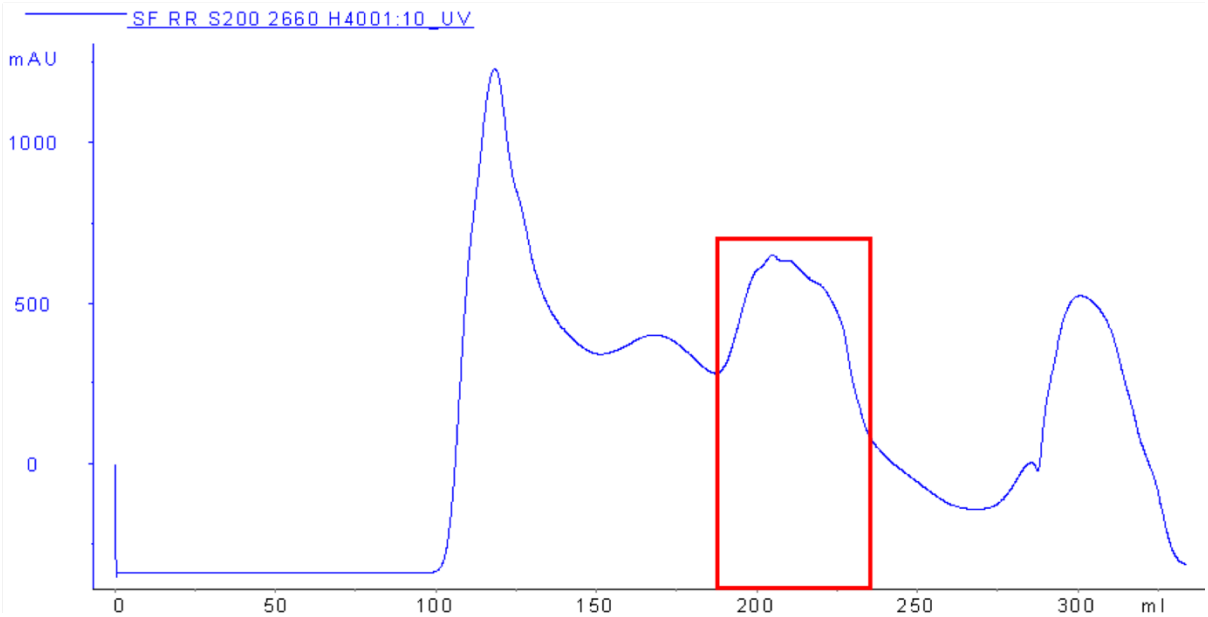
Fractions highlighted in red correspond to those containing H2A.

Note: H2A yield was unusually low. This was because FPLC was left to run overnight, but stopped before H2A had eluted; as a result, H2A had time to diffuse through the column and was lost.

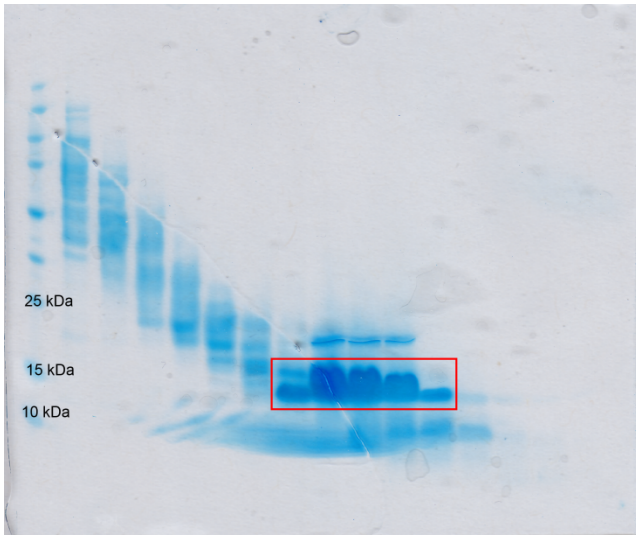
Fig. 2.4.4.2.4 *Characterization of H2A after SEC purification*

H4

SEC UV trace



SDS-PAGE



Fractions highlighted in red correspond to those containing H4

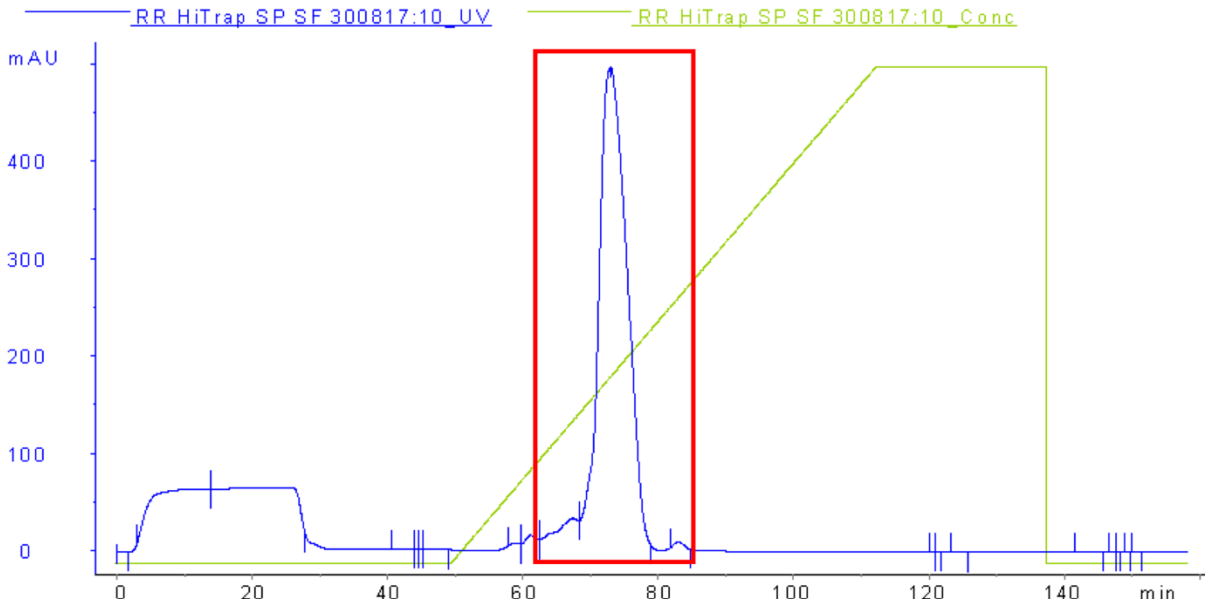
Fig. 2.4.4.2.5 *Characterization of H4 after SEC purification*

2.4.4.3 Cation exchange purification.

Two histone proteins were typically purified sequentially in one afternoon. The dialysed crude histone solutions were loaded onto a 5 mL HiTrap SP HP column pre-equilibrated with buffer A (IEX binding buffer; this was done manually with flow rate= 1.00 mL/ min, pressure limit 0.5 MPa until UV, pressure and cond% were constant for at least 1 C.V.). In the case of H4, which had an initial volume of approx. 60 mL, concentration using two 20 mL 3 kDa vivaspin columns using centrifugation (5,000 x g, 30 min) was first necessary to fit all protein solution in the 50 mL SuperLoop. Purification was then carried out using the protocol RR SP HP; in summary, the column was washed with 12 C.V. buffer A to remove unbound proteins, before eluting histone in a gradient with buffer B (IEX elution buffer; 7M Urea, 10 mM Tris.HCl pH 7.5, 1M NaCl, 1 mM EDTA 5 mM β-ME. Elution gradient was 0 to 100% B over 15 C.V.). The fractions were analysed by SDS-PAGE, and those containing pure histone protein were combined and dialysed against 5 mM β-ME in MilliQ H_2O (3.5 kDa dialysis cassette). Note that H4 crashed out after this step, while the others did not. These were then lyophilized and initially stored at -20 °C for several days to two weeks, before storing at -80 °C.

H3 WT

UV trace from Hitrap



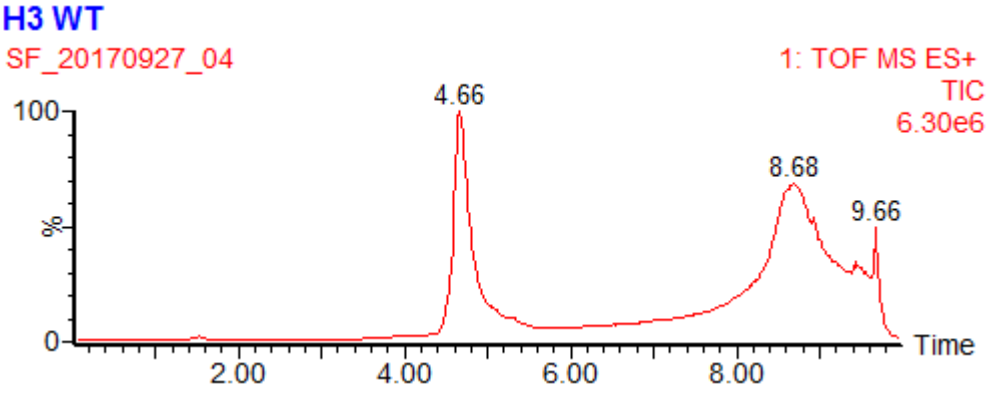
SDS-PAGE



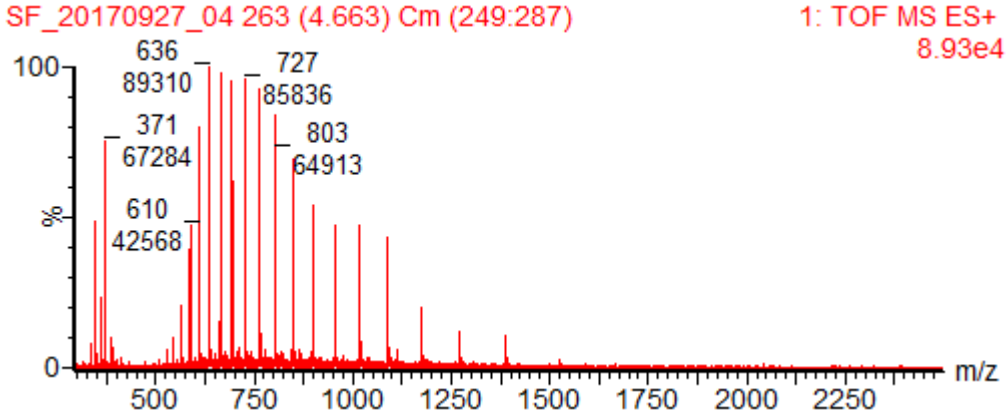
Fractions highlighted in red correspond to those containing H3 WT

Fig. 2.4.4.3.1 *Characterization of H3 WT after Hitrap purification- UV trace and SDS-PAGE. From fig. 2.2.1.2.*

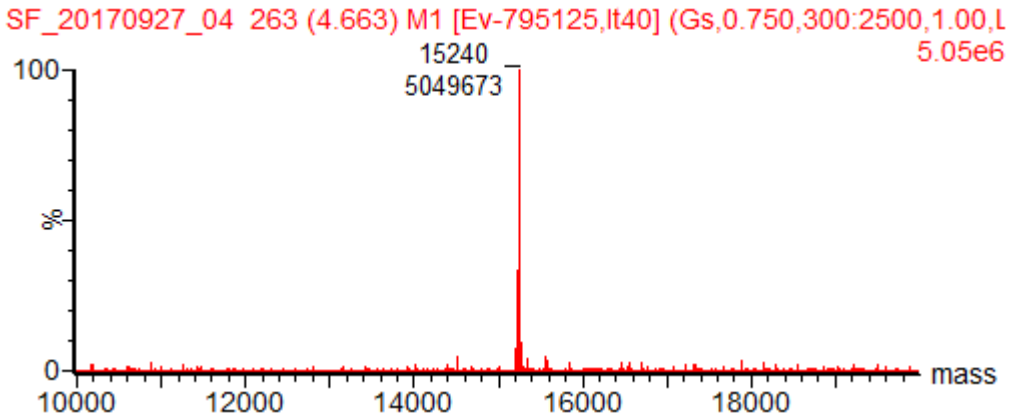
LCMS-ESI+ characterization- chromatogram



Ion series



Deconvoluted mass spectrum



MS-ESI+: Expected 15238.80 Da ; Found 15240 Da

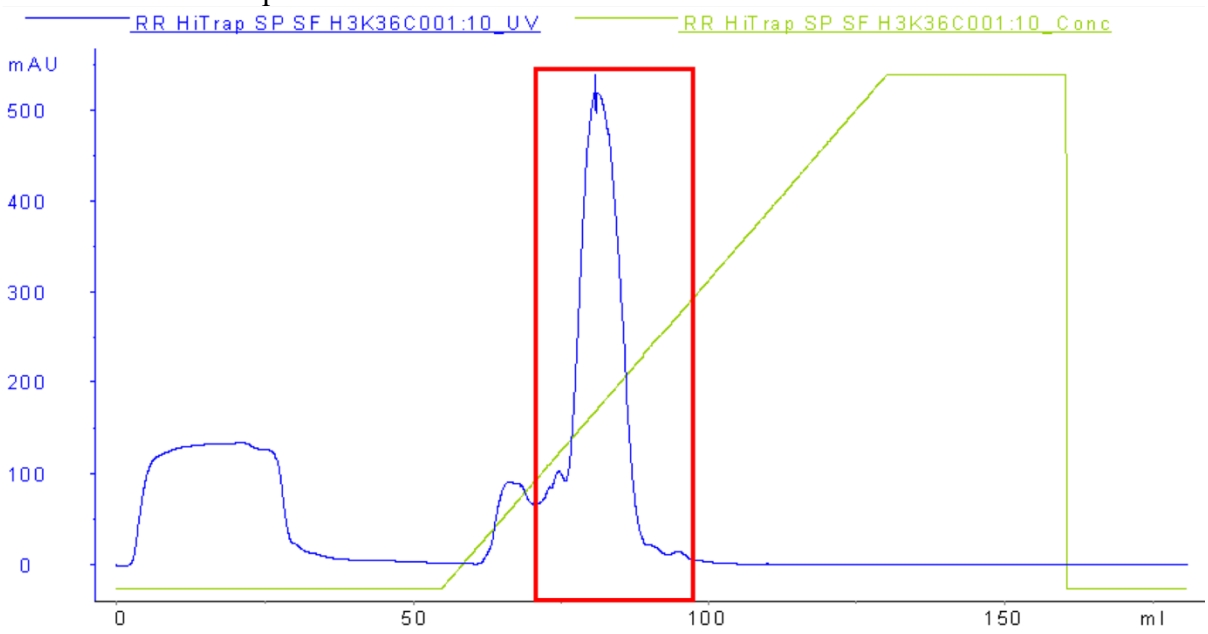
(Note that M from start codon is not present in final protein)

Fig. 2.4.4.3.2 *Characterization of H3 WT after Hitrap purification- LCMS-ESI+. From fig. 2.2.1.2.*

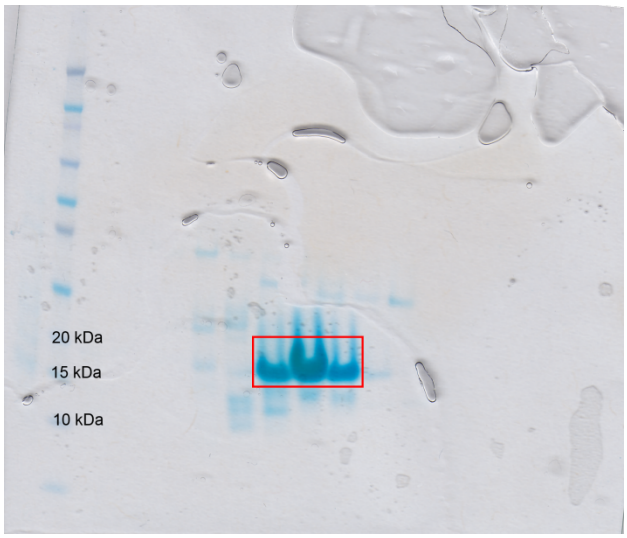
Yield = 101 mg (42 mg/ L of culture)

H3K36C

UV trace from Hitrap



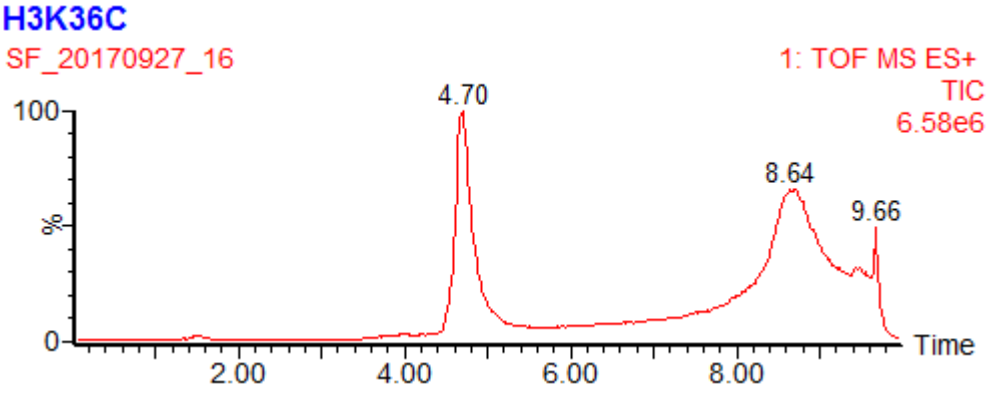
SDS-PAGE



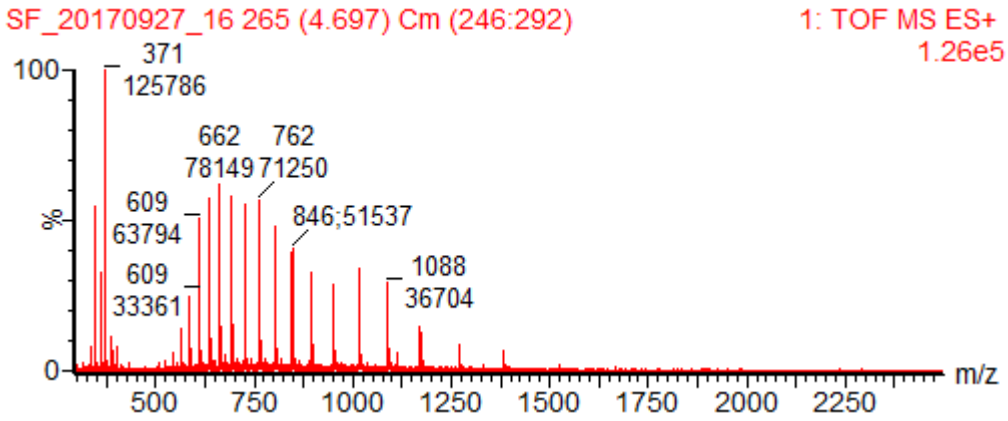
Fractions highlighted in red correspond to those containing H3K36C

Fig. 2.4.4.3.3 *Characterization of HK36C after Hitrap purification- UV trace and SDS-PAGE. From fig. 2.2.1.2.*

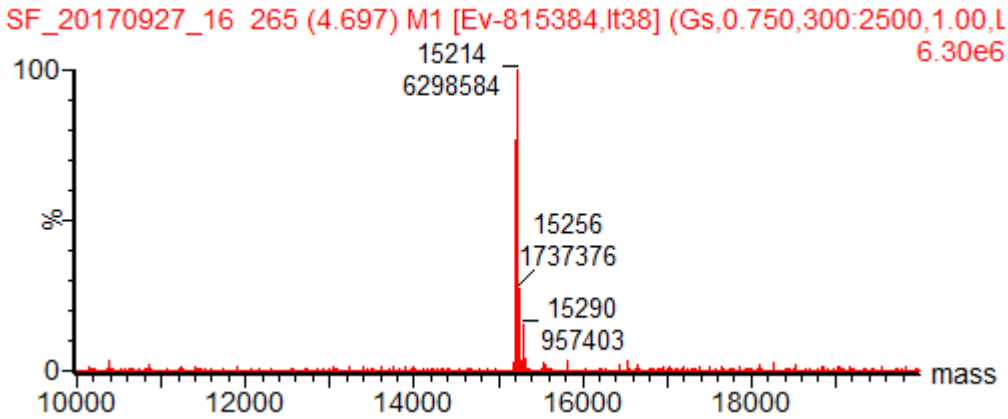
LCMS-ESI+ characterization- Chromatogram



Ion series



Deconvoluted mass spectrum



MS-ESI+: Expected 15213.76 Da ; Found 15214 Da (15256 peak= +42= acetonitrile adduct)

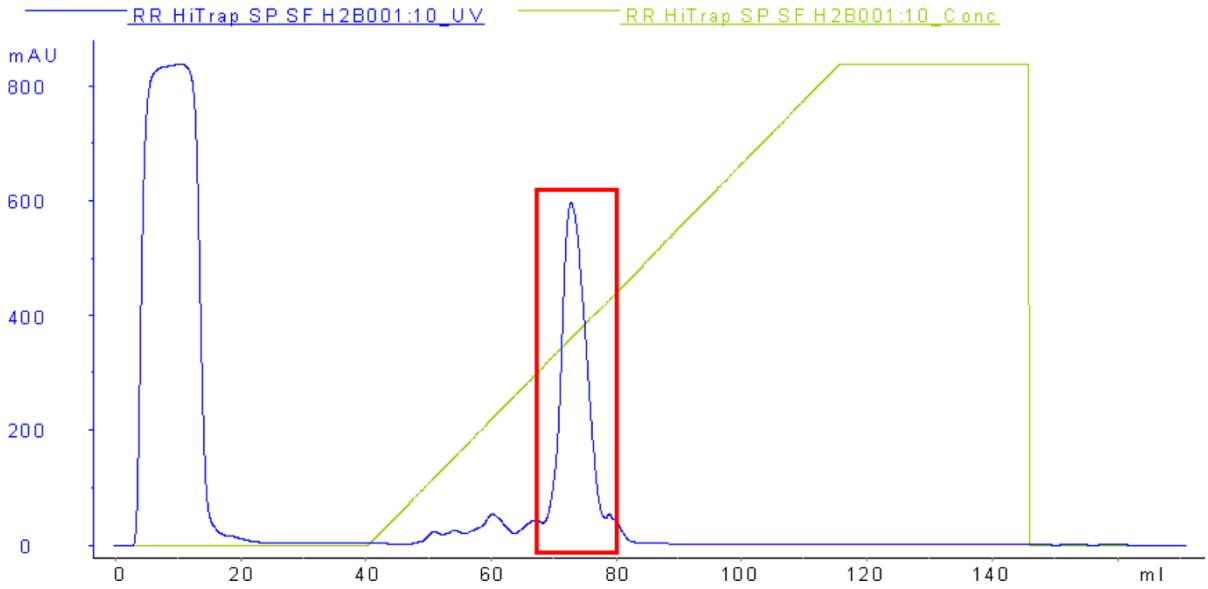
(Note that M from start codon is not present in final protein)

Fig. 2.4.4.3.4 *Characterization of H3K36C after Hitrap purification- LCMS-ESI+. From fig. 2.2.1.2.*

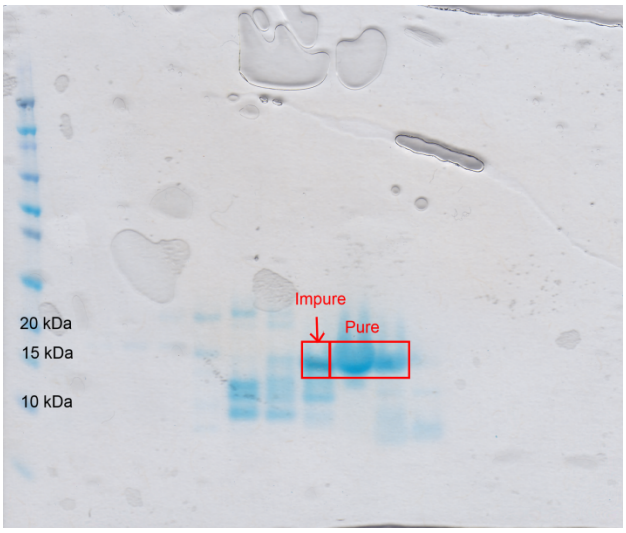
Yield = 66 mg (27.5 mg/ L of culture)

H2B

UV trace from Hitrap



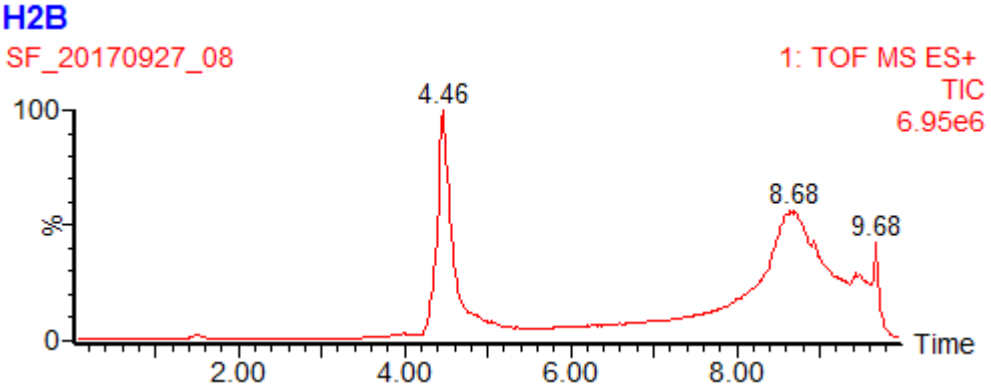
SDS-PAGE



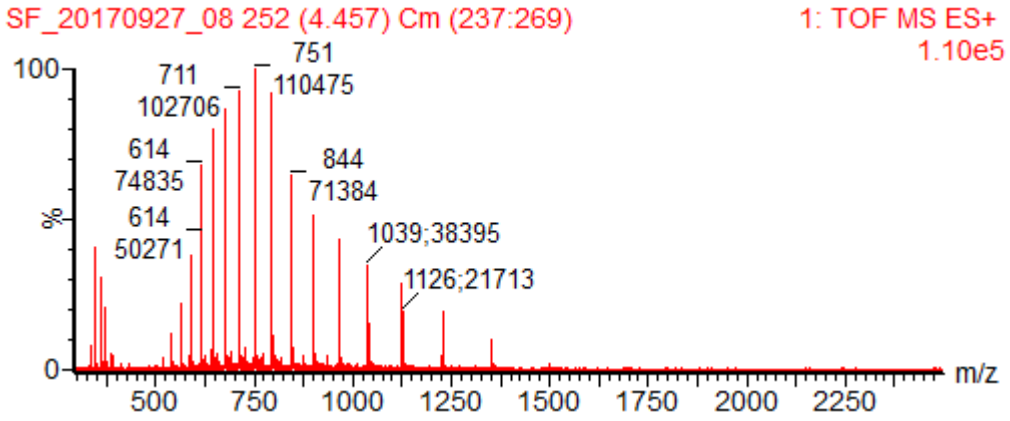
Fractions highlighted in red correspond to those containing H2B

Fig. 2.4.4.3.5 *Characterization of H2B after Hitrap purification- UV trace and SDS-PAGE. From fig. 2.2.1.2.*

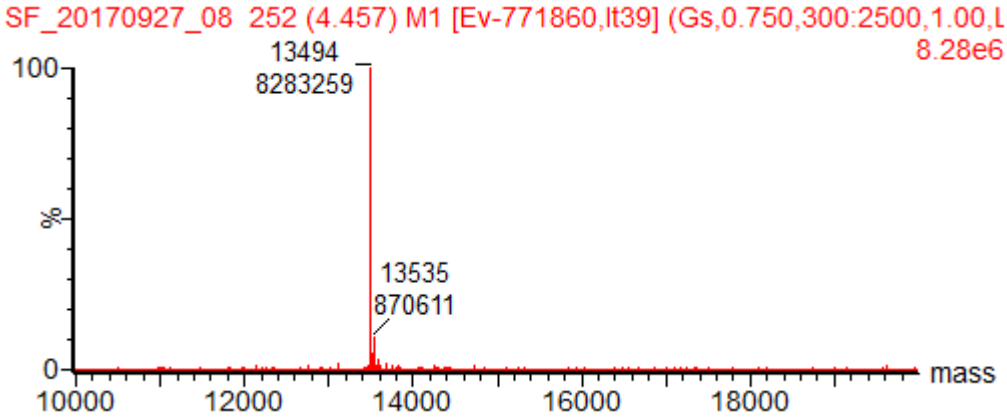
LCMS-ESI+ characterization- Chromatogram



Ion series



Deconvoluted mass spectrum



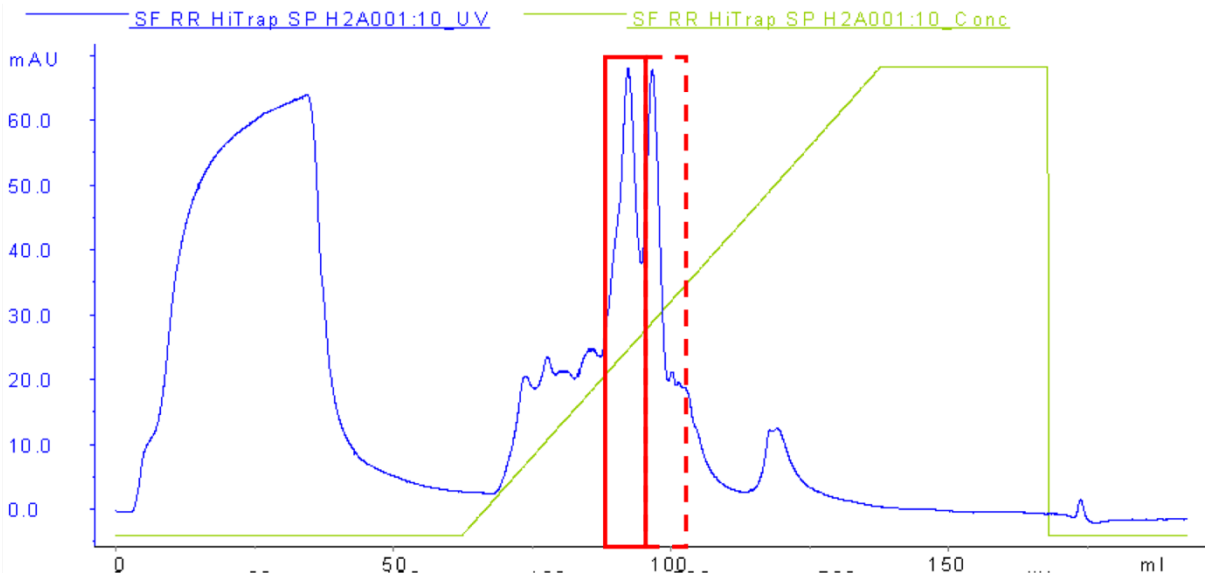
MS-ESI+: Expected 13493.68 Da ; Found 13494 Da (Note that M from start codon is not present in final protein)

Fig. 2.4.4.3.6 *Characterization of H2Bafter Hitrap purification- LCMS-ESI+. From fig. 2.2.1.2.*

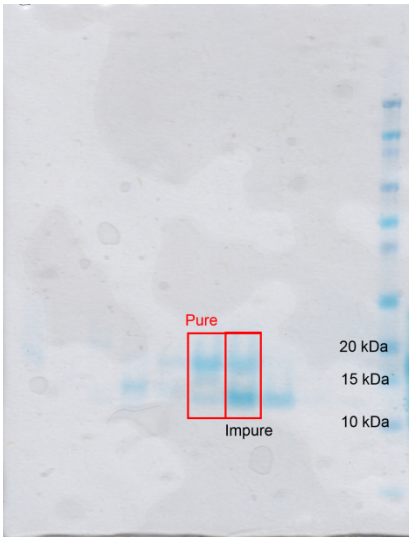
Yield = 60 mg (25 mg/ L of culture)

H2A

UV trace from Hitrap



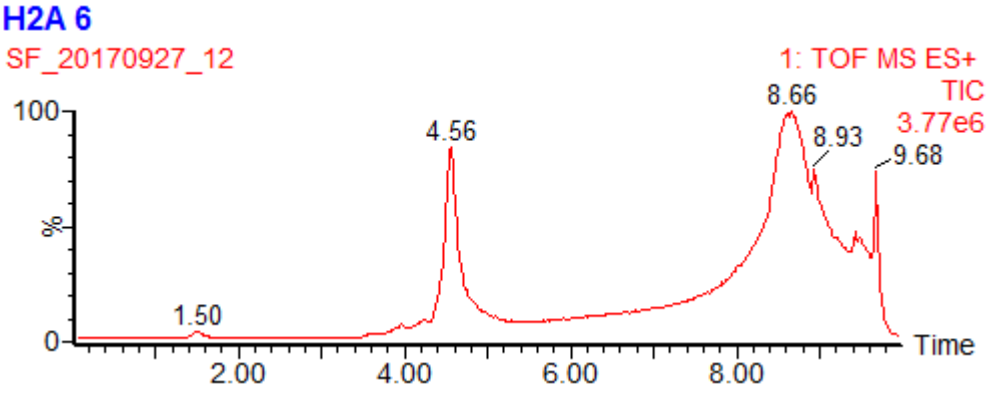
SDS-PAGE



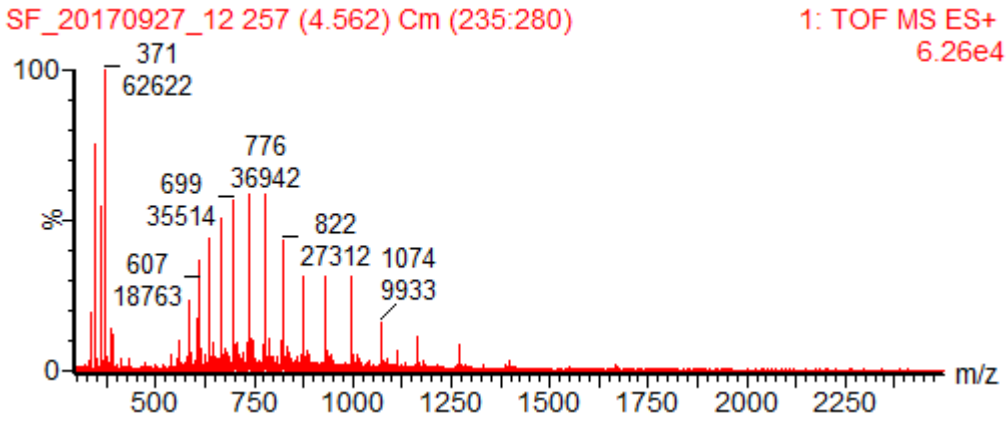
Fractions highlighted in red correspond to those containing H2A

Fig. 2.4.4.3.7 *Characterization of H2A after Hitrap purification- UV trace and SDS-PAGE. From fig. 2.2.1.2.*

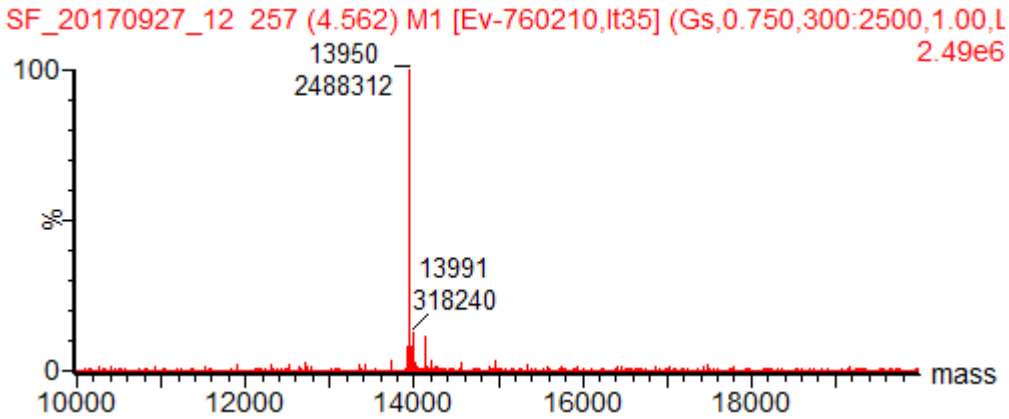
LCMS-ESI+ characterization- Chromatogram



Ion series



Deconvoluted mass spectrum



MS-ESI+: Expected 13950.20 Da ; Found 13950 Da.

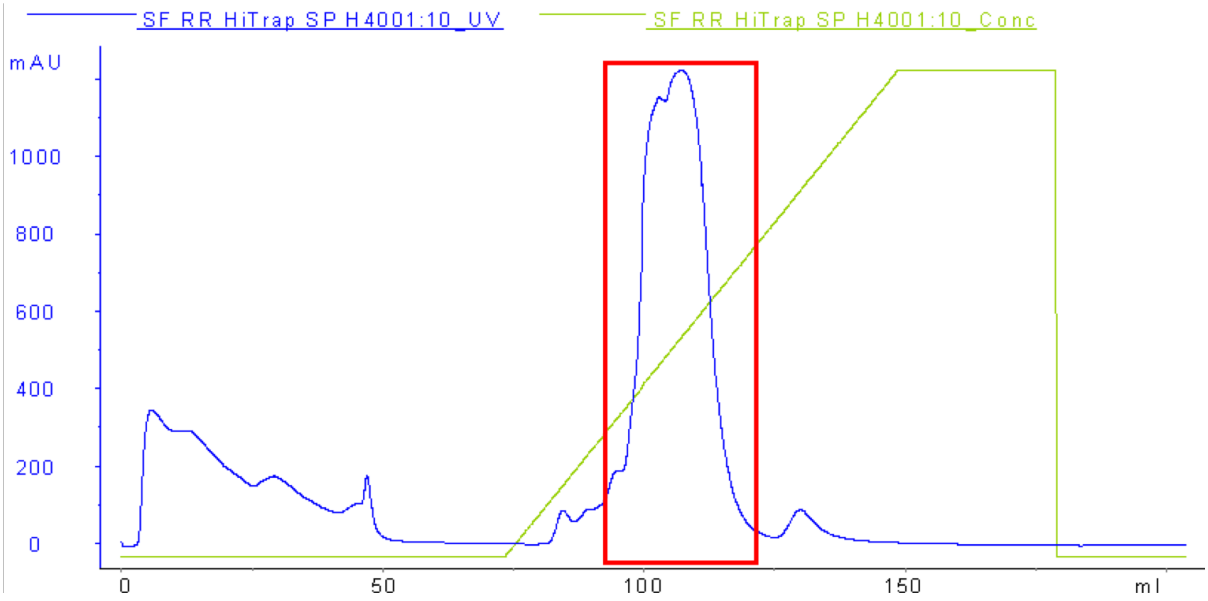
(Note that M from start codon is not present in final protein)

Fig. 2.4.4.3.8 *Characterization of H2A after Hitrap purification- LCMS-ESI+. From fig. 2.2.1.2.*

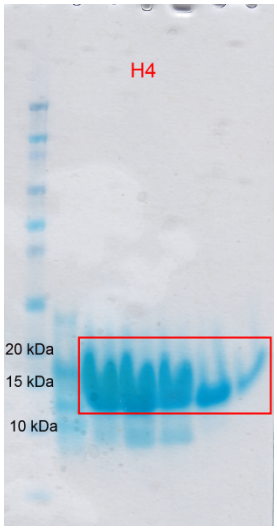
Yield = 13 mg (5.4 mg/ L of culture). Low yield is due to loss on SEC column.

H4

UV trace from Hitrap



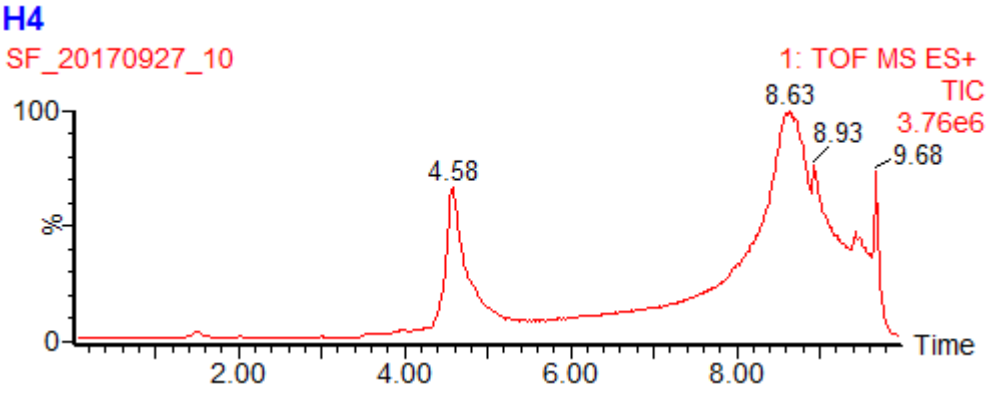
SDS-PAGE



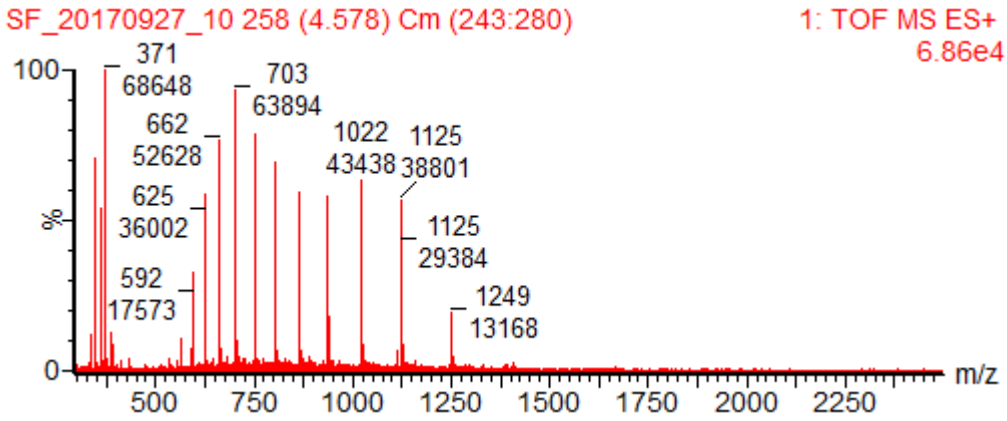
Fractions highlighted in red correspond to those containing H4

Fig. 2.4.4.3.9 *Characterization of H4 after Hitrap purification- UV trace and SDS-PAGE. From fig. 2.2.1.2.*

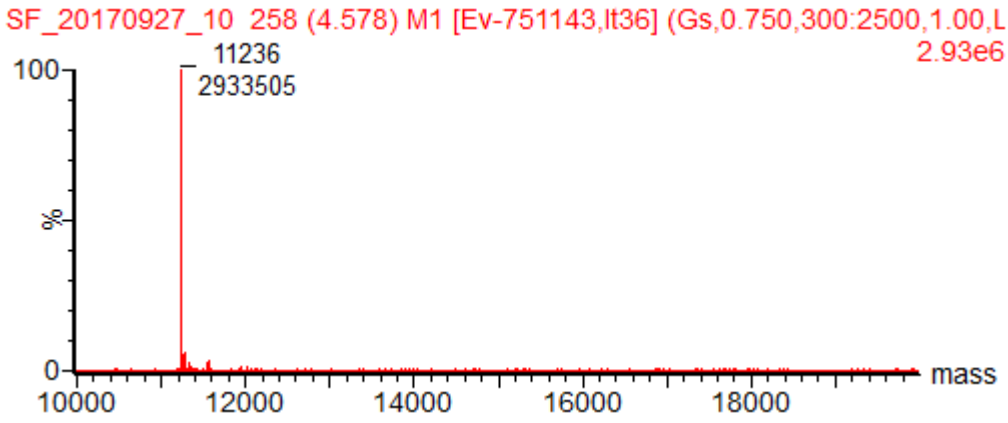
LCMS-ESI+ characterization- Chromatogram



Ion series



Deconvoluted mass spectrum



MS-ESI+:

Expected 11236.15 Da ; Found 11236 Da.

(Note that M from start codon is not present in final protein)

Fig. 2.4.4.3.10 *Characterization of H4 after Hitrap purification- LCMS-ESI+. From fig. 2.2.1.2.*

Yield= 127 mg (53 mg/ L of culture)

2.4.5 Optimization of the Dha formation reaction on H3K36C

2.4.5.1 Pre-treatment of H3K36C with DTT to reduce disulfides

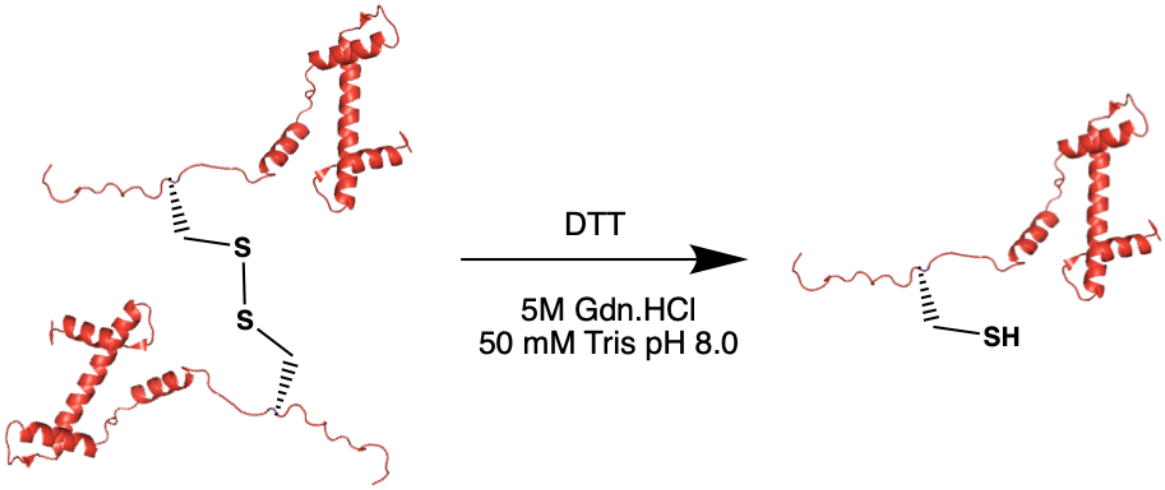
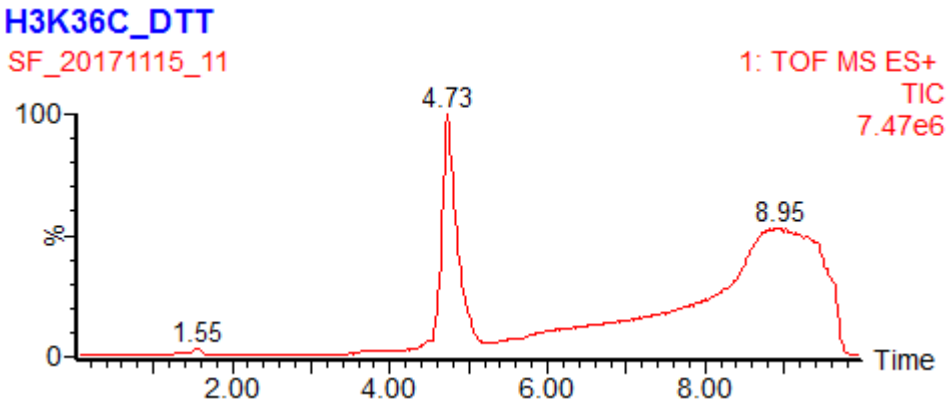


Fig. 2.4.5.1.1 *Scheme for DTT reduction of intermolecular disulfides in H3K36C.*

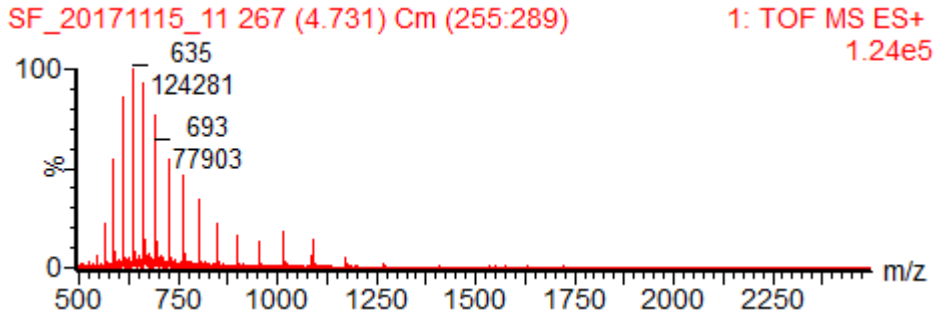
H3K36C (4.7 mg, 0.347 µMol, 1 eq) was weighed into a 1.5 mL Eppendorf tube and dissolved in 500 µL reaction buffer (5M Gdn.HCl, 50 mM Tris pH 8.0). Approx. 6 mg DTT was then added and the reaction was shaken for 45 minutes at 37 °C. During this time a GE PD G-25 minitrap column was equilibrated with reaction buffer (3 x 2 mL). The H3K36C solution was then passed through the minitrap and eluted in 1 mL reaction buffer.

LCMS-ESI+ characterization

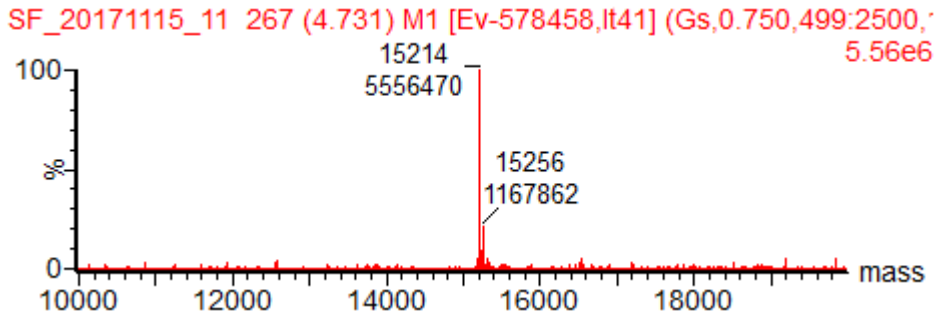
Chromatogram



Ion series



Deconvoluted mass spectrum



H3K36C: Expect 15214Da [M], 15256 Da[M+ MeCN]: found 15214 Da [M], 15256 Da [M+ MeCN]

Fig. 2.4.5.1.2 *LCMS- ESI+ characterization of H3K36C treated with DTT.*

2.4.5.2 Reaction with MDBP

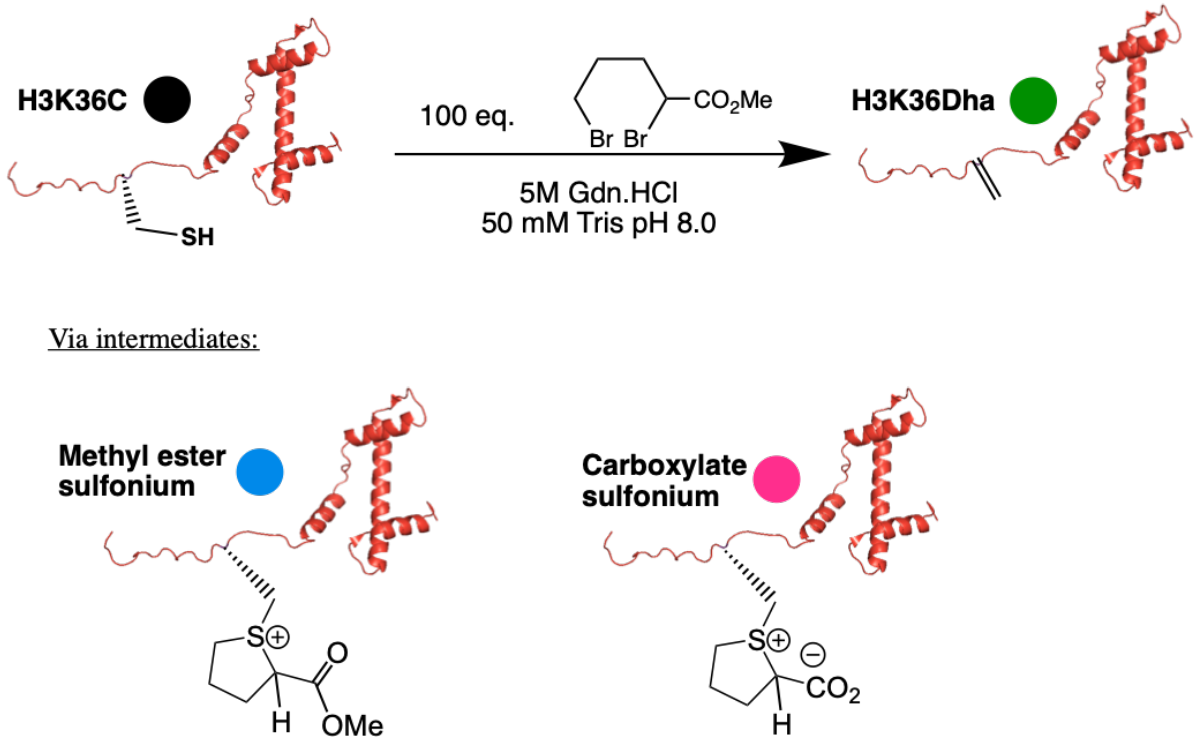


Fig. 2.4.5.2.1 *Scheme for expected starting material, product and intermediates for reaction of H3K36C with MDBP.*

Immediately after **2.4.5.1**, MDBP (5.13 µL in 50 µL DMSO, 100 eq.) was added, upon which the reaction became a cloudy suspension. The reaction was incubated at 37° C and aliquots taken for LCMS-ESI+ analysis at half-hourly or hourly timepoints. After 5 hours when the reaction had gone to completion as suggested by LCMS-ESI+ data, it was quenched by passing through a GE PD G-25 miditrap pre-equilibrated with Luche buffer (3 x 5 mL; 5M Gdn.HCl, 500 mM NH_4OAc pH 6.0) and eluted in 1.5 mL Luche buffer into a 2 mL Eppendorf tube. Protein concentration was measured by nanodrop (ε_0= 4470 $M^{-1}cm^{-1}$, MW= 15.18 kDa). Solutions were diluted to approx. 1 mg/ mL in Luche buffer, re-measured by nanodrop, divided into 0.75 mL aliquots and placed in a nitrogen glovebox to degas overnight at room temperature.

Percentages of each species present at each timepoint were calculated from LCMS-ESI+ spectra and plotted using PRISM graphing software. LCMS-ESI+ data for the first timepoint (30 min) is shown here; for all other timepoint data, see appendix **2.6.2**

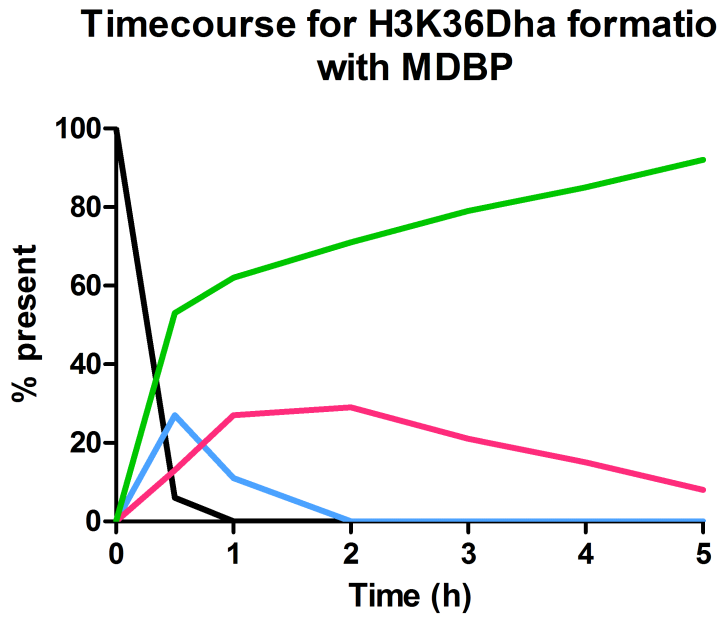
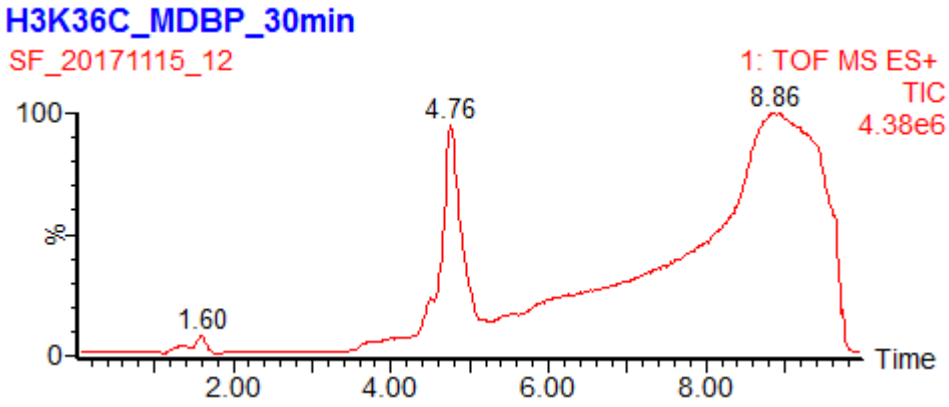


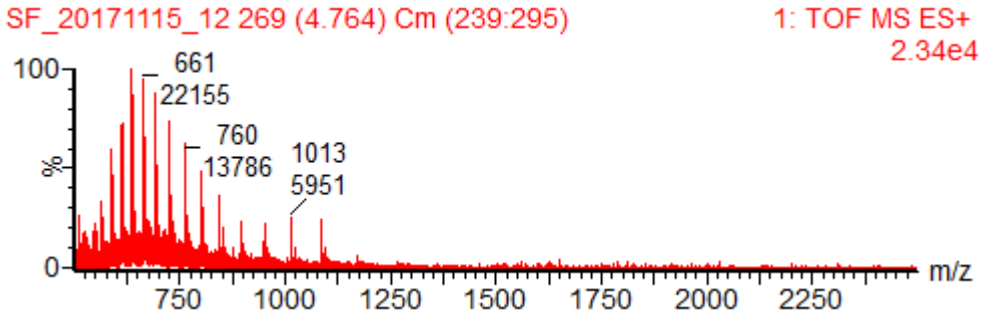
Fig. 2.4.5.2.2. *Graph showing timecourse for H3K36Dha formation with MDBP: percentages for each species are calculated by taking the relative height of the corresponding mass peak from the deconvoluted LCMS-ESI+ spectrum (see below for example). From fig. 2.2.2.2.*

Example LCMS-ESI+ characterization from 30 min timepoint

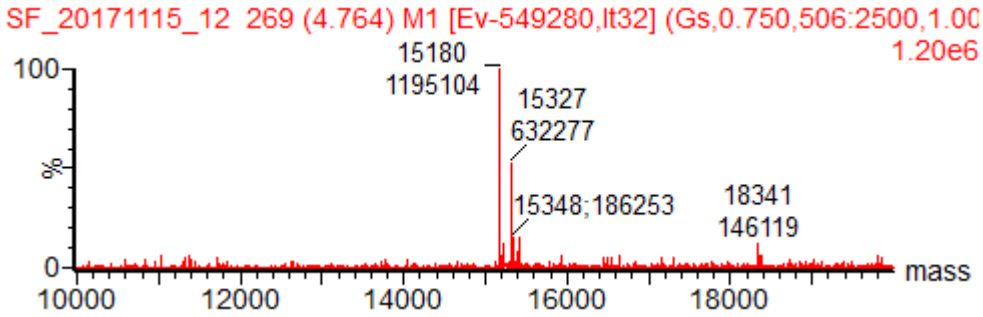
Chromatogram



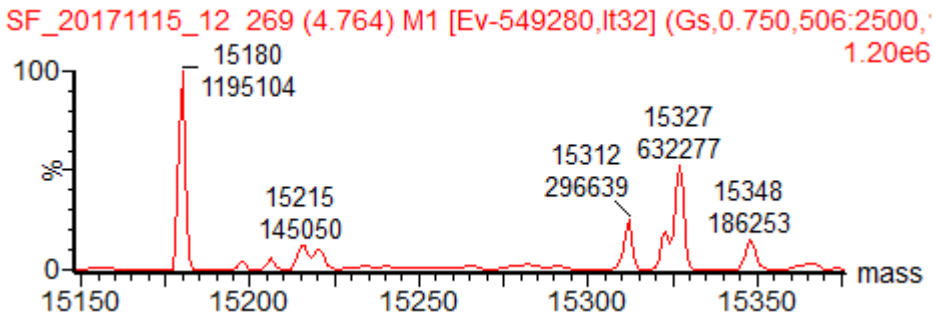
Ion series



Deconvoluted mass specteum (full)



Deconvoluted mass spectrum (zoomed view)



H3K36Dha: Expect 15181 Da, found 15180 Da, 53%

H3K36C: Expect 15214 Da, found 15215 Da, 6% (noise)

H3K36[carboxylate sulfonium]: Expect 15312 Da, found 15312 Da, 13%

H3K36[methyl ester sulfonium]: Expect 15326 Da, found 15327 Da, 27%

Fig. 2.4.5.2.3. *Example LCMS-ESI+ quantification of 30 min timepoint in H3K36C reaction with MDBP.*

2.4.6 Post-translational mutagenesis

2.4.6.1 Synthesis of H3K36Me3; screening protein and (iodide) concentrations.

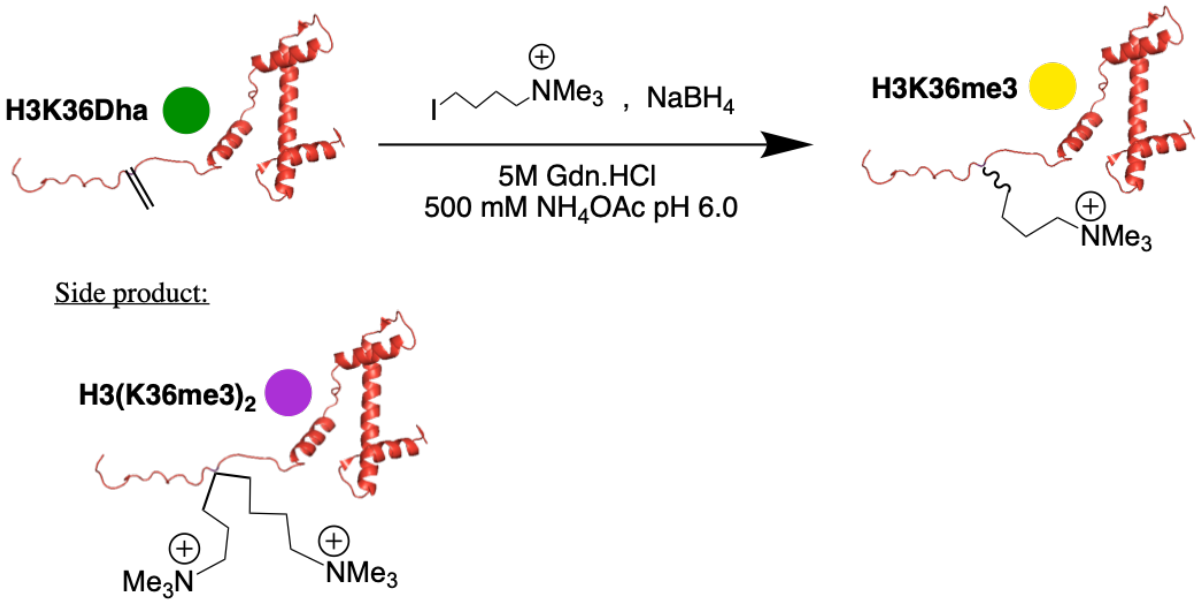


Fig. 2.4.6.1.1 *Scheme for radical reaction of H3K36Dha to form H3K36me3.*

Following equilibration of **2.4.5.1** with N_2 (g) overnight in anaerobic glovebox, H3K36Dha in Luche buffer (1.2 mg/ mL or 0.3 mg/ mL in 5M Gdn.HCl, 500 mM NH_4OAc pH 6.0; note some buffer evaporated overnight) was taken in 200 µL aliquots and mixed first into 3-iodopropyltrimethylamine then into $NaBH_4$ pre-portioned into 2 mL eppendorf tubes in the amounts shown in the table below. These were shaken at 4 °C for 30 mins before being passed through GE PD G-25 spintraps pre-equilibrated with unfolding buffer (7M Gdn.HCl, 10 mM Tris pH 7.5, 1 mM DTT). Reactions were then analysed by LCMS-ESI+ to determine reaction conversion.

See appendix **2.6.3.2** for LCMS-ESI+ spectra.

	Low [Protein, $NaBH_4$]		High [Protein, $NaBH_4$]	
H3K36Dha (mg/ mL)	0.3		1.2	
$NaBH_4$ (mg)	0.2		0.5	
(3-iodopropyl) trimethylamine (mg)	2.5	5	10	15

Table 2.4.6.1.1 *Concentrations of reagents screened.*

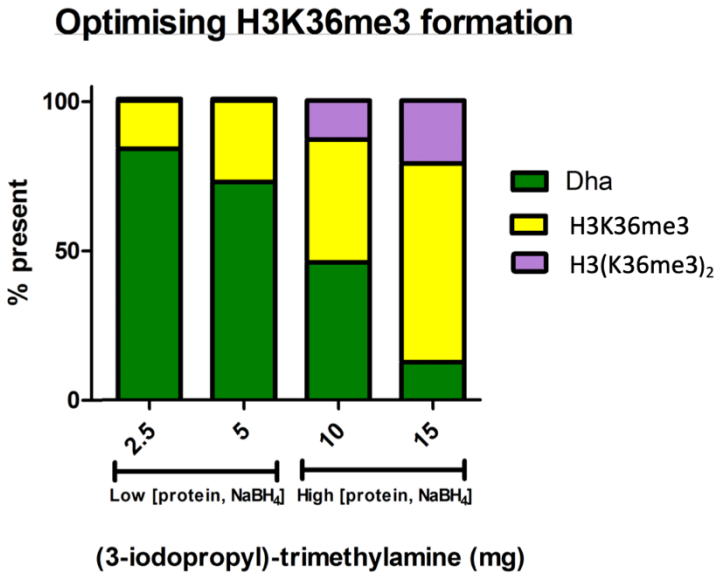


Fig. 2.4.6.1.2. *Graph of screening results- conversions quantified from LCMS-ESI+ spectra (appendix 2.6.3.2) were plotted using PRISM graphing software. From fig. 2.2.2.3.*

2.4.6.2 Synthesis of H3K36me2 and H3K36me3 histone used in nucleosome assembly

These experiments were carried out by Brian Josephson (DPhil, B.G.D. group), and the following data was supplied by him.

H3K36me3

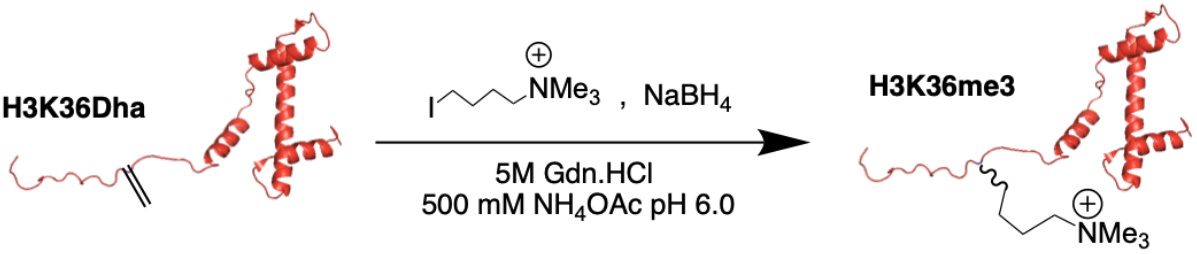
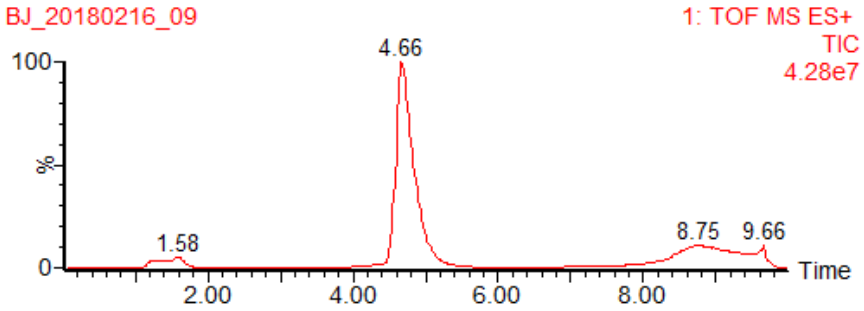


Fig. 2.4.6.2.1. *Scheme for synthesis of H3K36me3from H3K36Dha.*

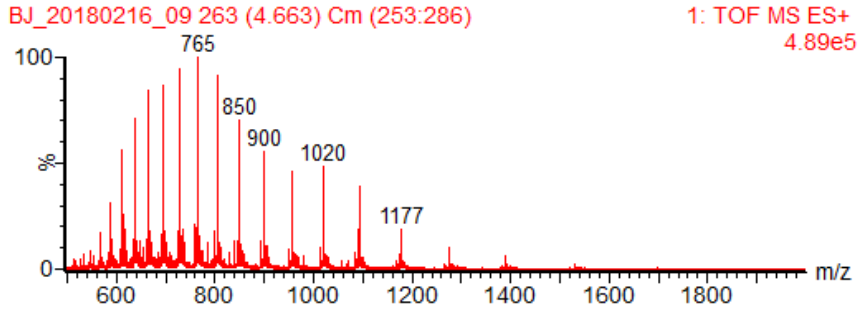
This reaction was carried out in duplicate. Following equilibration with N_2 (g) overnight in anaerobic glovebox, 100 µL H3K36Dha in Luche buffer (1 mg/ mL in 5M Gdn.HCl, 500 mM NH_4OAc pH 6.0) were mixed first into (3-iodopropyl)-trimethylamine (20 mg) then into $NaBH_4$ (2 mg) pre-portioned into a 2 mL eppendorf tube. This was shaken at 4 °C for 30 mins before being passed through GE PD G-25 spintraps pre-equilibrated with unfolding buffer (7M Gdm.HCl, 10 mM Tris pH 7.5, 1 mM DTT). The reactions were then analysed by LCMS-ESI+.

H3K36me3 ‘batch 1’ LCMS-ESI+ characterization

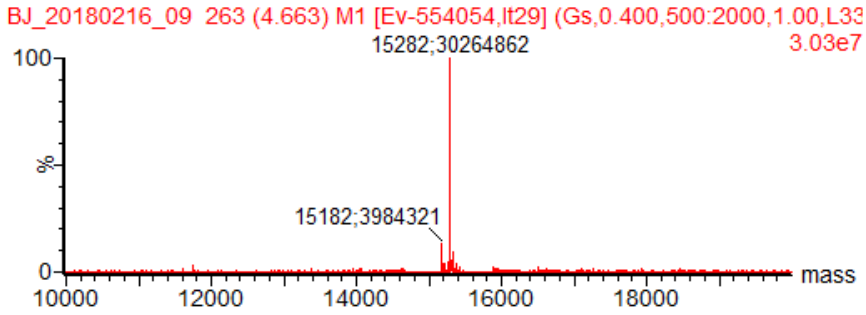
Chromatogram



Ion series



Deconvoluted mass spectrum



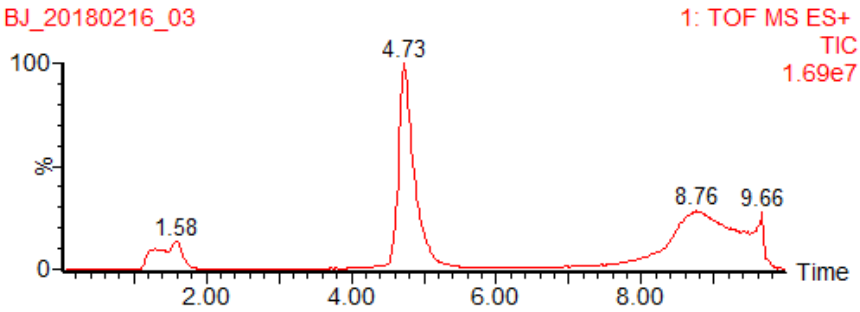
H3K36Dha: Expected 15180 Da, observed 15182 Da.

H3K36me3: Expected 15280 Da, observed 15282 Da.

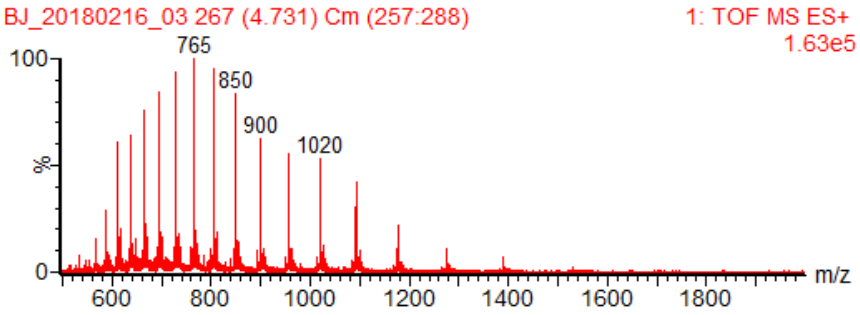
Fig. 2.4.6.2.2. *LCMS-ESI+ characterization data for H3K36me3 (batch 1)*

H3K36me3 ‘batch 2’ LCMS-ESI+ characterization

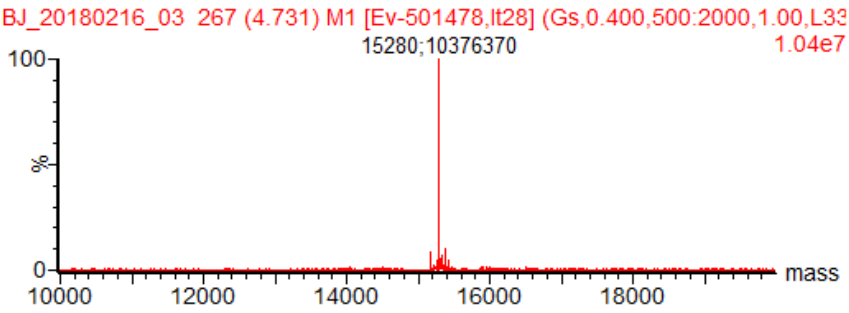
Chromatogram



Ion series



Deconvoluted mass spectrum



H3K36me3: Expected 15280 Da, observed 15280 Da. This batch was used for subsequent nucleosome assembly.

Fig. 2.4.6.2.3. *LCMS-ESI+ characterization data for H3K36me3 (batch 2). From fig. 2.2.2.4.*

H3K36me2

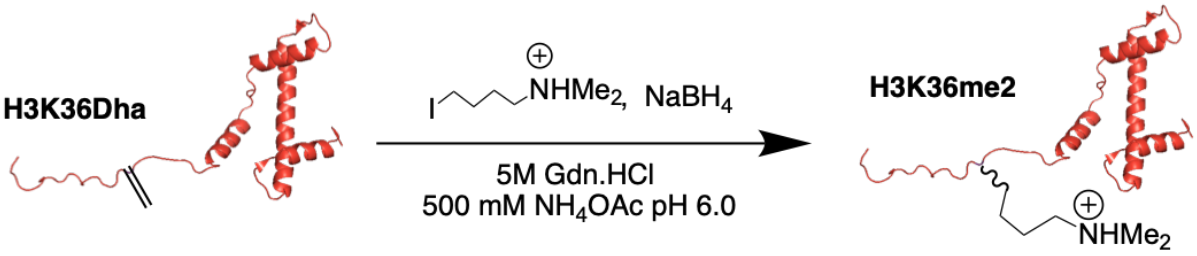
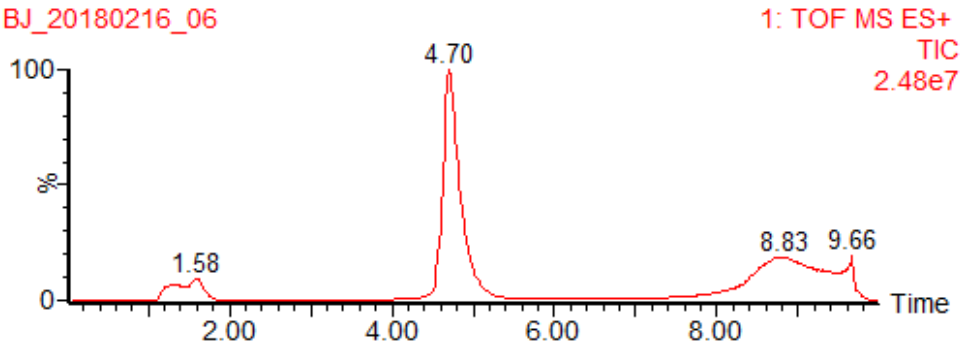


Fig. 2.4.6.2.4. *Scheme for the synthesis of H3K36me2 from H3K36Dha.*

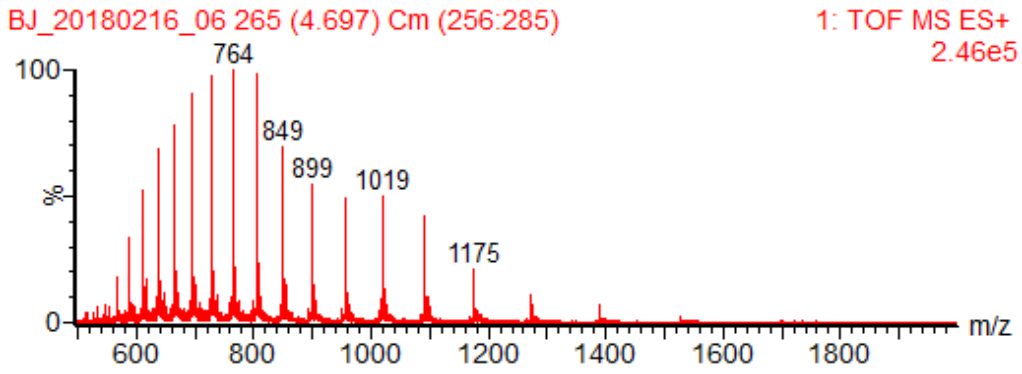
Following equilibration with N_2 (g) overnight in anaerobic glovebox, 100 µL H3K36Dha in Luche buffer (1 mg/ mL in 5M Gdn.HCl, 500 mM NH_4OAc pH 6.0) was mixed first into (3-iodopropyl)dimethylamine hydroiodide (20 mg) then into $NaBH_4$ (2 mg) pre-portioned into a 2 mL eppendorf tube. This was shaken at 4 °C for 30 mins before being passed through GE PD G-25 spintraps pre-equilibrated with unfolding buffer (7M Gdn.HCl, 10 mM Tris pH 7.5, 1 mM DTT). The reaction was then analysed by LCMS-ESI+.

LCMS-ESI+ characterization

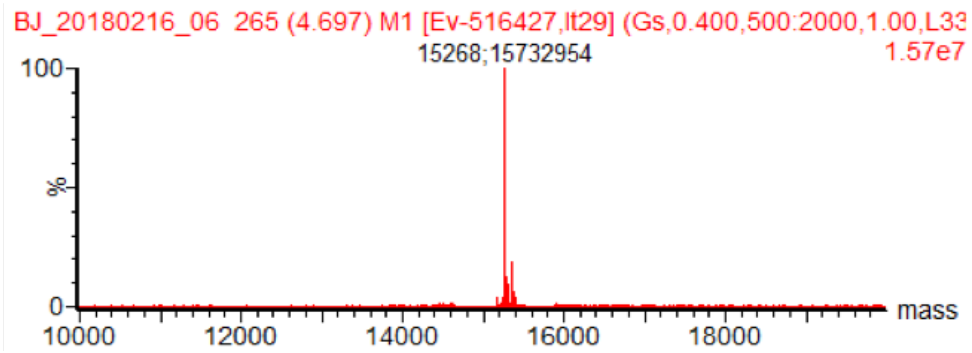
Chromatogram



Ion series



Deconvoluted mass spectrum



H3K36me2: Expected 15267 Da, observed 15268 Da.

Fig. 2.4.6.2.4. *LCMS-ESI+ characterization data for H3K36me2. From fig. 2.2.2.4.*

2.4.7 Octamer refolding

2.4.7.1 Refolding with H3 WT

Histones H2A, H2B, H3 WT and H4 (~ 2 mg) were weighed out into 2 mL Eppendorf tubes and dissolved in 1 mL unfolding buffer (6 M Gdn. HCl, 20 mM Tris. HCl pH 7.5, 5 mM DTT, prepared on a 20 mL scale immediately beforehand) by gentle pipetting. The solutions were then incubated at room temperature for 1 hour before measuring concentrations via nanodrop.

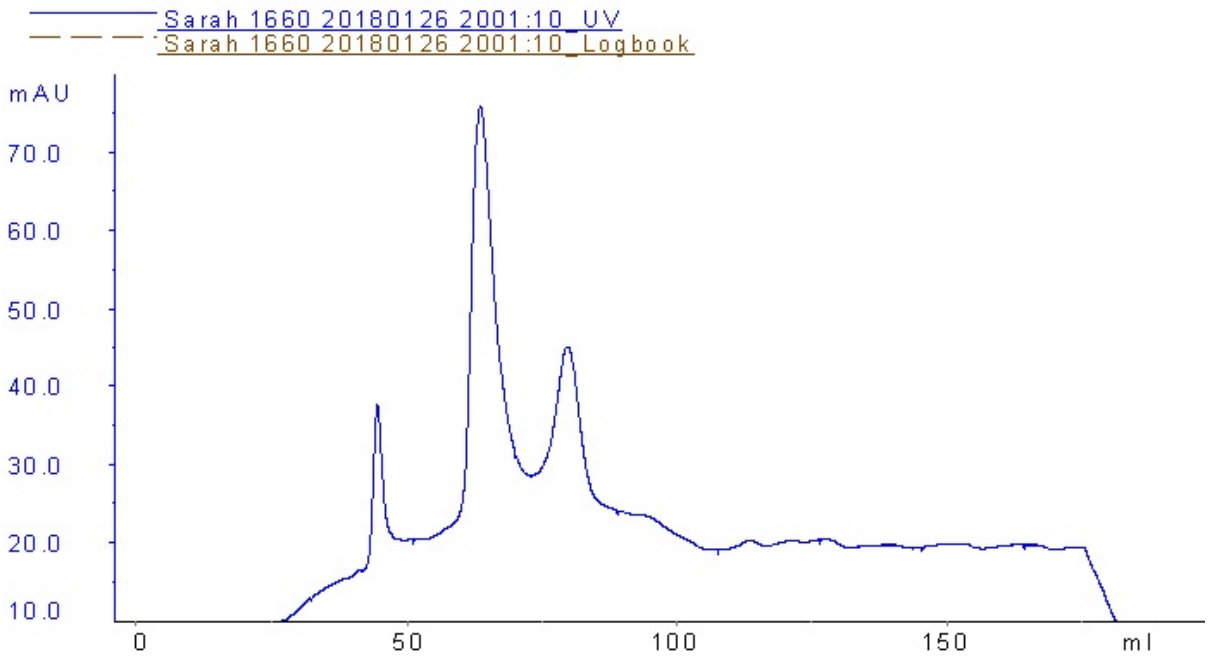
H2A, H2B, H3 WT and H4 were then mixed in an equimolar ratio in a 50 mL falcon tube to a final concentration of approx. 1 mg/ mL in unfolding buffer (see appendix 2.6.5.2 for calculations and final quantities used). The combined histones were then transferred to a 6 kDa Pur-a-Lyzer Mega dialysis cup, and dialysed against 600 mL of pre-chilled refolding buffer (2M NaCl, 10 mM Tris. HCl pH 7.5, 1 mM EDTA and 5 mM β-ME) for 6 hours at 4 °C, then another 14 hours in a fresh 600 mL refolding buffer (overnight).

The next day, the final dialysis step was carried out (600 mL refolding buffer, 4 °C, 6 hours), during which time a HiLoad 16/60 size exclusion column (GE healthcare) was equilibrated with refolding buffer. The octamer was transferred to a 5 kDa MWCO Vivaspin concentrator and centrifuged (3750 rpm, 20 min, 4°C) until volume was ≤1 mL, then transferred to an Eppendorf and centrifuged again (7000 rpm, 10 min, 4°C) to pellet any precipitate. The supernatant was then loaded onto the SEC 16/60 and eluted in refolding buffer over 1.2 CV at 0.2 mL/ min.

Octamer-containing fractions were identified by retention time and/ or denaturing SDS-PAGE (4-12% Bis-Tris); fractions containing all 4 histones gave 4 corresponding bands on the gel. These fractions were combined and concentrated to a volume of 1 mL using a 6 mL 6 kDa cutoff vivaspin concentrator at 4°C, avoiding excess mechanical stress to prevent dissociation. The octamer was diluted with an equivalent volume of 100% glycerol (to a final concentration of 50% glycerol) for storage at -20 °C.

WT octamer characterization

UV trace from SEC



SDS-PAGE

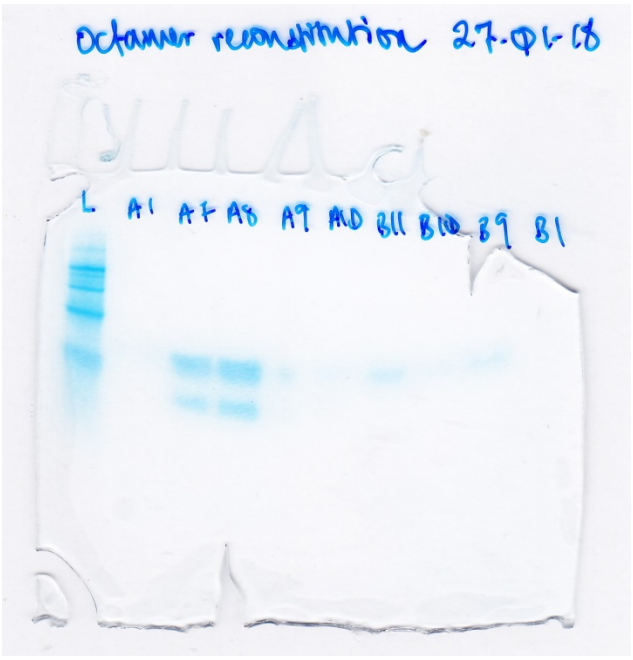
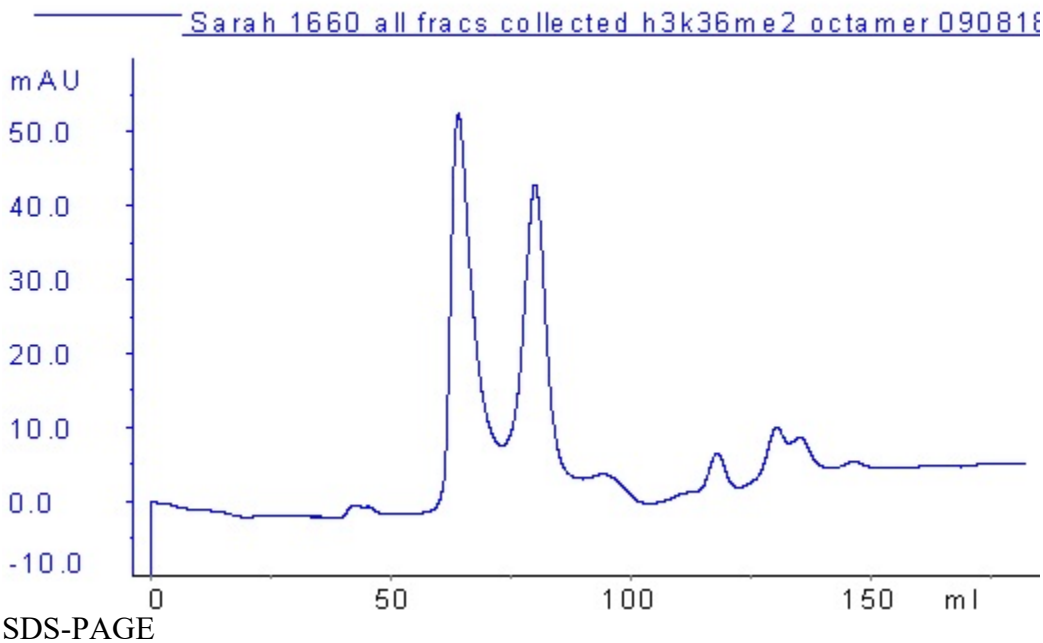


Fig. 2.4.7.1.1. *Characterization data for H3 WT containing nucleosome. SEC trace from fig. 2.2.3.1.A.*

H3K36me2-containing octamer characterization

UV trace from SEC



SDS-PAGE

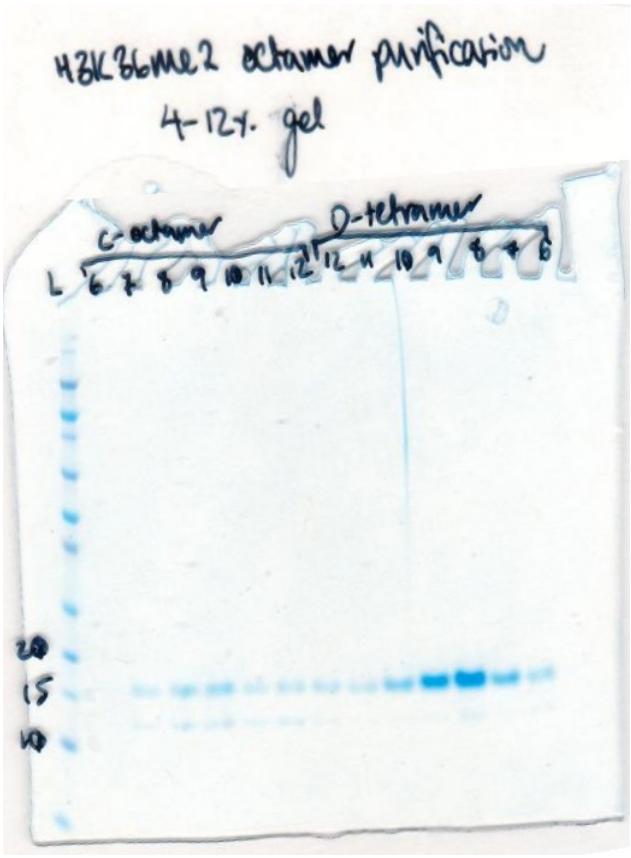


Fig. 2.4.7.1.2. *Characterization data for H3K36me2 containing nucleosome. SEC trace from fig. 2.2.3.1.A.*

H3K36me3-containing nucleosome characterization

UV trace from SEC

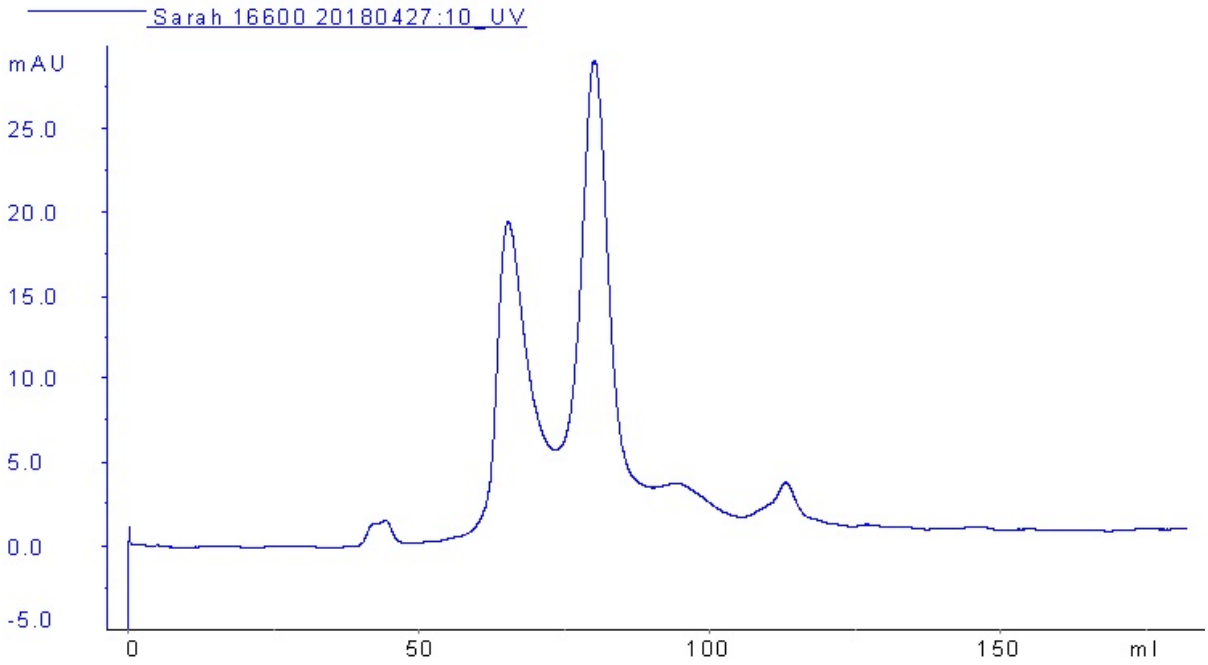


Fig. 2.4.7.1.3. *Characterization data for H3K36me2 containing nucleosome. SDS-PAGE is absent due to unexpected removal from destaining area- the octamer peak was identified by its retention volume. SEC trace from fig. 2.2.3.1.A.*

2.4.8 Nucleosome refolding with 145 bp DNA

2.4.8.1 Optimizing DNA: octamer ratio

When necessary, an aliquot of octamer (200 µL) was dialysed against 500 mL refolding buffer (2M KCl, 10 mM Tris. HCl pH 7.5, 1 mM DTT, 5 mM β-ME) overnight in a 6 kDa Pur-a-Lyzer mini dialysis membrane to remove glycerol.

Octamer concentration was measured by ND1000 using MW= 108485 Da and E 276 nm = 39120. Concentration of 145 bp DNA (prepared by Simon Nadal) was also determined.

Octamer and DNA were then mixed in a 1:0.85, 1:1 and 1:1.15 ratio (see appendix 2.6.6.1 for concentrations and calculations)

Importantly, the octamer was added last on each instance, and mixed by pipetting once only to avoid damaging the assembled octamer. Each of the three DNA/ octamer mixtures were then pipetted into a 6 kDa Pur-a-Lyzer mini dialysis membrane and dialysed sequentially into RRTCS (10 mM Tris. HCl pH 7.5, 1 mM EDTA, 1 mM DTT with decreasing KCl concentrations of 2M, 0.85 M and 0.65 M) for 2 hours each, then RRTCS 0.25 M (0.25 M KCl, 10 mM Tris. HCl pH 7.5, 1 mM EDTA, 1 mM DTT) overnight.

Heat shifting was then performed as follows. Each DNA/ octamer mixture was divided into 3 aliquots and incubated for 2 hours at 4°C, 37°C or 55 °C. During this time a Native-PAGE gel (Novex TBE 6%) was pre-run in 1x TBE buffer (150 V, 90 min, 4°C), after which the TBE buffer was mixed by emptying and re-filling the gel tank.

Samples were prepared by mixing 8 µL nucleosome solution with 2 µL 30% sucrose to a final concentration of 6%; 6 µL of each sample was loaded into the TBE gel wells and run (150 V, 90 min, 4°C). The gel was then rinsed, stained with SYBR safe (15 µL in 50 mL MilliQ H_2O for 15 mins) or SYBR gold (6 uL in 30 mL MilliQ for 20 mins) then InstantBlue overnight.



Fig. 2.2.4.1). *Uncropped gels from Fig. 2.2.4.1; 6% TBE gel characterization of nucleosome reconstitution samples (all with WT octamer and 145 bp DNA). Left= InstantBlue, right= SYBR safe.*

2.4.8.2 Assembly with WT, me2 and me3-containing octamers

Protocol 2.4.8.1 was repeated with a 1:1 DNA: octamer ratio and heat shifting at 55 °C

WT TBE gel

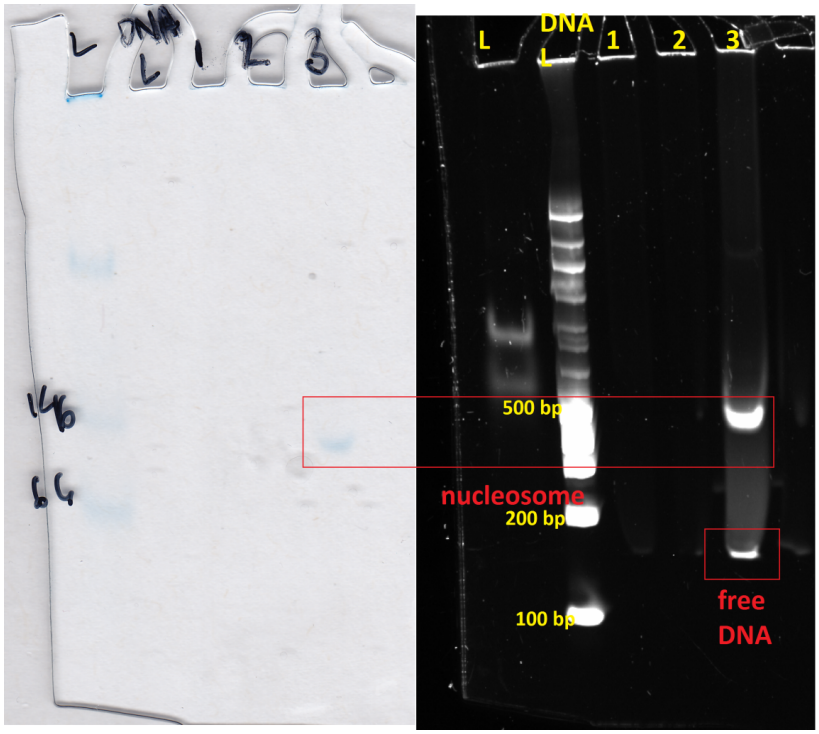


Fig. 2.4.8.2.1. *Uncropped gels from Fig. 2.2.4.2.A. 6% TBE gel of three nucleosome reconstitution samples with WT octamer and 145 bp DNA. Left= InstantBlue, right= SYBR safe*

Me2 TBE gel



Left= InstantBlue, right= SYBR safe. Cropped from full gel below:

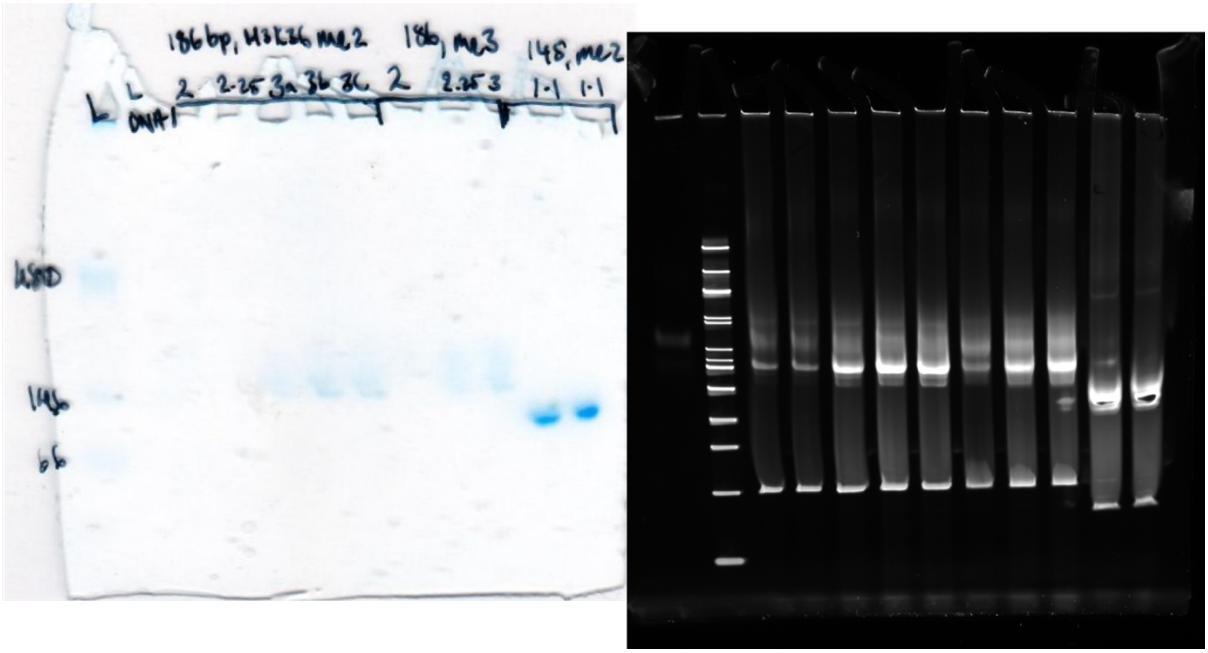


Fig. 2.4.8.2.2. *Uncropped gels from Fig. 2.2.4.2.C. 6% TBE gel of three nucleosome reconstitution samples with H3K36me2 octamer and 145 bp DNA.*

Me3 TBE gel

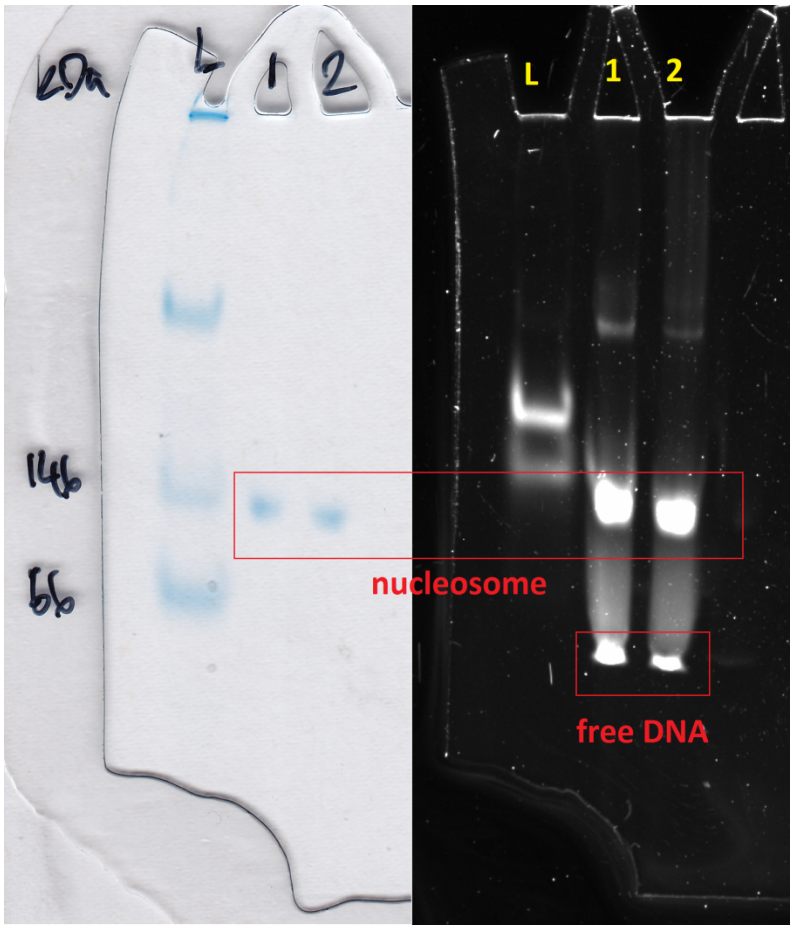


Fig. 2.4.8.2.3. *Uncropped gels from Fig. 2.2.4.2.B. 6% TBE gel of three nucleosome reconstitution samples with H3K36me3 octamer and 145 bp DNA. Left= InstantBlue, right= SYBR gold.*

2.4.9 186 bp DNA preparation

2.4.9.1 186 bp DNA plasmid expression and purification

Transformation was carried out; briefly, 0.5 µL of pUC19 plasmid containing a single copy of the 186 bp DNA sequence (LL39, from S. Nadal) was transformed into 10 µL of XL10-gold cells before plating 30 µL onto an LBA plate (LB agar supplemented with 100 µg/ mL ampicillin) and incubating overnight at 37 °C.

After 16 hours (the next morning), a single colony was used to inoculate 100 mL of 2 XYT media supplemented with 100 µg/ mL ampicillin. This was incubated for 6 hours before using to inoculate 9 x 800 mL large-scale cultures (8 mL starter culture into 800 mL). The large cultures were incubated at 37 °C for 20 hours (overnight).

The next morning, each 800 mL culture was split into 2 x 400 mL in 500 mL centrifuge bottles and cells pelleted by centrifugation (F500 rotor, 7000 rpm, 8 minutes at 4 °C). This was done in 3 rounds with the pellet retained in the centrifuge bottle after each round (6 pellets total at the end). Each pellet was resuspended in 120 mL ALS I (50 mM glucose, 25 mM Tris.HCl pH 8.0, 10 mM EDTA). 120 mL ALS II (200 mM NaOH, 1% SDS) was then added to each pellet, shaken vigorously until no clumps remained, and incubated for 10 minutes on ice with occasional shaking and pipette mixing to disrupt any remaining clumps. 210 mL pre-chilled ALS III (4M NaOAc, 2M acetic acid) was then poured carefully into each bottle, mixed by inverting 10 times, and incubated on ice for 20 minutes. Insoluble debris was then pelleted by centrifugation (F500 rotor, 10,000 rpm, 30 minutes at 4 °C) The rotor was then warmed to 20 °C by running for 20 minutes in preparation for the next centrifugation step.

The supernatant was then decanted and filtered through Miracloth which was folded 4 times and secured in a large plastic funnel with paperclips. 0.52 volumes of IPA were then added to the filtrate to precipitate the DNA, which was stood at room temperature for 15 minutes before DNA was pelleted by centrifugation (F500 rotor, 8000 rpm, 30 minutes, 20 °C). The supernatant was discarded, and pellets air-dried by standing centrifuge bottles upside-down (carefully) for 1 hour. The pellets were then redistributed between two 50 mL falcon tubes. Centrifuge bottles were then washed out with 5 mL TE 10/50 each to remove any remaining DNA, and the volume in each falcon tube adjusted to 50 mL with TE 10/50. The solutions were

mixed until homogeneous; Ribonuclease A (Sigma, from bovine pancreas; 120 µL of freezer stock) was then added to each tube, and incubated overnight at 37 °C.

Plasmid was then purified by NaCl/ PEG 6000 precipitation. DNA solutions in each falcon tube were diluted up to 40 mL in TE 10/50 and centrifuged to remove insoluble material (x rotor, 20,000 rpm, 15 min room temp). The supernatant was then decanted into a new falcon tube, and DNA precipitated with 4 mL 4M NaCl and 8 mL 20% PEG 6000 (note this was a mistake; was meant to be 40%). The solutions were incubated for 30 mins at 37 °C with shaking, followed by incubation on ice for 30 mins. Plasmid was then pelleted by centrifugation; the supernatant was decanted and taken iteratively through 3 more rounds of precipitation by adding 1 mL NaCl and 2 mL PEG 6000 each round to increase yield. The DNA pellets were dissolved in TE 10/ 0.1 (20 mL and 30 mL) by shaking overnight at 37 °C or room temp. DNA concentration was measured by Nanodrop with 10 x diluted samples. Total yield= 95.53 mg plasmid (1 x 1.237 mg/ mL in 30 mL, 1 x 2.891 mg/ mL in 20 mL)

2.4.9.2 186 bp DNA excision

A small-scale digestion reaction using NEB HF restriction enzymes was carried out to test isolation of the 186 bp plasmid. The reaction was incubated at 37 °C for 4 hours and subsequently analysed by gel electrophoresis.

	EcoRV HF	**EcoRI HF**	**Both**
10 x CutSmart buffer	2 µL		
Acetylated BSA (10 ug/ uL)	0.2 µL		
EcoRV	0.5 µL	-	0.5 µL
EcoRI	-	0.5 µL	0.5 µL

Plasmid	2 µL (~4 µg)	
MilliQ	15.3 µL	14.8 µL
Total volume	**20 µL**	

Table 2.4.9.2.1. *Components of 186 bp DNA excision reaction (double digest with EcoRV and EcoRI).*

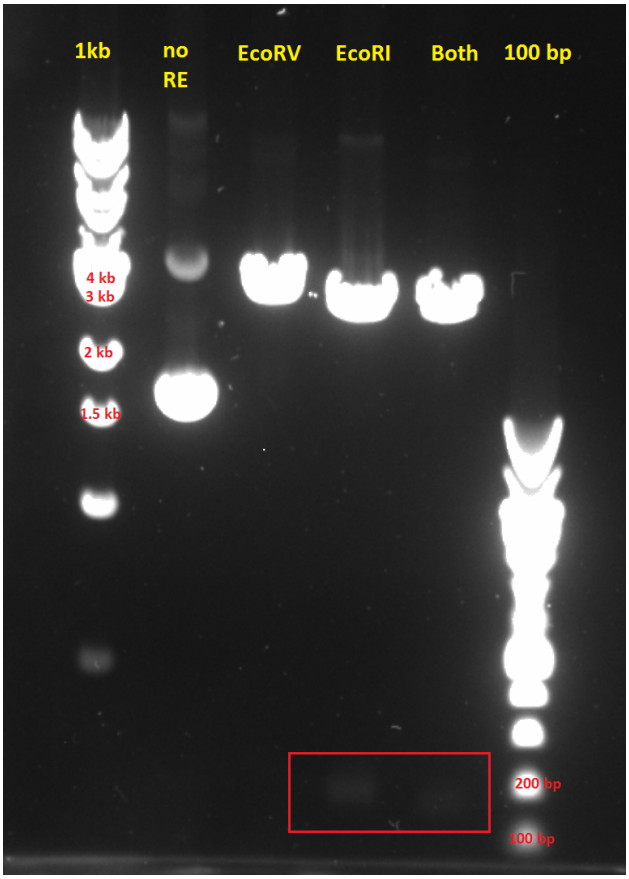


Fig. 2.4.9.2.1. *Characterization of the excision reaction by agarose gel electrophoresis.*

2.4.9.3 186 bp DNA biotinylation and extraction

These were carried out according to protocols by group members Ritu Raj [70] and Simon Nadal, and are described briefly here.

Biotinylation reaction

A solution was prepared as follows:

Component	**Initial concentration**	**Volume added**	**Final concentration**
186 bp DNA	1.2 mg/ mL	4 mL	0.4 mg/ mL
10x NEBuffer 2	10x	1.2 mL	1x
dATP	100 mM	12 µL	100 µM
Biotin-11-dUTP	10 mM	480 µL	40 µM
Klenow 3' to 5' exo	5,000 U/ mL	80 µL	33.3 U/ mL
MilliQ H_2O	6.23 mL	-	-

Table 2.4.9.3.1. *Composition of 186 bp DNA biotinylation reaction.*

This was incubated at 25 °C for 2 h (Buchi bath), then quenched by adding EDTA (480 µL of 0.5 M stock to a final concentration of 20 mM).

Phenol extraction

Phenol (4.8 mL, 2/5 V) and chloroform (2.4 mL, 1/5 V) were mixed with the DNA solution in a 50 mL falcon tube by shaking vigorously, before spinning down in a benchtop centrifuge (8000 rpm, 20 mins) until phases separated. The upper aqueous layer was then removed by pipette into a new 50 mL falcon tube and mixed 1:1 with chloroform (12 mL), shaken, and spun down as before. The aqueous layer was then removed again and chloroform extraction repeated a second time. Finally, to the aqueous layer (11 mL after the 3x extraction steps) was added NaOAc (3M pH 5.2, 1.1 mL, 1/10 V) and EtOH (33 mL, 3V), mixed until homogeneous and stored at 4°C for 4 days, during which time the biotinylated DNA precipitated. The precipitate was resuspended and divided between 2x opaque 40 mL centrifuge tubes, spun down (13,000 rpm, JA25.50 rotor, 4°C, 30 min) to pellet the biotinylated DNA, which was

transferred to a 2mL Eppendorf tube and dissolved in MilliQ H_2O to a final concentration of 2345.4 ng/ µL as measured by nanodrop.

2.4.9.4 Testing 186 bp DNA for degradation..

A 1% agarose gel (50 mL with 15 µL SYBR-safe) was prepared. Samples from several aliquots of 186 bp DNA were prepared (1 µL 6x purple loading dye, 5 µL DNA), loaded onto the gel and run (100 V 30 min).

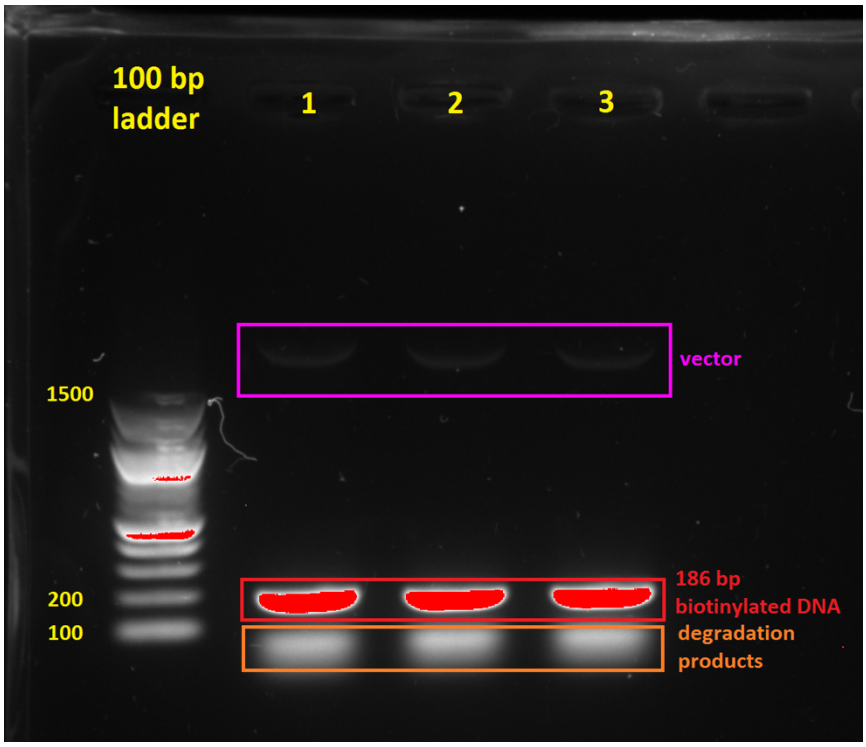


Fig. 2.4.9.4.1. *Agarose gel electrophoresis analysis of three aliquots of 186 bp DNA after biotinylation reaction and freeze-thawing (stored at -20 ℃)*

2.4.9.5 Attempt to re-purify DNA

Phenol / chloroform extraction

Phenol (572 µL, 2/5 V) and chlofororm (286 µL, 1/5 V) were mixed with the DNA solution in a 50 mL falcon tube by shaking vigorously, before spinning down in a benchtop centrifuge

(8000 rpm, 20 mins) until phases separated. The upper aqueous layer was then removed by pipette into a new 50 mL falcon tube and mixed 1:1 with chloroform (1.3 mL), shaken, and spun down as before. The aqueous layer was then removed again and chloform extraction repeated a second time. Finally, to the aqueous layer (1.1 mL after the 3x extraction steps) was added NaOAc (3M pH 5.2, 110 µL, 1/10 V) and EtOH (3.3 mL, 3V), mixed until homogeneous and stored at 4°C for 4 days, during which time the biotinylated DNA precipitated. The precipitate was resuspended and transferred to an opaque 40 mL centrifuge tubes, spun down (13,000 rpm, JA25.50 rotor, 4°C, 30 min) to pellet the biotinylated DNA, which was transferred to a 2mL Eppendorf tube and dissolved in MilliQ H_2O to a final concentration of 3343 ng/ µL as measured by nanodrop.

2.4.5.6 Southern blot of 186 bp DNA

A 6% TBE gel was pre-run in 1x TBE buffer (150 V, 90 min, 4°C), after which the buffer was mixed and samples loaded:

Sample	186 bp DNA (µL)	MilliQ (µL)	186 bp DNA (µg)	Glycerol (µL)	6x loading dye (µL)	Final V (µL)
1	10	-	200	2	3	15
2	1	9	20	2	3	15

Table 2.4.5.6.1. *Composition of samples for Southern blot*

The gel was run until bands from the loading dye reached the bottom (150 V, 75 min, 4°C), stained with SYBR Gold (8 µL in 30 mL MilliQ), and imaged. DNA was then transferred to a membrane using the iBlot kit according to manufacturer's instructions for Southern blotting. Biotin was then visualised on the membrane using a Biotin Chromogenic Detection Kit

(Thermo Scientific); manufacturer's protocol was followed for all steps except blocking, which was performed overnight at 4°C.

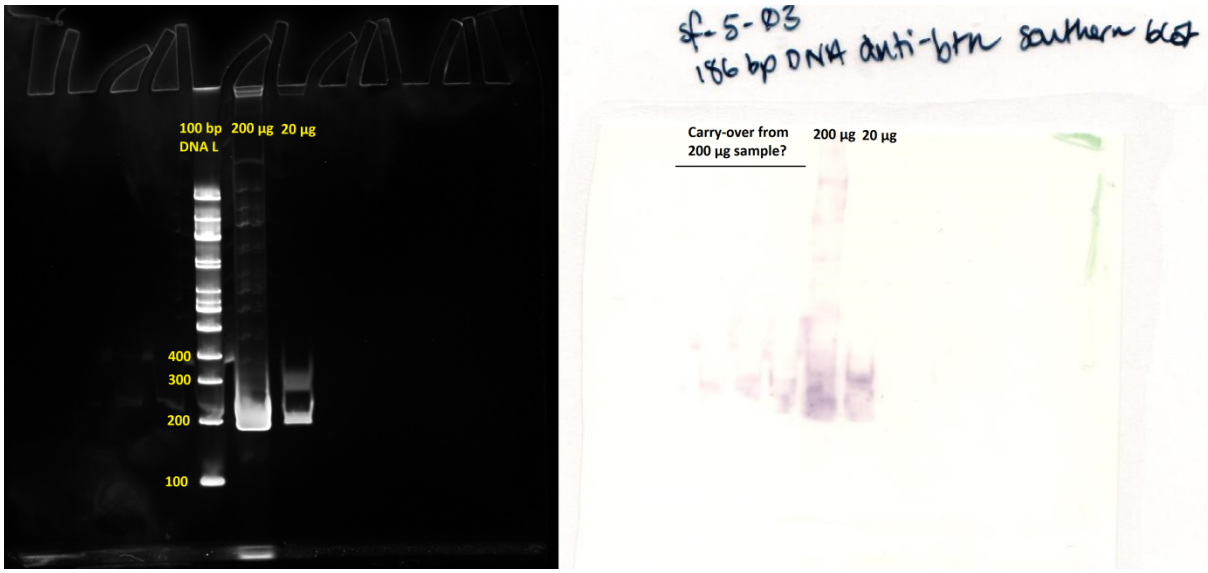


Fig. 2.4.5.6.1. *Anti-biotin Southern Blot of 186 bp biotinylated DNA after chloroform/ phenol re-extraction (uncropped gels from fig. 2.2.5.1.)*

Agarose gel analysis of 186 bp DNA after phenol/ chloroform re-extraction

A 3.5% agarose gel (50 mL with 15 µL SYBR-safe) was prepared. Samples from 186 bp DNA before and after phenol/ chloroform re-extraction were prepared (3 µL 6x purple loading dye, 15 µL DNA), loaded onto the gel and run (50 V 40 min). Note that much of the "before" sample diffused out of the well on loading so concentrations are not consistent between samples.

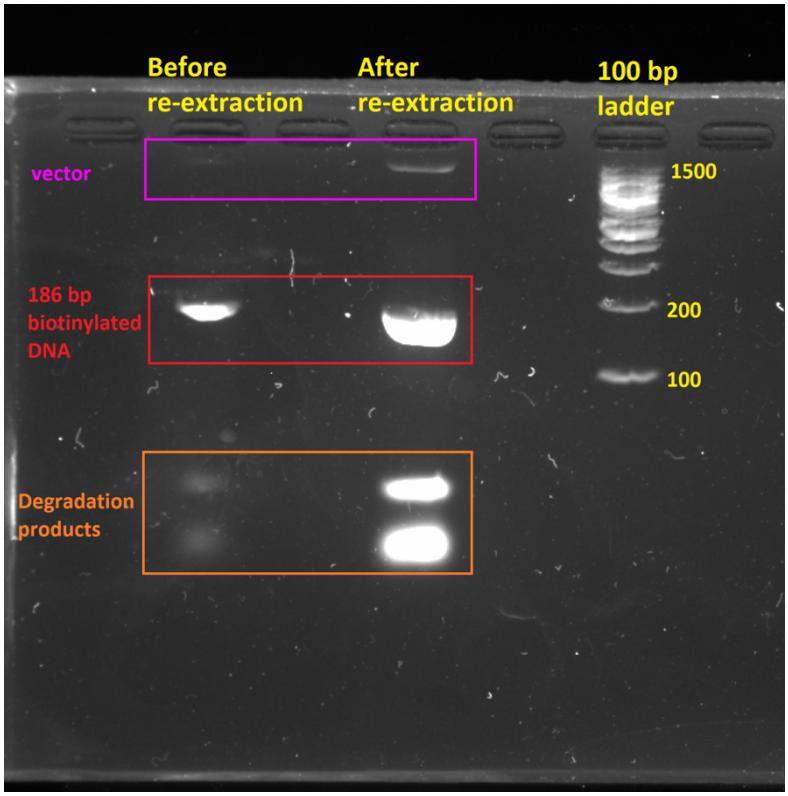


Fig. 2.2.5.3 (Reproduced from Fig. 2.2.5.3.) *Agarose gel electrophoresis analysis of 186 bp biotinylated DNA before and after re-extraction.*

2.4.10 Nucleosome assembly with 186 bp DNA

2.4.10.1 Initial test with standard DNA: octamer ratios

When necessary, an aliquot of octamer (200 µL) was dialysed against 500 mL refolding buffer (2M KCl, 10 mM Tris. HCl pH 7.5, 1 mM DTT, 5 mM β-ME) overnight in a 6 kDa Pur-a-Lyzer mini dialysis membrane to remove glycerol.

Octamer concentration was measured by ND1000 using MW= 108485 Da and E 276 nm = 39120. Concentration of 145 bp DNA (prepared by Simon Nadal) was also determined. Octamer and DNA were then mixed in a 1:0.9, 1:1 and 1:1.1 ratio. 186 bp DNA prepared by another member of the Davis group was also tested as a positive control (A. Lakshmiranayanan, labelled "AL") (see appendix 2.6.6.2. for concentrations and calculations)

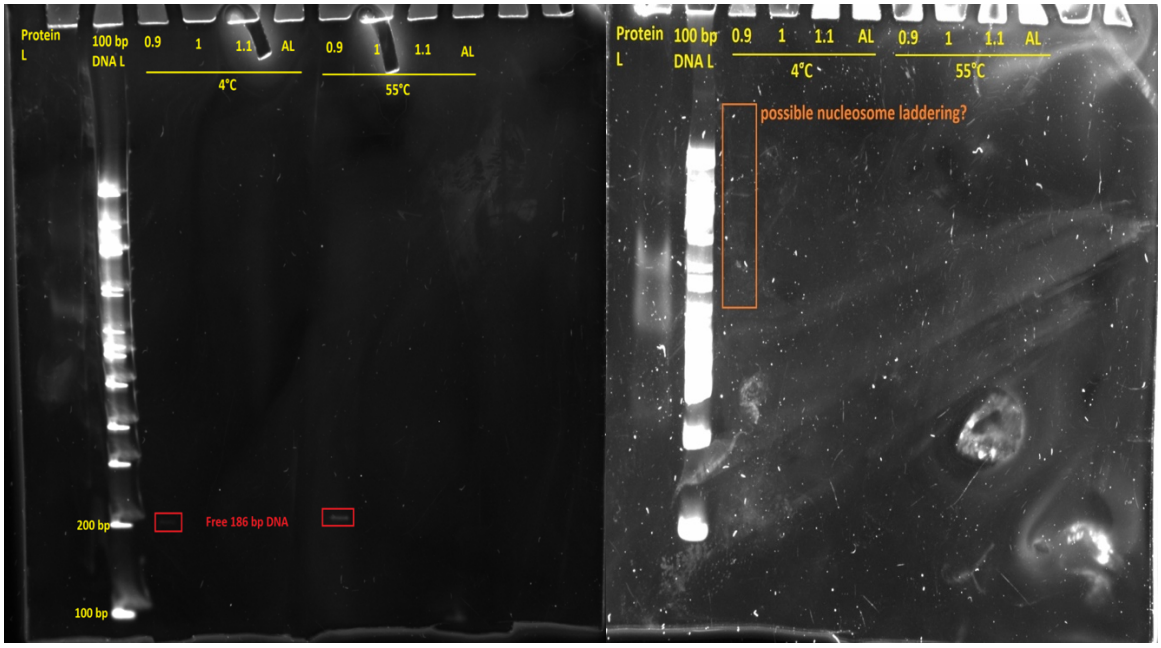
Importantly, the octamer was added last on each instance, and mixed by pipetting once only to avoid damaging the assembled octamer. Each of the DNA/ octamer mixtures were then pipetted into a 6 kDa Pur-a-Lyzer mini dialysis membrane and dialysed sequentially into RRTCS (10 mM Tris. HCl pH 7.5, 1 mM EDTA, 1 mM DTT with decreasing KCl concentrations of 2M, 0.85 M and 0.65 M) for 2 hours each, then RRTCS 0.25 M (0.25 M KCl, 10 mM Tris. HCl pH 7.5, 1 mM EDTA, 1 mM DTT) overnight.

Heat shifting was then performed as follows. Each DNA/ octamer mixture was divided into 3 aliquots and incubated for 2 hours at 4°C, 37°C or 55 °C. During this time a Native-PAGE gel (Novex TBE 6%) was pre-run in 1x TBE buffer (150 V, 90 min, 4°C), after which the TBE buffer was mixed by emptying and re-filling the gel tank.

Samples were prepared by mixing 8 µL nucleosome solution with 2 µL 30% sucrose to a final concentration of 6%; 6 µL of each sample was loaded into the TBE gel wells and run (150 V, 90 min, 4°C). The gel was then rinsed, stained with SYBR gold (6 uL in 30 mL MilliQ for 20

mins) . As no bands were present in SYBR gold staining, the gel was then silver stained as a more sensitive alternative to Coomassie.

SYBR gold visualization



Silver stain visualization

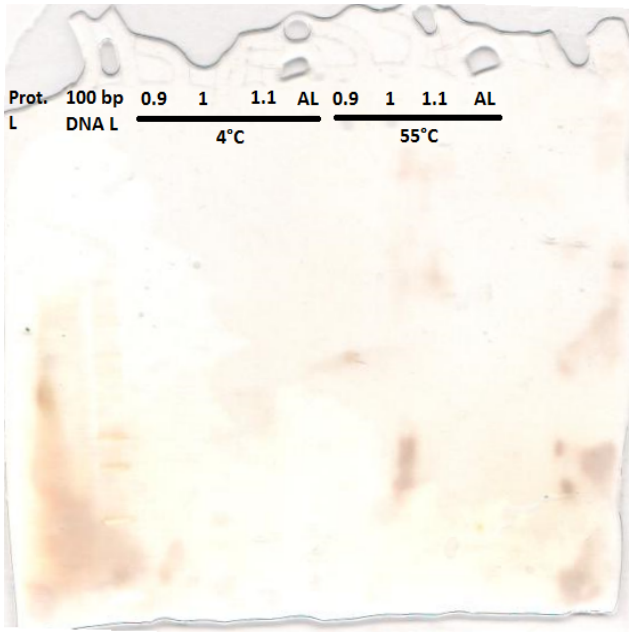


Fig. 2.4.10.1.1. *6% TBE gel of nucleosome reconstitution samples with standard DNA: octamer ratios 0.9:1, 1:1 and 1.1:1. Top= SYBR gold staining (left= standard exposure, right= long exposure), bottom= silver staining).*

2.4.10.2 Testing higher DNA: octamer ratios part 1

Protocol 2.4.10.2 was repeated with higher DNA: octamer ratios included (0.9, 1.0, 1.1, 1.1 and 2). Samples were either incubated at 4°C or heat shifted at 55°C for 2 hours.

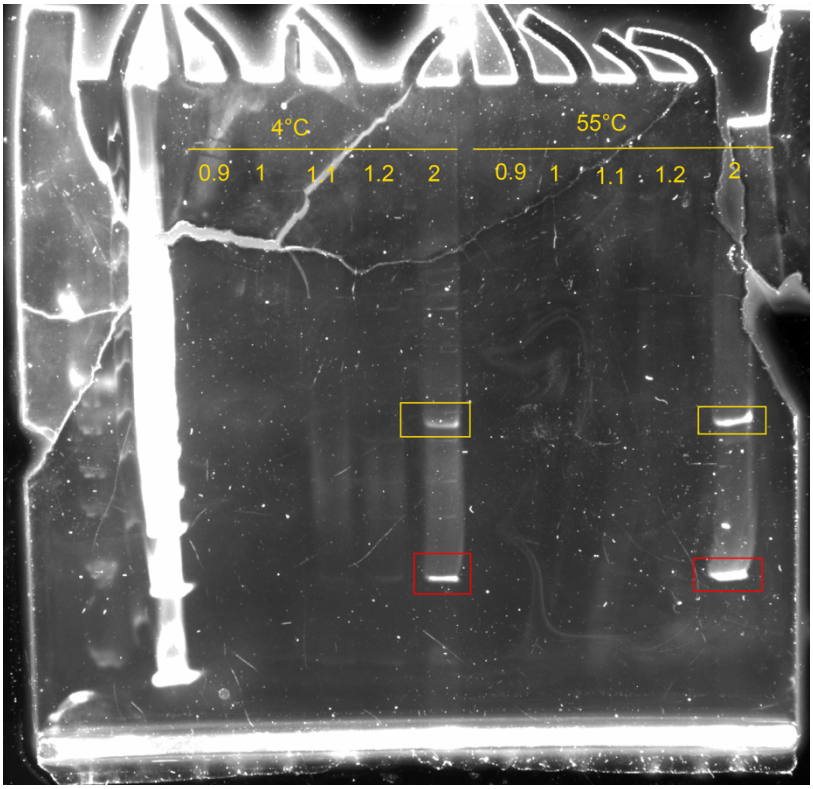


Fig. 2.4.10.2.1*. 6% TBE gel of nucleosome reconstitution samples with higher DNA: octamer concentrations (numbering denotes DNA: octamer concentration screened and temperature of heat shift/ incubation). Yellow box indicates nucleosome, red box indicates free 186 bp DNA.*

2.4.10.3 Testing higher DNA: octamer ratios part 2

Protocol 2.4.10.2 was repeated with higher DNA: octamer ratios included (2, 2.25, 2.5, 2.75 and 3- see appendix 2.6.6.3)). Samples were either incubated at 4°C or heat shifted at 55°C for 2 hours.

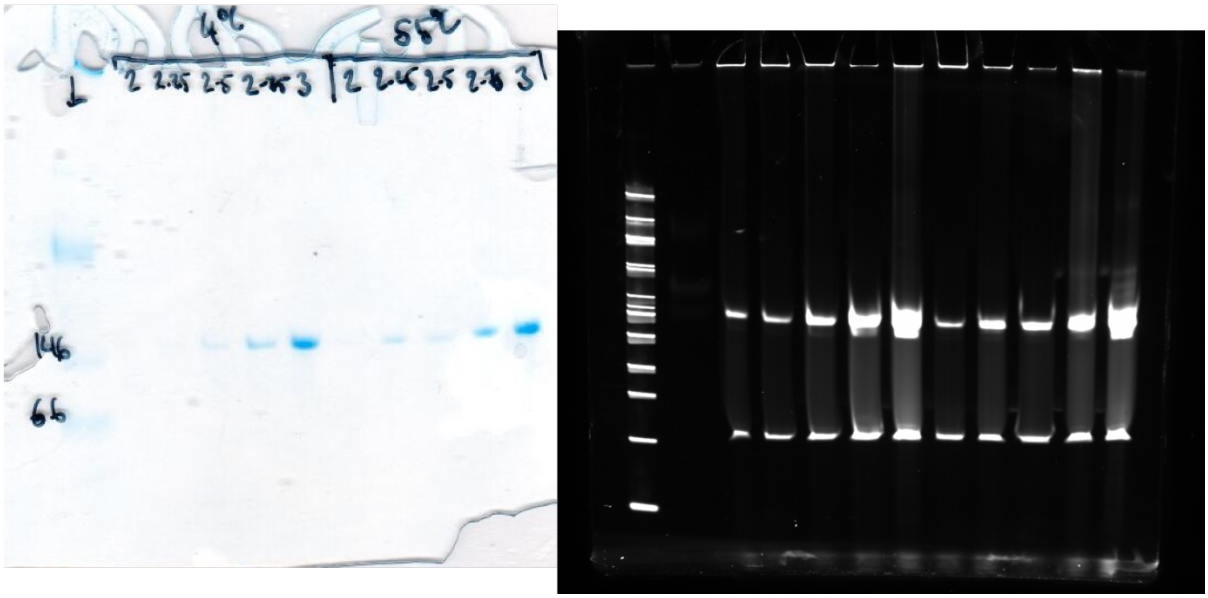


Fig. 2.4.10.3.1. 6% TBE gel of nucleosome reconstitution samples with WT octamer and 186 bp DNA (2:1 to 3:1 DNA- octamer ratios)- left= Coomassie stain, right= SYBR gold. Reproduced from fig. 2.2.5.4.B.

2.4.10.4 Assembly with 186 bp DNA me2 and me3 nucleosomes

Protocol 2.4.10.2 was repeated with H3K36me2 and H3K36me3-containing nucleosomes with DNA: octamer ratios of 2, 2.5 and 3 (2.6.6.4). Samples were heat shifted at 55°C for 2 hours.

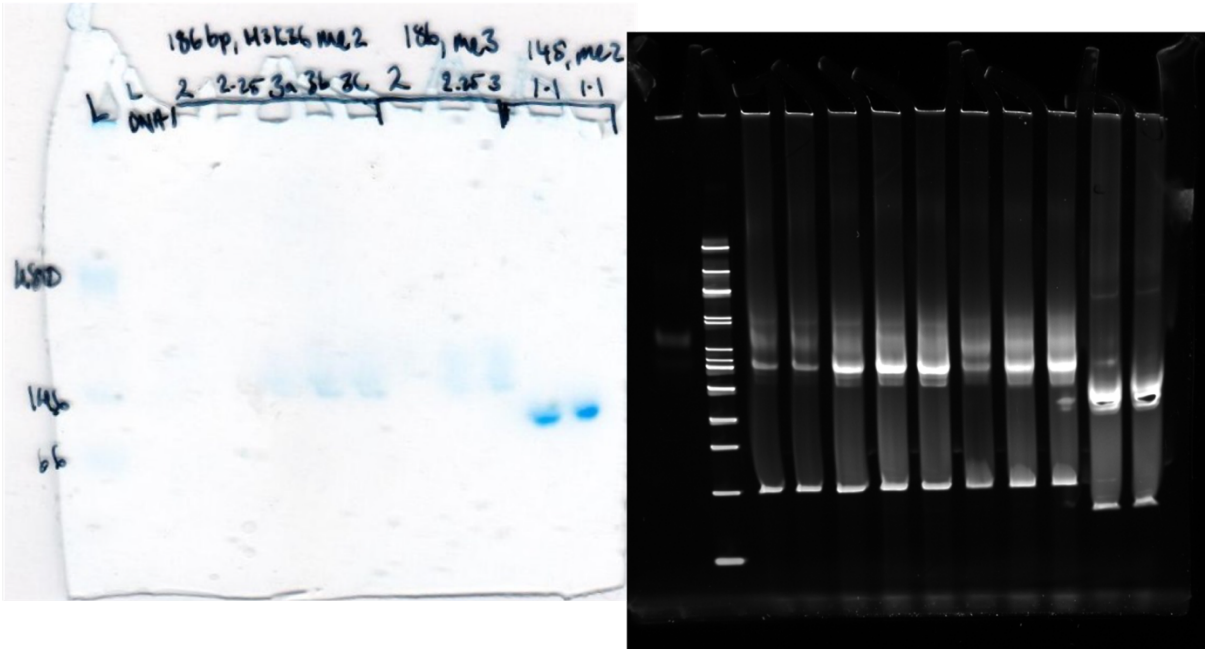


Fig. 2.4.10.4.1*. 6% TBE gel of nucleosome reconstitution samples with me2/ me3 octamer and 186 bp DNA (2:1 to 3:1 DNA- octamer ratios)- left= Coomassie stain, right= SYBR gold. Reproduced from fig. 2.2.5.4.C.*

2.4.11. Gel to test nucleosome composition

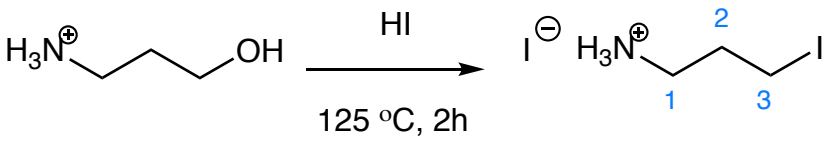
0.5 µg of each 186 bp nucleosome or free H3 was loaded onto a 4-12% Bis-Tris gel (NuPage) and run at 200V for 35 min at 4 °C in MES-SDS. The gel was stained with Coomassie (specs) for 1 hour followed by destaining overnight.



Fig. 2.4.11.1 *SDS-PAGE analysis of nucleosomes. H3 is loaded as control. Uncropped gel from fig. 2.2.5.5.*

2.4.11 Reagent synthesis

(3-iodopropyl) amine hydriodide



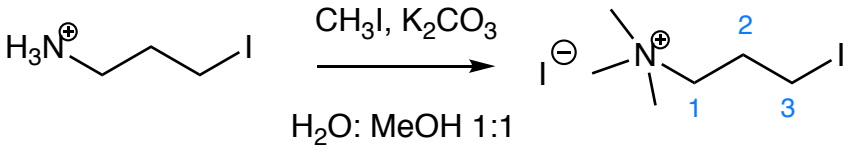
Hydriodic acid (13 mL, 55.9 mMol, 2 equiv.) was chilled to 0 °C on an ice/ brine bath for 5 min. 3-amino-1-propanol (2.0 mL, 26.6 mMol, 1 equiv) was added, and the solution stirred for 5 mins before heating to 125 °C for 2 hours. The reaction was then allowed to cool slowly to room temperature, giving a dark red residue from which the product was isolated by two rounds of recrystallization in $CHCl_3$/ EtOH as colourless prisms (3.96g, 16 mMol, 159-160°C, 60%)

^{1}H NMR (400 MHz, Methanol-*d*$_4$) δ 3.31 (t, *J*= 6.8 Hz, 2H, **H1**), 3.06 (t, *J*= 7.5 Hz, 2H, **H3**), 2.19 (q, *J*= 7.3 Hz, 2H, **H2**)

^{13}C NMR (101 MHz, Methanol-*d*$_4$) δ 41.49 (**C1**), 32.18 (**C2**), 0.47 (**C3**)

LCMS-ESI+ [M]+ Expect 186.02, found 185.98

(3-iodopropyl)- trimethylammonium iodide



3-iodopropyl) amine hydriodide (3.96 g, 16 mMol, 1 equiv.) was dissolved in 1:1 H_2O: MeOH (20 mL). K_2CO_3)10.83 g, 78.5 mMol, 5 equiv) was then added, on which the solution changed from pale yellow to colourless. Iodomethane (4.76 mL, 67 mMol, 4 equiv) was then added dropwise, and the reaction stirred for 1 h at room temperature. Solvents were then removed *in-vacuo* to leave an off-white powder which was re-dissolved into acetonitrile. The solvent was again evaporated *in-vacuo* followed by trituration in $CHCl_3$ to give the product as a white solid which turned yellow on exposure to light (3.41 g, 9.6 mMol, 188-190 °C, 60%).

^{1}H NMR (400 MHz, DMSO-*d*$_6$) δ 3.34 (m, 2H, **H1**) 3.23 (t, *J*= 6.8 Hz, 2H, **H3**), 3.08 (s, 9H, ($C\mathbf{H_3}$)$_3$), 2.23 (dt, *J*= 14.8, 7.0, 2H, **H2**)

^{13}C NMR (101 MHz, DMSO-*d*$_6$) δ 65.86 (**C1**), 52.48 ((CH_3)$_3$), 26.13 (**C2**), 1.80 (**C3**)

LCMS-ESI+ [M]+ Expect 228.02, found 228.0

The values for these compounds are in agreement with literature[63]

2.5 References

1. Castillo, J., López-Rodas, G. & Franco, L. Histone Post-Translational Modifications and Nucleosome Organisation in Transcriptional Regulation: Some Open Questions. in *Protein Reviews: Volume 18* (ed. Atassi, M. Z.) 65–92 (Springer Singapore, 2017). doi:10.1007/5584_2017_58
2. Olins, D. E. & Olins, A. L. Chromatin history: our view from the bridge. *Nat. Rev. Mol. Cell Biol.* **4**, 809–814 (2003).
3. Kornberg, R. D. & Lorch, Y. Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome. *Cell* **98**, 285–294 (1999).
4. Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.* **13**, 115–126 (2012).
5. Widom, J. Toward a Unified Model of Chromatin Folding. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 365–395 (1989).
6. Babu, A. & Verma, R. S. Chromosome structure: euchromatin and heterochromatin. *Int. Rev. Cytol.* **108**, 1–60 (1987).
7. Grunstein, M. *et al.* The regulation of euchromatin and heterochromatin by histones in yeast. *J. Cell Sci. Suppl.* **19**, 29–36 (1995).
8. Thoma, F., Koller, T. & Klug, A. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.* **83**, 403–27 (1979).
9. Robinson, P. J. J., Fairall, L., Huynh, V. A. T. & Rhodes, D. EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6506–11 (2006).
10. Luger, K., Mä Der, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. *Crystal*

structure of the nucleosome core particle at 2.8 A ˚ resolution. *Nature © Macmillan Publishers Ltd* **389**, (1997).

11. Dorigo, B., Schalch, T., Bystricky, K. & Richmond, T. J. Chromatin Fiber Folding: Requirement for the Histone H4 N-terminal Tail. *J. Mol. Biol.* **327**, 85–96 (2003).
12. Kornberg, R. D. Structure of Chromatin. *Annu. Rev. Biochem.* **46**, 931–954 (1977).
13. Dillon, N. Heterochromatin structure and function. *Biol. Cell* **96**, 631–637 (2004).
14. Histone Inheritance. *Science (80-.).* **328**, 15 LP – 15 (2010).
15. Whitlock, J. P. & Simpson, R. T. Localization of the sites along nucleosome DNA which interact with NH2-terminal histone regions. *J. Biol. Chem.* **252**, 6516–6520 (1977).
16. Ausio, J., Dong, F. & van Holde, K. E. Use of selectively trypsinized nucleosome core particles to analyze the role of the histone "tails" in the stabilization of the nucleosome. *J. Mol. Biol.* **206**, 451–463 (1989).
17. Wells, D. E. Compilation analysis of histones and histone genes. *Nucleic Acids Res.* **14**, r119 (1986).
18. Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. Structure of the nucleosome core particle at 7 Å resolution. *Nature* **311**, 532–537 (1984).
19. Hayes, J. J. & Lee, K.-M. In VitroReconstitution and Analysis of Mononucleosomes Containing Defined DNAs and Proteins. *Methods* **12**, 2–9 (1997).
20. Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. . & Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J. Mol. Biol.* **272**, 301–311 (1997).
21. Luger, K., Suto, R. K., Clarkson, M. J. & Tremethick, D. J. Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat. Struct. Biol.* **7**, 1121–1124 (2000).

22. Zhao, Y. & Garcia, B. A. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harb. Perspect. Biol.* **7**, a025064 (2015).

23. Rothbart, S. B. & Strahl, B. D. Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839**, 627–643 (2014).

24. Lawrence, M., Daujat, S. & Schneider, R. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.* **32**, 42–56 (2016).

25. Su, Z. & Denu, J. M. Reading the Combinatorial Histone Language. *ACS Chem. Biol.* **11**, 564–574 (2016).

26. Tan, M. *et al.* Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* **146**, 1016–1028 (2011).

27. Jiang, T., Zhou, X., Taghizadeh, K., Dong, M. & Dedon, P. C. *N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage*. (2006).

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

29. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res. 2011 213* **21**, 381 (2011).

30. Lercher, L. *et al.* Generation of a synthetic GlcNAcylated nucleosome reveals regulation of stability by H2A-Thr101 GlcNAcylation. *Nat. Commun.* **6**, 7978 (2015).

31. Byvoet, P., Shepherd, G. R., Hardin, J. M. & Noland, B. J. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Arch. Biochem. Biophys.* **148**, 558–567 (1972).

32. Murray, K. The Occurrence of iε-N-Methyl Lysine in Histones. *Biochemistry* **3**, 10–15 (1964).

33. Fischle, W., Franz, H., Jacobs, S. A., Allis, C. D. & Khorasanizadeh, S. Specificity of

the Chromodomain Y Chromosome Family of Chromodomains for Lysine-methylated ARK(S/T) Motifs. *J. Biol. Chem.* **283**, 19626–19635 (2008).

34. Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol. Cell* **48**, 491–507 (2012).

35. Paik, W. K. & Kim, S. E-N-dimethyllysine in histones. *Biochem. Biophys. Res. Commun.* **27**, 479–483 (1967).

36. Hempel, K., Lange, H. W. & Birkofer, L. ∈-N-Trimethyllysin, eine neue Aminosäure in Histonen. *Naturwissenschaften* **55**, 37 (1968).

37. Zhang, Y. & Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**, 2343–60 (2001).

38. Hall, I. M. *et al.* Establishment and Maintenance of a Heterochromatin Domain. *Science (80-.).* **297**, 2232 LP – 2237 (2002).

39. Luco, R. F. *et al.* Regulation of Alternative Splicing by Histone Modifications. *Science (80-.).* **327**, 996 LP – 1000 (2010).

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

41. Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13**, 343–57 (2012).

42. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Exp. &Amp; Mol. Med.* **49**, e324 (2017).

43. Bernstein, B. E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci.* **99**, 8695 LP – 8700 (2002).

44. Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature*

419, 407–411 (2002).

45. Wiles, E. T. & Selker, E. U. H3K27 methylation: a promiscuous repressive chromatin mark. *Curr. Opin. Genet. Dev.* **43**, 31–37 (2017).

46. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311 (2007).

47. Sharifi-Zarchi, A. *et al.* DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *BMC Genomics* **18**, 964 (2017).

48. Schulze, J. M. *et al.* Linking Cell Cycle to Histone Modifications: SBF and H2B Monoubiquitination Machinery and Cell-Cycle Regulation of H3K79 Dimethylation. *Mol. Cell* **35**, 626–641 (2009).

49. Mohan, M. *et al.* Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes Dev.* **24**, 574–89 (2010).

50. Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**, 96–99 (2006).

51. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315–326 (2006).

52. Li, B., Howe, L., Anderson, S., Yates, J. R. & Workman, J. L. The Set2 Histone Methyltransferase Functions through the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. *J. Biol. Chem.* **278**, 8897–8903 (2003).

53. Sanidas, I. *et al.* Phosphoproteomics screen reveals akt isoform-specific signals linking RNA processing to lung cancer. *Mol. Cell* **53**, 577–90 (2014).

54. Fang, R. *et al.* Human LSD2/KDM1b/AOF1 Regulates Gene Transcription by Modulating Intragenic H3K4me2 Methylation. *Mol. Cell* **39**, 222–233 (2010).

55. Yoh, S. M., Lucas, J. S. & Jones, K. A. The Iws1:Spt6:CTD complex controls

cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev.* **22**, 3422–3434 (2008).

56. Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **27**, 406–420 (2008).

57. Nimura, K. *et al.* A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf–Hirschhorn syndrome. *Nature* **460**, 287–291 (2009).

58. Ho, T. H. *et al.* High-resolution profiling of histone h3 lysine 36 trimethylation in metastatic renal cell carcinoma. *Oncogene* **35**, 1565–1574 (2016).

59. Zhu, X. *et al.* Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nat. Genet.* **46**, 287 (2014).

60. Newbold, R. F. & Mokbel, K. Evidence for a tumour suppressor function of SETD2 in human breast cancer: a new hypothesis. *Anticancer Res.* **30**, 3309–11 (2010).

61. Hu, M. *et al.* Histone H3 lysine 36 methyltransferase Hypb/Setd2 is required for embryonic vascular remodeling. *Proc. Natl. Acad. Sci.* **107**, 2956–2961 (2010).

62. Zhang, P. *et al.* Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Res.* **34**, 6621–6628 (2006).

63. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science (80-.).* **354**, (2016).

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

65. Dyer, P. N. *et al. [2] Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA*. (2004).

66. Raj, R., Lercher, L., Mohammed, S. & Davis, B. G. Synthetic Nucleosomes Reveal

that GlcNAcylation Modulates Direct Interaction with the FACT Complex. *Angew. Chemie Int. Ed.* **55**, 8918–8922 (2016).

67. Luger, K. Structure and dynamic behavior of nucleosomes. *Curr. Opin. Genet. Dev.* **13**, 127–135 (2003).

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

69. Flaus, A. Principles and practice of nucleosome positioning in vitro. *Front. Life Sci.* **5**, 5–27 (2011).

70. Raj, R. *Final Report*. (2017).

2.6 Appendix

2.6.1 Sequencing

2.6.1.1 General sequencing protocol

All manipulations were carried out in a laminar flow hood. For transformation into competent cells, plasmid (0.5 µL) and XL-10 gold cells (10µL) were incubated together on ice for 30 mins. The cells were then heat-shocked at 42°C for 42 seconds before incubating for 2 minutes on ice to allow recovery, followed by addition of SOC media (300 µL) and incubation at 37°C for 1 hour. The cells were then spread on LBA plates (75 µL on LB agar supplemented with 100 µg/ mL ampicillin) and incubated at 37°C for approximately 20 hours.

For the miniprep, LB broth (10 mL, agar supplemented with 100 µg/ mL ampicillin) was inoculated with a single colony from the previous transformation; this was carried out for each of the three plasmid types. Cultures were incubated for 16 hours before pellets were harvested by centrifugation (8000 rpm for 10 minutes). DNA was then extracted using a QIAprep Spin Miniprep kit according to manufacturer's protocol with some modifications; the final elution step was carried out using a reduced volume of MilliQ H2O (40 µL) with incubation for an extended period of time (10 minutes) before centrifugation to increase final concentration and yield. Plasmid concentration was determined by nanodrop. Samples of each (6 µL of 100 ng/µL) were submitted for Sanger Sequencing using T7F primers. Alignments were carried out using the MUSCLE server and translation from DNA to protein sequence was carried out using the exPASY Translate Tool.

2.6.1.2 Sequencing data for plasmids taken directly from Davis group plasmid bank (H2A, H2B and H3 WT)

H2A

H2A forward DNA sequence

NNNNNNNNNATTTNNTTNACTTTANGAAGGAGATATACATATGTCAGGAAGAGGCAAACAAGGCGGTAAAACCCGCGCTA
AGGCCAAGACTCGCTCATCTCGGGCTGGGCTACAGTTCCCTGTTGGCCGTGTTCACCGGCTGTTAAGGAAAGGCAATTAT
GCAGAGCGGGTGGGAGCTGGAGCTCCAGTCTATCTGGCTGCAGTGTTGGAGTATCTGACCGCTGAGATTTTGGAATTGGC
CGGGAATGCGGCCCGTGATAACAAGAAGACTCGCATTATCCCCAGACACCTGCAGCTCGCTGTGCGCAACGATGAGGAAC
TGAACAAACTGCTCGGAAGAGTCACTATCGCTCAGGGCGGGGTCCTGCCCAACATCCAGTCCGTGCTGCTGCCCAAGAAA
ACCGAGAGTTCCAAGTCGGCCAAGAGCAAGTGATCGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGC
TGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAA
CTATATCCGGATATCCACAGGACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAG
GACTGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTA
GCAGCACGCCATAGTGACTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAG
CCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCATACACGGTGCCTGACT
GCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAATTCTTGAAGA
CGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTT
TCGGGGAAATGTGCGCGGAACCCNTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGANACAATAAC
CCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTT
GCGGCATTTTGCCTTCCTGTTTTTGCTCACCCNGAAACGCTGGTGAAAGTAAAANATGCTGAANNTCAGTTGGGTGCACA
AGTGGGTTACATCGAACTGGATCTCAACAGCGGTAANATCCTTGAANNTTTNNNCCCNGAAAANNGTTTTCCAANNNTNN
NCCNTTTTAAAGTTCTGCNNNNNGGCCNGGNNTTNNCCNNTGTTGACCCCGGNNAAAANCAACTCGNNCGCCNNNANCTN
NTTCNCAAAAGACTTGGTTGAANNNNCCNGNTCCNNAAAANCNNCTTNNGGATGGCNGNNNNNATAAAANTATNNNNGNN
NNNNAANCCTANNNNNAAANCTNGGGNCCANNNATTTTNAANNNTNNNNNGNANCCNAANGTNNNNCCNTTTTTNNNCAA
NNGGGGGNAATTNNGANCCCCNNNNNTNTTNGNAANCCCNTCNAAAANTGTNNTNNNANNNNGNANNTNNGTNNNNNNGC
GTTCAAANTTA

H2A translation

MSGRGKQGGKTRAKAKTRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYLAAVLEYLTAEILELAGNAARDNKKTRII
PRHLQLAVRNDEELNKLLGRVTIAQGGVLPNIQSVLLPKKTESSKSAKSK*

2. Probing H3K36 methylation with synthetically installed PTMs- nucleosome construction

H2A translation alignment with H2A WT reference

```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)


H2A          MSGRGKQGGKTRAKAKTRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYLAAVLEYLT
H2Aref
MSGRGKQGGKTRAKAKTRSSRAGLQFPVGRVHRLLRKGNYAERVGAGAPVYLAAVLEYLT

************************************************************

H2A          AEILELAGNAARDNKKTRIIPRHLQLAVRNDEELNKLLGRVTIAQGGVLPNIQSVLLPKK
H2Aref
AEILELAGNAARDNKKTRIIPRHLQLAVRNDEELNKLLGRVTIAQGGVLPNIQSVLLPKK

************************************************************

H2A             TESSKSAKSK
H2Aref          TESSKSAKSK
                **********
```

H2B

H2B forward DNA sequence

NNNNNNNNNNTNNNNNNTTTGTTTACTTTAAGAAGGAGATATATATGGCCAAGTCCGCTCCAGCCCCGAAGAAAGGCTCC

AAGAAAGCGGTGACCAAGACTCAGAAGAAAGACGGGAAAAAGCGCAGGAAGACAAGGAAGGAGAGTTATGCCATTTACGT

GTACAAGGTGCTGAAGCAGGTGCACCCCGATACCGGCATCTCGTCCAAGGCCATGAGCATCATGAACTCCTTTGTCAACG

ATGTGTTTGAGCGCATCGCAGGGGAAGCCTCCCGCCTGGCTCATTACAACAAGCGCTCCACCATCACCTCCCGGGAGATC

CAGACCGCGGTCCGACTGCTGCTGCCTGGGGAGTTGGCCAAACACGCCGTGTCCGAGGGCACCAANGCTGTCACCAAGTA

AACCAGCGCCCAGTAACTGCTGCTGCCGGATCCGGCTGCTAACAAAGCCCGAAANGAANCTGANTTGGCTGCTGCCATCG

CTGAGCAACAACGAGCATAACCNCTTGGGGCCTCTAAACGGGTCNNGANAGCTNTATCGGTNNNNNANAANNANNNGCNG

GAAAGNTAACNAAACTCNANCCACCNANANGT

2. Probing H3K36 methylation with synthetically installed PTMs- nucleosome construction

H2B_translation

MAKSAPAPKKGSKKAVTKTQKKDGKKRRKTRKESYAIYVYKVLKQVHPDTGISSKAMSIMNSFVNDVFERIAGEASRLAH
YNKRSTITSREIQTAVRLLLPGELAKHAVSEGTXAVTK*

H2B translation alignment with H2B reference

```
LL14_seq        MAKSAPAPKKGSKKAVTKTQKKDGKKRRKTRKESYAIYVYKVLKQVHPDTGISSKAMSIM
H2B_ref         MAKSAPAPKKGSKKAVTKTQKKDGKKRRKTRKESYAIYVYKVLKQVHPDTGISSKAMSIM
                ************************************************************

LL14_seq        NSFVNDVFERIAGEASRLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTXAVTK--
H2B_ref         NSFVNDVFERIAGEASRLAHYNKRSTITSREIQTAVRLLLPGELAKHAVSEGTKAVTKYT
                ********************************************************* ****

LL14_seq        ---
H2B_ref         SAK
```

H3 WT

H3 WT forward DNA sequence

GNNNCNNNNNANATTTTNTTTACTTTAAGAAGGAGATATACCATGGCCCGTACCAAGCAGACCGCCCGTAAATCCACCGG
AGGGAAGGCTCCCCGCAAGCAGCTGGCCACCAAGGCAGCCAGGAAGTCCGCTCCTGCTACCGGCGGAGTCAAGAAACCTC
ACCGTTACCGGCCCGGCACAGTCGCTCTCCGCGAGATCCGCCGCTACCAGAAATCCACCGAGCTGCTCATCCGCAAACTG
CCTTTCCAGCGCCTGGTCCGGGAGATCGCTCAGGACTTCAAGACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGCTCT
GCAGGAGGCCAGCGAGGCTTATCTGGTCGCTCTCTTTGAGGACACCAACCTGGCCGCCATCCACGCCAAGAGGGTCACCA
TCATGCCCAAGGACATCCAGCTGGCCCGCAGAATCCGAGGCGAGAGGGCTTAGATCCGGCTGCTAACAAAGCCCGAAAGG
AAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTT
TTGCTGAAAGGAGGAACTATATCCGGATATCCACAGGACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAA
GTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTGCATC
AACGCATATAGCGCTAGCAGCACGCCATAGTGACTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAG
TACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCA
TACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACA

2. Probing H3K36 methylation with synthetically installed PTMs- nucleosome construction

TGAGAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGA

CGTCNGNGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGC

TCATGAGACANAACCCTGATAAATGCTTCANAATATTGAAAAANNAGAGTATGAGTATTCAACATTTCNNGTCGCCCTTA

TTCCCTTTTTGCGGCATTTTGCCTNCNGTTTTGCTCNCCCAGAAACNCTGNNGAAAGTAAANAATGCTGAAGATNANTNG

NNGCNNNANTGGGNNCNTCNNNNGNNNTNACANNGNAANATCNNGAAANNTTTNNCCCNANANNNNNNNNNATGACNNTTT

NANNNNNCNNTGNNNGCNNNNNNTNANNNNNTNNNNNCNNGNNNNANNNNNNGNNNNNNANANNNNNTTNNNNNNN

A

H3 WT translation and alignment with H3 reference

```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)


   LL1_seq1
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE
   LL1_ref
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE

************************************************************


   LL1_seq1
LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAIHAKRVTI
   LL1_ref
LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAIHAKRVTI

************************************************************


   LL1_seq1       MPKDIQLARRIRGERA
   LL1_ref        MPKDIQLARRIRGERA
                  ****************
```

2.6.1.3 Sequencing data for plasmids generated by site-directed mutagenesis

H3K36C

Four colonies were prepared for sequencing (denoted M1, M2, M3 and M4). M1,2 and 3 all did not sequence

H3K36C M4F DNA sequence

NNNNNNNNNNNNNNNATTTNGTTTACTTTANGAAGGAGATATACCATGGCCCGTACCAAGCAGACCGCCCGTAAATCCAC
CGGAGGGAAGGCTCCCCGCAAGCAGCTGGCCACCAAGGCAGCCAGGAAGTCCGCTCCTGCTACCGGCGGAGTCTGCAAAC
CTCACCGTTACCGGCCCGGCACAGTCGCTCTCCGCGAGATCCGCCGCTACCAGAAATCCACCGAGCTGCTCATCCGCAAA
CTGCCTTTCCAGCGCCTGGTCCGGGAGATCGCTCAGGACTTCAAGACCGACCTGCGCTTCCAGAGCTCGGCCGTTATGGC
TCTGCAGGAGGCCAGCGAGGCTTATCTGGTCGCTCTCTTTGAGGACACCAACCTGGCCGCCATCCACGCCAAGAGGGTCA
CCATCATGCCCAAGGACATCCAGCTGGCCCGCAGAATCCGAGGCGAGAGGGCTTAGATCCGGCTGCTAACAAAGCCCGAA
AGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGT
TTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAGGACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTGGCTC
CAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTGC
ATCAACGCATATAGCGCTAGCAGCACGCCATAGTGACTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGG
CAGTACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATT
TCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAA
ACATGAGAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTT
AGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTNATTTTTCTAAATACATTCAAATATGTAT
CCGCTCATGAANCAATAACCCTGATAAATGCTTCAATAATATTGAAAANGGAANAGTATGANTATTCAACATTTCCNGGN
TCCCCNTTATTCCCTTTTTNGCGGNATTTNGCCTTCCTGTTTTNGCTCACCNAAAANCGCTGGGNAAANGTAAAAAANTN
CTGAAAANNNNATTNGGGTNNNCCAANTGGGGTTACATCNAANCTGGGATCTNAANNNNNGNNAAAANTCCTTNNGANNT
TTNCCCCCCNAAAAANNNGTTTTCCAANNNNNAANCCTTTTNAAANTNNNNNNNTNNTGGGNCCGGNATANNNCCCCGTN
NNGAACCCNGGNTAAAANNAAANCNNGGNNNGCCCNANANNANNNTTNCCNAAAANGGANNNGGNNGNANNNNNCCCCNT
TNCCNAGAAAAANNNNNNTNNNGNAAGGGGNNNNNNCNNAAAAANGATTNNNNNNGGNNGNNNANNNACCGGGGGGNNNA

H3K36C M4F protein sequence

MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVCKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFK
TDLRFQSSAVMALQEASEAYLVALFEDTNLAAIHAKRVTIMPKDIQLARRIRGERA*

H3K36C M4F DNA sequence alignment with H3WT reference

Alignment of Sequence_1: [439681901_H3K36CM4F_T7F_H02.seq.xdna] with
Sequence_2: [438886901_H3WT1F_T7F_C06.seq.xdna]

Similarity : 1343/1600 (83.94 %)

```
Seq_1  1     NNNNNNNNNNNNNNNATTTNGTTTACTTTANGAAGGAGATATACCATGGCCCGTACCAAG  60
             :   |||:|||||:|||||::|||||||||:||||||||||||||||||||||||||||
Seq_2  1     G---NNNCNNNNNANATTTTNTTTACTTTAAGAAGGAGATATACCATGGCCCGTACCAAG  57

Seq_1  61    CAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAGCTGGCCACCAAGGCA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  58    CAGACCGCCCGTAAATCCACCGGAGGGAAGGCTCCCCGCAAGCAGCTGGCCACCAAGGCA  117

Seq_1  121   GCCAGGAAGTCCGCTCCTGCTACCGGCGGAGTC-TGCAAACCTCACCGTTACCGGCCCGG  179
             |||||||||||||||||||||||||||||||  | ||||||||||||||||||||||||
Seq_2  118   GCCAGGAAGTCCGCTCCTGCTACCGGCGGAGTCAAG-AAACCTCACCGTTACCGGCCCGG  176

Seq_1  180   CACAGTCGCTCTCCGCGAGATCCGCCGCTACCAGAAATCCACCGAGCTGCTCATCCGCAA  239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  177   CACAGTCGCTCTCCGCGAGATCCGCCGCTACCAGAAATCCACCGAGCTGCTCATCCGCAA  236
```

```
Seq_1  240   ACTGCCTTTCCAGCGCCTGGTCCGGGAGATCGCTCAGGACTTCAAGACCGACCTGCGCTT  299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  237   ACTGCCTTTCCAGCGCCTGGTCCGGGAGATCGCTCAGGACTTCAAGACCGACCTGCGCTT  296

Seq_1  300   CCAGAGCTCGGCCGTTATGGCTCTGCAGGAGGCCAGCGAGGCTTATCTGGTCGCTCTCTT  359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  297   CCAGAGCTCGGCCGTTATGGCTCTGCAGGAGGCCAGCGAGGCTTATCTGGTCGCTCTCTT  356

Seq_1  360   TGAGGACACCAACCTGGCCGCCATCCACGCCAAGAGGGTCACCATCATGCCCAAGGACAT  419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  357   TGAGGACACCAACCTGGCCGCCATCCACGCCAAGAGGGTCACCATCATGCCCAAGGACAT  416

Seq_1  420   CCAGCTGGCCCGCAGAATCCGAGGCGAGAGGGCTTAGATCCGGCTGCTAACAAAGCCCGA  479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  417   CCAGCTGGCCCGCAGAATCCGAGGCGAGAGGGCTTAGATCCGGCTGCTAACAAAGCCCGA  476

Seq_1  480   AAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCC  539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  477   AAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCC  536

Seq_1  540   TCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAG  599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  537   TCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAG  596

Seq_1  600   GACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAG  659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  597   GACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAG  656
```

```
Seq_1  660   GACTGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTG  719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  657   GACTGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTG  716

Seq_1  720   CATCAACGCATATAGCGCTAGCAGCACGCCATAGTGACTGGCGATGCTGTCGGAATGGAC  779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  717   CATCAACGCATATAGCGCTAGCAGCACGCCATAGTGACTGGCGATGCTGTCGGAATGGAC  776

Seq_1  780   GATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATCCAG  839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  777   GATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATCCAG  836

Seq_1  840   GGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCATACACGGTGCCTGACT  899
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  837   GGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCATACACGGTGCCTGACT  896

Seq_1  900   GCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCA  959
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  897   GCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCA  956

Seq_1  960   AACATGAGAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATG  1019
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  957   AACATGAGAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATG  1016

Seq_1  1020  TCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAA  1079
             ||||||||||||||||||||||||||||||: |:||||||||||||||||||||||||||||
```

```
Seq_2  1017  TCATGATAATAATGGTTTCTTAGACGTCN-GNGGCACTTTTCGGGGAAATGTGCGCGGAA  1075

Seq_1  1080  CCCCTATTTGTTNATTTTTCTAAATACATTCAAATATGTATCCGCTCATGA-ANCAATAA  1138
             ||||||||||||:|||||||||||||||||||||||||||||||||||| | ||: ||
Seq_2  1076  CCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGA-CAN-AA  1133

Seq_1  1139  CCCTGATAAATGCTTCAATAATATTGAAAANGGAANAGTATGANTATTCAACATTTCCNG  1198
             ||||||||||||||||                  || |||  : |  ::| | |   :|
Seq_2  1134  CCCTGATAAATGCTTCA------------------NAATATTGAAAAANNAGAGTATGAG  1175

Seq_1  1199  GNTCCCCNTTATTCCCTTTTTNGCGGNATTTNGCCTTCCTGTTTTNGCTCACCNAAAANC  1258
              :| |  :   |||:: |   :    :    :   | |    |||:||  :|::    :|
Seq_2  1176  TATTCAAC-ATTTCNNGTCGCCCTTATTCCCTTTTTGCGGCATTTTGCCTNCNGTTTTGC  1234

Seq_1  1259  GCTGGGNAAANGTAAAAAANTNCTGAAAANNNNATTNGGGTNNNCCAANTGGGGTTACAT  1318
              |:   : ||: :   :: : : | ||:|::::   :|  ::|:: ::: ::: :|   :
Seq_2  1235  TCNCCCAGAAACNCTGNNGAAAGTAAANAATGCTGAAGATNANTNGNNGCNNNANTGGGN  1294

Seq_1  1319  CNAANCTGGGATCTNAANNNNNGNNAAAANTCCTTNNGANNTTTNCCCCCCNAAAAANNN  1378
             ::: :::::|:::||| :||:| :||  :|    :|:  ||   | : :::|::::|:::
Seq_2  1295  NCNTCNNNNGNNNTNACANNGNAANATCNNGAAANNTTTNNCCCNANANNNNNNNNATGA  1354

Seq_1  1379  GTTTTCCAANNNNNAANCCTTTTNAAANTNNNNNNNTNNTGGGNCCGGNATANNNCCCCG  1438
              ::|| :|:||||:::: :::  |:::|::|:||||::|::::::::|:|::||||::: :
Seq_2  1355  CNNTTTNANNNNNCNNTGNNNGCNNNNNNTNANNNNNTNNNNNCNNGNNNNANNNNNNGN  1414

Seq_1  1439  TNNNGAACCCNGGNTAAAANNAAANCNNGGNNNGCCCNANANNANNNTTNCCNAAAANGG  1498
```

```
             :|||:|: ::|: || :::||::|
Seq_2  1415  NNNNNANANNNN-NTTNNNNNNNA----------------------------------  1437


Seq_1  1499  ANNNGGNNGNANNNNNCCCCNTTNCCNAGAAAAANNNNNNTNNNGNAAGGGGNNNNNNCN  1558

Seq_2  1438  ------------------------------------------------------------  1437


Seq_1  1559  NAAAAANGATTNNNNNNGGNNGNNNANNNACCGGGGGGNNNA  1600

Seq_2  1438  ----------------------------------------  1437
```

H3K36C M4F protein sequence alignment with H3WT reference

```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)



   H3K36CM4
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVCKPHRYRPGTVALREIRRYQKSTE
   H3WT
MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRYQKSTE
                  ************************************
**********************


   H3K36CM4
LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAIHAKRVTI
   H3WT
LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAIHAKRVTI

************************************************************


   H3K36CM4       MPKDIQLARRIRGERA
```

2. Probing H3K36 methylation with synthetically installed PTMs- nucleosome construction

```
H3WT            MPKDIQLARRIRGERA
                ****************
```

H3K36C M4F protein sequence alignment with H3K36C reference

```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)


H3K36CM4        MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVCKPHRYRPGTVALREIRRYQKSTE
H3K36Cref       MARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVCKPHRYRPGTVALREIRRYQKSTE
                ************************************************************

H3K36CM4        LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAIHAKRVTI
H3K36Cref       LLIRKLPFQRLVREIAQDFKTDLRFQSSAVMALQEASEAYLVALFEDTNLAAIHAKRVTI
                ************************************************************

H3K36CM4        MPKDIQLARRIRGERA
H3K36Cref       MPKDIQLARRIRGERA
                ****************
```

H4 WT (C17R)

Four colonies were selected for sequencing (denoted M1, M2, M3 and M4). M2, 3 and 4 did not sequence.

H4 WT M1F DNA sequence

```
ATGTCTGGTCGTGGTAAAGGTGGTAAAGGTCTGGGTAAAGGTGGTGCTAAACGCCACCGTAAAGTTCTGCGTGACAACAT
CCAGGGTATCACCAAGCCGGCTATCCGTCGTCTGGCTCGTCGTGGTGGTGTTAAACGTATCTCCGGTCTGATCTACGAAG
AAACCCGCGGTGTTCTGAAAGTTTTCCTGGAAAACGTTATCCGTGACGCTGTTACCTACACCGAACACGCTAAACGTAAA
ACCGTTACCGCTATGGACGTTGTTTACGCTCTGAAACGTCAGGGTCGTACCCTGTACGGTTTCGGTGGTTAAAGATCCGG
CTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCT
AAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAGGACGGGTGTGGTCGCCATGATCG
CGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCGAGAACG
```

GGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACGCCATAGTGACTGGCGATGCTGTCGGAATGGACGAT
ATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGA
TGAGCGCATTGTTAGATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTA
TCGATGATAAGCTGTCAAACATGAGAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCA
TGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAA
ATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAANGGAAGAGTATGAG
TATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTNGCTCACCCAGAAACGCTGN
TGAAAGTAAAANATGCTGAAANTCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCANCNGCGGTAANNTCNTT
GAGAGTTTTNNCCCCNAAAGAACGTTTTCCAATGATGAACACTTTTAAANTNCNNGCNNNNNGGNNNCGGNNTNNNNTCC
GTGTTGNCCNNGGGCAAGAACAATTNGNTNGCNNNNTAACCTTATTNCCAAANNNNNNTNGGNTTNNNNNNNCCCCNNGN
CCNNGAAAANNNNNNTTNNGGNATGGNNNNNCNNNTNAAAAANNTTATNNNNGGGNTNNCNAAANCCNNNNNGNNAAAAN
CNNGGGNNNNANNNNNNTTTTNGAAAANNNNTNNNNNNNNNNAANGNNNTTAACCNNTTTTTTNNCCAAAANNGGGGGNN
NNNNNNAANNCCCNTNNNNNNTTNNNNAANCCGNNNNTNA

H4 WT M1F protein sequence

MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRK
TVTAMDVVYALKRQGRTLYGFGG*

H4 WT M1F DNA sequence alignment with RV H4 17C reference

Alignment of Sequence_1: [439681901_H4WTM1F_T7F_A02.seq.xdna] with Sequence_2: [439450301_RVH417CF_T7F_C08.seq.xdna]

Similarity : 1491/1871 (79.69 %)

```
Seq_1  1     NNNNNNNNNNNNNNNATTTTGTTTACTTTA-GAAGGAGATATACATATGTCTGGTCGTGG  59
                ||||||||:|||||||:|||:|||||| |||||||||||||||||||||||||||||
Seq_2  1     ---NNNNNNNNANNNATTTNGTTNACTTTANGAAGGAGATATACATATGTCTGGTCGTGG  57

Seq_1  60    TAAAGGTGGTAAAGGTCTGGGTAAAGGTGGTGCTAAACGCCACCGTAAAGTTCTGCGTGA  119
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Seq_2  58    TAAAGGTGGTAAAGGTCTGGGTAAAGGTGGTGCTAAATGCCACCGTAAAGTTCTGCGTGA  117
```

```
Seq_1  120   CAACATCCAGGGTATCACCAAGCCGGCTATCCGTCGTCTGGCTCGTCGTGGTGGTGTTAA  179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  118   CAACATCCAGGGTATCACCAAGCCGGCTATCCGTCGTCTGGCTCGTCGTGGTGGTGTTAA  177

Seq_1  180   ACGTATCTCCGGTCTGATCTACGAAGAAACCCGCGGTGTTCTGAAAGTTTTCCTGGAAAA  239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  178   ACGTATCTCCGGTCTGATCTACGAAGAAACCCGCGGTGTTCTGAAAGTTTTCCTGGAAAA  237

Seq_1  240   CGTTATCCGTGACGCTGTTACCTACACCGAACACGCTAAACGTAAAACCGTTACCGCTAT  299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  238   CGTTATCCGTGACGCTGTTACCTACACCGAACACGCTAAACGTAAAACCGTTACCGCTAT  297

Seq_1  300   GGACGTTGTTTACGCTCTGAAACGTCAGGGTCGTACCCTGTACGGTTTCGGTGGTTAAAG  359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  298   GGACGTTGTTTACGCTCTGAAACGTCAGGGTCGTACCCTGTACGGTTTCGGTGGTTAAAG  357

Seq_1  360   ATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAAT  419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  358   ATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAAT  417

Seq_1  420   AACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAG  479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  418   AACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAG  477

Seq_1  480   GAACTATATCCGGATATCCACAGGACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTG  539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Seq_2  478   GAACTATATCCGGATATCCACAGGACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTG  537

Seq_1  540   GCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCG  599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  538   GCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCG  597

Seq_1  600   AGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACGCCATAGTGA  659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  598   AGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACGCCATAGTGA  657

Seq_1  660   CTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACC  719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  658   CTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACC  717

Seq_1  720   AAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTA  779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  718   AAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTA  777

Seq_1  780   GATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAA  839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  778   GATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAA  837

Seq_1  840   AGCTTATCGATGATAAGCTGTCAAACATGAGAATTCTTGAAGACGAAAGGGCCTCGTGAT  899
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  838   AGCTTATCGATGATAAGCTGTCAAACATGAGAATTCTTGAAGACGAAAGGGCCTCGTGAT  897

Seq_1  900   ACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCAC  959
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  898   ACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCAC  957

Seq_1  960   TTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT  1019
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  958   TTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT  1017

Seq_1  1020  GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAANGGAAGAG  1079
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1018  GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAANGGAAGAG  1077

Seq_1  1080  TATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCC  1139
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  1078  TATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCC  1137

Seq_1  1140  TGTTTTNGCTCACCCAGAAACGCTGNTGAAAGTAAAANATGCTGAAANTCAGTTGGGTGC  1199
            ||||||:||||||||||||||||:||||||||||:||||||||  :||||||||||||
Seq_2  1138  TGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGC  1197

Seq_1  1200  ACGAGTGGGTTACATCGAACTGGATCTCANCNGCGGTAANNTCNTTGAGAGTTTTNNCCC  1259
            |||||||||||||||||||||||||||||:|:|||||||::||:||||||||||::|||
Seq_2  1198  ACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCC  1257

Seq_1  1260  CNAAAGAACGTTTTCCAATGATGAACACTTTTAAANTNCNNGCNNNNNGGNNNCGGNNTN  1319
            |: ||||||||||||||||||||          |:|: ::  ::::: |:::  |::
Seq_2  1258  CG-AAGAACGTTTTCCAATGATGA------GC--ACTTTTAAAGTTCT-GCTAT-GTGG-  1305
```

```
Seq_1  1320  NNNTCC-GTGTTGNCCNNGGGCAAGAACAATTNGNTNGCNNNNTAACCTTATTNCCAAAN  1378
              |  : || || :||:: |   |   :      : |  : :  ::|       ||  |:
Seq_2  1306  --N-GNGGTATTATCCCGTGTTGACCCNGGGC-AA-NAAC-A-ANNCGG-TCG-CCGCAT  1356

Seq_1  1379  NNNNNTNGGNTTNNNNNNNCCCCNNGNCCNNGAAAANNNNNNTTNNGGNATGGNNNNNCN  1438
             ||:|: :  :  :::::::    ::|:  ::   | :::::  :|:  ::   |::::|:
Seq_2  1357  NNCNTATTCTCAGAATGACTTGGTTGAATACT-CACCAGTC-CNNGAAANGCTNCTTACG  1414

Seq_1  1439  NNTNAAAAANNTTATNNNNGGGNTNNCNAAANCCNNNNNGNNAAAANCNNGGGNNNNANN  1498
             ::|: : ::||:   :|::   | |:|: :::  ::||: :|:::| :::  :|||:|:
Seq_2  1415  GATGGNTNNNNNGTAANGAAATNANGCAGNNCTGCCNNACCNNNNNNTGANAACNNNNNG  1474

Seq_1  1499  NNNNTTTTNGAAAANNNNTNNNNNNNNNNAANGNNNTTAACCNNTTTTTTNNCCAAAANN  1558
             |::: ||:|:   :::::: :||:|:::::: |: ::|::::| ::  ::||:: :  :|::
Seq_2  1475  NCCAATTNNNTTCNGAACACNNTNCGGAGGACCAANNNNNCTAACCNNTTTTTNGCNAAA  1534

Seq_1  1559  GGGGGNNNNNNNNAANNCCCNTNNNNNNTTNNNNAANCCGNNNNTNA-------------  1605
             : |||::|||||| ::|:::: |:|||: |:|:: |:||||||:|:
Seq_2  1535  NTGGGGGNNNNNNGNANNNNGCNTNNNTCTTNGGGAACCGNNNNNNNNNNNNCCTANNAA  1594

Seq_1  1606  ------------------------------------------------------------  1605

Seq_2  1595  NNNNCAGAGGNNNANCCCNANTNNNNNNNAANNGGGAAANNNTNNGGCAAAATATNANNG  1654

Seq_1  1606  ------------------------------------------------------------  1605

Seq_2  1655  GGNAAANTNNNNNNNNNGTTNCNNGGNNNNNNNAANNNNCGGGNNNGGNNNNNAAAANTT  1714
```

```
Seq_1  1606  ------------------------------------------------------------  1605

Seq_2  1715  NNNNNNNNTNNCNNCCCNNNNNNNNNNGNNNGGGNNNNNNNNGGNAAAANNNNNNNNCNN  1774


Seq_1  1606  ------------------------------------------------------------  1605

Seq_2  1775  NNNANGGGGNNTNCCNGNNNNNTNNNNNNNNNGGGGNNNAANNNANNNCNTNNNNNNNNN  1834


Seq_1  1606  ------------------------------------  1605

Seq_2  1835  TNNNTNNNNNNCNNNGGNANNNNNNNNNNNANNNTNA  1871
```

H4 WT M1F protein sequence alignment with H4 WT reference

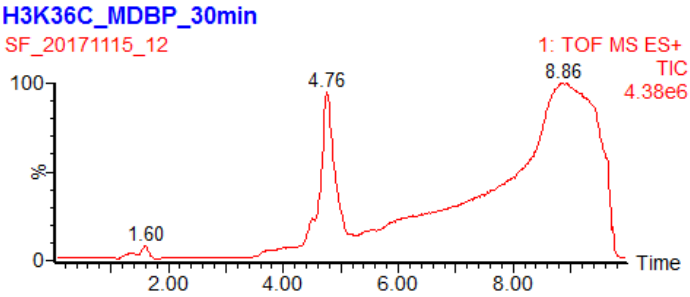
```
CLUSTAL multiple sequence alignment by MUSCLE (3.8)


H4WTM1          MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLK
H4WTref         MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLK
                ************************************************************

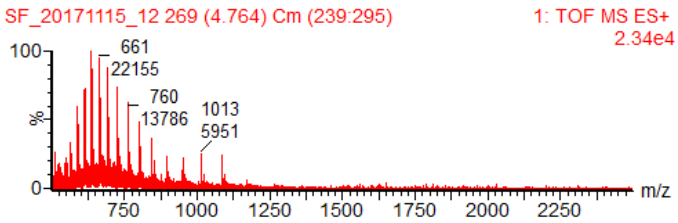
H4WTM1          VFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG
H4WTref         VFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG
                *******************************************
```

2.6.2 LCMS-ESI+ spectra for Dha formation timecourse

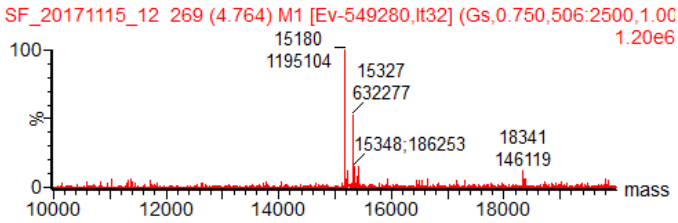
30 min timepoint LCMS-ESI+ timepoint: chromatogram



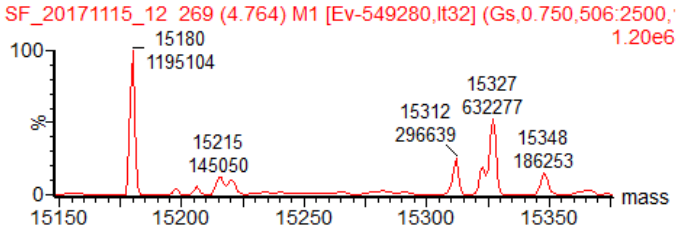
Ion series



Deconvoluted mass spectrum



Deconvoluted mass spectrum (zoom)



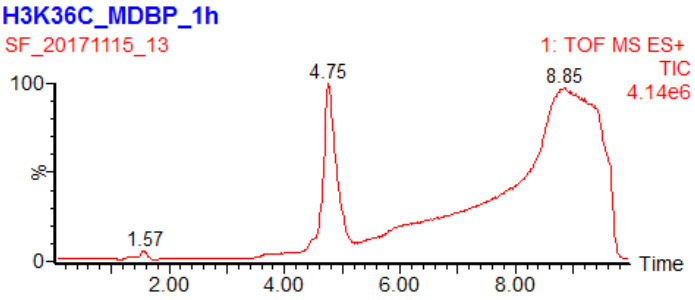
H3K36Dha: Expect 15181 Da, found 15180 Da, 53%

H3K36C: Expect 15214 Da, found 15215 Da, 6% (noise)

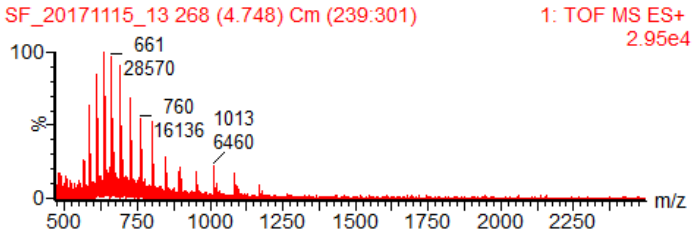
H3K36[carboxylate sulfonium]: Expect 15312 Da, found 15312 Da, 13%

H3K36[methyl ester sulfonium]: Expect 15326 Da, found 15327 Da, 27%

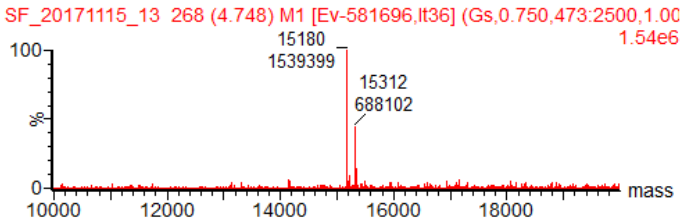
1 h LCMS-ESI+ timepoint: chromatogram



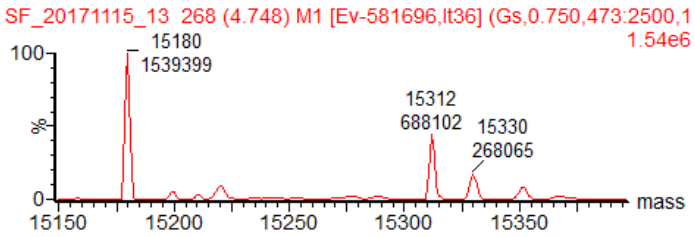
Ion series



Deconvoluted mass spectrum



Deconvoluted mass spectrum (zoom)

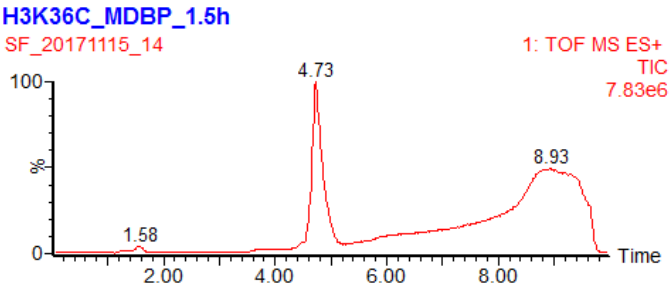


H3K36Dha: Expect 15181 Da, found 15180 Da, 62%

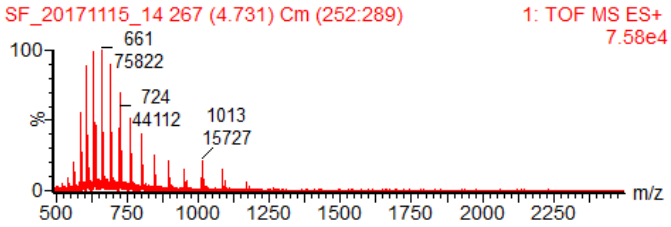
H3K36[carboxylate sulfonium]: Expect 15312 Da, found 15312 Da, 27%

H3K36[methyl ester sulfonium]: Expect 15326 Da, found 15330 Da, 11%

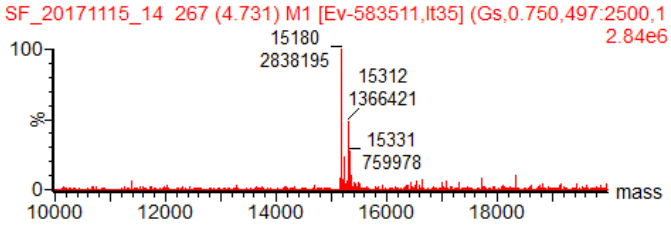
1.5 h timepoint LCMS-ESI+ : chromatogram



Ion series



Deconvoluted mass spectrum

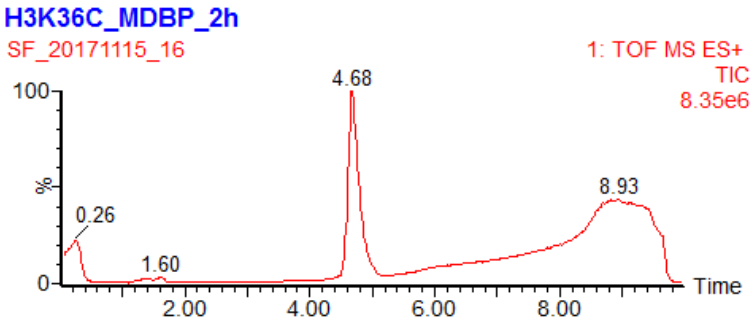


H3K36Dha: Expect 15181 Da, found 15180 Da, 57%

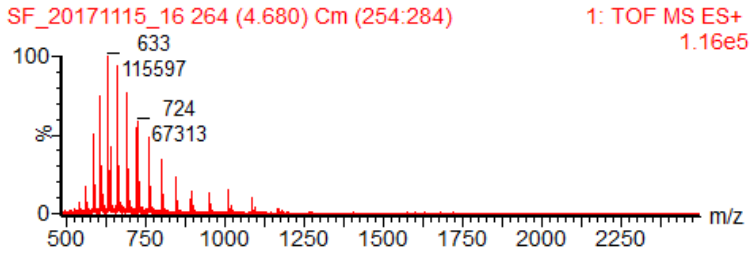
H3K36[carboxylate sulfonium]: Expect 15312 Da, found 15312 Da, 28%

H3K36[methyl ester sulfonium]: Expect 15326 Da, found 15331 Da (?), 15%

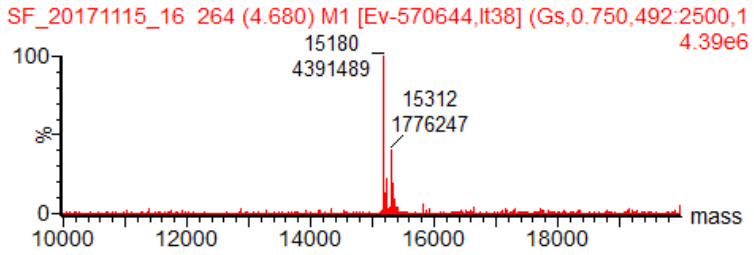
2h timepoint LCMS-ESI+ characterization: chromatogram



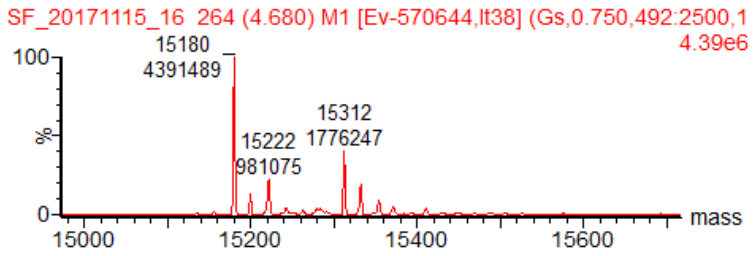
Ion series



Deconvoluted mass spectrum



Deconvoluted mass spectrum (zoom)

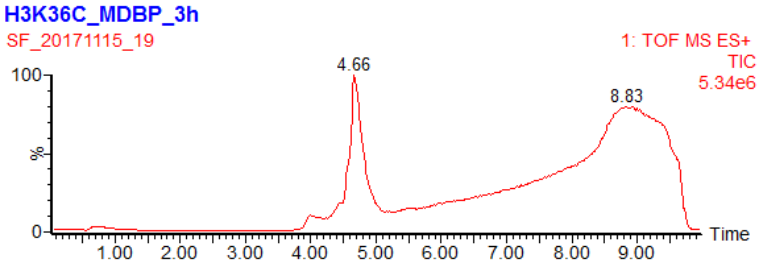


H3K36Dha: Expect 15181 Da, found 15180 Da, 71%

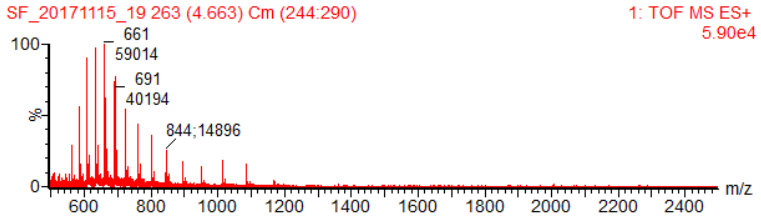
H3K36[carboxylate sulfonium]: Expect 15312 Da, found 15312 Da, 29%

3 h timepoint LCMS-ESI+ characterization: chromatogram

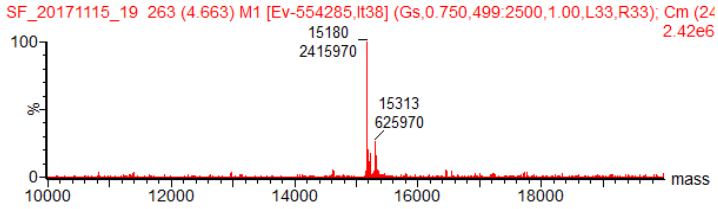
2. Probing H3K36 methylation with synthetically installed PTMs- nucleosome construction



Ion series



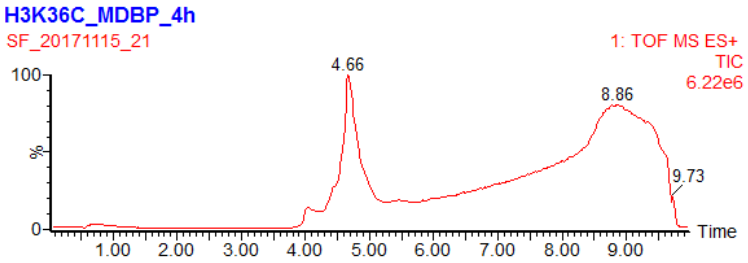
Deconvoluted mass spectrum



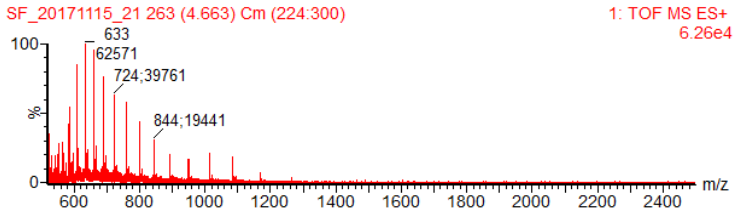
H3K36Dha: Expect 15181 Da, found 15180 Da, 79%

H3K36[carboxylate sulfonium]: Expect 15312 Da, found 15312 Da, 21%

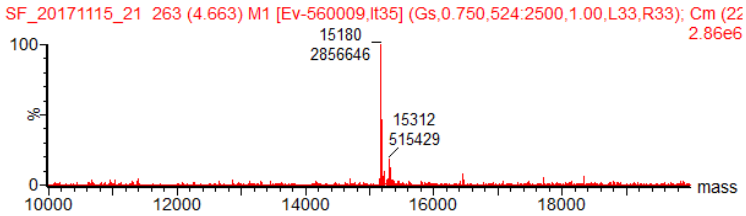
4 h timepoint LCMS-ESI+ characterization: chromatogram



Ion series



Deconvoluted mass spectrum



H3K36Dha: Expect 15181 Da, found 15180 Da, 85%

H3K36[carboxylate sulfonium]: Expect 15312 Da, found 15312 Da, 15%

Note: Reaction was quenched after 5 hours; LCMS-ESI+ spectra from 2.6.3.1 show no remaining carboxylate sulfonium, indicating the reaction went to completion.

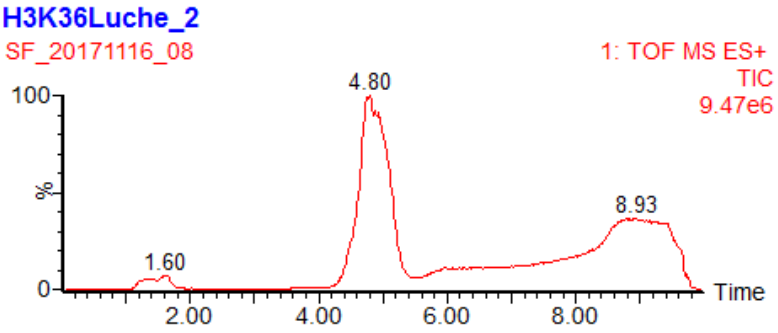
2.6.3 Post-translational mutagenesis optimization

2.6.3.1 Initial uattempt at synthesis of H3K36Me3

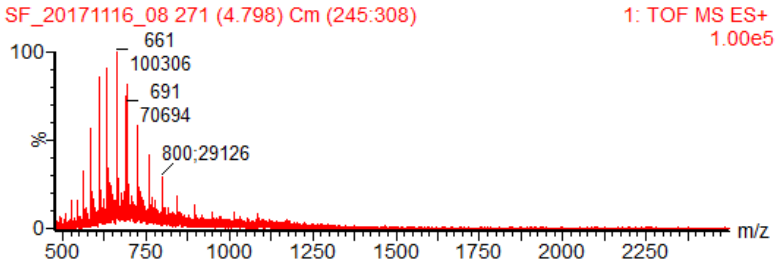
Reactions were done according to protocol established by T. Wright et. al. Following equilibration with N_2 (g) overnight in anaerobic glovebox H3K36Dha (initially 1 mg/ mL, although some water had evaporated overnight) was taken in 11 x 200 µL aliquots and mixed first into 3-iodopropyltrimethylamine (11 x 8 mg) then into $NaBH_4$ (11 x0.5 mg) pre-portioned into 2 mL eppendorf tubes. These were shaken at 4 °C for 30 mins before being passed through GE PD G-25 spintraps pre-equilibrated with unfolding buffer (7M Gdm.HCl, 10 mM Tris pH 7.5, 1 mM DTT). Reactions were then analysed by LCMS-ESI+ to determine reaction conversion. Product was not observed in one reaction out of eleven: LCMS-ESI+ spectrum is displayed below.

LCMS-ESI+ characterization of initial radical addition to Dha using T. Wright protocol

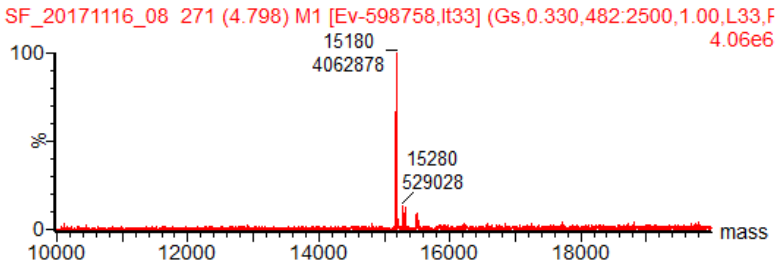
Chromatogram



Ion series



Deconvoluted mass spectrum



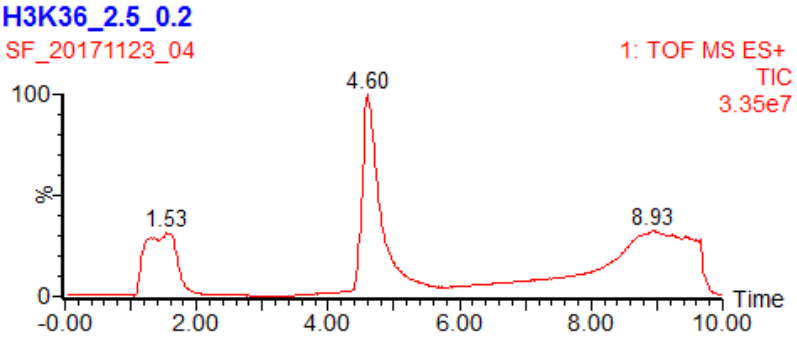
H3K36Dha: Expect 15181 Da, found 15180 Da (88%)

H3K36me3: Expect 15283 Da, found 15280 Da (12%)

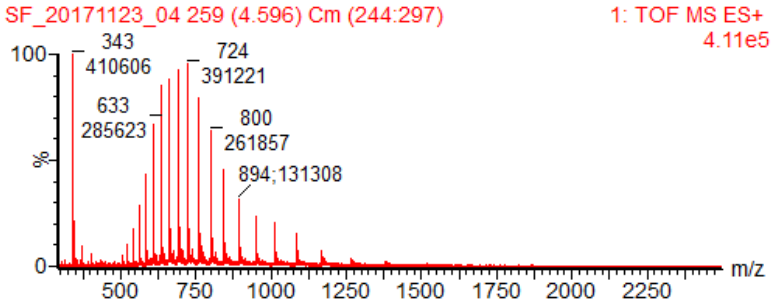
2.6.3.2 LCMS-ESI+ spectra for post-translational mutagenesis radical reaction optimization

2.5 mg iodide/ 0.2 mg $NaBH_4$ LCMS-ESI+ characterization

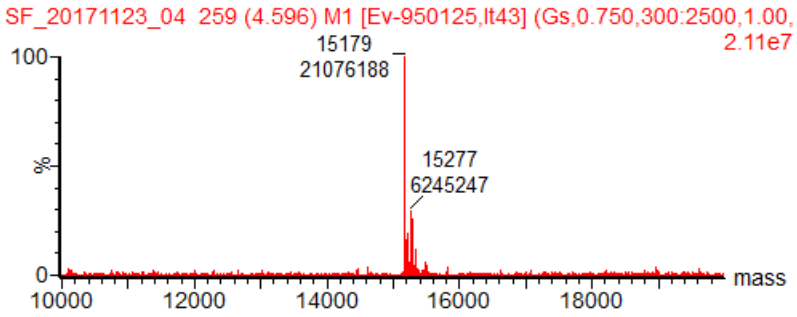
Chromatogram



Ion series



Deconvoluted mass spectrum

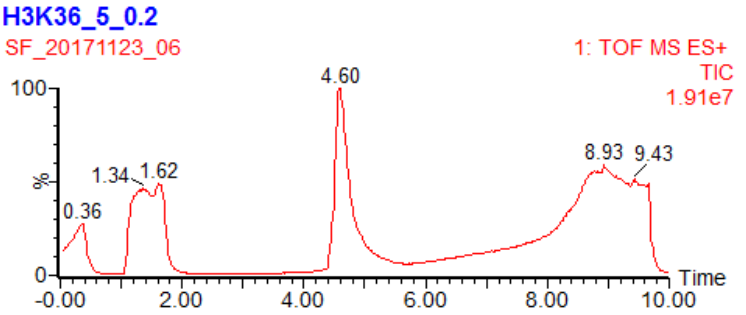


H3K36Dha: Expect 15181 Da, found 15179 Da, 77%

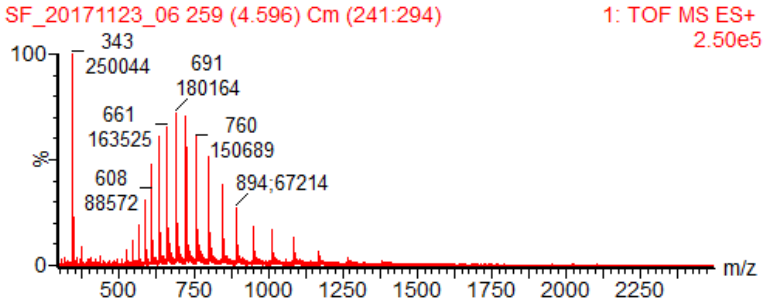
H3K36me3: Expect 15283 Da, found 15277 Da: 22%

0.5 mg iodide, 0.2 mg $NaBH_4$ LCMS-ESI+ characterization

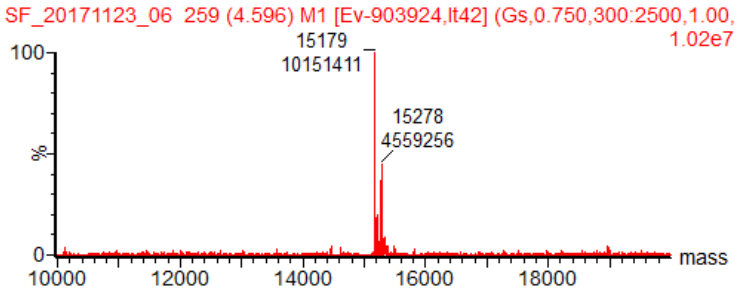
CHromatogram



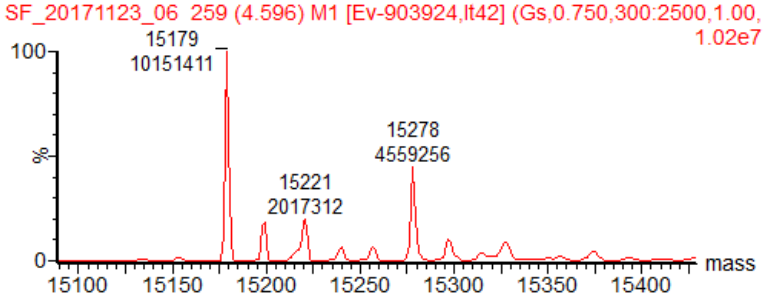
Ion series



Deconvoluted mass spectrum



Deconvoluted mass spectrum (zoom).



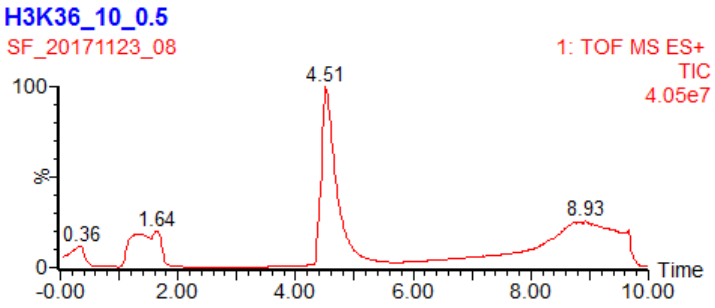
H3K36Dha: Expect 15181 Da, found 15179 Da, 69%

H3K36me3: Expect 15283 Da, found 15277 Da; 31%

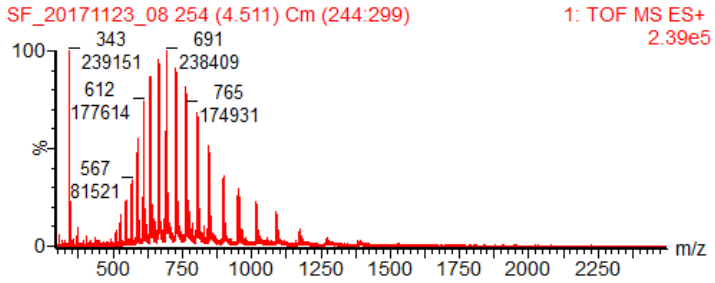
10 mg iodide, 0.5 mg $NaBH_4$ LCMS-ESI+ characterization

2. Probing H3K36 methylation with synthetically installed PTMs- nucleosome construction

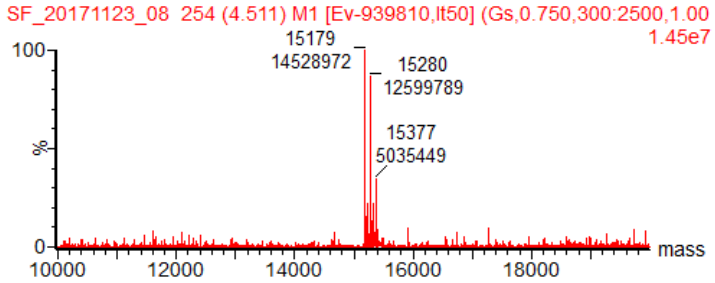
Chromatogram



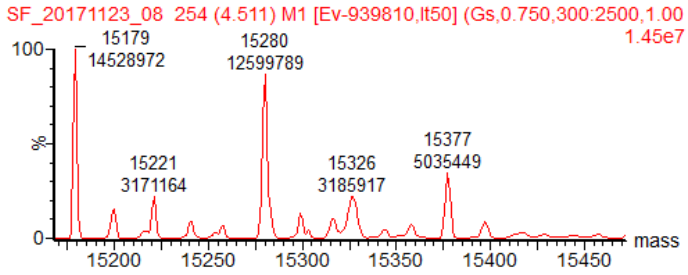
Ion series



Deconvoluted mass spectrum



Deconvoluted mass spectrum (zoom)



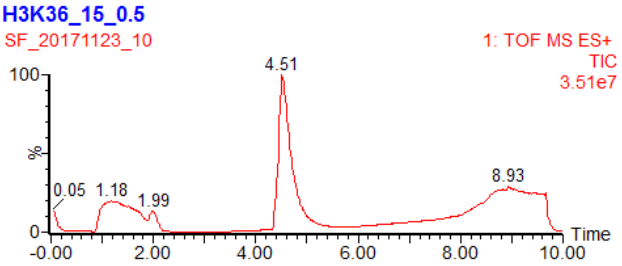
H3K36Dha: Expect 15181 Da, found 15179 Da, 45%

H3K36me3: Expect 15283 Da, found 15277 Da: 39%

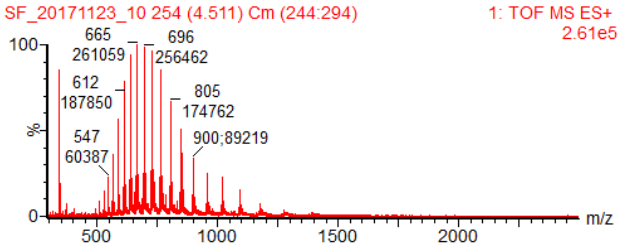
H3K36me3 double addition? : Expect 15385 Da, found 15377 Da; 16%.

<u>15 mg iodide, 0.5 mg $NaBH_4$ LCMS-ESI+ characterization</u>

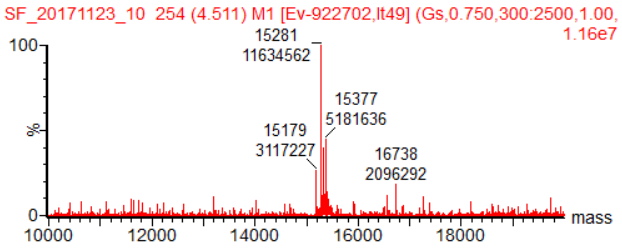
Chromatogram



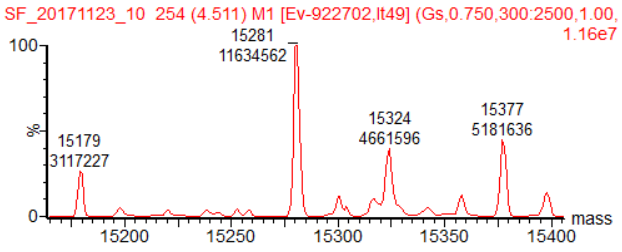
Ion series



Deconvoluted mass spectrum



Deconvoluted mass spectrum (zoom).



H3K36Dha: Expect 15181 Da, found 15179 Da, 16%

H3K36me3: Expect 15283 Da, found 15277 Da: 58%

H3K36me3 double addition? : Expect 15385 Da, found 15377 Da; 26%.

2.6.4 Single histone refolding and CD analysis

Isolated histones were used in some experiments described in Chapter 3; CD was therefore used to examine the structure of the histones.

2.6.4.1 Refolding procedure

Approx. 1 mg each of WT histones H3 and H4 was weighed out into a 1.5 mL Eppendorf tube and dissolved in 800 µL unfolding buffer (6 M Gdn. HCl, 20 mM Tris. HCl pH 7.5, 5 mM DTT, prepared on a 20 mL scale immediately beforehand) by gentle pipetting. The solutions were then incubated at room temperature for 2 hours before dialysing in pre-chilled refolding buffer (2M NaCl, 10 mM Tris. HCl pH 7.5, 1 mM EDTA and 5 mM β-ME) in 3 steps (2L overnight for 15.5 hours, then 2 x 1L for 2h). The refolded histones were then dialysed into CD buffer (250 mM KF, 25 mM NaPi pH 8.6) in 3 steps (2L overnight, then 2 x 1L for 2h). 50 mL of CD buffer was reserved for buffer blanks. Histone concentrations were then measured by nanodrop:

Histone	MW	$\varepsilon_{276\ nm}$	Conc (mg mL^{-1})
H3	15273	4040	1.09
H4	11236	5400	3.16

Table 2.6.4.1. *Extinction coefficients and concentrations for histones H3 and H4*

Note these values are for 276 nm, not the default 280 nm for the ND_{1000} nanodrop.

2.6.4.2 CD analysis of refolded histones- H3 WT and H4

For each histone, a 1 mL sample was made up to approximately 0.3 mg/ mL, with concentration measured accurately via nanodrop:

Histone	MW	$\varepsilon_{276\ nm}$	Conc (mg mL^{-1})
H3	15273	4040	0.36
H4	11236	5400	0.27

Table 2.6.4.1. *Extinction coefficients and concentrations for histones H3 and H4*

Note: these values are for 276 nm, not the default 280 nm for the ND_{1000} nanodrop.

CD spectra of buffer, H3 and H4 were taken using 300 µL samples in a 1mm path length quartz cuvette. Spectra were taken in triplicate from 180 to 280 nm with step=1 and time per point 0.5 s (this range was confirmed to be appropriate by monitoring HT for buffer alone) and processed by averaging the 3 runs and taking smooth (s(0)) with r=4. The data for H3 and H4 was then further processed by subtracting the buffer blank and converting to molar ellipticity [θ] using the formula

$[\theta] = \theta/(10.c.l)$

Where θ = raw CD value, c = concentration in mol L^{-1} and l = path length in cm (1 mm for this cuvette).

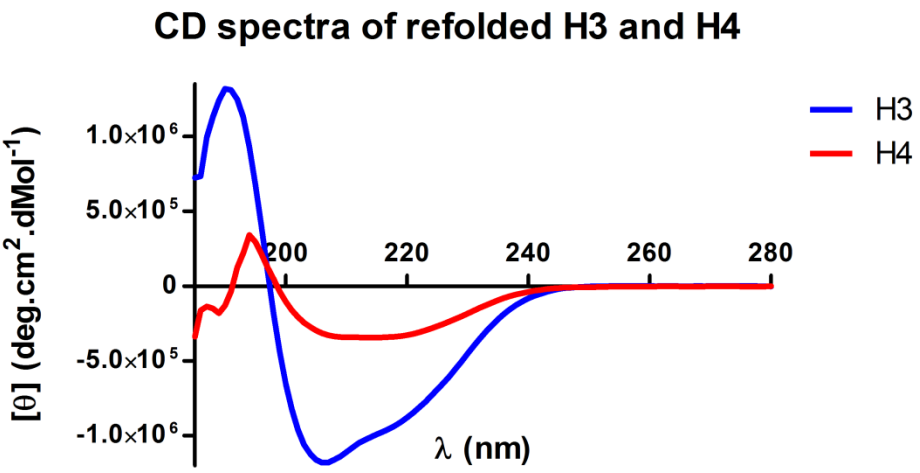


Fig. 2.6.4.2.1. *CD spectra for refolded H3 and H4.*

2.6.4.3 Refolding procedure- H3K36me2 and H3K36me3

These experiments were carried out by Brian Josephson (DPhil, B.G.D. group), and the following data was supplied by him.

H3K36me2 and H3K36me3 from post-translational mutagenesis were dialysed in pre-chilled refolding buffer (2M NaCl, 10 mM Tris. HCl pH 7.5, 1 mM EDTA and 5 mM β-ME) in 3 steps (2L overnight for 15.5 hours, then 2 x 1L for 2h). The refolded histones were then dialysed into CD buffer (250 mM KF, 25 mM NaPi pH 8.6) in 3 steps (2L overnight, then 2 x 1L for 2h). 50 mL of CD buffer was reserved for buffer blanks.

2.6.4.4 CD analysis of refolded histones- H3K36me2 and H3K36me3

These experiments were carried out by Brian Josephson (DPhil, B.G.D. group), and the following data was supplied by him.

For each histone, a 1 mL sample was made up to approximately 0.3 mg/ mL. Note precise concentrations were not determined at this point, and may account for discrepancies in CD curves.

CD spectra of buffer, H3K36me2 and H3K36me3 were taken using 300 µL samples in a 1mm path length quartz cuvette. Spectra were taken in triplicate from 180 to 280 nm with step=1 and time per point 0.5 s (this range was confirmed to be appropriate by monitoring HT for buffer alone) and processed by averaging the 3 runs and taking smooth (s(0)) with r=4. The data for H3K36me2 and H3K36me3 was then further processed by subtracting the buffer blank and converting to molar ellipticity [θ] using the formula

$[\theta] = \theta/(10.c.l)$

Where θ = raw CD value, c = concentration in mol L^{-1} and l = path length in cm (1 mm for this cuvette).

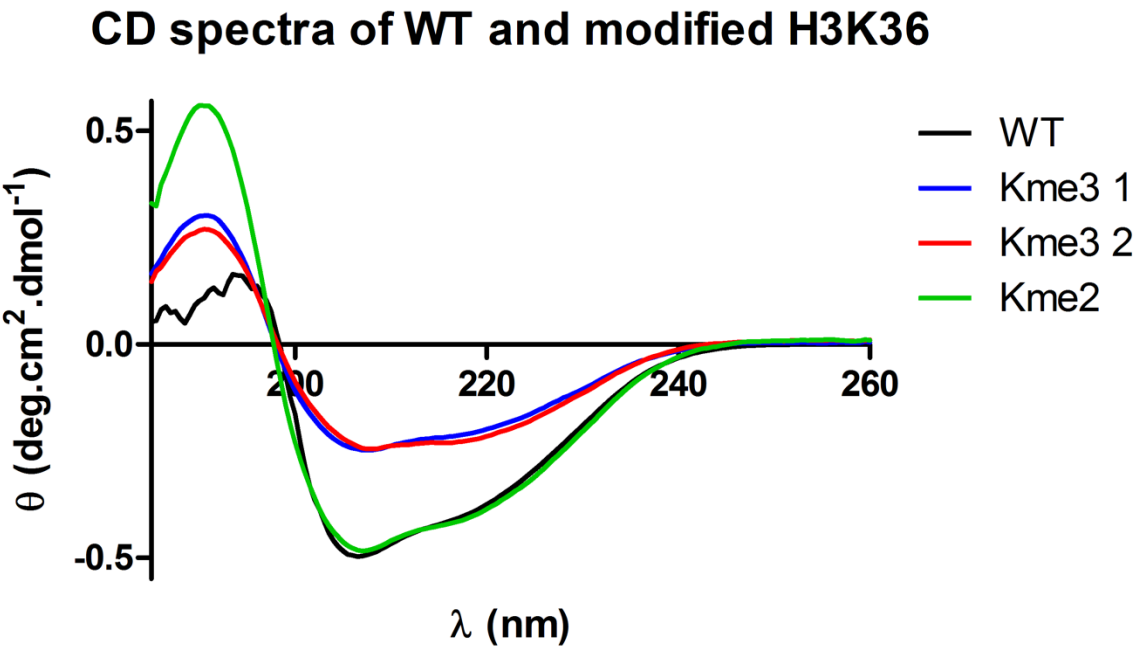


Fig. 2.6.4.2. *CD spectra of WT and modified H3K36.*

2.6.5 Octamer refolding

Histone	MW	$\varepsilon_{276\ nm}$
H2A	13960	4050
H2B	13774	6070
H3	15273	4040
H4	11236	5400

Table 2.6.5.1 *Extinction coefficients used to calculate histone concentrations.*

Histone	MW	ε276	Final μM	conc mg/mL	μmol in 1 mL	final mL for 20 μM	Final V (mL)	μmol histone in final V	μL histone needed for final V
H2A	13960	4050	20	1.65	0.118	5.910	3.6	0.072	609.2
H2B	13774	6070	20	1.33	0.097	4.828	3.6	0.072	745.7
H3 WT	15273	4040	20	1.11	0.073	3.634	3.6	0.072	990.7
H4	11236	5400	20	2.73	0.243	12.148	3.6	0.072	296.3
				Input from ND			*Decide based on lowest amt*		

Final V μL	Additional buffer μL	Average MW	Average ε276	Final g/L
2641.8	958.2	13560.75	4890	1.085

Table 2.6.5.2. *Calculated quantities needed for octamer reconstitution with H3 WT.*

2.6.6 Nucleosome reconstitution

145 bp, 1:0.85

	MW	Conc (mg/ mL)	Conc (μM)	Ratio	Final vol (μL)	μg needed	μL needed	Final conc (mg/ mL)
DNA	90875	5.504	60.57	0.85	*65*	17.8	3.2	*0.27*
octamer	*108486*	0.430	3.96	1	*65*	*25.0*	58.1	
complex	199361					42.8		
TCS buffer							0.4	
4M KCl							3.2	

145 bp, 1:1

	MW	mg/ mL	μM	Ratio	Final vol (μL)	μg needed	μL needed	Final conc (mg/ mL)
DNA	90875	5.504	60.57	1	66	20.9	3.8	0.32
octamer	108486	0.43	3.96	1	66	25.0	58.1	
complex	199361					45.9		
TCS buffer							0.3	
4M KCl							3.8	

145 bp, 1:1.15

	MW	mg/ mL	μM	Ratio	Final vol (μL)	μg needed	μL needed	Final conc (mg/ mL)
DNA	90875	5.504	60.57	1.15	67	24.1	4.4	0.36
octamer	108486	0.43	3.96	1	67	25.0	58.1	
complex	199361					49.1		
TCS buffer							0.1	
4M KCl							4.4	

Table 2.6.6.1. *Calculated quantities for nucleosome reconstitution with 145 bp DNA and WT octamer.*

186 bp, 1:0.9								
	MW	**Conc (mg/ mL)**	**Conc (μM)**	**Ratio**	**Final vol (μL)**	**μg needed**	**μL needed**	**Final conc (mg/ mL)**
DNA	114974	3.340	29.08	0.9	50	33.4	10.0	0.67
octamer	108486	7.380	68.03	1	50	35.0	4.7	
complex	223460					68.4		
TCS buffer							25.3	
4M KCl							10.0	
186 bp, 1:1								
	MW	**mg/ mL**	**μM**	**Ratio**	**Final vol (μL)**	**μg needed**	**μL needed**	**Final conc (mg/ mL)**
DNA	114974	3.340	29.08	1	50	37.1	11.1	0.74
octamer	108486	7.380	68.03	1	50	35.0	4.7	
complex	223460					72.1		
TCS buffer							23.0	
4M KCl							11.1	
186 bp, 1:1								
	MW	**mg/ mL**	**μM**	**Ratio**	**Final vol (μL)**	**μg needed**	**μL needed**	**Final conc (mg/ mL)**
DNA	114974	3.340	29.08	1.1	50	40.8	12.2	0.82
octamer	108486	7.380	68.03	1	50	35.0	4.7	
complex	223460					75.8		
TCS buffer							20.9	
4M KCl							12.2	
186 bp, 1:1;2								
	MW	**mg/ mL**	**μM**	**Ratio**	**Final vol (μL)**	**μg needed**	**μL needed**	**Final conc (mg/ mL)**
DNA	114974	3.340	29.08	1.2	50	44.5	13.3	0.89
octamer	108486	7.380	68.03	1	50	35.0	4.7	
complex	223460					79.5		
TCS buffer							18.6	
4M KCl							13.3	
186 bp, 1:2								
	MW	**mg/ mL**	**μM**	**Ratio**	**Final vol (μL)**	**μg needed**	**μL needed**	**Final conc (mg/ mL)**
DNA	114974	3.340	29.08	2	50	74.2	22.2	1.48
octamer	108486	7.380	68.03	1	50	35.0	4.7	
complex	223460					109.2		
TCS buffer							0.9	
4M KCl							22.2	

Table 2.6.6.2. *Calculated quantities for nucleosome reconstitution with186 bp DNA and WT octamer using standard octamer: DNA ratios.*

186 bp, 1:2

	MW	Conc (mg/ mL)	Conc (μM)	Ratio	Final vol (μL)	μg needed	μL needed	Final conc (mg/ mL)
DNA	114974	3.340	29.08	2	80	74.0	22.2	0.93
octamer	108486	0.980	68.03	1	80	35.0	35.7	
complex	223460					109.2		
TCS buffer							0.0	
4M KCl							22.2	

186 bp, 1:2.25

	MW	mg/ mL	μM	Ratio	Final vol (μL)	ug needed	μL needed	Final conc (mg/ mL)
DNA	114974	3.340	29.08	2.25	85	83.5	24.5	0.98
octamer	108486	0.980	68.03	1	85	35.0	35.7	
complex	223460					109.2		
TCS buffer						118.5	0.0	
4M KCl							24.5	

186 bp, 1:2.5

	MW	mg/ mL	μM	Ratio	Final vol (μL)	μg needed	μL needed	Final conc (mg/ mL)
DNA	114974	3.340	29.08	2.5	91	92.7	27.7	1.02
octamer	108486	0.980	68.03	1	91	35.0	35.7	
complex	223460					127.7		
TCS buffer							0.0	
4M KCl							27.7	

186 bp, 1:2.75

	MW	mg/ mL	μM	Ratio	Final vol (μL)	μg needed	μL needed	Final conc (mg/ mL)
DNA	114974	3.340	29.08	2.75	97	102.0	30.5	1.05
octamer	108486	0.980	68.03	1	97	35.0	35.7	
complex	223460					137.0		
TCS buffer							0.0	
4M KCl							30.5	

186 bp, 1:3

	MW	mg/ mL	μM	Ratio	Final vol (μL)	μg needed	μL needed	Final conc (mg/ mL)
DNA	114974	3.340	29.08	3	102	111.3	33.3	1.09
octamer	108486	0.980	68.03	1	102	35.0	35.7	
complex	223460					146.3		
TCS buffer							0.0	
4M KCl							33.3	

Table 2.6.6.3. *Calculated quantities for nucleosome reconstitution with186 bp DNA and WT octamer using higher octamer: DNA ratios.*

Me2 186 bp, 1:2

	MW	Conc (mg/ mL)	Conc (μM)	Ratio	Final vol (μL)	μg needed	μL needed	Final conc (mg/ mL)
DNA	114974	3.340	29.08	2	102	74.0	22.2	0.727
octamer	108486	5.65	52.06	1	102	35.0	6.2	
complex	223460					109.2		
TCS buffer							51.4	
4M KCl							22.2	

Me2 186 bp, 1:2.5

	MW	mg/ mL	μM	Ratio	Final vol (μL)	ug needed	μL needed	Final conc (mg/ mL)
DNA	114974	3.340	29.08	2.5	102	92.7	27.7	0.909
octamer	108486	5.65	52.06	1	102	35.0	6.2	
complex	223460					127.7		
TCS buffer							40.3	
4M KCl							27.7	

Me2 186 bp, 1:3

	MW	mg/ mL	μM	Ratio	Final vol (μL)	μg needed	μL needed	Final conc (mg/ mL)
DNA	114974	3.340	29.08	3	102	111.2	33.3	1.09
octamer	108486	5.65	52.06	1	102	35.0	6.2	
complex	223460					146.2		
TCS buffer							29.3	
4M KCl							33.3	

Me3 186 bp, 1:2

	MW	Conc (mg/ mL)	Conc (μM)	Ratio	Final vol (μL)	μg needed	μL needed	Final conc (mg/ mL)
DNA	114974	3.340	29.08	2	102	74.0	22.2	0.727
octamer	108486	2.4	22.11	1	102	35.0	14.6	
complex	223460					109.2		
TCS buffer							43.06	
4M KCl							22.2	

Me3 186 bp, 1:2.5

	MW	mg/ mL	μM	Ratio	Final vol (μL)	ug needed	μL needed	Final conc (mg/ mL)
DNA	114974	3.340	29.08	2.5	102	92.7	27.7	0.909

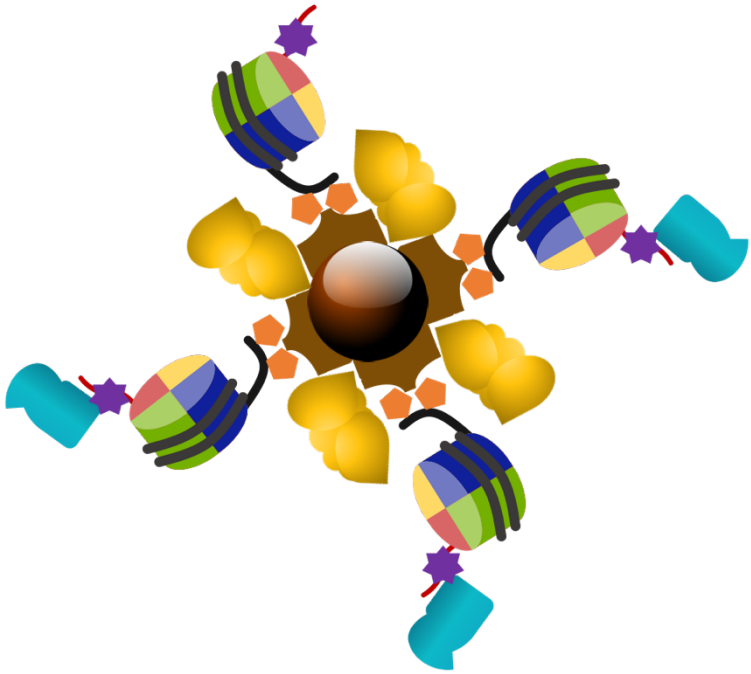
octamer	108486	2.4	22.11	1	102	35.0	14.6	
complex	223460					127.7		
TCS buffer							32.0	
4M KCl							27.7	

Me3 186 bp, 1:3

	MW	mg/ mL	μM	Ratio	Final vol (μL)		μg needed	μL needed	Final conc (mg/ mL)
DNA	114974	3.340	29.08	3	102		111.2	33.3	1.09
octamer	108486	2.4	22.11	1	102		35.0	14.6	
complex	223460						146.2		
TCS buffer								21.0	
4M KCl								33.3	

Table 2.6.6.4. *Calculated quantities for nucleosome reconstitution with186 bp DNA and me2/ me3 octamers using higher octamer: DNA ratios.*

3



Chapter 3: Probing H3K36 methylation with synthetically installed PTMs

PART 2: INTERACTIONS WITH MRG15

TABLE OF CONTENTS FOR CHAPTER 3

LIST OF ABBREVIATIONS FOR CHAPTER 3 187

3.1 INTRODUCTION 191

3.1.1 CHROMODOMAIN PROTEINS ARE A CLASS OF READER PROTEINS RECOGNISING PTMS ON HISTONES 191

3.1.2 MRG15 RECOGNISES H3K36 METHYLATION 192

3.1.3 MRG15 PREFERS H3K36ME3 OVER -ME2 IN VIVO BUT NOT IN VITRO 194

3.1.4 AIMS 198

3.2 RESULTS AND DISCUSSION 199

3.2.1 PULLDOWN STUDIES WITH MRG15 199

3.2.1.1 GST-MRG15 IMMOBILISED VIA ANTIBODY 200

3.2.1.2 GST-MRG15 IMMOBILISED ON GSH SEPHAROSE BEADS 202

3.2.1.3 186 BP NUCLEOSOMES IMMOBILISED ON STREPTAVIDIN DYNABEADS 205

3.2.1.4 SCREENING EFFECT OF SALT IDENTITY AND CONCENTRATION ON MRG15-NUCLEOSOME INTERACTION 207

3.2.2 SURFACE PLASMON RESONANCE (SPR) STUDIES ON MRG15 BINDING TO NUCLEOSOMES 209

3.2.3 PULLDOWN STUDIES WITH W53A NEGATIVE CONTROL 215

3.2.4 DNASE I ASSAYS 216

3.3 CONCLUSIONS 219

3.4 EXPERIMENTAL 222

3.4.1 GENERAL WESTERN BLOT PROTOCOL 222

3.4.1 GST-MRG15 IMMOBILISED VIA ANTIBODY 223

3.4.2 GST-MRG15 IMMOBILISED ON GSH SEPHAROSE BEADS 225

3.4.3 PULLDOWN ASSAYS WITH GSH-SEPHAROSE BEADS 226

3.4.4 PULLDOWN FROM 186 BP NUCLEOSOMES IMMOBILIZED ON STREPTAVIDIN BEADS- GST-MRG15 231

4.4.5 PULLDOWN FROM 186 BP NUCLEOSOMES IMMOBILIZED ON STREPTAVIDIN BEADS- BUFFER SCREEN 234

3.4.6 INITIAL SPR MEASUREMENTS OF MRG15 BINDING TO IMMOBILISED 186 BP NUCLEOSOME 239

3.4.7 BLUE NATIVE PAGE ANALYSIS OF HMRG15 AND GST-MRG15 242

3.4.8 EXPRESSION OF GST-HMRG15 243

3.4.9 PURIFICATION OF HMRG15 244

3.4.10 Further SPR experiments 248

3.4.11 Generation of MRG15 W53A mutant 251

3.4.12 DNAse I digest experiments 256

4.5 REFERENCES 260

3.6. APPENDIX 267

3.6.1. Additional traces for SPR in HBS-EP (3.4.6) 267

3.6.2. Additional traces for SPR in HBK (round 1) 270

3.6.3. Additional traces for SPR in HBK (round 2) 274

LIST OF ABBREVIATIONS FOR CHAPTER 3

Abbreviation	Meaning
2x YT	2x Yeast Tryptone
ADP	Adenosine diphosphate
Alp13	altered polarity 13
Aly	Aly/REF export factor (THO complex subunit 4)
β-ME	β- mercaptoethanol
bp	Base pair
BRCA2	Breast Cancer 2
BSA	Bovine Serum Albumin
C. elegans	*Caenorhabditis elegans*
Dha	Dehydroalanine
DNA	Deoxyribonucleic Acid
DNAse I	Deoxyribonuclease I
dUTP	Deoxyuridine triphosphate
E. coli	*Escherichia coli*
Eaf3p	Esa1p- associated factor-3 protein
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
EHMT2	Euchromatic histone-lysine N-methyltransferase 2
FPLC	Fast protein liquid chromatography
GSH	Glutathione
GST	Glutathione S-transferase
GST-MRG15	GST-tagged MRG15
H1	Histone H1

H2A	Histone 2A
H2A.Z	Histone 2A (Z variant)
H2B	Histone 2B
H3	Histone H3
H3K9	Histone 3 lysine 9
H3K9me3	Histone 3 trimethylated at lysine 9
H3K37	Histone 3 lysine 27
H3K36me2	Histone H3 dimethylated at lysine 36
H3K36me3	Histone H3 trimethylated at lysine 36
H4	Histone H4
HAT	Histone Acetyltransferase
HBS	Hepes-buffered saline
HPS-EP	Hepes-buffered saline with 3 mM EDTA, 0.005% P20
HDAC	Histone Deacetylase Complex
HMT	Histone Methyltransferase
hMRG15	His_6-MRG15
HP-1	Heterochromatin protein 1
HRP	Horseradish peroxidase
IgG_1	Immunoglobin G, class 1
IPTG	Isopropyl-β-D-thiogalactoside
Iws1	Interacts with Spt6 1
KCl	Potassium chloride
KDM4A	Lysine-specific Demethylase 4A
K_D	Equilibrium dissociation constant
k_{on}	Rate constant for ligand binding

k_{off}	Rate constant for ligand dissociation
LBA	Luria-Bertani media with Ampicillin
LB	Luria-Bertani
LSD2	Lysine-specific histone demethylase 2
MDBP	2,5-methyldibromopentonate
MES	2-(N-morpholino)ethane sulfonic acid
MOPS	3-(N-morpholino)propane sulfonic acid
MORF4L1	Mortality factor 4-like protein 1
MRG1	MORF-related gene on chromosome 1
MRG15	MORF-related gene on chromosome 15
MRG2	MORF-related gene on chromosome 2
MRGX	MORF-related gene on chromosome X
mRNA	Messenger ribonucleic acid
MWCO	Molecular weight cut-off
NaCl	Sodium Chloride
Ni-NTA	Nickel- nitriloacetic acid
NHS	N-Hydroxysuccinimide
NuA4	Nucleosome acetyltransferase of H4
OD_{600}	Optical density at 600 nm
PALB2	Partner and Localiser of BRCA2
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline with 0.05% Tween20
Pc	Polycomb
PTM	Post-translational modification
RNA	Ribonucleic acid

RNAPII	Ribonucleic acid polymerase II
Rpd3s	Histone deacetylase Rpd3 small complex
RU	Resonance Units
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate Polyacrylamide gel electrophoresis
SETD2	SET domain containing protein 2 (human)
SOC	Super Optimal broth with Catabolite repression
Spt6	Transcription elongation factor Spt6 (SUPT6H)
SPR	Surface Plasmon Resonance
TBE	Tris/ Borate/ EDTA
TBU	TBE-Urea
Tris	tris(hydroxymethyl)aminomethane
TRRAP	Transformation/transcription domain-associated protein
UV	Ultraviolet
WT	Wild type
X. laevis	*Xenopus laevis*

3.1 Introduction

3.1.1 Chromodomain proteins are a class of reader proteins recognising PTMS on histones

Compared to the overall size of a nucleosome (approx. 200 kDa), a methyl group is relatively small (15 Da) and does not dramatically alter the charge of the histone lysine to which it is appended.[1] This means that, unlike other modifications such as acetylation, methylation does not effect change in chromatin structure by directly modifying the strength of the interaction between histone octamer and DNA. Instead, the function of methylated lysine residues on histones depends heavily on the reader proteins which they recruit.

Many methyllysine reader proteins fall within the Tudor Royal family, a family of proteins folds ubiquitous across eukaryotic species which is thought to have arisen from an ancient methyl substrate binding ancestor.[2,3] This family encompasses the Tudor, plant Agenet, PWWP, Malignant Brain Tumour (MBT), and, most importantly for the purposes of this thesis, chromo- domains.[4]

The chromodomain is a 37 to 50 amino acid motif of three N-terminal anti-parallel β-sheets alongside a C-terminal α-helix.[5] As well as being a member of the Tudor Royal family, it is also a member of the OB class of folds common to archaea, bacteria and eukaryotes.[6,7]

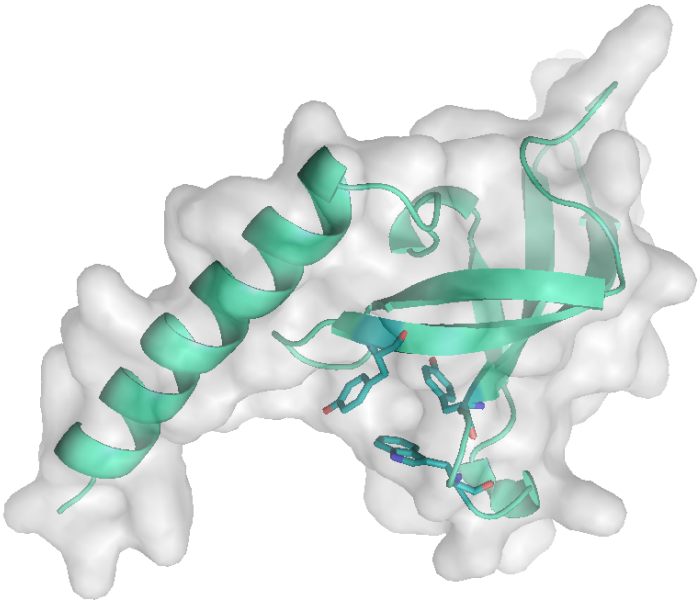
The prefix 'chromo', a portmanteau of 'chromatin organisation modifier', comes from its initial discovery in two *Drosophila* proteins capable of remodelling chromatin into a heterochromatic state to facilitate gene silencing, Polycomb (Pc) and Heterochromatin Protein 1 (HP-1).[5] While chromodomain-proteins in general function to remodel chromatin, this can either activate or repress transcription.[8910] Typically, a chromodomain- containing reader will bind to a particular methyllysine mark on H3 and recruit a larger multimeric complex to alter chromatin structure.[11,12] Often, this multiprotein complex will encompass components capable of

modifying other PTMs such as Histone Acetyl Transferase (HAT) or Histone De-Acetylase Complex (HDAC), which then directly alter the interaction between histone octamer and DNA via modulation of the former's charge.[13–1716] However, many other mechanisms for altering transcription such as recruitment of other complexes or direct binding to RNA have also been found.[19,20]

3.1.2 MRG15 recognises H3K36 methylation

Because chromodomain-containing reader proteins act through recruitment of large complexes, it is possible for one reader protein recognising one methyllysine mark to have a range of possible effects depending on the identity of the other proteins with which it forms a complex.

Exemplifying this is the chromodomain-containing reader MRG15 (MORF-related gene on chromosome 15), also known as MORF4L1 (Mortality factor 4-like protein 1), which is known to display at least four broadly different function depending on which proteins it associates with.[2122]



Fig, 3.1.2.1. *Crystal structure of the chromodomain from MRG15 (PDB ID 2F5K).*

The protein consists of an *N*-terminal chromodomain tethered to a C-terminal MRG domain, with the latter comprised of the nuclear localisation signal, helix-loop-helix, and leucine zipper motifs common to many transcription factors.[23] While the structure of the entire protein has not been solved, each of these domains has been crystallised separately.[24,25]

In humans, the MRG family also encompasses MRGX (MORF-related gene on chromosome X) and MORF4 (Mortality Factor on chromosome 4).[23,26,27] All members share a common *C*-terminal region but differ in their *N*-terminal regions, leading to the conclusion that their common MRG domain mediates protein-protein interactions, while their unique *N*-terminal domains confer different specificities to each of the three proteins. MRG15 in particular is highly conserved with homologs known across a range of species, from *Drosophila* Msl-3 (male specific lethal-3), *C. elegans* MRG-1 (MORF- related gene on chromosome 1), *Arabidopsis* MRG1 and MRG2 (MORF- related genes on chromosomes 1 and 2) and the yeast *Saccharomyces* Alp13 (altered polarity 13), and Eaf3p (Esa1p- associated factor-3 protein), suggesting it is involved in a range of crucial functions.[28–3233]

Indeed, MRG15 has at least four distinct functions. The first identified, and best characterised, is their involvement in chromatin remodelling through recruitment of HATs or HDACs to activate or repress transcription, respectively.[31,34,35] This function has been identified in a range of organisms; Eaf3 forms part of the NuA4 HAT complex in yeast, and MRG15 is similarly implicated in the human TRRAP/ Tip60 complex.[11,28,32,36,37] Likewise for deacetylation, Eaf3 can also associate with Rpd3s in yeast or Sin3b in the human HDAC.[14–17]

A novel mechanism for transcription repression has also recently been identified, as Mrg15 in *Drosophila* stimulates the H3K36 methyltransferase activity of Ash1 to generate the repressive H3K36me2 mark.[38] Evidence has also emerged for a distinct role for MRG15 in facilitating alternative splicing of genes, recruiting PTB (polypyrimidine- tract binding protein) to the alternative exons. [21,39,40]

Finally, MRG15 also plays a role in genome stability, recruiting known homologous recombination (HR) factors PALB2 (Partner and Localiser of BRCA2) and BRCA2 (Breast Cancer 2) to transcriptionally active loci marked by H3K36me3. [41,42]

3.1.3 MRG15 prefers H3K36me3 over -me2 in vivo but not in vitro

In-vivo evidence suggests MRG15 interacts preferentially with H3K36me3 over other methylation states of the residue. MRG15 has been shown to localise to sites containing hyperphosphorylated Ser^2-*p*-RNAPII (RNA Polymerase II phosphorylated at Serine 2), as does the dedicated H3K36me3 methyltransferase SETD2, from which it can be inferred that the reader is specifically binding at sites enriched in H3K36me3.[43] It is also known that modulating the levels of H3K36me3 through manipulating expression of SETD2 or demethylase KDM4A (Lysine-specific Demethylase 4A) consequently impacts the function of MRG15. In the

context of viral splicing, MRG15 recruits PTB to alternatively spliced exons; when H3K36me3 levels are increased by overexpression of SETD2, MRG15- mediated PTB recruitment consequently increases.[44,45] The reverse scenario has also been demonstrated in the context of HR, where MRG15 forms a complex with PALB2 and BRCA2 at sites of recent replication where H3K36me3 is present; upregulating KDM4A reduces the localization of PALB2 to these sites.[42] Moreover, ChIP-seq data has shown a direct link between H3K36me3 and MRG15. Taken together, these observations strongly suggest a preference for H3K36me3 *in-vivo*.[42] However, this specificity has not been replicated *in-vitro*. The yeast homolog Eaf3 has been tested with mono-, di- and tri-methylated H3K36-containing peptides, and no difference in binding was observed between the three; indeed, the di-and trimethyl states both displayed relatively weak KD's of 200-400 μM.[4613] Similarly, when the MRG15 chromodomain was initially characterised, it was found to bind both H3K36me2 and -me3 -containing peptides indistinguishably by Western blot.[24] This result was reinforced recently in the Esashi lab, where studies by Anthony Couturier using full-length synthetic H3K36me2 and -me3 – containing H3 (synthesised in the Davis lab according to methods in Ch. 3) showed no discrimination in MRG15 binding.

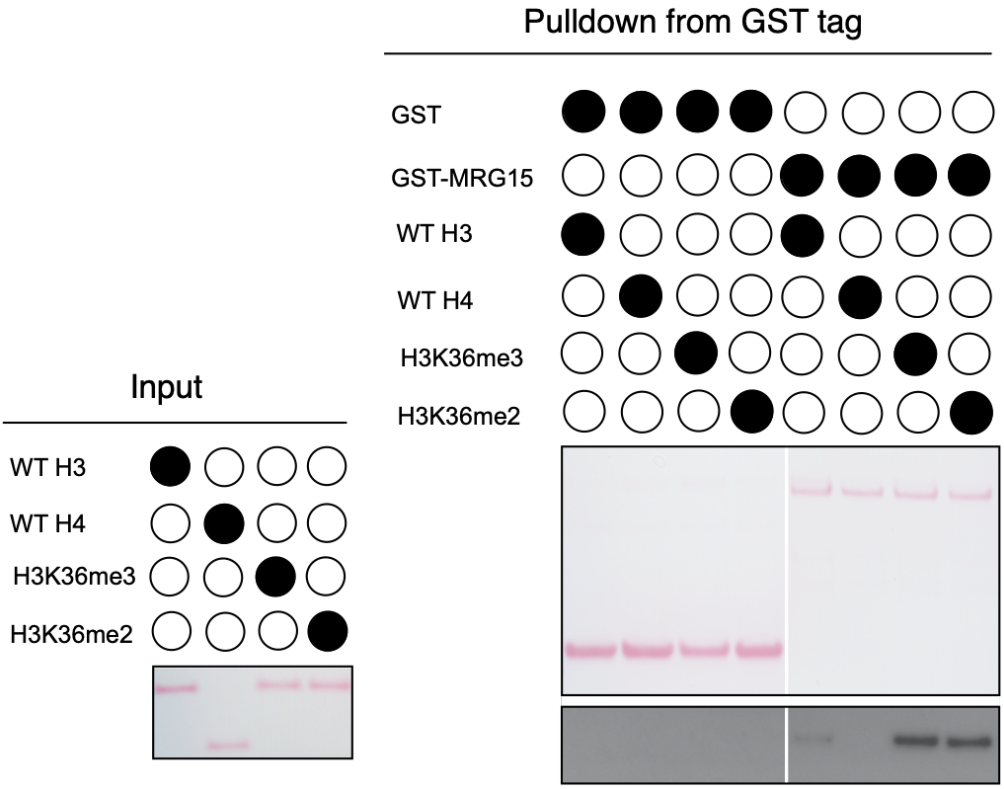


Fig. 3.1.3.1 *Pulldown by A. Couturier using GST-MRG15 immobilized onto sepharose beads as bait to bind the free histones H3 WT, synthetic H3K36me2 or H3K36me3. GST was used as a negative control for bait, and WT H4 as a negative control for the histones.*

This lack of discrimination between di- and trimethyl is in line with other studies on chromodomain protein binding to methyllysine residues nearer the beginning of H3's *N*-terminal tail. A strong molecular mechanism for distinguishing between different methyllysine residues has been elucidated for Pc. It identifies its target residue, methyl H3K27, via recognition of 5 amino acids *N*-terminal to K27; consequently, the difference in K_D with its target, methyl H3K27, and non-target methyl H3K9, is substantial (5μM vs 125 μM). [47]

Conversely, studies on HP-1 interacting with peptides containing its target K9me3 and K9me2 showed little difference between the two methylation states (10 μM vs 15 μM).[48] This is consistent with the mechanism of di- and trimethyllysine recognition by chromodomains,

which has been extensively studied in HP-1 and is thought to occur via a combination of two factors.[49,50] The chromodomain envelops the methyllysine, and the -NMe3 side chain forms interactions with a triad of aromatic tyrosine and tryptophan residues. A small part of this stabilizing interaction is the hydrophobic component between the methyl groups and the aromatic residues; however, binding studies with the non-charged trimethyllysine analog tert-butylnorleucine gave a much weaker affinity for the binding site, suggesting that the greatest contribution is from the cation-pi interactions between the positive charge of *N*-methyllysine and the quadrupole moment of tryptophan.[48] Importantly, neither of these contributions provide a mechanism for distinguishing between di- and tri-methyl. In both cases, the dimethyl state possesses much of the hydrophobicity and similar charge to the trimethyl, and one might therefore expect that it binds with similar affinity.

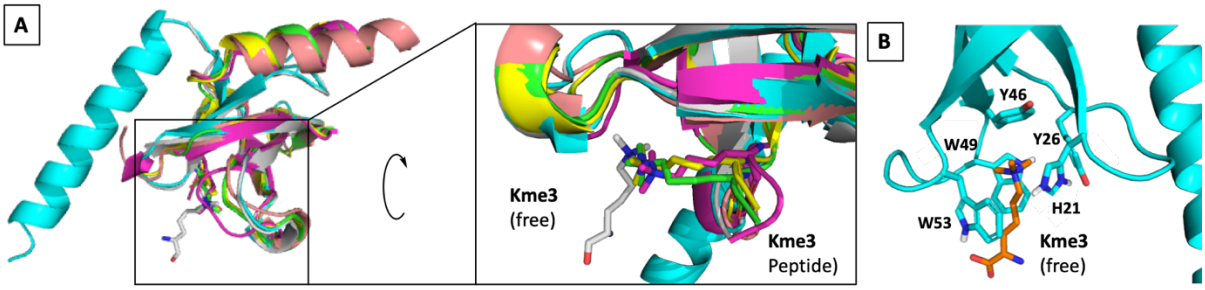


Fig. 3.1.3.2. (A) *Overlay of chromodomains from MRG15 (cyan, 2f5k), HP-1 with H3K9me3 tail (green, 1kne),Chp1 with H3K9me3 peptide (pink, 2rsn), Chp1 with H3K9me3 peptide (yellow, 3g71), MPP8 with with H3K9me3 peptide (salmon, 3qo2), MRG2 with free Kme3 amino acid (grey, 4pl6).* ***(B)*** *Docking structure of free Kme3 in the binding site of MRG15 (performed by T. Watts) with key aromatic residues highlighted.*

The crystal structure of MRG15's chromodomain reveals a similar aromatic triad of residues Y26, Y46 and W49, which are thought to bind H3K36me3 in an analogous manner.[24] Moreover, the MRG15 chromodomain aligns closely with the crystal structures of other

chromodomains, including HP-1, and docking studies suggest free lysine binds in a similar fashion to HP-1.

3.1.4 Aims

How, then, can the strong preference for H3K36me3 over -me2 inferred from *in-vivo* observations be reconciled with the apparent lack of a molecular mechanism for discrimination? In previous studies on chromodomain binding, this question answered by invoking the principle of multivalency, with overall discrimination arising from the cumulative slight preference for trimethyllysine over its dimethyl counterpart over several nucleosome binding sites.

However, inspection of the nucleosome crystal structure reveals another possibility.[1] Unlike the extensively studied HP-1 binding partner H3K9, K36 is located at the base of the H3 tail and is sandwiched between H3K37 and the DNA phospho-ribose backbone. This means that MRG15 would normally be blocked from accessing the residue by the DNA, and in order to surround the lysine side chain in its binding site, a change in conformation must occur to make K36 more accessible. Previous *in-vitro* studies, which used only peptides or isolated H3, would be unable to capture this shielding effect. Indeed, it is known from previous examples in the literature with the NSD1 lysine methyltransferase that mechanisms for substrate recognition can be apparent only on the scale of the entire nucleosome.[51]

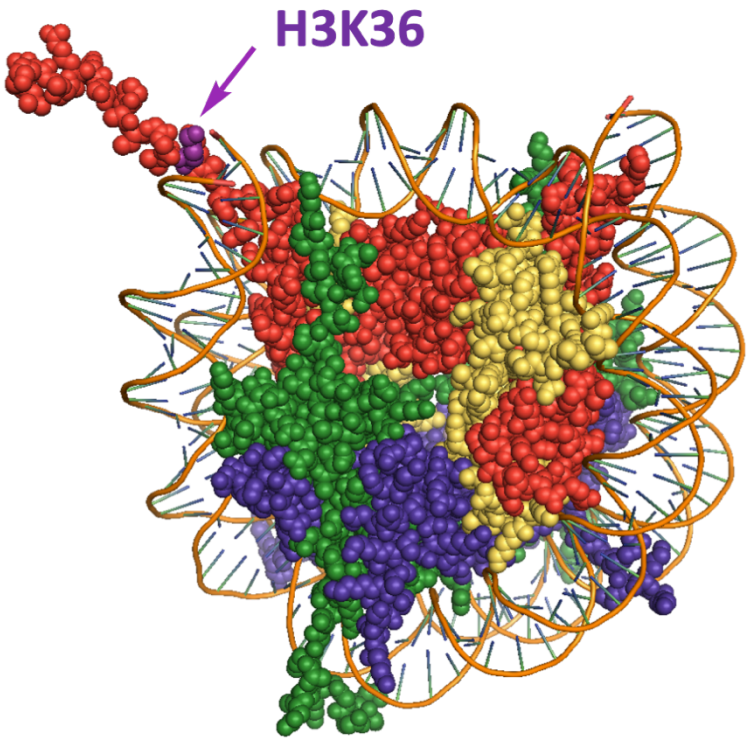


Fig. 3.1.4.1 *Crystal structure of nucleosome with position H3K36 highlighted. From PDB ID 1aoi.*

It is therefore hypothesised that discrimination between methylation states by MRG15 is due to a change in the overall nucleosome conformation, and through binding studies with intact H3K36me2 and -me3 bearing nucleosomes, it would be possible to reconcile the *in-vivo* and *in-vitro* observations.

3.2 Results and discussion

3.2.1 Pulldown studies with MRG15

In order to qualitatively compare the binding to MRG15 of the different nucleosome methylation states described in Chapter 3, a pulldown assay strategy was employed. In general, one component (MRG15 or nucleosome) was immobilised on a solid bead surface as the 'bait'.

The beads were then incubated with the second component for a number of hours so that binding equilibrium might be reached, before stringently washing away any unbound components. All proteins attached to the beads were denatured, run on SDS- PAGE for separation, and visualised by Western blot detection. Where there was a strong interaction between MRG15 and a methylation state, a correspondingly more intense band was expected for whichever component was pulled down from solution; conversely, a weak interaction should correspond to a fainter or absent band.

3.2.1.1 GST-MRG15 immobilised via antibody

As previous initial tests on free H3 were carried out using GST-MRG15 as the immobilised component (fig. 3.1.3.1), initial pulldown attempts replicated this system. Here, MRG15 was tethered to beads via a GST fusion tag, with GST-only beads used as an appropriate negative control to test whether any binding was to MRG15 itself or a non-specific interaction with the GST tag or beads themselves.

First attempted was an antibody capture method, whereby an anti-GST antibody was chemically crosslinked to Protein A beads, and GST-MRG15 (expressed and purified by A. Couturier) was subsequently captured onto the beads via the tag.

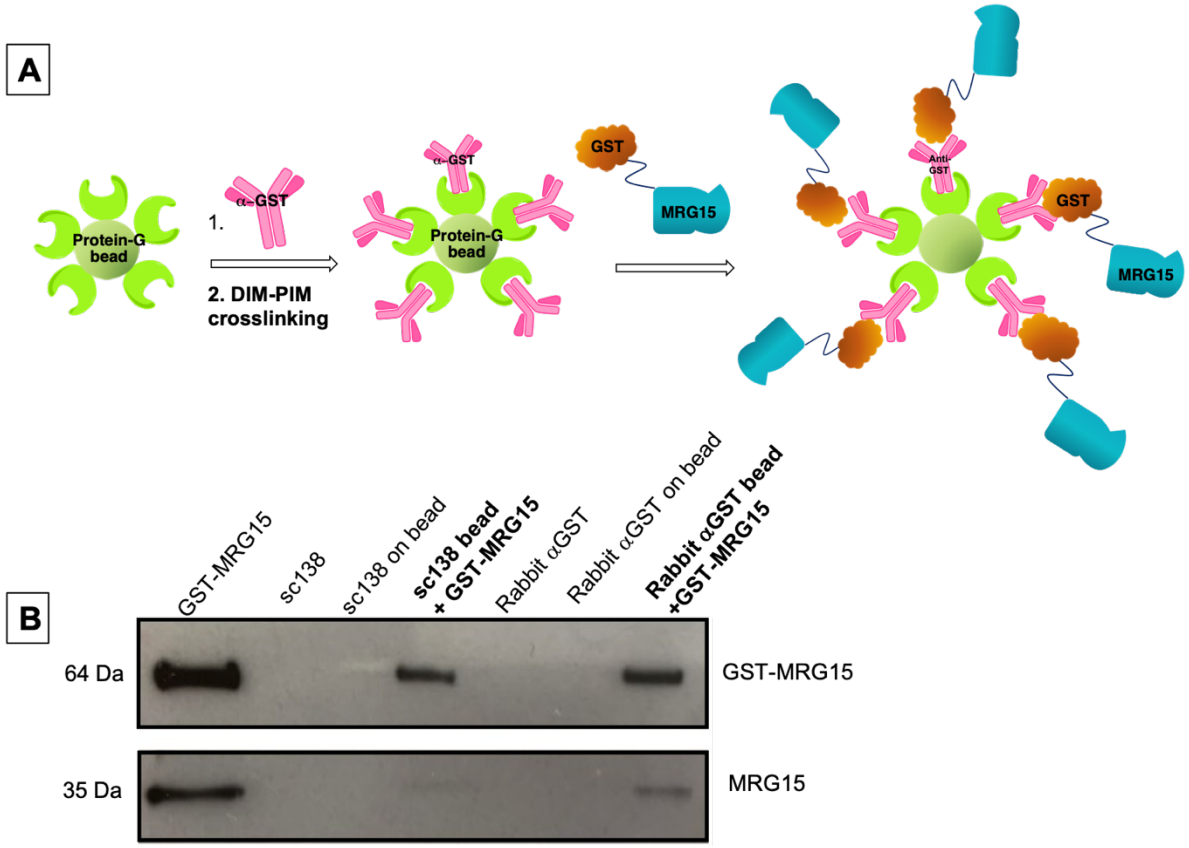


Fig. 3.2.1.1.1. (A) *Schematic for the generation of antibody beads; α-GST antibodies were first crosslinked to Protein G beads, which were then used to capture GST-MRG15.* ***(B)*** *Western blot analysis of α-GST antibodies (sc138 and rabbit) alone, antibodies on beads, and antibodies on beads after GST-MRG15 capture. GST-MRG15 was loaded as a positive control.*

This was tested using two anti-GST antibodies; sc138, a commercial mouse monoclonal IgG_1 α-GST HRP conjugate, and Rabbit α-GST, a polyclonal antibody. Western blot analysis gave bands only for the GST-MRG15 control and the fully assembled beads, indicating success. No cross-reactivity with either of the antibodies used for capture was observed; this was attributed to the resilience of the covalent amide crosslinking to SDS denaturation, preventing the antibodies from dissociating from the beads and subsequently entering the SDS-PAGE gel matrix. Unexpectedly, a band at 35 Da was observed, which was attributed to free MRG15 cross-reacting with the anti-GST primary antibody used in the Western blot. As GST runs at

28 kDa, and longer exposure revealed a third faint band at this molecular weight, it was concluded that the 35 kDa free MRG15 band must appear due to degradation of the GST tag.

While this initial test was successful, reproducibility proved an issue in subsequent attempts to make larger quantities of the antibody-linked beads. Additionally, it was nontrivial to accurately measure the concentration of GST-MRG15 captured on the beads, as they contained a mixture of MRG15 and antibody. Therefore, alternative strategies for pulldown experiments were pursued.

3.2.1.2 GST-MRG15 immobilised on GSH Sepharose beads

Next, GST-MRG15 was immobilised directly to glutathione (GSH) Sepharose beads using a strategy adopted from resin purification methods. The purified protein was incubated with the beads overnight, and unbound GST-MRG15 removed from the sample in a series of careful wash steps.

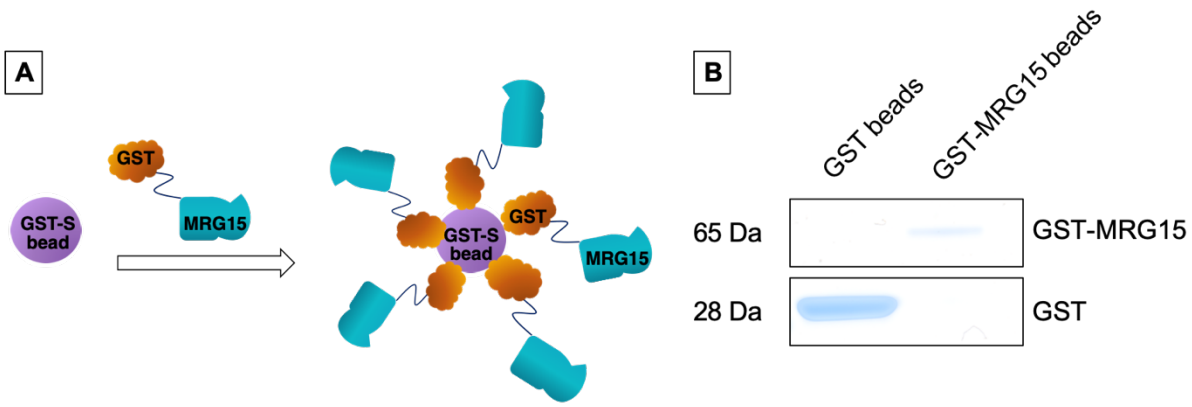


Fig. 3.2.1.2.1. ***(A)*** *Schematic of the immobilisation of GST-MRG15 onto GSH-sepharose beads.* ***(B)*** *SDS-PAGE analysis of the beads. GST beads generated previously by A. Couturier were also loaded for comparison.*

This method of GST-MRG15 capture on beads was reproducible, albeit with consistently low immobilised protein concentration which was not improved on increasing the amount of GST-MRG15 used in preparation relative to beads. This may be due to a low affinity of the GST tag for the beads, particularly when binding may be sterically hindered in the context of the fusion protein, or traces of remaining glutathione from the initial protein purification out-competing the interaction with the beads.

A pulldown assay was subsequently attempted with the sepharose beads and 145 bp WT, H3K36me2 and -me3 containing nucleosomes. The beads were blocked with BSA, and additional BSA and salmon sperm DNA were included in excess in the binding buffer used to out-compete any non-specific interactions. A salt concentration of 250 mM KCl was also chosen, as this was found in previous screens with free H3 to give the best discrimination between the WT and methylated variants.

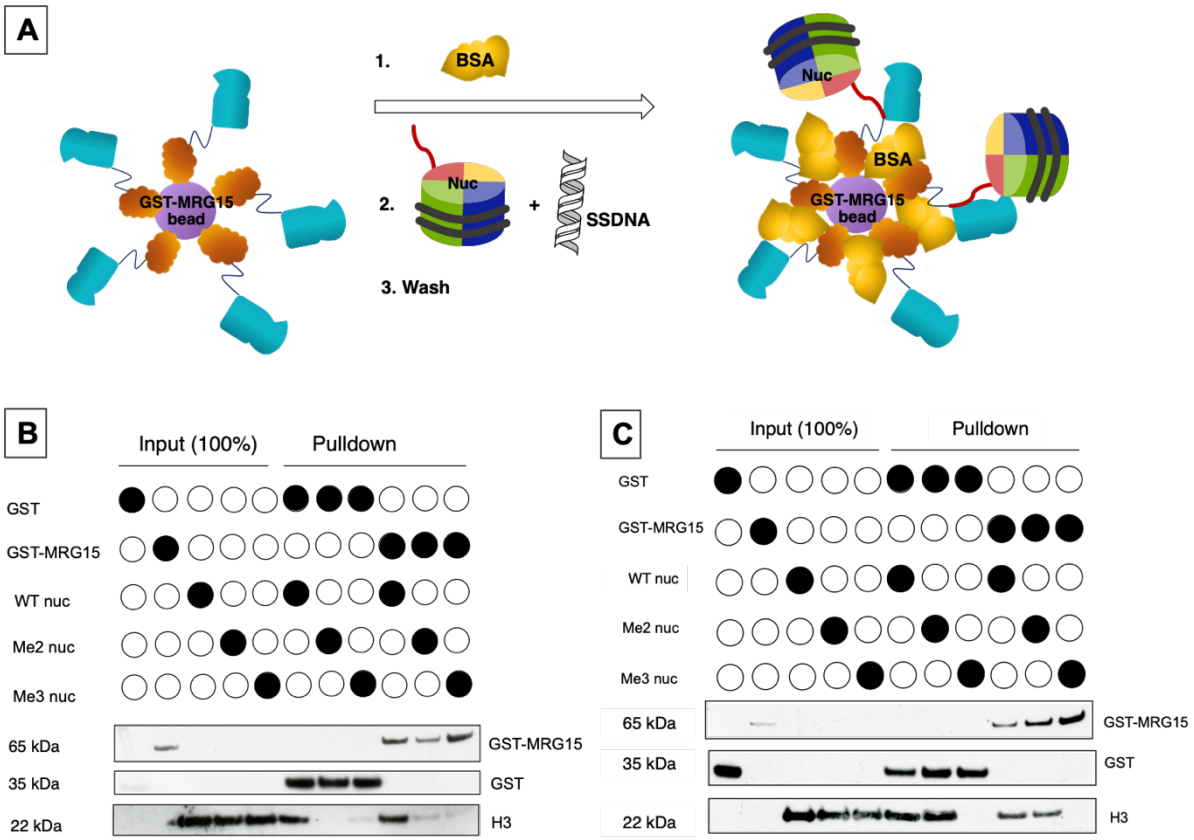


Fig. 3.2.1.2.2. (A) *Schematic for the pulldown experiments using GST-MRG15 immobilised on GSH- sepharose beads as bait. Beads were first blocked with Bovine Serum Albumin (BSA) before incubating with nucleosome and salmon-sperm DNA (ssDNA). Unbound nucleosomes were then removed with wash steps, before analysing the beads with SDS-PAGE and Western blot.*

(B). *Western blot analysis of pulldown with 145 bp nucleosomes.* ***(C).*** *Western blot analysis of pulldown with 186 bp nucleosomes.*

However, binding was observed only with the WT nucleosome, to both GST and GST-MRG15 beads. Re-analysis of the 145 bp nucleosomes by 6% TBE gel revealed that the WT and H3K36me3 samples may have degraded; the pulldown was thus repeated with the intact 186 bp variants. Again, similar results were observed with binding to the negative GST controls.

This suggested that the nucleosomes might be sticking to the beads themselves, or to the GST tag.

More generally, a potential for error in any strategy where MRG15 is immobilised and the amount of nucleosome bound is variable was also recognised. Here, binding is quantified by the intensity of the H3 band in Western blot, but the H3 substrate in each sample is subtly different due to the difference in K36 methylation state. Moreover, as mentioned in earlier sections, the chemically produced methyl H3 variants are unavoidably and inseparably present in two epimers; the natural L and unnatural D. Previous studies using an anti-H3K9me3 antibody to detect a methyl histone variant only recognised 50% of the total protein present, suggesting that the antibody was only able to recognise the natural epimer.[52] If this was occurring in these studies, the methyl H3 bands would be expected to be approximately half the intensity of the recombinantly expressed WT band; this was however not the case, suggesting that the anti-H3 antibody recognises an epitope distal to the tail.
Nonetheless, it was reasoned that employing a strategy where the nucleosomes were captured on beads and bound MRG15 was the variable to be measured would yield more robust results, as the antibody would be recognising identical MRG15 structures in each sample.

3.2.1.3 186 bp nucleosomes immobilised on streptavidin Dynabeads

Given the challenges with using immobilised GST-MRG15, a different strategy was pursued whereby the nucleosomes were used as the immobilised 'bait' component instead. Here, the biotinylated 186 bp nucleosomes were captured on Dynabeads functionalised with streptavidin, which were then blocked with BSA as per the manufacturer's recommendation. The beads were then incubated with MRG15 for several hours to allow binding to equilibrate as before,

washed stringently to remove any unbound MRG15, and analysed by SDS-PAGE and Western blot.

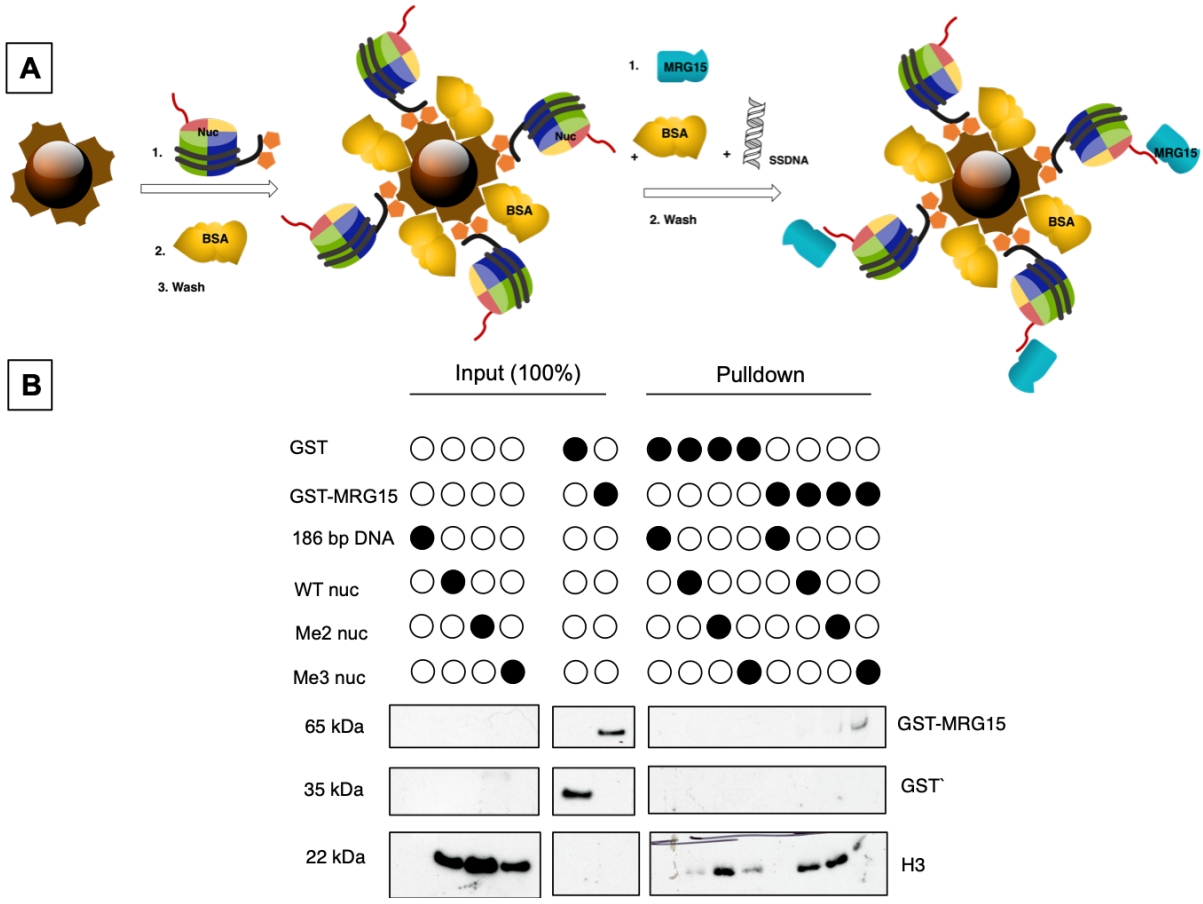


Fig. 3.2.1.3.1. (A). *Schematic of the pulldown experiment with 186 bp nucleosomes immobilised on streptavidin dynabeads. Nucleosomes were immobilised, and beads blocked with BSA in a blocking step. Unbound nucleosomes and BSA were then washed from the beads. Next, the beads were incubated with GST-MRG15 (MRG15 shown- see experiments 4.2.1.4) or GST alone (negative control, not shown) with BSA and ssDNA to out-compete non-specific interactions. Unbound species were then removed in wash steps, and the beads analysed by SDS-PAGE and Western blot.* ***(B).*** *Western blot analysis of the pulldown.*

Initially, conditions were kept as similar to those previously used as possible, with the same buffer composition and GST-MRG15 paired with GST only negative controls. Free 186 bp

DNA was also on streptavidin beads as a control to ensure that MRG15 was binding selectively to the H3 PTM and not simply the 601 DNA sequence.

Now, selective binding could be observed. The nucleosomes did not bind GST but only GST-MRG15, indicating true binding to the reader protein instead of the fusion tag. In all pulldown lanes, the intensities of H3 bands were variable between samples and noticeably less than in the input control. This was attributed to human error in handling, as the quantities of beads used in each case were very small; 5 µL in a 1 mL pulldown. However, in the GST negative controls no bands were observed, even in the H3K36me2-nucleosome lane where the H3 band was most intense. Moreover, although the H3K36me3 band was the faintest in the GST-MRG15 pulldowns, it was the only lane where GST-MRG15 was present. The intensity of the H3 band here was also comparable to the intensity of the H3 band in the GST negative control, but no GST band was observed in the H3K36me3 nucleosome lane. Finally, the selectivity between nucleosome methylation states was upheld in later experiments.

Taken together, these observations suggested that the selective binding of H3K36me3-containing nucleosomes to GST-MRG15 was indeed valid, and via the MRG domain rather than the GST domain. Now, GST-MRG15 bound H3K36me3-containing nucleosomes, but not -me2, WT, or free 186 bp DNA.

3.2.1.4 Screening effect of salt identity and concentration on MRG15-nucleosome interaction

Previous results suggested that the selectivity of MRG15 binding to free H3 was highly dependent on the salt concentration in the buffer. Therefore, alternative salt types (NaCl) and concentrations (150 mM KCl and NaCl) were screened to investigate their influence on selectivity in the context of the whole nucleosome. MRG15 cleaved from the GST tag was also used to eliminate the need for a GST-only negative control.

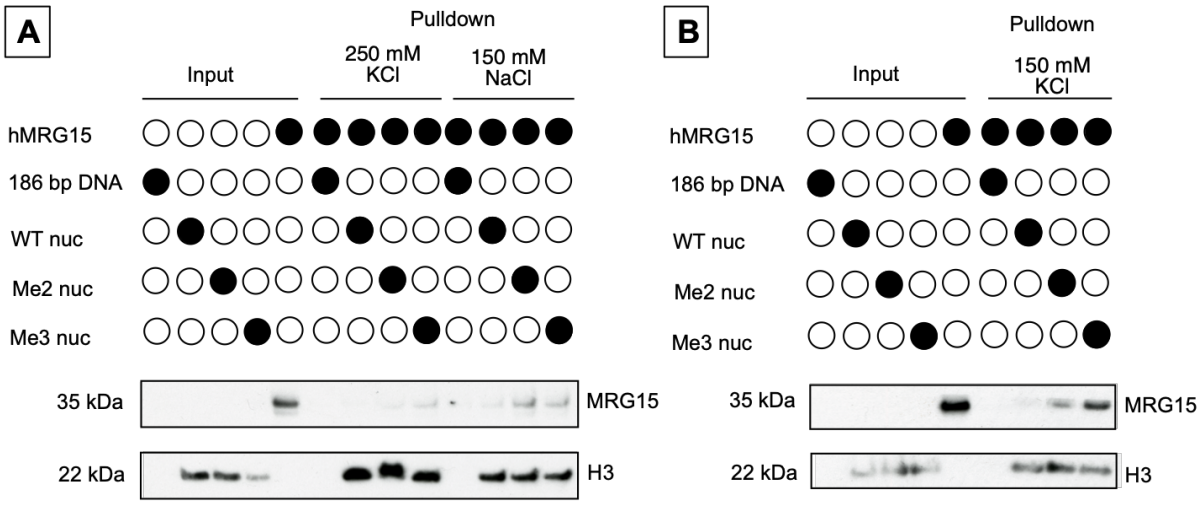


Fig. 3.2.1.3.2. *Results of pulldowns of MRG15 using 186 bp nucleosomes immobilized on streptavidin dynabeads as bait.* ***(A)*** *In 250 mM KCl and 150 mM NaCl buffers* ***(B)*** *in 150 mM KCl buffer.*

In general, it was found that the strength of binding improved with lower salt concentration. Interestingly, while selectivity was preserved with 150 mM KCl, when 150 mM NaCl was used equal binding to both H3K36me2 and -me3 nucleosomes was seen. As a caveat to this observation, it should also be noted that the H3 bands for all nucleosomes were fainter in NaCl compared to KCl, suggesting some nucleosome degradation may have taken place; indeed, nucleosome core particles are known to degrade at 150 mM NaCl.[53]

A salt concentration of 150 mM NaCl was broadly used in similar pulldown assays to represent physiological conditions, as it is the concentration of a saline solution isotonic to blood. However, measurements of ionic concentrations in the nucleus have found the concentration of potassium ions is several fold higher than that of sodium ions, with values of approximately 7.5 to 20 mM [Na+] and 120 to 270 mM [K+]. [54,55]

It is therefore argued that the measurements in KCl are the more physiologically relevant, suggesting that MRG15 does discriminate in the nucleus.

The responsiveness of MRG15 binding to the salt concentration of the buffer suggested the mechanism of selectivity is dependent on a charge-charge effect in some way. In general, cations are known to abrogate the interactions between DNA and octamer as they bind to the negative DNA backbone, and the smaller Na^+ is known to do this more effectively than K^+.[56] The dependence of the discrimination effect on cation concentration and identity suggests it might arise from different modes of DNA-octamer interactions between methylation states.

3.2.2 Surface Plasmon Resonance (SPR) studies on MRG15 binding to nucleosomes

From the Western blot based studies, a qualitative picture of the binding interaction with MRG15 has emerged. The next goal was to quantify equilibrium dissociation constants (K_Ds) for binding with the different variants.

For this, Surface Plasmon Resonance (SPR) was employed. This technique involved immobilising a substrate onto a chip, over which an analyte was then flowed in increasing concentrations. The mass of the chip was constantly measured (in the arbitrary Resonance Units) and increases in mass detected corresponded to the binding of the analyte on the surface.

Here, a chip was functionalised with streptavidin, and 186 bp biotinylated nucleosomes immobilised onto its surface to form the substrate. MRG15 with the GST fusion tag removed (previously purified by A. Couturier) was then flowed over the chip as the analyte. Initially, this was attempted in a buffer containing 250 mM KCl, as this had given the greatest discrimination between methylation states in the pulldown experiments and literature suggested this was an appropriate representation of *in-nucleo* salt composition. However, no

binding was observed even at the highest MRG15 concentrations used. As previous observations suggest there is an ionic component to the interaction, it was hypothesised this could be disrupted by the high salt, and so measurements were subsequently attempted with the lower salt 150 mM KCl buffer. Here, binding to MRG15 was observed, allowing further SPR optimization.

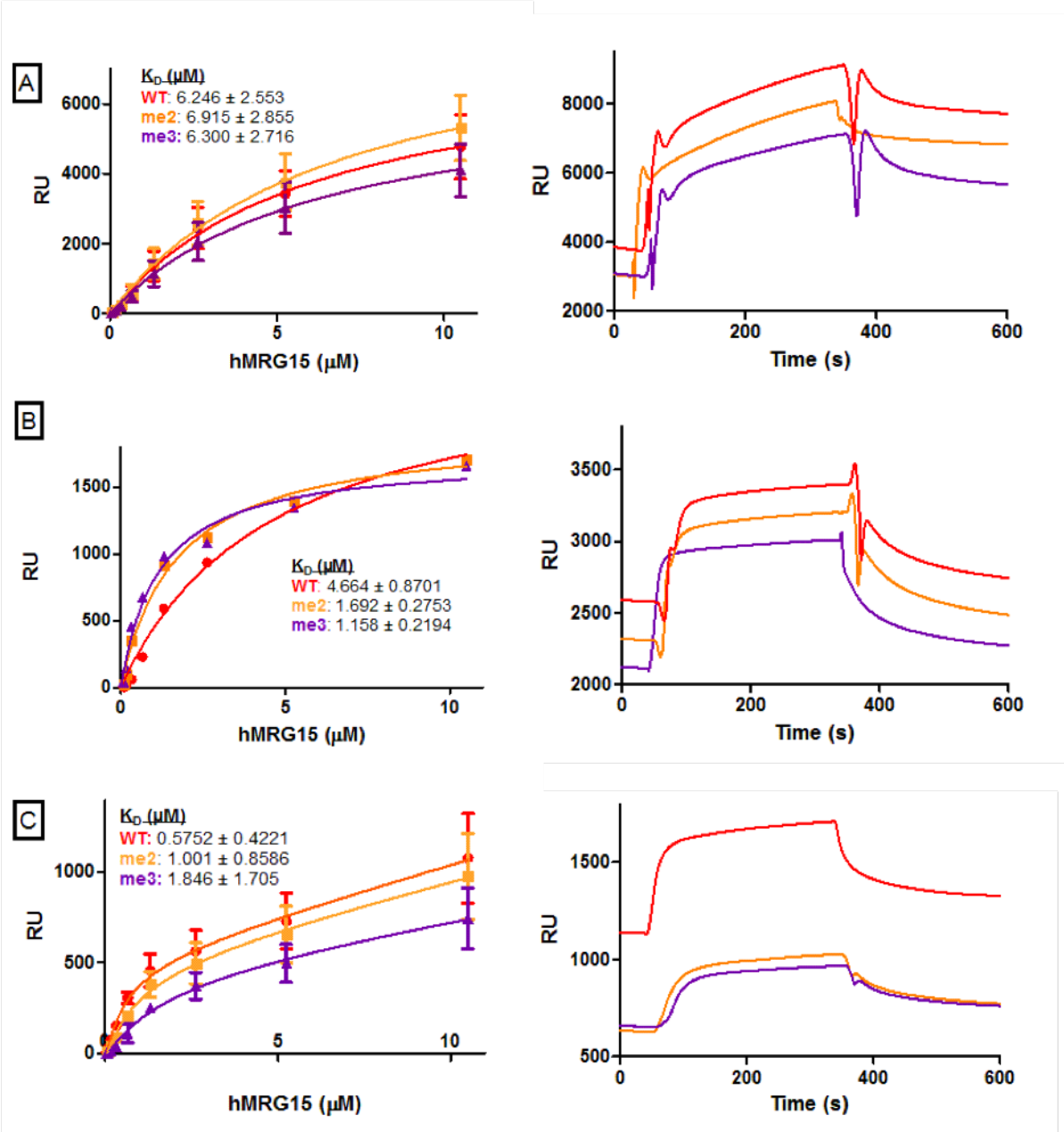


Fig. 3.2.2.1. *Curves used to calculate Kd values (left) and examples of SPR profiles acquired at 10.5 μM MRG15 (right) for* ***(A)*** *initial attempt in 150 mM NaCl buffer* ***(B)*** *second attempt*

with purified MRG15 in 150 mM KCl buffer, and ***(C)*** *third attempt with purified MRG15 (3 repeats, but issues with nucleosome immobilization caused variability between samples. 150 mM KCl buffer).*

A broad screen of MRG15 concentrations was then performed to determine the optimum range for subsequent experiments, i.e. beginning at the lowest concentration needed for a response. Initial experiments were then performed in order to gain an estimate of K_D. However, the response curves in the raw measurement data did not reach a plateau; instead, an initial rapid binding phase followed by a slower phase was observed (fig. 4.2.2.1 A).

This two-phase binding is characteristic of multiple analyte molecules binding to the same substrate molecule in succession. This may occur if the analyte sample is not homogeneous, and aggregate species are present. The thawing of previously flash-frozen MRG15 samples was identified as a potential cause for aggregate formation, and indeed native-PAGE analysis of the samples confirmed that such higher-order species were present.

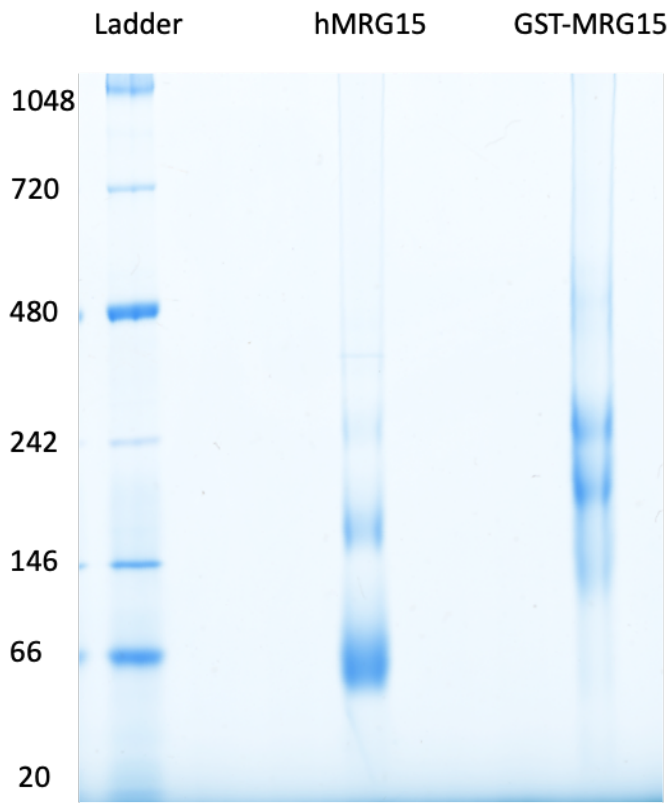


Fig. 3.2.2.2. *Native-PAGE analysis of MRG15 cleaved from GST tag (hMRG15) and GST-MRG15 thawed after flash-freezing and storage at -80 ℃.*

To overcome this issue, a strategy of freshly purifying MRG15 via size-exclusion chromatography immediately before SPR analysis was pursued in order to ensure only pure, monomeric species was present during measurement. Briefly, a construct containing a GST tag, PreScission protease cleavage site, and a His_6 tag N-terminal to MRG15 was expressed from glycerol stock in Arctic (details) *E. coli*. The fusion protein was then isolated by binding to GST resin, cleaved from the resin using PreScission protease, isolated again by Ni-NTA resin binding to the His_6 tag, eluted dialysed, and purified on an S75 gel filtration column.

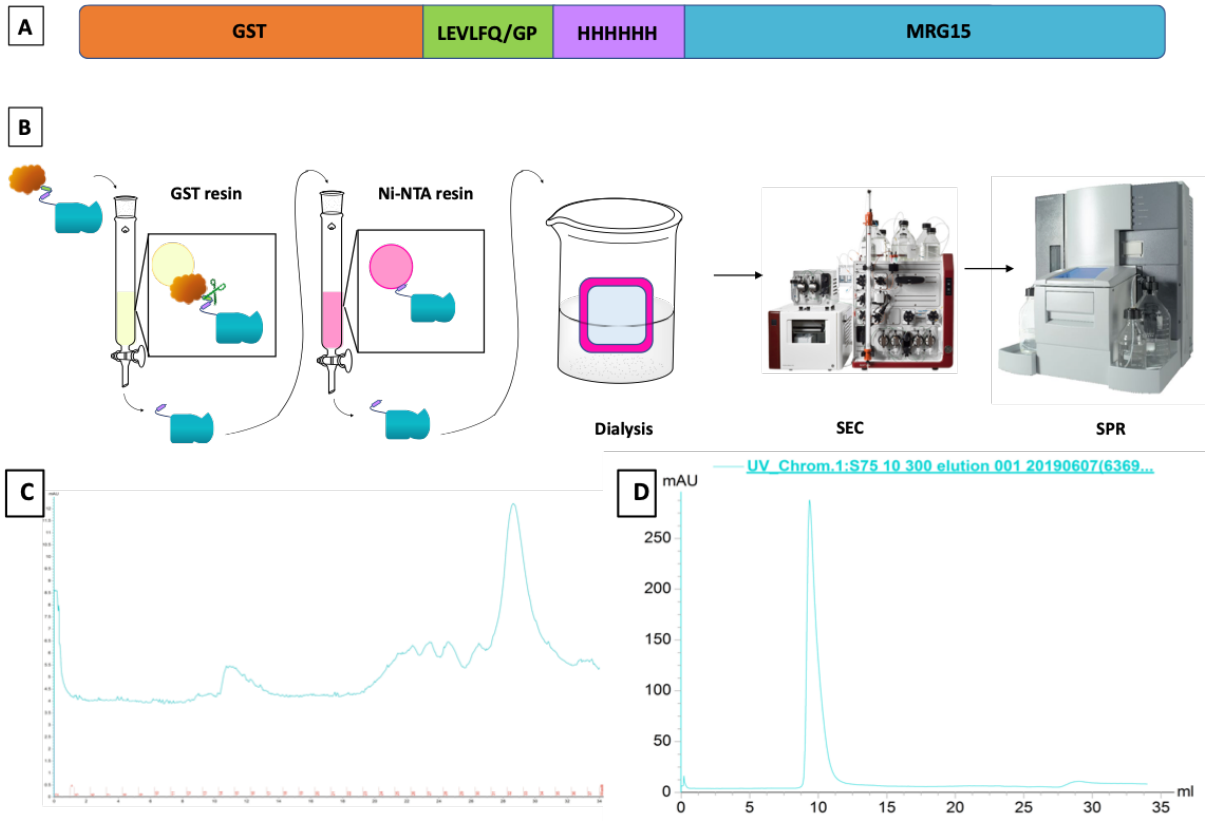


Fig. 3.2.2.3. (A) *Schematic representation of the GST-MRG15 construct expressed, containing a N-terminal GST tag (orange), a Prescission Protease cleavage site (green), a* His_6 *tag (purple) and MRG15 (blue).* ***(B)****. Schematic of the MRG15 purification protocol.* ***(C)*** *SEC trace for the first round of MRG15 purification* ***(D)*** *SEC trace for the second round of MRG15 purification*

Initially, excess imidazole present after the Ni-NTA step was removed via dialysis only. However, the subsequent SEC trace showed a large amount of degradation and imidazole still present (fig. 4.2.2.3 C), suggesting that dialysis was not rapid enough to remove imidazole before it began to damage the protein structure. This was overcome by adding a desalting step using PD-10 columns immediately after elution to rapidly remove most imidazole, followed by overnight dialysis to remove what remained. This altered protocol resulted in a single, clean peak for monomeric MRG15 in the size exclusion trace (fig. 4.2.2.3 D)

In addition, BSA and Salmon Sperm DNA were also added to the SPR buffer in order to out-compete any non-specific interactions. However, while lower apparent K_D values were obtained with the purified MRG15, a sloped plateau characteristic of multi-phasic binding was still observed (fig. 4.2.2.1 C), and so true K_D values could not be extracted.

The MRG domain is known to facilitate dimerization in plant MRG15 homologs, and is thought to do the same in MRG15.[57] It is likely that the initial rapid binding phase represents a single monomer of MRG15 binding to a single nucleosome, while the second slow phase represents recruitment of a second MRG15 via the MRG domain. In order to overcome this in future experiments, SPR studies using only the chromodomain to eliminate MRG15's capacity for dimerization are suggested.

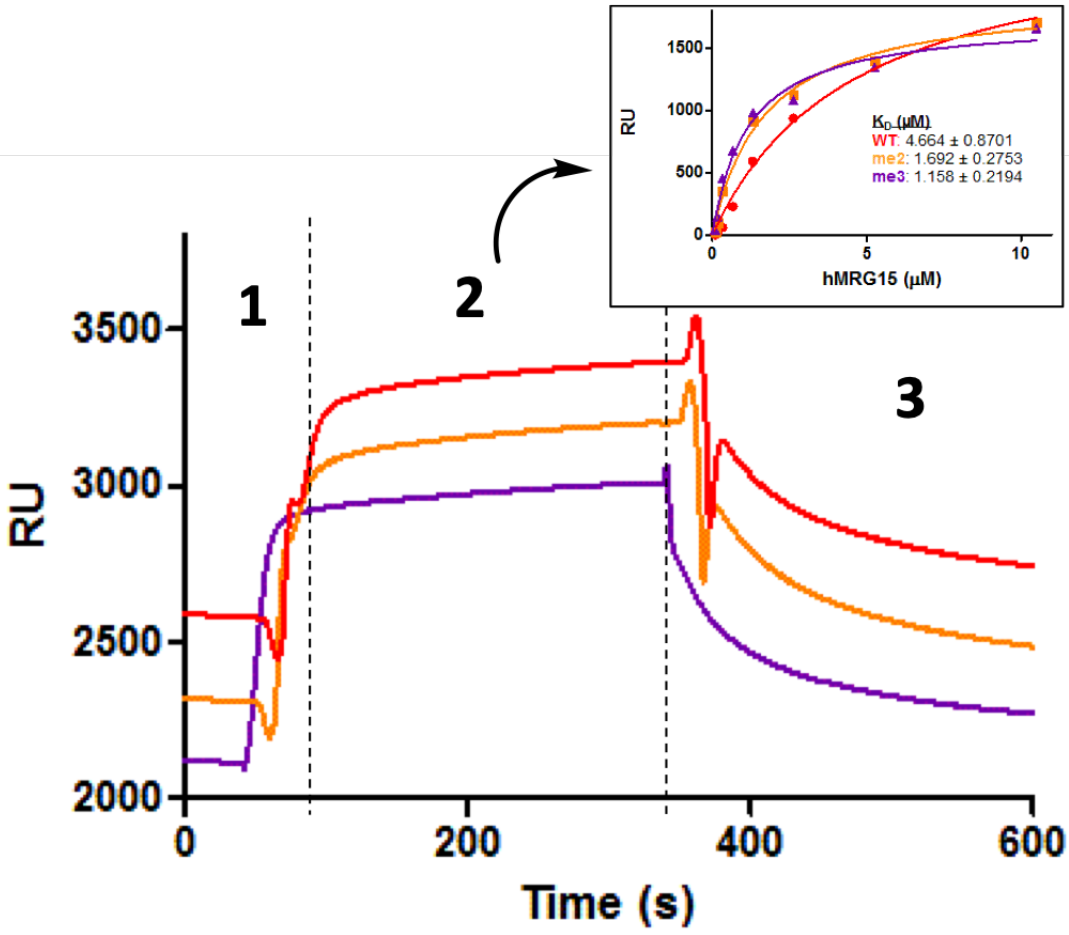


Fig. 3.2.2.4. *Phases in analyte binding in SPR. Phase 1: $k_{on} > k_{off}$. Phase 2: In monophasic binding, we expect to see equilibrium here (plateau) as $k_{on} = k_{off}$. The curve used to calculate K_D is extracted from this phase. Phase 3: Analyte is no longer injected, $k_{on} < k_{off}$ as analyte dissociates.*

Importantly, data was obtained in the low sample consumption suitable for determining thermodynamic parameters at equilibrium but not kinetic parameters, where higher accuracy is required. This means accurate k_{on} or k_{off} parameters for MRG15 binding to the nucleosomes cannot be determined using this data. However, on qualitative examination of the raw data obtained at 10.5 μM (fig. 4.2.2.1; left hand column) it is observed that the slope for dissociation appears to be less steep for the methylated variants compared to wild type. As mentioned previously, it is known that there are two components affecting the observed binding in the pulldown assays: the equilibrium dissociation constant for MRG15 binding to nucleosomes which impacts the incubation step, and the k_{off} which impacts the wash step, analogous to what is seen in the SPR trace. It is possible that MRG15 has similar equilibrium K_D values for the different nucleosomes but different k_{off} rates, and it is the latter which gives rise to the different binding we see. To investigate whether this is the case, full kinetic SPR experiments are suggested in future.

3.2.3 Pulldown studies with W53A negative control

Our structural alignment and docking studies suggest that MRG15 binds to H3K36me3 in nucleosomes via its chromodomain, while in plant homologs the MRG domain is known to facilitate dimerization; it is thought to do the same in MRG15. However, this had not yet been experimentally validated. In light of the potential co-operative binding observed in SPR, we

performed further pulldown studies using the MRG15 W53A mutant as a negative control. Here, a key aromatic residue in the binding pocket was mutated to alanine, which has previously been shown to abrogate its binding ability.

MRG15 W53A was generated using standard site-directed mutagenesis techniques, and expressed using the same protocol as for MRG15 WT. As expected, binding to H3K36 across all methylation states was noticeably impaired in the mutant.

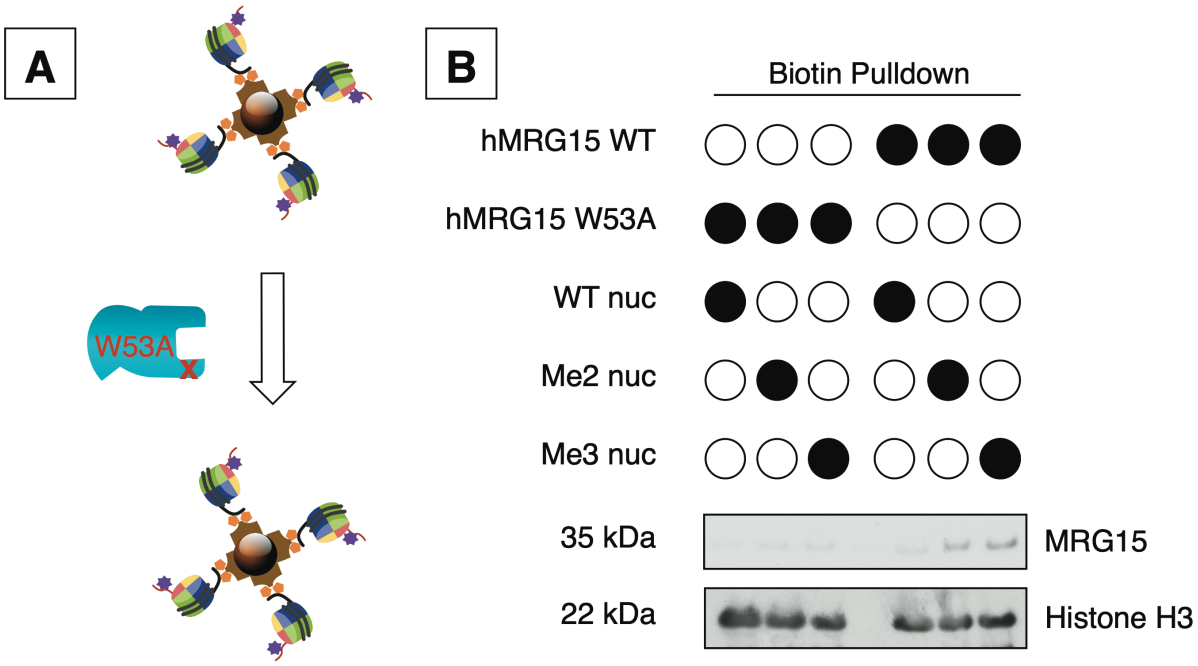


Fig. 3.2.3.1 *(A) Schematic of the pulldown with MRG15 W53A. (B) Western blot results of the pulldown. Studies were performed in 250 mM KCl buffer.*

These results suggest that, as predicted, the chromodomain alone facilitates binding to H3K36 in nucleosomes, whilst the MRG domain likely facilitates a co-operative interaction.

3.2.4 DNAse I assays

As noted previously, when K36 bears at least one hydrogen on the side chain nitrogen (as in all methylation states bar -me3), it has the capacity for hydrogen bonding to the DNA phosphate backbone. One potential cause for MRG15 binding more readily to the -me3 state,

therefore, could be a loosened contact between H3 and DNA; if the DNA was further from the H3 tail, or was able to move away more readily, it might provide a better opportunity for MRG15 to envelop K36.

In this model, an increase in the flexibility of the DNA ends might be expected as methylation state increases. As these flexible ends would face greater exposure to solvent, they would also be more susceptible to endonuclease digest; therefore, studying the digestion of the nucleosomes in different methylation states was proposes as a method to test this hypothesis. As a preliminary experiment, DNAse I was used to perform digests, as it was readily available on hand. Initial screens comparing 186 bp nucleosomes to 145 bp nucleosomes showed the latter gave clearer digestion product bands, so the 145 bp variants were used in subsequent experiments.

The digest was first performed across all nucleosome, and lengths of DNA fragments present at one timepoint were analysed by denaturing nucleosomes and running on a TBU gel. The analysis was subsequently expanded to a timecourse in an attempt to capture differences in the rate of DNA digestion which might occur.

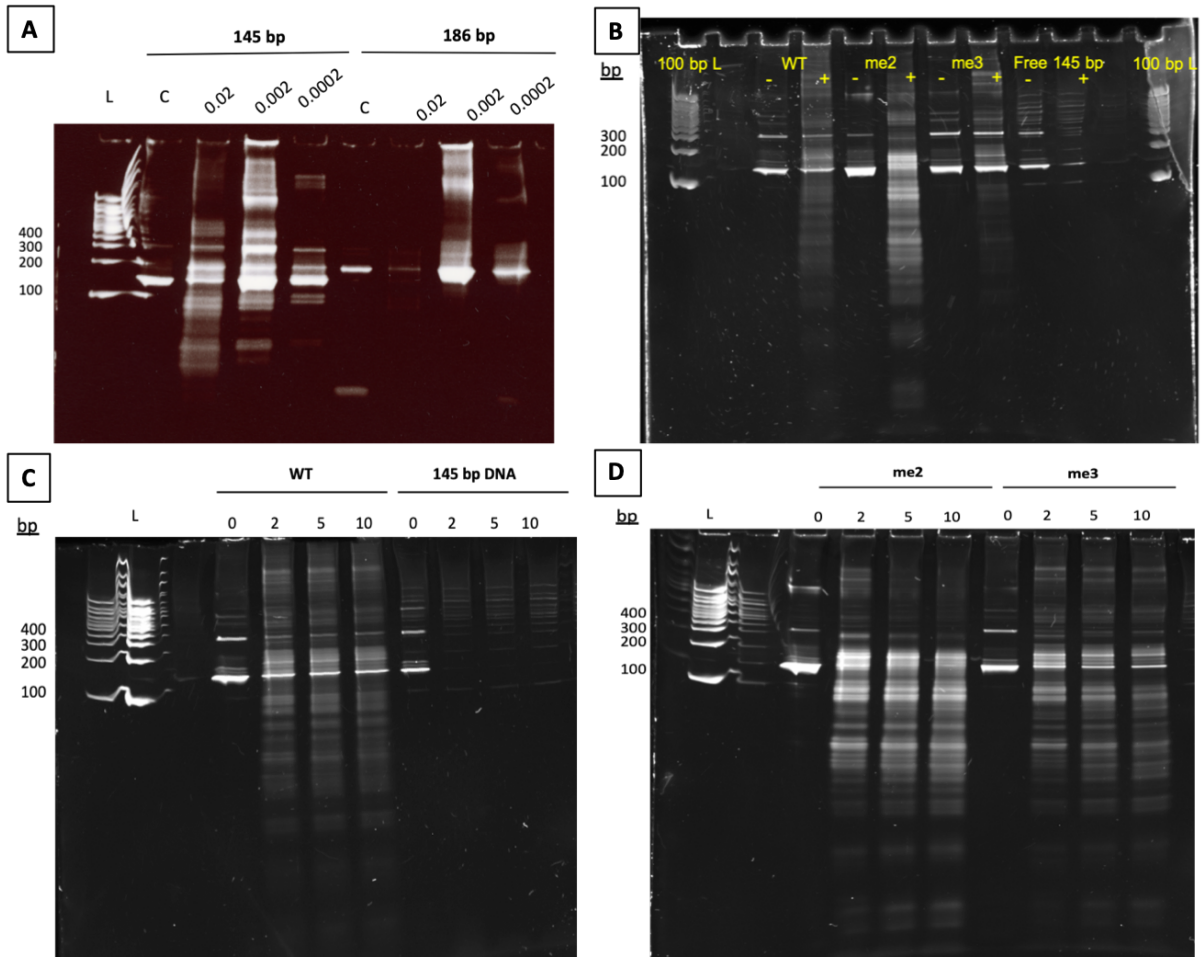


Fig. 3.2.4.1. ***(A)*** *Preliminary assay of 186 bp vs 145 bp nucleosomes. Numbers refer to DNAse 1 quantity in KU.* ***(B)*** *Initial DNase I assay with 145 bp DNA, all samples quenched after 30 min. += DNAse I added, -= no DNAse I added (control).* ***(C)*** *and* ***(D)*** *Timecourse assay; numbers refer to the number of minutes incubated with DNAse I before quenching (0= control).*

However, several fundamental errors with these experiments were encountered, rendering it impossible to draw reliable mechanistic conclusions from the data. It was hoped that bands of characteristic length would be identified, which would correspond to a specific, looser end of the DNA being cleaved first. Such a band could not be distinguished consistently across the samples, likely due to differences in nucleosome conformations between methylation states. Moreover, on monitoring the decrease in intensity of our un-digested full length DNA bands, great inconsistencies between repeats were also observed.

One fundamental issue was that, as an unavoidable consequence of nucleosome assembly, some free DNA remained in each of the samples, which would obscure our results from the digestion of on-nucleosome DNA. It was also noted that the older nucleosome stock (me2, synthesised several months before WT and me3) was more prone to degradation.

In order to get more reliable data, several alterations to the experiment are proposed. First, micrococcal nuclease (MNase) should be used instead of DNAse I, as the former is more traditionally used for nucleosome digest experiments. Second, nucleosomes should be purified by SEC before the experiment to remove any contaminating free DNA. However, it is noted that the absence of characteristic bands suggests a mechanism beyond the simplistic model of greater freedom in the DNA ends to account for the greater affinity of MRG15 for the me3 nucleosomes.

3.3 Conclusions

In vivo evidence has suggested a biological need for trimethyl H3K36me3 in particular over the dimethyl state. However, this has previously not been supported by in-vitro binding studies of MRG15 on peptides or free H3. It was therefore hypothesised that discrimination between methylation states by MRG15 is due to a change in the overall nucleosome conformation, and through binding studies with intact H3K36me2 and -me3 bearing nucleosomes, it would be possible to reconcile the *in-vivo* and *in-vitro* observations.

On assembling the nucleosomes, evidence for conformational differences between methylation states was observed (Chapter 2). Previous pulldown-based binding assays were revisited with these synthetic nucleosomes, and it was now possible to observe the clear preference for -me3 nucleosomes over -me2 that one would expect from *in-vivo* data.

Moreover, the strength of binding and magnitude of discrimination appears dependent on the salt type and concentration of the buffer, with weaker binding observed on increasing salt concentration from 150 mM KCl to 250 mM, and weaker selectivity observed on changing the salt to NaCl. This suggests the mechanism of selectivity might arise from different modes of DNA-octamer interactions between methylation states.

Attempts at gaining insight into this by DNAse I digest of the nucleosomes did not give any clear indication of a weakening DNA-octamer interaction. Therefore, the following model is proposed to explain MRG15 discrimination between di- and trimethylated nucleosomes. From the crystal structure, H3K36 is ideally positioned to form a hydrogen bond between a side chain -NH and the phosphate of the DNA backbone. In this state, H3K36 is shielded and inaccessible to MRG15. With a predicted pKa of 9.40 (ProPKA) this residue is expected to exist primarily in the +1 state as $-NH_3^+$; up to two of these hydrogens can be replaced by methyl groups, as in H3K36me2, without disrupting the bond to the phosphate backbone. However, when the third is added the residue no longer has the capacity for this hydrogen bond with the DNA backbone; it is hypothesised that this broken interaction, along with the additional bulk and hydrophobicity of the trimethyl state, causes a conformational shift in the H3 tail where K36me3 becomes exposed, and therefore accessible to MRG15.

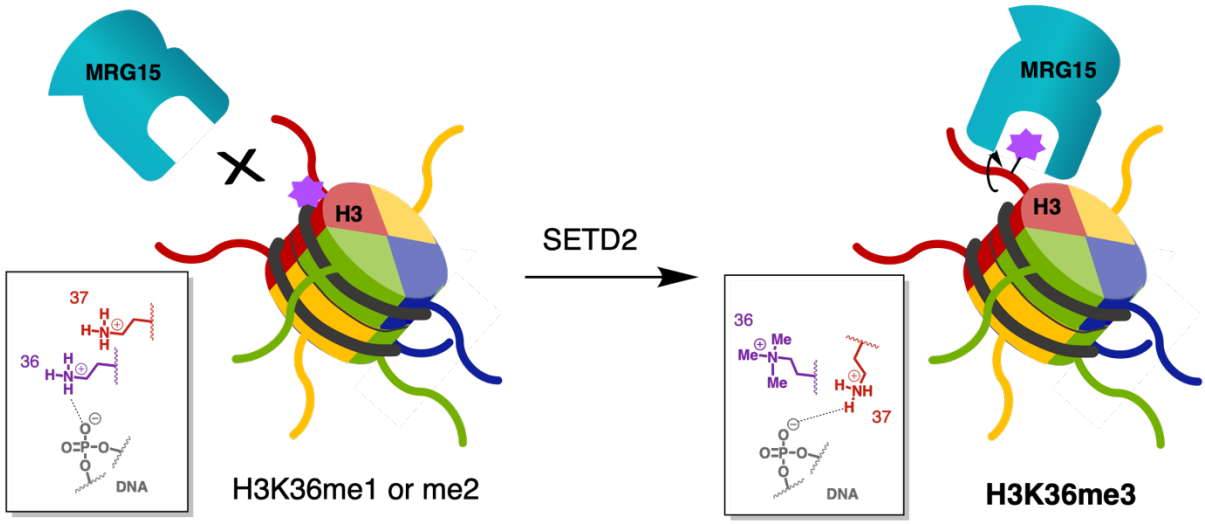


Fig. 3.3.1 Model for MRG15 selectivity for H3K36me3 over H3K36me2.

The binary nature of the hydrogen bonding switch on going from H3K36me2- to -me3 provides an attractive explanation for the necessity of the latter *in vivo*. The buried nature of H3K36me2 also has implications for the mode of SETD2 methylation, as SETD2 may not be able to access its substrate when it is part of a whole nucleosome.

SETD2 binds to the C-terminal domain tail of RNAPII, and has previously been assumed to methylate H3K36me2 in nucleosomes which have re-assembled after the polymerase has passed through them.[58,59]

However, previous studies have shown that under conditions of moderate transcription, RNAPII is able to transcribe through nucleosomes by displacing only H2A-H2B dimers, leaving the $H3_2$-$H4_2$ tetramer intact.[60] This $H3_2$-$H4_2$ tetramer-only state is expected to leave H3K36 more exposed; it is suggested, then, that SETD2 may take advantage of this and methylate H3K36me2 concurrently with RNAPII transcription.

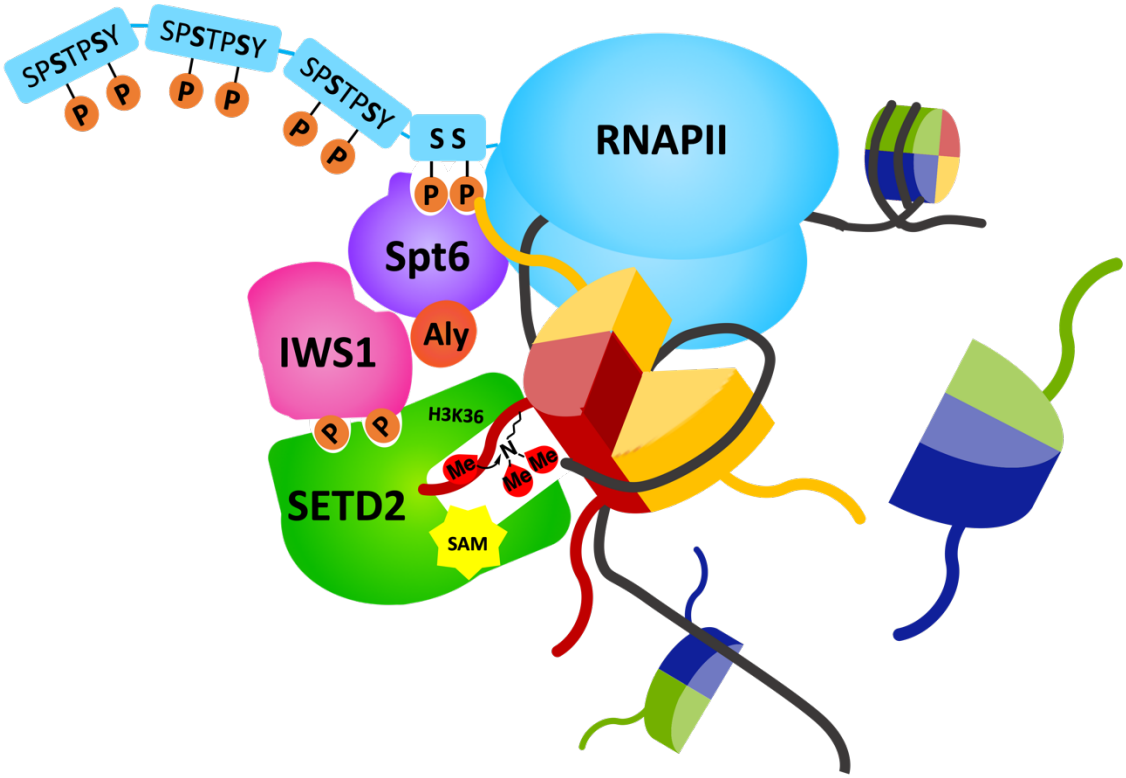


Fig. 3.3.2 Proposed alternative model for SETD2 mechanism.

3.4 Experimental

3.4.1 General Western blot protocol

Proteins were transferred from the SDS-PAGE gel to a 0.2 μm Amersham Protran Nitrocellulose blotting membrane (GE life sciences) using a BioRad Trans-blot cell gel tank filled with transfer buffer (192 mM glycine, 25 mM Tris base pH 8.3, 20% methanol). Transfer was performed for 1 h at 90 V, 4°C. Bands on the membrane were then visualised using Ponceau stain (0.1% Ponceau S, 1% acetic acid) and membrane cut into sections as appropriate for primary antibody application. The membrane was blocked with 30 mL 5% milk powder in PBST (PBS with 0.05% Tween20) for 30 minutes at room temperature with orbital rotation, then incubated with primary antibody in 5% milk/ PBST at 4°C with orbital rotation for 16 h.

The membrane was then washed three times with PBST (30 mL, 5 min, room temp) before incubation with secondary antibody in 5% milk/ PBST (polyclonal goat α rabbit IgG HRP conjugate, Dako, lot 20066477, 1:1000 dilution, 1 h, room temp), after which the membrane was washed three times with PBST.

ECL (Amersham Biosciences) was used to visualise the bound secondary antibody. The membrane was covered in ECL and incubated in the dark (2 min room, room temp) before exposing in a cassette (Hypercasette, Amersham Biosciences) and developing onto chemiluminescence film (Amersham Hyperfilm ECL) using a Xenograph SRX-101A developer.

3.4.1 GST-MRG15 immobilised via antibody

Two anti-GST antibodies were screened; sc138 mouse monoclonal IgG_1 α-GST HRP conjugate (Lot D2208) and rabbit polyclonal α-GST (prepared in the lab of Tim Hunt). 30 µL of Protein A bead slurry (Affiprep Protein A bead support, BioRad) was prepared by washing with 100 µL PBS via resuspension, centrifugation to pellet the beads (3000 rpm, 3 min, room temp) and removing the supernatant. For the antibody crosslinking step, the beads were then resuspended in 400 µL PBS and split into two 200 µL aliquots. To each aliquot was added one of the anti-GST antibodies, and the beads were incubated for 3 h with gentle rotation (20 rpm, room temp). The beads were then each washed three times with 200 µL PBS, twice with 200 µL coupling buffer (20 mM Boric acid, 20 mM sodium tetraborate pH 9.0) and once with 200 µL DIM-PIM (56 mM dimethyl pimelimidate dihydrochloride, 100 mM sodium tetraborate pH 8.5), all at room temperature. The beads were then resuspended in 400 µL DIM-PIM and incubated at 4° C 16 h with gentle rotation (20 rpm) to allow the coupling reaction to proceed overnight.

The antibody coupling reaction was then quenched by washing beads with 200 µL 1M Tris pH 9.0 at room temperature. A further 400 µL 1M Tris pH 9.0 was then added, and the beads incubated with gentle rotation at room temperature for 10 minutes. The beads were then washed with 400 µL PBS.

For antibody binding to GST-MRG15, the beads were resuspended in 400 µL wash buffer (200 mM NaCl, PBS pH 7.5, 0.05% Triton X-100, 2 mM DTT, 1 mM EDTA) and GST-MRG15 was added (5 µL of 0.8 mg/ mL). Samples were incubated for 2h at 4° C with gentle rotation (20 rpm). Unbound GST-MRG15 was then rinsed away by washing once with 200 µL wash buffer, twice with 200 µL dialysis buffer (50 mM Tris. HCl pH 7.8, 150 mM NaCl, 1 mM EDTA, 2 mM DTT).

Samples were made up in SDS sample buffer (NuPAGE LDS sample buffer + 10 mM DTT), run on SDS-PAGE in MOPS-SDS buffer (130 V, 1 h 15 min, room temp) and imaged via Western blot.

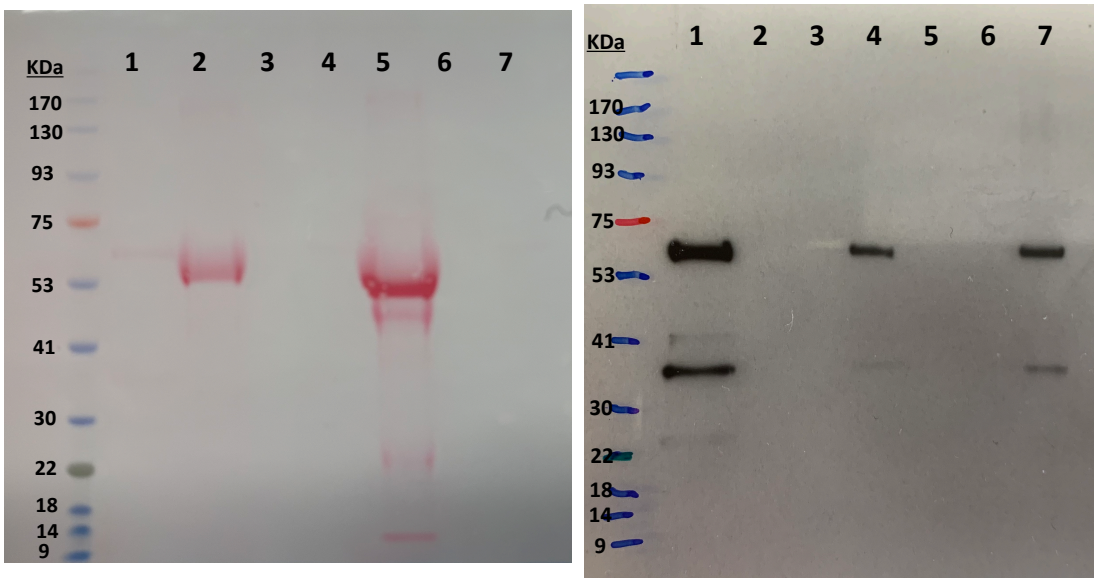


Fig. 3.4.1.1 *Uncropped Ponceau-stained membrane (left) and developed Western blot (right) for antibody immobilization of GST-MRG15 on beads shown in Fig. 3.2.1.1.1. See Table 4.4.1.1. for description of lanes.*

Number	*Description*	*µL sample*	*µL 2 x SDS*
	Ladder	3.5	
1	GST-MRG15	1.25	18.75
2	Antibody 1	1	19
3	Ab 1 beads	-	15
4	GST-MRG 15 on Ab 1 beads	-	15
5	Ab 2	1	19
6	Ab 2 beads	-	15
7	GST-MRG 15 on Ab 2 beads	-	15

Table 3.4.1.1. *Lanes in fig. 3.4.1.1.*

3.4.2 GST-MRG15 immobilised on GSH Sepharose beads

50 µL of GSH-Sepharose 4 Fast Flow beads (Quiagen) was prepared by washing three times with 300 µL wash buffer (200 mM NaCl, PBS pH 7.5, 0.05% Triton X-100, 2 mM DTT, 1 mM EDTA) via resuspension, centrifugation to pellet the beads (3000 rpm, 3 min, room temp) and removing the supernatant. The beads were then resuspended in 400 µL wash buffer and 40 µL of 0.8 mg/ mL GST-MRG15 was added. The beads were incubated for 16 h (overnight) with gentle rotation (20 rpm, 4°C). The beads were then washed once with 200 µL wash buffer, twice with 200 µL dialysis buffer (50 mM Tris. HCl pH 7.8, 150 mM NaCl, 1 mM EDTA, 2 mM DTT), and resuspended in 225 µL dialysis buffer.

Samples were made up in SDS sample buffer (NuPAGE LDS sample buffer + 10 mM DTT) and run on SDS-PAGE in MOPS-SDS buffer (130 V, 1 h 15 min, room temp). Bands were visualised by staining with Coomassie (SafeBlue protein stain, NBS biologicals) for 1 hour and destaining with MilliQ water overnight. Protein concentration on the beads was quantified by comparison with BSA band intensity in ImageJ and determined to be 0.035 mg/ mL.

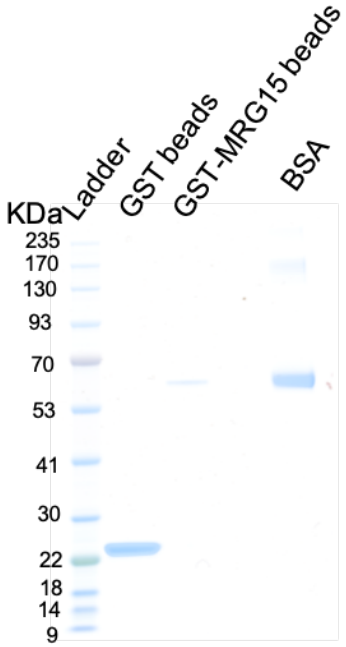


Fig. 3.4.2.1. *SDS-PAGE for GST-MRG15 immobilization on sepharose beads in Fig. 3.2.1.2.1.*

3.4.3 Pulldown assays with GSH-sepharose beads

GSH sepharose beads bearing 1.2 µg of GST-MRG15 (33.5 µL of 0.035 mg/ mL) or GST (10x dilution of stock, 4.8 µL of 0.26 mg/ mL) were washed with 200 µL of NET250 buffer (50 mM Tris. HCl pH 7.5, 250 mM NaCl, 0.5% NP40, 2 mM DTT) via resuspension, centrifugation to pellet the beads (3000 rpm, 3 min, 4°C) and removing the supernatant. They were then each resuspended in 600 µL BSA buffer (NET250 with 5 mg/ mL BSA) and split into three 195 µL aliquots for each of the three nucleosome variants. The volume was then adjusted with a further 200 µL BSA buffer in each tube, and blocking was carried out by incubating for 1h with gentle rotation (4°C, 20 rpm). Nucleosomes and salmon sperm DNA (Invitrogen; to out-compete non-specific DNA binding of MRG15) were then added:

Component	*Quantity (μL)*					
	WT/ GST	**Me2/ GST**	**Me3/ GST**	**WT/ MRG15**	**Me2/ MRG15**	**Me3/ MRG15**
186 wt	3.64			3.64		
186 me2		2.22			2.22	
186 me3			6.67			6.67
GST beads	1.56	1.56	1.56			
GST- MRG15 beads				20	20	20
NET250 + BSA	400	400	400	400	400	400
Salmon sperm DNA (10 mg/ mL stock, 10 x excess)	0.4	0.4	0.4	0.4	0.4	0.4

Table 3.4.3.1. *Composition of pulldown assay samples with GST-sepharose beads.*

The samples were then incubated for a further 1h with gentle rotation (4°C, 20 rpm). Samples were then spun down and washed three times with 200 μL NET250 (3000 rpm, 3 min, 4°C) to remove unbound nucleosome. They were then made up in 2 x SDS (NuPAGE LDS sample buffer + 80 mM DTT) and run on SDS-PAGE in MES-SDS buffer (200 V, 40 min, 4°C) and imaged via Western blot. PBS with 0.05% Tween20 was used for washing steps, and PBS-tween with 5% milk powder was used for blocking steps. Primary antibodies used were: Santa Cruz 138 α GST-HRP conjugate (1:1000 dilution) for GST portion of membrane, and rabbit α H3 for histone portion of membrane (Bethyl, A300-B22 A-1, 1:100 dilution). Secondary antibody used in both cases was goat α rabbit IgG (1:1000 dilution).

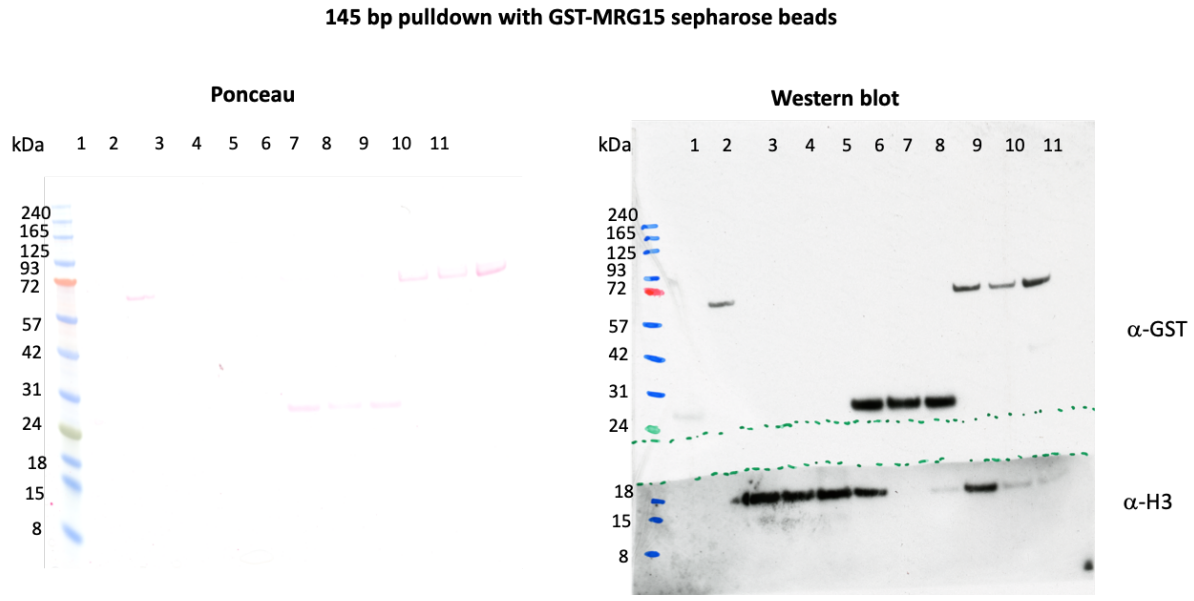


Fig. 3.4.3.1. *Uncropped Ponceau-stained membrane (left) and developed Western blot (right) for attempted pulldown of 145 bp nucleosomes with GST-sepharose beads in fig. 3.2.1.2.2.B. See table 4.4.3.2 for decription of lanes.*

Number	*Description*	*µL sample*	*µL 2 x SDS*
	Ladder	3.5	
1	GST beads	0.975	19
2	GST-MRG15 beads	12.5	15
3	145 bp WT nucleosome	2.35	17.65
4	145 bp me2 nucleosome	1.39	18.61
5	145 bp me3 nucleosome	4.46	15.54
6	WT/GST	-	15
7	Me2/GST	-	15
8	Me3/GST	-	15
9	WT/MRG15	-	15
10	Me2/MRG15	-	15
11	Me3/MRG15	-	15

Table 3.4.3.2. *Lanes in fig. 3.4.3.1.*

186 bp nucleosomes

GST sepharose beads bearing 1.2 μg of GST-MRG15 (33.5 μL of 0.035 mg/ mL) or GST (10x dilution of stock, 4.8 μL of 0.26 mg/ mL) were washed with 200 μL of NET250 buffer (50 mM Tris. HCl pH 7.5, 250 mM NaCl, 0.5% NP40, 2 mM DTT) via resuspension, centrifugation to pellet the beads (3000 rpm, 3 min, 4°C) and removing the supernatant. They were then each resuspended in 600 μL BSA buffer (NET250 with 5 mg/ mL BSA) and split into three 195 μL aliquots for each of the three nucleosome variants. The volume was then adjusted with a further 200 μL BSA buffer in each tube, and blocking was carried out by incubating for 1h with gentle rotation (4°C, 20 rpm). Nucleosomes and salmon sperm DNA (to out-compete non-specific DNA binding of MRG15) were then added:

	Quantity (μL)					
Component	**WT/ GST**	**Me2/ GST**	**Me3/ GST**	**WT/ MRG15**	**Me2/ MRG15**	**Me3/ MRG15**
186 wt	4.92			4.92		
186 me2		5.4			5.4	
186 me3			5.33			5.33
GST beads	1.53	1.53	1.53			
GST- MRG15 beads				11.3	11.3	11.3
NET250 + BSA	400	400	400	400	400	400
Salmon sperm DNA (10 mg/ mL stock, 10 x excess)	0.4	0.4	0.4	0.4	0.4	0.4

Table 3.4.3.3. *Composition of pulldown assay samples with 186 bp nucleosomes.*

Note that in all cases nucleosomes assembled with a 3:1 apparent DNA: octamer ratio were used. The samples were then incubated for a further 1h with gentle rotation (4°C, 20 rpm). Samples were then spun down and washed three times with 200 μL NET250 (3000 rpm, 3 min, 4°C) to remove unbound nucleosome. They were then made up in 2 x SDS (NuPAGE LDS

sample buffer + 80 mM DTT) and run on SDS-PAGE in MES-SDS buffer (200 V, 40 min, 4°C) and imaged via Western blot. PBS with 0.05% Tween20 was used for washing steps, and PBS-tween with 5% milk powder was used for blocking steps. Primary antibodies used were: Santa Cruz 138 α GST-HRP conjugate (1:1000 dilution) for GST portion of membrane, and rabbit α H3 for histone portion of membrane (1:100 dilution). Secondary antibody used in both cases was goat α rabbit IgG (1:1000 dilution).

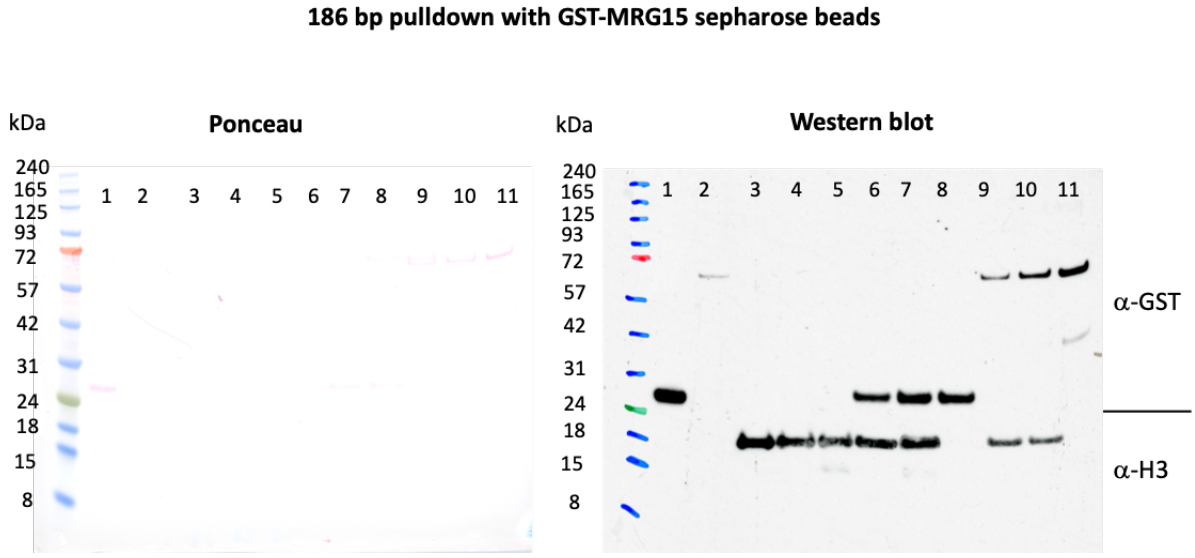


Fig. 3.4.3.2. *Uncropped Ponceau-stained membrane (left) and developed Western blot (right) for attempted pulldown of 186 bp nucleosomes with GST-sepharose beads in fig. 3.2.1.2.2.C. See table 3.4.3.4 for description of lanes.*

Number	*Description*	*µL sample*	*µL 2 x SDS*
	Ladder	3.5	
1	GST beads	0.77	19.24
2	GST-MRG15 beads	5.65	14.35
3	186 bp WT nucleosome	2.46	17.54
4	186 bp me2 nucleosome	2.70	17.30
5	186 bp me3 nucleosome	2.67	17.34
6	WT/GST	-	15
7	Me2/GST	-	15
8	Me3/GST	-	15
9	WT/MRG15	-	15
10	Me2/MRG15	-	15
11	Me3/MRG15	-	15

Table 3.4.3.4. *Lanes in fig. 3.4.3.2.*

3.4.4 Pulldown from 186 bp nucleosomes immobilized on streptavidin beads- GST-MRG15

Biotinylated 186 bp nucleosomes were immobilised onto Dynabeads Streptavidin MyOne T1 beads according to the manufacturer's protocol with modifications. Briefly, four 10 µL aliquots of beads were washed in three times 1 mL binding/ washing buffer (10 mM Tris. HCl pH 7.5, 250 mM KCl, 1 mM EDTA, 1 mM DTT, 0.1% NP40) at 4°C. Each aliquot of beads was then resuspended in 1 mL binding/ washing buffer and 0.2 µg of each 186 bp nucleosome variant or free 186 bp DNA was added. The beads were then incubated with gentle rotation (15 rpm, 30 min room temp) to allow nucleosome or DNA to bind. The beads were then washed carefully three times with 1 mL binding/ washing buffer with 0.1% BSA, and resuspended in 200 µL binding/ washing buffer with 5% BSA. Each of the resuspended bead samples was then divided into two 95 µL aliquots for GST and GST-MRG15 pulldowns, and made up to 1 mL with 900 µL binding/ washing buffer with 5% BSA. The beads were incubated for a further 1h with gentle rotation (15 rpm, 4°C) to ensure complete blocking; salmon sperm DNA and either

GST or GST-MRG15 were then added as appropriate before incubating for a further 4h (15 rpm, 4°C) to allow binding interactions between protein and/or DNA to equilibrate.

Component	*Amount needed (μg)*	*Concentration (mg/ mL)*
186 bp DNA (20 x dil)	0.2	0.167
186 wt	0.2	0.081
186 me2	0.2	0.074
186 me3	0.2	0.075
GST (20 x dilution)	0.1	0.325
GST-MRG15 (10 x dilution)	0.1	0.080
Salmon Sperm DNA (10 x dilution)	2.0	1.000

Table 3.4.4.1. *Quantities and concentrations of components in pulldown of GST-MRG15 using 186 bp nucleosomes immobilised on Streptavidin beads.*

	Quantity (μL)							
Component	**186 bp DNA/ GST**	**WT/ GST**	**me2/ GST**	**me3/ GST**	**186 bp DNA/ MRG15**	**WT/ MRG15**	**me2/ MRG15**	**me3/ MRG15**
186 bp DNA	0.598				0.598			
186 WT		1.23				1.23		
186 Me2			1.35				1.35	
186 Me3				1.333				1.333
GST	0.308	0.308	0.308	0.308				
GST-MRG15					1.25	1.25	1.25	1.25
BSA buffer	1000	1000	1000	1000	1000	1000	1000	1000
Salmon Sperm DNA	2	2	2	2	2	2	2	2

Table 3.4.4.2. *Composition of samples in pulldown of GST-MRG15 using 186 bp nucleosomes immobilised on Streptavidin beads.*

The beads were then washed four times with binding/ washing buffer to remove any unbound GST or GST-MRG15. They were then made up in 2 x SDS (NuPAGE LDS sample buffer + 80 mM DTT) and run on SDS-PAGE in MES-SDS buffer (200 V, 40 min, 4°C) and imaged

via Western blot. PBS with 0.05% Tween20 was used for washing steps, and PBS-tween with 5% milk powder was used for blocking steps. Primary antibodies used were: Santa Cruz 138 α GST-HRP conjugate (1:1000 dilution) for GST portion of membrane, and rabbit α H3 for histone portion of membrane (1:100 dilution). Secondary antibody used in both cases was goat α rabbit IgG (1:1000 dilution).

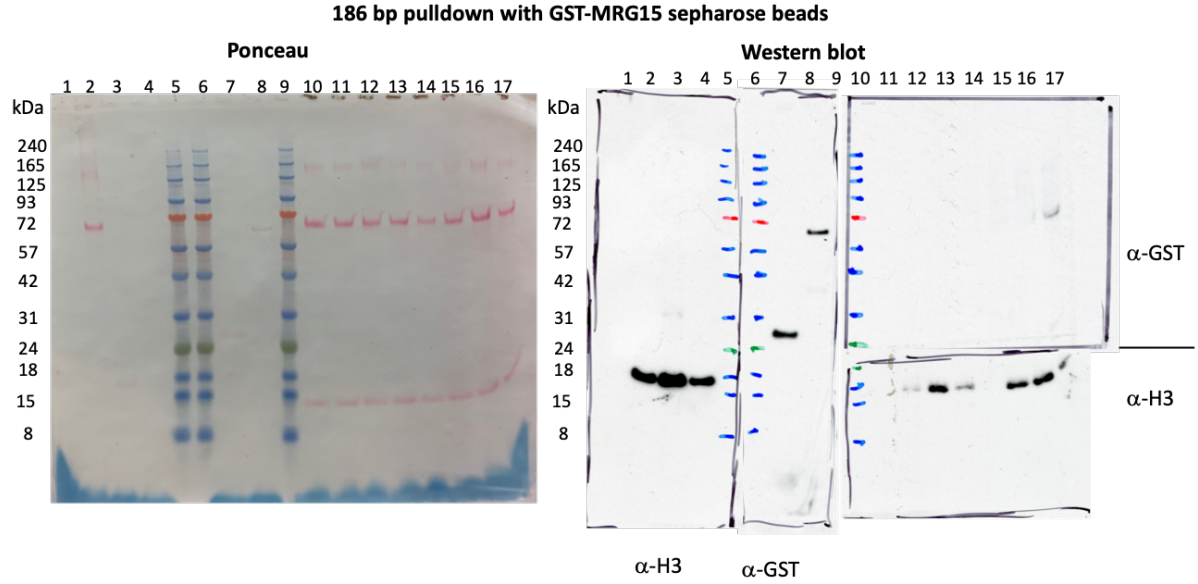


Fig. 3.4.4.1. *Uncropped Ponceau-stained membrane (left) and developed Western blot (right) for pulldown of GST-MRG15 using 186 bp nucleosomes immobilised on Streptavidin beads in fig. 3.2.1.3.1. See table 3.4.4.3. for description of lanes.*

Number	*Description*	*µL sample*	*µL 2 x SDS*
1	186 bp DNA (20 x dil)	0.60	9.40
2	186 bp WT nucleosome	1.23	8.77
3	186 bp me2 nucleosome	1.350	8.65
4	186 bp me3nucleosome	1.33	8.67
5	Ladder	3.5	-
6	Ladder	3.5	-
7	GST	0.31	9.69
8	GST-MRG15 (10 x dil)	1.25	8.75
9	Ladder	3.5	-
10	186 bp DNA/GST		
11	WT nucleosome/GST	-	15
12	me2 nucleosome/GST	-	15
13	Me3 nucleosome/GST	-	15
14	186 bp DNA/GST-MRG15	-	15
15	WT nucleosome/GST-MRG15	-	15
16	me2 nucleosome/GST-MRG15	-	15
17	Me3 nucleosome/GST-MRG15	-	15

Table 3.4.4.3. *Lanes in Fig. 3.4.4.1.*

4.4.5 Pulldown from 186 bp nucleosomes immobilized on streptavidin beads- buffer screen

Screen 1: 250 mM KCl buffer vs 150 mM NaCl buffer.

Biotinylated 186 bp nucleosomes were immobilised onto Dynabeads Streptavidin MyOne T1 beads according to the manufacturer's protocol with modifications. Briefly, eight 5 µL aliquots of beads were washed in three times 1 mL RRTCS or HBS (HBS: 10 mM HEPES pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.1% NP40) at 4°C. Each aliquot of beads was then resuspended in 1 mL of the appropriate buffer and 0.2 µg of each 186 bp nucleosome variant or free 186 bp DNA was added. The beads were then incubated with gentle rotation (15 rpm, 30 min room temp) to allow nucleosome or DNA to bind. The beads were then washed carefully three times with 1 mL binding/ washing buffer or HBS with 0.1% BSA, and resuspended in 1 mL binding/ washing buffer or HBS with 5% BSA. The beads were incubated for a further 1h with gentle rotation (15 rpm, 4°C) to ensure complete blocking; salmon sperm

DNA and MRG15 was added as appropriate before incubating for a further 4h (15 rpm, 4°C) to allow binding interactions between protein and/or DNA to equilibrate.

Component	*Amount needed (μg)*	*Concentration (mg/ mL)*
186 bp DNA (20 x dil)	0.2	0.167
186 wt	0.2	0.081
186 me2	0.2	0.074
186 me3	0.2	0.075
hMRG15	0.1	0.213
Salmon Sperm DNA (10 x dilution)	2.0	1.000

Table 3.4.5.1. *Quantities and concentrations of components in pulldown to screen 250 mM KCl buffer vs 150 mM NaCl buffer.*

Component	**186 bp DNA/ MRG15**	**WT/ MRG15**	**me2/ MRG15**	**me3/ MRG15**
186 bp DNA	0.598			
186 WT		1.23		
186 Me2			1.35	
186 Me3				1.333
GST-MRG15	1.174	1.174	1.174	1.174
BSA buffer	1000	1000	1000	1000
Salmon Sperm DNA	2	2	2	2

Table 3.4.5.2. *Composition of samples in pulldown to screen 250 mM KCl buffer vs 150 mM NaCl buffer.*

The beads were then washed four times with binding/ washing buffer to remove any unbound hMRG15. They were then made up in 2 x SDS (NuPAGE LDS sample buffer + 80 mM DTT) and run on SDS-PAGE in MES-SDS buffer (200 V, 40 min, 4°C) and imaged via Western blot. PBS with 0.05% Tween20 was used for washing steps, and PBS-tween with 5% milk powder

was used for blocking steps. Primary antibodies used were: rabbit monoclonal α MRG15 (Cell Signalling Technology, lot 1, 1:1000 dilution) for MRG15 portion of membrane, and rabbit α H3 for histone portion of membrane (1:100 dilution). Secondary antibody used in both cases was goat α rabbit IgG (1:1000 dilution).

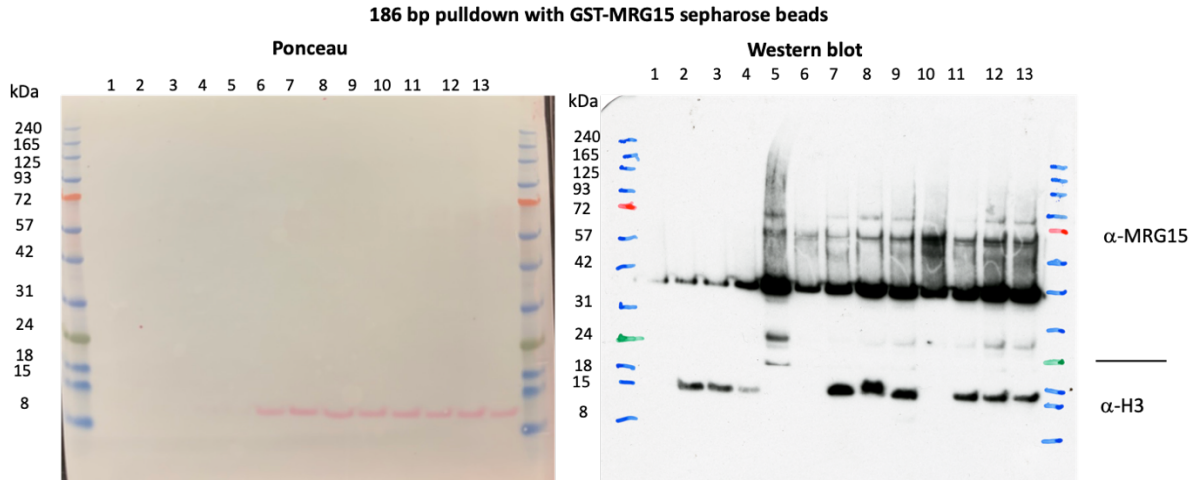


Fig. 3.4.5.1. *Uncropped Ponceau-stained membrane (left) and developed Western blot (right) for pulldown to screen 250 mM KCl buffer vs 150 mM NaCl buffer in fig. 3.2.1.3.2.A. See table 4.4.5.3. for description of lanes.*

Number	*Description*	*μL sample*	*μL 2 x SDS*
1	Ladder	3.5	-
2	186 bp DNA (20 x dil)	0.60	9.40
3	186 bp WT nucleosome	1.23	8.77
4	186 bp me2 nucleosome	1.350	8.65
5	186 bp me3 nucleosome	1.33	8.67
6	hMRG15 (2.5 x dil)	1.17	
7	186 bp DNA/hMRG15 (KCl)	-	8
8	WT nucleosome/hMRG15 (KCl)	-	8
9	me2 nucleosome/ hMRG15 (KCl)	-	8
10	Me3 nucleosome/ hMRG15 (KCl)	-	8
11	186 bp DNA/hMRG15 (NaCl)	-	8
12	WT nucleosome/hMRG15 (NaCl)	-	8
13	me2 nucleosome/ hMRG15 (NaCl)	-	8
14	Me3 nucleosome/ hMRG15 (NaCl)	-	8
15	Ladder	3.5	-

Table 3.4.5.3. *Lanes in fig. 3.4.5.1.*

Screen 2: 150 mM KCl buffer

Biotinylated 186 bp nucleosomes were immobilised onto Dynabeads Streptavidin MyOne T1 beads according to the manufacturer's protocol with modifications. Briefly, eight 5 μL aliquots of beads were washed in three times 1 mL HBK (10 mM HEPES pH 7.5, 150 mM KCl, 1 mM EDTA, 1 mM DTT, 0.1% NP40) at 4°C. Each aliquot of beads was then resuspended in 1 mL of the appropriate buffer and 0.2 μg of each 186 bp nucleosome variant or free 186 bp DNA was added. The beads were then incubated with gentle rotation (15 rpm, 30 min room temp) to allow nucleosome or DNA to bind. The beads were then washed carefully three times with 1 mL HBK with 0.1% BSA, and resuspended in 1 mL HBK with 5% BSA. The beads were incubated for a further 1h with gentle rotation (15 rpm, 4°C) to ensure complete blocking; salmon sperm DNA and MRG15 was added as appropriate before incubating for a further 4h (15 rpm, 4°C) to allow binding interactions between protein and/or DNA to equilibrate.

Component	*Amount needed (μg)*	*Concentration (mg/ mL)*
186 bp DNA (20 x dil)	0.2	0.167
186 wt	0.2	0.081
186 me2	0.2	0.074
186 me3	0.2	0.075
hMRG15	0.1	0.213
Salmon Sperm DNA (10 x dilution)	2.0	1.000

Table 3.4.5.4. *Quantities and concentrations of components in pulldown to screen 150 mM KCl buffer.*

Component	**186 bp DNA/ MRG15**	**WT/ MRG15**	**me2/ MRG15**	**me3/ MRG15**
186 bp DNA	0.598			
186 WT		1.23		
186 Me2			1.35	
186 Me3				1.333
GST-MRG15	1.174	1.174	1.174	1.174
BSA buffer	1000	1000	1000	1000
Salmon Sperm DNA	2	2	2	2

Table 3.4.5.5. *Composition of samples in pulldown to screen 150 mM KCl buffer.*

The beads were then washed four times with binding/ washing buffer to remove any unbound hMRG15. They were then made up in 2 x SDS (NuPAGE LDS sample buffer + 80 mM DTT) and run on SDS-PAGE in MES-SDS buffer (200 V, 40 min, 4°C) and imaged via Western blot. TBS with 0.05% Tween20 was used for washing steps, and TBS-tween with 5% BSA was used for blocking steps. Primary antibodies used were: α MRG15 (1:1000 dilution) for MRG15 portion of membrane, and rabbit α H3 for histone portion of membrane (1:100 dilution). Secondary antibody used in both cases was goat α rabbit IgG (1:1000 dilution).

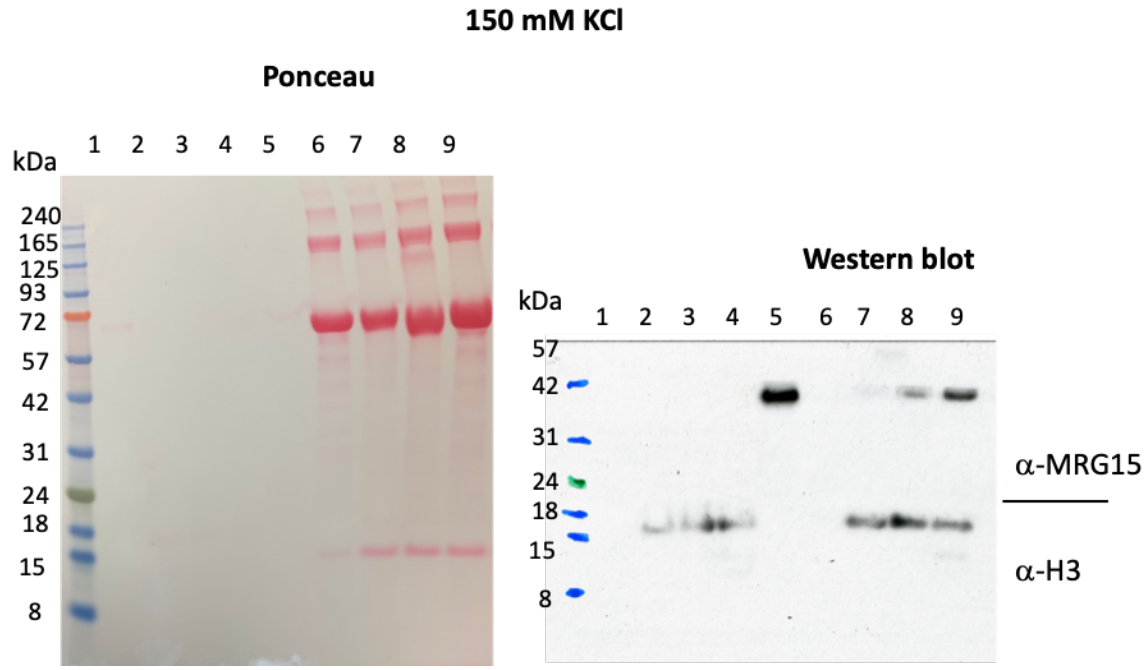


Fig. 3.4.5.2. *Uncropped Ponceau-stained membrane (left) and developed Western blot (right) for pulldown to screen 150 mM KCl buffer in fig. 3.2.1.3.2.B. See table 3.4.5.6. for description of lanes.*

Number	*Description*	*µL sample*	*µL 2 x SDS*
1	Ladder	3.5	-
2	186 bp DNA (20 x dil)	0.60	9.40
3	186 bp WT nucleosome	1.23	8.77
4	186 bp me2 nucleosome	1.350	8.65
5	186 bp me3 nucleosome	1.33	8.67
6	hMRG15 (2.5 x dil)	1.17	
7	186 bp DNA/hMRG15 (KCl)	-	8
8	WT nucleosome/hMRG15 (KCl)	-	8
9	me2 nucleosome/ hMRG15 (KCl)	-	8
10	Me3 nucleosome/ hMRG15 (KCl)	-	8

Table 3.4.5.6. *Lanes in fig. 3.4.5.2.*

3.4.6 Initial SPR measurements of MRG15 binding to immobilised 186 bp nucleosome

All measurements were performed using a Biocore T200 (GE Healthcare). Biotinylated 186 bp nucleosomes were immobilised as the substrate for SPR measurements with hMRG15 as the

analyte. A CM5 chip was first functionalised with streptavidin using the following regime at 5 µL/min with HBS-EP as the buffer. First, a 1:1 mixture of NHS and EDC was run over (600s contact time) to generate a reactive amine surface, followed by 0.5 mg/ mL Streptavidin in acetate buffer (10 mM Sodium Acetate pH 5.0, 600s contact time), then ethanolamine (contact time 300s) to quench the activated surface, and glycine (50 mM pH 2.5 in 10 mM sodium acetate buffer) to displace any unbound streptavidin.

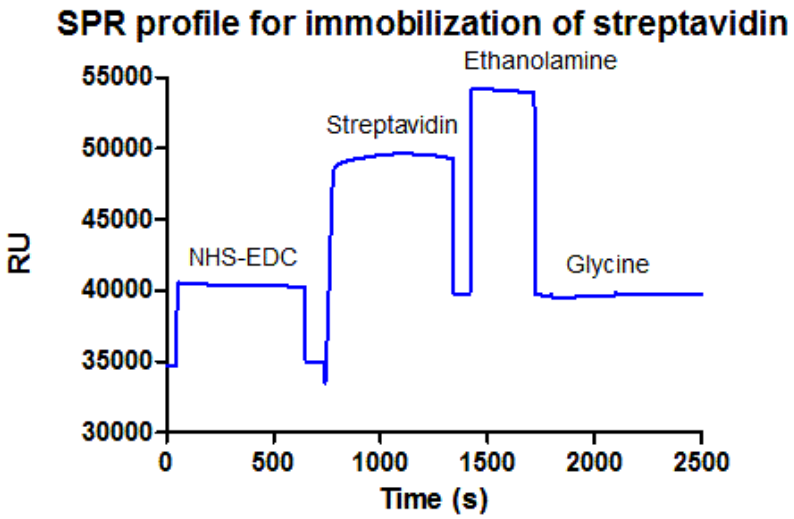


Fig.3.4.6.1. *A typical SPR profile for immobilization of streptavidin onto CM5 chip by amide crosslinking.*

Buffer was then switched to RRTCS (10 mM Tris pH 7.5, 250 mM KCl, 1 mM EDTA, 1 mM DTT, 0.1% NP40) for nucleosome immobilization. Biotinylated 186 bp WT, me2 and me3 nucleosomes (all assembled with a 3:1 DNA: octamer ratio) were immobilised in channels 2,3 and 4 respectively; channel 1 was reserved for a control. Nucleosomes were made up to 10 µg/ mL and flowed over a single channel at 5 µL/ min for 300s contact time, followed by 300s wait time before switching to the next channel and next nucleosome. Final RU values measured for each immobilised nucleosome are shown below.

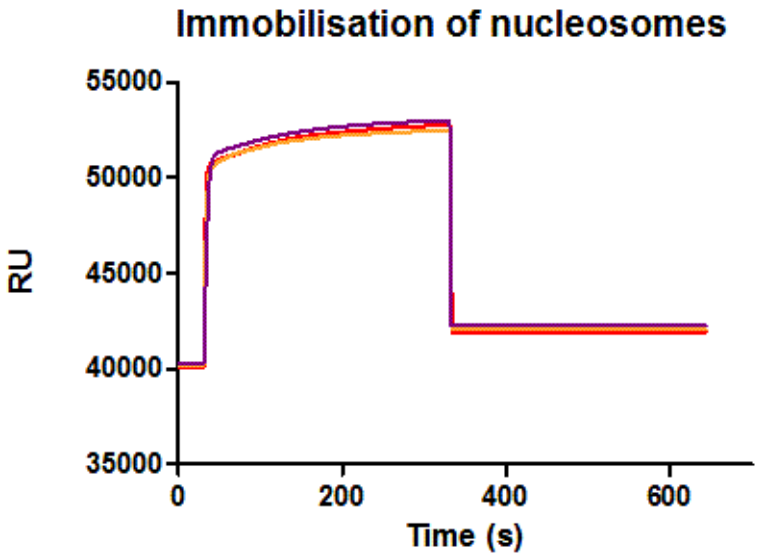


Fig. 3.4.6.2. *SPR profile for nucleosome immobilization. Each nucleosome is flowed over the appropriate channel (50- 350 s) then unbound nucleosomes are washed away (350 – 650 s).*

Nucleosome variant	*RU immobilised*
WT	1670
Me2	1850
Me3	1870

Table 3.4.6.1. *Final RU values for each immobilised nucleosome*

For an initial screen to find the ideal concentration range, the analyte hMRG15 was made up in a 2x dilution series of twenty samples beginning from 10.5 μM, and samples were analysed sequentially from lowest concentration to highest (5 μL/min, 300s contact time followed by 300s wait time between samples). Initial measurements were performed in RRTCS buffer; however, when no binding was observed even at highest concentration, the experiment was repeated in HBS-EP+ 1 mM DTT. Measurements were then repeated twice more with the highest 9 concentrations, with the chip regenerated between runs with 10 mM Tris pH 7.5, 400

mM KCl, 1 mM EDTA, 1 mM DTT, 0.1% NP40 contact time 7200 s. 300s was taken as the point where equilibrium was reached.

A global baseline was taken as the RU pre-injection of the lowest concentration sample and subtracted during data processing. Apparent equilibrium dissociation constants ($K_{D\ app}$) were determined by fitting a one-site binding-saturation curve in Prism software.

Processed data is shown below: for the raw data, see Appendix 3.6.1.

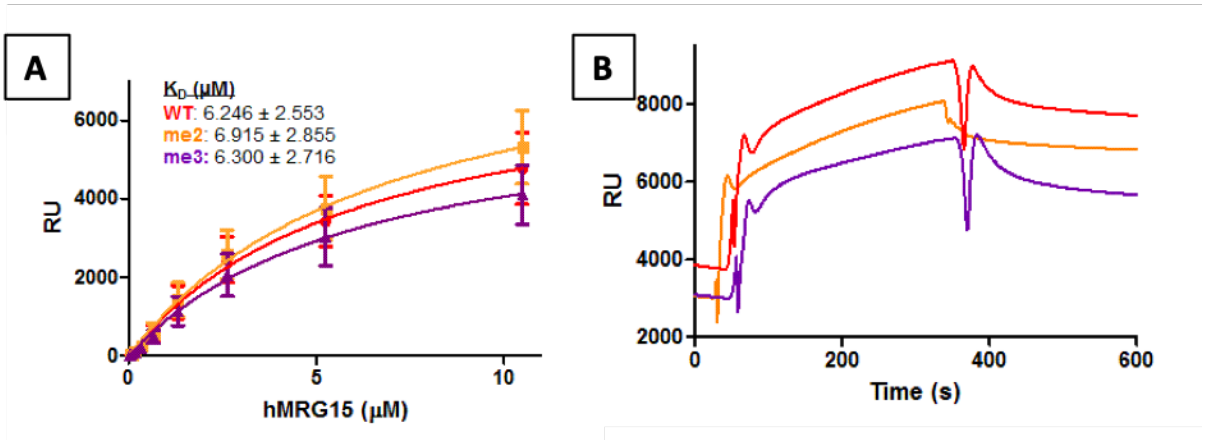


Fig. 3.4.6.3 (A) *Curve used to calculate* K_{Dapp} *values* ***(B)*** *Raw SPR curves for highest [hMRG15]. Reproduced from fig. 3.2.2.1.A.*

3.4.7 Blue Native PAGE analysis of hMRG15 and GST-MRG15

GST-MRG15 and hMRG15 were analysed using the Blue NativePAGE Novex Bis-Tris gel system according to manufacturer's protocol optimised for Coomassie staining with Dark Blue cathode buffer. Samples were made up with 5 μL thawed protein stock and 5 μL 100% glycerol, and loaded onto a Novex Bis Tris Native PAGE gel alongside NativeMark Unstained Protein Standard (Thermo Fisher). The gel was run at 250 V for 60 min, after which the gel was fixed and destained according to manufacturer's protocols.

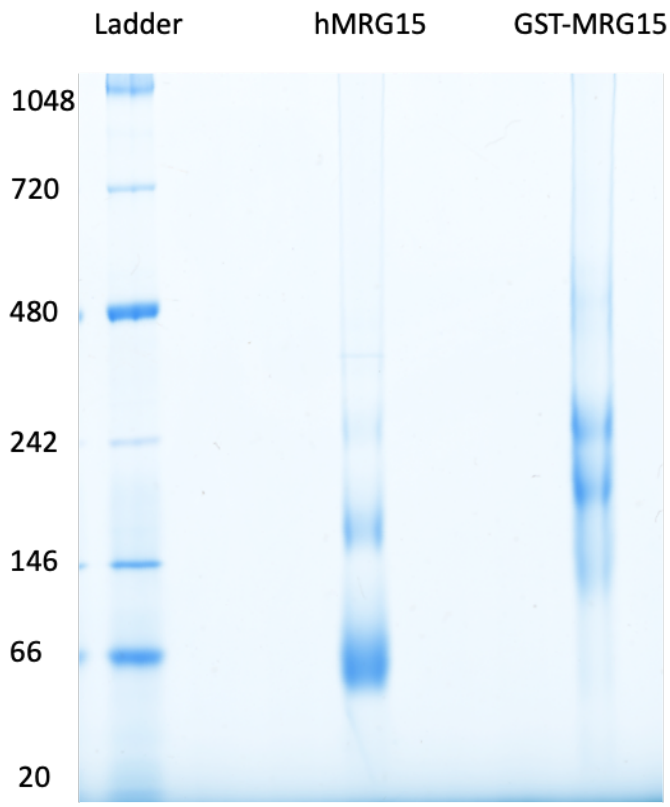


Fig. 3.4.7.1. *Blue Native PAGE analysis of hMRG15 and GST-MRG15 after freeze-thawing. Reproduced from fig. 3.2.2.2.*

3.4.8 Expression of GST-hMRG15

For each 1 L culture, precultures were prepared by inoculating 1 mL LB supplemented with ampicillin (100 µg/ mL) and gentamycin (20 µg/ mL) with Pgex6P3 containing GST-His_6-MRG15 from glycerol stock (Anthony Couturier) and incubated at 37 °C (1 h, 180 rpm) before transferring to 50 mL LB/ amp/ gent. The 50 mL precultures were further incubated for 16h (overnight; 37 °C, 180 rpm).

50 mL preculture was then used to incubate 1L LB/ amp/ gent. Each 1L culture was grown at 30°C (200 rpm) until an OD_{600} of approx 0.5 was reached, at which point the temperature was lowered to 13 °C before induction with the addition of IPTG to 0.2 mM. The culture was

incubated for a further 24 h to allow protein to express, before pellet was harvested by centrifugation (4°C, 8000 rpm, 10 min in Beckman Coulter Avanti JXN-26 centrifuge, JLA 8.1 rotor), flash frozen in N_2 (l) and stored at -80 °C.

3.4.9 Purification of hMRG15

The pellet was thawed at room temp and made up to 30 mL with the addition of Tritbuffer (PBS pH 7.4 supplemented with 150 mM NaCl, 0.05% Triton X100) and Protease Inhibitor Cocktail (PIC, 150 µL of 200 x stock, Calbiochem). Cells were lysed on ice with sonication (three 30s pulses at force 12, name sonicator type), after which DNAse I (1 µL, Roche) and $MgCl_2$ (to 1 mM, 30 µL of 1 mM stock) were added. The lysate was incubated at 4 °C with gentle rotation before centrifugation (45 min, 23,000 rpm, 4°C) to pellet insoluble debris. The supernatant was decanted, filtered (45 µm, specs), and added to GST- Sepharose beads (2 mL of slurry, pre-washed three times with 50 mL tritbuffer by resuspending in buffer, pelleting by centrifugation at 4°C 4000 rpm 10 min, and decanting supernatant to remove). The GST beads were incubated (1 h 30 min, 4 °C, gentle rotation) to allow the GST tag to bind, before loading onto a gravity column and washing once with tritbuffer (50 mL) and twice with wash buffer (50 mL of 50 mM Na_2HPO_4, 50 mM NaH_2PO_4, pH 7.4, 0.05% Triton X100, 200 mM NaCl, 5 mM Imidazole). The beads were then resuspended in tritbuffer (30 mL) and supplemented with 5 mM ATP (300 µL 0.5 M stock) and 15 mM $MgCl_2$ (450 µL of 1 M stock) and incubated (1 h, 4 °C, gentle rotation) before washing twice with tritbuffer (30 mL) and twice with p5 buffer (30 mL of 50 mM Na_2HPO_4, 50 mM NaH_2PO_4, pH 7.4, 0.05% Triton X100, 200 mM NaCl, 5 mM Imidazole). PreScission protease (expressed by A. Couturier) was then added to cleave hMRG15 from the immobilised GST tag, and incubated for 16 h (4 °C, gentle rotation).

The GST bead slurry was then loaded onto a gravity column and the supernatant was eluted directly onto Talon beads (600 μL of slurry, pre-washed three times with 50 mL p5 by resuspending in buffer, pelleting by centrifugation at 4°C 4000 rpm 10 min, and decanting supernatant to remove). The talon beads were incubated with the supernatant (1 h 30 min, 4 °C, gentle rotation) to allow the His_6 tag to bind. The beads were then loaded onto a gravity column and washed with p5 (50 mL) and p30 (50 mL of 50 mM Na_2HPO_4, 50 mM NaH_2PO_4, pH 7.4, 0.05% Triton X100, 200 mM NaCl, 30 mM Imidazole) before elution in p500 (1 mL aliquots of 50 mM Na_2HPO_4, 50 mM NaH_2PO_4, pH 7.4, 0.05% Triton X100, 200 mM NaCl, 500 mM Imidazole). All loaded directly onto PD10 columns (pre-washed with SEC buffer according to manufacturer's instructions: 50 mM Tris. HCl pH 7.5, 150 mM NaCl, 2 mM DTT, 1 mM EDTA) and eluted in 3.5 mL SEC buffer to remove imidazole. Fractions were then pooled and imidazole was further removed by overnight dialysis in SEC buffer (4 °C, 16 h, 6 kDa membrane dialysis cassette).

The dialysed hMRG15 was then concentrated to 500 μL (10 kDa vivaspin, 4 °C, x rpm, details) and loaded onto a S75 gel filtration column (GE life technologies) pre-equilibrated with SEC buffer. Protein was then eluted over 1 CV and protein-containing fractions identified by UV trace. A yield of 5.84 mg/ mL in 1 mL was obtained (0.97 mg/ L culture).

Round 1

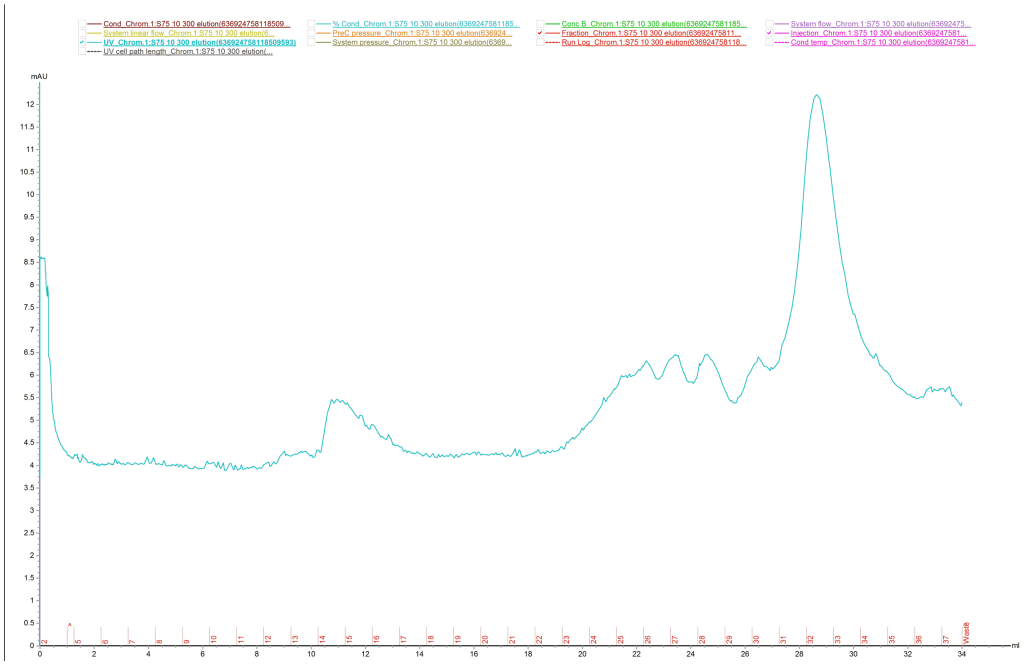
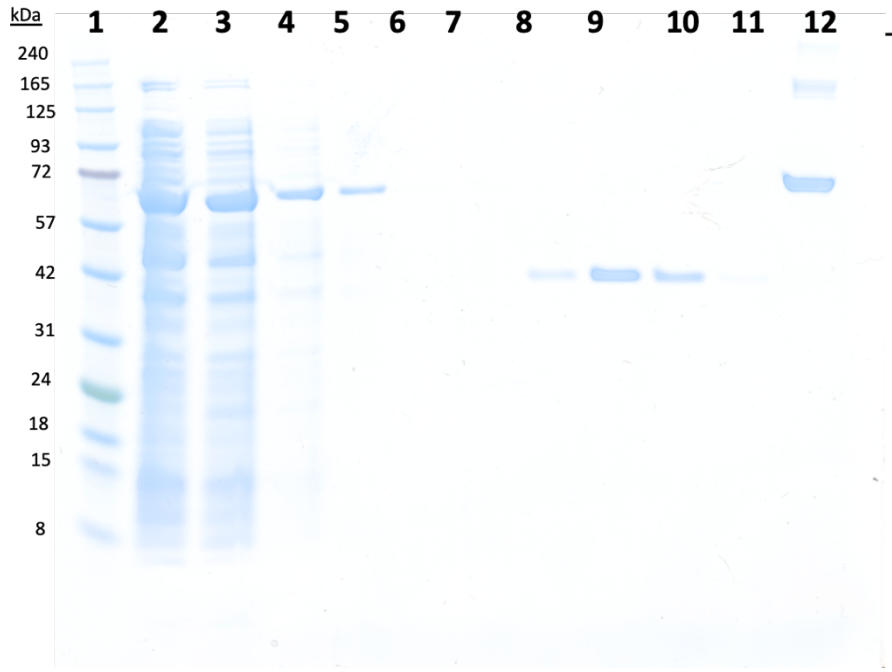


Fig. 3.4.9.1 *SEC UV trace for hMRG15 purification. Reproduced from fig. 3.2.2.3.A.*



	Sample name
1	Ladder
2	Pre-induction
3	Post-induction
4	GST beads 1
5	GST beads 2
6	p30 FT
7	E1
8	E2
9	E3
10	E4
11	p500 wash
12	BSA

Fig. 3.4.9.2 *SDS-PAGE analysis of purification (4-12% Bis-Tris gel, MES buffer); Ladder= BluEye prestained protein standard. Pre-induction= sample taken from culture before IPTG added. Post-induction= sample taken after IPTG added. GST beads 1 and 2= samples of GST-MRG15 bound to GSH sepharose beads. P30 FT= flowthrough from wash with p30 buffer. E1,2,3 and 4= elution with p500 rounds 1, 2, 3 and 4. P500 wash= final wash with p500 buffer. BSA= positive loading control (1μg)*

Round 2

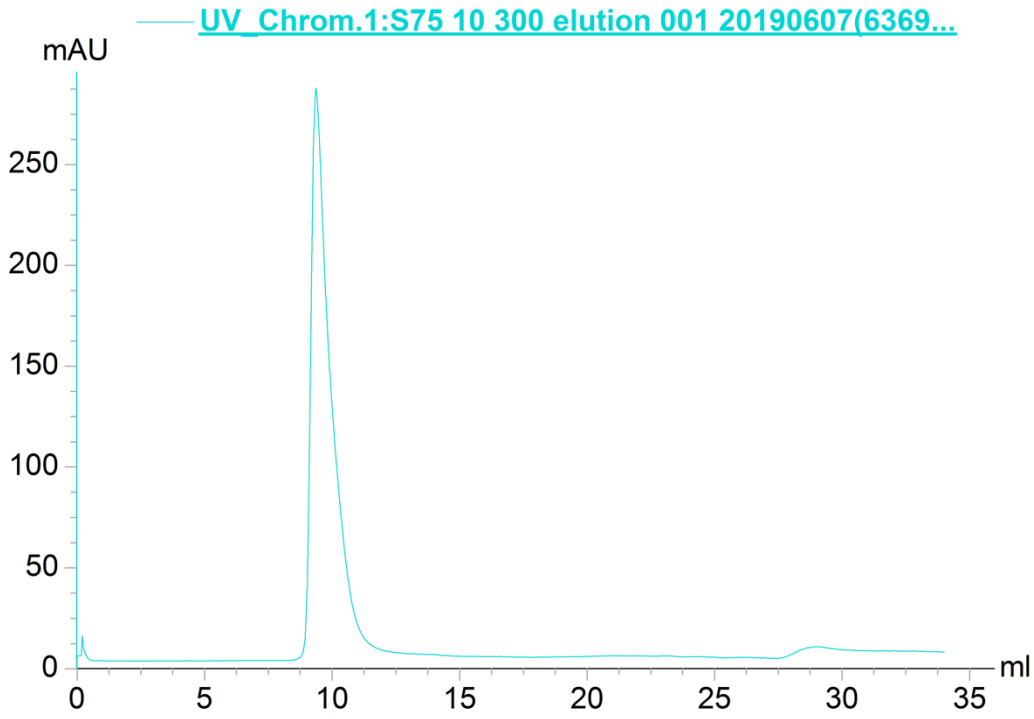


Fig. 3.4.9.3 *SEC UV trace for hMRG15 purification. Reproduced from fig. 3.2.2.3.B.*

3.4.10 Further SPR experiments

MRG15 purification and immediate SPR- round 1

A new CM5 chip was functionalised with streptavidin using the same regime as previously described at 5 µL/min with HBS-EP as the buffer. First, a 1:1 mixture of NHS and EDC was run over (600s contact time) followed by 0.5 mg/ mL Streptavidin in acetate buffer (10 mM sodium acetate pH 5.0, 600s contact time), then ethanolamine (contact time 300s), and glycine (50 mM pH 2.5 in 10 mM sodium acetate buffer).

Buffer was then switched to RRTCS (10 mM Tris pH 7.5, 250 mM KCl, 1 mM EDTA, 1 mM DTT, 0.1% NP40) for nucleosome immobilization. Biotinylated 186 bp WT, me2 and me3 nucleosomes (all assembled with a 3:1 DNA: octamer ratio) were immobilised in channels 2,3 and 4 respectively; channel 1 was reserved for a control. Nucleosomes were made up to 10 µg/ mL and flowed over a single channel at 5 µL/ min for 300s contact time, followed by 300s wait time before switching to the next channel and next nucleosome. Final RU values measured for each immobilised nucleosome are shown below:

Nucleosome variant	*RU immobilised*
WT	1940
Me2	1900
Me3	1800

Table 3.4.10.1 *Final RU values for immobilised nucleosomes- round 1*

The analyte hMRG15 was made up in a 2x dilution series of nine samples beginning from 10.5 µM, and samples were analysed sequentially from lowest concentration to highest (5 µL/min, 300s contact time followed by 300s wait time between samples) in HBK-BSA (10 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT, 0.1% NP40, 5% BSA). Measurements were then repeated

twice more, with the chip regenerated between runs with 10 mM Tris pH 7.5, 400 mM KCl, 1 mM EDTA, 1 mM DTT, 0.1% NP40 contact time 7200 s. 300s was taken as the point where equilibrium was reached.

A global baseline was taken as the RU pre-injection of the lowest concentration sample and subtracted during data processing. Apparent equilibrium dissociation constants ($k_{D\ app}$) were determined by fitting a one-site binding-saturation curve in Prism software.

Processed data is shown below: for the raw data, see Appendix 3.6.2.

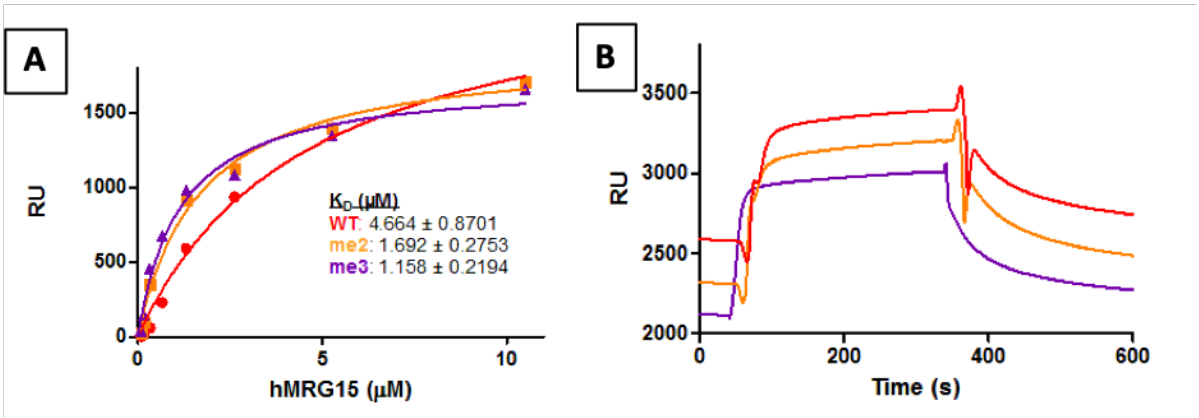


Fig. 3.4.10.1 (A) *Curve used to calculate K_{Dass} values* ***(B)*** *Raw SPR curves for highest [hMRG15]. Reproduced from fig. 3.2.2.1.B.*

MRG15 purification and immediate SPR- round 2

A new CM5 chip was functionalised with streptavidin using the same regime as previously described at 5 μL/min with HBS-EP as the buffer. First, a 1:1 mixture of NHS and EDC was run over (600s contact time) followed by 0.5 mg/ mL Streptavidin in acetate buffer (10 mM sodium acetate 5.0, 600s contact time), then ethanolamine (contact time 300s), and glycine (50 mM pH 2.5 in 10 mM sodium acetate buffer).

Buffer was then switched to RRTCS (10 mM Tris pH 7.5, 250 mM KCl, 1 mM EDTA, 1 mM DTT, 0.1% NP40) for nucleosome immobilization. Biotinylated 186 bp WT, me2 and me3 nucleosomes (all assembled with a 3:1 DNA: octamer ratio) were immobilised in channels 2,3

and 4 respectively; channel 1 was reserved for a control. Nucleosomes were made up to 10 μg/ mL and flowed over a single channel at 5 μL/ min for 300s contact time, followed by 300s wait time before switching to the next channel and next nucleosome. Final RU values measured for each immobilised nucleosome are shown below:

Nucleosome variant	*RU immobilised*
WT	980
Me2	700
Me3	630

Table 3.4.10.2 *Final RU values for immobilised nucleosomes- round 2*

The analyte hMRG15 was made up in a 2x dilution series of ten samples beginning from 21 μM, and samples were analysed sequentially from lowest concentration to highest (5 μL/min, 300s contact time followed by 300s wait time between samples) in HBK-BSA-SSDNA (10 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT, 0.1% NP40, 5% BSA, 0.01% Salmon Sperm DNA). As the highest concentration gave a sloped 'equilibrium' binding phase, it was not used in subsequent repeats. Measurements were then repeated twice more, with the chip regenerated between runs with 10 mM Tris pH 7.5, 400 mM KCl, 1 mM EDTA, 1 mM DTT, 0.1% NP40 contact time 7200 s. 300s was taken as the point where equilibrium was reached.

A global baseline was taken as the RU pre-injection of the lowest concentration sample and subtracted during data processing. Apparent equilibrium dissociation constants ($k_{D\ app}$) were determined by fitting a one-site binding-saturation curve in Prism software.

Processed data is shown below: for the raw data, see Appendix 3.6.3.

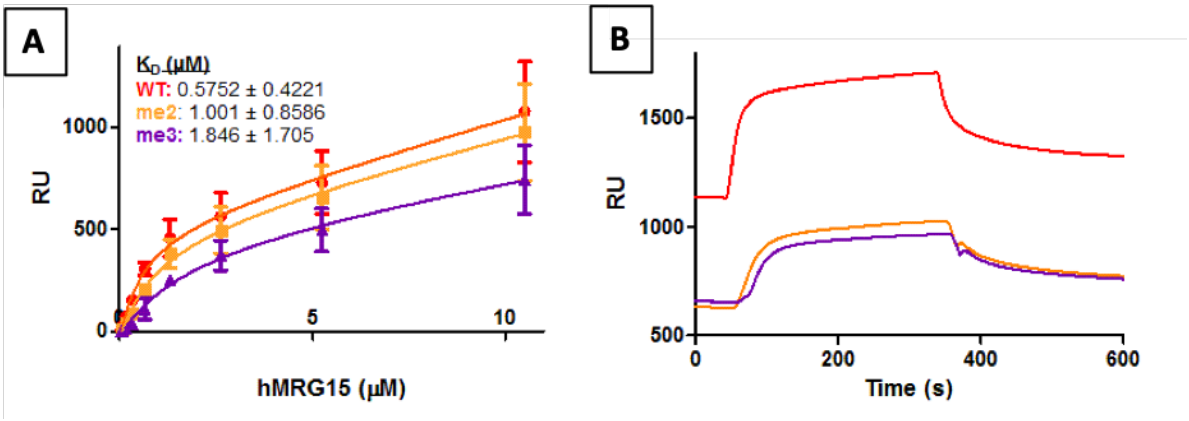


Fig. 3.4.10.2 (A) *Curve used to calculate K_D values* ***(B)*** *Raw SPR curves for highest [hMRG15]. Reproduced from fig. 3.2.2.1.C.*

3.4.11 Generation of MRG15 W53A mutant

Site-directed mutagenesis to generate MRG15 W53A mutant

MRG15 W53A was generated by site-directed mutagenesis using an Agilent Quikchange XL Site-Directed Mutagenesis kit (Agilent #200517) with the MRG15 WT in pGEX-6P-3 vector. Primers were taken from Bleuyard et. al.[42]

F: 5‘-CTTCATACATTACAGTGGTTGGAATAAAAATGCGGATGAATGGGTTCCG-3‘

R: 5‘-CGGAACCCATTCATCCGCATTTTTATTCCAACCACTGTAATGTATGAAG-3‘

Successful generation of the mutant was verified by sequencing.

Expression and purification of MRG15 W53A

Expression and purification of MRG15 W53A was carried out according to the protocol for MRG15 WT. However, the final SEC purification step was eliminated; instead, aliquots were flash-frozen immediately after dialysis.

Pulldown from biotinylated nucleosomes with MRG15 W53A negative control

Biotinylated 186 bp nucleosomes were immobilised onto Dynabeads Streptavidin MyOne T1 beads according to the manufacturer’s protocol with modifications. Briefly, four 10 µL aliquots

of beads were washed in three times 1 mL RRTCS (10 mM Tris. HCl pH 7.5, 250 mM KCl, 1 mM EDTA, 1 mM DTT, 0.1% NP40) at 4°C. Each aliquot of beads was then resuspended in 1 mL RRTCS and 0.4 μg of each 186 bp nucleosome variant or free 186 bp DNA was added. The beads were then incubated with gentle rotation (15 rpm, 30 min room temp) to allow nucleosome or DNA to bind. The beads were then washed carefully three times with 1 mL RRTCS with 0.1% BSA, divided into two, and resuspended in 1 mL RRTCS with 5% BSA. The beads were incubated for a further 1h with gentle rotation (15 rpm, 4°C) to ensure complete blocking; salmon sperm DNA and hMRG15 WT or W53A were added as appropriate before incubating for a further 4h (15 rpm, 4°C) to allow binding interactions between protein and/or DNA to equilibrate.

Component	*Amount needed (μg)*	*Concentration (mg/ mL)*
186 bp DNA (20 x dil)	0.4	0.167
186 wt	0.4	0.022
186 me2	0.4	0.074
186 me3	0.4	0.075
hMRG15	0.4	0.213
Salmon Sperm DNA (10 x dilution)	4.0	1.000

Table 3.4.11.1. *Quantities and concentrations of components in pulldown with MRG15 W3A*

Component	*186 bp DNA/ WT MRG15*	*WT/ WT MRG15*	*me2/ WT MRG15*	*me3/ WT MRG15*	*186 bp DNA/ WT MRG15*	*WT/ WT MRG15*	*me2/ WT MRG15*	*me3/ WT MRG15*
186 bp DNA (20 x dil)	0.598				0.598			
186 WT		8.809				8.809		
186 Me2			2.7				2.7	
186 Me3				2.665				2.665
hMRG15 (10x dil)	2.347	2.347	2.347	2.347				

hMRG15 W53A					5.356	5.356	5.356	5.356
BSA buffer	1000	1000	1000	1000	1000	1000	1000	1000
Salmon Sperm DNA (10 x dil)	4	4	4	4	4	4	4	4

Table 3.4.11.2. *Composition of samples in pulldown with WT MRG15 vs W53A*

The beads were then washed four times with binding/ washing buffer to remove any unbound hMRG15. They were then made up in 2 x SDS (NuPAGE LDS sample buffer + 80 mM DTT) and run on SDS-PAGE in MES-SDS buffer (200 V, 40 min, 4°C) and imaged via Western blot. TBS with 0.05% Tween20 was used for washing steps, and TBS-tween with 5% BSA was used for blocking steps. Primary antibodies used were: α MR15 (1:1000 dilution) for MRG15 portion of membrane, and rabbit α H3 for histone portion of membrane (1:100 dilution). Secondary antibody used in both cases was goat α rabbit IgG (1:1000 dilution).

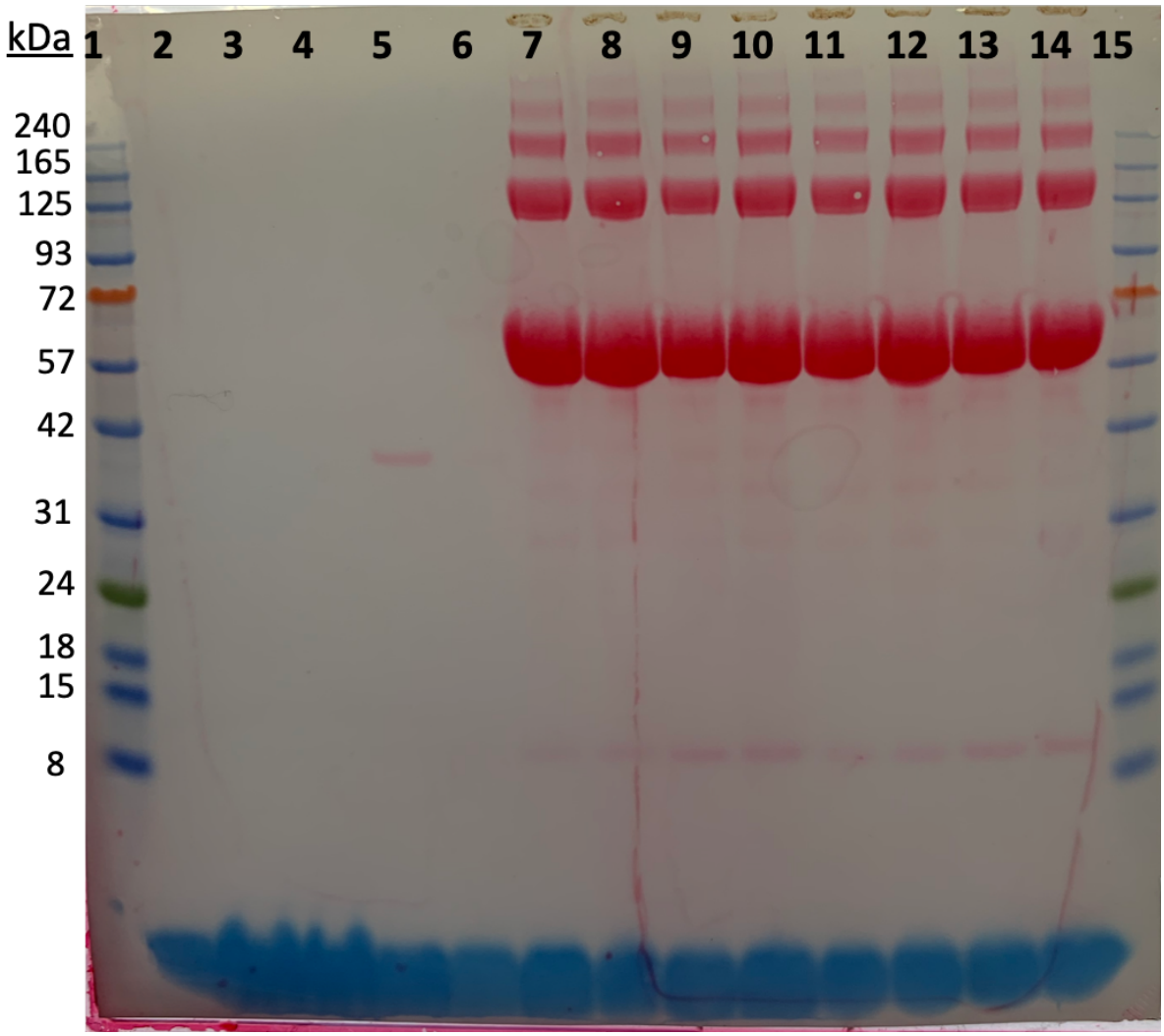


Fig. 3.4.11.1 Ponceau stained membrane for pulldown with MRG15 WT and W53A in fig. 3.2.3.1.B. See Table 3.4.11.3 for description of lanes.

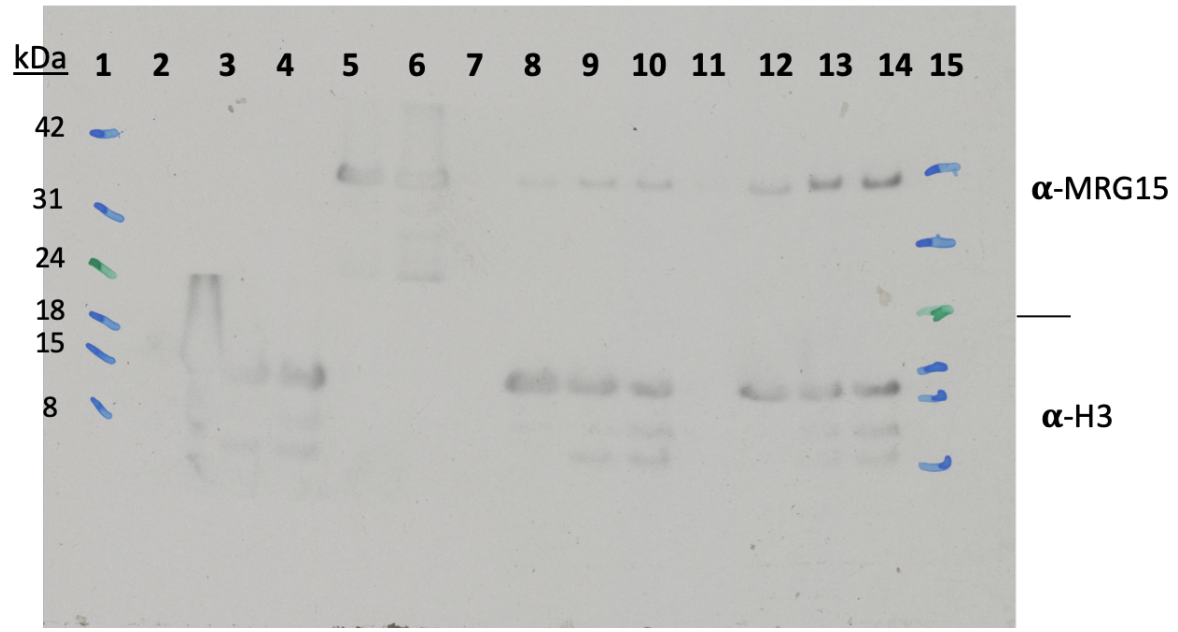


Fig. 3.4.11.2 Developed western blot of pulldown with MRG15 WT and W53A in fig. 3.2.3.1.B. (short exposure). See Table 3.4.11.3 for description of lanes.

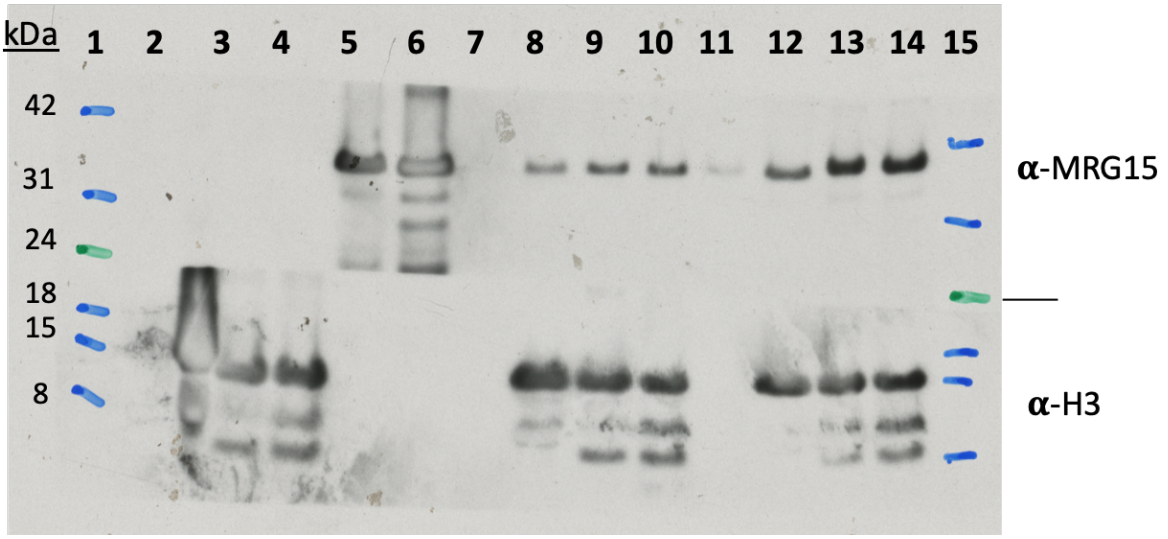


Fig. 3.4.11.3 Developed western blot of pulldown with MRG15 WT and W53A in fig. 3.2.3.1.B. (long exposure). See Table 3.4.11.3 for description of lanes.

	Sample name	*Sample vol (µL)*	*SDS vol (µL)*
1	Ladder	2.5	
2	WT	8.809	1.191

3	Me2	2.7	7.3
4	Me3	2.665	7.335
5	hMRG15 W53A	5.3	4.7
6	hMRG15 5x dil	2.347	7.653
7	186 bp DNA/ MRG15 W53A		8.000
8	WT/ MRG15 W53A		8.000
9	Me2/ MRG15 W53A		8.000
10	Me3/ MRG15 W53A		8.000
11	186 bp DNA/ MRG15 WT		8.000
12	WT/ MRG15 WT		8.000
13	Me2/ MRG15 WT		8.000
14	Me3/ MRG15 WT		8.000
15	ladder	2.5	

Table 3.4.11.3 Description of lanes in pulldown with MRG15 W53A and WT.

3.4.12 DNAse I digest experiments

Initial optimization of conditions using me2 nucleosomes

Nucleosome and DNAse I (Roche) were combined in DNAse I buffer (10 mM Tris. HCl pH 7.5, 2.5 mM $MgCl_2$, 500 µM $CaCl_2$) to a final volume of 7 µL and incubated at room temperature for 30 min. 7 µL denaturing buffer (8M urea, 2M NaCl) was then added and samples heated at 95 °C for 10 min to stop the reaction and denature nucleosomes.

	Quantity							
Component	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
145 bp DNA (µL)	1	1	1	1				
186 bp DNA (µL)					1	1	1	1
DNAse I (KU)	0	0.002	0.0002	0.00002	0	0.002	0.0002	0.00002

Table 3.4.12.1 *Composition of DNAse I digest reactions*

3 μL 6X purple loading dye (NEB) was then added, and samples loaded onto a 15% TBE-Urea gel (Novex) and run un 1x TBE buffer at 180 V, 4 °C. The gel was stained using SYBR gold (6 μL in 20 mL MilliQ for 20 min, destained 2 x 5 min in 20 mL MilliQ) and imaged.

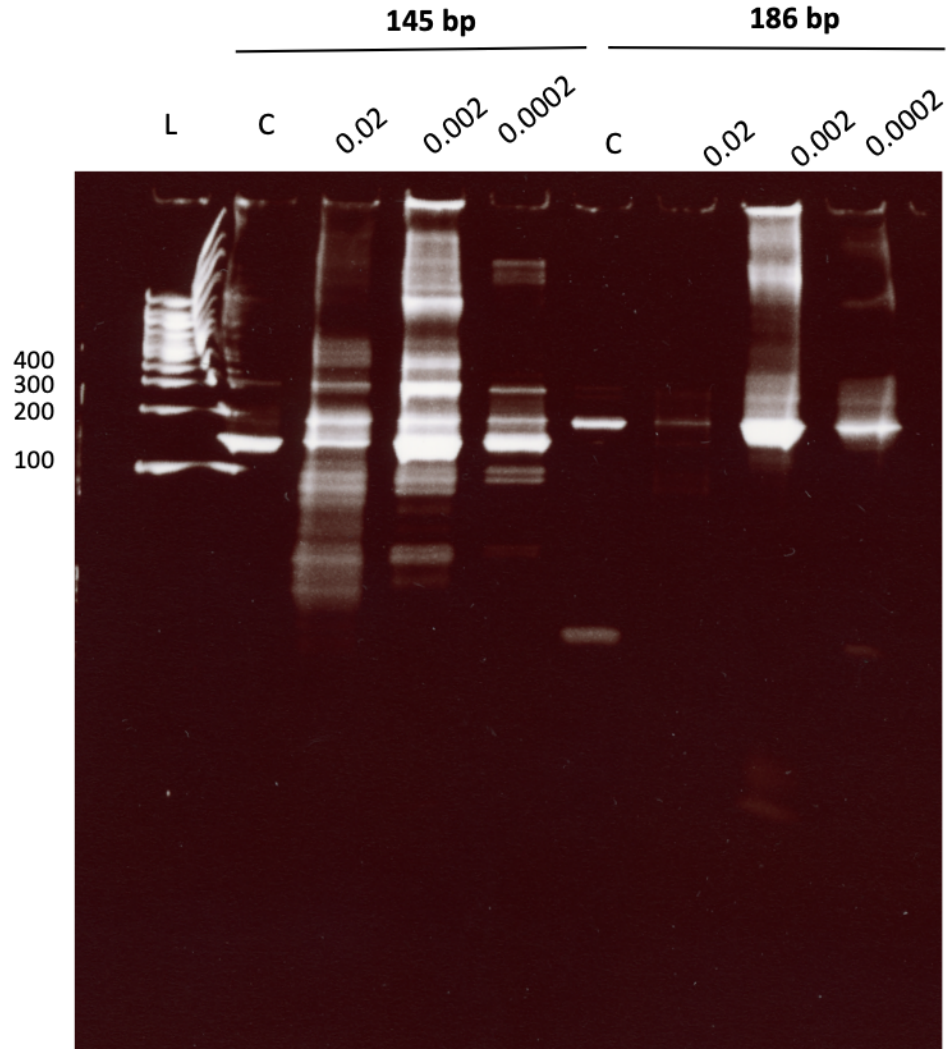


Fig. 3.4.12.1. *Preliminary assay of 186 bp vs 145 bp nucleosomes. Numbers refer to DNAse 1 quantity in KU. (uncropped gel). Reproduced from fig. 3.2.4.1.A.*

Test with three nucleosome variants

145 bp WT, me2 and me3 nucleosomes were combined with either 0.0002 KU DNAse I (Roche) for digest reactions or MilliQ for controls in in DNAse I buffer (10 mM Tris. HCl pH 7.5, 2.5 mM $MgCl_2$, 500 μM $CaCl_2$) to a final volume of 7 μL and incubated at room

temperature for 30 min. 7 µL denaturing buffer (8M urea, 2M NaCl) was then added and samples heated at 95 °C for 10 min to stop the reaction and denature nucleosomes. 3 µL 6X loading dye (NEB) was then added, and samples loaded onto a 15% TBE-Urea gel (Novex) and run un 1x TBE buffer at 180 V, 4 °C. The gel was stained using SYBR gold (6 µL in 20 mL MilliQ for 20 min, destained 2 x 5 min in 20 mL MilliQ) and imaged.

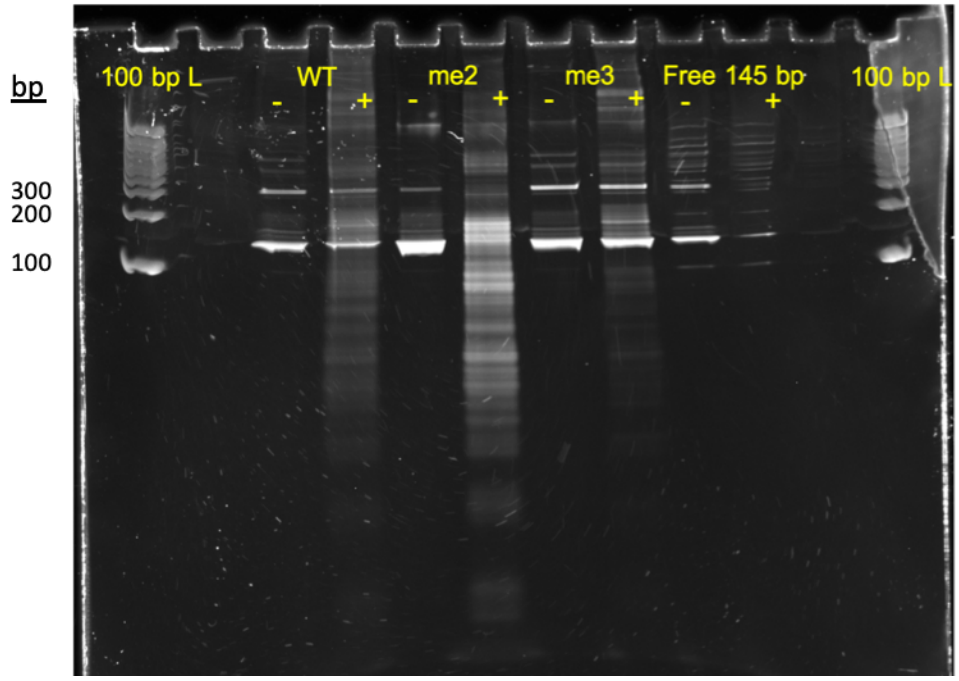


Fig. 3.4.12.2 *Initial DNase I assay with 145 bp DNA, all samples quenched after 30 min. += DNAse I added, -= no DNAse I added (control). (uncropped gel). Reproduced from fig. 3.2.4.1.B.*

Timecourse experiments

145 bp WT, me2 and me3 nucleosomes were combined with 0.0002 KU DNAse I (Roche) for digest reactions in DNAse I buffer (10 mM Tris. HCl pH 7.5, 2.5 mM $MgCl_2$, 500 µM $CaCl_2$) to a final volume of 28 µL and incubated at room temperature. 7 µL aliquots were taken at 2 min, 5 min and 10 min timepoints, combined with 5 µL denaturing buffer (8M urea, 2M NaCl)

and heated for 95 °C for 10 min to stop the reaction and denature nucleosomes. Control samples lacking DNAse I were similarly prepared. 3 μL 6X loading dye (NEB) was then added, and samples loaded onto a 15% TBE-Urea gel (Novex) and run un 1x TBE buffer at 180 V, 4 °C. The gel was stained using SYBR gold (6 μL in 20 mL MilliQ for 20 min, destained 2 x 5 min in 20 mL MilliQ) and imaged.

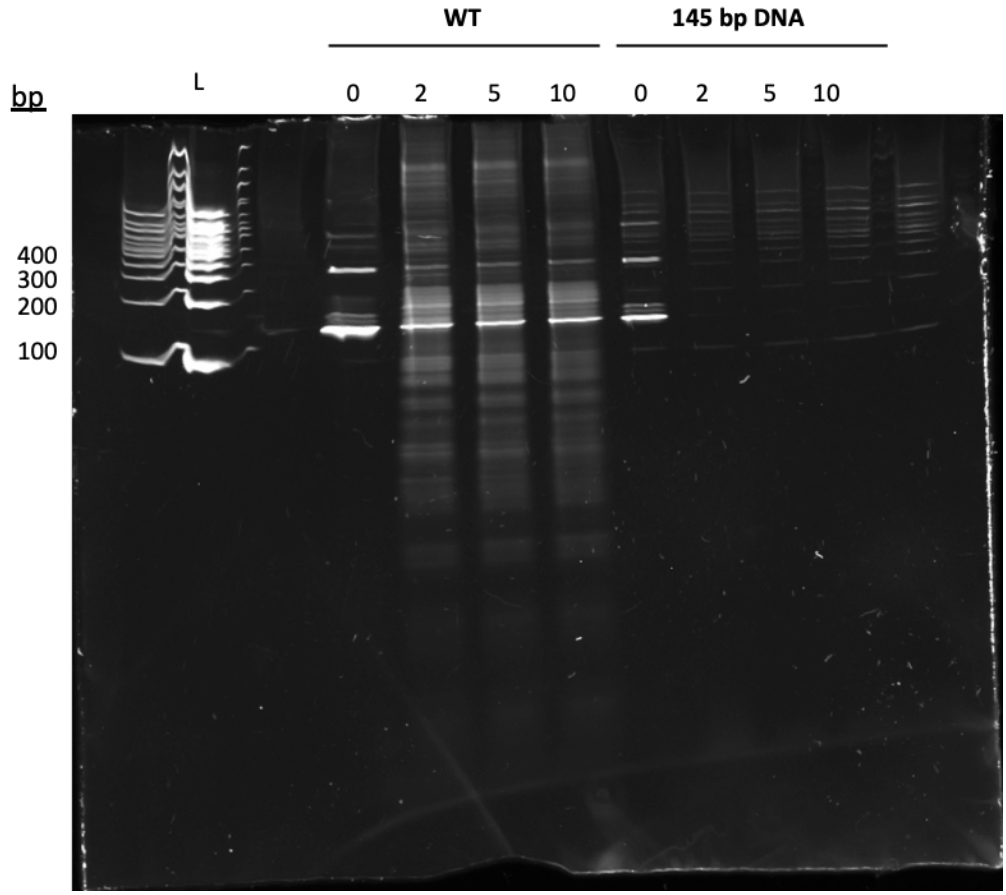


Fig. 3.4.12.3 *Timecourse assay for WT nucleosome and free 145 bp DNA; numbers refer to the number of minutes incubated with DNAse I before quenching (0= control).(uncropped gel). Reproduced from fig. 3.2.4.1.C.*

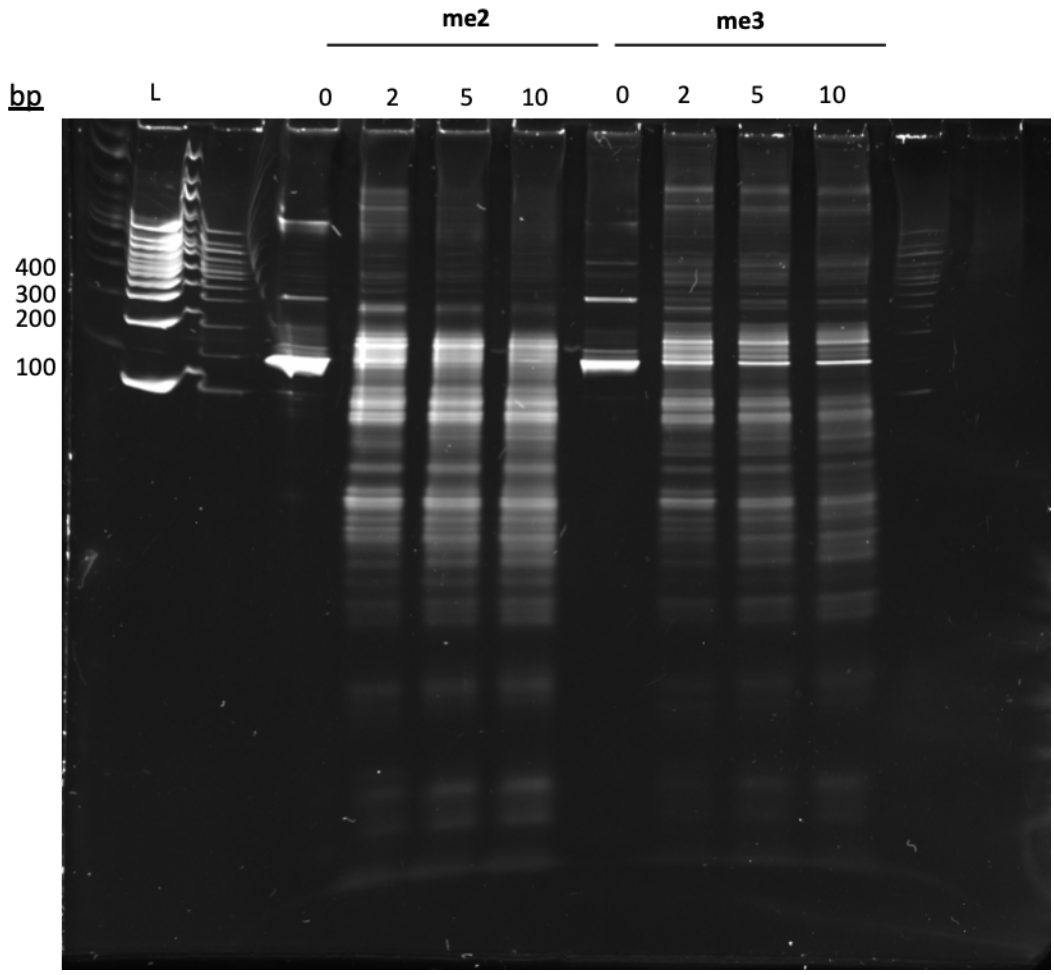


Fig. 3.4.12.4 *Timecourse assay for me2 and me3 nucleosomes; numbers refer to the number of minutes incubated with DNAse I before quenching (0= control).(uncropped gel). Reproduced from fig. 3.2.4.1.D.*

4.5 References

1. Luger, K., Mä Der, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. *Crystal structure of the nucleosome core particle at 2.8 A ˚ resolution*. *Nature © Macmillan Publishers Ltd* **389**, (1997).
2. Maurer-Stroh, S. *et al.* The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem. Sci.* **28**, 69–74 (2003).
3. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket

pickers. *Nat. Struct. &Amp; Mol. Biol.* **14**, 1025 (2007).

4. Yap, K. L. & Zhou, M.-M. Keeping it in the family: diverse histone recognition by conserved structural folds. *Crit. Rev. Biochem. Mol. Biol.* **45**, 488–505 (2010).
5. Paro, R. & Hogness, D. S. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 263–7 (1991).
6. Theobald, D. L., Mitton-Fry, R. M. & Wuttke, D. S. Nucleic Acid Recognition by OB-Fold Proteins. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 115–133 (2003).
7. Arcus, V. OB-fold domains: a snapshot of the evolution of sequence, structure and function. *Curr. Opin. Struct. Biol.* **12**, 794–801 (2002).
8. Cavalli, G. & Paro, R. Chromo-domain proteins: linking chromatin structure to epigenetic regulation. *Curr. Opin. Cell Biol.* **10**, 354–360 (1998).
9. Ball, L. J. *et al.* Structure of the chromatin binding (chromo) domain from mouse modifier protein 1. *EMBO J.* **16**, 2473–2481 (1997).
10. Jones, D. O., Cowell, I. G. & Singh, P. B. Mammalian chromodomain proteins: their role in genome organisation and expression. *BioEssays* **22**, 124–137 (2000).
11. Kusch, T. *et al.* Acetylation by Tip60 Is Required for Selective Histone Variant Exchange at DNA Lesions. *Science (80-.).* **306**, 2084 LP – 2087 (2004).
12. Workman, J. L. & Kingston, R. E. ALTERATION OF NUCLEOSOME STRUCTURE AS A MECHANISM OF TRANSCRIPTIONAL REGULATION. *Annu. Rev. Biochem.* **67**, 545–579 (1998).
13. Joshi, A. A. & Struhl, K. Eaf3 Chromodomain Interaction with Methylated H3-K36 Links Histone Deacetylation to Pol II Elongation. *Mol. Cell* **20**, 971–978 (2005).
14. Carrozza, M. J. *et al.* Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell* **123**, 581–592

(2005).

15. Keogh, M.-C. *et al.* Cotranscriptional Set2 Methylation of Histone H3 Lysine 36 Recruits a Repressive Rpd3 Complex. *Cell* **123**, 593–605 (2005).
16. Lee, N., Erdjument-Bromage, H., Tempst, P., Jones, R. S. & Zhang, Y. The H3K4 demethylase lid associates with and inhibits histone deacetylase Rpd3. *Mol. Cell. Biol.* **29**, 1401–10 (2009).
17. Jelinic, P., Pellegrino, J. & David, G. A novel mammalian complex containing Sin3B mitigates histone acetylation and RNA polymerase II progression within transcribed loci. *Mol. Cell. Biol.* **31**, 54–62 (2011).
18. Jones, D. O., Cowell, I. G. & Singh, P. B. Mammalian chromodomain proteins: their role in genome organisation and expression. *BioEssays* **22**, 124–137 (2000).
19. Brehm, A. *et al.* Retinoblastoma protein recruits histone deacetylase to repress transcription. *Nature* **391**, 597–601 (1998).
20. Akhtar, A., Zink, D. & Becker, P. B. Chromodomains are protein–RNA interaction modules. *Nature* **407**, 405–409 (2000).
21. Pardo, P. S., Leung, J. K., Lucchesi, J. C. & Pereira-Smith, O. M. MRG15, a novel chromodomain protein, is present in two distinct multiprotein complexes involved in transcriptional activation. *J. Biol. Chem.* **277**, 50860–6 (2002).
22. Xie, T. *et al.* Structural Basis for Molecular Interactions Involving MRG Domains: Implications in Chromatin Biology. *Structure* **20**, 151–160 (2012).
23. Bertram, M. J. & Pereira-Smith, O. M. Conservation of the MORF4 related gene family: identification of a new chromo domain subfamily and novel protein motif. *Gene* **266**, 111–21 (2001).
24. Zhang, P. *et al.* Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Res.* **34**, 6621–6628 (2006).

25. Bowman, B. R. *et al.* Multipurpose MRG Domain Involved in Cell Senescence and Proliferation Exhibits Structural Homology to a DNA-Interacting Domain. *Structure* **14**, 151–158 (2006).

26. Bertram, M. J. *et al.* Identification of a Gene That Reverses the Immortal Phenotype of a Subset of Cells and Is a Member of a Novel Family of Transcription Factor-Like Genes. *Mol. Cell. Biol.* **19**, 1479 LP – 1485 (1999).

27. Zhang, P. *et al.* The MRG domain of human MRG15 uses a shallow hydrophobic pocket to interact with the N-terminal region of PAM14. *Protein Sci.* **15**, 2423–2434 (2006).

28. Eisen, A. *et al.* The Yeast NuA4 and Drosophila MSL Complexes Contain Homologous Subunits Important for Transcription Regulation . *J. Biol. Chem.* **276**, 3484–3491 (2001).

29. Fujita, M. *et al.* MRG-1, a mortality factor-related chromodomain protein, is required maternally for primordial germ cells to initiate mitotic proliferation in C. elegans. *Mech. Dev.* **114**, 61–9 (2002).

30. Reid, J. L., Moqtaderi, Z. & Struhl, K. Eaf3 regulates the global pattern of histone acetylation in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **24**, 757–64 (2004).

31. Nakayama, J. -i. *et al.* Alp13, an MRG family protein, is a component of fission yeast Clr6 histone deacetylase required for genomic integrity. *EMBO J.* **22**, 2776–2787 (2003).

32. Doyon, Y., Selleck, W., Lane, W. S., Tan, S. & Côté, J. Structural and Functional Conservation of the NuA4 Histone Acetyltransferase Complex from Yeast to Humans. *Mol. Cell. Biol.* **24**, 1884–1896 (2004).

33. Xu, Y. *et al.* Arabidopsis MRG domain proteins bridge two histone modifications to elevate expression of flowering genes. *Nucleic Acids Res.* **42**, 10960–10974 (2014).

34. Morales, V. *et al.* Functional integration of the histone acetyltransferase MOF into the dosage compensation complex. *EMBO J.* **23**, 2258–2268 (2004).

35. Galarneau, L. *et al.* Multiple links between the NuA4 histone acetyltransferase complex and epigenetic control of transcription. *Mol. Cell* **5**, 927–37 (2000).

36. Ikura, T. *et al.* Involvement of the TIP60 Histone Acetylase Complex in DNA Repair and Apoptosis. *Cell* **102**, 463–473 (2000).

37. Cai, Y. *et al.* Identification of New Subunits of the Multiprotein Mammalian TRRAP/TIP60-containing Histone Acetyltransferase Complex. *J. Biol. Chem.* **278**, 42733–42736 (2003).

38. Huang, C. *et al.* Mrg15 stimulates Ash1 H3K36 methyltransferase activity and facilitates Ash1 Trithorax group protein function in Drosophila. *Nat. Commun.* **8**, 1649 (2017).

39. Tominaga, K. *et al.* MRG15 Regulates Embryonic Development and Cell Proliferation. *Mol. Cell. Biol.* **25**, 2924–2937 (2005).

40. Luco, R. F. *et al.* Regulation of Alternative Splicing by Histone Modifications. *Science (80-.).* **327**, 996 LP – 1000 (2010).

41. Hayakawa, T. *et al.* MRG15 binds directly to PALB2 and stimulates homology-directed repair of chromosomal breaks. *J. Cell Sci.* **123**, 1124 LP – 1130 (2010).

42. Bleuyard, J.-Y. *et al.* MRG15-mediated tethering of PALB2 to unperturbed chromatin protects active genes from genotoxic stress. *Proc. Natl. Acad. Sci.* **114**, 7671–7676 (2017).

43. Hayakawa, T. *et al.* RBP2 is an MRG15 complex component and down-regulates intragenic histone H3 lysine 4 methylation. *Genes to Cells* **0**, 070606122915002-??? (2007).

44. Gautam, D., Johnson, B. A., Mac, M. & Moody, C. A. SETD2-dependent H3K36me3

plays a critical role in epigenetic regulation of the HPV31 life cycle. *PLOS Pathog.* **14**, e1007367 (2018).

45. Graham, S. V. & Faizo, A. A. A. Control of human papillomavirus gene expression by alternative splicing. *Virus Res.* **231**, 83–95 (2017).

46. Xu, C., Cui, G., Botuyan, M. V. & Mer, G. Structural Basis for the Recognition of Methylated Histone H3K36 by the Eaf3 Subunit of Histone Deacetylase Complex Rpd3S. *Structure* **16**, 1740–1750 (2008).

47. Fischle, W. *et al.* Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870–81 (2003).

48. Hughes, R. M., Wiggins, K. R., Khorasanizadeh, S. & Waters, M. L. Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. *Proc. Natl. Acad. Sci.* **104**, 11184 LP – 11188 (2007).

49. Nielsen, P. R. *et al.* Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103–107 (2002).

50. Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–3 (2002).

51. Li, Y. *et al.* The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J. Biol. Chem.* **284**, 34283–95 (2009).

52. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science (80-.).* **354**, (2016).

53. Hazan, N. P. *et al.* Nucleosome Core Particle Disassembly and Assembly Kinetics Studied Using Single-Molecule Fluorescence. *Biophys. J.* **109**, 1676–1685 (2015).

54. Dick, D. A. The distribution of sodium, potassium and chloride in the nucleus and cytoplasm of Bufo bufo oocytes measured by electron microprobe analysis. *J. Physiol.*

284, 37 (1978).

55. Century, T. J., Fenichel, I. R. & Horowitz, S. B. *THE CONCENTRATIONS OF WATER, SODIUM AND POTASSIUM IN THE NUCLEUS AND CYTOPLASM OF AMPHIBIAN OOCYTES. J. Cell Set* **7**, (1970).

56. Allahverdi, A., Chen, Q., Korolev, N. & Nordenskiöld, L. Chromatin compaction under mixed salt conditions: Opposite effects of sodium and potassium ions on nucleosome array folding. *Sci. Rep.* **5**, 8512 (2015).

57. Liu, Y., Wu, H., Yu, Y. & Huang, Y. Structural studies on MRG701 chromodomain reveal a novel dimerization interface of MRG proteins in green plants. *Protein Cell* **7**, 792–803 (2016).

58. Li, J. *et al.* SETD2: An epigenetic modifier with tumor suppressor functionality. *Oncotarget* **7**, (2016).

59. Yoh, S. M., Lucas, J. S. & Jones, K. A. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev.* **22**, 3422–3434 (2008).

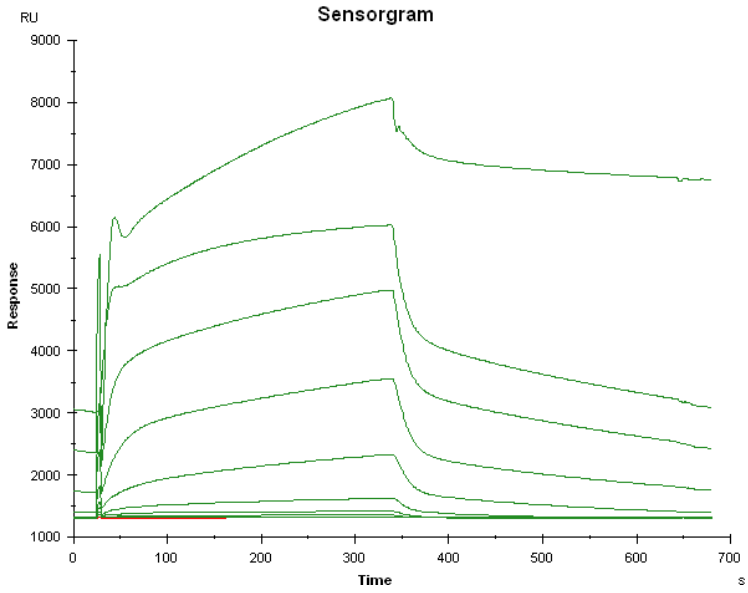
60. Kulaeva, O. I., Hsieh, F.-K., Chang, H.-W., Luse, D. S. & Studitsky, V. M. Mechanism of transcription through a nucleosome by RNA polymerase II. *Biochim. Biophys. Acta* **1829**, 76–83 (2013).

3.6. Appendix

3.6.1. Additional traces for SPR in HBS-EP (3.4.6)

3.6.1.1 Raw SPR data- WT

Run 1



Runs 2 and 3

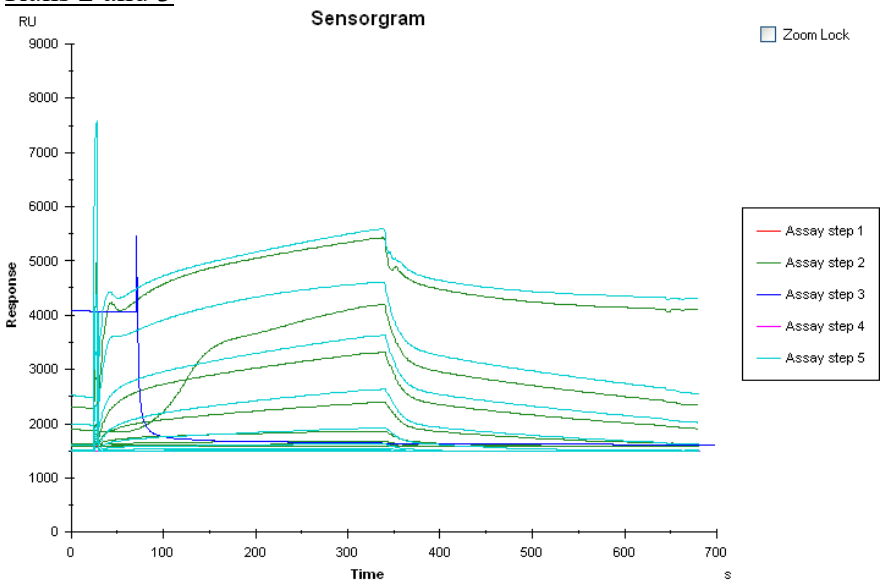
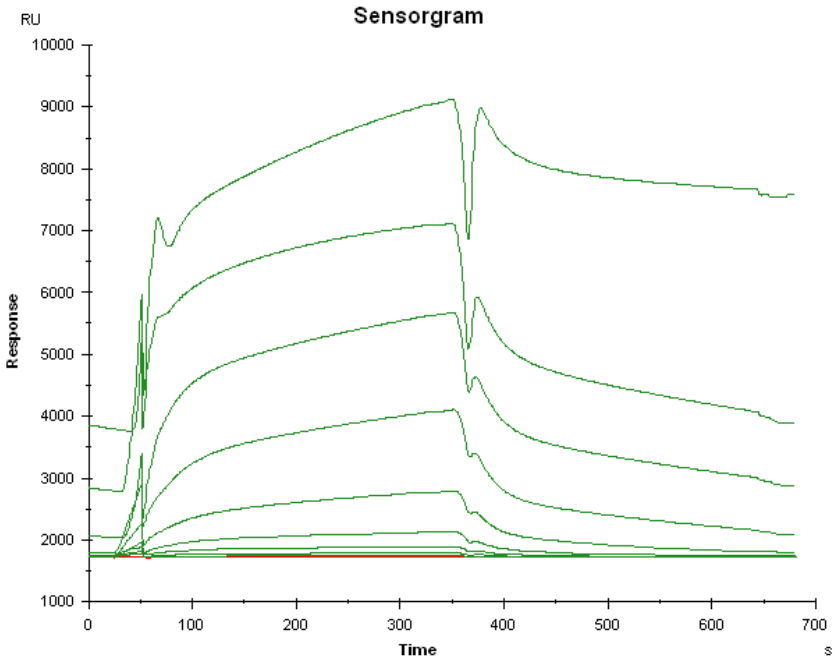


Fig. 3.6.1.1.1. *Raw SPR traces for hMRG15 with WT nucleosome*

3.6.1.2 Raw SPR data- me2

Run 1



Runs 2 and 3

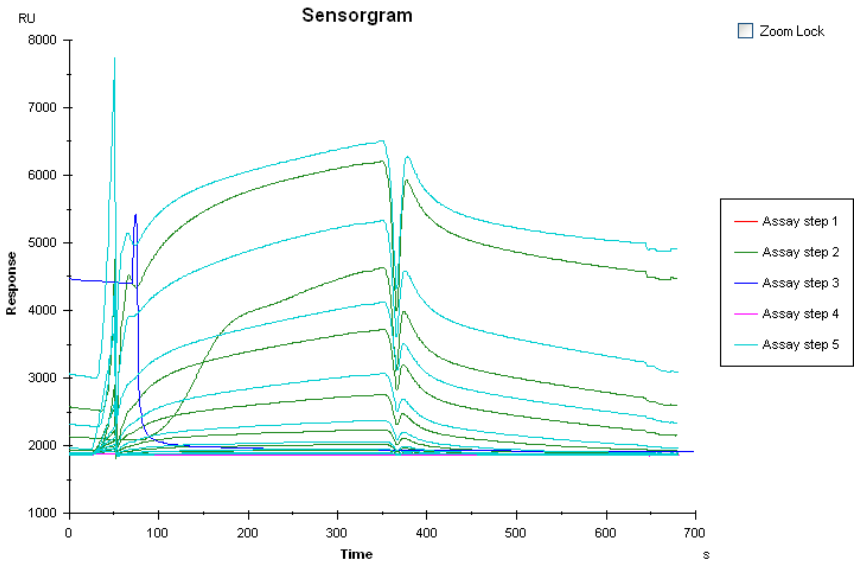
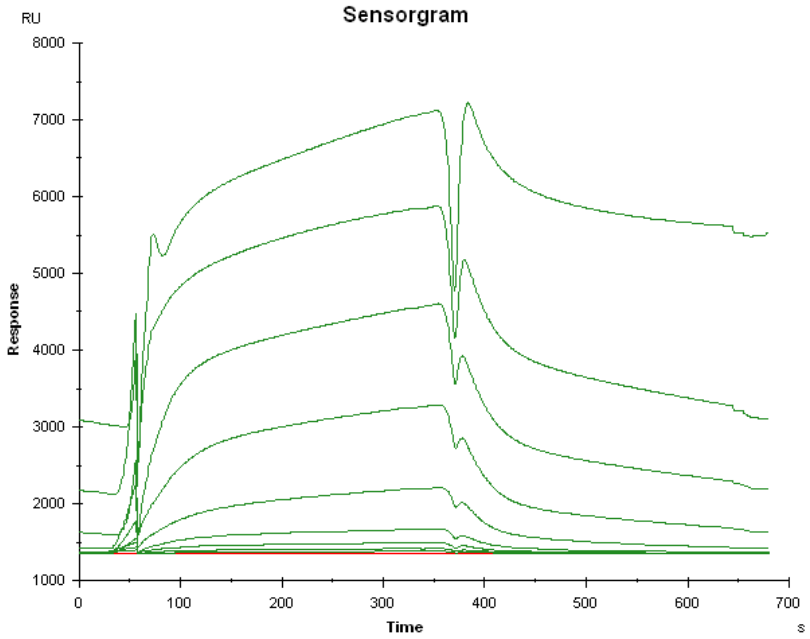


Fig. 3.6.1.2.1. *Raw SPR traces for hMRG15 with me2 nucleosome*

3.6.1.3 Raw SPR data- me3

Run 1



Runs 2 and 3

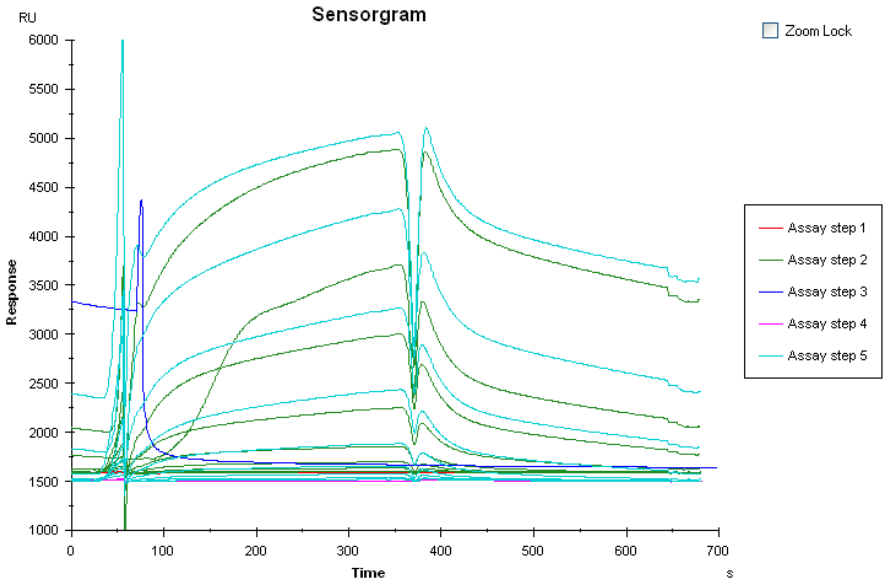


Fig. 3.6.1.3.1. *Raw SPR traces for hMRG15 with me2 nucleosome*

3.6.2. Additional traces for SPR in HBK (round 1)

3.6.2.1. Chip preparation

Streptavidin functionalization

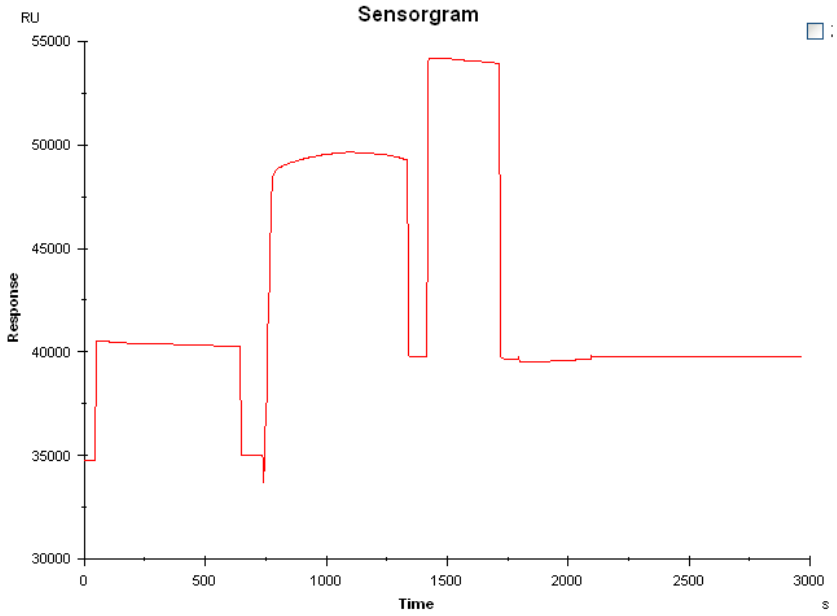
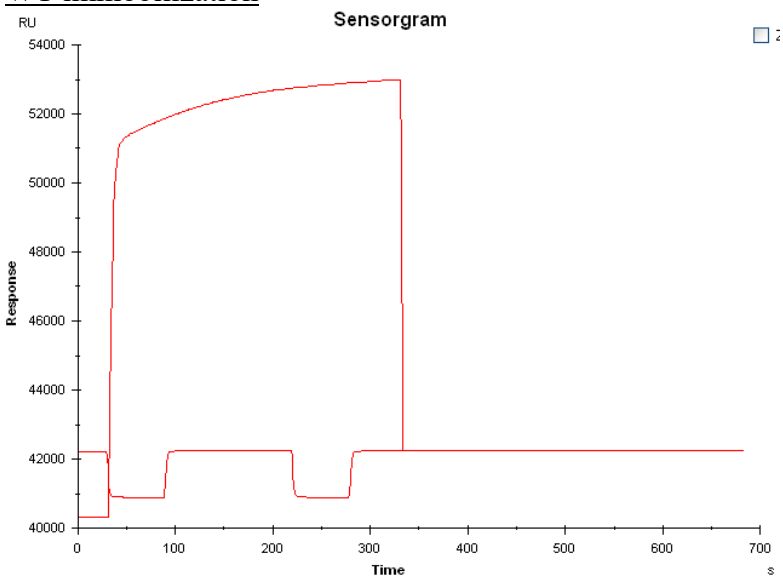
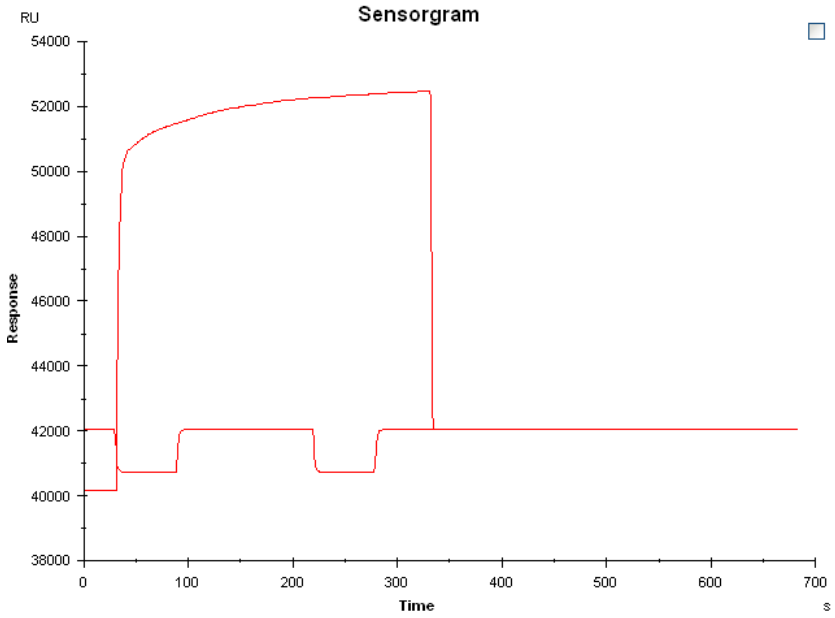


Fig. 3.6.2.1.1. *Raw SPR trace for streptavidin immobilization*

WT immobilization



Me2 immobilization



Me3 immobilization

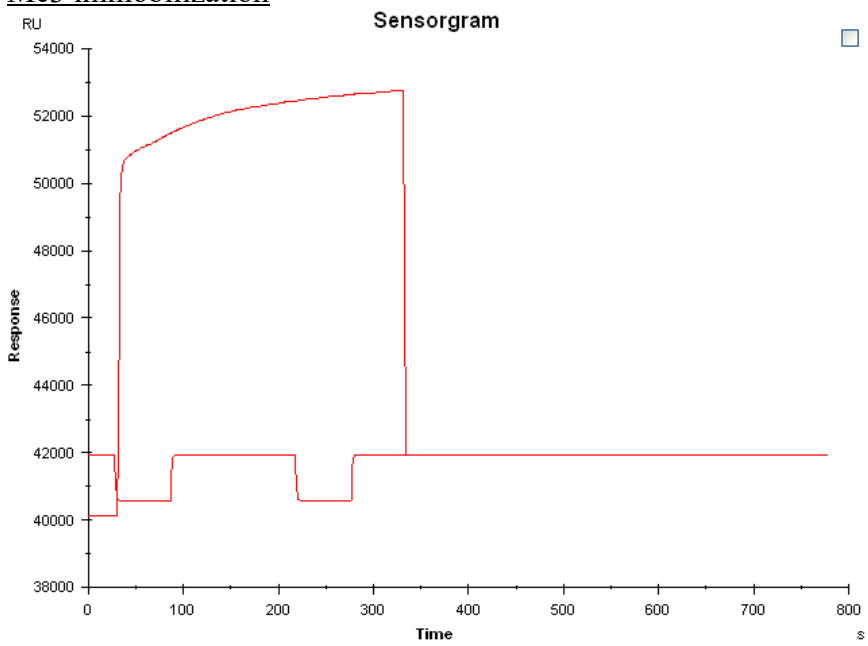


Fig. 3.6.2.1.2. *Raw SPR traces for nucleosome immobilization*

3.6.2.2. Raw SPR data- WT

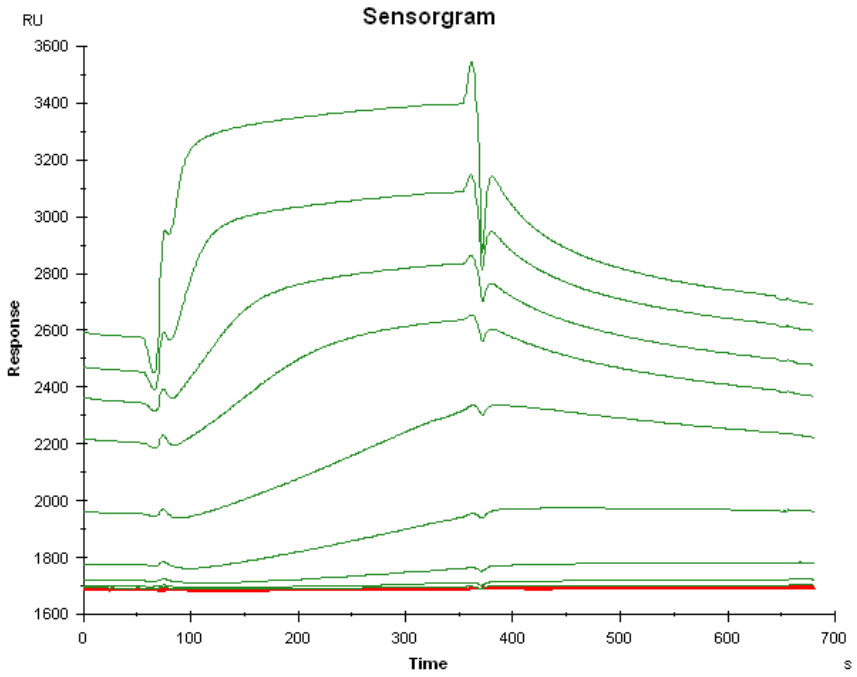


Fig. 3.6.2.2.1. *Raw SPR trace for hMRG15 with WT nucleosome*

3.6.2.3. Raw SPR data- me2

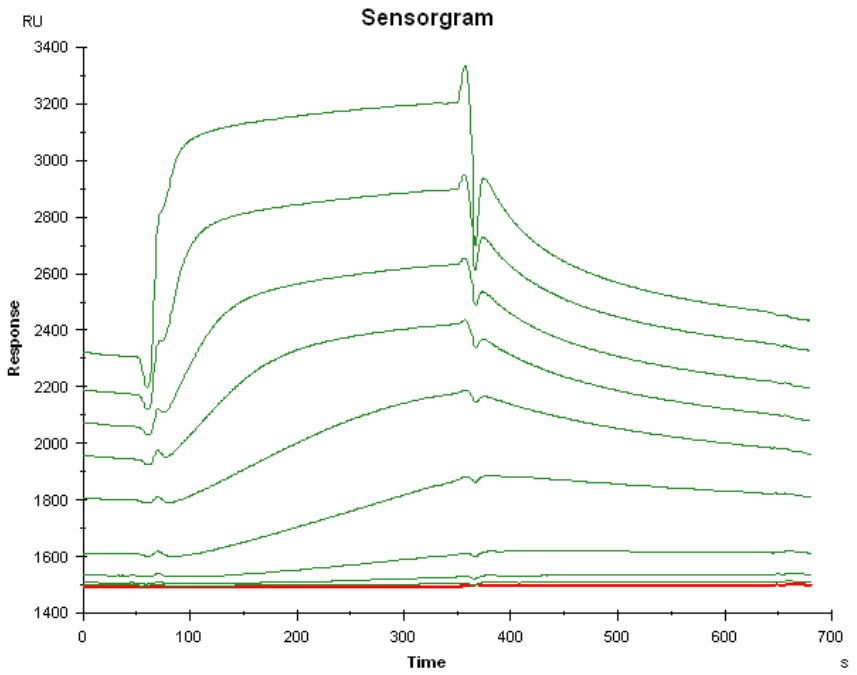


Fig. 3.6.2.3.1. *Raw SPR trace for hMRG15 with me2 nucleosome*

3.6.2.4. Raw SPR data- me3

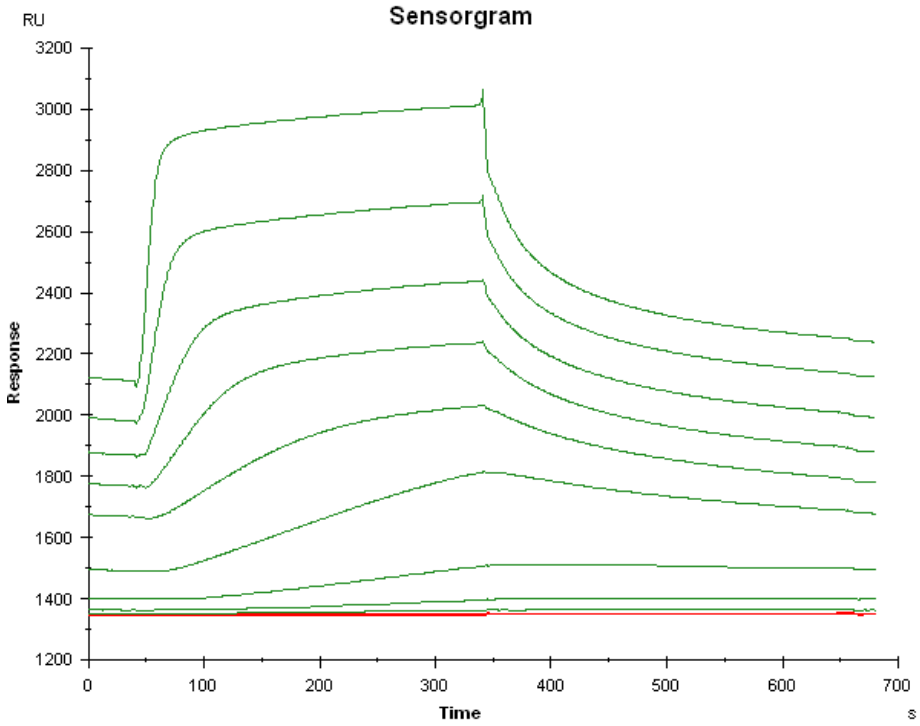


Fig. 3.6.2.4.1. *Raw SPR trace for hMRG15 with me3 nucleosome*

3.6.3. Additional traces for SPR in HBK (round 2)

3.6.3.1. Chip preparation

Streptavidin functionalization

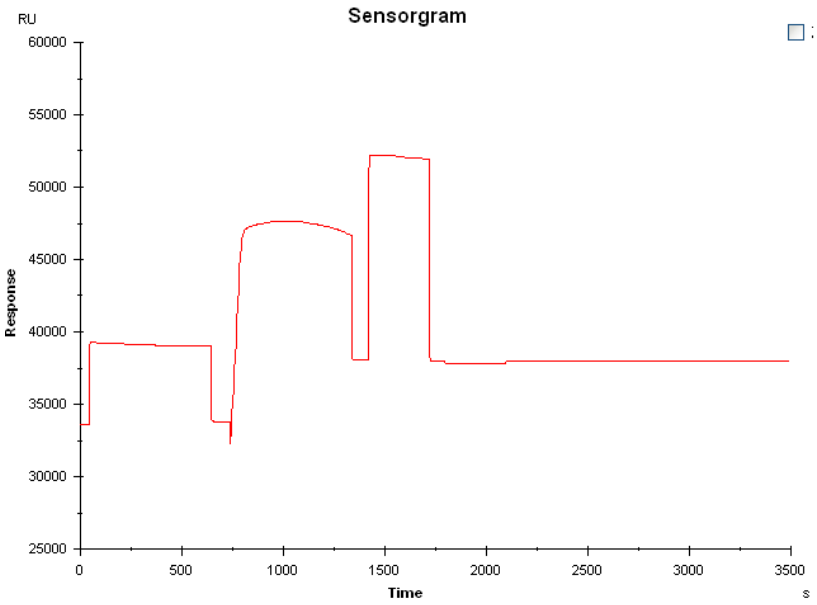
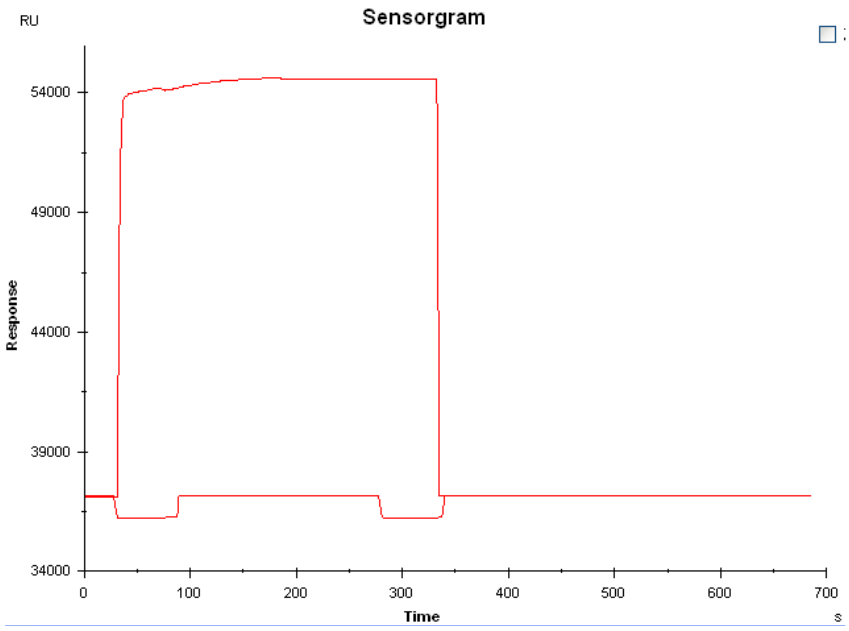
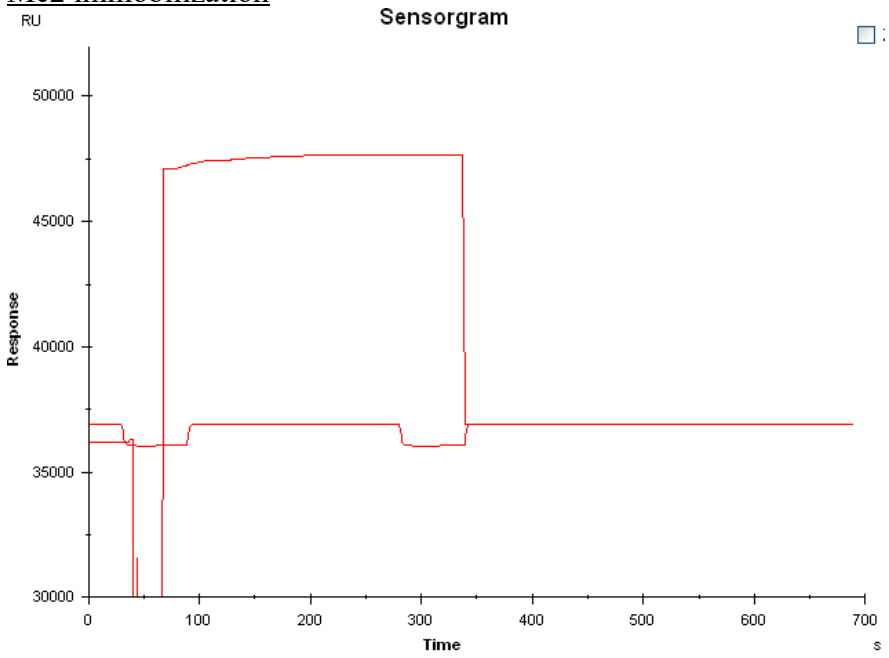


Fig. 3.6.3.1.1. *Raw SPR trace for streptavidin immobilization*

WT immobilization



Me2 immobilization



Me3 immobilization

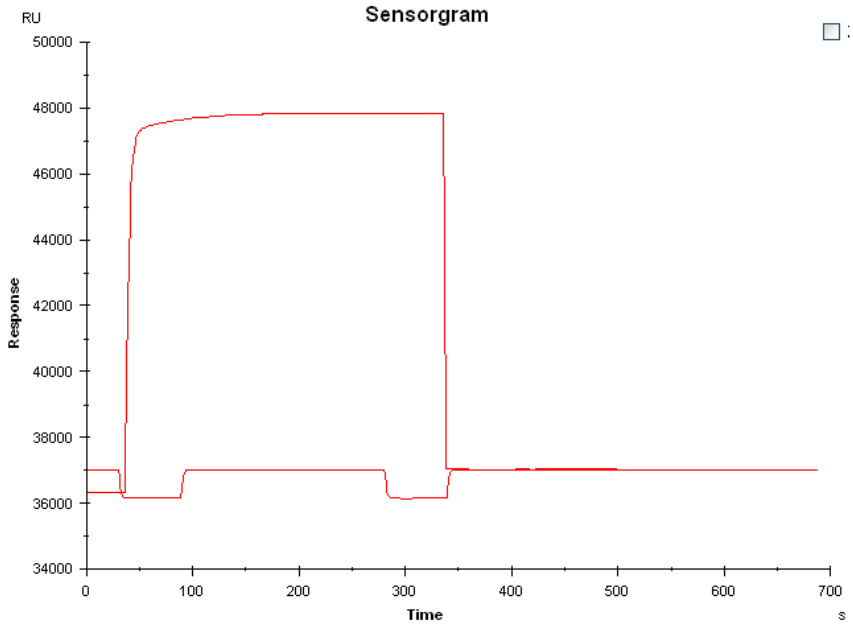
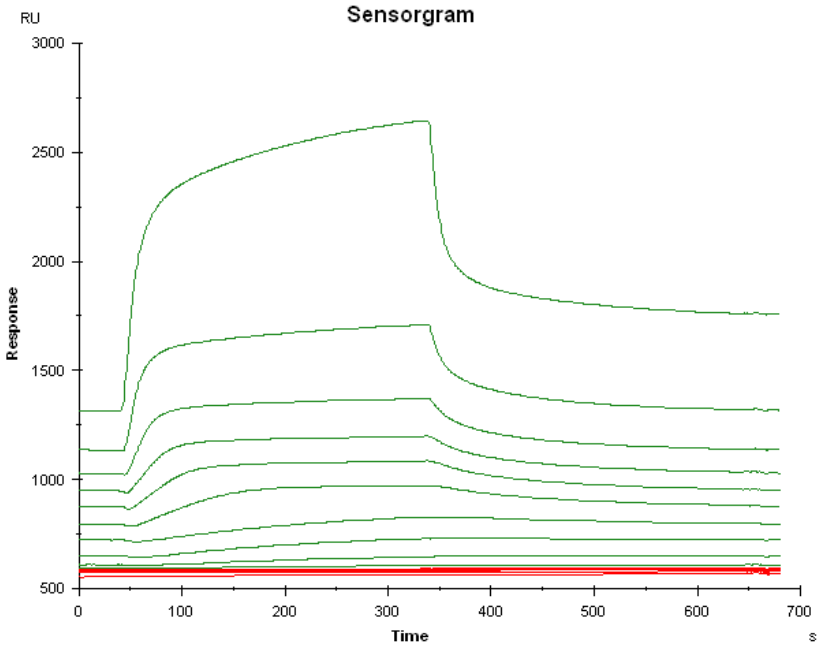


Fig. 3.6.3.1.2. *Raw SPR traces for nucleosome immobilization*

3.6.3.2. Raw SPR data- WT

Run 1



Runs 2 and 3

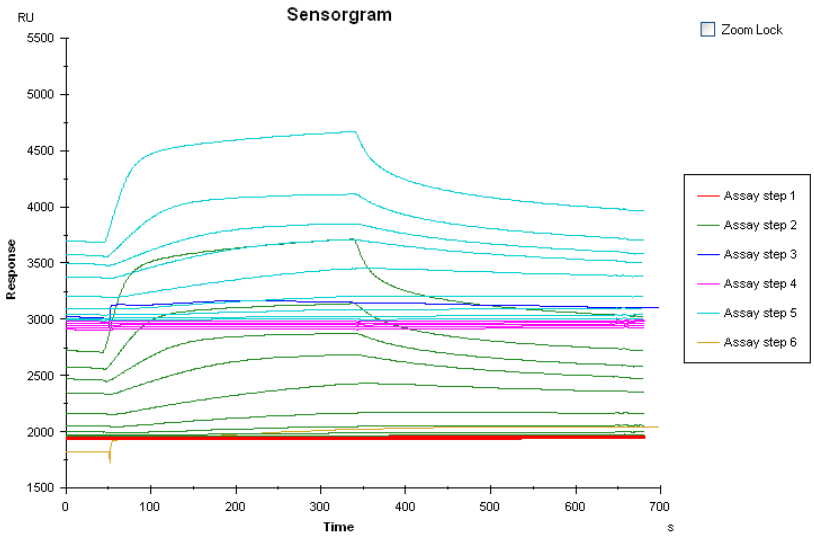
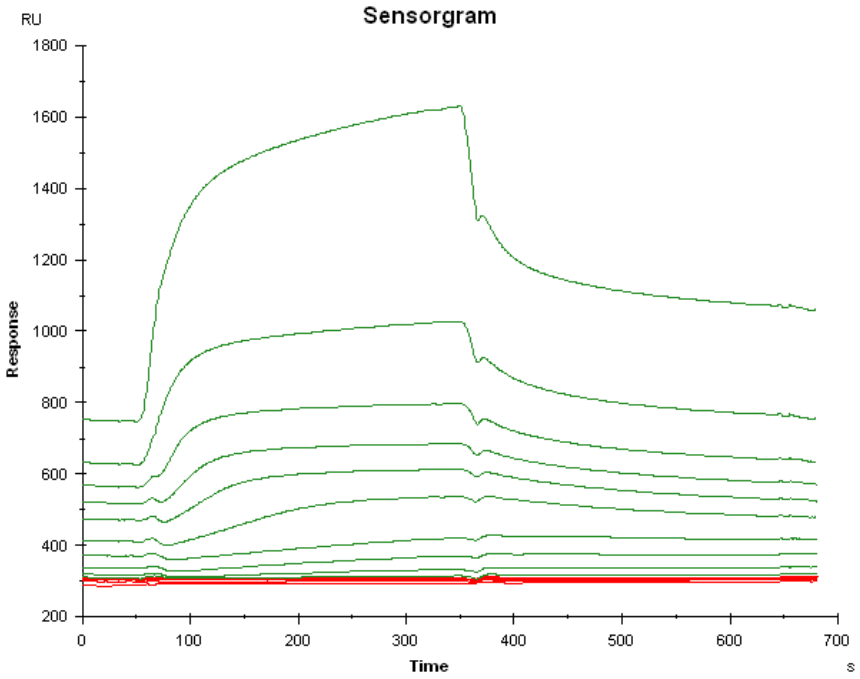


Fig. 3.6.3.2.1. *Raw SPR traces for hMRG15 with WT nucleosome*

3.6.3.3. Raw SPR data- me2

Run 1



Runs 2 and 3

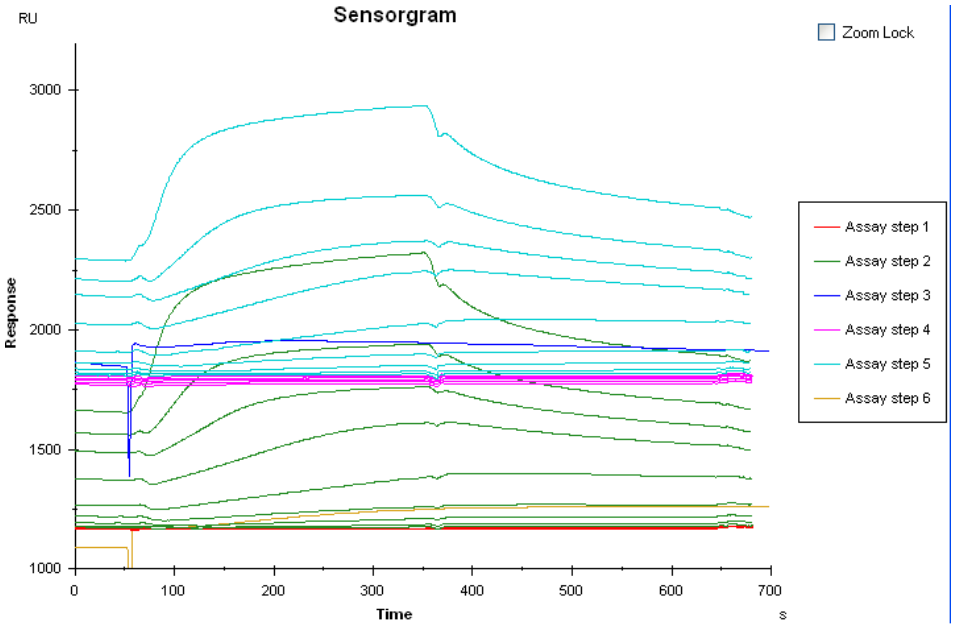
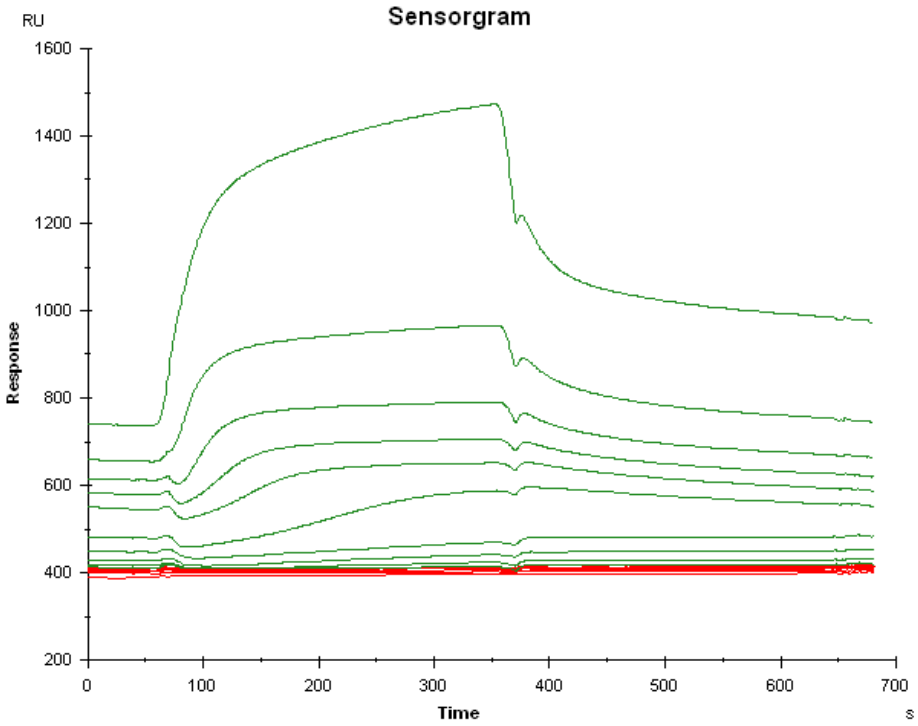


Fig. 3.6.3.3.1. *Raw SPR traces for hMRG15 with me2 nucleosome*

3.6.3.4. Raw SPR data- me3

Run 1



Runs 2 and 3

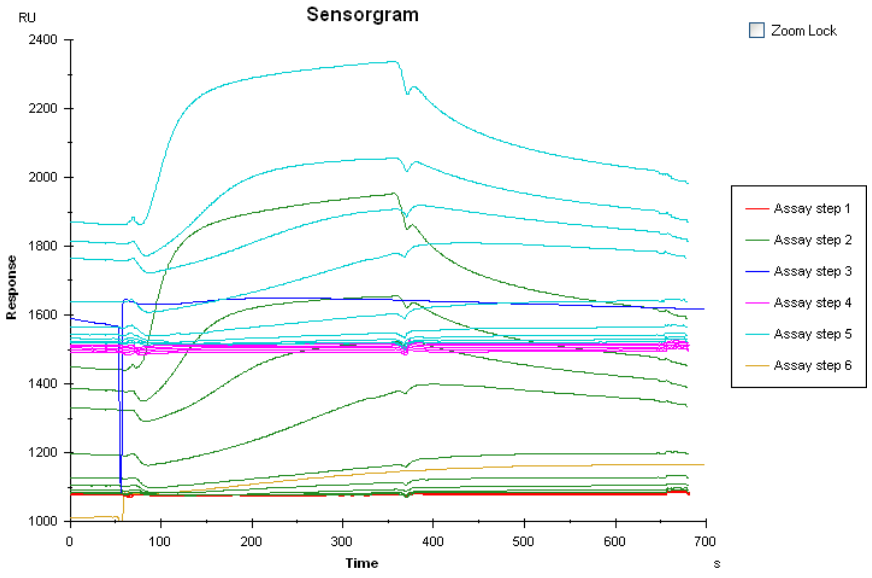
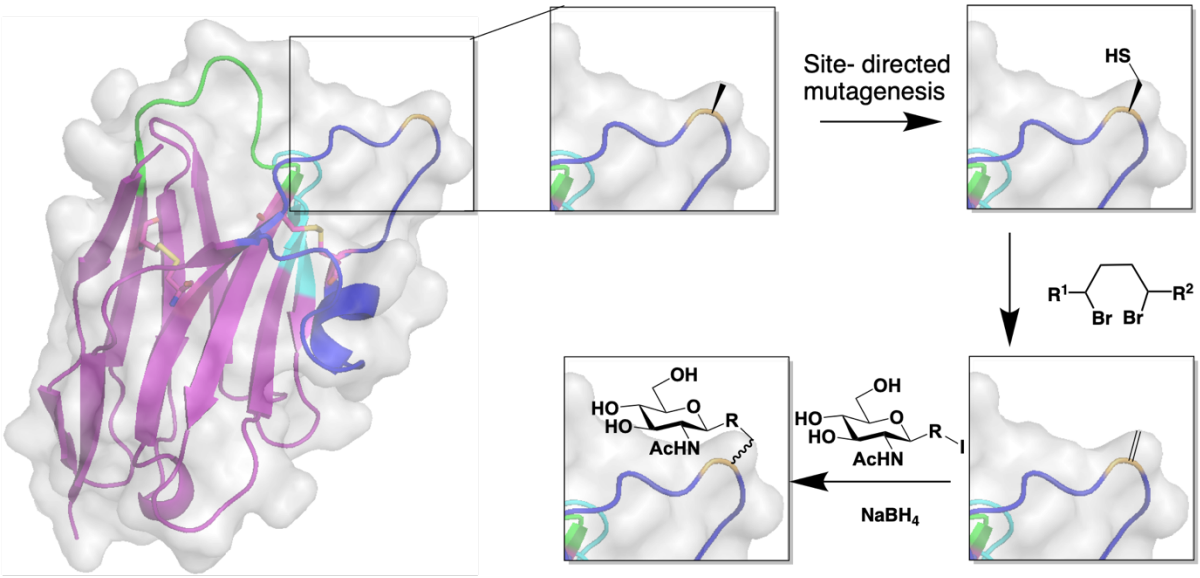


Fig. 3.6.3.4.1. *Raw SPR traces for hMRG15 with me3 nucleosome*

4



Chapter 4: Constructing switchable antibodies through site-selective chemical glycosylation

TABLE OF CONTENTS FOR CHAPTER 4

ABBREVIATIONS FOR CHAPTER 4 282

4.1 INTRODUCTION 284

4.1.1 INTRODUCING UNNATURAL PTMS AS A SWITCH FOR PROTEIN FUNCTION 284

4.1.2 ANTIBODIES AS CANDIDATES FOR SWITCHABLE PTMS 285

4.1.3 NANOBODIES AS A MORE ROBUST AND MANIPULATABLE ALTERNATIVE TO ANTIBODIES 287

4.1.4 *O*-LINKED GLCNAC 291

4.1.5 *N*-LINKED GLCNAC 293

4.1.5 AIMS 293

4.2 RESULTS AND DISCUSSION 295

4.2.1 SUGAR IODIDE SYNTHESIS 295

4.2.2 DHA FORMATION 297

4.2.3 RADICAL REACTIONS; N AND O LINKED SUGARS 298

4.2.4 TESTING SCOPE OF THE RADICAL REACTION WITH OTHER SUBSTRATES 300

4.2.5 TESTING FOR DISRUPTION OF DISULFIDES 301

4.2.6 SCREENING *O*-GLCNACASES FOR CLEAVAGE 303

4.2.7 PURIFICATION OF CABLYS3-BIS-HOMOSER 304

4.2.8 TESTING *N*-LINKED CLEAVAGE WITH PNGASE 305

4.2.9 ELISA TO ASSAY BINDING OF CABLYS3 VARIANTS 307

4.2.10 TESTING ADDITION OF LARGER GLYCANS TO CABLYS3-*O*-GLCNAC 309

4.3 CONCLUSION 311

4.4 EXPERIMENTAL 314

4.4.1.GENERAL EXPERIMENTAL CONSIDERATIONS 314

4.4.2. PREPARATION OF CABLYS3 A104DHA 318

4.4.3 RADICAL REACTIONS WITH CABLYS3-A104DHA 320

4.4.4 Synthesis of cAbLys3 ELISA controls....328

4.4.5 Demonstrating retention of disulphide bonds in cAb-Lys3 during chemical mutagenesis....330

4.4.6 Enzymatic reactions with modified cAbLys3....339

4.4.7. Preparation and purification of cAbLys3-*bis*-homoSer for ELISA....350

4.4.8. Enzyme-Linked Immunoabsorbent Assays (ELISAs)....352

4.4.9. Testing addition of decasaccharide to cAb-Lys3-*O*-GlcNAc using EndoA and EndoS....359

4.4.10. Small molecule synthesis....364

4.5 REFERENCES....378

4.6 APPENDIX....390

4.6.1. Replicate LCMS-ESI+ spectra from the synthesis of cAbLys3-A104-N-GlcNAc....390

4.6.2. Replicate LCMS-ESI+ spectra from enzymatic reactions with modified cAbLys3....392

4.6.3. Additional data from testing addition of decasaccharide to cAb-Lys3-*O*-GlcNAc using EndoA and EndoS....406

4.6.3 Raw NMR spectra....410

ABBREVIATIONS FOR CHAPTER 4

Abbreviation	Meaning
Ala	Alanine
Asn	Asparagine
CDR	Complementarity-determining region
C_H	Heavy chain
C_L	Light chain
COSY	Correlation spectroscopy
CV	Column volume
Cys	Cysteine
DBHDA	2,5-dibromohexane diamide
DEPTQ	Distortionless enhancements by polarisation transfer with retention of quaternaries
Dha	Dehydroalanine
GlcNAc	N-acetyl glucosamine
EndoA	Endoglycosidase A
EndoS	Endoglycosidase S
ELISA	Enzyme-linked immunoabsorbent assay
ESI	Electrospray ionization
Fc	Fragment crystallizable
FcγR	Fcγ receptor.
FPLC	Fast protein liquid chromatography
FR	Framework region
Fv	Fragment variable
GlcNAc	*N*-Acetylglucosamine

Glu	Glutamine
HEL	Hen- egg lysozyme
HSQC	heteronuclear single quantum correlation
HV	Hyper Variable (domain)
IgG	Immunoglobin G
LC-MS	Liquid-chromatography-mass spectrometry
Met	Methionine
N-GlcNAc	*N*-linked *N*-Acetylglucosamine
O-GlcNAc	*O*-linked *N*-Acetylglucosamine
OGA	*O*-GlcNAcase
OGT	*O*-GlcNAcyltransferase
PNGase F	Peptide:N-Glycosidase F
PTM	Post-Translational Modification
scFv	Single chain fragment variable
SDS-PAGE	Sodium dodecyl sulfate Polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
Ser	Serine
TLC	Thin-layer chromatography
TOF	Time of Flight
UDP	Uridine diphosphate
UV	Ultraviolet
V_H	Variable heavy
V_L	Variable light
WT	Wild-type

4.1 Introduction

4.1.1 Introducing unnatural PTMs as a switch for protein function

In the case study described in chapters two and three, we used the radical Dha chemistry to append PTMs which resemble natural states of a protein, mimicking methylation of H3K36 which is generated by a natural enzyme within the nucleus.[1–3] Consequently, these modifications interacted with the natural binding partner MRG15, and allowed us to make inferences about the mechanism of interaction *in vivo*.[4]

However, one significant advantage of the radical-Dha method for introducing PTMs is that, in principle, any modification can be introduced at any position.[5,6] We are not limited to mimicking nature, but can instead modify proteins in entirely novel ways to produce a new function.

As described in chapter 1, PTMs are used by nature on many proteins to switch their functions on or off in response to stimuli; prime examples being the labile modifications phosphorylation and GlcNAcylation.[7–9] We envisaged co-opting this to tune the function of a protein in response to certain signals. Such tuning would be particularly promising for reducing the off-target effects of pharmaceutical proteins, as they could be conditionally activated only in the presence of disease biomarkers. As off-target effects can represent a significant challenge to engineering a clinically useful treatment, this concept may have practical applications in future.[10–12]

4.1.2 Antibodies as candidates for switchable PTMs

One area in which this has enormous potential for medical applications is in antibody therapies for diseases where the antigens targeted are also present in other places in the body.[13] The best known example of this is cancer immunotherapy.[14–16] On a simplistic level, tumours are comprised of aberrant forms of the body's own cells; in practice, this means that many of the antigens targeted are over-expressed in tumours, rather than ones exclusive to them.[17] Key challenges therefore, are identifying appropriate antigens and monitoring the biodistribution of immunotherapeutic antibodies, which can have significant off-target effects. An example of significant off-target binding affecting the viability of antibodies treatments can be found in the mutant EGFR specific antibodies 528 and 225, murine precursors to cetuximab from early trials, which were radio-traced to the liver 'antigen sink' where wild-type EGFR was highly expressed.[18]

Antibodies are part of the immunoglobin superfamily, and are produced as part of the adaptive immune response against specific antigens.[19] Antibodies exist in several different isotypes. IgG, the most common in humans, are comprised of two heavy chains and two light chains linked by disulfide bonds; other isotypes can vary greatly in their composition, but all contain the basic structural motif shown in fig. 4.1.2.1.[20] Early structural studies by papain and pepsin proteolytic digests revealed that these are comprised of two antigen binding (Fab) arms and the 'crystallizable fragment' (Fc) region which mediates activation of other components of the immune system.[21–23] This activation is achieved through Asn-linked glycans. These are known to mediate complement activation via the lectin pathway, and are also expected to influence the classical complement pathway. The glycans can also bind directly to Fc receptors such as FcγIIIA, which triggers antibody-dependent cellular cytotoxicity (ADCC).[24–26] The extent and

type of glycosylation varies not only between isotypes but within them; moreover, the glycosylation profile also varies greatly with expression system due to differences in glycosyltransferase and transporter levels between cell lines. As such, isolating truly homogeneous antibodies is a challenge.[27]

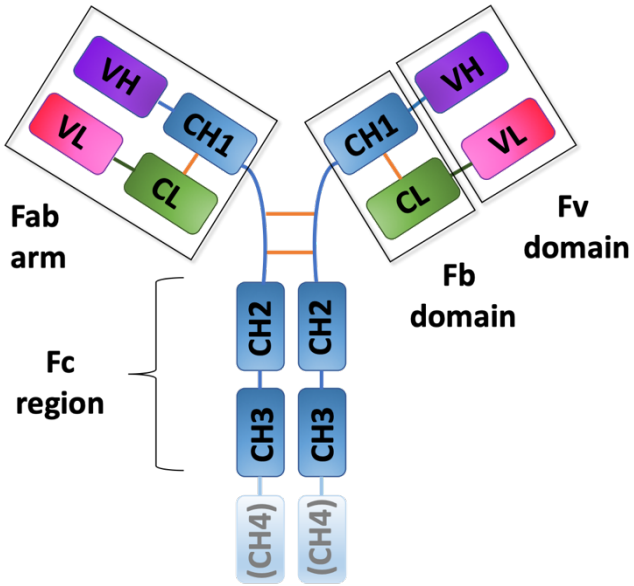


Fig. 4.1.2.1. *Diagram of IgG1, also representing the core motif found in other isotypes. Orange lines represent disulfide linkages. The CH4 domain is present in IgM and IgE isotypes only, and absent in IgG, IgA and IgD.*[28]

The Fab region, which recognises antigens, is comprised of fragment variable (Fv) domains on the light (V_L) and heavy (V_H) domains, as well as the constant fragment domains (Fb) of the light (C_L) and heavy (C_{H1}) chains.[20] The variable domains consist of three hyper-variable (HV) loop regions where amino acid sequence is shuffled during antibody generation to produce a library in which one might bind to the antigen, surrounded by more conserved framework (FR) β-sheet regions.[29,30]

Full, native antibodies are large (ca. 150 kDa), which has significant disadvantages; chiefly, they cannot penetrate some regions of the body or bind small clefts in antigens, and they are very long-lived, which may be undesirable if the antibody has significant off-target effects or toxicity.[31] In 1985, the technique of phage display was discovered, which enabled engineering of Fv fragments alone to reproduce characteristics of the parent antibody.[32,33] This opened up the possibility of using smaller fragments in place of whole antibodies, and was further enabled by the discovery of nanobodies.

4.1.3 Nanobodies as a more robust and manipulatable alternative to antibodies

The first antibody fragments generated were single chain Fvs (scFvs), consisted of the V_H and V_L fragments with an engineered linker joining them.[34] While these have found success in some clinical applications, they suffer the disadvantage of misfolding in some variants where hydrophobic regions can lead to aggregation.[35,36]

Antibodies lacking the light chain and comprising only heavy chains have since been discovered; first in camelids, then in a variety of shark species, with surprising convergent evolution between them.[37–40] As the Fv region naturally lacked the V_L component, a smaller and more robust V_H component (V_HH) could be generated by immunising an appropriate camelid, cloning these fragments produced from circulating B cells, and selection by phage display to produce the V_HH nanobody.[41,42] Since then, humanised versions have been generated for clinical applications. Similarly, variable new antigen receptor (V_{NAR}) fragment nanobodies have been produced from sharks.[43]

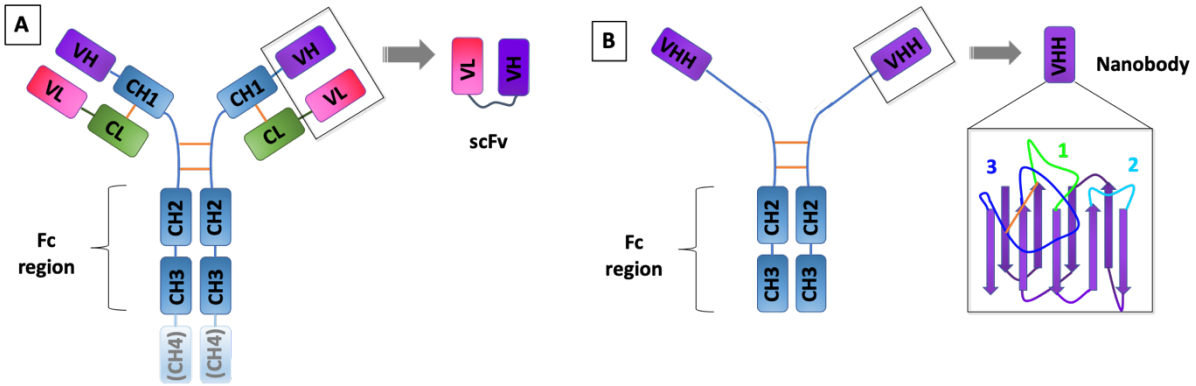


Fig. 4.1.3.1 (A) *Structure of an antibody and scFv derived from the VL and VH domains with an engineered linker.* ***(B)*** *Structure of a camelid antibody consisting of heavy chains only, and the nanobody derived from its V_HH domain. The three CDR loops are also illustrated.*[44]

Several key structural variations were found in these nanobodies. Significantly, in order to achieve sufficient affinity for the antigen, the CDR loops need a combined surface area of 600–800 Å^2.[45] In normal antibodies, there are six CDR loops in total from the VH and VL regions, whereas nanobodies, derived from the heavy chain only, possess only three. In order to compensate for this reduced number of CDR loops, each loop is therefore longer; approx. 20 amino acids compared to 8-10 in a full antibody.[46]

This, however, engenders another issue in that larger loops have a greater entropic penalty for antigen binding.[47] In order to overcome this, disulfide bonds pre-organising the loops have been found. In camelid- and shark-derived nanobodies, this is typically found between the H1 and H3 loops.[48–52] In variants possessing an extended H3 loop, a second disulfide bridge between the H2 and H3 loops is also present.[53,54]

Nanobodies as alternatives to traditional full antibodies have gained traction in the pharmaceutical industry, and in 2019 caplacizumab became the first nanobody to be approved

by the FDA.[55] Although traditional full-length antibodies typically have higher potency, nanobody-based therapies have found success in niches where a smaller, modular and highly tunable solution is required.[56]

One significant advantage nanobodies possess over their full length counterparts is their smaller size, which allows entry into more areas of the body, such as the receptor-mediated crossing of the blood-brain barrier developed by Ossianix.[57] Moreover, as nanobodies can target smaller antigen crevices than full antibodies, they have also found use in binding proteins such as GPCRs, either as drugs or to stabilise these conformationally flexible molecules for structural studies and further drug discovery.[58] Their smaller size also makes engineering across the entire antibody sequence feasible, compared to full-length antibodies where the structure is often unknown and in parts heterogeneous. This has allowed, for example, engineering of gut-protease resistant nanobodies for oral delivery.[59]

Perhaps the most powerful aspect of nanobodies is their ability to act as building blocks for larger, multifunctional constructs. The utility of this modularity was highlighted in the development of a universal flu vaccine by conjugating four different VHH nanobodies, each capable of binding a different epitope of influenza A and B.[60]

Nanobodies have some limitations, but their amenability to engineering makes it possible to overcome these. Being comprised of a Fv fragment only, nanobodies lack the ability to recruit other components of the immune system and longer serum half-life which come from the Fc domain.[61,62] This is particularly significant for applications such as immunotherapy, where antibody binding alone is often insufficient to produce the desired therapeutic effect and generation of a full immune response is required.[15,63–65] However, this has in part been overcome by conjugating nanobodies to an Fc fragment.[66]

Moreover, although they are small, nanobodies do not possess additional mechanisms for being taken up into cells beyond the receptor-mediated endocytosis common to full-length antibodies.[67] Cell-penetrating nanobodies have thus been synthetically generated, and work to improve on this concept is ongoing.[68–70]

In 2014, the Davis group produced a synthetic conditionally active nanobody.[71] In the work pioneered by S. Gunnoo, the α-lysozyme camelid VHH, cAbLys3, was modified with a dehydroalanine at position 104 within the CDR3 loop. The Dha served as a bio-orthogonal electrophile for reaction with the PTM mimic thiophosphate to block its binding ability. In the presence of a phosphatase, the thiophosphate could be removed to reactivate the nanobody. This formed an AND gate, where both lysozyme and phosphatase were necessary to activate binding.

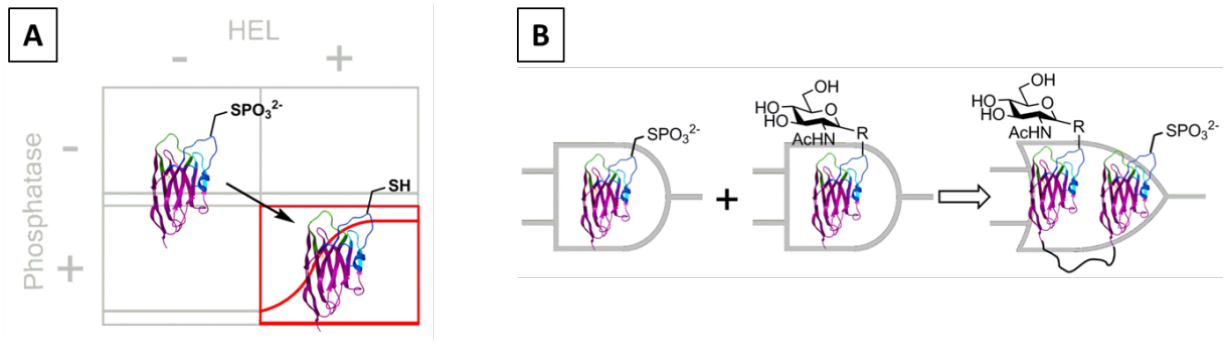


Fig. 4.1.3.2 (A) *The thiophosphate AND gate. The curves shown represent stylised ELISA binding data in the presence or absence of phosphatase and antigen HEL.* ***(B)*** *Fusion of two cAbLys3 AND gates with a linker to create an OR gate.* ***R***= N- *or* O- *linker to mimic linkages to Asn or Ser.*

In order to build a library of operators to allow more complex "computing", it was necessary to develop a second, orthogonal activity switch system. Fusing two units of cAb-Lys3, each with a different modification to block binding, would generate an OR antibody which would work when an enzyme to remove either of these switches was present. Glycosylation was identified as a potential candidate for a second, orthogonal modification; nearly half of all human proteins are glycoproteins, and O-GlcNAcylation in particular has been identified as a PTM which works in a similar but mutually exclusive fashion to phosphorylation.[72]
Generation of such a GlcNAcylated cAbLys3 variant was trialled by use of a thio-sugar in a thiolene coupling reaction to Dha.[73] Unlike the thiophosphate, however, the resulting thiosugar could not be removed by O-GlcNAcase enzymes to restore activity.

Subsequently, a radical reaction has been developed within the Davis group which allows the formation of C-C bonds.[74] This removes the necessity of using a thiosugar and presented the possibility of generating O- and N-linkages more representative of those found *in-vivo*. Both of these linkages have different biological purposes, and both were investigated in this work.

4.1.4 *O*-linked GlcNAc

In humans, 2-5% of the glucose consumed each day enters the hexosamine biosynthetic pathway to generate UDP- β-N-Acetylglucosamine (UDP-GlcNAc) for the formation of *O*-GlcNAcylated serine or threonine residues, or N-GlcNAcylated asparagine.[75] *O*-GlcNac is unusual in that, unlike other biological glycopeptides, it is not elongated to form more complex structures but instead acts as a labile PTM which can be rapidly attached and removed multiple times within the lifetime of a single protein. [76,77] Moreover, while other glycoproteins are typically synthesised on the endoplasmic reticulum, *O*-GlcNacylation occurs in the nucleus as

well as the cytoplasm, and has been found as a PTM on histones.[78,79] In these respects GlcNAcylation resembles phosphorylation, and indeed some interplay has been found between the two, as they can occur on the same sites but are mutually exclusive.[80] *O*-GlcNAcylation is highly responsive to nutrient levels, as its synthesis depends on the metabolism of glucose, amino acids, fatty acids and nucleotides, and in general it is known as a sensor for metabolic state and stress.[81,82]

Unlike the other histone PTMs mentioned in chapters 2 and 3, a single pair of enzymes is responsible for adding and removing *O*-GlcNAc from every protein substrate; *O*-GlcNAc transferase (OGT) for its addition, and *O*-GlcNAcase (OGA) for its removal.[77] Because of this, OGA has been proposed to recognise flexible loops in general, which can adopt a conformation exposing the amide backbone, and is hence independent of the peptide sequence.[83] While a full crystal structure has not been resolved, partial structures available support this hypothesis.[84–86] Moreover, the hOGA homolog NagJ has been crystallised and shown to fit this binding model, further confirmed by molecular dynamics simulations.[87,88] This promiscuity supports the hypothesis that an unnatural *O*-GlcNAc mimic on cAbLys3 could be recognised and removed.

Aberrant *O*-GlcNAcylation is associated with a range of diseases caused by the breakdown of the body's own molecular processes, such as metabolic disorders.[89–91] *O*-GlcNAcylation has also been linked to Alzheimer's disease; the protein tau which forms the aggregates characteristic of the disease is normally O-GlcNAcylated, preventing aggregation.[92,93] When these O-GlcNAc groups are replaced by phosphates, which may occur with overactive OGA activity, aggregation occurs.[94] Moreover, O-GlcNAcylation in the hypoxic tumour environment has been implicated in the metabolic reprogramming found in cancer, and OGA is known to be up-regulated in cancerous cells.[95,96]

4.1.5 *N*-linked GlcNAc

Mammalian cells contain at least 7000 different glycans, built from 10 monosaccharide building blocks.[97–99] In animal cells, the base of this glycan is almost always a GlcNAc linked to asparagine, the first sugar in the common core $Man_3GlcNAc_2$ pentasaccharide. This forms a biantennary glycan which is further extended into one of three general glycan classes which are found in many different functions across many different species; high mannose, complex, or mixed.[100,101] As synthesis of each glycan is determined by the balance of activities of 200 different glycosyltransferases, microheterogeneities are ubiquitous across glycosylation sites.[102,103]

Glycans are found in myriad different locations, with multiple roles which have broadly been divided into four categories; structural and modulatory, intrinsic recognition within a species, extrinsic recognition between species, and in the case of pathogens, molecular mimicry of their host glycans.[104] As such, many specific enzymes for N-glycan trimming exist for the different structures. A ubiquitous one is PNGase F, which is secreted by some bacterial species and capable of processing a range of sugars. [105,106] Glycans can also be delivered into the cell by endocytosis, and degraded within the lysozyme.[107]

4.1.5 Aims

The aim of this work was to test the suitability of *O*- and *N*-linked GlcNAc as a second switchable modification for building a library of 'logic gated' PTM mimics for reversibly blocking the CDR of cAbLys3. To achieve this, it was first necessary to establish the synthesis

of cAbLys3A104-*O*-GlcNAc and cAbLys3A104-*N*-GlcNAc; the 'tag and modify' strategy was once again adopted. As cAbLys3A104C had been generated previously, this was used as the starting material; it was converted to Dha and then GlcNAc added via the radical post-translational mutagenesis reaction.

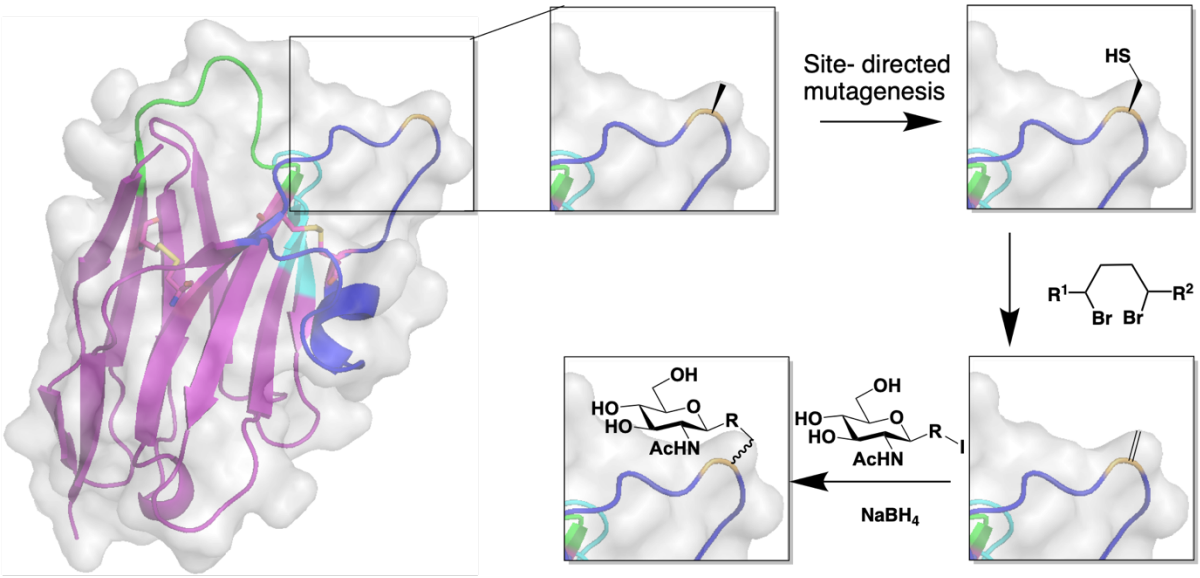


Fig. 4.1.5.1 *Strategy for installation of* N- *and* O- *linked GlcNAc mimics at position 104 in cAbLys3. The mutation A104C is first made using site-directed mutagenesis. Cys is then converted to Dha using an appropriate alkylating agent:* $\boldsymbol{R^1}$**=** $\boldsymbol{R^2}$= $CONH_2$ *for DBHDA. A radical reaction with* $NaBH_4$ *and an appropriate iodide installs the sugar.* ***R***= N- *or* O- *linker to mimic linkages to Asn or Ser.*

For the PTM mimic to be useful as a switch for cAbLys3's binding, it was necessary to determine whether it could be removed by appropriate glycosidases; to this end, enzymatic cleavage of cAbLys3A104-*O*-GlcNAc and cAbLys3A104-*N*-GlcNAc was tested by an OGA and PNGase respectively.

Finally, it was necessary to determine whether cAbLys3 showed differential binding to HEL before and after removal of the PTM; this was evaluated by ELISA.

4.2 Results and Discussion

4.2.1 Sugar iodide synthesis

The synthesis of β-D-1-(2-iodoethoxy)-GlcNAc had previously been established within the Davis group using a two-step protocol.[108] Starting from per-acetylated β-GlcNAc, a Lewis acid catalysed substitution reaction proceeded at the anomeric carbon, with retention of stereochemistry due to anchimeric assistance of the neighbouring N-acetyl group.[109] The acetyl groups were then readily removed by sodium methoxide treatment.

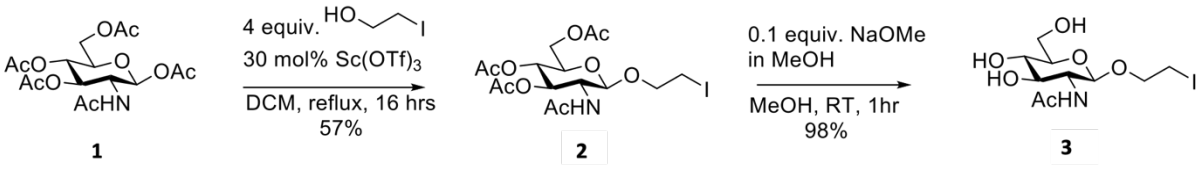


Fig. 4.2.1.1. *Scheme for the sythesis of β-D-1-(2-iodoethoxy)-GlcNAc* ***3*** *from per-acetylated β-GlcNAc* ***1*** *via intermediate per-acetylated β-D-1-(2-iodoethoxy)-GlcNAc* ***2****.*

The synthesis of an appropriate *N*-linked iodo sugar was more complex, and several alternatives were trialled. As the primary amide was a poorer nucleophile than the hydroxyl of the *O*-linked sugar, an amine nucleophile was instead generated at the anomeric position on the GlcNAc so that standard amide coupling methods to an appropriate carbonyl electrophile could be used. Generation of the amide was readily achieved by treatment of a racemic mixture of unprotected GlcNAc with acetyl chloride, acetylating the sugar and chlorinating at the anomeric carbon. Here, the α anomer was thermodynamically favoured due to the anomeric

effect at the C-Cl bond.[110] From there, an S_N2 reaction with sodium azide yielded an azide group at the anomeric position, which could be reduced with 1,3 propane dithiol to generate the desired amine.

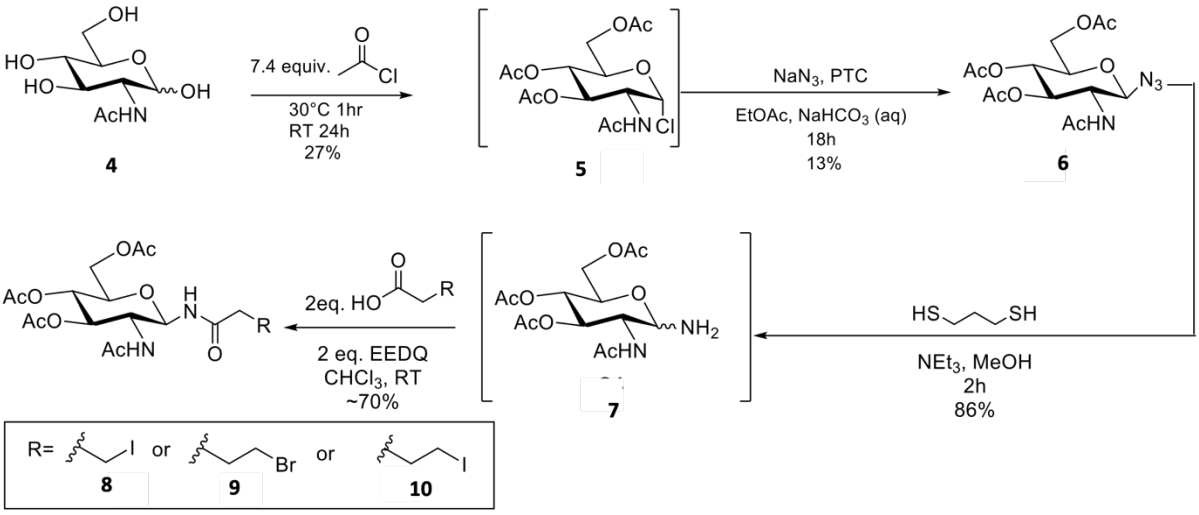


Fig. 4.2.1.2 *Scheme for the synthesis of peracetyl-β-D-1-(3-iodopropionamido)-GlcNAc* ***8****, peracetyl- β-D-1-(4-bromobutyramido)-GlcNAc* ***9,*** *and peracetyl- β-D-1-(4-iodobutyramido)-GlcNAc* ***10****. Synthesis begins from GlcNAc* ***4*** *and goes via Peracetyl-α-D-1-chloro-GlcNAc* ***5****, Peracetyl-β-D-1-azido-GlcNAc* ***6*** *and Peracetyl-D-1-amino-GlcNAc 7.*

Initially, (3-iodopropionamido)-GlcNAc was formed as this would give a shorter homoglutamine linker when used in reaction with Dha. However, while the EEDQ coupling proceeded without issue, even the mildest of basic conditions necessary for deprotection of the acetyl groups caused degradation of the product via E1cB elimination of the iodine.

The longer (4-bromobutyramido)-GlcNAc and (4-iodobutyramido)-GlcNAc were therefore trialled instead, and both synthesised successfully.

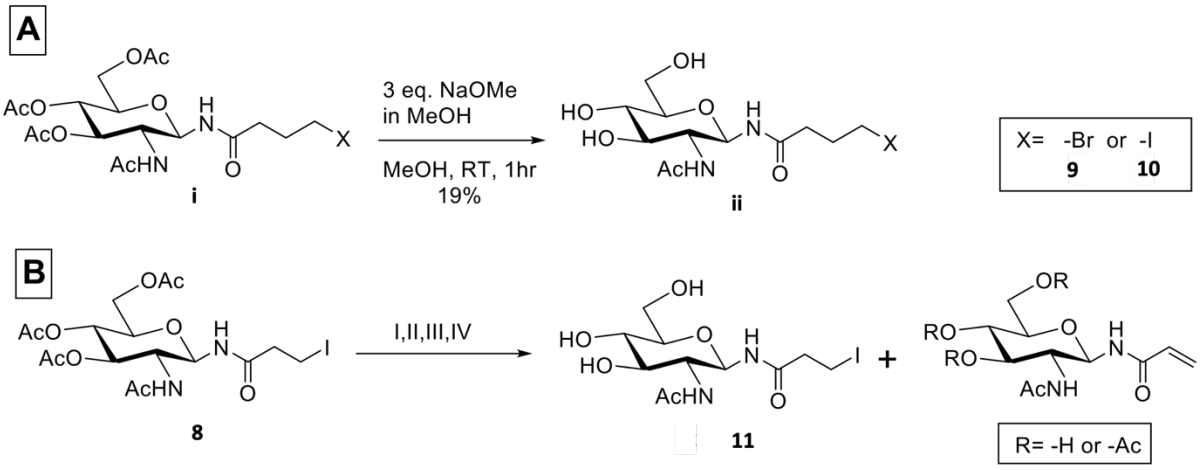


Fig. 4.2.1.3. ***(A)*** *Scheme for the deprotection of peracetyl-β-D-1-(4-bromobutyramido)-GlcNAc* ***9i*** *and peracetyl-β- D-1-(4-iodobutyramido)-GlcNAc* ***10i*** *to form β-D-1-(4-bromobutyramido)-GlcNAc* ***9ii*** *and β-D-1-(4-iodobutyramido)-GlcNAc* ***10ii*** *respectively.* ***(B)*** *Attempted deprotection of peracetyl-β-D-1-(3-iodopropionamido)-GlcNAc* ***8*** *to form β-D-1-(3-iodopropionamido)-GlcNAc* ***11****. The conditions for B are as follows: I= 0.5 M NaOMe in MeOH, 30 mins RT, II = 5 mM NaOMe in MeOH, 30 mins RT, III = Guanidinium nitrate, 62 h RT 2 h 60 ˚C, IV=0.2 M Mg(OMe)2, 45 mins RT.*

4.2.2 Dha formation

Dha formation at cAbLys3-A104C using DBHDA had been established previously within the Davis group, and was repeated successfully with reactions typically achieving 100% conversion after 5 h incubation at 37 °C.



Fig. 4.2.2.1 *cAbLys3A104Dha formation.*

DBHDA is only slightly soluble in aqueous buffer, and most was observed to remain solid by the end of the 5 h reaction period. In previous protocols, remaining DBHDA was removed by centrifugation to pellet solid material, followed by decanting supernatant. However, it was found that passing cAblLys3 through a PD-10 desalting column at this stage to remove any dissolved DBHDA resulted in improved radical reaction efficiency downstream.

4.2.3 Radical reactions; N and O linked sugars

The reaction with the *O*-GlcNAc iodide proceeded to 100% conversion readily, as had been previously reported.[108] However, conversion with both *N*-linked halides was the same within error, and never reached more than 60%.

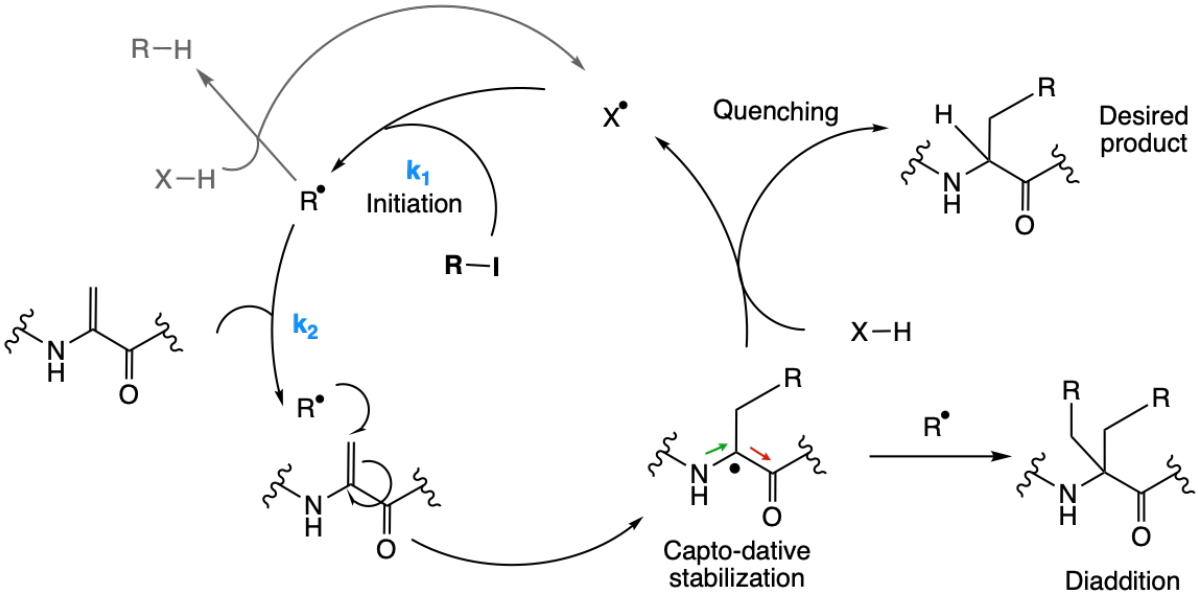


Fig. 4.2.3.1 *Catalytic cycle for radical post-translational mutagenesis reaction.*

The first step in the proposed catalytic cycle is cleavage of the R-X bond to form the requisite GlcNAc radical. As C-I bonds are weaker than C-Br bonds (234 kJ mol^{-1}/ 56 kCal mol^{-1} for C-I compared to 293 kJ mol^{-1}/ 70 kCal mol^{-1} for C-Br), this step was expected to occur faster in

the iodo-*N*-GlcNAc compared to the bromo-*N*-GlcNAc, with both species forming an identical *N*-GlcNAc radical.[111] However, as there was no difference in conversion between the two N-linked halides, in comparison to the dramatic difference in conversion between the N- and O-linked GlcNAc, it was postulated that the nature of the radical, contributing to the attack on Dha was more important than the nature of the R-X bond to be broken; k_2, rather than k_1, was the step limiting conversion.

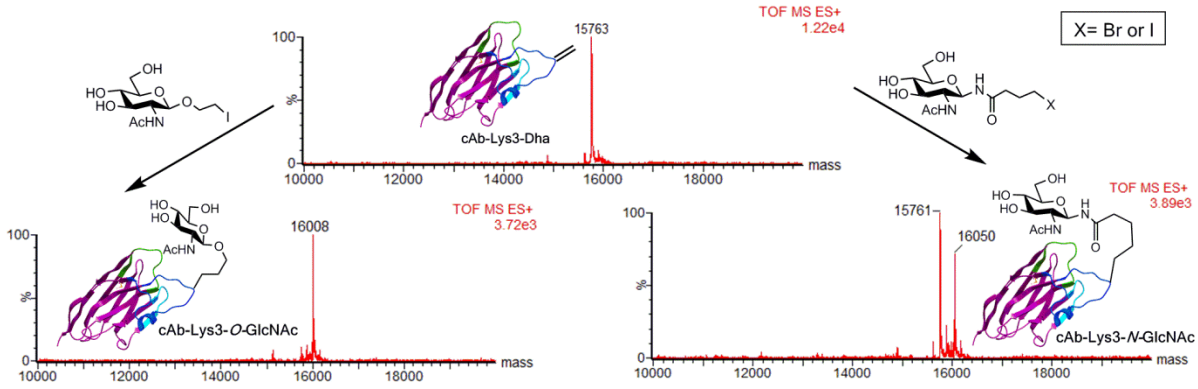


Fig. 4.2.3.2 *Formation of cAbLys3 A104 O-GlcNAc and cAbLys3 A104 N-GlcNAc.*

This was attributed to the inductive electron withdrawing effect of the oxygen two bonds away from the radical in the O-linked GlcNac, destabilizing the radical and thereby enhancing its reactivity. Conversely, in the N-linked GlcNAc there is an extra carbon between the radical and amide centre, and the electron density of the oxygen is resonance stabilised across the amide. As the inductive effect decreases sharply with through-bond distance, the destabilization is expected to be less and the radical formed correspondingly less reactive.

The difference in size between the side product cAbLys3-$A104_{rac}$ and the product cAbLys3-GlcNAc was negligible on the scale of the entire protein, meaning that separation by size exclusion chromatography (SEC) was unfeasible. Moreover, the reaction did not yield a tag

through which the desired product and side product could be separated; although lectin columns are routinely used for the purification of glycoproteins, this is typically successful with larger glycans as the lectin-monosaccharide interaction is relatively weak. Therefore, in practice successful further application of the GlcNAcylated cAbLys3 was dependent on achieving 100% conversion so that no separation was necessary. While the N-linked species was still studied further, the O-linked species was chosen for assays where pure sample was necessary to deconvolute the effects of each species.

In addition, controls for subsequent biochemical analyses were also synthesised. Most interesting was cAbLys3A104A$_{rac}$, in which cAbLys3A104Dha was treated with $NaBH_4$ only to reduce it back to alanine, but in a racemic mixture of D and L epimers.[112] As position 104 is in the middle of cAbLys3's CDR3 loop, it was expected to be crucial in recognition. As the reaction goes via a planar Dha species and radical attack can occur from either side, a 50:50 mixture of natural L and unnatural D epimers was expected to form. It was important therefore to assess whether racemization at the Cα centre alone would impact binding.

4.2.4 Testing scope of the radical reaction with other substrates

The success of the radical post-translational mutagenesis reaction on cAbLys3 was particularly exciting because it represented the first example of successful modification without disruption of the internal disulfide linkages, maintaining the structural integrity of the nanobody. The scope of the reaction on cAbLys3 was probed with other, smaller alkyl halides; as it was hypothesised that the lower conversion in the case of cAbLys3-*N*-GlcNAc was due to the *N*-GlcNAc radical being more alkyl-like in nature than the O-GlcNAc radical, it was reasoned that these results could shed light on this hypothesis.

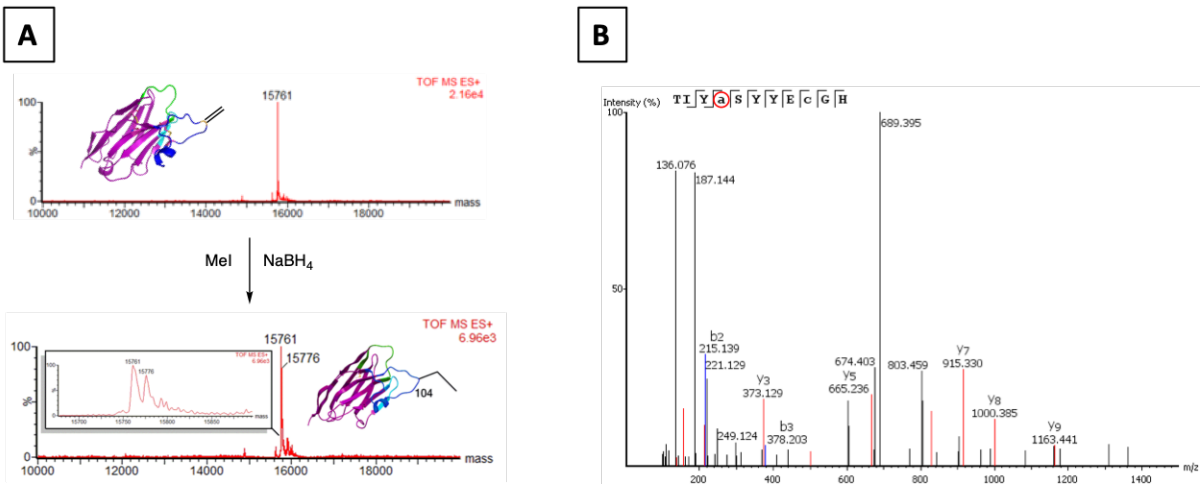


Fig. 4.2.4.1 ***(A)*** *Synthesis of cAbLys3-homoAla from cAbLys3-Dha.* ***(B).*** *MS/MS spectrum for cAbLys3-homoAla peptide confirms the PTM.*

Installation of homoalanine and bis-homoalanine with MeI and EtI, respectively, was successful, although consistent with the hypothesis for the GlcNAc addition, low in yield in both cases.

4.2.5 Testing for disruption of disulfides

In addition to the cysteine introduced for post-translational mutagenesis at position 104, cAbLys3 contains four other native cysteines (C22, C33, C96, and C109) forming two disulfide bridges. In studies on cAbLys3-A104C, only one Dha was observed to form, and previous LC-MS/MS analysis confirmed this to be at position 104 only.[113] Moreover, DTT-treated cAbLys3 WT proved inert to Ellman's reagent, again supporting the resilience of the native disulfides to the step most likely to disrupt them.[113]

Nonetheless, this was confirmed for the post-translational mutagenesis method by a systematic study on cAbLys3's reactivity with iodoacetamide on native, denatured and denatured, reduced protein at every stage of the chemical modification process.

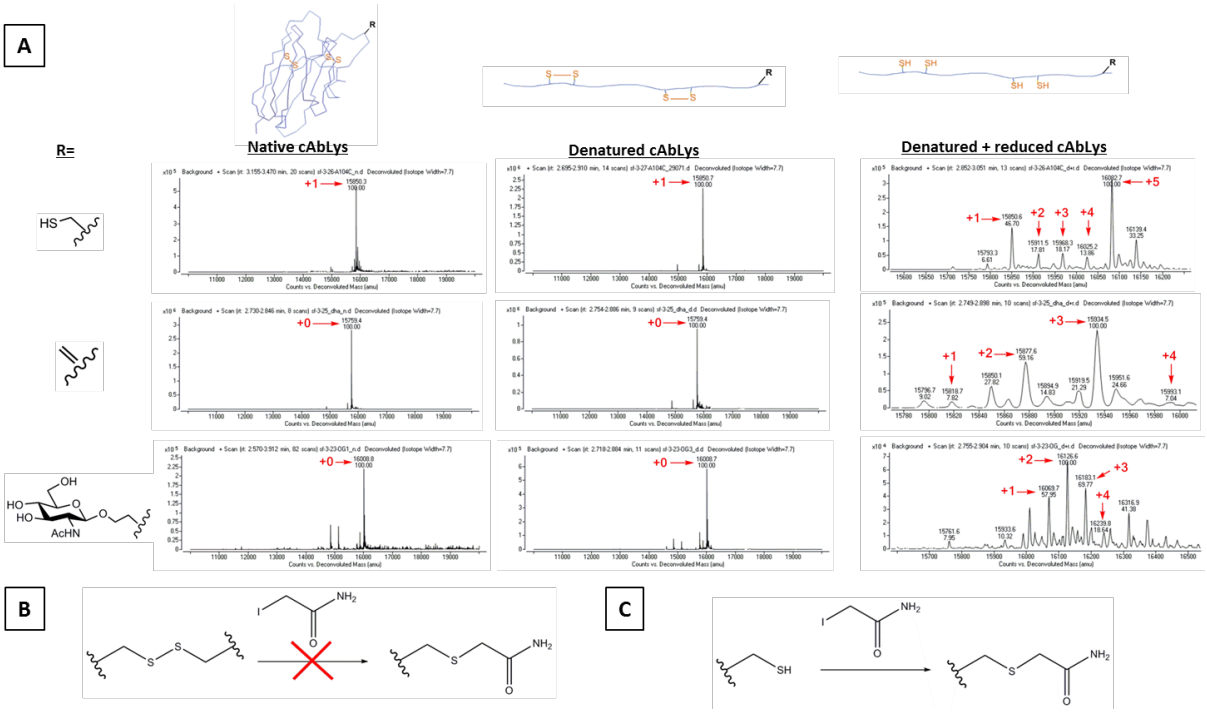


Fig. 4.2.5.1 (A) *ESI-MS showing reaction of free cysteines in the cAbLys variants A104C, A104Dha and A104-O-GlcNAc with iodoacetamide in native, denatured, and denatured + reduced states. Peaks corresponding to the added acetamides are indicated.* ***(B)*** *Reaction with iodoacetamide cannot occur when cysteines are in disulfide bonds.* ***(C)*** *The reaction of free cysteines with iodoacetamide.*

Iodoacetamide was chosen because it could only react with free cysteines, not those sequestered in disulphide linkages; this could be monitored by an increase mass as seen by LCMS-ESI+. In the cAbLys3 A104C native and denatured samples, only one iodoacetamide was added, while five iodoacetamides were added in the denatured and reduced sample in which the disulfide linkages were broken. Similarly for the cAbLys3A104 Dha and -*O*-GlcNAc samples, no reaction with iodoacetamide was observed until the protein was both denatured and reduced, at which point four iodoacetamides were added. As the protein is normally only subjected to the conditions needed to make the 'native' sample, it was confirmed that the treatment did not harm the native disulfides.

It was not possible to conclusively distinguish whether the disulfides were never affected by the treatment, or whether they temporarily broke and always re-formed in the correct configuration. However, previous LCMS/MS results have observed reaction only at position 104, and controls have shown that cAbLys3 WT subjected to the same treatment retain activity, so it is highly likely that the former was the case here.

In any case, retention of the native disulfides was crucial to maintain cAbLys3's binding capacity, as these were known to position the antigen recognition loops in the correct conformation for binding.

4.2.6 Screening *O*-GlcNAcases for cleavage

Having installed the GlcNAc PTMs to initially block the binding activity of cAbLys3, it was then necessary to establish an enzymatic system for removing them so that binding could be 'switched' back on. *In-vivo*, this occurs via GlcNAcases. For the O-linked sugar, partial conversion (less than 50%) had previously been established with a Jackbean O-GlcNAcase (OGA); however, in order to accurately assess the binding capacities of the glycosylated and deglycosylated species, we sought conditions where 100% conversion could be achieved so as to obtain the pure deglycosylated form.

This was achieved by use of the bacterial *X.manhotis* OGA (xmOGA), which yielded 100% conversion after 24 hours. As we expect 50% of the cAbLys3-A10- O-GlcNAc to be present in the unnatural D epimer, this suggests that the enzyme is able to handle both epimers, whereas jbOGA's conversion of less than 50% suggests that it was not; the sensitivity is species specific.

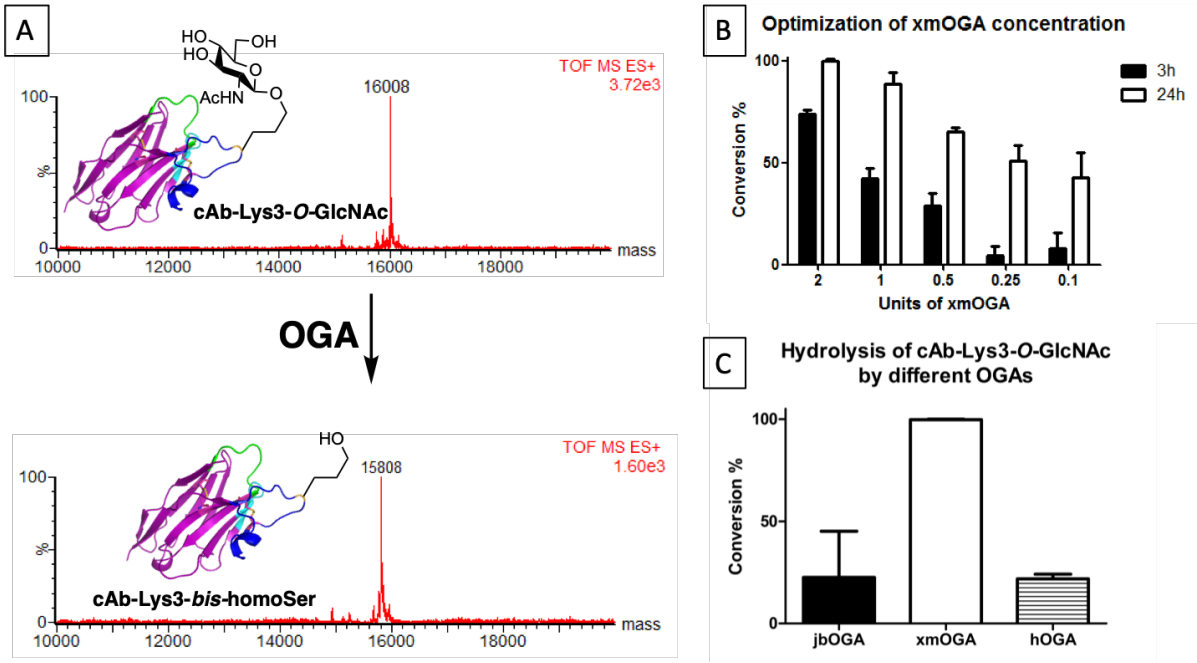


Fig. 4.2.6.1 ***(A)*** *Deconvoluted LCMS-ESI+ spectra for OGA cleavage.* ***(B)*** *Results of X. manhotis OGA reaction optimization. Error bars represent the error of measurement, and are taken from the standard error of the mean (SEM) of values calculated from ESI-MS from each third of the protein peak in the LC chromatogram.* ***(C)*** *Comparison of xmOGA cleavage with jackbean OGA (jbOGA) and human OGA (hOGA). Data for these enxymes from refs* 74,108.

It has also been shown previously that human OGA was capable of cleaving the sugar, albeit with low conversion.[74] This was promising for potential therapeutic *in-vivo* applications, where hOGA would be responsible for removing the sugar.

4.2.7 Purification of cAbLys3-bis-homoSer

Following successful OGA cleavage, it was necessary to separate the cAbLys3-hhSer from the OGA and free GlcNAc. As only small quantities were generated, a method which preserved as much cAbLys3 as possible was needed.

As OGA is much larger than cAbLys3 (approx. 100 kDa for OGA compared to approx.15 kDa for cAbLys3), an initial attempt was made to purify cAbLys3 by passing through a vivaspin

concentrator, with the aim that OGA would remain in the concentrator and cAbLys3 would pass through. However, it was found that cAbLys3 remained stuck to the membrane.

Instead, an analytical S75 size exclusion column was used for FPLC purification of the modified cAbLys3. This method was successful for the separation, giving pure cAbLys3-bis-homoSer for subsequent steps.

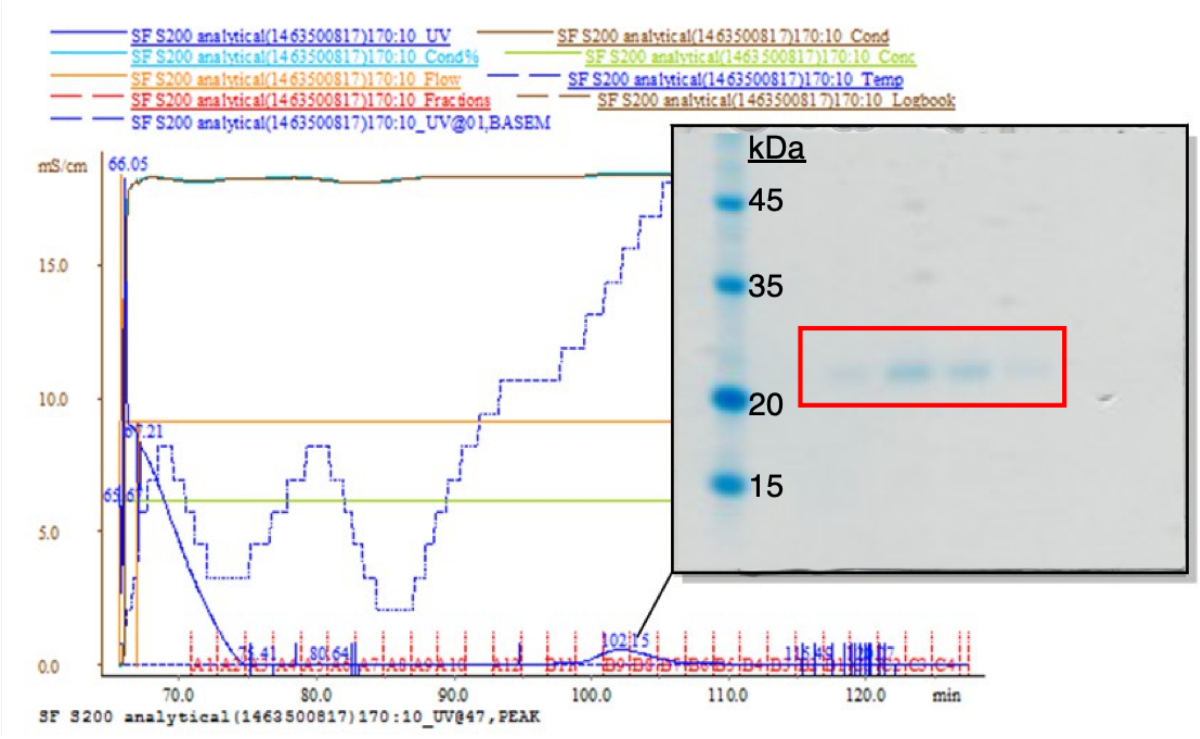


Fig. 4.2.7.1*. SEC UV trace and SDS-PAGE for cAbLys3 purification.*

4.2.8 Testing *N*-linked cleavage with PNGase

Unlike with the O-linked GlcNAc, cleavage of the N-linked sugar from cAbLys3 with PNGase F was unsuccessful, with no conversion observed. In contrast, the branched polysaccharides of Herceptin were readily cleaved by PNGase F.

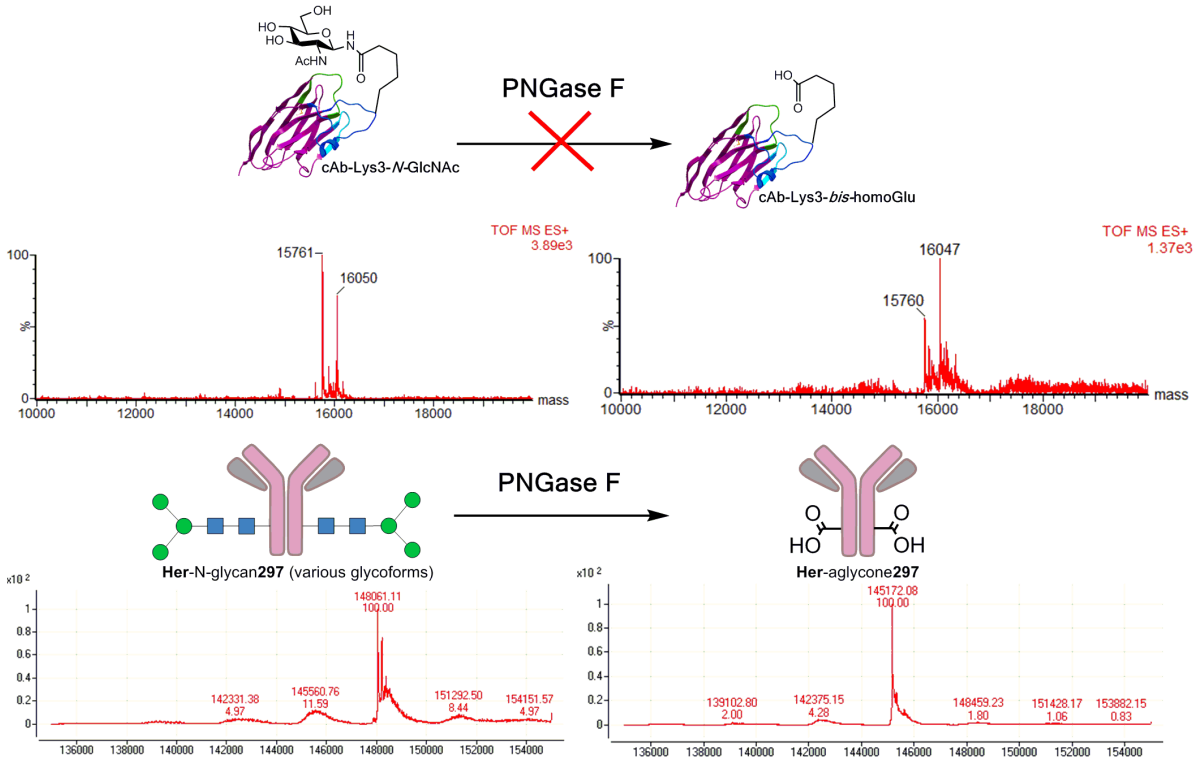


Fig. 4.2.8.1*. Treatment of cAbLys3-N-GlcNAc with PNGase F does not give cleavage of the sugar as shown by LCMS-ESI+. In comparison, the glycan of Herceptin (positive control) is readily cleaved.*

It was possible that this was due to the unnatural linker, which was 2 carbons longer than of the natural glutamine. This result was consistent with the biological roles of the two sugar modifications. O-GlcNAc typically exists *in-vivo* as a monosaccharide, and is involved in rapid signalling processes in the same vein as phosphorylation. Conversely, N-linked GlcNAc typically forms the base of a larger glycan for stabilizing proteins and aiding cell-cell recognition. Indeed, literature reports show that PNGase is unable to process any N-linked monosaccharide, not just the unnatural variants tested here.[114]

4.2.9 ELISA to assay binding of cAbLys3 variants

An Enzyme-Linked Immunoabsorbance Assay (ELISA) was used to quantify binding in the three different cAbLys3 states. Conditions were initially optimised using cAbLys3 WT. Initially, the para-nitrophenolphosphate (pNPP) substrate system with a phosphatase-conjugate secondary antibody was used as in previous studies; however, better results were obtained with the more sensitive 3,3',5,5'-Tetramethylbenzidine (TMB) substrate system with a horseradish peroxidase (HRP) conjugate secondary antibody. The effects of varying antigen Hen Egg Lysozyme (HEL) coating concentration and Bovine Serum Albumin (BSA) percentage used in blocking were also tested, although less variation was observed.

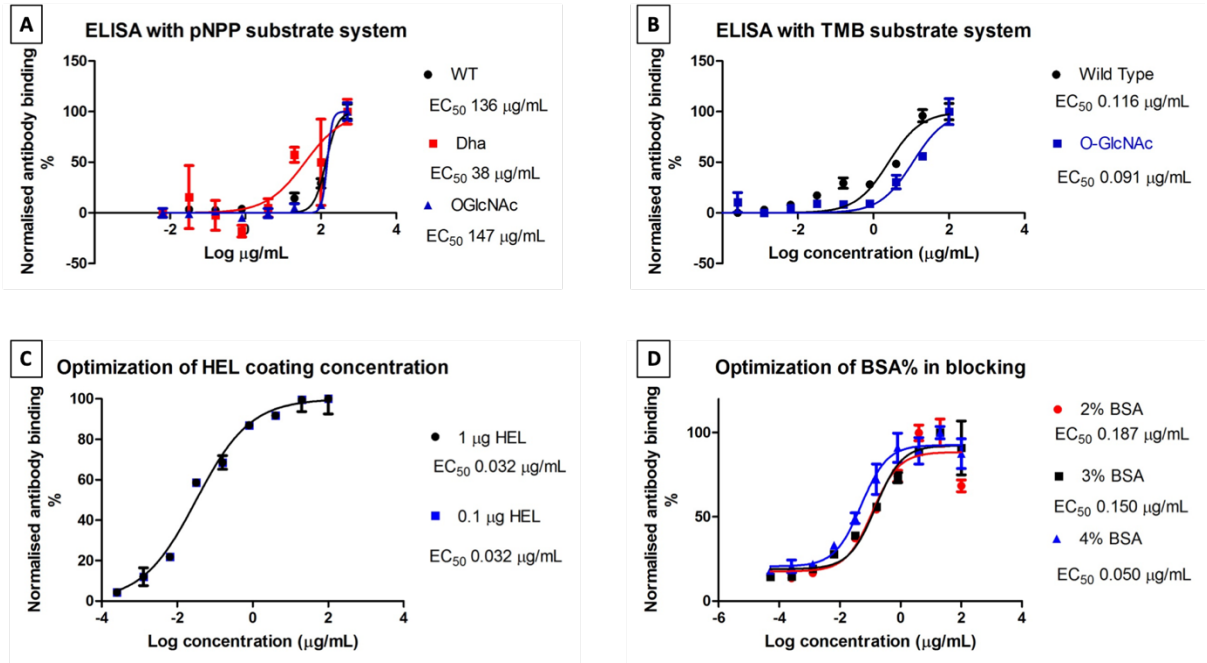


Fig. 4.2.9.1 *Optimization of parameters for ELISA* ***(A)*** *Initial test with pNPP substrate system* ***(B)*** *Test with TMB substrate system* ***(C)*** *Optimization of HEL (antigen) coating concentration* ***(D)*** *Optimization of BSA% used in blocking.*

With optimised conditions in hand, the antigen binding capacities of cAbLys3 WT, cAbLys3 A104-*O*-GlcNAc and cAbLys3A104-hhSer were measured. The addition of the *O*-GlcNAc

was successful in ameliorating much of cAbLys3's binding ability, showing a successful system for switching off the nanobody. However, removal of the sugar was insufficient to switch binding back on, as binding for cAbLys3-bis-homoSer was the same within error.

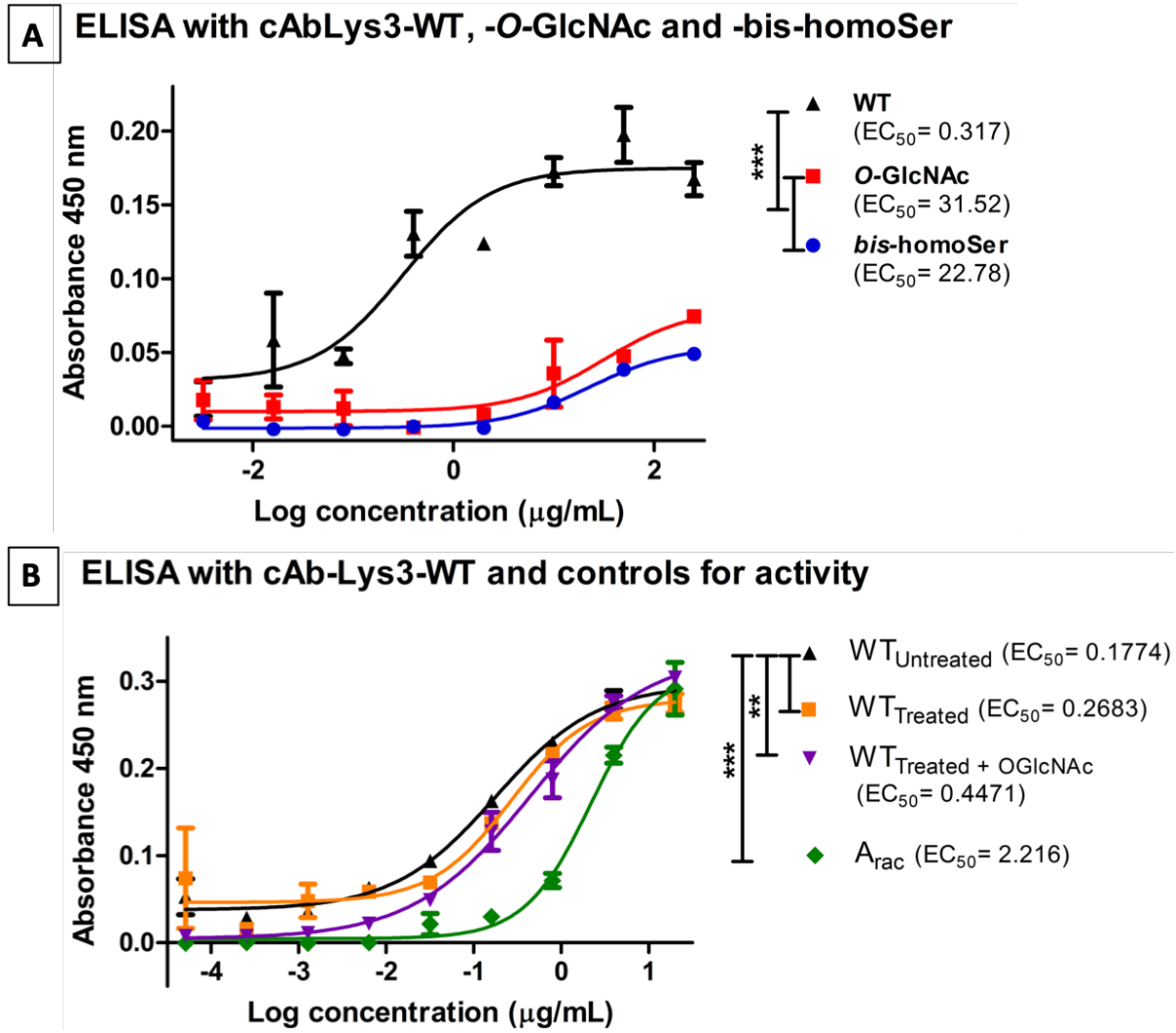


Fig. 4.2.9.2. (A) *ELISA with cAb-Lys3-WT, -O-GlcNAc and –bis-homoSer.* ***(B)*** *ELISA with cAb-Lys3-WT, WT treated with DTT, DBHDA and $NaBH_4$ ($WT_{treated}$), cAb-Lys3-A104A$_{rac}$, and WT treated with DTT, DBHDA, $NaBH_4$ and OGlcNAc ($WT_{treated+OGlcNAc}$). Statistical significance is determined by a sum-of-squares F test (***p<0.001, **p<0.005, blank indicates no significant difference).*

A control study was also carried out to ensure this loss of binding ability was not simply due to degradation of cAbLys3 during treatment. While the WT controls subjected to chemical treatment gave binding curves which were again within error of the untreated WT, the sample where alanine was generated by reaction of cAbLys3-A104Dha with $NaBH_4$, and therefore racemised, showed a slightly reduced binding affinity. However, the EC_{50} value measured here was still 2.216 μg/ mL, 10-fold lower than those measured with cAbLys3-A104-*O*-GlcNAc or -bis-homoSer. This is consistent with previous studies, where binding was restored on cleavage of a thiophosphate moiety at position 104 to leave a similarly racemised cysteine.

Two explanations are proposed for this result. The first is that the linker between cAbLys3 and the GlcNAc is too long and flexible, allowing the sugar to move away from the binding site to the same extent that the bis-homoSer can. The second is that the O-GlcNAc alone is not bulky enough to decrease the binding affinity with respect to the cleaved bis-homoSerine, and the decreased binding affinity results from the change from the apolar alanine to a polar group. In the previous system where switching was successful, a phosphate group was used as the PTM; this carried a charge of -2, which was returned to 0 on phosphatase cleavage. It is also likely that these two factors acted in concert to prevent any differential binding between cAbLys3-*O*-GlcNAc and cAbLys3-bis-homoSer.

4.2.10 Testing addition of larger glycans to cAbLys3-*O*-GlcNAc

While switching the binding of cAbLys3 on and off using a GlcNAc PTM had proved unsuccessful, it provided a potential base for further modification. In particular, the addition of larger sugars characteristic of full antibodies was investigated, as this might provide a means of recruiting FcγRs which nanobodies alone are unable to do. This would be interesting as a

proof of concept towards bifunctional nanobody-based therapeutics, retaining the advantages of using the smaller nanobody whilst also bearing some functionality of traditional, full length antibody therapies.

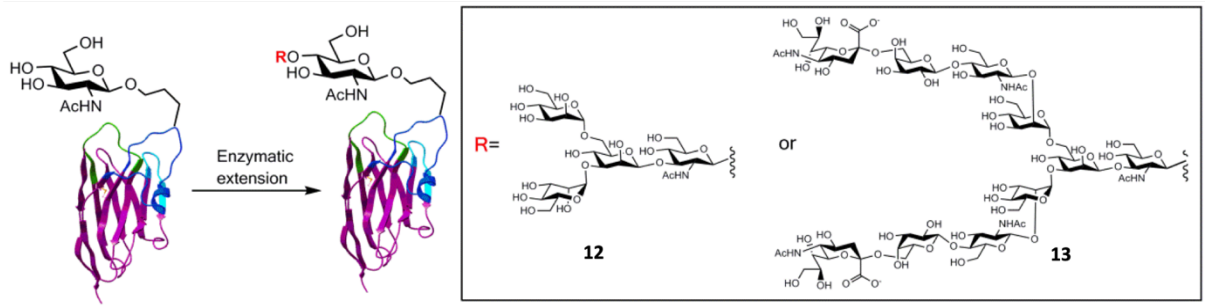


Fig. 4.2.10.1. *Enzymatic extension of cAb-Lys3-O-GlcNAc with tetrasaccharide* ***12*** *or decasaccharide* ***13****. Extension with the tetrasaccharide* ***12*** *has been previously tested; extension with decasaccharide* ***13*** *was tested in this work.*

Preliminary studies had established the addition of a core tetrasaccharide to cAbLys3-O-GlcNAc using engineered glycosyltransferase enzymes EndoS D233Q and EndoA Y299F and previous work within the Davis group had shown these enzymes could be used to install a structurally precise biantennary A2F glycan on the antibody Herceptin after its native glycans were removed by endoglycosidase trimming.[115]

Both EndoA and EndoS naturally possess glycosidase activity to cleave glycosidic linkages, with EndoA wild type also displaying slight glycosyltransferase activity. In both cases, glycosyltransferase activity was favoured by using an oxazoline donor substrate and using enzyme active site mutant variants. EndoA naturally processes GlcNAc-Asn linkages, but has previously been shown to tolerate a wide range of acceptor sugar-protein linkages.[116] However, the donor sugars it is known to add are much smaller than the decasaccharide; indeed, it has been reported that EndoA is incapable of cleaving the core decasaccharide.[117]

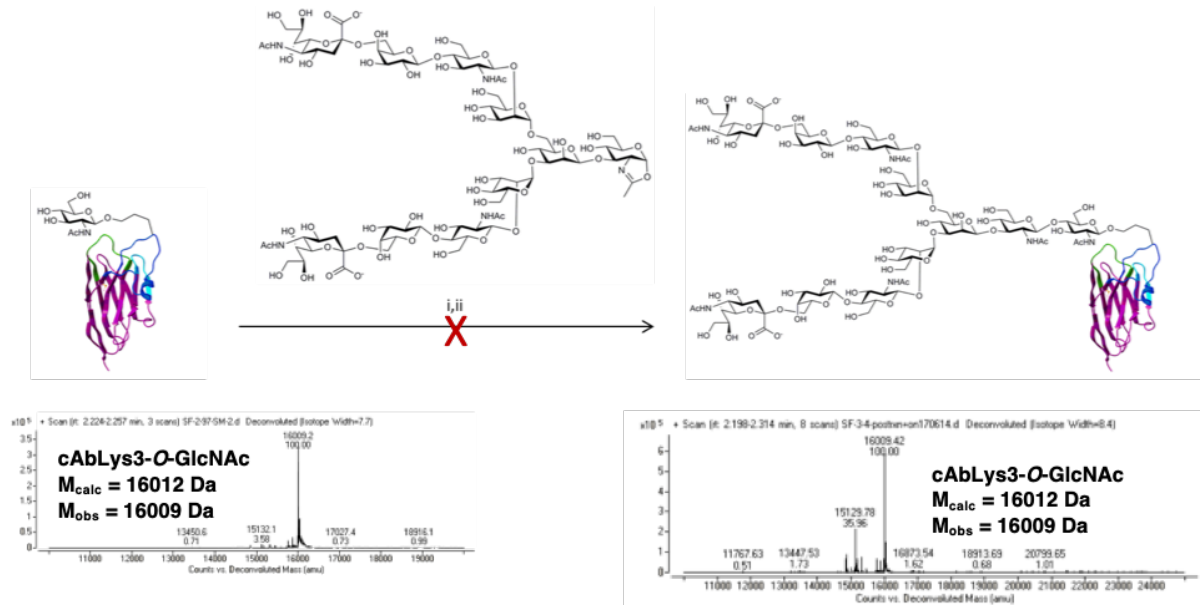


Fig. 4.2.10.2. *No extension from cAbLys3-O-GlcNAc was observed with either EndoA Y299F or Endo S D233Q.*

In contrast, EndoS WT functions to trim glycans from IgG substrates, recognising the core $GlcNAc_2Man_3$ tetrasaccharide.[117] However, this is in the context of an entire antibody and from an N-GlcNAc link to Asp, and the enzyme is known for its substrate selectivity. It is therefore likely that the EndoS mutant is incapable of processing the *O*-GlcNAc linkage at least, and may be further unable to recognise cAbLys3 as a substrate entirely.

4.3 Conclusion

Synthesis of the *O*-linked sugar β-D-1-(2-iodoethoxy)-GlcNAc **1** had previously been established, and was able to readily modify cAb-Lys3. Initial attempts to synthesize β-D-1-(3-iodopropionamido)-GlcNAc failed at the final deprotection stage, so β-D-1-(4 bromobutyramido)-GlcNAc and β-D-1-(4-iodobutyramido)-GlcNAc were made instead to generate a cAb-Lys3-*N*-GlcNAc and screen different halides for conversion rate.

The *O*-linked β-6-(2-iodoethoxy)-GlcNAc and the *N*-linked butyramido-GlcNAcs were used to modify the mutant cAb-LysA104C using a radical coupling reaction with dehydroalanine (Dha) generated from the introduced cysteine. In addition to the free cysteine, cAb-Lys3-A104C contains four other cysteine residues in two disulfide bonds. To test whether these were retained during this process, different folding states were reacted with iodoacetamide. The results confirmed that the conditions for generating cAb-Lys3-Dha and –*O*-GlcNAc did not disrupt these disulfide bonds.

Separation of starting material and product was unfeasible due to the negligible difference in relative sizes. The previously established *O*-linked sugar reaction typically went to completion, circumventing this issue. However, at most 58% conversion was achieved with the *N*-linked sugars, with greater conversion achieved from the more reactive iodo variant ELISAs could therefore only be performed with the *O*-linked sugar.

The *O*-GlcNAcase (OGA) enzymes from several organisms were tested with the cAb-Lys3-*O*-GlcNAc, and PNGase F was likewise tested with cAb-Lys3-*N*-GlcNAc. The bacterial *X. manhotis* OGA gave 100% conversion to cAb-Lys3-*bis*-homoSerine within 24 hours, providing the means for generating the cleaved nanobody on a large scale for binding assays. The Family 84 human OGA (hOGA) tested also gave some conversion. As it is the only known hOGA, this is particularly promising for future clinical applications such as detecting and treating cancerous cells where it is overexpressed. The *N*-linked sugar proved resistant to PNGase F cleavage. This is consistent with literature reports showing PNGase F typically has relatively low activity with single GlcNAc moieties, preferring as an endoglycosidase to cleave polysaccharide substrates. It is also possible that the enzyme is not promiscuous enough to process the extended carbon linker in the synthetically modified cAb-Lys3.

The binding affinities of cAbLys3-WT, -*O*-GlcNAc, and –*bis*-homoSer were tested in an ELISA against the nanobody's natural antigen, Hen-egg lysozyme (HEL). Surprisingly, both the *O*-GlcNAc and *bis*-homoSer variants were found to have a reduced but non-zero binding affinity for HEL, and no statistically significant difference between them was observed. Control studies showed that the process of synthesizing the modified nanobodies is not enough to account for the lowered activity. The possibility of HEL itself removing the *O*-GlcNAc during binding was also investigated, but test reactions conducted under the conditions of the assay concluded that no such hydrolysis was occurring. This was hypothesized to be due to insufficient steric bulk of cAbLys3-*O*-GlcNAc compared to cAbLys3-bis-homoSer, lack of overall change in charge as was seen in the successful thiophosphate AND gate, or a combination of the two factors.

Extension of the cAbLys3-*O*-GlcNAc with a core tetrasaccharide has been previously established, and attention was subsequently turned to modifying this with the Fc decasaccharide as an alternative means of tuning a nanobody's effects for potential therapeutic applications. However, this was also unsuccessful, likely due to fundamental limitations of the glycosyltransferase enzymes available. A key future aim for this work is testing the extension using Endo on cAbLys3-*N*-GlcNAc. An advantage here is that the reaction could be performed even with incomplete conversion to cAbLys3-*N*-GlcNAc from Dha, as the large tetrasaccharide added would facilitate separation by size-exclusion chromatography.

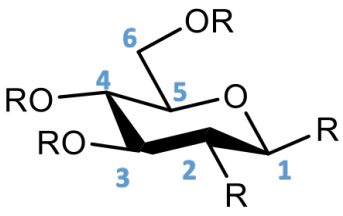
4.4 Experimental

4.4.1.General experimental considerations

Chemicals and solvents were purchased from Acros UK, Alfa Aesar UK, Carbosynth, Fischer UK or Sigma-Aldrich UK and used as delivered unless stated otherwise. Thin layer chromatography (TLC) was carried out using plastic 0.20 mm Polygram® Sil g/UV254 plates that were dried under a stream of nitrogen and visualized under UV (λ_{max} 254 nm), or by use of anisaldhyde, anisidinium phthalate, sulfuric acid or ninhydrin dip. Melting points were recorded using a Leica hotstage microscope and are uncorrected. Flash column chromatography was performed using Geduran® Si 60.8.2 or a Biotage Flash Purification System with a Kinesis Telos or Biotage Snap columns.

NMR Spectroscopy General

Deuterated solvents were used as the lock, and the residual protonated solvent as the internal reference peak. Spectra were analysed using MestReNova. Correlation spectroscopy (COSY) and heteronuclear single quantum correlation (HSQC) were used when necessary to assign peaks. Samples were prepared by dissolving approx. 20 mg in 0.5 mL of solvent. The following numbering system was used when assigning pyranose rings. Carbon chains at the anomeric position are labelled α, β, γ etc with increasing distance from the pyranose ring.



^{1}H NMR Spectroscopy

Chemical shift information (δH) is given in parts per million (ppm) relative to the centre of the solvent signal ($CDCl_3$ 7.26 ppm, D_2O 4.79 ppm DMSO-d6 2.50 ppm, or MeOH-d4 3.31 ppm). Spectra were recorded on a Bruker AVIIIHD 400 nanobays (400.1 MHz) or AVII 500 (500.1 MHz). The number of protons for a resonance signal is given by nH based on integral values. Multiplicity is denoted by the bracketed symbol. Coupling constants (J) are reported to the nearest 0.1 Hz and are determined using MestReNova.

^{13}C NMR Spectroscopy

Chemical shift information (δC) is given in parts per million (ppm) relative to the centre of the solvent signal ($CDCl_3$ 77.16 ppm, DMSO-d6 39.52 ppm, MeOH-d4 49.00 ppm). Spectra were recorded on a Bruker AVIIIHD 400 nanobays (100.6 MHz) or AVII 500 (125.7 MHz). Unless otherwise stated, signals are assumed to be singlets. Distortionless enhancements by polarisation transfer with retention of quaternaries (DEPTQ) pulse sequence with broadband proton decoupling were used to record spectra.

IR Spectroscopy

Infrared (IR) spectra were recorded on a Bruker Tensor 27 Fourier Transform spectrophotometer using an attenuated total reflectance (ATR) attachment, where crystals of the compound were placed directly against the stage. Background measurement was subtracted from each spectrum, and a new background measurement was recorded before recording each spectrum.

Small Molecule Mass Spectrometry

Low-resolution mass spectra were recorded on an Agilent 6120 Quadrupole spectrometer using electrospray ionization (ESI). Positive and/ or negative mode results are given as appropriate. High-resolution mass spectra were recorded on a Bruker MicroTOF ESI mass spectrometer**.**

Optical Rotation

Optical rotation values were recorded using a Schmidt Haensch Unipol polarimeter. Samples were dissolved in spectrophotometric grade chloroform (Sigma-Aldrich UK), D_2O (Apollo Scientific), or methanol (Sigma-Aldrich UK). The polarimeter was blanked with the corresponding solvent before use.

UV-Vis Spectroscopy

UV-Vis spectroscopy was performed using a BMG Labtech Spectrostar Nano. Data was analysed in BMG Labtech Spectrostar Nano MARS. Grenier flat bottom 96 well plates (normal binding unless stated otherwise) were used for sample containment.

Glovebox

Reactions requiring an anaerobic atmosphere were carried out within a Belle Technology glovebox fitted with the BASF R3-11G catalyst. The oxygen level was monitored and measured ≤6 ppm during all reactions.

Protein mass spectrometry

Liquid-chromatography-mass spectrometry (LC-MS) was performed using:

a) Waters LCT Classic coupled to a Shimadzu 20 Series HPLC using a Thermo Proswift (250 x 4.6 mm x 5 μm) column. The gradient was programmed as follows: 95% A (5 min isocratic) to 100% B after 15 min then isocratic for 3 min.

b) Waters LCT Premier XE coupled to a Waters 1525 Micro HPLC using a Merck Chromolith C18 guard cartridge. Water (solvent A) and acetonitrile (solvent B), both containing 0.1% formic acid, were used as mobile phase at a variable flow rate. The gradient program was: 95% A (1 min isocratic, 0.4 ml/min) to 100% B after 5 min (0.4 ml/min) then isocratic for 3 min (flow rate increasing to 1.0 ml/min). The spectra were calibrated automatically. The electrospray source parameters were: capillary voltage: 3000 V, cone voltage: 160 V. Nitrogen was used as the desolvation gas and nebuliser at a total flow of 600 l/hr in both spectrometers.

c) Agilent 6230 TOF MS coupled to a 1260 Infinity HPLC using a poroshell 5 μm 3000 B C3 column (21 x 75 mm). Water (solvent A) and acetonitrile (solvent B), both containing 0.1% formic acid, were used as a mobile phase at a flow rate of 0.8 mL/min. The gradient program was: 90% A (2 min isocratic) to 90% B after 4 min, then isocratic for 1.5 min before returning to 90% A over 0.1 min. The electrospray source parameters were: capillary voltage 4000 V, cone voltage 2000 V. The TOF was run in 1 GHz extended mass range mode (m/z range 20000) or 4GHz high resolution mode (m/z range 3200) as appropriate. Data was processed using the MassHunter BioConfirm software tool (Agilent technologies) using the Maximum Entropy deconvolution algorithm.

4.4.2. Preparation of cAbLys3 A104Dha

4.4.2.1. Reduction of intermolecular disulfides in cAbLys3-A104C with DTT

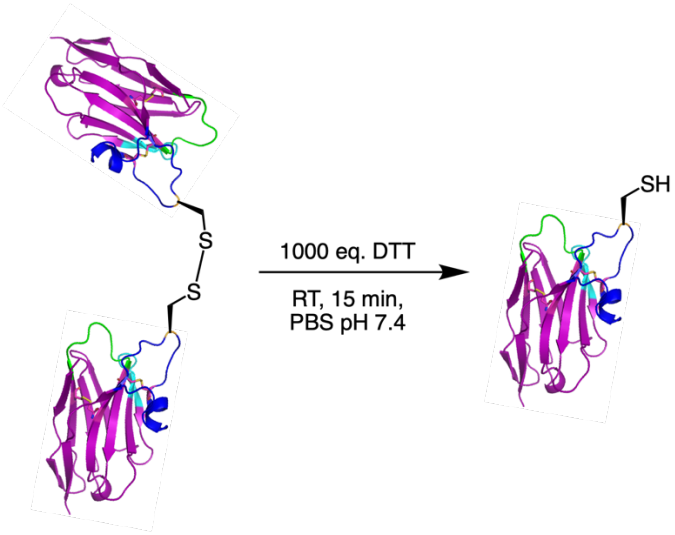
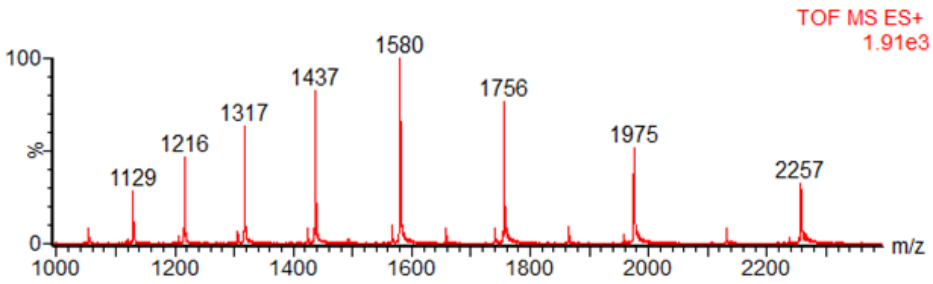


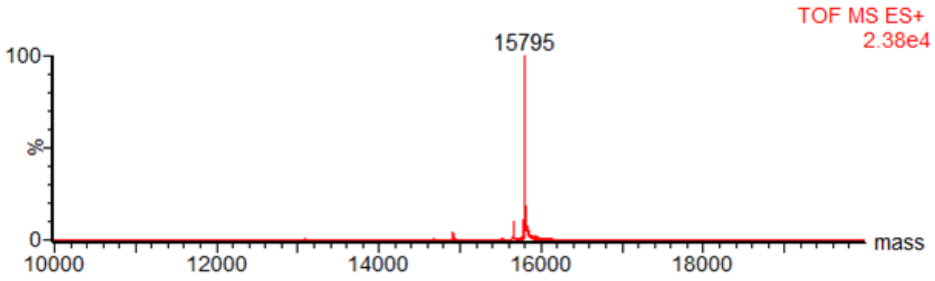
Fig. 4.4.2.1.1. *Scheme for reduction of cAbLys3-A104C with DTT*

DTT (10 mg, 65 µMol; 1000 equivalents) was added to cAb-Lys3-A104C (500 µL of 2.74 mg/mL, prepared by O. Coleman), which was then incubated at room temperature with rotation for 15 minutes. Excess DTT was then removed by loading onto a PD Minitrap G-25 pre-equilibrated with 4x 2 mL PBS pH 8.0, and eluted in 1 mL PBS pH 8.0. The product was characterized by LCMS-ESI+.

LCMS-ESI+ characterization: Ion series



Deconvoluted mass spectrum



cAbLys3A104C: Expect 15797 Da, observe 15795 Da.

Fig. 4.4.2.1.2*. LCMS-ESI+ characterization of reduced cAbLys3-A104C. Also in fig. 4.2.2.1.*

4.4.2.2. Conversion to cAbLys3-A104Dha

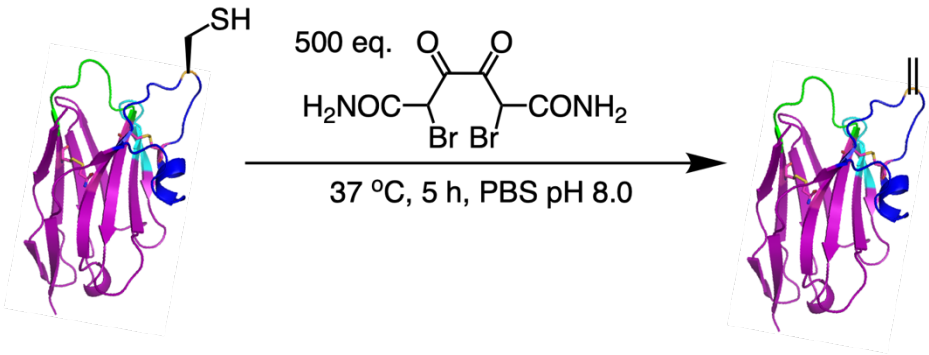
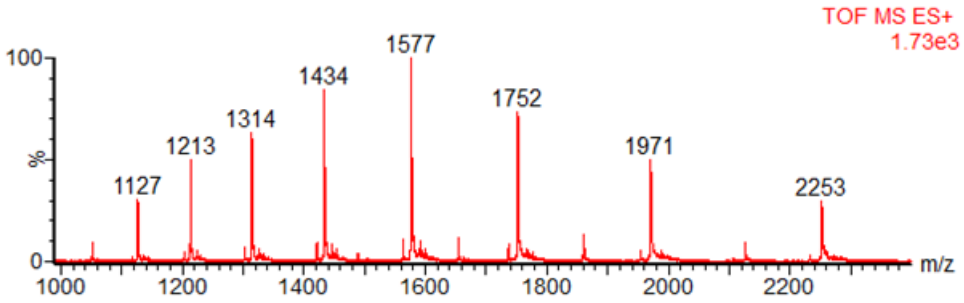


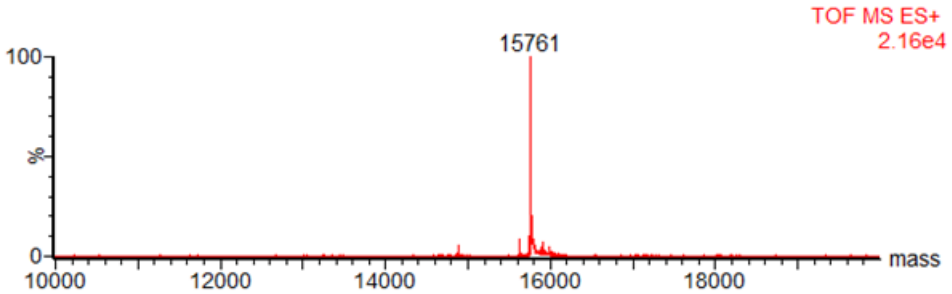
Fig. 4.4.2.2.1*. Scheme for conversion of cAbLys3-A104C to cAbLys3-A104Dha.*

Dibromohexanediamine (DBHDA, synthesised within the Davis group; 10 mg, 500 equivalents) was added to reduced cAb-Lys3-A104C from the previous step. This was then incubated at 37˚C for 5 hours with shaking at 500 rpm. Excess DBHDA was removed by passing through a PD Miditrap G-25 pre-equilibrated with 4x 5 mL PBS pH 8.0, and eluted in 1.5 mL PBS pH 8.0. The product was characterized by LCMS-ESI+.

LCMS-ESI+ characterization: Ion series



Deconvoluted mass spectrum



cAbLys3-A103Dha: Expect 15763 Da, observe 15761 Da.

Fig. 4.4.2.2. *LCMS-ESI+ characterization for cAbLys3-A104Dha. Also in fig. 4.2.2.1.*

4.4.3 Radical reactions with cAbLys3-A104Dha

4.4.3.1. Radical addition to cAb-Lys3-A104Dha- general procedure A

1.5 mL cAb-Lys3-A104Dha solution in PBS pH 8.0 was divided into two 750 µL aliquots and placed in an inert atmosphere glovebox via rapid port. The solution was equilibrated to anaerobic conditions for at least 12 hours.

Immediately before reactions were carried out, the iodide of interest was ported into the glovebox in pre-weighed aliquots along with 0.5 mg aliquots of $NaBH_4$. 200 µL of cAb-Lys3-A104Dha solution was added to each aliquot of the iodide, and pipette mixed to suspend the latter. The mixture was then transferred into the $NaBH_4$-containing Eppendorf and shaken at 800 rpm for 30 minutes at 5°C. Evolution of gas was observed during the reaction. Successful

reactions were identified by LCMS-ESI+ and $NaBH_4$ removed by passing through a PD spintrap G-25 column pre-equilibrated with the desired buffer.

Where applicable, conversion was calculated by taking the ratio of product to starting material and/or side product peak intensities in the deconvoluted LCMS-ESI+ mass spectrum.

4.4.3.2. Synthesis of cAb-Lys3-A104-O-GlcNAc

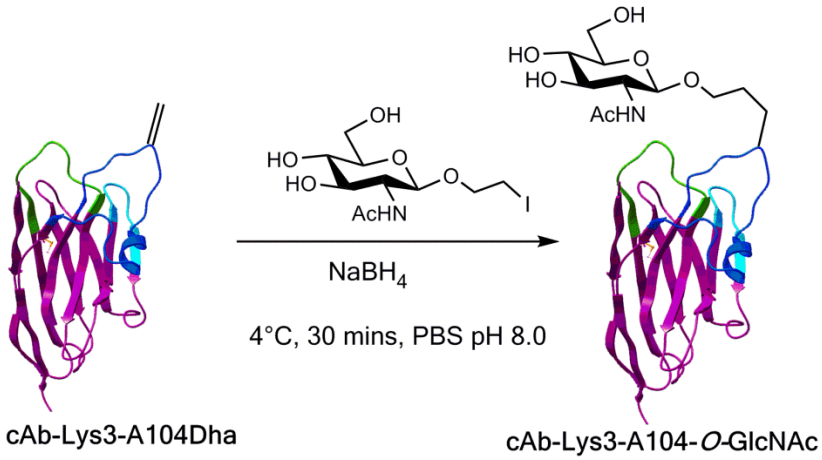
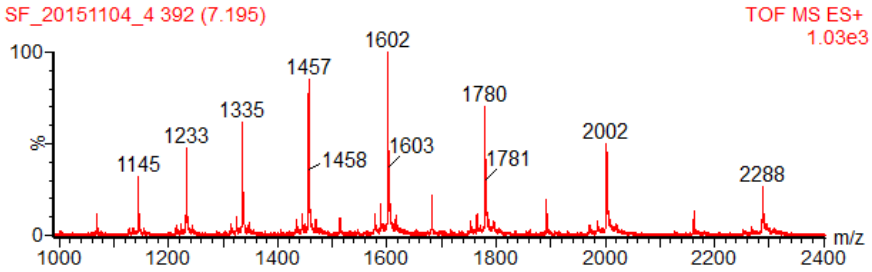


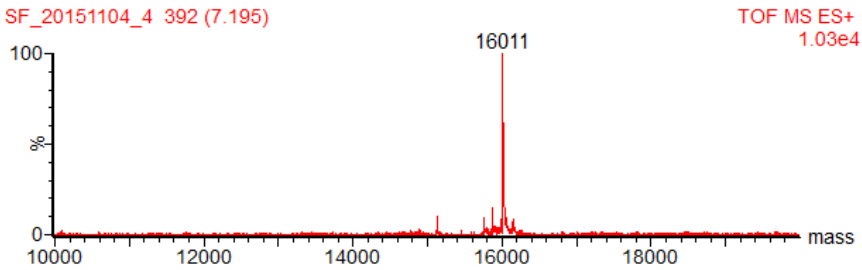
Fig. 4.4.3.2.1*. Scheme for synthesis of cAbLys3-A104-*O*-GlcNAc.*

Following general procedure A, β-6-(2-iodoethoxy)-GlcNAc (2.5mg, 6.7 μMol) was added to 200 μL cAb-Lys3-A104Dha, and subsequently the mixture was added to 0.5 mg $NaBH_4$. Some β-6-(2-iodoethoxy)-GlcNAc was observed to remain undissolved in the mixture. Successful reactions were passed through a spintrap into citrate buffer (100 mM Citric acid, 200 mM NaCl, pH 6.0; for jbOGA cleavage studies), G1 buffer (5mM $CaCl_2$, 50 mM sodium acetate pH 5.5; for xmOGA studies), PBS pH 6.0 (EndoA D233Q reactions) or PBS pH 7.4.

LCMS-ESI+ characterization: ion series



Deconvoluted mass spectrum



cAbLys3-A104-O-GlcNAc: Expected mass 16012 Da, observe 16011 Da.

Fig. 4.4.3.2.2. *LCMS-ESI+ characterization for cAbLys3-A104-*O*-GlcNAc. Also in fig. 4.2.3.2.*

4.4.3.3. Synthesis of cAb-Lys3-A104-N-GlcNAc using β-6-(4-iodobutyramido)-GlcNAc

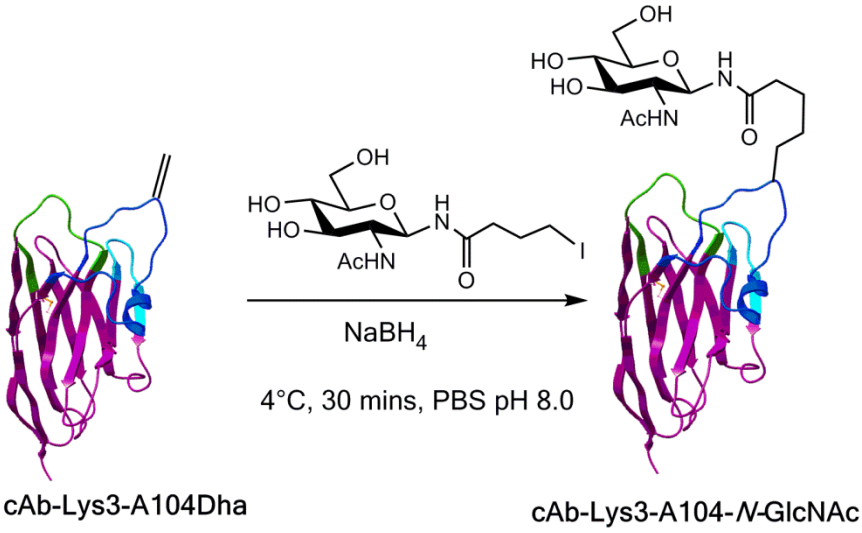
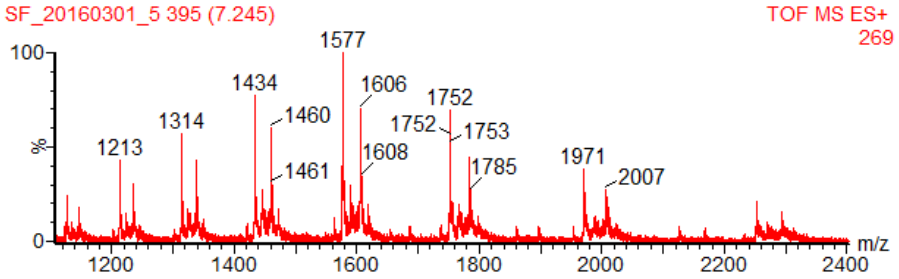


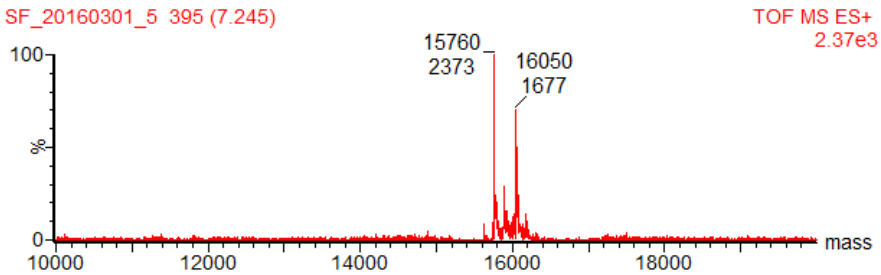
*Fig. 4.4.3.3.1. Scheme for the synthesis of cAb-Lys3-A104-*N*-GlcNAc using β-6-(4-iodobutyramido)-GlcNAc.*

Following general procedure A in duplicate, β-6-(4-iodobutyramido)-GlcNAc (5 mg, 12.4 μMol) was added to 200 μL cAb-Lys3-A104Dha, and subsequently the mixture was added to 0.5 mg $NaBH_4$. Some β-6-(4-iodobutyramido)-GlcNAc was observed to remain undissolved in the mixture. Successful reactions were passed through a spintrap into PBS pH 7.4. LCMS-ESI+ characterization data for one representative reaction is shown below: for additional replicate data, see appendix 4.6.1.1.

LCMS-ESI+ characterization; ion series



Deconvoluted mass spectrum



cAb-Lys3-N-GlcNAc: Expected mass 16049 Da, observed mass 16050 Da (58%)

cAb-Lys3-A_{rac}: Expected mass 15765 Da, observed 15760 Da.

Fig. 4.4.3.3.2. LCMS-ESI+ characterization data for the synthesis of cAb-Lys3-A104-N-GlcNAc using β-6-(4-iodobutyramido)-GlcNAc. Also in fig. 4.2.3.2.

4.4.3.4. Synthesis of cAb-Lys3-A104-*N*-GlcNAc using β-6-(4-bromobutyramido)-GlcNAc

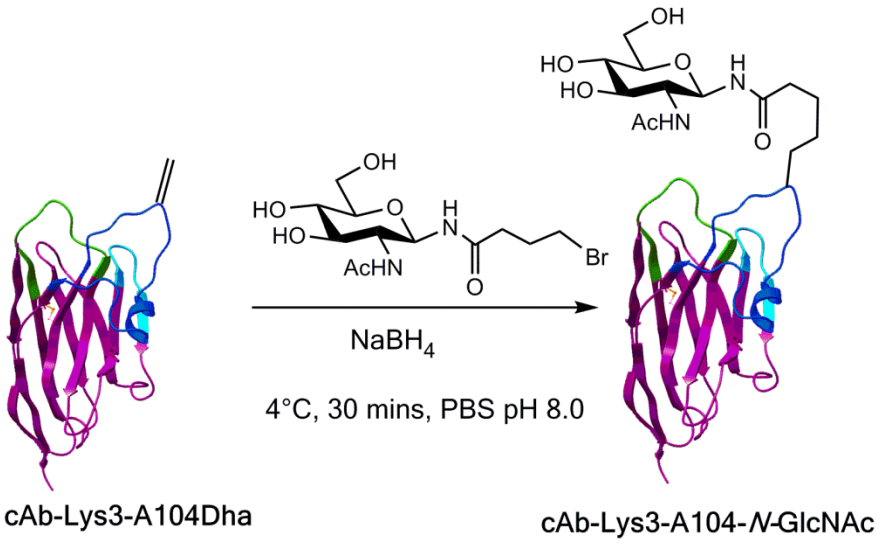
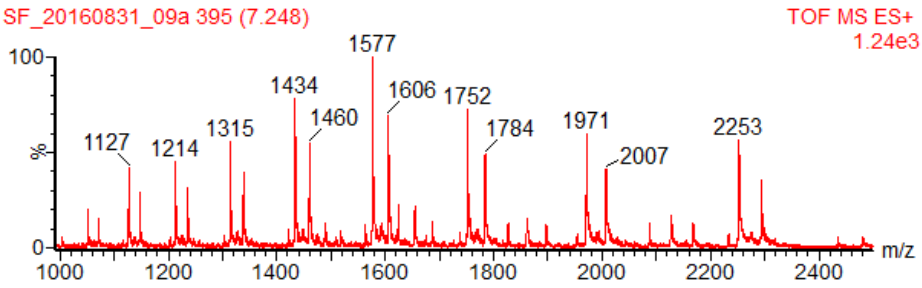


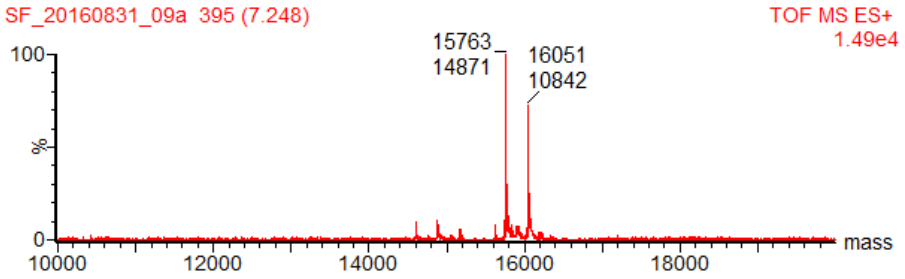
Fig. 4.4.3.4.1. *Scheme for the synthesis of cAb-Lys3-A104-N-GlcNAc using β-6-(4-bromobutyramido)-GlcNAc*

Following general procedure A in triplicate, β-6-(4-bromobutyramido)-GlcNAc (5 mg, synthesised within the Davis group) was added to 200 µL cAb-Lys3-A104Dha, and subsequently the mixture was added to 0.5 mg $NaBH_4$. Successful reactions were passed through a spintrap into PBS pH 7.4. Representative LCMS-ESI+ data for one reaction is shown below: for additional replicates, see appendix 4.6.1.2.

LCMS-ESI+ characterization: ion series



Deconvoluted mass spectrum



cAb-Lys3-*N*-GlcNAc: Expected mass 16049 Da, observed mass 16051 Da (42%)

cAb-Lys3-A_{rac}: Expected mass 15765 Da, observed 15763 Da.

Fig. 4.4.3.4.2. *LCMS-ESI+ characterization data for the synthesis of cAb-Lys3-A104-N-GlcNAc using β-6-(4-bromobutyramido)-GlcNAc*

4.4.3.5. Installation of homoalanine from cAb-Lys3-A104Dha

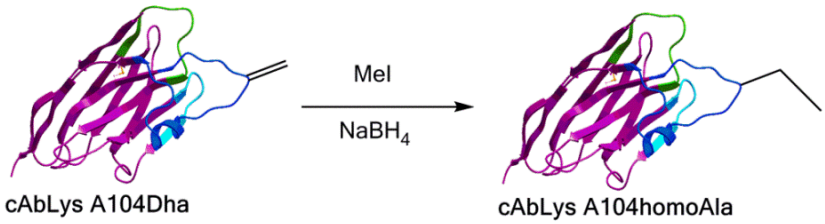
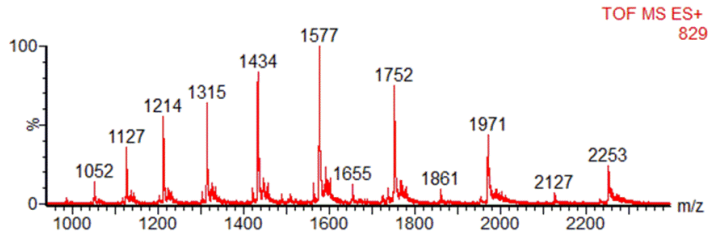


Fig.4.4.3.5.1. *Scheme for synthesis of cAbLys3-homoAla*

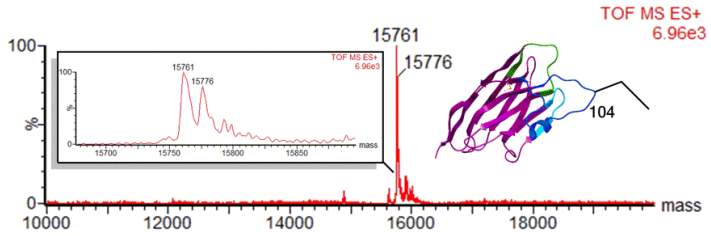
General procedure A was followed. A 200 µL aliquot of cAb-Lys3-A104Dha (0.7 mg/mL solution in PBS pH 8.0) was transferred to an anaerobic atmosphere glovebox and incubated for 12 hr at 4°C. The protein solution was then combined with 6 µL iodomethane and 0.5 mg of sodium borohydride. The reaction was gently shaken at 4°C for 30 minutes before being quenched by elution through a G25 size-exclusion spin column into PBS pH 7.4 and analysed by LC-MS. Conversion of 40% or less was observed, with the rest remaining as cAb-Lys-A104Dha or being reduced to cAb-Lys-A104Ala (15761 Da peak in MS below).

ESI-MS for the modified nanobody is shown below.

LCMS-ESI+ characterization: ion series



Deconvoluted mass spectrum



Expected mass (cAbLysA104homoAla): 15778, Observed mass: 15776.

MS/MS spectrum for peptide $_{100}$TIYhASYYECGH$_{110}$

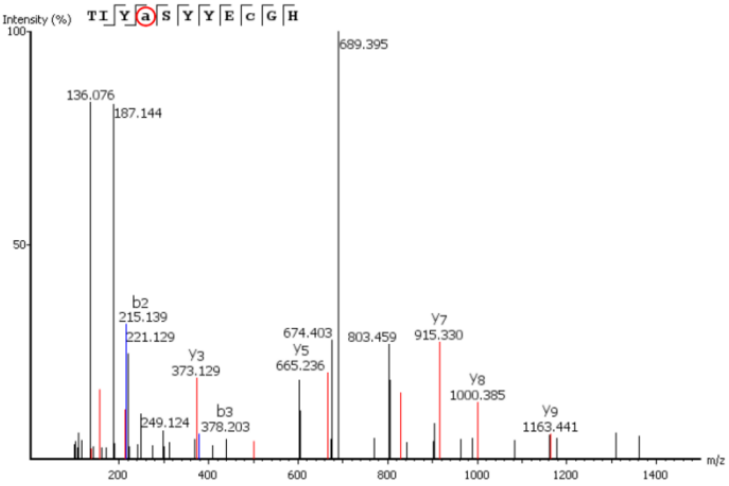


Fig.4.4.3.5.2. *MS characterization of cABLys3-A104homoAla. Also in fig. 4.2.4.1.*

4.4.3.6. Installation of *bis*-homoalanine from cAb-Lys3-A104Dha

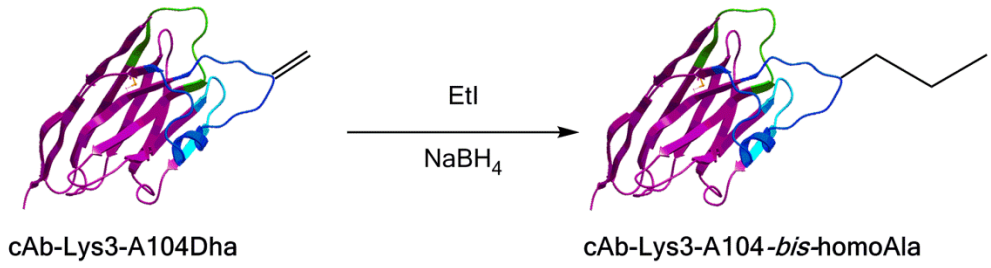
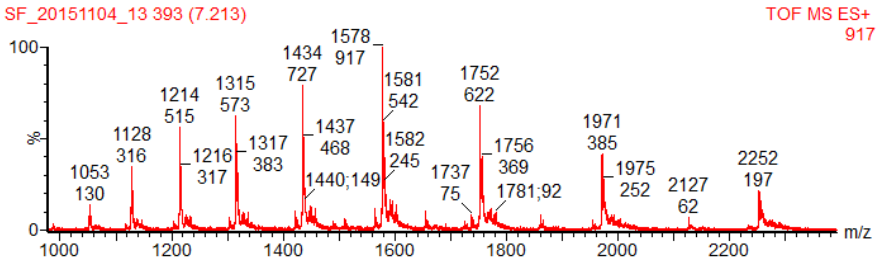


Fig.4.4.3.6.1. *Scheme for synthesis of cAbLys3-bis-homoAla*

General procedure A was followed. A 200 µL aliquot of cAb-Lys3-A104Dha (0.7 mg/mL solution in PBS pH 8.0) was transferred to an anaerobic atmosphere glovebox and incubated for 12 hr at 4°C. The protein solution was then combined with 4 µL 2-iodoethane and 0.5 mg of sodium borohydride. The reaction was gently shaken at 4°C for 30 minutes before being quenched by elution through a G25 size-exclusion spin column into PBS pH 7.4 and analysed by LC-MS. Conversion of 40% or less was observed, with the rest remaining as cAb-Lys-A104Dha or being reduced to cAb-Lys-A104Ala (15765 Da peak in MS below).

LCMS-ESI+ characterization: ion series



Deconvoluted mass spectrum

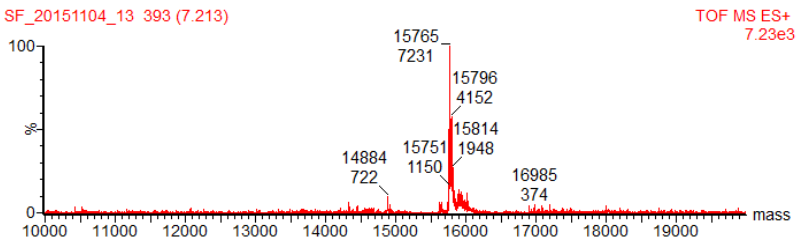


Fig.4.4.3.6.2. *MS characterization of cABLys3-A104-bis-homoAla*

4.4.4 Synthesis of cAbLys3 ELISA controls

4.4.4.1. Synthesis of ELISA controls: cAb-Lys3-$WT_{Treated}$

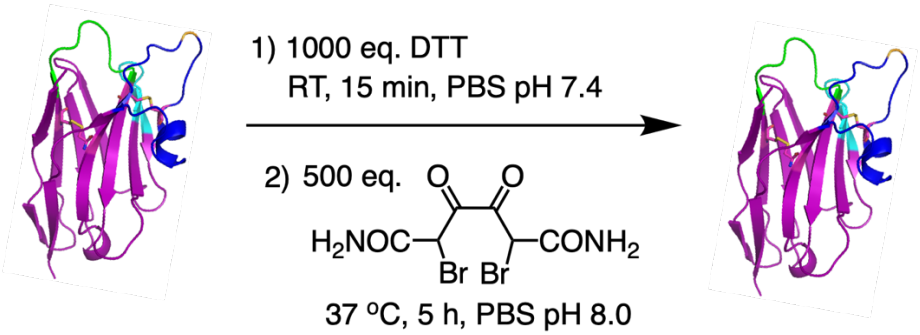


Fig. 4.4.4.1.1. *Scheme for generation of cAb-Lys3-$WT_{Treated}$ ELISA control*

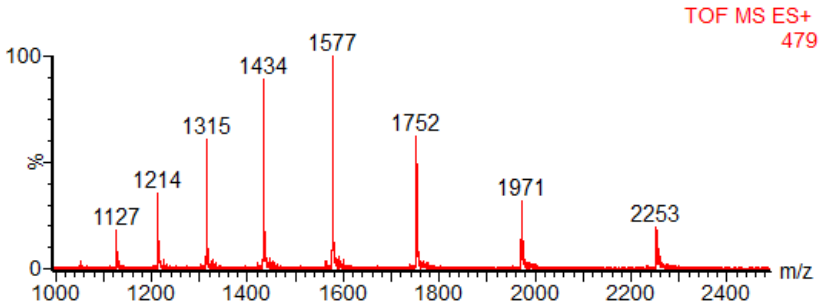
0.5 mL of 1 mg/mL cAb-Lys3-WT was incubated with 10 mg of DTT at room temperature with rotation before eluting through a PD MiniTrap G-25 column into 1 mL PBS pH 8.0. The cAb-Lys3-WT sample was then incubated with 5 mg DBHDA with gentle shaking at 37°C for 5 hours before eluting through a PD MidiTrap G-25 column into 1.5 mL PBS pH 8.0 and analysing by LC-MS.

After incubating for 12 hr in an anaerobic atmosphere glovebox at room temperature, the cAb-Lys3-WT was then combined with 2.5 mg of β-6-(2-iodoethoxy)-GlcNAc and 0.5 mg of sodium borohydride and gently shaken at 4°C for 30 minutes before being quenched by elution through a G25 size-exclusion spin column. It was then dialysed against PBS pH 7.4 to remove excess β-6-(2-iodoethoxy)-GlcNAc and analysed by LC-MS.

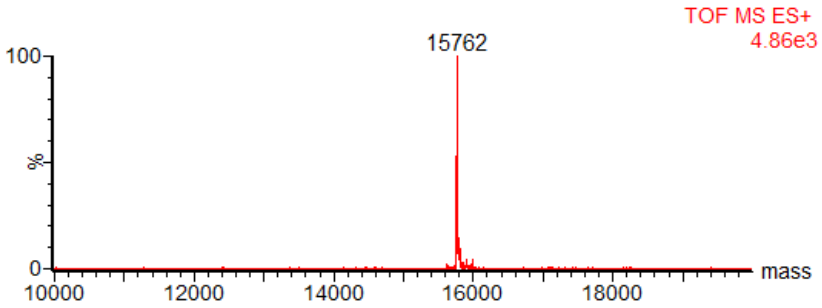
ESI-MS for the modified nanobody are shown below.

LCMS-ESI+ characterization After incubation with DTT and DBHDA

Ion series



Deconvoluted mass spectrum

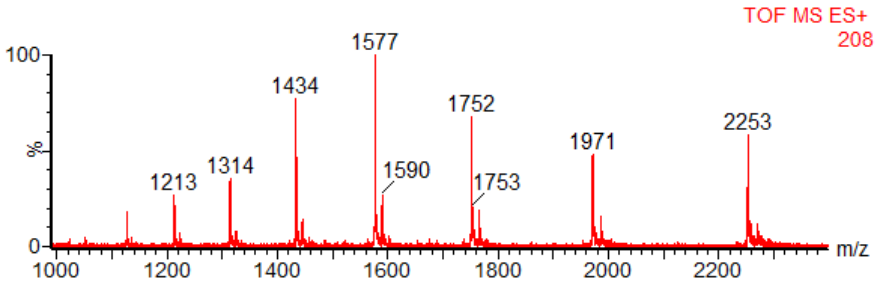


cAbLys3-WT expected mass: 15765 Da, Observed: 15762 Da.

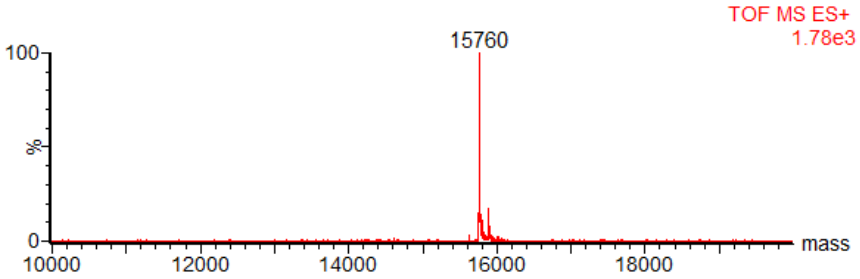
Fig. 4.4.4.1.2. *LCMS-ESI+ characterization data after incubation with DTT and DBHDA*

LCMS-ESI+ characterization After incubation with $NaBH_4$

Ion series



Deconvoluted mass spectrum



cAbLys3-WT expected mass: 15765 Da, Observed: 15760 Da.

Fig. 4.4.4.1.3. *LCMS-ESI+ characterization data after incubation with $NaBH_4$.*

4.4.4.2. Synthesis of ELISA controls: cAb-Lys3-$WT_{unreated}$

An aliquot of the cAb-Lys3-WT stock used in the preparation of cAb-Lys3-$WT_{Treated}$ and cAb-Lys3-$WT_{Treated + OGlcNAc}$ was reserved and stored at 4 ˚C to ensure that any increased activity for cAb-Lys3-$WT_{Untreated}$ was not simply due to the sample being more recently thawed.

4.4.5 Demonstrating retention of disulphide bonds in cAb-Lys3 during chemical mutagenesis

4.4.5.1. General procedure B

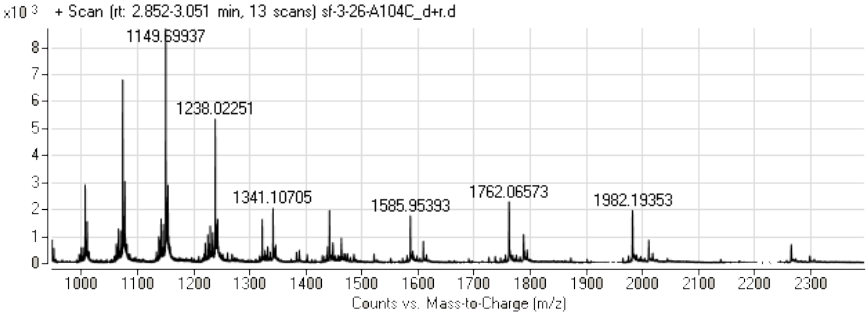
Three aliquots of cAb-Lys3 variant (10µL in PBS pH 5.5) were taken. To the first was added DTT (1mg; 6.5 µMol) and 10 µL of 8M urea to reduce and denature the protein; to the second, 10 µL of 8M urea to denature the protein only, and to the third, 10 µL of PBS pH 5.5 was added to retain the protein's native structure. The reduced+ denatured and denatured samples were incubated at 60°C with shaking at 700 rpm for 20 minutes, whilst the native sample was retained on ice. Approximately 250 equivalents of 2-iodoacetamide per cysteine residue in the protein were then added, and the reaction incubated for 35 minutes at room temperature before analysing by LC-MS ESI+.

4.4.5.2. Demonstrating presence of disulphide bonds in cAb-Lys3-A104C

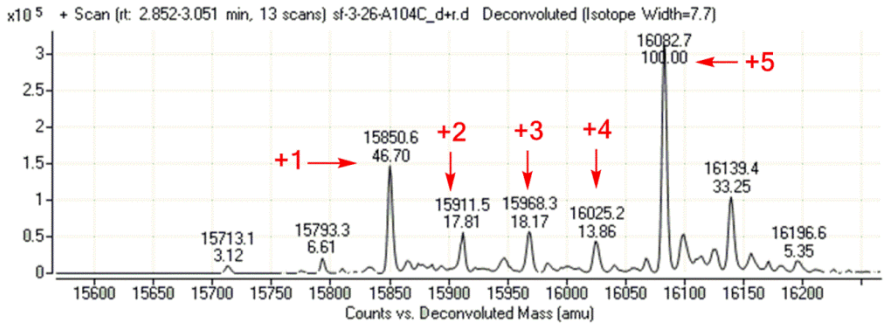
Following general procedure B, three aliquots of freshly-thawed cAb-Lys3-A104C (10 µL of 2.55 mg/mL, 1.59 nMoles) were taken. After the 20 minute denaturation or incubation period for the native and denatured + reduced reactions, 1295 equivalents of 2-iodoacetamide (1.95 µL of 0.95 M stock; 2.06 mMoles) were added, and the reactions incubated at room temperature before analysing by LC-MS ESI+. For the denatured only reaction, cAb-Lys3-A104C, 8M urea and 265 equivalents of 2-iodoacetamide (0.4 µL of 0.95 M stock; 0.265 mMoles) were combined and incubated at room temperature for 5 minutes before temperature before analysing by LC-MS ESI+.

Denatured + reduced cAbLys3-A104C LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



Starting material (cAb-Lys3-A104C); expected mass 15797 Da, observed 15793.3 Da.

+1 acetamide; expected mass 15854 Da, observed 15850.6 Da.

+2 acetamides; expected mass 15911 Da, observed 15911.5 Da.

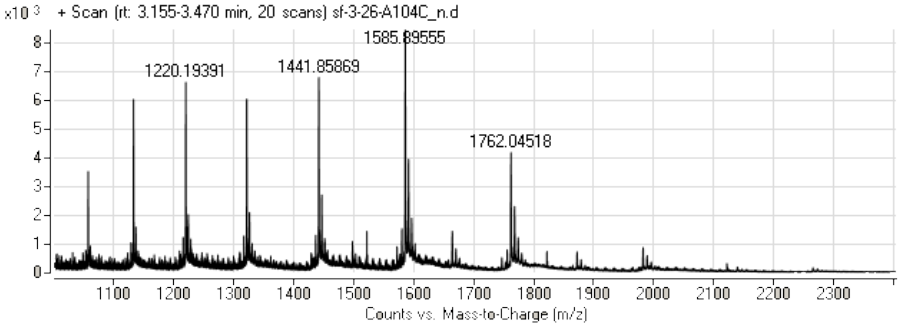
+3 acetamides; expected mass 15968 Da, observed 15968.3 Da.

+4 acetamides; expected mass 16025 Da, observed 16025.2 Da.

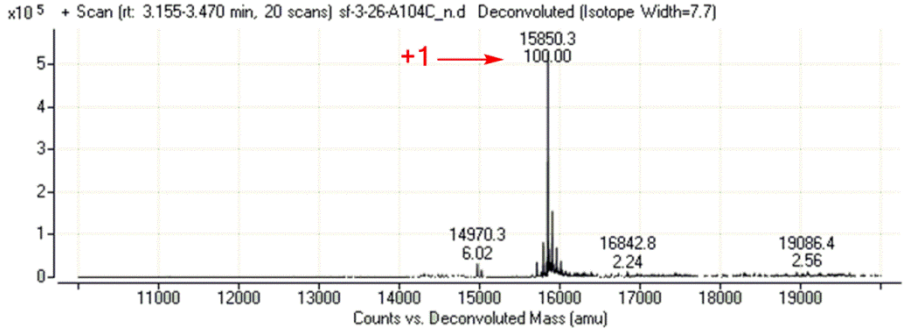
Fig. 4.4.5.2.1. *LCMS-ESI+ characterization of denatured + reduced cAbLys3-A104C. Also in fig. 4.2.5.1.A.*

Native cAbLys3-A104C LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum

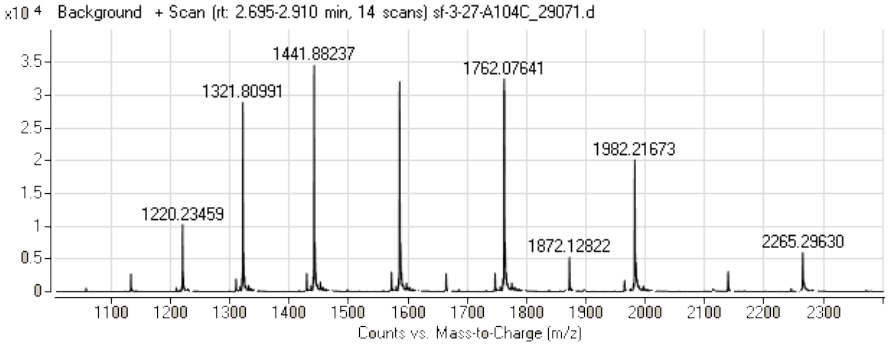


cAb-Lys3-A104C +1 acetamide; expected mass 15854 Da, observed 15850.3 Da.

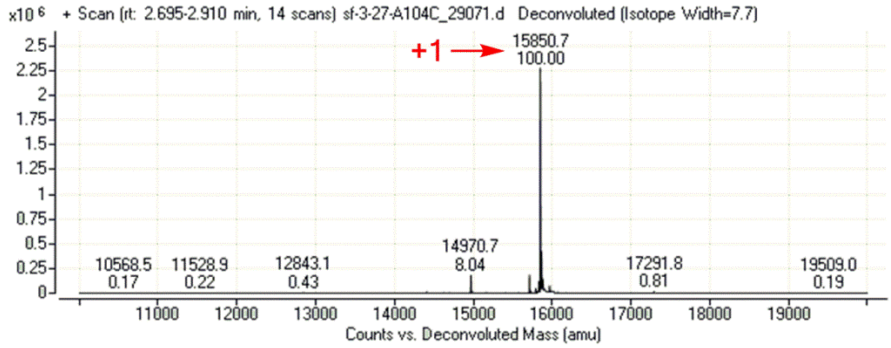
Fig. 4.4.5.2.2. *LCMS-ESI+ characterization of native cAbLys3-A104C. Also in fig. 4.2.5.1.A.*

Reduced cAbLys3-A104C LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-A104C +1 acetamide; expected mass 15854 Da, observed 15850.7 Da.

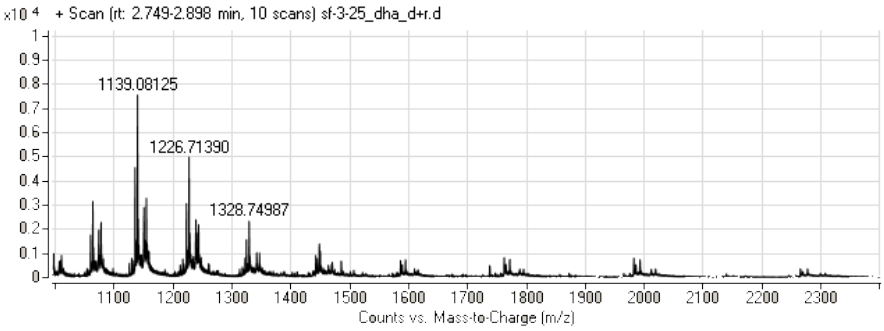
Fig. 4.4.5.2.3. *LCMS-ESI+ characterization of reduced cAbLys3-A104C. Also in fig. 4.2.5.1.A.*

4.4.5.3. Demonstrating retention of disulphide bonds in cAbLys after Dha formation.

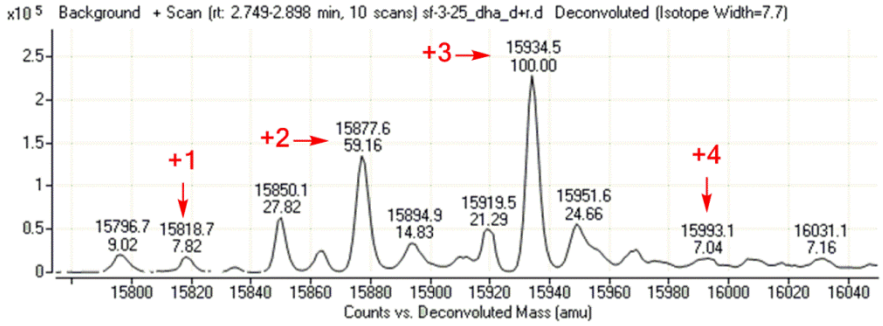
Following general procedure B, three aliquots of freshly prepared cAb-Lys3-A104Dha (10 µL of 0.75 mg/mL, 0.468 nMoles) were taken. After the 20 minute denaturation or incubation period, 1000 equivalents of 2-iodoacetamide (0.5 µL of 0.95 M stock; 0.473 mMoles) were added, and the reactions incubated at room temperature before analysing by LC-MS ESI+.

Denatured + reduced cAbLys3-A104Dha LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-A104Dha +1 acetamide; expected mass 15820 Da, observed 15818.7 Da.

+2 acetamides; expected mass 15877 Da, observed 15877.6 Da.

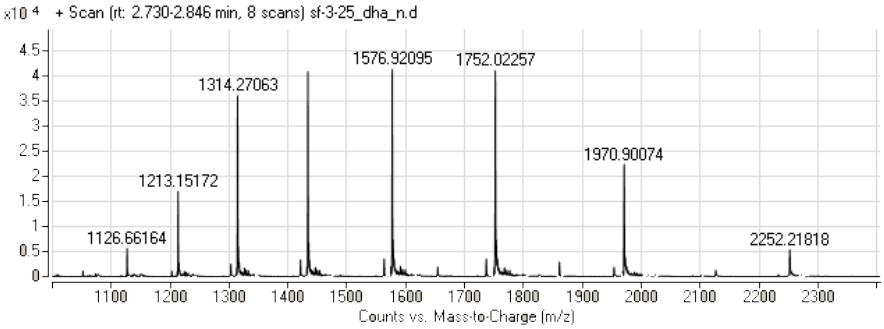
+3 acetamides; expected mass 15934 Da, observed 15934.5 Da.

+4 acetamides; expected mass 15991 Da, observed 15993.1 Da.

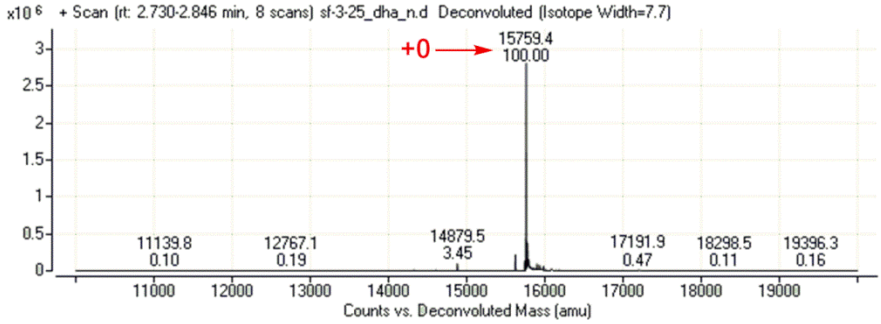
Fig. 4.4.5.3.1. *LCMS-ESI+ characterization of denatured + reduced cAbLys3-A104Dha. Also in fig. 4.2.5.1.A.*

<u>Native cAbLys3-A104Dha LCMS-ESI+ characterization</u>

Ion series



Deconvoluted mass spectrum

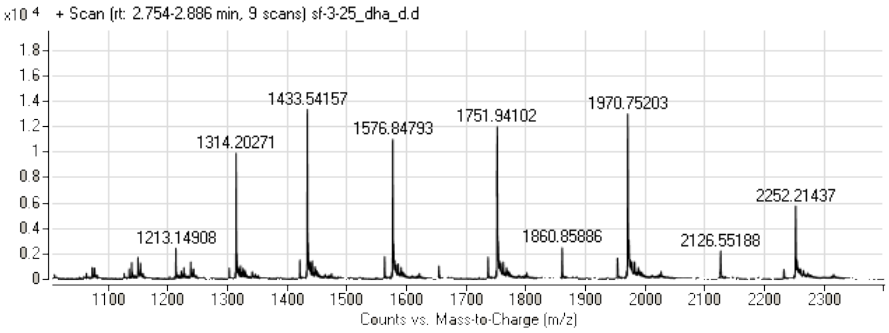


Starting material (cAb-Lys3-A104-Dha); expected mass 15763 Da, observed 15759.4 Da.

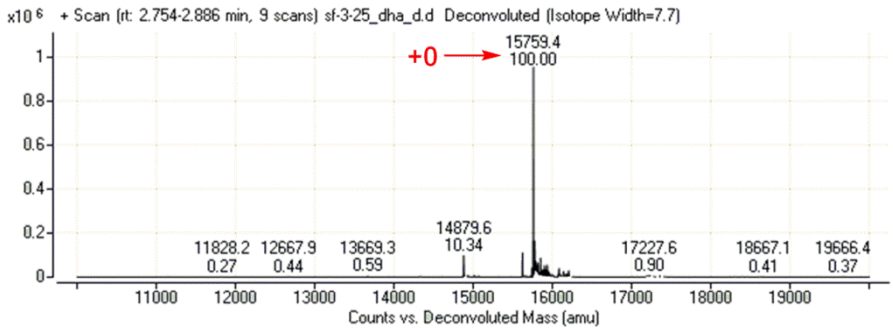
Fig. 4.4.5.3.2. *LCMS-ESI+ characterization of native cAbLys3-A104Dha. Also in fig. 4.2.5.1.A.*

<u>Denatured cAbLys3-A104Dha LCMS-ESI+ characterization</u>

Ion series



Deconvoluted mass spectrum



Starting material (cAb-Lys3-A104Dha); expected mass 15763 Da, observed 15759.4 Da

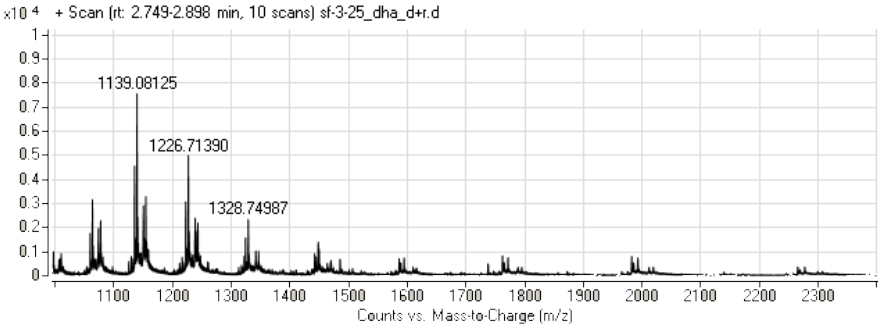
Fig. 4.4.5.3.3. *LCMS-ESI+ characterization of denatured cAbLys3-A104Dha. Also in fig. 4.2.5.1.A.*

4.4.5.4. Demonstrating retention of disulphide bonds in cAbLys after radical reaction.

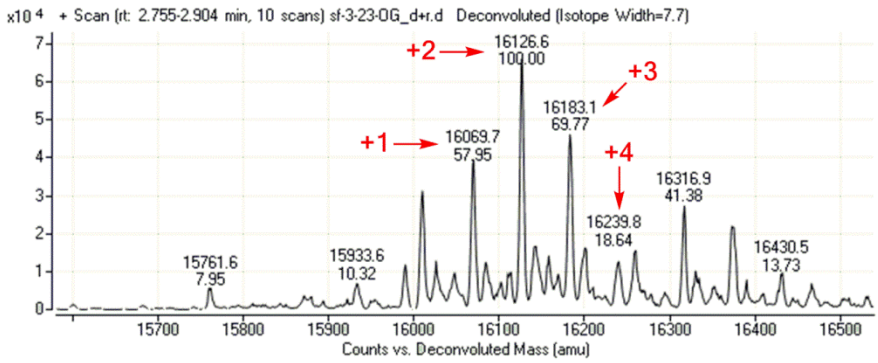
Following general procedure B, three aliquots of freshly prepared cAb-Lys3-A104-*O*-GlcNAc (10 µL of 1.55 mg/mL, 0.968 nMoles) were taken. After the 20 minute denaturation or incubation period, 855 equivalents of 2-iodoacetamide (0.875 µL of 0.95 M stock; 0.828 mMoles) were added, and the reactions incubated at room temperature before analysing by LC-MS ESI+.

Denatured + reduced cAbLys3-A104-*O*-GlcNAc *LCMS*-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-A104-*O*-GlcNAc +1 acetamide; expected mass 16069 Da, observed 16069.7 Da.

+2 acetamides; expected mass 16126 Da, observed 16126.6 Da.

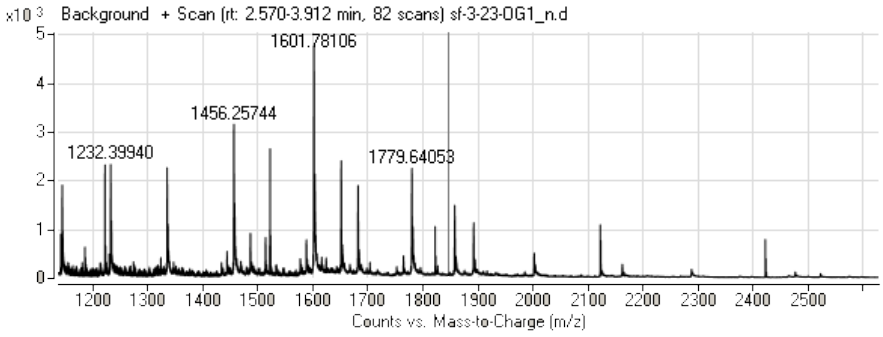
+3 acetamides; expected mass 16183 Da, observed 16183.1 Da.

+4 acetamides; expected mass 16240 Da, observed 16239.8 Da.

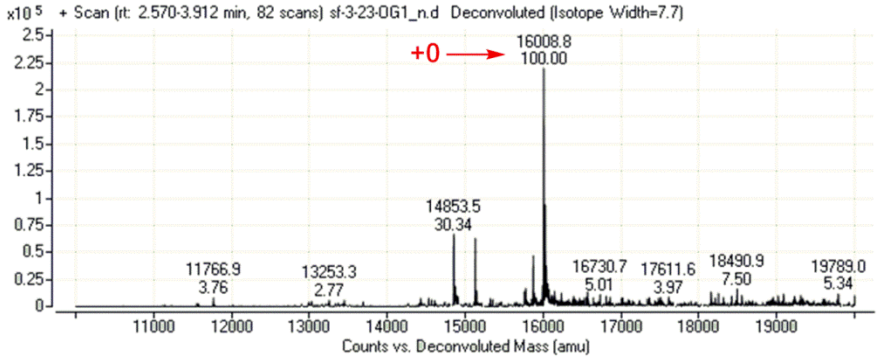
Fig. 4.4.5.4.1. *LCMS-ESI+ characterization of denatured + reduced cAbLys3-A104-*O*-GlcNAc. Also in fig. 4.2.5.1.A.*

Native cAbLys3-A104-*O*-GlcNAc *LCMS*-ESI+ characterization

Ion series



Deconvoluted mass spectrum

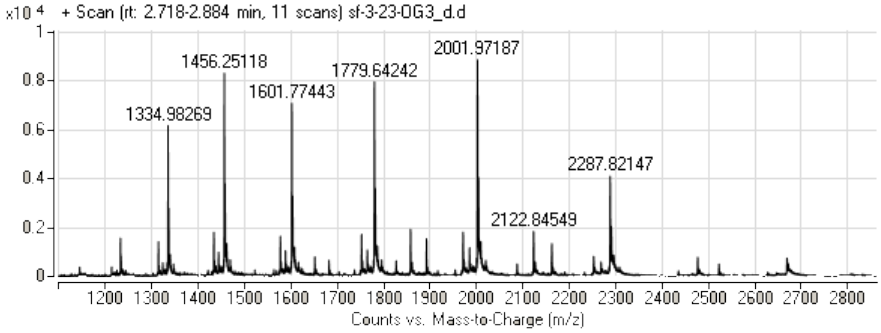


Starting material (cAb-Lys3-A104-*O*-GlcNAc); expected mass 16012 Da, observed 16008.8 Da.

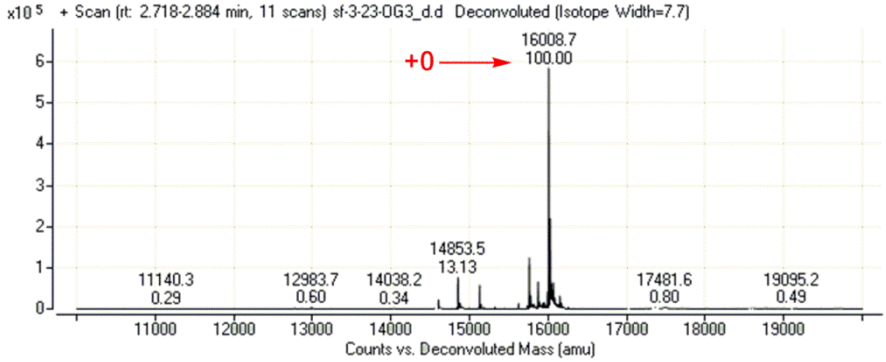
Fig. 4.4.5.4.2. *LCMS-ESI+ characterization of native cAbLys3-A104-*O*-GlcNAc. Also in fig. 4.2.5.1.A.*

Denatured cAbLys3-A104-*O*-GlcNAc *LCMS*-ESI+ characterization

Ion series



Deconvoluted mass spectrum



Starting material (cAb-Lys3-A104-*O*-GlcNAc); expected mass 16012 Da, observed 16008.7 Da.

Fig. 4.4.5.4.3. *LCMS-ESI+ characterization of denatured cAbLys3-A104-*O*-GlcNAc. Also in fig. 4.2.5.1.A.*

4.4.6 Enzymatic reactions with modified cAbLys3

4.4.6.1. Hydrolysis of cAb-Lys3-*O*-GlcNAc by jbOGA

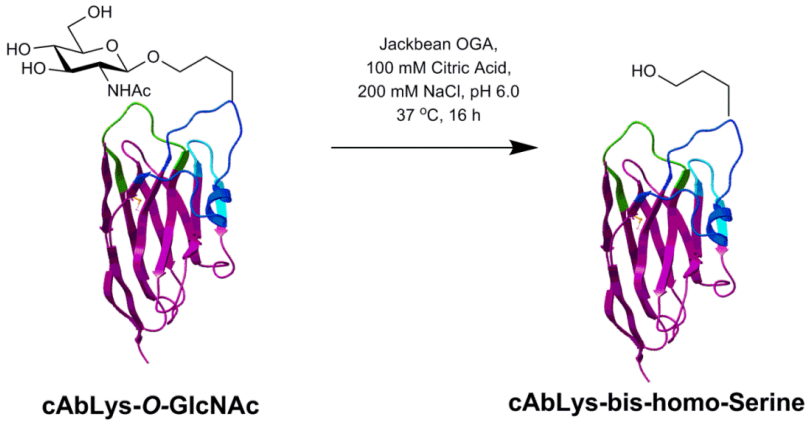
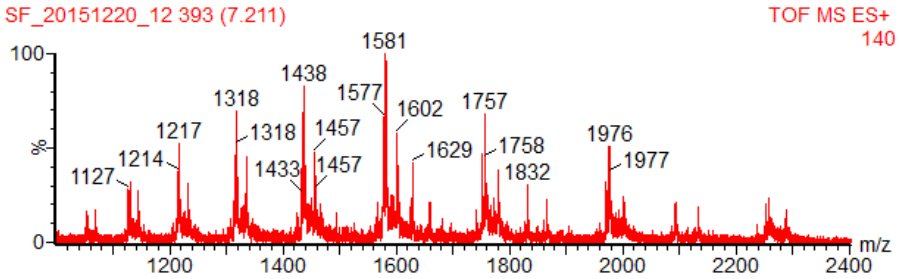


Fig. 4.4.6.1.1. *Scheme for jbOGA hydrolysis of cAbLys3-A104-*O*-GlcNAc.*

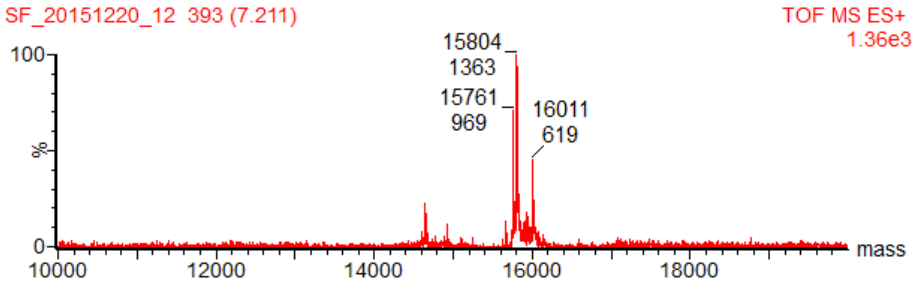
10 µL of 1 unit/mL β-N-Acetylglucosaminidase (from *Canavalia ensiformis*, Jack bean) solution in citrate buffer (100 mM Citric acid, 200 mM NaCl, pH 6.0) was added to 10 µL of 0.4 mg/mL cAb-Lys3-*O*-GlcNAc in citrate buffer. The reaction was incubated for 20 hours at 37˚C before quenching by adding 15 µL borate buffer to raise pH and analysing through LC-MS ESI+. Representative data from one reaction is shown below: for additional repeats, see appendix 4.6.2.1.

LCMS-ESI+ characterization of representative reaction

Ion series



Deconvoluted mass spectrum



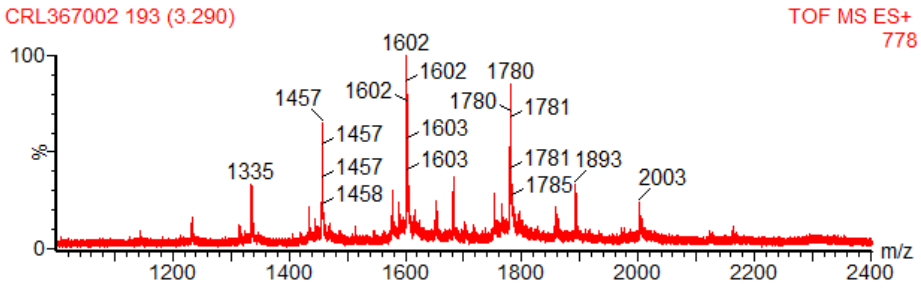
cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16011 Da.

cAb-Lys3-A104hhSer: expected mass 15808 Da, observed 15804 Da (46%).

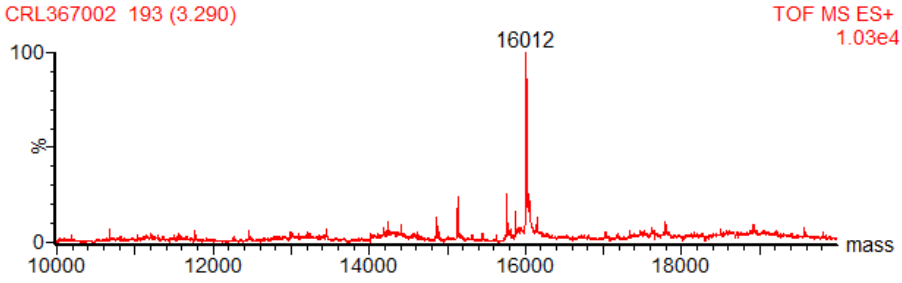
Fig. 4.4.6.1.2. *LCMS-ESI+ characterization data for jbOGA cleavage of cAbLys3-O-GlcNAc.*

<u>LCMS-ESI+ characterization of control (no jbOGA)</u>

Ion series



Deconvoluted mass spectrum



cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 160012 Da.

cAb-Lys3-A104hhSer: expected mass 15808 Da, not observed.

Fig. 4.4.6.1.3. *LCMS-ESI+ characterization data for no jbOGA control.*

4.4.6.2. Hydrolysis of cAb-Lys3-*O*-GlcNAc by X.manhotis OGA

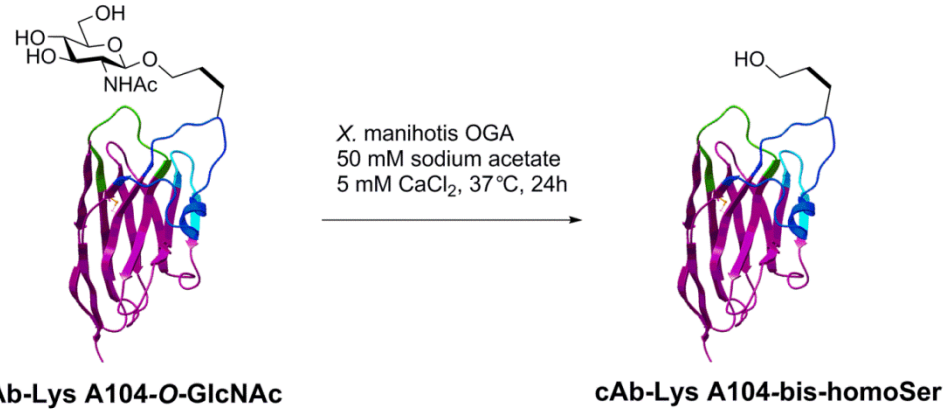
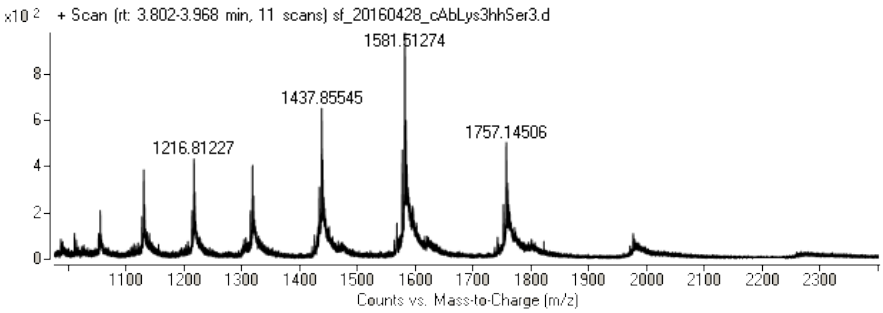


Fig. 4.4.6.2.1. *Scheme for xmOGA hydrolysis of cAbLys3-A104-*O*-GlcNAc.*

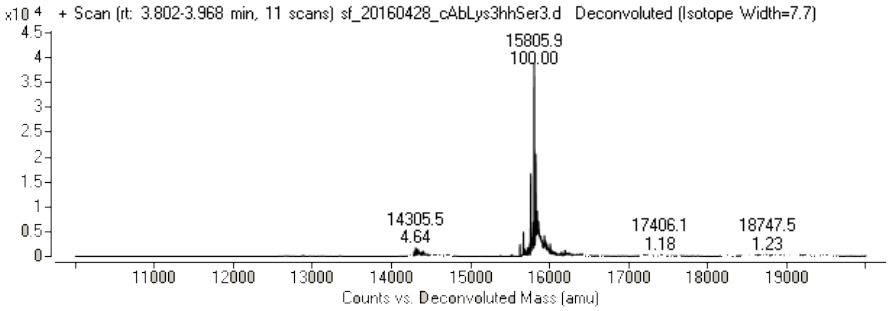
Small scale reactions

β-N-Acetylglucosaminidase (New England Biolabs) was used according to manufacturer's protocol. cAb-Lys3-*O*-GlcNAc (4.68 µL of 3.96 mg/ mL, 0.8 µg total) was added to G1 buffer (2 µL of 10 x stock in 12.62 µL MilliQ water) and BSA (0.2 µL of 100 x stock). 0.5 µL of OGA solution was then added (4 U/ µL stock, 2U total). This was then incubated for 37˚C for 24 hours. Samples were analysed after 24 h by LCMS-ESI+. Conversion was expected by taking the intensity of the cAb-Lys3-hhSer peak as a percentage of combined cAb-Lys3-*O*-GlcNAc (if present) and cAb-Lys3-hhSer peak intensities. The reaction was done in triplicate; representative data from one replicate is shown below, for additional data see appendix 4.6.2.2.

LCMS-ESI+ characterization: ion series



Deconvoluted mass spectrum



cAb-Lys3-A104-hhSer: expected mass 15808 Da, observed 15805.9 Da.

Fig. 4.4.6.2.2. *LCMS-ESI+ characterization data for xmOGA cleavage of cAbLys3-O-GlcNAc.*

4.4.6.3. Optimization of OGA concentration for large-scale reactions

β-N-Acetylglucosaminidase (New England Biolabs) was used according to manufacturer's protocol. cAb-Lys3-*O*-GlcNAc (4.68 µL of 3.96 mg/ mL, 0.8 µg total) was added to G1 buffer (2 µL of 10 x stock in 12.62 µL MilliQ water) and BSA (0.2 µL of 100 x stock). 0.5 µL of OGA solution was then added from one of the 4, 2, 1, 0.5 or 0.2 U/ µL stock solutions to give in total 2, 1, 0.5, 0.25 and 0.1 U in the five samples. These were then incubated for 37˚C for 24 hours. Samples were taken at 3h and 24h time points and analysed by LCMS-ESI+. Conversion was expected by taking the intensity of the cAb-Lys3-hhSer peak as a percentage of combined cAb-Lys3-*O*-GlcNAc and cAb-Lys3-hhSer peak intensities. Graph of results is shown below; for LCMS-ESI+ spectra used to calculate conversion, see appendix 4.6.2.3.

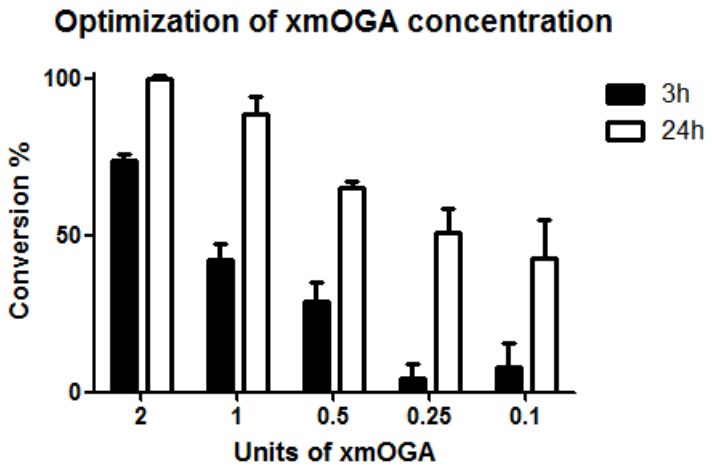


Fig. 4.4.6.3.1. *Results of X. manhotis OGA reaction optimization. Error bars represent the error of measurement, and are taken from the standard error of the mean (SEM) of values calculated from ESI-MS from each third of the protein peak in the LC chromatogram.*

4.4.6.4. Hydrolysis of cAb-Lys3-*O*-GlcNAc by human OGA

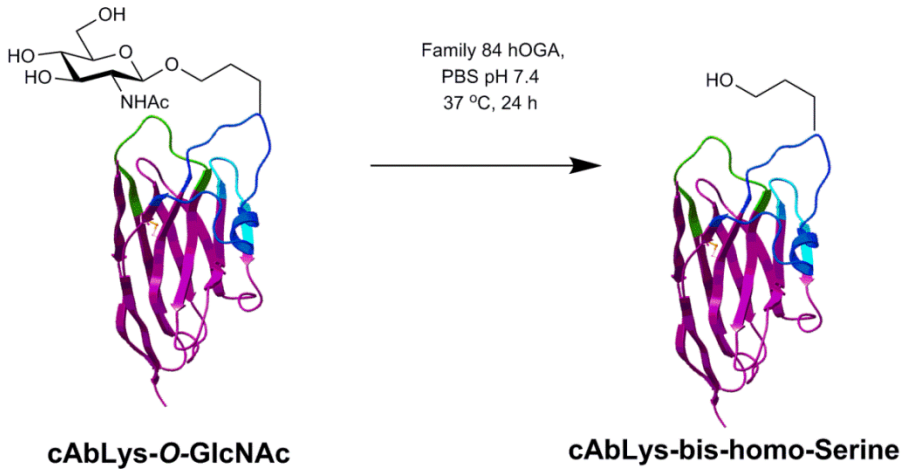
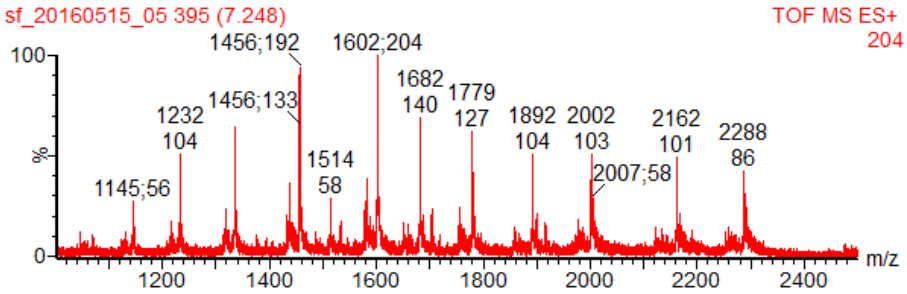


Fig. 4.4.6.4.1. *Scheme for the hydrolysis of cAbLys3-*O*-GlcNAc by hOGA*

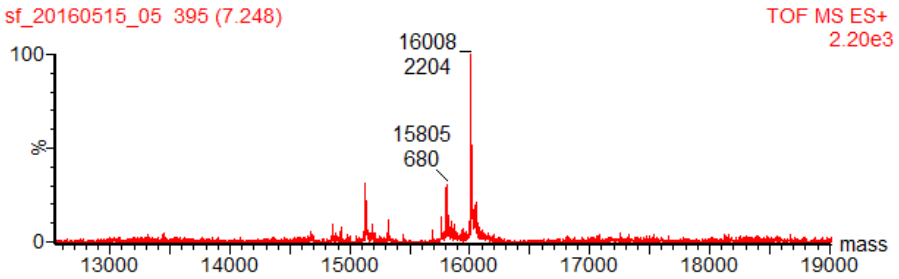
cAb-Lys3-A104-*O*-GlcNAc (20.0 µL, 1.0 mg/mL) was treated with Family 84 hOGA (0.5 µL, 10 mg/mL) in PBS pH 7.4. The reaction mixture was shaken at 25 °C for 2 h. A second dose of hOGA was added (0.5 µL, 10 mg/mL) then shaken at 25 °C for 24 hours. Samples were taken and analysed by LCMS-ESI+ at 3 hour and 24 hour timepoints. The reaction was performed in triplicate; representative data from one replicate is shown below. For additional data see appendix 4.6.2.4.

24 hour timepoint LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-*O*-GlcNAc: expected mass 16012 Da, observed 16008 Da.

cAb-Lys3-hhSer: expected mass 15808 Da, observed 15805 Da. (26.2%)

Fig. 4.4.6.4.2. *LCMS-ESI+ characterization data for hydrolysis of cAbLys3-O-GlcNAc by hOGA. Also in fig. 4.2.6.1.A.*

4.4.6.5. Attempted PNGase cleavage of cAb-Lys3-*N*-GlcNAc

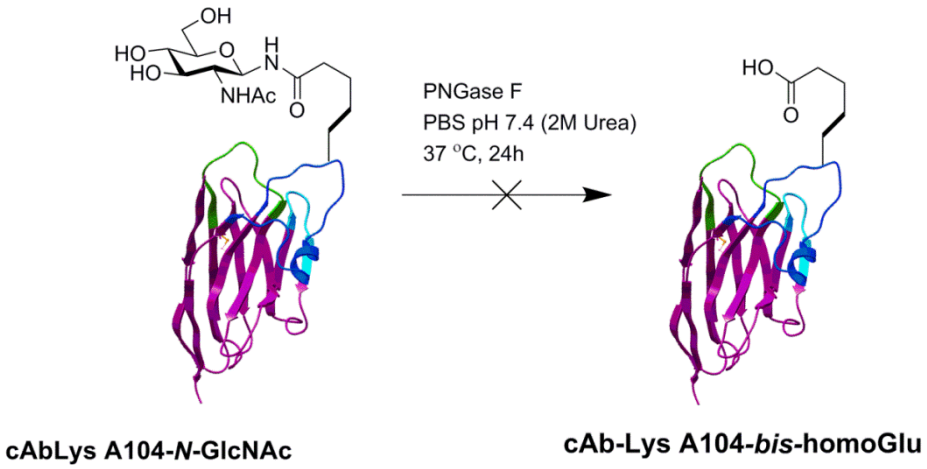
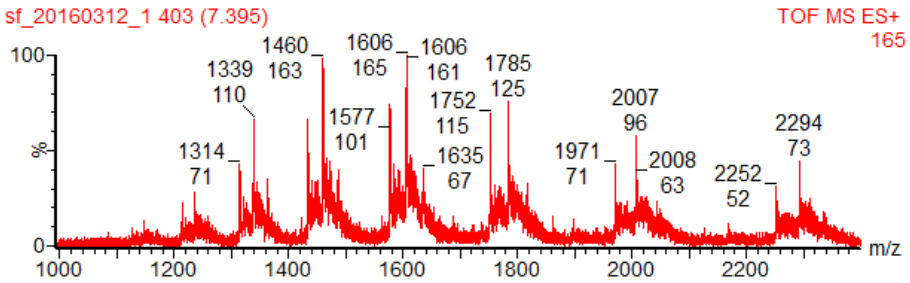


Fig. 4.4.6.5.1. *Scheme for attempted hydrolysis of cAbLys3-*N*-GlcNAc by PNGase F*

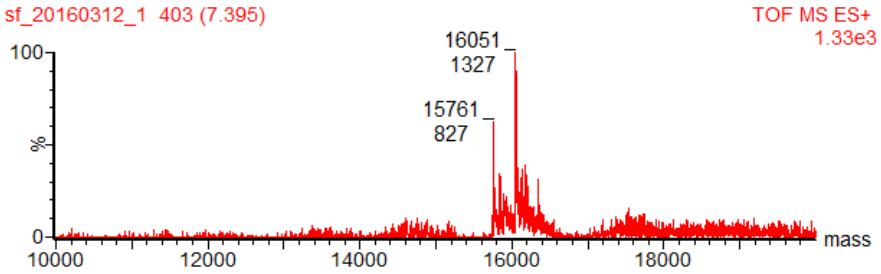
cAb-Lys3-*N*-GlcNAc (0.9 mg/ mL, 15 µL in PBS pH 7.4) was combined with 5 µL of either 8M urea (denaturing conditions, 2M final concentration) or MilliQ water (non-denaturing conditions). 25 U of PNGase F (0.5 µL of 50 U/ µL, New England Biolabs) was then added to each reaction, and these were then incubated at 37˚C for 24 hours. Samples were analysed by LCMS-ESI+.

Denaturing conditions: LCMS-ESI+ characterization

Ion series



Deconvoluted mass specturm



cAb-Lys3-*N*-GlcNAc: expected mass 16049 Da, observed mass 16051 Da.

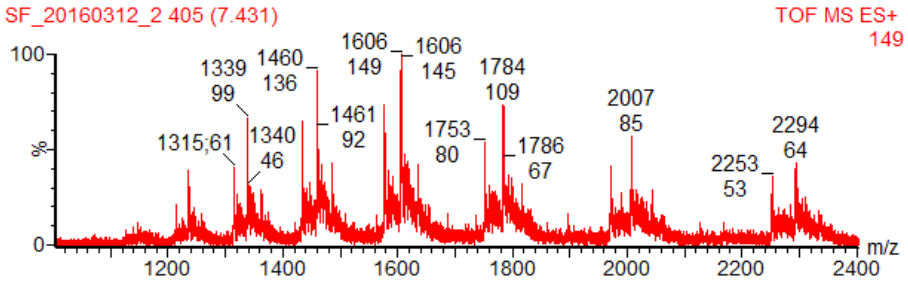
cAb-Lys3-A_{rac}: expected mass 15765 Da, observed 15761 Da.

cAb-Lys3-hhhGlu: expected mass 15885, no significant peaks above baseline observed.

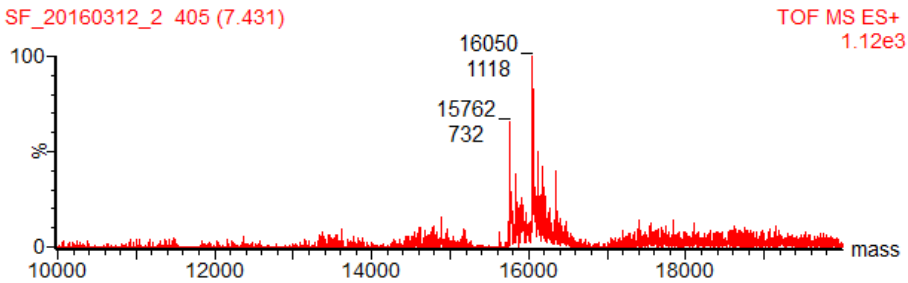
Fig. 4.4.6.5.2. *LCMS-ESI+ characterization data for attempted PNGase F cleavage of cAbLys3-*N*-GlcNAc under denaturing conditions.*

<u>Non-denaturing conditions: LCMS-ESI+ characterization</u>

Ion series



Deconvoluted mass spectrum



cAb-Lys3-*N*-GlcNAc: expected mass 16049 Da, observed mass 16050 Da.

cAb-Lys3-A_{rac}: expected mass 15765 Da, observed 15762 Da. Also in fig. 4.2.8.1.

cAb-Lys3-hhhGlu: expected mass 15885, no significant peaks above baseline observed.

Fig. 4.4.6.5.3. *LCMS-ESI+ characterization data for attempted PNGase F cleavage of cAbLys3-*N*-GlcNAc under denaturing conditions.*

4.4.6.6. PNGase F activity control with Herceptin

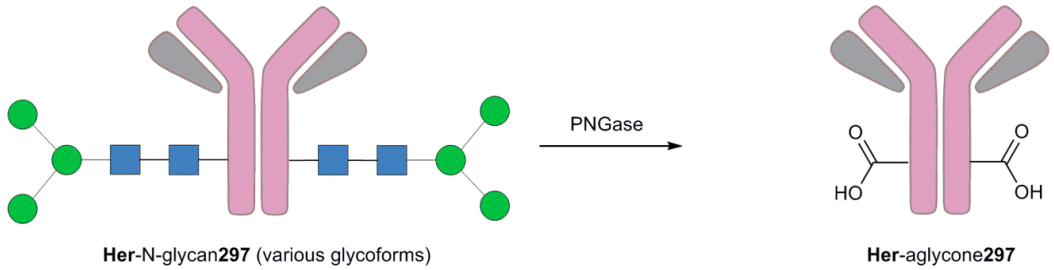
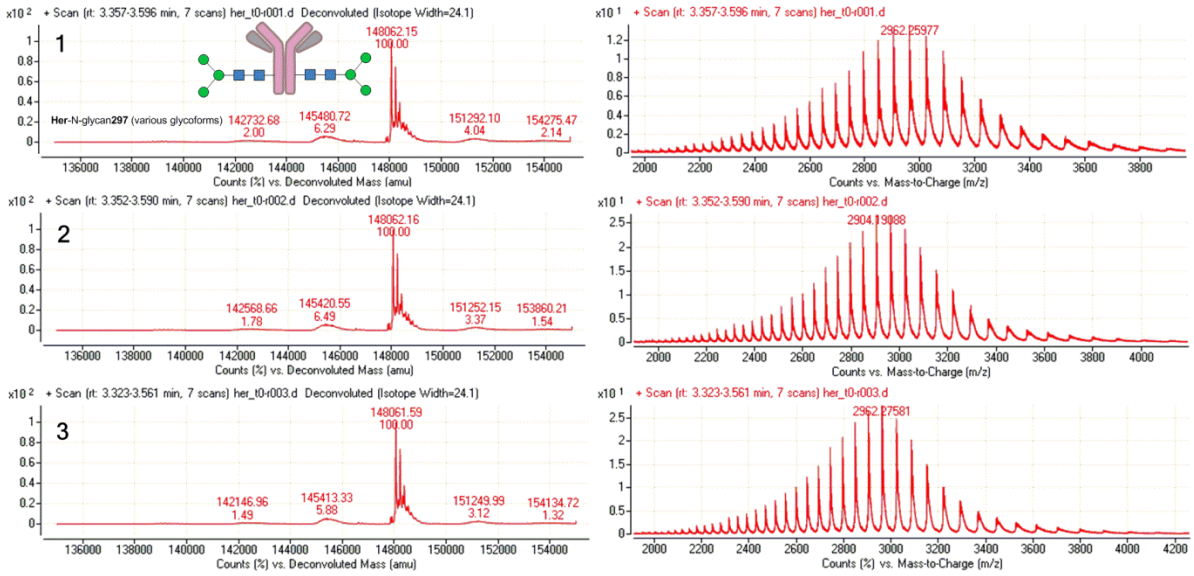


Fig. 4.4.6.6.1. *Scheme for hydrolysis of Herceptin glycans by PNGase F*

In triplicate, herceptin (5.0 µL, 4.0 mg/mL) was mixed with GlycoBuffer 2 (2.0 µL), MilliQ H_2O (11.0 µL) and PNGase F (2 µL, 500 U/µL) and incubated at 37°C for 24 hours. Samples were taken at given time-points and analysed by intact LCMS-ESI+. Full conversion was observed after 1 h; data for this timepoint is shown here. Data for additional timepoints can be found in appendix 4.6.2.5.

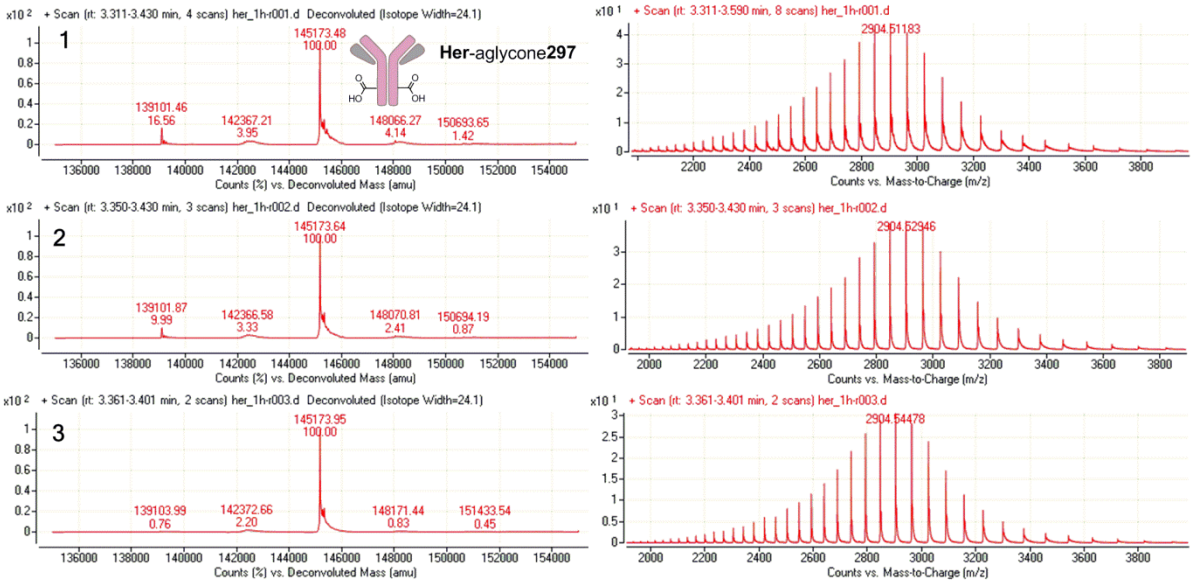
t0 (before PNGase F added) LCMS-ESI+ characterization



Her-*N*-glycan297 expected mass: 148057, observed: 148062

Fig. 4.4.6.6.2. *LCMS-ESI+ deconvoluted mass spectra (left) and ion series (right) for Herceptin at timepoint t0 (no PNGase F).*

t1h LCMS-ESI+ characterization



Her-aglycone297 expected mass: 145166, observed mass:145174

Fig. 4.4.6.6.3. *LCMS-ESI+ deconvoluted mass spectra (left) and ion series (right) for*

Herceptin at timepoint t1h; 1 hour after PNGase F added. Also in fig. 4.2.8.1.

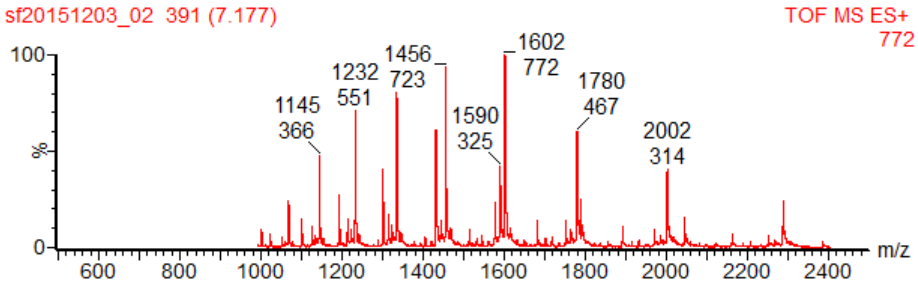
4.4.6.7. Test reaction to determine whether HEL is hydrolysing cAb-Lys3-*O*-GlcNAc in ELISAs.

10 µL of 0.2 mg/ mL HEL in PBS pH 7.4 was added to cAb-Lys3-A104GlcNAc (10 µL of 0.5 mg/mL in PBS pH 7.4) and incubated for 2 hours at room temperature with shaking at 550 rpm. After this time the reaction was analysed by LC-MS ESI+.

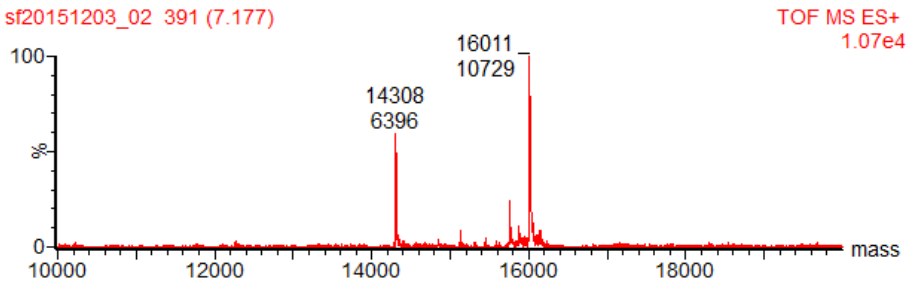
ESI-MS for the reaction and control are shown below.

Test reaction LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16011 Da.

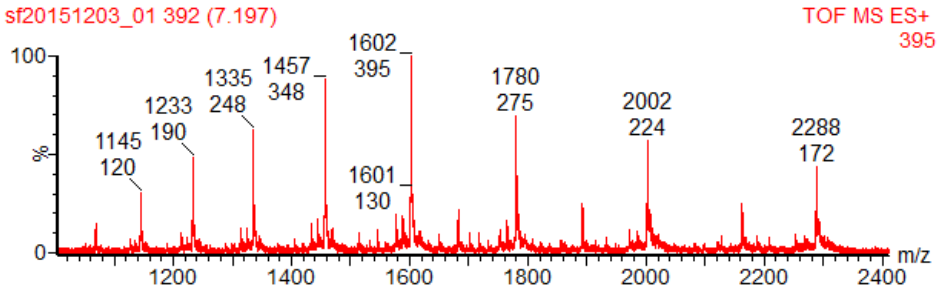
cAb-Lys3-A104hhSer: expected mass 15808 Da, none observed.

HEL: expected mass 14307Da, observed 14308 Da.

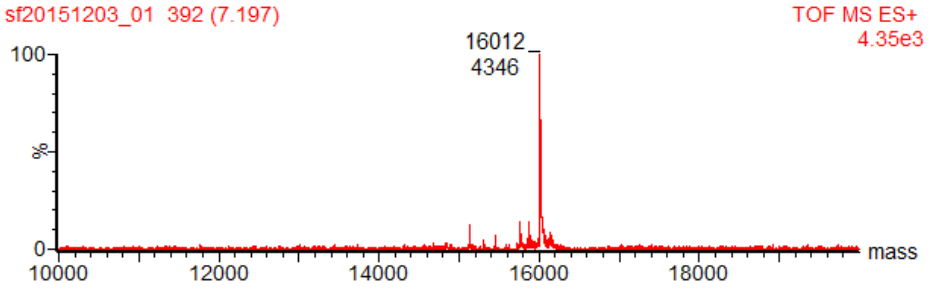
Fig. 4.4.6.7.1. *LCMS-ESI+ data for cAbLys3-O-GlcNAc with HEL.*

Control LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16012 Da.

cAb-Lys3-A104hhSer: expected mass 15808 Da, none observed.

Fig. 4.4.6.7.2. *LCMS-ESI+ data for cAbLys3-O-GlcNAc control (no HEL).*

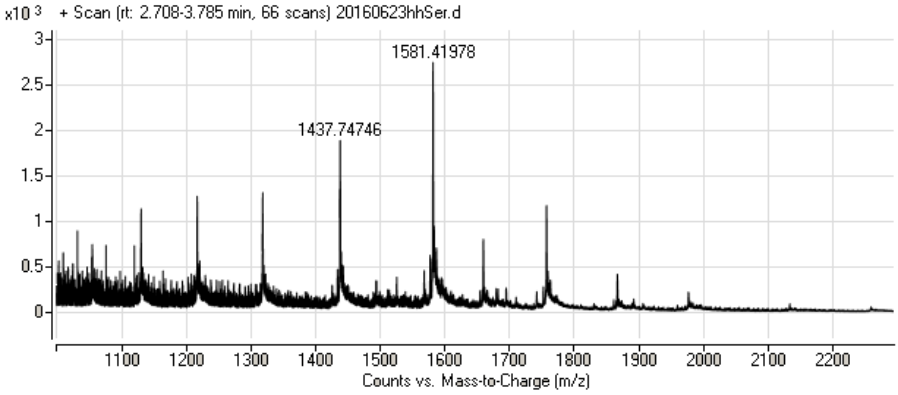
4.4.7. Preparation and purification of cAbLys3-*bis*-homoSer for ELISA

4.4.7.1. Large-scale preparation of cAb-Lys3-*bis*-homoSer

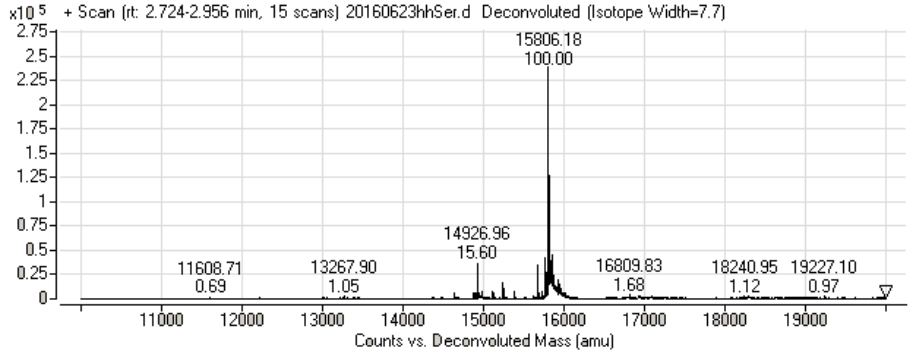
The protocol for hydrolysis of cAb-Lys3-O-GlcNAc by X.manhotis OGA was followed with modifications. In triplicate, cAb-Lys3-*O*-GlcNAc (0.5 mL of 0.78 mg/mL, 24 nMol in G1 buffer) was combined with BSA (15 µL of 100 x stock), OGA solution (24 µL), and G1 buffer (961 µL) in a total volume of 1.5 mL. After 24 hours the reaction was confirmed to have gone to completion by LCMS-ESI+ analysis. A precipitate was observed to have formed, which re-dissolved into solution when the reaction was shaken. The three reactions were combined and

concentrated to 2 mL using a Vivaspin 5kDa mwco centrifuge concentrator (Sartorius) before purifying by size exclusion chromatography.

LCMS-ESI+ characterization: ion series



Deconvoluted mass spectrum



cAb-Lys3-*bis*-homoSer expected mass: 15808 Da; observed 15806.18 Da.

Fig. 4.4.7.1.1. *LCMS-ESI+ characterization of large-scale cAbLys3-bis-homoSer preparation.*

4.4.7.2. Purification of cAb-Lys3-*bis*-homoSerine

A size exclusion chromatography column (GE Superdex 75 10/300 Gl) was equilibrated with 1.1 column volume PBS pH 7.4. 2mL cAb-Lys3-*bis*-homoSerine preparation was then loaded onto the column using a 2 mL sample loop, and eluted over 1.1 CV. Protein-containing fractions as identified from the UV trace were analysed by SDS-PAGE; those containing pure

cAb-Lys3-*bis*-homoSerine were combined and concentrated using a Vivaspin 5kDa mwco centrifuge concentrator (Sartorius) to a final volume of approximately 500 µL.

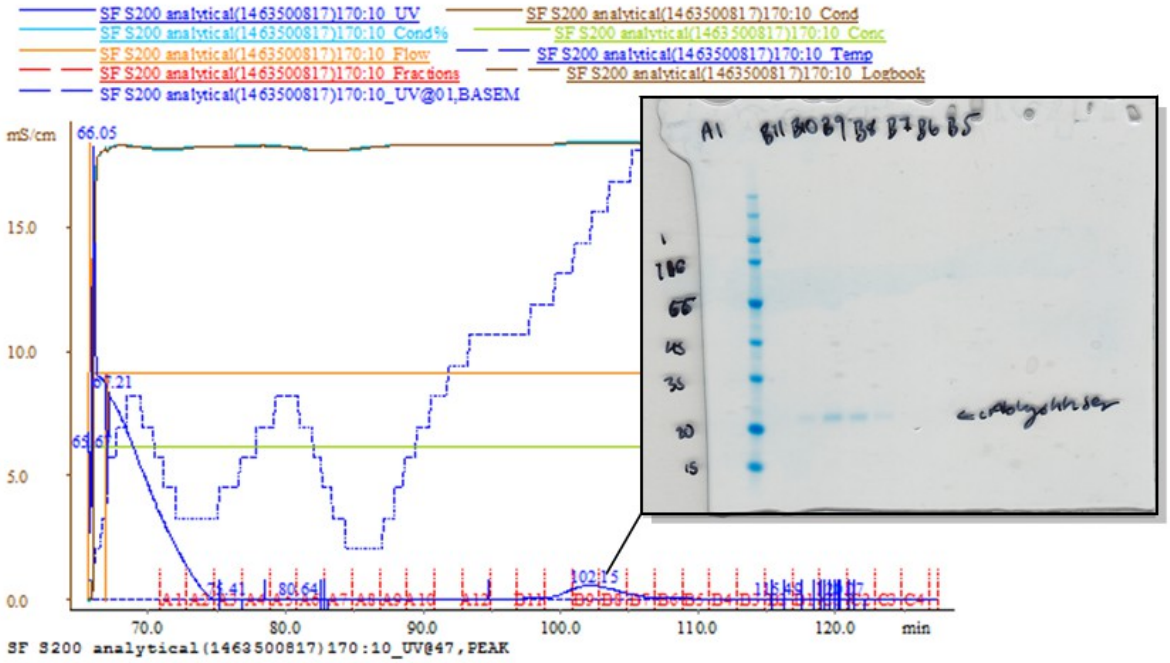


Fig. 4.4.7.2.1. *SEC UV trace and SDS-PAGE characterization for cAbLys3-bis-homoSer purification.Reproduced from fig. 4.2.7.1.*

4.4.8. Enzyme-Linked Immunoabsorbent Assays (ELISAs)

4.4.8.1. General procedure B: ELISAs

Hen-egg lysozyme (HEL; 50 µL of a solution in PBS pH 7.4) was added to each well of a 96-well plate bar 3 for a no HEL negative control. The plate was incubated at 4˚C overnight, then washed three times with PBST (PBS pH 7.4 with 0.05% Tween20; 200 µL per well). Blocking was performed by adding Bovine Serum Albumin (BSA) in PBST (200 µL per well) and incubating at room temperature for 1 hour with shaking at 550 rpm. The plate was washed once with PBST (200 µL per well). Primary antibody was then added in a dilution series into 1% BSA in PBST in triplicate unless otherwise stated. 1% BSA in PBST (50 µL per well) was added to wells with no primary antibody. The primary antibody was then incubated at room

temperature for 2 hours with shaking at 550 rpm. The plate was washed three times with PBST (200 μL per well). Secondary antibody was then added to all wells bar three, which were reserved for a no secondary antibody negative control; 1% BSA in PBST (50 μL per well) was added to these. The plate was then incubated at room temperature with shaking at 550 rpm for 1 hour before washing five times with PBST (200 μL per well). The appropriate developing substrate was then added. When the plate was sufficiently developed, quenching was carried out if necessary and absorbance read at the appropriate wavelength using a platereader.

4.4.8.2. Optimization of conditions for ELISAs

Initial ELISA with pNPP substrate system

General procedure B was followed with modifications. A standard 96-well plate was used; HEL for plate coating was used at a concentration of 20 μg/mL to give a total of 1 μg per well, and 2% BSA solution was used for blocking. The initial concentration of primary antibody was 500 μg/ mL, and dilution was done in a 5x series across 10 wells. Primary antibodies studied were cAb-Lys3-WT, -A104Dha and –A104-*O*-GlcNAc. The secondary antibody used was Anti-polyHistidine-alkaline phosphatase conjugate in a 1:1000 dilution, and the development substrate was Alkaline Phosphatase Yellow (pNPP) Liquid Substrate System; when sufficient yellow colour had developed in some wells absorbance was read at 405 nm.

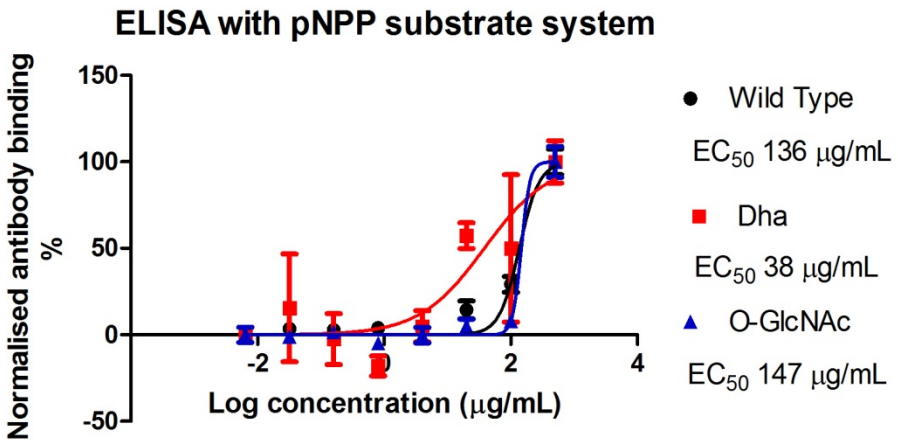


Fig. 4.4.8.2.1 *Optimization.ELISA with cAbLys3 WT, -DHa and -O-GlcNAc using pNPP substrate system. Reproduced from fig. 4.2.9.1.A.*

Initial ELISA with TMB substrate system

General procedure B was followed with modifications. A standard 96-well plate was used; HEL for plate coating was used at a concentration of 20 µg/mL to give a total of 1 µg per well, and 2% BSA solution was used for blocking. Primary antibodies studied were cAb-Lys3-WT and –A104-*O*-GlcNAc. The initial concentration of primary antibody was 100 µg/ mL, and dilution was done in a 5x series across 10 wells. The secondary antibody used was HRP anti-His tag antibody IgG1 (BioLegend) in a 1:10000 dilution and the development substrate was TMB substrate set (BioLegend); development was quenched using 50 µL 1M HCl before absorbance at 450 nm was read.

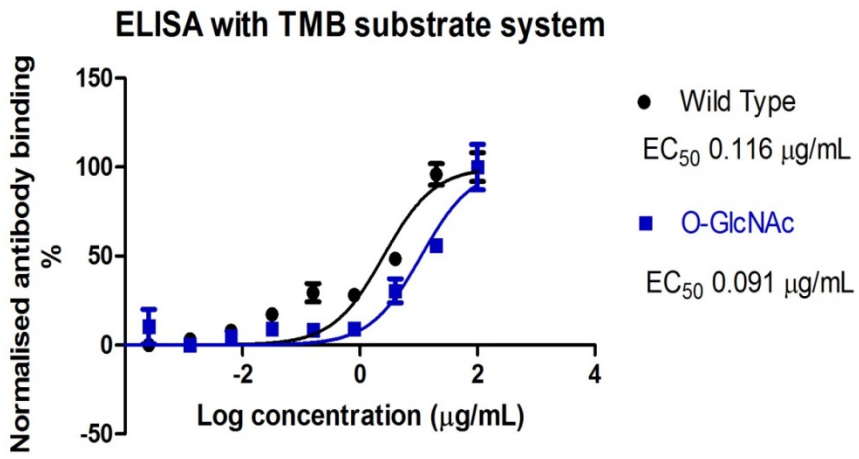


Fig. 4.4.8.2.2 *Optimization.ELISA with cAbLys3 WT, and -O-GlcNAc using TMB substrate system. Reproduced from fig. 4.2.9.1.B.*

ELISA to optimize HEL concentration

General procedure B was followed with modifications. A 96 well high-binding ELISA plate was used. Wells were coated with either 1 µg HEL (50 µL of 20 mg/ mL) or 0.1 µg HEL (50 µL of 20mg/ mL) in the initial step. cAb-Lys3-WT was used as the primary antibody for all wells; dilution series were done in triplicate for each HEL concentration (six in total). The initial concentration of primary antibody was 100 µg/ mL, and dilution was done in a 5x series across 10 wells. The secondary antibody used was HRP anti-His tag antibody IgG1 (BioLegend) in a 1:10000 dilution and the development substrate was TMB substrate set (BioLegend); development was quenched using 50 µL 1M HCl before absorbance at 450 nm was read.

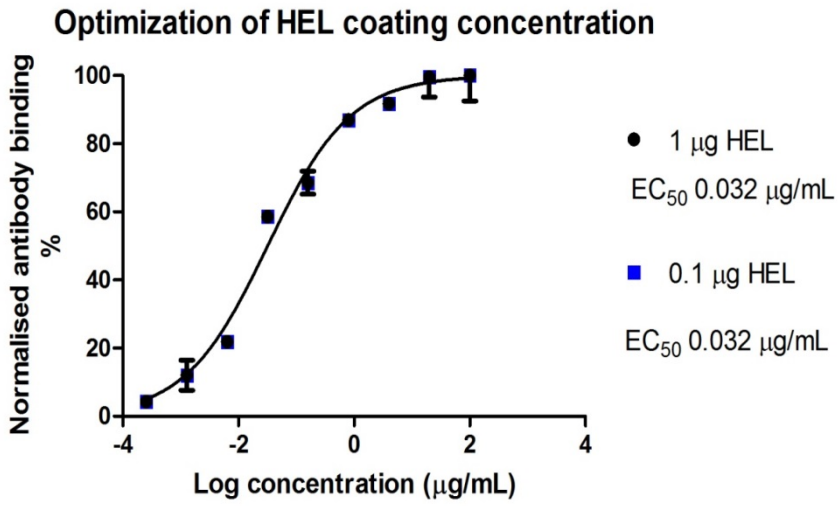


Fig. 4.4.8.2.3. *Optimization.ELISA with cAbLys3 WT, and -O-GlcNAc using TMB substrate system. Reproduced from fig. 4.2.9.1.C.*

ELISA to optimize BSA% in blocking

General procedure B was followed with modifications. A 96 well high-binding ELISA plate was used; HEL for plate coating was used at a concentration of 2 μg/mL to give a total of 0.1 μg per well. Blocking was done with 2% BSA in PBST (triplicate), 3% BSA in PBST (triplicate) or 4% BSA in PBST (duplicate). cAb-Lys3-WT was used as the primary antibody for all wells; dilution series were done in triplicate for each HEL concentration (six in total). The initial concentration of primary antibody was 100 μg/ mL, and dilution was done in a 5x series across 10 wells. The secondary antibody used was HRP anti-His tag antibody IgG1 (BioLegend) in a 1:10000 dilution and the development substrate was TMB substrate set (BioLegend); development was quenched using 50 μL 1M HCl before absorbance at 450 nm was read.

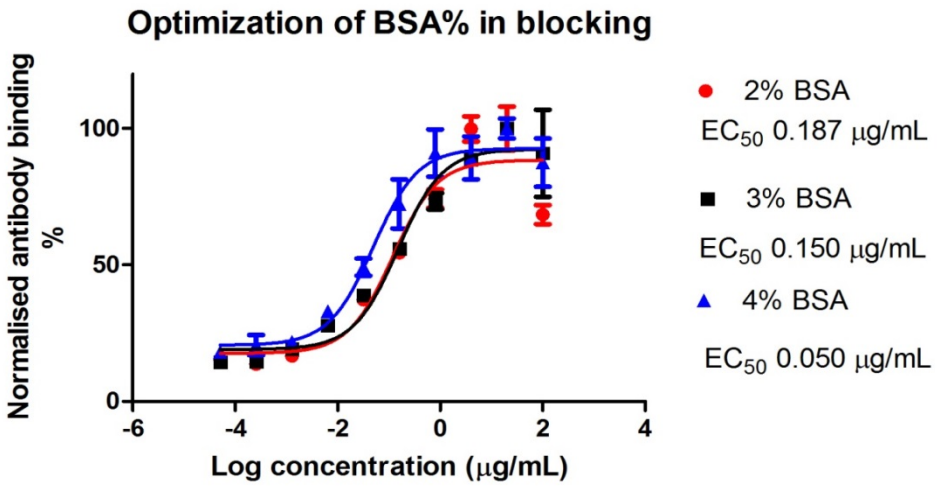


Fig. 4.4.8.2.3. *Optimization.ELISA with cAbLys3 WT varying BSA %. Reproduced from fig. 4.2.9.1.D.*

ELISA with controls cAb-Lys3-$WT_{Untreated,}$ -$WT_{Treated}$, -A_{rac} and –$WT_{Treated + OGlcNAc}$

General procedure B was followed. All samples were thawed at the same time and stored at 4˚C for approximately 2 weeks. As the measured concentration ratio of –$WT_{Treated + OGlcNAc}$ with the other three samples was markedly lower by nanodrop compared to BCA, the sample was dialysed (x dialysis kit) for 4 hours into PBS pH 7.4 to remove excess *O*-GlcNAc before repeating concentration measurements. After dialysis the concentration ratio by nanodrop had increased and was consistent with BCA measurement. Two plates were needed to accommodate all samples; cAb-Lys3-$WT_{Untreated,}$ -$WT_{Treated}$ were on one plate, and -A_{rac} and –$WT_{Treated + OGlcNAc}$ were on another.

HEL for plate coating was used at a concentration of 2 µg/mL to give a total of 0.1 µg per well, and 3% BSA solution was used for blocking. The initial concentration of primary antibody was 100 µg/ mL, and dilution was done in a 5x series across 10 wells. The secondary antibody used was HRP anti-His tag antibody IgG1 (BioLegend) and the development substrate was TMB substrate set (BioLegend); development was quenched using 50 µL 1M HCl before absorbance at 450 nm was read.

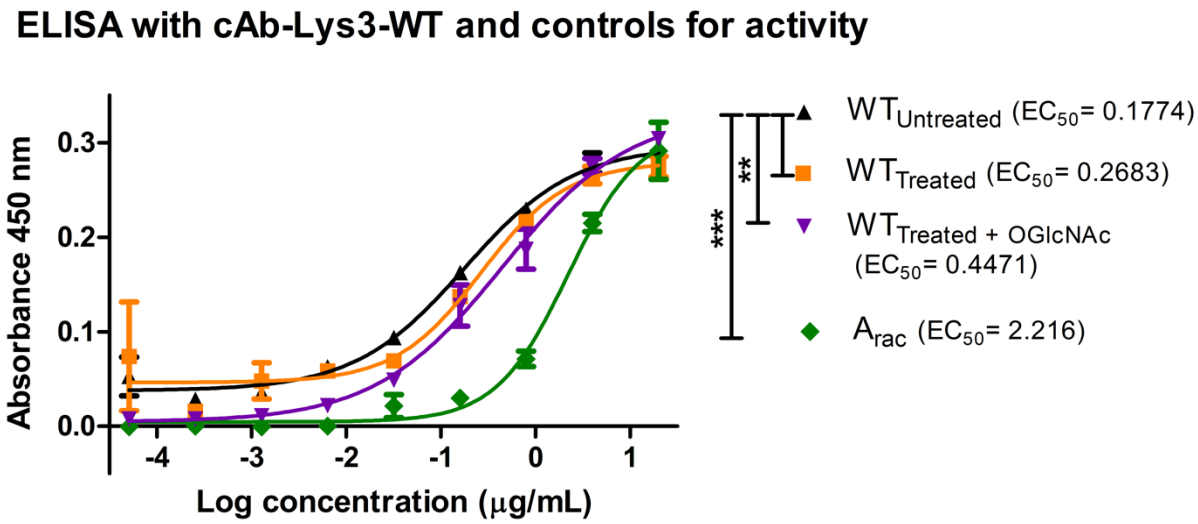


Fig. 4.4.8.2.4. *ELISA with controls cAb-Lys3-$WT_{Untreated,}$ -$WT_{Treated}$, -A_{rac} and –$WT_{Treated\ +\ OGlcNAc.}$ Reproduced from fig. 4.2.9.2.B.*

ELISA with cAb-Lys3-WT, -*O*-GlcNAc and –*bis*-homoSerine.

General procedure B was followed. HEL for plate coating was used at a concentration of 2 μg/mL to give a total of 0.1 μg per well, and 3% BSA solution was used for blocking. The initial concentration of primary antibody was 250 μg/ mL, and dilution was done in a 5x series across 7 wells. The secondary antibody used was HRP anti-His tag antibody IgG1 (BioLegend) and the development substrate was TMB substrate set (BioLegend); development was quenched using 50 μL 1M HCl before absorbance at 450 nm was read.

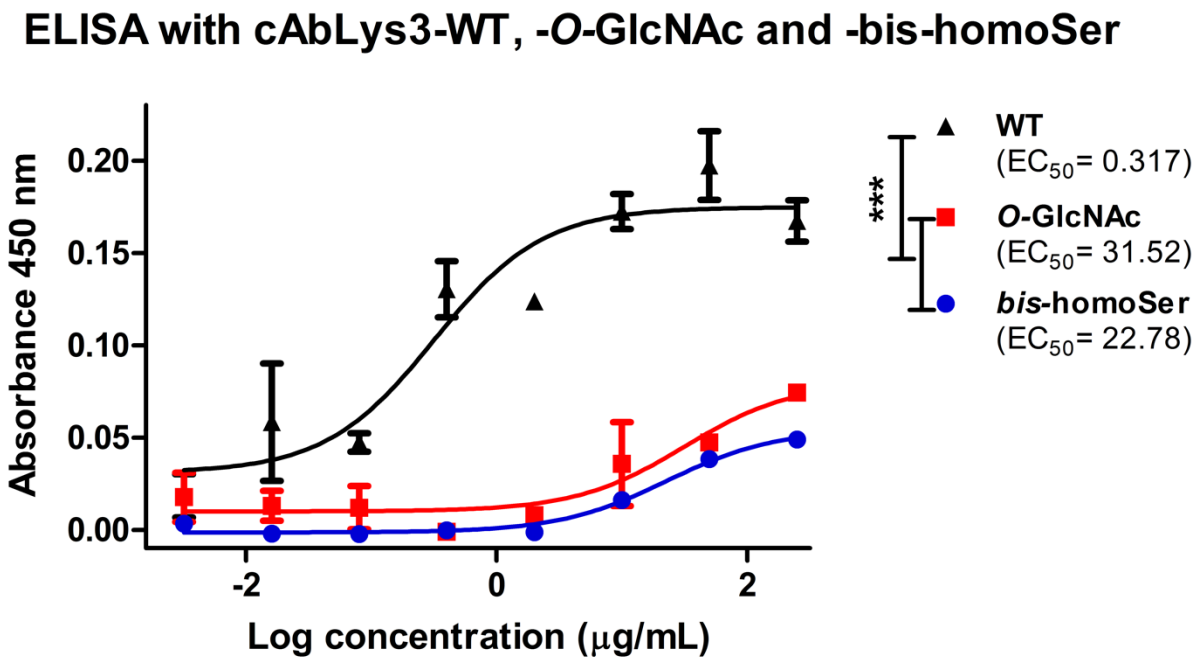


Fig. 4.4.8.2.5. *ELISA with cAb-Lys3-WT, -O-GlcNAc and –bis-homoSerine. Reproduced from fig. 4.2.9.2.A.*

4.4.9. Testing addition of decasaccharide to cAb-Lys3-*O*-GlcNAc using EndoA and EndoS

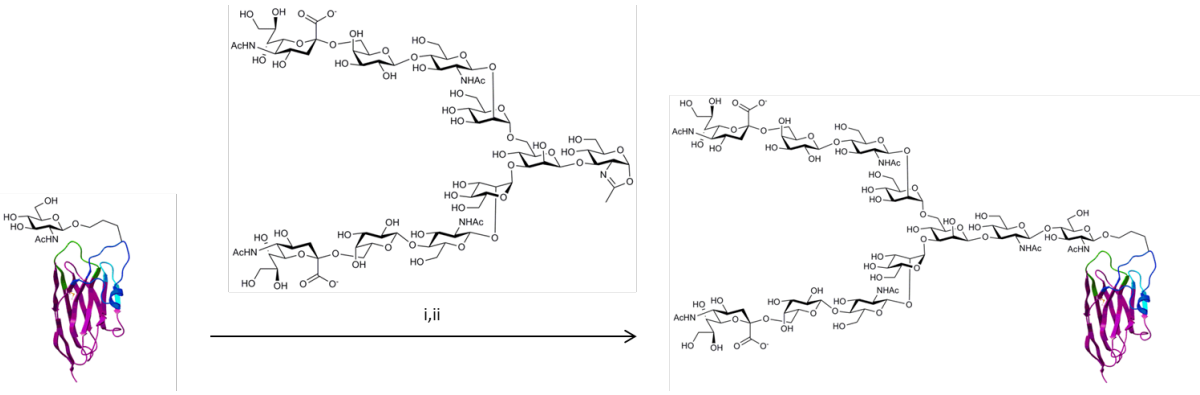


Fig. 4.4.9.1. *Scheme for the addition of decasaccharide to cAbLys3-O-GlcNAc using (i) Endo A Y299F or (ii) Endo S D233Q*

i) 4.4.9. **<u>EndoA:</u>** cAb-Lys3-*O*-GlcNAc (20 µL of a 0.3 mg/mL stock in PBS pH 6.0) was mixed with decasaccharide oxazoline (7.2 µL of a 10 mg/mL stock; synthesized

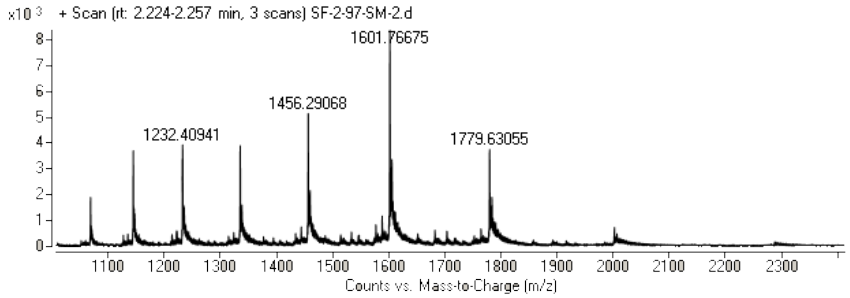
by M. Richard), and PBS pH 6.0 (21.8 µL) to a final volume of 50 µL. EndoA Y299F (1 µL of a 0.5 mg/mL stock; prepared in the Davis group) was then added. The reaction was incubated at 37 °C and 8 µL injected directly onto an LCMS-ESI+ for analysis at 1.5, 3 and 18 hr timepoints.

ii) **EndoS:** cAb-Lys3-*O*-GlcNAc (13 µL of a 1.53 mg/mL stock in PBS pH 7.4; 20 µg total) was mixed with 4 µL 500 mM sodium phosphate buffer pH 6.5, one aliquot of the decasaccharide oxazoline (2 µL of a 10 mg/ mL stock) and one aliquot of EndoS D233Q (0.6 µL of a 1.67 mg/mL stock; from M. Richard). The reaction was incubated at 30 °C. One additional aliquot of decasaccharide oxazoline was added every 5 minutes over a period of 100 minutes (19 additions in total). After 50 minutes an additional aliquot of EndoS was added. Analysis was performed by LCMS-ESI+. As no product was observed after 120 minutes, an additional 12 µL of EndoS was added and the reaction left overnight.

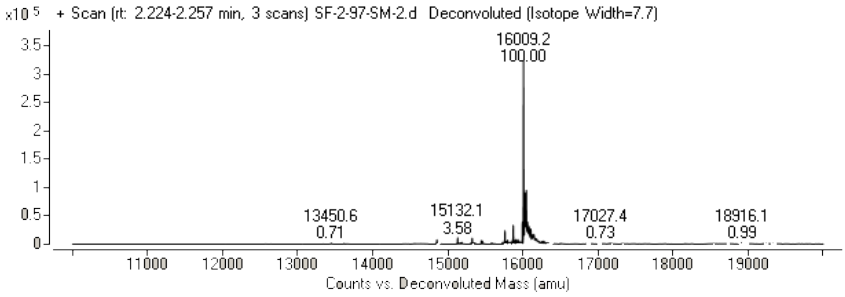
LCMS-ESI+ spectra for starting material and final timepoints are shown below. For additional timepoints, see appendix 4.6.3.

i. EndoA Y299F reaction; starting material LCMS-ESI+ characterization

Ion series



Deconvoluted ion series

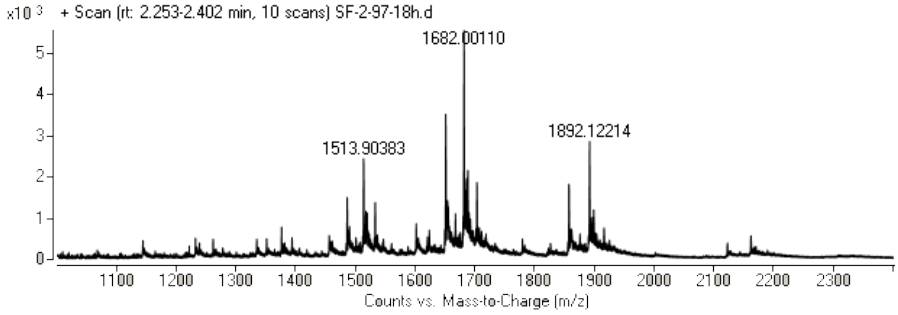


cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16009.2 Da.

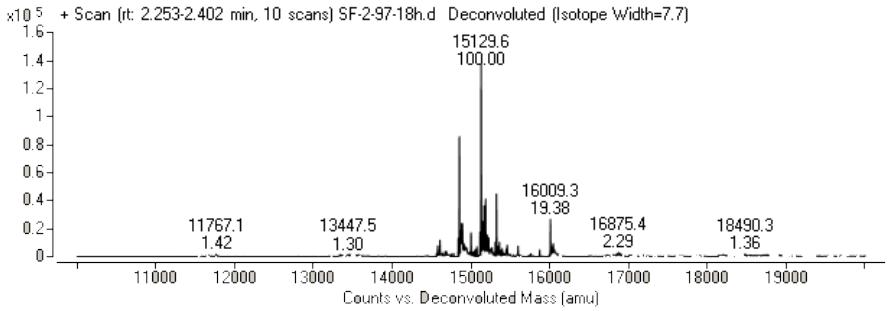
Fig. 4.4.9.1.2. *LCMS-ESI+ characterization of EndoA Y299F reaction starting material, cABLys3-O-GlcNAc. Also in fig. 4.2.10.2.*

<u>Endo A Y299F reaction; 18h LCMS-ESI+ characterization</u>

Ion series



Deconvoluted mass spectrum



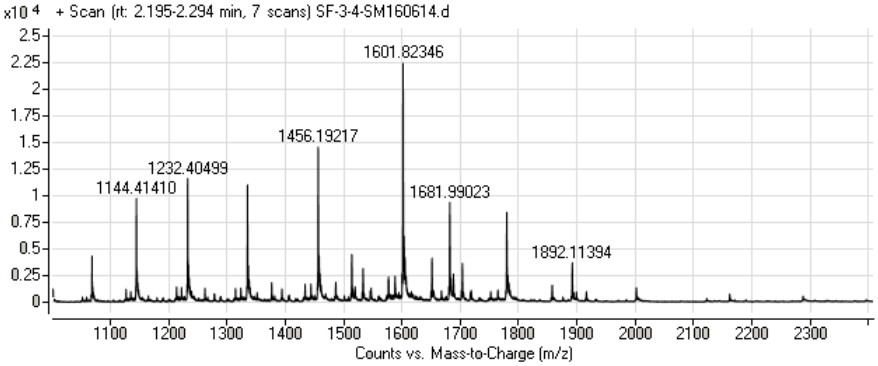
cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16009.3 Da.

cAb-Lys3-A104-glycan: expected mass 18159 Da, none observed.

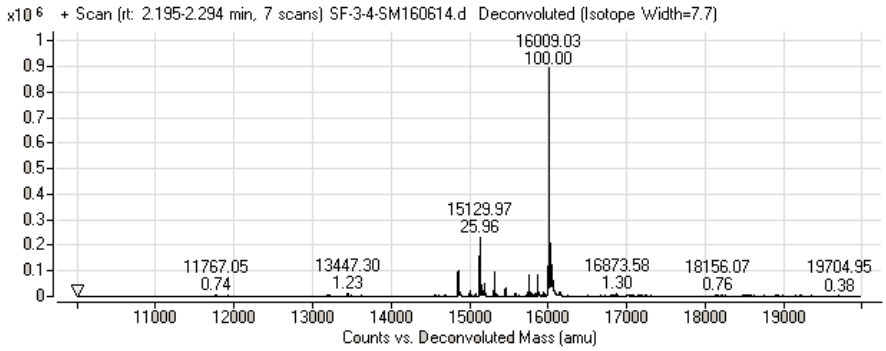
Fig. 4.4.9.1.3. *LCMS-ESI+ characterization of EndoA Y299F reaction- 18h. Degradation of cAbLys3-O-GlcNAc is observed.*

ii. EndoS D233Q reaction; starting material LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum

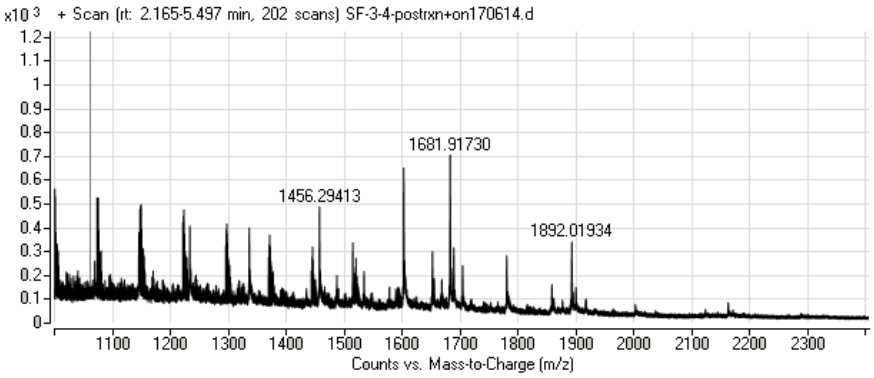


cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16009.03 Da.

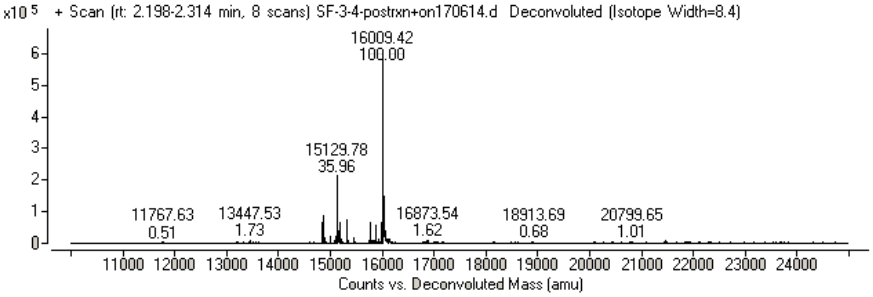
Fig. 4.4.9.1.4. *LCMS-ESI+ characterization of EndoS D233Q reaction starting material, cABLys3-O-GlcNAc.*

Endo S D233Q reaction; 18h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16009.42 Da.

cAb-Lys3-A104-glycan: expected mass 18159 Da, none observed.

Fig. 4.4.9.1.5. *LCMS-ESI+ characterization of EndoS D233Q reaction- 18h. Degradation of cAbLys3-O-GlcNAc is observed. Also in fig. 4.2.10.2.*

4.4.10. Small molecule synthesis

4.4.10.1 Synthesis of β-D-1-(2-iodoethoxy)-GlcNAc

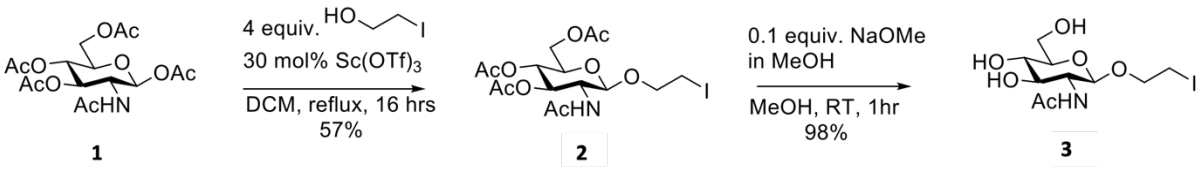
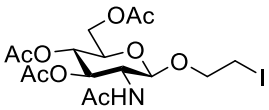


Fig. 4.2.1.1. (reproduced from results and discussion section) *Scheme for the sythesis of β-D-1-(2-iodoethoxy)-GlcNAc* ***3*** *from per-acetylated β-GlcNAc* ***1*** *via intermediate per-acetylated β-D-1-(2-iodoethoxy)-GlcNAc* ***2****.*



4.4.10.1.1. Peracetyl- β-D-1-(2-iodoethoxy)-GlcNAc (**2**)

Peracetyl-β-D-GlcNAc (150mg, 0.384 mMol, purchased from Carbosynth) and $Sc(OTf)_3$ (60 mg, 0.115 mMol, 30 mol%) were dissolved in DCM (7 mL) under argon. 2-iodo-ethanol (120 µL, 1.54 mMol) was then added, the reaction heated to reflux under N_2, and left overnight. Consumption of starting material was monitored by TLC (80% EtOAc in 40-60 petroleum ether, H_2SO_4 stain) and mass spectrometry. When all starting material was consumed, the reaction was diluted with DCM (20 mL). It was then washed with sat. $NaHCO_3$ (20 mL) followed by H_2O (20 mL), and dried over Na_2SO_4 and in-vacuo. The resulting yellow-tinted off-white solid was dissolved in minimal DCM and purified by flash column chromatography (12g column, Pet Ether/EtOAc 40 % -> 80 % gradient elution over 20 CV) to afford the title product as a colourless amorphous solid (110 mg, 0.219 mMol, 57%)

^{1}H NMR (400 MHz, Chloroform-*d*) δ 5.57 (d, *J* = 8.7 Hz, 1H, N**H**), 5.31 (dd, *J* = 10.7, 9.3 Hz, 1H, **H3**), 5.07 (t, *J* = 9.6 Hz, 1H, **H4**), 4.75 (d, *J* = 8.3 Hz, 1H, **H1**), 4.25 (dd, *J* = 12.3, 4.8 Hz, 1H, **H6**), 4.17 – 4.05 (m, 2H, **H6'** and OC**H$_2$**CH_2I), 3.86 (dt, *J* = 10.6, 8.5 Hz, 1H, **H2**), 3.78

(ddd, *J* = 11.3, 7.8, 6.6 Hz, 1H, OC**H₂**CH₂I), 3.71 (ddd, *J* = 10.1, 4.8, 2.4 Hz, 1H, **H5**), 3.37 – 3.17 (m, 2H, C**H₂**I), 2.09 (s, 3H, N**Ac**), 2.04 (s, 3H, O**Ac**), 2.03 (s, 3H, O**Ac**), 1.98 (s, 3H, O**Ac**).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 170.76 (NH**C**(=O)Me), 170.37 (O**C**(=O)Me), 169.51 (O**C**(=O)Me), 169.49 (O**C**(=O)Me), 100.89 (**C1**), 72.76 (**C5**), 72.06 (**C3**), 71.15 (**C4**), 70.62 (O**C**H₂CH₂I), 68.43(**C6**), 61.96 (**C2**), 21.02 (NH**Ac**), 20.89 (O**Ac**), 20.74 (O**Ac**), 20.73 (O**Ac**), 2.19 (OCH₂**C**H₂I).

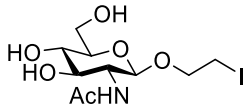
Matches reported peaks in literature[74]

MS+ [M+ H] expected: 502.0, observed: 502.1

HRMS m/z (ESI+) [M+ Na] expected: 524.03880, observed: 524.03837

$[\alpha]_D^{20}$ - 91.0

IR (ν_{max}/cm^{-1}) 3266, 3097, 1743, 1650, 1566, 1431, 1367, 1223, 1166, 1122, 1078, 1031, 979.0, 946.9, 903.6, 738.0, 695.7 and 631.3



4.4.10.1.1. β-D-1-(2-iodoethoxy)-GlcNAc (**3**)

Compound **2** (75 mg, 0.137 mMol) was suspended in dry MeOH (1.8 mL) under argon. 25% NaOMe in MeOH (26 µL, 120 µMol, 0.8 equivalents) was then added. The reaction was stirred and monitored by TLC (100% EtOAc, H_2SO_4 stain). After approximately 45 minutes the reaction had reached completion and was quenched by the addition of Amberlite H+ until neutral pH was reached. The solution was then filtered, and solvent removed *in vacuo* to yield the title product as an amorphous white solid (51.5 mg, 0.137mMol, 156.8 °C (decomp.) 98%).

^{1}H NMR (700 MHz, Deuterium Oxide) δ 4.40 (dd, *J* = 8.5, 1.0 Hz, 1H, **H1**), 3.97 (ddd, *J* = 11.9, 5.9, 4.7 Hz, 1H, **H6**), 3.77 (dd, *J* = 12.4, 1.7 Hz, 1H, **H6'**), 3.69 (ddd, *J* = 12.1, 7.8, 4.5

Hz, 1H, **H5**), 3.62 – 3.55 (m, 2H, **H3** and **H4**), 3.43 – 3.37 (m, 1H, **H2**), 3.31 – 3.29 (m, 2H, $OCH_2C\mathbf{H_2}I$), 3.21 (ddd, *J* = 10.6, 5.9, 4.5 Hz, 1H, $C\mathbf{H_2}CH_2I$), 3.14 (ddd, *J* = 10.9, 7.8, 4.7 Hz, 1H, $C\mathbf{H_2}CH_2I$'), 1.93 (s, 3H, NH**Ac**).

Matches reported peaks in literature[74]

MS+ [M+ H] expected: 376.1, observed: 376.0

HRMS m/z (ESI+) [M+ Na] Expected: 398.00710, observed: 398.00713

[α]$_D^{20}$ -125.7

IR (ν_{max}/cm^{-1}) 3259 (br), 2956, 1650, 1553, 1422, 1376, 1325, 1281, 1213, 1143, 1112, 1084, 1064, 1035, 961.2, 912.9, 891.3, 749.3, 705.6, 671.0 and 633.1

4.4.10.2. Synthesis of *N*-linked GlcNAc halides.

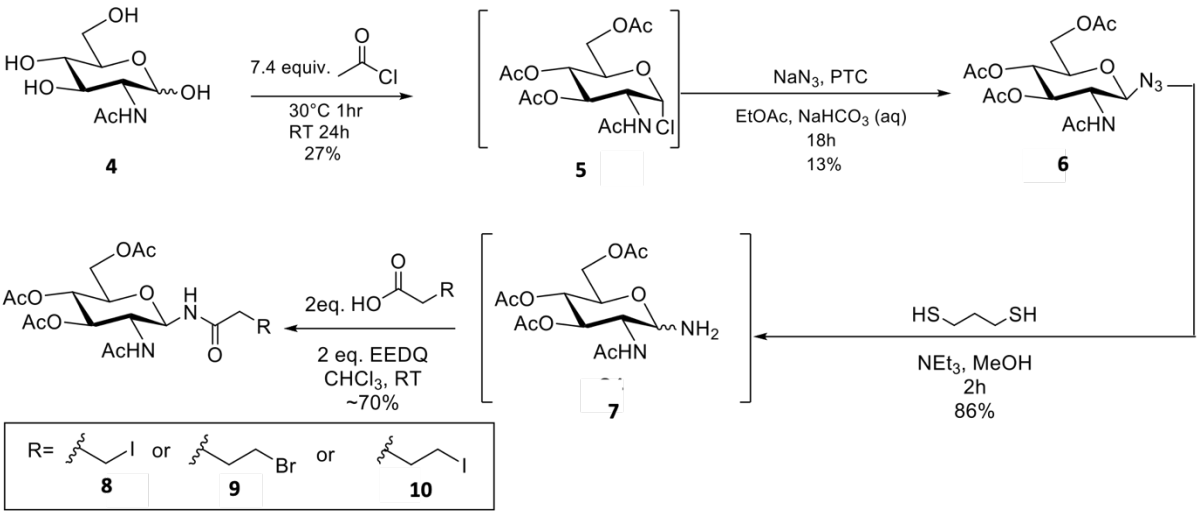


Fig. 4.2.1.2 (reproduced from results and discussion) *Scheme for the synthesis of peracetyl-β-D-1-(3-iodopropionamido)-GlcNAc* ***8****, peracetyl- β-D-1-(4-bromobutyramido)-GlcNAc* ***9,*** *and peracetyl- β-D-1-(4-iodobutyramido)-GlcNAc* ***10****. Synthesis begins from GlcNAc* ***4*** *and goes via Peracetyl-α-D-1-chloro-GlcNAc* ***5****, Peracetyl-β-D-1-azido-GlcNAc* ***6*** *and Peracetyl-D-1-amino-GlcNAc 7.*

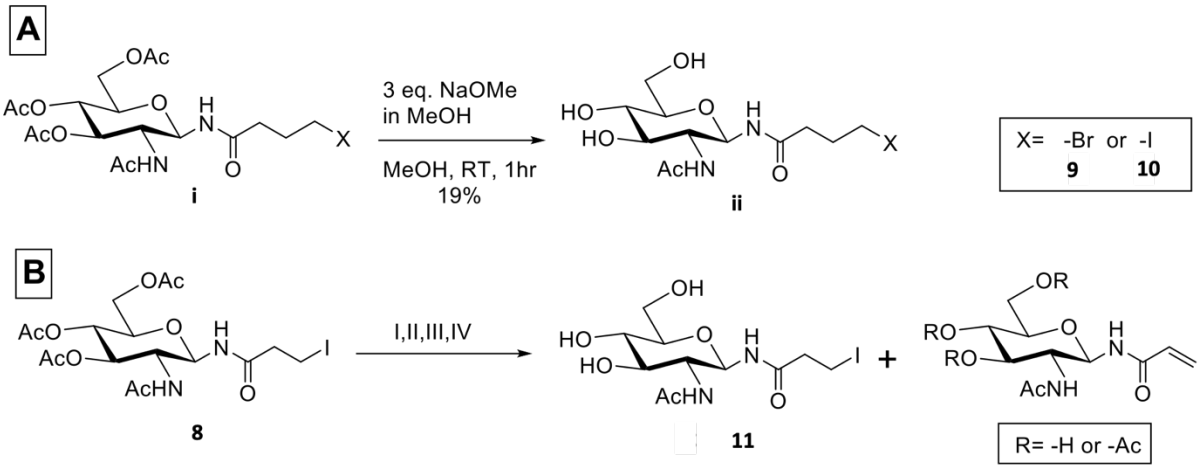
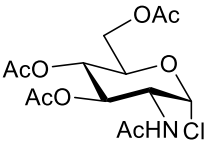


Fig. 4.2.1.3. (reproduced from results and discussion) ***(A)*** *Scheme for the deprotection of peracetyl-β-D-1-(4-bromobutyramido)-GlcNAc* ***9i*** *and peracetyl-β- D-1-(4-iodobutyramido)-GlcNAc* ***10i*** *to form β-D-1-(4-bromobutyramido)-GlcNAc* ***9ii*** *and β-D-1-(4-iodobutyramido)-GlcNAc* ***10ii*** *respectively.* ***(B)*** *Attempted deprotection of peracetyl-β-D-1-(3-iodopropionamido)-GlcNAc* ***8*** *to form β-D-1-(3-iodopropionamido)-GlcNAc* ***11****. The conditions for B are as follows: I= 0.5 M NaOMe in MeOH, 30 mins RT, II = 5 mM NaOMe in MeOH, 30 mins RT, III = Guanidinium nitrate, 62 h RT 2 h 60 ˚C, IV=0.2 M Mg(OMe)2, 45 mins RT.*



4.4.10.2.1. Peracetyl-α-D-1-chloro-GlcNAc (**5**)

GlcNAc (5g, 22.6 mMol) was suspended in acetyl chloride (12 mL, 168 mMol) under nitrogen. The reaction was stirred overnight and observed to change from an opaque white suspension to a clear orange-red gel. The reaction was monitored by TLC (100% EtOAc, H_2SO_4 stain). After 48 hours no further progress was observed the reaction was quenched by diluting with DCM (40 mL) and pouring onto a slurry of ice in H_2O (20 mL) with vigorous stirring. This was transferred into a separating funnel and the organic layer drawn off into ice in 10%

$NaHCO_3$ (30 mL) with vigorous stirring. The organic layer was once again separated and washed over sat. $NaHCO_3$ (2x 30 mL) before drying over Na_2SO_4. The orange solution was then reduced to approximately 10 mL *in-vacuo* and precipitated by adding diethyl ether (20 mL) and blowing argon gas over the surface. Filtration afforded the crude title product as a pale pink solid (2.248g, 6.15 mMol, 27%)

^{1}H NMR (400 MHz, Chloroform-*d*) δ 6.18 (d, *J* = 3.7 Hz, 1H, **H1**), 5.85 (d, *J* = 8.7 Hz, 1H, N**H**), 5.37 – 5.28 (m, 1H, **H3**), 5.22 (t, 1H, **H4**), 4.53 (ddd, *J* = 10.6, 8.7, 3.8 Hz, 1H, **H2**), 4.34 – 4.24 (m, 2H, **H5** and **H6**), 4.17 – 4.05 (m, 1H, **H6'**), 2.10 (s, 3H, NH**Ac**), 2.05 (s, 6H, 2x O**Ac**), 1.98 (s, 3H, O**Ac**).

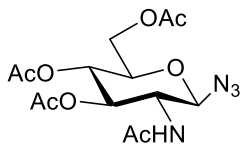
^{13}C NMR (101 MHz, Chloroform-*d*) δ 171.60 (NH**C**(=O)Me), 170.71 (OC(=O)Me), 170.25 (OC(=O)Me), 169.27 (OC(=O)Me), 93.76 (**C1**), 71.01 (**C3**), 70.25 (**C5**), 67.08 (**C4**), 61.27 (**C6**), 53.60 (**C2**), 23.20 (NH**Ac**), 20.80 (2x O**Ac**), 20.67 (O**Ac**)

Matches reported peaks in literature[118]

MS+ [M+ H] expected: 366.8 , observed: 366.1

HRMS m/z (ESI+) [M+Na] Expected: 388.07697, observed: 388.07708.

IR (ν_{max}/cm^{-1}) 3242, 1739, 1643, 1542, 1435, 1375, 1290, 1269, 1227, 1210, 1118, 1034, 984.3, 912.1, 893.0, 874.8, 821.2, 761.4, 730.8, 643.3, and 606.7.



4.4.10.2.2. Peracetyl-β-D-1-azido-GlcNAc (**6**)

Compound **5** (374 mg, 1.03 mMol) and tetrabutyl ammonium hydrogen sulfate (348 mg, 1.03 mMol) were dissolved in EtOAc (5 mL) to form an off-white semi-opaque solution. $NaHCO_3$ (5 mL) was added, followed by NaN_3 (200 mg, 3.07 mMol). The solution was stirred and reaction progress monitored by TLC (100% EtOAc, H_2SO_4 stain). After 4 hours all starting material was consumed and the reaction quenched by washing with sat. $NaHCO_3$ (3x 10 mL). Organic fractions were combined and dried over $MgSO_4$ and in-vacuo. The resultant yellow-tinted off-white solid was re-dissolved in minimal EtOAc and purified by flash column chromatography (12 g column, Pet Ether/EtOAc 50 % -> 80 % gradient elution over 10 CV) to afford the title compound as a white amorphous solid (152.3 mg, 0.134 mMol, 172.0 °C (decomp.), 13.3%)

^{1}H NMR (400 MHz, Chloroform-*d*) δ 5.82 (d, *J* = 8.9 Hz, 1H, N**H**), 5.33 – 5.20 (m, 1H, **H3**), 5.09 (t, *J* = 9.7 Hz, 1H, **H4**), 4.77 (d, *J* = 9.2 Hz, 1H, **H1**), 4.26 (dd, *J* = 12.4, 4.8 Hz, 1H, **H6**), 4.16 (dd, *J* = 12.5, 2.3 Hz, 1H, **H6'**), 3.91 (q, *J* = 9.5 Hz, 1H, **H2**), 3.80 (ddd, *J* = 10.1, 4.8, 2.3 Hz, 1H, **H5**), 2.10 (s, 3H, NH**Ac**), 2.03 (d, *J* = 3.1 Hz, 6H,2x O**Ac**), 1.97 (s, 3H, O**Ac**).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 171.17 (NH**C**(=O)Me), 170.83 (O**C**(=O)Me), 170.54 (O**C**(=O)Me), 169.41 (O**C**(=O)Me), 88.57 (**C1**), 74.13 (**C5**), 72.25 (**C3**), 68.11 (**C4**), 61.96 (**C6**), 54.31 (**C2**), 23.40 (NH**Ac**), 20.88 (O**Ac**), 20.78 (O**Ac**), 20.73 (O**Ac**).

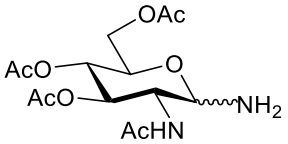
Matches reported peaks in literature[118]

MS + [M+ H] expected: 373.3, observed: 373.1

HRMS m/z (ESI+) Expected: 395.11733 Found: 395.11676

$[\alpha]_D^{20}$ – 414.0

IR (ν_{max}/cm^{-1}) 3272, 2119, 1744, 1655, 1556, 1434, 1367, 1216, 1152, 1032, 975.3, 897.8, 735.5, and 620.7.



4.4.10.2.3.Peracetyl-D-1-amino-GlcNAc (**7**)

Compound **6** (1g, 2.69 mMol) was dissolved in 16 mL dry MeOH in a round- bottomed flask cooled under argon. Triethylamine (0.9 mL, 6.46 mMol) and 1,3 propane dithiol (0.6 mL, 5.98 mMol) were added sequentially, and the reaction was then stirred and monitored by TLC (10% MeOH in $CHCl_3$, anisidine phthalate stain). After 2.5 hours the reaction had gone to completion and was concentrated *in-vacuo* and purified by flash column chromatography (40g column, MeOH/$CHCl_3$ 0 % -> 10 % gradient elution over 10 CV) to afford the title product as a glassy solid (801mg, 2.31 mMol, 86%. Both anomers present but β predominant.)

^{1}H NMR (400 MHz, Chloroform-d) δ 5.65 (d, J = 9.1 Hz, 1H, N**H**), 5.19 – 4.85 (m, 2H, **H3** and **H4**), 4.18 (dd, J = 12.3, 4.9 Hz, 1H, **H6**), 4.08 (d, J = 9.18 Hz, 1H, **H1**), 4.07 (dd, J =12.31, 2.14, 1H, **H6'**) 4.02 – 3.93 (m, 2H, **H2** and **H5**), 2.06 (s, 3H, O**Ac**), 2.01 (s, 3H, O**Ac**), 2.00 (s, 3H, O**Ac**) and 1.94 (s, 3H, NH**Ac**).

^{13}C NMR (101 MHz, Chloroform-*d*) δ 171.67 (NH**C**(=O)Ac), 170.93 (**C**(=O)Me), 170.91 (**C**(=O)Me), 169.47 (**C**(=O)Me), 86.49(**C1**), 73.55 (**C5**), 72.85 (**C3**), 68.65 (**C4**), 62.54 (**C6**), 54.98 (**C2**), 23.43 (NH**Ac**), 20.92 (O**Ac**), 20.84(O**Ac**), 20.75(O**Ac**).

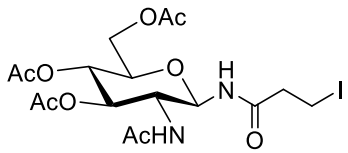
Matches reported peaks in literature[118]

MS+ [M+Na] Observed 369.2, Expected 369.0

HRMS m/z (ESI+) Expected: 369.12684 [M+Na], Observed 369.12696.

$[\alpha]_D^{20}$ - 92.0

IR (ν_{max}/cm^{-1}) 3400, 3342, 1749, 1733, 1660, 1537, 1428, 1367, 1231, 1111, 1088, 1044, 1018, 955.6, 902.6, and 690.6.



4.4.10.2.4. Peracetyl-β-D-1-(3-propionamido)-GlcNAc **(8)**

Compound **S4** (930 mg, 2.687 μMol) was added to a solution of N-Ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline (1.328 g, 5.374 mMol) and 3-iodopropionic acid (1.075 g, 5.374 mMol) in DCM (20 mL). The reaction was stirred and monitored by TLC (20% acetone in $CHCl_3$, anisaldehyde stain). After 6 hours the reaction had gone to completion; it was loaded directly onto a biotage column and purified by flash column chromatography (40g column, MeOH/$CHCl_3$ 0 % -> 7 % gradient elution over 10 CV) to afford the title product as a white solid (835.1 mg, 1.58 mMol, 70.7%)

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.21 (d, *J* = 8.3 Hz, 1H, Bu N**H**C(=O)), 6.43 (d, *J* = 8.3 Hz, 1H, N**H**Ac), 5.18 – 5.06 (m, 3H, **H1, H3** and **H4**), 4.29 (dd, *J* = 12.5, 4.4 Hz, 1H, **H6**), 4.13 (td, *J* = 10.0, 8.4 Hz, 1H, **H2**), 4.07 (dd, *J* = 12.5, 2.3 Hz, 1H, **H6'**), 3.80 (ddd, *J* = 9.6, 4.4, 2.2 Hz, 1H, **H5**), 3.45 – 3.24 (m, 2H, Bu **Hα**), 2.97 – 2.68 (m, 2H, Bu **Hβ**), 2.07 (s, 3H, O**Ac**), 2.06 (s, 3H, O**Ac**), 2.03 (s, 3H, O**Ac**), 1.97 (s, 3H, NH**Ac**).

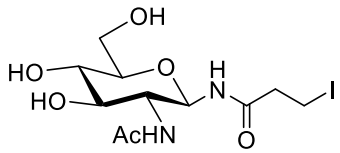
^{13}C NMR (101 MHz, Chloroform-*d*) δ 172.41 (Bu NH**C**(=O)), 171.91(NH**C**(=O)Me), 171.16(O**C**(=O)Me), 170.80(O**C**(=O)Me), 169.43 (O**C**(=O)Me), 80.30(**C1**), 73.58(**C5**), 72.94 (**C3**), 68.00(**C4**), 61.91(**C6**), 53.40(**C2**), 40.29(Bu **Cα**), 23.60(NH**Ac**), 20.89(O**Ac**), 20.856O**Ac**), 20.74(O**Ac**), -2.53 (Bu **Cβ**).

MS+ [M+ H] expected: 529.3, observed: 529.1. [M+ Na] expected: 551.0, observed: 551.1

HRMS m/z (ESI+) Expected: 529.06775 [M+H], found: 529.06815.

$[\alpha]_D^{25}$= -12.7

IR (ν_{max}/cm^{-1}) 3312, 1743, 1661, 1534, 1370, 1302, 1226, 1168, 1119, 1040, 967.6, 922.5, 898.1, and 660.3.



4.4.10.2.5. β-D-1-(3-propionamido)-GlcNAc **(11)**

Method 1: 0.5 M NaOMe in MeOH

Compound **S5** (997.3 mg, 1.8878 mMol) was suspended in MeOH (10 mL) under argon. 0.5 M NaOMe in MeOH (3.8 mL, 1.8878 mMol) was then added, and the reaction stirred and monitored by TLC (100% EtOAc, anisaldehyde stain). The suspended solid was observed to dissolve within 2 minutes, after which a white precipitate formed at approximately 5 minutes. After 30 mins the reaction appeared to have gone to completion by TLC, and was quenched by adding Amberlite H+ to lower the pH to neutral. The mixture was then filtered dried *in vacuo*. The crude product was dry-loaded onto a biotage column pre-washed with MeOH (10 CV), and purified by column chromatography (6 to 50% MeOH in $CHCl_3$ over 10 CV). The white precipitate was again observed forming in product-containing fractions. Only elimination product was isolated.

Method 2: 5 mM NaOMe in MeOH

Compound **S5** (100 mg, 188.9 µMol) was suspended 5 mM NaOMe in MeOH (6.3 mL, 31.5 µMol) under argon. The reaction was stirred and monitored by TLC (100% EtOAc, anisaldehyde stain). After 40 minutes no significant change had occurred, so NaOMe in MeOH was added to bring to 10 mM (60 µL 0.5 M NaOMe in MeOH). The white precipitate was again observed, and only elimination product was isolated.

Method 3: Guanidinium nitrate

Guanidinium nitrate (622 mg, 5.09 mMol) was dissolved in MeOH (5 mL) and $CHCl_3$ (45 mL) under argon, and NaOMe in MeOH (2 mL of 0.5M) was then added. To this solution (2 mL) was added **S5** under argon. The reaction was stirred and monitored by TLC (100% EtOAc, anisaldehyde stain). After 24 and 48 hours additional aliquots of guanidinium nitrate stock

solution (2 mL) were added to the reaction as it had still not gone to completion; it was then heated (60 ˚C, 2 hours). As it did not go to completion, it was discarded.

Method 4: 0.2 M $Mg(OMe)_2$

Compound **S5** (100 mg, 189 µMol) was suspended in $Mg(OMe)_2$ in MeOH (2.85 mL of 0.2 M, 0.57 mMol) under argon. The reaction was stirred and monitored by TLC (25% MeOH in $CHCl_3$, anisaldehyde stain) and MS-ESI-. After 30 minutes the reaction had gone to completion, and at 45 minutes it was quenched by adding Amberlite H+ to neutralise pH. A gel formed on addition of Amberlite, and dry MeOH (3 mL) was added to dilute the mixture. Filtration afforded the crude mixture of products as an off-white solid (79 mg).

^{13}C NMR (101 MHz, Methanol-d_4) δ 174.42 (-NH**C**OCH_2CH_2I), 168.29(-NH**C**OCH_3), 131.77(-NHCO**C**HCH_2), 128.20 (-NHCOCH**C**H_2), 80.41 (C1), 79.82 (C5), 76.24 (C3), 71.77 (C4), 62.23 (C6) 56.24 (C2), 47.97 (-NHCO**C**H_2CH_2I), 22.69 (-NHCO**C**H_3), 9.31(-NHCOCH_2**C**H_2I)

^{1}H NMR (400 MHz, Methanol-d_4) δ 6.28 – 6.19 (m, 2H, impurity –NHCOCHC**H_2**), 5.72 (dd, J = 9.2, 2.7 Hz, 1H impurity –NHCOC**H**CH_2), 5.06 (d, J = 9.8 Hz, 1H, H1), 3.84 (dd, J = 12.0, 1.7 Hz, 1H, H6), 3.80 (t, J = 9.9 Hz, 1H, H2), 3.68 – 3.62 (m, 1H, H6'), 3.48-3.52 (m, 1H, H3), 3.35 – 3.31 (m, 2H, H4 and H5), 3.23 (q, J = 7.3 Hz, 1H-NHCOC**H_2**CH_2I), 1.95 (s, 3H, -NHCOC**H_3**), 1.32 (t, J = 7.3 Hz, 1H, -NHCOCH_2C**H_2**I).

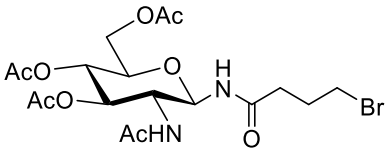
MS- [M+ H] expected: 401.2, observed: 401.0

Elimination product: LCMS-ESI- [M-H] expected: 273, observed: 273.1

Deprotection method	Went to completion?	Ratio of **4**: side product isolated
0.5 M NaOMe in MeOH	Yes	0:1
5 mM NaOMe in MeOH	Yes	0:1

Guanidinium nitrate	No	-
0.2 M $Mg(OMe)_2$	Yes	2:1

Comparison of deprotection methods tested.



4.4.10.2.6. Peracetyl-β-D-1-(4-bromobutyramido)-GlcNAc **(9i)**

Compound **S4** (480 mg, 1.384 µMol) was added to a solution of N-Ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline (686 mg, 2.77 mMol) and 4-bromobutyric acid (460 mg, 2.77 mMol) in DCM (20 mL). The reaction was stirred and monitored by TLC (20% acetone in $CHCl_3$, anisaldehyde stain). After 5 hours the reaction had gone to completion; it was loaded directly onto a biotage column and purified by flash column chromatography (40g column, MeOH/$CHCl_3$ 0 % -> 7 % gradient elution over 10 CV) to afford the title product as a white solid (459 mg, 0.929 mMol , 225.2 °C (decomp.), 67%)

^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.11 (d, *J* = 8.4 Hz, 1H, -N**H**), 6.27 (d, *J* = 8.2 Hz, 1H, N**H**), 5.14 – 5.03 (m, 3H, H1, H3 and H5), 4.29 (dd, *J* = 12.5, 4.4 Hz, 1H, H6), δ 4.17 – 4.05 (m, 2H, H2 and H6'), 3.77 (ddd, *J* = 9.6, 4.2, 2.1 Hz, 1H, H5), 3.42 (t, *J* = 6.4 Hz, 2H, Bu **Hγ**), 2.36 (dq, *J* = 8.4, 10.1 Hz, 2H, Bu **Hβ**), 2.26 – 2.10 (m, 2H, Bu **Hα**), 2.09 (s, 3H, O**Ac**), 2.06 (s, 3H, O**Ac**), 2.03 (s, 3H, O**Ac**), 1.95 (s, 3H, -NH**Ac**).

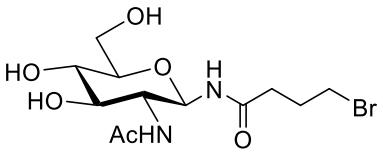
^{13}C NMR (101 MHz, Chloroform-*d*) δ 172.55 (Bu NH**C**(=O)), 172.16 (NH**C**(=O)Me), 171.98 (O**C**(=O)Me), 170.82 (O**C**(=O)Me), 169.41(O**C**(=O)Me), 80.38 (**C1**), 73.60 (**C5**), 73.04(**C3**), 67.93 (**C4**), 61.88 (**C6**), 53.44 (**C2**), 34.47(Bu **Cα**), 32.87 (Bu **Cγ**), 27.81 (Bu **Cβ**), 23.27 (NH**Ac**), 20.88 (O**Ac**), 20.84(O**Ac**), 20.72 (O**Ac**).

MS + [M+ Na] expected: 517.0, observed: 517.0

HRMS m/z (ESI+) Expected: 517.07921; 519.07717 [M + Na, Br^{79} and Br^{81} isotopes], Observed: 517.07932; 519.07723.

$[\alpha]_D^{25}$= -11.8

IR (ν_{max}/cm^{-1}) 3328, 1742, 1659, 1532, 1430, 1371, 1303, 1233, 1214, 1109, 1043, 968.8, 898.85 and 673.0.



4.4.10.2.7. β-D-1-(4-bromobutyramido)-GlcNAc (**9ii)**

Compound **S6** (400 mg, 0.808 mMol) was suspended in 5 mM NaOMe in MeOH (28 mL, 0.14 mMol in a round-bottomed flask under argon. The reaction was stirred and monitored by TLC (30% MeOH in $CHCl_3$, anisaldehyde stain), and the suspended solid was observed to dissolve over time. After 30 minutes, the reaction had gone to completion and was quenched by adding Dowex H+ until neutral pH was reached. The solution was filtered and dried *in vacuo* to afford the title product as a white solid (276 mg, 0.748 mMol, 188.5 °C (decomp.), 92.6%)

^{1}H NMR (400 MHz, Deuterium Oxide) δ 5.07 (d, *J* = 9.7 Hz, 1H, **H1**), 3.88 (dd, *J* = 12.4, 2.0 Hz, 1H, **H6**), 3.82 (t, *J* = 9.9 Hz, 1H, **H2**), 3.75 (dd, *J* = 12.3, 4.7 Hz, 1H, **H6'**), 3.60 (dd, *J* = 10.2, 8.4 Hz, 1H, **H3**), 3.49 (d, *J* = 8.4 Hz, 1H, **H4**), 3.57-3.40 (m, 3H, **H5** and Bu **Hβ**) 2.44 (td, *J* = 7.2, 4.1 Hz, 2H, Bu **Hα**), 2.12 (dtd, *J* = 13.1, 7.1, 5.6 Hz, 2H, Bu **Hγ**), 2.01 (s, 3H, **Ac**).

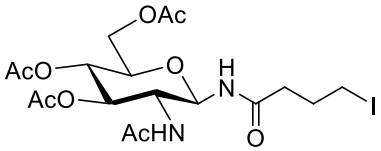
^{13}C NMR (101 MHz, Deuterium Oxide) δ 176.36 (NH**C**(=O)Bu), 174.68 (NH**C**(=O)Me), 78.26 (**C1**), 77.60 (**C5**), 74.22 (**C3**), 69.43 (**C4**), 60.46 (**C6**), 54.27 (**C2**), 34.16 (Bu **Cα**), 32.95 (Bu **Cβ**), 27.89 (Bu **Cγ**), 22.08 (**Ac**).

MS + [M+ Na] expected: 391.0, observed: 391.0

HRMS m/z (ESI+) Expected: 391.04752; 393.04547 [M+ Na, Br^{79} and Br^{81} isotopes], found: 391.04749, 393.04539.

$[\alpha]_D^{25}$= +26.9

IR (ν_{max}/cm^{-1}) 3254, 2913, 2361, 1651, 1542, 1438, 1374, 1322, 1264, 1166, 1109, 1085, 1036, 918.9, 889.8, 668.8, and 613.1.



4.4.10.2.8. Peracetyl-β-D-1-(4-iodobutyramido)-GlcNAc **(10i)**

Compound **S4** (613 mg, 1.77 mMol) was added to N-Ethoxycarbonyl-2-ethoxy-1,2-dihydroquinoline (497 mg, 2.00 mMol) and 4-iodobutyric acid (429 mg, 2.00 mMol) dissolved in $CHCl_3$ (10 mL) under argon. The reaction was stirred and monitored by TLC (5% MeOH in $CHCl_3$, anisaldehyde stain). After 5 hours the reaction had gone to completion; it was loaded directly onto a biotage column and purified by flash column chromatography (40g column, MeOH/$CHCl_3$ 0 % -> 7 % gradient elution over 10 CV) to afford the title product as a pale yellow solid which turned orange when exposed to light for extended periods (681.3 mg, 1.26 mMol , 71%)

^{1}H NMR (500 MHz, Chloroform-*d*) δ 7.08 (d, *J* = 8.4 Hz, 1H, BuN**H**), 6.26 (d, *J* = 8.3 Hz, 1H, AcN**H**), 5.13 – 5.03 (m, 3H, **H1, H3** and **H4**), 4.28 (dd, *J* = 12.5, 4.4 Hz, 1H, **H6**), 4.13 (dq, *J* = 9.9, 8.3 Hz, 1H, **H2**), 4.08 (dd, *J* = 12.5, 2.3 Hz, 1H, **H6'**), 3.77 (ddd, *J* = 9.6, 4.3, 2.2 Hz, 1H, **H5**), 3.19 (t, *J* = 6.7 Hz, 2H, **Bu γ**), 2.31(tt, *J* =7.5, 6.8 Hz, 2H, **Bu β**), 2.16-2.05 (m, 2H, **Bu α**) 2.07 (s, 3H, O**Ac**), 2.06 (s, 3H, O**Ac**), 2.03 (s, 3H, O**Ac**), 1.96 (s, 3H, NH**Ac**).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 172.34 (NH**C**(=O)Bu), 172.15 (NH**C**(=O)Me), 171.99 (O**C**(=O)Me), 170.82 (O**C**(=O)Me), 169.41 (O**C**(=O)Me), 80.31 (**C1**), 73.60 (**C5**), 73.04 (**C3**),

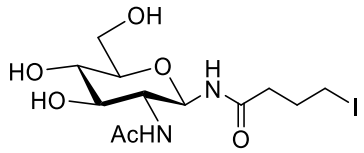
67.91 (**C4**), 61.87 (C6), 53.45 (**C2**), 36.82 (Bu **Cα**), 28.57 (Bu **Cβ**), 23.33 (NH**Ac**), 20.88 (O**Ac**), 20.84 (O**Ac**), 20.73 (O**Ac**), 5.61 (Bu **Cγ**).

MS+ [M+ NA^+] Expected: 565.1 found: 566.0

HRMS m/z (ESI+) Expected: 565.06534 Found: 565.06506

$[\alpha]_D^{20}$ -98.45

IR (ν_{max}/cm^{-1}) 3295, 1745, 1658, 1534, 1371, 1305, 1225, 1176, 1036, 965.5, and 916.0



4.4.10.2.9. β-D-1-(4-iodobutyramido)-GlcNAc **(10ii)**

This compound was synthesised by others within the BGD group- reported here as in ref 118 for reference only.

Compound **S7** (450 mg, 0.830 mMol) was suspended in 5 mM NaOMe in MeOH (32 mL, 0.16 mMol) in a round-bottomed flask cooled under argon. The reaction was stirred and monitored by TLC (30% MeOH in $CHCl_3$, anisaldehyde stain), and the suspended solid was observed to dissolve over time. After 30 minutes, the reaction had gone to completion and was quenched by adding Dowex H+ until neutral pH was reached. The solution was filtered and dried *in-vacuo* to afford the title product as a white solid (299 mg, 0.718 mMol, 86.5%)

^{1}H NMR (400 MHz, Deuterium Oxide) δ 4.99 (d, J = 9.7 Hz, 1H, **H1**), 3.79 (dd, J = 12.4, 1.9 Hz, 1H, **H6**), 3.73 (t, J = 9.9 Hz, 1H, **H2**), 3.66 (dd, J = 12.3, 4.7 Hz, 1H, **H6'**), 3.51(dd, J = 10.2, 8.3 Hz, 1H, **H3**), 3.44-3.35 (m, 2H, **H**5, **H**5), 3.20-3.05 (m, 2H, Bu **Hγ**), 2.31 (td, J = 7.1, 2.7 Hz, 2H, Bu **Hα**), 2.01-1.95 (m, 2H, Bu **Hβ**) and 1.94 (s, 3H, **Ac**)

^{13}C NMR (100.6 MHz, Deuterium Oxide) δ 176.11 (Bu **C**(=O)), 174.59 (NH**C**(=O)Me), 78.20 (**C1**), 77.55 (**C5**), 74.18 (**C3**), 69.37 (**C4**), 60.40 (**C6**), 54.22 (**C2**), 36.30 (Bu **Cα**), 28.59 (Bu **Cβ**), 22.18 (NH**Ac**) and 5.28 (Bu **C**ϒ)

MS+ [M + Na+] Expected: 439.0, found 439.0

HRMS m/z (ESI+) Expected: 439.03365 Found: 439.03314

$[\alpha]_D^{20}$ = + 53.4

IR (ν_{max}/cm^{-1}) 3284, 2952, 1655, 1543, 1439, 1373, 1314, 1279, 1256, 1206, 1170, 1122, 1106, 1083, 1042, 956.2, 914.3, 894.3, 694.0, 633.5, and 616.9

4.5 References

1. Fang, R. *et al.* Human LSD2/KDM1b/AOF1 Regulates Gene Transcription by Modulating Intragenic H3K4me2 Methylation. *Mol. Cell* **39**, 222–233 (2010).
2. Yoh, S. M., Lucas, J. S. & Jones, K. A. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev.* **22**, 3422–3434 (2008).
3. Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **27**, 406–420 (2008).
4. Zhang, P. *et al.* Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic Acids Res.* **34**, 6621–6628 (2006).
5. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science (80-.).* **354**, (2016).
6. Wright, T. H. & Davis, B. G. Post-translational mutagenesis for installation of natural and unnatural amino acid side chains into recombinant proteins. *Nat. Protoc.* **12**, 2243

(2017).

7. Vlastaridis, P. *et al.* Estimating the total number of phosphoproteins and phosphorylation sites in eukaryotic proteomes. *Gigascience* **6**, 1–11 (2017).
8. DENHARDT, D. T. Signal-transducing protein phosphorylation cascades mediated by Ras/Rho proteins in the mammalian cell: the potential for multiplex signalling. *Biochem. J.* **318**, 729–747 (1996).
9. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu. Rev. Biochem.* **80**, 825–858 (2011).
10. Lin, A. *et al.* Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci. Transl. Med.* **11**, (2019).
11. Rudmann, D. G. On-target and off-target-based toxicologic effects. *Toxicol. Pathol.* **41**, 310–4 (2013).
12. Bowes, J. *et al.* Reducing safety-related drug attrition: The use of in vitro pharmacological profiling. *Nature Reviews Drug Discovery* **11**, 909–922 (2012).
13. Scott, A. M. *et al.* A phase I clinical trial with monoclonal antibody ch806 targeting transitional state and mutant epidermal growth factor receptors. *Proc. Natl. Acad. Sci.* **104**, 4071 LP – 4076 (2007).
14. Hughes, B. Antibody–drug conjugates for cancer: poised to deliver? *Nat. Rev. Drug Discov.* **9**, 665–667 (2010).
15. Weiner, L. M., Surana, R. & Wang, S. Monoclonal antibodies: versatile platforms for cancer immunotherapy. *Nat. Rev. Immunol.* **10**, 317–327 (2010).
16. Schliemann, C. & Neri, D. Antibody-Based Vascular Tumor Targeting. in 201–216 (Springer, Berlin, Heidelberg, 2010). doi:10.1007/978-3-540-78281-0_12
17. Scott, A. M., Wolchok, J. D. & Old, L. J. Antibody therapy of cancer. *Nat. Rev.*

Cancer **12**, 278–287 (2012).

18. Herbertson, R. A. *et al.* Phase I Biodistribution and Pharmacokinetic Study of Lewis Y–Targeting Immunoconjugate CMD-193 in Patients with Advanced Epithelial Cancers. *Clin. Cancer Res.* **15**, 6709–6715 (2009).
19. Williams, A. F. & Barclay, A. N. The Immunoglobulin Superfamily—Domains for Cell Surface Recognition. *Annu. Rev. Immunol.* **6**, 381–405 (1988).
20. Schroeder, H. W., Cavacini, L. & Cavacini, L. Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125**, S41-52 (2010).
21. Rousseaux, J., Rousseaux-Prévost, R. & Bazin, H. Optimal conditions for the preparation of Fab and F(ab′)2 fragments from monoclonal IgG of different rat IgG subclasses. *J. Immunol. Methods* **64**, 141–146 (1983).
22. Rousseaux, J., Rousseaux-Prevost, R. & Bazin, H. [63] Optimal conditions for the preparation of proteolytic fragments from monoclonal IgG of different rat IgG subclasses. *Methods Enzymol.* **121**, 663–669 (1986).
23. Rousseaus, J., Biserte, G. & Bazin, H. The differential enzyme sensitivity of rat immunoglobulin G subclasses to papain and pepsin. *Mol. Immunol.* **17**, 469–482 (1980).
24. Normansell, D. E. Human immunoglobulin subclasses. *Diagn. Clin. Immunol.* **5**, 115–28 (1987).
25. Torres, M. & Casadevall, A. The immunoglobulin constant region contributes to affinity and specificity. *Trends Immunol.* **29**, 91–97 (2008).
26. Gayle E. Ritchie, † *et al.* Glycosylation and the Complement System. (2002). doi:10.1021/CR990294A
27. Yu, X., Baruah, K., Scanlan, C. N. & Crispin, M. Antibody Glycosylation. *Antib. Fc* 179–194 (2014). doi:10.1016/B978-0-12-394802-1.00010-8

28. Charles A Janeway, J., Travers, P., Walport, M. & Shlomchik, M. J. The distribution and functions of immunoglobulin isotypes. (2001).

29. Sela-Culang, I., Kunik, V. & Ofran, Y. The structural basis of antibody-antigen recognition. *Front. Immunol.* **4**, 302 (2013).

30. Padlan, E. A. Anatomy of the antibody molecule. *Mol. Immunol.* **31**, 169–217 (1994).

31. Heidelberger, M. & Pedersen, K. O. The molecular weight of antibodies. *J. Exp. Med.* **65**, 393–414 (1937).

32. Smith, G. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science (80-.).* **228**, 1315–1317 (1985).

33. Smith, K. A. *et al.* Demystified... recombinant antibodies. *J. Clin. Pathol.* **57**, 912–917 (2004).

34. Ahmad, Z. A. *et al.* scFv antibody: principles and clinical application. *Clin. Dev. Immunol.* **2012**, 980250 (2012).

35. Huehls, A. M., Coupet, T. A. & Sentman, C. L. Bispecific T-cell engagers for cancer immunotherapy. *Immunol. Cell Biol.* **93**, 290–296 (2015).

36. Argos, P. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. *J. Mol. Biol.* **211**, 943–958 (1990).

37. Hamers-Casterman, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1993).

38. Roux, K. H. *et al.* Structural analysis of the nurse shark (new) antigen receptor (NAR): Molecular convergence of NAR and unusual mammalian immunoglobulins. *Proc. Natl. Acad. Sci.* **95**, 11804–11809 (1998).

39. Flajnik, M. F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.* **11**, 47–59 (2010).

40. Flajnik, M. F., Deschacht, N. & Muyldermans, S. A Case Of Convergence: Why Did a Simple Alternative to Canonical Antibodies Arise in Sharks and Camels? *PLoS Biol.* **9**, e1001120 (2011).

41. Nguyen, V. K., Desmyter, A. & Muyldermans, S. Functional heavy-chain antibodies in camelidae. *Adv. Immunol.* **79**, 261–296 (2001).

42. Holliger, P. & Hudson, P. J. Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol.* **23**, 1126–1136 (2005).

43. Feng, M. *et al.* Construction and next-generation sequencing analysis of a large phage-displayed VNAR single-domain antibody library from six naïve nurse sharks. *Antib. Ther.* **2**, 1–11 (2019).

44. Muyldermans, S. Nanobodies: Natural Single-Domain Antibodies. *Annu. Rev. Biochem.* **82**, 775–797 (2013).

45. De Genst, E. *et al.* Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4586–91 (2006).

46. Nguyen, V. K., Hamers, R., Wyns, L. & Muyldermans, S. Camel heavy-chain antibodies: diverse germline VHH and specific mechanisms enlarge the antigen-binding repertoire. *EMBO J.* **19**, 921–930 (2000).

47. Govaert, J. *et al.* Dual Beneficial Effect of Interloop Disulfide Bond for Single Domain Antibody Fragments. *J. Biol. Chem.* **287**, 1970–1979 (2012).

48. Vu, K. B., Ghahroudi, M. A., Wyns, L. & Muyldermans, S. Comparison of llama VH sequences from conventional and heavy chain antibodies. *Mol. Immunol.* **34**, 1121–1131 (1997).

49. Harmsen, M. M. *et al.* Llama heavy-chain V regions consist of at least four distinct subfamilies revealing novel sequence features. *Mol. Immunol.* **37**, 579–590 (2000).

50. Maass, D. R., Sepulveda, J., Pernthaner, A. & Shoemaker, C. B. Alpaca (Lama pacos)

as a convenient source of recombinant camelid heavy chain antibodies (VHHs). *J. Immunol. Methods* **324**, 13–25 (2007).

51. Stanfield, R. L., Dooley, H., Verdino, P., Flajnik, M. F. & Wilson, I. A. Maturation of Shark Single-domain (IgNAR) Antibodies: Evidence for Induced-fit Binding. *J. Mol. Biol.* **367**, 358–372 (2007).

52. Stanfield, R. L., Dooley, H., Flajnik, M. F. & Wilson, I. A. Crystal Structure of a Shark Single-Domain Antibody V Region in Complex with Lysozyme. *Science (80-.).* **305**, 1770 LP – 1773 (2004).

53. Kirchhofer, A. *et al.* Modulation of protein properties in living cells using nanobodies. *Nat. Struct. Mol. Biol.* **17**, 133–138 (2010).

54. Sircar, A., Sanni, K. A., Shi, J. & Gray, J. J. Analysis and Modeling of the Variable Region of Camelid Single-Domain Antibodies. *J. Immunol.* **186**, 6357–6367 (2011).

55. Duggan, S. Caplacizumab: First Global Approval. *Drugs* **78**, 1639–1642 (2018).

56. Morrison, C. Nanobody approval gives domain antibodies a boost. *Nat. Rev. Drug Discov.* **18**, 485–487 (2019).

57. Walsh, F. S. Ossianix and Novo Nordisk sign Research and Option Agreement on Blood Brain Barrier delivery technology for proteins in diabetes and other metabolic diseases [Press release]. (2018). Available at: www.ossianix.com. (Accessed: 2nd September 2019)

58. Manglik, A., Kobilka, B. K. & Steyaert, J. Nanobodies to Study G Protein–Coupled Receptor Structure and Function. *Annu. Rev. Pharmacol. Toxicol.* **57**, 19–37 (2017).

59. Crowe, J. S. *et al.* Preclinical Development of a Novel, Orally-Administered Anti-Tumour Necrosis Factor Domain Antibody for the Treatment of Inflammatory Bowel Disease. *Sci. Rep.* **8**, 4941 (2018).

60. Laursen, N. S. *et al.* Universal protection against influenza infection by a multidomain

antibody to influenza hemagglutinin. *Science (80-.).* **362**, 598–602 (2018).

61. Woof, J. M. & Burton, D. R. Human antibody–Fc receptor interactions illuminated by crystal structures. *Nat. Rev. Immunol.* **4**, 89–99 (2004).

62. Ward, E. S. *et al.* From Sorting Endosomes to Exocytosis: Association of Rab4 and Rab11 GTPases with the Fc Receptor, FcRn, during Recycling. *Mol. Biol. Cell* **16**, 2028–2038 (2005).

63. Wu, A. M. & Senter, P. D. Arming antibodies: prospects and challenges for immunoconjugates. *Nat. Biotechnol.* **23**, 1137–1146 (2005).

64. Hodi, F. S. *et al.* Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N. Engl. J. Med.* **363**, 711–723 (2010).

65. Weiner, G. J. Rituximab: Mechanism of Action. *Semin. Hematol.* **47**, 115–123 (2010).

66. Bobkov, V. *et al.* Nanobody-Fc constructs targeting chemokine receptor CXCR4 potently inhibit signaling and CXCR4-mediated HIV-entry and induce antibody effector functions. *Biochem. Pharmacol.* **158**, 413–424 (2018).

67. Van den Eynde, B. J. & Scott, A. M. Encylopedia of Immunology. (1998).

68. Herce, H. D. *et al.* Cell-permeable nanobodies for targeted immunolabelling and antigen manipulation in living cells. *Nat. Chem.* **9**, 762–771 (2017).

69. Slastnikova, T. A., Ulasov, A. V., Rosenkranz, A. A. & Sobolev, A. S. Targeted Intracellular Delivery of Antibodies: The State of the Art. *Front. Pharmacol.* **9**, 1208 (2018).

70. Bruce, V. J., Lopez-Islas, M. & McNaughton, B. R. Resurfaced cell-penetrating nanobodies: A potentially general scaffold for intracellularly targeted protein discovery. *Protein Sci.* **25**, 1129–37 (2016).

71. Gunnoo, S. B. *et al.* Creation of a gated antibody as a conditionally functional synthetic protein. *Nat. Commun.* **5**, 4388 (2014).

72. Comer, F. I. & Hart, G. W. O-Glycosylation of nuclear and cytosolic proteins. Dynamic interplay between O-GlcNAc and O-phosphate. *J. Biol. Chem.* **275**, 29179–82 (2000).

73. Gunnoo, S. B. Site-Specific Chemical Modification of Antibodies for the Modulation of Function. (University of Oxford, 2013).

74. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science (80-.).* **354**, (2016).

75. Marshall, S., Bacote, V. & Traxinger, R. R. Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role of hexosamine biosynthesis in the induction of insulin resistance. *J. Biol. Chem.* **266**, 4706–12 (1991).

76. Comer, F. I. & Hart, G. W. *O* -Glycosylation of Nuclear and Cytosolic Proteins. *J. Biol. Chem.* **275**, 29179–29182 (2000).

77. Hart, G. W. & Akimoto, Y. *The O-GlcNAc Modification*. *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2009).

78. Varki, A., Esko, J. D. & Colley, K. J. *Cellular Organization of Glycosylation*. *Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 2009).

79. ZACHARA, N. & Hart, G. W. O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. *Biochim. Biophys. Acta - Gen. Subj.* **1673**, 13–28 (2004).

80. Hart, G. W., Slawson, C., Ramirez-Correa, G. & Lagerlof, O. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu. Rev. Biochem.* **80**, 825–858 (2011).

81. Lucena, M. C. *et al.* Epithelial Mesenchymal Transition Induces Aberrant Glycosylation through Hexosamine Biosynthetic Pathway Activation. *J. Biol. Chem.*

291, 12917–12929 (2016).

82. Vasconcelos-dos-Santos, A. *et al.* Biosynthetic Machinery Involved in Aberrant Glycosylation: Promising Targets for Developing of Drugs Against Cancer. *Front. Oncol.* **5**, 138 (2015).

83. Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564 (2011).

84. Li, B., Li, H., Lu, L. & Jiang, J. Structures of human O-GlcNAcase and its complexes reveal a new substrate recognition mode. *Nat. Struct. Mol. Biol.* **24**, 362–369 (2017).

85. Elsen, N. L. *et al.* Insights into activity and inhibition from the crystal structure of human O-GlcNAcase. *Nat. Chem. Biol.* **13**, 613–615 (2017).

86. Dennis, R. J. *et al.* Structure and mechanism of a bacterial β-glucosaminidase having O-GlcNAcase activity. *Nat. Struct. Mol. Biol.* **13**, 365–371 (2006).

87. Schimpl, M., Borodkin, V. S., Gray, L. J. & van Aalten, D. M. F. Synergy of Peptide and Sugar in O-GlcNAcase Substrate Recognition. *Chem. Biol.* **19**, 173–178 (2012).

88. Martin, J. C., Fadda, E., Ito, K. & Woods, R. J. Defining the structural origin of the substrate sequence independence of O-GlcNAcase using a combination of molecular docking and dynamics simulation. *Glycobiology* **24**, 85–96 (2014).

89. Ruan, H.-B., Singh, J. P., Li, M.-D., Wu, J. & Yang, X. Cracking the O-GlcNAc code in metabolism. *Trends Endocrinol. Metab.* **24**, 301–309 (2013).

90. Bond, M. R. & Hanover, J. A. *O*- GlcNAc Cycling: A Link Between Metabolism and Chronic Disease. *Annu. Rev. Nutr.* **33**, 205–229 (2013).

91. Yang, X. *et al.* Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature* **451**, 964–969 (2008).

92. Yuzwa, S. A. *et al.* Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat. Chem. Biol.* **8**, 393–399 (2012).

93. Zhu, Y., Shan, X., Yuzwa, S. A. & Vocadlo, D. J. The emerging link between O-GlcNAc and Alzheimer disease. *J. Biol. Chem.* **289**, 34472–81 (2014).

94. Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W. & Gong, C.-X. O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10804–9 (2004).

95. Yi, W. *et al.* Phosphofructokinase 1 Glycosylation Regulates Cell Growth and Metabolism. *Science (80-.).* **337**, 975 LP – 980 (2012).

96. Fardini, Y., Dehennaut, V., Lefebvre, T. & Issad, T. O-GlcNAcylation: A New Cancer Hallmark? *Front. Endocrinol. (Lausanne).* **4**, 99 (2013).

97. Cummings, R. D. The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **5**, 1087 (2009).

98. Nairn, A. V *et al.* Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *J. Biol. Chem.* **283**, 17298–313 (2008).

99. Nairn, A. V, Moremen, K. W., Glycotranscriptomics, R. & Cummings, M. P. Handbook of Glycomics. (2009).

100. Pratt, M. R. & Bertozzi, C. R. Synthetic glycopeptides and glycoproteins as tools for biology. *Chem. Soc. Rev.* **34**, 58 (2005).

101. Stanley, P., Schachter, H. & Taniguchi, N. N-Glycans. (2009).

102. Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13**, 448–462 (2012).

103. Schachter, H. The 'yellow brick road' to branched complex *N* -glycans. *Glycobiology* **1**, 453–461 (1991).

104. Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3–49 (2017).

105. Norris, G. E., Stillman, T. J., Anderson, B. F. & Baker, E. N. The three-dimensional structure of PNGase F, a glycosyl-asparaginase from Flavobacterium

meningosepticum. *Structure* **2**, 1049–1059 (1994).

106. Imaki, H. *et al.* Identification and characterization of a novel secreted glycosidase with multiple glycosidase activities in Streptococcus intermedius. *J. Bacteriol.* **196**, 2817–26 (2014).

107. Winchester, B. Lysosomal metabolism of glycoproteins. *Glycobiology* **15**, 1R-15R (2005).

108. Coleman, O. D. & Davis, B. G. *Development of a Conditionally Active 'Logic Gated' Anti- body through Site-Selective Chemical Modification [Rotation project report]*. (2015).

109. Collins, P. & Ferrier, R. Monosaccharides: Their Chemistry and Their Roles in Natural Products. in 60–71 (John Wiley & Sons, Ltd, 1995).

110. Davis, B. G. & Fairbanks, A. J. *Carbohydrate Chemistry*. (Oxford University Press, 2002).

111. Drago, R. S. *C-X and Sl-X Bond Energy Analyses with One-Parameter Scales and ECT. J. Phys. Chem* **95**, (W+, 1991).

112. Wright, T. H. Post-translational mutagenesis: radical methods for protein modification. (University of Oxford, 2017).

113. Gunnoo, S. B. *et al.* Creation of a gated antibody as a conditionally functional synthetic protein. *Nat. Commun.* **5**, 4388 (2014).

114. Fan, J. Q. & Lee, Y. C. Detailed studies on substrate structure requirements of glycoamidases A and F. *J. Biol. Chem.* **272**, 27058–64 (1997).

115. Parsons, T. B. *et al.* Optimal Synthetic Glycosylation of a Therapeutic Antibody. *Angew. Chemie Int. Ed.* **55**, 2361–2367 (2016).

116. Fern andez-Gonz alez, M. *et al.* Site-selective chemoenzymatic construction of synthetic glycoproteins using endoglycosidases †. doi:10.1039/c0sc00265h

117. Goodfellow, J. J. *et al.* An Endoglycosidase with Alternative Glycan Specificity Allows Broadened Glycoprotein Remodelling. *J. Am. Chem. Soc* **134**, 8033 (2012).

118. Macmillan, D. *et al* Solid-Phase Synthesis of Thioether-Linked Glycopeptide Mimics for Application to Glycoprotein Semisynthesis. *Org. Lett.* **4,** 1467-1470 (2002).

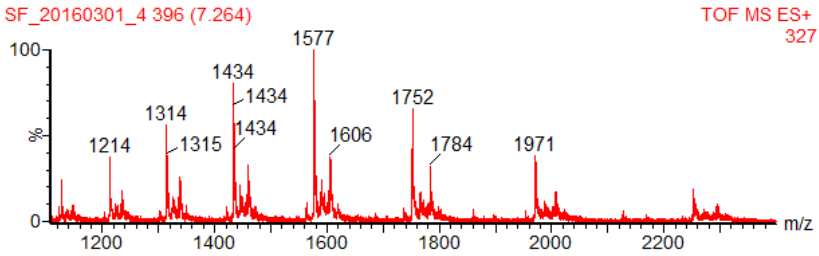
4.6 Appendix

4.6.1. Replicate LCMS-ESI+ spectra from the synthesis of cAbLys3-A104-N-GlcNAc

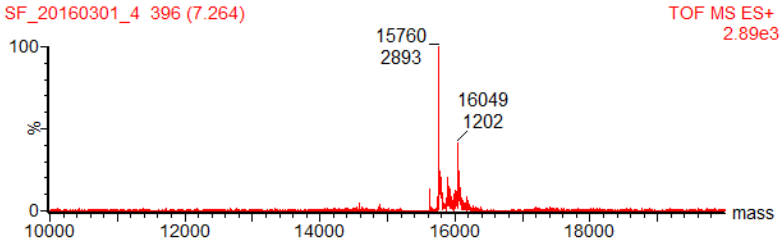
4.6.1.1. Synthesis of cAb-Lys3-A104-*N*-GlcNAc using β-6-(4-iodobutyramido)-GlcNAc

See section 4.4.3.3. for additional data.

LCMS-ESI+ characterization- ion series



Deconvoluted mass spectrum



cAb-Lys3-*N*-GlcNAc: Expected mass 16049 Da, observed mass 16049 Da (26.6%)

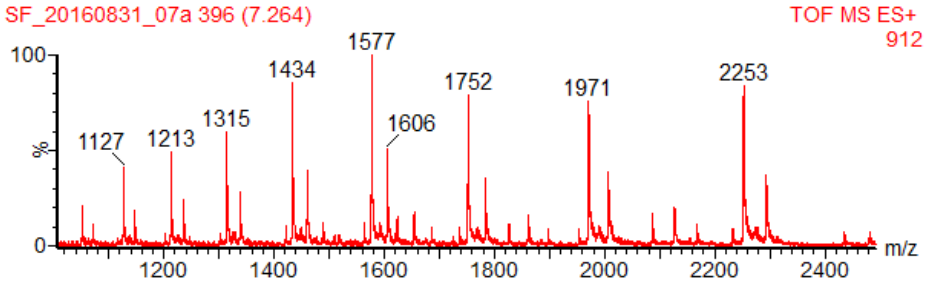
cAb-Lys3-A_{rac}: Expected mass 15765 Da, observed 15760 Da.

4.6.1.2. Synthesis of cAb-Lys3-A104-*N*-GlcNAc using β-6-(4-bromobutyramido)-GlcNAc

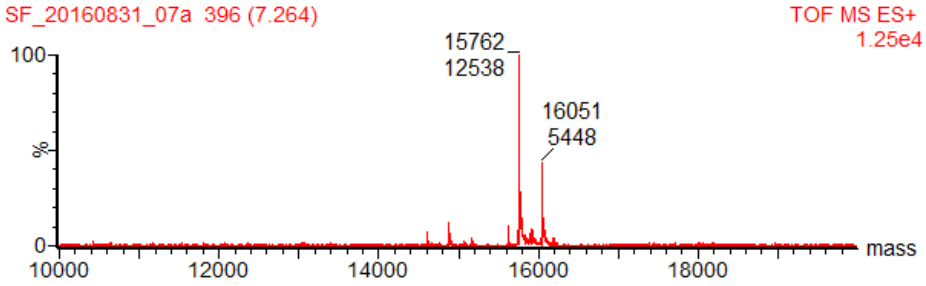
See section 4.4.3.4. for additional data.

Reaction 1

LCMS-ESI+ characterization- ion series



Deconvoluted mass spectrum

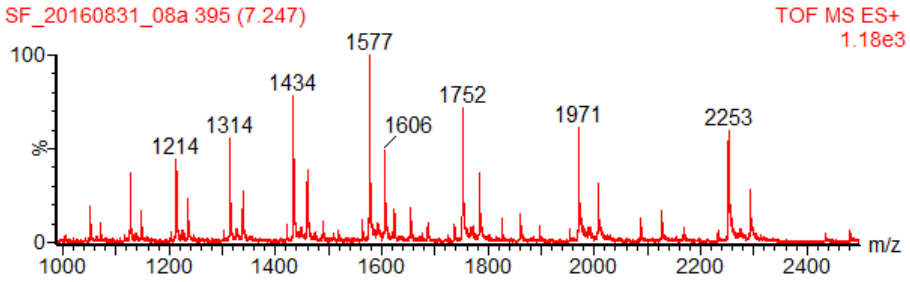


cAb-Lys3-*N*-GlcNAc: Expected mass 16049 Da, observed mass 16051 Da (30%)

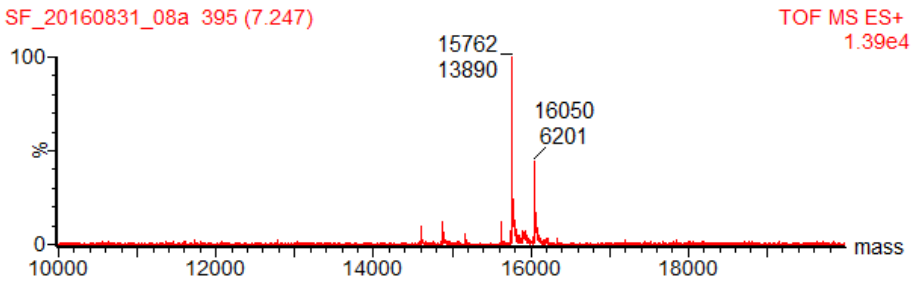
cAb-Lys3-A_{rac}: Expected mass 15765 Da, observed 15762 Da.

Reaction 2

LCMS-ESI+ characterization- ion series



Deconvoluted mass spectrum



cAb-Lys3-*N*-GlcNAc: Expected mass 16049 Da, observed mass 16050 Da (31%)

cAb-Lys3-A_{rac}: Expected mass 15765 Da, observed 15762 Da.

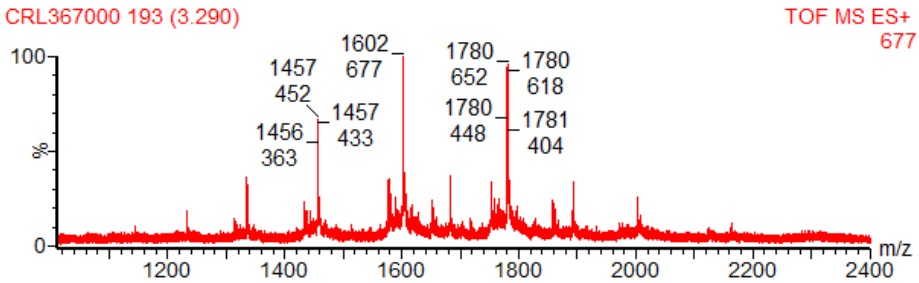
4.6.2. Replicate LCMS-ESI+ spectra from enzymatic reactions with modified cAbLys3

4.6.2.1. Hydrolysis of cAb-Lys3-*O*-GlcNAc by jbOGA

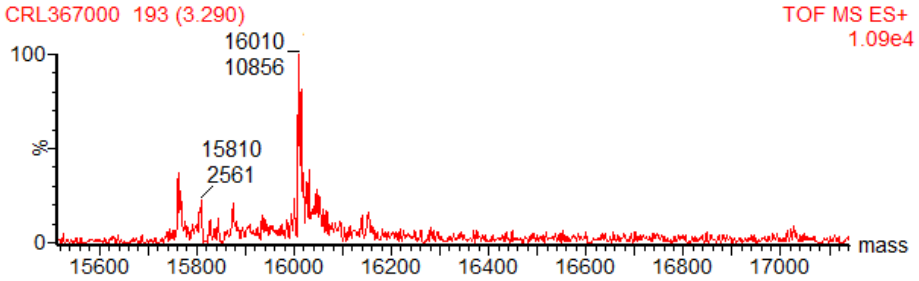
See section 4.4.6.1 for additional data.

Reaction 1

LCMS-ESI+ characterization- ion series



Deconvoluted mass spectrum

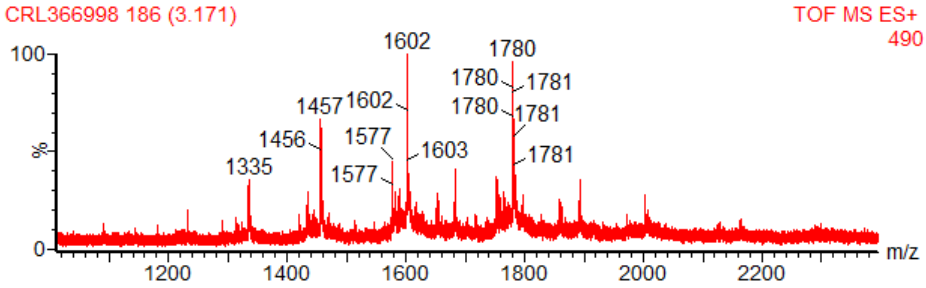


cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16010 Da.

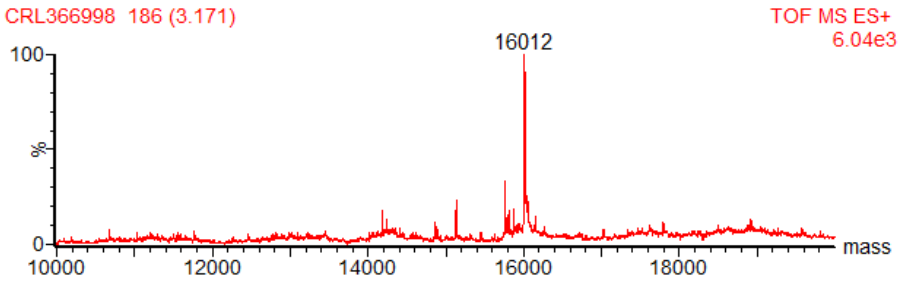
cAb-Lys3-A104hhSer: expected mass 15808 Da, observed 15810 Da (19%)

Reaction 2

LCMS-ESI+ characterization- ion series



Deconvoluted mass spectrum

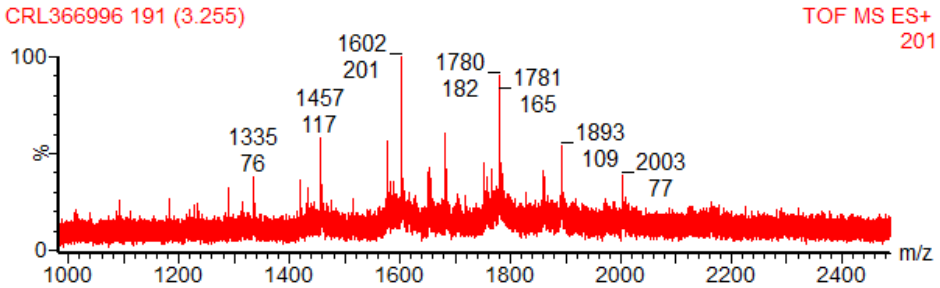


cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 160012 Da.

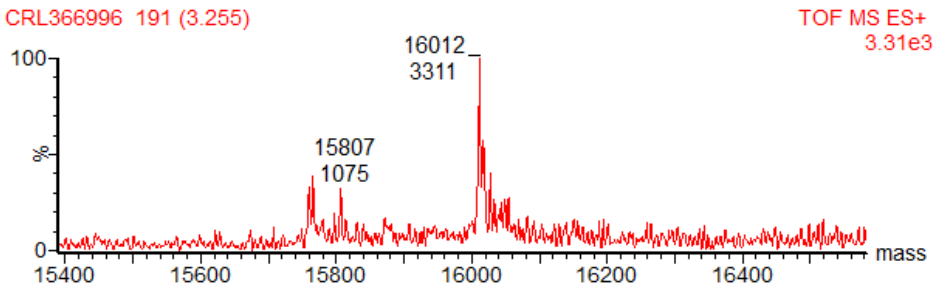
cAb-Lys3-A104hhSer: expected mass 15808 Da, not observed.

Reaction 3

LCMS-ESI+ characterization- ion series



Deconvoluted mass spectrum



cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 160012 Da.

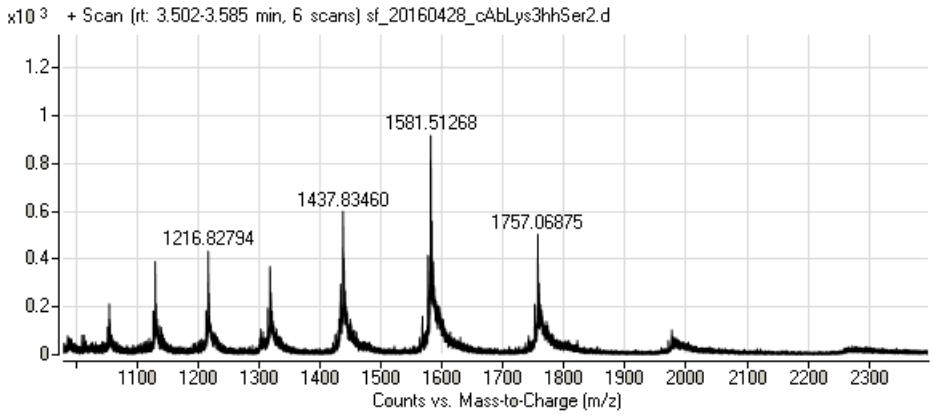
cAb-Lys3-A104hhSer expected mass 15808 Da, observed 15807 Da. (25%)

4.6.2.2. Hydrolysis of cAb-Lys3-*O*-GlcNAc by xmOGA

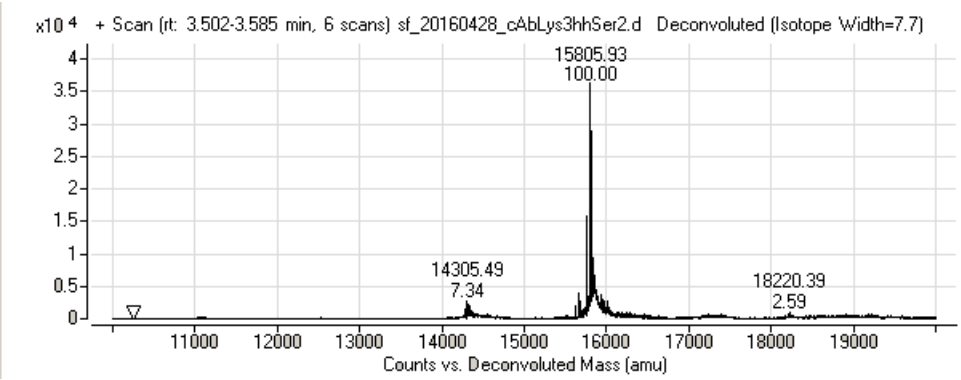
For additional data, see 4.4.6.2.

Reaction 1

LCMS-ESI+ characterization- ion series



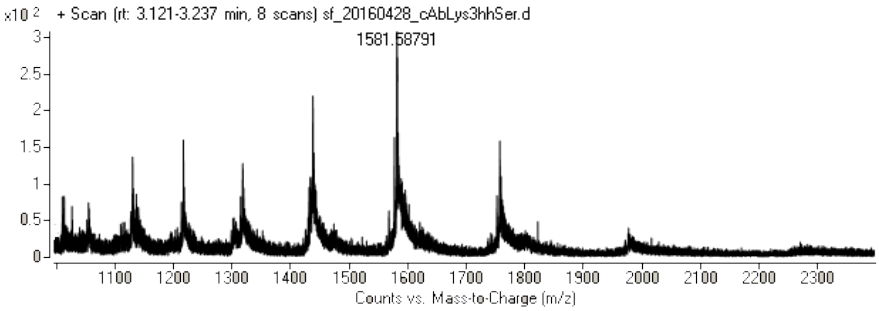
Deconvoluted mass spectrum



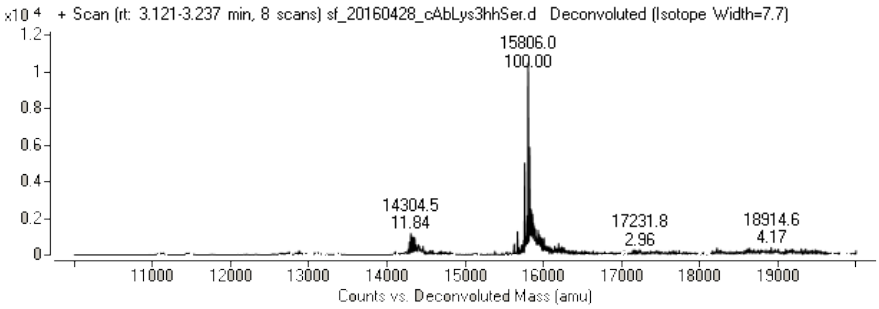
cAb-Lys3-A104-hhSer: expected mass 15808 Da, observed 15805.93 Da.

Reaction 2

LCMS-ESI+ characterization- ion series



Deconvoluted mass spectrum



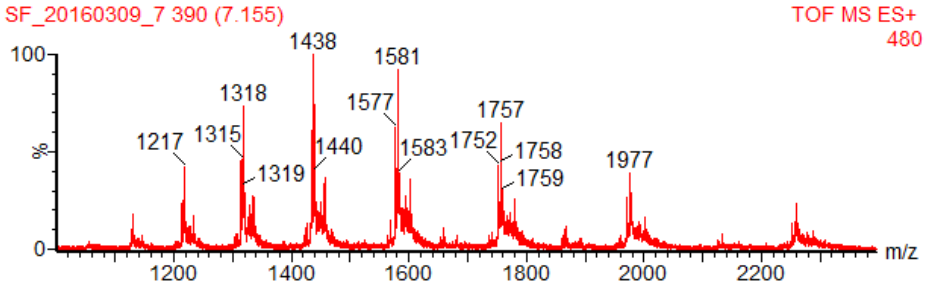
cAb-Lys3-A104-hhSer: expected mass 15808 Da, observed 15806.0 Da.

4.6.2.3. Optimization of OGA concentration for large-scale reactions

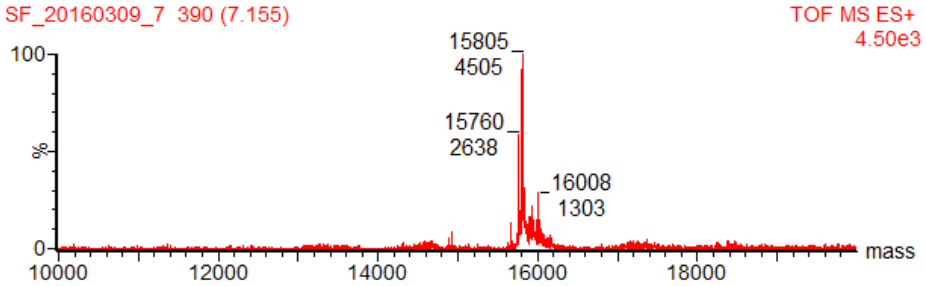
See 4.4.6.3 for further data.

2 U, 3h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-*O*-GlcNAc: expected mass 16012 Da, observed 16008 Da.

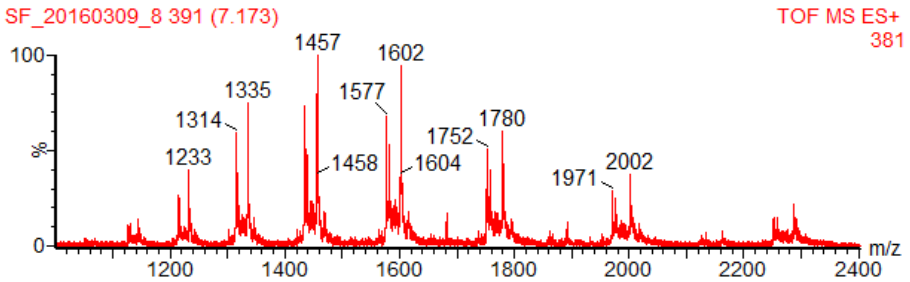
cAb-Lys3-hhSer: expected mass 15808 Da, observed 15805 Da.

cAb-Lys3-A_{rac} (from impure starting material): expected mass 15765 Da, observed 15760 Da.

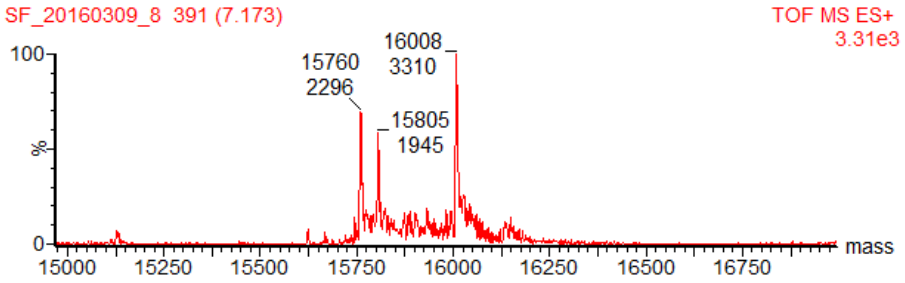
Conversion= 77%

1 U, 3h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-*O*-GlcNAc: expected mass 16012 Da, observed 16008 Da.

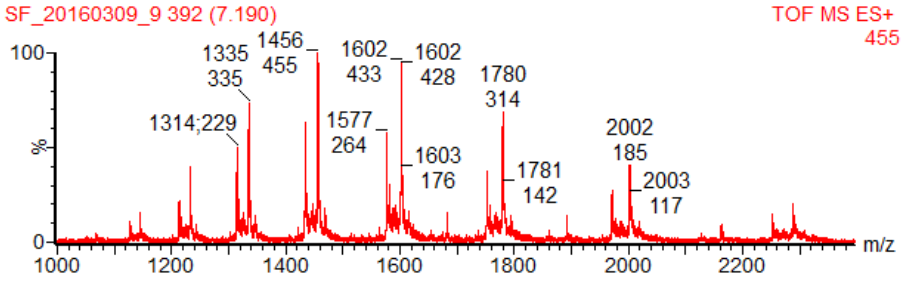
cAb-Lys3-hhSer: expected mass 15808 Da, observed 15805 Da.

cAb-Lys3-A_{rac} (from impure starting material): expected mass 15765 Da, observed 15760 Da.

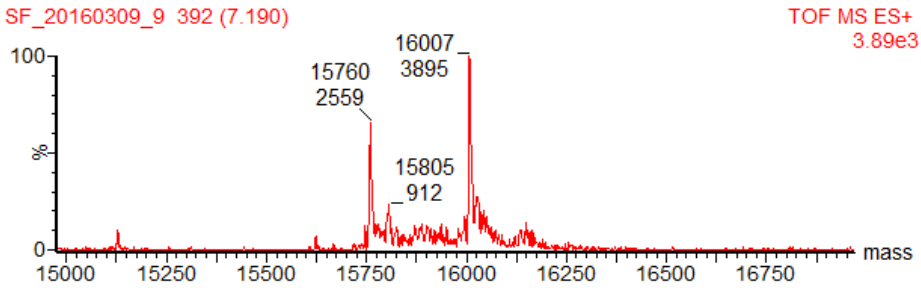
Conversion=37%

<u>0.5 U, 3h LCMS-ESI+ characterization</u>

Ion series



Deconvoluted mass spectrum



cAb-Lys3-*O*-GlcNAc: expected mass 16012 Da, observed 16007 Da.

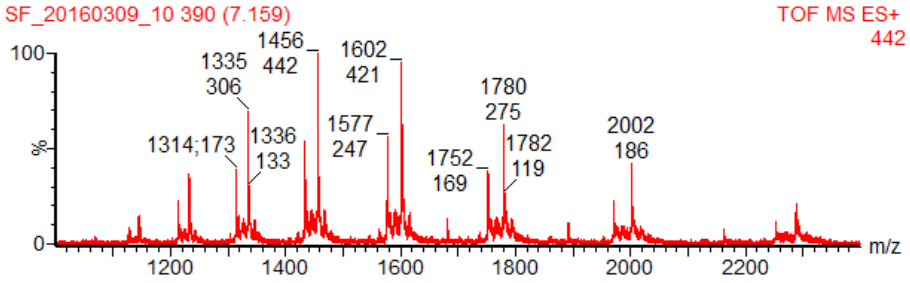
cAb-Lys3-hhSer: expected mass 15808 Da, observed 15805 Da.

cAb-Lys3-A_{rac} (from impure starting material): expected mass 15765 Da, observed 15760 Da.

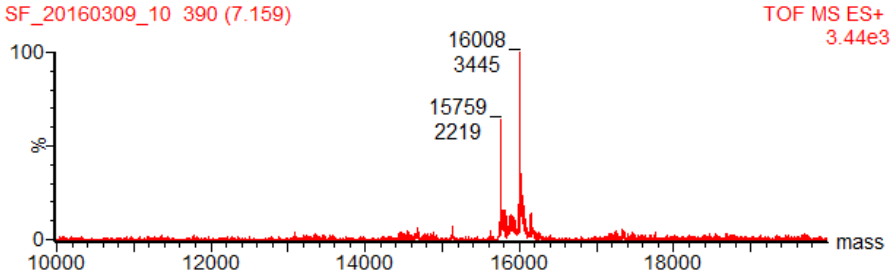
Conversion= 18%

0.25 U, 3h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



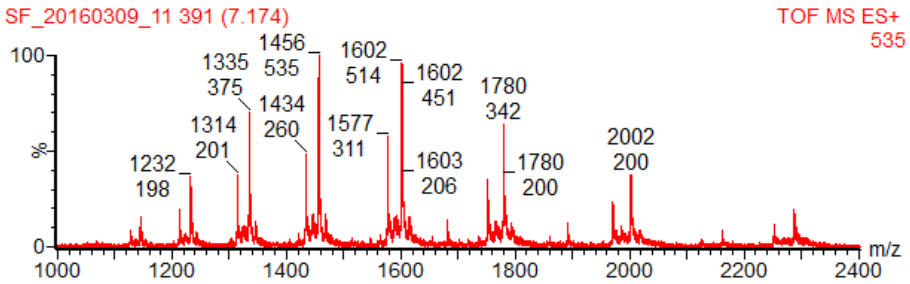
cAb-Lys3-*O*-GlcNAc: expected mass 16012 Da, observed 16008 Da.

cAb-Lys3-hhSer: expected mass 15808 Da, not observed.

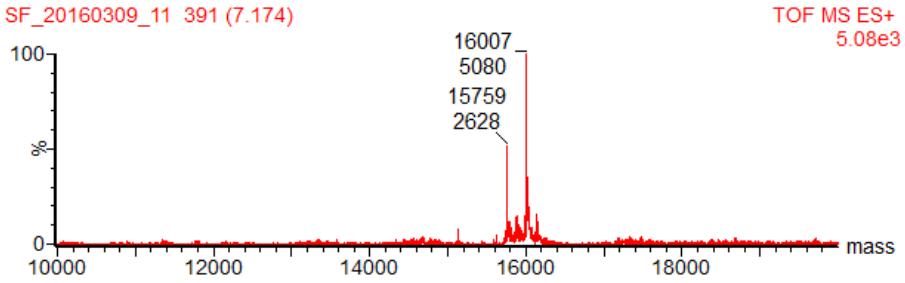
cAb-Lys3-A_{rac} (from impure starting material): expected mass 15765 Da, observed 15759 Da.

0.1 U, 3h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



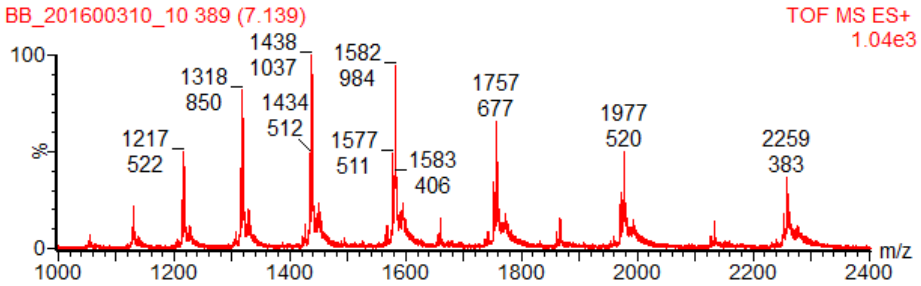
cAb-Lys3-*O*-GlcNAc: expected mass 16012 Da, observed 16007 Da.

cAb-Lys3-hhSer: expected mass 15808 Da, not observed.

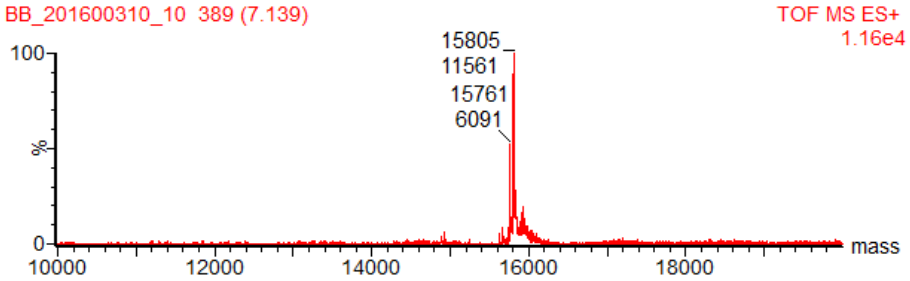
cAb-Lys3-A_{rac} (from impure starting material): expected mass 15765 Da, observed 15759 Da.

2 U, 24 h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



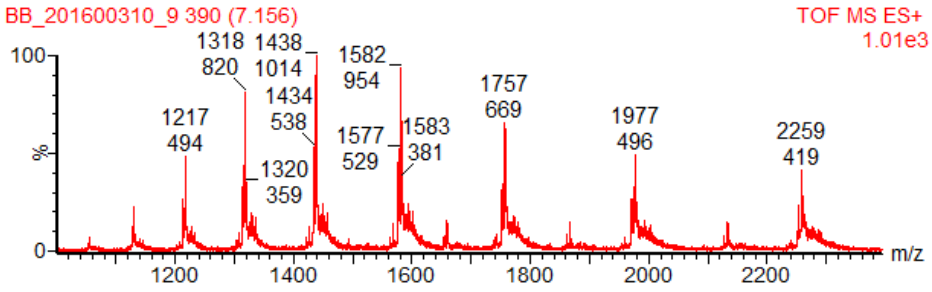
cAb-Lys3-hhSer: expected mass 15808 Da, observed 15805 Da.

cAb-Lys3-A_{rac} (from impure starting material): expected mass 15765 Da, observed 15760 Da.

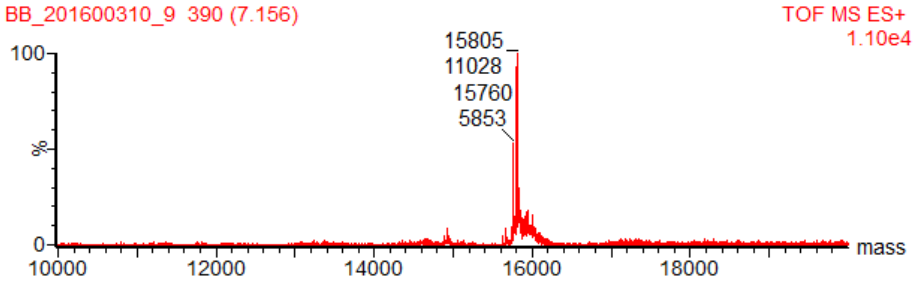
Conversion= 100%

1 U, 24 h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



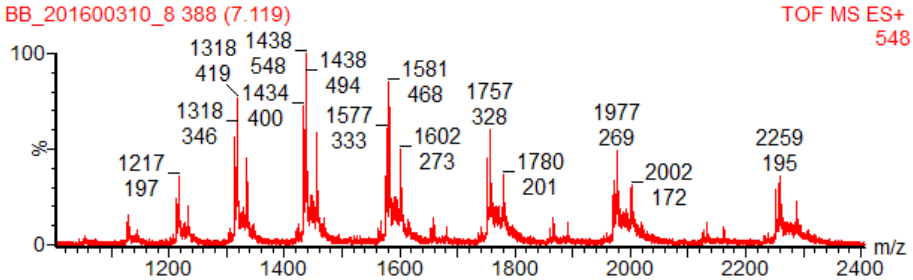
cAb-Lys3-hhSer: expected mass 15808 Da, observed 15805 Da.

cAb-Lys3-A_{rac} (from impure starting material): expected mass 15765 Da, observed 15760 Da.

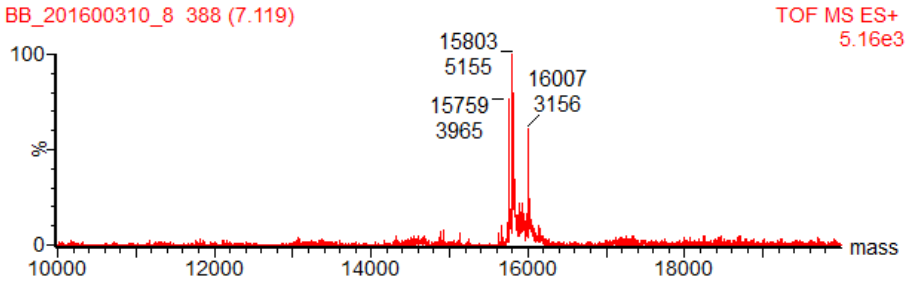
Conversion= 100%

0.5 U, 24 h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-*O*-GlcNAc: expected mass 16012 Da, observed 16007 Da.

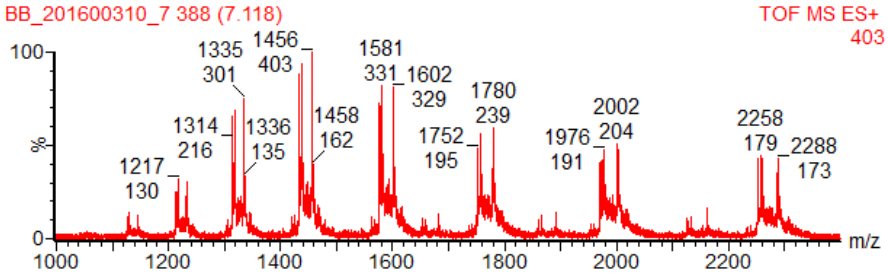
cAb-Lys3-hhSer: expected mass 15808 Da, observed 15803 Da.

cAb-Lys3-A_{rac} (from impure starting material): expected mass 15765 Da, observed 15759 Da.

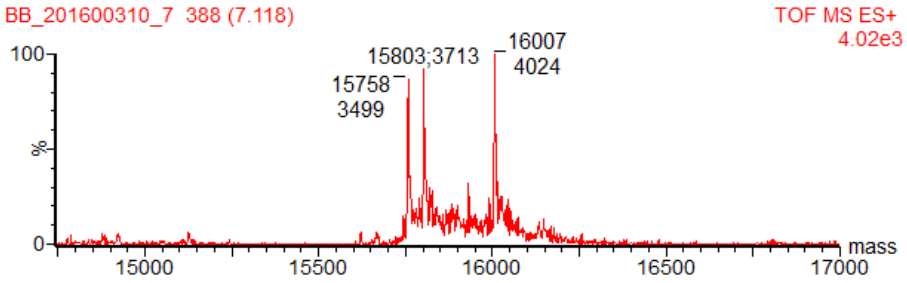
Conversion= 62%

<u>0.25 U, 24 h LCMS-ESI+ characterization</u>

Ion series



Deconvoluted mass spectrum



cAb-Lys3-*O*-GlcNAc: expected mass 16012 Da, observed 16007 Da.

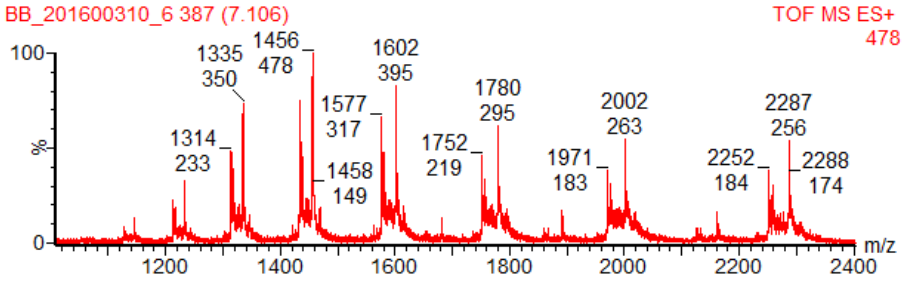
cAb-Lys3-hhSer: expected mass 15808 Da, observed 15803 Da.

cAb-Lys3-A_{rac} (from impure starting material): expected mass 15765 Da, observed 15758 Da.

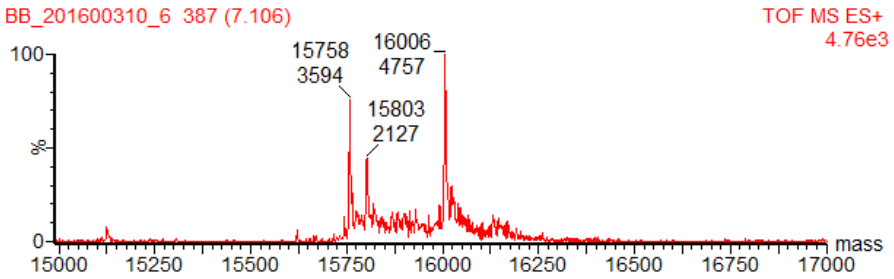
Conversion= 48%

0.1 U, 24 h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-*O*-GlcNAc: expected mass 16012 Da, observed 16006 Da.

cAb-Lys3-hhSer: expected mass 15808 Da, observed 15803 Da.

cAb-Lys3-A_{rac} (from impure starting material): expected mass 15765 Da, observed 15758 Da.

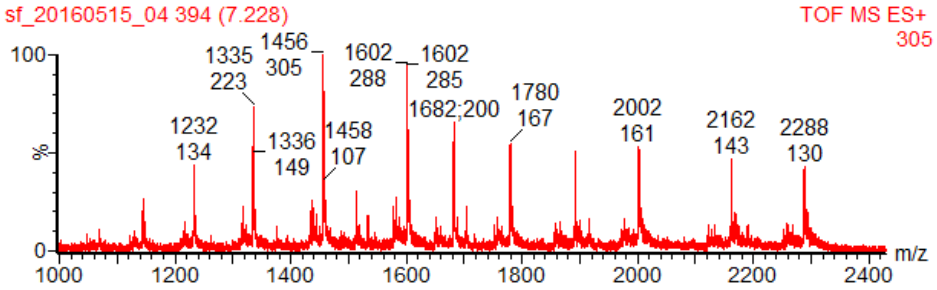
Conversion= 31%

4.6.2.4. Hydrolysis of cAbLys3-*O*-GlcNAc by hOGA

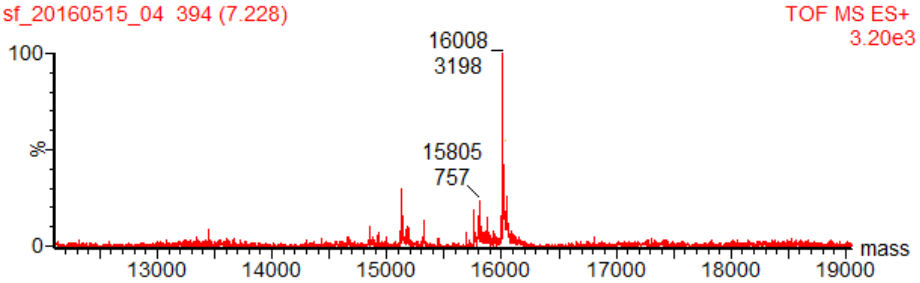
See 4.4.6.4 for further data.

Reaction 1, 24 h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



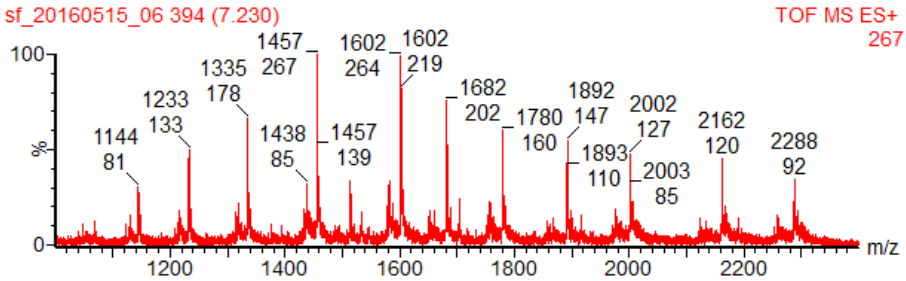
cAb-Lys3-*O*-GlcNAc: expected mass 16012 Da, observed 16008 Da.

cAb-Lys3-hhSer: expected mass 15808 Da, observed 15805 Da.

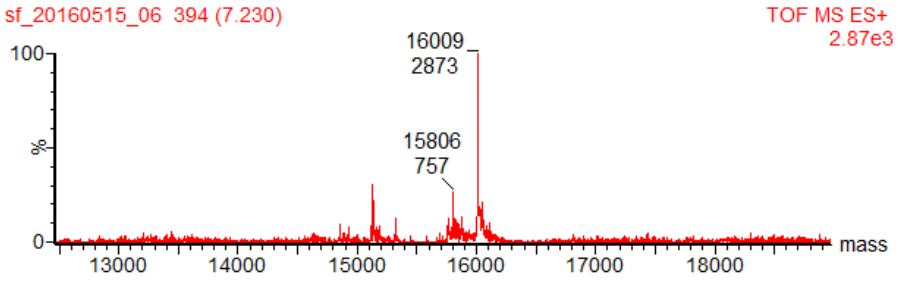
Conversion = 18.4%

Reaction 2, 24 h LCMS-ESI+ characterization

Ion series



Deconvoluted ion series



cAb-Lys3-*O*-GlcNAc: expected mass 16012 Da, observed 16009 Da.

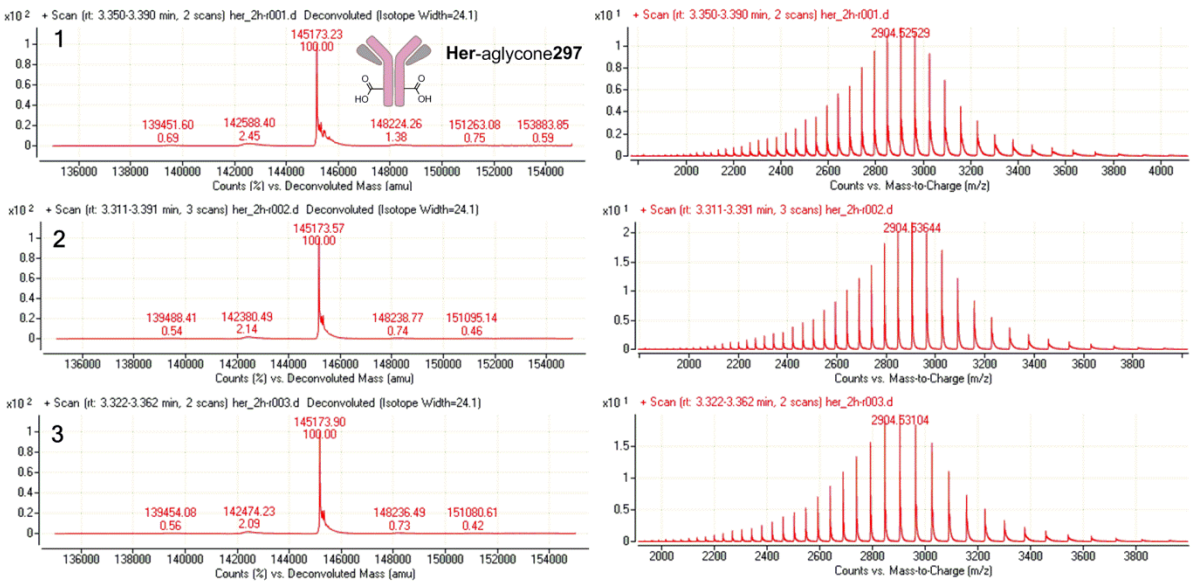
cAb-Lys3-hhSer: expected mass 15808 Da, observed 15806 Da.

Conversion= 21.2%

4.6.2.5. Control reactions with PNGase F cleavage of Herceptin; additional timepoints

For additional data, see 4.4.6.6.

t2 h LCMS-ESI+ characterization

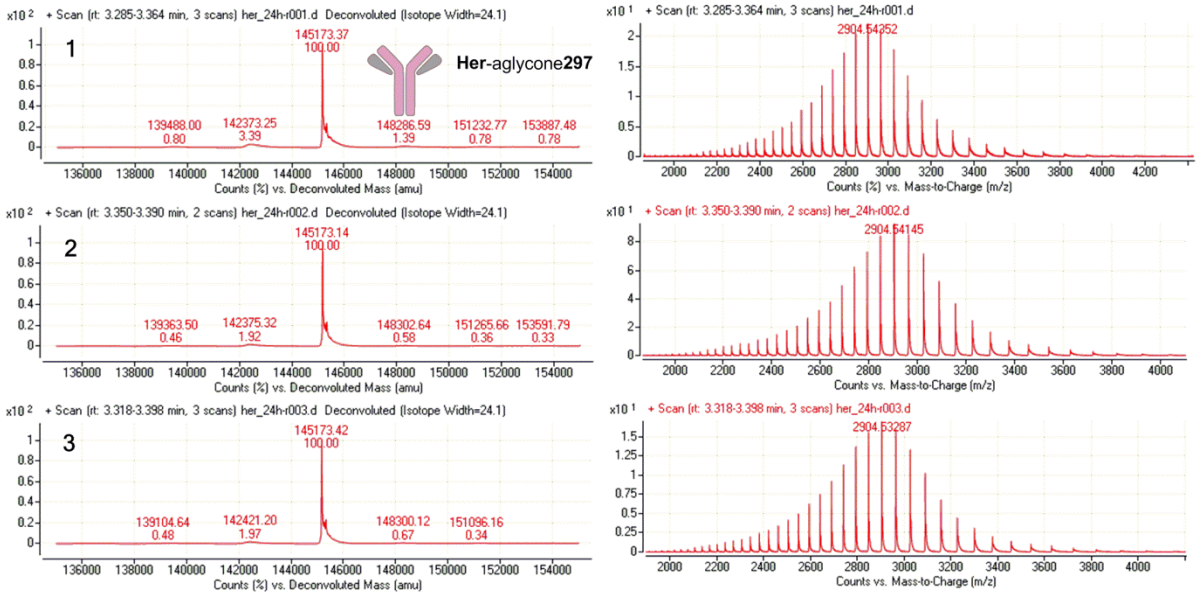


Her-aglycone297 expected mass: 145166

observed mass:145174

4. Constructing switchable antibodies through site-selective chemical glycosylation

t24 h LCMS-ESI+ characterization



Her-aglycone297 expected mass: 145166

observed mass: 145173

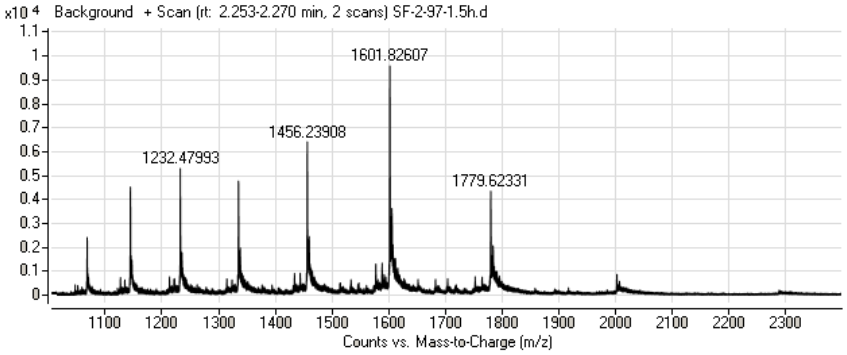
4.6.3. Additional data from testing addition of decasaccharide to cAb-Lys3-*O*-GlcNAc using EndoA and EndoS

For additional information see 4.4.9.

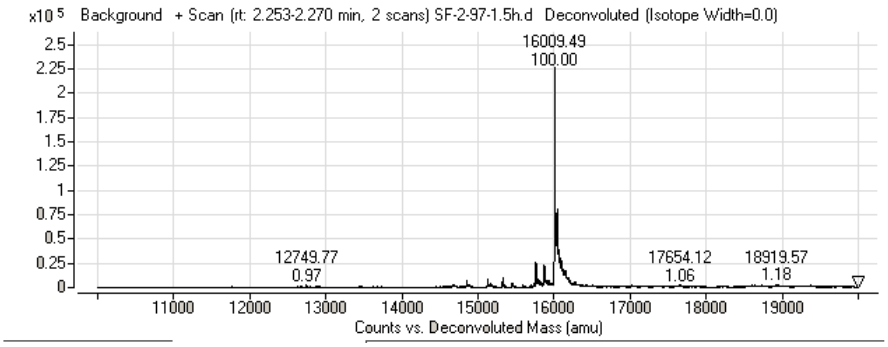
4.6.3.1. Endo A Y299F

1.5 h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum

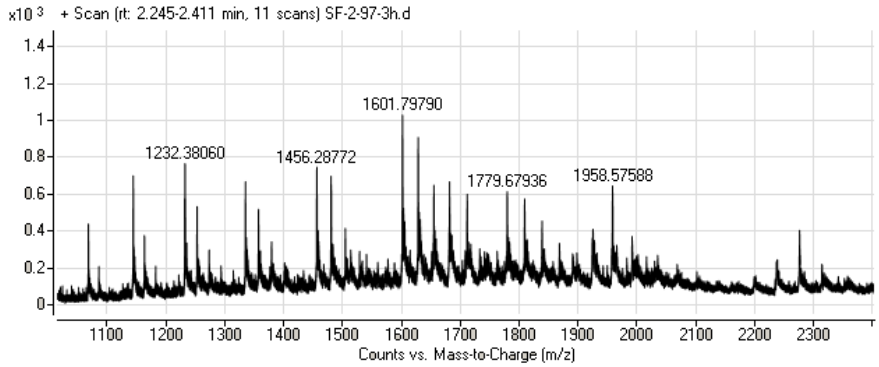


cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16009.49 Da.

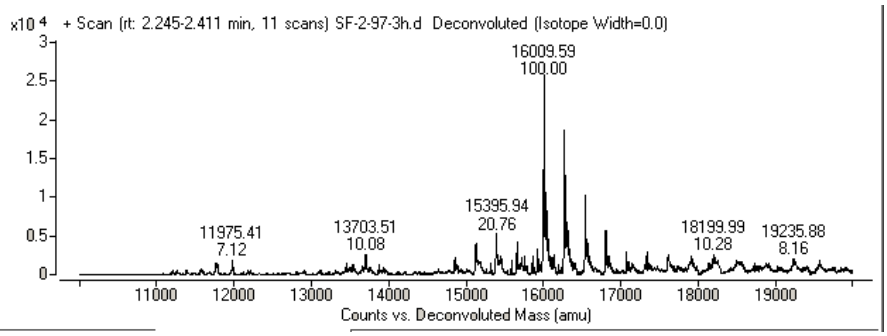
cAb-Lys3-A104-glycan: expected mass 18159 Da, none observed.

3 h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum



cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16009.59 Da.

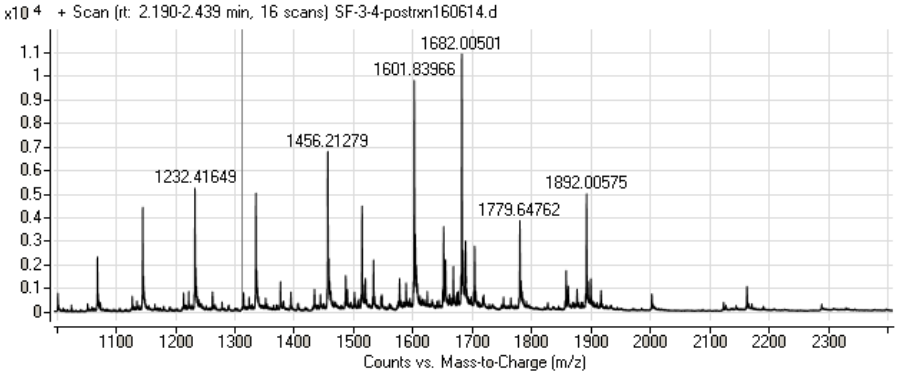
cAb-Lys3-A104-glycan: expected mass 18159 Da, none observed.

Note: Additional peaks are due to +266 adducts from column contamination.

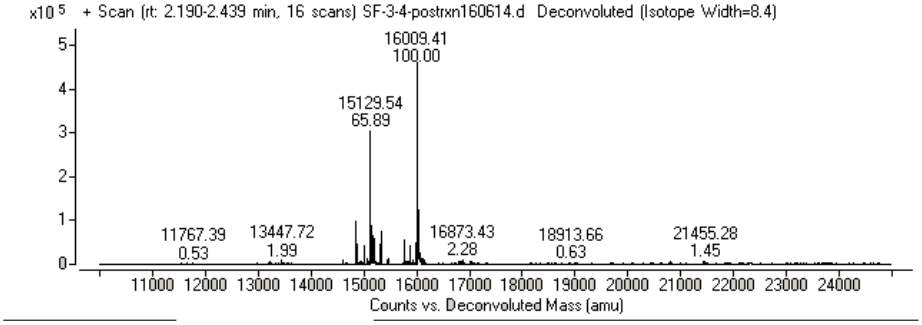
4.6.3.2. Endo S D233Q

1 h LCMS-ESI+ characterization

Ion series



Deconvoluted mass spectrum

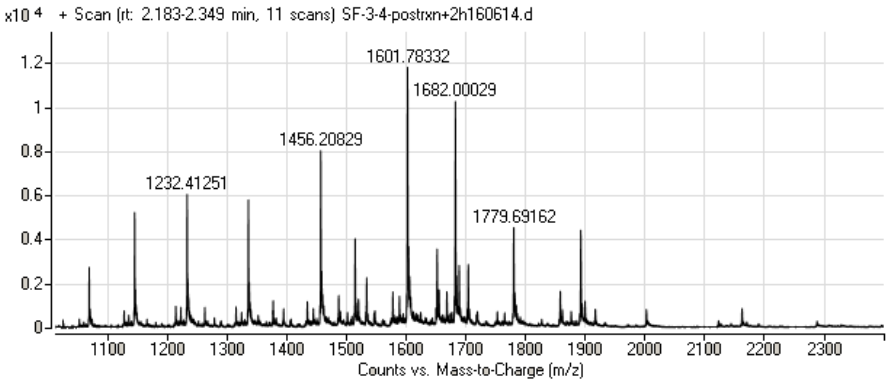


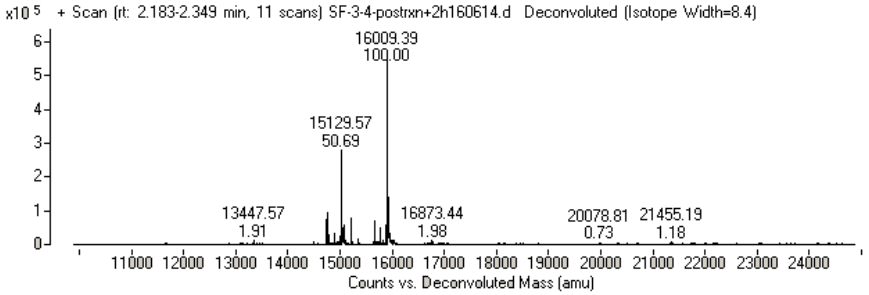
cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16009.41 Da.

cAb-Lys3-A104-glycan: expected mass 18159 Da, none observed.

3 h LCMS-ESI+ characterization

Ion series

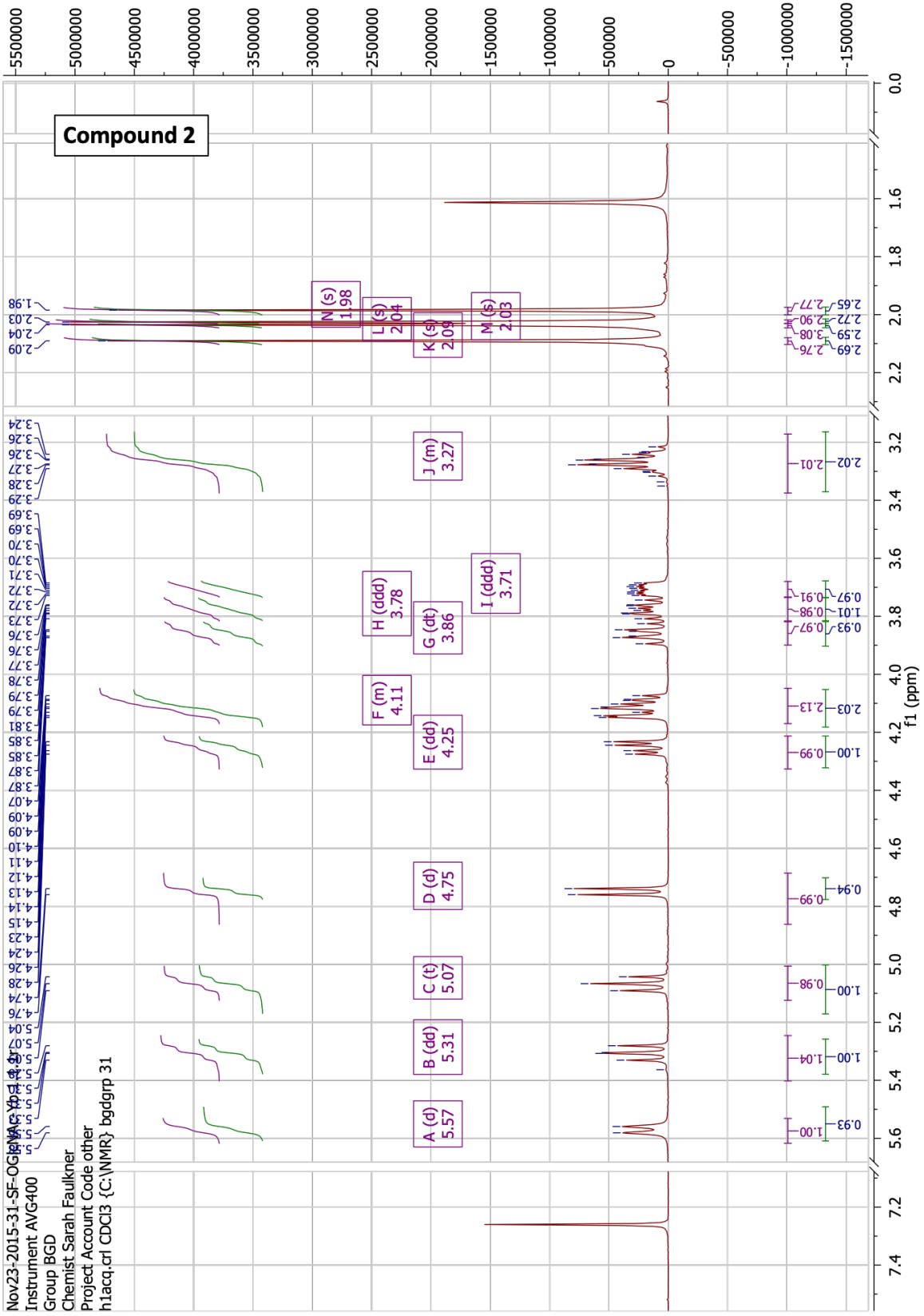


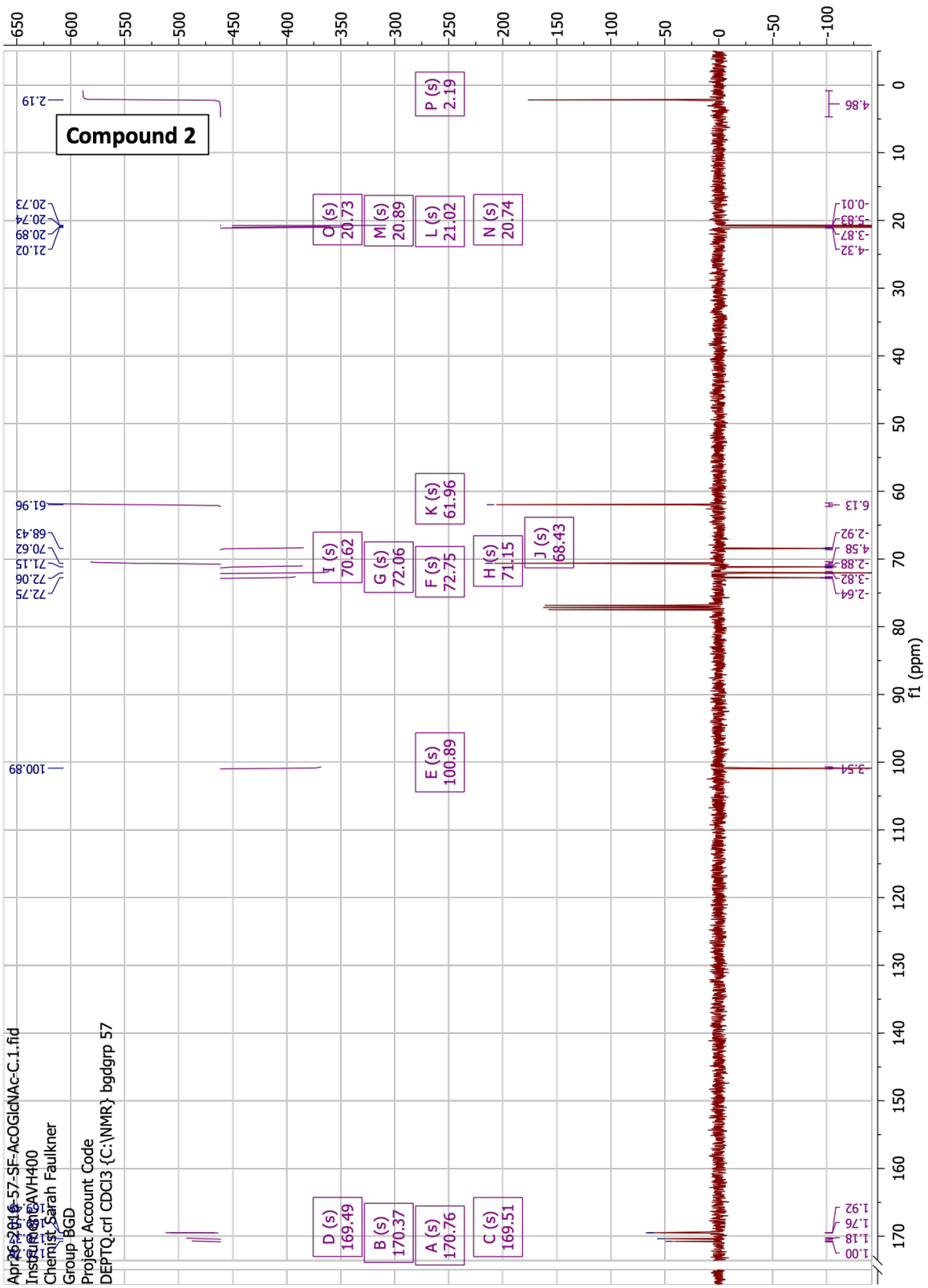


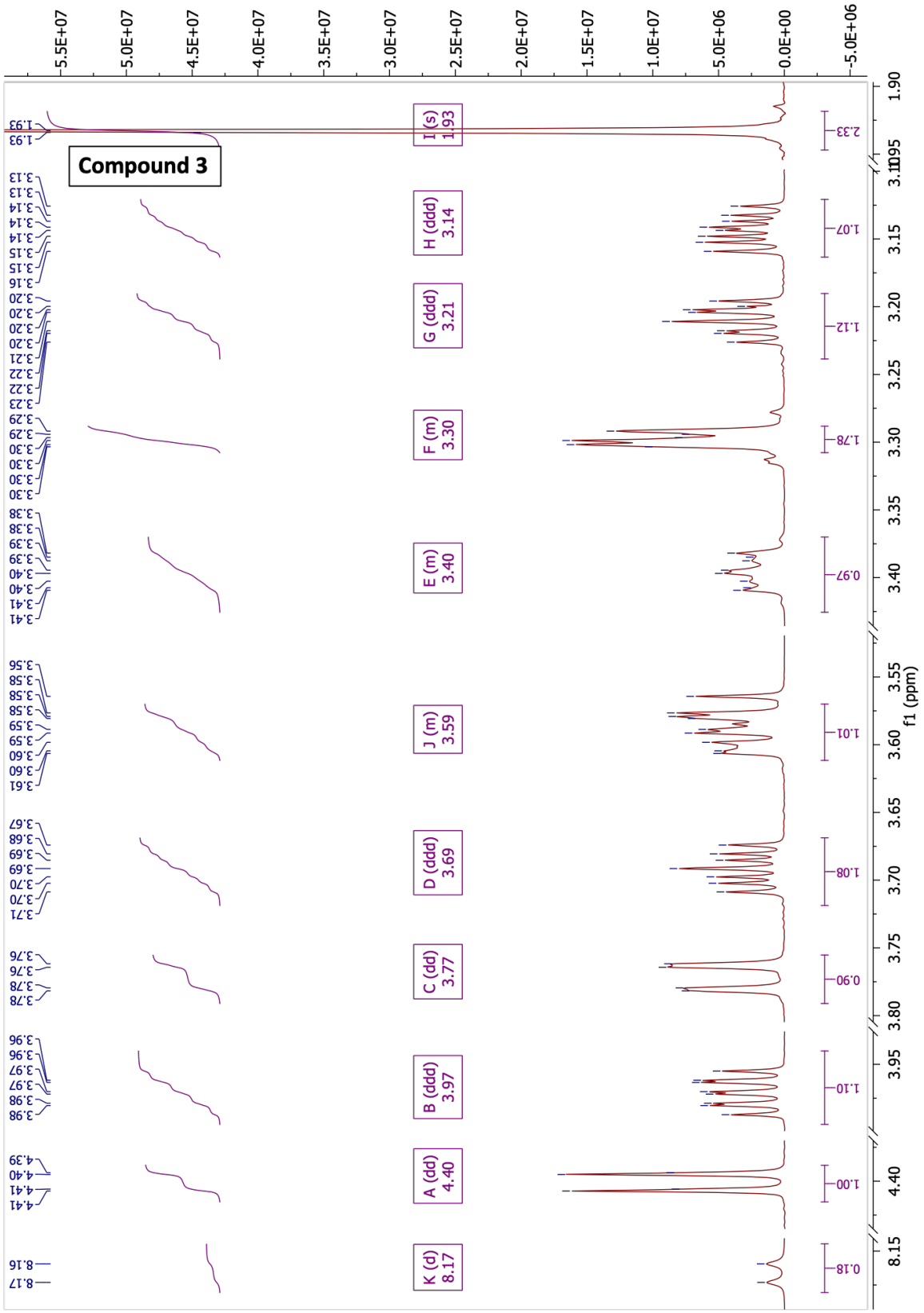
cAb-Lys3-A104-*O*-GlcNAc: expected mass 16012 Da, observed 16009.39 Da.

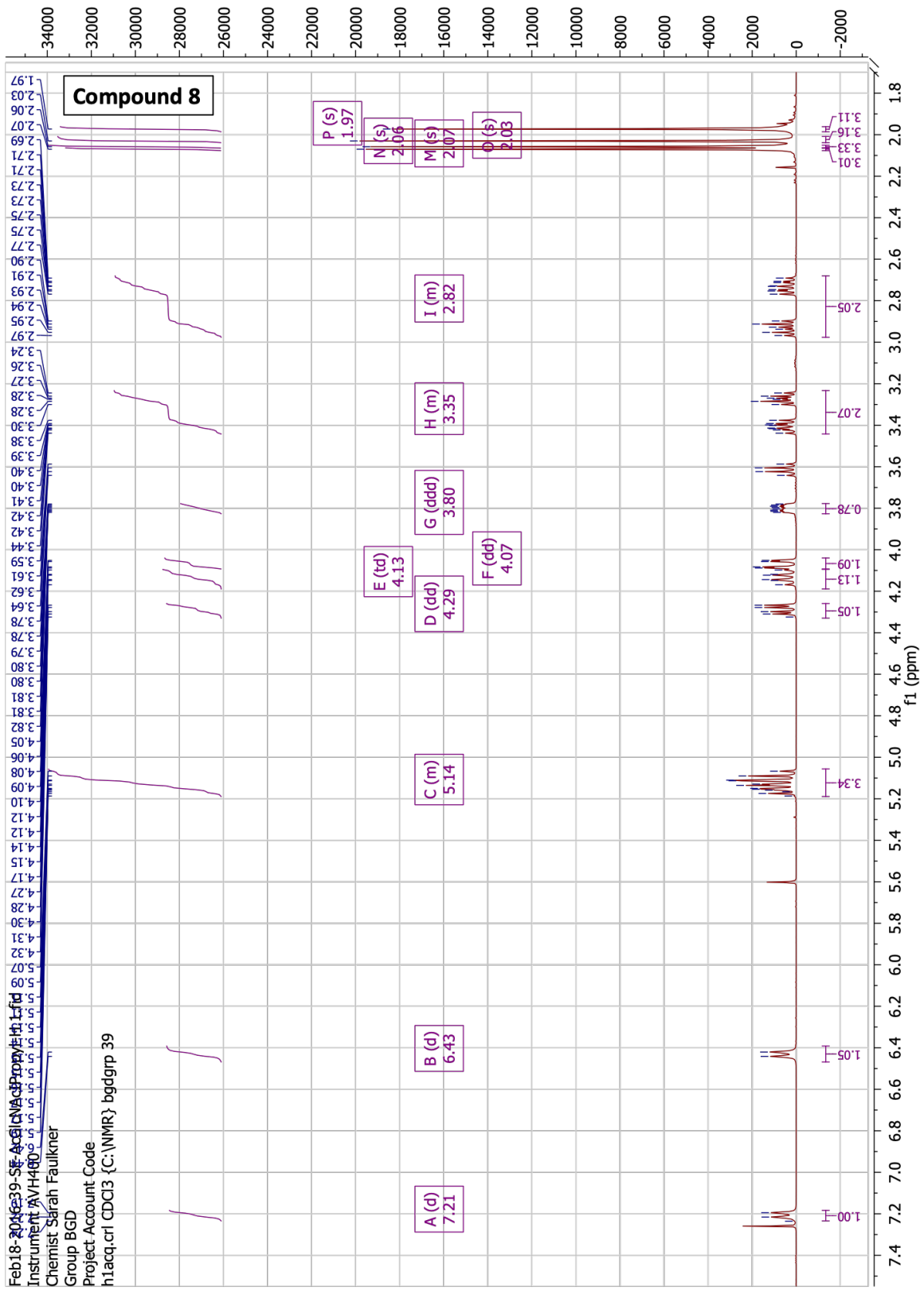
cAb-Lys3-A104-glycan: expected mass 18159 Da, none observed.

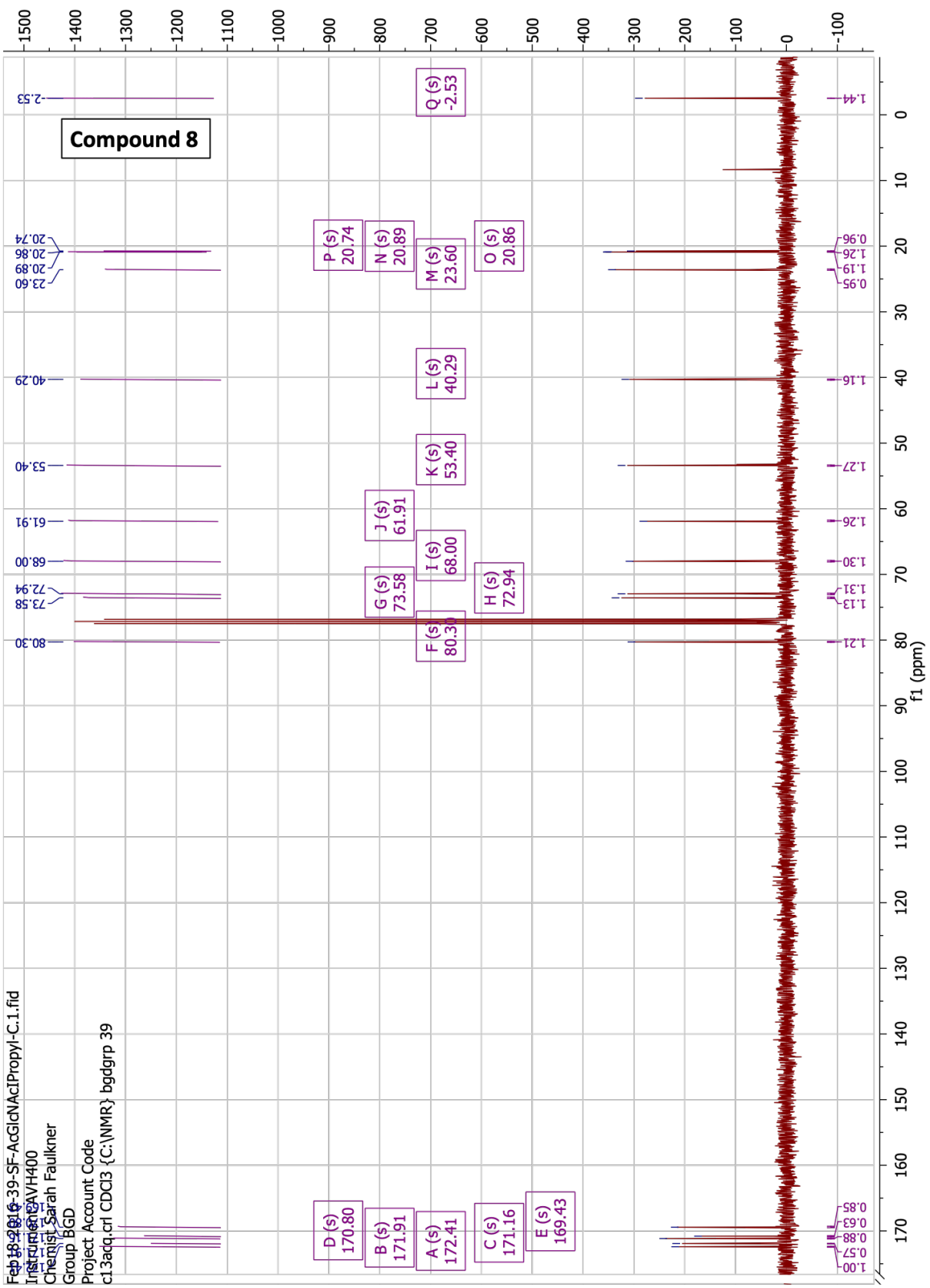
4.6.3 Raw NMR spectra

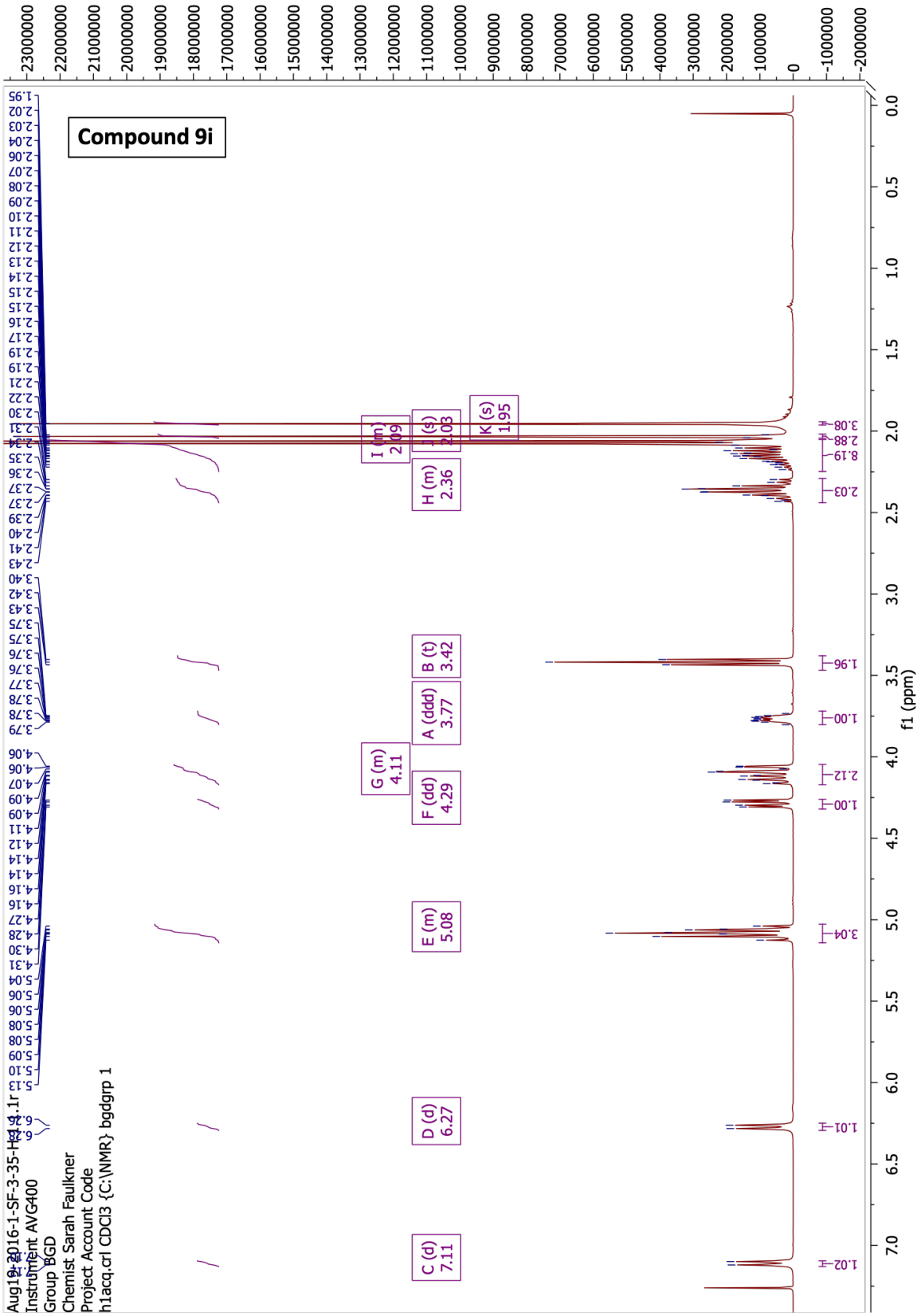


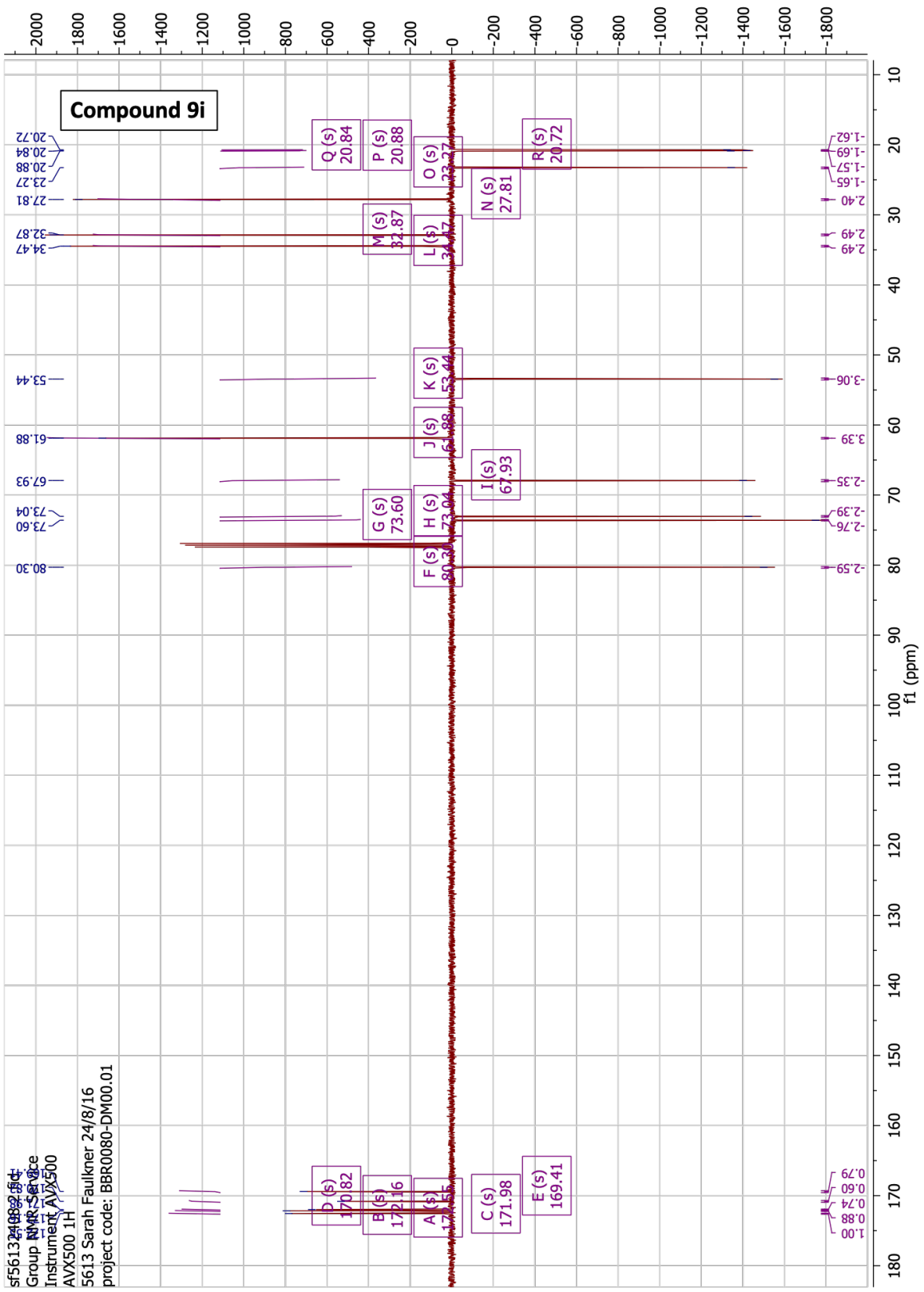

Compound 2
P (s) 2.19
O (s) 20.73
M (s) 20.89
L (s) 21.02
N (s) 20.74
K (s) 61.96
I (s) 70.62
G (s) 72.06
F (s) 72.75
H (s) 71.15
J (s) 68.43
E (s) 100.89
D (s) 169.49
B (s) 170.37
A (s) 170.76
C (s) 169.51
f1 (ppm)
Apr 26 2016 57-SF-AcOGlcNAc-C.1.fid
Project Account Code
Group BGD
Chemist Sarah Faulkner
DEPTQ.crl CDCl3 {C:\NMR} bgdgrp 57

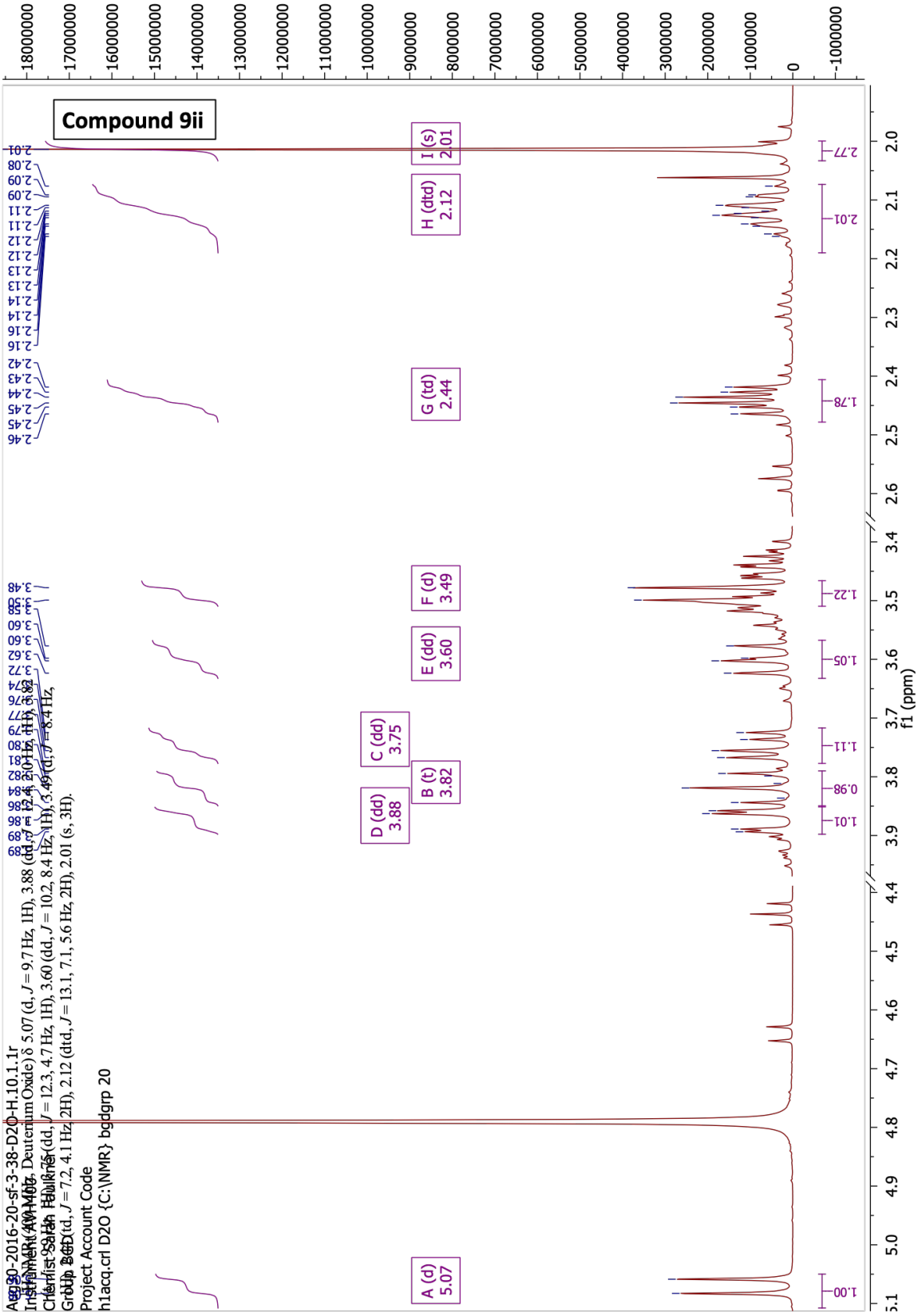

Compound 3
I (s) 1.93
H (ddd) 3.14
G (ddd) 3.21
F (m) 3.30
E (m) 3.40
J (m) 3.59
D (ddd) 3.69
C (dd) 3.77
B (ddd) 3.97
A (dd) 4.40
K (d) 8.17
f1 (ppm)

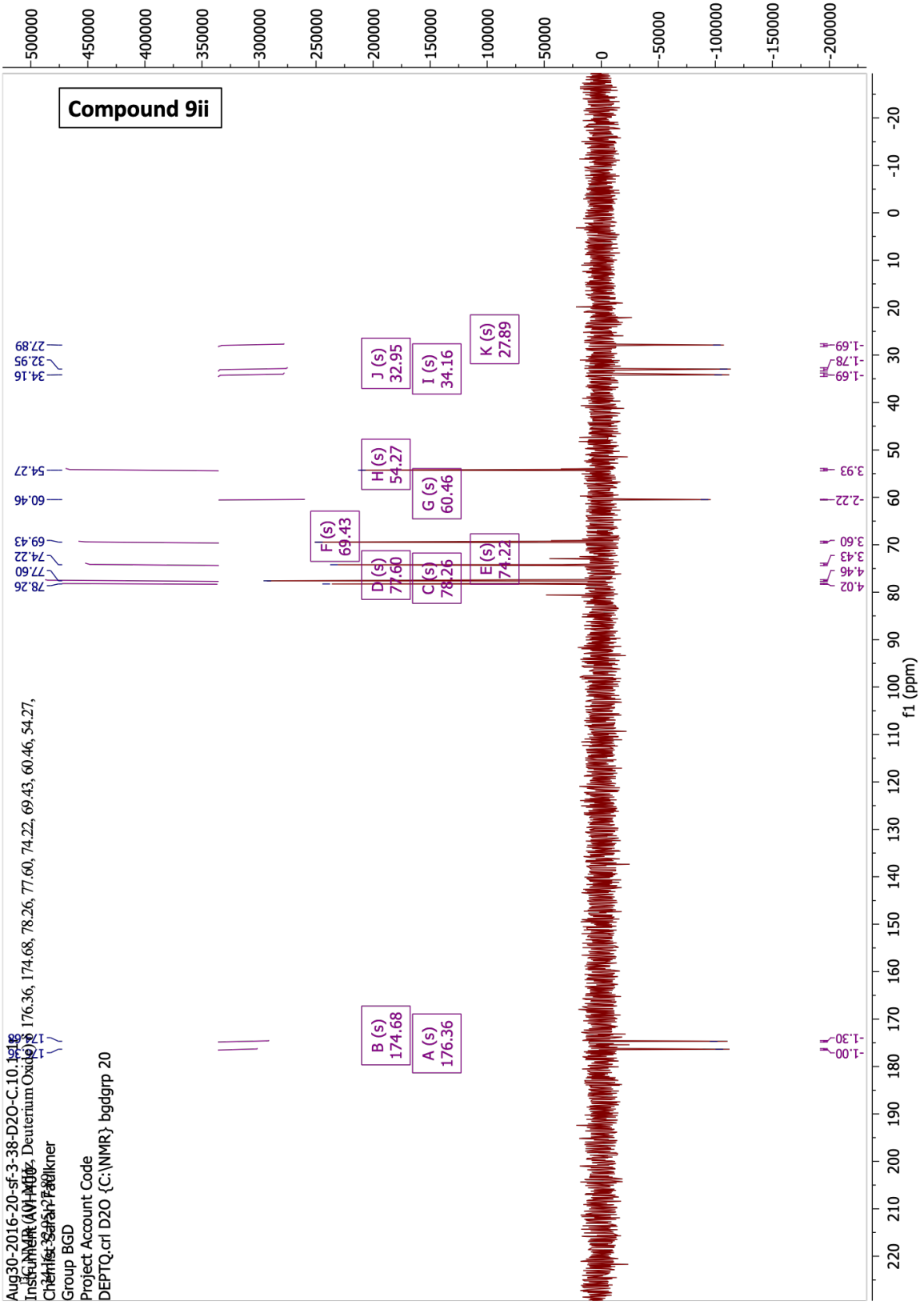

Compound 8
A (d) 7.21
B (d) 6.43
C (m) 5.14
D (dd) 4.29
E (td) 4.13
F (dd) 4.07
G (ddd) 3.80
H (m) 3.35
I (m) 2.82
M (s) 2.07
N (s) 2.06
O (s) 2.03
P (s) 1.97
f1 (ppm)
Feb18-2016e39-SE-AcGlcNAcPropyl-1.1.fid
Instrument AVH400
Chemist Sarah Faulkner
Group BGD
Project Account Code
h1acq.crl CDCl3 {C:\NMR} bgdgrp 39

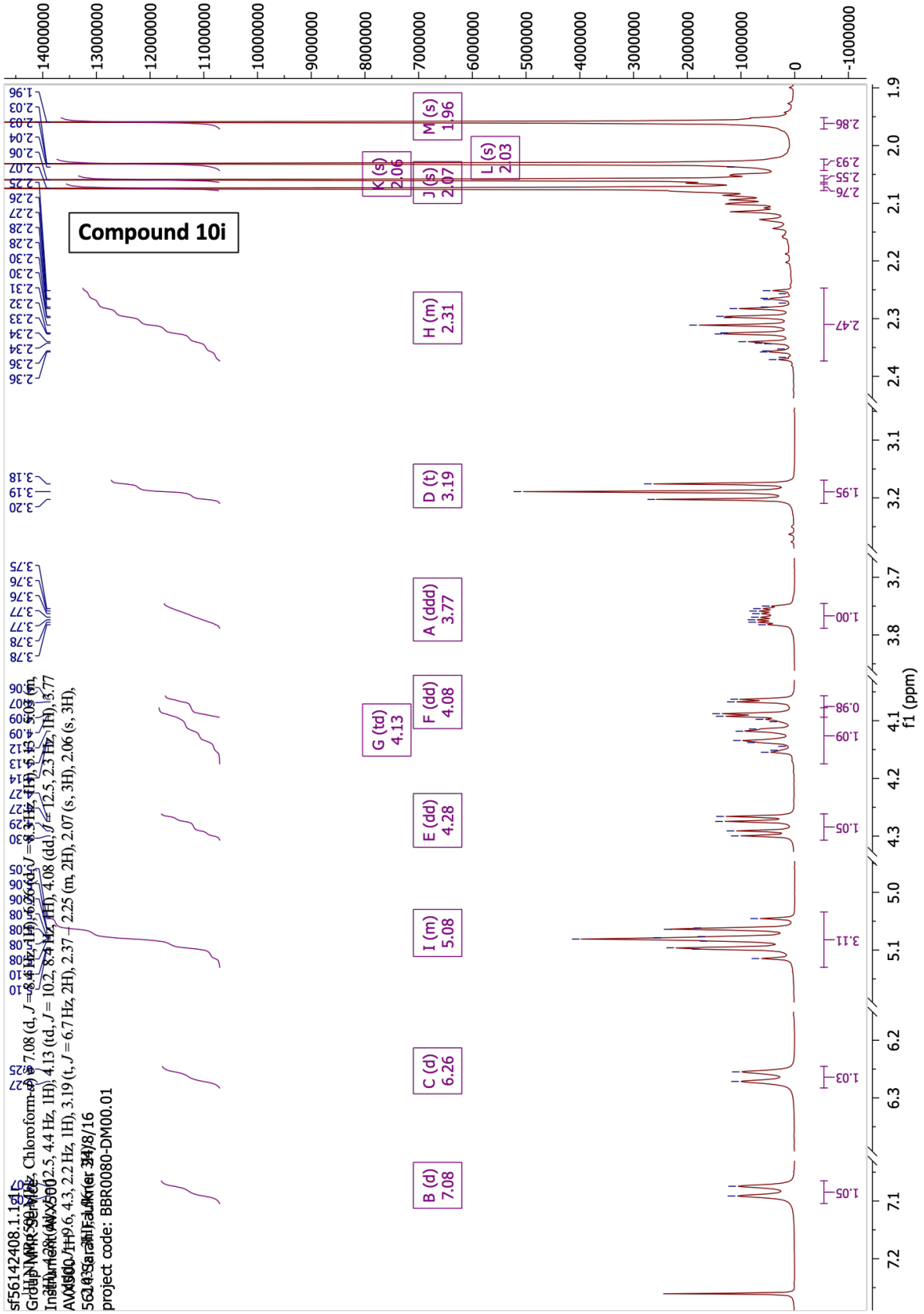

Compound 8
Feb18-2016-39-SF-AcGlcNAcIPropyl-C.1.fid
Instrument AVH400
Chemist Sarah Faulkner
Group BGD
Project Account Code
c13adq.crl CDCl3 {C:\NMR} bgdgrp 39
A (s) 172.41
B (s) 171.91
C (s) 171.16
D (s) 170.80
E (s) 169.43
F (s) 80.30
G (s) 73.58
H (s) 72.94
I (s) 68.00
J (s) 61.91
K (s) 53.40
L (s) 40.29
M (s) 23.60
N (s) 20.89
O (s) 20.86
P (s) 20.74
Q (s) -2.53
f1 (ppm)

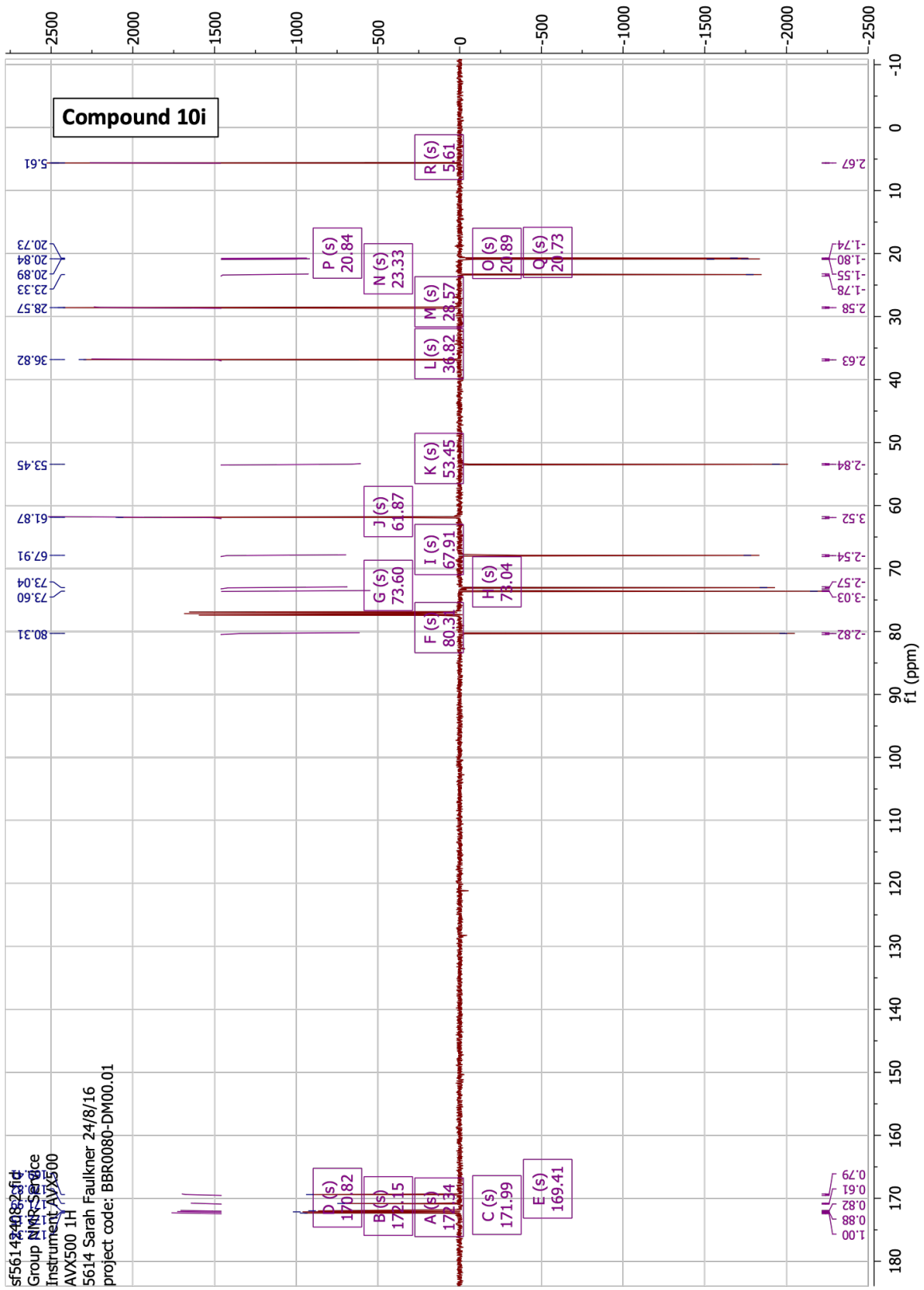

Compound 9i
C (d) 7.11
D (d) 6.27
E (m) 5.08
F (dd) 4.29
G (m) 4.11
A (ddd) 3.77
B (t) 3.42
H (m) 2.36
J (s) 2.03
I (m) 2.09
K (s) 1.95
f1 (ppm)
Aug19_2016-1-SF-3-35-H4.4.1r
Instrument AVG400
Group BGD
Chemist Sarah Faulkner
Project Account Code
h1acq.crl CDCl3 {C:\NMR} bgdgrp 1

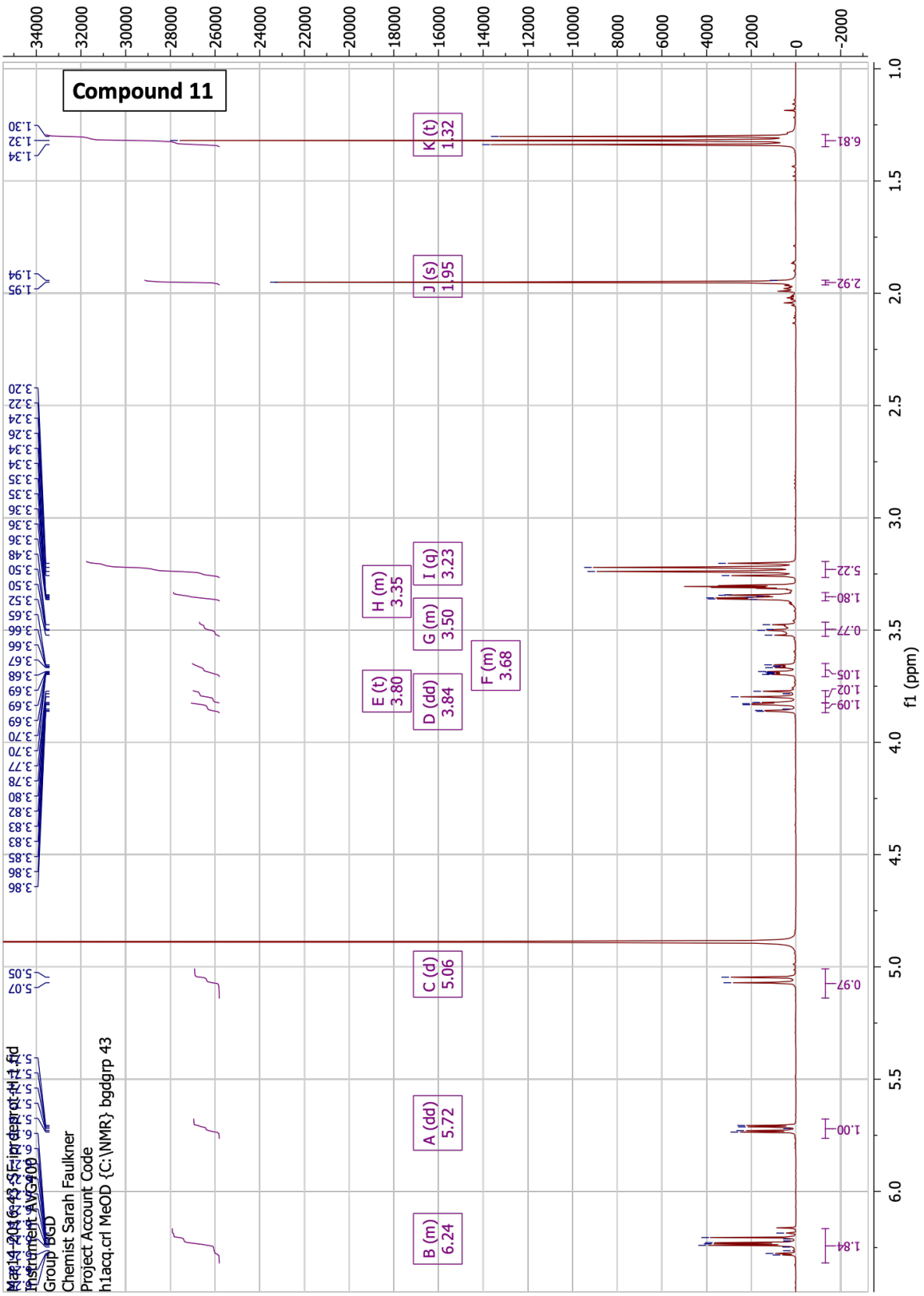

Compound 9i
Q (s) 20.84
P (s) 20.88
O (s) 23.27
R (s) 20.72
N (s) 27.81
M (s) 32.87
L (s) 34.47
K (s) 53.44
J (s) 61.88
I (s) 67.93
G (s) 73.60
H (s) 73.04
F (s) 80.30
D (s) 170.82
B (s) 172.16
A (s) 172.55
C (s) 171.98
E (s) 169.41
f1 (ppm)
sf5613 24082.fid
Group NMR Service
Instrument AVX500
AVX500 1H
5613 Sarah Faulkner 24/8/16
project code: BBR0080-DM00.01

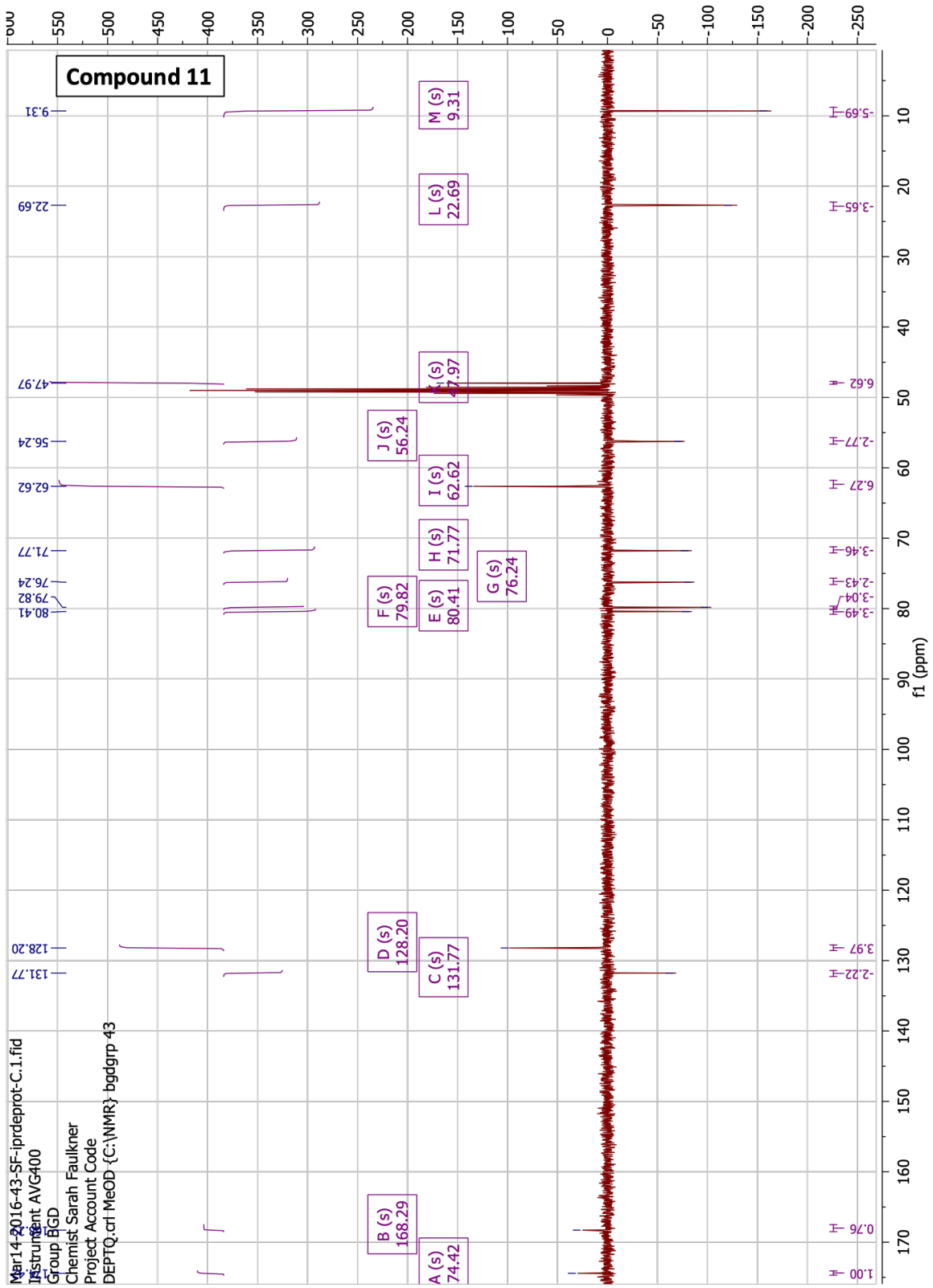

Compound 9ii
A (d) 5.07
B (t) 3.82
C (dd) 3.75
D (dd) 3.88
E (dd) 3.60
F (d) 3.49
G (td) 2.44
H (dtd) 2.12
I (s) 2.01
f1 (ppm)


Compound 9ii
A (s) 176.36
B (s) 174.68
C (s) 78.26
D (s) 77.60
E (s) 74.22
F (s) 69.43
G (s) 60.46
H (s) 54.27
I (s) 34.16
J (s) 32.95
K (s) 27.89
f1 (ppm)

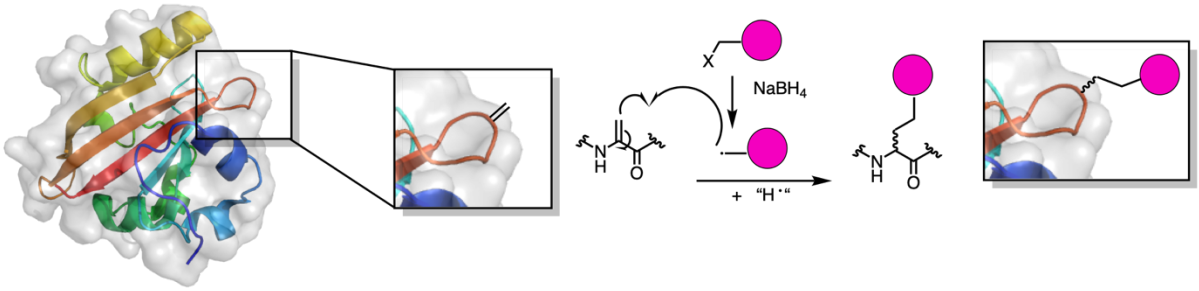

Compound 10i
B (d) 7.08
C (d) 6.26
I (m) 5.08
E (dd) 4.28
G (td) 4.13
F (dd) 4.08
A (ddd) 3.77
D (t) 3.19
H (m) 2.31
K (s) 2.06
J (s) 2.07
L (s) 2.03
M (s) 1.96
f1 (ppm)


Compound 10i
f1 (ppm)
sf561432408.2.fid
Group NMR Service
Instrument AVX500
AVX500 1H
5614 Sarah Faulkner 24/8/16
project code: BBR0080-DM00.01


Compound 11
A (dd) 5.72
B (m) 6.24
C (d) 5.06
D (dd) 3.84
E (t) 3.80
F (m) 3.68
G (m) 3.50
H (m) 3.35
I (q) 3.23
J (s) 1.95
K (t) 1.32
f1 (ppm)
Chemist Sarah Faulkner
Project Account Code
h1acq.crl MeOD {C:\NMR} bgdgrp 43


Compound 11
M (s) 9.31
L (s) 22.69
K (s) 47.97
J (s) 56.24
I (s) 62.62
H (s) 71.77
G (s) 76.24
F (s) 79.82
E (s) 80.41
D (s) 128.20
C (s) 131.77
B (s) 168.29
A (s) 74.42
9.31
22.69
47.97
56.24
62.62
71.77
76.24
79.82
80.41
128.20
131.77
-5.69
-3.65
6.62
-2.77
6.27
-3.46
-2.43
-3.04
-3.49
3.97
-2.22
0.76
1.00
f1 (ppm)
Mar14-2016-43-SF-iprdeprot-C.1.fid
Instrument AVG400
Group EGD
Chemist Sarah Faulkner
Project Account Code
DEPTQ.crl MeOD {C:\NMR} bgdgrp 43

5



Chapter 5: Concluding remarks

MICROSOFT OFFICE USER

LIST OF ABBREVIATIONS FOR CHAPTER 5

Abbreviation	Meaning
Asn	Asparagine
Cys	Cysteine
Dha	Dehydroalanine
DNA	Deoxyribonucleic Acid
ELISA	Enzyme-linked Immunoabsorbent Assay
Fc	Fragment crystallizable
H2A	Histone 2A
H2B	Histone 2B
H3	Histone H3
H3K36me2	Histone H3 dimethylated at lysine 36
H3K36me2	Histone H3 trimethylated at lysine 36
H4	Histone H4
MRG15	MORF-Related on Gene 15
N-GlcNAc	*N*-linked N-Acetylglucosamine
O-GlcNAc	*O*-linked N-Acetylglucosamine
PTM	Post-translational modification
RNAPII	Ribonucleic acid polymerase II
SETD2	SET domain containing protein 2 (human)
WT	Wild type

5. Concluding remarks

Post-translational modifications are able to tune the properties of proteins, allowing certain functions to be switched on or off. Recently, a method of selectively converting cysteine to a variety of PTM mimics via conversion to Dha followed by a radical reaction has been developed in the Davis group. This new method represents the first example of generating PTM mimics by the formation of a C-C bond, allowing the synthesis of PTM mimics which had been previously inaccessible. In this thesis, two such PTM mimics were explored in case studies which broadly reflected two key biological functions in eukaryotes where the switchable nature of PTMs is important; in controlling the interactions of chromatin in the nucleus, and in rapid cell signalling in the cytoplasm.

Chapters 2 and 3 of this thesis detailed the study of the histone PTM H3K36me3. In vivo evidence has suggested a biological need for trimethyl H3K36me3 in particular over the dimethyl state. However, this has previously not been supported by in-vitro binding studies of chromodomain reader protein MRG15 on peptides or free H3. It was therefore hypothesised that discrimination between the di- and tri-methylation states by MRG15 is due to a change in the overall nucleosome conformation, and through binding studies with intact H3K36me2 and -me3 bearing nucleosomes, it would be possible to reconcile the *in-vivo* and *in-vitro* observations.

In Chapter 2, the synthesis of modified histones H3K36me2 and H3K36me3 using radical-Dha chemistry for 'post-translational mutagenesis', and subsequent incorporation of these into nucleosomes, was described. In the octamer assembly, a striking difference in the ratio between octamer and tetramer peaks in SEC purification was observed, suggesting a destabilization of the interaction between $H3_2$-$H4_2$ tetramer and H2A-H2B dimers.

Upon nucleosome assembly, further differences between the methylation states were also noticed. The appearance of different bands in the methyl variants, a lower 'WT-like' and a higher 'alternative' band suggested different conformations, and the higher ratio of 'alternative' to 'WT-like' in the trimethyl variant compared to dimethyl suggested that methylation stabilised a new conformation in the nucleosome.

The altered interactions with the H2A-H2B dimer and the new conformation on addition of DNA suggested that methylation affected the preferred position of the H3 tail. This was confirmed in Chapter 3, where the interactions of modified nucleosomes with MRG15 were studied. Previous pulldown-based binding assays were revisited with these synthetic nucleosomes, and it was now possible to observe the clear preference for -me3 nucleosomes over -me2 that one would expect from *in-vivo* data.

Moreover, the strength of binding and magnitude of discrimination appeared dependent on the salt type and concentration of the buffer, with weaker binding observed on increasing salt concentration from 150 mM KCl to 250 mM, and weaker selectivity observed on changing the salt to NaCl.

A model was proposed to explain MRG15 discrimination between di- and trimethylated nucleosomes. From the crystal structure, H3K36 is ideally positioned to form a hydrogen bond between a side chain -NH and the phosphate of the DNA backbone. In this state, H3K36 is shielded and inaccessible to MRG15. However, in the trimethyl state, H3K36 no longer has the capacity to hydrogen bond with the DNA backbone; it was hypothesised that this broken interaction, along with the additional bulk and hydrophobicity of the trimethyl state, causes a conformational shift in the H3 tail where K36me3 becomes exposed, and therefore accessible to MRG15.

These results also had implications for the mode of SETD2 binding to H3K36me2; if it is normally inaccessible due to burial behind DNA, methylation may occur when the DNA is partially unwound from the nucleosome by RNAPII and H2A-H2B have dissociated.

Altogether, this first case study highlighted the utility of using radical Dha chemistry to faithfully reproduce two PTMs found in vivo. With these in hand, it was possible to deconstruct how switching between the two subsequently turns on important interactions in the nucleus.

One of the exciting aspects of chemical protein modification is that any residue can in theory be altered to bear a PTM, not merely those which already do in biology. This allows the possibility for generating completely novel constructs. The second case study, therefore, explored this concept by installing a natural PTM in an unnatural location to produce a novel sensor. Dha chemistry was once again employed to install *O*- and *N*-linked GlcNAc mimics in the complement determining region (CDR) of a nanobody cAbLys3, with the aim of reversibly blocking its antigen binding capacity. As up-regulated *O*-GlcNAcases are a known cancer biomarker, it was anticipated that this could form the basis for further medicinal applications.

Both *O*- and *N*-linked GlcNAc could be installed using this method, but only the *O*-linked variant proceeded to 100% conversion, important for generating pure product for later assays. Likewise, whilst *O*-GlcNAc could be removed by enzymatic cleavage, *N*-GlcNAc could not; only the *O*-linked variant and its cleavage product cAbLys3-bis-homoSer, therefore, could be assayed using ELISA.

However, whilst addition of the *O*-linked sugar proved successful in abrogating cAbLys3's binding capacity, removing the sugar was unsuccessful for switching it back on, likely due to

the unnatural bulk of the remaining *bis*-homoSer group. Enzymatic extension of cAbLys3-*O*-GlcNAc to produce an Fc decasaccharide was also investigated as an alternative means of tuning a nanobody's effects for potential therapeutic applications. However, this was also unsuccessful, likely due to fundamental limitations of the glycosyltransferase enzymes available. A key future aim for this work is testing the extension on cAbLys3-*N*-GlcNAc, which is closer to the natural Asn-GlcNAc structure found at the base of the tetrasaccharide *in-vivo*.

Overall, the aim of this thesis has been to demonstrate the use of radical-Dha chemistry to install mimics of labile PTMs, both to study their function in a natural setting, and to impart a novel function to a protein by appending an unnatural PTM. The first was successful, allowing the synthesis and study of H3K36me2 and -me3 containing nucleosomes with unprecedented molecular precision, and giving insight into the biological need for H3K36me3. The second did not provide the GlcNAc-based on/off switch which was hoped for, but did demonstrate the utility of Dha-radical chemistry for installing sugar PTMs in a completely novel location on a folded, disulfide-bearing protein substrate, which has traditionally been a challenge for other methods. Moreover, it has provided a starting point for further studies which may allow the generation of hybrid nanobodies with the Fc glycans of larger antibodies.

Altogether, it is hoped that this thesis has provided an insight into the utility of the radical Dha chemistry across a range of biological applications.