

STRUCTURAL AND BIOCHEMICAL INVESTIGATION OF THE
SIN3A : HDAC1 : SDS3 TRANSCRIPTIONAL CO-REPRESSOR
COMPLEX

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

Irene Nigi MSc (University of Florence)
BSc (University of Florence)
Department of Molecular and Cell Biology
University of Leicester

February 2018

Irene Nigi

Structural and biochemical investigation of the Sin3a : HDAC1 : SDS3 transcriptional co-repressor complex

Abstract

The Sin3a complex is a large multi-protein complex that is involved in embryonic development, cell cycle regulation, proliferation, oncogene regulation and senescence. The Sin3a protein acts as a scaffold protein and the complex contains several proteins including HDAC1 and HDAC2 as catalytic subunits, SDS3 and various SAPs.

The Sin3a complex is one of the most evolutionarily conserved HDAC complexes regulates transcription by altering chromatin condensation. To date only few structures of Sin3a-Paired Amphipathic Helices domains interacting with various proteins have been published, and it therefore remains unknown how Sin3a interacts with HDAC1 and how the activity of the complex is regulated.

In this study a stable Sin3a ternary complex containing Sin3a : HDAC1 : SDS3 was expressed using a mammalian expression system and purified using two alternative methods. A structural model was built by combining data from cross-linking, Small Angle X-Ray Scattering and negative stain Electron Microscopy providing the first structural information on the assembly of the Sin3a complex. SDS3 was found to be essential to allow the Sin3a complex to assemble as a dimer. The model also suggests that Sin3a interacts with HDAC1 through a structured central region, while the N- and C-termini of the HID domain wrap around the catalytic subunits to secure a stable interaction.

Enzymatic assays showed that the Sin3a complex, unlike other HDAC complexes, is not regulated by inositol tetrakis phosphate and suggest a distinct regulatory mechanism that might involve the presence of different proteins within the Sin3a complex. In conclusion, both structural and functional results suggest that the Sin3a complex appears to be distinct from other Class I HDACs with both its structure and regulatory mechanism.

Acknowledgements

This PhD has been a truly rewarding and life-changing experience and there are many people that I would like to thank for contributing into making this adventure unforgettable.

First of all, I would like to thank my supervisor Professor John Schwabe, who despite a huge amount of duties and meetings has always found the time to be supportive and give me his precious guidance for this project. I would also like to thank my co-supervisor Dr Shaun Cowley for our helpful discussions and for sharing with me his passion for science. A grateful thanks also goes to my committee members Dr Sue Shackleton and Dr Andrey Revyakin for the discussion and advice given.

A special and warm thank you goes to Dr Louise Fairall and Dr Peter Watson for their invaluable help, endless support, time and patience. A big appreciation goes also to Dr Christopher Millard, Dr Rob Turnbull and to the entire Schwabe group, from the advises to the big laughs, it has been a real pleasure to work in such a friendly and encouraging environment. During this PhD I have been so lucky to meet such sweet and brilliant people like Dr Yun Song, friends who always cheered me up like Dr Charles Milano and who collected funny travelling stories with me like Beatriz Romartinez.

The Department of Molecular and Cell biology has provided the perfect surroundings to allow me learning many different techniques and this could have not be possible without the help of Dr Cyril Dominguez (thank you for your suggestions and help throughout these years), Dr Andrew Jamieson, Dr Naomi Robertson, Dr Olga Makarova. Thank you also to the Cowley's group for their friendly help and support.

I feel that something that made this PhD special was also the chance to meet a brilliant group of scientist within the Nuclear Receptor Marie Curie Network. Not only I had the best time travelling and visiting some of the best labs in Europe but it also has been a huge inspiring reality to live. So, thank you to all the PIs and students involved in the network for sharing your science with me and inspiring me in so many ways!

Living in another country would have not been the same without meeting Dr Jaelle Foot, Dr Saleha Patel and Karolin Kroese. You made my days brighter by sharing with me the love for fitness, food and animals. I will do my best to make our beautiful friendship to last over time and despite distance.

A last thank, full of my love, goes to my family. To the family who grew me up strong and cheerful, who always backed me up in my crazy ideas and supported me when I decided to live this experience in another country, thank you Mum and Dad, I love you with all my heart, I would be nothing without you. And a last praise to the family that I am making now, Nicola, you truly changed my life and I could not feel luckier. Diana Iris doesn't know yet how special her dad is, but I can tell you that you are the most loving, intelligent and strongest person that I ever knew. Thank you for making my life what it is.

Table of Titles

Abstract.....	ii
Acknowledgements	iii
Table of Titles.....	iv
Table of Contents.....	iv
Table of Figures.....	ix
Table of Tables	xiii
List of Abbreviations	xv

Table of Contents

Chapter 1 - Introduction	1
1.1 - The chromatin architecture	3
1.2 - Histone post-translational modifications	8
1.2.2. Acetylation	10
1.2.3. Phosphorylation	11
1.2.4. Methylation	12
1.2.5. Ubiquitination	14
1.3 - The histone deacetylase enzymes	15
1.3.2. Class I	17
1.3.3. Class II	18
1.3.4. Class III	19
1.3.5. Class IV	19
1.4 - Class I HDAC co-repressor complexes	20
1.4.1. The Mi-2/NuRD complex	23
1.4.2. The MiDAC complex	24

1.4.3. The CoREST complex	24
1.4.4. The SMRT/NCoR complex	25
1.4.5. The Sin3a complex	25
1.5 - Thesis overview and project aims	31
Chapter 2. Materials and Methods	33
2.1. Materials	33
2.1.1. Standard chemicals and reagents	33
2.1.2. Bacterial and mammalian cell lines	33
2.2. Generating constructs	34
2.2.1. Bioinformatics	34
2.2.2. Primer design	34
2.2.3. Amplification from cDNA library of the genes of interest	34
2.2.4. Plasmid sequencing	35
2.3. Plasmid preparation	36
2.3.1. DH5 α competent cells preparation	36
2.4. Transformation	36
2.4.1. Small scale plasmid preparation	37
2.4.2. Large scale plasmid preparation	37
2.4.3. Quantification of plasmid DNA	39
2.5. Protein expression and purification (E. coli expression)	39
2.5.1. Rosetta DE3 transformation	39
2.5.2. GST tagged protein purification	39
2.5.3. His tagged protein purification	40
2.6. Protein expression and purification in mammalian HEK 293F	41
2.6.1. Culture of HEK 293F cells	41
2.6.2. Transient transfection in HEK 293F	42

2.6.3. Flag-tagged protein purification	43
2.7. Further purification methods	44
2.7.1. Gel filtration chromatography	44
2.7.2. Sucrose density gradient	44
2.8. Protein analysis	45
2.8.1. SDS PAGE	45
2.8.2. Protein quantification	46
2.8.3. Size Exclusion Chromatography – Multi-angle Light Scattering (SEC-MALS)	46
2.8.4. Liquid chromatography-tandem mass spectrometry (LC-MS/MS)	46
2.9. Tissue culture and transfection in monolayer HEK 293T	47
2.9.1. Culture of HEK 293T	47
2.9.2. Reporter assay transient transfection	47
2.10. β -galactosidase	48
2.11. Luciferase assay	49
2.12. Stapled peptide synthesis and purification	49
2.13. Fluorescence polarization assay	49
2.14. Histone deacetylase assay	50
2.15. Microfluidic Mobility Shift Assay (MSA)	50
2.16. Chemical protein cross-linking coupled with mass spectrometry assay	51
2.17. Small Angle X-Ray Scattering (SAXS)	51
2.18. Protein crystallisation trials	52
2.19. Negative stain Electron Microscopy	52
Chapter 3. Expression and purification of the Sin3a complex	54
3.1 - Introduction	54
3.2 - Bioinformatics: protein constructs designing	55
3.2.1 - The Sin3a protein constructs	55

3.2.2 - The HDAC1 protein constructs	59
3.2.3 - The SDS3 protein constructs	60
3.2.4 - The Sin3a Associated Proteins (SAPs) protein constructs	61
3.3 - Protein expression	62
3.3.1 - Investigating the effect of different affinity tags on Sin3a	63
3.3.2 - The Sin3a ₆₀₁₋₇₄₂ : HDAC1 : SDS3 ₂₀₁₋₂₃₄ complex is successfully expressed in bacteria	65
3.3.3 - The Sin3a ₆₀₁₋₇₄₂ : HDAC1 : SDS3 ₂₀₁₋₂₃₄ complex was expressed by combining bacterial and mammalian expression systems.	68
3.3.4 - Different combinations of complex components were tested to find a stoichiometric Sin3a complex	70
3.3.5 - Different combination of components were expressed in mammalian HEK 293F cells to find the most stable Sin3a complex	73
3.4 - Protein purification methods	75
3.4.1 - Size exclusion chromatography for protein purification - Investigating the temperature conditions to improve protein complex purification	76
3.4.2 - The use of sucrose gradient as an alternative method to purify the Sin3a complex	78
3.5 - Initial characterization of the Sin3a complex - Investigation of substrate specificity with microfluidic mobility shift assay (MSA).	80
3.6 - Result summary and discussion	85
Chapter 4. Functional Studies	88
4.1 - The use of stapled peptides to specifically target the Sin3a complex	88
4.2 - Stapled peptides synthesis	90
4.3 - The use of SDS3 native and stapled peptides to disrupt the Sin3a complex	90
4.3.1. The SDS3 native and stapled peptides disrupt the interaction between Sin3a and SDS3 <i>in vitro</i>	92

4.3.2. Measuring the K _d for the SDS3 native and staple peptides	94
4.3.3. The SDS3 native and stapled peptides fail to disrupt the Sin3a complex <i>in cellulo</i>	97
4.3.4. GFP tagged SDS3 ²⁰¹⁻²³⁴ fails to disrupt the Sin3a complex	103
4.4 - The used of Mxd1 native and stapled peptides to impede the recruitment of the Sin3a complex	105
4.4.1. Mxd1 native and stapled peptide disrupt the Sin3a : Mxd1 complex <i>in vitro</i>	106
4.4.2. Mxd1 native and stapled peptides fail to disrupt the interaction between the native Sin3a complex and Mxd1 <i>in cellulo</i>	108
4.4.3. GFP tagged Mxd1 ¹¹⁻²⁷ disrupt the interaction between the endogenous Sin3a complex and Mxd1	111
4.5. Results summary and conclusions	114
Chapter 5. Structural studies	116
5.1. Understanding how HDAC1 is recruited by Sin3a	116
5.2. Confirmation that the Sin3a ⁵³¹⁻⁷⁷² : HDAC1 : SDS3 complex can exist as a dimer	117
5.2.2. Mass spectrometry analysis	122
5.3. Crystallization trials on the Sin3a complex	123
5.4. Mapping interactions between HDAC1, Sin3a and SDS3	126
5.4.1. Cross-linking analysis of the Sin3a ⁵³¹⁻⁷⁷² : HDAC1 _{FL} : SDS3 _{FL} complex	127
5.4.2. Structural modelling of the Sin3a complex	134
5.5. Using SAXS to study the structure of the Sin3a ⁵³¹⁻⁷⁷² : HDAC1 _{FL} : SDS3 _{FL} dimeric complex	138
5.6. Negative stain electron microscopy	142
5.7. Cryo - electron microscopy	147
5.8. Results summary and discussion	149

Chapter 6. Discussion	152
6.1. A structural model for the Sin3a complex and suggested regulatory mechanism	152
Appendix	158
References	168

Table of Figures

Figure 1.1: The central dogma of molecular biology.	1
Figure 1.2: Structural arrangement of histones in the chromatin assembly.	5
Figure 1.3: The compaction of eukaryotic chromosome.	7
Figure 1.4: HATs and HDACs regulate the acetylation state of chromatin.	15
Figure 1.5: Schematic division of HDAC enzymes in their phylogenetic classes.	17
Figure 1.6: The inositol phosphate molecule is not required to the SANT domain to recruit the HDAC enzyme.	22
Figure 1.7: Best known Class I HDACs containing complexes.	23
Figure 1.8: Schematic representation of the Sin3a complex.	28
Figure 1.9: Schematic representation of the Sin3a complex with the published boundaries of the HID.	30
Figure 3.1: Sequence alignment of the PAH3-HID domains of the Sin3a protein from different model organisms.	57
Figure 3.2: Disorder and secondary structure prediction analysis for Sin3a₄₅₀₋₈₅₁.	58
Figure 3.3: Disorder and secondary structure prediction analysis for HDAC1_{FL}.	59
Figure 3.4: Disorder and secondary structure prediction analysis for SDS3_{FL}.	60
Figure 3.5: Sequence alignment of the SAP proteins.	
Figure 3.6: Preliminary attempt for the expression and purification of Sin3a₄₅₀₋₇₇₂ using different affinity tags.	64
Figure 3.7: Bacterial expression of Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄.	66

Figure 3.8. Investigating the activity of HDAC1 expressed in bacteria.	67
Figure 3.9. Combination of bacterial and mammalian expression of Sin3a ₆₀₁₋₇₄₂ : HDAC1 _{FL} : SDS3 ₂₀₁₋₂₃₄ .	69
Figure 3.10. Reconstitution of the Sin3a complex by combining bacterial and mammalian expression - expression trials using different complex components and constructs.	72
Figure 3.11. Mammalian expression using different complex components and constructs.	74
Figure 3.12. SEC purification of the Sin3a complex at different experimental temperatures.	77
Figure 3.13. Sucrose gradient purification of the Sin3a complex.	79
Figure 3.14. Chemical structures of the histone peptides.	81
Figure 3.15. BOC-lysine activity assay of the HDAC containing complex.	82
Figure 3.16. Initial deacetylation activity rate of Sin3a with different histone tails peptides in the presence or absence of IP ₆ .	83
Figure 3.17. Initial deacetylation activity rate of different HDAC containing complexes with H4K16ac in the presence or absence of IP ₆ .	84
Figure 4.1: The SDS3 stapled peptide, mechanism of action and amino acid sequence.	91
Figure 4.2: The fluorescent SDS3 stapled peptide. Chemical structure of the fluorescent SDS3 stapled peptide.	91
Figure 4.3: The SDS3 peptides successfully disrupt the Sin3a ₆₀₁₋₇₄₂ : SDS3 ₂₀₁₋₂₃₄ complex. <i>In-vitro</i> binding assay.	93
Figure 4.4: Fluorescent polarization assay with SDS3 native peptide.	95
Figure 4.5: Fluorescent polarization assay with SDS3 stapled peptide.	95
Figure 4.6: The apparent K _d of the SDS3 peptide is not affected by different constructs and components of the Sin3a complex.	96
Figure 4.7: Possible outcome of the SDS3 peptide on the luciferase assay.	98

Figure 4.8: The shorter construct Gal4.SDS3₂₀₁₋₂₃₄ fails to recruit the endogenous Sin3a complex.	99
Figure 4.9: Higher peptide concentration does not affect the outcome of the luciferase assay.	100
Figure 4.10: Lipofectamine fails to facilitate the entrance of the peptides into the cells.	101
Figure 4.11: Digitonin fails to help the peptides entering into the cells.	102
Figure 4.12: (GFP)-tagged SDS3₂₀₁₋₂₃₄ in HEK 293T cells.	103
Figure 4.13: (GFP)-SDS3₂₀₁₋₂₃₄ fails to disrupt the endogenous Sin3a complex.	104
Figure 4.14: Mxd1 native and stapled peptides.	106
Figure 4.15: The Mxd1₁₋₂₇ peptides successfully disrupt the Sin3a₂₉₁₋₃₈₈ : Mxd1₁₋₂₇ complex.	107
Figure 4.16: Expected effect of the Mxd1 stapled peptide on the luciferase assay.	108
Figure 4.17: Increasing amount of Mxd1 peptides do not affect the luciferase activity.	109
Figure 4.18: Varying the incubation time for the Mxd1 native/stapled peptides.	110
Figure 4.19: (GFP)-Mxd1₁₋₂₇ transfected in HEK 293T.	112
Figure 4.20: (GFP)- Mxd1₁₋₂₇ successfully disrupt the interaction between the endogenous Sin3a complex and Gal4.Mxd1₁₋₂₇.	113
Figure 5.1: SEC-MALS analysis on the Sin3a complex.	119
Figure 5.2: SEC-MALS analysis on the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.	120
Figure 5.3: SEC-MALS analysis on the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} complex.	121
Figure 5.4: Mass spectrometry analysis of Sin3a₅₃₁₋₇₇₂ and the degraded Sin3a.	122
Figure 5.5: Crystallization phase diagram and sitting drop crystallization method.	124
Figure 5.6: Images from Sin3a crystal trials.	125
Figure 5.7: Homobifunctional cross-linkers used for the XL-MS experiments.	127
Figure 5.8: Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} cross-linked with CBDPSS.	128

Figure 5.9: Mapping model created with the CDPSS inter-crosslinking results for the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.	131
Figure 5.10: Mapping model created with the CDPSS intra-crosslinking results for the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.	131
Figure 5.11: Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} cross-linked with DSSG.	132
Figure 5.12: Model of the DSSG inter-crosslinking results for the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.	133
Figure 5.13: Model of the DSSG intra-crosslinking results for the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.	133
Figure 5.14: Mapping the cross-linking results on the structure of the Sin3a and SDS3 (mouse).	134
Figure 5.15: Docking Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} with HADDOCK.	136
Figure 5.16: Model for the Sin3a complex based on the cross-linking experiments.	137
Figure 5.17: SEC-SAXS workflow.	139
Figure 5.18: SAXS analysis for the Sin3a complex.	140
Figure 5.19: SAXS structural modelling of the Sin3a complex, SAXS envelope created with P1 symmetry.	141
Figure 5.20: SDS-PAGE analysis of the sucrose gradient fractions of the cross-linked Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄ complex.	143
Figure 5.21: Deconstruction of a carbon film 400 mesh grid.	144
Figure 5.22: Negative stain analysis of the Sin3a complex.	145
Figure 5.23: Initial 3D models for the negative stained Sin3a complex.	146
Figure 5.24: Negative stain EM 3D model of the Sin3a complex with C2 symmetry.	147
Figure 5.25: Cryo-EM experiment for the Sin3a complex.	148

Table of Tables

Table 1.1: Reader domains of histone modifications.	9
Table 1.2: Lysine residues on histone tails that act as substrates for acetylation.	11
Table 1.3: List of residues on histone tails found to be substrates of phosphorylation.	12
Table 1.4: Residues on histone tails that are known as substrates of methylation.	14
Table 1.5: Lysine residues on histone tails that are known to be substrates of ubiquitynation.	15
Table 1.6: Classification of the HDAC superfamily.	16
Table 2.7: PCR protocol.	35
Table 2.8: Transfection reaction recipes.	42
Table 3.9: Comparison between the localisation of the SID and the region that remained conserved between the SAPs.	
Table 4.10: SDS3 native and staple peptides, fluorescent polarization assays details.	96
Table 4.11: Improving the accessibility of the peptides to the cells.	111
Table 5.12: Predicted molecular weight of the components of the Sin3a complex. ..	118
Table 5.13: Inter-crosslinks from the CBDPSS analysis of the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.	129
Table 5.14: Intra-crosslinks from the CBDPSS analysis of the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.	130
Table 5.15: Table of inter/intra-crosslinks from the DSSG cross-linked Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.	133
Table 5.16: Finding the best model for docking Sin3a₆₀₁₋₇₃₇ and HDAC1.	135
Appendix 7: Disorder prediction for the SAPs proteins.	164
Appendix 8: Details of protein constructs used to express the Sin3a complex components in <i>E. coli</i>.	165

Appendix 9: Details of protein constructs used to express the Sin3a complex components in HEK 293F cells.	166
---	------------

List of Abbreviations

aa: amino acid

Ada2: adaptor

ATP: Adenosine triphosphate

Boc-K: Boc-acetyl-lysine

Bp: base pair

BRMS: breast cancer metastasis suppressor

Bromo: Bromo domain

BSA: Bovine Serum Albumin

CBDPSS: Cyanurbiotindipionylsuccinimide

c-Myc: cellular-myelocytomatosis oncogene

C-terminal: carboxyl terminal

CDK4/6: cyclin-dependent kinases 4/6

CHD: chromo-domain helicase DNA

Chromo: Chromo domain

CoREST: Co-repressor of REST

CtBP: carboxyl-terminal binding protein

D-PBS: Dulbecco's modified PBS

DAD: deacetylase activation domain

DMEM: Dulbecco's modified Eagle's Medium

DMSO: dimethyl-sulphoxide

DNA: deoxyribonucleic acid

DNMT: DNA methyl transferase

DNTTIP1: deoxynucleotidyltransferase, terminal, interacting protein 1

DSSG: Disuccinidyl glutarate

E. coli: *Escherichia coli*

E(n.n): embryonic day (n.n)

EDTA: ethylenediaminetetraacetic acid

ELM2: Egl-27 and MTA1 homology 2 domain

ES: Embryonic Stem

EtBr: ethidium bromide

FBS: fetal bovine serum

FDA: Food and Drug Administration

FP: Fluorescence Polarization

GPS2: G protein pathway suppressor

GST: Glutathione S-transferase

H: histone

H(n)K(n): histone (n) lysine (n)

H(n)R(n): histone (n) arginine (n)

H(n)S(n): histone (n) serine (n)

HAT: histone acetyl transferase

HCR: highly conserved region

Hda1: yeast histone deacetylase 1

HDAC: histone deacetylase

HDACi: HDAC inhibitor

HEK 293F: human embryonic kidney 293 freestyle

HEK 293T: human embryonic kidney 293

HPLC: High Performance Liquid Chromatography

HRP: Horseradish Peroxidase

SV40 large T-antigen

HEPES: 4-(2-hydroxyethyl)piperazine-1-ethanesulphonic acid

HID: HDAC Interaction Domain

HRP: horseradish peroxidase

Ing2: inhibitor of growth

IP: immuno-precipitation

IP(n): inositol n phosphate

ISWI: imitation switch

JHDM; Jumonji domain-containing histone demethylase

K-Acetate: potassium acetate

Kbp: kilo base pair

kDa: kilo Daltons

KDAC: lysine deacetylase

L: liter

LB: Lauria-Bertani

LC-MS/MS: liquid chromatography mass spectrometry

LSD1: lysine specific demethylase 1

M: molar

M10: Medium 10% FBS

Maldi-Tof: matrix-assisted laser desorption/ionization time of flight

MBD2/3: methylated DNA binding protein

MBT: malignant brain tumor domain (chromatin reader)

MeCP2: methylated CpG binding protein

2 MEF2: myocyte enhancer factor

MEFs: mouse embryonic fibroblasts

mg: milligram

MiDAC: mitotic deacetylase complex

MIDEAS: mitotic deacetylase-associated SANT protein

ml: milliliter

mM: millimolar

MRG15: mortality factor related gene 15

mRNA: messenger RNA

MS/MS: Tandem Mass Spectrometry

MSA: Mobility Shift Assay

MTA1-3: metastasis associated protein

N-terminal: amino terminal

NAD⁺: nicotinamide adenine dinucleotide

NCoR: nuclear receptor co-repressor

nm: nanometer

NMR: nuclear magnetic resonance

NuRD: nucleosome remodeling deacetylase complex

O/N: Over night

OD: Optical Density

°C: degrees centigrade

°K: degrees kelvin

p53: protein 53 kDa

PAH: paired amphipathic helix

PBS: phosphate buffer saline

pCMV: cytomegalovirus promoter

PNACL: Protein Nucleic Acid Chemistry Laboratory (University of Leicester)

PCR: polymerase chain reaction

PEG: polyethylene glycol PEI: polyethylenimine

PHD: plant homeo domain

PPAR γ : peroxisome proliferator-activated receptor gamma

PPG: polypropylene glycol

PSG: penicillin streptomycin glutamine

PTM: post-translational modification

Ras: rat sarcoma gene

RBBP: retinoblastoma binding protein

RIP140: receptor interacting protein 140

RNA: ribonucleic acid

RNase A: Ribonuclease A

RONN; Regional order neural network

Rpd3: reduced potassium dependency

Rpd3L: large Rpd3 complex (yeast)

Rpd3S: small Rpd3 complex (yeast)

Rpm: rounds per minute

RUNX2: runt-related transcription factor 2

S200: SuperdexTM200 gel filtration column

S6: Superose[®]6 gel filtration column

SANT: Swi3, Ada2, NCoR, TFIIIB domain

SAP: Sin3 associated protein

SAXS: Small-angle X-ray Scattering

SDS-PAGE: SDS polyacrylamide gel electrophoresis

SDS: sodium dodecyl sulphate

SDS3: suppressor of defective silencing 3

SEC-MALS: Size-exclusion Chromatography Multi-angle light Scattering

SID: Sin3a Interaction domain

Sin3a/b: Switch independent3 a/b

siRNA: short interfering RNA

Sirt1-7: sirtuin 1-7

SMRT: silencing mediator of retinoid or thyroid hormone receptor

SWI/SNF: switch/ sucrose non fermentable

Swi3: switch 3

SWR1: SWI2/SNF2 related ATPase

T-X100: triton-X100

TAE: Tris-acetate-EDTA

TB: Tris-Borate

TBE: Tris-Borate-EDTA

TBL1: transducin (beta)-like protein

TCEP: Tris (2-carboxyethyl) phosphine

TE: Tris-EDTA

TEV: Tobacco Etch Virus

Tris: Tris (hydroxymethyl) aminomethane

TSA: Trichostatin A

v/v: volume/volume

v/w: volume/weight

WD40: tryptophan-aspartic acid repeat

x g: times gravity

XL-MS: Cross-linking combined tandem mass spectrometry

CC: coiled-coil domain

μm : micrometer

μM : micro molar

Chapter 1. Introduction

It has been 60 years since Francis Crick presented the central dogma for the first time in 1957 during a talk. The central dogma describes the possible directions of informational flow that can occur between DNA, RNA and proteins (Crick 1958; Crick 1970). In his first publication about the dogma, Crick stated: “Once information has passed into protein, it cannot get out again” (Crick 1958). This means that there is no route of reverse information to transfer information from protein to nucleic acids (figure 1.1).

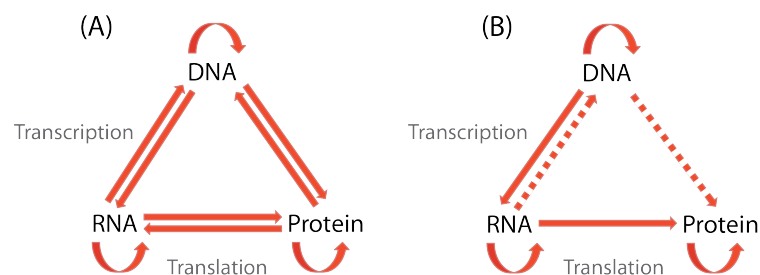


Figure 1.1: **The central dogma of molecular biology.**

All cellular life forms share the same fundamental scheme of genome replication and expression. (A) shows the possible directional flows of information, whereas (B) shows that, as proposed by Crick, no route of reverse information is allowed from protein to nucleic acids. Picture adapted from (Crick 1958).

In eukaryotes, the process of DNA transcription into RNA by RNA polymerase II can be divided into 3 phases: 1) recognition and binding to promoter sequences, 2) elongation and 3) termination. However, the description of informational flow does not explain how it is possible that the cells in a human body, or in general in a multicellular organism, can be so different from one another as they all share the same DNA. This process requires the help of several transcription factors and its regulation is essential for a normal function of cell cycle and development.

Developmental studies have shed some light on how these differences develop and are maintained. In a cell, the information that flows from DNA to RNA and finally to protein needs to be regulated to control the function of the cell. Basically, each step of this information flow can be regulated, but the first control point resides at the

initiation of transcription. Regulatory mechanisms for transcription and translation are found in both prokaryotes and eukaryotes, but the eukaryotic regulatory mechanism for transcription is more intricate than for prokaryotes.

DNA polymerase binding to DNA on a promoter sequence dictates the initiation of transcription. In eukaryotic cells the presence of “enhancer sequences” can affect polymerase activity. These sequences act as binding sites for regulatory proteins and transcription factors (J.-D. Wen et al. 2008; Cornish et al. 2008; Spitz & Furlong 2012). Transcriptional regulation is finely tuned during development in different ways. The activity of multiple enhancers, co-binding of co-factors or complexes and the architecture of sequence motifs and enhances are only part of the possible strategies to modulate the transcriptional regulation (Zinzen et al. 2009; Giorgetti et al. 2010; Biggar & Crabtree 2001).

This regulatory process is important for the self-regulation of the single cell, but it is also fundamental for the development of multicellular organisms. In a multicellular organism different cell types differ both structurally and functionally one another. During the development from single cell to multicellular organism, different sets of genes need to be turned on and off by transcription factors and regulatory proteins. In this way, during successive generation of cells different cell types express different combination of regulatory proteins in order to differentiate cells to play different roles.

In eukaryotes the structure of chromatin plays a large role in the control of gene expression. Chromatin structure is controlled at many levels by a wide variety of protein complexes, histone modifications, DNA modifications etc. Histone deacetylases (HDACs), as part of repression complexes, play a crucial role in many biological processes mainly by altering the status of chromatin and repressing transcription.

HDACs are a large family of enzymes and Class I HDACs (further described in paragraph 1.4.1) play a role in specific gene expression control. In particular, the deletion of HDAC1 is lethal to the embryo (E10.5) and shows general delayed growth (Montgomery et al. 2007; Lagger et al. 2002). A HDAC1 containing co-repressor complex that has been extensively studied is the Sin3a (Switch Independent 3 homologue “a”) co-repressor complex. The deletion of Sin3 in mouse embryos results

in early lethality (E3.5), showing the important role of this complex in early development (Cowley et al. 2005; McDonel et al. 2012). In mouse, the deletion of this complex in embryonic stem cells (ES) leads to apoptosis caused by high levels of DNA damage (McDonel et al. 2012). These experiments together with many other functional studies, confirm the importance of the Sin3a complex as a gene regulator in a large variety of functions, as in embryonic development, cell cycle regulation, oncogene regulation and metabolism.

The aim of this introduction is to introduce chromatin, describing its structure and the possible modifications that can occur to it. In addition, the transcriptional co-repressor HDAC complexes will be described, focusing on the Sin3a complex, which will be the main topic of this thesis.

1.1. The chromatin architecture

The human genome contains around 6 billion base pairs, which is about 2 metres of DNA. In order to fit into an approximately 6 μm diameter eukaryotic nucleus, the genomic material must be folded tightly and at the same time it must remain accessible for DNA replication and gene transcription. Within the nucleus, human DNA is organised into 46 chromosomes. The arrangement of genes within the chromosomes collects together genes in order to coordinate expression and allow transcription in gene clusters (Blumenthal 1998).

Chromatin is composed of repeating units of a protein core made of 8 histone molecules and 146 bp of DNA, called the nucleosome (Figure 1.2B) (Kornberg 1974). In 1965, five different histone molecules were extracted and analysed: H1, H2A, H2B, H3 and H4 (Phillips & Johns 1965). In addition, there are variants for H2A, H2B and H3 with additional sequences domains or small sequence variations (Bonenfant et al. 2006). The amino acid sequence of the canonical histones is very well conserved during evolution. X-ray diffraction was used to elucidate the structure of chromatin and the interaction of histones with DNA was deduced from the properties of the histone core (Kornberg & Thomas 1974).

The canonical histones share a common polypeptide chain fold made of 3 α helices (1 long central and 2 shorter helices on the side) separated by 2 loops (L1 and L2) that are essential for interaction with DNA (Figure 1.2A) (Finch et al. 1977; Arents & Moudrianakis 1995; Luger, Rechsteiner, et al. 1997). At their N-terminus 15-30 residues are called “tails” and these are known to be unstructured and thereby play a role in gene regulation. This common structural domain is called “histone fold” and it facilitates the heterodimerization of histones (H2A with H2B and H3 with H4) through a head to tail “handshake” motif (Figure 3A) (Luger, Mäder, et al. 1997). In the nucleosome, the histones pairs are then coupled together to form a four-helix bundle in an ordered manner. A central core is formed by a H3-H4 tetramer and two H2A-H2B are positioned on each side where they can be removed more easily for chromatin remodelling (Smith & Stillman 1991; Kulaeva et al. 2010).

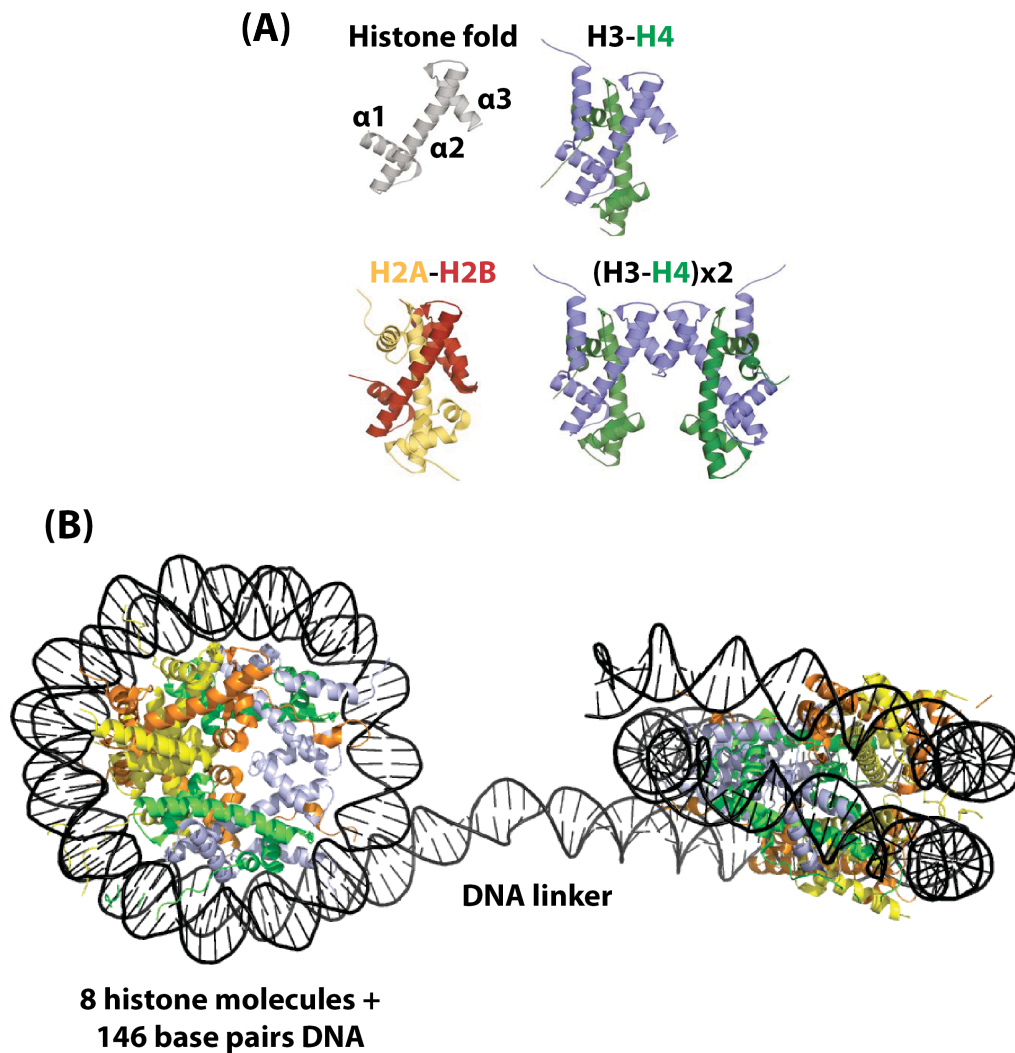


Figure 1.2: **Structural arrangement of histones in the chromatin assembly.**

(A) The “histone fold” consists of 3 helices separated by 2 loops. H2A/H2B and H3/H4 heterodimerize through a “handshake motif”. Picture is adapted from (Mattioli et al. 2015).

(B) Schematic representation of how the nucleosome beads assemble. Each nucleosome contains 8 histone molecules (histone core) and 146 base pairs DNA wrapped around the histone core. Each nucleosome is connected to the next one through a DNA linker. (Picture was made with Pymol software and adapted from PDB file: 1ZBB (Schalch et al. 2005)).

Histone H1 binds the nucleosome around the entry and exit sites of the DNA (Syed et al. 2010; B.-R. Zhou et al. 2013; Song et al. 2014). The structure of the nucleosome was visualised for the first time by electron microscopy in 1971, revealing a “beaded” structure of chromatin fibres (Henson & Walker 1971; Woodcock et al. 1980; A. L. Olins & D. E. Olins 1983). After digesting chromatin with micrococcal nuclease, the beads were identified to be nucleosomes (Oudet et al. 1975). Between the nucleosome beads the length of DNA (called “linker”) can vary (Spadafora et al. 1976). Importantly,

nucleosomes in the proximity of promoters and regulatory elements can modulate the accessibility to DNA (Thoma 1992). The function of histone tails extending from nucleosomes is thought to be to promote fibre formation and making contact with the neighbour nucleosomes (Schwarz et al. 1996). Although the DNA folding in nucleosome has been described, the interaction among themselves is still unclear (Luger, Mäder, et al. 1997).

The complexity of chromatin organisation relies on different mechanisms as subdomain organization, long-range interactions and enhancer-promoter interactions. Furthermore, recent studies have found that chromatin can also be organised in structural domains that represent distinct functional genomic units (Sexton et al. 2012). There are mainly three types of spatial organization of chromatin: nucleosomes, loops and topologically associating domains (TADs).

Chromatin loops bring together cis-regulatory elements (i.e. enhancers) with their target promoters (Tan-Wong et al. 2012; Rao et al. 2014).

Topologically associating domains (known as TADs) are domains of approximately 185 kb characterized by high self-interaction. Regions within the same TAD interact within themselves more frequently than with other regions located in adjacent domains (Sexton et al. 2012; Dixon et al. 2012). Figure 1.3 shows the canonical condensational phases of chromatin.

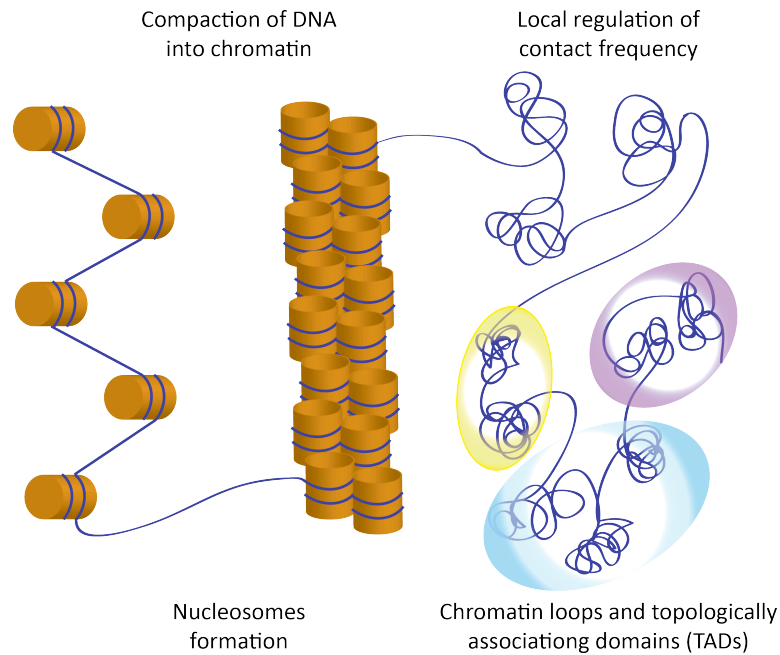


Figure 1.3: **The compaction of eukaryotic chromosome.**

Several degrees of compaction occur to DNA and histones to allow the all DNA to fit in the small volume of the nucleus. The double helix DNA gets compacted into nucleosomes, which are important for epigenetic modifications. Chromatin then gets organized in loops and topologically associating domains (TADs). Picture is adapted from (Hansen et al. 2018).

The resulting tight structure of nucleosomes hinders the interaction between RNA Polymerase II (Pol II) and DNA. To allow the transcription of DNA, the interactions between histones and histones-DNA can be altered by several factors: Post Translational Modifications (PTMs), ATP dependent chromatin remodellers and histone chaperones (Smolle & Workman 2013; Becker & Workman 2013; De Koning et al. 2007). These factors cooperate in the regulation of nucleosome dynamics to enhance the advancement of Pol II on DNA (Hassan et al. 2001; Segal et al. 2006; Jiang & Pugh 2009; Hughes & Rando 2014).

1.2. Histone post-translational modifications

Histone Post Translational Modification (PTMs) have been described as “epigenetic changes” since they are able to alter gene expression without affecting the DNA sequence. PTMs such as phosphorylation, acetylation, methylation, sumoylation and ubiquitynation on histones are fundamental to regulate the chromatin structure (Allfrey et al. 1964; Goto et al. 2002; Sugiyama et al. 2002; Z. Wang et al. 2008; Smolle & Workman 2013).

By modifying the chemical interactions between and within nucleosomes, some PTMs can destabilise the nucleosome and switch the state of chromatin from a “closed” to a “open” state (Petruk et al. 2012). For example, by targeting histone tails, PTMs can control fibre formation and modulate the accessibility to chromatin (Tse et al. 1998). Furthermore, depending on the combination of epigenetic marks applied, PTMs can regulate transcription by either promoting or inhibiting the interaction with effector proteins. “Writers” and “erasers” enzymes can apply or remove these epigenetic marks to histones and modulate the final effect on transcription (Smolle & Workman 2013).

PTMs can occur on either the same histone tail or on different histone tails and their possible combinations are thought to describe a particular “histone code” (Strahl & Allis 2000; Suganuma & Workman 2011). The histone code is interpreted by “reader” proteins that contain specific domains for each modification (as described in table 1.1) (Suganuma & Workman 2011; Ruthenburg et al. 2007). Usually, reader proteins are part of larger protein complexes that contain components able to add or remove other PTMs to chromatin (Taverna et al. 2007; Yun et al. 2011). In this way, the recognition of particular modifications from the readers will lead to an active modification of the chromatin status.

Table 1.1: Reader domains of histone modifications.

The table has been adapted from (Yun et al. 2011).

PTM mark	Reader domain	Histone residue
Acetylation	Bromo	H3K14; H3K56; H4K16; H4K5; H4K8
	PHD	H3K14
Lysine-Methylation	PHD	H3K4; H3K9
	WD40	H3K4; H3K9; H3K27; H4K20; H1K26
	ADD	H3K4
	Chromo	H3K4; H3K9; H3K23; H3K27; H3K36
	Tudor	H3K4; H3K9; H4K20
	MBT	H3K4; H4K20; H1K26
	Zf-CW	H3K4
	Ankyrin repeats	H3K9
	PWWP	H3K36; H4K20
Arginine-Methylation	Tudor	H3R17; H4R3
	ADD	H4R3
Phosphorylation	(Gcn5)	H3S10
	14-3-3	
Ubiquitination	?	H2AK119; H2BK120/123

1.2.2. Acetylation

The condensation of the chromatin fibre can be regulated by the opposing activity on the ϵ -group of lysine residues of histone tails by histone acetyl transferases (HAT) and histone deacetyl transferases complexes (HDAC) (Shahbazian & Grunstein 2007). The lysine residues of histone tails that are known targets of acetylation are listed in table 1.2. Histone acetyl transferases add acetyl residues to histone tails. By removing the positive charge on the histones, the interaction of histones with the negatively charged DNA is weakened. This effect tends to relax the state of chromatin and increase its accessibility (Grunstein 1997; Howe et al. 1999; Berger 2001). HDAC enzymes catalyse the opposite reaction, removing the acetyl groups and lead to the condensation of chromatin.

Several studies have demonstrated the relationships between histone acetylation and transcriptional activity (Sternier & Berger 2000; H. Chen et al. 2001; Roth et al. 2001) and also between histone deacetylation and gene repression (Garcia-Ramirez et al. 1995; Tse et al. 1998; Grunstein 1997; Struhl 1998). However, acetylation by itself is not sufficient to enable full accessibility to chromatin. Indeed, acetyl residues lead to partial decondensation and more importantly label regions of competence (Brown et al. 2000). Specific acetylation patterns represent marks for the recruitment of other chromatin modulator factors that can rearrange histones and methylate and phosphorylate them (X. Wang et al. 2000; Gräff & Tsai 2013; T. Zhang et al. 2015). The specific combination of acetylation, methylation and phosphorylation marks can be recognised by domains in chromatin remodellers and other chromatin targeted complexes (Turner 2005; Nathan et al. 2006; Camporeale et al. 2007).

In contrast, histone deacetylation enzymes (HDACs) promote condensation by recruiting further repressive factors for a synergetic action to switch chromatin to its repressive state. Depending on their position, methylation of lysine residues that are otherwise substrates of acetylation can result in more permanent repression and lock chromatin in its “closed” state. Histone deacetylation is described in more detail in section 1.3.

Table 1.2: Lysine residues on histone tails that act as substrates for acetylation.

Histone	Residue
H2A	K5
H2B	K5; K12; K15; K20
H3	K9; K14; K18; K23; K27; K36
H4	K5; K8; K12; K16

1.2.3. Phosphorylation

Histone phosphorylation is a dynamic process that targets serines, threonines and tyrosines from both histone tails and histone cores (table 1.3) (Nowak & Corces 2004). The phosphorylation state is regulated by kinases and phosphatases that add and remove phosphate groups respectively (Brehove et al. 2015). By adding a negatively charged phosphate group to histones, kinases can alter the chromatin structure and affect its accessibility.

Phosphorylation on histone H3 was initially associated to chromosome condensation during mitosis and meiosis (Sauvé et al. 1999; Wei et al. 1999), but subsequent studies showed the involvement of phosphorylation in chromatin relaxation. Phosphorylation of promoters of genes enhance their activation (Lau & Cheung 2011). Additionally, phosphorylation of H3S10 is associated with acetylation (in particular of H3K9ac and H3K14ac) and this leads to transcription activation (Cheung et al. 2000; Clayton et al. 2000). Phosphorylation of Serine 28 of Histone 3 (H3S28) leads to the dissociation of Polycomb repressive complexes from chromatin and promotes the synergetic action of acetylation and demethylation of H3K27 to activate transcription (Metzger et al. 2008; Metzger et al. 2010).

During mitosis and meiosis phosphorylation of residues threonine 3, serine 10, threonine 11 and serine 2 on H3 lead to chromosome condensation and segregation (Gopalan et al. 1997; Tatsuka et al. 1998).

Table 1.3: List of residues on histone tails found to be substrates of phosphorylation.

Histone	Residue
H1	Ser27
H2A	Ser1; Ser139; Thr120; Thr142
H2B	Ser14; Ser36
H3	Ser10; Ser28; Thr3; Thr6; Thr11; Thr41; Thr45
H4	CK2

1.2.4. Methylation

Methylation can occur on all the basic residues: arginine, lysine and histidine (Murray 1964; Byvoet et al. 1972; Fischle et al. 2008). Lysine and arginine residues can also be mono-, di- or tri-methylated on their ϵ -amine group. There are histone methylation sites in all the canonical histones: H1, H2A, H2B, H3 and H4 (table 1.4). The best known targets of methylation are on histone H3: lysines K4, K9, K27, K79 and K20, and arginines R2, R8, R17, R26 and R3 (Tan et al. 2011). The methylation state is maintained by the opposing action of methyl transferases and demethylases.

Initially, methylation was thought to be a permanent modification (Byvoet et al. 1972), but in 2004, the discovery of the demethylase complex LSD1 revealed the possibility that methylation is a dynamic process (Y. Shi et al. 2004). More recent studies also shown that methylation follows different turn-over rates depending on the biological process (Zee et al. 2010). Indeed, in certain situations the methylation state has to be maintained in for example the heterochromatin state, whereas in other situations it needs to remain flexible in for example cell differentiation (Khavari et al. 2010).

Three protein families are known to transfer a methyl group from S-adenosyl methionine to histones: SET-domain-containing proteins, DOT1-like proteins and protein-arginine N-methyl transferase (PRMT) proteins (Rea et al. 2000; Feng et al. 2002; Bannister & Kouzarides 2011). These enzymes are able to methylate both histones and non-histone proteins (J. Huang & Berger 2008). Two families of demethylase enzymes act on both histones and non-histone proteins to remove methyl groups: amine oxidases and Jumonji C-domain containing dioxygenases (Tsukada et al. 2006; Cloos et al. 2006).

Methylated histones are recognised by binding domain such as: chromodomain, chromobarrels, Tudor domains and MBT repeats (table 1) (Flanagan et al. 2005; P. Zhang et al. 2006; Y. Yang et al. 2010; Jeeseun Kim et al. 2006).

The presence of other PTMs can influence the methylation of certain residues. For example, phosphorylation of H3S10 prevents the methylation of H3K9, and mono-ubiquitination of H2B is required for methylation of H3K4 and H3K79 (Rea et al. 2000).

Methylation on certain histones plays a fundamental role in many fundamental biological processes: cell cycle, DNA damage, stress response and development (Kouzarides 2007; Pedersen & Helin 2010; Eissenberg & Shilatifard 2010). Alteration in histone methylation can lead to tumorigenesis and ageing (Pollina & Brunet 2011; Conway et al. 2015). The histone residues that are recognised as target of methylation are listed in table 1.4.

Table 1.4: Residues on histone tails that are known as substrates of methylation.

Histone	Residue
H1	K26
H2A	Arg3
H3	Arg2; Arg8; Arg17; Arg26; Arg42; K4; K9; K27; K36; K79
H4	Arg3; K20; K59

1.2.5. Ubiquitination

Multiple enzymes catalyse the addition of ubiquitin and the most frequent targets of ubiquitination are H2A and H2B (table 1.5) (Goldknopf et al. 1975; Robzyk et al. 2000; Pham & Sauer 2000; Jones et al. 2011). The ubiquitination state of histones is also regulated by the activity of deubiquitinating enzymes (DUBs) (Hochstrasser 1996; Weake et al. 2008).

There are two types of ubiquitination: monoubiquitination and polyubiquitination. Common targets for monoubiquitination are K119 of H2A and K120 and K123 of H2B (Goldknopf et al. 1975; West & Bonner 1980).

In the nucleus the ubiquitination state plays an important role in transcription, DNA repair and maintenance of chromatin structure. Indeed, during embryogenesis, ubiquitination of H2A is required for viability and an efficient histone H3 Lys27 trimethylation by PRC2. Whereas ubiquitination of H2B was shown to be linked to gene activation (Joo et al. 2007; W. Zhou et al. 2008; Minsky et al. 2008; Pengelly et al. 2015) (Xiao et al. 2005; Joo et al. 2007; Zhu et al. 2007; W. Zhou et al. 2008; Minsky et al. 2008). Furthermore, both mono- and poly-ubiquitination occurring at the site of

DNA damage act as recognition element for the recruitment of DNA repair machine proteins (Moyal et al. 2011; Jaehoon Kim et al. 2009).

Table 1.5: Lysine residues on histone tails that are known to be substrates of ubiquitynation.

Histone	Residue
H2A	K119
H2B	K120; K123

1.3. The histone deacetylase enzymes

Histone deacetylation is classically thought to repress transcription, packing nucleosomes tighter and to hinder the access of the transcription machinery to DNA (figure 1.4). Deacetylation can lead to methylation, enhancing in this way the repression effect on transcription (C. Martin & Y. Zhang 2005). However, HATs and HDACs are both found at the promoters of active genes (Z. Wang et al. 2009; Signolet & Hendrich 2015). This suggests that HDACs are involved in resetting chromatin at active genes during transcription.



Figure 1.4: HATs and HDACs regulate the acetylation state of chromatin.

On the left, HATs enzymes add acetyl groups to lysine residues of the histone tails leading to a relaxation of the state of chromatin. HDACs remove these acetyl groups and lead to chromatin condensation and transcription repression.

In yeast, the deacetylase proteins can be differentiated by their functional roles. Sir2 is associated to amino-acid biosynthesis, HDA1 with carbon metabolism and Rdp3 with cell cycle genes (Blander & Guarente 2004). In eukaryotic cells, the HDAC superfamily

is divided into 2 families based upon the cofactor needed to transfer the acetyl group: The NAD⁺ dependent Sirtuins (Class III) and the Zn²⁺ dependent classical HDACs (Class I, II and IV) (de Ruijter et al. 2003; Blander & Guarente 2004). These families can be further divided into 4 classes: Class I, II, III and IV. The HDAC classification is described in detail in table 1.6 and shown in figure 1.5.

Table 1.6: Classification of the HDAC superfamily.

In eukaryotic cells, deacetylase proteins are divided into 4 classes. Their classification and main features are listed above.

Class	Isoform	Subcellular localization	Cofactor	Yeast counterparts
Class I	HDAC1	Nucleus	Zn ²⁺	Rdp3 Hos1 Hos2
	HDAC2	Nucleus	Zn ²⁺	
	HDAC3	Nucleus→Cytoplasm	Zn ²⁺	
	HDAC8	Nucleus→Cytoplasm	Zn ²⁺	
Class IIa	HDAC4	Nucleus, Cytoplasm	Zn ²⁺	HDA1
	HDAC5	Nucleus, Cytoplasm	Zn ²⁺	
	HDAC7	Nucleus, Cytoplasm, Mitochondria	Zn ²⁺	
	HDAC9	Nucleus, Cytoplasm	Zn ²⁺	
Class IIb	HDAC6	Cytoplasm→Nucleus	Zn ²⁺	HDA2
	HDAC10	Cytoplasm→Nucleus	Zn ²⁺	
Class III	Sirt1-6	Nucleus, Cytoplasm, Mitochondria	NAD ⁺	Sir2
Class IV	HDAC11	Nucleus	Zn ²⁺	Rdp3/HDA1

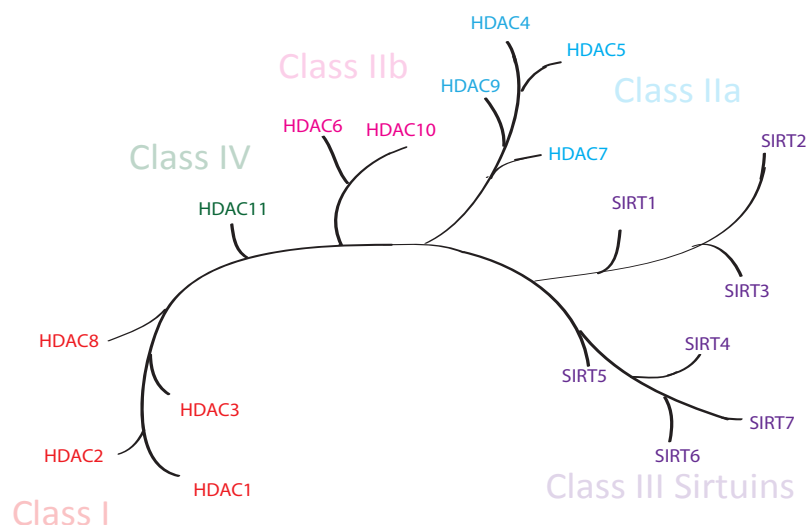


Figure 1.5: **Schematic division of HDAC enzymes in their phylogenetic classes.**

Class I, IIa, IIb and IV belong to the classical HDACs family, whereas Class III belongs to the Sirtuins family.

1.3.2. Class I

Class I HDACs are homologous with the yeast proteins Rpd3, Hos1 and Hos2 (de Ruijter et al. 2003). This class includes isoforms HDAC 1, 2, 3 and 8. HDAC1 and 2 are exclusively nuclear and share 82% sequence identity (Johnstone 2002).

HDAC1 and 2 are incorporated into Sin3a, NuRD and CoREST co-repressor complexes (X.-J. Yang & Seto 2003). They share overlapping functions and are both involved in the regulation of cell cycle, but at the same time they show specialised functions (Zhu et al. 2004). For example, HDAC1 appears to be unique and essential, its deletion during embryonic development is lethal and it cannot be compensated by an up-regulation of any other Class I HDACs enzymes (Lagger et al. 2002). Whereas the deletion of HDAC2 appears not to be lethal (Kelly & Cowley 2013).

HDAC3 shares approximately 52% similarity with the catalytic domain of HDAC1 and 2 (Cress & Seto 2000). This isoform is involved in the regulation of gene expression by nuclear hormone receptors and it is part of the SMRT and NCoR multiprotein co-repressor complexes (X.-J. Yang & Seto 2003). HDAC3 possesses both nuclear import

signal and nuclear export signal (NES), and can be found in both nucleus and cytoplasm (Fischle et al. 2002).

HDAC8 is a unique isoform within Class I since it does not have to be recruited to co-repressor complexes to be active. HDAC8 has been shown to be involved in muscle contractility in smooth muscle (Waltregny et al. 2005). Additionally, it also regulates telomerase activity and it could participate to tumorigenesis (H. Lee et al. 2006; Kang et al. 2014; Z.-T. Wang et al. 2016).

1.3.3. Class II

Class II HDACs share homologies with the Hda1 and Hda2 yeast proteins and includes HDAC 4, 5, 6, 7, 9 and 10. This class is further divided into Class IIa (HDAC4, 5, 7 and 9) and Class IIb (HDAC6 and 10) (Bertos et al. 2001; Verdin et al. 2003; Gregoret et al. 2004).

Class IIa HDACs show sequence similarities in their N- and C-termini, where important domains for their functions and regulation of these enzymes are contained. This specific class of HDACs play an important role in cell differentiation and they are expressed in specific tissues. They can shuttle between the nucleus and cytoplasm according to stimuli due to the presence of both nuclear import and export signals (McKinsey et al. 2000; McKinsey et al. 2001; Dressel et al. 2001; Chang et al. 2004; Halkidou et al. 2004). The location of Class IIa HDACs is determined by site-specific phosphorylation, which is controlled by a variety of kinases (Verdin et al. 2003). HDAC5 is located in the nucleus during pre-differentiation of muscle cells but during differentiation it is phosphorylated by Ca^{2+} /calmodulin dependent kinase (CaMK) and relocates to the cytoplasm (Bertos et al. 2001). Interestingly, class IIa HDACs possess a low deacetylase activity due to a substitution of a tyrosine to a histidine in their active site, but they are able to bind acetylated lysines and recruit other transcription factors (Fischle et al. 2002; Bradner et al. 2010).

Class IIb HDACs include HDAC6 and HDAC10. HDAC6 contains two deacetylase domains and a zinc finger motif (Grozing et al. 1999) and can affect cell motility by

deacetylating α -tubulin (Hubbert et al. 2002; Haggarty et al. 2003). In addition, the HUB finger contained in HDAC6 is able to bind ubiquitin and possesses E3 ligase activity (Kovacs et al. 2004). HDAC10 was one of the last identified Class IIb HDAC and little is known about this enzyme. HDAC10 is found both in the nucleus and cytoplasm and it shows low transcriptional repression activity (Fischer et al. 2002; Guardiola & Yao 2002).

1.3.4. Class III

Class III histone deacetylases comprise a highly conserved family of oxidized nicotinamide adenine dinucleotide (NAD^+) dependent enzymes called Sirtuins. This class of enzymes share homologies with the yeast Silent Mating-type Information Regulation 2 (Sir2). In mammals there are seven sirtuin genes, Sirt1-7 (Frye 2000; Vinciguerra et al. 2010). The sirtuins all share the same mechanism of action by targeting acetylated lysines from various proteins. Sirtuins transfer the acetyl group from the acetylated lysine to the ADP ribose moiety of NAD^+ , releasing nicotinamide and 2'-O-acetyl-ADP-ribose (X.-X. Kong et al. 2009). Since Sirtuins depends upon the availability of NAD^+ , the cellular metabolism affect a wide range of cellular processes (Hildmann et al. 2007). Sirtuins can respond to stress and toxicity, and in addition they are involved in aging and longevity. Their localisation vary upon isoforms: Sirt2, 6 and 7 are located in the nucleus, Sirt1 and 2 are found in the nucleus and in the cytoplasm and Sirt3, 4 and 5 are found in mitochondria (Michishita et al. 2005; Tennen et al. 2010).

1.3.5. Class IV

The only member of Class IV is HDAC11. HDAC11 is homologous to Class I and II HDACs and it is phylogenetically similar to HDAC3 and 8 (Gao et al. 2002; de Ruijter et al. 2003). HDAC11 is expressed in specific tissues like kidney, brain and skeletal muscle (Gao et al. 2002). Recent studies revealed its implication in immune tolerance and activation but little is known about this isoform (Villagra et al. 2010).

1.4. HDACs and cancer

Aberrant expression of HDAC enzymes has been found to characterize a variety of tumors, including colon and prostate cancer (Patra et al. 2001; Mariadason 2008). Interestingly, key outcomes of tumorigenesis are linked to the overexpression of HDAC enzymes. For example, the epigenetic repression of the tumour suppressor genes cyclin-dependent kinase inhibitor p21 (*CDKN1A*), breast cancer 1 (*BRCA1*) and ataxia telangiectasia and Rad3 related (*ATR*) were shown to be linked to the overexpression of HDACs (Glozak & Seto 2007; Eot-Houllier et al. 2009). Further studies have also shown that the expression of HDAC1, HDAC5 and HDAC7 can be used as biomarkers of tumors (Ozdağ et al. 2006). These findings raised the interest into developing anti-cancer agents capable to bind to the active site of the HDAC enzymes in order to target their action.

Nowadays, 5 HDAC inhibitors (HDACi) are approved by the U.S. FDA (food and drug administration). HDAC inhibitors are not only used for treating cancer but they are also involved in the treatment of neurodegenerative diseases (Saha & Pahan 2006). Indeed, it was shown that HDACi function as neuroprotectors by strengthening synaptic plasticity and therefore improving memory and learning (Xu et al. 2011).

Unfortunately, the use of HDACi brings severe side effects. By acting on the HDAC enzyme's active site, HDACi target all HDAC complexes. As described in paragraph 1.5, different HDAC complexes cover a variety of targets. By striking the HDAC enzymes, all the complexes' targeted would be disturbed. Therefore, in order to avoid any off-target effect, current studies are aiming to develop isoform selective HDAC inhibitors.

1.5. Class I HDAC co-repressor complexes

As previously mentioned in paragraph 1.3.1 Class I HDAC enzymes, with the only exception of HDAC8, need to be recruited to co-repressor complexes for activity. Interestingly, most Class I HDAC complexes share a common mechanism to recruit their catalytic subunits and regulatory mechanism.

With the exception of Sin3a, a SANT domain is contained in all the best-known Class I HDAC complexes. The SANT (Swi3 Ada2, NCoR and TFIIB) domain is characterised by a 50 amino acids motif that can be found in several chromatin remodelling complexes (Aasland et al. 1996). The role of the SANT domain in the different proteins (SMRT, CoREST, MTA1 and MIDEAS) is to mediate the interaction with the HDAC enzymes and it is essential for the deacetylase activity of the complexes (Guenther et al. 2001; Humphrey et al. 2001; Yu et al. 2003). Furthermore, SMRT/NCoR and CoREST have two SANT domains. In the CoREST complex the second SANT domain is responsible for the recruitment of the demethylase LSD1 (Guenther et al. 2001; Humphrey et al. 2001; Y.-J. Shi et al. 2005; Itoh et al. 2015).

When the structure of interaction between HDAC3 and SMRT was solved, it revealed the presence of a D-myo-inositol-1,4,5,6-tetrakisphosphate (inositol tetrakisphosphate, IP₄) located at the interface of the basic surfaces of the two interacting proteins (Watson et al. 2012). In 2013, the structure of interaction between MTA1 and HDAC1 was also solved, revealing a similar arrangement of the helices bundle of SANT interacting with HDAC1. No molecule of inositol phosphate was found in the structure due to high salt crystallization condition (Millard et al. 2013). But inositol phosphates were shown to activate the deacetylase activity (Watson et al. 2016).

Sequence conservation revealed high conservation in the residues involved in the interaction with IP₄ in the different Class I HDAC co-repressor complexes (Watson et al. 2012; Millard et al. 2013; Watson et al. 2016). However, it was shown that IP₄ is not essential for the complex assembly since the SANT domain and N-terminal domain wrap around the HDACs (Figure 1.6). Functional studies have shown instead that IP₄ plays a fundamental role in modulating the catalytic activity of the complexes (Watson et al. 2012; Millard et al. 2013). This finding is particularly interesting as inositol phosphates are known regulatory molecules in gene regulation, cell growth and differentiation (Majerus 1992; Odom et al. 2000).

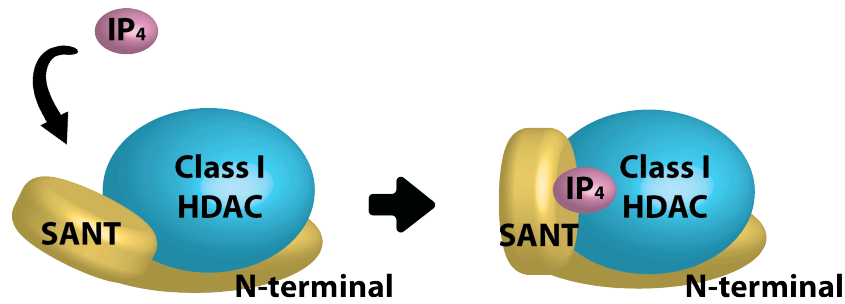


Figure 1.6: **The inositol phosphate molecule is not required to the SANT domain to recruit the HDAC enzyme.**

The SANT domain and the N-terminal domain wrap around the HDAC enzyme. Picture is adapted from Millard *et al.* 2013.

Despite sharing the same regulatory mechanism and the same architecture to recruit the HDAC enzymes, the different Class I HDAC complexes have separate roles depending on the different components and how the complexes are recruited to chromatin. Figure 1.7 shows the five best-known Class I HDAC containing complexes

Furthermore, within the Class I HDAC co-repressor complexes, Sin3a is the only complex that does not share any homology with the SANT domain in any of its complex components. Therefore, the mechanism of HDAC incorporation in the complex remains unclear. For the same reason, its regulatory mechanism remains unclear since it is not known if inositol phosphate is involved in the regulation of the Sin3a complex.

The main details for the best-known Class I HDAC complex will be described and since Sin3a is the main focus of this thesis more details will be given about this complex.

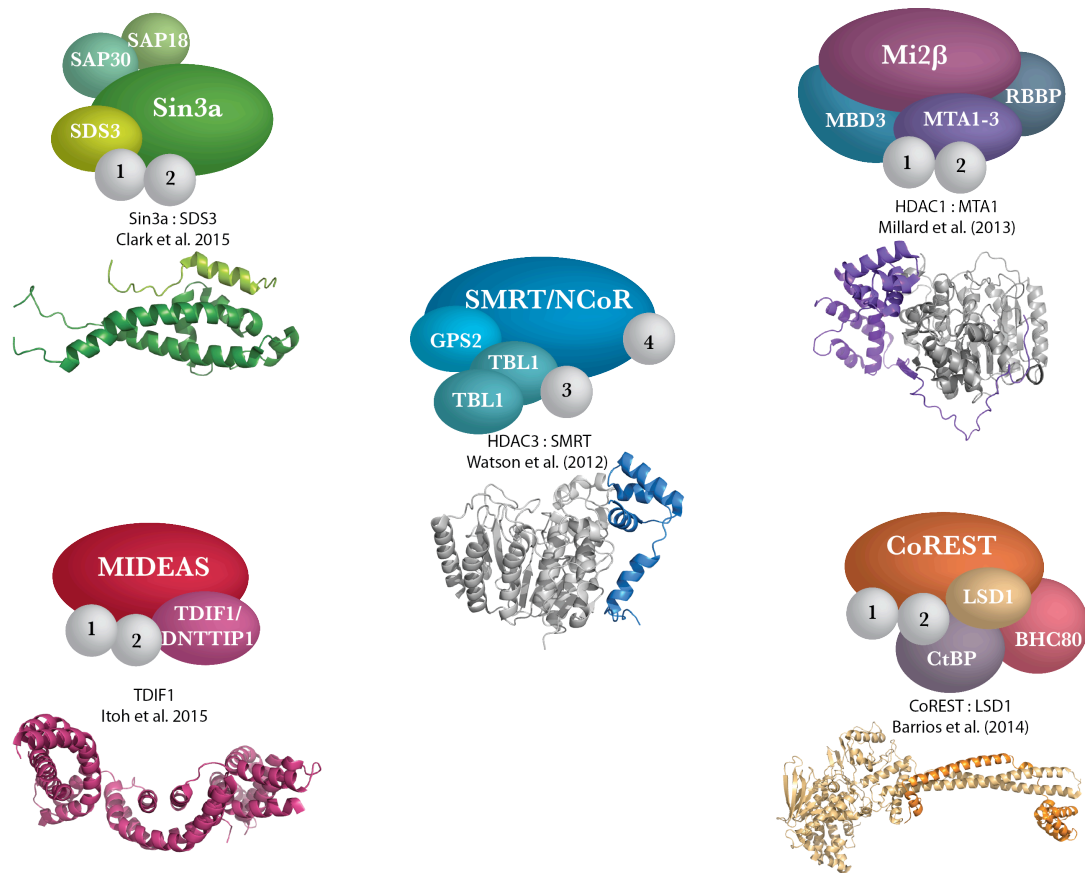


Figure 1.7: **Best known Class I HDACs containing complexes.**

Schematic representation of the five most characterized Class I HDAC complexes with their main components. Partial structures of the complexes were previously published (references for each structure are shown for each complex).

1.5.1. The Mi-2/NuRD complex

The NuRD complex comprises HDAC1 and HDAC2, RBBP4/7 (Retinoblastoma associated protein 46/48), Mi-2 α/β , MBD2/3, MTA1/2/3, and p66 α/β proteins (Ahringer 2000; Guschin et al. 2000; Bowen et al. 2004; Millard et al. 2016). NuRD is involved in development, hematopoiesis and cancer (Hendrich et al. 2001; Yoshida et al. 2008; Ramírez & Hagman 2009). This complex was originally identified due to its nucleosome remodeling effect. Indeed, the NuRD complex is able to remodel chromatin due to the ATPase activity of Mi-2 (Ge et al. 1995). A variety of specific functions are achieved by varying the components of the NuRD complex (Bowen et al. 2004; Denslow & Wade 2007). For example, the presence of MBD3 is required for the

correct complex formation and its deacetylase activity (Y. Zhang et al. 1999). MBD2 in the NuRD complex was originally thought to allow the targeting of methylated DNA (Kaji et al. 2006). The core complex stability depends greatly on the MTA proteins contained in the NuRD complex due to the SANT domain. As mentioned in paragraph 1.4, the SANT domain appears to be a shared feature among most of Class I HDAC complex and it enables the recruitment of HDACs enzymes (Humphrey et al. 2001).

1.5.2. The MiDAC complex

The role of the MiDAC complex is yet to be fully clarified but it is thought to be involved in mitosis (Bantscheff et al. 2011). HDAC1 and/or HDAC2 are contained in this complex. Within the core of this multiprotein complex, MIDEAS (Mitotic Deacetylase-associated SANT domain) acts as a scaffold and it recruits HDAC1/2 through its SANT domain. DNTTIP1 (terminal deoxynucleotidyl transferase-interacting protein), also known as TDIF1, was also identified as part of the core complex (Bantscheff et al. 2011). DNTTIP1 was shown to enable the MiDAC complex to interact with DNA and nucleosome due to a SKI/SNO/Dach domain located at its C-terminus (Itoh et al. 2015). The MiDAC complex assembles as a tetrameric complex, with four copies for each component (Itoh et al. 2015).

1.5.3. The CoREST complex

The CoREST (Co-repressor of RE1-Silencing transcription factor) complex shares the common mechanism of recruitment of HDAC1 and HDAC2 through its SANT domain with the other Class I HDAC complexes. This complex has both deacetylation and demethylation activity by interacting with HDAC1/2 and LSD1 (Lysine Specific Demethylase 1) (M. G. Lee et al. 2005). In this complex other components have also been identified: C-terminal binding protein (CtBP1), PHD finger protein 21A (also known as BHC80), and zinc finger protein 217 ZNF217 (Kuppuswamy et al. 2008; Hayakawa & Nakayama 2011). The specific functions of these proteins are still unknown but it is possible that they might be involved in protein-protein interaction

and in the recruitment of other components to the CoREST complex. One role of the CoREST complex is in neural development, indeed this complex was also known as neuronal restricted silencing factor, NRSF (Andrés et al. 1999; Abrajano et al. 2009).

1.5.4. The SMRT/NCoR complex

In contrast to the other known Class I HDAC co-repressor complexes, the Silencing Mediator of Retinoid and Thyroid Receptor (SMRT) and the Nuclear Receptor Co-Repressor 1 (NCoR) complex recruits HDAC3 instead of HDAC1/2. SMRT and NCoR proteins share 40% of sequence identity and recruit HDAC3 through the Deacetylation Activation Domain (DAD) which contains a SANT domain very similar to that in many other class I HDAC co-repressors. SMRT and NCoR act as scaffold for the recruitment of several proteins to assemble a large multi-protein complex (Y. D. Wen et al. 2000; Guenther et al. 2001; Oberoi et al. 2011) (Y. D. Wen et al. 2000; Guenther et al. 2001). Previous studies have also identified Transducin β -like Protein1 (TBL1) and G-protein Pathway Suppressor2 (GPS2) as part of the core complex (Guenther et al. 2001; Oberoi et al. 2011). Interestingly, SMRT and NCoR contain predicted disordered regions called Repression Domain 3 (RD3) which have been shown to be involved in the interaction with Class IIa HDACs (Hudson et al. 2015). The SMRT/NCoR complex is involved in transcriptional regulation and it plays a fundamental role in cardiogenesis, neuronal differentiation and its deletion is lethal during embryonic development (Jepsen et al. 2000; Jepsen et al. 2008).

1.5.5. The Sin3a complex

The discovery

In 1987, two independent groups identified the Sin3 complex while investigating mating-type switch in *Saccharomyces cerevisiae* (Sternberg et al. 1987; Nasmyth et al. 1987). In yeast, mating-type switching happens in mother cells during G1 phase and it is regulated by the HO endonuclease. The name “Switch Independent” reflects the

ability of Sin3 to circumvent the requirement of SWI5 to activate the transcription of HO endonuclease that is necessary for the gene conversion. Therefore, the observed final effect of Sin3 mutants was a release of repression of the HO endonuclease in the daughter cells.

Phylogenetic connections

The Sin3 complex is well conserved throughout evolution and it is possible to identify different isoforms and paralogs from yeast to mammals. In yeast, *Saccharomyces cerevisiae* as two distinct complexes: Rpd3L and Rdp3S. The two complexes share the same core subunits: Sin3, Rpd3 and Ume1. Their differences are based on the additional components to the complexes and their localization, Rdp3L acts mainly on promoter regions whereas Rpd3S is localized on transcribed regions (Carrozza et al. 2005; Keogh et al. 2005; Youdell et al. 2008). *Saccharomyces pombe* has three different genes encoding for three isoforms: Pst1, Pst2 and Pst3 (Dang et al. 1999; Silverstein & Ekwall 2005). In *Drosophila*, 3 isoforms exist with different C-termini as a result from alternate splicing: Sin3 187, Sin3 190 and Sin3 220. These isoforms regulate different sets of genes depending on the stage of development (Sharma et al. 2008). For example, Sin3 187 is found in differentiated cells whereas Sin3 220 is involved in the proliferation stage. Unlike the other two isoforms, Sin3 190 appears to be maternally inherited and it can be found only in adult females and embryos (Sharma et al. 2008).

Mammals contain 2 paralogs: Sin3a and Sin3b. Sin3a/b are encoded from different genes and they both have splice variants (Ayer et al. 1995). The two paralogs are 57% identical and they have overlapping localization (Silverstein & Ekwall 2005). Functional studies have shown that the different isoforms have non-overlapping roles and they can bind distinct transcriptional co-repressors (Alland et al. 1997; Q. Yang et al. 2000; Brubaker et al. 2000; David et al. 2008; van Oevelen et al. 2008; Jelinic et al. 2011).

The presence of several isoforms and paralogs, associated with distinct functions for each isoform, contributes to the variety of Sin3 functions in fine-tuning gene the regulation process (Spronk et al. 2000; Brubaker et al. 2000; Jelinic et al. 2011).

Functions

The Sin3 protein acts as a scaffold within the complex and it interacts with HDAC1/2 and several transcription factors. The Sin3 complex does not have a chromatin binding motif, but it is recruited by sequence-specific DNA-binding transcription factors (including Rest, Mxd1, Mad1, Klf proteins as well as N-CoR and SMRT) to target genes (Silverstein & Ekwall 2005).

From yeast to mammals the Sin3 complex is involved in embryonic development, stem cell differentiation, cell cycle regulation, proliferation, oncogene regulation and senescence. The role of Sin3 in these functions is to either repress or activate gene transcription via its multiple protein–protein interactions depending on the cellular state. Gene repression is linked to a change in the acetylation state of histones both at promoter and transcribed regions, however the gene activation mechanism is still unclear (Kadamb et al. 2013; Baymaz et al. 2015).

In mammals the two paralogs Sin3a and Sin3b show high similarity in both their sequence and expression pattern. However, they also possess non-overlapping functions. While Sin3a is essential for the early stage of embryonic development, Sin3b is important for the late-gestation stage (Dannenbergh et al. 2005; David et al. 2008). It was also shown that Sin3B can interact with an E3 ubiquitin ligase and it is involved in proteasomal degradation (Q. Kong et al. 2010). Sin3a was found to be involved in DNA replication and repair, apoptosis, cell cycle regulation and chromatin modification. Further analysis showed that Sin3A modulates the transcriptional network controlled by Myc-Mad, E2F and p53 and is involved in senescence induction (Garcia & Pereira-Smith 2008; Grandinetti et al. 2009; Bansal et al. 2011). Further studies revealed the involvement of Sin3a in muscle and sarcomere development, as well as in reproduction since apoptosis-induced sterility can be induced by the down-regulation of Sin3a in male germline (van Oevelen et al. 2008; Pellegrino et al. 2012).

Tumorigenesis studies have implicated Sin3a in cancer. Since Sin3 can have a dual effect (repression or activation) of gene transcription depending on the cell conditions, Sin3a can have either proliferative or non-proliferative effects on transformed cells. Indeed, Sin3a can interact with Breast Cancer Metastasis Suppressor 1 (BRMS1) and

repress metastasis, but it can also induce the survival of transformed cells (Dannenberg et al. 2005; Hurst 2012).

Complex components

The Sin3a protein consists of 1273 amino acids and acts as a scaffold by providing multiple protein binding interfaces through its multiple domains (figure 1.8). Sin3 is characterized by six very well conserved domains: 4 Paired Amphipathic Helices (PAH), one HDAC Interaction Domain (HID) and one Highly Conserved Region (HCR) (Laherty et al. 1997; Brubaker et al. 2000; Le Guezennec et al. 2006; Sahu et al. 2008). The central region (PAH3-HID) is dedicated to the complex assembly, while the N-terminal region (PAH1-2) is required for chromatin targeting and the C-terminal region (PAH4-HCR) allows protein binding and intra-molecular folding (Grzenda et al. 2009).

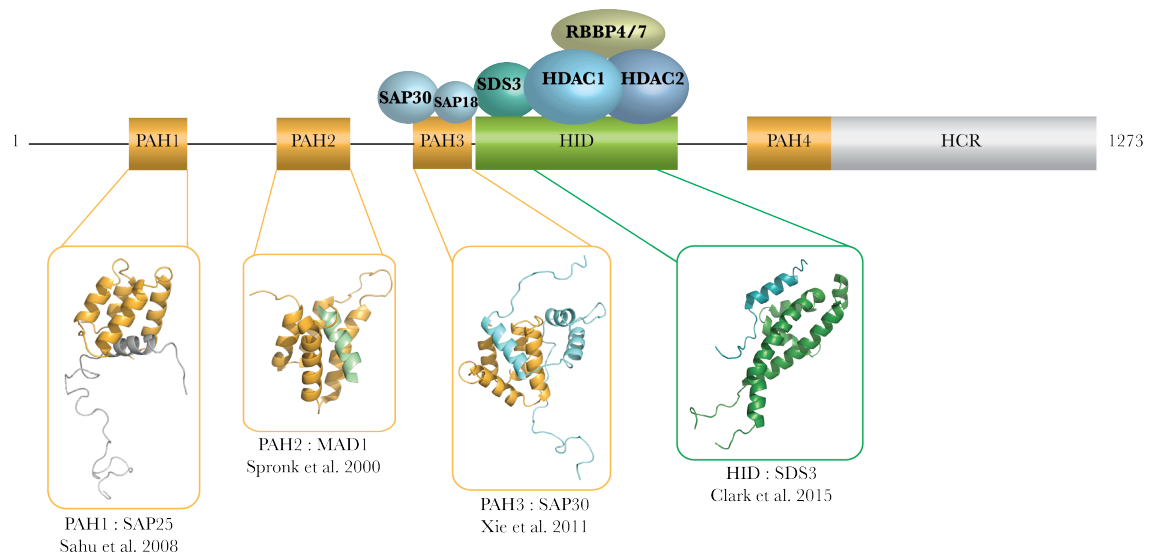


Figure 1.8: Schematic representation of the Sin3a complex.

The scaffold Sin3a protein contains 4 Paired Amphipathic Helices (PAH) domains, HDAC interaction Domain (HID) and a Highly Conserved Region (HCR). The structure of interaction between Sin3a and Sin3 Associated Proteins (SAPs) are shown in the boxes below the corresponding region of interaction.

The PAH domain comprises approximately 100 amino acids and it is arranged in a four paired amphipathic helix motif that shares structural similarity with the helix-loop-helix dimerization domain of the Myc transcription factors (He & Radhakrishnan 2008). PAH1-3 share this pre-folded binding module, but PAH4 sequence differs to the other PAH domains and it is likely to adopt a distinct fold (Le Guezennec et al. 2006; van

Ingen et al. 2006). PAH1 and PAH2 share high sequence similarity and they provide a binding interface for several DNA-binding transcription factors. However, these domains recognize different sequence motifs allowing different target specificity (Le Guezennec et al. 2006; Sahu et al. 2008). PAH1 interacts with SAP25 (Sahu et al. 2008) and PAH2 interacts with the transcription factor Mxd1 (previously known as Mad1) (structures are shown in figure 9) (Alland et al. 1997; Heinzel et al. 1997; Spronk et al. 2000).

The central region comprising PAH3 and HID domains interacts with the core complex components HDAC1/2, SDS3/SAP45, SAP30, SAP18, SAP130 and SAP180 (Y. Zhang et al. 1997; Hassig et al. 1997; Fleischer et al. 2003; Alland et al. 1997; Lechner et al. 2000). In 2011, the structure of PAH3 interacting with SAP30 was solved by NMR (shown in figure 1.9), whereas the structure of the interaction of SAP18, SAP130 and SAP180 with Sin3a has to be solved but the boundaries of interactions are known to fall in this region (Y. Zhang et al. 1997; Fleischer et al. 2003; Xie et al. 2011).

SAP30 is involved in gene repression by linking Sin3a to other complex members as well as to DNA directly (Viiri et al. 2009). The specific role of the other Sin3a Associated Proteins (SAPs) is still to be clarified but it is thought that they help stabilizing the interaction between complex members. RBBP4 and RBBP7 have also been identified as part of the Sin3a complex but they interact with Sin3a through HDAC1/2 and SAP30 (Hassig et al. 1997; Y. Zhang et al. 1998; Y. Zhang et al. 1999). RBBP4/7 are not necessary for the deacetylase activity but they are involved in cell cycle regulation and can interact with histones H2A, H3 and H4 (Qian & E. Y. Lee 1995; Taunton et al. 1996; Yoon et al. 2005).

Recently, Streubel *et al.* demonstrated that the Sin3a complex composition varies between embryonic stem (ES) and differentiated cells. In ES cells Fam60s, Tet1 and Ogt are defining subunits of the complex and it helps maintaining rapid cellular proliferation (Streubel et al. 2017).

The HID domain was shown to interact with the catalytic subunits HDAC1 and HDAC2 (Laherty et al. 1997; Portolano et al. 2014). Within the same region, Suppressor of Defective Silencing (SDS3, previously known as SAP45) interacts with Sin3a and

recently the NMR structure of the interaction between Sin3a and SDS3 was solved (Clark, Marcum, et al. 2015).

Several publications have demonstrated SDS3 to be an important component of the core complex. Previous studies in the Schwabe group have shown that the addition of SDS3 to the binary complex Sin3a : HDAC1 helps stabilizing the complex and leads to an increased protein yield (Portolano et al. 2014). Functional studies have also implicated SDS3 in the formation of pericentric heterochromatin. Deletion of SDS3 leads to chromosome missegregation, probably due to the disruption of the complex integrity and failure of HDAC recruitment (Lechner et al. 2000; David et al. 2003). The recently solved structure of Sin3a and SDS3 shows 4 helices from Sin3a₆₀₁₋₇₄₂ interacting with a small single helix from SDS₂₀₁₋₂₃₄ (Clark, Marcum, et al. 2015). This structure confirmed that SDS3 interacts within the HID domain of Sin3a, but as only a small portion of SDS3 is in the structure, it is still possible that SDS3 has multiple points of interaction with Sin3a along its full length.

The exact boundaries of interaction of HDAC1/2 with Sin3a are still unknown. Laherty *et al.* were the first to define the HID to the region incorporating amino acids 524 to 899 (Laherty et al. 1997). Subsequently, Portolano *et al.* (2014) found that Sin3a₅₃₁₋₇₇₂ was sufficient to interact stably with HDAC1 (Figure 1.9).

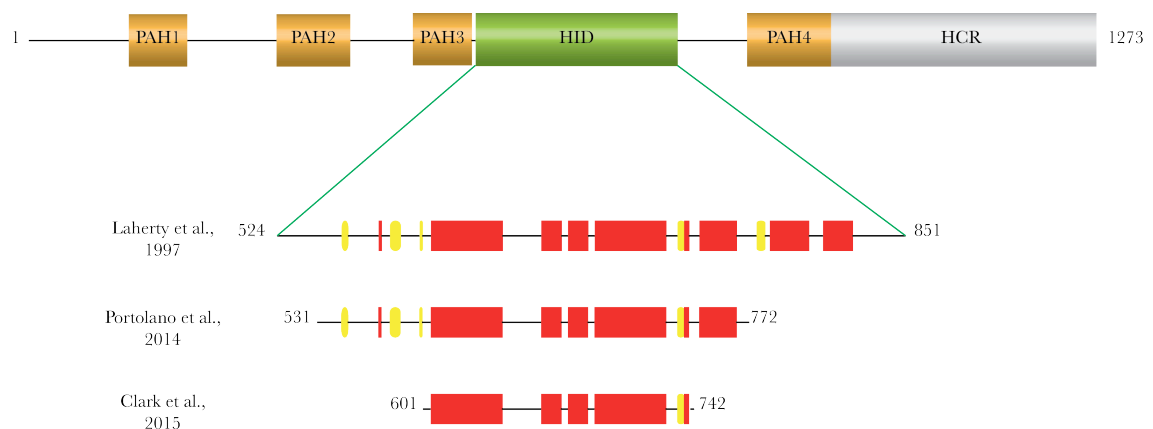


Figure 1.9: Schematic representation of the Sin3a complex with the published boundaries of the HID.

Over the years, several studies have characterized the HID domain. The predicted secondary structure for the HID domain is shown. The red boxes represent alpha helices and the yellow boxes represent beta strands.

The secondary structure prediction analysis of the HID predicts an alpha helix rich region along most of the length of the domain. However, the HID is not homologous to the SANT domain contained in the other Class I co-repressor complexes which is necessary for HDAC recruitment.

The lack of a SANT domain combined with the importance of SDS3 in the stability of the complex suggests a unique mode of HDAC recruitment for the Sin3a complex.

1.6. Thesis overview and project aims

In order to obtain a better understanding of both the Sin3a complex and how it is related to the other Class I HDAC complexes, the overall aims of this project are:

- To use a variety of biochemical techniques to elucidate the architecture of the core complex containing Sin3a, HDAC1 and SDS3.
- To understand the regulatory mechanism of the Sin3a complex with functional studies.
- To use stapled peptides to target specifically the core complex assembly and its recruitment by transcription factors.

A deeper insight into the structure and function of the core complex would show how the Sin3a complex is related to the other Class I HDAC complexes and could be used to target the Sin3a complex without inhibiting the other complexes.

Chapter 3 describes the details of the optimization trials performed for protein complex expression and purification. Different protein constructs were designed for each member of the protein complex in order to find the most stable sample and to map the regions of interaction required between the complex components. Two expression systems were also compared: bacterial and mammalian expression. The bacterial expression (*E. coli*) and the mammalian expression (HEK 293F) were compared with each other in terms of the protein yield achieved and the resulting deacetylase activity of the complex. In order to find the most stable complex suitable for structural studies, a variety of affinity tags and combinations of complex components were also tested. Furthermore, it was found that maintenance of low temperature during the gel filtration process in the last purification step was

important. Additionally, an alternative method of purification was found using sucrose gradients in order to purify the complex in a gentle manner. Finally, as part of the initial characterization of Sin3a, microfluidic mobility shift assays were performed to understand any histone substrate preference of the Sin3a complex.

Chapter 4 describes the use of stapled peptide to specifically target the Sin3a complex. The stapled peptides were designed based upon two Sin3a interacting proteins: SDS3 and Mxd1. If the designed peptides were able to mimic the interaction between Sin3a and SDS3 or Mxd1, then they could potentially displace the native protein interaction to disrupt its function. Displacement of SDS3 could potentially lead to the disruption of the core complex, affecting the deacetylase activity of the complex. If Mxd1 could be effectively displaced by the stapled peptides, then the core complex would remain intact but the specific mechanism of recruitment of Sin3a to chromatin would be disrupted.

The efficiency of the stapled peptides was assessed both *in vitro* and *in cellulo*. Different treatments were also performed on the cells in order to enhance peptide delivery. Finally, the presence of the peptides within the cells was assessed by adding a GFP tag to the peptides and by visualizing the transfected cells with a Nikon TE300 inverted microscope.

Chapter 5 describes the experiments performed in order to understand the structure of the Sin3a complex. To date a few structures have been published of the Sin3a PAH domains. The size of the complex, and hence the stoichiometry, was measured with the use of Size Exclusion Chromatography coupled with Multi-Angle Light Scattering (SEC-MALS). A low-resolution envelope was obtained using Small-Angle X-Ray Scattering (SAXS) and information about proximity of the proteins within the complex was obtained using chemical cross-linking. A structural model was built using a combination of the information obtained using these different techniques in addition to negative stain electron microscopy.

Chapter 2. Materials and Methods

2.1. Materials

2.1.1. Standard chemicals and reagents

All chemicals and reagents were of analytical grade or higher and were purchased from Sigma-Aldrich, Fisher Scientific, Merk Millipore or Melford unless stated otherwise. KOD Hot Start DNA Polymerase A was purchased from Novagen. Lysozyme and porcine trypsin were supplied by Sigma Aldrich. Heat-treated RNase A (Sigma) and DNase I (Sigma) were prepared in house, TEV protease was expressed and purified by Jacquie Greenwood. Pre-cast Novex® NuPAGE® 4-12% gradient bis-tris SDS PAGE gels and NuPAGE® MES-SDS running buffer were purchased from Life Technologies (UK).

2.1.2. Bacterial and mammalian cell lines

Competent DH5α *E. coli* strain for plasmid amplification, Freestyle™ HEK 293 (HEK 293F) and Rosetta DE3 cell lines for protein expression were purchased from Invitrogen. HEK 293F suspension cells are a Life Technologies product and they were cultured in FreeStyle™ 293F Expression Medium which was purchased from Gibco® Life technologies. HEK 293T monolayer cells were provided by ACTT® and they were cultured in M10 media (See below for recipe).

M10 media

DMEM

10% FBS

1% PS

2.2. Generating constructs

2.2.1. Bioinformatics

Protein sequences were acquired from the Uniprot database. For each protein, the disorder prediction was analysed using RONN software (Z. R. Yang et al. 2005). The secondary structure prediction was analysed using Jpred (Cuff et al. 1998). Sequences were aligned using MultiAlign and Clustal software (Corpet 1988; Sievers et al. 2011).

2.2.2. Primer design

Primers were manually designed using Enzyme X software (nucleobytes.com). All primers were designed to have 21 bp annealing region to the gene of interest plus a vector homology region (16 bp on forward primers and 18 bp on reverse primers) for cloning the PCR product into an expression vector. Primers were purchased from Eurofins MWG operon. Depending on the experiment, the vector homology regions were designed to either include or exclude an affinity tag for protein purification, either at the N- or at the C-terminus of the protein. The vectors were selected based on the expression system of choice.

2.2.3. Amplification from cDNA library of the genes of interest

Genes of interest were amplified by PCR from either a cDNA library (Megaman Transcriptome Library, Agilent) or from plasmid templates previously designed and purified by Dr Nicola Portolano.

Amplification PCR was performed using the commercially available KOD Hot Start DNA Polymerase kit (Novagen® Toyobo) using the protocol in table 2.7.

Table 2.7: PCR protocol.

Cycle parameters and reagents required to amplify genes from the cDNA library.

Mastermix (50 µl reaction)	Thermic Profile	
30 µl MilliQ H ₂ O 5 units 10x Buffer (KOD hot start) 200 µM dNTPs 1 mM MgSO ₄ 20-200 ng Template DNA 300 nM Forward Primer 300 nM Reverse Primer 1 unit KOD hot start polymerase	2 min 94 °C	Enzyme pre-activation
	15 sec 94 °C	Denaturation
	15 sec 55 °C, 65 °C	Annealing
	30 sec 68 °C	Extension
	X 30 cycles	
	∞ 10°C	

The PCR products were run on a 1% agarose gel containing 0.005% Ethidium Bromide. Electrophoresis was carried out in an electrophoresis apparatus containing 1 X TBE buffer for 45 minutes at 120 V. DNA bands were visualised and photographed using a UV light. Hyperladder 1kb (Bioline) was used as DNA molecular weight marker. The gel band containing the PCR product of the right size was purified using a QIAquick Gel Extraction kit (QIAGEN) following the manufacturer's instructions. The purified PCR products were then cloned into the required expression vector by Dr Xiaowen Yang and Dipti Vashi at the PROTEX cloning facility of the University of Leicester using the In-fusion® HD Cloning kit (Takara Bio).

2.2.4. Plasmid sequencing

The plasmid concentration was determined by nanodrop using an IMPLEN nanophotometer in accordance with the manufacturer's instructions. All the protein expression vectors were sequenced by the Protein and Nucleic Acid Chemistry Lab

(PNAOL) at the University of Leicester. Sequencing results were then analysed using MacVector software (macvector.com).

2.3. Plasmid preparation

2.3.1. DH5 α competent cells preparation

DH5 α cells were purchased from Invitrogen. To prepare competent cells, DH5 α cells were spread on a TYE plate and incubated at 37 °C overnight. The next day, a streak of cells was inoculated into 0.5 litre of 2xTY media. Cells were grown in a shaking incubator at 120 rpm, 37 °C until the O.D.₆₀₀ value reached 0.3. Cells were then harvested at 3500 xg (Sorvall SLC-6000), for 10 min at 4 °C. Cells were gently re-suspended in 100 ml of pre-chilled 0.1 M MgCl₂ and incubated on ice for at least 30 minutes. Cells were then spun at 4 °C at 3220 x g for 10 min (Eppendorf 5810R Rotor A-4-81). The cell pellet was gently re-suspended in 50 ml pre-chilled 0.1 M CaCl₂ 14% glycerol. Finally, the cells were aliquoted into pre-chilled 600 μ l Eppendorf tubes. Competent cells were stored at -80 °C.

0.1 MgCl₂

Filtered sterilized

0.1 CaCl₂

14% glycerol

Filtered sterilized

2xTY media

15 g/L Bacto Tryptone

10 g/L Bacto Yeast Extract

5 g/L NaCl

TYE plate

15 g/L agarose

8 g/L NaCl

10g/L Bacto Tryptone

5 g/L Bacto Yeast Extract

2.4. Transformation

DH5 α competent cells were transformed to amplify the plasmid DNA. After adding 1 μ l of plasmid to 100 μ l of competent cells, the DNA : cells mixture was incubated on ice for 20 min. The transformed DH5 α cells were then spread on a TYE (plus ampicillin 30 μ g/ml) plate, and incubated at 37 °C overnight. When using kanamycin or chloramphenicol (30 μ g/ml), the *E. coli* were incubated for 30 min with 2xTY at 37 °C before plating.

TYE plate + 30 µg/ml ampicillin

15 g/L agarose

8 g/L NaCl

10g/L Bacto Tryptone

5 g/L yeast extract

2.4.1. Small scale plasmid preparation

Small scale of plasmid DNA purification used commercially available kits based on the alkaline lysis method. A single colony from a transformed TYE plate is inoculated into 10 ml of 2xTY culture plus the required antibiotic(s) for the plasmid of interest. The culture is incubated for approximately 16 hours at 200 rpm, 37 °C in a shaking incubator, before being harvested. The plasmid DNA is purified with a QIAprep Spin miniprep Kit (QUIAGEN) following the manufacturer's instructions.

2.4.2. Large scale plasmid preparation

0.75 L of 2xTY media with the appropriate antibiotic(s) (recipes for media and buffers are described below) was inoculated with a streak from a freshly transformed plate. Cells were grown for approximately 16 hours in a shaking incubator at 120 rpm at 37 °C. Cells were then harvested at 3500 x g (Sorvall SLC-6000) for 10 minutes at 4 °C. After re-suspending the pellet into 10 ml of starting buffer, 5 ml of lysozyme buffer were added to the mix and the lysate was incubated at room temperature for 5 minutes. 30 ml of alkaline buffer was added and the mixture was incubated at room temperature for 5 minutes, followed by an addition of 22.5 ml of neutralization buffer. The denaturation and neutralization steps result in soluble plasmid. The mixture was incubated at room temperature for 5 minutes and then the precipitant was removed by spinning down the lysate for 10 minutes, 3220 x g (Eppendorf 5810R Rotor A-4-81), at 4 °C. The supernatant was filtered through miracloth (22-25 µm pore size, Calbiochem) and then 0.6 volume of isopropanol was added to the filtered

supernatant. After 5 minutes incubation at room temperature the mixture was spun down for 10 minutes, 3220 x g, at 4 °C. The pellet was re-suspended into a volume of 15 ml with re-suspension buffer 1. To precipitate RNA and proteins, an equal volume of pre-chilled 5 M LiCl was added and the sample was incubated on ice for 5 minutes before spinning down the mixture for 10 minutes, 3220 x g, at 4 °C. To precipitate the plasmid DNA, 0.6 volume of ethanol was added to the supernatant, and after 5 minutes incubation the sample was spun down for 10 minutes, 3220 x g, at 4 °C. The pellet was then re-suspended in a total volume of 10 ml of re-suspension buffer 2. To remove any remaining RNA, 50 µl of heat treated RNase A (stock solution 10 mg/ml) was added to the solution. The sample was incubated at room temperature for 15 minutes, followed by the addition of 2.5 ml of PEG buffer to precipitate the plasmid DNA again. An incubation of 30 minutes on ice was followed by spinning the sample for 10 minutes, 3220 x g, at 4 °C. The pellet was re-suspended in 10 ml of re-suspension buffer 2 and 2 ml of chloroform was added, followed by a brief vortexing of the sample to remove both PEG and residual proteins. The sample was spun down for 10 minutes, 3220 x g, at 4 °C, and then the top aqueous layer containing the DNA was transferred to a fresh tube. 1/10 volume of 5 M NaCl and 3 volumes of 100% ethanol were added to precipitate DNA, followed by spinning the DNA sample for 10 minutes, 3220 x g at 4 °C. To remove the excess of NaCl, 5 ml of 70% ethanol was added before spinning down the mixture for an additional 10 minutes, 3220 x g, at 4 °C. The supernatant was discarded and the pellet was left to dry overnight at room temperature, with the tube upside down on clean paper tissue. The dried pellet was re-suspended with 5 ml of sterile filtered TE buffer and stored at -20 °C.

Starting buffer

25 mM Tris pH 8.0
10 mM EDTA pH 8.0

Lysozyme buffer

25 mM Tris pH 8.0
10 mM EDTA pH 8.0
10 mg/ml lysozyme

Alkaline buffer

0.2 M NaOH
1% SDS

Neutralization buffer

3 M Potassium acetate pH 5.0

Re-suspension buffer 1

10 mM Tris pH 8.0
10 mM EDTA pH 8.0

Re-suspension buffer 2

10 mM Tris pH 8

1 mM EDTA

TE buffer

10 mM Tris pH 8

0.1mM EDTA

PEG buffer

30% PEG 6000

2.5 M NaCl

2.4.3. Quantification of plasmid DNA

The final concentration of the purified plasmid DNA concentration was determined using a NanoPhotometer (IMPLEN) following the manufacturer's instructions.

2.5. Protein expression and purification (E. coli expression)**2.5.1. Rosetta DE3 transformation**

Competent Rosetta DE3 cells were transformed with the required plasmid DNA and plated out onto TYE agar containing chloramphenicol plus the appropriate antibiotic(s) (see recipe in section 3b) for the plasmid resistance. Plates were incubated at 37 °C overnight. Subsequently cultures were prepared by inoculating 0.75 L of 2xTY (see recipe in section 3a) with the appropriate antibiotics and a streak of freshly transformed colonies. Flasks were then incubated at 37 °C, 120 rpm in a shaking incubator. Protein expression was induced with 40 µM IPTG once the O.D.₆₀₀ value reached 0.2. Cells were incubated overnight at 20 °C and then harvested at 3500 x g (Sorvall SLC-6000) for 10 min at 4 °C. Cell pellets were stored at -80 °C.

2.5.2. GST tagged protein purification

The cell pellet from 0.75 litre of cells was defrosted and re-suspended in 15 ml of Glutathione S-transferase (GST) lysis buffer (see recipe below). The resuspended cells were sonicated for 3 cycles (15 seconds on and 15 seconds off) with an amplitude of 10 (MSE, SoniPrep150) and centrifuged at 108,000 x g at 4 °C for 30 seconds. The pellet was discarded and the cleared lysate was incubated with 1 ml Glutathione sepharose 4B (GE healthcare) slurry for every 0.75 L culture (the GST resin was equilibrated in GST wash buffer prior use, recipe below), the protein : resin incubation was carried out rotating the sample for 30 minutes at 4°C. The resin was then spun down for 5 minutes at 3000 x g and then it was washed 3 times with 10 ml of GST wash buffer (recipe below). The concentration of proteins was determined with the Bio-Rad protein assay dye following the manufacturer's protocol, as described in section 7b. The protein was finally then eluted from the resin by TEV cleavage (40 µg of TEV protease for litre of culture) in 4 volumes of GST-wash buffer at 4 °C overnight.

GST lysis buffer

1 x PBS

0.5 mM DTT

Roche complete protease inhibitor - EDTA free (1 tablet per 50 ml)

GST wash buffer

1 x PBS

0.5 mM DTT

2.5.3. His tagged protein purification

A cell pellet from 0.75 litre of cells was defrosted and re-suspended in 15 ml of Ni-NTA lysis buffer (see recipe below). The resuspended cells were sonicated for 3 cycles (15 seconds on and 15 seconds off) with amplitude of 10 (MSE, SoniPrep150). The insoluble fractions were removed by centrifugation at 108,000 x g at 4 °C for 30 minutes. 1.2 ml Ni-NTA agarose slurry for every 0.75 litre of culture. The Ni-NTA (Qiagen) resin was equilibrated in Ni-NTA wash buffer (see recipe below) prior use. After the centrifugation of the lysate, the supernatant was incubated with the washed

Ni-NTA resin for 30 minutes at 4 °C. The sample as then spun down for 5 minutes at 3000 x g. The resin was washed 3 times with 10 ml of Ni-NTA wash buffer. The protein was finally eluted with TEV (40 µg of TEV protease for litre of culture) in 5 resin volumes of Ni-elution buffer (recipe below).

Ni-NTA lysis buffer

100 mM Potassium acetate

50 mM Tris pH 7.5

Roche complete protease inhibitor - EDTA free (1 tablet per 50 ml)

Ni-NTA wash buffer

100 mM Potassium acetate

50 mM Tris pH 7.5

2.6. Protein expression and purification in mammalian HEK 293F

2.6.1. Culture of HEK 293F cells

HEK 293F suspension cells (Gibco®) were grown, maintained and transfected in sterile conditions in a Class II laminar flow hood. Both the hood and working materials were sterilized with 70% ethanol to prevent infections. Serum free FreeStyle™ 293 expression media was pre-warmed in a 37 °C water bath prior any cell culture work.

HEK 293F cells were typically sub-cultured when reaching a final density of 2×10^6 cells/ml. HEK 293F cells double approximately every 24 hours. The cell density was counted using a glass haemocytometer. The cell stock was divided to a density of 0.35×10^6 cells/ml using a volume of 30-60 ml in a 250 ml Erlenmeyer conical flask (Corning, Fisher Scientific) sealed with vented caps. Larger scale cultures (300ml) were grown in 1 litre (490 cm^3) roller bottles (Corning®, Fisher Scientific) sealed with vented caps (Corning®). The small cultures were incubated in a humidified orbital shaking incubator (N-BIOTEK) at 8% CO₂, 120 rpm, 37 °C, and the larger cultures were incubated with at the same conditions but at 135 rpm.

2.6.2. Transient transfection in HEK 293F

Small, medium and large scale transient transfections and purifications were carried out using branched polyethylenimine (PEI) as a transfection reagent and following the transfection protocol from Portolano in 2014 (Portolano et al. 2014; Nigi et al. 2017). A stock solution of PEI was made at 0.5 mg/ml and the pH was adjusted to 7.0 with hydrochloric acid. The volumes and quantities of reagents were scaled proportionally to the working volumes. 24 hours before transfection experiment, cells were split to a final density of 0.5×10^6 cells/ml into an appropriate number of 250 ml conical cell culture flasks (for small scale transfections), or into 1 L roller bottles (for medium and large scale transfections). Multiple bottles were used when a larger volume was needed. Cells were incubated for 24 hours into a humidified shaking incubator at 37 °C, 8% CO₂ and 135 rpm. After 24 hours the cells would have reached a density of 1×10^6 cells/ml. Transfection reactions were prepared accordingly with the cell volume transfected (as shown in table 2).

Table 2.8: Transfection reaction recipes.

Volumes are calculated according to the final volume of cells to be transfected.

60 ml cells	300 ml cells
60 ml 1x PBS	30 ml 1x PBS
240 µl PEI	1.2 ml PEI
60 µg DNA	300 µg DNA

The appropriate amount of DNA was diluted into 1 x PBS and then mixed with PEI. The DNA : PBS : PEI mixture was briefly vortexed and incubated at room temperature for 20-50 minutes before being poured into the cells. The transfected cells were incubated for 48 hours into a humidified shaking incubator at 37 °C, 8% CO₂ and 135 rpm. Cells were harvested after 48 hours from the transfection by centrifuging for 5 minutes at 3500 x g at 4 °C, the pellet could be directly used for purification, or stored at -80 °C.

PEI solution

0.5 mg/ml PEI pH 7.0

1x PBS

Dulbecco's phosphate buffered saline (SIGMA)

2.6.3. Flag-tagged protein purification

All the recipes for the buffers used for the protein purification are listed below. A cell pellet from 1.2 L (fresh or defrosted) was re-suspended to a final volume of c. 40 ml in FLAG lysis buffer by pipetting up and down. The cell suspension was then glass homogenized five times, sonicated for 3 cycles (30 seconds on, 30 seconds off) with an amplitude of 10 (MSE, SoniPrep150) and centrifuged at 108,000 x g at 4 °C for 30 minutes to remove the insoluble fraction. 1.25 ml of anti-FLAG[®] affinity resin slurry per litre of culture was washed three times with 10 ml of FLAG equilibration buffer. The cleared lysate was incubated and bound to the resin with gentle rotation for approximately 1 hour at 4 °C. The protein sample was then spun down for 5 minutes at 3000 x g and the supernatant was discarded. The protein bound to the resin was then washed with 10 ml of FLAG lysis buffer, followed by two washes with 10 ml of wash buffer, high salt buffer and two final washes with 10 ml of cleavage buffer.

Subsequently the resin was re-suspended in 6-8 ml of TEV cleavage buffer and 20-30 µl of 1.2 mg/ml TEV protease was added to the sample. The tag was cleaved overnight by gently rotating the sample at 4°C. After the overnight TEV cleavage, the sample was centrifuged for 5 minutes at 3000 x g and the resin pellet was discarded. The eluted protein was then ready for further purification. 10 µl of sample was taken from both the sample with the protein bound to the resin, the TEV-cleaved resin and the supernatant after TEV cleavage. Each 10 µl sample was diluted into 10 µl of Protein loading buffer (2X) (recipe in section 7a) for SDS-PAGE analysis.

FLAG lysis buffer

50 mM Tris pH 7.5

100 mM Potassium acetate

5% Glycerol

0.3% Triton X-100

Wash buffer

50 mM Tris pH 7.5

50 mM Potassium acetate

5% Glycerol

Roche complete protease inhibitor - EDTA free (1 tablet per 50 ml)

Wash buffer high salt

25 mM Tris pH 7.5

50 mM Potassium acetate

5% Glycerol

TEV Cleavage buffer

50 mM Tris pH 7.5

25 mM Potassium acetate

5% Glycerol

0.5 mM TCEP

2.7. Further purification methods

2.7.1. Gel filtration chromatography

The TEV eluted protein complex was concentrated to 500 µl using a 15 ml Amicon ultra centrifugal concentrator (Millipore) with a 10.0 kDa molecular weight cut off. The concentrated protein complex was then filtered through a 0.22 µm filter and then purified through size exclusion chromatography (Superose 6 (10/300) or Superdex-200 (10/300) GL columns, GE Healthcare). Prior loading the sample, the column was equilibrated with one column volume of gel filtration buffer (See recipe below). The concentrated sample was then loaded on the column and gel filtrated at a 0.5 ml/min flow rate. 500 µl fractions were collected and 10 µl of sample from each fraction was taken and diluted into 10 µl of Protein loading buffer (2X) (recipe in section 7a) for later analyses by SDS PAGE. The fractions containing the protein complex were pooled for subsequent experiments.

Gel filtration buffer

50 mM Tris pH 7.5

150 mM Potassium acetate

0.5 mM TCEP

2.7.2. Sucrose density gradient

The sucrose gradient purification method was used either as an alternative or as a further purification step from the gel filtration method described in section 6a. The

protein complex was concentrated to a final volume of 50-200 µl using a 15 ml Amicon ultra centrifugal concentrator (Millipore) with a 10.0 kDa molecular weight cut off. A 5-25% sucrose gradient (with/without 0%-0.1% glutaraldehyde) was prepared in a 4.4 ml Thin-Walled tube (Sorvall) using a Gradient Master IP (BIOCOMP). The protein sample was then added directly on top of the gradient and centrifuged at 166,000 x g for 15 hours in a swing-out rotor (TH660, Thermofisher®). The gradient was manually fractionated into 175 µl fractions starting from the top of the tube. The collected fractions were run on SDS PAGE and visualized with Coomassie stain.

5% Sucrose buffer

20 mM HEPES pH 7.5

5% Sucrose

40 mM NaCl

25% Sucrose buffer

20 mM HEPES pH 7.5

25% Sucrose

40 mM NaCl

(±)0.1% glutaraldehyde

2.8. Protein analysis

2.8.1. SDS PAGE

Protein samples were analysed by SDS-PAGE. 10 µl of protein sample was diluted into 10 µl of protein loading buffer (2X) (see recipe below) and then loaded on NuPAGE® 4-12% Bis-Tris gel (Novex) with NuPAGE® MES SDS running buffer. Electrophoresis was performed at 125 mA, 200 V for 35 minutes. SeeBlue2 protein pre-stained standards (Invitrogen) were used molecular weight markers. The protein was then visualised by staining with Instant Blue Coomassie Stain (Expedeon) for at least 1 hour before de-staining with deionised water.

Protein loading buffer (2X)

0.125 M Tris pH 6.8

1% (w/v) SDS

20% (v/v) glycerol

20 mM DTT

0.01% bromophenol blue

2.8.2. Protein quantification

The protein quantification was performed by using the Bio-Rad® protein assay dye reagent. 200 µl of Bio-Rad reagent was diluted into 800 µl of deionised H₂O to set the blank. 2 µl of protein sample was added to the blank and after mixing the solution by pipetting several times, the absorbance was measured at 595 nm. The concentration in mg/ml was calculated by multiplying the reading at 595 nm by 17 (factor calculated from a BSA standard curve).

2.8.3. Size Exclusion Chromatography – Multi-angle Light Scattering (SEC-MALS)

SEC-MALS was used to analyse the size of the Sin3a complex. A Superose 6 (10/300) GL column was used attached to an AKTA purifier which was in turn attached to the MALS instrumentation (Dawn® HELEOS® II and Optilab® T-rEX Wyatt Technology). The column was equilibrated with the gel filtration buffer (see recipe in section 6.a). The experiment set up and the light scattering data analysis were done using the ASTRA software (Wyatt Technology) following the manufacturer's instructions.

2.8.4. Liquid chromatography-tandem mass spectrometry (LC-MS/MS)

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was used for protein identification. After analysing the sample using SDS PAGE (details in section 7.a), the bands of interest were excised and sent to PNACL (University of Leicester Core Biotechnology services) for LC-MS/MS. Data were analysed using Mascot Server (Matrix Science).

2.9. Tissue culture and transfection in monolayer HEK 293T

2.9.1. Culture of HEK 293T

HEK 293T monolayer cells (Thermofisher®) were grown, maintained, and transfected in sterile conditions. Both the Class II laminar flow hood and the required working materials were sterilized with 70% ethanol prior to use. HEK 293T cell stocks were maintained in 10 cm dish cell culture (Corning). To enhance the adherence of the cells to the tissue culture plates, the plates were pre-coated with gelatine solution (recipe below) 20 minutes prior to use. The M10 media (see below for recipe) was pre-warmed at 37 °C before starting the cell culture work.

When the monolayer cells reached an 85-90% confluence, the media was aspirated from the plates and the cells were washed once with 7 ml of PBS. Cells were subsequently lifted from the plate with 2 ml of trypsin solution (see below for buffer recipe). After 3 minutes incubation, the trypsin was neutralized by re-suspending the cells into 8 ml of pre-warmed M10 media. The cells were then counted using trypan blue staining (Thermofisher®). Stocks of cells were maintained by splitting 1 in 5 approximately every 2-3 days.

Gelatine solution	M10 media	Trypsin solution
99.9% D-PBS	DMEM	PBS
0.1% porcine gelatine	10% FBS	0.5 mM EDTA;
	1% PS	1% v/v chicken serum;
		2.5% w/v trypsin

2.9.2. Reporter assay transient transfection

All HEK 293T transfections were done in 48 well Nunc™ cell culture treated multidishes (Thermofisher). 24 hours before a transfection experiment, 2.5×10^6 cells were seeded in a final volume of 500 µl into the required number of pre-gelatinised wells and incubated for 24 hours into a humidified incubator at 37 °C and 5% CO₂ (details in section 8.a). Next, 25 µl of Optimem (Thermofisher®) was incubated with 1

µl of Lipofectamine 2000 (Thermofisher®) in a 1.5 ml centrifuge tube for 5 minutes for each transfection. At the same time, the DNA plasmid(s) were diluted into 25 µl Optimem in another 1.5 ml centrifuge tube. A total amount of 0.5 µg of DNA was added to each transfection, using equimolar amount of DNA for each plasmid. The Optimem : DNA and the Optimem : lipofectamine mixtures were then mixed together and incubated at room temperature for 20 minutes. The Optimem : DNA : lipofectamine mixture was pipetted into the cells seeded 24 hours earlier and the transfected cells were incubated for 48 hours into a humidified incubator at 37 °C and 5% CO₂. After 48 hours the media was removed. Cells were gently washed with PBS, followed by 140 µl of luciferase lysis buffer (1X) (recipe below). The 48 well plate was sealed with tape and incubated with the luciferase lysis buffer for two hours at room temperature on a plate shaker to ensure the complete coverage of the cells. Then, the lysed cells were frozen at - 80 °C for at least 30 minutes prior performing the luciferase assay.

Lysis buffer (5X)

5 mM Tris pH 7.8

10 mM DTT

10% 0.1 M EDTA

50% glycerol

5% Triton x-100

2.10. β-galactosidase

In a fresh vinyl 96 well plate (Costar®) 100 µl of β-galactosidase substrate (1X) was added to the diluted lysate (1:8 dilution factor). The β-galactosidase substrate : cell mixture was incubated at 37 °C for 5 minutes. In a Victor X5 plate reader (Perkin Elmer) the plate was briefly shaken and the absorbance was read at 420 nm.

2.11. Luciferase assay

The luciferase assay was performed in a fresh 96 well black plate (Costar®). The assay was performed with Luciferase Reporter Assay Kit (BioVision®) following the manufacturer's instructions. This assay included the addition to the cell lysate of buffer containing CoA, ATP and Mg²⁺. After 5 minutes incubation, the luminescent reaction was started with an injection of luciferin. In a Victor X5 plate reader (Perkin Elmer) the plate was briefly shaken and the absorbance was read at 454 nm.

2.12. Stapled peptide synthesis and purification

All the peptides used in the following experiments were synthesized and purified by Dr Naomi Robertson and Dr Andrew Jamieson in the Department of Chemistry, University of Leicester (Currently at the University of Glasgow).

2.13. Fluorescence polarization assay

The fluorescent polarization (FP) assay was performed in a fresh 96 well black plate (Greiner Bio-one), in a final volume of 25 µl. The Sin3a complex was serially diluted from 50 µM to 0.05 µM in FP buffer (see recipe below). The fluorescent native or stapled peptides were added at the fixed concentration of 100 nM. The plate was shaken for one minute to mix the samples in a Victor X5 plate reader (Perkin Elmer). Subsequently, fluorescence was excited at 480 and measured at 535 nm. The FP data analysis was done using GraphPad Prism 6.0 (GraphPad Software), where the K_d were calculated by nonlinear curve fitting with a one-site binding (hyperbola) model, $Y = B_{max} \times X / (K_d + X)$.

FP Buffer

25 mM Tris pH 7.5

50 mM NaCl

0.01% Triton X-100

0.1 mg/ml BSA

2.14. Histone deacetylase assay

The histone deacetylase assay was performed in a 96 well black plate (Corning) in a final volume of 50 μ l. In a shaking incubator, 50 nM of the Sin3a complex was incubated with 100 μ M of Boc-Lys(Ac)-AMC (BaChem) substrate at 37 °C for 30 minutes. 2 μ M of Trichostatin A was added to the developer buffer (see recipe below) prior adding 50 μ l of the developer buffer to develop the assay. Fluorescence was measured at an excitation wavelength of 335 and an emission wavelength of 460 nm using Victor X5 plate reader (Perkin Elmer) after incubating for 10 minutes. The HDAC assay data analysis was done using GraphPad Prism 6.0 (GraphPad Software). Kd values were calculated by nonlinear curve fitting with a one-site binding (hyperbola) model, $Y = b_{max} \times X / (Kd + X)$

Buffers ingredients:

Reaction buffer	Developer buffer
50 mM Tris pH7.5	50 mM Tris pH7.5
50 mM NaCl	100 mM NaCl
5% Glycerol	10 μ g/ μ l Trypsin

2.15. Microfluidic Mobility Shift Assay (MSA)

Microfluidic Mobility Shift Assay (MSA) was used to determine substrate turnover on a Caliper EZ reader II (Caliper Life Sciences). The assay was performed on a chip system where electrophoretic separation is used to characterise the protein sample.

The protein sample at the required concentration was aliquoted into a 384 wells microtiter plate (Perkin Elmer). Through a system of capillaries the protein was mixed with the substrate. If the substrate was deacetylated the resulting positive charge on the lysine residues determined its migration through the capillaries at a faster rate compared the acetylated substrates. The appearance of products (deacetylated substrates) and disappearance of reactants (acetylated substrates) were measured in

real time by the Caliper reader. Through the Caliper software it was possible to determine enzymatic rates and kinetic analysis.

2.16. Chemical protein cross-linking coupled with mass spectrometry assay

The Sin3a complex was prepared as described in section 2.6.3. The sucrose gradient fractions containing the protein were concentrated to a final concentration of 4 mg/ml using a 500 µl Amicon® Ultra centrifugal concentrator (Merck Millipore) with a 10 kDa cut-off. The cross-linking experiments were based on the protocol defined in 2013 by Lariviere (Larivière et al. 2013). Two different cross-linkers were used for these experiments: Isotopically-coded cross-linker Di-(SulfoSuccinimidyl)-Glutarate (DSSG) and Cyanur Biotin Dimercapto Propionyl Sulfo Succinimide (CBDPSS). Several dilutions of the cross-linkers were used to determine the ideal ratio of DSSG/CBDPSS to protein: 0.1-5 mM of DSSG/CBDPSS was added to 1 mg/ml of protein in a final volume of 10 µl. The protein : cross-linker mixture was incubated at room temperature for 1 hour and the cross-linking reaction was stopped by adding 40 mM (NH₄)₂SO₄ to the mixture and incubating for 10 minutes at room temperature. The cross-linked protein samples were run on a SDS-PAGE and visualized with Instant Blue Coomassie Stain. The protein bands corresponding to the cross-linked complex were cut and submitted for mass spectrometry to the PNACL facility at the University of Leicester. The initial data analysis was run by Dr Andrew Bottrill (PNACL, University of Leicester) and further data analyses were performed using xQuest by Dr Sidhu Khushwant.

2.17. Small Angle X-Ray Scattering (SAXS)

The Sin3a complex was prepared as described in section 2.6.3 with a Superose 6 (10/300) column. The protein complex was concentrated using a 500 µl Amicon® Ultra centrifugal concentrator (Merck Millipore) 10 kDa cut-off, to a final volume of 200 µl. The samples were loaded into a 96 well plate to a Size-Exclusion Chromatography (SEC) coupled SAXS with the Shodex 404 SEC column controlled by an

Agilent HPLC. The Shodex column was equilibrated with gel filtration buffer (recipe in section 6.a) prior use. With the help of Dr Peter Watson and Dr Robert Rambo, data were collected at Diamond Light Source B21 SAXS beamline, Didcot, Oxford. Data analysis was performed with ScÅtter (Biosis).

2.18. Protein crystallisation trials

The protein sample was expressed and purified as described in section 2.6.3 the purified protein was then concentrated to 2-9 mg/ml. Initial crystallization studies were carried on using the commercial crystallization screens: JCSG⁺, ProPlexTM, PACT and MORPHEUS. Crystallization experiments were performed in MRC 2 drop 96-well plates using the sitting drop vapor diffusion method. Each reservoir well of the plate was manually filled with 80 µl of mother liquor. Plates were set up using a Mosquito LCP crystallization robot (TTP Labtech) by mixing 100 nl protein sample with 100 nl reservoir solution creating 200 nl drops. Plates were manually sealed with a transparent sheet and stored at room temperature or 4°C. The plates were regularly checked with an optical microscope.

2.19. Negative stain Electron Microscopy

The Sin3a complex was expressed and purified as described in section 2.6.3. The complex was cross-linked during the sucrose gradient step, 0.1% glutaraldehyde was added to the 25% sucrose buffer prior making the gradient with the gradient master. The sucrose gradient fractions were run on SDS-PAGE and analysed with Coomassie stained. A single fraction containing the cross-linked sample showing the least amount of impurities was incubated 20 minutes on ice with 4 volumes of gel filtration buffer (recipe in section 6a) to stop the cross-linking reaction. The sample was then concentrated and diluted twice to remove sucrose, then concentrated to a final volume of 50 µl using a 500 µl Amicon[®] Ultra centrifugal concentrator (Merck Millipore) 10 KDa cut-off.

Carbon Films on 400 Mesh Copper Grids (agar scientific®) were glow discharged for 30 seconds at 20 milliAmps. 10 µl of Sin3a complex at 0.05 mg/ml was applied to the grid. The excess liquid was blotted away with filter paper after incubating for a minute. The sample was stained with 2% filtered uranyl acetate. The grid was then left to dry on filter paper before visualizing. With the help of Dr Nobuhiro Morone and Almutasem Saleh, data was collected with Falcon™ II Direct Detector and the micrographs were analysed with EMAN2 software.

Chapter 3. Expression and purification of the Sin3a complex

3.1. Introduction

As discussed in paragraph 1.4.5, the Sin3a complex plays an important role in normal development as histone deacetylase complex regulating gene expression by altering the chromatin structure (Das et al. 2013). Sin3a protein is a large multi-domain protein that acts as a scaffold for interacting with numerous transcription factors and co-factors, including HDAC1, SDS3, SAP proteins and RBBP4/7. As important regulator of the transcription network, the Sin3a complex can be involved in a variety of tumours when misregulated (Farias et al. 2010; McDonel et al. 2012). While the assembly and the regulatory mechanism of two of the Class I HDAC complexes, the SMRT : HDAC3 and MTA1 : HDAC1 complexes, has been studied (Watson et al. 2012; Millard et al. 2013; Watson et al. 2016), little is known about how the Sin3a protein recruits HDAC1 or how the HDAC activity of the complex is regulated.

Previous studies from our laboratory revealed that the Sin3a complex differs from the other Class I HDAC complexes both with its structure and its regulatory mechanism (Portolano et al. 2014). Since Sin3a remains highly conserved throughout the evolution (Alland et al. 1997; Dang et al. 1999; Q. Yang et al. 2000; Silverstein & Ekwall 2005; Sharma et al. 2008), understanding the evolution of the Class I HDAC complexes will be particularly interesting.

In order to study the structure and function regulation of the Sin3a complex, a wide range of protein expression and purification methods were performed. Since the Sin3a protein interacts with several different proteins, different combinations of complex components were tested in order to produce a stable and well-behaved protein complex.

3.2. Bioinformatics: protein constructs designing

To express a well-behaved Sin3a complex suitable for structural and functional studies protein sequence conservation, published interaction domains, predicted secondary structure and disorder prediction parameters were taken in consideration to design the protein constructs.

3.2.1. The Sin3a protein constructs

The Sin3a protein contains six very well conserved domains: 4 PAH (Paired Amphipathic Helix) domains, 1 HID (HDAC Interaction Domain) and 1 HCR (Highly Conserved Region).

The HID domain provides the interaction region for HDAC1/2. In 1997 Laherty indicated residues 524-821 as the region implicated in this interaction (Laherty et al. 1997). Subsequently, Portolano in 2014 narrowed down this region to residues 531-772 (Portolano et al. 2014). More recently, the structure of the Sin3a in complex with SDS3 was solved and it was shown to require residues 601-742 of the HID domain (Clark, Marcum, et al. 2015).

At the N-terminus of the HID domain is located the PAH3 domain (residues 459-526), which acts as an interaction platform for the majority of the Sin3a Associated Proteins (SAPs) interacting with Sin3a (Y. Zhang et al. 1997; Laherty et al. 1998; Le Guezennec et al. 2006; Viiri et al. 2006).

Since the goal of this study was to identify the minimal components between the large numbers of proteins able to interact with Sin3a required to form a stable complex, the SAPs were tested to help stabilising the complex. Therefore, a variety of constructs lengths for Sin3a were designed to include both the HID and the PAH3 domains to co-express Sin3a with or without the SAPs.

At the C-terminus of the HID domain is located the Highly Conserved Region (HCR) but since there was no reason to believe that the SAPs interact with the HCR, it was not included in the Sin3a constructs.

For the Sin3a constructs designing, MultiAlign and Clustal software were used to create sequence alignments and to study the protein conservation (figure 3.1, appendix 1 and 2) (Corpet 1988; Z. R. Yang et al. 2005), Jpred (secondary structure prediction software) was used for secondary structure prediction (Cuff et al. 1998) (Figure 3.2, appendix 3).

Figure 1 shows that the region including the PAH3 and the HID domains is very well conserved between different species. Jpred software suggests a four helical structure in the PAH3 domain and this was confirmed by Xie *et al.* who solved the left handed four helix bundle structure of PAH3 in interaction with SAP30 (Xie et al. 2011).

The HID domain is predicted to have an unstructured N-terminal region (approximately 80 amino acids long), while the central region is predicted to form 6 alpha helices and the C terminal region is predicted to have 2 alpha helices and to terminate with an unstructured and poorly conserved region. The four helices predicted within residues 605 and 729 were confirmed by the NMR structure solved by Clark *et al.* in 2015. Unfortunately the whole structure of the HID domain has not been solved yet.

In figure 3.2, the disorder prediction calculated using RONN software confirms the predicted secondary structures to be corresponding to the regions with low entropy.

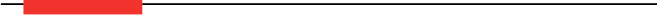
S.cerevisiae	-TEPIENNISLNEEVTFFKAKRYIGNKHLYTEFLKILNLYSQDILDLDDLVEKVDFYLG	711
D.melanogaster	VSFSEASSKCTISDAAFFDKVRKALRSPEVDNFLRCLTLFNQEIIVSKTELLGLVSPFLM	967
D.rerio	SSGAEGKHGGGTESLFFKVRKVLRSAAEYDNFLRCLVIFNEEIIISRAELVQLVVPFLG	512
H.sapiens	SSMADASKHGGGTESLFFDKVRKALRSAAEYENFLRCLVIFNQEVISRAELVQLVSPFLG	509
M.musculus	SSMADASKHGVGTESLFFDKVRKALRSAAEYENFLRCLVIFNQEVISRAELVQLVSPFLG	510
		
S.cerevisiae	SNKELFTWFKNFVGYQEKTKCIENI-----VHEKHR	742
D.melanogaster	KFPDLLRNFTDFLPPSGQPAGGLIDGMPLAATQRQGGSSNSSHGRGTSHQSAAEYVQD	1027
D.rerio	KFPPELFTWFKNFLGYKEMSHLET---YP-----KERATEGIAME	548
H.sapiens	KFPPELFNWFKNFLGYKESVHLET---YP-----KERATEGIAME	545
M.musculus	KFPPELFNWFKNFLGYKESVHLES---FP-----KERATEGIAME	546
		
S.cerevisiae	LDLDCFAFGPSYKRLPKSDTFMPCSGRDDMCWEVLNDEWVGHPVWASEDSGFIARHKNQ	802
D.melanogaster	VDLSCKRLGASYCALPOSTVPKKCSGRALCREVLNDKNVSFETWASEDSTFVTSRKQTQ	1087
D.rerio	IDYASCKRLGSSYRALPKSYQLPKCTGRTAICKEVLNDTNVSFSPWS-EDSTFVSSKKTQ	607
H.sapiens	IDYASCKRLGSSYRALPKSYQQPKCTGRTPLCKEVLNDTNVSFSPWS-EDSTFVSSKKTQ	604
M.musculus	IDYASCKRLGSSYRALPKSYQQPKCTGRTPLCKEVLNDTNVSFSPWS-EDSTFVSSKKTQ	605
		
S.cerevisiae	YEETLFKIEERHEYDFYIESNLRTIQCLETIVNKIENMTENEKANFKLPPGLHTSMTI	862
D.melanogaster	FEETIYRTEDERFELDLVIEVNSATIRVLENLQKMSRMSTEELSKFHLDDHGGTSQTI	1147
D.rerio	YEEHMYRCEDERFELDVVLETNLATIRVLETTQKKLSRMSAEEQAKFRLDNTLGGFSEII	667
H.sapiens	YEEHIYRCEDERFELDVVLETNLATIRVLEAIQKKLSRLSAEEQAKFRLDNTLGGTSEVI	664
M.musculus	YEEHIYRCEDERFELDVVLETNLATIRVLEAIQKKLSRLSAEEQAKFRLDNTLGGTSEVI	665
		
S.cerevisiae	YKKVIRKVDKERGFETIDALHEHPAVTAPVVLKRLKQKDEEWRAQREWNKVVRELEQK	922
D.melanogaster	HQRAIHRIVGDKS-GEIITGMKKNPFAVPIVLKRLKVKEEWEWAQKTFNKQWRQNEK	1206
D.rerio	HRKAIQRIYADKA-PDIIDGLKKNPAASVPIVLKRLKLKEEWEWAQRFNKKVWRQNEK	726
H.sapiens	HRKALQRIYADKA-ADIIDGLRKNPSIAVPIVLKRLKMKEEWEWAQRFNKKVWRQNEK	723
M.musculus	HRKALQRIYADKA-ADIIDGLRKNPSIAVPIVLKRLKMKEEWEWAQRFNKKVWRQNEK	724
		
S.cerevisiae	VFFKSLDHLGLTFKQADKLLITTKQLISEISSIKVDQTNKKIHW--LTPPKPSQLDFDFP	980
D.melanogaster	YYLKSLDHQAINFKNPNDMKALRSKSLFNEIETLYDERHQED---DAMEPFGPHLVLPHY	1263
D.rerio	YYLKSLDHQGINFKQNDTKVLRSLTLNEIESIFDERQEQVSEDNSSTASSPHLTLYE	786
H.sapiens	YYLKSLDHQGINFKQNDTKVLRSLTLNEIESIYDERQEQATEENAG-VPVGPHLSLAYE	782
M.musculus	YYLKSLDHQGINFKQNDTKVLRSLTLNEIESIYDERQEQATEENAG-VPVGPHLSLAYE	783
		
S.cerevisiae	DKNIFYDILCLADTFITHTTAYSNPDKERLKDLLKYFISLFFSISFEKIEES-----LY	1034
D.melanogaster	DKTILDDAANLLIHHVKRQTGIQKQKQIKQIIRQFVPDLFFAPRQPLSDDERDDAFFF	1323
D.rerio	DHQILEDAAAALIIHHVKRQSSINKEDKYKIQIYVHFIPDMLFAQRGVLSDVEEEDDEE--	844
H.sapiens	DQKILEDAAAALIIHHVKRQTGIQKEDKYKIQIMHFIIPDLLFAQRGDLSDVEEEDDEE--	841
M.musculus	DQKILEDAAAALIIHHVKRQTGIQKEDKYKIQIMHFIIPDLLFAQRGDLSDVEEEDDEE--	842
		
S.cerevisiae	SHK----QNVSESSGSDDGSSIASRKRPYQ-----QEMSL-----DILHRS-----	1072
D.melanogaster	LVDDNTKMDVDSPLGRTESSRNKSTPSESASPARSNASTSSVTPAGIKKETDDSKATT	1383
D.rerio	-----EDMELDEGASKKH-----	857
H.sapiens	-----MDVDEATGAVKKH-----	854
M.musculus	-----MDVDEATGAPKKH-----	855
		

Figure 3.1: Sequence alignment of the PAH3-HID domains of the Sin3a protein from different model organisms.

Clustal software was used to align Sin3a₄₅₀₋₈₅₁ from human and other four model organisms. The sequence includes both the PAH3 domain (residues 459-526) and HID domain (residues 524-851). Residues highlighted in green are positions that have a single and fully conserved residue, in yellow are indicated the residues with strongly similar properties and in grey are indicated residues with weakly similar properties. Underneath the sequence, a schematic representation of the secondary structure prediction calculated with Jpred software is shown. Red boxes indicate alpha helical structures and blue arrows indicate beta strands.

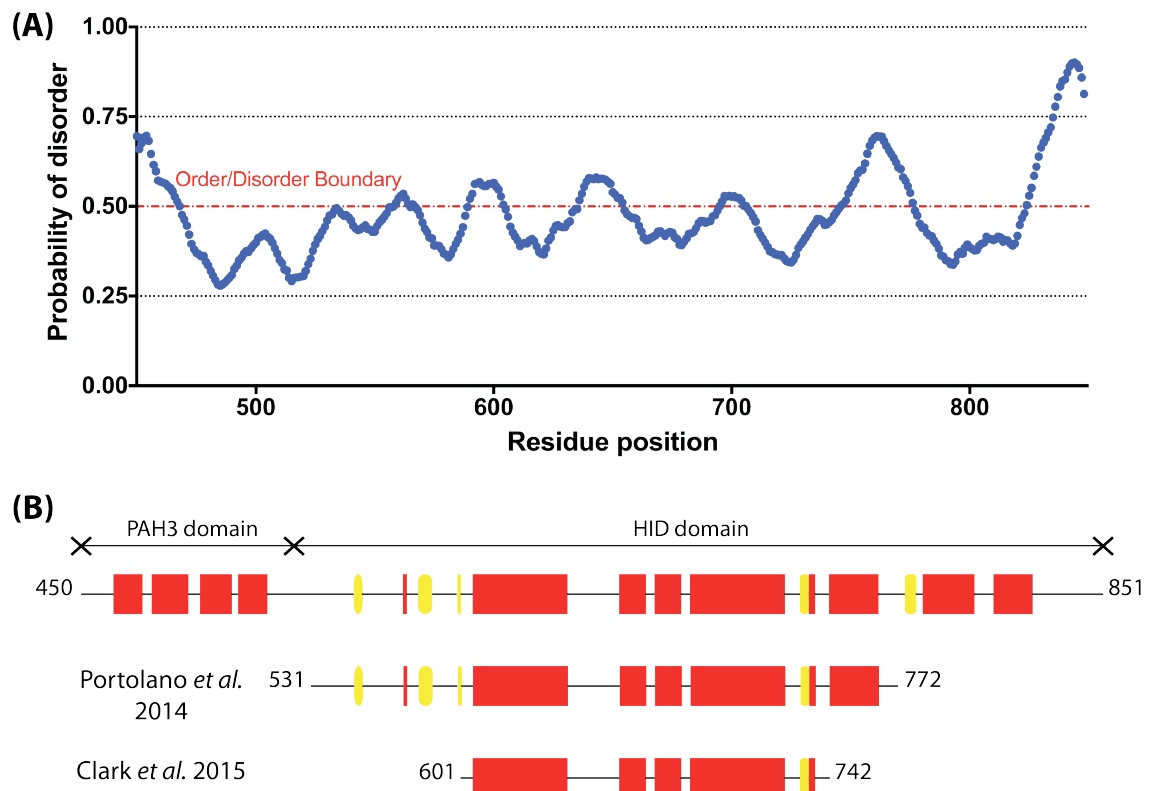


Figure 3.2: **Disorder and secondary structure prediction analysis for Sin3a₄₅₀₋₈₅₁.**

(A) Disorder prediction analysis performed with RONN software. The blue line represents the predicted disorder probability. The red line represents the order/disorder boundary, values above 0.5 are considered to be disordered. (B) Secondary structure prediction of Sin3a₄₅₀₋₈₅₁ calculated with Jpred software. The black line shows the disordered region, red boxes represent helix regions and yellow boxes represents strands. The secondary structure prediction is also shown for the constructs that were published in previous studies are also shown.

The Sin3a constructs were designed by taking into account previously published studies and secondary structure prediction. Construct boundaries were located in highly unstructured regions or within regions of low conservation (all the constructs details for Sin3a are listed in appendix 8 and 9).

Sequence conservation, secondary structure and disorder prediction was also calculated for HDAC1, SDS3 and the SAPs involved in the expression and purification trials.

3.2.2. The HDAC1 protein constructs

HDAC1 is very well conserved between different species (appendix 4). Only the C-terminal region is less well conserved compared to the whole sequence. The C-terminal region is also predicted to be highly unstructured by both RONN and Jpred software (figure 3.3). Previous studies have shown that the C-terminal flexible tail of HDAC1 can carry modifications (phosphorylation and acetylation) that regulate the HDAC activity (Qiu et al. 2006; Luo et al. 2009).

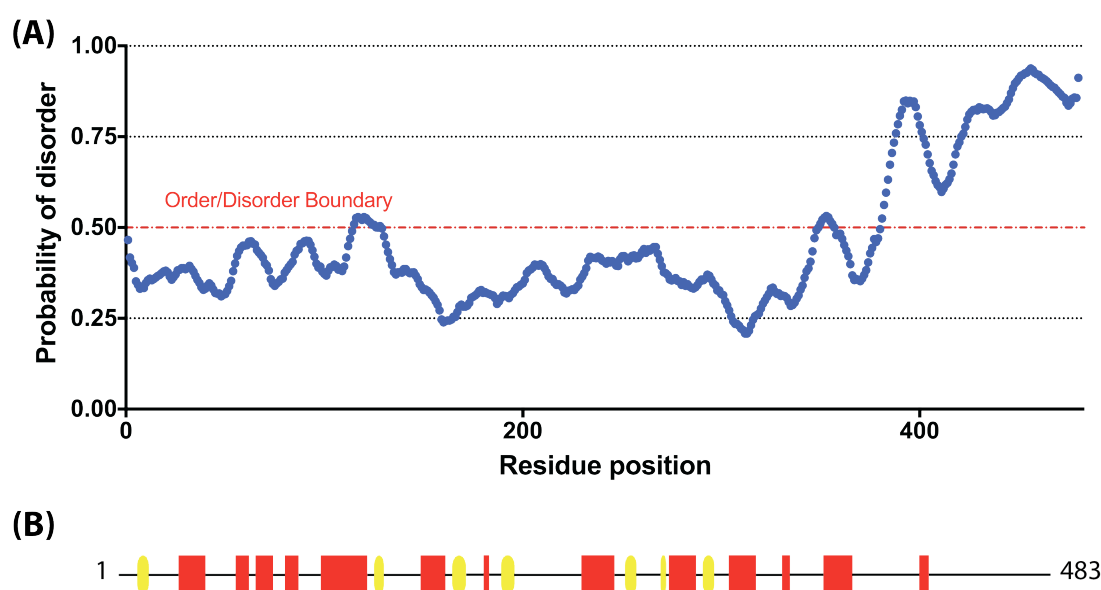


Figure 3.3: **Disorder and secondary structure prediction analysis for HDAC1_{FL}.**

(A) Disorder prediction analysis performed with RONN software. The blue line represents the predicted disorder probability. The red line represents the order/disorder boundary, values above 0.5 are considered to be disordered. (B) Secondary structure prediction of HDAC1_{FL} calculated with Jpred software. The black line shows the disordered region, red boxes represent helix regions and yellow boxes represents strands.

For this study, HDAC1 was always expressed as a full length protein (all tags and expression systems used to express HDAC1 are listed in appendix 8 and 9).

3.2.3. The SDS3 protein constructs

When comparing different species, it is clear that SDS3 was very well conserved N- and C-terminal regions with a lower conservation within the central region of the protein (appendix 5). Jpred software was used to predict secondary structures in SDS3 (figure 3.4A), while RONN software was used to study its disorder prediction (Figure 3.4B).

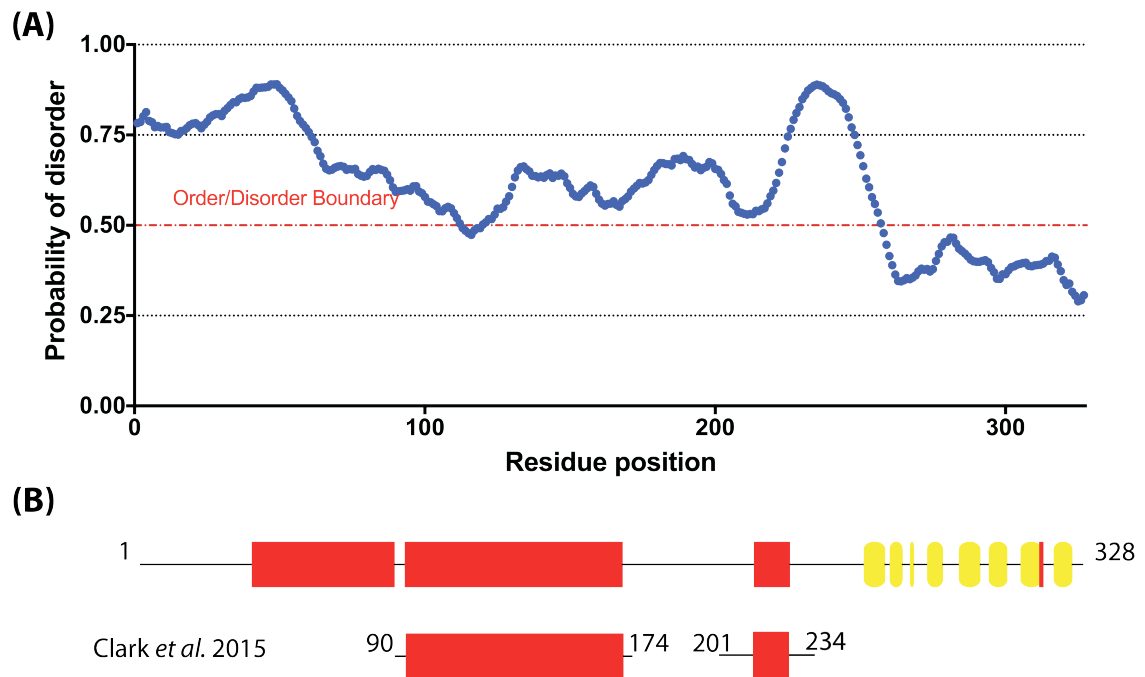


Figure 3.4: **Disorder and secondary structure prediction analysis for SDS3_{FL}.**

(A) Disorder prediction analysis performed with RONN software. The blue line represents the predicted disorder probability. The red line represents the order/disorder boundary, values above 0.5 are considered to be disordered. (B) Secondary structure prediction of SDS3FL calculated with Jpred software. The black line shows the disordered region, red boxes represent helix regions and yellow boxes represents strands. The secondary structure prediction is also shown for the constructs that were published by Clark *et al.*

Previously, the internal domain organization of SDS3 was studied by Alland *et al.* and it was shown that SDS3 contains two coiled-coil (CC) domains (CC1 and CC2) in its N-terminal region, a central structured region responsible for the interaction with Sin3a and a C-terminal secondary-structure-rich domain that appears to have a globular fold.

The secondary structure prediction performed with Jpred is partially confirmed by both the structures of the Sin3a Interaction Domain (SID) and the CC2 domain solved by Clark *et al.* (Clark, Marcum, et al. 2015).

The full structure of SDS3 is still unknown. Alland *et al.* previously located the region involved in the interaction with Sin3a within residues 188 and 229. Only recently Clark *et al.* confirmed that a small alpha helix located within residues 201-234 is sufficient for SDS3 to interact with Sin3a (PDB code: 2NHN) (Alland et al. 2002; Clark, Marcum, et al. 2015). Furthermore, the same laboratory solved the crystal structure of the CC2 domain, located within residues 90-174 (PDB code: 4ZQA). Through its CC domains SDS3 can form homodimers and it is possible that the homodimerization of SDS3 could enable the Sin3a complex to dimerise.

To obtain better insight into the structural role of SDS3 within the complex, different constructs were designed. The details of boundaries, affinity tags and expression systems for the SDS3 constructs are listed in appendix 8 and 9.

3.2.4. The Sin3a Associated Proteins (SAPs) protein constructs

Sequence conservation, sequence alignments and secondary structure prediction were also taken in consideration when designing constructs for the SAPs proteins.

The alignment of the full sequences between SAP18, SAP30, SAP30L and SAP130 is shown in appendix 6. The sequence alignment shows that no conservation is found between the different SAPs proteins. Since for most of the SAPs the residues involved in the interaction with Sin3a are published (Fleischer et al. 2003; Viiri et al. 2006; Xie et al. 2011), it was possible to map those residues onto the conserved regions. As shown in appendix 10, the Sin3a Interaction Domains (SID) of the SAPs may have some common features between these proteins. Therefore, the SID within the SAPs was always included when the SAPs constructs were designed. Disorder prediction for the SAPs proteins was calculated with RONN software (appendix 7). Details for SAPs constructs are listed in appendix 8 and 9.

3.3. Protein expression

Heterologous protein production is the expression of recombinant proteins in cell hosts where they would not naturally be expressed. Between the large variety of expression systems available, four systems are the most commonly used: bacteria, yeast, insect and mammalian (Gecchele et al. 2015; Rosano & Ceccarelli 2014; Verma et al. 1998; Jarvis 2009; Hartley 2012). Between these four, *E. coli* remains the most frequently used to produce recombinant proteins due to its high productivity, easy manipulation and the availability of a large number of cloning vectors and host strains. Unfortunately, despite these advantages, bacterial expression can produce insoluble or non-functional protein due to its inability to carry out Post Translational Modifications (PTMs) and the production of misfolded or aggregated forms.

The specific character of the protein that has to be expressed dictates the choice of the expression system that needs to be used to produce a soluble and active protein. Features like protein structure, activity, PTMs and folding requirements are essentials to avoid altering the nature of the target protein. Many mammalian proteins need to be modified in order to be properly folded and to be active.

Most laboratories use Human Embryonic Kidney (HEK) or Chinese Hamster Ovaries (CHO) cells when expressing protein that need PTMs. Mammalian cells are more complex and they can carry out PTMs, assist protein folding with chaperones and accomplish all the natural modification found in mammalian cells. In our laboratory a cheaper and simple method was developed in the last few years where HEK 293F cells are transfected with the inexpensive polymeric reagent polyethylenimine (PEI) to produce milligram quantities of protein (Portolano et al. 2014; Nigi et al. 2017).

Since previous published studies successfully expressed Sin3a in *E. coli*, in this chapter both bacterial and mammalian expressions are tested to express the Sin3a complex.

3.3.1. Investigating the effect of different affinity tags on Sin3a

In 2015, bacterial expression was used to successfully express the Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄ complex and to solve the structure of the complex by NMR (Clark, Marcum, et al. 2015). Accordingly, preliminary experiments were performed to express Sin3a in *E. coli* with various N- and C- terminus boundaries and with an affinity tag followed by a TEV (Tobacco Etch Virus) cleavage site. For each protein tested, details of boundaries, tag and expression system used are shown in appendix 8 and 9.

The use of affinity tags is a common method to identify and simplify protein isolation. Different tag systems not only require different types of resin to purify the protein, but they also affect the purity and the protein yield achieved by preventing proteolysis of the protein and increasing its expression and solubility. When a protein is expressed without tag, it was co-expressed and co-purified with a tagged interacting protein. Instead, when a protein was designed to have an affinity tag, preliminary attempts to express and purify the tagged protein with different affinity tag were performed before performing the interacting experiments.

Figure 3.5 shows a comparison of the expression of Sin3a₄₅₀₋₇₇₂ using three different tags: GST, His₆-S and GB₁-His₆. The GST tagged protein was purified with Glutathione sepharose 4B (GE healthcare) resin and the His₆-S and GB₁-His₆ tagged protein were purified with Ni-NTA (Qiagen) resin. Buffers and purification methods are described in materials and methods. The pre-TEV cleavage samples were analysed on SDS-PAGE in order to assess the purity and the expression efficiency for each tagged Sin3a₄₅₀₋₇₇₂.

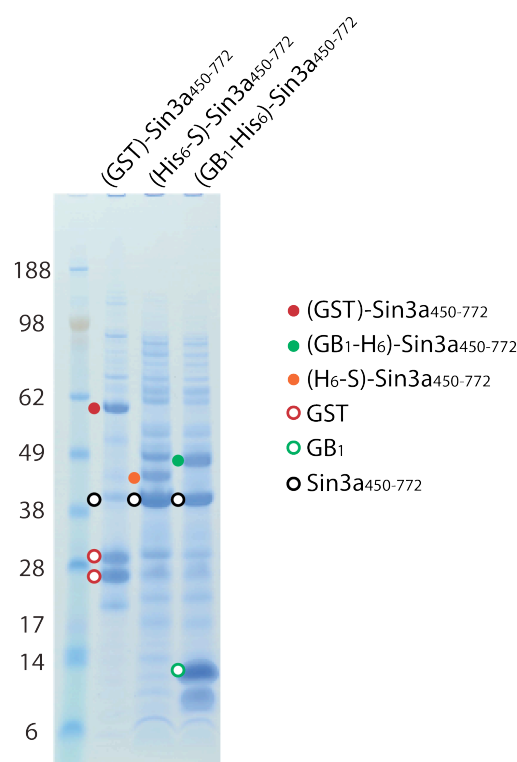


Figure 3.5. **Preliminary attempt for the expression and purification of Sin3a₄₅₀₋₇₇₂ using different affinity tags.**

Sin3a₄₅₀₋₇₇₂ was expressed and purified from *E. coli* using three different tags: GST-, H₆-S and GB₁-H₆. Samples from pre-TEV cleavage were analysed on SDS-PAGE. Bands of interest are indicated with coloured dots.

Figure 3.5 shows that Sin3a₄₅₀₋₇₇₂ was successfully expressed with GST, His₆-S and GB₁-His₆ tag. In each case endogenous proteases cleave hinge regions of the constructs and produce degradation products containing either GST or GB₁ tag and Sin3a₄₅₀₋₇₇₂ freed by the endogenous cut. Unlabelled bands are contaminants. By comparing the expression of the three different tagged Sin3a₄₅₀₋₇₇₂, the (GST)-Sin3a₄₅₀₋₇₇₂ appears to have higher purity (showing fewer contamination bands and lower background) than when expressed with His₆-S or GB₁-His₆ tag. This result could be the result of better protein expression given by the presence of the GST tag or to a higher specificity of the GST resin compared to the Ni-NTA resin.

Given the trial protein purification results, the Sin3a constructs expressed with GST tag were selected for large-scale protein expression and complex formation analyses.

3.3.2. The Sin3a₆₀₁₋₇₄₂ : HDAC1 : SDS3₂₀₁₋₂₃₄ complex is successfully expressed in bacteria

As previously mentioned, bacterial expression systems have been widely used to express proteins for structural studies. Such expression systems have the advantage of being versatile, economical, straightforward and fast (Verma et al. 1998; Rosano & Ceccarelli 2014; Gecchele et al. 2015). Since the main goal of this study is to understand the structural organisation and regulation of the Sin3a complex, HDAC1 was added to the structurally characterised Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄ complex to analyse the ternary complex.

Given the successful expression of expressed the Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄ complex in *E. coli* from the structural studies of Clark *et al.*, the ternary complex (GST)-Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : (GB₁-H₆)-SDS3₂₀₁₋₂₃₄ was initially expressed in *E. coli* (figure 3.6A). The complex was purified with Glutathione sepharose 4B (GE healthcare) resin. Details for protein expression and purification with GST resin are specified in materials and methods.

The SDS-PAGE analysis of the protein purification (figure 3.6A) shows that this small ternary complex expresses well in bacteria. The protein band corresponding to the expected molecular weight for HDAC1 was sent for mass-spec analysis and the identity of HDAC1 was confirmed. Furthermore, the SDS-PAGE analysis shows that the construct Sin3a₆₀₁₋₇₄₂ is sufficient to interact with HDAC1. Indeed, since HDAC1 was expressed without an affinity tag and the purification of the complex was performed using GST resin, the presence of HDAC1 in the post-TEV sample must be due to interaction with Sin3a.

After cleaving the affinity tags by TEV cleavage, the complex was further purified with size exclusion chromatography with Superdex-200 (10/300) GL columns, GE Healthcare (Figure 3.6B). Gel filtration buffer is specified in materials and methods.

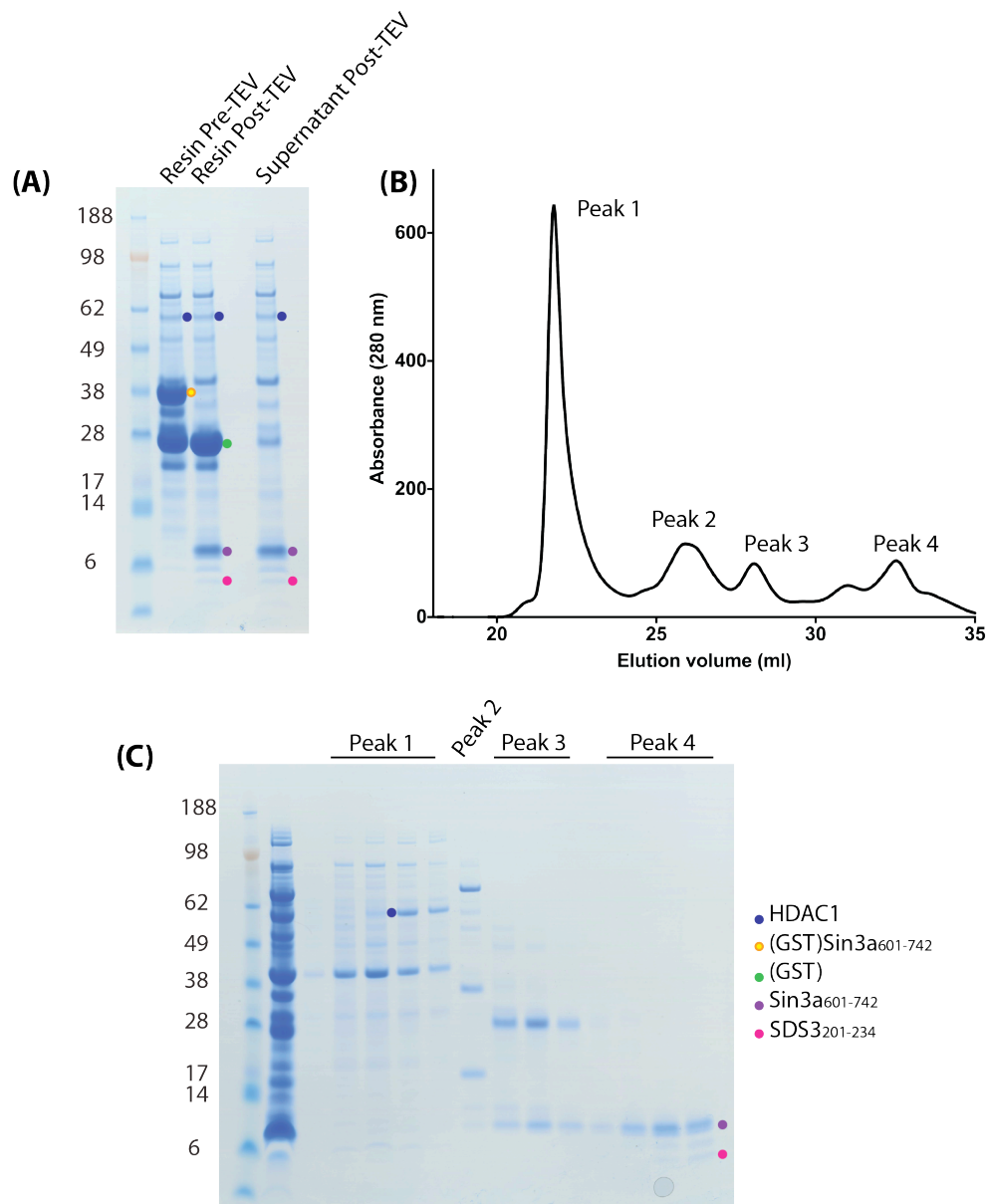


Figure 3.6. Bacterial expression of Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄.

(GST)-Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : (GB₁-H₆)-SDS3₂₀₁₋₂₃₄ was expressed and purified from *E. coli*. Bands of interest are indicated with coloured dots. (A) Samples from pre and post-TEV cleavage were analysed on SDS-PAGE. (B) Gel filtration purification of the TEV cleaved sample with a Superdex-200 (10/300) GL column. The absorbance was measured at 280 nm. (C) SDS-PAGE analysis of the fractions after gel filtration. Unfortunately the complex falls apart during the gel filtration step and HDAC1 is separated from Sin3a₆₀₁₋₇₄₂ and SDS3₂₀₁₋₂₃₄.

Figure 3.6C shows the SDS-PAGE analysis of the fractions corresponding to the gel filtration peaks. Unfortunately the complex is not stable during the gel filtration step and HDAC1 gets separated from the binary complex Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄. Before performing other purification trials with HDAC1 expressed in bacteria, the HDAC activity was tested. The histone deacetylase assay was performed as described in

materials and methods. The HDAC activity of the HDAC1 expressed in *E. coli* was compared to the HDAC activity of HDAC3 : DAD-SMRT and HDAC1 : MTA1, both complexes were expressed in HEK 293F cells. HDAC3 : DAD-SMRT was expressed and purified by Dr Watson and HDAC1 : MTA1 was expressed and purified by Dr Millard in our laboratory (Figure 3.8).

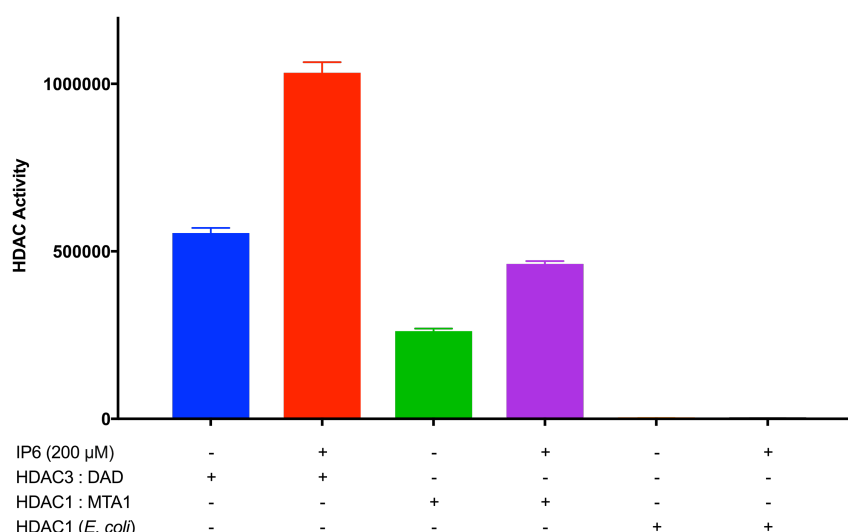


Figure 3.7. Investigating the activity of HDAC1 expressed in bacteria.

HDAC activity measured with or without $\text{Ins}(1,4,5,6)\text{P}_4$. The activity of HDAC1 expressed in *E. coli* was compared to the activity of HDAC1 and HDAC3 expressed in mammalian HEK 293F cells. No HDAC activity is shown from the bacterial version of HDAC1. Experiments were performed in triplicate. Error bars indicate $\pm$ SEM. (n = 3)

Figure 3.7 shows that HDAC1 is not active when expressed in bacteria. The biochemical environment inside different host cells has a major impact on the successful protein expression (Jarvis 2009). The availability of biochemical co-factors, folding machinery and enzymes able to introduce the necessary post-translational modifications (PTMs) plays a key role in the expression of a properly folded protein (Verma et al. 1998; Hutt et al. 2009; Mattanovich et al. 2012; Rosano & Ceccarelli 2014; Gecchele et al. 2015). The necessity of PTMs and special co-factor to produce an active form for HDAC1 determines the inability of bacterial hosts to be used to produce HDAC1 for biochemical studies.

Accordingly, for structural and functional studies, HDAC1 was subsequently expressed in mammalian HEK 293F.

3.3.3. The Sin3a₆₀₁₋₇₄₂ : HDAC1 : SDS3₂₀₁₋₂₃₄ complex was expressed by combining bacterial and mammalian expression systems.

As shown in paragraph 3.3.2, HDAC1 requires expression in mammalian cells in order to be active. Accordingly, to analyse the ternary complex Sin3a : HDAC1 : SDS3, (GST)-Sin3a₆₀₁₋₇₄₂ : (GB₁-His₆)-SDS3₂₀₁₋₂₃₄ were co-expressed in *E. coli* (0.75 L) and (Flag)-HDAC1_{FL} was expressed in HEK 293F cells (0.60 L). Bacterial and mammalian pellets were re-suspended together in GST lysis buffer and after the sonication step the mixture was incubated for one hour at 4°C. The protein complex was purified with Glutathione sepharose 4B (GE healthcare) resin as described in materials and methods. After washing the resin from unbound materials, the sample was analysed with SDS-PAGE.

The ternary complex was successfully reconstituted by combining bacterial and mammalian expression. Figure 3.8A shows (Flag)-HDAC1 successfully expressed in HEK 293F and interacting with the (GST)-Sin3a₆₀₁₋₇₄₂ : (GB₁-His₆)-SDS3₂₀₁₋₂₃₄ complex (unlabelled bands are contaminants). Indeed, after the TEV-cleavage of the resin sample, the protein band corresponding to (Flag)-HDAC1 (60 kDa) shows a band shift to 55 kDa, which is the expected molecular weight for the cleaved HDAC1.

However, from the density of the bands corresponding to the complex components, it appears that the stoichiometry of the complex is not 1:1:1 as it would be expected but HDAC1 is sub-stoichiometric compared to Sin3a and SDS3. The experiment was repeated with 0.75 L of *E. coli* expressing (GST)-Sin3a₆₀₁₋₇₄₂ : (GB₁-His₆)-SDS3₂₀₁₋₂₃₄ and double amount of mammalian pellet expressing HDAC1 (1.2 L of HEK 293F expressing (Flag)-HDAC1). Unfortunately, the ratio of the complex components did not change upon the increased volume of cells used for the protein complex purification.

Furthermore, during the purification step performed by gel filtration, HDAC1 lose its interaction with Sin3a and SDS3 and the ternary complex falls apart (figure 3.8C).

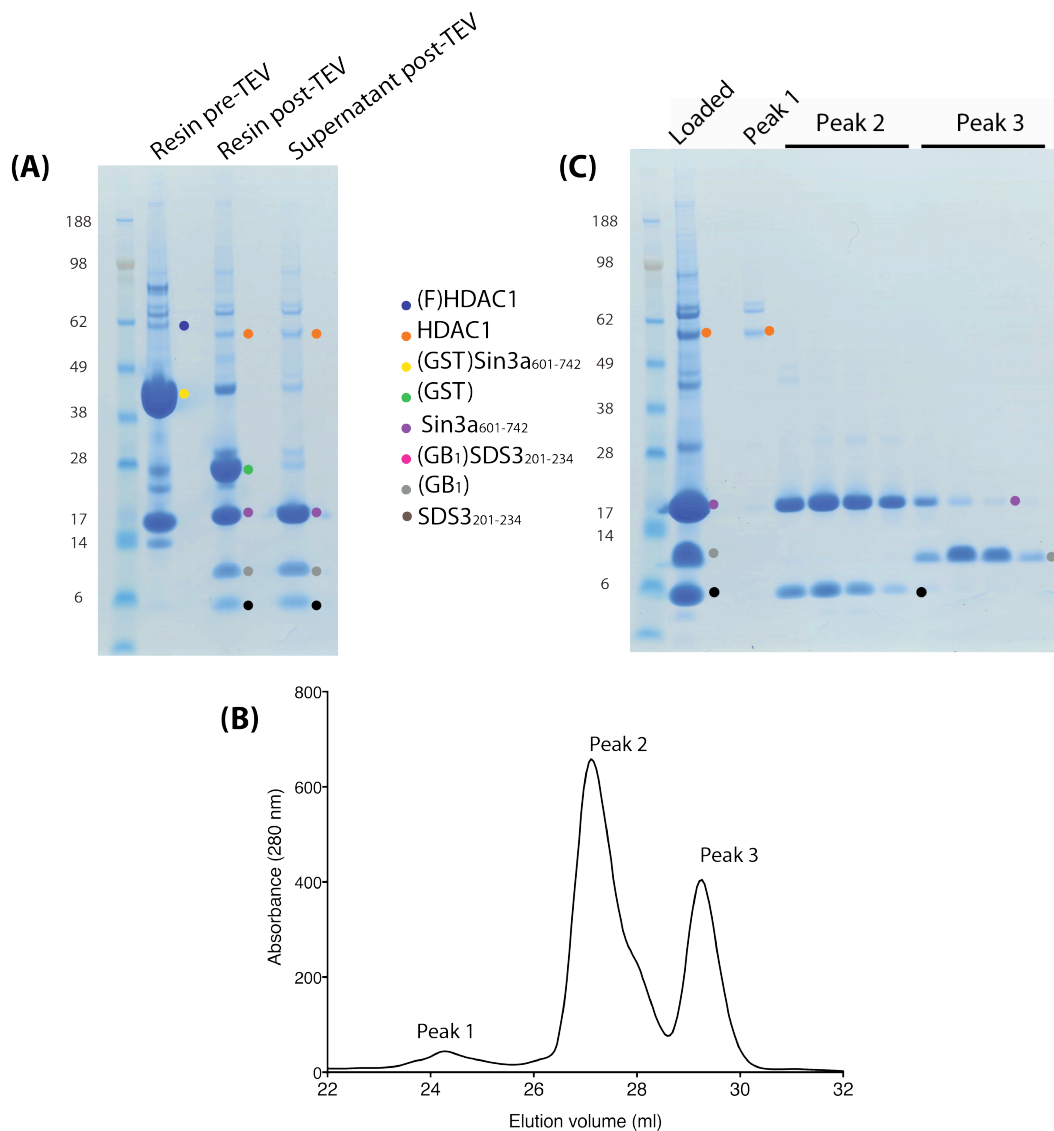


Figure 3.8. **Combination of bacterial and mammalian expression of Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄.**

(GST)-Sin3a₆₀₁₋₇₄₂ : (GB₁-His₆)-SDS3₂₀₁₋₂₃₄ was expressed in *E. coli*. (Flag)-HDAC1_{FL} was expressed in HEK 293F cells. The cell pellets were combined during the cell lyse step and the complex was purified with Glutathione sepharose 4B resin. Bands of interest are indicated with dots or lines. (A) Samples from pre and post-TEV cleavage were analysed on SDS-PAGE. (B) The purified complex was further purified with gel filtration with a Superdex-200 (10/300) GL column. The absorbance was measured at 280 nm. (C) Fractions from each peak were analysed on SDS-PAGE.

The dissociation of the complex through multiple rounds of purification could be due to the complex components concentration, if this is below the dissociation constant it will not be able to hold a tight interaction between the components of the complex. Another possibility would be that the short length of the Sin3a and SDS3 constructs could cause the low HDAC1 yield and/or the instability of the complex. The short

construct Sin3a₆₀₁₋₇₄₂ might be necessary to interact with HDAC1 but not sufficient to hold a tight interaction.

Additionally, it is not known how SDS3 interacts in the Sin3a : HDAC1 complex. It is possible that SDS3_{FL} interacts with Sin3a with a larger surface compared to the structure solved by Clark (Clark, Marcum, et al. 2015). An extended interaction surface with multiple points of interaction could help to further stabilise Sin3a and consequently its interaction with HDAC1. SDS3 itself could also interact directly with HDAC1 and help to stabilise the overall complex by interacting with both Sin3a and HDAC1. If at least one of these hypotheses is verified, the short SDS3₂₀₁₋₂₃₄ construct might fail to strengthen the interaction between Sin3a and HDAC1, leading to a weaker recruitment of HDAC1 and the failure to hold together the ternary complex through the gel filtration step.

Since the small ternary complex disassociated during gel filtration, it was not possible to take this complex forward for structural studies. Therefore, a variety of components and constructs were tested in order to find a better-behaved complex suitable for structural and biochemical studies.

3.3.4. Different combinations of complex components were tested to find a stoichiometric Sin3a complex

Proteins often undergo structural rearrangements when interacting with protein partners. Sin3a interacts with multiple proteins and its structure is likely to change upon these interactions. The overall stability of the complex is likely to be affected by the presence or absence of certain interactions. Therefore, different constructs for Sin3a and several other complex components were tested to find a well-behaved and stoichiometric Sin3a complex suitable for structural studies.

Previous studies in our laboratory showed that Sin3a₅₃₁₋₇₇₂ is sufficient to interact with HDAC1 and that the addition of SDS3_{FL} to the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} complex stabilises the complex and leads to higher protein yield (Portolano et al. 2014). The HID domain was always included in the Sin3a constructs and different boundaries were tested.

Additionally, some Sin3a constructs were designed to include both the HID and the adjacent N-terminal PAH3 domain. The PAH3 domain includes residues 459-526 and allows Sin3a to interact with the Sin3 Associated Proteins (SAPs). The following SAPs were identified as integral components of the Sin3a complex (Fleischer et al. 2003). The following SAPs were tested for interaction: SDS3 (previously known as SAP45), SAP18, SAP30, SAP30L and SAP130 were tested. Suppressor of Defective Silencing 3 (SDS3) plays an essential role in the structure and function of the Sin3a complex, its importance is further described in paragraph 1.4.5. SAP30 is widely conserved from yeast to mammals as a component of the Sin3/HDAC complex. It is thought to interact through its C-terminus region to the PAH3 domain of Sin3a and to act as a bridge stabilizing the complex together with other subunits (Lai et al. 2001). SAP30L is a homolog of SAP30 and they both share a functional nucleolar localization signal (NoLS), which targets Sin3 to the nucleolus (Y. Zhou et al. 2002). SAP18 interacts with both Sin3a and HDAC1 and it is thought to act as an adaptor between Sin3/HDAC complex and other proteins (Espinás et al. 2000; Wada et al. 2002). The C-terminus of SAP130 was shown to be able to interact with both mSin3a and HDAC1, from binding studies it was shown to be able to interact with different fragments of Sin3a, suggesting that more than one single binding site might be involved in this interaction (Fleischer et al. 2003). The details of the constructs designed for Sin3a, SAP30, SAP18, SAP30L and SAP130 are described in appendix 8 and 9.

Figure 10 shows a preliminary attempt to express the Sin3a complex using different components. In this experiment, bacterial and mammalian expression systems were combined as described in paragraph 1.3.3. (GST)-Sin3a and the (GB₁)-SAPs were expressed in *E. coli* and (Flag)-HDAC1 was expressed in HEK 293F cells. The cell pellets were combined in GST lysis buffer and after the sonication step the mixture was incubated for 1 hour at 4°C. The GST protein purification step was performed as described in materials and methods. The pre-TEV sample for each purification trial was analysed on SDS-PAGE (figure 3.9). Unlabelled bands are contaminants.

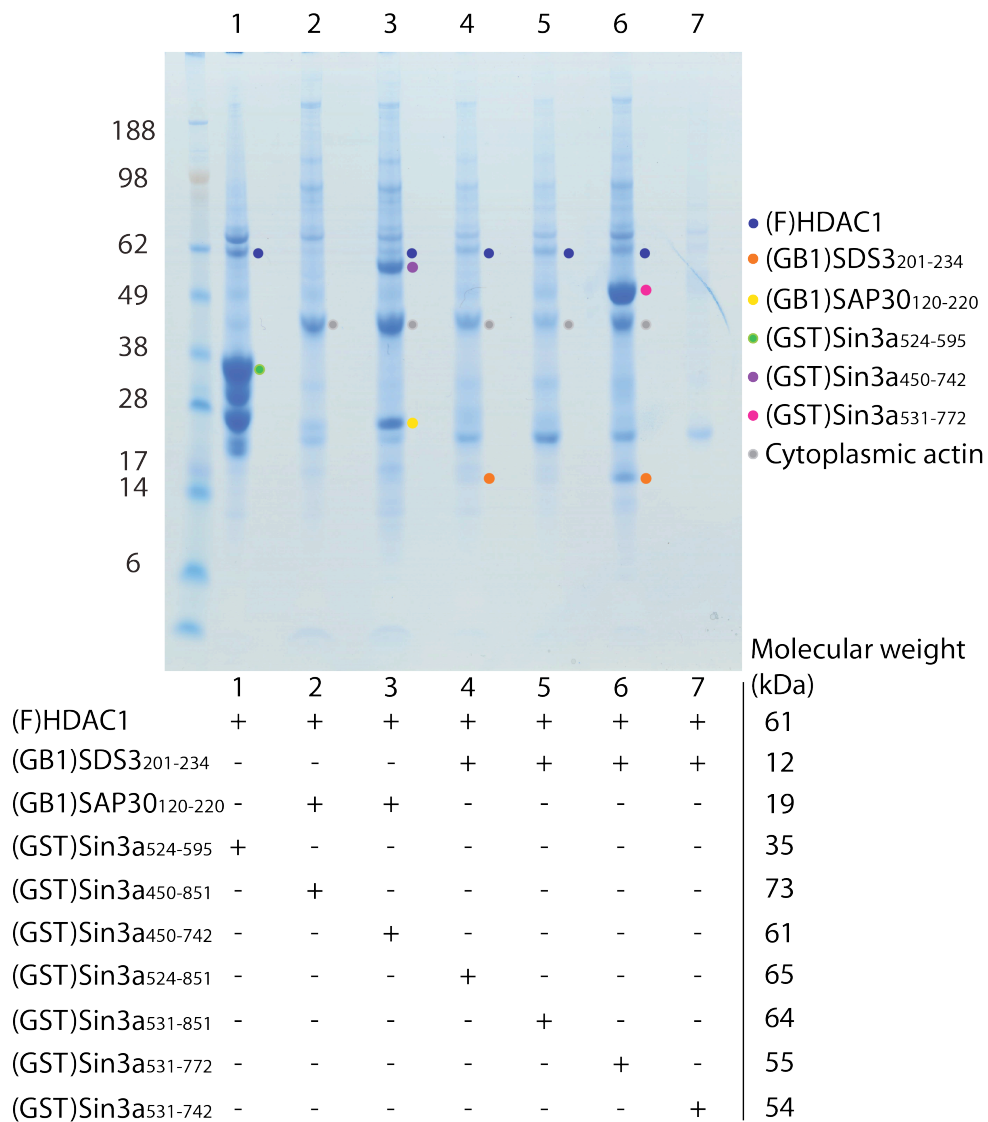


Figure 3.9. Reconstitution of the Sin3a complex by combining bacterial and mammalian expression- expression trials using different complex components and constructs.

Sin3a and SAPs were expressed from *E. coli* using three different tags: GST-, and GB₁-His₆. (Flag)-HDAC1 was expressed in HEK 293F cells. The mammalian and bacterial cell pellets were mixed together and the complex components were co-purified with Glutathione sepharose 4B resin. Samples from pre-TEV cleavage were analysed on SDS-PAGE. Bands of interest are indicated with coloured dots.

Unfortunately, none of the Sin3a complexes obtained by combining mammalian and bacterial expression showed a stoichiometry 1:1 between HDAC1 and the other complex components.

Since (paragraph 3.3.2) it was shown that in order to have a properly folded and active HDAC1 is necessary to express HDAC1 in mammalian cells, it was decided to fully express the Sin3a complex in HEK 293F mammalian cells.

3.3.5. Different combination of components were expressed in mammalian HEK 293F cells to find the most stable Sin3a complex

In the attempt to express a stoichiometric complex, a variety of components and constructs for the Sin3a components were co-transfected in HEK 293F cells. Cell transfections and FLAG® purifications were performed as described in materials and methods.

Figure 3.10 shows the comparison of two Sin3a constructs (Sin3a₅₃₁₋₇₇₂ and Sin3a₄₅₀₋₇₇₂) when co-expressed with SDS3, SAP18, SAP18₄₄₋₁₅₃, SAP30₆₄₋₂₂₀ and SAP30₁₂₀₋₂₂₀. Sin3a₅₃₁₋₇₇₂ contains the HID domain that was previously shown to be able to interact with both HDAC1 and SDS3 (Portolano et al. 2014). Sin3a₄₅₀₋₇₇₂ includes both the HID and the PAH3 domain, this construct should allow the interaction of Sin3a with HDAC1, SDS3 and the SAPs proteins.

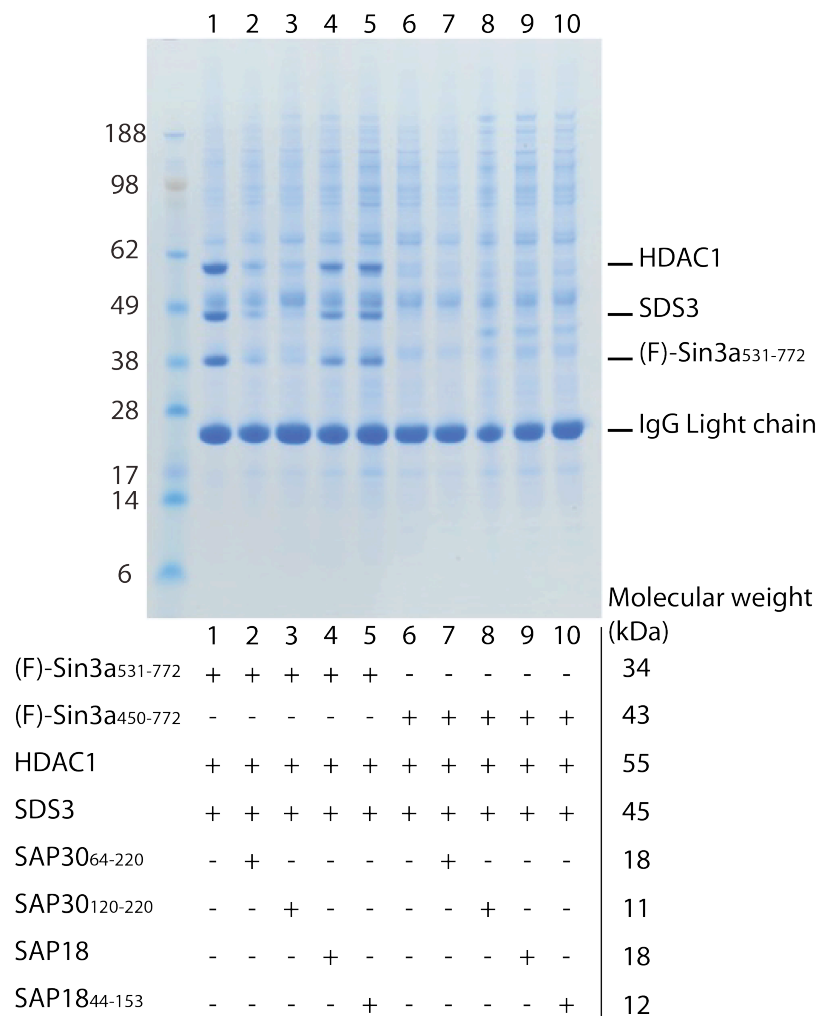


Figure 3.10. **Mammalian expression using different complex components and constructs.**

Sin3a, SDS3, HDAC1 and SAP18/30 were co-transfected in HEK 293F cells and the complex was purified with FLAG® resin. Samples from pre-TEV cleavage were analysed on SDS-PAGE. Bands of interest are indicated with lines.

From the SDS-PAGE analysis in figure 3.10, Sin3a₅₃₁₋₇₇₂ appears to be successfully expressed in HEK 293F cells and to create a well-behaved and stoichiometric complex with HDAC1_{FL} and SDS3_{FL}. The absence of the SAPs when co-expressed with Sin3a₅₃₁₋₇₇₂ was expected since the lack of PAH3 domain in the Sin3a construct would impair the interaction of the SAPs with Sin3a.

Sin3a₄₅₀₋₇₇₂ in contrast to Sin3a₅₃₁₋₇₇₂, did not express well and there is no evidence of co-purification of any of the SAPs. A variety of different constructs for the SAPs (listed in appendix 9) were used to improve the complex expression, but unfortunately with no success. Since the SAPs were expressed without an affinity tag it is not possible to

say whether they were simply failing to be expressed or expressed but not co-purified with the complex.

Since from the expression trials, the ternary complex (Flag)-Sin3a₅₃₁₋₇₇₃ : HDAC1 : SDS3 appeared to be the only stoichiometric complex, it was decided to take this complex up to a larger scale of expression and purification.

3.4. Protein purification methods

When a protein or a protein complex is expressed with an affinity tag, the use of the appropriate affinity resin is the classical approach to isolate the target protein from unbound material. However, resin purification alone is not usually sufficient to fully separate the selected protein complex from contaminants. Therefore, a second round of purification is usually needed to provide a good separation between the recombinant protein and higher- and lower-molecular weight contaminants. This additional step will also help to separate aggregated from physically stable forms of protein complexes.

The gel filtration method is one of the most commonly used methods to purify recombinant protein due to the large variety of resins and sizes of column available. Additionally, density gradient purification method has recently become more popular to purify and stabilize protein complexes for structural studies (Kastner et al. 2008; Stark 2010).

In order to purify the Sin3a complex both gel filtration and density gradient approaches were performed to find the most suitable method to produce a stable complex suitable for structural and biochemical studies.

3.4.1. Size exclusion chromatography for protein purification - Investigating the temperature conditions to improve protein complex purification

Size exclusion chromatography (SEC) purification method is one of the most common technique used to separate molecules based on their molecular weight. Nowadays several size exclusion columns are available. Each column is filled with molecular sieve (or resin) that comes in different sizes, the resin is uncharged, inert and it should not react with the studied material. The molecular sieve is composed of beads and represents the stationary phase. The mobile phase is instead liquid, and it occupies the space between the beads. The main working principle of SEC is to separate molecules by their size and shape. The beads in the matrix play a fundamental role in the separation thanks to their size and the presence of pores on their surface. The size of the beads goes from large to superfine, finer beads often enable higher resolution separation on the column.

The molecules of interest are suspended in the liquid phase, molecules larger than the pores will only be able to go through the space between the beads. Smaller molecules instead are more likely to diffuse through the pores and therefore it will require them longer time (and a larger elution volume) to cross the column compared to larger molecules. Different shapes will also affect the elution time for molecules with the same molecular weight. Indeed, elongated molecules are less likely to access the pores of the beads compared to spherical molecules (Andrews 1964; Janson & Rydén 1990).

In our laboratory, SEC is commonly used to purify HDAC containing complexes. Accordingly, a first attempt to purify the Sin3a complex after TEV elution was performed at room temperature on a Superose 6 (10/300) GE Healthcare as described in materials and methods (Figure 3.12A). Unfortunately, as shown in Figure 12A, despite the presence of the complex in the loaded sample, the complex tends to fall apart through the size exclusion column. This result could be caused by insufficient material loaded on the SEC column, if the protein complex is not sufficiently concentrated it might be below the dissociation constant (K_d) of the complex (Giddings 1968).

Furthermore, it was noticed that when purified through a sucrose gradient (as described in paragraph 3.4.2) the Sin3a complex did not fall apart. The entire sucrose gradient process is performed at 4 °C and it was noticed that the protein sample containing the Sin3a complex formed aggregates forms (resulting in precipitation forms) when left at room temperature. Therefore, it was decided to perform the SEC analysis loading higher protein yield (to ensure the complex to be above its K_d) and changing the experimental temperature to 4 °C (figure 3.11B).

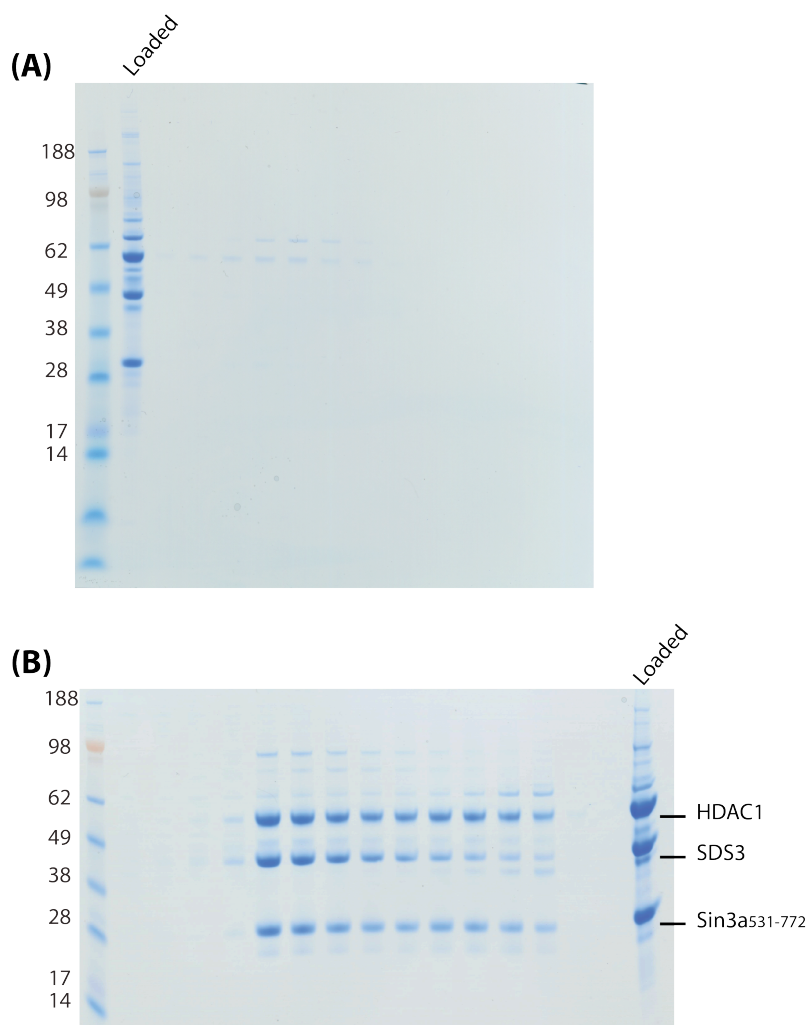


Figure 3.11. SEC purification of the Sin3a complex at different experimental temperatures. The Sin3a complex was purified with a Superose 6 (10/300) GE Healthcare column at (A) room temperature, and (B) at 4°C. Protein of interest are indicated with lines.

The SDS-PAGE analysis in figure 11B indicates that the Sin3a complex is stable at 4 °C through the Superose 6 (10/300) GE Healthcare column.

Thanks to these changes made on protein concentration and experimental temperature it was possible to consistently purify the Sin3a complex, allowing the complex to be studied with different biochemical and structural techniques.

3.4.2. The use of sucrose gradient as an alternative method to purify the Sin3a complex

Density gradient sedimentation is a commonly used method to separate protein complexes and molecules based on their sedimentation properties. This method allows to separate protein complexes from a mixture of components and to preserve the protein activity (R. G. Martin & Ames 1961; Tanese 1997; Stark 2010; Fernandez-Martinez et al. 2016). In our laboratory the rate zonal method is commonly used to prepare protein samples for electron microscopy studies.

A sample solution containing the concentrated protein complex is loaded on a pre-formed gradient of increasing density (as described in materials and methods). The range of densities varies upon the protein density that has to be selected. Ideally the highest density should not be higher than the densest component in the mixture. The particles within the mixture will be then separated upon centrifugation. Both the strength of the centrifugal field and the density of the gradient play a fundamental role in the resolution of the particles separation.

In parallel to the SEC purification (described in paragraph 3.4.1), the density gradient purification method was tested after the TEV cleavage step to separate the Sin3a complex from contaminants. The protein sample was loaded on a pre-formed gradient of 5-25% sucrose and centrifuged at 166,000 x g for 15 hours in a swing-out rotor (TH660, Thermofisher®) (as described in materials and methods). Figure 3.12B shows the SDS-PAGE analysis of the gel filtration fractions from the sucrose gradient.

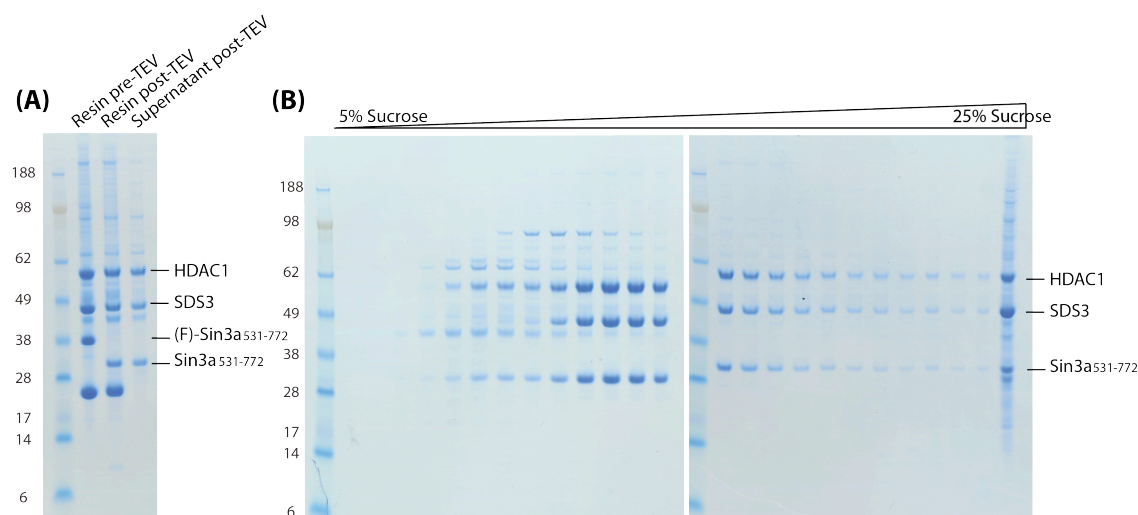


Figure 3.12. Sucrose gradient purification of the Sin3a complex.

After TEV cleavage (shown in A), the Sin3a complex was further purified through a 5-25% sucrose gradient. (B) SDS-PAGE analysis of the sucrose gradient fractions. Proteins of interest are indicated with lines, unlabelled bands correspond to contaminants.

In our laboratory, the sucrose gradient purification is usually performed as a further purification method after SEC purification to prepare the protein sample for electron microscopy analysis. In this case instead, the protein sample was directly loaded on the sucrose gradient after the TEV cleavage step (shown in figure 3.12A). As shown by the SDS-PAGE analysis in figure 3.12B the Sin3a complex is successfully separated by contaminants.

Therefore, the sucrose gradient purification method represents an alternative way to the SEC to purify the Sin3a complex for biochemical studies. Indeed, it is possible to choose which purification technique will be more appropriate depending on the experiment to be performed.

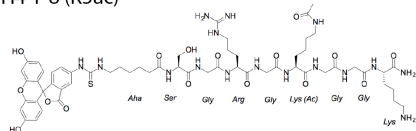
3.5. Initial characterization of the Sin3a complex - Investigation of substrate specificity with microfluidic mobility shift assay (MSA).

Previous studies revealed that the lysine residues of the tails of histone H3 and H4 contain the majority of known targets for the HDAC enzymes (Waterborg & Matthews 1984; Drogaris et al. 2012; Tessarz & Kouzarides 2014). To investigate the role of the Sin3a complex in the regulation of transcription and to identify any substrate preference, microfluidic Mobility Shift Assays (MSA) were performed on nine HDACs targets using the Caliper LabChip platform (Perkin Elmer) as described in the material and methods.

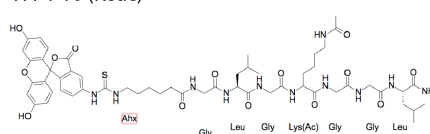
The substrates were synthesized as fluorescently labeled histone tail peptides in the laboratory of Dr Andrew Jamieson at the University of Leicester. The histone tail peptides were synthesized as 7-mers and they all included residues N- and C- terminal to the acetylated lysine. The synthesized peptides include the following acetylated lysine residues: 5, 8, 12 and 16 of histone H4 and 4, 9, 14, 27 and 56 of histone H3 (Figure 3.13).

H4 peptides

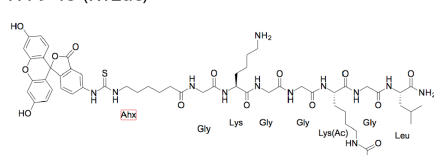
H4 1-8 (K5ac)



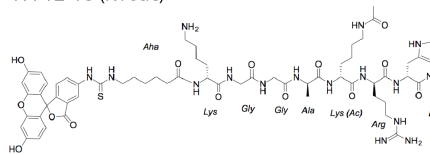
H4 4-10 (K8ac)



H4 9-15 (K12ac)

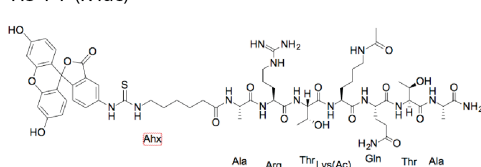


H4 12-18 (K16ac)

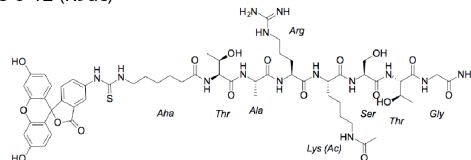


H3 peptides

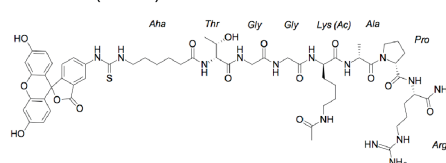
H3 1-7 (K4ac)



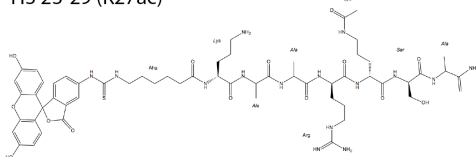
H3 6-12 (K9ac)



H3 11-17 (K14ac)



H3 23-29 (K27ac)



H3 53-60 (K56ac)

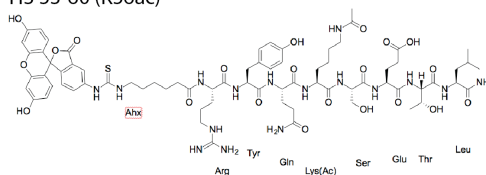


Figure 3.13. **Chemical structures of the histone peptides.**

The five H3 and four H4 tail peptides used in the MSA to test the Sin3a substrate specificity are shown. The structures also show the fluorescein isothiocyanate (FITC) attached to the peptide with a short linker.

Microfluidic Mobility Shift Assay (MSA) was performed to investigate any potential substrate specificity of the Sin3a complex for any of these nine different peptides tested. The MSA experiment was carried out on Caliper EZ reader (Perkin Elmer) in collaboration with Dr Watson. The Sin3a₅₃₁₋₇₇₂ : HDAC1 : SDS3 complex was expressed

in HEK 293F cells, purified with FLAG[®] resin and subsequently further purified through a Superose 6 (10/300) GE Healthcare column at 4°C as described in materials and methods. To verify the HDAC activity of the Sin3a complex and to compare it to the activity of other HDAC containing complexes, a BOC-lysine activity assay was performed as described in materials and methods (Figure 3.14).

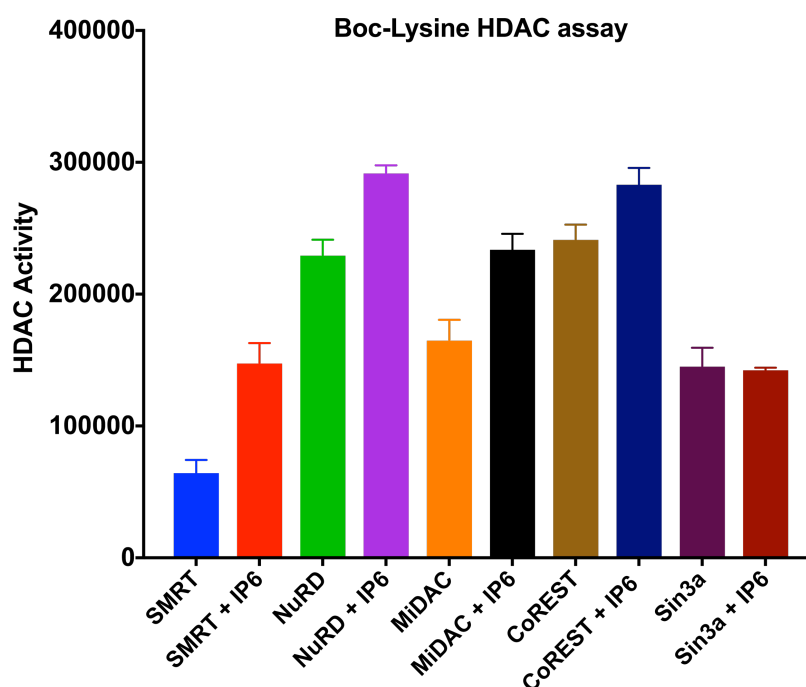


Figure 3.14. **BOC-lysine activity assay of the HDAC containing complex.**

The HDAC activity of 10 nM SMRT, NuRD, MiDAC, CoREST and Sin3a complex was tested in the presence or absence of 100 μ M IP₆.

The graph in figure 3.14 shows that the presence of IP₆ does not affect the deacetylase activity against any of the substrate. This result confirms previous preliminary studies in our laboratory (unpublished), where it was shown that the activity of the Sin3a complex is not affected by the presence of IP₆.

After finding that the Sin3a complex was active, the substrate specificity of the complex was tested. The Sin3a complex was tested in triplicate, at a final concentration of 125 nM with or without 100 μ M IP₆. The peptides were added to a final concentration of 4 μ M before analyzing the sample with the Caliper. The initial deacetylation rate of Sin3a with the different histone tails peptides is shown in figure 3.15.

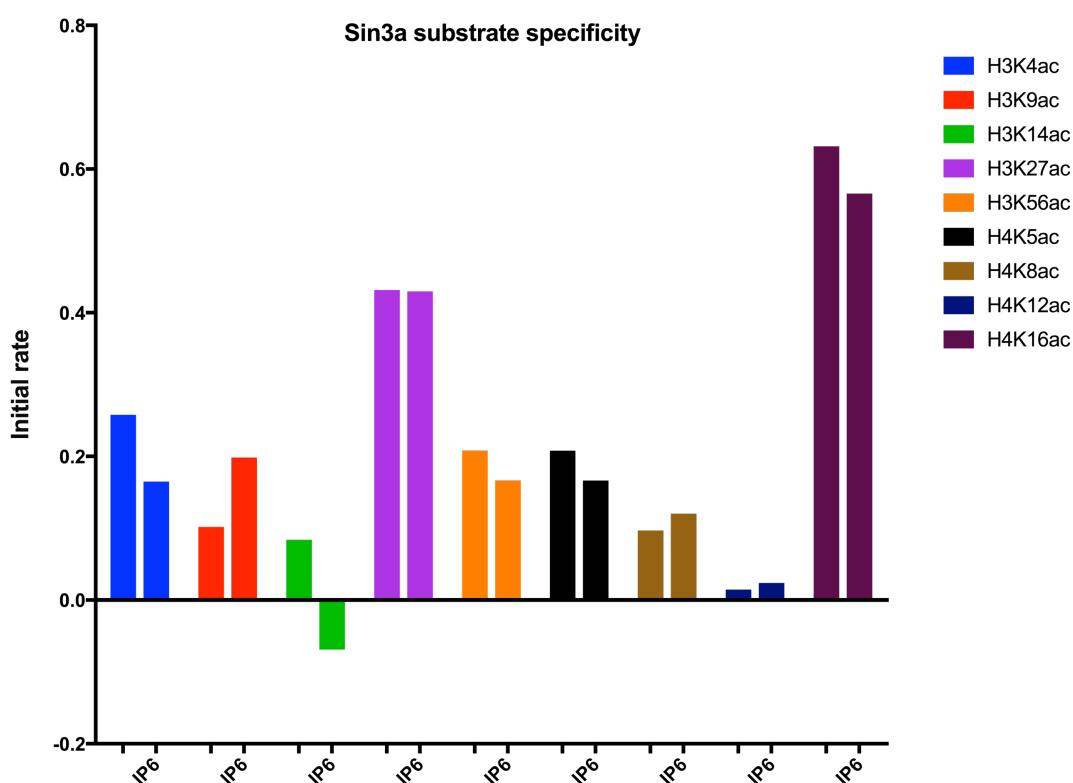


Figure 3.15. **Initial deacetylation activity rate of Sin3a with different histone tails peptides in the presence or absence of IP₆.**

The HDAC activity of the Sin3a complex was tested in the presence or absence of IP₆.

The results show that the most significant differences are seen between H4K12ac, H3K27ac and H4K16ac (Figure 3.15). Compared to the other substrate, Sin3a appears almost inactive against H4K12ac. Instead the Sin3a complex is twice as active against H3K27ac and three times as active against H4K16ac in comparison to the H4K12ac substrate. However, the artificial nature of the assay or the large fluorescent group on the peptide substrate could affect the specificity of the Sin3a activity.

Since the Sin3a complex seems to be most active against the H4K16ac peptide, to compare the Sin3a activity with other HDAC containing complexes this peptide was chosen as common target. The SMRT, NuRD, CoREST and MiDAC complexes were kindly expressed and purified by Dr Watson, Dr Millard, Dr Song and Dr Milano respectively.

In order to normalise the quantity of HDAC enzyme for each complex, several dilutions

of each protein complex was analysed by SDS-PAGE followed by densitometry on the gel bands corresponding to the HDAC enzymes. The initial rate of HDAC activity of the different HDAC complexes tested is shown in figure 3.16.

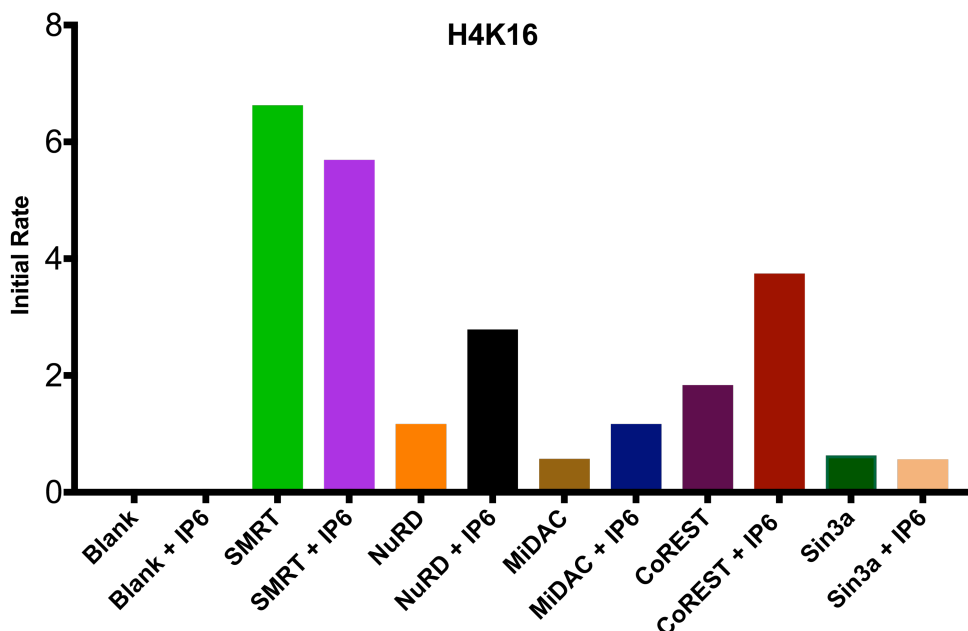


Figure 3.16. **Initial deacetylation activity rate of different HDAC containing complexes with H4K16ac in the presence or absence of IP₆.**

The deacetylase activity on H4K16ac of the Sin3a complex, MiDAC complex, CoREST complex was tested in the presence or absence of IP₆. Sin3a appears to be the least active compared to the other HDAC complexes.

From figure 3.15, H4K16ac was shown to be the target with the highest Sin3a activity, but when comparing the activity of different HDAC complexes on H4K16ac (figure 3.16), the Sin3a complex appears to be less active compared to the other complexes tested.

The Sin3a complex certainly shows differential activity against the substrates tested, which could be interpreted as substrate specificity. However, the Sin3a complex used for this assay comprises only Sin3a₅₃₁₋₇₇₂, HDAC1 and SDS3. *In vivo*, Sin3a is a large complex comprising many different components. It is possible that different complex components could change the substrate specificity to the endogenous Sin3a complex *in vivo*. A possible structural change or the direct recruitment of substrate by one of the complex components could affect the specificity of the Sin3a complex dramatically.

Additionally, it should be considered that the peptides used for this assay are only 7 mers with single modifications. In the cell, these peptides would be attached on histones and contain multiple modifications (not just acetylation). Therefore, the *in vivo* specificity could be very different compared to the one showed by this assay.

3.6. Result summary and discussion

In this chapter the expression and purification of the Sin3a complex was discussed. First of all, bacterial and mammalian expressions were tested and compared. Despite the successful expression of Sin3a₆₀₁₋₇₄₂ and SDS3₂₀₁₋₂₃₄ in bacteria, when HDAC1 is expressed in *E. coli* the complex fails to show deacetylation activity. When HDAC1 was instead expressed in mammalian cells, the complex resulted to be active (Fig 3.8 and 3.12). Post Translational Modification (PTMs) are therefore required in order to express an active form of HDAC1, determining the necessity to express HDAC1 in HEK 293F suspension cells.

However, when Sin3a₆₀₁₋₇₄₂ and SDS3₂₀₁₋₂₃₄ were expressed in bacteria and HDAC1 in mammalian cells and the complex was reconstituted during the purification step, despite several attempts a stoichiometric complex could not be produced. Furthermore, the resulting complex fell apart during multiple purification steps, showing a weak interaction between the components. Only by expressing all the complex components together in mammalian cells could a stoichiometric complex hold together through multiple rounds of purification.

From the expression trials performed in this study it was revealed that in contrast with what was published by Clark *et al.*, Sin3a₆₀₁₋₇₄₂ is able to interact with HDAC1. In every purification trials performed Sin3a₆₀₁₋₇₄₂ was always able to co-purify HDAC1 either it was expressed in bacteria or in mammalian cells. However, since the Sin3a₆₀₁₋₇₄₂ : HDAC1 interaction did not survive multiple rounds of purification, it became clear that the additional residues in the C- and N-termini of Sin3a₅₃₁₋₇₇₂ are necessary to hold a tight interaction with HDAC1 and enable the complex to hold together through multiple rounds of purification.

After assessing the expression of a stoichiometric and well-behaved complex, a suitable purification method was needed to obtain a high purity sample for further biochemical and structural studies. Two different methods were tested to purify the Sin3a₅₃₁₋₇₇₂ : HDAC1 : SDS3 complex. By using the commonly used Size Exclusion Chromatography (SEC) method, it was shown that the experimental temperature plays a fundamental role in the maintenance of the stability of the complex. To avoid the degradation of the complex it is necessary to perform the purification step and all further experiments at 4°C. If SEC purification cannot be performed at 4°C, an alternative method that can be used is the sucrose gradient density separation method. During this study it was shown that by purifying the complex through a sucrose gradient it was possible to successfully purify the Sin3a complex from other proteins and at the same time the complex remains stable.

Preliminary experiments were also performed to characterize the Sin3a complex. 9 different lysine residues of the tails of histone H3 and H4 (known targets for the HDAC enzymes) were used to test possible substrate specificity for the Sin3a complex. Overall, the deacetylase activity shown on any peptides was low, but between the peptides target tested Sin3a showed some preferences H4K16ac. However, by comparing the activity on H4K16ac of Sin3a with four other Class I HDAC complexes Sin3a appeared to be the least active between all the complexes tested. As a large multiprotein complex, Sin3a interacts with many different co-repressor proteins that target specifically the activity of the complex to chromatin. The substrate specificity of the complex could be influenced by the presence of different proteins interacting with Sin3a. Furthermore, cells in vivo present a variety of PTMs on their histone tails. It is possible that different patterns of PTMs are recognised by Sin3a interacting proteins that target the activity of the complex to that specific site. Therefore, to have a better understanding of a possible substrate selectivity of the Sin3a complex the same experiment could be repeated including different components of the complex. Additionally, peptides of lysine residues showing more PTMs could be designed and tested to verify if the presence of different modifications affects the resulting activity of the complex.

Overall, from the expression and purification trials performed, a stoichiometric and

well behaved complex was successfully produced, facilitating the study of the Sin3a complex using biochemical and structural studies. Furthermore, more information about the Sin3a minimal residues involved in the interaction with HDAC1 was obtained. Residues 601-742 of Sin3a appear to be sufficient to interact with HDAC1, but residues 531-601 at the N-terminus and 742-772 at the C-termini of Sin3a allow a further stabilization of this interaction. These data provide insight into the structure of the interaction between Sin3a and HDAC1, suggesting multiple points of interaction between the two proteins. Further structural studies to characterize the Sin3a : HDAC1 interaction are described in Chapter 5.

Chapter 4. Functional Studies

4.1. The use of stapled peptides to specifically target the Sin3a complex

The Sin3a complex plays a fundamental role in the cell by activating and repressing target genes to regulate multiple aspects of both cellular and biological functions (Kadamb et al. 2013). Since the role of Sin3a is critical in cellular functions, if the regulation of the complex is compromised it can result in malignant transformation (S. J. Chen et al. 1993; Lin et al. 1998). In case the malignancy would be caused by an over activation of the complex (or by the failure of repressing its activity), the possibility to specifically target the Sin3a complex would allow to specifically block its action without affecting the activity of other Class I HDAC complexes.

Currently, there are only 5 HDAC inhibitors (HDACi) approved by the *U.S. FDA* (Food and Drug Administration). However, the use of HDACi is limited because of their serious side effects. For example, Sodium Valproate, an HDACi used for epilepsy treatment, was recently shown to be dangerous if taken during pregnancy. Indeed, it brings a 10% risk to bring physical abnormalities and 40% risk of developing neuro-developmental disorders in unborn babies (<http://www.bbc.com/news/health-41345647>). Since they all share the same mechanism of action by targeting the active site of HDACs enzymes, these severe side effects are caused by the lack of selectivity of these HDACi on their targets. HDACs enzymes share high structural similarities and it is very difficult to design isoform specific HDACi. Also HDAC1/2 are found in multiple complexes.

A better understanding of the complexes that recruit the enzymes, how they recruit their catalytic subunits and how they are regulated is required to develop more specific drugs. The ability to target the complexes instead of the HDACs' active sites would allow to target the specific activity of the complexes avoiding the side effects caused by HDACi.

To date, drugs directly directed against Sin3a are not available on the market and treatments are limited to the use of HDAC inhibitors. Potentially, there would be two

ways of targeting the Sin3a activity: either the disruption of the core complex or the prevention of its recruitment to chromatin.

The importance of the presence of SDS3 to assemble and to stabilise the Sin3a dimeric complex was discussed and demonstrated in Chapters 3. Previous functional studies have also shown that the deletion of SDS3 leads to a cell-lethal condition and its presence is required for the deacetylase activity of the complex (Lechner et al. 2000; David et al. 2003; David et al. 2006). The ability to displace SDS3 from its interaction with Sin3a would lead to the disruption of the core complex and the impairment of the complex activity.

Furthermore, as described in Chapter 1, paragraph 1.3.1, since Sin3a does not possess a DNA binding motif it needs to be recruited by co-repressor factors in order to bind chromatin. Sin3a can interact with chromatin-associated factors both directly and indirectly through other members of the complex. One of the most studied Sin3a interacting proteins is Mxd1 (previously known as Mad1). Mxd1 forms a heterodimer with Max and recruits the Sin3a complex to mediate gene repression (Laherty et al. 1997; Ayer et al. 1993). By hindering the interaction between Mxd1 and Sin3a, the inhibitory action of Sin3a to the selected targets will be relieved.

The use of small molecules to target the activity of proteins and protein complex is growing in significance for the treatment of several disorders. The use of hydrocarbon-stapled α -helical peptides for such treatments is a novel approach that seeks to combine the recognition capabilities of protein therapeutics with the ability of small molecules to penetrate cells (Jenssen & Aspö 2008; Karle et al. 1989; Karle 2001; Verdine & Hilinski 2012).

The aim of this chapter is to use stapled peptides to specifically target the Sin3a complex. In this study two different staple peptides were designed and tested. The first one was designed based upon the Sin3a Interaction Domain (SID) of SDS3 and the second one was designed on the SID of Mxd1. The SDS3 peptide should have the effect to disrupt the core complex, while the Mxd1 peptide should interfere with the recruitment of the complex to chromatin.

4.2. Stapled peptides synthesis

SDS3 and Mxd1 native and stapled peptides were synthesised by Dr Naomi Robertson and Dr Andrew Jamieson from the Chemistry Department at the University of Leicester (Currently at the University of Glasgow). The stapled peptide differs from the native form by the presence of an all-hydrocarbon staple. This chemical staple is introduced to the sequence to stabilise the α -helical structure of the peptide and to enhance target affinity, cell penetration and proteolytic resistance of the peptide.

4.3. The use of SDS3 native and stapled peptides to disrupt the Sin3a complex

Previous studies have shown that the deletion of SDS3 is lethal since its presence is required for the deacetylase activity of the Sin3a complex (Lechner et al. 2000; David et al. 2003). Furthermore, previous purification trials have shown that the presence of SDS3 in the Sin3a : HDAC1 complex tends to stabilize the complex leading to higher purification yields (Portolano et al. 2014). The use of stapled peptides as potential agents to target specifically Sin3a and disrupt the complex will bring two main benefits: a better understanding of the Sin3a : SDS3 interaction and the potential use of the SDS3 stapled peptide as a specific drug or tool compound to target Sin3a.

The SDS3 peptide was designed to contain 21 amino acids based on the sequence of the Sin3a Interaction Domain (SID) (figure 4.1) and comprises amino acids residues from 201 to 234. The choice of these amino acid boundaries was based on the recent publication of Clark *et al.* where the interaction between Sin3a₆₀₁₋₇₄₂ and SDS3₂₀₁₋₂₃₄ was solved by NMR studies (Clark, Marcum, et al. 2015; Clark, Zhang, et al. 2015). Figure 4.1A shows the displacement mechanism expected by the SDS3 stapled peptide and figure 4.1B shows the amino acid sequence for both the SDS3 peptides (native and stapled).

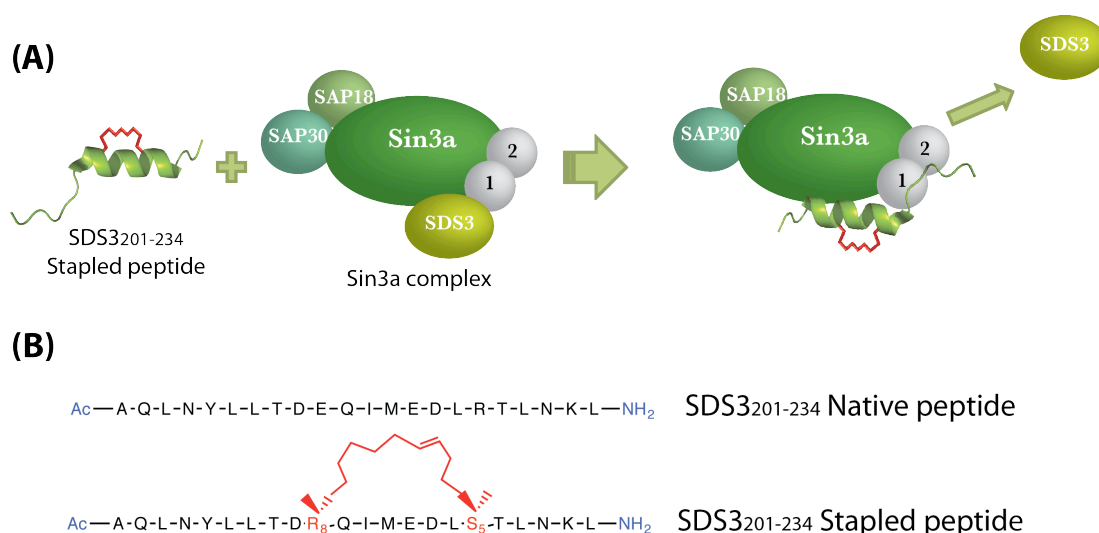


Figure 4.1: The SDS3 stapled peptide, mechanism of action and amino acid sequence.

(A) Predicted mechanism of action of the SDS3 stapled peptide. The addition of the SDS3 stapled peptide should cause the displacement of SDS3 (shown in light green) from its pre-existing interaction with Sin3a (shown in dark green). (B) SDS3 native and stapled peptides amino acid sequences. The chemical staple is shown in red.

Fluorescent versions of both native and stapled peptides were synthesised. The fluorescent peptides were synthesized, with fluorescein isothiocyanate (FITC) at their N-terminus and purified (> 95%) by Dr Naomi Robertson at the University of Leicester, Department of Chemistry (figure 4.2).

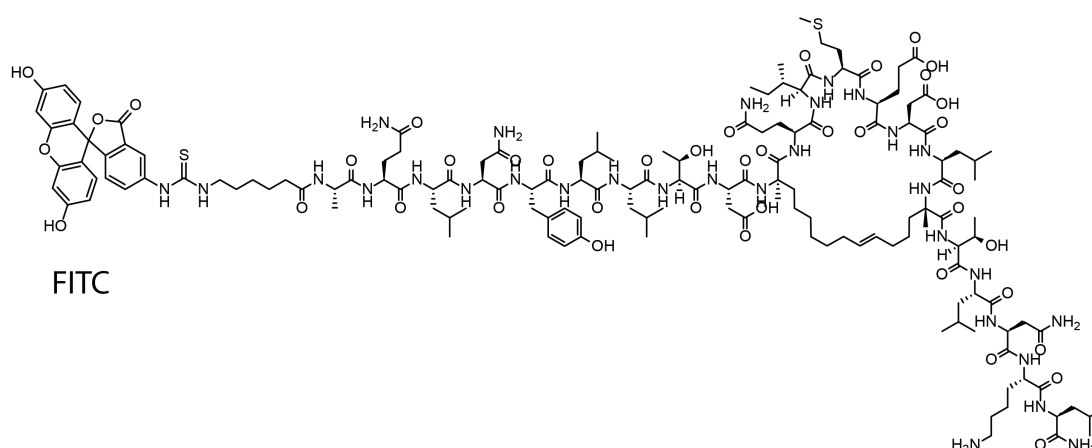


Figure 4.2: The fluorescent SDS3 stapled peptide. Chemical structure of the fluorescent SDS3 stapled peptide.

The fluorescein isothiocyanate (FITC) is positioned at the N-terminus of the sequence.

4.3.1. The SDS3 native and stapled peptides disrupt the interaction between Sin3a and SDS3 *in vitro*

To assess the ability of the SDS3 native and stapled peptides to disrupt the Sin3a : SDS3 complex, *in vitro* binding assays were performed.

(GST)-Sin3a₆₀₁₋₇₄₂ and (GB₁)-SDS3₂₀₁₋₂₃₄ were co-expressed and purified from bacterial cells as described in materials and methods. *In vitro* binding assays were performed by adding increasing amounts of native/stapled SDS3 peptides (0-40 nM) to a fixed amount of Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄ complex. The SDS3 native/stapled peptides were incubated with the GST resin to bind the Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄ complex for 1 hour at 4 °C. The GST resin was subsequently washed 3 times with GST wash buffer to remove any unbound material. Finally, the samples were analysed by SDS-PAGE. For this experiment the TEV-cleavage step was not performed, leaving Sin3a₆₀₁₋₇₄₂ with the GST tag and SDS3₂₀₁₋₂₃₄ with the GB₁ tag at their N-terminus. Figure 4.3 shows the results of the *in vitro* displacement experiment for the SDS3 native peptide and stapled peptide.

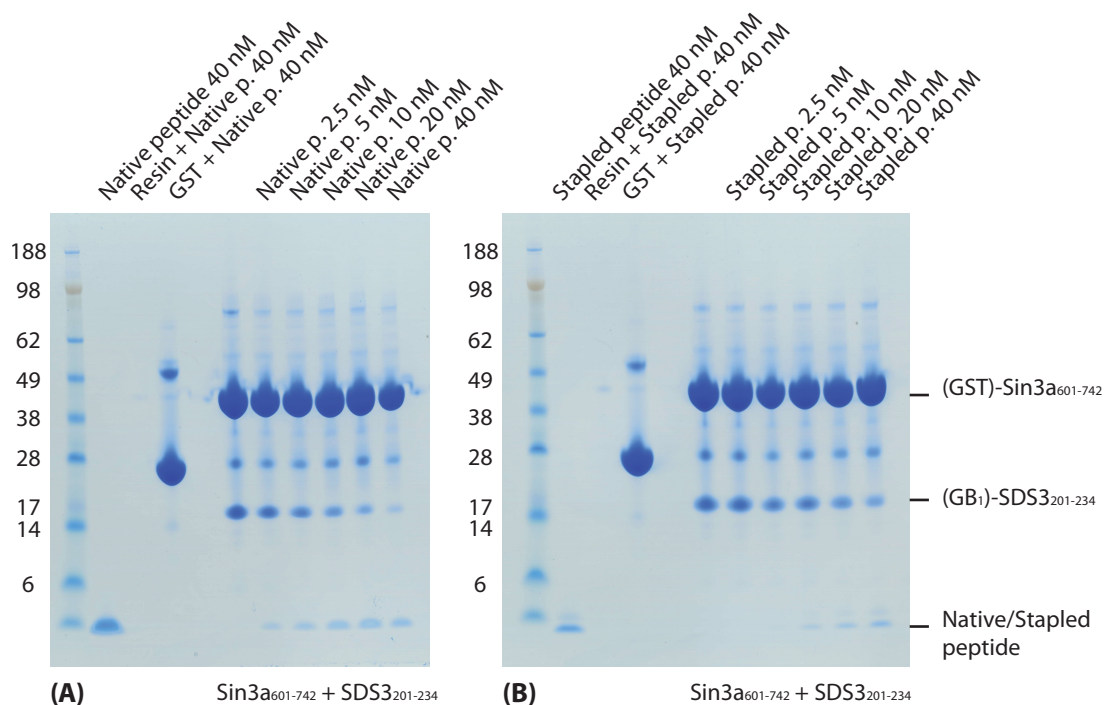


Figure 4.3: The SDS3 peptides successfully disrupt the Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄ complex. *In-vitro* binding assay.

(A) Increasing amount of native peptide (0-40 nM) are added to fixed concentration of Sin3a₆₀₁₋₇₃₇-SDS3₂₀₁₋₂₃₄ complex. (B) Increasing amounts of stapled peptide (0-40 nM) are added to fixed concentration of Sin3a₆₀₁₋₇₃₇-SDS3₂₀₁₋₂₃₄ complex.

Figure 4.3 shows that when either the SDS3 native or stapled peptides are added to the pre-existing binary complex Sin3a₆₀₁₋₇₄₂-SDS3₂₀₁₋₂₃₄, they both successfully displace SDS3₂₀₁₋₂₃₄ and interact with Sin3a₆₀₁₋₇₄₂. Increasing the amount of SDS3 peptides leads to a reduction of the intensity of the band corresponding to SDS3₂₀₁₋₂₃₄.

These *in vitro* experiments show that the SDS3 peptides can actively displace SDS3₂₀₁₋₂₃₄ from its pre-existing interaction with Sin3a₆₀₁₋₇₄₂. From these encouraging results, the actual dissociation constant (K_d) for fluorescent SDS3 stapled and native peptides was measured to compare the binding affinities between the peptides and the SDS3₂₀₁₋₂₃₄ construct used by Clark (Clark, Marcum, et al. 2015).

4.3.2. Measuring the K_d for the SDS3 native and stapled peptides

The chemical staple existing on the stapled peptide was designed to strengthen the stability of the helical structure to enhance its target affinity. By measuring the actual K_d of the SDS3 peptides binding to Sin3a₆₀₁₋₇₄₂, it will be possible to compare the binding affinity of the native and stapled peptides. Furthermore, it will be possible to verify if the displacement capabilities of the peptides could be affected by different constructs lengths by measuring the K_d of different constructs for the Sin3a complex components.

In 2015, Clark *et al.* found SDS3₂₀₁₋₂₃₄ to be the construct with the lowest K_d when comparing the interactions between Sin3a and different SDS3 constructs. Using a luciferase assay, the actual K_d for SDS3₂₀₁₋₂₃₄ was measured to be 0.93 μM (Clark, Marcum, et al. 2015). A fluorescent polarization (FP) direct binding assay was used to measure the actual K_d with three different complexes: Sin3a₆₀₁₋₇₄₂ alone, Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄ and Sin3a₅₃₁₋₇₇₂ : SDS3 : HDAC1.

Sin3a₆₀₁₋₇₄₂ and the binary complex Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄ were expressed and purified from *E. coli*, whereas the ternary complex Sin3a₅₃₁₋₇₇₂ : HDAC1 : SDS3 was expressed and purified from HEK 293F cells. Protein expression, purification and fluorescent polarization assays were performed as described in the materials and methods. The concentration of the SDS3 native and stapled peptides was fixed at 100 nM, whereas the concentration of Sin3a varied from 0.05 to 50 μM. Fluorescent polarization assays are shown in figures 4.4 and 4.5, the actual K_d measured for each experiments are shown in table 4.10 and compared in figure 4.6.

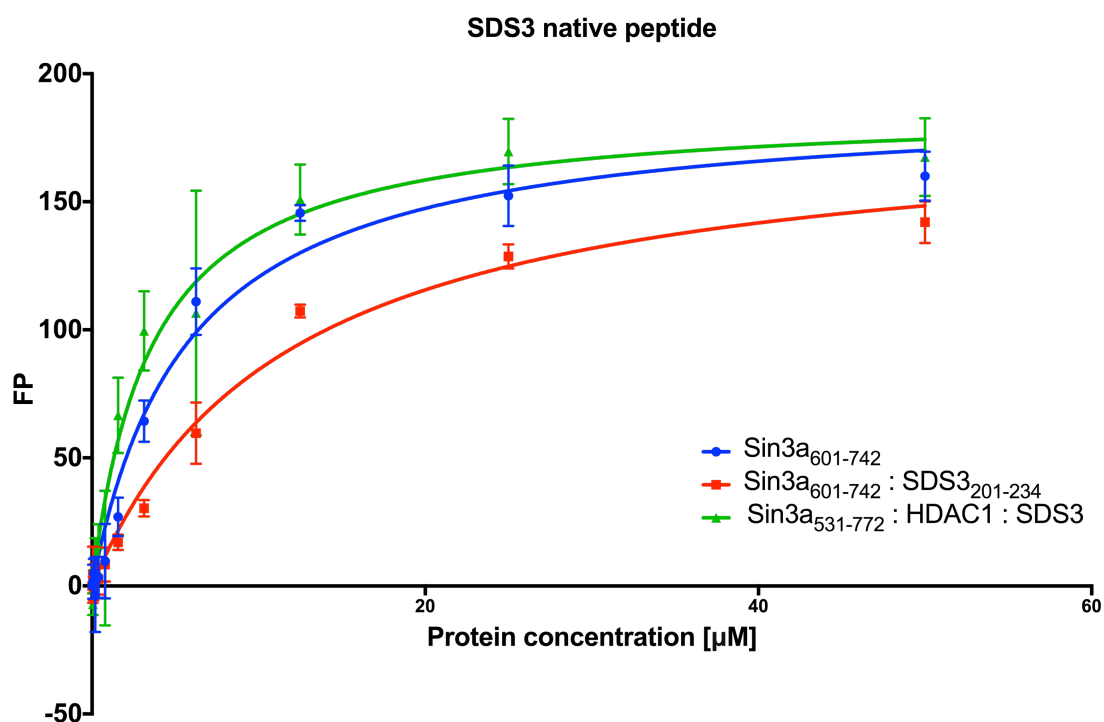


Figure 4.4: **Fluorescent polarization assay with SDS3 native peptide.**

The assay was performed with mixed amount of peptides (100 μM) and varying the amount of Sin3a (0.05 - 50 μM). Error bars indicate the mean with SEM, each experiment was performed in triplicate.

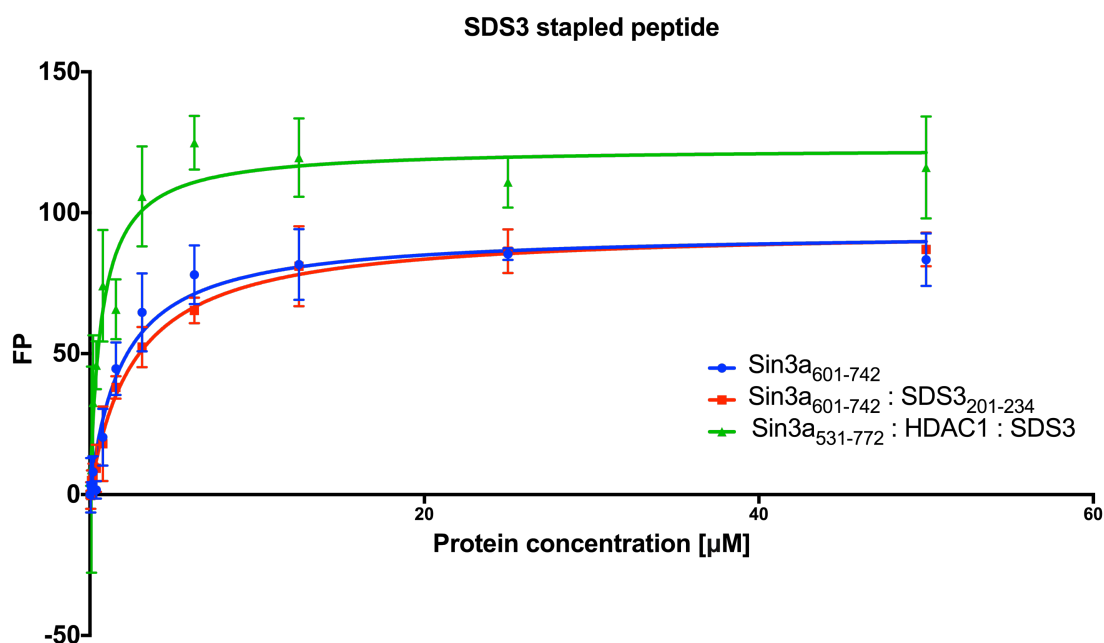


Figure 4.5: **Fluorescent polarization assay with SDS3 stapled peptide.**

The assay was performed with fixed amount of peptides (100 μM) and varying the amount of Sin3a (0.05 - 50 μM). Error bars indicate the mean with SEM, each experiment was performed in triplicate.

Table 4.9: SDS3 native and staple peptides, fluorescent polarization assays details.

The apparent K_d of the peptides with the different Sin3a complex constructs and components were calculated with GraphPad Prism. All data were calculated from triplicate measurements.

Constructs	Native SDS3 peptide			Stapled SDS3 peptide		
	Bmax (%) Normalised	Kd	R ²	Bmax (%) Normalised	Kd	R ²
Sin3a ₆₀₁₋₇₃₇	189.5 ± 16.5	5.7 ± 1.5	0.9	93.2 ± 7.9	1.9 ± 0.6	0.9
Sin3a ₆₀₁₋₇₃₇ + SDS3 ₂₀₁₋₂₃₄	183.2 ± 16.2	11.7 ± 2.7	0.9	94.4 ± 6.4	2.6 ± 0.6	0.9
Sin3a ₅₃₁₋₇₇₂ + HDAC1 + SDS3	186.9 ± 19.9	3.5 ± 1.3	0.9	123 ± 11.7	0.6 ± 0.2	0.8

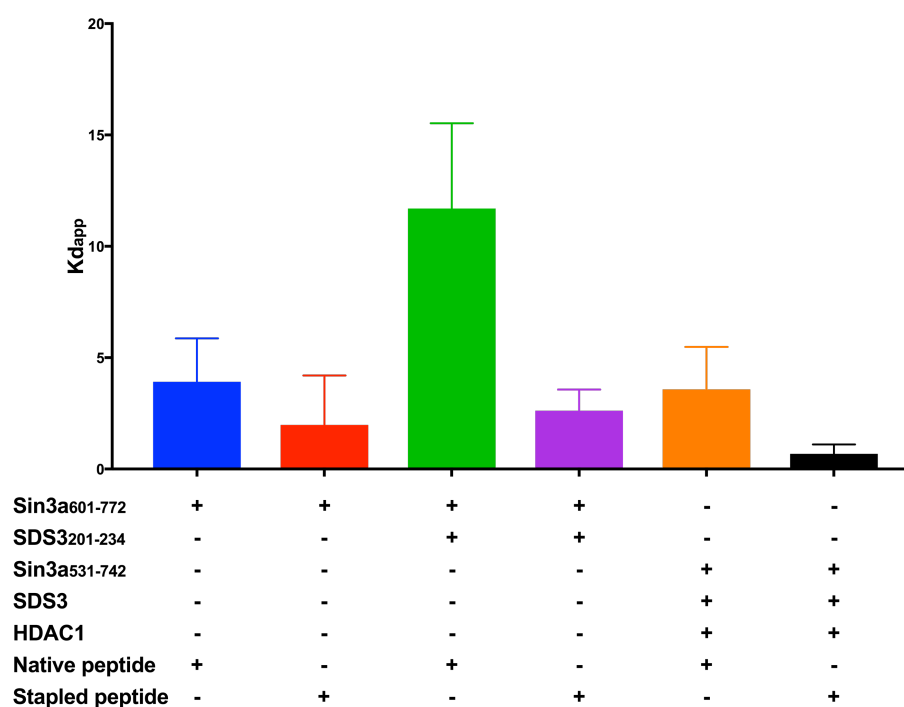


Figure 4.6: The actual K_d of the SDS3 peptide is not affected by different constructs and components of the Sin3a complex.

Comparison of the apparent K_d between SDS3 native and stapled peptides using different construct reactants.

From the results shown in figure 4.4 and 4.5, both native and stapled peptides show similar binding curves when binding different complexes. Therefore, the binding capabilities of the peptides are not affected by the different construct lengths used or

by the presence of other proteins that are already binding to Sin3a. It is also possible that the complex is not very stable at these low concentrations or more likely a fast exchange could happen in the interaction between Sin3a and the SDS3 peptides.

The SDS3 stapled peptide appears to have a trend of lower actual K_d values (higher affinity) compared to the native peptide. This result might be due to the presence of the chemical staple that stabilises the alpha helical structure of the peptide and enhances the target affinity of the peptide through reducing entropic cost.

After the promising *in vitro* results, it was decided to test the displacement abilities of the stapled peptides *in cellulo*.

4.3.3. The SDS3 native and stapled peptides fail to disrupt the Sin3a complex *in cellulo*

The ability to disrupt the Sin3a : SDS3 complex could open a new possibility for the use of the SDS3 stapled peptide as a tool to target specifically the activity of the endogenous Sin3a complex. However, the peptides are not only required to interact specifically with their target but it is also necessary that they can enter the cells. Therefore, *in cellulo* luciferase assays were performed in HEK 293T mammalian cells to further test the capabilities of the SDS3 stapled and native peptides.

HEK 293T cells were transfected with a luciferase reporter plasmid containing both Gal4- and LexA-binding sites (as described in materials and methods). Two constructs, full length SDS3 and SDS3₂₀₁₋₂₃₄, were fused to the heterologous DNA-binding domain of Gal4 to measure the ability of the fusions to repress transcription through the recruitment of Sin3a. The ability of Gal4.SDS3 or Gal4.SDS3₂₀₁₋₂₃₄ to repress Lex.VP16-activated transcription was measured as the resultant normalized luciferase activity. If the native and/or stapled peptides are able to interact with the endogenous Sin3a complex, the interaction between Sin3a and Gal4.SDS3 and/or Gal4.SDS3₂₀₁₋₂₃₄ would be prevented or displaced. Consequently, the Sin3a complex would fail to repress transcription.

Figure 4.7 represents the possible outcome of the peptides on the luciferase assay. As a preliminary experiment, the efficiency of both the Gal4.SDS3 constructs was tested. Figure 4.8 shows the comparison of the luciferase assay with Gal4.SDS3₂₀₁₋₂₃₄ and Gal4.SDS3.

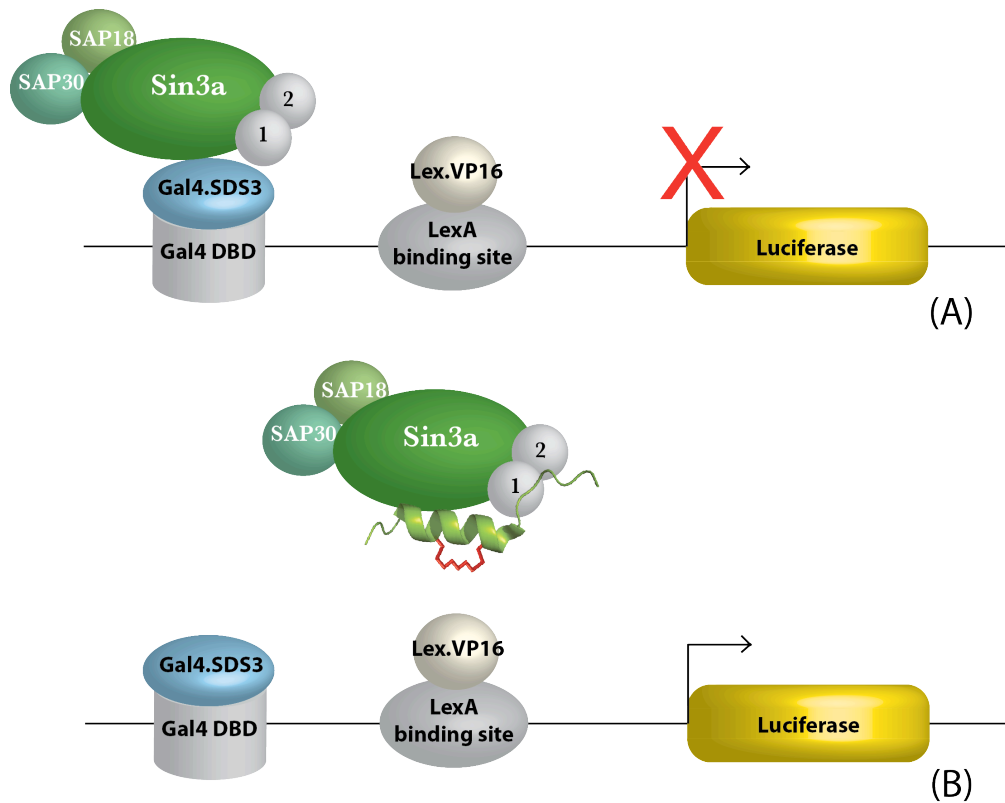


Figure 4.7: Possible outcome of the SDS3 peptide on the luciferase assay.

(A) The interaction of the Sin3a complex with the Gal4.SDS3 will lead to the repression of the luciferase activity. (B) When the peptide is added to the cells, if capable of recruiting the endogenous Sin3a complex, will displace the latter by its interact with Gal4.SDS3, leading to a release of luciferase activity.

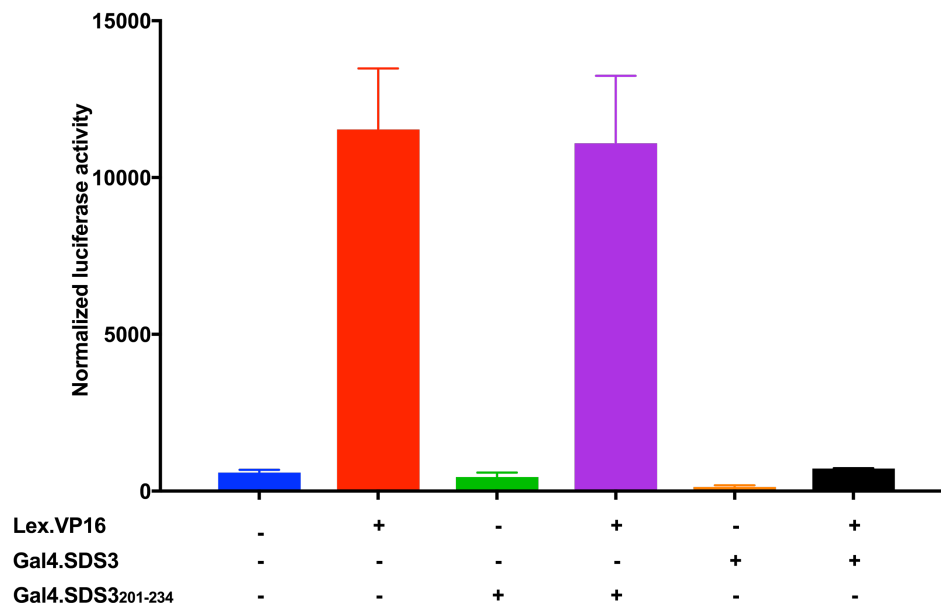


Figure 4.8: **The shorter construct Gal4.SDS3₂₀₁₋₂₃₄ fails to recruit the endogenous Sin3a complex.**

Luciferase assay to comparison the ability of Gal4.SDS3 and Gal4.SDS3₂₀₁₋₂₃₄ to recruit Sin3a. The shorter SDS3 construct fails to recruit the endogenous Sin3a complex and to repress the luciferase activity.

Figure 4.8 shows that full length Gal4.SDS3 successfully recruits Sin3a leading to the repression of the luciferase activity. However, the shorter SDS3 construct (Gal4.SDS3₂₀₁₋₂₃₄) fails in the recruitment of the endogenous Sin3a complex.

Therefore, the full length Gal4.SDS3 was used in all the following experiments to test the efficiency of the SDS3 native and stapled peptides to disrupt the endogenous Sin3a complex.

As preliminary experiment, the effects of both native and stapled peptides were tested by adding increasing amount of peptides to the assay (figure 4.9).

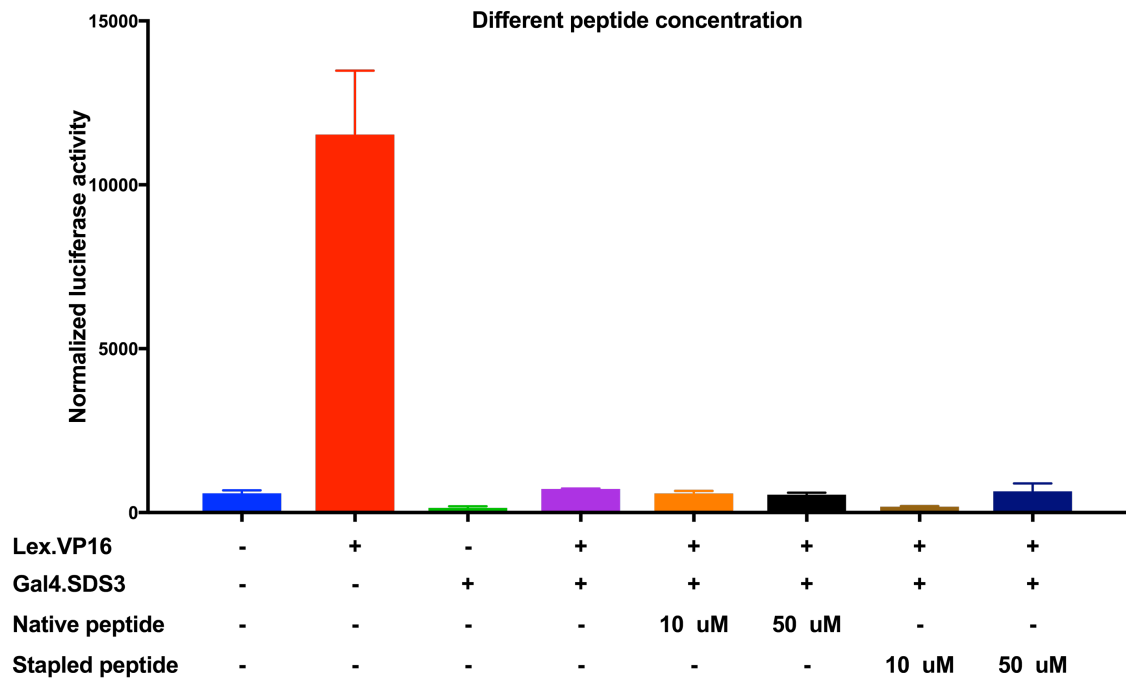


Figure 4.9: **Higher peptide concentration does not affect the outcome of the luciferase assay.** Two concentrations for the SDS3 peptides were tested (10 and 50 μ M). No release of luciferase activity is shown when either the SDS3 native or stapled peptides are added to the cells.

Figure 4.9 shows that both the SDS3 native and stapled peptides fail to disrupt the endogenous Sin3a complex containing Gal4.SDS3 despite the increasing concentrations used. A release of luciferase activity cannot be detected even when higher peptide concentrations are added to the cells. This result could be caused by a failure to compete or poor uptake of the peptides into the cells.

To facilitate the entrance of the peptides to the cells, 10 μ M of SDS3 native and stapled peptides were incubated with lipofectamine during the transfection stage (figure 4.10). Lipofectamine forms liposomes in an aqueous environment and it interacts with negatively charged molecules. This reagent is routinely used in other systems to facilitate DNA to enter the cells thanks to the fusion of the positive charged liposome with the cell membrane. Since the SDS3 peptides also negatively charged lipofectamine was tested to assist the entrance of the peptides into the cells.

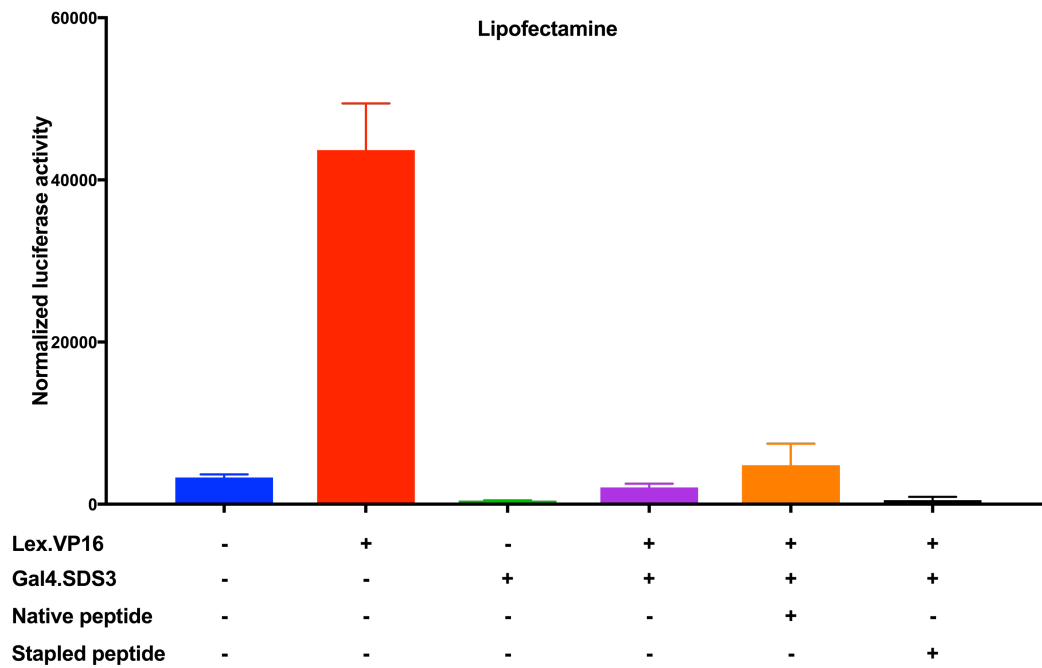


Figure 4.10: **Lipofectamine fails to facilitate the entrance of the peptides into the cells.**

10 μ M SDS3 native and stapled peptides were incubated with lipofectamine during the transfection. No effects can be detected on the luciferase activity when the peptides are added.

Figure 4.10 shows that when the peptides are added to the cells after being incubated with lipofectamine no change in luciferase activity can be detected.

Therefore, a different method to promote the access of the peptides into the cells was tested. In 2013, Ryan and Letai (Ryan & Letai 2013) described a new method to help peptides entering the cells by treating the cells with digitonin. Accordingly to the published protocol, 16 hours before harvesting the cells, 0.001% digitonin was added along with a final concentration of 10 μ M of native and stapled, after 20 minutes the cells were washed with PBS and left in complete media until harvesting (figure 4.11).

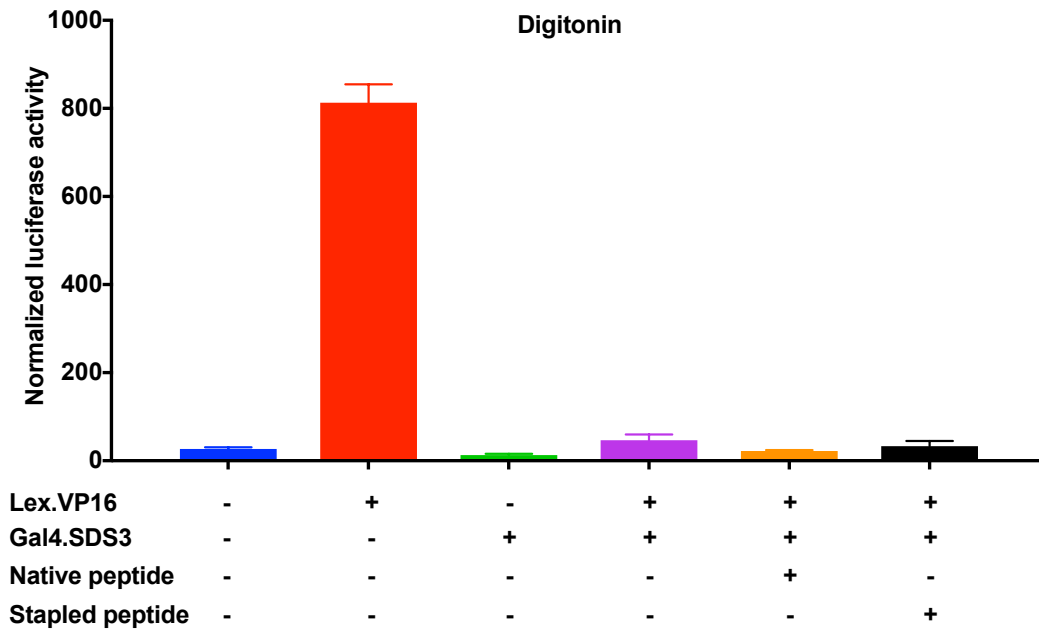


Figure 4.11: **Digitonin fails to help the peptides entering into the cells.**

16 hours before harvesting the SDS3 peptides were incubated with digitonin to facilitate the access of the peptides into the cells.

Unfortunately, figure 4.11 shows that the digitonin treatment showed results consistent with the previous experiments and neither the SDS3 native or stapled peptides seem to be able to disrupt the endogenous Sin3a.

Assuming that the peptides were able to enter the cell, the assays consistently showed that the SDS3 peptides fail to interact with the endogenous Sin3a complex. However, it remains possible that the peptides successfully enter the cells but they are targeted by proteases and fail to show their effects because degraded once entered the cell. To explore this possibility, a GFP tagged SDS3 construct was tested to verify the capabilities of the peptide sequence to disrupt the Sin3a complex.

4.3.4. GFP tagged SDS3₂₀₁₋₂₃₄ fails to disrupt the Sin3a complex

Since the SDS3 peptides comprise only 21 amino acids, this small peptide could be target of proteases. In this case, the failure in realising the repression of the Sin3a complex on the transcription of the luciferase could be caused by the early degradation of the peptides and not by the failure of interaction. In a further attempt to ensure that the SDS3 peptide sequence can enter the cells, a GFP tagged SDS3 peptide was transfected within HEK 293T cells. With this experiment, the GFP tag reduces the chances of degradation of the peptide sequence and allows the cells expressing the target sequence to be identified.

The immunofluorescence experiment was set as described in materials and methods. Figure 4.12 shows HEK 293T cells transfected with (GFP)-SDS3₂₀₁₋₂₃₄ that were visualised with Nikon TE300 inverted microscope.

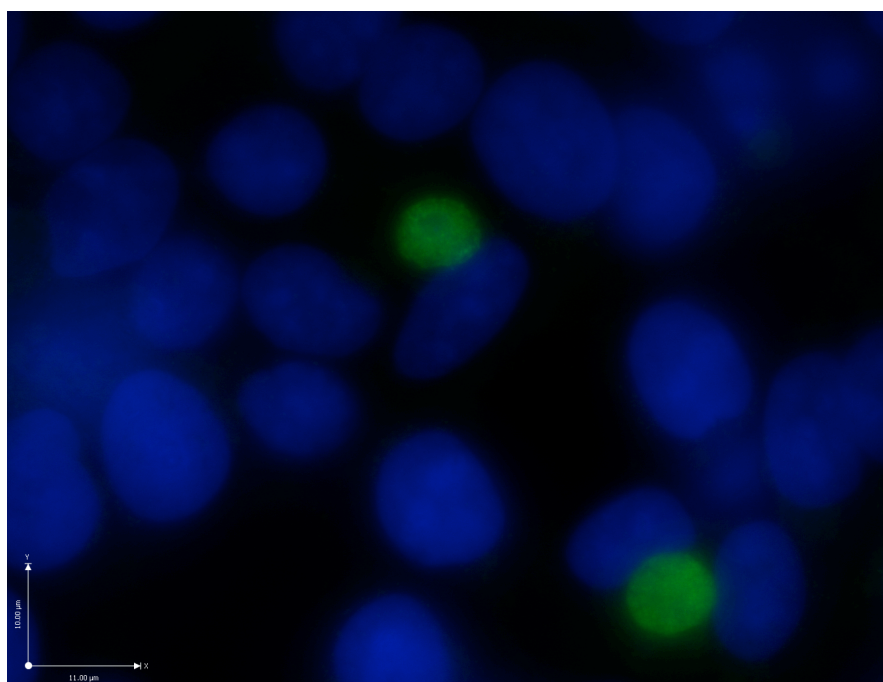


Figure 4.12: **(GFP)-tagged SDS3₂₀₁₋₂₃₄ in HEK 293T cells.**
Cells expressing (GFP)-SDS3₂₀₁₋₂₃₄ fluoresce green under the microscope.

GFP tagged SDS3₂₀₁₋₂₃₄ is successfully expressed in these cells, although the expression appears to be low. After assessing the presence of the construct in the cells, luciferase

assays were performed to verify ability of the (GFP)-SDS3₂₀₁₋₂₃₄ to interact with the endogenous Sin3a complex (figure 4.13).

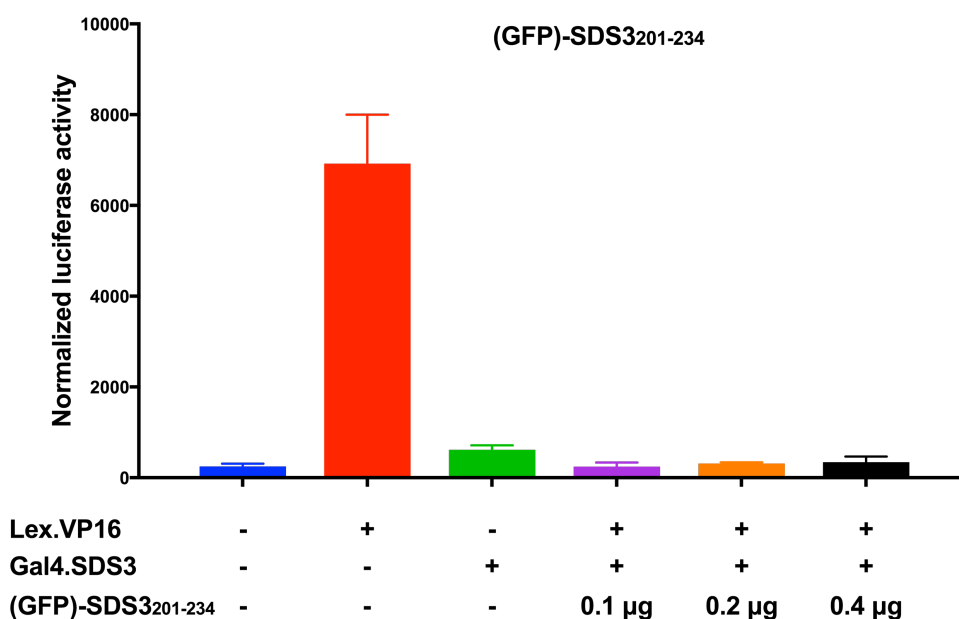


Figure 4.13: **(GFP)-SDS3₂₀₁₋₂₃₄ fails to disrupt the endogenous Sin3a complex.**

Luciferase assay, the addition of increasing amounts of (GFP)-SDS3₂₀₁₋₂₃₄ does not affect the repression of luciferase activity.

As shown in figure 4.13, the presence of (GFP)-SDS3₂₀₁₋₂₃₄ does not cause any change in the regulation of the reporter gene. Since the presence of SDS3₂₀₁₋₂₃₄ in the cells is verified by visualising the GFP tag, this assay confirms that SDS3₂₀₁₋₂₃₄ fails to disrupt the Sin3a complex.

From these results we could conclude that even if SDS3₂₀₁₋₂₃₄ was shown to interact with Sin3a by the structural studies performed by Clark in 2015 (Clark, Marcum, et al. 2015), this construct seems not to be sufficient to displace the full length SDS3 from its interaction with the endogenous Sin3a complex. The SDS3₂₀₁₋₂₃₄ represents only a small part of the full-length protein and it is possible that there are multiple points of interaction between Sin3a and SDS3. This more extended interaction would explain the inability of the stapled peptides to displace SDS3 *in vivo*.

However, after immunofluorescence (IF) the expression of (GFP)-SDS3₂₀₁₋₂₃₄ appears to be low with only few cells expressing the GFP tagged protein. Before applying IF to the cells, the transfected cells were visualised and it was noticed that many cells

expressing the GFP tagged protein were dying. When HEK 293T cells are dying they assume a round shape, they are less attached to the other cells and they tend to float in the media. Since IF comprises many cell washes, it appears that most of the cells expressing (GFP)-SDS3₂₀₁₋₂₃₄ were washed away, therefore they were not analysed with luciferase assay.

The reason for this (supposed) shorter life of the cells expressing (GFP)-SDS3₂₀₁₋₂₃₄ is unknown and it remains unclear whether the presence of the peptides is the cause for the death of the cells. To better follow the cells expressing the (GFP)-peptides and to compare them to the other cells, the cells could be monitored from the time of the transfection until the IF. In this way it would be possible to quantify the dying cells and to evaluate the percentage of cells expressing the peptides.

Since a different approach to target the Sin3a complex with the Mxd1 stapled peptides was possible, it was decided to bring forward the study of the use of stapled peptides by targeting the recruitment of the complex by chromatin associated factors.

4.4. Using Mxd1 native and stapled peptides to impede the recruitment of the Sin3a complex

Mxd1 (previously known as Mad1) is a basic helix-loop-helix leucine zipper protein (bHLH-Zip) that interacts with its related bHLH-Zip protein Max to form a heterodimer that binds to DNA. Sin3a can be recruited by Mxd1 through its PAH2 domain. In this way, the Mxd1 : Max heterodimer is able to target specific promoters and to function as a transcriptional repressor by forming a ternary complex with the Sin3a complex (Ayer et al. 1995). The repression by the Mad1 : Max DNA-binding heterodimer requires the interaction with the Sin3a complex, therefore the disruption of the interaction between Sin3a and Mxd1 leads to the failure of its inhibitory function (Kasten et al. 1996). Mxd1 recruits the PAH2 domain of Sin3a with its N-terminal alpha helical structure Mxd₁₋₂₇ (Cowley et al. 2004).

The use of a stapled peptide designed on Mxd₁₋₂₇ to obstruct the Sin3a recruitment from the Mxd1 : Max heterodimer will give a better understanding of the possibilities

to specifically target the recruitment of the Sin3a complex. Dr Naomi Robertson synthesized the Mxd1 peptide as a 21 amino acids peptide which includes the Sin3a interaction helix within amino acids 1-27 of Mxd1 (Kumar et al. 2011). Figure 4.14 shows the amino acid sequence for both the native and stapled Mxd1 peptides.

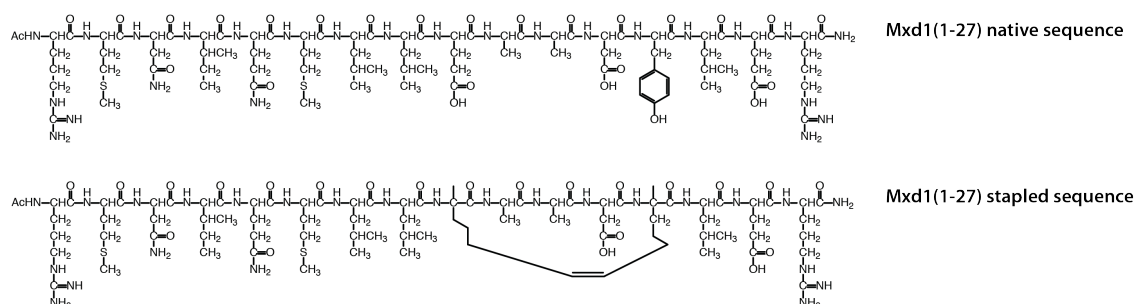


Figure 4.14: Mxd1 native and stapled peptides.
Amino acid sequence for both Mxd1 native and stapled peptides.

4.4.1. Mxd1 native and stapled peptide disrupt the Sin3a : Mxd1 complex *in vitro*

In vitro binding assays were performed to assess the capabilities of the Mxd1 native and stapled peptides to disrupt the pre-existing interaction with Mxd1 and to bind the PAH2 domain Sin3a₂₉₁₋₃₈₈.

(GST)-Sin3a₂₉₁₋₃₈₈ and (GB1)-Mxd1₁₋₂₇ were co-expressed and purified from bacterial cells as described in materials and methods. Increasing amounts of peptide (0-1.25 mM) were incubated with the GST resin binding the binary complex Sin3a₂₉₁₋₃₈₈ : Mxd1₁₋₂₇ for one hour. The resin was then washed three times and samples were analysed on SDS-PAGE (Figure 4.15).

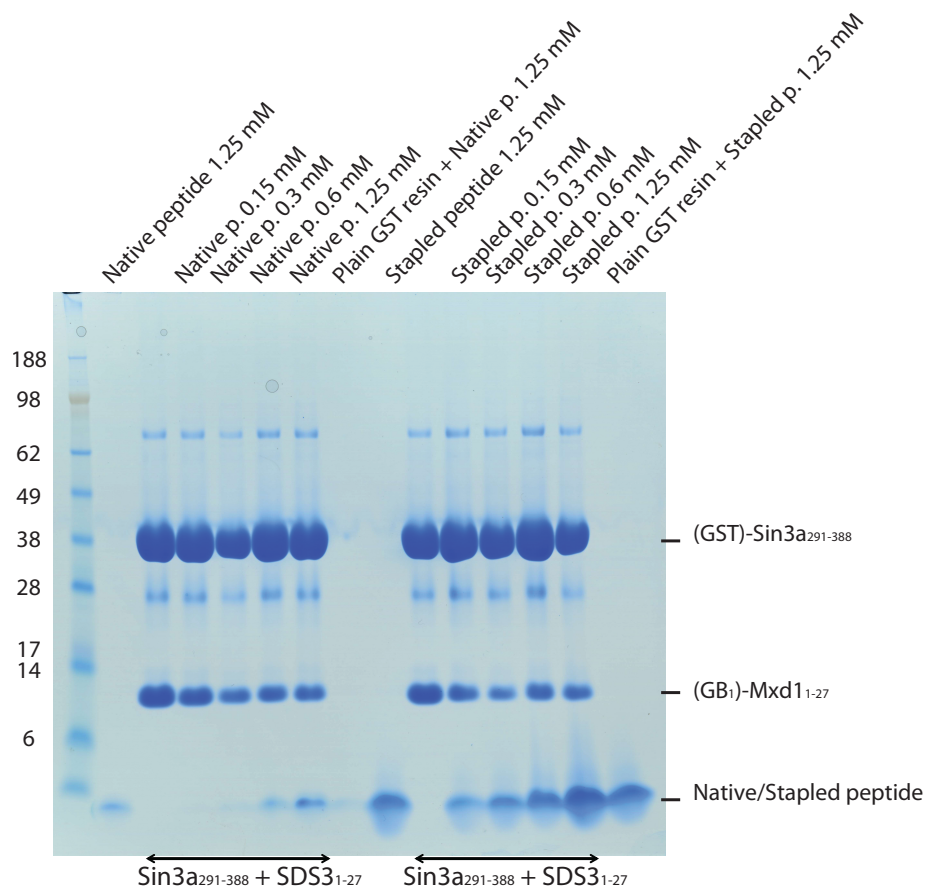


Figure 4.15: **The Mxd1₁₋₂₇ peptides successfully disrupt the Sin3a₂₉₁₋₃₈₈ : Mxd1₁₋₂₇ complex.** SDS-PAGE of the in vitro displacement experiment with the Mxd1 native and stapled peptides. Increasing amounts of peptides (0-1.25 nM) are added to fixed concentration of the Sin3a₂₉₁₋₃₈₈-Mxd1₁₋₂₇ binary complex. Both native and stapled peptides successfully disrupt the complex.

Figure 4.15 shows that when either the Mxd1 native or stapled peptides are added to the pre-existing binary complex Sin3a₂₉₁₋₃₈₈ : Mxd1₁₋₂₇, they both successfully displace Mxd1₁₋₂₇ and interact with Sin3a₂₉₁₋₃₈₈. On the SDS-PAGE it is clear that increasing amount of Mxd1 peptide leads to a more efficient displacement of Mxd1₁₋₂₇. The increased intensity of the peptide band on the gel corresponds to a Mxd1₁₋₂₇ protein band with reduced intensity.

This result shows that the Mxd1 peptides can disrupt the pre-existing interaction between Mxd1₁₋₂₇ and Sin3a₂₉₁₋₃₈₈. Therefore, from these encouraging results, the displacement capabilities of the peptides were tested *in cellulo*.

4.4.2. Mxd1 native and stapled peptides fail to disrupt the interaction between the native Sin3a complex and Mxd1 *in cellulo*

To confirm the displacement capabilities of both the native and stapled Mxd1 peptides, *in cellulo* luciferase assay experiments were performed in HEK 293T mammalian cells.

HEK 293T cells were transfected with a luciferase reporter plasmid containing both Gal4- and LexA-binding sites. Mxd1₁₋₂₇ was fused to the heterologous DNA-binding domain of Gal4. The ability of Gal4.Mxd1₁₋₂₇ to repress Lex.VP16-activated transcription was measured as the resultant normalized luciferase activity. If the native and/or stapled peptides would be able to interact with the endogenous Sin3a complex, the interaction between Sin3a and Gal4.Mxd1₁₋₂₇ would be prevented or displaced. Consequently, the Sin3a complex would fail to repress transcription (figure 4.16).

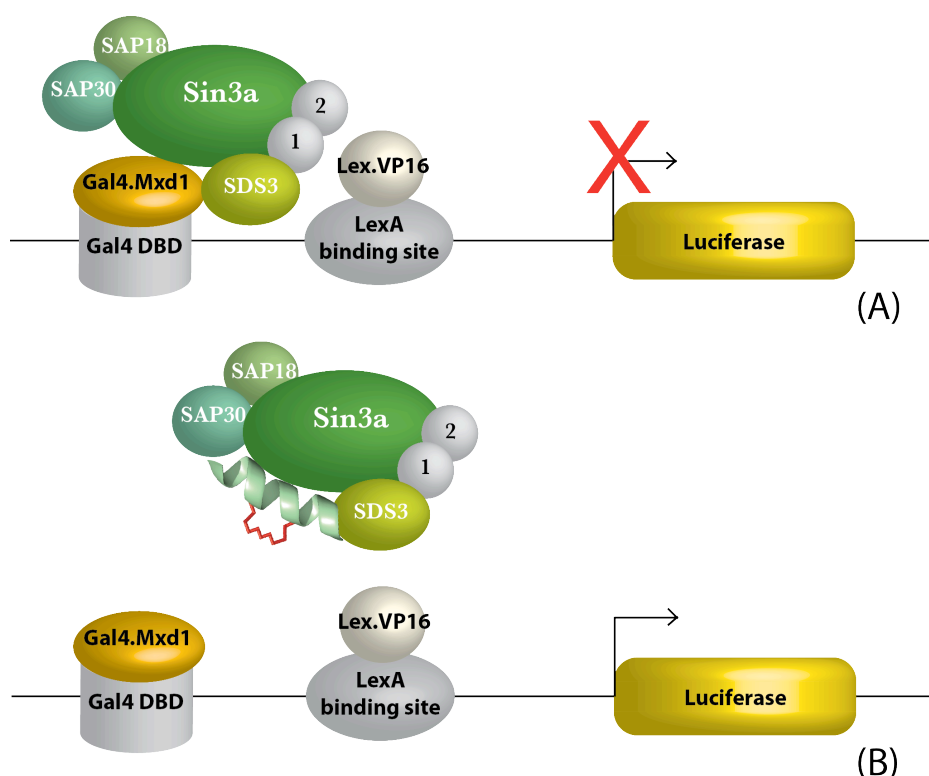


Figure 4.16: **Expected effect of the Mxd1 stapled peptide on the luciferase assay.**

(A) The interaction of the Sin3a complex with the Gal4.Mxd1₁₋₂₇ will lead to the repression of the luciferase activity. (B) When the peptide is added to the cells, if capable of recruiting the endogenous Sin3a complex, will displace the latter by its interaction with Gal4.Mxd1₁₋₂₇, leading to a release of luciferase activity.

As a preliminary experiment, the efficiency of the Gal4.Mxd1₁₋₂₇ was tested by adding increasing concentrations of peptides (10-100 μ M) to the transfected cells (figure 4.17). As control, the native peptide was also added to the cells without Gal4.Mxd1₁₋₂₇ and with Lex.VP16.

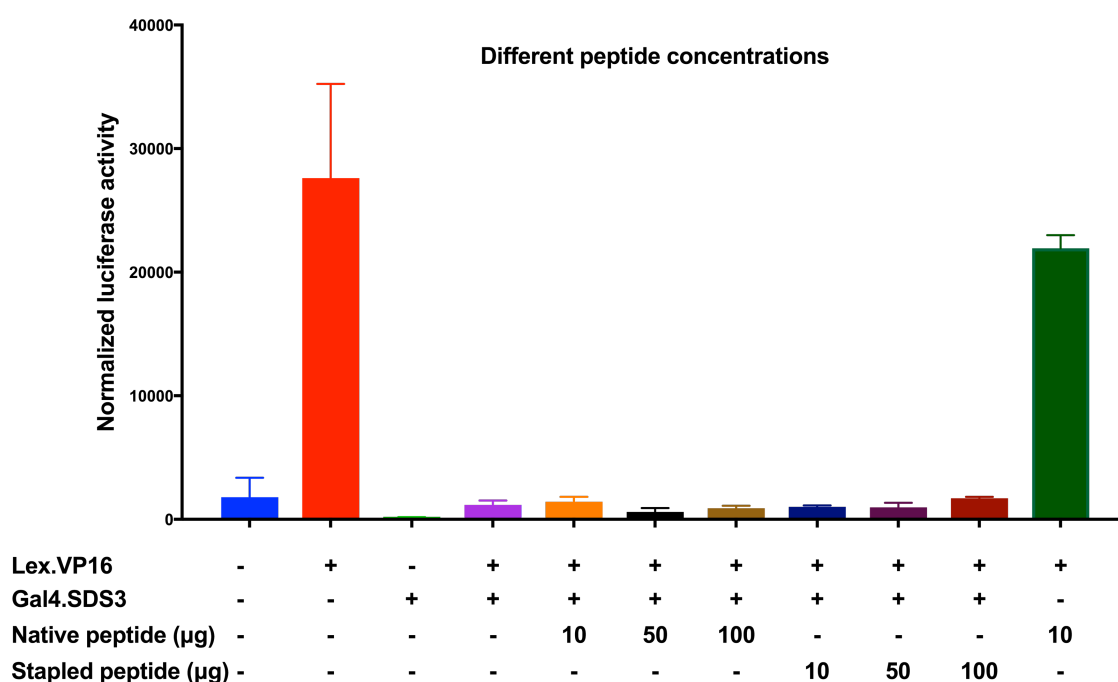


Figure 4.17: **Increasing amount of Mxd1 peptides do not affect the luciferase activity.**

Luciferase assay with Gal4.Mxd1₁₋₂₇, both native and stapled peptides were added to the cells at three different concentrations, 10-50-100 μ M. No difference in the luciferase activity can be measured when the peptides are added to the cells.

As we can see from figure 4.17, both the Mxd1 native and stapled peptides fail to prevent the endogenous Sin3a complex from interacting with Gal4.Mxd1₁₋₂₇. A release of luciferase activity cannot be seen even when higher peptide concentrations are added. This result could be caused either by the failure of the interaction of the peptide with the Sin3a complex or by a wrong choice of incubation time for the peptides.

In the previous experiment the peptides were incubated for 16 hours prior harvesting the cells. This incubation time could be too long and the peptides could be degraded by the harvesting time, or it could be too short and more time is needed to see the effects of the peptides on the luciferase transcription.

Therefore, a range of times between 3 and 24 hours was tested to find the best incubation time to show the effects of the peptides. A final concentration of 50 μ M Mxd1 peptides was added to the cells (Figure 4.18).

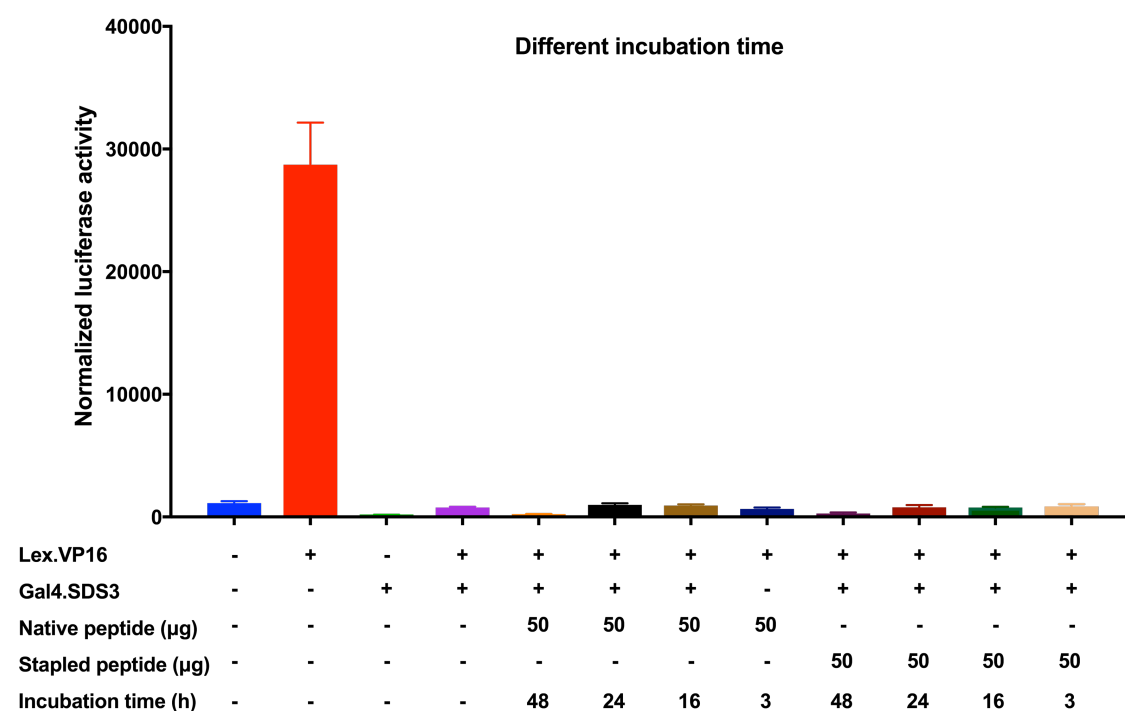


Figure 4.18: **Varying the incubation time for the Mxd1 native/stapled peptides.**

Luciferase assay with Gal4.Mxd1₁₋₂₇, both native and stapled peptides were added to the cells at 10 μ M. The peptides were added 3-16-24-48 hours before harvesting.

Figure 4.18 shows that despite the different incubation time used, neither the Mxd1 native or stapled peptides prevent the endogenous Sin3a complex from interacting with Gal4.Mxd1₁₋₂₇.

Since it could possible that the lack of effects on the luciferase activity is caused by the inability of the Mxd1 peptides to access the cells, different methods were used to facilitate the entrance of the peptides to the cells. As described in paragraph 4.3.3, both lipofectamine and digitonin were tested to improve the accessibility of the peptides into the cells (table 2). Unfortunately, no effects can be detected when either of the reagents is incubated with the peptides (table 4.11).

Table 4.10: Improving the accessibility of the peptides to the cells.

Lipofectamine and digitonin were used to incubate the peptides with the cells in order to improve access to the cells. Both the reagents fail to show any effects.

	Peptide concentration	Reagent	Effects shown
Mxd1 native peptide	50 μ M	Lipofectamine	No
Mxd1 stapled peptide	50 μ M	Digitonin	No

With the previous results, the lack of effects on the luciferase activity even when additional reagents were used to help the peptides to access the cells suggest that the Mxd1 stapled/native peptides fail to displace Mxd1 *in cellulo* and/or to access the cell.

However, it remains possible that the Mxd1 peptides could be target of proteases and fail to show their effects because they get degraded once entered the cell. A GFP tagged Mxd1 construct was therefore tested to verify the capabilities of the peptide sequence to impede the recruitment of the Sin3a complex.

4.4.3. GFP tagged Mxd1₁₋₂₇ disrupt the interaction between the endogenous Sin3a complex and Mxd1

Following the *in cellulo* luciferase assay, to assess the presence of Mxd1₁₋₂₇ within the cell, the Mxd1₁₋₂₇ construct was expressed with a N-terminal GFP tag. The immunofluorescence (IF) experiment was set up as described in materials and methods. Figure 4.19 shows HEK 293T cells transfected with (GFP)-Mxd1₁₋₂₇ that were visualised with Nikon TE300 inverted microscope.

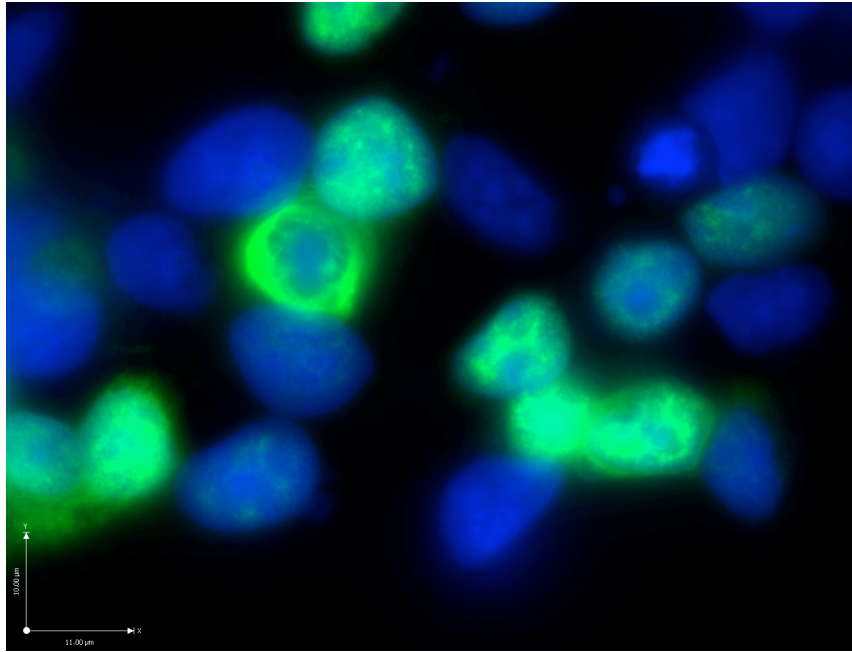


Figure 4.19: **(GFP)-Mxd1₁₋₂₇ transfected in HEK 293T.**
Cells expressing (GFP)-Mxd1₁₋₂₇ are shown in green.

Figure 4.19 shows (GFP)-Mxd1₁₋₂₇ successfully expressed in the HEK 293T cells. After assessing the presence of the construct in the cells, luciferase assay was performed to verify the ability of the (GFP)-Mxd1₁₋₂₇ to interact with the endogenous Sin3a complex (figure 4.20).

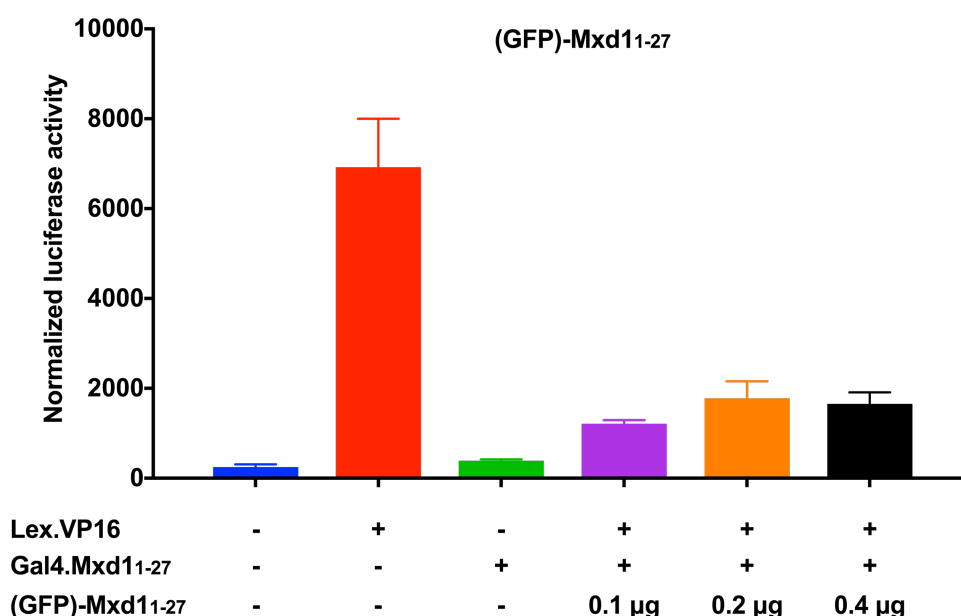


Figure 4.20: **(GFP)- Mxd1₁₋₂₇ successfully disrupt the interaction between the endogenous Sin3a complex and Gal4.Mxd1₁₋₂₇.**

The luciferase assay shows a release of luciferase activity when expressing (GFP)-Mxd1₁₋₂₇. The luciferase activity increases when increasing the DNA for (GFP)-Mxd1₁₋₂₇ used in the transfection.

Figure 4.20 shows that the presence of (GFP)-Mxd1₁₋₂₇ causes the release of luciferase activity. Increasing concentrations of (GFP)-Mxd1₁₋₂₇ leads to higher luciferase activity until finding a plateau above 0.2 µg of DNA for (GFP)-Mxd1₁₋₂₇ used in the transfection. This effect could be caused by the saturation of the transcription machinery or more likely the percentage of cells expressing the GFP-Mxd1.

This result shows the ability of (GFP)-Mxd1₁₋₂₇ to impede the Gal4.Mxd1₁₋₂₇ recruitment of the endogenous Sin3a complex. The same effect was shown by the *in vitro* displacement assays. However, the *in cellulo* experiment could not confirm the ability of the peptides to target the endogenous Sin3a. A reason for the failure of the *in cellulo* experiments could be the inability of the peptides to access the cells or their possible degradation by proteases. Indeed, when Mxd1₁₋₂₇ is expressed with either a N-terminal GFP tag, or with Gal4, it successfully recruits the Sin3a complex. Therefore, the presence of a N-terminal tag could prevent the degradation of the peptides and ensure their correct mechanism of action.

4.5. Results summary and conclusions

The aim of this chapter was to investigate the use of stapled peptides as potential future drugs to specifically target the Sin3a complex. The results of these studies have revealed important information about the mechanism of interaction existing between Sin3a and SDS3 and they have pointed out a possible weakness for the use of the stapled peptides as drugs.

The displacement experiments using SDS3 peptides gave promising results *in vitro* and they have shown a similar apparent K_d to the SDS3₂₀₁₋₂₃₄ construct measured by Clark (Clark, Marcum, et al. 2015). However, *in cellulo* experiments could not verify the displacement activity of the SDS3 stapled/native peptides on the endogenous Sin3a complex. To explore the possibility that the peptides could not enter the cells, GFP tagged SDS3 constructs were tested. The use of (GFP)-SDS3₂₀₁₋₂₃₄ would have mimicked the activity of the peptides and it would have allowed the detection of the fluorescent protein within the cell. Unfortunately, even though it was possible to detect a low fraction of the GFP-SDS3₂₀₁₋₂₃₄ in the cells, no alterations were shown on the regulation of transcription of the reporter gene.

The structural details of the interactions between the core complex components of Sin3a are still unknown and specifically, it remains unclear whether SDS3 has multiple points of contact with Sin3a or it only binds through its small alpha helix in the SID. Therefore, a better understanding of the structure of interaction is needed in order to draw a conclusion on the success or failure of the use of stapled peptides to target the Sin3a complex. When the interaction between Sin3a and SDS3 is fully understood, changes in the structure of the stapled peptide could be made and/or tandem stapled peptides could be employed to target multiple points of interaction (this approach might be successful to fully displace the endogenous SDS3) but would be unlikely to be a “drug”.

Experiments with Mxd1 native and stapled peptide gave a different outcome to the SDS3 native/stapled peptides. Promising *in vitro* displacement assays were followed by unsuccessful *in cellulo* displacement experiments. To verify the presence of the

peptides within the cells, GFP tagged Mxd1₁₋₂₇ constructs were tested. The presence of the GFP-Mxd1₁₋₂₇ was confirmed in the cells and interestingly the presence of the tagged protein was coupled with the release of repression on the transcription of the gene reporter. This result shows that the sequence of the peptides successfully displace Mxd1 from its interaction with the endogenous Sin3a complex, but at the same time it highlights possible weaknesses in the use of stapled peptides as drugs.

The lack of effects on transcription in the *in cellulo* experiments shows that either the peptides are incapable to access the cells or they are target of endogenous proteases which destroy the peptides and impede their interaction with the Sin3a complex. If any of these hypotheses are true, improvement on the stability and cell penetrance of the peptides are needed in order to use stapled peptides to selectively target protein complexes.

Chapter 5. Structural studies

5.1. Understanding how HDAC1 is recruited by Sin3a

As mentioned in paragraph 1.4, with the only exception of the Sin3a complex, Class I HDAC complexes seem to share the same mechanism of recruitment of the HDAC enzymes and the same regulatory mechanism. The overall aim of this chapter is to study the structural interaction of the core complex of Sin3a. In particular, to investigate how Sin3a interacts with HDAC1 and SDS3.

Within Class I HDACs, structural studies have shown that HDAC1 interacts with MTA1 and HDAC3 interacts with SMRT through their SANT domains. Both these interactions involve an inositol phosphate (IP) molecule located between the two interacting proteins. Functional studies have shown that this IP molecule plays an important role in the regulation of the deacetylase activity of the complexes (Watson et al. 2012; Millard et al. 2013; Watson et al. 2016). This recruitment mechanism of HDAC enzymes through a SANT domain involving an inositol phosphate molecule appears to be a shared feature among Class I HDAC complexes (Watson et al. 2016). Sin3a represent the only exception within the Class since it does not contain a SANT domain and it appears not to be regulated by the variation of inositol phosphate levels.

To date, the structure of the interaction between Sin3a and HDAC1 is still uncharacterised. Previous purification trials performed in our laboratory have shown that the presence of SDS3 in the Sin3a : HDAC1 binary complex stabilises the complex leading to higher purification yields (Portolano et al. 2014). Functional studies on SDS3 also revealed its importance for the activity of the complex and its deletion leads to a cell-lethal condition (David et al. 2003).

In this study, a variety of different methods have been used to achieve a better understanding of the structural arrangement of the ternary complex Sin3a : HDAC1 : SDS3. First of all, the size of the complex has been analysed using SEC-MALS (Size Exclusion Chromatography coupled with Multi-Angle Light Scattering). Subsequently, a low-resolution envelope was acquired by SAXS (Small Angle X-Ray Scattering) and

details about the inner organisation of the complex components were analysed by XL-MS (cross-linking coupled with mass spectrometry). Finally, a structural model of the Sin3a complex was built by combining all these data as well as a negative stain electron microscopy analysis.

Understanding the structural arrangement of the Sin3a complex will help to elucidate its mechanism of action and will help to better comprehend Class I HDAC complexes and their evolution.

5.2. Confirmation that the Sin3a₅₃₁₋₇₇₂ : HDAC1 : SDS3 complex is a dimer

The Sin3a ternary complex used for structural studies contains Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL}. This specific complex was chosen because it appears to be the most stable during the expression and purification process (as shown in chapter 3). The full-length SDS3 contains an N-terminal coiled-coil dimerization domain. Therefore, it is expected that the Sin3a complex may assemble as a dimer (Alland et al. 2002). The calculated molecular weight of Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} as a monomer is 128 kDa, whereas the calculated molecular weight for the dimeric form is 256 kDa.

To confirm that the Sin3a complex is dimeric and to confirm its molecular weight, the complex was analysed with SEC-MALS (Size exclusion chromatography with multi-angle light scattering). The SEC-MALS experiments were performed as described in materials and methods. The Sin3a complex was analysed using a Superose 6 (10/300) GL column attached to the MALS (Multi-Angle Light Scattering) instrumentation, an 8-angle MALS light scattering detector Dawn® HELEOS® II and a differential refractive detector Optilab® T-rEX Wyatt Technology. The gel filtration buffer was kept on ice during the entire experimental procedure since the Sin3a complex quickly degrades at room temperature. The expected molecular mass for each complex component is specified in table 5.12.

Table 5.11: Predicted molecular weight of the components of the Sin3a complex.

Protein / Protein complex	Expected molecular weight (kDa)
Sin3a ₅₃₁₋₇₇₂ : HDAC1 _{FL} : SDS3 _{FL} dimer	256
Sin3a ₅₃₁₋₇₇₂ : HDAC1 _{FL} : SDS3 _{FL} monomer/ ternary complex	128
Sin3a ₅₃₁₋₇₇₂ : HDAC1 _{FL} binary complex	83
HDAC1 _{FL}	55
SDS3 _{FL}	45
Sin3a ₅₃₁₋₇₇₂	28

Interestingly, SEC-MALS analysis of the Sin3a complex produced two peaks (figure 5.1A). The SDS-PAGE (figure 5.1B) of this experiment shows that the first peak corresponds to the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} ternary complex, ranging between 256 and 165 kDa (Figure 5.1C). The second peak corresponds to a binary complex of Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} ranging between 80 and 60 kDa (Figure 5.1C). Therefore, the first peak corresponds to a range of molecular weight where the majority of the complex is a dimer, with some dissociation. The molecular weight detected in the second peak covers the range of expected molecular weight of the Sin3a₅₃₁₋₇₇₂ : HDAC1 monomer (80 kDa) with some complex dissociation.

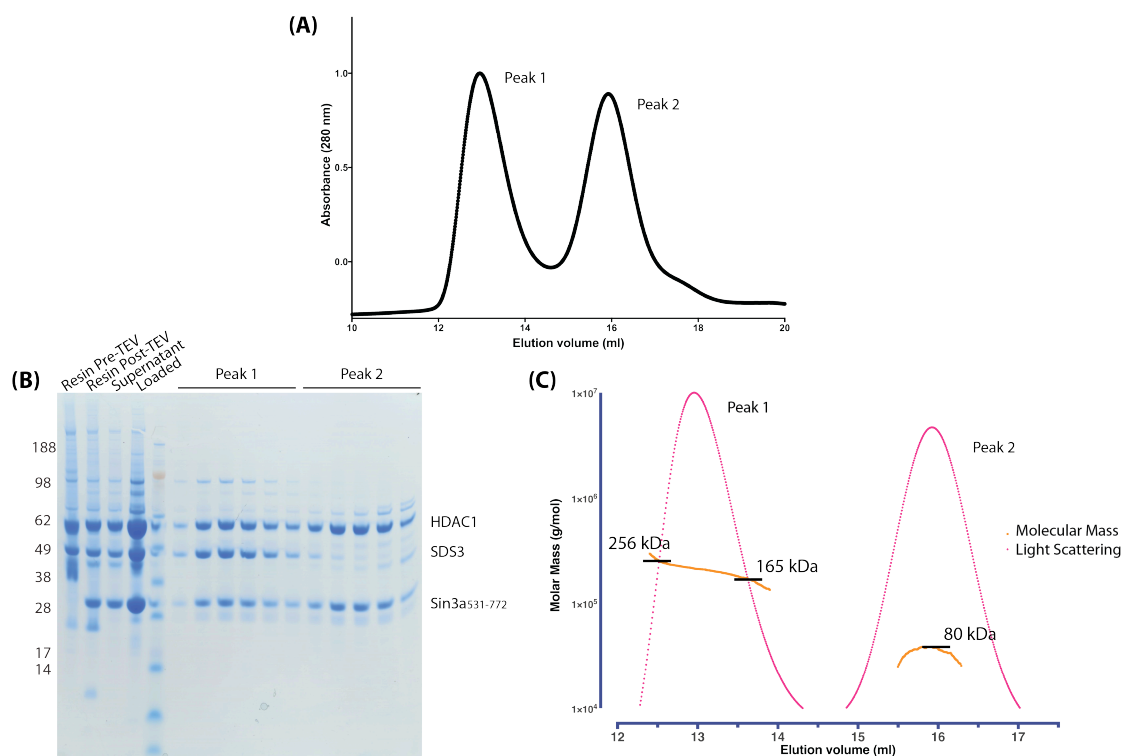


Figure 5.1: SEC-MALS analysis on the Sin3a complex.

(A) Chromatogram showing the A_{280} protein trace by size exclusion chromatography (Superose 6) of the SEC-MALS analysis of the Sin3a complex. (B) SDS-PAGE analysis of the two peaks from SEC-MALS analysis. The first peak includes all three components Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL}. The second peak includes the binary complex Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL}. (C) Light scattering curve and corresponding molecular mass of the first SEC-MALS analysis of the Sin3a complex.

To have a better understanding of the two peaks, the two central fractions from each peak were combined and reanalyzed. For the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex, the second SEC-MALS analysis produced only one peak (figure 5.2A). This peak has a longer tail compared with the first SEC-MALS run, suggesting higher dissociation of the complex, likely caused by the lower protein concentration loaded on the column. The SDS-PAGE analysis of the peak confirmed the presence of all three components Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS_{FL} at the beginning of the peak (Figure 5.2B) and the molecular weight of the peak ranges from 258 to 184 kDa (Figure 5.2C).

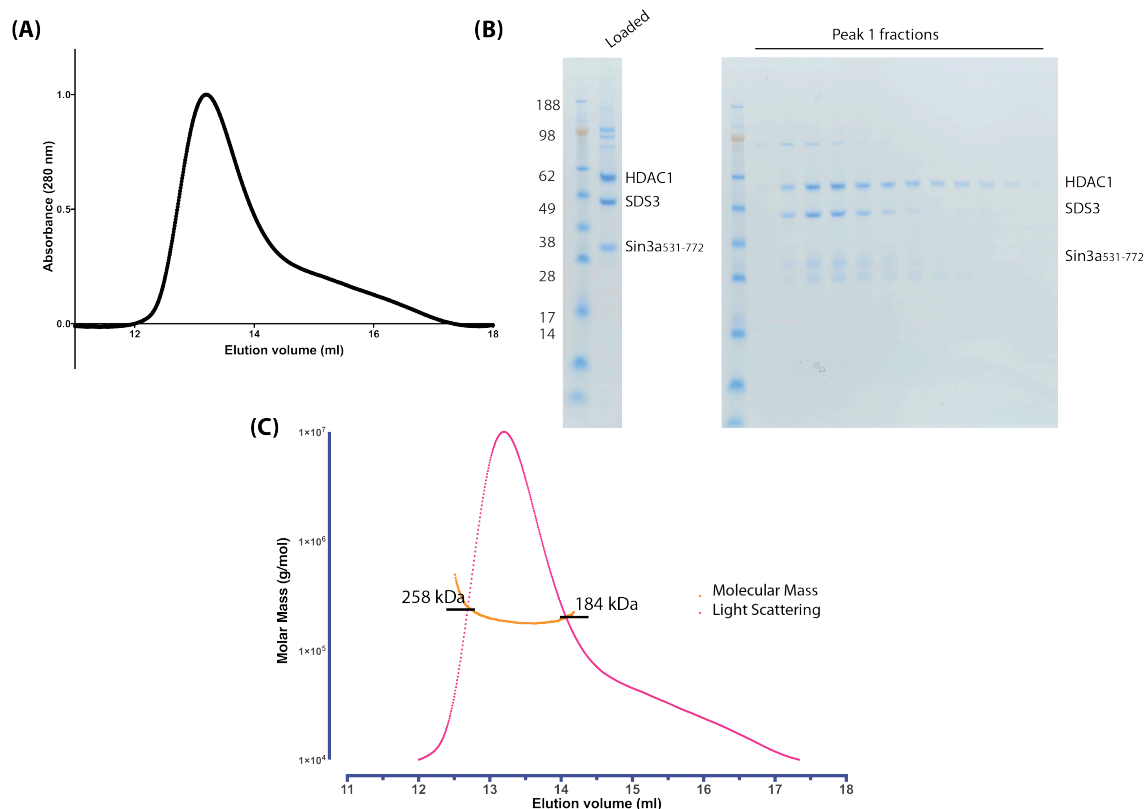


Figure 5.2: SEC-MALS analysis on the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.

Two central fractions from the first peak from the SEC-MALS analysis of the Sin3a complex are analysed with SEC-MALS for a second time. (A) A_{280} protein trace from the size exclusion chromatography (Superose 6) of the SEC-MALS analysis for the first peak. (B) SDS-PAGE analysis of the SEC-MALS peak. The peak contains the three complex components Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL}. An additional fourth band appears underneath Sin3a₅₃₁₋₇₇₂. (C) Light scattering curve and corresponding molecular mass of Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL}. The calculated molecular weight, ranging between 258 and 184 kDa, includes both dimeric form and monomeric form of the ternary complex Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL}.

The second analysis performed on the first peak confirmed the results found in the first round of analysis. The Sin3a ternary complex appears to assemble as a dimer but it tends to dissociate. Hence the Sin3a complex is unstable at room temperature on a gel filtration column despite keeping the gel filtration buffer on ice throughout the experiment.

In addition to the dissociation of the complex, a fourth band appears underneath the band corresponding to Sin3a₅₃₁₋₇₇₂, which probably corresponds to a proteolytic fragment of Sin3a. A sample from this band was sent for further mass-spectrometry analysis (paragraph 5.2.1).

A second SEC-MALS analysis was also performed on the second peak containing the binary complex Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL}. This second round of analysis produced one peak at the same elution volume as the first run (Figure 5.3A). SDS-PAGE analysis confirmed the presence of the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} complex in the peak (Figure 5.3B) with a molecular weight range between 80 and 58 kDa (Figure 5.3C).

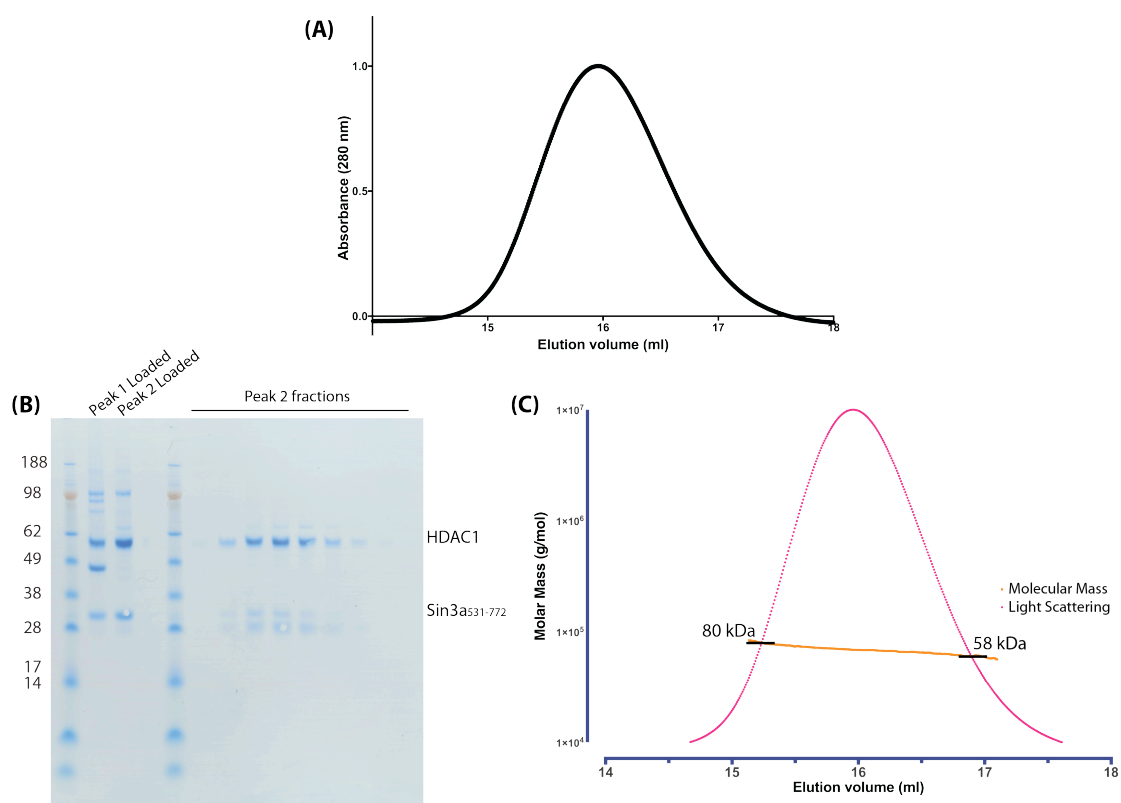


Figure 5.3: SEC-MALS analysis on the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} complex.

Two central fractions from the second peak corresponding to the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} complex are analysed with a second SEC-MALS run. (A) A₂₈₀ protein trace from the size exclusion chromatography (Superose 6) of the SEC-MALS analysis for the second peak. (B) SDS-PAGE analysis of the peak fractions. The peak contains the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} complex. An additional third band appears underneath Sin3a₅₃₁₋₇₇₂. (C) Light scattering curve and molecular mass of the second SEC-MALS analysis of the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} complex.

With the second SEC-MALS analysis of peak 2, it was confirmed that, in the absence of SDS3, the binary complex Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} forms a monomeric complex.

Furthermore, an additional band appeared underneath the band corresponding to Sin3a₅₃₁₋₇₇₂. A sample from this new protein band was sent for further mass-spectrometry analysis and the results are discussed in the next paragraph.

5.2.2. Mass spectrometry analysis

From the SDS-PAGE following the second SEC-MALS analyses of both the ternary and binary complex (figure 5.2B and 5.3B), a fourth band appears beneath the band corresponding to Sin3a₅₃₁₋₇₇₂. This could be a degradation form of the Sin3a construct. To confirm this, samples from this band and from the Sin3a₅₃₁₋₇₇₂ protein band were sent for mass-spectrometry analysis. The mass-spectrometry results confirmed that the additional protein band corresponds to a shorter (degraded) construct of Sin3a.

By mapping the peptides onto the sequence of Sin3a, more information about the Sin3a domain interacting with HDAC1 was obtained. This was used in combination with the secondary structure prediction to design a new Sin3a construct (figure 5.4). The new Sin3a construct includes amino acids 601-737, it is shorter at both the N and C termini compared to the previous construct, and it consists of five predicted alpha helices.

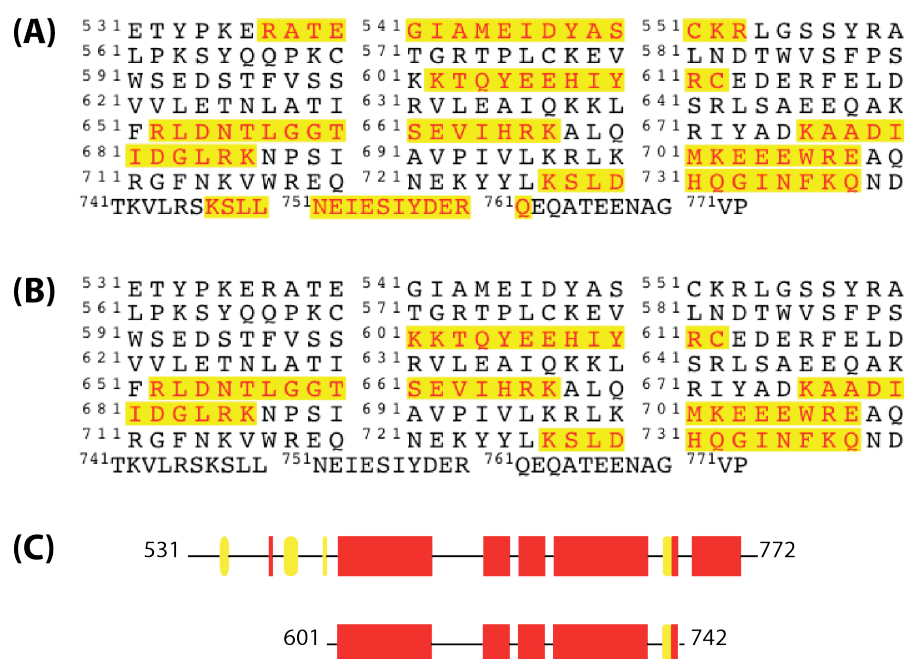


Figure 5.4: **Mass spectrometry analysis of Sin3a₅₃₁₋₇₇₂ and the degraded Sin3a.**

The peptides found by mass spectrometry analysis were mapped for both (A) Sin3a₅₃₁₋₇₇₂ that was used for the SEC-MALS experiment and for (B) the degraded Sin3a band. (C) The secondary structure prediction for Sin3a₅₃₁₋₇₇₂ was taken in consideration to identify boundaries for a new Sin3a construct. Alpha helices are shown as red boxes and beta sheets are shown as yellow boxes.

A search of the number of protease sites was done using the ExPASy PeptideCutter server to analyze what caused the degradation of Sin3a. A total of 570 protease sites were found in the Sin3a₅₃₁₋₇₄₂ and they appear to be equally distributed along the length of the protein. Since the distribution of protease sites is homogenous along the construct, the degradation of the termini of the construct is probably due to the fact that both the N- and the C-termini interact with HDAC1 either weakly or not at all.

From these results, it seems likely that the region comprising the amino acids 601-737 is sufficient to enable Sin3a to interact with HDAC1 and to form a stable interaction in the ternary complex.

5.3. Crystallization trials on the Sin3a complex

To achieve the main goal of this project and to solve the three-dimensional structure of the core complex of Sin3a, it would be ideal if it were possible to obtain diffraction-quality crystals of the complex. X-ray crystallography allows the determination of protein structure at atomic resolution and hence to understand the interactions between complex components. Unfortunately, this technique is as challenging as it is powerful.

Since proteins are delicate and the process of growing crystals must happen very slowly. To enhance the probability of growing crystals, a highly concentrated protein sample is mixed with a variety of reagents that should help to reduce the protein solubility near to spontaneous precipitation. When the right nucleation conditions are met, the protein sample should be able to crystallize (Figure 5.5A).

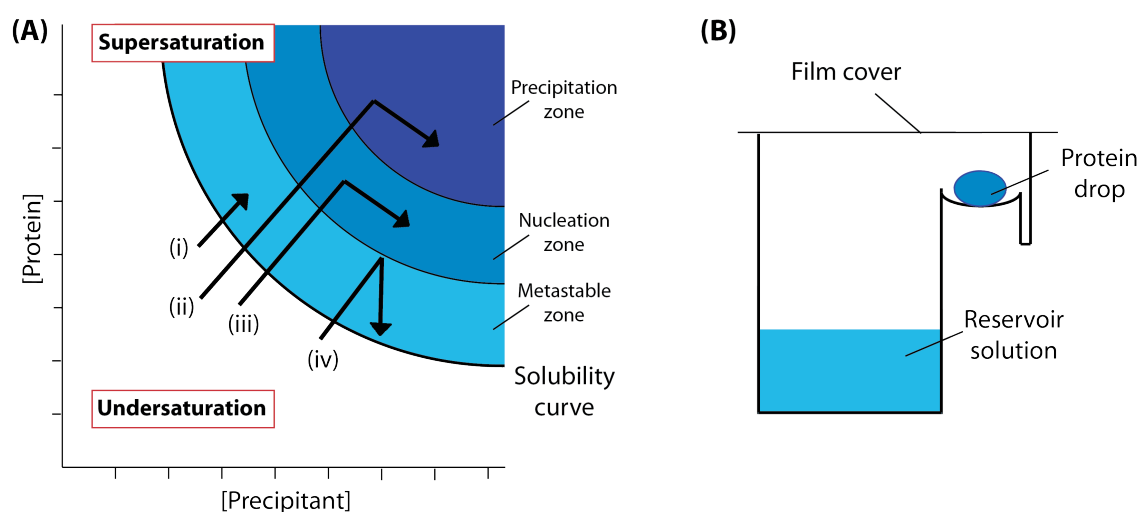


Figure 5.5: **Crystallization phase diagram and sitting drop crystallization method.**

(A) Schematic illustration of a crystallization phase diagram, precipitant agent and protein concentrations are respectively on the X and Y axes. Depending on the protein sample the different regions can have variable shapes and sizes. (B) Schematic representation of a protein drop in relation to the reservoir solution.

Many different factors can determine whether protein crystallisation is successful. The protein sample must be homogenous, stable and highly pure (>90%). Additionally, the proteins contained in the sample must precipitate in an ordered manner so they adopt an ordered 3D lattice. If all these conditions are satisfied, the protein crystal will be able to diffract X-rays and create a scattering pattern from which it is possible to determine the protein structure.

The successful creation of crystals is also dependent on the choice of precipitation reagents. Nowadays there are several commercially available crystallization screens that cover a wide range of pH, precipitant reagents, salt concentrations and additives. If a crystal can be generated by any of these conditions, the conditions can be optimised to grow crystals that are more suitable for structural analysis.

All crystallization trials attempted for this work were set up in sitting drop vapour diffusion plates (Figure 5.5B). Protein and precipitant mixture were mixed in equal volumes (100nl protein + 100nl precipitant to produce a 200nl drop) in a chamber next to the reservoir solution. In this way the two chambers will slowly reach equilibrium by diffusion.

The Sin3a complex was produced as described in chapter 3. To purify the complex for the crystallization trials both gel filtration and sucrose gradient purifications were tested to verify if a different purification method could enhance the protein complex stability. The protein sample was concentrated to various concentrations between 2 and 9 mg/ml and dispensed with a Mosquito LCP crystallization robot (TTP Labtech). Initial trials were performed using the commercially available screens JCSG⁺, ProPlexTM, PACT and MORPHEUS. The crystallization plates were incubated at both room temperature and 4°C .

An average of 60% of the crystallization conditions did not produce any type of precipitation, less than 10% resulted in phase separation while the remaining 30% generated light and amorphous precipitation (Figure 5.6).

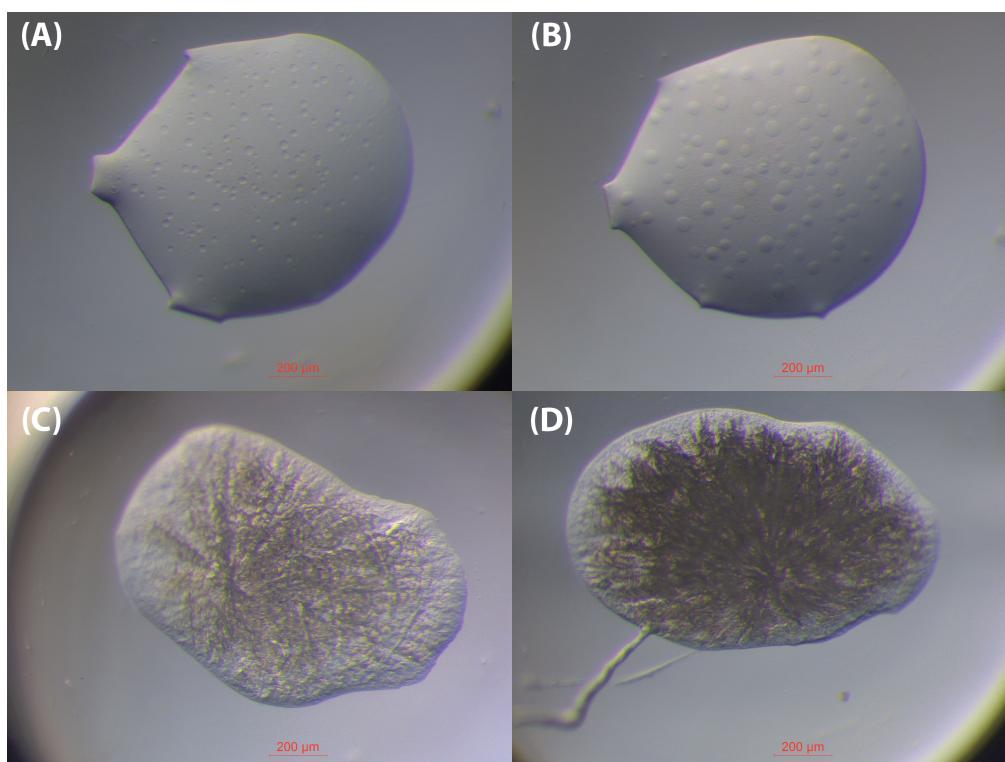


Figure 5.6: Images from Sin3a crystal trials.

(A) and (B) show phase separation, the round shapes are protein aggregates that are packed together but not in a crystal lattice. This result could be caused by the wrong temperature, pH or choice of reagents. (C) and (D) show amorphous precipitate formed during the crystallization trials probably due to too high protein and/or precipitant concentration, which lead to the protein forming large aggregates.

Unfortunately, the crystallization trials attempted only produced amorphous precipitation and no hits could be observed across the multiple conditions on the plates. The lack of hits could be addressed to several factors. By SEC-MALS analyses, the protein complex was shown to be unstable through variation of temperature and furthermore the complex tends to degrade over time (a non-homogeneous sample could impede the crystallization process). Furthermore, protein flexibility within the complex could hinder the crystallization process. As well as the sample characteristics, the lack of crystals could be caused by the use of non-optimal precipitant agents and additives.

Since none of the tested conditions provided any hint of crystallization, no further optimizations were performed. The Sin3a complex could have a high degree of flexibility. Indeed HDAC1 possesses an unstructured C-terminal tail of 106 amino acids (Millard et al. 2013). Furthermore, the Sin3a dimer assembles through the SDS3 dimerization domain, and the structure of the dimer could be flexible. The percentage of unstructured part of the complex and its overall flexibility could hinder the crystallization process and impede crystal growth. Therefore, since crystallizing the complex proved to be very challenging and time consuming, it was decided to adopt different approaches to gain structural information about the complex.

5.4. Mapping interactions between HDAC1, Sin3a and SDS3

In order to get a better insight into the interactions that occur between the components of the Sin3a complex, an approach was used that combines cross-linking and mass spectrometry LC-MS/MS (Holding 2015). This method, called XL-MS, involves cross-linking a protein complex and analysing it using mass spectrometry.

For this study, two different cross-linkers, CBDPSS (Cyanur Biotin Dimercapto Propionyl Sulfo Succinimide) and DSSG (DiSulfo Succinimidyl Glutarate) were used (figure 5.7). Both the cross-linkers are water-soluble, homobifunctional and isotopically-coded. With their activated esters they react mainly with primary amino groups to form a

stable amine bond. CBDPSS has a spacer arm length of 16 Å, whereas DSSG has a slightly shorter spacer arm length of 10 Å. When the lengths of two lysine side chains (6.5 Å each) are added to the overall length, the maximum inter C α distance for CBDPSS is 27 Å and for DSSG 23.7 Å.

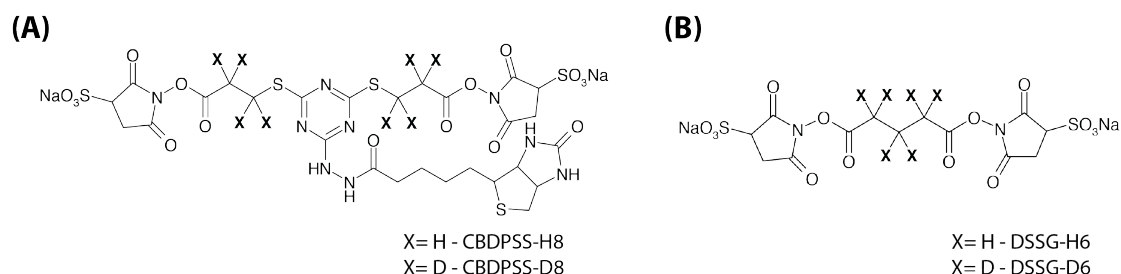


Figure 5.7: Homobifunctional cross-linkers used for the XL-MS experiments.

Chemical structure of the (A) CBDPSS-H8/D8 and (B) DSSG-H6/D6 molecules. Both CBDPSS and DSSG are water-soluble, homobifunctional and isotopically-coded. The cross-linkers differ in their spacer arm length, 16 Å for CBDPSS and 10 Å for DSSG.

The cross-linkers can fix transient interactions that occur between two proteins. When the purified protein complex is exposed to the bifunctional cross-linker, both ends of the cross-linker react with a primary amine of lysine side chains and link two lysine residues that are in close proximity to each other. Subsequently, the cross-linked protein complex is digested, typically using trypsin, and then analysed by mass spectrometry.

Deuterated cross-linker reagents were used to identify the linked peptides. A mix of light and heavy cross-linkers mix at 1:1 ratio was used in the cross-linking process. In this way, the linked peptides would appear as doublets in the mass-spectrometry analysis (Leitner et al. 2012). The cross-linked peptides and their position are identified, and the resulting location of the cross-linked residues can be used to create a structural model of the protein complex.

5.4.1. Cross-linking analysis of the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex

The first XL-MS experiment on the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex was carried out with the CBDPSS cross-linker. For this experiment 3 µg of Sin3a complex was

incubated with increasing amount of cross-linker (0-1.5 mM). All the cross-linking experiments followed the protocol described in material and methods. Figure 5.8 shows the SDS-PAGE of the cross-linking experiment with the CBDPSS cross-linker.

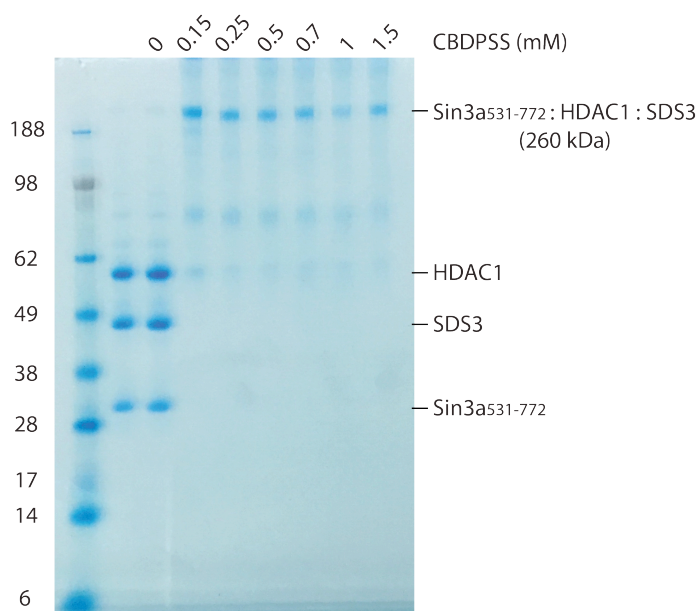


Figure 5.8: **Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} cross-linked with CBDPSS.**

SDS-PAGE analysis, increasing amount of CBDPSS cross-linker (0-1.5 mM) are used to cross-link the Sin3a ternary complex.

The CBDPSS cross-linker successfully cross-linked the ternary complex. Even at the lowest concentration of cross-linker (0.15 mM), the three individual protein bands disappear and in their place a band corresponding to the expected size of the Sin3a complex (260 kDa) appears. This band was excised and analyzed by mass-spectrometry.

The XL-MS experiments and data analysis were performed with the help of Dr. Andrew Bottrill and Dr. Sidhu Khushwant. The data were analysed in X-Quest ID score. Table 5.13 shows the crosslinks found between different proteins (inter-crosslinks) and table 5.14 shows the crosslinks found within the same protein (intra-crosslinks). The cross-links that include residues from the flexible tail of HDAC1 (residues 377-482) were discarded. Only the cross-links with a X-Quest Id-Score above 14.0 were considered valid.

Table 5.12: Inter-molecular crosslinks from the CDBPSS analysis of the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex. Cross-links found between residues of different proteins. The residues that belong to structurally known regions of the Sin3a complex are shown in bold. Residues that were confirmed to be in close proximity by the known structures are shown in green, residues that were not confirmed are shown in orange.

Protein 1	Residue No.	Protein 2	Residue No.	ID score	Confirmed by known structures
Sin3a	569	SDS3	130	14.4	
Sin3a	650	SDS3	307	20	
Sin3a	702	SDS3	266	15	
Sin3a	563	SDS3	271	19.9	
Sin3a	569	SDS3	225	20.4	
Sin3a	569	SDS3	271	17.9	
Sin3a	569	SDS3	81	17	
Sin3a	569	SDS3	133	16.6	
Sin3a	569	SDS3	101	15	
Sin3a	602	SDS3	307	16.4	
Sin3a	639	SDS3	142	20.5	No
Sin3a	686	SDS3	225	31.2	Yes
Sin3a	715	SDS3	147	19.2	No
Sin3a	715	SDS3	301	19	
Sin3a	715	SDS3	225	17.9	No
Sin3a	737	SDS3	130	18	
Sin3a	742	SDS3	130	20.6	
Sin3a	747	SDS3	266	17.7	
HDAC1	165	Sin3a	702	17.5	No
HDAC1	220	Sin3a	563	33.4	
HDAC1	220	Sin3a	552	18.3	
HDAC1	279	Sin3a	552	14.1	
HDAC1	283	Sin3a	650	16	No
HDAC1	50	SDS3	107	14.5	
HDAC1	58	SDS3	157	14.6	
HDAC1	218	SDS3	266	17.2	
HDAC1	218	SDS3	225	22.5	
HDAC1	218	SDS3	280	25.8	
HDAC1	283	SDS3	100	15.4	

Table 5.13: Intra-molecular crosslinks from the CDPSS analysis of the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex. Cross-links found between residues of the same protein. The residues belonging to structurally known regions of the Sin3a complex are highlighted in bold, residues that were confirmed to be in close proximity by known structures are shown in green.

Protein	Residue No.	Residue No.	ID score	Confirmed by known structure
SDS3	76	301	17	
SDS3	101	81	28.9	
SDS3	130	266	14.5	
SDS3	137	81	14.8	
SDS3	142	266	21.5	
SDS3	157	307	16	
SDS3	159	266	16	
SDS3	198	225	14.9	
SDS3	271	307	38.5	
SDS3	280	142	21.3	
SDS3	280	301	17.1	
SDS3	280	157	15.8	
SDS3	307	142	16.1	
SDS3	307	266	14.6	
SDS3	307	142	18	
Sin3a	569	702	15.7	
Sin3a	686	650	41.5	Yes
Sin3a	747	742	15	

The cross-linked residues were mapped onto linear representations of the proteins. Figure 5.9 shows the map for the inter-crosslinks and figure 5.10 shows the map for the intra-molecular crosslinks.

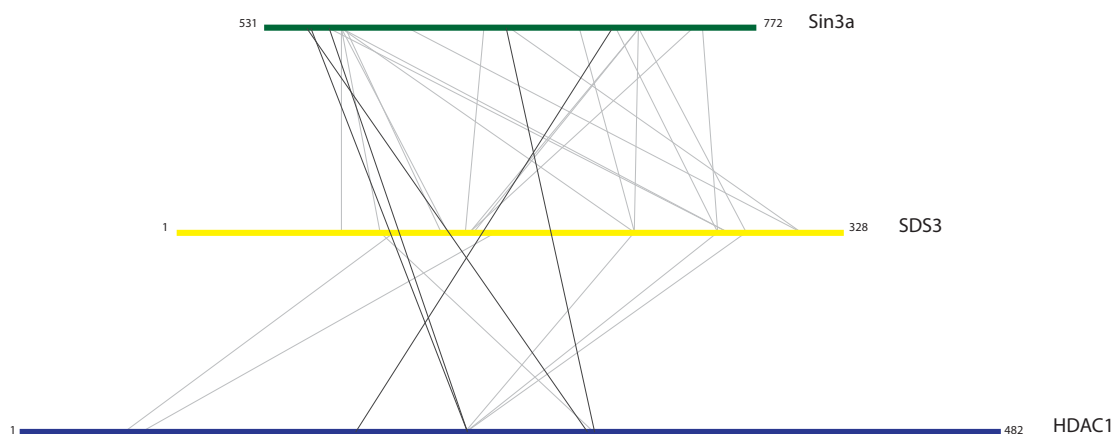


Figure 5.9: **Mapping schematic model created based on the CBDPSS inter-crosslinking results for the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.**

Sin3a is shown in green, SDS3 is in yellow, and HDAC1 is in blue. Each line represents a cross-link between two lysine residues on different proteins. The black lines in bold indicate the cross-links between HDAC1 and Sin3a.

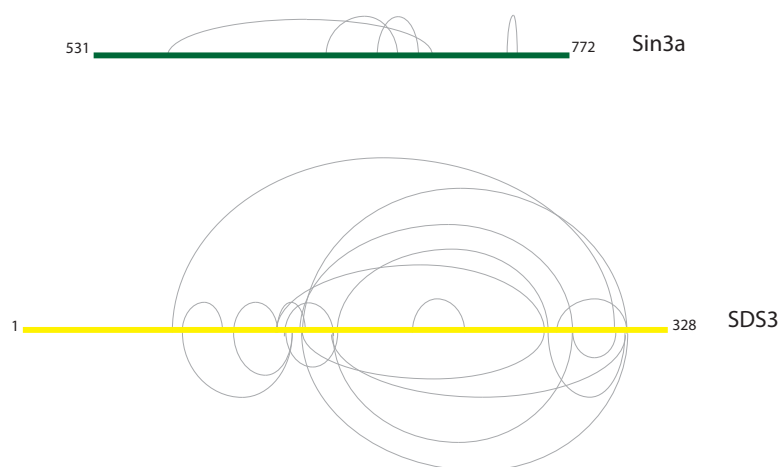


Figure 5.10: **Mapping schematic model created based on the CBDPSS intra-crosslinking results for the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.**

Sin3a is shown in green, SDS3 is in yellow. Each line represents a cross-link between two lysine residues on the same protein.

A second XL-MS experiment was carried out on the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex using the DSSG cross-linker. DSSG, with a spacer arm length of 10 Å, is shorter in length compared to the CBDPSS cross-linker. This should reveal closer range interactions than those found using CBDPSS. 3 µg of Sin3a complex was incubated with

increasing amount of DSSG cross-linker (0-1.5 mM). Figure 5.11 shows the SDS-PAGE of the cross-linking experiment carried out with DSSG.

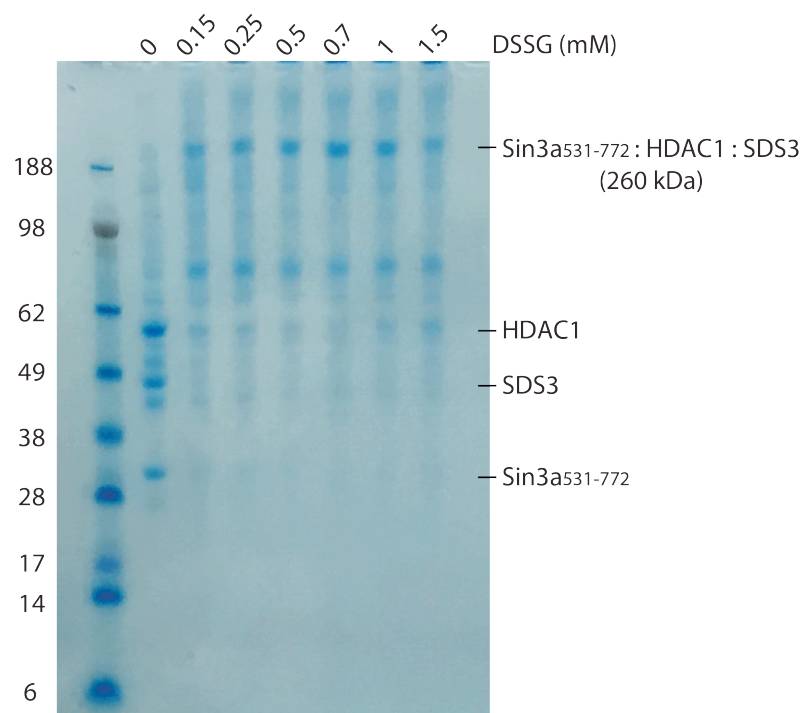


Figure 5.11: **Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} cross-linked with DSSG.**

SDS-PAGE analysis, the ternary complex is cross-linked with increasing amount of cross-linker DSSG. The appearance of a new band corresponding to 260 kDa confirms that DSSG successfully cross-link the Sin3a complex.

DSSG successfully cross-linked the ternary complex. A band corresponding to the expected size of the Sin3a complex (260 kDa) appears and at the same time the three bands corresponding to the Sin3a complex components disappear. Again, the 260 kDa band was excised and analyzed using mass-spectrometry. The crosslinks found by crosslinking the Sin3a complex with DSSG are shown in table 5.15.

Table 5.14: Table of inter/intra-molecular crosslinks from the DSSG cross-linked Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.

The residues that are included in structurally known region of the Sin3a complex are shown in bold. Residues that were confirmed to be in close proximity are highlighted in green.

Protein 1	Residue No.	Protein 2	Residue No.	Type of link	ID score	Confirmed by known structure
SDS3	97	SDS3	101	Intra	42.3	
SDS3	271	SDS3	307	Intra	41.2	
Sin3a	686	SDS3	225	inter	41.44	Yes

Figure 5.12 shows the mapped cross-links for the inter-molecular crosslinking and figure 5.13 shows the map for the intra-molecular crosslinks.

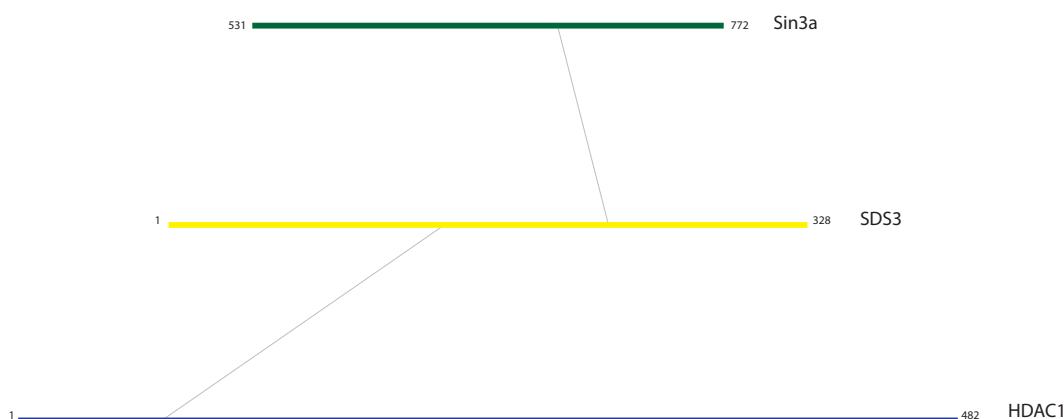


Figure 5.12: Schematic model of the DSSG inter-crosslinking results for the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.

Sin3a is shown in green, SDS3 is in yellow, and HDAC1 is in blue. Each line represents a cross-link between two lysine residues on different proteins.



Figure 5.13: Schematic model of the DSSG intra-crosslinking results for the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex.

SDS3 is shown in yellow. Each line represents a cross-link between two lysine residues.

The NMR structure of the binary complex Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄ that was solved by NMR in 2015 (Clark, Marcum, et al. 2015) was used to validate those cross-links that were present in this structure and found in the XL-MS experiments. For example, as shown in figure 5.14, both the intra-molecular crosslink between Sin3a₆₅₀ and Sin3a₆₈₆ (shown in magenta, in figure 5.14A) and the inter-molecular crosslink between Sin3a₆₈₆ and SDS3₂₂₅ (shown in magenta, in figure 5.14B) were in close proximity in the NMR structure.

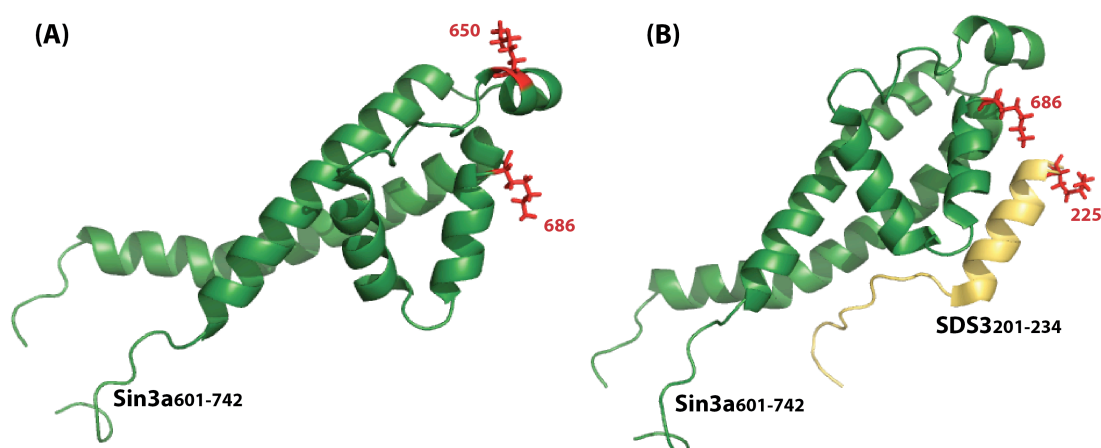


Figure 5.14: **Mapping the cross-linking results on the structure of the Sin3a and SDS3 (mouse).**

Sin3a is shown in green, SDS3 in yellow. (A) The residues Sin3a₆₅₀ and Sin3a₆₈₆ that are cross-linked by CBDPSS are shown in magenta. (B) The residues Sin3a₆₈₆ and SDS3₂₂₅ that are cross-linked together by CBDPSS are shown in magenta.

The fact that the crosslinks found within the known structure were validated increased the confidence in the other crosslinks and meant that a structural model based on the crosslinks and known structures could be generated with some confidence.

5.4.2. Structural modelling of the Sin3a complex

The known structures of the Sin3a complex components and the cross-links were used to create a model for the Sin3a complex. To create a model, Sin3a and HDAC1 were docked using HADDOCK (High Ambiguity Driven protein-protein Docking) docking software (Dominguez et al. 2003). Docking was performed and analyzed with the help of Dr. Cyril Dominguez. HADDOCK calculates various structural arrangements of the

protein complex starting from the pdb files and using constraints derived from the cross-linking analysis. The program assists the building of structural models by calculating the energy of the different arrangements based on van der Waals forces and electrostatic interactions. The crystal structure of HDAC1 (PDB: 4BKX) (Millard et al. 2013), and the NMR structure of Sin3a₆₀₁₋₇₄₂ (PDB: 2NHN) (Clark, Marcum, et al. 2015) were constrained using a distance of 16 Å (± 2) for the following crosslinks: Sin3a₇₀₂-HDAC1₁₆₅ and Sin3a₆₅₀-HDAC1₂₈₃. 200 models were generated independently and clustering of solutions, after water refinement, was set to an rmsd (root-mean-square deviation of atomic position) of 5 Å. This generated two clusters. The two clusters show small differences in the structural arrangement of the HDAC1 and Sin3a interaction (figure 5.15A). Cluster 1 is significantly more populated and shows lower average energy compared to Cluster 2 (table 5.16). The best 10 structures from Cluster 1 (Figure 5.15B) present very similar structural arrangement of the complex components, resulting in a lower average energy. The best structure generated by HADDOCK is shown in figure 5.15C. All the details regarding the two clusters are described in table 5.16.

Table 5.15: Finding the best model for docking Sin3a₆₀₁₋₇₃₇ and HDAC1.

Two Clusters were generated by HADDOCK software. Rmsd cut-off was set at 5 Å. Cluster 1 is more populated than Cluster 2 and shows lower average energy.

Cluster No.	Structures No.	Average energy (kcal/mol)	Standard deviation	Average energy 10 best structures (kcal/mol)	Standard deviation 10 best structures
1	164	- 213.0	± 54.7	- 306.4	± 32.3
2	32	- 139.2	± 27.5	- 134.5	± 22.6

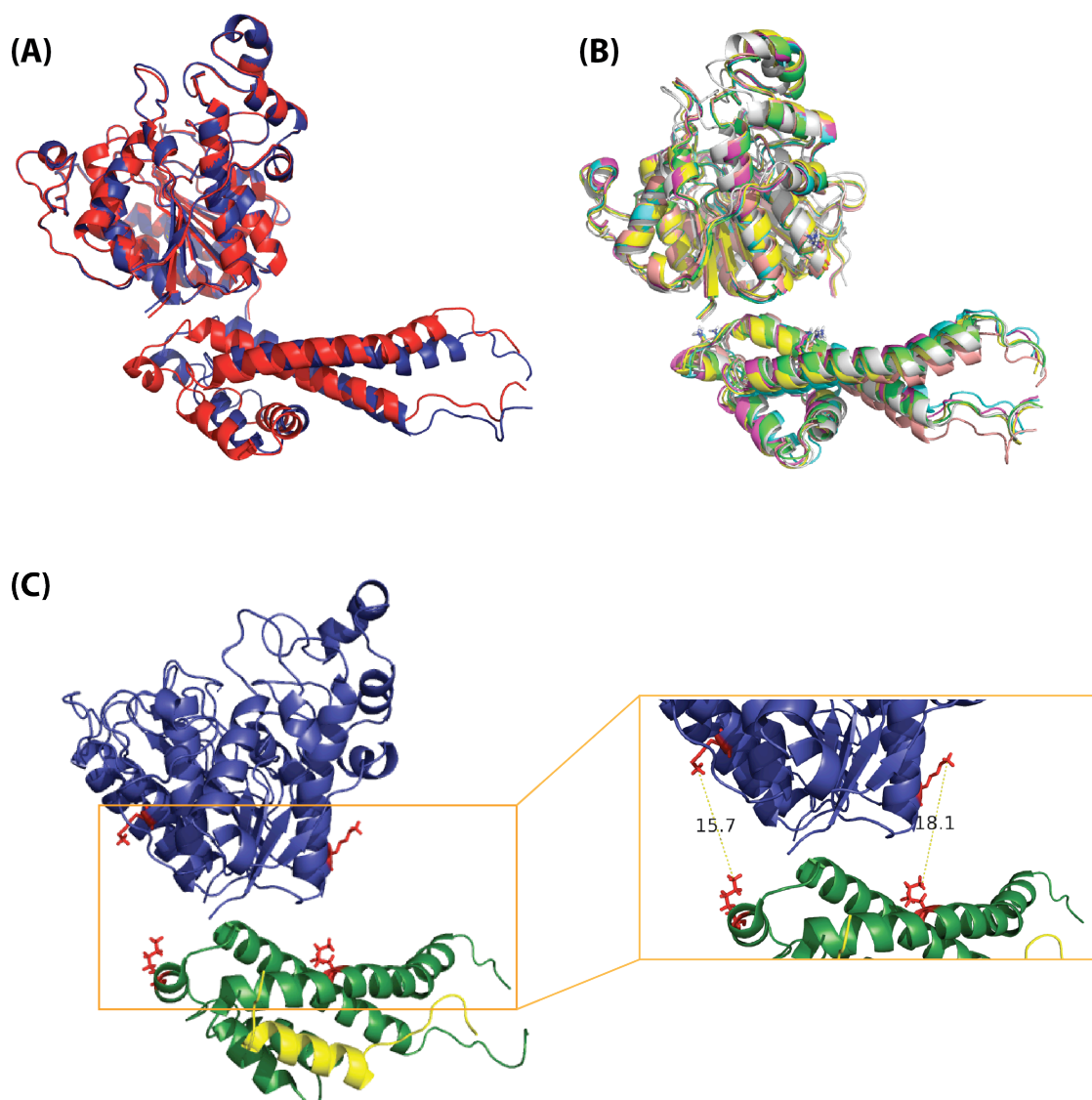


Figure 5.15: **Docking Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} with HADDOCK.**

(A) Comparison of the two clusters created with HADDOCK software. Cluster 1 is shown in red, Cluster 2 is shown in blue. (B) The 10 best models from Cluster 1, the overall arrangement of HDAC1 and Sin3a is similar in all models. (C) Best model with the lowest energy from Cluster 1. The residues that were modelling are shown in magenta.

Full length SDS3 is expected to dimerise (Lechner et al. 2000) and the dimeric form of the ternary complex was confirmed by SEC-MALS experiments. To create a dimeric model for the Sin3a complex, two Sin3a-SDS3 structures (PDB code: 2N2H. Clark, 2015) were manually positioned close to each other using the crosslink Sin3a₇₁₆-SDS3₂₂₅ (shown in magenta in figure 5.16A). The distance between the residues was set to the length of the CBDPSS cross-linker. The HDAC1 catalytic subunits were positioned on Sin3a as calculated using HADDOCK (figure 5.15C).

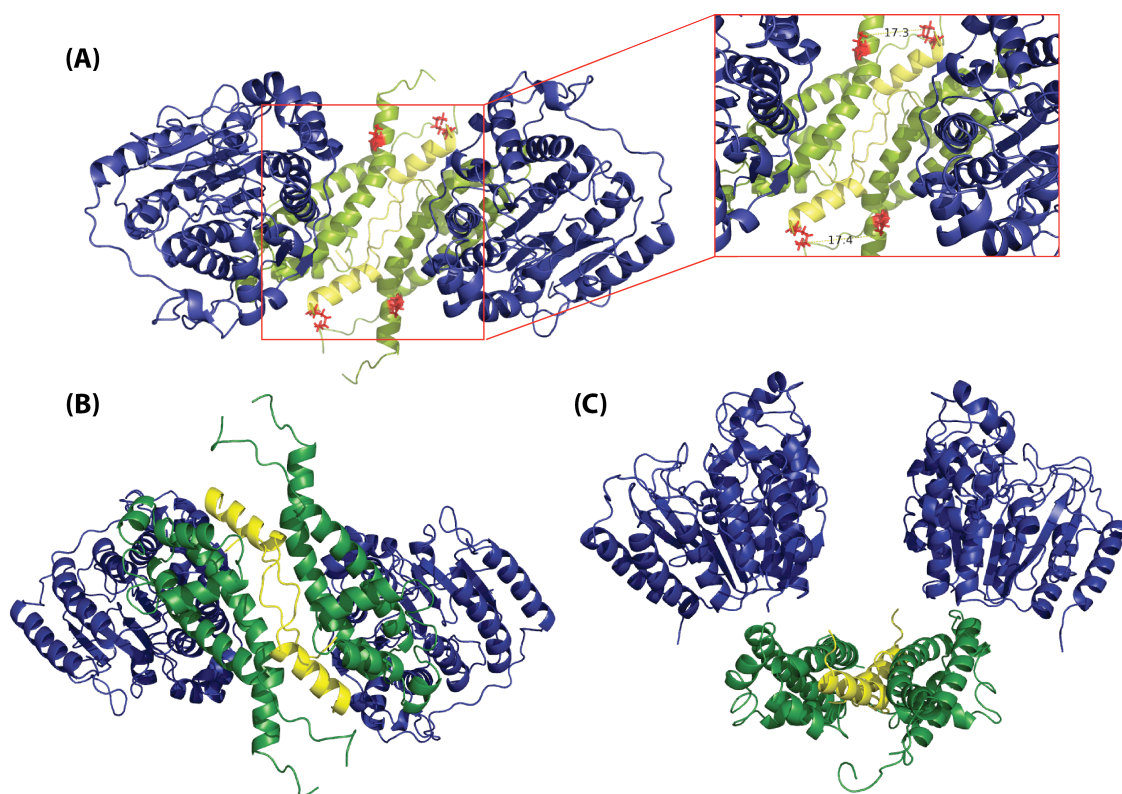


Figure 5.16: Model for the Sin3a complex based on the cross-linking experiments.

The Sin3a₆₀₁₋₇₃₇ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄ ternary complex was manually arranged as dimer following the coordinates of the crosslinks. HDAC1 is shown in blue, Sin3a is shown in green, SDS3 is shown in yellow. (A) Top view of the model, the crosslinks Sin3a₇₁₆-SDS3₂₂₅ are highlighted in magenta, the measurements between the residues are shown. (B) Bottom view of the cross-linking model. (C) Front view of the cross-linking model.

A model for the Sin3a dimeric complex was created using the HADDOCK docked HDAC1 Sin3a together with cross-linking results from the XL-MS analysis (Figure 5.16). In this model of the dimer, Sin3a and SDS3 create a bridge structure between the two HDAC subunits at the bottom of the complex. The two HDAC1 enzymes are instead located at the top of the structure.

Since it is still unclear whether the complex adopt a more open or closed conformation and the degree of flexibility of the complex remains unknown, SAXS (Small Angle Light Scattering) analysis was performed on the complex to obtain a low-resolution envelope of the complex and to gain a better understanding of the structural arrangement of Sin3a.

5.5. Using SAXS to study the structure of the Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} dimeric complex

In 1947 Guinier and Fournet developed the principle for a new method able to characterize macromolecular and nanostructured systems (Gurnier & Fournet 1947; Franke & Svergun 2009). Small Angle X-Ray Scattering (SAXS) has the advantage of being able to study samples in solution and can also analyze a large range of sample sizes (5kDa - 100MDa). In this study, SAXS analysis was used to obtain a low-resolution shape, also called an envelope, of the dimeric Sin3a complex.

The specific method used was SEC-SAXS since it combines Size Exclusion Chromatography (SEC) with SAXS analysis. The specific steps for the protocol used are described in the material and methods section.

In SEC-SAXS, the concentrated protein sample is loaded on a SEC column to reduce contaminations and/or aggregation in the sample (Figure 5.17A). The flow from the SEC column gets directed to the Small Angle X-Ray instrument where the protein sample is exposed to a monochromatic focused X-ray beam (Figure 5.17B). The scattered photons generated from the X-ray beam that hits the protein sample are then subtracted from a buffer blank to generate a scattering pattern. The resulting scattering pattern allows determination of a low-resolution envelope that corresponds to the volume occupied by the analysed protein.

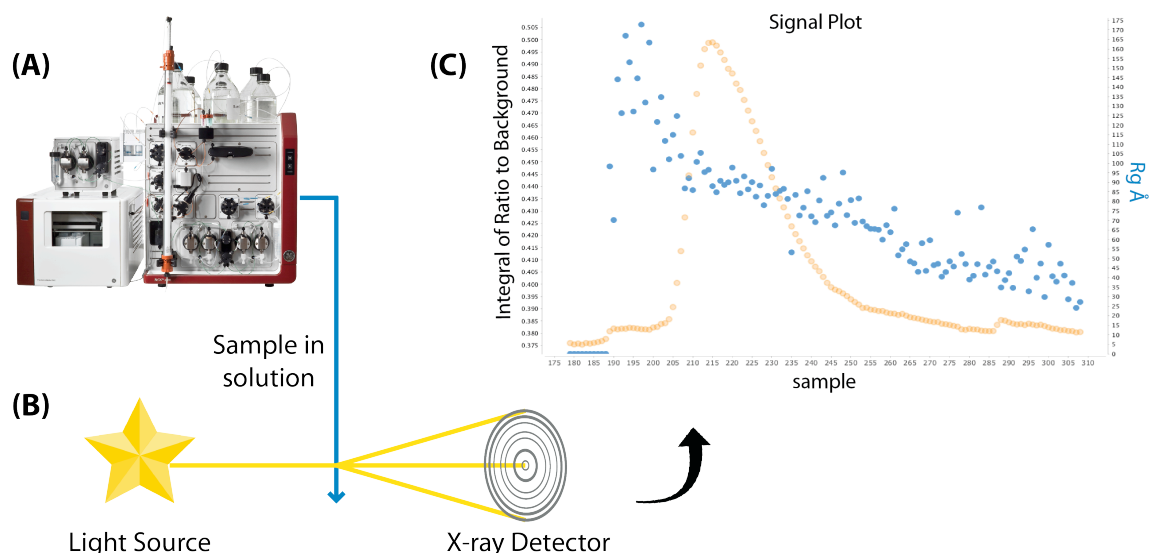


Figure 5.17: **SEC-SAXS workflow.**

(A) The protein sample is filtrated though a SEC column. (B) The protein in solution is directed to a X-ray beam that passes through the protein sample and generate a diffraction pattern. (C) The diffraction pattern generates a scattering curve, which will be analysed to obtain the protein envelope.

The Sin3a complex comprising Sin3a₅₃₁₋₇₄₂ : HDAC1_{FL} : SDS3_{FL} was expressed in HEK 293F and purified as described in materials and methods. Data were collected at Diamond Light Source B21 SAXS beamline, Didcot, Oxford. The complex was loaded at a final concentration of 10.5 mg/ml to a Shodex 404 SEC column coupled to the SAXS beamline. Since the Sin3a complex quickly degrades at room temperature, the temperature of the experiment was reduced by keeping the gel filtration buffer on ice during the entire experimental procedure.

Data analysis was performed with ScÅtter (Biosis) and FoXS web service. Two sets of SAXS data were collected for the Sin3a complex. The first set of data, called “flow mode”, was collected by running the entire gel filtration analysis at the same flow rate (A_{280} trace for the flow mode is shown in figure 5.18A). The second set of data was collected using the “capture method”, where the flow rate was stopped for 15 minutes when the A_{280} absorbance on the peak corresponding to the Sin3a complex reached its highest point (A_{280} trace for the capture method is shown in figure 5.18B). Using the flow mode it is possible to collect 10 to 20 frames of the sample, which correspond to 30-60 seconds of SAXS data. The capture method allows higher resolution SAXS information to be collected by exposing the protein peak for a longer time.

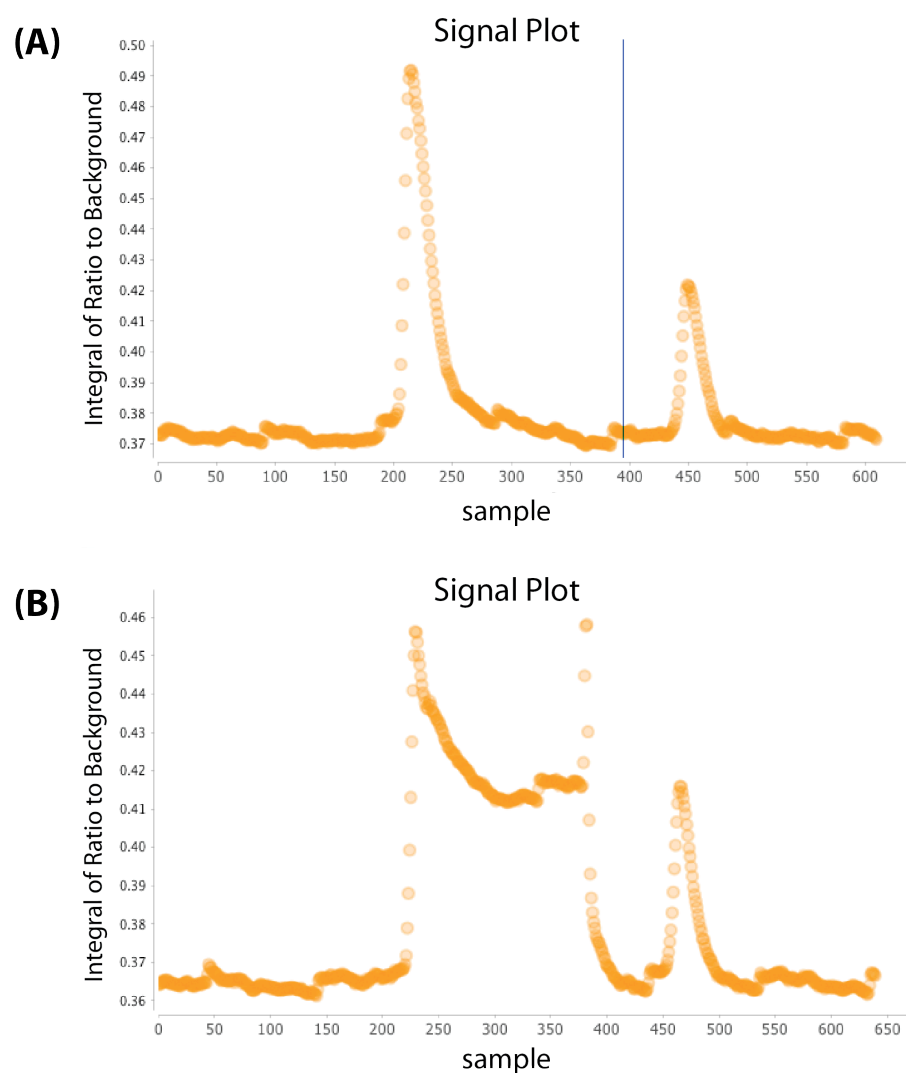


Figure 5.18: **SAXS analysis for the Sin3a complex.**

(A) A280 absorbance trace for the Sin3a complex when analysed with the "flow mode". (B) A280 absorbance trace for the Sin3a complex when analysed with the "captured mode", the protein peak was trapped for 15 minutes.

The SAXS envelope of the Sin3a complex was generated with ScÅtter (Biosis). The structures of HDAC1 (PDB: 4BKX, (Millard et al. 2013)), the Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄ complex and the SDS3₉₀₋₁₇₄ coil-coiled domain (PDB: 2N2H and 4ZQA, (Clark, Marcum, et al. 2015)) were used for data fitting and modelling.

The size of the envelope generated from the experimental curve confirmed that the Sin3a complex is a dimer and shows that the overall shape is an asymmetric "V" structure.

Two models were fitted in the SAXS envelope, one including only the ternary complex Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄ (shown in figure 5.19A) and the second one adding the SDS3₉₀₋₁₇₄ coil-coiled domain to the ternary complex based on the envelope shape (shown in figure 5.19C). The Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄ model created by HADDOCK (shown in figure 5.15C) was used for both SAXS envelopes.

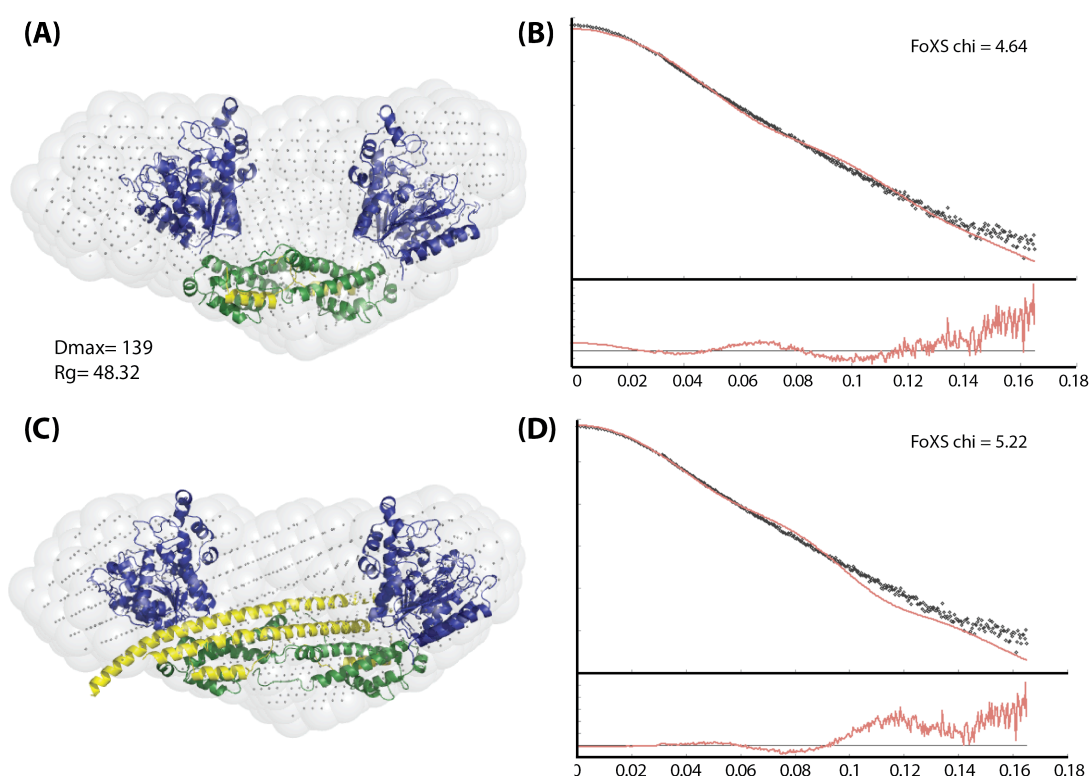


Figure 5.19: SAXS structural modelling of the Sin3a complex, SAXS envelope created with P1 symmetry.

HDAC1 is shown in blue, Sin3a in green and SDS3 in yellow. (A) The generated P1 SAXS envelope is fitted with the docked Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄ complex. (B) Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄ model is analysed with FoXS server, the experimental intensity is shown in black and the theoretical intensity from the fitted Sin3a model is shown in red. The intensities $I(q)$ are represented versus the momentum transfer q . The χ^2 value is 4.64. (C) SDS3₉₀₋₁₇₄ is added to the docked Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄ complex for the envelope fitting. (D) The Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₉₀₋₁₇₄ : SDS3₂₀₁₋₂₃₄ model is analysed with FoXS server, the experimental intensity is shown in black and the theoretical intensity is shown in red. The χ^2 value is 5.22.

With FoXS server the experimental data were compared to the models fitted in the SAXS envelope. To determine how closely the experimental data fits the theoretical one, Chi square (χ^2) calculation has been applied. When the experimental and

theoretical data coincide $\chi^2=0$, while the increasing of the χ^2 value determines the increasing of the magnitude of discrepancy between the two sets of data. The following formula is used to calculate Chi Square: $\chi^2 = \sum \frac{(O-E)^2}{E}$.

Where O= observed frequency and E= Expected or Theoretical Frequency.

In the data analysis of the SAXS for both models, the theoretical scattering curve calculated from the Sin3a models shows a poor agreement to the experimental scattering curve, with a Chi square (χ^2) value of 4.64 for the model including only the ternary complex Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄ (Figure 19B) and a $\chi^2= 5.22$ for the model including the SDS3₉₀₋₁₇₄ to the ternary complex (Figure 19D). The low-confidence fitting may be due to not having a complete model of the complex. From the Sin3a coordinated used 100 amino acids are missing, 294 amino acids are missing from SDS3 and 100 amino acids are missing from HDAC1.

The model with the coil-coiled domain of SDS3 shows a higher χ^2 value compared to the model without the coil-coiled domain. This is probably because the SDS3 coil-coiled domain does not fit completely within the SAXS envelope or that the coil-coiled was not properly placed. It is likely that the dimerization domain of SDS3 in the context of the full-length protein will have a different structure.

Overall, the model fitting produced a SAXS envelope that agrees with the cross-linking model, confirming the structure where Sin3a : SDS3 form a bridge below the two HDAC1 subunits located at the top of the structure.

5.6. Negative stain electron microscopy

To gain a better resolution image of the complex, the next step was to use negative stain Electron Microscopy (EM). Negative stain EM is used to visualize biological samples by increasing the contrast with the use of dried heavy metal salts (Brenner & Horne 1959).

The Sin3a₅₃₁₋₇₇₂ : HDAC1_{FL} : SDS3_{FL} complex was fixed with glutaraldehyde in a sucrose gradient (as described in materials and methods). Interestingly, two clear bands were separated by size from the cross-linked sucrose gradient fractions (shown in figure 5.20). The monomeric form of the Sin3a complex is expected to be approximately 130 kDa and the dimeric form 260 kDa. As fraction 10 and 15 seemed to correspond to the expected molecular weights, they were chosen for negative stain analysis (figure 5.20). The other bands appearing above fraction 15 were assumed to be aggregated forms of the complex.

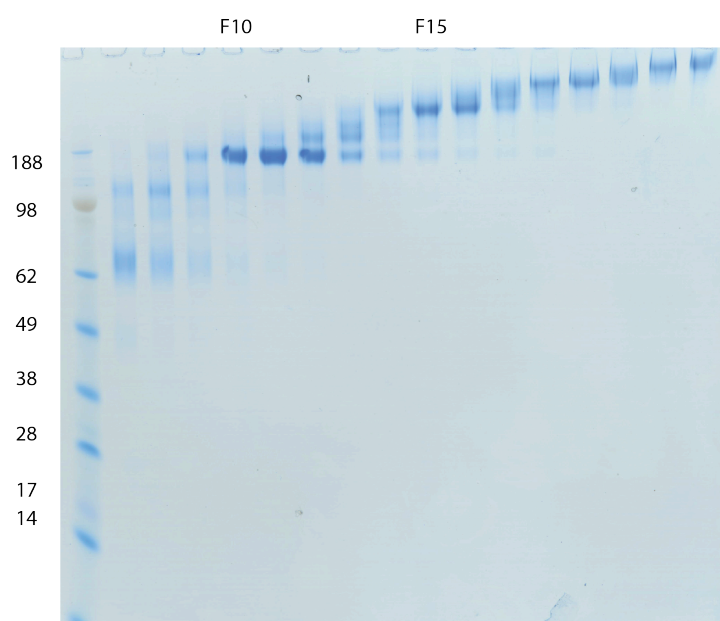


Figure 5.20: **SDS-PAGE analysis of the sucrose gradient fractions of the cross-linked Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄ complex.**

5%-25% sucrose gradient was used to purify the Sin3a₆₀₁₋₇₄₂ : HDAC1_{FL} : SDS3₂₀₁₋₂₃₄ complex, with a 0%-0.1% glutaraldehyde gradient. Fractions F10 and F15 were selected for negative stain analysis.

The grids were prepared and visualized with the help of Dr Nobuhiro Morone at the MRC toxicology unit. The conditions for preparing the grids were optimised by varying the protein concentration and the staining with 2% uranyl acetate. By visualizing the grids, it was shown that the Sin3a complex was in its dimeric conformation in both the samples, but the grid containing F15 was highly aggregated. Therefore, data collection

was performed only on fraction F10.

Data collection was performed using the negative stained grid containing the F10 fraction, using FalconTM II Direct Detector at the University of Leicester. Figure 5.21 shows a 400-mesh carbon film grid progressively scaling to show the detailed micrograph containing the negatively stained Sin3a complex.

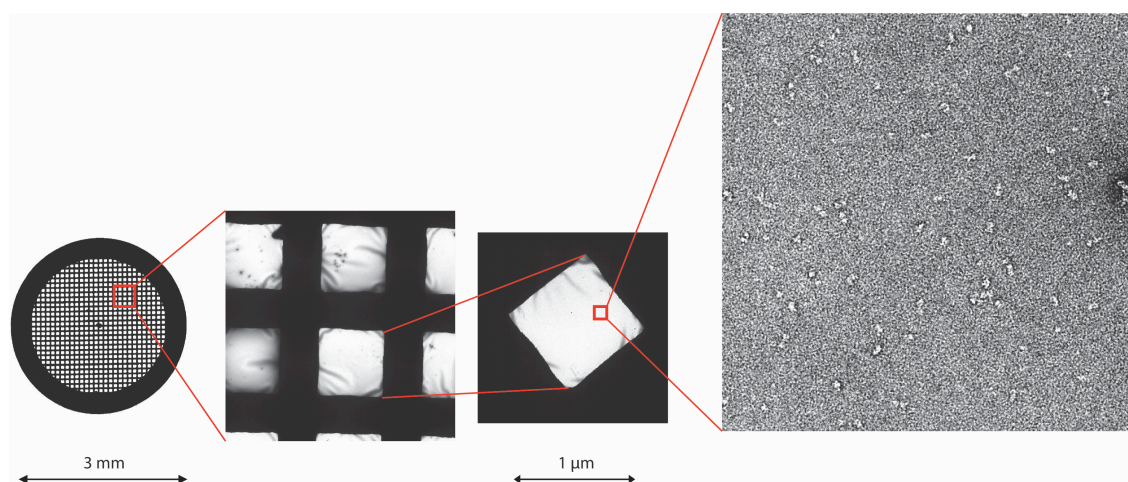


Figure 5.21: Close up of a carbon film 400 mesh grid.

Close up of a 400 mesh grid. The Sin3a complex in the picture on the right was negatively stained with 2% uranyl acetate.

The EM micrographs were analyzed using EMAN2 software. 525 micrographs were collected with a range of defocus values from -0.5 to -4 μm. A subset of 447 micrographs was selected for data analysis based upon their Thon ring profile, astigmatism and particle distribution.

10710 particles were manually picked from this subset of micrographs and they were used to create 100 2D class averages (one micrograph with manually picked particles is shown in figure 5.22A). Based upon the quality of the images contained in these 100 classes, 60 classes were selected (figure 5.22B) to create the first 3D initial model.

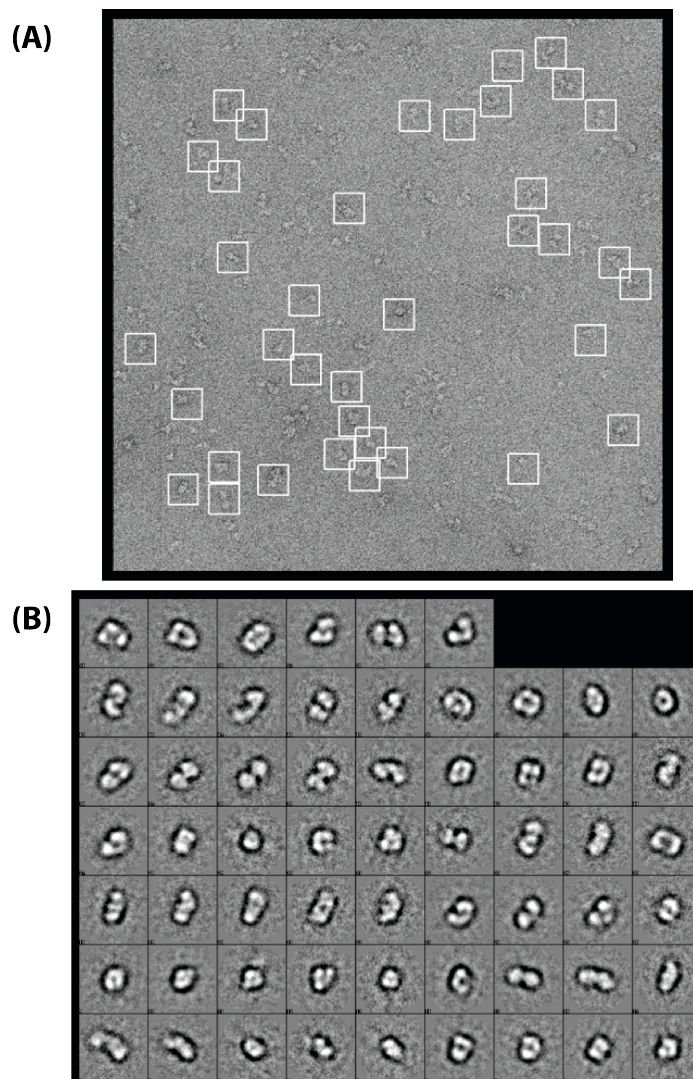


Figure 5.22: **Negative stain analysis of the Sin3a complex.**

(A) One micrograph showing the Sin3a complex that was stained with 2% uranyl acetate. Particles were picked manually with the EMAN2 software. (B) Refined class 2D averages created from the manually picked particles. 60 class 2D averages were selected to create 3D initial models.

Two initial 3D models were created from the 60 refined class 2D averages, one with C1 symmetry and one with C2 symmetry (Figure 5.23).

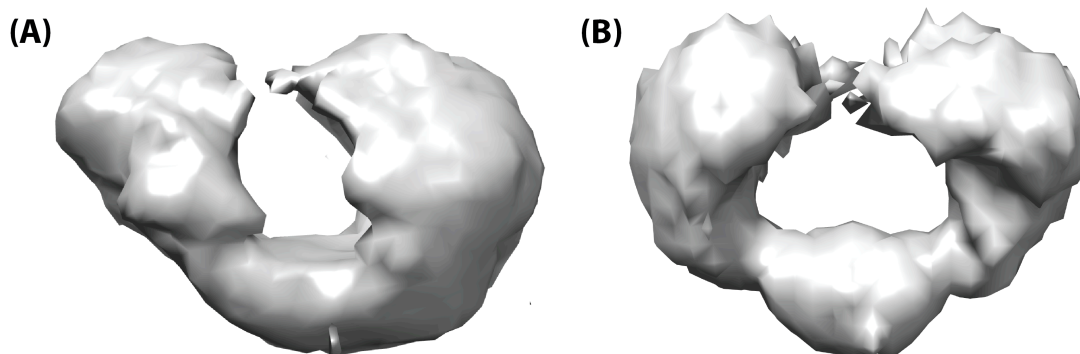


Figure 5.23: Initial 3D models for the negative stained Sin3a complex.

With EMAN2 software 3D models were created from the class 2D averages. (A) 3D model of the Sin3a complex created where no symmetry imposed to the model. (B) 3D model imposing C2 symmetry to the Sin3a complex.

The overall structural arrangement is similar between the two initial 3D models (either with C1 or C2 symmetries). Two globular structures at the top are linked by a bridge structure positioned at the bottom.

To create a structural model, the structures of HDAC1 (PDB: 4BKX, (Millard et al. 2013)), Sin3a₆₀₁₋₇₄₂ : SDS3₂₀₁₋₂₃₄ (PDB: 2N2H, (Clark, Marcum, et al. 2015)) and the SDS3₉₀₋₁₇₄ coil-coiled domain (PDB: 4ZQA, (Clark, Marcum, et al. 2015)) were manually fitted into the map.

The 3D initial map for the C2 model was refined with EMAN2 software to 21 Å. The best docking model (figure 5.15C) was used for model fitting. The two HDAC1 were located within the two globular structures at the top, while Sin3a and SDS3 were placed in the bridge structure at the bottom (figure 5.24).

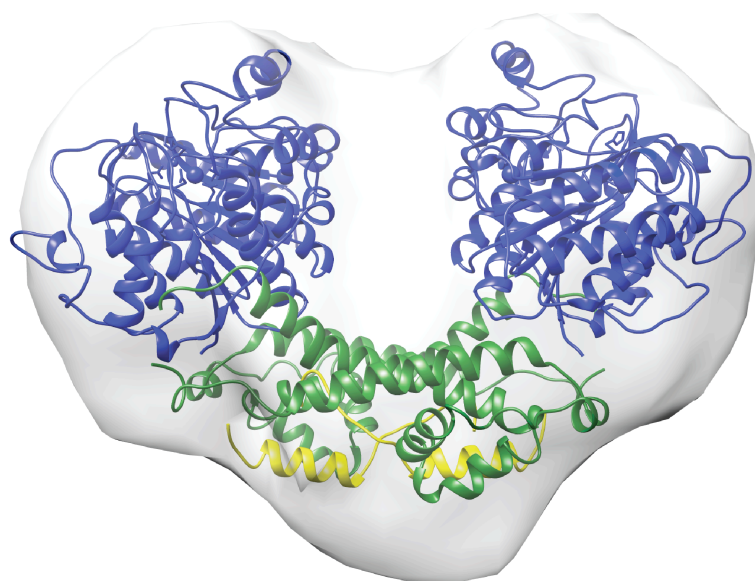


Figure 5.24: **Negative stain EM 3D model of the Sin3a complex with C2 symmetry.** 21 Å structure created with EMAN2 software from the class 2D averages from the analysis of fraction 10. Model fitting include the Sin3a₆₀₁₋₇₄₂ : HDAC1 : SDS3₂₀₁₋₂₃₄ structures. HDAC1 is shown in blue, Sin3a is shown in green and SDS3 is shown in yellow. The two HDAC1 subunits are located on top of the structure, while Sin3a and SDS3 act as a bridge at the bottom.

The two HDAC1s fit well in the two globular structures at the top of the model. Furthermore, both Sin3a and SDS3 fit in the bridge structure at the bottom. The SDS3 coil-coiled domain was not included in the model because it appears to be too big to fit into the envelope. This confirms the idea that the SDS3 coiled-coil might change conformation in the full-length construct when it interacts with its protein partners.

Overall, the negative stain EM results further validate the previous SAXS and cross-linking results by producing a map that corresponds to the structural arrangement calculated from the previous analyses.

5.7. Cryo - electron microscopy

At this stage, to take the structural analysis forward and to obtain a high-resolution structure of the complex, cryo-Electron Microscopy (cryo-EM) was used.

Cryo-EM grids were prepared with a Vitrobot™ Vitrification Robot (FEI) with the help of Dr Richard Collins at the University of Manchester. The conditions for making the grids

were optimized by varying the concentration of the protein complex and the blotting time. After screening the different conditions data collection was performed from the grid that was prepared with 0.05 mg/ml protein, blotted for 6 seconds at force 5. Data collection was performed at Diamond, Oxford. 3224 micrographs were collected with a range of defocus values from -0.5 to -4 μm . The collected data was analysed with Relion software. 1600 particles were manually picked and used to set the parameters for an automated particle-picking job (Figure 5.25A shows a micrograph where particles were manually picked). 193651 particles were automatically picked and figure 5.25B shows the 2D class averages calculated from the particles automatically picked.

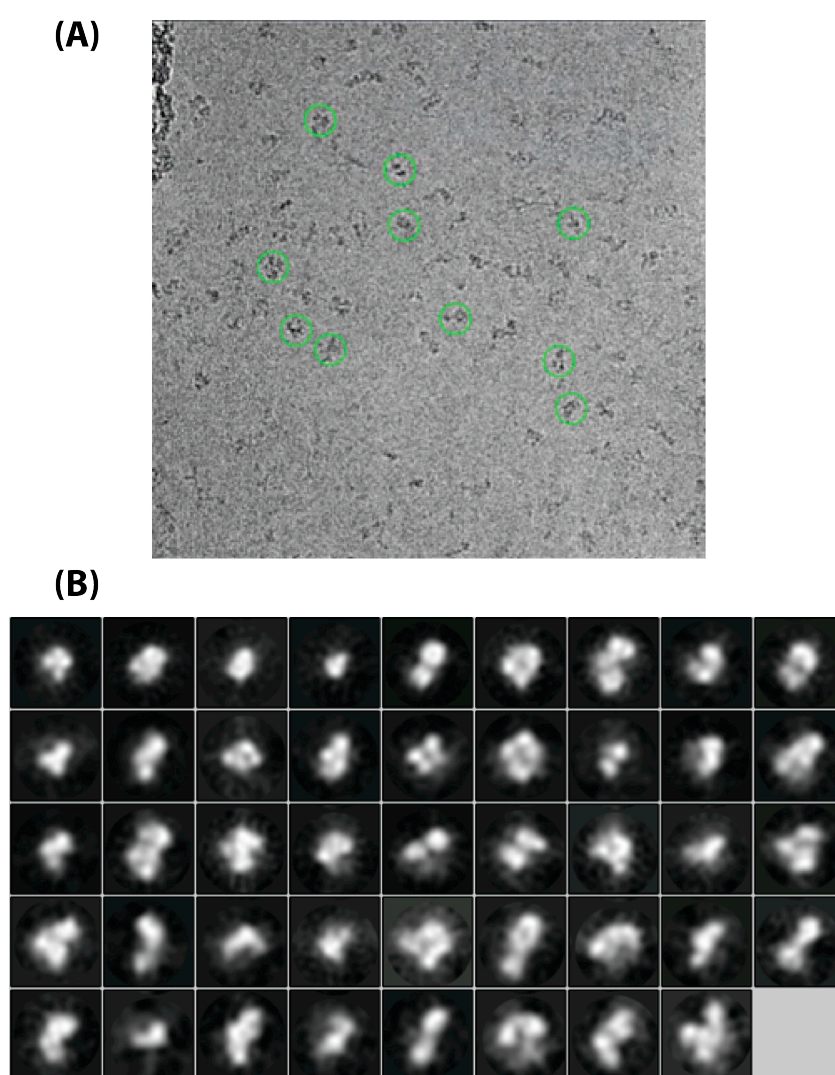


Figure 5.25: Cryo-EM experiment for the Sin3a complex.

Data analysis was performed with Relion software. (A) Cryo-em micrograph showing manually picked particles. (B) Class 2D averages from 193651 particles that were both manually and automatically picked.

The 2D class averages show that the Sin3a complex was aggregated. The micrographs contained aggregates and large particles, only few particles could fit in the green circle that defines the expected particle size. Most of the 2D class averages also show aggregated forms of the protein complex, with large and variable shapes.

For future cryo-EM experiments, it will be necessary to optimize the protein preparation and the conditions used to prepare the grid in order to avoid the aggregation problems.

5.8. Results summary and discussion

Integrative approaches using data from different methods were combined to create a model structure of the Sin3a core complex.

SEC-MALS analyses have shown the Sin3a complex to be a dimer when SDS3 is included in the complex. Previous structural studies had already shown that the presence of SDS3 in the complex leads to an overall stabilisation of the complex and to an increase of protein yield (Portolano et al. 2014). Furthermore, functional studies have also shown the importance of SDS3 for the activity of the complex (Lechner et al. 2000; David et al. 2003). The SEC-MALS analysis shows that the SDS3 dimerization domain is necessary to mediate the dimerization of the complex. It is possible that in order to be active, Sin3a needs to be in its dimeric conformation and to present two HDAC enzymes to the target site. In this way, SDS3 would have both a structural and functional relevance to the complex, and it would explain why its deletion is associated with a cell-lethal condition (David et al. 2003; David et al. 2006).

XL-MS experiments have provided some constraints for modelling the complex. A model for the ternary complex Sin3a₅₃₁₋₇₇₂ : HDAC1 : SDS3 was created by docking the previously published structures with the cross-links that were found between the complex components. The docking model was then used for model fitting in both the SAXS envelope and in the negative stain EM structure. With this study a better understanding of the inner organisation of the Sin3a complex was achieved. For the first time, a structural arrangement for the Sin3a : HDAC1 interaction was modelled

based on experimental data.

From the crosslinking data, Sin3a appears to wrap around HDAC1 through multiple points of interaction (figure 5.9). An analogous structure of interaction was revealed for MTA1 interacting with HDAC1 (Millard et al. 2016). However, since Sin3a does not share a SANT domain with MTA1, the exact recruitment manner for Sin3a cannot be hypothesised based on the MTA1 : HDAC1 complex. Since the cross-linking analyses revealed multiple points of interaction between Sin3a and HDAC1, it is possible that Sin3a wraps around HDAC1 in a similar manner to MTA1, possibly creating contacts with a different part of the HDAC1 surface.

SAXS and negative stain EM have provided the first image for the Sin3a complex in its dimeric form. The SAXS envelope confirmed the dimeric assembly of the complex and suggests a structure with two globular shapes positioned at the top of a bridge structure. By fitting the docking model created with the cross-linking results, it appears that the two globular shapes contain the HDAC1 subunits and that Sin3a and SDS3 create the bridge structure. The presence of Sin3a and SDS3 in the bridge structure could be also justified by presence of the SDS3 dimerization domain. Moreover, the fact that the cross-linking docked model fits in the envelope increases the credibility of this model structure.

When comparing the SAXS and the EM results, the SAXS envelope results to be conserved to many 2D class averages that were generated by negative stain EM. Since the structure created by negative stain EM analysis correspond to the overall structural arrangement of the complex suggested by the cross-linking and SAXS model, it strengthens the possibility that the structural model built on the data achieved in this study corresponds to the real structure.

During these studies it was shown that the SDS3 coil-coiled dimerization domain could not be included in the model. This structure was achieved by crystallising the only coil coiled domain 2 from SDS3 (PDB code: 4ZQA. (Clark, Marcum, et al. 2015)). It is possible that this structure could change conformation when interacting in the complex. The fact that the structure of the dimerization domain can not fit in any of the models created by the different techniques used in this study suggest that it is

possible that the helix would undergo a sort of break, maybe resulting in a four-helix bundle. This observation increases the importance of solving the structure of the ternary complex in order to understand how its components interact with each other and which structural rearrangements they experience for the complex assembly.

Until now, XL-MS, SAXS and negative stain EM have provided a structural model for the Sin3a ternary complex. However, these data only allowed building a low resolution model and the exact details about how Sin3a recruit HDAC1 remain unknown. A better resolution structure could be achieved by cryo-EM analysis. Preliminary results have shown the Sin3a complex to aggregate during the process of sample preparation. It is possible that by varying the conditions of the sample preparation, the structure of the Sin3a complex could be determined.

Chapter 6. Discussion

6.1. A structural model for the Sin3a complex and suggested regulatory mechanism

In this thesis a variety of methods have been used to gain a better understanding of the Sin3a complex. The understanding of other Class I HDAC complexes is more advance. Structural and functional studies performed in the Schwabe group have revealed a common recruitment mechanism of deacetylase enzymes and regulatory mechanism. The SANT domain contained in SMRT, CoREST, MTA1 and NCoR allows the recruitment of the HDAC enzymes. This interaction involves the presence of a molecule of inositol phosphate that regulates the activity of the complexes. However, the same cannot be said for the Sin3a complex. To date only approximate boundaries of interactions with HDAC1 are known and a few structures have been published regarding the PAH domains of Sin3a interacting with different co-factors.

In **Chapter 3**, by testing a variety of constructs and performing different purification trials, a stable and stoichiometric Sin3a complex suitable for structural studies was produced. These studies also provided a more detailed insight into the interaction between Sin3a and HDAC1. From both mapping experiments (**Chapter 3**) and mass spectrometry characterization of a degraded form of Sin3a identified during the SEC-MALS analysis (described in **Chapter 5**), the central region of the HID domain comprising amino acids 601-742 was found to be sufficient to recruit HDAC1. However, additional residues are needed for a tight interaction. The requirement of residues 531-601 at the N-terminal and 742-772 at the C-terminal for a stable interaction suggest the presence of multiple interaction motifs. By analogy to the interaction manner of the SANT domain and the HDAC enzymes, it could be possible that the central four helices in Sin3a₆₀₁₋₇₄₂ represents a critical interaction surface, while the extended N- and C-termini wrap around HDAC1 and secure a stable interaction. The flexibility of the N- and C-termini of the Sin3a construct used for these experiments is also likely to facilitate a larger conformational change that could be observed *in vivo*. When expressed *in vivo*, the full length Sin3a interacts with multiple proteins and these interactions are likely to change the scaffold protein conformation. A

conformational change in the proximity of the HID domain could have the effect of modulating the interaction with HDAC1 and therefore the enzymatic activity of the complex.

The initial characterisation of the Sin3a complex described in **Chapter 3** also allowed testing histone substrate preference of the Sin3a complex and comparing its activity with other Class I HDACs. From the MSA results, the Sin3a complex did not show particular preference for any of the histone targets tested (Figure 3.16). Yet, when comparing the Sin3a activity to other Class I HDACs, the baseline activity of Sin3a is very similar to the baseline activity of the MTA1 and MiDAC complexes (figure 3.17) suggesting that there is scope for further activation of the complex. In addition to these findings, HDAC assays carried out in presence or absence of inositol phosphate confirm that the inositol molecule is not involved in the regulation of the Sin3a complex. The lack of substrate specificity and the insensitivity to the presence of inositol phosphate suggest a distinct regulatory mechanism for the Sin3a complex.

As mentioned in **Chapter 1**, the fine tuning of transcription relies on the dynamic recruitment of multiprotein complexes. Furthermore, these multiprotein complexes can themselves be quite dynamic. Within large multiprotein complexes, only part of the interacting partners are essential for the complex stability, whereas the remaining interacting proteins are able to modulate the enzymatic activity of the complex. Both the location and timing of their enzymatic activity may depend upon the gain or loss of specific subunits (Mittag et al. 2010). As a large multiprotein complex, the variability of complex components that can assemble into the Sin3a complex could be a determinant of its specific functions.

As mentioned in **Chapter 1**, the Sin3a complex is very well conserved during evolution and a different regulatory mechanism could be a reflection that the Sin3a complex has followed a different evolutionary pattern from the other Class I HDACs. When compared to simpler organisms like prokaryotes, more complex organisms require more complicated regulatory mechanisms in order to ensure the correct cell functions. The rise of multiple Class I HDAC complexes mirrors the necessity of a finely tuned and sensitive network of complexes able to regulate transcription in more specific ways.

Individual proteins that assemble into multiprotein complexes allow many levels of regulation. Conformational changes upon binding, the presence of different complex subunits and their order of assembly represent only part of the dynamic aspects of a multiprotein complex regulation.

As discussed in paragraph 1.3.3, HDAC1 and HDAC2 are recruited in different multiprotein complexes resulting in a wide spectrum of distinct functions and biological roles. Sin3a represents a distinct example of diversification within the same class of enzymes. As described in paragraph 1.4.5, the HID domain of Sin3a does not share homology with the SANT domain, which is found instead within the other class I HDAC complexes. The distinct way of recruiting the HDAC enzymes is also paired to a different response to the presence of inositol phosphate. As Inositol phosphate was found to be able to regulate the deacetylase activity of the class I complexes containing the SANT domain, the regulation of the Sin3a complex remains to be elucidated. It is possible that the activity of the Sin3a complex could be based upon the different subunits that can be part of the complex. Different complex subunits could determine structural changes on the Sin3a scaffold protein and together with the recruitment of the complex by different regulators could be the key of the regulation to the Sin3a complex. For example, previous studies have shown that different regulator proteins, like REST and SMRT, are able to recruit Sin3a (Nagy et al. 1997; Y. Huang et al. 1999). The variety of proteins that can participate in the Sin3a complex assembly and the different regulators that can recruit the complex establish a wide scenario of different possibilities that can resolve in different downstream effects.

Understanding this intricate multilevel regulation of these complexes would represent a major advance for developing complex specific drugs. The severe side effects associated with the use of HDAC inhibitors and the lack of alternative drugs to target the activity of HDAC complexes, emphasise the necessity to find drugs that allow individual complexes to be specifically targeted without affecting the activity of the other HDACs. Since the use of stapled peptides as potential drugs has recently gained prominence, the studies described in **Chapter 4** exploit the possibility to target the Sin3a complex specifically with the use of stapled peptides. Two different peptides were designed to target the Sin3a complex with two distinct mechanisms of action.

Aiming to disrupt the Sin3a core complex, SDS3 stapled peptides were designed based on the published structure of Clark et al. in 2015 (PDB code: 2N2H). From testing the SDS3 stapled peptides a better understanding of the interaction between Sin3a and SDS3 was gained. Since the SDS3 stapled peptide successfully displaced SDS3₂₀₁₋₂₃₄ *in vitro*, but it failed in showing its effects *in cellulo*, it is possible that multiple binding motifs are present between Sin3a and the full length SDS3 in the endogenous full length Sin3a complex. In a similar manner to the hypothesized interaction between Sin3a and HDAC1, SDS3 could present a critical interaction region composed by the alpha helix in residues 201-234 with multiple points of interaction at the N- and C-termini that support and strengthen the interaction with Sin3a. Additionally, since SDS3 is part of the core complex, it is not known whether other components interacting in the complex might shield SDS3 and hinder the access of the stapled peptides. A better understanding of the interaction between Sin3a and the full-length SDS3 is therefore required to understand if the use of stapled peptides is to be successful in targeting the Sin3a complex. Perhaps a combination of two stapled peptides, acting on two different regions of SDS3 would be an effective way to disrupt the complex.

As the disruption of the core complex by using SDS3 stapled peptides appeared to require further structural studies, a different approach was tested to verify if stapled peptides could be used for targeting the Sin3a complex. In **Chapter 4** the use of Mxd1 stapled peptide to target the recruitment of the Sin3a complex to chromatin was discussed. In this case the interaction between Mxd1 and the PAH2 is well established. *In vitro* results were confirmed by the use of GFP constructs showing the ability of the Mxd1 stapled peptide to displace the interaction of Mxd1 with the endogenous Sin3a. However, these promising results could not be repeated *in cellulo*. This could be due to failure of the stapled peptides to enter the cell or of endogenous proteases that destroy them. Since the structure of the peptides was shown to be capable of actively displace the Sin3a endogenous complex, it will be interesting to modify the Mxd1 stapled peptide structure in order to enhance its cell penetrance and resistance to proteases. Perhaps a small tag could be added to the peptide to enhance its solubility or mutations on the peptides could be tested to reduce the proteolytic sensitivity

without altering the interaction with the target.

In this study a structural model for the Sin3a complex was built by using a variety of techniques (**Chapter 5**). The Sin3a complex was established to assemble as a dimer through the dimerization domain of SDS3. The dimeric form of the complex was verified both by SEC-MALS, SAXS and negative stain EM analyses. The complex appears to adopt a “V” shape, where two HDAC1 are located into two globular structures while Sin3a and SDS3 connect them creating a bridge at the bottom of the structure.

From the cross-linking results it is possible to map the region of interaction between Sin3a and HDAC1. Interestingly, both Sin3a and SDS3 appear to make extensive contact with HDAC1 by wrapping around its structural core. However, when comparing the docking model created with the cross-linking results to the structure of interaction of the MTA1:HDAC1 complex, it appears that Sin3a makes contact with a different region of HDAC1. The binding interface between MTA1 and HDAC1 is characterized by a basic pocket that allocates an inositol tetraphosphate molecule (as observed for the SMRT:HDAC3 complex), which has been shown to play a fundamental role for the regulation of the complex activity. Since Sin3a appears to recruit HDAC1 by making contact with a different region of the enzyme, the highly basic pocket required to allocate the inositol phosphate molecule might not be created. This particular feature could determine the immunity of the Sin3a complex from inositol phosphate molecules. Therefore, further studies (both structural and functional) will be required to elucidate a detailed structure of the Sin3a:HDAC1 interaction and to understand the regulatory mechanism of action of the complex.

A negative stain EM model was acquired by studying the cross-linked complex, therefore, it is not clear if the complex can shift between an “open” and “closed” state by increasing or reducing the distance between the two HDAC1. It may be important to determine the structure without the use of cross-linkers. An attractive approach would be cryo-EM. Preliminary experiments of cryo-EM showed a highly aggregated protein sample. Different conditions in protein purification, protein concentration and blotting time could be tested to be able to analyze the complex with this powerful technique. By optimizing the sample it may be possible to achieve a high-resolution structure of the Sin3a complex.

The possible flexibility of the complex between an “open” and “closed” state would also explain the failure of the crystallization trials. Furthermore, heterogeneity in the protein sample was shown by the SEC-MALS analyses (**Chapter 5**). The creation of two distinct peaks (figure 5.1), the shape of the chromatograms and the size of the complex detected by the MALS analysis indicate a heterogeneous sample that is not ideal for crystallization trials. The heterogeneity of the sample could be mainly due to two factors: the dissociation of the complex and protein degradation. The effects of both can be directly verified on the SEC-MALS results. The two distinct peaks formed by the separation of the binary from the ternary complex show the disassembly of the Sin3a complex. While the appearance of an additional band corresponding to a degraded form of Sin3a showed that Sin3a is target of proteases. One of the reasons for this outcome could be that the HID might not have been fully folded and hence vulnerable to proteases.

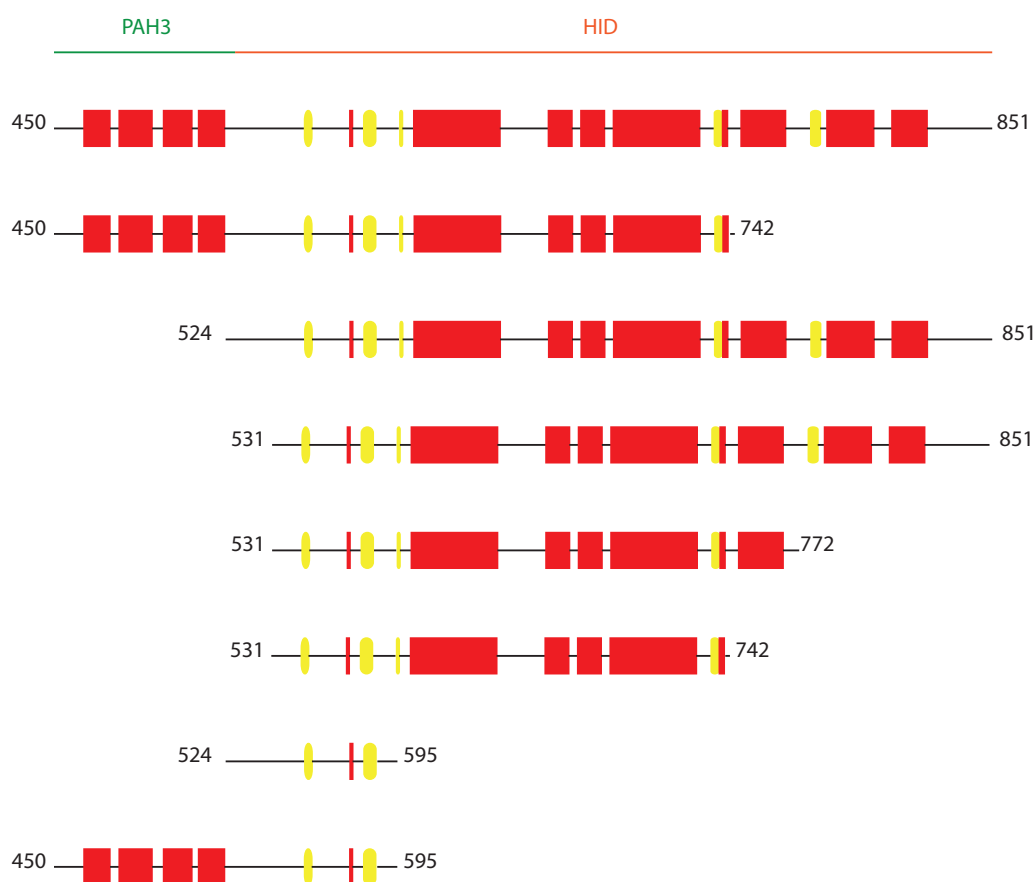
Overall, as a result of this study, we now have a better understanding of the structure of the Sin3a complex. Furthermore, a regulatory mechanism that involves the variable spectrum of proteins that can interact with Sin3a and alter the composition and the recruitment of the complex has been hypothesized. However, the structure of interaction between Sin3a and HDAC1 is yet to be solved. Understanding the recruitment of HDAC1 that occur in a different fashion from the SANT domain could bring a breakthrough in our understanding into the regulation of the complex and could be helpful to generate drugs able to specifically target the Sin3a enzymatic activity. It would be also interesting to identify the key factors that determine the substrate specificity of the Sin3a complex. The role of several components of the Sin3a complex is still unknown, exploring their role and effects on the complex might reveal further regulatory mechanisms on the complex activity. Hopefully, with the constant advance in technologies such as the advances in cryoEM we will be able to address many of these questions.

Appendix 1: **Sequence alignment of Sin3a₁₋₅₄₅**. Clustal software was used to align Sin3a from human and other four model organisms.

158

Appendix 2: Sequence alignment of Sin3a⁵⁴⁵⁻¹²⁷³. Clustal software was used to align Sin3a from human and other four model organisms.

S.cerevisiae	LDLDELAFGSPYKRLLPKSDTFMPCSGRDMCMWEVLNDEWGHVPWASEDSGFIHRRKNQ	802	S.cerevisiae	IYIFRHWITTYIRLLEIKOMNERTKEINTRS-----IVTFAKD--LDLSSQLSMGL	1196
D.melanogaster	VDLSSCKRLGASVCAIPQSTVPRKCSGRTALCEVINDRWKVSFPTWASEDSTFVSSKTYQ	1087	D.melanogaster	WYLFRLHAILCDRLHVMYRARIILATEERC-----VNRSTVATALRKPKP	1553
D.erio	IDYASCKRLGSSYRALPKSYQLPKCTGTGTALCKEVLNDTWVSFPSWS--EDSTFVSSKTKQ	607	D.erio	WYLFRLHAILCGLSRLLRLYQAQAEQIEEENREREREVILGLGRDNENTAVQLRKEPM	958
H.sapiens	IDYASCKRLGSSYRALPKSYQQPKCTGTGTPLCKEVLNDTWVSFPSWS--EDSTFVSSKTKQ	604	H.sapiens	WYTFMRLHILICLRLIRLCQAQEQIEEENREREREVILGIRKDKSDSPATQLRKEPM	951
M.musculus	IDYASCKRLGSSYRALPKSYQQPKCTGTGTPLCKEVLNDTWVSFPSWS--EDSTFVSSKTKQ	605	M.musculus	WYTFMRLHILICLRLIRLCQAQEQIEEENREREREVILGIRKDKSDSPATQLRKEPM	952
* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *			* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *		
S.cerevisiae	YEETLFKIEERHEYDFYIESNLRTIQGLETVINKEINMTEKANKFLPGLCHTSMTI	862	S.cerevisiae	DFVEDAYKQVLRKSRRLINGLGHQWFESIRQVNNKAFKLYTIDKVTQSLVKHAIITL	1256
D.melanogaster	FEETIYHTEDERFELDLVLENSATIRVLENLOKKMRSMTESKFLHDLHGGTSOTI	1147	D.melanogaster	ELOVEDYPTFLDMKLVNLDGNMNSDFTDMREMTGIVAYISFTLDKVVSNVAPQLQYC	1613
D.erio	YEEHMYCDERFELDLVLENLATIRVETQKILSRMSAEQAKERLDNTLGGFSELI	667	D.erio	DVEVDYSAFLDKVRLNLDGNKEASOYEDSIREMTTHAYIAFTMDKLIQNLAQLOHC	1018
H.sapiens	YEEHLYCDERFELDLVLENLATIRVLEAIQKKLSRLSABEQAKERLDNTLGGTSSEVI	664	H.sapiens	DVDVDYTPALDMDVRSLLDGNLSSQYEDSIREMTTHAYIAFTMDKLIQSLVIRQAQHI	1011
M.musculus	YEEHLYCDERFELDLVLENLATIRVLEAIQKKLSRLSABEQAKERLDNTLGGTSSEVI	665	M.musculus	DVDVDYTPALDMDVRSLLDGNLSSQYEDSIREMTTHAYIAFTMDKLIQSLVIRQAQHI	1012
** : : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *			: : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *		
S.cerevisiae	YKVKIRKVDKERGFETIDALHHPFAVTPAVPLVKRLKQKDEWRRRAQRWNVKWRLEPKQ	922	S.cerevisiae	MTDAKTAETMALFVKORNASTTSAK-----DOIITYRLOVRSHMSTENNFRIFDK	1307
D.melanogaster	HORALIRHLYGDKS--GELITGMKNPFAVPDVLVKRLKVEEWEREAQKTFHKQWRQNEK	1206	D.melanogaster	VTERAALDCELEFANTEQHRCTGCGPCGCHDAIKTFDREMSYQK--AESILNENCCKVLYK	1672
D.erio	HRKALQRIYADKA--PDIIDGLKKNPAASVPVLVKRLKKEEWEREAQGFNKKIWRQNEK	726	D.erio	VSEELICQVTELLIYVESHSSTGSGSLTSSBAQAQASVQK--AEQLMSDENGCKFLMFLK	1077
H.sapiens	HRKALQRIYADKA--ADIIDGLKKNPSIAVPVLVKRLKMKKEEWEREAQGFNKKVWRQNEK	723	H.sapiens	VSDEICQVQVTELLYLAENNINGCATGGQJATQNSRLLESTYQK--AEQLMSDENGCKFLMFLK	1070
M.musculus	HRKALQRIYADKA--ADIIDGLKKNPSIAVPVLVKRLKMKKEEWEREAQGFNKKVWRQNEK	724	M.musculus	VSDEVCQVQVTELLYLAENNINGCATGGQJATQNSRLSLESAYQK--AEQLMSDENGCKFLMFLK	1071
: : : : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *			: : : : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *		
S.cerevisiae	VFFKSLDLGLTEKQADKKLLITTKOLISEISSIKVDQTNKKLIHW--LTQPKKSQLDFFP	980	S.cerevisiae	RTLHVSII--OYVIALDOLLKEP--KADEDKKVVYVTSYVALPHPTTEGIPHEKLIPTLERLI	1364
D.melanogaster	YYLKSILDHQAENKPNMKAALRSKSLFNEIETLYDERHQED--DAMEPGPHLVLYPK	1263	D.melanogaster	IDC--KVITIELLSEPEVVKPAALKQAQKFSKYVERLANPALGGGNTGKSDSALGDSV	1730
D.erio	YYLKSILDHQENKPNQNDTKVRSKTLNLETESIYDERQKQVSEUNSSYVASSPHLTYTE	786	D.erio	SRGVOLITVELLDTSEENSQGP--WETERWSDYVRYLSADFTSPDLAD---HLSQKPV	1131
H.sapiens	YYLKSILDHQENKPNQNDTKVRSKTLNLETESIYDERQKQATEENAG--VPVGPHLSLAYE	782	H.sapiens	SQGVOLITVELLDTSEENSQGP--VEAERWSDYVRYMNSDTTSPDLAE---HLAQKPV	1124
M.musculus	YYLKSILDHQENKPNQNDTKVRSKTLNLETESIYDERQKQATEENAG--VPVGPHLSLAYE	783	M.musculus	SQGVOLITVELLDTSEENSQGP--VEAERWSDYVRYMNSDTTSPDLAE---HLAQKPV	1125
: : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *			: : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *		
S.cerevisiae	DNKIFVDILCLADPTFTHHTTAYSNDPKERLKLQKLYFISIFESTIFEKTFES-----LY	1034	S.cerevisiae	EFQGDID-----GTEVDEEFSDEGI-----SVSTLIKTIQPIYQLH--IENGSS	1406
D.melanogaster	DRTILDDAANLLIHHVKRQTGIQKQKQKIKQI IQGVFDPDLPFAQPLSDDERDDATPF	1323	D.melanogaster	VDGSDIK-----TEADEDTRAHTQS-----QQWT-----LLESNEK	1762
D.erio	DHQILEDAAALI IHHVKRQSSINKEKDYKIKQI VYVFTI PMILFAQRGLVLSDEEEDEE--	844	D.erio	FLPNRLRIRKQRCQCAKELKDGERKENAD---SNRMECPKLNISKYKMYVYKSEID	1187
H.sapiens	DKQILEDAAALI IHHVKRQTGIQKEDKYIKQIIMHIFIPOLLFAQRGLSDVEEEEEE--	841	H.sapiens	FLPNRLRIRKQRCQRCQKQKQKGNKKTMTENVDSLDKCEKFKLNISKYKMYVYKSEID	1184
M.musculus	DKQILEDAAALI IHHVKRQTGIQKEDKYIKQIIMHIFIPOLLFAQRGLSDVEEEEEE--	842	M.musculus	FLPNRLRIRKQRCQRCQKQKQKGNKKTMTENVESLDKCEKFKLNISKYKMYVYKSEID	1185
* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *			* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *		
S.cerevisiae	SHK-----QNVSESSGSDGSSIASRKRQYQ-----QKMSLLI-----DLHRS	1072	S.cerevisiae	YDVFTKATNKYPTTANDNTQKQWQSKKELISKFLDCAVGLNNILDEAQKLSMQKWN	1466
D.melanogaster	LVDNNTKMDVDSPIGRKTESSTRNAKSTPSESASPARNASSTSVTPAGIKKETDSDKAIT	1383	D.melanogaster	YI-WSS-----P--IKTHVQ-----QWWT-----LLESNEK	1776
D.erio	EDMELDEGASKKH-----MDVDEATGAKKH-----	857	D.erio	YM-YRR-----TALLRAHQSHERSV--TRLHQRF-----HAMSQWAK	1222
H.sapiens	MDVDEATGAKKH-----MDVDEATGAKKH-----	854	H.sapiens	YM-YRR-----TALLRAHQSHERSV--KRLHQRF-----QAWVDKWTK	1219
M.musculus	MDVDEATGAKKH-----MDVDEATGAKKH-----	855	M.musculus	YM-YRR-----TALLRAHQSHERSV--KRLHQRF-----QAWVDKWTK	1220
: : : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *			* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *		
S.cerevisiae	-RYQ-----KLVK-----KLVK-----KLVK-----KLVK-----	1084	S.cerevisiae	LKDSIAKTSAGNQGIESETEKCKITKQESQSNLSDSTASVLPASITVTPQDDMIETGNT	1526
D.melanogaster	GSFAPASSATASSATPVDDATPSTSSAAAAAASSTVSSTGTEKPKDDPLSSHKEGAG	1443	D.melanogaster	EHVTRMAAETSRLWLMGEGRE-----GLIPGTSCTDPEILHFQRI--NK	1776
D.erio	-----EDMELDEGASKKH-----MDVDEATGAKKH-----	861	D.erio	EHVTRMAAETSRLWLMGEGRE-----GLIPGTSCTDPEILHFQRI--NK	1264
H.sapiens	-----MDVDEATGAKKH-----MDVDEATGAKKH-----	858	H.sapiens	EHVTRMAAETSRLWLMGEGLE-----GLIPGTTCTDPEILHFVSI--NK	1261
M.musculus	-----MDVDEATGAKKH-----MDVDEATGAKKH-----	859	M.musculus	EHVTRMAAETSRLWLMGEGLE-----GLIPGTTCTDPEILHFVSI--NK	1262
: : : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *			: : : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *		
S.cerevisiae	GKVQLSEPEEPPEPNTIEEELIDGAEKNPWITGN--LVEEAANSQGLIQNRSLFNLFANN	1143	S.cerevisiae	ESSDKGAKIQ--	1536
D.melanogaster	STSSGVATSPRAQDITAGAVDVEIKLHPHADFNSNPKLLPHAHQRCDESYTLFANN	1503	D.melanogaster	-----	1776
D.erio	G---GASAPTKSK-----LIFSAS-----TAQRLRSCEDAYNLFYVANN	898	D.erio	YRKVYGTNKPQ	1276
H.sapiens	G-----SPPKSK-----LIFSNT-----AAQKLKGMDEVNLFYVANN	891	H.sapiens	YRKVYGTNKPQ	1273
M.musculus	G-----SPPKSK-----LIFSNT-----AAQKLKGMDEVNLFYVANN	892	M.musculus	YRKVYGTNKPQ	1274
* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *			* : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *		



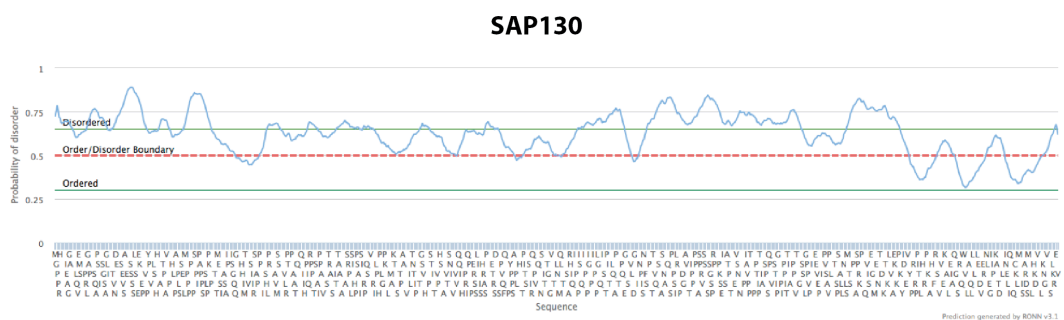
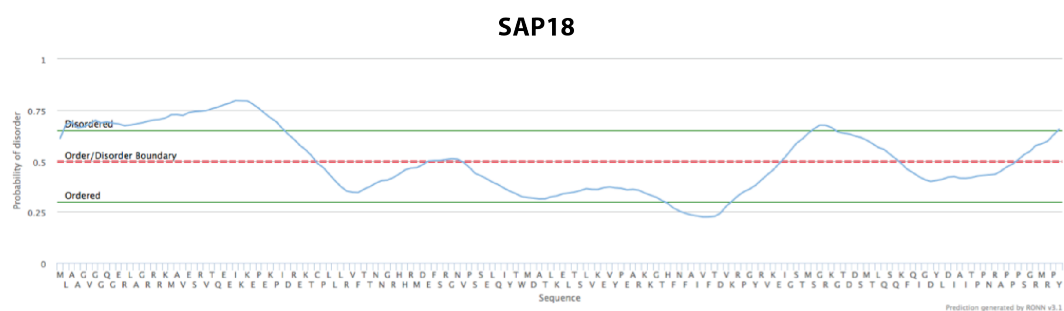
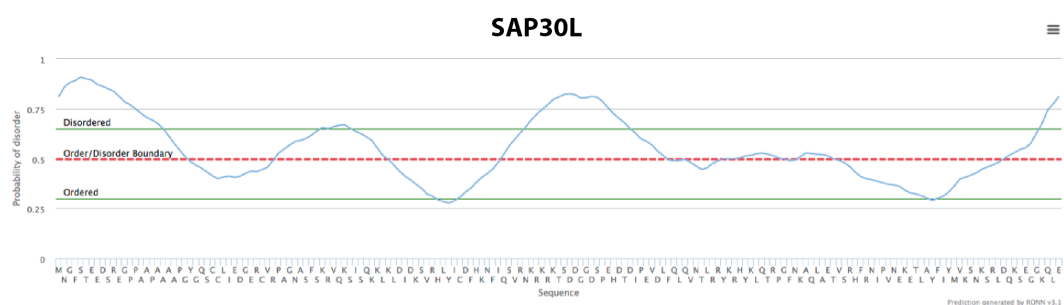
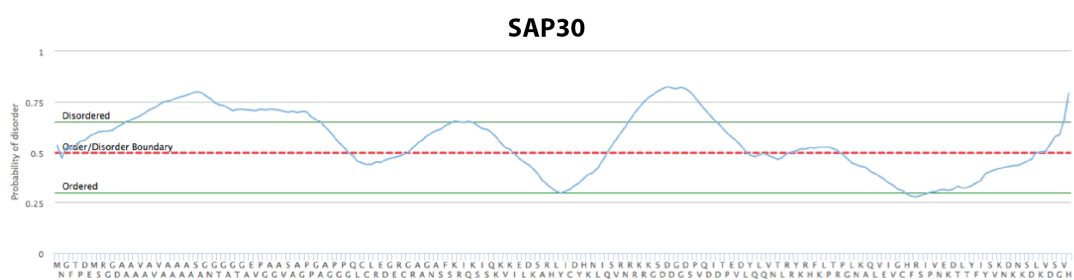
Appendix 3: **Secondary structure prediction for the Sin3a protein constructs.** Jpred software was used to make secondary structure prediction on the Sin3a protein sequence. Red boxes represent alpha helices and yellow boxes represent beta strands.

S.cerevisiae	-----MAIQKVSNKDLSRK	14
D.rerio	-----MVDYYNDEEELDSVDEDEDRSFRGRDSEEDTEDASETDLAKH	42
H.sapiens	MSAAGLLAPAPAQAGAPPAPEYYPEEDEELESADDDERSCRGREDEDTEDASETDLAKH	60
M.musculus	MSAAGLLAPAPAPAAAPAAPEYYPEDEEELESADDDERSCRGREDEDTEDASETDLAKH	60
	...*.***:	
S.cerevisiae	DKRRFNIESKVNKIYQNFYSERDNQYKDRLTALQTDLTSLHQGDNGQYARQVRDLEEERD	74
D.rerio	D-----EDDYVEIKEQMYQDKLASLKRQLQQLQEGTLQEYQKRMKKLDQQYK	89
H.sapiens	D-----EEDYVEMKEQMYQDKLASLKRQLQQLQEGTLQEYQKRMKKLDQQYK	107
M.musculus	D-----EEDYVEMKEQMYQDKLASLKRQLQQLQEGTLQEYQKRMKKLDQQYR	107
	* : : . : : * : * : * : * : * . * : * : * : : . * : : :	
S.cerevisiae	LELVRLRLFEEYRVSRSGIEFQEDIEKAKAEHEKLIKLCERLYSSIEQKIKKLQEERLL	134
D.rerio	ERLRNADLFLQLETEQVERNYIKEKKAAVKEFDDKKVELKENLIAELEKKKKMVENEKL	149
H.sapiens	ERIRNAELFLQLETEQVERNYIKEKKAAVKEFEDKKVELKENLIAELEKKKKMIENEKL	167
M.musculus	ERIRNAELFLQLETEQVERNYIKEKKAAVKEFEDKKVELKENLIAELEKKKKMIENEKL	167
	. : . * * : . . : : : * * . . * . * . : : * * : : * * :	
S.cerevisiae	MDVANVHSYAMNYSRPQYQKNTRSHTVSGW--DSSSNEYGRDTANESATDTGAGNDRRTL	
192		
D.rerio	MELTGDSMEV---KPIMTRKLRRLRRPNDPVPIDPKRRKPAPAQLNYLLSDEQIMEDLRTL	205
H.sapiens	MELTGDSMEV---KPIMTRKLRRLRRPNDPVPIDPKRRKPAPAQLNYLLTDEQIMEDLRTL	223
M.musculus	MELTGDSMEV---KPIMTRKLWRRPNDPVPIDPKRRKPAPAQLNYLLTDEQIMEDLRTL	223
	* : : . . : * : : : . . . : . * : * : * * *	
S.cerevisiae	RRRNASKDTRGNNNNQESDFQTGNNGSGSNGHSGRQGSQFPFNNLTYKSGMNSDSDFLQ	252
D.rerio	NKLKSPKRQ-G---MDVLSFRASPSS--PE-----HLPS-----TPVDTPSQRYEA	245
H.sapiens	NKLKSPKRP-----ASPSS--PE-----HLPA-----TPAESPAQRFEA	255
M.musculus	NKLKSPKRP-----ASPSS--PE-----HLPA-----TPAESPAQRFEA	255
	. : : * : . * : : * . . . :	
S.cerevisiae	GINEGTDLYAFLFGEKNPKDNANGNEKKKNRGAQRYSTKTAPPLQSLKPDEV---TEDI	308
D.rerio	RIEKGKLYDKRWYHKSQAIYLES---KENT-----KISCVISSVGTNEIWRKTSDS	295
H.sapiens	RIEDGKLYDKRWYHKSQAIYLES---KDNQ-----KLSCVISSVGANEIWRKTSDS	305
M.musculus	RIEDGKLYDKRWYHKSQAIYLES---KDNQ-----KLSCVISSVGANEIWRKTSDS	305
	* : : . * : . * : . * * : : * : : * : * :	
S.cerevisiae	SLIRELTGQP-PAPFRLRSD---	327
D.rerio	TKMRIYLGQLQRGAFFIRRRSAA	318
H.sapiens	TKMRIYLGQLQRGLFVIRRRSAA	328
M.musculus	TKMRIYVGQLQRGLFVIRRRSAA	328
	: : * * * . * : *	

Appendix 5: **Sequence alignment of SDS3_{FL}**.Clustal software was used to align SDS3 from human and other four model organisms.

SAP30	-----	0
SAP30L	-----	0
SAP18	-----	0
SAP130	MGPPRHPQAGEIEAGGAGGRRLLQVEMSSQQFPRLGAPSTGLSQAPSQIANSQSAGLINP	60
SAP30	-----	0
SAP30L	-----	0
SAP18	-----	0
SAP130	AATVNDESGRDSEVSAREHMSSSSLQSREEKQEPVVVRYPYPQVQMLSTHHAVASATPVA	120
SAP30	-----	0
SAP30L	-----	0
SAP18	-----	0
SAP130	VTAPPAHLTFAPVLSFSFSEGLMKPPPKPTMPSRPIAPAPPSTLSLPPKVPQGQVVTMESSI	180
SAP30	-----	0
SAP30L	-----	0
SAP18	-----	0
SAP130	PQASATPVATISGQQGHPSNLHHIMTTNVQMSIIRSNAPGPPLHIGASHLPRGAAAAVM	240
SAP30	-----	0
SAP30L	-----	0
SAP18	-----	0
SAP130	SSSKVTTVLRPTSQLPNAATAQPAVQHIIHQPIQSRPPTTSNAIPPAVVATVSATRAQS	300
SAP30	-----	0
SAP30L	-----	0
SAP18	-----	0
SAP130	PVITTTAAHATDSALSRPTLSIQHPPSAATSIQRPASRDVTTTRITLPSHPALGTPKQQL	360
SAP30	-----	0
SAP30L	-----	0
SAP18	-----	0
SAP130	HTMAQKTIISTGTPEAAATVAPILATNTIPSATTAGSVSHQAPTSTIVTMTVPSSSHA	420
SAP30	-----	0
SAP30L	-----	0
SAP18	-----MAVESRVTOEEIKKEPEKPIDREKTCPL-----	28
SAP130	TAVTTSNIPVAKVVPQQITHSTPRIQPDYPAERSSLIPISGHRASPNPVAMETRSNRP	480
SAP30	-----	0
SAP30L	-----	0
SAP18	-----	28
SAP130	VPVQFQYFLPTYPPSAYPLAAHTYTPITSSVSTIRQYPVSAQAPNSAITAQTGVGVA	540
SAP30	-----MNGFTPEMSRGG-----	13
SAP30L	-----MNGFTEEDSR-----	11
SAP18	-----	28
SAP130	HLNPMQLMTVDASHARHIQGIQPAPISTQGIQPAPITGPGIQAPLGTQGIHSATPINTQ	600
SAP30	-----	13
SAP30L	-----	11
SAP18	-----	28
SAP130	GLQPAPMGTQQPQPEGKTSVVVLADGATIVANPISNPFSAAPAAATTVQTHSQSASTNAP	660
SAP30	-----DAA--	16
SAP30L	-----	11
SAP18	-----	28
SAP130	AQGSSPRPSILRKKPATDGMVARKTLIPQPPDVASPRVSSMRSTSGSPRPAGAKPKSE	720
SAP30	-----AAVAAVAAAAAASAGNGTGAGT	40
SAP30L	-----	11
SAP18	-----	28
SAP130	IHVSMATPVTVSMETVSNQNDQPTIAVPPTAQPPPTIPTMIAAASP-----SQPAVA	775
SAP30	GAEVPGAGAVSAAGPPGAAG-----	60
SAP30L	-----EGPPAAPAAAA	22
SAP18	-----	28
SAP130	LSTIPGAVPITPPIITIAAAPPSVTVGGLSSVLGPPVPEIKVKEEVEPMDIMRPVSAV	835
SAP30	-----PGPGQLCCLREDGERCGRAAGNASF	85
SAP30L	-----PGYGQSCCLIEDGERCVRPAGNASF	47
SAP18	-----	28
SAP130	PPLATNTVSPSLALLANLSMPTSDLPFGASPRKKPRKQHVISTEEGDMMETNSTDD--	893
SAP30	SKRIQKSISQKKVKIELDKSARHLYICDYHKNLIQSVRNRRKRKGSDDGGDSPVQDIDT	145
SAP30L	SKRVQKSISQKKLKLIDKSVRHLIYICDFHKNFIQSVRNKRKRKTS-DDGGDSPHEDTDI	106
SAP18	-----	28
SAP130	EKSTAKSL---LVKAERKSP-----PKEYIDE	918
SAP30	PEVDLYQLQVNTLRRYKRHKLPTRPGLNKAQ-----LVEI-----VGCHF	186
SAP30L	PEVDLEQLQVNTLRRYKRHKLPTRPGFNKAQ-----LAET-----VSRHF	147
SAP18	-----LIRVFTT--NNGRHHRMDEFSGRNVPS--	58
SAP130	EGVRYVPVRPRPITLLRHRYRNPWKAAYHHFQRYSDVRVKEEKAMLQEIANKQKGVSCRA	978
	* : . :: :	:
SAP30	RSIPVNE-----KDTLTYFIYS-----VKNDK	208
SAP30L	RNIPVNE-----KETLAYFIYM-----VKSNDK	169
SAP18	YTWM---DATLKELTSLVKEVYPEARKKGTGFNFIVFTDVKRPGYRVKEIGSTMSGRK	114
SAP130	QGWKVLCAAQLQLNLNLEHDVYERLTNLQ---EGIIPKKKAATDDDLHRINELIQGNM	1034
	. * : *	...

Appendix 6: Sequence alignment of the Sin3a Associated Proteins (SAPs). Clustal software was used to align SAP18, SAP30, SAP30L and SAP130 from human and other four model organisms.



Appendix 7: Disorder prediction for the SAPs proteins. The disorder prediction for the SAPs was calculated by using RONN software.

Appendix 8: **Details of protein constructs used to express the Sin3a complex components in *E. coli*.** Secondary structure and disorder prediction were taken in consideration to decide the boundaries for each construct. A variety of affinity tags were tested to optimise protein expression in *E. coli*.

Protein	Included Domain(s)	Residues	Affinity tag	Successfully expressed?	Protein yield
Sin3a	PAH3-HID	450-772	N-GST ; N-His ₆ ; N-GB ₁ /His ₆	Yes (all)	High (all)
	PAH3-HID	450-595	N-GST	No	
	HID	651-741	N-GST	No	
	HID	601-737	N-GST	Yes	High
	HID	609-742	N-GST	Yes	Low
	HID	601-742	N-GST	Yes	High
	HID	609-737	N-GST	Yes	High
	HID	450-851	N-GST	Yes	Low
	HID	450-742	N-GST	Yes	High
	HID	524-851	N-GST	No	
	HID	531-851	N-GST	Yes	Low
	HID	531-742	N-GST	Yes	Low
	HID	524-595	N-GST	Yes	High
	HID	531-772	N-GST	Yes	High
HDAC1	Full length	1-483	No Tag	Yes	Low
SDS3	Full length	1-328	GB ₁ /His ₆	No	
	SID	201-234	GB ₁ /His ₆	Yes	High
SAP30	SID	64-220	N-GB ₁ /His ₆	No	
	SID	120-220	N-GB ₁ /His ₆	Yes	High

Appendix 9: **Details of protein constructs used to express the Sin3a complex components in HEK 293F cells.** Secondary structure and disorder prediction were taken in consideration to decide the boundaries for each construct. A variety of affinity tags were tested to optimise protein expression in HEK 293F.

Protein	Included Domain(s)	Residues	Affinity tag	Successfully expressed?	Protein yield
Sin3a	PAH3-HID	450-772	N-Flag ; C-Flag	No	
	PAH3-HID	438-772	N-Flag	No	
	PAH3-HID	438-855	N-Flag	No	
	PAH3-HID	450-855	N-Flag	No	
	HID	601-742	N-Flag	Yes	Low
	HID	531-742	N-Flag	Yes	Low
	HID	524-742	N-Flag	No	
	HID	524-772	N-Flag	No	
	HID	531-772	N-Flag	Yes	High
HDAC1	Full length	1-483	C-Flag ; No Tag	Yes	Low ; High
SDS3	Full length	1-328	No Tag	Yes	High
	SID	201-234	No Tag	Yes	Low
	SID	89-328	No Tag	No	
	SID	89-234	No Tag	No	
SAP30	SID	64-220	No tag	No	
	SID	120-220	No tag	No	
SAP18	Full length	1-153	No tag	No	
	SID	44-153	No tag	No	
SAP30L	Full length	1-184	N-Flag	No	
	SID	24-184	No tag	No	
SAP130	Full length	1-1084	No tag	No	
	SID	836-1084	No tag	No	
	SID	836-1047	No tag	No	
	SID	913-1084	No tag	No	
	SID	913-1047	No tag	No	

Appendix 10: **Sequence alignment of the SAP proteins.** Clustal software was used to align SAP18, SAP30, SAP30L and SAP130 from human. Residues highlighted in green are positions that have a single and fully conserved residue, in yellow are indicated the residues with strongly similar properties and in grey are indicated residues with weakly similar properties.

SAP30	SKRIQKSIQKKVKIELDKSARHLYICDYHKNLIQSVNRNRKRKGSDDDDGGDSPVQDIDT	145
SAP30L	SKRVQKSIQKKLKLDIDKSVRHLYICDFHKNFIQSVNRNRKRKTS-DDGGDSPEHDTDI	106
SAP18	-----	28
SAP130	EKSTAKSL---LVKAERKSP-----PKEYIDE	918
SAP30	PEVDLYQLQVNTLRRYKRFEKLPTRPGLNKAQ-----LVEI----VGCHF	186
SAP30L	PEVDLFQLQVNTLRRYKRHYKLQTRPGFNKAQ-----LAET----VSRHF	147
SAP18	-----LLRVFTT-NNGRHHRMDEFSGNVPSS-----ELQI	58
SAP130	EGVRYVPVRPRPPITLLRYRNPWKAAYHFFQRYSDVRVKEEKAMLQEIANQKGVSCRA	978
	* : . :: :	:
SAP30	RSIPVNE-----KDTLLYFYYS-----VKNDK	208
SAP30L	RNIPVNE-----KETLAYFYIM-----VKSNK	169
SAP18	YTWM----DATLKELTSLVKEVYFEARKKGTHFNFAIVFTDVKRPGYRVKEIGSTMSGRK	114
SAP130	QGWKVHLCAAQLQLTNLEHDVYERLTNLQ---EGII PKKKAATDDDLHRINELIQGNM	1034
	. * : *	...

References

- Aasland, R., Stewart, A.F. & Gibson, T., 1996. The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. *Trends in biochemical sciences*, 21(3), pp.87–88.
- Abrajano, J.J. et al., 2009. REST and CoREST modulate neuronal subtype specification, maturation and maintenance. H. Okazawa, ed. *PloS one*, 4(12), p.e7936.
- Ahringer, J., 2000. NuRD and SIN3 histone deacetylase complexes in development. *Trends in genetics : TIG*, 16(8), pp.351–356.
- Alland, L. et al., 2002. Identification of mammalian Sds3 as an integral component of the Sin3/histone deacetylase corepressor complex. *Molecular and cellular biology*, 22(8), pp.2743–2750.
- Alland, L. et al., 1997. Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. *Nature*, 387(6628), pp.49–55.
- Allfrey, V.G., Faulkner, R. & Mirsky, A.E., 1964. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 51, pp.786–794.
- Andrews, P., 1964. Estimation of the molecular weights of proteins by Sephadex gel-filtration. *The Biochemical journal*, 91(2), pp.222–233.
- Andrés, M.E. et al., 1999. CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 96(17), pp.9873–9878.
- Arents, G. & Moudrianakis, E.N., 1995. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proceedings of the National Academy of Sciences of the United States of America*, 92(24), pp.11170–11174.
- Ayer, D.E., Kretzner, L. & Eisenman, R.N., 1993. Mad: a heterodimeric partner for Max that antagonizes Myc transcriptional activity. *Cell*, 72(2), pp.211–222.
- Ayer, D.E., Lawrence, Q.A. & Eisenman, R.N., 1995. Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell*, 80(5), pp.767–776.
- Bannister, A.J. & Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell research*, 21(3), pp.381–395.
- Bansal, N. et al., 2011. Tumor suppressor protein p53 recruits human Sin3B/HDAC1 complex for down-regulation of its target promoters in response to genotoxic stress. A. L. Gartel, ed. *PloS one*, 6(10), p.e26156.

- Bantscheff, M. et al., 2011. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nature biotechnology*, 29(3), pp.255–265.
- Baymaz, H.I., Karemaker, I.D. & Vermeulen, M., 2015. Perspective on unraveling the versatility of “co-repressor” complexes. *Biochimica et biophysica acta*, 1849(8), pp.1051–1056.
- Becker, P.B. & Workman, J.L., 2013. Nucleosome remodeling and epigenetics. *Cold Spring Harbor perspectives in biology*, 5(9), pp.a017905–a017905.
- Berger, S.L., 2001. An embarrassment of niches: the many covalent modifications of histones in transcriptional regulation. *Oncogene*, 20(24), pp.3007–3013.
- Bertos, N.R., Wang, A.H. & Yang, X.J., 2001. Class II histone deacetylases: structure, function, and regulation. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 79(3), pp.243–252.
- Biggar, S.R. & Crabtree, G.R., 2001. Cell signaling can direct either binary or graded transcriptional responses. *The EMBO journal*, 20(12), pp.3167–3176.
- Blander, G. & Guarente, L., 2004. The Sir2 family of protein deacetylases. *Annual review of biochemistry*, 73(1), pp.417–435.
- Blumenthal, T., 1998. Gene clusters and polycistronic transcription in eukaryotes. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 20(6), pp.480–487.
- Bonenfant, D. et al., 2006. Characterization of histone H2A and H2B variants and their post-translational modifications by mass spectrometry. *Molecular & cellular proteomics : MCP*, 5(3), pp.541–552.
- Bowen, N.J. et al., 2004. Mi-2/NuRD: multiple complexes for many purposes. *Biochimica et biophysica acta*, 1677(1-3), pp.52–57.
- Bradner, J.E. et al., 2010. Chemical phylogenetics of histone deacetylases. *Nature chemical biology*, 6(3), pp.238–243.
- Brehove, M. et al., 2015. Histone core phosphorylation regulates DNA accessibility. *The Journal of biological chemistry*, 290(37), pp.22612–22621.
- Brenner, S. & Horne, R.W., 1959. A negative staining method for high resolution electron microscopy of viruses. *Biochimica et biophysica acta*, 34, pp.103–110.
- Brown, C.E. et al., 2000. The many HATs of transcription coactivators. *Trends in biochemical sciences*, 25(1), pp.15–19.
- Brubaker, K. et al., 2000. Solution structure of the interacting domains of the Mad-Sin3 complex: implications for recruitment of a chromatin-modifying complex. *Cell*, 103(4), pp.655–665.

- Byvoet, P. et al., 1972. The distribution and turnover of labeled methyl groups in histone fractions of cultured mammalian cells. *Archives of biochemistry and biophysics*, 148(2), pp.558–567.
- Camporeale, G. et al., 2007. K12-biotinylated histone H4 marks heterochromatin in human lymphoblastoma cells. *The Journal of nutritional biochemistry*, 18(11), pp.760–768.
- Carrozza, M.J. et al., 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell*, 123(4), pp.581–592.
- Chang, S. et al., 2004. Histone deacetylases 5 and 9 govern responsiveness of the heart to a subset of stress signals and play redundant roles in heart development. *Molecular and cellular biology*, 24(19), pp.8467–8476.
- Chen, H., Tini, M. & Evans, R.M., 2001. HATs on and beyond chromatin. *Current opinion in cell biology*, 13(2), pp.218–224.
- Chen, H.Y. et al., 1998. Ubiquitination of histone H3 in elongating spermatids of rat testes. *The Journal of biological chemistry*, 273(21), pp.13165–13169.
- Chen, S.J. et al., 1993. Rearrangements of the retinoic acid receptor alpha and promyelocytic leukemia zinc finger genes resulting from t(11;17)(q23;q21) in a patient with acute promyelocytic leukemia. *The Journal of clinical investigation*, 91(5), pp.2260–2267.
- Cheung, P. et al., 2000. Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Molecular cell*, 5(6), pp.905–915.
- Clark, M.D., Marcum, R., et al., 2015. Structural insights into the assembly of the histone deacetylase-associated Sin3L/Rpd3L corepressor complex. *Proceedings of the National Academy of Sciences of the United States of America*, 112(28), pp.201504021–E3678.
- Clark, M.D., Zhang, Y. & Radhakrishnan, I., 2015. Solution NMR Studies of an Alternative Mode of Sin3 Engagement by the Sds3 Subunit in the Histone Deacetylase-Associated Sin3L/Rpd3L Corepressor Complex. *Journal of molecular biology*, 427(24), pp.3817–3823.
- Clayton, A.L. et al., 2000. Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. *The EMBO journal*, 19(14), pp.3714–3726.
- Cloos, P.A.C. et al., 2006. The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. *Nature Publishing Group*, 442(7100), pp.307–311.

- Conway, E., Healy, E. & Bracken, A.P., 2015. PRC2 mediated H3K27 methylations in cellular identity and cancer. *Current opinion in cell biology*, 37, pp.42–48.
- Cornish, P.V. et al., 2008. Spontaneous intersubunit rotation in single ribosomes. *Molecular cell*, 30(5), pp.578–588.
- Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic acids research*, 16(22), pp.10881–10890.
- Cowley, S.M. et al., 2004. Functional analysis of the Mad1-mSin3A repressor-corepressor interaction reveals determinants of specificity, affinity, and transcriptional response. *Molecular and cellular biology*, 24(7), pp.2698–2709.
- Cowley, S.M. et al., 2005. The mSin3A chromatin-modifying complex is essential for embryogenesis and T-cell development. *Molecular and cellular biology*, 25(16), pp.6990–7004.
- Cress, W.D. & Seto, E., 2000. Histone deacetylases, transcriptional control, and cancer. *Journal of cellular physiology*, 184(1), pp.1–16.
- Crick, F., 1970. Central dogma of molecular biology. *Nature*, 227(5258), pp.561–563.
- Crick, F.H., 1958. On protein synthesis. *Symposia of the Society for Experimental Biology*, 12, pp.138–163.
- Cuff, J.A. et al., 1998. JPred: a consensus secondary structure prediction server. *Bioinformatics (Oxford, England)*, 14(10), pp.892–893.
- Dang, V.D. et al., 1999. A new member of the Sin3 family of corepressors is essential for cell viability and required for retroelement propagation in fission yeast. *Molecular and cellular biology*, 19(3), pp.2351–2365.
- Dannenberger, J.-H. et al., 2005. mSin3A corepressor regulates diverse transcriptional networks governing normal and neoplastic growth and survival. *Genes & development*, 19(13), pp.1581–1595.
- Das, T.K. et al., 2013. Sin3a acts through a multi-gene module to regulate invasion in *Drosophila* and human tumors. *Oncogene*, 32(26), pp.3184–3197.
- David, G. et al., 2006. Haploinsufficiency of the mSds3 chromatin regulator promotes chromosomal instability and cancer only upon complete neutralization of p53. *Oncogene*, 25(56), pp.7354–7360.
- David, G. et al., 2003. mSin3-associated protein, mSds3, is essential for pericentric heterochromatin formation and chromosome segregation in mammalian cells. *Genes & development*, 17(19), pp.2396–2405.
- David, G. et al., 2008. Specific requirement of the chromatin modifier mSin3B in cell cycle exit and cellular differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 105(11), pp.4168–4172.

- De Koning, L. et al., 2007. Histone chaperones: an escort network regulating histone traffic. *Nature structural & molecular biology*, 14(11), pp.997–1007.
- de Ruijter, A.J.M. et al., 2003. Histone deacetylases (HDACs): characterization of the classical HDAC family. *The Biochemical journal*, 370(Pt 3), pp.737–749.
- Denslow, S.A. & Wade, P.A., 2007. The human Mi-2/NuRD complex and gene regulation. *Oncogene*, 26(37), pp.5433–5438.
- Dixon, J.R. et al., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature Publishing Group*, 485(7398), pp.376–380.
- Dominguez, C., Boelens, R. & Bonvin, A.M.J.J., 2003. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, 125(7), pp.1731–1737.
- Dressel, U. et al., 2001. A dynamic role for HDAC7 in MEF2-mediated muscle differentiation. *The Journal of biological chemistry*, 276(20), pp.17007–17013.
- Drogaris, P. et al., 2012. Histone deacetylase inhibitors globally enhance h3/h4 tail acetylation without affecting h3 lysine 56 acetylation. *Scientific reports*, 2, p.220.
- Eissenberg, J.C. & Shilatifard, A., 2010. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Developmental biology*, 339(2), pp.240–249.
- Eot-Houllier, G. et al., 2009. Histone deacetylase inhibitors and genomic instability. *Cancer letters*, 274(2), pp.169–176.
- Espinás, M.L. et al., 2000. The GAGA factor of Drosophila interacts with SAP18, a Sin3-associated polypeptide. *EMBO reports*, 1(3), pp.253–259.
- Farias, E.F. et al., 2010. Interference with Sin3 function induces epigenetic reprogramming and differentiation in breast cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, 107(26), pp.11811–11816.
- Feng, Q. et al., 2002. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Current biology : CB*, 12(12), pp.1052–1058.
- Fernandez-Martinez, J., LaCava, J. & Rout, M.P., 2016. Density Gradient Ultracentrifugation to Isolate Endogenous Protein Complexes after Affinity Capture. *Cold Spring Harbor Protocols*, 2016(7), pp.pdb.prot087957–4.
- Finch, J.T. et al., 1977. Structure of nucleosome core particles of chromatin. *Nature*, 269(5623), pp.29–36.
- Fischer, D.D. et al., 2002. Isolation and characterization of a novel class II histone deacetylase, HDAC10. *The Journal of biological chemistry*, 277(8), pp.6656–6666.

- Fischle, W. et al., 2002. Enzymatic activity associated with class II HDACs is dependent on a multiprotein complex containing HDAC3 and SMRT/N-CoR. *Molecular cell*, 9(1), pp.45–57.
- Fischle, W. et al., 2008. Specificity of the chromodomain Y chromosome family of chromodomains for lysine-methylated ARK(S/T) motifs. *The Journal of biological chemistry*, 283(28), pp.19626–19635.
- Flanagan, J.F. et al., 2005. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature Publishing Group*, 438(7071), pp.1181–1185.
- Fleischer, T.C., Yun, U.J. & Ayer, D.E., 2003. Identification and characterization of three new components of the mSin3A corepressor complex. *Molecular and cellular biology*, 23(10), pp.3456–3467.
- Franke, D. & Svergun, D.I., 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of applied crystallography*, 42(Pt 2), pp.342–346.
- Frye, R.A., 2000. Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins. *Biochemical and biophysical research communications*, 273(2), pp.793–798.
- Gao, L. et al., 2002. Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family. *The Journal of biological chemistry*, 277(28), pp.25748–25755.
- Garcia, S.N. & Pereira-Smith, O., 2008. MRGing chromatin dynamics and cellular senescence. *Cell biochemistry and biophysics*, 50(3), pp.133–141.
- Garcia-Ramirez, M., Rocchini, C. & Ausio, J., 1995. Modulation of chromatin folding by histone acetylation. *The Journal of biological chemistry*, 270(30), pp.17923–17928.
- Ge, Q. et al., 1995. Molecular analysis of a major antigenic region of the 240-kD protein of Mi-2 autoantigen. *The Journal of clinical investigation*, 96(4), pp.1730–1737.
- Gecchele, E. et al., 2015. A comparative analysis of recombinant protein expression in different biofactories: bacteria, insect cells and plant systems. *Journal of visualized experiments : JoVE*, (97), pp.e52459–e52459.
- Giddings, J.C., 1968. Resolution and optimization in gel filtration and permeation chromatography. pp.1–7.
- Giorgetti, L. et al., 2010. Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Molecular cell*, 37(3), pp.418–428.
- Glozak, M.A. & Seto, E., 2007. Histone deacetylases and cancer. *Oncogene*, 26(37), pp.5420–5432.

- Goldknopf, I.L. et al., 1975. Isolation and characterization of protein A24, a “histone-like” non-histone chromosomal protein. *The Journal of biological chemistry*, 250(18), pp.7182–7187.
- Gopalan, G., Chan, C.S. & Donovan, P.J., 1997. A novel mammalian, mitotic spindle-associated kinase is related to yeast and fly chromosome segregation regulators. *The Journal of cell biology*, 138(3), pp.643–656.
- Goto, H. et al., 2002. Aurora-B phosphorylates Histone H3 at serine28 with regard to the mitotic chromosome condensation. *Genes to cells : devoted to molecular & cellular mechanisms*, 7(1), pp.11–17.
- Grandinetti, K.B. et al., 2009. Sin3B expression is required for cellular senescence and is up-regulated upon oncogenic stress. *Cancer research*, 69(16), pp.6430–6437.
- Gräff, J. & Tsai, L.-H., 2013. Histone acetylation: molecular mnemonics on the chromatin. *Nature reviews. Neuroscience*, 14(2), pp.97–111.
- Gregoret, I.V., Lee, Y.-M. & Goodson, H.V., 2004. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. *Journal of molecular biology*, 338(1), pp.17–31.
- Grozinger, C.M., Hassig, C.A. & Schreiber, S.L., 1999. Three proteins define a class of human histone deacetylases related to yeast Hda1p. *Proceedings of the National Academy of Sciences of the United States of America*, 96(9), pp.4868–4873.
- Grunstein, M., 1997. Histone acetylation in chromatin structure and transcription. *Nature*, 389(6649), pp.349–352.
- Grzenda, A. et al., 2009. Sin3: master scaffold and transcriptional corepressor. *Biochimica et biophysica acta*, 1789(6-8), pp.443–450.
- Guardiola, A.R. & Yao, T.-P., 2002. Molecular cloning and characterization of a novel histone deacetylase HDAC10. *The Journal of biological chemistry*, 277(5), pp.3350–3356.
- Guenther, M.G., Barak, O. & Lazar, M.A., 2001. The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Molecular and cellular biology*, 21(18), pp.6091–6101.
- Gurnier, A. & Fournet, G., 1947. Facteurs de correction dans les mesures de la diffusion des rayons X aux faibles angles J. P. Radium, ed., 8(12), pp.345–351.
- Guschin, D. et al., 2000. ATP-Dependent histone octamer mobilization and histone deacetylation mediated by the Mi-2 chromatin remodeling complex. *Biochemistry*, 39(18), pp.5238–5245.
- Haggarty, S.J. et al., 2003. Domain-selective small-molecule inhibitor of histone deacetylase 6 (HDAC6)-mediated tubulin deacetylation. *Proceedings of the National Academy of Sciences of the United States of America*, 100(8), pp.4389–

4394.

- Halkidou, K. et al., 2004. Nuclear accumulation of histone deacetylase 4 (HDAC4) coincides with the loss of androgen sensitivity in hormone refractory cancer of the prostate. *European urology*, 45(3), pp.382–9– author reply 389.
- Hansen, A.S. et al., 2018. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus (Austin, Tex.)*, 9(1), pp.20–32.
- Hartley, J.L., 2012. Why proteins in mammalian cells? *Methods in molecular biology (Clifton, N.J.)*, 801(Chapter 1), pp.1–12.
- Hassan, A.H., Neely, K.E. & Workman, J.L., 2001. Histone acetyltransferase complexes stabilize swi/snf binding to promoter nucleosomes. *Cell*, 104(6), pp.817–827.
- Hassig, C.A. et al., 1997. Histone deacetylase activity is required for full transcriptional repression by mSin3A. *Cell*, 89(3), pp.341–347.
- Hayakawa, T. & Nakayama, J.-I., 2011. Physiological roles of class I HDAC complex and histone demethylase. *Journal of biomedicine & biotechnology*, 2011, p.129383.
- He, Y. & Radhakrishnan, I., 2008. Solution NMR studies of apo-mSin3A and -mSin3B reveal that the PAH1 and PAH2 domains are structurally independent. *Protein science : a publication of the Protein Society*, 17(1), pp.171–175.
- Heinzel, T. et al., 1997. A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature*, 387(6628), pp.43–48.
- Hendrich, B. et al., 2001. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes & development*, 15(6), pp.710–723.
- Henson, P. & Walker, I.O., 1971. The Structure of nucleohistone. pp.1–4.
- Hildmann, C., Riester, D. & Schwienhorst, A., 2007. Histone deacetylases--an important class of cellular regulators with a variety of functions. *Applied microbiology and biotechnology*, 75(3), pp.487–497.
- Hochstrasser, M., 1996. Ubiquitin-dependent protein degradation. *Annual review of genetics*, 30(1), pp.405–439.
- Holding, A.N., 2015. XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods (San Diego, Calif.)*, 89, pp.54–63.
- Howe, L. et al., 1999. Histone acetyltransferase complexes and their link to transcription. *Critical reviews in eukaryotic gene expression*, 9(3-4), pp.231–243.
- Huang, J. & Berger, S.L., 2008. The emerging field of dynamic lysine methylation of non-histone proteins. *Current opinion in genetics & development*, 18(2), pp.152–158.

- Huang, Y., Myers, S.J. & Dingledine, R., 1999. Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. *Nature neuroscience*, 2(10), pp.867–872.
- Hubbert, C. et al., 2002. HDAC6 is a microtubule-associated deacetylase. *Nature*, 417(6887), pp.455–458.
- Hudson, G.M. et al., 2015. Insights into the Recruitment of Class IIa Histone Deacetylases (HDACs) to the SMRT/NCOR Transcriptional Repression Complex. *The Journal of biological chemistry*, 290(29), pp.18237–18244.
- Huen, M.S.Y. et al., 2007. RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. *Cell*, 131(5), pp.901–914.
- Hughes, A.L. & Rando, O.J., 2014. Mechanisms underlying nucleosome positioning in vivo. *Annual review of biophysics*, 43(1), pp.41–63.
- Humphrey, G.W. et al., 2001. Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *The Journal of biological chemistry*, 276(9), pp.6817–6824.
- Hurst, D.R., 2012. Metastasis suppression by BRMS1 associated with SIN3 chromatin remodeling complexes. *Cancer metastasis reviews*, 31(3-4), pp.641–651.
- Hutt, D.M., Powers, E.T. & Balch, W.E., 2009. The proteostasis boundary in misfolding diseases of membrane traffic. *FEBS letters*, 583(16), pp.2639–2646.
- Itoh, T. et al., 2015. Structural and functional characterization of a cell cycle associated HDAC1/2 complex reveals the structural basis for complex assembly and nucleosome targeting. *Nucleic acids research*, 43(4), pp.2033–2044.
- Janson, J.C. & Rydén, L., 1990. Protein purification. Principles, high resolution methods and applications., pp.207–208.
- Jarvis, D.L., 2009. Baculovirus–Insect Cell Expression Systems. In *Guide to Protein Purification, 2nd Edition*. Methods in Enzymology. Elsevier, pp. 191–222.
- Jelinic, P., Pellegrino, J. & David, G., 2011. A novel mammalian complex containing Sin3B mitigates histone acetylation and RNA polymerase II progression within transcribed loci. *Molecular and cellular biology*, 31(1), pp.54–62.
- Jenssen, H. & Aspö, S.I., 2008. Serum stability of peptides. *Methods in molecular biology (Clifton, N.J.)*, 494(Chapter 10), pp.177–186.
- Jepsen, K. et al., 2000. Combinatorial roles of the nuclear receptor corepressor in transcription and development. *Cell*, 102(6), pp.753–763.
- Jepsen, K. et al., 2008. Cooperative regulation in development by SMRT and FOXP1. *Genes & development*, 22(6), pp.740–745.

- Jiang, C. & Pugh, B.F., 2009. Nucleosome positioning and gene regulation: advances through genomics. *Nature Reviews Genetics*, 10(3), pp.161–172.
- Johnstone, R.W., 2002. Histone-deacetylase inhibitors: novel drugs for the treatment of cancer. *Nature reviews. Drug discovery*, 1(4), pp.287–299.
- Jones, A. et al., 2011. Purification of histone ubiquitin ligases from HeLa cells. *Methods (San Diego, Calif.)*, 54(3), pp.315–325.
- Joo, H.-Y. et al., 2007. Regulation of cell cycle progression and gene expression by H2A deubiquitination. *Nature Publishing Group*, 449(7165), pp.1068–1072.
- Kadamb, R. et al., 2013. Sin3: insight into its transcription regulatory functions. *European journal of cell biology*, 92(8-9), pp.237–246.
- Kaji, K. et al., 2006. The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nature cell biology*, 8(3), pp.285–292.
- Kang, Y. et al., 2014. HDAC8 and STAT3 repress BMF gene activity in colon cancer cells. *Cell death & disease*, 5(10), pp.e1476–e1476.
- Karle, I.L., 2001. Controls exerted by the Aib residue: helix formation and helix reversal. *Biopolymers*, 60(5), pp.351–365.
- Karle, I.L. et al., 1989. Modular design of synthetic protein mimics. Characterization of the helical conformation of a 13-residue peptide in crystals. *Biochemistry*, 28(16), pp.6696–6701.
- Kasten, M.M., Ayer, D.E. & Stillman, D.J., 1996. SIN3-dependent transcriptional repression by interaction with the Mad1 DNA-binding protein. *Molecular and cellular biology*, 16(8), pp.4215–4221.
- Kastner, B. et al., 2008. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nature methods*, 5(1), pp.53–55.
- Kelly, R.D.W. & Cowley, S.M., 2013. The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. *Biochemical Society transactions*, 41(3), pp.741–749.
- Keogh, M.-C. et al., 2005. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell*, 123(4), pp.593–605.
- Khavari, D.A., Sen, G.L. & Rinn, J.L., 2010. DNA methylation and epigenetic control of cellular differentiation. *Cell cycle (Georgetown, Tex.)*, 9(19), pp.3880–3883.
- Kim, Jaehoon et al., 2009. RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell*, 137(3), pp.459–471.
- Kim, Jeeseun et al., 2006. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO reports*, 7(4), pp.397–403.

- Kong, Q. et al., 2010. RNF220, an E3 ubiquitin ligase that targets Sin3B for ubiquitination. *Biochemical and biophysical research communications*, 393(4), pp.708–713.
- Kong, X.-X. et al., 2009. Function of SIRT1 in physiology. *Biochemistry. Biokhimiia*, 74(7), pp.703–708.
- Kornberg, R.D., 1974. Chromatin Structure: A repeating unit of histones and DNA. 184, pp.1–5.
- Kornberg, R.D. & Thomas, J.O., 1974. Chromatin Structure: Oligomers of the histones. (184), pp.1–5.
- Kouzarides, T., 2007. Chromatin modifications and their function. *Cell*, 128(4), pp.693–705.
- Kovacs, J.J., Hubbert, C. & Yao, T.-P., 2004. The HDAC complex and cytoskeleton. *Novartis Foundation symposium*, 259, pp.170–7– discussion 178–81– 223–5.
- Kulaeva, O.I., Hsieh, F.-K. & Studitsky, V.M., 2010. RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25), pp.11325–11330.
- Kumar, G.S. et al., 2011. Solution structure of the mSin3A PAH2-Pf1 SID1 complex: a Mad1/Mxd1-like interaction disrupted by MRG15 in the Rpd3S/Sin3S complex. *Journal of molecular biology*, 408(5), pp.987–1000.
- Kuppuswamy, M. et al., 2008. Role of the PLDLS-binding cleft region of CtBP1 in recruitment of core and auxiliary components of the corepressor complex. *Molecular and cellular biology*, 28(1), pp.269–281.
- Lagger, G. et al., 2002. Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *The EMBO journal*, 21(11), pp.2672–2681.
- Laherty, C.D. et al., 1997. Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell*, 89(3), pp.349–356.
- Laherty, C.D. et al., 1998. SAP30, a component of the mSin3 corepressor complex involved in N-CoR-mediated repression by specific transcription factors. *Molecular cell*, 2(1), pp.33–42.
- Lai, A. et al., 2001. RBP1 recruits the mSIN3-histone deacetylase complex to the pocket of retinoblastoma tumor suppressor family proteins found in limited discrete regions of the nucleus at growth arrest. *Molecular and cellular biology*, 21(8), pp.2918–2932.
- Larivière, L. et al., 2013. Model of the Mediator middle module based on protein cross-linking. *Nucleic acids research*, 41(20), pp.9266–9273.

- Lau, P.N.I. & Cheung, P., 2011. Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), pp.2801–2806.
- Le Guezennec, X., Vermeulen, M. & Stunnenberg, H.G., 2006. Molecular characterization of Sin3 PAH-domain interactor specificity and identification of PAH partners. *Nucleic acids research*, 34(14), pp.3929–3937.
- Lechner, T. et al., 2000. Sds3 (suppressor of defective silencing 3) is an integral component of the yeast Sin3[middle dot]Rpd3 histone deacetylase complex and is required for histone deacetylase activity. *The Journal of biological chemistry*, 275(52), pp.40961–40966.
- Lee, H. et al., 2006. Histone deacetylase 8 safeguards the human ever-shorter telomeres 1B (hEST1B) protein from ubiquitin-mediated degradation. *Molecular and cellular biology*, 26(14), pp.5259–5269.
- Lee, M.G. et al., 2005. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature Publishing Group*, 437(7057), pp.432–435.
- Leitner, A. et al., 2012. Expanding the chemical cross-linking toolbox by the use of multiple proteases and enrichment by size exclusion chromatography. *Molecular & cellular proteomics : MCP*, 11(3), pp.M111.014126–M111.014126.
- Lin, R.J. et al., 1998. Role of the histone deacetylase complex in acute promyelocytic leukaemia. 391, p.811(4).
- Liu, Z., Oughtred, R. & Wing, S.S., 2005. Characterization of E3Histone, a novel testis ubiquitin protein ligase which ubiquitinates histones. *Molecular and cellular biology*, 25(7), pp.2819–2831.
- Luger, K., Mäder, A.W., et al., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), pp.251–260.
- Luger, K., Rechsteiner, T.J., et al., 1997. Characterization of nucleosome core particles containing histone proteins made in bacteria. *Journal of molecular biology*, 272(3), pp.301–311.
- Luo, Y. et al., 2009. Trans-regulation of histone deacetylase activities through acetylation. *The Journal of biological chemistry*, 284(50), pp.34901–34910.
- Majerus, P.W., 1992. Inositol phosphate biochemistry. *Annual review of biochemistry*, 61(1), pp.225–250.
- Mariadason, J.M., 2008. HDACs and HDAC inhibitors in colon cancer. *Epigenetics : official journal of the DNA Methylation Society*, 3(1), pp.28–37.
- Martin, C. & Zhang, Y., 2005. The diverse functions of histone lysine methylation. *Nature reviews. Molecular cell biology*, 6(11), pp.838–849.

- Martin, R.G. & Ames, B.N., 1961. **Sucrose Gradient Centrifugation for Low Molecular Weight Substances: the Work of Bruce N. Ames**
 . pp.1372–1379.
- Mattanovich, D. et al., 2012. Recombinant protein production in yeasts. *Methods in molecular biology (Clifton, N.J.)*, 824(Chapter 17), pp.329–358.
- Mattioli, F., D'Arcy, S. & Luger, K., 2015. The right place at the right time: chaperoning core histone variants. *EMBO reports*, 16(11), pp.1454–1466.
- McDonel, P. et al., 2012. Sin3a is essential for the genome integrity and viability of pluripotent cells. *Developmental biology*, 363(1), pp.62–73.
- McKinsey, T.A. et al., 2000. Signal-dependent nuclear export of a histone deacetylase regulates muscle differentiation. *Nature*, 408(6808), pp.106–111.
- McKinsey, T.A., Zhang, C.L. & Olson, E.N., 2001. Identification of a signal-responsive nuclear export sequence in class II histone deacetylases. *Molecular and cellular biology*, 21(18), pp.6312–6321.
- Metzger, E. et al., 2008. Phosphorylation of histone H3 at threonine 11 establishes a novel chromatin mark for transcriptional regulation. *Nature cell biology*, 10(1), pp.53–60.
- Metzger, E. et al., 2010. Phosphorylation of histone H3T6 by PKC β (I) controls demethylation at histone H3K4. *Nature Publishing Group*, 464(7289), pp.792–796.
- Michishita, E. et al., 2005. Evolutionarily conserved and nonconserved cellular localizations and functions of human SIRT proteins. *Molecular biology of the cell*, 16(10), pp.4623–4635.
- Millard, C.J. et al., 2013. Class I HDACs share a common mechanism of regulation by inositol phosphates. *Molecular cell*, 51(1), pp.57–67.
- Millard, C.J. et al., 2016. The structure of the core NuRD repression complex provides insights into its interaction with chromatin. *eLife*, 5, p.e13941.
- Minsky, N. et al., 2008. Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nature cell biology*, 10(4), pp.483–488.
- Mittag, T., Kay, L.E. & Forman-Kay, J.D., 2010. Protein dynamics and conformational disorder in molecular recognition. *Journal of molecular recognition : JMR*, 23(2), pp.105–116.
- Montgomery, R.L. et al., 2007. Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes & development*, 21(14), pp.1790–1802.

- Moyal, L. et al., 2011. Requirement of ATM-dependent monoubiquitylation of histone H2B for timely repair of DNA double-strand breaks. *Molecular cell*, 41(5), pp.529–542.
- Murray, K., 1964. The occurrence of epsilon-N-methyl lysine in histones. *Biochemistry*, 3, pp.10–15.
- Nagy, L. et al., 1997. Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase. *Cell*, 89(3), pp.373–380.
- Nasmyth, K., Stillman, D. & Kipling, D., 1987. Both positive and negative regulators of HO transcription are required for mother-cell-specific mating-type switching in yeast. *Cell*, 48(4), pp.579–587.
- Nathan, D. et al., 2006. Histone sumoylation is a negative regulator in *Saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications. *Genes & development*, 20(8), pp.966–976.
- Nigi, I., Fairall, L. & Schwabe, J.W.R., 2017. Expression and Purification of Protein Complexes Suitable for Structural Studies Using Mammalian HEK 293F Cells. *Current protocols in protein science*, 90, pp.5.28.1–5.28.16.
- Nowak, S.J. & Corces, V.G., 2004. Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends in genetics : TIG*, 20(4), pp.214–220.
- Oberoi, J. et al., 2011. Structural basis for the assembly of the SMRT/NCOR core transcriptional repression machinery. *Nature structural & molecular biology*, 18(2), pp.177–184.
- Odom, A.R. et al., 2000. A role for nuclear inositol 1,4,5-trisphosphate kinase in transcriptional control. *Science (New York, N.Y.)*, 287(5460), pp.2026–2029.
- Olins, A.L. & Olins, D.E., 1983. Spheroid chromatin units (v bodies)., pp.1–1.
- Oudet, P., Gross-Bellard, M. & Chambon, P., 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, 4(4), pp.281–300.
- Ozdağ, H. et al., 2006. Differential expression of selected histone modifier genes in human solid cancers. *BMC genomics*, 7(1), p.90.
- Patra, S.K., Patra, A. & Dahiya, R., 2001. Histone deacetylase and DNA methyltransferase in human prostate cancer. *Biochemical and biophysical research communications*, 287(3), pp.705–713.
- Pedersen, M.T. & Helin, K., 2010. Histone demethylases in development and disease. *Trends in cell biology*, 20(11), pp.662–671.
- Pellegrino, J., Castrillon, D.H. & David, G., 2012. Chromatin associated Sin3A is essential

- for male germ cell lineage in the mouse. *Developmental biology*, 369(2), pp.349–355.
- Pengelly, A.R. et al., 2015. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes & development*, 29(14), pp.1487–1492.
- Petruk, S. et al., 2012. TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. *Cell*, 150(5), pp.922–933.
- Pham, A.D. & Sauer, F., 2000. Ubiquitin-activating/conjugating activity of TAFII250, a mediator of activation of gene expression in *Drosophila*. *Science (New York, N.Y.)*, 289(5488), pp.2357–2360.
- Phillips, D.M. & Johns, E.W., 1965. A fractionation of the histones of group F2A from calf thymus. *The Biochemical journal*, 94(1), pp.127–130.
- Pollina, E.A. & Brunet, A., 2011. Epigenetic regulation of aging stem cells. *Oncogene*, 30(28), pp.3105–3126.
- Portolano, N. et al., 2014. Recombinant Protein Expression for Structural Biology in HEK 293F Suspension Cells: A Novel and Accessible Approach. *Journal of visualized experiments : JoVE*, (92), pp.e51897–e51897.
- Qian, Y.W. & Lee, E.Y., 1995. Dual retinoblastoma-binding proteins with properties related to a negative regulator of ras in yeast. *The Journal of biological chemistry*, 270(43), pp.25507–25513.
- Qiu, Y. et al., 2006. HDAC1 Acetylation Is Linked to Progressive Modulation of Steroid Receptor-Induced Gene Transcription. *Molecular cell*, 22(5), pp.669–679.
- Ramírez, J. & Hagman, J., 2009. The Mi-2/NuRD complex: a critical epigenetic regulator of hematopoietic development, differentiation and cancer. *Epigenetics : official journal of the DNA Methylation Society*, 4(8), pp.532–536.
- Rao, S.S.P. et al., 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7), pp.1665–1680.
- Rea, S. et al., 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(6796), pp.593–599.
- Robzyk, K., Recht, J. & Osley, M.A., 2000. Rad6-dependent ubiquitination of histone H2B in yeast. *Science (New York, N.Y.)*, 287(5452), pp.501–504.
- Rosano, G.L. & Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in microbiology*, 5(631607), p.172.
- Roth, S.Y., Denu, J.M. & Allis, C.D., 2001. Histone acetyltransferases. *Annual review of biochemistry*, 70(1), pp.81–120.
- Ruthenburg, A.J. et al., 2007. Multivalent engagement of chromatin modifications by

- linked binding modules. *Nature reviews. Molecular cell biology*, 8(12), pp.983–994.
- Ryan, J. & Letai, A., 2013. BH3 profiling in whole cells by fluorimeter or FACS. *Methods (San Diego, Calif.)*, 61(2), pp.156–164.
- Saha, R.N. & Pahan, K., 2006. HATs and HDACs in neurodegeneration: a tale of disconcerted acetylation homeostasis. *Cell death and differentiation*, 13(4), pp.539–550.
- Sahu, S.C. et al., 2008. Conserved themes in target recognition by the PAH1 and PAH2 domains of the Sin3 transcriptional corepressor. *Journal of molecular biology*, 375(5), pp.1444–1456.
- Sauvé, D.M. et al., 1999. Phosphorylation-induced rearrangement of the histone H3 NH2-terminal domain during mitotic chromosome condensation. *The Journal of cell biology*, 145(2), pp.225–235.
- Schalch, T. et al., 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature Publishing Group*, 436(7047), pp.138–141.
- Schwarz, P.M. et al., 1996. Reversible oligonucleosome self-association: dependence on divalent cations and core histone tail domains. *Biochemistry*, 35(13), pp.4009–4015.
- Segal, E. et al., 2006. A genomic code for nucleosome positioning. *Nature Publishing Group*, 442(7104), pp.772–778.
- Sexton, T. et al., 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148(3), pp.458–472.
- Shahbazian, M.D. & Grunstein, M., 2007. Functions of site-specific histone acetylation and deacetylation. *Annual review of biochemistry*, 76(1), pp.75–100.
- Sharma, V. et al., 2008. Drosophila SIN3 is required at multiple stages of development. *Developmental dynamics : an official publication of the American Association of Anatomists*, 237(10), pp.3040–3050.
- Shi, Y. et al., 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, 119(7), pp.941–953.
- Shi, Y.-J. et al., 2005. Regulation of LSD1 histone demethylase activity by its associated factors. *Molecular cell*, 19(6), pp.857–864.
- Sievers, F. et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), pp.539–539.
- Signolet, J. & Hendrich, B., 2015. The function of chromatin modifiers in lineage commitment and cell fate specification. *The FEBS journal*, 282(9), pp.1692–1702.

- Silverstein, R.A. & Ekwall, K., 2005. Sin3: a flexible regulator of global gene expression and genome stability. *Current genetics*, 47(1), pp.1–17.
- Smith, S. & Stillman, B., 1991. Stepwise assembly of chromatin during DNA replication in vitro. *The EMBO journal*, 10(4), pp.971–980.
- Smolle, M. & Workman, J.L., 2013. Transcription-associated histone modifications and cryptic transcription. *Biochimica et biophysica acta*, 1829(1), pp.84–97.
- Song, F. et al., 2014. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science (New York, N.Y.)*, 344(6182), pp.376–380.
- Spadafora, C. et al., 1976. The DNA repeat lengths in chromatins from sea urchin sperm and gastrule cells are markedly different. *FEBS letters*, 69(1), pp.281–285.
- Spitz, F. & Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, 13(9), pp.613–626.
- Spronk, C.A. et al., 2000. The Mad1-Sin3B interaction involves a novel helical fold. *Nature structural biology*, 7(12), pp.1100–1104.
- Stark, H., 2010. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods in enzymology*, 481, pp.109–126.
- Sternberg, P.W. et al., 1987. Activation of the yeast HO gene by release from multiple negative controls. *Cell*, 48(4), pp.567–577.
- Sterner, D.E. & Berger, S.L., 2000. Acetylation of histones and transcription-related factors. *Microbiology and molecular biology reviews : MMBR*, 64(2), pp.435–459.
- Stewart, G.S. et al., 2009. The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. *Cell*, 136(3), pp.420–434.
- Strahl, B.D. & Allis, C.D., 2000. The language of covalent histone modifications. *Nature*, 403(6765), pp.41–45.
- Streubel, G. et al., 2017. Fam60a defines a variant Sin3a-Hdac complex in embryonic stem cells required for self-renewal. *The EMBO journal*, 36(15), pp.2216–2232.
- Struhl, K., 1998. Histone acetylation and transcriptional regulatory mechanisms. *Genes & development*, 12(5), pp.599–606.
- Suganuma, T. & Workman, J.L., 2011. Signals and combinatorial functions of histone modifications. *Annual review of biochemistry*, 80(1), pp.473–499.
- Sugiyama, K. et al., 2002. Aurora-B associated protein phosphatases as negative regulators of kinase activation. *Oncogene*, 21(20), pp.3103–3111.
- Syed, S.H. et al., 2010. Single-base resolution mapping of H1-nucleosome interactions

- and 3D organization of the nucleosome. *Proceedings of the National Academy of Sciences of the United States of America*, 107(21), pp.9620–9625.
- Tan, M. et al., 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, 146(6), pp.1016–1028.
- Tan-Wong, S.M. et al., 2012. Gene loops enhance transcriptional directionality. *Science (New York, N.Y.)*, 338(6107), pp.671–675.
- Tanese, N., 1997. Small-Scale Density Gradient Sedimentation to Separate and Analyze Multiprotein Complexes. pp.224–234.
- Tatsuka, M. et al., 1998. Multinuclearity and increased ploidy caused by overexpression of the aurora- and lpl1-like midbody-associated protein mitotic kinase in human cancer cells. *Cancer research*, 58(21), pp.4811–4816.
- Taunton, J., Hassig, C.A. & Schreiber, S.L., 1996. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science (New York, N.Y.)*, 272(5260), pp.408–411.
- Taverna, S.D. et al., 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature structural & molecular biology*, 14(11), pp.1025–1040.
- Tennen, R.I., Berber, E. & Chua, K.F., 2010. Functional dissection of SIRT6: identification of domains that regulate histone deacetylase activity and chromatin localization. *Mechanisms of ageing and development*, 131(3), pp.185–192.
- Tessarz, P. & Kouzarides, T., 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nature reviews. Molecular cell biology*, 15(11), pp.703–708.
- Thoma, F., 1992. Nucleosome positioning. *Biochimica et biophysica acta*, 1130(1), pp.1–19.
- Tse, C. et al., 1998. Disruption of higher-order folding by core histone acetylation dramatically enhances transcription of nucleosomal arrays by RNA polymerase III. *Molecular and cellular biology*, 18(8), pp.4629–4638.
- Tsukada, Y.-I. et al., 2006. Histone demethylation by a family of JmjC domain-containing proteins. *Nature Publishing Group*, 439(7078), pp.811–816.
- Turner, B.M., 2005. Reading signals on the nucleosome with a new nomenclature for modified histones. *Nature structural & molecular biology*, 12(2), pp.110–112.
- van Ingen, H. et al., 2006. Role of structural and dynamical plasticity in Sin3: the free PAH2 domain is a folded module in mSin3B. *Journal of molecular biology*, 358(2), pp.485–497.
- van Oevelen, C. et al., 2008. A role for mammalian Sin3 in permanent gene silencing.

Molecular cell, 32(3), pp.359–370.

Verdin, E., Dequiedt, F. & Kasler, H.G., 2003. Class II histone deacetylases: versatile regulators. *Trends in genetics : TIG*, 19(5), pp.286–293.

Verdine, G.L. & Hilinski, G.J., 2012. Stapled peptides for intracellular drug targets. *Methods in enzymology*, 503, pp.3–33.

Verma, R., Boleti, E. & George, A.J., 1998. Antibody engineering: comparison of bacterial, yeast, insect and mammalian expression systems. *Journal of immunological methods*, 216(1-2), pp.165–181.

Viiri, K.M. et al., 2009. DNA-binding and -bending activities of SAP30L and SAP30 are mediated by a zinc-dependent module and monophosphoinositides. *Molecular and cellular biology*, 29(2), pp.342–356.

Viiri, K.M. et al., 2006. SAP30L interacts with members of the Sin3A corepressor complex and targets Sin3A to the nucleolus. *Nucleic acids research*, 34(11), pp.3288–3298.

Villagra, A., Sotomayor, E.M. & Seto, E., 2010. Histone deacetylases and the immunological network: implications in cancer and inflammation. *Oncogene*, 29(2), pp.157–173.

Vinciguerra, M. et al., 2010. SirT1 in muscle physiology and disease: lessons from mouse models. *Disease models & mechanisms*, 3(5-6), pp.298–303.

Wada, M. et al., 2002. The second largest subunit of mouse DNA polymerase epsilon, DPE2, interacts with SAP18 and recruits the Sin3 co-repressor protein to DNA. *Journal of biochemistry*, 131(3), pp.307–311.

Waltregny, D. et al., 2005. Histone deacetylase HDAC8 associates with smooth muscle alpha-actin and is essential for smooth muscle cell contractility. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 19(8), pp.966–968.

Wang, X. et al., 2000. Acetylation increases the alpha-helical content of the histone tails of the nucleosome. *The Journal of biological chemistry*, 275(45), pp.35013–35020.

Wang, Z. et al., 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics*, 40(7), pp.897–903.

Wang, Z. et al., 2009. Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes. *Cell*, 138(5), pp.1019–1031.

Wang, Z.-T. et al., 2016. Histone deacetylase inhibitors suppress mutant p53 transcription via HDAC8/YY1 signals in triple negative breast cancer cells. *Cellular signalling*, 28(5), pp.506–515.

- Waterborg, J.H. & Matthews, H.R., 1984. Patterns of histone acetylation in *Physarum polycephalum*. H2A and H2B acetylation is functionally distinct from H3 and H4 acetylation. *European journal of biochemistry*, 142(2), pp.329–335.
- Watson, P.J. et al., 2016. Insights into the activation mechanism of class I HDAC complexes by inositol phosphates. *Nature communications*, 7, p.11262.
- Watson, P.J. et al., 2012. Structure of HDAC3 bound to co-repressor and inositol tetraphosphate. *Nature*, 481(7381), pp.335–340.
- Weake, V.M. et al., 2008. SAGA-mediated H2B deubiquitination controls the development of neuronal connectivity in the *Drosophila* visual system. *The EMBO journal*, 27(2), pp.394–405.
- Wei, Y. et al., 1999. Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell*, 97(1), pp.99–109.
- Wen, J.-D. et al., 2008. Following translation by single ribosomes one codon at a time. *Nature Publishing Group*, 452(7187), pp.598–603.
- Wen, Y.D. et al., 2000. The histone deacetylase-3 complex contains nuclear receptor corepressors. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13), pp.7202–7207.
- West, M.H. & Bonner, W.M., 1980. Histone 2B can be modified by the attachment of ubiquitin. *Nucleic acids research*, 8(20), pp.4671–4680.
- Woodcock, C.L., Frado, L.L. & Wall, J.S., 1980. Composition of native and reconstituted chromatin particles: direct mass determination by scanning transmission electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America*, 77(8), pp.4818–4822.
- Xiao, T. et al., 2005. Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Molecular and cellular biology*, 25(2), pp.637–651.
- Xie, T. et al., 2011. Structure of the 30-kDa Sin3-associated protein (SAP30) in complex with the mammalian Sin3A corepressor and its role in nucleic acid binding. *The Journal of biological chemistry*, 286(31), pp.27814–27824.
- Xu, C. et al., 2011. Improved Histone Deacetylase Inhibitors as Therapeutics for the Neurodegenerative Disease Friedreich's Ataxia: A New Synthetic Route. *Pharmaceuticals (Basel, Switzerland)*, 4(12), pp.1578–1590.
- Yang, Q. et al., 2000. The winged-helix/forkhead protein myocyte nuclear factor beta (MNF-beta) forms a co-repressor complex with mammalian sin3B. *The Biochemical journal*, 345 Pt 2, pp.335–343.
- Yang, X.-J. & Seto, E., 2003. Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression. *Current opinion in genetics & development*, 13(2), pp.143–153.

- Yang, Y. et al., 2010. TDRD3 is an effector molecule for arginine-methylated histone marks. *Molecular cell*, 40(6), pp.1016–1023.
- Yang, Z.R. et al., 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics (Oxford, England)*, 21(16), pp.3369–3376.
- Yoon, H.-G. et al., 2005. Reading and function of a histone code involved in targeting corepressor complexes for repression. *Molecular and cellular biology*, 25(1), pp.324–335.
- Yoshida, T. et al., 2008. The role of the chromatin remodeler Mi-2beta in hematopoietic stem cell self-renewal and multilineage differentiation. *Genes & development*, 22(9), pp.1174–1189.
- Youdell, M.L. et al., 2008. Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. *Molecular and cellular biology*, 28(16), pp.4915–4926.
- Yu, J. et al., 2003. A SANT motif in the SMRT corepressor interprets the histone code and promotes histone deacetylation. *The EMBO journal*, 22(13), pp.3403–3410.
- Yun, M. et al., 2011. Readers of histone modifications. *Cell research*, 21(4), pp.564–578.
- Zee, B.M. et al., 2010. In vivo residue-specific histone methylation dynamics. *The Journal of biological chemistry*, 285(5), pp.3341–3350.
- Zhang, P. et al., 2006. Structure of human MRG15 chromo domain and its binding to Lys36-methylated histone H3. *Nucleic acids research*, 34(22), pp.6621–6628.
- Zhang, T., Cooper, S. & Brockdorff, N., 2015. The interplay of histone modifications - writers that read. *EMBO reports*, 16(11), pp.1467–1481.
- Zhang, Y. et al., 1999. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes & development*, 13(15), pp.1924–1935.
- Zhang, Y. et al., 1997. Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell*, 89(3), pp.357–364.
- Zhang, Y. et al., 1998. SAP30, a novel protein conserved between human and yeast, is a component of a histone deacetylase complex. *Molecular cell*, 1(7), pp.1021–1031.
- Zhou, B.-R. et al., 2013. Structural insights into the histone H1-nucleosome complex. *Proceedings of the National Academy of Sciences of the United States of America*, 110(48), pp.19390–19395.
- Zhou, W. et al., 2008. Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Molecular cell*, 29(1),

pp.69–80.

Zhou, Y., Santoro, R. & Grummt, I., 2002. The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. *The EMBO journal*, 21(17), pp.4632–4640.

Zhu, P. et al., 2007. A histone H2A deubiquitinase complex coordinating histone acetylation and H1 dissociation in transcriptional regulation. *Molecular cell*, 27(4), pp.609–621.

Zhu, P. et al., 2004. Induction of HDAC2 expression upon loss of APC in colorectal tumorigenesis. *Cancer cell*, 5(5), pp.455–463.

Zinzen, R.P. et al., 2009. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature Publishing Group*, 462(7269), pp.65–70.