Investigation of the Sin3a-HDAC1-SDS3 transcriptional co-repressor complex

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

by

Nicola Portolano
College of Medicine, Biological Science and Psychology
Department of Biochemistry
University of Leicester

September 2014
Nicola Portolano

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

**Abstract**

The transcriptional co-repressor Sin3a is a ubiquitous eukaryotic protein complex that has a multitude of critical functions, including the regulation of embryonic development, cell division and maintenance of genomic integrity. It incorporates the highly related HDAC1 and HDAC2 enzymes as its catalytic subunits, which interact with the complex through the HID domain of the Sin3a co-repressor. Sin3a is responsible for deacetylating lysines of Histone tails, condensing chromatin and consequently repressing the transcription of genes. The enzymatic activities of HDAC1 and 2 within Sin3a depend on the association of the Sin3a-specific SDS3 protein, which also interacts with the complex via the HID domain. The mechanism by which Sin3a recruits HDAC1, HDAC2 and SDS3 remains unknown, and elucidating it would represent a big step forward in understanding the epigenetic regulation of genes through the deacetylation of chromatin.

The aim of this thesis is to understand how Sin3a recruits its catalytic subunits into the complex as well as to get a deeper insight into the role of SDS3 in the HID domain by using both structural (X-ray crystallography) and biochemical approaches.

Our data suggests that HDAC1 may interact with Sin3a through an extended surface of the co-repressor and that SDS3 stabilizes this interaction by simultaneously binding to HDAC1 and Sin3a. Enzymatic and kinetic assays indicate that Sin3a may be the only Class I HDAC containing complex that is not regulated by IP$_4$. IP$_4$ is a co-factor that regulates the activity of the Class I HDAC-dependent complexes NuRD and SMRT-NCoR. Thus, our results suggest that Sin3a may have followed a separate evolutionary pattern and its activity may be regulated in a different way.
Acknowledgments

I would like to sincerely thank my Ph.D. supervisors Dr. Shaun Cowley and Professor John Schwabe for supporting me under every technical and personal aspect of the course as well as my committee members Dr. Cyril Dominguez and Dr. Nick Barlev. A special thank you goes to Dr. Peter Watson, Dr. Louise Fairall and Dr. Christopher Millard for their invaluable help in the lab and for always keeping my moral and enthusiasm high. I would also like to extend my gratitude to Dr. Oliver Dovey and Dr. Charles Fosters for the warmest welcome a starting Ph.D. student can ever receive. You two guys made Leicester my second home, taught me to always stay positive no matter what life throws at you! I must also acknowledge all the lab members in 3/37, 1/40 and 1/45. In particular Dr. Kayoko Tanaka for her huge passion for science, which she transmitted daily to everyone in the lab. I am immensely grateful to my fellow Ph.D. students for four unforgettable years. In particular, I would like to mention Jesvin Samuel, my dear friend and housemate, and Aditya Chandru for bringing a good dose of healthy craziness to the lab! I would also like to thank Dr. Deep Dhani, Dr. Alessio Bortoluzzi and (almost Dr.) Robert Weinmeister for being really good friends and all the ones I won’t have enough space to mention. My deepest and most sincere appreciation for the unconditional support received goes to my parents, Luciano and Susy, and my sister, Mariagiovanna. I love you and it goes without question that none of this would have been possible without you.
List of contents:

Abstract: ........................................................................................................... II
Acknowledgments: ......................................................................................... III
List of contents: .............................................................................................. IV
Table of contents: ........................................................................................... IV
List of figures: .................................................................................................. X
List of tables: .................................................................................................. XIII
Abbreviations: ............................................................................................... XIV

Table of contents:

Chapter 1: Introduction: ............................................................................... 1

1.1 Introduction to the topic and chapter aims: ............................................ 1

1.1.1 Chromatin structure and epigenetic regulation: ............................... 3

1.1.2 Histone modifications and the regulation of transcription: ............... 9

1.1.3 Aberrant epigenome alterations and disease: ................................. 12

1.2 The Histone deacetylase superfamily: .................................................. 14

1.2.1 Classification of HDACs: ................................................................. 15

1.2.2 Class I HDACs: ................................................................................ 16

1.2.3 Class IIa HDACs: .............................................................................. 21

1.2.4 Class IIb HDACs: .............................................................................. 23

1.2.5 Class IV HDACs: .............................................................................. 23

1.2.6 Class III HDACs (Sirtuins): ............................................................ 24

1.3 Class I HDACs transcriptional co-repressor complexes: ..................... 25
1.3.1 The NuRD transcriptional co-repressor complex: .............................. 28
1.3.2 The CoREST transcriptional co-repressor complex: .......................... 29
1.3.3 The SMRT/NCoR transcriptional co-repressor complex: ..................... 29
1.3.4 The MiDAC transcriptional co-repressor complex: ............................ 30
1.3.5 The RIP140 transcriptional co-repressor complex: ............................ 31
1.4 The Sin3a transcriptional co-repressor complex: .................................. 31
1.4.1: The HDAC Interaction Domain (HID): ............................................. 34
1.4.2: Functions of the Sin3a complex: ...................................................... 36
1.5 Project aims and experiments overview: .............................................. 38

Chapter 2: Materials and Methods: ....................................................... 40
2.1.1 Mammalian expression vectors: ....................................................... 40
2.1.2 Primers: .......................................................................................... 41
2.1.3 Bacterial cell lines: .......................................................................... 41
2.1.4 Mammalian cell lines: ...................................................................... 42
2.1.5 Standard chemicals and reagents: ..................................................... 42
2.1.6 Commercial crystallization screens: .................................................. 42
2.1.7 Buffers and media: ............................................................................ 43
2.2 Monolayer HEK 293T cells: tissue culture and transfections: ................. 45
2.2.1 Preparation: ...................................................................................... 45
2.2.2 Starter cultures: ................................................................................ 45
2.2.3 Seeding cultures growing in 10 cm² tissue culture plates: .................... 45
2.2.4 Seeding cultures growing in 6 cm² tissue culture plates: ...................... 46
2.2.5 Transient transfections: .................................................................... 46
2.3 Suspension HEK 293F cells: tissue culture and transfections: ............... 47
2.3.1 Preparation: ........................................................................................................47
2.3.2 Starter cultures: ..................................................................................................47
2.3.3 Seeding cells using 250 ml conical cell culture flasks and 1 L roller bottles: ......................................................................................................................48
2.3.4 Gassing cells: .....................................................................................................48
2.3.5 Small-scale transient transfections: ...................................................................49
2.3.5 Large-scale transient transfections: .................................................................49
2.4 Immuno-precipitations (IPs) and Western blot analysis: ......................................50
2.4.1 Preparation of HEK 293T whole cell extracts: ................................................50
2.4.2 Preparation of HEK 293F whole cell extracts: ................................................50
2.4.3 Equilibration and immunization of protein-G Sepharose affinity resin: ..........51
2.4.4 Equilibration of anti-FLAG® affinity resin: .....................................................51
2.4.5 Pull-downs from HEK 293T whole cell extracts: .............................................51
2.4.6 Pull-downs from HEK 293F whole cell extracts: .............................................52
2.4.7 HRP Western blot analysis: ...............................................................................52
2.4.8 Fluorescence Western blot analysis: .................................................................54
2.5 Large-scale protein complex purifications: ..........................................................54
2.5.1 Transfections of HEK 293F suspension cells: ................................................54
2.5.2 Whole cell extracts from HEK 293F suspension cells: .....................................54
2.5.3 Equilibration of anti-FLAG® affinity resin: .....................................................55
2.5.4 Pull-down: .........................................................................................................55
2.5.5 TEV protease elution: ........................................................................................56
2.5.6 Size exclusion chromatography: .......................................................................56
2.5.7 Protein complex concentration and quantification: .........................................56
2.6 Deacetylase activity assay: .................................................................57
2.6.1 Protein complex expression and purification: ..............................57
2.6.2 Activity assay: ............................................................57
2.7 Crystallization trials: .................................................................57
2.7.1 Protein complex expression and purification: ..............................57
2.7.2 Setting of sitting drops crystallization plates: ............................58
2.7.3 Crystals cryoprotection, collection and freezing: ..........................58
2.7.4 Data collection: .....................................................................59
2.8 Binding assays: ........................................................................59
2.8.1 Protein complex expression and purification: ..............................59
2.8.2 Fluorescent polarization assay: ................................................59
2.8.3 Bio-layer interferometry assay: ................................................59
2.9 Cloning, DNA sequencing and DNA purification methods: ..........60
2.9.1 PCRs: ..............................................................................60
2.9.2 Agarose gel electrophoresis and PCR product extraction: ..........60
2.9.3 Cloning: ............................................................................61
2.9.4 Transformation of competent cells: ..........................................61
2.9.5 Small-scale plasmid DNA purifications: ....................................62
2.9.6 Large-scale plasmid DNA purifications: ....................................62
2.9.7 DNA quantification: ..............................................................64
2.9.8 Plasmid DNA sequencing: .....................................................64
2.10 Mass spectrometry: analysis .....................................................64

Chapter 3: Interactions among HDAC1, SDS3 and the HID domain of
Sin3a: .............................................................................................................65

3.1 Chapter aims: ............................................................................................65

3.2 HDAC1 interacts with the HID through multiple surfaces: .......................65
3.2.1 Expression of affinity tagged HDAC1: .......................................................65
3.2.2 Bioinformatics: designing constructs of the HID: ....................................66
3.2.3 Mapping the region of the HID responsible for interacting with HDAC1: .......
..........................................................................................................................70

3.3 SDS3 stabilized the HDAC1 interaction with Sin3a: .....................................74

3.4 Results summary and conclusions: ...............................................................78

Chapter 4: Complex purifications and crystallization trials: .........................80

4.1 Chapter aims: ............................................................................................80

4.2 Binary complex (HDAC1/Sin3a): .................................................................81
4.2.1 Small-scale purifications: .........................................................................81
4.2.2 Large-scale purifications and crystallization trials: ...................................90

4.3 Ternary complex: (HDAC1/Sin3a/SDS3): .......................................................93
4.3.1: Large-scale purifications and crystallization trials of the full ternary complex ...........................................................................................................95
4.3.2 Large-scale purifications of the ternary complex: analysis of nucleic acids impurities: .................................................................................................97
4.3.3 Crystallization trials with the full ternary complex cleared from the nucleic acid contaminants: .................................................................................105
4.3.4 Crystallization trials with the full ternary complex gel filtrated with a Superose®6 (S6) size exclusion chromatography column: ..............................114
4.3.5 Small-scale purifications of the ternary complex using shorter versions of the HID:...........................................................................................................118
4.3.6: Large-scale purifications and crystallization trials of the ternary complex with a construct of SDS3 that lacks the coiled-coil domain:.................................121
4.4 Results summary and conclusions:...............................................................132

Chapter 5: Activity and binding assays:.........................................................134

5.1 Chapter aims:..............................................................................................134
5.2 HDAC activity assays:..................................................................................134
5.2.1 Binary and ternary complex activity assays:..............................................136
5.3 Sin3a-IP₅ binding assays:............................................................................142
5.3.1 Fluorescent Polarization (FP) assay of IP₅ binding to the Sin3a ternary complex:.................................................................................................................143
5.3.2 Bio-layer interferometry assay of IP₅ binding to the Sin3a ternary complex:....................................................................................................................145
5.4 Results summary and conclusions:.............................................................147

Chapter 6: Discussion:.....................................................................................148

6.1 HDAC1 and its interaction within the HID domain of Sin3a:.......................148
6.1.1 HDAC1/HID/SDS3 interaction mapping experiments:............................148
6.1.2 Protein complex purifications and crystallization trials:..........................151
6.2 Sin3a: IP₅ dependent or independent?:.......................................................155
6.3 Summary and conclusions: transcriptional co-repressor complexes, chromatin and the regulation of life:.................................................................158

List of references:...........................................................................................161
List of figures:

1.1 Phases of chromatin condensation: ............................................................5
1.2 Opposing roles of HATs and HDACs in chromatin condensation: ..............15
1.3 Classification and domain organization of zinc-dependent HDACs: ...........16
1.4 Structures HDAC8 and HDAC2: ...............................................................17
1.5 Topology diagram of HDAC1: ..................................................................18
1.6 Best characterized Class I HDACs-containing complexes .......................20
1.7 Schematic diagram of Sirtuins: .................................................................25
1.8 Structures of HDAC3/SMRT and HDAC1/MTA1: ....................................27
1.9 IP$_4$ is indispensable for the interaction with the SANT domain but not for the interaction with the co-repressor: ....................................................28
1.10 Sin3a: domains schematic: .................................................................32
1.11 Sin3a PAH3 domain: .............................................................................32
1.12 Sin3a Associated protein 18: .................................................................36
2.1 PROTEX mammalian expression vector: .................................................41
2.2 Blotting scheme for protein transfer: ....................................................53
3.1 Expression of C-terminal flag-tagged HDAC1: .......................................66
3.2 Multiple sequence alignment of Sin3a (1-1273): ....................................67
3.3 Disorder and secondary structure predictions of the full length HID 524-855 (RONN and Jpred): .................................................................69
3.4 Interaction of HDAC1 with the C-terminal truncated HID constructs: ....71
3.5 Interaction of HDAC1 with C-terminal and N-terminal truncated HID constructs: .........................................................................................72
3.6 Interaction of HDAC1 with alpha-helices 4 to 7 of the HID domain of Sin3a:

3.7 Summary of the HID constructs used for the mapping experiments:

3.8 SDS3 interacts with the full length HID domain of Sin3a:

3.9 SDS3 cannot interact with the N-terminal portion of the HID unless HDAC1 is also overexpressed:

3.10 SDS3 cannot interact with the C-terminal portion of the HID unless HDAC1 is also overexpressed:

3.11 Summary of the interaction of SDS3 with the HID:

4.1 Small-scale purifications of binary and ternary complexes:

4.2 Large-scale binary complex purification (1):

4.3 Optimization of small-scale binary complex purification (1):

4.4 Optimization of small-scale binary complex purification (2):

4.5 Optimization of small-scale binary complex purification (3):

4.6 Large-scale binary complex purification (2):

4.7 Optimization of Stura & MacroSol B-10 condition:

4.8 Ternary complex small-scale versus-medium scale purifications:

4.9 Large-scale ternary complex purification:

4.9bis Large-scale ternary complex purification (bis):

4.10 Nucleic acids contaminants:

4.11 Full ternary complex S200 gel filtrations at different salt concentrations:

4.12 UV spectra of the gel filtration peaks at 18 ml (figure 4.11 C):

4.13 Effect of salt concentration and RNase A on the purity of the ternary
4.14 Ternary complex crystallization trials: first round of optimization:..........106
4.15 Ternary complex crystallization trials: second round of optimization:.......107
4.16 Crystals obtained with the second round of optimization:......................108
4.17 Ternary complex crystallization trials: third round of optimization:..........109
4.18 Crystals obtained with the third round of optimization:........................110
4.19 3 weeks old ternary complex from a JCSG plus screen (12 conditions): .................................................................112
4.20 Mass spectrometry analysis of 3 weeks old ternary complex (figure 4.19): ........................................................................................................................................................................................................................113
4.21 S6 gel filtrations at different salt concentrations:......................................115
4.22 Nucleic acid contaminants co-purified with the ternary complex:.............116
4.23 Comparison of 3 weeks old ternary complex from a JCSG plus screen and a sample stored at room temperature for the same amount of time:.........................117
4.24 Short HIDs: small-scale purifications:....................................................120
4.25 Disorder and secondary structure prediction of SDS3 (RONN and Jpred): ........................................................................................................................................................................................................122
4.26 Summary of alternative SDS3 constructs:.............................................123
4.27 Small-scale purifications of the ternary complex with alternative SDS3 constructs:........................................................................................................................................................................................................125
4.28 Example of a large-scale Δ-CC ternary complex purification:..................126
4.29 Crystals obtained with the Δ-CC ternary complex crystallization trials:......128
4.30 Δ-CC ternary complex crystallization trials: first round of optimization:....129
4.31 Δ-CC ternary complex crystal obtained with the original MIDEAS A7
condition and diffraction pattern:.........................................................................................................................130

4.32 Δ-CC ternary complex crystallization trials: second round of optimization:........................................................................................................................................................................131

4.33 3 weeks old Δ-CC ternary complex stored at room temperature:.........................132

5.1 SMRT and NuRD activation profiles:.................................................................................................................................135

5.2 HDAC assays: effect of high salt and IP₄ on the activity of the Sin3a complex:..................................................................................................................................................................................................................137

5.3 SMRT and Sin3a: direct comparison of IP₄-induced activation:........................................138

5.4 Effect of pre-incubating the ternary complex with IP₄ at different temperatures and for different amounts of time:.................................................................................................................................140

5.5 Comparison of the baseline activity of different transcriptional co-repressor complexes:...........................................................................................................................................................................................................142

5.6 Molecular details of the IP₄ interaction with the HDAC3/DAD complex:.................143

5.7 Fluorescent polarization assay:.................................................................................................145

5.8 Bio-layer interferometry assay:.................................................................................................................................147

6.1 SANT domains of SMRT and MTA1 (NuRD):.........................................................................................150

**List of tables:**

1.1 Post-translational modifications of Histone tails:.................................................................9

1.2 Reader domains of histone PTMs:.................................................................................................11

2.1 List of buffers and media and their components:.......................................................................43
**Abbreviations:**

aa: amino acid

Ada2: adaptor 2

ADD: ATRX-Dnmt3-Dnmt3L domain

BHC80: BRAF-35 HDAC complex protein

Boc-K: Boc-acetyl-lysine

Bp: base pair

BRMS: breast cancer metastasis suppressor

Bromo: Bromo domain

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

*E. coli: Escherichia coli*
E(n.n): embryonic day (n.n)
E2F: E2 transcription factor
ECL: enhanced chemioluminescence
EDTA: ethylenediaminetetraacetic acid
ELM2: Egl-27 and MTA1 homology 2 domain
ERR-α: estrogen-related receptor alpha
EtBr: ethidium bromide
FBS: foetal bovine serum
FOXP3: forkhead box P3
Gcn5: general control non repressed (HAT protein)
GPS2: G protein pathway suppressor
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
HEK 293F: human embryonic kidney 293 freestyle
HEK 293T: human embryonic kidney 293 SV40 large T-antigen
HEPES: 4-(2-hydroxyethyl)piperazine-1-ethanesulphonic acid
HID: HDAC Interaction Domain
HMT: homocystein S-methyltransferase
HRP: horseradish peroxidase
Ing2: inhibitor of growth
IP: immuno-precipitation
IP(n): inositol n phosphate
ISWI: imitation switch
K-Acetate: potassium acetate
Kbp: kilo base pair
kDa: kilo Daltons
KDAC: lysine deacetylase
L: litre
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 tumour 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: millilitre
MLL: mixed lineage leukaemia
mM: millimolar
MMP2/9: matrix metalloproteinases 2/9
MRG15: mortality factor related gene 15
mRNA: messenger RNA
MTA1-3: metastasis associated protein
N-terminal: amino terminal
NAD+: nicotinamide adenine dinucleotide
NCoR: nuclear receptor co-repressor
nm: nanometre
NMR: nuclear magnetic resonance
NuRD: nucleosome remodelling deacetylase complex
ºC: degrees centigrade
ºK: degrees kelvin
p16: protein 16 kDa
p53: protein 53 kDa
PAH: paired amphipathic helix
PBS phosphate buffer saline
PBS-T: phosphate buffer saline - tween
pCMV: cytomegalovirus promoter
PCR: polymerase chain reaction
PEG: polyethylene glycol
PEI: polyethylenimine
PHD: plant homeo domain

PPARγ: peroxisome proliferator-activated receptor gamma

PPG: polypropylene glycol

PSG: penicillin streptomycin glutamine

PTM: post translational modification

PWWP: proline-tryptophan-tryptophan-proline domain

Ras: rat sarcoma gene

RBBP: retinoblastoma binding protein

RIP140: receptor interacting protein 140

RNA: ribonucleic acid

Rpd3: reduced potassium dependency

Rpd3L: large Rpd3 complex (yeast)

Rpd3S: small Rpd3 complex (yeast)

RUNX2: runt-related transcription factor 2

S200: Superdex™200 gel filtration column

S6: Superose®6 gel filtration column

SANT: Swi3, Ada2, NCoR, TFIIIB domain

SAP: Sin3 associated protein

SDS-PAGE: SDS polyacrylamide gel electrophoresis

SDS: sodium dodecyl sulphate

SDS3: suppressor of defective silencing 3

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

SMRTER: SMRT-related ecdysone receptor-interacting factor

SWI/SNF: switch/ sucrose non fermentable

Swi3: switch 3

SWR1: SWI2/SNF2 related ATPase

T-X100: triton-X100

TAE: Tris-acetate-EDTA

TBL1: transducin (beta)-like protein

TCEP: tris (2-carboxyethyl) phosphine

tDMR: tissue-specific differentially methylated regions

TFIIB: transcription factor IIIB

Tris: tris (hydroxymethyl) aminomethane

v/v: volume/volume

v/w: volume/weight

WD40: tryptophan-aspartic acid repeat

x g: times gravity

Zf-CW: zinc finger CW

Δ-CC: coiled-coil domain deletion

µm: micrometre

µM: micromolar
CHAPTER 1: Introduction

1.1 Introduction to the topic and chapter aims.

Among the first people who ever wondered about the secrets of life were the ancient Greeks. Among them, Aristotle was most curious to know what life was all about and how it gets transmitted through the living generations. He was very keen to pass his knowledge about life to the people of his time, so much that today we sometimes refer to him as the “Father of Biology”. If, in a hypothetical world, we could time-travel and meet Aristotle in person, how could we communicate to him our understanding of life at a biological level? Probably, the vast majority of us would start by saying that it all begins with a molecule called DNA that contains all of the information necessary to create life. We would then continue explaining that DNA generates fac simile copies of itself, called RNA, and that RNA is then translated in to proteins that make us move, think, feel emotions and shape us. In other words, we would tell him that proteins are the muscles of life and are made thanks to the instructions stored into the DNA. He would probably be surprised to hear that organs of our body that are so different from each other, such us our brains and livers, actually share the exact same genetic information. Therefore, he might wonder and ask to us: “Then what controls the DNA? What tells it what to do and when?” Little would he know that this is a question that in 2014 we are still trying to answer! Indeed, each one of us may respond to this question in slightly different ways. Personally, however, I would say that DNA regulates its own destiny thanks to series of intricate mechanisms that today we refer to as epigenetics, which, paradoxically, means
beyond the genes, beyond DNA itself. Indeed, shortly after Watson and Crick’s discovery of the DNA double helix in 1953, it became clear that in order to explain certain phenotypical differences in a living population, biochemists had to shift their attention from the mere genetic sequence of DNA to some chemical and physical modifications of this molecule that did not affect its sequence of nucleotides. In fact, epigenetics studies the chemical and physical alterations of DNA that control its function without altering its genetic sequence.

It was discovered that the eukaryotic genome gets packed and organised in a structure known as chromatin. Nevertheless, it was found that chromatin was not to be considered as a mere tool only able to pack DNA, but it has a dynamic structure that can change shape and conformation in response to various requirements of the cell. Indeed, chromatin can exist in two main states: a condensed form known as “heterochromatin”, generally associated with gene inactivity, and a relaxed/open form generally characterised by transcriptional activation known as “euchromatin”. In fact, as will be discussed in the following paragraphs, structural and chemical changes on chromatin can determine which genes will be expressed and which will be silenced. The expression of genes is highly regulated by a vast array of enzymes and protein complexes that can shape and modify chromatin in such ways to enhance or inhibit transcription. Later in this chapter it will be discussed how the correct functioning of these enzymes is crucial for the maintenance of essential cellular functions.

The study of chromatin has been given increasing importance by the scientific community, which is gradually learning to look at the genome no longer as a meagre container of genetic information, but as a sort of biological highway with
a vast number of traffic signals and rules to control eukaryotic life at the cellular level. The aim of this chapter is to give an overview of chromatin and its properties as well as to gradually introduce a class of transcriptional co-repressor complexes that regulate gene expression through the deacetylation of histone tails. These complexes, and in particular the Sin3a complex, will be the main focus of this thesis.

1.1.1 Chromatin structure and epigenetic regulation
In 1978, it was discovered that inactive genes were less sensitive to DNase I degradation, while active genes were more susceptible to be digested by the enzyme (Sealy and Chalkley 1978; Vidali et al. 1978). This was the first evidence that inactive genes were being protected by some kind of structure that today we call the nucleosome. This consists of histone proteins (H) and DNA and it is a repeating unit of chromatin itself. The core nucleosome is formed by the interaction of 4 dimers of H2A, H2B, H3 and H4, which together form an octamer that can wrap 146 DNA base pairs (bp) as represented in Figure 1.1 (Luger, et al. 1997b; Luger, et al. 1997a; Zhou et al. 1998). Histone H1 is not part of the core nucleosome, but its function is to lock the DNA in place on the core (Daujat et al. 2005). N-terminal histone tails make up approximately 30% of the histones mass and are largely unstructured (Zheng and Hayes 2003). They protrude from the nucleosomes allowing for a large number of post-translational modifications (PTMs) that, by changing the conformation of chromatin, alter the accessibility of the transcription machineries to the DNA (Norton et al. 1989; Allfrey et al. 1964; LI Littau et al. 1964).
The “beads on a string” conformation is a model of a relaxed chromatin state that consists of nucleosomes separated from each other by strings of linker DNA, which vary between 10-80 bp depending on tissue type (Figure 1.1) (Felsenfeld and Groudine 2003). When PTMs on the lysines of histone tails are removed, lysine positive charges favor nucleosome/nucleosome interactions and, in synergy with H1 and other scaffold proteins, contribute to the condensation of chromatin into a structure called a poly-nucleosome array (Figure 1.1) (Hong et al. 1993; Zhou et al. 1998; Thoma et al. 1979). Poly-nucleosome arrays can in turn fold in to a helical conformation consisting of 6 nucleosomes per turn, forming a 30 nm fiber, which can finally fold into the metaphase chromosomal conformation (Figure 1.1) (Thoma et al. 1979; Finch and Klug 1976; Marsden and Laemmli 1979).
Indeed, the structure of chromatin is modified by the cell according to specific chromosomal functions as well as the different requirements during the various phases of the cell cycle. Euchromatin is mostly found in gene-rich genomic regions that are transcriptionally active (Gilbert et al. 2004). Genomic regions that can be differentially activated or silenced according to tissue type and during cell differentiation are referred to as facultative heterochromatin (Goto and Monk 1998). In contrast, genomic regions that have a low density of genes or that are
rich in interspersed repeated DNA sequences are normally repressed to avoid
dangerous recombinations that can lead to genome instability and disease.
These regions are referred to as constitutive heterochromatin (Weintraub and
Groudine 1976; Grewal and Jia 2007). Nevertheless, heterochromatin is not
always associated with a complete lack of activity. Indeed, it has been reported
that some siRNAs are generated from heterochromatic regions (Reinhart and
Bartel 2002).
Analogously to how DNase I is unable to attack DNA of condensed chromatin
conformations, transcription machineries can bind DNA only when this is
released from the nucleosome core particle (Allfrey et al. 1964; Littau et al. 1964).
Therefore, it appears clear that chemical alterations of chromatin that affect the
structure of chromatin itself play a crucial role in the regulation of gene
expression. It is possible to summarize that such alterations are implemented at
three different levels: nucleosome relocations, DNA methylation and histone
modifications.
Alterations in nucleosome positioning (often referred to as ‘remodeling’) are
regulated by 4 families of large ATP-dependent protein complexes, also known
as chromatin remodeling complexes. 1) The SWI/SNF family are master
regulators of gene expression (Carlson and Laurent 1994) and include the
SWI/SNF and RSC complexes (Clapier and Cairns 2009). 2) The ISWI family
mainly consists of transcriptional repressors (Tsukiyama et al. 1995). Members of
the ISWI family include ISW1a, ISW1b and ISW2 (Clapier and Cairns 2009). 3) Members of the CHD family contain chromo and helicase domains and are
involved in moving nucleosomes around chromatin (Stokes and Perry 1995) and
include the CHD1 complex in yeast and CHD1 and Mi2/NuRD complexes in insects and vertebrates (Clapier and Cairns 2009). 4) Finally the INO80 family includes the SWR1 and INO80 complexes and are involved in many important biological processes, including maintenance of genomic integrity, chromosomal segregation and DNA replication (Jin et al. 2005).

DNA methylation is probably the most studied epigenetic mark at present date and it is a strong signal for transcriptional inhibition. Indeed, it has been shown that DNA methylation can repress transcription by limiting the access of transcription factors to DNA (Molloy and Watt 1988). DNA methylation predominantly occurs on DNA regions known as CpG dinucleotides (Bird 1992). A CpG dinucleotide consists of a cytosine base followed by a guanine, where “p” indicates a single phosphodiester bond. DNMTs (DNA methyl transferases) are a class of enzymes responsible for adding one methyl group to the cytosines of CpG dinucleotides. Clustered sequences of CpG dinucleotides extending from 200-2000 bp in length are known as CpG islands (Bird 1986). CpG islands found in proximity of gene promoters are mostly unmethylated in mature somatic cells, but can be differentially methylated according to tissue type, forming genomic regions also known as “tissue differentially methylated regions”, or tDMRs (Rakyan et al. 2008). The vast majority of tissue-specific DNA methylation, however, occurs at CpG-rich regions of the genome that are found at approximately 2 kb away from CpG islands and that are known as CpG island shores (Doi et al. 2009). The inhibition of transcription induced by methylated CpG islands has several important biological functions, including genomic imprinting. This is caused by the hyper-methylation and silencing of one allele
and the consequent observation of recessive phenotypical traits (Reik and Lewis 2005). Moreover, methylated CpG islands can recruit chromatin modifying complexes (Wakefield et al. 1999; Free et al. 2001), indirectly causing chromatin condensation and transcriptional repression. For example, MeCP2 (Methylated CpG binding protein 2) is the best described methylated DNA binding protein and it has been shown to bind a number of chromatin modifiers such as Sin3a and NCoR (Cukier et al. 2008). MBD2 (Methylated DNA Binding Protein 2) is part of the NuRD complex (discussed in paragraph 1.3.1), which condenses chromatin through the deacetylation of histone tails (Millard et al. 2013). The interaction of methylated DNA binding proteins and chromatin modifiers is particularly relevant for methylated CpG islands located near transposon elements of the genome. Indeed, consequently to these interactions, transposons are condensed in a heterochromatin conformation, avoiding DNA damage and ensuring genome integrity (Weintraub and Groudine 1976; Grewal and Jia 2007). Haemimethylated DNA strands paired with newly synthesized DNA play an important role in DNA proofreading following DNA replication thanks to the interaction with Dnmt1 (DNA Methyl Transferase 1) and grant the inheritability of methylation markers in the daughter cell (Bashtrykov et al. 2012).

A different type of regulation of transcription and gene expression occurs by the action of histone modifying enzymes. Indeed, as previously mentioned, post-translational modifications of histone tails, in synergy with other chromatin modifying complexes, affect transcription by favoring chromatin condensation or relaxation. Histone PTMs include phosphorylation, acetylation, methylation and ubiquitination. These will be discussed in the following paragraph.
1.1.2 Histone modifications and the regulation of transcription

Histones possess unstructured N-terminus tails that protrude from the nucleosomes and are target for a number of PTMs that can either cause transcriptional activation or repression and are involved in functions such as DNA repair, chromatin remodeling and alternative splicing (Kouzarides 2007; Luco et al. 2010; Huertas et al. 2009). These modifications are specific to histone type and occur at arginine, lysine, serine, threonine and tyrosine residues. A summary of PTMs occurring on core histones is reported in table 1.1.

<table>
<thead>
<tr>
<th>Histone</th>
<th>PTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2A</td>
<td>S1 (Ph); K5 (Ac); K9 (Ac); K13 (Ac); K15 (Ac); K36 (Ac); K99 (Me); K119 (Me, Ub).</td>
</tr>
<tr>
<td>H2B</td>
<td>K5 (Ac, Me); K12 (Ac); S14 (Ph); K15 (Ac); K20 (Ac, Ub); K43 (Me); K85 (Ac); K108 (Ac); K116 (Ac); K120 (Ac, Ub).</td>
</tr>
<tr>
<td>H3</td>
<td>R2 (Me); T3 (Ph); K4 (Ac, Me); R8 (Me); K9 (Ac, Me); S10 (Ph); T11 (Ph); K14 (Ac); R17 (Me); K18 (Ac); K23 (Ac, Me); R26 (Me); K27 (Ac, Me); S28 (Ph); K36 (Ac, Me); K37 (Me); Y41 (Ph); T45 (Ph); K56 (Ac, Me); K79 (Me).</td>
</tr>
<tr>
<td>H4</td>
<td>S1 (Ph); R3 (Me); K5 (Ac); K8 (Ac); K12 (Ac, Me); K16 (Ac); S47 (Ph); K59 (Me); K77 (Ac); K79 (Ac); K91 (Ac); K92 (Me).</td>
</tr>
</tbody>
</table>

*Table 1.1: Post-translational modifications of histone tails.* Residues are represented with the one-letter code and are followed by their position within the histone. PTM are in brackets: Ac=Acetylation; Me=Methylation; Ph=Phosphorylation; Ub=Ubiquitination. Table derived from (Portela and Esteller 2010).

Heterochromatin shows low general levels of acetylation, while H3K9, H3K27 and H4K20 are highly methylated. In contrast, transcriptionally active chromatin is rich in H3K4, H3K36, H3K79 tri-methylation and acetylation, with active genes promoters being heavily tri-methylated at H3K4, mono-methylated at H4K20 and acetylated at H2BK5 and H3K27 (Li et al. 2007; Karlić et al. 2010). The body of
active genes are rich in methylated H3K79 and H4K20 (Karlić et al. 2010). Single post-translational modifications occurring on histone tails, however, cannot result in an immediate response in terms of gene expression or chromatin structure alterations. Indeed, outcomes are specified thanks to a crosstalk of co-existing PTMs in different histone tails or on the same tail or even on the same site (Nakanishi et al. 2009; Duan et al. 2008; Wang et al. 2008). Embryonic stem cells, which may require rapid transcriptional activation of genes in specific stages of development or differentiation, can present contradicting epigenetic marks in close proximity, such as tri-methylated H3K27 (repressive) and tri-methylated H3K4 (active). These are known as bivalent domains and are lost upon cellular differentiation (Bernstein et al. 2006).

Different PTMs of histone tails are read and interpreted by a class of proteins known as “reader proteins”, which bear domains able to recognize specific epigenetic marks. Although the list of proteins and domains may quickly grow in the next few years, a list of reader domains and their “targets” is reported in table 1.2. Among other functions, these proteins can recruit chromatin modifying complexes that can add or remove further histone tail PTMs, regulating transcription and, in turn, gene expression. An example is represented by MRG15, which can specifically recognize methylated H3K36 and recruits the HDAC-dependent Sin3a complex (Zhang et al. 2006; Yochum and Ayer 2002), deacetylating the 3’ ends of active genes and thus preventing the activation of cryptic promoters.
<table>
<thead>
<tr>
<th>PTM Mark</th>
<th>Reader domain</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetylation</td>
<td>Bromo</td>
<td>H3K14; H3K56; H4K16; H4K5; H4K8</td>
</tr>
<tr>
<td></td>
<td>PHD</td>
<td>H3K14</td>
</tr>
<tr>
<td>K-Methylation</td>
<td>PHD</td>
<td>H3K4; H3K9</td>
</tr>
<tr>
<td></td>
<td>WD40</td>
<td>H3K4; H3K9; H3K27; H4K20; H1K26</td>
</tr>
<tr>
<td></td>
<td>ADD</td>
<td>H3K4</td>
</tr>
<tr>
<td></td>
<td>Chromo</td>
<td>H3K4; H3K9; H3K23; H3K27; H3K36</td>
</tr>
<tr>
<td></td>
<td>Tudor</td>
<td>H3K4; H3K9; H4K20</td>
</tr>
<tr>
<td></td>
<td>MBT</td>
<td>H3K4; H4K20; H1K26</td>
</tr>
<tr>
<td></td>
<td>Zf-CW</td>
<td>H3K4</td>
</tr>
<tr>
<td></td>
<td>Ankyrin repeats</td>
<td>H3K9</td>
</tr>
<tr>
<td></td>
<td>PWWP</td>
<td>H3K36; H4K20</td>
</tr>
<tr>
<td>R-Methylation</td>
<td>Tudor</td>
<td>H3R17; H4R3</td>
</tr>
<tr>
<td></td>
<td>ADD</td>
<td>H4R3</td>
</tr>
<tr>
<td>Phosphorylation</td>
<td>(Gcn5)</td>
<td>H3S10</td>
</tr>
<tr>
<td></td>
<td>14-3-3</td>
<td>H3S10</td>
</tr>
</tbody>
</table>

Table 1.2: Reader domains of histone PTMs. The table is derived from Yun et al. (2011).

While histone methylation marks are read and regulated by high specificity enzymes that are able to target particular residues of specific histones, acetylation marks are added and removed by HATs and HDACs respectively, which do not have such a high specificity and are able to target multiple histone residues (Kouzarides 2007; Chi et al. 2010).
1.1.3 Aberrant epigenome alterations and disease

Epigenetic marks and multi-subunit protein complexes that coordinate to control gene expression must function in a very delicate equilibrium. If this equilibrium is lost, diseases ranging from cancer, neurodevelopmental and neurodegenerative disorders as well as autoimmune diseases may arise.

So far, cancer has been widely studied under the genetic point of view. However, it is becoming increasingly evident that the epigenetic regulation of the genome plays a crucial role also in this disease and, therefore, a new generation of drugs are being developed and used for the treatment for a large number of neoplasms (Kerl et al. 2013). The cancer epigenome is profoundly altered at the nucleosome positioning, DNA methylation and histone modification levels. Overall, cancer cells are heavily hypo-methylated in comparison to normal cells (Goelz et al. 1985). While transposons-rich DNA is hyper-methylated in healthy cells, cancer genomes are hypo-methylated at these regions, which are consequently found in euchromatin conformations promoting spurious chromosome recombinations and activation of endo-parasitic sequences (Goelz et al. 1985; Gaudet et al. 2003). Specific promoters of imprinted genes that require to be hyper-methylated in specific tissues result to be hypo-methylated in some cancers (Futscher et al. 2002; Bettstetter et al. 2005). However, cancer cells can also show hyper-methylation in CpG islands that would normally be required to be hypo-methylated, compromising important cellular functions such as apoptosis, DNA repair, cell cycle control and Ras signalling (as reviewed by Esteller 2007). The typical example of epigenetic aberration in cancer is the hyper-methylation of the
promoter of the p16 tumour suppressor gene, which negatively regulates the formation of the CyclinD-CDK4/6 complex (Bhagat et al. 2014). As mentioned in the previous paragraph, CpG island shores represent the main region of tissue specific methylation (Irizarry et al. 2009; Doi et al. 2009). It is therefore not particularly surprising that abnormal DNA methylation at CpG island shores are observed in a large number of cancers (Irizarry et al. 2009).

Histone modification marks are also altered in cancer, mainly characterized by massive loss of acetylated H4K16 and tri-methylated H3K4 (Fraga et al. 2005; Hamamoto et al. 2004). The loss of the acetylated H4K16 mark has been attributed to the mis-regulation of HDAC-containing complexes, while aberrations of tri-methylated H3K4 and mono-methylated H3K9 marks is attributed to the alteration of the expression profiles of both methyltransferase and demethylase enzymes (Chi et al. 2010). The H4K20 tri-methylation mark is also lost while there is a general gain of methylated H3K9 and tri-methylated H3K27 (Fraga et al. 2005; Kondo et al. 2007; Vire et al. 2006). In addition, aberrant histone tail methylation is thought to be caused by the action of aberrant MLL (Mixed Lineage Leukemia) fusion proteins which, in turn, are generated by spurious chromosomal recombination or DNA damage (Wang et al. 2009).

Tumor suppressor SWI/SNF complexes have been found to be down-regulated in approximately 1 out of 5 non-small cells lung carcinomas, indicating that mis-regulation of chromatin remodeling complexes can also favor the formation and progression of a number of malignancies (Medina and Sanchez-Cespedes 2008; Lin et al. 2007). Indeed, ARID1, a component of the SWI/SNF complex, is one of the most mutated proteins in cancer, suggesting that it acts as a tumor
suppressor (Wu and Roberts 2013).

Aside from cancer, epigenetic mis-regulation may also lead to neurological disorders and autoimmune diseases. For instance, aberrations in DNA methylation and histone modification patterns are involved in serious disorders such as the Alzheimer’s disease, Parkinson disease and Amyotrophic Lateral Sclerosis (Pieper et al. 2008; Urdinguio et al. 2009). Rheumatoid arthritis patients show reduced global levels of DNA methylation (Karouzakis et al. 2009), while abnormally increased levels of di-methylated H3K9 are observed in patients suffering from type I diabetes (Miao et al. 2008).

1.2 The Histone Deacetylase superfamily

When in 1968 the acetylation of ε-amino groups of histone lysines was observed for the first time (Gershey et al. 1968), very little was known about gene regulation. One year later, however, a class of enzymes responsible for removing the acetyl group from histone lysines was observed by Inoue and Fujimoto (1969). Because histones were the first observed substrates of these enzymes, Inoue and Fujimoto (1969) decided to name them Histone Deacetylases, or HDACs. Nevertheless, as it will be discussed later, today we know that some HDACs are able to catalyze the deacetylation of non-histone substrates; this is the reason why these enzymes are also known as KDACs, or Lysine Deacetylases. Taunton et al. (1996) purified the first HDAC (HDAC1) and found that it was related to the yeast Rpd3 enzyme, which had previously been identified to have a role in gene regulation by Vidal and Gaber (1991).

HDACs catalyze the opposite reaction catalyzed by HATs, or histone acetyl
transferases, condensing chromatin and consequently repressing transcription (as reviewed by Hassig and Schreiber 1997) (Figure 1.2).

**Figure 1.2:** Opposing roles of HATs and HDACs in chromatin condensation.

Due to their potential as drug targets for cancer and also thanks to the fast progress made in the chromatin field, HDACs have received a great deal of attention by the scientific community.

### 1.2.1 Classification of HDACs

De Ruijter et al. (2003), grouped the 18 mammalian HDAC enzymes into two main families: the *classical* HDACs (1-11), defined as the family of zinc-dependent histone deacetylases, and the seven NAD$^{+}$-dependent Sirtuins, or Sir2-like proteins that were discovered by Landry et al. (2000). Khochbin et al. (2001) and Gregoretti et al. (2004), subsequently made a phylogenetic classification of HDACs and have divided them into four main classes of zinc-dependent Histone deacetylases: Class I, Class II, and Class IV, with Class II enzymes being subdivided in Classes IIa and IIb (Figure 1.3). NAD$^{+}$-dependent Sirtuins were put in Class III. The following paragraphs will discuss the structural and functional diversities of the above-mentioned classes of HDACs.
1.2.2 Class I HDACs

Class I HDACs are ubiquitous enzymes and comprise HDAC1, HDAC2, HDAC3 and HDAC8, which are orthologues of the yeast HDAC, Rpd3. As visible from Figure 1.3, the four Class I HDACs share a common domain organization, with a structured HDAC domain and an unstructured C-terminal tail (Vidal and Gaber 1991). These tails bear numerous sites for post-translational modifications that might be involved in the activity regulation of these enzymes (Luo et al. 2009; H. Lee et al. 2004). A common characteristic of these enzymes is that, with the sole exception of HDAC8, they require to be recruited and activated by larger transcriptional co-repressor complexes (Watson et al. 2012b; Millard et al. 2013; Codina et al. 2005). From a structural point of view, they are characterized by a central β-sheet (consisting of the interactions of eight parallel β-strands) sandwiched by a 12 or more of alpha-helices that confer them their characteristic
globular shape (Figures 1.4 and 1.5) (Somoza et al. 2004; Bressi et al. 2010; Watson et al. 2012b; Millard et al. 2013).

**Figure 1.4:** *Structures of HDAC8 and HDAC2.* The red arrows point at the active sites of the enzymes.
HDAC1 and 2 are structurally and functionally highly similar enzymes so that they are often referred to as sister enzymes. They share 85% sequence identity and are recruited to the same transcriptional co-repressor complexes (Grozinger and Schreiber 2002) (Figure 1.6). Unsurprisingly, they have several overlapping functions and often substantial phenotypical changes can be observed only when both of the enzymes are deleted (Kelly and Cowley 2013). Indeed, deleting either one of the enzymes results in mRNA-independent up-regulation of the other (Senese et al. 2007). Both enzymes play a role in the DNA damage response and double knock out models show an increased incidence of DNA double strand breaks (Miller et al. 2010). Overlapping functions of HDAC1 and HDAC2 include...
roles in haematopoiesis, cell cycle regulation, accurate cell division and embryonic stem cells homeostasis (Wilting et al. 2010; Dovey et al. 2013; Jamaladdin et al. 2014), although distinct functions can also be observed. 

*HDAC1*, for example, appears to have a crucial role in the differentiation of mouse embryonic stem cells (Dovey et al. 2010b) and its deletion is embryonic lethal at E.9.5 (Lagger et al. 2002). HDAC2, however, does not seem to play an indispensible role in mouse embryonic stem cell differentiation (Dovey et al., 2010b), but gene knock out studies have shown that deletion of *HDAC2* results in lethal cardio-myopathies and that it is therefore not compatible with life (Montgomery et al. 2007). Nevertheless, Zimmermann et al. (2007), reported that mice expressing a catalytically inactive HDAC2 (also unable to interact with the Sin3b co-repressor complex) result in viable animals with decreased intestinal tumor rates and reduced body size. Guan et al. (2009), observed that neuron-specific deletion of *HDAC2* results in mice with increased learning capabilities and memory function. HDAC2 also appears to be involved in the down-regulation of genes that have a role in synaptogenesis and was found to be up-regulated in patients suffering from Alzheimer’s, confirming its importance for nervous tissues (Guan et al. 2009; Gräff et al. 2012).
**HDAC3** deletion impairs cardiac development, the DNA repair machinery and cell cycle control, as well as resulting in gastrulation defects, causing early embryonic lethality (Bhaskara et al. 2008; Montgomery et al. 2008). Liver-specific deletion of this gene resulted in viable mice with up-regulated PPARγ-2 and consequent severe hepatosteatosis due to an imbalance of the carbohydrate and lipid metabolisms (Knutson et al. 2008). HDAC3 has been reported to play an important role in mitosis and it is essential for correct cell division (Li et al. 2006). Analogous to the results of HDAC2, decreased levels of HDAC3 activity result in gain of memory function in mice (McQuown et al. 2011). While these roles are related to the catalytic activity of HDAC3, this enzyme also plays important HDAC-independent metabolic functions that are preserved as long as even a catalytically-inactive HDAC3 enzyme is able to interact with the NCoR complex (Sun et al. 2013).

HDAC8 is the smallest Class I HDAC and it is the only enzyme of this class that seems to function free in solution without the need to interact with a transcriptional co-repressor complex (as reviewed by Watson et al., 2012a). Gene knock out studies have shown that HDAC8 plays important roles in gene

---

**Figure 1.6: Best characterized Class I HDACs-containing complexes.** The figure was kindly provided by Dr. Shaun Cowley.
control during skull morphogenesis (Haberland et al. 2009). Wilson et al. (2010) have found that HDAC8 is directly involved in deacetylating the estrogen-related receptor alpha (ERR-α) nuclear receptor complex, increasing its affinity with DNA and activating ERR-α targets.

1.2.3 Class Ila HDACs

Class Ila HDACs are orthologues of the yeast Hda1 (Histone deacetylase 1) protein and include HDAC4, HDAC5, HDAC7 and HDAC9. Their structures and catalytic rates are significantly different from Class I HDACs. As visible from Figure 1.3, they possess a single HDAC catalytic domain that follows a long N-terminus tail that bears MEF2 and 14-3-3 binding motifs. They possess a much lower level of HDAC activity attributed to a point mutation in the active site (Y→H) (Lahm et al. 2007). Interaction of Class II HDACs with members of the MEF2 transcription factor family promotes their nuclear localization and consequently causes the repression of MEF2-induced transcription (Lemercier et al. 2000). In contrast, interaction of Class Ila HDACs with 14-3-3 proteins promotes their accumulation in the cytoplasm (Grozinger and Schreiber 2000), although their cytoplasmic roles and substrates remain largely unexplored.

HDAC4 interacts with the SMRT/NCoR complex along with HDAC3, which is thought to be largely responsible for the observed deacetylase activities related the SMRT/NCoR complex (Fischle et al. 2002). HDAC4 can play both proliferative and anti-proliferative roles in cancer. Indeed, Geng et al. (2006) have demonstrated that knocking down HDAC4 results in decreased cell viability, thanks to the generation of apoptotic cells in p53-negative epithelial cancers.
In contrast, Berns et al. (2004) found that HDAC4-targeting shRNAs prevented p53-dependent cell cycle arrest, highlighting HDAC4 anti-proliferative role. HDAC5 and HDAC9 are two highly related enzymes whose activities are required for correct cardiac development (Chang et al. 2004). Differential expression of HDAC5, HDAC7 and HDAC9 can determine the prognosis in medulloblastoma, childhood acute lymphoblastic leukemia and glioblastoma patients (Milde et al. 2010; Moreno et al. 2010; Lucio-Eterovic et al. 2008). HDAC9 is highly expressed in skeletal muscles and in the brain in the form of two distinct transcript variants that can both repress the activity of MEF2 (Petrie et al. 2003; Sparrow et al. 1999). HDAC9 is involved in the development of cortical neurons (Sugo et al. 2010), the development and homeostasis of suppressor T-cells (Tao et al. 2007) and muscle homeostasis (Zhang et al. 2002). HDAC5 has roles in the skeletal muscle and bone development. In particular, HDAC5 arrests myoblast differentiation in a MEF2-dependent manner (Lu et al. 2000) and is recruited with HDAC4 to the transcription factor RUNX2, resulting in the control of osteoblast maturation (Jeon et al. 2006; Harada and Rodan 2003; Kang et al. 2005). HDAC7 has a fundamental role in cardiovascular development and homeostasis (Chang et al. 2006). HDAC7 knock out mice have an early embryonic lethal phenotype due the disruption of endothelial cell-cell adhesion, causing lethal hemorrhages as a consequence of the weakening of blood vessels. Analogous to HDAC5, HDAC7 is also involved in the regulation of osteoblasts maturation in a catalytically-independent manner by binding and repressing the RUNX2 transcription factor (Jensen et al. 2008). High expression levels of HDAC7 are associated with poor prognosis in acute lymphoblastic
leukemia and pancreatic cancers (Moreno et al. 2010).

1.2.4 Class IIb HDACs

Class IIb HDACs only comprises HDAC6 and HDAC10. HDAC6 has a tandem deacetylase domain and a C-terminal Zinc-finger domain able to bind ubiquitin. The N-terminal deacetylase domain replaces the N-terminal tail of Class IIa HDACs; therefore Class IIb HDACs do not have any MEF2 or 14-3-3 binding motifs. HDAC10 has only one deacetylase domain that is homologous to the N-terminal catalytic site of HDAC6. The C-terminal Leucine-rich region of HDAC10 has functions yet to be characterized but that may be involved in the cytoplasmic accumulation of the enzyme (Tong et al. 2002; Yang and Seto 2008). HDAC6 is probably the major cytoplasmic histone deacetylase and, among other functions, it was found to be involved in the regulation of microtubule-dependent cell mobility (Boyault et al. 2007; Hubbert et al. 2002). Jung et al. (2012), have reported that HDAC6 promotes apoptotic death in liver cancer cells, therefore functioning as a tumor suppressor. HDAC10 functions remain largely unexplored, but Song et al. (2013), have found that it plays a role as tumor suppressor in cervical cancer through the inhibition of the MMP2 and MMP9 matrix metalloproteinases.

1.2.5 Class IV HDACs

The sole member of this class is HDAC11, which was placed into a separate category because it bears sequence similarities to both Rpd3 (Class I HDACs) and Hda1 (Class II HDACs). HDAC11 is highly conserved from bacteria to
humans (Gregoretti et al. 2004). Nevertheless, HDAC11 roles in the cell largely remain to be elucidated. Sahakian et al. (2014) found that HDAC11 plays a role in hematopoiesis regulation, while, through the interaction with HDAC6, it controls the activity of interleukin 10, an important modulator of the immune response (Villagra et al. 2009; Cheng et al. 2014).

### 1.2.6 Class III HDACs (Sirtuins)

As previously mentioned, HDACs belonging to this class are orthologues of the yeast Sir2 protein. There are a total of 7 Sirtuins (Sirt1-Sirt7) in mammals and they all have a central deacetylase domain sandwiched by a Zinc-binding domain and a Rossman-fold NAD\(^+\)-binding domain (Figure 1.7). They carry out roles in the nucleus, in the cytoplasm and in the mitochondria. Because their activities depend on the presence of NAD\(^+\), they effectively act as metabolic sensors (Houtkooper et al. 2012). Sirtuins convert NAD\(^+\) into nicotinamide, which can then act as feedback-inhibitor of the Sirtuin enzymes (Bitterman et al. 2002; Anderson et al. 2003). They have received great attention from the scientific community as they have great potential for the prevention of ageing-related diseases, they are directly involved in the control of metabolic homeostasis and their activation is generally believed to be beneficial for human health (as reviewed by Houtkooper et al. 2012).
1.3 Class I HDACs transcriptional co-repressor complexes

Histone deacetylation can take place when HDACs are recruited to appropriate chromatin targets. However, Class I HDACs do not possess any DNA binding motif, therefore, with the sole exception of HDAC8, they must be recruited into large complexes that co-repress transcription synergistically with other chromatin modifying proteins. These complexes are indeed known as HDAC-dependent transcriptional co-repressor complexes and they include Sin3a/b, NuRD, CoREST, SMRT/NCoR, MiDAC and RIP140, which can deliver HDACs to their chromatin substrates. The way in which HDACs are recruited to their co-repressor complexes is of great interest because of the potential of discovering new drugs that can selectively target specific complexes and thus would not have the side effects associated with using potent pan-HDAC inhibitors as cancer chemotherapeutics. Aasland et al. (1996) discovered the SANT (Swi3, Ada2, NCoR, TFIIIB) domain and Guenther et al. (2001) found that SANT1 of SMRT and NCoR is indispensible for their interaction with HDAC3 as well as for its activation, therefore this domain was later defined as the DAD, or Deacetylase
Activation Domain. Codina et al. (2005) solved the NMR structure of the DAD and found that it is composed of 4 alpha helices. 7 years later, Watson et al. (2012b), solved the crystal structure of HDAC3 in complex with the DAD domain of SMRT (Figure 1.8) and observed a conformational change of the DAD helical bundle upon HDAC3 binding. However, the structural diversity between the bound and unbound DAD might not be of significant biological significance. The same structure revealed that the interaction between HDAC3 and the corepressor takes place thanks to the presence of a molecule of D-myoinositol- (1,4,5,6)-tetrakisphosphate (inositol tetraphosphate or IP₄), which was later found to be essential for both the activation of the enzyme and its interaction with the DAD domain of the corepressor (Millard et al. 2013). This is particularly significant, as inositol phosphates are well known gene regulators. IP₄ mediates the interaction of HDAC3 and SMRT-DAD by interposing between the basic surfaces of the two proteins, allowing them to come together as an active complex (Watson, et al. 2012b). Later studies have demonstrated that IP₄ is not required for the interaction of HDAC3 with the SMRT complex (Figure 1.9). Indeed, the corepressor can bind the enzyme through an extension of the SANT domain that wraps around it (Figures 1.6 and 1.7) (Millard et al. 2013). Millard et al. (2013) solved the crystal structure of HDAC1 in complex with the ELM2-SANT domain of MTA1 (NuRD complex). The structure reveals that the ELM2 domain wraps around HDAC1 giving a major contribution to the interaction between the two proteins (Figures 1.6 and 1.7). Despite that IP₄ was not present in the structure due to the high salt crystallization conditions, mass spectrometry and enzymatic assays revealed that IP₄ is a structural component of the NuRD complex and it
can modulate its catalytic activity (Millard et al. 2013). Sequence conservation and selective targeting of HDAC complexes with HDAC inhibitors suggest that HDAC1 and HDAC2 may interact in a similar way within the CoREST and MiDAC complexes (Bantscheff et al. 2011), although, at present, the structures of these interactions have not yet been solved. Sin3a, nor any of its core subunits, show any sequence homology with the SANT domain, therefore the recruitment mechanism of HDAC1 and HDAC2 into this complex remain to be elucidated.

**Figure 1.8:** Structures of HDAC3/SMRT and HDAC1/MTA1. The enzymes are represented in blue, while their co-repressors in green. The actives sites are indicated with a red arrow. The co-factor visible in the HDAC3 structure is a molecule of IP$_4$. 
Figure 1.9: IP₄ is indispensible for the interaction with the SANT domain but not for the interaction with the co-repressor. (Adapted from Millard et al. 2013).

1.3.1 The NuRD transcriptional co-repressor complex

The Nucleosome Remodeling Deacetylase complex (NuRD or Mi-2/NuRD) (of which a cartoon is represented in Figure 1.6) is characterized by a catalytic core formed by IP₄-activated HDAC1 and HDAC2 (Millard et al. 2013), while other core subunits include Mi-2α/β, MTA1, MTA2, MTA3, RBBP4, RBBP7 and MBD2/3. The core components of this complex are interchangeable and form distinct complexes with specific functions (as reviewed by Bowen et al. 2004). NuRD is responsible for the regulation of B and T lymphocytes (Fujita et al. 2004; Williams et al. 2004), maintenance of hematopoietic stem cells during hematopoiesis (Yoshida et al. 2008; Harju-Baker et al. 2008) and cancer (as reviewed by Ramirez and Hagman 2009). MTA1 (Metastasis Associated Protein 1) can promote deacetylation of p53 and suppress p53-mediated apoptosis, promoting cancer cell proliferation (Moon et al. 2007). NuRD has also been reported to have functions in DNA methylation (Wade et al. 1999).
1.3.2 The CoREST transcriptional corepressor complex

The CoREST complex, or co-repressor of RE1-Silencing-Transcription factor (Figure 1.6) (Andres et al. 1999), was identified by You et al. (2001) as an HDAC-dependent transcriptional co-repressor with HDAC1 and HDAC2 as catalytic subunits that interact with CoREST (complex scaffold) through a SANT domain. However, You et al. (2001) have challenged the possibility that CoREST may actually interact with the REST transcription factor. Yang et al. (2006), have solved the structure of LSD1 (Lysine specific demethylase 1) in complex with CoREST, and in fact the complex is involved in the regulation of histone tail methylation (M. G. Lee et al. 2005). Foster et al. (2010) have shown that LSD1 regulates CoREST stability. Another core component of CoREST is CtBP1 (Kuppuswamy et al. 2008). Like the other Class I HDAC-containing complexes, CoREST is involved in regulating transcription through the deacetylation and demethylation of histone tails and controls the expression of neuronal genes and neuronal development (Lakowski et al. 2006; Abrajano et al. 2009).

1.3.3 The SMRT/NCoR transcriptional co-repressor complex

The Silencing Mediator of Retinoid or Thyroid hormone receptor (SMRT or NCoR-2) and the Nuclear Receptor Co-repressor (NCoR) are homologous proteins that share 40% identity and constitute Mega-Dalton complexes involved in gene regulation (Chen and Evans 1995). Its catalytic subunits consist of HDAC3 and HDAC4, although HDAC3 is probably responsible for the largest part of deacetylation attributed to this complex. The catalytic activity of HDAC3
entirely depends on the association with the DAD domain of SMRT (Figure 1.8) (Watson et al. 2012b; Millard et al. 2013; Kelly and Cowley 2013), while core subunits TBL1 and GPS2 target HDAC3 and HDAC4 to chromatin substrates, promoting transcriptional regulation (Oberoi et al. 2011). Deletion of the SMRT gene results in an embryonic lethal phenotype in the mouse (E16.5) and evidence suggests that SMRT is involved in the regulation of cardiogenesis and neuronal stem-cell self-renewal (Jepsen et al. 2008; Jepsen et al. 2007). Similarly, loss of NCoR results in early embryonic lethality (E15.5) due to severe aberrations of neuronal differentiation, T-cell defects, under-developed livers and anemia (Jepsen et al. 2000).

1.3.4 The MiDAC transcriptional co-repressor complex

The MiDAC (Mitotic deacetylase complex) is a novel co-regulator of transcription discovered by Bantscheff et al. (2011). Evidence suggests that the MiDAC complex catalytic subunits are HDAC1 and HDAC2, while other interacting partners are the ELM2 and SANT-containing protein MIDEAS (Mitotic Deacetylase-Associated SANT domain), likely to form the scaffold of the MiDAC complex, and DNTTIP1 (deoxynucleotidyltransferase-interacting protein) (Bantscheff et al. 2011). At present, functions of the MiDAC complex and its non-catalytic subunits remain unexplored, but since this complex is specifically expressed during mitosis (Bantscheff et al. 2011), it could imply that it has a role in cellular division.
1.3.5 The RIP140 transcriptional co-repressor complex

The nuclear Receptor-Interacting Protein 140 (NRIP140 or RIP140) complex interacts with HDAC1 and HDAC2 (Wei et al. 2000), DNMTs (Kiskinis et al. 2007), CtBP (Carboxyl-terminal Binding Protein) (Vo et al. 2001) and HMTs (Histone Methyl Transferases) (Kiskinis et al. 2007), and RIP140 is the complex scaffold. The main regulatory pathway of the RIP140 complex is through the interaction with liganded nuclear receptors, mainly resulting in the transcriptional regulation of metabolic pathways. Indeed, RIP140 has a central role in a multitude of biological functions including the regulation of glucose and lipid metabolism in skeletal muscle, in the heart and in the liver (Seth et al. 2007; Fritah et al. 2010; Herzog et al. 2007) and the regulation of the circadian clock (Poliandri et al. 2011).

1.4 The Sin3a transcriptional co-repressor complex

The Switch-independent 3 (homologue “a”) complex, or Sin3a transcriptional co-repressor complex, has the most conserved sequence of all Class I HDAC-containing complex from yeast to humans (Knoepfler and Eisenman 1999). It was identified for the first time in 1987 by two laboratories that conducted independent studies on the mating type switch in yeast (Sternberg et al. 1987; Nasmyth et al. 1987). The central scaffold of the complex is Sin3a, which consists of 1273 amino acids in humans. Domains of Sin3a include four Paired Amphipathic Helix (PAH) domains, a HID (HDAC Interaction Domain) domain found between PAH3 and PAH4, and a Highly Conserved Region (HCR) found immediately after PAH4 (Figure 1.10). PAH domains are four-helical bundle structures of approximately
100 amino acids (Figure 1.11) (He and Radhakrishnan 2008) and are separated by unstructured, yet highly conserved amino acid regions (Wang and Stillman 1990; Grzenda et al. 2009).

**Figure 1.10:** *Sin3a: domains schematic.* Domain amino acid boundaries: (PAH1: 119-189; PAH2: 295-383; PAH3: 459-526; HID: ~531-855; PAH4: 855-955; HCR: ~955-1273). (Boundaries derived from Cowley et al. 2004; Grzenda et al. 2009).

**Figure 1.11:** *Sin3a PAH3 domain.* This structure was solved by Xie et al. (2011) in complex with the SAP30 protein (not shown in the figure).

The main role of PAH domains is to form protein-protein interactions that recruit various gene-regulatory proteins. PAH1 was shown to interact with SMRT/NCoR (Alland et al. 1997; Heinzel, et al. 1997), although this interaction has not been confirmed. PAH2, on the other hand, is known to interact with the Sin3 Interaction Domain (SID) of the transcription factor Mad1. The NMR structure of this
complex was solved by Brubaker et al. (2000). PAH3 binds the Sin3a Associated Protein 30 (SAP30) and the NMR structure of this complex was solved by Xie et al. (2011). Yang et al. (2002) have shown that PAH4 may not be folding as a 4-helical bundle but may have a distinct structure and act as a scaffold for a number of other proteins to promote transcriptional repression. The protein sequence of Sin3a does not show homology with any known DNA binding motif, suggesting that chromatin targeting occurs in DNA-independent mechanisms (Wang and Stillman 1993) or through transient interactions with other transcription factors. Indeed, Sin3a is recruited to chromatin targets by a multitude of DNA-binding transcription factors, including SMRT/NCoR, p53, Mad-Max, Ikaros and Aiolos (Alland et al. 1997; Heinzel, et al. 1997; Murphy et al. 1999; (Schreiber-Agus et al. 1995; Ayer et al. 1995; Koipally et al. 1999). The catalytic core subunits of the Sin3a complex were identified by Laherty et al. (1997) to be HDAC1 and HDAC2, which interact with the complex scaffold through the HID domain. Suppressor of Defective Silencing 3 (SDS3 or SAP45) was identified as an integral component of the Sin3 complex by Lechner et al. (2000), Alland et al. (2002) and Fleischer et al. (2003). SDS3 is essential for the deacetylase activity of Sin3a and it has also been reported that the conditional deletion of the SDS3 gene impairs chromosome segregation and the formation of pericentric heterochromatin, probably due to the disruption of Sin3a deacetylase activity (Lechner et al. 2000; David et al. 2003). The mechanisms through which SDS3 ensures Sin3a deacetylase activity is currently unknown, although it has been suggested that SDS3 may stabilize the HID and HDAC1.
Sin3b is a paralogue of Sin3a and it is also very well conserved from yeast to humans and shows significant sequence similarities with Sin3a, particularly at the PAH domains, the HID and the HCR (Ayer et al. 1995; Halleck et al. 1995). The functions and interacting partners of Sin3b compared to Sin3a remain largely unexplored.

All of the core subunits of Sin3a cluster into a small region of the Sin3a scaffold that includes PAH3 and the HID. In addition to HDAC1, HDAC2 and SDS3, other core proteins are RBBP4, RBB7, SAP30 and SAP18 (Hassig et al. 1997; Zhang et al. 1997; Zhang et al. 1998). The role in the Sin3a context of other subunits, including SAP180, SAP130, SAP25, ING1, ING2, BRMS1 and RBP1 is largely unexplored and it remains unclear whether any of this may actually be part of a core complex. However, the existence of distinct Sin3 complex in yeast (Rpd3L and Rpd3S) suggests that these proteins may bind Sin3 to form complex variants with different specialized roles (Carrozza et al. 2005).

1.4.1 The HDAC Interaction Domain (HID)

As already mentioned, HDAC1, HDAC2 and SDS3 directly interact with Sin3a through the HID. Laherty et al. (1997) defined the HID as the Sin3a fragment that spanning amino acid 524 to amino acid 899. However, the exact boundaries of a minimal interaction region for SDS3 and HDAC1/2 are unknown. Currently there are no solved structured of the HID and the absence of homology regions with the SANT domain raises the question of how HDAC1 and HDAC2 are recruited to the Sin3a complex. As it will be extensively described in chapter 3,
the HID appears to be an alpha-helix-rich region of Sin3a, although their three-dimensional arrangement remains unknown. In contrast to HDAC1, HDAC2 and SDS3, which form direct interactions with the HID, RBBP4 and RBBP7 interaction with Sin3a is mediated by HDAC1, HDAC2 and SAP30 (Hassig et al. 1997; Zhang et al. 1999; Zhang et al. 1998). RBB4 and RBBP7 are dispensable for HDAC activity of the Sin3a complex, but they have been reported to be involved in the regulation of cell cycle through the suppression of Ras activity in yeast and can interact with histones H2A, H3 and H4 via a β-propeller motif (Taunton et al. 1996; Qian and Lee 1995); Vermeulen et al. 2006; Yoon et al. 2005). SAP18 is a core component of the Sin3a complex, and it has been reported to stabilize the interaction of HDAC1 with Sin3a (Zhang et al. 1997; Grzenda et al. 2009). Its NMR structure was solved by McCallum et al. (2006) (Figure 1.12). The exact position of its interaction with Sin3a is unknown although, since it stabilizes the interaction of HDAC1, it is possible to hypothesize that it may take place within the HID.
Figure 1.12: *Sin3 Associated Protein 18*. Loops are labeled in green; beta-strands are labeled in yellow and alpha helices in red.

Other proteins, including SAP130, SAP180 and MRG15 have been reported to interact with the HID domain and confer specific functions to Sin3a (Grzenda et al. 2009; (Fleischer et al. 2003; Nakayama et al. 2003; Carrozza et al. 2005).

### 1.4.2 Functions of the Sin3a complex

Functions of the Sin3a complex are probably the most studied among all of the HDAC-dependent transcriptional co-repressor complexes. Functions related to its regulation of transcription include cell survival, cell cycle regulation, protein stabilization, regulation of development, oncogene regulation, cellular senescence and metabolism.

Deletion of the *Sin3a* gene is early embryonic lethal (E3.5) in the mouse due to impairment of the DNA repairing mechanism and consequent high levels of DNA damage (Cowley et al. 2005; Dannenberg et al. 2005; McDonel et al. 2012). Deletion of *Sin3a* in mouse embryonic stem cells also causes high levels of DNA damage, which in turn cause G2 arrest and apoptosis (McDonel et al. 2012), indicating that Sin3a has a role in cell survival and cell cycle regulation. Cowley et al. (2005) have shown that Sin3a plays a critical role in embryonic and T-cell
development. Mouse embryonic fibroblasts (MEFs) lacking Sin3a show reduced cellular division and, similarly, T-cells-specific deletion of Sin3a in the mouse results in a massively reduced number of CD8 T-Cells and underdeveloped thymus (Cowley et al. 2005). Sin3a is also known to regulate male reproduction. Indeed, down regulating Sin3a in the male germline results in apoptosis-induced sterility (Pellegrino et al. 2012). Van Oevelen et al. (2008), have observed that Sin3a regulates genes related to muscle development and sarcomere function. Moreover, they found that deletion of Sin3b results in the up-regulation of Sin3a, in a similar fashion observed for HDAC1 and HDAC2 by (Dovey et al. 2010b).

The interaction of p53 with Sin3a prevents proteolytic turnover of p53 in response to cellular stress (Zilfou et al. 2001). A similar phenomenon is observed with the SMRTER complex, which is stabilized by the interaction with Sin3a (Pile et al. 2002). In contrast, c-Myc deacetylation by Sin3a promotes its proteolytic degradation, affecting the regulation of c-Myc targets (Nascimento et al. 2011). Sin3a interaction with p53 can induce senescence by repressing p53 targets (Bansal et al. 2011). Moreover, the interaction of Sin3a with MRG15 and E2F (proteins with established roles in senescence) are further evidence of the involvement of Sin3a in senescence (Grandinetti et al. 2009; Garcia and Pereira-Smith 2008). Barnes et al. (2010) showed evidence that Sin3a is involved in the regulation of mitochondrial metabolism. Finally, Sin3a is also involved in oncogene transformation and it has been implicated in a number of malignancies. Indeed, Dannenberg et al. (2005) report that Sin3a is indispensible for the survival of transformed cells, contributing to the proliferation of cancer cells. In contrast, Sin3a interaction with the Brest Cancer Metastasis suppressor 1
(BRMS1) mediates the suppression of metastasis (as reviewed by Hurst 2012). This evidence suggests that Sin3a can act both as a proliferative or anti-proliferative complex, similarly to HDAC4.

1.5 Project aims and experiments overview

The overall goal of the project is to characterize the interactions existing among HDAC1, the HID domain of Sin3a and SDS3. Understanding the molecular details of this interaction would benefit the general understanding of how Class I HDACs are recruited into their transcriptional co-repressor complexes. One of the main questions to answer is whether Sin3a is catalytically activated by IP4. Indeed, although there is no homology between the HID and SANT domains, the possibility that Sin3a can interact with inositol phosphate molecules remains to be explored. Therefore, the main aim of this study was to understand the activation mechanism of the Sin3a complex using X-ray crystallography.

In chapter 3 mapping experiments were carried out as an attempt to find a smaller region of interaction between HDAC1 and the HID to avoid using the full length HID for the crystallization trials. Chapter 4 reports the crystallization trials and purifications of a binary complex formed by HDAC1 and HID and a ternary complex formed by HDAC1, HID and SDS3 in mammalian cells. Finally, chapter 5 reports HDAC activity and kinetic assays performed on Sin3a ought to establish the activation mechanism of the complex and its affinity for IP4. Expression of protein complexes in mammalian systems (HEK 293F suspension cells) and their affinity purification using an immunized resin has been extensively proven to be a successful approach. Indeed, other members of the departments successfully
expressed and crystallized HDAC3 in complex with the DAD domain of the SMRT co-repressor (Watson et al. 2012b) and HDAC1 in complex with MTA1 as part of the NuRD complex (Millard et al. 2013). Their findings represented a major breakthrough in the chromatin remodeling and epigenetics fields, resulting in the publication of a Nature article and a Molecular Cell paper.
CHAPTER 2: Materials and Methods

2.1 Materials

2.1.1 Mammalian expression vectors

Mammalian expression vectors were provided by the PROTEX cloning facility of the University of Leicester and were based on pcDNA 3.1 plasmids with a pCMV promoter that allowed protein expression in mammalian systems. Amino-terminal affinity tags were spaced from the inserts by a TEV protease recognition sequence and in addition to a selective antibiotic resistance gene, the vectors allowed selecting for positive clones thanks to the presence of a sacB gene, which is lethal for the bacterial cells expressing it in sucrose agar plates. The vectors were also modified to allow ligase-free ligation using a BD In-Fusion kit. The example schematic of a PROTEX vector is shown in Figure 2.1.

A mammalian expression vector containing a carboxyl-terminal flag-tag was kindly provided by Dr. Oliver Dovey, who also designed it and purified it. This vector is also based on pcDNA 3.1 and has a pCMV promoter. Dr. Shaun Cowley kindly provided some pCS2+N-myc vectors used in chapter 3 with an sCMV IE94 promoter for the expression in mammalian systems.
Figure 2.1: PROTEX mammalian expression vector. pLEICS 12 is a mammalian expression vector with an N-terminal flag tag. By digesting the vector with EcoRI and EcoRV the tag can be fused to the insert. Instead, by digesting it with KpnI and EcoRI the tag is excised from the plasmid and it is possible to clone an untagged construct. pLEICS 20 shares the same structure of pLEICS 12 but has an N-terminal myc tag and was used for the experiments reported in chapter 3.

2.1.2 Primers

Primers were manually designed using the Enzyme X software and purchased either from Invitrogen or Eurofins MWG Operon. They were designed to have a 21 bp annealing region to the gene of interest plus a vector homology region (16 bp on forward primers and a 18 bp on reverse primers) for cloning the PCR product into an expression vector.

2.1.3 Bacterial cell lines

Competent DH5α E. coli cells were kindly provided by Dr. Oliver Dovey.
Alpha Select Bronze Efficiency competent cells were purchased from Bioline.

2.1.4 Mammalian cell lines

HEK 293T monolayer cells were provided by ACTT® and were cultured in M10 media (90% DMEM; 10% FBS; 1% PSG).

FreeStyle™ HEK 293F suspension cells are a Life Technologies™ product and were cultured in Gibco® FreeStyle™ 293F Expression Medium from Life Technologies™.

2.1.5 Standard chemicals and reagents

All chemicals and reagents were of analytical grade or higher and were purchased from Sigma-Aldrich, Fisher Scientific or Melford unless otherwise specified. Restriction enzymes were purchased from New England Biolabs. KOD Hot Start DNA Polymerase was supplied by Toyobo-Novagen. Lysozyme and porcine trypsin were supplied by Sigma Aldrich, while TEV protease, heat-treated RNase A and DNase were expressed and purified in house by Mrs 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).

2.1.6 Commercial crystallisation screens

The following crystallisation screens were purchased from Molecular Dimensions: JCSG-plus; ProPlex™; Stura and MacroSol; MIDAS™; PACT premier;
Morpheus®; NuR LBD™. Chemicals to make optimised screens were also purchased from Molecular Dimensions or as specified in paragraph 2.1.5.

### 2.1.7 Buffers and media

<table>
<thead>
<tr>
<th>Buffers and Media</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>2yt media</td>
<td>15 g/L Bacto Tryptone; 10 g/L Bacto Yeast Extract; 5 g/L NaCl</td>
</tr>
<tr>
<td>Alkaline lysis buffer</td>
<td>0.2 M NaOH; 1% SDS</td>
</tr>
<tr>
<td>Buffer res1</td>
<td>25 mM Tris pH 8.0; 50 mM glucose; 10 mM EDTA</td>
</tr>
<tr>
<td>Buffer res2</td>
<td>10 mM Tris pH 8.0; 10 mM EDTA</td>
</tr>
<tr>
<td>Buffer res3</td>
<td>10 mM Tris pH 8.0; 1 mM EDTA</td>
</tr>
<tr>
<td>Developer buffer</td>
<td>10 mg/ml porcine trypsin; 100 mM NaCl; 5% glycerol</td>
</tr>
<tr>
<td>Freezing media (HEK 293F)</td>
<td>10% DMSO; 90% serum-free media</td>
</tr>
<tr>
<td>Freezing media (HEK 293T)</td>
<td>10% DMSO; 40% FBS; 50% DMEM</td>
</tr>
<tr>
<td>Gel filtration buffer (50 mM salt)</td>
<td>50 mM K-acetate; 50 mM Tris pH 7.5; 0.5 mM TCEP</td>
</tr>
<tr>
<td>Gel filtration buffer (150 mM salt)</td>
<td>150 mM K-acetate; 50 mM Tris pH 7.5; 0.5 mM TCEP</td>
</tr>
<tr>
<td>Gel filtration buffer (300 mM salt)</td>
<td>300 mM K-acetate; 50 mM Tris pH 7.5; 0.5 mM TCEP</td>
</tr>
<tr>
<td>Gelatine solution</td>
<td>99.9% D-PBS; 0.1% porcine gelatine</td>
</tr>
<tr>
<td>HDAC Assay buffer</td>
<td>50 mM NaCl; 50 mM Tris pH 7.5; 5% glycerol</td>
</tr>
<tr>
<td>IP buffer</td>
<td>250 mM NaCl; 10 mM HEPES pH 8.0; 1 mM EDTA; 0.5% Igepal; protease inhibitors</td>
</tr>
<tr>
<td>LB media</td>
<td>10 g/L Bacto tryptone; 10 g/L NaCl; 5 g/L Bacto yeast extract</td>
</tr>
<tr>
<td>Lysis buffer</td>
<td>100 mM K-acetate; 50 mM Tris pH 7.5; 5% glycerol; 0.3% T-X100; protease Inhibitors</td>
</tr>
<tr>
<td>Lysozyme solution</td>
<td>Buffer res1 plus 10 mg of Lysozyme/ml</td>
</tr>
<tr>
<td>M10 media</td>
<td>DMEM (500 ml); 10% FBS; 1% PSG</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Milk blocking buffer</td>
<td>3% dry and fat-free milk made in PBS</td>
</tr>
<tr>
<td>Neutralisation buffer</td>
<td>3 M K-acetate/acetic acid pH 5.5</td>
</tr>
<tr>
<td>PEG solution</td>
<td>30% PEG 6000; 2.5 M NaCl</td>
</tr>
<tr>
<td>PEI solution</td>
<td>0.5 mg/ml PEI; H_2O (adjust pH to 7.0 with HCl)</td>
</tr>
<tr>
<td>Protein loading buffer (2 X)</td>
<td>0.5 M Tris pH 6.8; 4.4% (w/v) SDS; 20% (v/v) glycerol; 2% (v/v) β-mercaptoethanol; 0.01% bromophenol blue</td>
</tr>
<tr>
<td>Resin equilibration buffer</td>
<td>100 mM K-acetate; 50 mM Tris pH 7.5</td>
</tr>
<tr>
<td>TAE Buffer</td>
<td>40 mM Tris; 20 mM acetic acid; 1 mM EDTA</td>
</tr>
<tr>
<td>TEV cleavage buffer</td>
<td>50 mM K-acetate; 50 mM Tris pH 7.5; 0.5 mM TCEP</td>
</tr>
<tr>
<td>Tobin's buffer</td>
<td>400 mM glycine; 50 mM glycine; 20% methanol</td>
</tr>
<tr>
<td>Trypsin solution</td>
<td>D-PBS plus: 0.5 mM EDTA; 1% v/v chicken serum; 2.5% w/v trypsin</td>
</tr>
<tr>
<td>Wash buffer 1 (low salt)</td>
<td>50 mM K-acetate; 50 mM Tris pH 7.5; 5% glycerol; 0.3% T-X100</td>
</tr>
<tr>
<td>Wash buffer 2 (low salt)</td>
<td>150 mM K-acetate; 50 mM Tris pH 7.5; 5% glycerol</td>
</tr>
<tr>
<td>Wash buffer 3 (low salt)</td>
<td>25 mM K-acetate; 50 mM Tris pH 7.5; 5% glycerol</td>
</tr>
<tr>
<td>Wash buffer 1 (normal salt)</td>
<td>100 mM K-acetate; 50 mM Tris pH 7.5; 5% glycerol; 0.3% T-X100</td>
</tr>
<tr>
<td>Wash buffer 2 (normal salt)</td>
<td>300 mM K-acetate; 50 mM Tris pH 7.5; 5% glycerol</td>
</tr>
<tr>
<td>Wash buffer 3 (normal salt)</td>
<td>50 mM K-acetate; 50 mM Tris pH 7.5; 5% glycerol</td>
</tr>
</tbody>
</table>

**Table 2.1:** List of buffers and media and their components.
2.2 Monolayer HEK 293T Cells: tissue culture and transfections.

2.2.1 Preparation
A Class II laminar flow hood was utilised for growing, maintaining and transfecting HEK 293T monolayer cells in sterile conditions. Both the hood and the required working materials were carefully sterilised with 70% ethanol. An appropriate number of tissue culture plates was pre-coated with gelatine solution for at least 20 minutes before plating the cells and an aliquot of M10 media was pre-warmed at 37ºC for at least 3 hours prior to starting any cell culture work.

2.2.2 Starter Cultures
Cryovials containing 1 ml of $8 \times 10^6$ cells/ml stored in freezing media were removed from the liquid N$_2$ and rapidly thawed at 37ºC. Cells were re-suspended into 8-10 ml of media and spun down at 1000 x g for 1 minute. After discarding the supernatant, the cells were suspended again into 1 ml of media, pipetted into a 10 cm$^2$ tissue culture plate containing 14 ml of fresh pre-warmed M10 and placed in a humidified incubator at 37ºC and 5% CO$_2$. Media was replaced after 24 hours.

2.2.3 Seeding cultures grown into 10 cm$^2$ tissue culture plates
HEK 293T monolayer cells were typically sub-cultured when they had reached an approximate confluence of 85-90%. The media was removed from the plates and the cells were washed once with 10 ml D-PBS before they were detached from the plate with 2 ml of trypsin buffer (see table 2.1 for buffer details) pre-warmed
at 37°C. Trypsin was neutralised by re-suspending the cells into 8 ml of media. Finally, cells were split 1 in 5 into one or more fresh pre-gelatinised 10 cm² tissue culture plates containing 14 ml of media and then incubated in a humidified incubator at 37°C and 5% CO₂.

2.2.4 Seeding cultures grown into 6 cm² tissue culture plates

Once an 85-90% confluence was reached, the media was aspirated from the plates and the cells were washed once with 5 ml of D-PBS prior to detaching them from the plate with 1 ml of trypsin solution (see table 2.1 for buffer details) pre-warmed at 37°C. Trypsin was neutralised by re-suspending the cells into 6 ml of M10 media. Finally, cells were split 1 in 5 into one or more fresh pre-gelatinised 6 cm² tissue culture plates containing 7 ml of media and then incubated in a humidified incubator at 37°C and 5% CO₂.

2.2.5 Transient transfections

24 hours before a transfection experiment, 1 x 10⁶ cells were seeded into 1 or more pre-gelatinised 6 cm² tissue culture plates containing 6 ml of media and incubated for 24 hours into a humidified incubator at 37°C and 5% CO₂. After 24 hours, one or more plasmids were diluted into 500 µl of plain DMEM into a 1.5 ml centrifuge tube (2 µg of plasmid per construct were used). An appropriate amount of Lipofectamine 2000 or PEI (table 2.1) was diluted into 500µl of plain DMEM into a separate 1.5 ml centrifuge tube (Lipofectamine 2000 was used in a 1:2 ratio per µg of DNA; PEI was used in a 1:4 ratio per µg of DNA). The diluted DNA and the transfection reagent were mixed together into a
fresh 1.5 ml centrifuge tube, vortexed briefly and incubated at room temperature for 20-40 minutes. The reaction mix was then pipetted into the cells seeded 24 hours earlier and mixed carefully by gently swirling the plate(s) for a few seconds. Transfected cells were incubated for 48 hours into a humidified incubator at 37°C and 5% CO₂.

2.3 Suspension HEK 293F Cells: tissue culture and transfections.

2.3.1 Preparation

A Class II laminar flow hood was utilised for growing, maintaining and transfecting HEK 293F monolayer cells in sterile conditions. Both the hood and the required working materials were carefully sterilised with 70% ethanol. Serum-free expression media was used to grow this cell line and was pre-warmed in a 37°C water bath at least one hour prior to starting any cell culture work.

2.3.2 Starter Cultures

Cryovials containing 1 ml of 2 x 10⁷ cells/ml stored in freezing media (10% DMSO, 90% serum-free media) were removed from the liquid N₂ and rapidly thawed at 37°C. Any clumps were broken by gentle pipetting and the cells were transferred into a 250 ml conical cell culture flask containing 34 ml of media. The flask was sealed with a vented cap and the culture incubated in a humidified shaking incubator at 37°C, 5% CO₂ and 120 rpm. Once they had reached a density of approximately 1 x 10⁶ cells/ml, cells were split to a final count of 0.35 x 10⁶ cells/ml into a 250 ml conical cell culture flask containing 60 ml of media and
passaged at least 3 times before any transfection experiment was attempted.

2.3.3 Seeding cells using 250 ml conical cell culture flasks and 1 L roller bottles

HEK 293F suspension-grown cells were typically sub-cultured every 48 hours or when they had reached a final density of $2 \times 10^6$ cells/ml. Cell density was calculated using a glass haemocytometer and viability was determined by trypan-blue staining (typical cell viability was above 95%).

30-60 ml cultures were grown in 250 ml conical cell culture flasks sealed with vented caps. A volume of cells sufficient to reach a final count of approximately $0.35 \times 10^6$ cells/ml was pipetted into one or more flasks and the final volume was made up to 30 to 60 ml by pipetting an appropriate volume of media. The cultures were then incubated into a humidified orbital shaking incubator at 37°C, 5% CO$_2$ and 120 rpm.

300 ml cultures were grown into 1 L roller bottles sealed with vented caps and placed into a humidified orbital shaking incubator at 37°C, 5% CO$_2$ and 135 rpm. An appropriate volume of cells was poured into one or more bottles and a final density of approximately $0.35 \times 10^6$ of cells/ml was reached by pouring fresh media to bring the volume of the culture up to 300 ml.

2.3.4 Gassing cells

Occasionally, standard non-CO$_2$ and non-humidified shaking incubators were used. In this case, the atmosphere above the cultures was replaced with 5% CO$_2$ gas by the use of a 5 ml Stripette connected to a 0.22 μm-filtered line of tubing.
fed by a 5% CO₂ gas cylinder. Cultures were typically gassed for 1 or 2 minutes.

2.3.5 Small-scale transient transfections

24 hours before a transfection experiment cells were seeded to a final count of 0.5 x 10⁶ viable cells/ml into an appropriate number of 250 ml conical cell culture flasks and incubated for 24 hours into a humidified shaking incubator at 37°C, 5% CO₂ and 120 rpm. After 24 hours, transfection reactions were prepared by diluting a total of 30 µg of DNA into 3 ml of D-PBS and 120 µl of PEI. The mix was vortexed briefly, incubated at room temperature for 20-50' and then poured into the cells seeded the previous day -that by then would have reached a density of 1 x 10⁶ cells/ml-. The transfected cells were incubated for 48 hours into a humidified shaking incubator at 37°C, 5% CO₂ and 120 rpm before harvesting.

2.3.6 Large-scale transient transfections

24 hours before a transfection experiment cells were seeded to a final count of 0.5 x 10⁶ viable cells/ml into 6 or 8 1 L roller bottles and incubated for 24 hours into a humidified shaking incubator at 37°C, 5% CO₂ and 135 rpm. Transfection reactions were prepared 24 hours later by diluting a total of 300 µg of DNA into 30 ml of D-PBS and 1.2 ml of PEI. The mix was vortexed briefly and incubated at room temperature for 20-50'. The DNA/PEI solution was then poured into the cells seeded the previous day -that by then would have reached a density of 1 x 10⁶ cells/ml- and incubated for 48 hours into a humidified shaking incubator at 37°C, 5% CO₂ and 135 rpm.
2.4 Immuno-precipitations (IPs) and Western blot analysis

2.4.1 HEK 293T monolayer whole cell extracts

Cells were harvested after 48 hours post-transfection (2.2.5). Media was aspirated from each plate and cells were washed once with 1 ml of PBS. This was then aspirated and the cells were scraped off the plate into 1 ml of fresh PBS using a plastic cell scraper. Cells were transferred into an appropriate number of 1.5 ml centrifuge tubes and spun down for 1 minute at 1000 x g. Supernatant was discarded and the cell pellets lysed into 400 µl of IP buffer (see table 2.1 for buffer details) by pipetting up and down 10-15 times. The lysates were then centrifuged at 25,000 x g at 4ºC for 30 minutes and the pellets were discarded.

2.4.2 Preparation of HEK 293F whole cell extracts

Cells were harvested 48 hours post-transfection (2.3.5) by transferring them into an appropriate number of 50 ml centrifuge tubes and centrifuging for 5 minutes at 3000 x g. The supernatant was discarded and the cell pellet(s) were lysed into 1.0 ml of lysis buffer by pipetting up and down 10-15 times. The lysates were sonicated for 3 cycles (15” on, 15” off), centrifuged at 25,000 x g at 4ºC for 30 minutes and the pellets discarded.
2.4.3 Equilibration and immunisation of protein-G sepharose affinity resin

Protein-G sepharose resin (GE Healthcare Life Sciences) was normally used to pull down proteins from HEK 293T whole cell extracts. 40 µl of resin slurry per experiment were washed 3 times with 1.0 ml of IP buffer (see table 2.1 for buffer details) by re-suspending the resin with a pipette and centrifuging for 10" at 400 x g. 1 µg of antibody per IP was diluted into 500 µl of resin/buffer suspension and incubated on a rotating wheel at 4ºC for 20'. The immunised resin was then rinsed 3 times with IP buffer by suspending the resin with a pipette and centrifuging for 10" at 400 x g.

2.4.4 Equilibration of anti-FLAG® affinity resin

Anti-Flag® affinity resin was typically used to pull down proteins from HEK 293F whole cell extracts. 40 µl of resin slurry per experiment were washed 3 times with 1 ml of resin equilibration buffer (see table 2.1 for buffer details) by re-suspending the resin and centrifuging for 5' at 6000 x g.

2.4.5 Pull-downs from HEK 293T whole cell extracts

Whole cell extracts (2.4.1) were mixed with the pre-immunised affinity resin (2.4.3) and incubated for 2 hours or overnight on a rotating wheel at 4ºC. Following the immuno-precipitation, crude lysates were discarded by centrifuging the samples for 10" at 400 x g and by removing the supernatants. The resin with the bound protein was rinsed 3 times with 1.0 ml of IP buffer in order to wash away non-specifically bound proteins and finally diluted with 1 volume of 2 X
protein loading buffer (see table 2.1 for buffer details).

2.4.6 Pull-downs from HEK 293F whole cell extracts

Whole cell extracts (2.4.2) were mixed with the equilibrated anti-flag affinity resin (2.4.4) and incubated for 2 hours on a roller at 4ºC. Following the incubation, crude lysates were discarded by centrifuging the samples for 5’ at 6000 x g and aspirating the supernatant. The resin bound to the protein was rinsed 3 times with 1.0 ml of lysis buffer or 1 ml of each Wash buffer (see table 2.1 for buffer details) to wash away non-specifically bound proteins and was finally diluted with an equal volume of 2 X protein loading buffer lacking the reducing agent (to avoid releasing large amounts of antibody from the resin).

2.4.7 HRP Western blot analysis

Whole cell extracts (2.4.1; 2.4.2) and/or immuno-precipitates (2.4.5; 2.4.6) were run on SDS-PAGE gels for 35’ at 200 V in commercially available MES-SDS running buffer. Protein bands were transferred on a nitrocellulose membrane using a fully wet transfer apparatus and the blots were assembled in pre-chilled (-20ºC) Tobin’s buffer (see Table 2.1 for buffer details) as described in Figure 2.2. The transfer was normally conducted at 100 V for 50’ and, when two blots were run simultaneously, the tank was placed into a tray filled with ice to prevent overheating. Following the transfer, the membrane(s) were blocked overnight with milk blocking buffer (see table 2.1 for buffer details) on a rocker at 4ºC. The membranes were then washed 3 times for 5’ in PBS and incubated for 1 hour at room temperature into heat-sealed envelopes containing the primary antibodies.
diluted in milk blocking buffer. Following the incubation, the antibody solutions were discarded and the membranes washed 4 times for 5’ with PBS. Secondary antibodies were diluted into milk blocking buffer and transferred into clean plastic envelopes with the probed membranes and incubated at room temperature for 30’. The antibody solutions were discarded and the membranes were washed 4 times for 5’ with PBS.

5 ml of Pierce® ECL Western Blotting substrate (GE Healthcare) were prepared by mixing equal amounts of reagents “1” and “2”, according to the manufacturer’s instructions, and poured onto the membranes. Excess developer was drained from the membranes and these were taped into an exposure cassette between 2 layers of claying film. ECL films (GE Healthcare) were exposed to the membranes in a dark room for a time that varied from 1’ to 2 hours depending on the intensity of the signal.

**Figure 2.2: Blotting scheme for protein transfer.**
2.4.8 Fluorescence Western blot analysis
Protein samples were run into SDS-PAGE gels and transferred onto nitrocellulose membranes as described in paragraph 2.4.6. The blot(s) (see paragraph 2.4.6) were blocked overnight with a 0.5 X Odyssey® blocking buffer (LiCOR) on a rocker at 4°C and washed 3 times for 5' in PBS-T. They were then incubated for 1 hour at room temperature into heat-sealed envelopes containing the primary antibodies diluted into 0.5 X Odyssey® blocking buffer (LiCOR). Following the incubation, the antibody solutions were discarded and the membranes washed 4 times for 10' with PBS-T.
LiCOR secondary antibodies were diluted into 0.5 X blocking buffer Odyssey® (LiCOR), transferred into clean plastic envelopes with the probed membranes and incubated in the dark at room temperature for 30'. The antibody solutions were then discarded and the membranes washed in the dark 4 times for 10' with PBS-T. A final 15' wash with plain PBS was conducted to remove the detergent from the blot. Blots were scanned and visualised with a LiCOR scanner.

2.5 Large-scale protein complex purifications

2.5.1 Transfection of HEK 293F suspension cells
Transfections were performed as described in paragraph 2.3.6.

2.5.2 Whole cell extracts from HEK 293F suspension cells
Cells were harvested 48 hours post-transfection (2.3.6) by centrifuging the cultures for 5' at 3000 x g. Cell pellet(s) were lysed into ~40 ml of lysis buffer per litre of culture by using a combination of pipetting up and down and a glass
homogeniser. The crude lysates were then sonicated for 3 cycles (15” on, 15” off) and centrifuged at 108,000 x g at 4°C for 30 minutes and the pellets were discarded.

2.5.3 Equilibration of anti-FLAG® affinity resin
1.25 ml of resin slurry per litre of culture were pipetted into a 15 ml centrifuge tube and washed 3 times with 10 ml of resin equilibration buffer (see table 2.1 for buffer details) by inverting the tube several times and centrifuging for 1’ at 3000 x g.

2.5.4 Pull-down
Approximately 40 ml of whole cell extracts (2.5.2) were mixed with the equilibrated resin (2.5.3) and rotated for approximately 2 hours at 4°C. Sample(s) were then spun down for 1’ at 3000 x g and the supernatant discarded. The resin bound to the protein complex was then transferred into a 15 ml centrifuge tube and washed 3 times with 10 ml of each of three wash-buffers (see table 2.1 for buffer details).

2.5.5 TEV protease elution
Following the washes described in paragraph 2.5.4, the resin was suspended into 8-10 ml of TEV-cleavage buffer (see table 2.1 for buffer details). Approximately 40 µl of 10 mg/ml TEV protease were added and the atmosphere above the samples was replaced with 100% N₂ gas to prevent oxidation of the protein complex. The tube(s) containing the samples were sealed with a strip of
Parafilm®M and gently rotated at 4°C overnight. The sample(s) were then centrifuged for 1’ at 3000 x g and the resin pellet discarded.

2.5.6 **Size exclusion chromatography**

Samples were prepared by transferring the diluted protein complex (2.5.5) into a 15 ml Amicon® ultracentrifugal filter (Millipore) with a 10.0 kDa molecular weight cut-off and centrifuged until the sample was concentrated down to approximately 500 µl. The concentrated protein complex was filtered through a 0.22 µm filter. A size exclusion chromatography column (Superdex®200 or Superose®6) mounted on a ÄKTA purifier was equilibrated 24 hours before the experiment with 50 ml of gel filtration buffer (see table 2.1 for buffer details). The sample was then loaded onto the column and separated by elution at a flow of 500 µl/minute. 500 µl fractions were collected into a fraction collector.

2.5.7 **Protein complex concentration and quantification**

Relevant gel filtration fractions (2.5.6) were merged together and transferred into a 15 ml Amicon® ultracentrifugal filter with a 10 kDa molecular weight cut-off and centrifuged at 3000 x g until the sample was concentrated down to a volume of 500 µl. The concentrated protein was transferred into a 500 µl Amicon® ultracentrifugal filter with a 10 kDa molecular weight cut off and centrifuged at 14,000 x g for a time sufficient to further concentrate the protein down to 20-100 µl. Protein concentration was determined using the BioRad Protein Assay following the manufacturer’s instructions.
2.6 Deacetylase activity assays

2.6.1 Protein complex expression and purification

The protein complex of interest was expressed and purified as described in paragraph 2.5.

2.6.2 Activity assay

10 µl of 100 µM boc-acetyl-lysine substrate (diluted from a 60X stock in HDAC assay buffer) were mixed into a black 96 well plate with 40 µl of purified protein (2.5.7) that had been previously diluted down in HDAC assay buffer (see table 2.1 for buffer details) to a final concentration of 25-100 nM. The plate was incubated in a shaking incubator at 37°C for 30’. 50 µl of developer buffer (see table 2.1 for buffer details) were added into each well and the plate was incubated for 10 minutes at room temperature. The plate(s) were read by a 58 Perkin Elmer Victor X5 plate reader that measured fluorescence levels with an excitation wavelength of 355 nm and an emission wavelength 460 nm. Readings were proportional to deacetylase activity of the purified complex.

2.7 Crystallisation trials

2.7.1 Protein complex expression and purification

The protein complex of interest was expressed and purified as described in paragraph 2.5.
2.7.2 Setting of sitting drop crystallisation plates
Crystallisation trials were conducted into duplicate MRC 96 well sitting drop crystallisation plates that were set up using a Cartesian robot. Each reservoir well of the plate was manually filled with 80 µl of mother liquor from commercially available crystallisation screens (see paragraph 2.1.6 for commercial screens used). The plate was then positioned into the Cartesian robot, which mixed 100 nl of mother liquor with 100 nl of protein sample creating 200 nl drops. The plates were then manually sealed with a transparent sheet that allowed checking for the formation of protein crystals with the aid of an optical microscope.

2.7.3 Crystals cryoprotection, collection and freezing
Freezing a crystal without an appropriate cryoprotectant would cause icing of the water surrounding it, producing a ring-like diffraction pattern that can interfere with the data generated by the protein crystal itself. To prevent this from happening, the 200 nl drops containing the crystal(s) were soaked into 2 µl of cryoprotectant solution made of mother liquor including the minimum required amount of a suitable cryoprotectant (glycerol; MPD; PPG 400). Cryoprotectants allow the water to freeze in the form of a disordered ‘glass’ that does not interfere with the data when the crystals are analysed under an X-ray source. After the crystals had been properly cryoprotected, they were collected with a nylon loop and flash frozen either by immersion into liquid N\textsubscript{2} or with a cryostream of N\textsubscript{2} gas flowing at 100ºK.
2.7.4 Data collection
The crystals were taken to the Diamond Light Source synchrotron (UK) where they were analysed at the microfocal X-ray beamline I-24.

2.8 Binding assays
2.8.1 Protein complex expression and purification
The protein complex of interest was expressed and purified as described in paragraph 2.5.

2.8.2 Fluorescent polarization assay
Protein sample was serial diluted to a final volume of 100 µl across a black 96 well plate and fluorescein-labelled IP$_{1,2,4,5,6}$ (IP$_5$) (kind gift from Professor Barry Potter, University of Bath) was added prior to incubating the plates at room temperature. Binding was measured by a 58 Perkin Elmer Victor X5 plate reader. Measurements were taken with an excitation wavelength of 480 nm and an emission wavelength of 535 nm at different time intervals.

2.8.3 Bio-layer interferometry assay
8 serial dilutions of the protein sample were made across a black 96 well plate. A biotinylated form of IP$_5$ (kind gift of Professor Barry Potter, University of Bath) was immobilised onto 8 biosensors connected to a FortéBio Octet®QK$^e$ system. Binding affinity was determined by following the manufacturer’s instructions.
2.9 Cloning, DNA sequencing and DNA purification methods

2.9.1 PCRs

Primers were designed as described in paragraph 2.1.2. Genes of interest were amplified from plasmid templates previously designed and purified by Dr. Shaun Cowley and Dr. Oliver Dovey. Cloning PCRs were performed using the commercially available KOD Hot Start DNA Polymerase kit (Novagen® Toyobo) following the manufacturer’s instructions:

**Reaction Mix (50 μl reaction)**
- 30 μl MilliQ H₂O
- 5 μl 10X Buffer
- 5 μl dNTPs (final concentration 200 μM)
- 2 μl MgSO₄ (final concentration 1 mM)
- 1 μl template DNA (20-200 ng)
- 3 μl forward primer (final concentration 300 nM)
- 3 μl reverse primer (final concentration 300 nM)
- 1 μl KOD Hot Start DNA Polymerase (1 unit)

**Thermal cycler settings**
- Enzyme pre-activation: 2’ at 94°C (one time only)
- Denaturation: 15” at 94°C
- Annealing: 30” at 60°C
- Extension: 20”/kbp at 72°C
  (35-40 cycles)
- Final extension: 5’ at 72°C (one time only)

If the thermal cycler was left to operate overnight, it was set to keep the samples at a temperature of 10°C at the end of the cycle.

2.9.2 Agarose gel electrophoresis and PCR product extraction

PCR products were purified from PCR ‘contaminants’ by agarose gel electrophoresis. 1% agarose gels were made in TAE buffer (see table 2.1 for buffer details) and pre-stained with an appropriate volume of 100,000 X EtBr (to a final concentration of 1 X). Electrophoresis was carried out in an electrophoresis apparatus containing 1 X TAE buffer for 30’ at 120 V. DNA bands of PCR
products were excised with the blade of a sterile scalpel under a UV lamp and placed into a clean 1.5 ml centrifuge tube. The fragment was then extracted from the gel using the commercially available QIAGEN gel extraction kit following the manufacturer’s instructions.

2.9.3 Cloning

The purified PCR products (2.9.2) were cloned into the required mammalian expression vectors by Dr. Xiaowen Yang at the PROTEX cloning facility of the University of Leicester. The expression vectors were restriction digested with different restriction enzymes to either include or exclude the affinity tag present on the vector (Figure 2.1). The insert was cloned into the plasmid using an In-Fusion cloning kit according to the manufacturer’s instructions and colonies were selected in LB agarose plates containing an appropriate selection antibody and 5% sucrose. Positive colonies were verified by colony PCR.

2.9.4 Transformation of competent cells

50 µl of competent DH5α E. coli cells stored at -80°C were defrosted on ice for 20’ and subsequently incubated on ice for 30’ with 1 µg of plasmid DNA. The cells were then heat-shocked at 42°C for 120” and rested for 5 minutes on ice. The transformed DH5α were then rescued in 1 ml of plain 1 X LB media (see table 2.1 for media details) for 1 hour at 37°C. The cells were then centrifuged at 6000 x g for 5 minutes and the supernatant was removed leaving only 30-50 µl of media in which the cells were re-suspended again and plated onto an LB-agarose plate made from a commercially available pre-mixed powder (Invivogen)
containing an appropriate selection antibiotic. The plates were then incubated at 37°C overnight. Resulting colonies were picked and expanded in larger liquid cultures containing the appropriate selection antibiotic.

Alpha-Select Bronze Efficiency competent \textit{E. coli} cells from Bioline were transformed according to the manufacturer’s instructions.

### 2.9.5 Small-scale plasmid DNA purifications

Small amounts of plasmid DNA were purified using commercially available Plasmid DNA purification kits based on the alkaline lysis method followed by an ion-exchange purification step.

5-8 ml cultures of transformed \textit{E. coli} cells were grown as specified by the Qiagen Miniprep protocol and the DNA was purified using a Qiagen Miniprep kit following the manufacturer’s instructions. DNA from larger cultures (up to 100 ml) was purified either with Qiagen Midiprep or with Nucleobond EF plasmid purification kits as per the manufacturer’s instructions.

### 2.9.6 Large-scale plasmid DNA purifications

Large-scale purifications were carried out with a LiCl and PEG based method optimised by Dr. Louise Fairall (see table 2.1 for buffer and media used in this paragraph).

750 ml of 2yt media containing 200 \(\mu g/ml\) of ampicillin were inoculated with a single colony of transformed \textit{E. coli} cells (2.9.4) containing the plasmid of interest and incubated for 20 hours in an orbital shaker incubator at 37°C and 160 rpm. Cells were harvested by centrifugation at 3000 \(x\) \(g\) for 10 minutes, re-suspended
in 15 ml of buffer res1 and lysed by mixing 5 ml lysozyme solution and 15 ml of alkaline lysis buffer. Genomic DNA and SDS were precipitated by adding 22.5 ml of neutralisation buffer and centrifuging at 3000 x g for 10' at 4°C. The supernatant was filtered through miracloth. Nucleic acids were precipitated out of solution by adding 0.6 volumes of isopropanol, then pelleted by centrifuging at 3000 x g for 10' at 4°C and re-dissolved into 15 ml of buffer res2. RNA and protein impurities were precipitated out of solution by adding 1 volume of 5 M LiCl kept at -20°C and discarded by centrifuging at 3000 x g for 10' at 4°C. Nucleic acids dissolved in the supernatant were pulled out of the LiCl solution by adding 0.6 volumes of isopropanol and centrifuging at 3000 x g for 10' at 4°C. The pellet was re-dissolved in 10 ml of buffer res3. Remaining RNA contaminants were digested by adding 50 µl of 10 mg/ml heat treated RNase A to the sample tube followed by an incubation of 15' at room temperature. Small RNA fragments resulting from the RNase A digestion were discarded by the addition of 0.25 volumes of PEG solution followed by a 30 minutes incubation on ice. PEG only precipitates large nucleic acid molecules leaving the small RNA contaminants in solution. The precipitated plasmid was then collected at the bottom of the centrifuge tube by centrifuging at 3000 x g for 10’ at 4°C and the pellet was re-dissolved into 10 ml of buffer res3. Remaining PEG impurities were extracted by adding 2 ml of chloroform to the sample, vortexing briefly and centrifuging for 1’ at 4°C. The aqueous layer was retained and the purified plasmid was precipitated out of solution by adding 0.1 volumes of 5 M NaCl and 3 volumes of absolute ethanol, followed by a centrifugation step at 3000 x g for 10’ at 4°C. The pellet was washed in 70% ethanol and let to dry upside down overnight. The plasmid
was finally dissolved into an appropriate volume of filter sterilised MilliQ H₂O and stored at -20°C.

2.9.7 DNA quantification

DNA concentration was calculated by an Implen NanoPhotometer® that was operated according to the manufacturer’s instructions.

2.9.8 Plasmid DNA Sequencing

PCR and sequencing reactions were executed by the University of Leicester PNACL and the results were analysed by 4Peaks, EnzymeX and Standard Nucleotide BLAST®.

2.10 Mass spectrometry analysis

Protein bands from Coomassie stained SDS PAGE gels were excised and submitted to the Protein and Nucleic Acid Chemistry Laboratory of the University of Leicester (PNACL) where they were analysed either by MALDI-ToF or LC-MS/MS.
CHAPTER 3: Characterization of the Interactions among HDAC1, SDS3 and the HID domain of Sin3a

3.1 Chapter aims

The first goal of the project was to gain a deeper understanding of the interaction between HDAC1 and the HID domain of Sin3a. Laherty et al. (1997) identified the HID (amino acids 524-899) as the region of Sin3a responsible for binding HDAC1. SDS3, which also interacts with the Sin3a complex through the HID, is indispensable for the deacetylase activity of the complex, ensuring the correct function of a large number of biological processes (Lechner et al. 2000; Alland et al. 2002; Fleischer et al. 2003). However, the mechanism by which SDS3 exerts its role remains largely unexplored. Having generated tagged constructs for mammalian expression (described in chapter 2), I attempted to find a minimal HDAC1 interacting region within the HID that would enable me to use a smaller protein region for the crystallization trials as well as getting an insight into the role of SDS3 within the HID.

3.2 HDAC1 interacts with the HID through multiple surfaces

3.2.1 Expression of affinity tagged HDAC1

In order to immuno-precipitate and visualize the expressed proteins by western blot, affinity tags were added to the proteins of interest as described in chapter 2. Early efforts to express and pull down N-terminally tagged HDAC1 in mammalian cells were unsuccessful due to the interference of the tag with the N-terminal active site of the enzyme. Since the C-terminal region of HDAC1 consists of an
unstructured tail, a C-terminal flag was added to HDAC1 (pcDNA 3.1-based vector designed and kindly provided by Dr. Oliver Dovey) which was then successfully expressed and immuno-precipitated as shown in Figure 3.1.

![Figure 3.1: Expression of C-terminal flag-tagged HDAC1](image)

**Figure 3.1**: Expression of C-terminal flag-tagged HDAC1. The figure shows an anti-flag western blot representing the expression and immuno-precipitation of a C-terminal tagged versus an N-terminal tagged HDAC1. The expression levels of the C-flag construct were significantly higher than the N-terminally tagged HDAC1 and therefore it was selected to be used for the rest of the mapping experiments.

### 3.2.2 Bioinformatics: designing constructs of the HID

The HID as characterized by Laherty et al. (1997) is a 375 amino acid long region of Sin3a that might be too large for crystallization trials. In order to design new and shorter constructs for the crystallization experiments, protein sequence conservation, predicted secondary structure and disorder prediction parameters were taken in consideration as follows.

Alignments of Sin3a from different species show that the HID domain is a highly conserved region of the protein, especially at the N-terminal and mid-section of the domain as shown in Figure 3.2.
Figure 3.2 Multiple sequence alignment of the full-length Sin3a from different species (MultAlin: Corpet, 1988). Amino acids from 1000 to 1360 have been highlighted in yellow and represent the HID domain.

Jpred (secondary structure prediction tool, Cole et al. 2008) indicates that the HID domain should be predominantly alpha-helical with an approximately 80
amino acid long N-terminal region that seems to be unstructured despite the high level of sequence conservation (Figures 3.2 and 3.3). The central region of the domain consists of 6 predicted alpha helices, four of which seem to be arranged in a way that might potentially resemble a SANT domain (helices 3-6, Figure 3.3), a protein fold known to form interactions with HDAC1 and HDAC3 in different transcriptional co-repressor complexes. The C-terminal region is predicted to have 3 alpha helices and it terminates with a poorly conserved and seemingly unstructured sequence (Figure 3.3). RONN (disorder prediction tool, Zheng et al, 2005) shows that the Jpred-predicted secondary structures coincide with regions of the domain with low entropy (Figure 3.3), adding reliability to the Jpred results.

As mentioned above, the N-terminal region of the HID (~531-600) seems to be mostly unstructured. This is partially confirmed by RONN, which assigns to this segment an average 50% probability of disorder. However, it remains possible that secondary structures may form upon binding to other proteins, potentially SDS3. The HID constructs used in this chapter and in chapter 4 were designed by paying particular attention to the predicted secondary structures and, when possible, boundaries were designed so as not to truncate the protein at low-entropy regions or regions of high conservation. Full-length HDAC1 and SDS3 were used for the experiments reported in this chapter.
Since it binds HDAC1/2 tightly, it is remarkable that the HID does not show any homology with the SANT domains of SMRT, NuRD, MiDAC and CoREST, implying that a crystal structure of the Sin3a/HDAC1 complex might reveal a novel fold able to bind HDACs. However, since the exact boundaries of the HID domain were unknown, we performed mapping experiments in order to find the region of the HID domain with the highest yield when expressed in mammalian cells and the most stable interaction with HDAC1.

**Figure 3.3:** Disorder and secondary structure predictions of the full length HID (524-855) (RONN and Jpred). The blue line represents the predicted entropy; values above 0.5 are considered to be disordered (RONN). Each red box represents a predicted alpha helix (Jpred) and was numbered for convenience. The level of confidence for each region of predicted secondary structure is indicated by the black and green bar, where confidence levels equal or above to 7 out of 9 are represented in green.
3.2.3 Mapping the region of the HID responsible for interacting with HDAC1

Having considered the results of the bioinformatics analysis (paragraph 3.2.2), the first step towards mapping the interaction between HDAC1 and Sin3a was to make a series of C-terminal truncations of the HID. This was aimed at determining the role of the predicted helices (Figure 3.3) in binding HDAC1. In these experiments, C-terminal flag-tagged HDAC1 was co-transfected with N-terminal myc-tagged constructs of the HID and co-immuno-precipitated with protein G sepharose beads pre-immunized with anti flag antibody. Anti-myc western blots followed to detect pulled-down HID fragments. Transient transfections in HEK 293T cells, were carried out to co-express Myc-tagged Sin3a constructs with Flag-tagged HDAC1. Anti-Flag co-IPs were then followed by anti-Myc western blots to detect pulled-down HID constructs.

Consistent with previous reports, the full length HID 531-855 showed a strong interaction with HDAC1 (Figure 3.4). Deletion of helices 8 and 9 (HID 531-772) does not alter the interaction with the enzyme (Figure 3.4). However, the stability of this interaction is lost when helices 3-7 are also truncated (HID 531-654). These results pointed to the necessity of further investigating the contribution of helices 3-7, which seem to be important although not essential for the interaction.
Our initial mapping data indicated that the interaction of HDAC1 with Sin3a occurs in the central region of the HID, including helices 3-7 (Figure 3.3), rather than in a more N-terminal or C-terminal portion of the protein. In order to test this, both an “N-terminal HID” (HID 524-680) and a “C-terminal HID” (HID 681-855) constructs were co-expressed with HDAC1 (pCS2+myc based constructs were kindly provided by Dr. Shaun Cowley). These constructs were aimed at understanding how the removal of a large portion of the N-terminus of the HID and the disruption of a sequence of alpha helices -that may potentially form a helical bundle- would affect the interaction of HDAC1. HID 524-680 was unable to interact with HDAC1 (Figure 3.5), possibly due to a folding issue that may occur by truncating the protein in the vicinity of helix 4 (Figure 3.3). In contrast, HDAC1 does interact with the C-terminal HID (HID 681-855) construct, although more weakly than with the full-length HID and HID 531-772.

These results strengthened the hypothesis that the interaction of HDAC1 with Sin3a occurs through the central region of the HID, in a region which is predicted...
to contain a 4 or 5 helical bundle (helices 3-7). Indeed, although the HID does not show any sequence homology with the SANT domains of other transcriptional co-repressors the disposition of helices 3-7 (Figure 3.3) suggests that a similar structure might form in this region of the protein, especially since it seems to be important for the interaction.

![Image](Figure 3.5: Interaction of HDAC1 with C-terminal and N-terminal truncated HID constructs. The figure represents an anti-flag-HDAC1 Co-IP followed by an anti-myc western blot to detect pulled-down myc-HID fragments.)

To test whether the “helical bundle” was sufficient to bind HDAC1, a HID 651-772 construct, including helices 3-7, was designed. This construct was shown to interact with the HDAC1 (Figure 3.6); however, the intensity of the IP lane compared to the input suggested that it does not form a stable interaction. Additional constructs, including HID 651-741 (helices 3-6), did not have a good yield but showed a stable interaction (Figure 3.6), possibly implying that this four-helical bundle is the main interacting surface of Sin3a and that the presence of a possibly unfolded helix 7 might perturb the binding of HDAC1. The amount of soluble Sin3a HID and Sin3a HID – representing helices 5-6 and 5-7 respectively – was so low it was not possible to determine whether they could interact with
HDAC1 at all (Figure 3.6). However, it was possible to conclude that the deletion of helices 3 and 4 causes both a dramatic decrease in yield and interaction stability with HDAC1.

**Figure 3.6 Interaction of HDAC1 with alpha helices 4 to 7 of the HID domain of Sin3a.** The figure represents an anti-flag-HDAC1 co-IP followed by an anti-myc western blot to detect pulled down myc-HID fragments. HDAC1 shows a strong interaction with HID 651-741, although expression of this construct is low.

To exclude the possibility that helices 8 and 9 (Figure 3.3) were also involved in binding HDAC1, two constructs (HID 741-855 and HID 773-855) expressing only this region of Sin3a were designed. However, the expression of these two fragments could not be detected (Figure 3.7) and determining the role of helices 8 and 9 was not possible.

In summary (Figure 3.7), these results suggested that HDAC1 interacts with Sin3a through multiple surfaces of the HID rather than solely with a single alpha helix or small region of the domain. Therefore, selecting a very short Sin3a construct for the crystallization trials was not feasible due to either loss of expression or interaction with the enzyme. Given that the shortest HID fragment
forming the most stable interaction was HID 531-772 (Figures 3.4 and 3.7), it was decided to use this construct for the crystallization trials in chapter 4.

**Figure 3.7:** *Summary of the HID constructs used for the mapping experiments.* Each red box represents an alpha helix and the strength of interaction with HDAC1 is reported on the right of the figure: ++++ (very strong); +++ (strong); ++ (medium); + (weak); n.d. (not detectable); p.d. (poor detection). The shortest construct with the strongest interaction is HID 531-772, which was used for the crystallization trials.

### 3.3 SDS3 stabilizes HDAC1 interaction with Sin3a

As reported in the introduction, SDS3 is a core component of the Sin3a complex as well as being an essential protein that can interact with a number of transcription factors, such as Foxk1, and regulates chromosomal segregation and heterochromatin formation (David et al. 2003; Shi et al. 2012). Lechner et al. (2000) have reported that SDS3 is essential for the deacetylase activity of the Sin3a complex, possibly by stabilizing the Sin3a complex. There are currently no
solved structures of SDS3, but the Sin3a Interaction Domain (SID) was mapped to amino acids 188-229 (Alland et al. 2002).

The following experiments were designed into get an insight to the role of SDS3 for the stability of the HID. Flag-IPs of Myc-HID 531-855 and Flag-HDAC1 overexpressed in HEK-293T cells formed a complex that could pull down endogenous SDS3 as shown in Figure 3.8. This interaction is maintained also if untagged SDS3 is co-expressed with the full length HID in the absence of HDAC1 (Figure 3.8), indicating that HDAC1 is dispensable for the interaction of SDS3 with the co-repressor.

![Figure 3.8: SDS3 interacts with the full length HID domain of Sin3a. The proteins (Myc-HID, Flag-HDAC1 and untagged SDS3) were co-expressed in HEK-293T cells, co-immuno precipitated with an appropriate antibody and detected by western blotting as shown.](image)

Two-hybrid screenings have shown that SDS3 is not able to repress genes if the 693-855 region of Sin3a is deleted (Alland et al. 2002). Protein constructs were therefore designed to test which region of Sin3a is essential for the interaction
with SDS3. As shown in Figure 3.9, SDS3 is not able to bind Sin3a 531-654 in the absence of HDAC1, which in fact “rescues” the interaction when it is co-transfected along with Sin3a. Indeed, HDAC1 pulled down both endogenous and exogenous untagged SDS3 (Figure 3.9). This was the first evidence that even though SDS3 could interact with the full-length HID independently of HDAC1, it may also contact the enzyme therefore making a “three-way” interaction that stabilizes the complex.

Figure 3.9: SDS3 cannot interact with the N-terminal portion of the HID unless HDAC1 is also overexpressed. The proteins (Myc-HID, Flag-HDAC1 and untagged SDS3) were co-expressed in HEK-293T cells, co-immuno precipitated with an appropriate antibody and detected by western blotting as shown.

Analogous results could be observed when the N-terminal region of the construct was deleted. Indeed, as shown in Figure 3.10 (and summarized in Figure 3.11), SDS3 is not able to form an interaction with HID 681-855 unless in the presence of overexpressed HDAC1. This was further evidence that SDS3 is indeed involved in the interaction between Sin3a and HDAC1, therefore the starting
hypothesis could not be confirmed.

From Figures 3.8, 3.9 and 3.10, it is also noticeable how the interaction of SDS3 with the HID seems not to be stable. In a first instance, this led me to believe that SDS3 was not a stoichiometric component of the complex. However, this was later contradicted by complex purification experiments reported in chapter 4.

**Figure 3.10:** SDS3 cannot interact with the C-terminal portion of the HID unless HDAC1 is also overexpressed. The proteins (Myc-HID, Flag-HDAC1 and untagged SDS3) were co-expressed in HEK-293T cells, co-immuno precipitated with an appropriate antibody and detected by western blotting as shown.
Figure 3.11: Summary of the interaction of SDS3 with the HID. When the central region of the HID is truncated, SDS3 loses the ability to interact with the HID domain of Sin3a. However, overexpressed HDAC1 can rescue the interaction.

3.4 Results summary and conclusions

Mapping experiments suggested that the interaction of HDAC1 with Sin3a might be taking place through multiple surfaces of the HID, rather than at a single location. Moreover, co-IP experiments suggest that the interaction of HDAC1 with Sin3a may be coordinated by the presence of SDS3, as a ternary complex. These results posed a challenging picture in regards to selecting a short HID construct, since the minimal HDAC1 interacting region was either expressed at low levels or did not interact with HDAC1. The shortest construct with the most stable binding of HDAC1 was HID 531-772, and this was therefore selected for the complex purification and crystallization experiments. This construct has 7
predicted alpha helices, including the region (helices 3-7; Figure 3.3) that seems to be particularly important for the interaction of the enzyme.
CHAPTER 4: Complex purifications and crystallization trials

4.1 Chapter aims

Recent publications have provided a major advance in understanding the mechanisms regulating the SMRT (Watson et al. 2012b) and NuRD (Millard et al. 2013) co-repressor complexes. In particular, it was discovered that HDAC1 and HDAC3 interact with the SANT domains of MTA1 and SMRT respectively through an IP_4 molecule, which acts as a connecting bridge between the enzymes and the co-repressors, triggering the catalytic activity of the complexes. As mentioned in the introduction and chapter 3, the HID domain of Sin3a does not share any sequence homology with these SANT domains and predicting whether it can also bind IP_4 is a difficult challenge. Obtaining the crystal structure of the Sin3a/HDAC1 complex would elucidate both the working mechanism of the Sin3 complexes and contribute to the general understanding of the HDAC-mediated regulation of transcription.

This chapter describes the purification and crystallization trials of the Sin3a binary complex consisting of HDAC1 and Sin3a-HID, as well as the ternary complex formed by HDAC1, Sin3a-HID and SDS3. This structural approach was aimed at understanding the molecular details of the Sin3a interaction with its catalytic subunit.
4.2 Binary complex (HDAC1/Sin3a)

4.2.1 Small-scale purifications

Following the mapping experiments reported in chapter 3, the next logical step to follow was to assess whether HDAC1, Sin3a and SDS3 could interact with each other in stoichiometric proportions, which would be indicative of their ability to form a stable complex suitable to be purified in large-scale for crystallisation trials. This was verified by small-scale purifications of the binary (HDAC1/Sin3a) and ternary (HDAC1/Sin3a/SDS3) complexes followed by Coomassie-staining analysis (chapter 2, paragraph 2.4.6: only IP buffer was utilised for all the steps of this experiment). The gel in Figure 4.1 contains samples from a small-scale purification from HEK-293F cells of the binary and ternary complexes purified using an anti-Flag resin. In this small-scale experiment, the HDAC1 and HID bands showed approximately equal intensities, indicating that each enzyme molecule bound to one molecule of co-repressor. Similarly, SDS3 seemed to be in a 1:1:1 stoichiometric ratio with the two other components of the complex (Figure 4.1). This was in contrast with the preliminary findings of the mapping experiments reported in paragraph 3.3 of chapter 3, which, on the contrary, suggested that SDS3 would not be part of a stoichiometric complex. Also visible in Figure 4.1, with the ternary complex an additional protein (similarly present at a 1:1:1:1 proportion with the other complex subunits) was pulled down. However, mass spectrometry analysis (MALDI-ToF) identified the protein to be cytoplasmic actin, which was later found to be an impurity that could be easily washed away with the use of different buffer conditions. These results encouraged me to attempt large-scale purifications of the ternary complex - in addition to the binary
complex- in order to maximise the possibilities of obtaining protein crystals suitable for X-ray analysis.

**Figure 4.1:** Small-scale purification of binary and ternary complexes. Flag-tagged HID 531-772 was co-transfected in 30 ml cultures of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1) and SDS3 (U-SDS3). The co-transfections were followed by anti-Flag affinity purifications using IP Buffer (table 2.1 of chapter 2).

The results of the binary complex shown in Figure 4.1 led me to attempt a large-scale purification adopting the methodology and buffer conditions described in Watson et al. (2012b) for the HDAC3/DAD complex. This was done before an appropriate screening of buffers and purification conditions was made (Figure 4.2). Unfortunately, this effort produced a total yield of approximately 135 µg of pure binary complex from 2 L of culture, barely sufficient for one crystallization
trial at a low protein concentration. Indeed, the faint protein bands visible in Figure 4.2 indicate the low level of expressed and purified protein. The complex elution began approximately 12.1 ml after the injection point and reached a maximum absorbance peak of 5.3 at 280 nm.
Figure 4.2: Large-scale binary complex purification (1). Proteins (Flag-HID 531-772 and U-HDAC1) were co-expressed in 2L of HEK-293F cultures and affinity purified with an anti-Flag affinity resin as described in paragraph 2.5 of chapter 2. The complex was passed through a Superdex™200 (S200) size exclusion chromatography column in gel filtration buffer with 50 mM K-acetate. A: First purification step. The gel shows the protein bound to the resin, the resin after the TEV elution and the TEV eluate. B: Gel filtration fractions: the protein bands appear faint and highly proteolyzed. C: S200 gel filtration chromatogram.

Following these results it was decided that prior to any other large-scale purification, further small-scale experiments should be carried out in order to define the best purification and buffer conditions for the Sin3a complex. A small-scale experiment was therefore designed to test the effect of salt concentration...
throughout the purification steps and the efficiency of a commercially available transfecting agent (293fectin™) was compared to the efficiency of PEI. Indeed, polyethylenimine (used as described in paragraph 2.3.5 of chapter 2) appeared to be more efficient than 293fectin™ as visible from the first two lanes of the gel shown in Figure 4.3. On the other hand, salt concentration (specified in the Figure legend) did not affect the complex in any visible way. The affinity purified protein was also cleaved with TEV and eluted as described in paragraph 2.5.5 of chapter 2 with protease and buffer volumes down-scaled directly for a small-scale purification. Remarkably, despite cleaving off the tag, most of the HID complex was insoluble (Figure 4.3, compare lanes “Post-TEV anti flag resin” and “Post-TEV eluate”).

Figure 4.3: Optimization of small-scale binary complex purification (1). Flag-tagged HID 531-772 was co-transfected in 30 ml cultures of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1). The complex was purified with an anti-Flag affinity resin as described in paragraph 2.4 of chapter 2. Lanes labelled as “normal salt” represent...
the complex purified with buffer conditions described in paragraph 2.4.6 of chapter 2 (Watson et al. 2012). The buffers used in the “low salt” lanes contained 50% less salt than the “normal salt” conditions, which had the following salt concentrations: (Lysis and Wash buffer 1: 100 mM K-acetate; Wash buffer 2: 300 mM K-acetate; Wash buffer 3: 50 mM K-acetate).

Watson et al. (2012b) had observed that salt concentrations greater than 100 mM K-acetate would disassemble the HDAC3/SMRT complex, which was the rationale for testing “low salt” purification conditions. However, these data (Figure 4.3) suggested that salt concentration might not affect the binary complex. Therefore, a second experiment was designed to confirm this phenomenon as well as to test the effect of a range of buffer conditions. In particular, in addition to the “normal salt” and “low salt” experiments already described in Figure 4.3, a sample was purified in “high glycerol” conditions, where all the buffers containing glycerol were made with 10% glycerol instead of the original 5% concentration. A “0.5% TX-100” experiment was also designed so that the lysis buffer and wash buffer 1 had a 0.2% increase in detergent concentration compared to the original 0.3%, while in the “0.5% Igepal” experiment the lysis buffer and wash buffer 1 contained Igepal instead of TX-100. Moreover, the volume of each wash buffer was 10 times higher than the standard protocol (that is 3 washes of 1 ml with each Wash buffer) in order to reduce the level of impurities non-specifically interacting with the anti-flag affinity resin. This experiment confirmed that salt concentration (specified in the figure legend) does not have any effect on the complex at least until the TEV elution step (Figure 4.4). In addition, glycerol and detergent concentration, and the detergent type, did not have any significant impact on the yield or purity of the complex. However, the general levels of purified protein was
greatly decreased compared to those visible from Figure 4.3, presumably because of the increased resin-wash volumes, which on the other hand did not improve the purity of the samples. Even in this case, the protein remained largely insoluble as visible from the “TEV eluate” lanes. This was likely to be caused by protein oxidation during the overnight TEV elution step, which was carried out at 4°C in PCR tubes sealed with atmospheric gas.

**Figure 4.4:** Optimization of small-scale binary complex purification (2). Flag-tagged HID 531-772 was co-transfected in 30 ml cultures of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1). The complex was purified with an anti-Flag affinity resin as described in paragraph 2.4 of chapter 2. “Normal salt”: Lysis and Wash buffer 1: 100 mM K-acetate; Wash buffer 2: 300 mM K-acetate; Wash buffer 3: 50 mM K-acetate, which were adapted from Watson et al., 2012a. “Low salt”: Lysis and Wash buffer 1: 50 mM K-acetate; Wash buffer 2: 150 mM K-acetate; Wash buffer 3: 25 mM K-acetate. “High glycerol”: all buffers containing glycerol were prepared with 10% glycerol instead of the original 5%. “0.5% TX-100”: lysis buffer and wash buffer 1 had a 0.5% detergent instead of the original 0.3%. “0.5% Igepal”: lysis buffer and wash buffer 1 contained Igepal instead of TX-100.

Despite the above experiment confirming that salt, glycerol and detergent may not affect the integrity of the complex, these results had to be verified in a
situation of optimal protein yield, which would allow a more accurate assessment of the effect of these variables. In order to improve the levels of purified protein complex, I decided to adopt the same methodology used in the previous experiment, with the difference that the wash buffer volumes from 10 ml were reduced back to 1 ml and the cells were lysed more thoroughly by pipetting them up and down in lysis buffer approximately 30 times. Moreover, to prevent protein oxidation occurring during the TEV elution step, the samples were sealed in PCR tubes with a 100% nitrogen atmosphere.

As in the two previous experiments, salt glycerol and detergent did not affect the complex in any visible way (Figure 4.5). However, the steps implemented to improve protein yield and solubility were successful. Indeed, as visible from Figure 4.5 (all lanes) the amount of yielded protein was significantly higher than in previous experiments (Figures 4.2, 4.3 and 4.4). Furthermore, the complex was finally detectable in the TEV eluate lanes, indicating that incubating the sample in 100% N₂ gas successfully prevented the oxidation and consequent precipitation of the protein complex.
Figure 4.5: Optimization of small-scale binary complex purification: (3). Flag-tagged HID 531-772 was co-transfected in 30 ml cultures of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1). The complex was purified with an anti-Flag affinity resin as described in paragraph 2.4 of chapter 2. “Normal salt”: Lysis and Wash buffer 1: 100 mM K-acetate; Wash buffer 2: 300 mM K-acetate; Wash buffer 3: 50 mM K-acetate, which were adapted from Watson et al., 2012a. “Low salt”: Lysis and Wash buffer 1: 50 mM K-acetate; Wash buffer 2: 150 mM K-acetate; Wash buffer 3: 25 mM K-acetate. “High glycerol”: all buffers containing glycerol were prepared with 10% glycerol instead of the original 5%. “0.5% TX-100”: lysis buffer and wash buffer 1 had a 0.5% detergent instead of the original 0.3%. “0.5% Igepal”: lysis buffer and wash buffer 1 contained Igepal instead of TX-100.

Since salt, glycerol and detergent did not have an effect on the complex but yield and solubility of the purified protein rather depended on harsher lysis and prevention of protein oxidation, it was decided that for the large-scale purifications the buffer conditions adapted from (Watson et al. 2012b) would be utilised as they also proved to be successful for the purification and crystallisation of the HDAC1/MTA1 complex (Millard et al. 2013).
4.2.2 Large-scale purifications and crystallization trials

Having successfully optimised the small-scale expression and purification of the binary complex, large-scale experiments were done in order to set up crystallisation trials. These experiments were carried out according to the protocols reported in paragraph 2.5 of chapter 2 with the variations at the lysis and TEV elution steps applied for the final round of optimisation of the binary complex (Figure 4.5).

Despite the progress made to improve the protein yield in small-scale preps, the highest amount of binary complex produced in a large-scale purification was \(~120 \mu g\) of pure protein per litre of cell culture. An example of these purifications is shown in Figure 4.6. The complex elution began approximately 12.1 ml post injection and reached a maximum absorbance peak of 11.4 at 280 nm. The double peak of the gel filtration chromatogram is due to the proteolysis affecting the complex visible in Figure 4.6 B. Even though this was not close to an optimal situation, it was just sufficient for a small number of crystallisation trials carried out at relatively low protein concentrations. In particular, 4 of the commercially available screens JCSG-plus, ProPlex™, PACT premiere and Stura & MacroSol were tested.
Figure 4.6: Large-scale binary complex purification (2). Flag-tagged HID 531-772 was co-transfected in a 2.4 L culture of HEK 293F suspension cells with untagged HDAC1 (U-HDAC1). The complex was affinity purified with an anti-Flag affinity resin as described in paragraph 2.5 of chapter 2 and passed through a S200 size exclusion chromatography column in gel filtration buffer with 50 mM salt. (See table 2.1 for buffer details). **A**: First purification step. The gel shows the protein bound to the resin, the resin after the TEV elution and the TEV eluate. **B**: Gel filtration fractions. **C**: S200 gel filtration chromatogram.

An average of 60% of the crystallization conditions did not produce any type of precipitation and only in less than 5% of conditions the precipitation was heavy. The remaining 35% of conditions generated light and amorphous precipitation. The only exception was represented by condition B-10 of the Stura and MacroSol screen, which, with the protein at 4.2 mg/ml, gave rise to a light crystalline
precipitate (Figure 4.7A) that was optimised around precipitant, pH, protein concentration and temperature as summarised in Figure 4.7B. The same type of precipitation was observed in the optimised screen (Figure 4.7 B) set up at 4°C with the protein complex at a 2.8 mg/ml concentration. The crystalline precipitates from the original and optimised screens were collected with nylon loops as described in paragraph 2.7.3 of chapter 2 and sent to the Diamond light source synchrotron in Oxford to be analysed in the I-24 beamline. Unfortunately no crystals could be detected in the loops and no X-ray diffraction was produced.

Figure 4.7: Optimization of Stura & MacroSol B-10 condition. A: Stura & MacroSol B-10, original condition. The 200 nl drop is observed at a 50 X magnification under a light polarizer. The largest piece of crystal-like material was of approximately 5 x 5 µm in size. B: The condition in A was optimised around PEG and pH and the protein concentration was increased in an attempt to produce larger crystals. Another optimised screen was set up with a lower protein concentration at a temperature of 4°C, which produced the same type of precipitation shown in A.
Overall, these results suggested that further optimisation of the large-scale purification method would be required to obtain a suitable protein sample for the crystallisation trials. However, since this would be a time consuming and expensive approach, it was decided that, instead, the effect of adding SDS3 to the binary complex should be explored with higher priority.

4.3 Ternary complex (HDAC1/Sin3a/SDS3)

As discussed in 4.2.2, protein yields obtained with the large-scale purifications of the binary complex were low and only allowed for a very limited number of crystallisation trials at relatively low protein concentrations. The next logical step was to test the effect of adding SDS3 to the HDAC1/Sin3a complex. Indeed, as already mentioned in the introduction and in chapter 3, this protein is crucial for the deacetylase activity of Sin3a as well as for other of biological functions (Lechner et al. 2000; Fleischer et al. 2003; Alland et al. 2002; David et al. 2003; Shi et al. 2012) and may therefore stabilise the HID allowing for higher purification yields and a more stable protein complex.

However, since the binary complex showed a remarkable difference in the amount of produced protein between small-scale and large-scale experiments, as a first step I carried out a medium scale purification (250 ml of cells, paragraph 2.5, chapter 2) side-by-side with a small-scale experiment (30 ml of cells, paragraphs 2.4.6, chapter 2) to look for any loss of yield. The results shown in Figure 4.8 clearly indicated that there was no difference in the expression or purification efficiency between medium and small-scale experiments and that the 3 co-expressed proteins appeared to be soluble and in a 1:1:1 ratio as previously
observed in Figure 4.1. This allowed me to up-scale the experiment to larger cell cultures aimed at producing enough protein for high throughput crystallisation trials.

Figure 4.8: Ternary complex: small-scale versus medium-scale purifications. Flag-tagged HID 531-772 was co-transfected in 30 and 250 ml cultures of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1) and SDS3 (U-SDS3). The complex was affinity purified with an anti-Flag affinity resin as described in paragraphs 2.4 and 2.5 of chapter 2 and passed through a S200 size exclusion chromatography column in gel filtration buffer with 50 mM salt. The gel shows the protein bound to the resin, the resin after the TEV elution and the TEV eluate coming from a 30 ml and a 250 ml cell culture protein preparations.

The complex formed by the full length HDAC1, the full length SDS3 and HID 531-772 will be from now on referred to as full ternary complex for convenience.
4.3.1 Large-scale purifications and crystallisation trials of the full ternary complex.

After the encouraging result shown in Figure 4.8, the purification was scaled-up to cultures ranging from 1.8 to 2.25 litres of HEK 293F suspension cell cultures as described in paragraph 2.5 of chapter 2.

Large-scale purifications of the ternary complex (an example is shown in Figure 4.9) produced an average yield of ~650 µg of purified protein complex per litre of cell culture, which is approximately 6 times higher than the yield obtained with the binary complex. The stoichiometry of the three complex components was 1:1:1, as already observed in previous purifications. However, as it is visible from Figure 4.9C, at the gel filtration stage, the complex eluted close to the void volume of the S200 size exclusion chromatography column, approximately 8.5 ml after the injection point and reached a maximum peak of 979 at 280 nm. Since it is known that SDS3 is able to homodimerise through its N-terminal coiled-coil domain (Alland et al. 2002), this result was not surprising as this could also cause the ternary complex to form a 260 kDa homodimer that elutes early from the gel filtration column. Moreover, Millard et al. 2013 have shown that the HDAC1/MTA1 complex also dimerises, which could be suggesting that these complexes function as homodimers.
Figure 4.9: Large-scale ternary complex purification. Flag-tagged HID 531-772 was co-transfected in a 2.0 L culture of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1) and untagged SDS3 (U-SDS3). The complex was affinity purified with an anti-Flag affinity resin as described in paragraph 2.5 of chapter 2 passed through a S200 size exclusion chromatography column in gel filtration buffer with 50 mM salt. A: First purification step. The gel shows the protein bound to the resin, the resin after the TEV elution and the TEV eluate. B: Gel filtration fractions. C: S200 gel filtration chromatogram.

Purifications of the ternary complex as described above allowed me to carry out a large number of crystallization trials with all the crystallization screens listed in paragraph 2.1.6 of chapter 2 with the exception of MIDAS™. Protein concentrations used ranged from 6.0 mg/ml to 9.0 mg/ml, and set up at room
temperature and at 4°C. Unfortunately these crystallisation trials only produced amorphous precipitation and no hits could be observed across the multiple conditions on the plates.

4.3.2 Large-scale purifications of full ternary complex: analysis of nucleic acids impurities.

Failure to crystallise the full ternary complex following the method used in paragraph 4.3.1 suggested that the early elution of the complex (close to the void volume, Figure 4.9) could have been caused by factors other than complex dimerisation in solution.

To check for the presence of impurities co-eluting with the complex during the gel filtration, the first 6 gel filtration fractions of a full ternary complex purification (Figure 4.9bis) were concentrated to a final volume of 300 µl and split into 3 different 1.5 ml centrifuge tubes. The first sample was treated with 5 µl of H₂O and represented the control, the second sample was treated with 5 µl 10 mg/ml DNase and the last one was treated with 5 µl of 10 mg/ml RNase A. These were incubated at 37°C for 1 hour and were then run on a 1% agarose gel stained with EtBr at 120 V for 30 minutes. The results visible from Figure 4.10 suggested that the complex may be mostly contaminated with RNA impurities and too a lesser extent with DNA. Indeed, since the nucleic acid band formed by the DNase lane is slightly fainter than the on in the control, it was possible to deduce that DNA contaminants may represent a very small portion of the impurities co-eluting with the complex, which instead seems to be in large part contaminated by RNA. From Figure 4.9bis it is possible to see that the complex elution began
approximately 8.5 ml after the injection point and reached a maximum absorbance peak of 307 at 280 nm. However, it may also be noted that SDS3 ratio seems to be progressively decreasing (Figure 4.9bis B). Attempts to change the ratio of plasmid DNAs at the transfection stage to tackle this problem were unsuccessful.

**Figure 4.9bis:** Large-scale ternary complex purification (bis). Flag-tagged HID 531-772 was co-transfected in a 2.0 L culture of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1) and untagged SDS3 (U-SDS3). The complex was affinity purified with an anti-Flag affinity resin as described in paragraph 2.5 of chapter 2 and passed
through a S200 size exclusion chromatography column in gel filtration buffer with 50 mM salt. **A**: S200 gel filtration chromatogram. **B**: Gel filtration fractions.

![Figure 4.10: Nucleic acid contaminants.](image)

Figure 4.10: Nucleic acid contaminants. The nucleic acid bands generated by the control (undigested) and DNase lanes did not properly enter the gel, suggesting that they could be still interacting with the protein complex. The RNase A lane only produced a very faint band which also did not enter the gel.

Watson *et al.*, 2012a found that when the HDAC3/SMRT-DAD complex is gel-filtrated in salt concentrations higher than 100 mM the interaction between the two proteins is lost due to the displacement of the IP$_4$ molecule that acts as a molecular bridge between the two proteins. Knowing how the Sin3a complex behaves in higher salt may give a partial insight to the nature of the interactions
among its subunits and establish whether the impurities detected in Figure 4.10 could be washed away with a higher salt molarity in the gel filtration buffer. A TEV eluted sample of ternary complex was purified from 2 L of cells split into three different vials and then sequentially further purified by size exclusion chromatography in three different salt concentrations: 50 mM K-acetate, 150 mM K-acetate and 300 mM K-acetate. While DNase and RNase A were not added at any point of the purification, the sonication step carried out for this purification consisted of 3 cycles of 30 seconds each, instead of 3 cycles of 15 seconds each.

From Figure 4.11 it is possible to see that salt concentration did not have an impact on the stability of the complex during the gel filtration. Nevertheless, the complex eluted with approximately an ~2 ml delay compared to previous gel filtrations of the full ternary complex (Figures 4.9 and 4.9bis). This may have been caused by the increased sonication step shredding the contaminant nucleic acids causing the complex to elute later as a smaller molecular weight complex. The absorbance peaks visible at 18 ml were increasingly bigger with increasing salt concentrations. While it was not possible to detect any proteins in these peaks through Coomassie staining, later UV spectra of the same fractions showed a very low absorbance at 280 nm and 230 nm, potentially indicating the presence of very small traces of protein (Figure 4.12). Nevertheless, since these proteins were present at very low concentrations it was decided not to give them much significance. The small peaks shouldering the main peak (at approximately 13 ml) also increased in size with increasing salt concentrations, but contained protein complex that could not be distinguished from the main peaks for either
quality or quantity (Figure 4.11). For this experiment it was not possible to visualize the elution fractions of the complex gel-filtrated in 300 mM salt because the protein would precipitate when mixed with the protein-loading buffer, possibly because of pH issues. This also seemed to be the reason why the 150 mM salt gel-filtration fractions appeared to be less intense than the 50 mM ones. Finally, it should be noted that the chromatogram of the complex eluted at 50 mM salt is smaller than the 150 and 300 mM salt gel filtrations peaks because of a technical reason concerning the ÄKTA purifier rather than the effect of the concentration of salt.

Figure 4.11: Full ternary complex S200 gel-filtrations at different salt concentrations. Flag-tagged HID 531-772 was co-transfected in a 2.0 L culture of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1) and untagged SDS3 (U-SDS3). The complex was affinity purified with an anti-Flag affinity resin as described in paragraph 2.5 of chapter 2. The protein sample was then split into three different samples and sequentially purified by size exclusion chromatography in a S200 column in gel filtration buffer with 50 mM, 150 mM and 300 mM salt. A: Elution fractions of the ternary complex
eluted in 50 mM K-acetate (from ml 10 to ml 18 of the corresponding chromatogram in C). B: Elution fractions of the ternary complex eluted in 150 mM K-acetate (from ml 10 to ml 18 of the corresponding chromatogram in C). C: S200 gel-filtration chromatograms at 50, 150 and 300 mM K-acetate concentrations.

Figure 4.12: UV spectra of the gel filtration peaks at 18 ml (Figure 4.11 C). The measurements were taken with an Implen NanoPhotometer®. A: 18 ml fraction of the 50 mM salt gel-filtration (Figure 4.11). B: 18 ml fraction of the 150 mM salt gel-filtration fraction (Figure 4.11). C: 18 ml fraction of the 300 mM salt gel-filtration fraction (Figure 4.11).

Having established that salt concentration has little or no effect on the stability of the complex and that longer sonication times may be able to reduce the amount of impurities contaminating the protein, I wanted to verify whether salt and RNase A could remove the RNA impurity of the complex independently on the sonication time. This experiment was carried out by splitting a sample of affinity-purified
protein complex (sonicated for 3 cycles of 15s each) into two different vials. The anti-flag resin of each of the samples was washed according to the protocol in paragraph 2.5.4 of chapter 2, with the difference that immediately after the wash in 300 mM salt, an extra wash in 50 mM K-acetate was carried out at 4°C for 20 minutes. During this extra wash, one of the vials was treated with 5 µl of H₂O, representing the control, while the second one was treated with 5 µl of 10 mg/ml RNase A. Both the samples were then gel-filtrated in a S200 column in 150 mM K-acetate. The control sample generated two major peaks: a void peak in correspondence to the elution volume of the complex shown in Figure 4.9 and 4.9bis and a second peak in correspondence to the elution volume of the complexes that were sonicated for a longer period of time (Figure 4.11 C). The same result could be observed when a separate protein sample (not treated with RNase A) was passed through a S200 size exclusion chromatography column in 300 mM salt (Figure 4.13 B). This demonstrated that higher salt molarities were indeed able to separate the nucleic acid impurities from the purified protein. On the other hand, the RNase A-treated sample produced a gel filtration chromatogram that was similar to those formed by the complexes that were sonicated for a longer period of time (Figure 4.11 C), therefore no absorbance peaks were generated within the void volume of the column. To confirm the nature of these contaminants, a small part of the void peak and the protein peak shown in Figure 4.13 B were mixed with 1 volume of phenol-chloroform and the aqueous phases were run in a 1% agarose gel stained with EtBr. As expected, the void peak contained a smear of nucleic acids (presumably RNA) while the protein peak did not contain any detectable amount of contamination.
**Figure 4.13:** Effect of salt concentrations and RNase A on the purity of the ternary complex. Flag-tagged HID 531-772 was co-transfected in 2.0 L cultures of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1) and untagged SDS3 (U-SDS3). The complex was affinity purified with an anti-Flag affinity resin as described in paragraph 2.5 of chapter 2 and passed through a S200 size exclusion chromatography column in gel filtration buffer with 150 mM and 300 mM salt. **A:** 150 mM salt gel filtration chromatograms: control vs. RNase A-treated. (1 protein sample split in to two different vials) **B:** 300 mM salt gel filtration chromatogram. **C:** 1% agarose gel stained with EtBr showing the nucleic acid content of the void and protein peaks shown in B. 200 µl of the void peak sample were taken from the fraction at 9.5 ml while 200 µl of the protein peak sample were taken from the fraction at 11.0 ml.

From these experiments it was possible to conclude that the stability of the complex was entirely preserved both at 150 and 300 mM salt (Figure 4.13 A, B; 4.11 A, B, C), which was significantly different from what Watson et al. (2012b) had observed with the HDAC3/SMRT-DAD complex. This could be suggesting that the nature of the interactions among the Sin3a ternary complex subunits may be at least partially hydrophobic or that these interactions occur through much
larger surfaces and therefore the amount of salt used was not sufficient to disrupt the integrity of the complex.

4.3.3 Crystallisation trials with the full ternary complex cleared form nucleic acid contaminants.

The experiments reported in 4.3.2 allowed me to perform a large number of crystallization trials using protein samples that lacked the nucleic acid contaminants. All the crystallization screens listed in paragraph 2.1.6 of chapter 2, with the exception of MIDAS™, were used at least once with protein samples purified by size exclusion chromatography in 150 mM K-acetate and once with protein samples in 300 mM K-acetate. Overall, less than 50% conditions of the PACT premier and Stura and MacroSol screens set up at room temperature managed to precipitate the protein complex at 8 mg/ml, therefore these screens were also set up at 4°C, unfortunately producing the same results.

Crystallization plates set up with the protein passed through a S200 size exclusion chromatography column in 300 mM K-acetate mostly showed heavy and amorphous precipitation (> 50%) while ~10-30% of the conditions produced phase separation. The remaining conditions failed to produce any kind of precipitation. On the other hand, trials set up with the complex passed through a S200 size exclusion chromatography column in 150 mM K-acetate buffer generally produced lighter precipitation, although this was dark and amorphous in approximately 50% of cases. This also produced phase separation in ~10-30% of the conditions. Most of the remaining conditions did not contain any form of precipitation, with the exception of 8, which produced a translucent but not
birefringent precipitate that seemed to present rather solid shapes. The 8 conditions were ProPlex™ E2; ProPlex™ E3; ProPlex™ E6; ProPlex™ E12; ProPlex™ D8; Morpheus® C2; Morpheus® C3; Morpheus® A2, in which the protein was set at a concentration of \(~8.5 \text{ mg/ml}\). The conditions were optimised around precipitant and pH as described in detail in Figure 4.14.

**Figure 4.14:** Ternary complex crystallisation trials: first round of optimization. The figure represents a 96 well MRC plate for crystallization trials and each square includes a total of 9 conditions (8 optimised plus the original condition of the screen reported in the figure).

While all of the outcomes of the optimised conditions shown in Figure 4.14 were the same as the originals, the crystalline precipitation generated by all of the 9 conditions set up around Morpheus® A2 presented a hint of birefringence that was previously absent. However, since the birefringence was present in all of the 9 optimised conditions set up around “Morpheus® A2”, this change was likely to be caused by very small differences in the protein sample rather than to the modifications applied to the conditions. I then decided to set up a second
optimisation plate with the aim of looking at the effect of different components of the Morpheus® A2 condition (Figure 4.15). Since the original condition contains two different types of divalent salts (CaCl$_2$ and MgCl$_2$) I designed two conditions in which one of the salts was missing, one condition lacking both of them and one condition containing a monovalent salt (NaCl) instead. Morpheus® A2 also contains two precipitants (ethylene glycol and PEG 4K), therefore I designed two conditions in which one of the precipitants was missing and one condition in which the concentration of both of the precipitants was halved. Since PEG and pH variations did not have any effect on the crystallisation of the complex, I decided to keep these two parameters unchanged from the original condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Row 1</th>
<th>Row 2</th>
<th>Row 3</th>
<th>Row 4</th>
<th>Row 5</th>
<th>Row 6</th>
<th>Row 7</th>
<th>Row 8</th>
<th>Row 9</th>
<th>Row 10</th>
<th>Row 11</th>
<th>Row 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>No CaCl$_2$</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>No MgCl$_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Salt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150 mM NaCl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No PEG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Ethylene Glycol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/2 Precipitants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.15: Ternary complex crystallisation trials: second round of optimization.** The condition was optimised across the rows of the plate as indicated by the blue box on the left of the figure. The crystals (an example is shown in Figure 4.16) were generated only in the “No CaCl$_2$”, “No PEG” and “1/2 Precipitant” rows. No crystals could be seen in the control row.

On day 7, a number of rectangular and highly birefringent crystals (Figure 4.16) were generated in the “No CaCl$_2$”, “No PEG” and “1/2 Precipitants” rows of the plate. Their size was approximately 10 x 15 µm and they appeared to be remarkably flat as visible from Figure 4.16. No crystals were generated in the
“control” row, which instead generated the same type of precipitation described for the optimised screen in Figure 4.14. The crystals were then flash frozen in liquid nitrogen using 30% MPD as cryoprotectant and brought to the Diamond light source synchrotron in Oxford where they were analysed in the I-24 microfocus beamline. However, probably due to the presence of ice in the loops, determining the nature of these crystals was not possible. Moreover, because of their small size and very flat shape, solving the protein structure from these crystals would have been very difficult.

![Figure 4.16: Crystals obtained with the second round of optimisation.](image)

In order to improve the size and shape of the crystals, the screen shown in Figure 4.15 was further optimised by the addition of molecules (additives) commonly used in commercially available crystallisation screens. There are a
large variety of additives that can be used in crystallisation trials and that improve the quality of protein crystals in different ways. The additives used for this experiments were selected to include two monosaccharide sugars (glucose and mannose), a polysaccharide (sucrose), a long carbohydrate (agarose), an alcohol sugar (sorbitol), a charged amino acid (lysine), two halide salts (NaI and LiCl), a divalent salt (ZnCl₂) and two monovalent salts (KCO₃ and NaNO₃). These were added to the previous optimised screen as shown in Figure 4.17

<table>
<thead>
<tr>
<th>Control</th>
<th>+ 150mM Glucose</th>
<th>+ 150mM Mannose</th>
<th>+ 150mM Sucrose</th>
<th>+ 0.001% Agarose</th>
<th>+ 150mM Sorbitol</th>
<th>+ 150mM Lysine</th>
<th>+ 150mM ZnCl₂</th>
<th>+ 150mM NaI</th>
<th>+ 150mM LiCl</th>
<th>+ 1.50mM KCO₃</th>
<th>+ 150mM NaNO₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>No ClO₂</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
</tr>
<tr>
<td>No MgCl₂</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
</tr>
<tr>
<td>No Salt</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
</tr>
<tr>
<td>150mM NaCl</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
</tr>
<tr>
<td>Control</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
</tr>
<tr>
<td>No PEG</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
</tr>
<tr>
<td>No Ethylene Glycol</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
</tr>
<tr>
<td>1/2 Precipitant</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
<td>Morpheus A2</td>
</tr>
</tbody>
</table>

At day 7, the additive screen generated a large number of crystals (3-7 crystals in each drop) with a 6-sided bi-pyramidal shape of approximately 30 x 20 x 20 µm in size. The distribution of the crystals throughout the plate (Figure 4.17) suggested that the presence of an additive was essential for the formation of the crystals (with the exception of ZnCl₂ that did not produce any crystals), although it did not matter which one. While the “no additive” control did not generate any crystals, 7 out of 11 additives generated crystals when mixed with the original Morpheus® A2 condition. The “No salt”, “NaCl” and “no ethylene glycol” rows
were unable to generate crystals, while it seemed evident that the most favourable condition for the crystal nucleation was the absence of PEG (Figure 4.17). The crystals (an example of which is shown in Figure 4.18 A) were very fragile when touched with the nylon loops.

After they were cryo-protected with 40% ethylene glycol and flash-frozen with a cryostream flowing at 100ºK, the crystals were transferred to the Diamond light source synchrotron in Oxford and analysed in the I-24 beamline. To our surprise, the X-ray diffraction pattern obtained by analysing the crystals (an example in Figure 4.18 B) was typical of a salt crystal, with only a few high-intensity diffraction spots mostly found at high resolution.

![Figure 4.18: Crystal obtained with the third round of optimisation. A: Snapshot of the crystal taken at the I-24 beamline of the Diamond synchrotron. B: Diffraction pattern of the crystal shown in A. The few high-intensity spots mostly located at high resolution indicate that the crystal was made of salt.](image)

To verify the condition of the protein after being incubated for 21 days at room temperature, I decided to take some of the protein from a JCSG plus screen set up 3 weeks earlier and run it on an SDS-PAGE gel followed by Coomassie staining. The results (Figure 4.19) showed that with the exception of HDAC1,
the complex was highly degraded. The Sin3a-HID could not be identified anywhere in the gel, while the mass-spec data indicated that SDS3 was reduced to small peptides of an average molecular weight of 10-14 kDa.

To avoid this degree of protein degradation in the future, I decided to double the concentration of protease inhibitors in the lysis buffer as well as keeping a 0.1 X concentration of protease inhibitors during the TEV elution step. Finally, after the gel filtration, protease inhibitors were added to the concentrated protein to a final concentration of 0.1 X. In fact, all the protein purifications performed in the following paragraphs were carried out with the addition of extra-protease inhibitors as described above.
Figure 4.19: 3 weeks old ternary complex from a JCSG plus screen (12 conditions). With the exception of HDAC1, which remained intact and was clearly visible in the gel, the remaining components of the complex had to be identified through mass spectrometry analysis (MALDI-ToF, Figure 4.20). The protein band visible below the 3 kDa mark could not be identified either by MALDI-ToF nor LC MS-MS.
Figure 4.20: Mass spectrometry analysis of 3 weeks old ternary complex (Figure 4.19). A: Peptides detected from the protein band at 17 kDa visible in Figure 4.19. These peptides do not belong to any component of the complex. B: Peptides detected from the protein band corresponding to 14 kDa visible from Figure 4.19. These peptides belong to SDS3. The peptide positions within the protein have been highlighted in red. C: Peptides detected from the protein band corresponding to approximately 8 kDa visible from Figure 4.19. These peptides also belong to SDS3. The peptide positions within the SDS3 protein have been highlighted in red.
4.3.4: Crystallisation trials with the full ternary complex gel-filtrated with a Superose®6 (S6) size exclusion chromatography column.

In order to produce a purer protein sample for additional crystallisation trials, it was decided to gel-filtrate the complex with a softer matrix (S6 10 x 300 mm). This would provide a better separation between the complex and the RNA impurities. The protein samples utilised for the crystallisation trials reported in this paragraph were prepared with the addition of extra protease inhibitors and RNase A as described in paragraph 4.3.3.

As a first experiment, a 50 mM salt gel filtration was done to observe whether the column provided a good separation even at low K-acetate molarities, since low salt concentrations might reduce the possibilities of crystallising salt in the crystallisation trials. As visible from Figure 4.21A, despite the treatment with RNase A, when the complex was further purified by size exclusion chromatography in 50 mM K-acetate, it eluted within the void volume of the column, approximately 8.5 ml after the injection point. Since the absorbance at 260 nm was higher than the absorbance at 280, it was possible to deduce that a large amount of nucleic acid contaminant was present in the sample. Later analysis of the peak by agarose gel electrophoresis (Figure 4.22) revealed that the contaminants were fully digested by the RNase A added during the purification but probably co-eluted with the purified protein due to electrostatic interactions with the complex which were not displaced by the low salt concentration of the gel filtration buffer. It was possible to deduce this because there was no difference between the control and the treated samples (Figure 4.22), thus it seemed clear that the RNA contaminants were interacting non-
specifically with the complex. Altogether, these results implied that gel filtrating the complex at low salt conditions is not advisable. When a second protein sample was eluted in 150 mM K-acetate, a 280 nm absorbance peak was observable at approximately 13 ml after the injection point (Figure 4.21 B). The peak was followed by two 260 nm absorbance peaks, indicating that the nucleic acid contaminants had been successfully separated from the complex.

**Figure 4.21:** *S6 gel filtrations at two different salt concentrations.* Flag-tagged HID 531-772 was co-transfected in 2.0 L cultures of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1) and untagged SDS3 (U-SDS3). The complex was affinity purified with an anti-Flag affinity resin as described in paragraph 2.5 of chapter 2 and in a S6 size exclusion chromatography column in 50 mM and 150 mM salt. **A:** Size exclusion chromatography trace of full-ternary complex eluted in 50 mM K-acetate. **B:** Size exclusion chromatography trace of full ternary complex eluted in 150 mM K-acetate. **C:** Gel filtration fractions coming from the chromatography shown in B.
Figure 4.22: Nucleic acid contaminants co-purified with the ternary complex. The contaminants were phenol/chloroform extracted from the protein sample and ran on an agarose gel, which was stained with EtBr. The control sample was untreated and came straight from the gel filtration peak (Figure 4.21 A). The other 2 samples were digested with RNase A and DNase respectively.

For the crystallisation trials, the first 7 fractions of the gel filtration were pooled together and concentrated. Indeed, starting from the 8\textsuperscript{th} fraction visible in Figure 4.21C, SDS3 was present in lower proportions, potentially compromising the homogeneity of the protein sample if mixed with the earlier fractions. With the exception of Morpheus\textsuperscript{®} and PACT premier, all crystallisation screens listed in chapter 2 were utilised with the complex used at a concentration of 8.0 mg/ml. Unfortunately, no significant hits were produced. In fact, over 50\% of the conditions resulted in phase separation or light and amorphous precipitation. The state of the protein degradation was assessed after it had been sitting for 3 weeks at room temperature in a crystallisation plate. The protein was extracted form 16 conditions of the JCSG plus screen and compared with a vial of the
same purification that was stored at the same temperature and for the same amount of time (Figure 4.23). From the gel shown in Figure 4.23 it appeared clear that the action taken to dramatically reduce the degree of protein degradation was successful.

**Figure 4.23:** Comparison of 3 weeks old ternary complex from a JCSG plus screen and a sample stored at room temperature for the same amount of time. There are no significant differences between the two samples but the addition of extra protease inhibitors to the protein sample dramatically improves the stability of the complex compared to standard amounts of protease inhibitors (Figure 4.19).
4.3.5 Small-scale purifications of the ternary complex using shorter versions of the HID

After the crystallisation trials reported in the previous paragraphs, more protein samples were prepared using the same constructs in the presence of extra protease inhibitors and RNase A (see paragraphs 4.3.3 and 4.3.4) but with inconclusive results.

As mentioned in chapter 3, HID 531-772 might be too long and it would be reasonable to think that not all of it is involved in the interaction with HDAC1 and SDS3. Indeed, a number of publications have reported that the HID also interacts with HDAC2 (Laherty et al. 1997), SAP180 and SAP130 (Fleischer et al. 2003) and MRG15 (Carrozza et al. 2005). Therefore, there might be large regions of the domain that in the ternary complex would be unoccupied and possibly unfolded, potentially hindering crystal nucleation and making the protein more susceptible to proteolytic digestion. Thus, I decided to test the shorter HID constructs already used in the mapping experiments reported in chapter 3 (Figures 3.6 and 3.7).

Despite these constructs expressing poorly, the importance of solving the structure of the complex pushed me to try all possibilities and carry out 2 small-scale experiments in HEK 293F suspension cells. Furthermore, PEI (which was used with these constructs for the mapping experiments) can transflect approximately 1-2% of HEK 293T monolayer cells and over 70% HEK 293F suspension cells (information kindly provided by Grace Hodson). Therefore, the amount of protein expressed in the suspension system might be significantly higher than in the monolayer cells.

The small-scale experiments were conducted according to the standard protocol
(2.4.6). As expected, when the short HID constructs were co-expressed with SDS3 and HDAC1 in the suspension cells system, the protein yield was significantly higher than the one obtained with the HEK 293T monolayer cells in the mapping experiments. Indeed, the HID fragments were expressed sufficiently to be easily visualised by Coomassie staining (Figure 4.24 A). However, the yields were still not sufficient to allow me to use the short version of the HID for crystallisation trials. Nevertheless, from the gel in Figure 4.24 A, it can be observed that the behaviours of HID 651-772 and HID 651-741 replicated the results of the mapping experiments (3.6 and 3.7). Indeed, HID 651-772 had a better yield than HID 651-741 but formed a very weak interaction with HDAC1. On the contrary, HID 651-741 expression was lower than HID 651-772 but it formed a relatively stronger interaction with the enzyme. HID 688-741 was the construct that showed the lowest expression levels and it is visible from Figure 4.24A in line with the 14 kDa marker. This construct pulled down a very low amount of untagged HDAC1, although this amount was in fact proportionally higher than the one pulled down by the other short HID constructs used in this series of experiments. This indicated that HID 688-741 can form a strong interaction with the enzyme and that it might be one of the main interacting regions of Sin3a with HDAC1 (HID predicted helices 5 and 6). Finally, of the four constructs, HID 688-741 is the fragment that showed the highest expression levels and that showed the weakest interaction with HDAC1, probably because of folding issues. A second experiment was designed to invert the IP and pull down Myc-tagged versions of the same small HID constructs with a C-terminally Flag-tagged HDAC1. This was an attempt to get a clearer answer of the interaction
profiles of the short HID constructs with HDAC1. However, the amount of HID pulled down by the enzyme was too low to even be detectable by Coomassie staining (Figure 4.24 B).

![Figure 4.24: Short HIDs: small-scale purifications. Flag and Myc and Flag-tagged HIDs were co-transfected in 30 ml cultures of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1), Flag-tagged HDAC1 (HDAC1) and untagged SDS3 (U-SDS3). The complexes were affinity purified with an anti-Flag affinity resin as described in paragraph 2.4 of chapter 2. A: Flag Co-IP through the short HID constructs. The constructs pulled down very low amounts of untagged HDAC1 similarly to the results of the mapping experiments shown in Figures 3.6 and 3.7. B: Flag Co-IP through HDAC1. The enzyme pulled down very low amounts of Myc-tagged HID constructs, which could not be seen in the gel. As already mentioned above, because of their poor expression, the short HIDs could not be employed for large-scale purifications and crystallisation trials.](image-url)
4.3.6 Large-scale purifications and crystallisation trials of the ternary complex with a construct of SDS3 that lacks the coiled-coil domain

Given that crystallisation trials with the full ternary complex produced inconclusive results and that it was not possible to express shorter versions of the HID in sufficient amounts to be used for large-scale purifications, it was decided to focus on SDS3 to understand whether it could be manipulated to ease the crystal formation process. As shown in paragraph 4.3.1, when this protein was added to the binary complex it significantly improved the final yield of purified protein, confirming its function as complex-stabiliser as suggested by (Alland et al. 2002).

There are currently no structures of SDS3 in the public databases but the internal domain organization of the protein is well known and the region able to interact with Sin3a has already been identified by (Alland et al. 2002). Indeed, this a 323 amino acid long protein with an N-terminal coiled-coil (CC) domain, an extended-secondary-structure-rich C-terminal region and a central (aa 188-229) Sin3a Interaction Domain (SID) (Alland et al. 2002). As previously mentioned in this chapter, SDS3 can form homodimers. Therefore, I hypothesised that the dimerisation of SDS3 may in turn cause the dimerisation of the Sin3a complex in vitro, similarly to the MTA1-HDAC1 complex (Millard et al. 2013). Depending on its relative position within the ternary complex in solution, the CC domain of SDS3 may hinder the nucleation of protein crystals. Moreover, a ternary complex dimer of ~260 kDa may create the need for larger crystals to generate a useful data set.

Keeping this in mind, alternative SDS3 constructs were generated with the aid of
bioinformatics tools to predict its disorder levels and secondary structures (Figure 4.25).

Figure 4.25: Disorder and secondary structure predictions of SDS3 (RONN and Jpred). The blue line represents the predicted disorder: values above 0.5 are considered to be disordered (RONN). Each red box represents a predicted alpha helix (Jpred) and was numbered for convenience. The level of confidence for each predicted secondary structure of the domain is indicated by the black and green bar, where confidence levels above 7/9 are represented in green. Each yellow box represents an extended secondary structure region, while the blue boxes indicate the presence of a predicted coiled-coil domain.

From the RONN prediction (Figure 4.25) it is possible to see that the first 50 N-terminal amino acids have a high probability of being unstructured, which is supported by the absence of Jpred-predicted secondary structures. The region comprising amino acids 50-170 seems to have low entropy due to the presence of one or two large helices that form the coiled-coil domain. The region between amino acids 210-230 is predicted to have a short alpha helix (helix 3). Interestingly, this helix is found within the known SID domain of the protein (aa
188-229), raising the hypothesis that SDS3 interacts with Sin3a through an alpha helix. The C-terminal region of the protein seems to be highly organized with extended secondary structures, which probably confer high globularity to this region of the protein. On the basis of these bioinformatics searches, 5 alternative SDS3 constructs were designed as follows. SDS3 51-328 was designed in order to lose the unstructured 50 N-terminal amino acid region of the protein; SDS3 61-328 was designed by focusing more on the prediction of the coiled-coil domain and trying to exclude a further 10 amino acids potentially not part of the domain; SDS3 91-238 removes the predicted alpha-helix 1, in correspondence of the beginning of a low-disorder region of the protein; SDS3 201-328 excludes the whole of the coiled-coil domain and start the construct with the region of the protein responsible for binding Sin3a; finally, SDS3 249-328 was designed to only include the C-terminal region of SDS3 and it was used as a negative control because it does not include the SID domain. A schematic of these constructs are summarized in Figure 4.26.

The following small-scale experiments were carried out according to the protocol for small-scale purifications (paragraph 2.4 of chapter 2).
With the exception of SDS3 249-328, all of the designed constructs interacted with the HID/HDAC1 complex (Figure 4.27). As expected, the protein bands of SDS3 51-323 and SDS3 61-328 partially overlapped with the Flag-HID 531-772 band. SDS3 91-328 is visible between the IgG light chain of the anti-Flag antibody and the HID, while SDS3 201-238 runs at approximately 16 kDa. Finally, SDS3 249-328 could not be detected in the gel, confirming that the region that spans from amino acid 201-248 is essential for the interaction with the co-repressor. The fact that SDS3 201-328 was successfully expressed and formed an interaction with Sin3a is in contrast with previous published findings that considered the coiled-coil of SDS3 domain as necessary but not sufficient for the interaction with Sin3a (Alland et al. 2002).
Figure 4.27: Small-scale purifications of the ternary complex with alternative SDS3 constructs. Flag-tagged HID 531-772 was co-transfected in 30 ml cultures of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1), and untagged SDS3 constructs (SDS3). The complexes were affinity purified with an anti-Flag affinity resin as described in paragraph 2.4 of chapter 2. The red stars indicate the various pulled-down SDS3 constructs. With the exception of SDS3 249-328, all of the other truncated versions of the protein interacted with Sin3a and HDAC1.

Since SDS3 201-328 (Δ-CC) lacked the coiled-coil domain and could still form a robust interaction with the complex, it was decided to use this construct for the next set of large-scale purification and crystallisation trials. These were carried out according to the protocol for large-scale purifications (paragraph 2.5 of chapter 2), plus the addition of RNase A and extra protease inhibitors as described in paragraph 4.3.3.

As shown in Figure 4.28 A and B, the Δ-CC complex was produced in sufficient amounts (approximately 450 µg per litre of cell culture) for a number of
crystallisation trials and eluted from the S200 column at approximately 12.5 ml after the injection point and reached a maximum absorbance peak of 35 at 280 nm. The shape of the peak slope (4.28 B and C) is probably due to the progressively decreasing amounts of SDS3 as already observed with the full length SDS3 ternary complex (Figure 4.9 bis). For these reasons, only fractions from ml 12.5 to 15.0 (first 6 fractions of the gel in Figure 4.28 B) were pooled together and concentrated for the crystallisation trials.

**Figure 4.28:** Example of a large-scale Δ-CC ternary complex purification. Flag-tagged HID 531-772 was co-transfected in 2.0 L cultures of HEK-293F suspension cells with untagged HDAC1 (U-HDAC1), and untagged SDS3 (U-SDS3 201-328). The complex
was passed through a S200 size exclusion chromatography column in 150 mM salt. A: First purification step. The gel shows the protein bound to the resin, the resin after the TEV elution and the TEV eluate. B: Gel filtration fractions (from ml 12.0 to ml 20.0). C: S200 gel filtration chromatogram.

All the commercial crystallisation screens listed in paragraph 2.1.6 were tested with the Δ-CC ternary complex using protein concentrations ranging from 6.0 mg/ml to 11.0 mg/ml. 24 hours after setting up the crystallisation plate, a protein sample concentrated at 10.2 mg/ml produced a birefringent crystalline material with variable and not definite geometrical shapes (Figure 4.29 A) in a condition containing PPG 400 as a precipitant (MIDAS™ A12). 6 days later, 10 x 10 µm and lightly birefringent crystals (Figure 4.29 B) formed in a related condition also containing PPG 400 (MIDAS™ A7). Optimization plates around MIDAS™ conditions containing PPG 400 were set up as shown in Figure 4.30. In the optimised plate, all conditions with over 30% v/v PPG 400 (50% of all conditions) produced large “oily” drops of phase separation. Conditions containing between 20-30% v/v PPG 400 occasionally showed phase separation with no apparent trend across the plate and mostly produced clear drops. Finally, conditions containing 10% and 15% v/v PPG 400 did not produce any kind of precipitation or phase separation with the exception of conditions also containing 10% v/v 1-propanol, which produced a light and amorphous precipitation across all of the plate. Therefore, the crystals shown in Figure 4.29 B could not be reproduced in the optimised screen.
Figure 4.29: Crystals obtained with the Δ-CC ternary complex crystallisation trials. A: MIDAS A12 condition, day 1: crystalline precipitation obtained with protein concentrated at 10.2 mg/ml. B: MIDAS A7 condition, day 6: ~10 x 10 µm crystals obtained with protein and concentrated at 10.2 mg/ml.
Since with the first round of optimizations it was not possible to reproduce the screen based on some of the MIDAS™ conditions containing PPG 400 (A7, B11, A12, D1). The boxes in green represent original MIDAS conditions. PPG 400 and 1-Propanol are represented as a v/v percentage.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5% PPG400</td>
<td>No buffer</td>
<td>10% PPG400</td>
<td>No buffer (A7)</td>
<td>15% PPG400</td>
<td>No buffer</td>
<td>20% PPG400</td>
<td>No buffer</td>
<td>25% PPG400</td>
<td>No buffer</td>
<td>30% PPG400</td>
<td>No buffer</td>
</tr>
<tr>
<td>B</td>
<td>5% PPG400</td>
<td>10% 1-Propanol</td>
<td>No Buffer</td>
<td>10% 1-Propanol</td>
<td>No Buffer</td>
<td>10% 1-Propanol</td>
<td>No Buffer</td>
<td>10% 1-Propanol</td>
<td>No Buffer</td>
<td>25% PPG400</td>
<td>No buffer</td>
<td>30% PPG400</td>
</tr>
<tr>
<td>C</td>
<td>5% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>10% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>15% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>20% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
</tr>
<tr>
<td>D</td>
<td>5% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>10% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>15% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>20% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
</tr>
<tr>
<td>E</td>
<td>5% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>10% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>15% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>20% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
</tr>
<tr>
<td>F</td>
<td>5% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>10% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>15% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>20% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
</tr>
<tr>
<td>G</td>
<td>5% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>10% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>15% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>20% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
</tr>
<tr>
<td>H</td>
<td>5% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>10% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>15% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
<td>20% PPG400</td>
<td>0.1M K/Na Phosphate</td>
<td>pH 7.0</td>
</tr>
</tbody>
</table>
crystals obtained with the original MIDAS™ screen, the crystals shown in Figure 4.29 B were cryoprotected in 40% v/v PPG 400, flash frozen in a nitrogen cryostream flowing a 100ºK and brought to the Diamond light source synchrotron in Oxford, where they were analysed in the I-24 beamline.

As shown in Figure 4.31 B, the crystals failed to generate any X-ray diffraction, indicating that they were made of protein material but that their internal structure was too heterogeneous for them to diffract the X-ray beam. While this result represented a significant progress in the attempt of generating crystals of the Sin3a ternary complex, they still required optimisation to generate a data set.

![Figure 4.31: Δ-CC ternary complex crystal obtained with the original MIDAS A7 condition and diffraction pattern. A: Snapshot of the crystal taken at the Diamond synchrotron. The red grid represents the “grid-scan” function available at the synchrotron’s I-24 beamline, ought to find the best diffraction area in a region of the loop or within a crystal. B: Diffraction pattern of the crystal shown in A. No diffraction was observed.](image)

Since the original MIDAS A7 condition was the only one that generated protein crystals, I decided to use a range of additives without altering the amount of precipitant and pH levels of the original condition. Protein complex was used in concentrations ranging from 5.0 mg/ml to 11.0 mg/ml. The additive screen
utilised for this optimization round was designed by Dr. Benjamin Goult in 2009 and it included a large variety of chemicals and amino acids commonly used in crystallisation trials (Figure 4.32). However, all of the optimised conditions resulted in clear drops with the exception of the conditions containing isopropanol and acetonitrile, which generated a light and amorphous precipitate.

<table>
<thead>
<tr>
<th>ADDITIVE-MIDAS A7</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>MIDAS A7 +10mM CaCl₂</td>
<td>MIDAS A7 +10mM NiCl₂</td>
<td>MIDAS A7 +100mM NaBr</td>
<td>MIDAS A7 +3% v/v DMSO</td>
<td>MIDAS A7 +3% v/v glucose +10μM Na₃</td>
<td>MIDAS A7 +4% v/v Acetone</td>
</tr>
<tr>
<td>B</td>
<td>MIDAS A7 +10mM BisCl₂</td>
<td>MIDAS A7 +100mM (NH₄)₂SO₄</td>
<td>MIDAS A7 +100mM KSCN</td>
<td>MIDAS A7 +100mM Glycine</td>
<td>MIDAS A7 +3% v/v Sucrose +10μM Na₃</td>
<td>MIDAS A7 +3% v/v Methanol</td>
</tr>
<tr>
<td>C</td>
<td>MIDAS A7 +10mM CoCl₂</td>
<td>MIDAS A7 +100mM KCl</td>
<td>MIDAS A7 +400mM KSCN</td>
<td>MIDAS A7 +30mM Gly-Gly-Gly</td>
<td>MIDAS A7 +3% v/v ethylene glycol</td>
<td>MIDAS A7 +4% 1-Butanol</td>
</tr>
<tr>
<td>D</td>
<td>MIDAS A7 +10mM CuCl₂</td>
<td>MIDAS A7 +100mM LiCl</td>
<td>MIDAS A7 +100mM KSCN</td>
<td>MIDAS A7 +10mM Co(NH₃)₃CO</td>
<td>MIDAS A7 +3% v/v glyceroil</td>
<td>MIDAS A7 +4% 1-Butanediol</td>
</tr>
<tr>
<td>E</td>
<td>MIDAS A7 +10mM MgCl₂</td>
<td>MIDAS A7 +200mM NaCl</td>
<td>MIDAS A7 +100mM K/Na tartarate</td>
<td>MIDAS A7 +10mM taurine</td>
<td>MIDAS A7 +5% v/v TMAO</td>
<td>MIDAS A7 +3% Ethanol</td>
</tr>
<tr>
<td>F</td>
<td>MIDAS A7 +10mM MnCl₂</td>
<td>MIDAS A7 +10mM Yb₂(SO₄)₃</td>
<td>MIDAS A7 +100mM HCO(OH)OK</td>
<td>MIDAS A7 +10mM urea</td>
<td>MIDAS A7 +3% v/v 1,6 hexanediol</td>
<td>MIDAS A7 +3% v/v isopropanol</td>
</tr>
<tr>
<td>G</td>
<td>MIDAS A7 +10mM ZnCl₂</td>
<td>MIDAS A7 +10mM CdCl₂</td>
<td>MIDAS A7 +100mM Na₂ citrate</td>
<td>MIDAS A7 +100mM Guanidine HCl</td>
<td>MIDAS A7 +3% v/v PEG400</td>
<td>MIDAS A7 +4% v/v MeCN</td>
</tr>
<tr>
<td>H</td>
<td>MIDAS A7 +10mM FeCl₃</td>
<td>MIDAS A7 +100mM Nal</td>
<td>MIDAS A7 +10mM L-proline</td>
<td>MIDAS A7 +10mM EDTA</td>
<td>MIDAS A7 +5% v/v EtOAc</td>
<td>MIDAS A7 +4% v/v 1,2 propanediol</td>
</tr>
</tbody>
</table>

**Figure 4.32:** Δ-CC ternary complex crystallisation trials: second round of optimisation. This is an additive screen based on condition MIDAS™ A7.

I also wanted to verify the degree of protein degradation in a 3 weeks old screen set up at room temperature. In order not remove any protein from the crystallisation trials, a small aliquot of 9.0 mg/ml concentrated protein was stored at the same temperature as the crystallisation plates and run on an SDS PAGE gel followed by Coomassie staining. As visible in Figure 4.33, the protein was well preserved with an only small amount of proteolysis observable for HDAC1 and the HID. Remarkably, SDS3 Δ-CC was not affected by proteolytic
degradation and remained intact, indicating that this portion of SDS3 might be very globular and structured as predicted by the bioinformatics analysis.

![Figure 4.33: 3 weeks old Δ-CC ternary complex stored at room temperature.](image)

4.4 Results summary and conclusions

The principal aim of the experiments reported in this chapter was to generate protein crystals suitable for X-ray crystallography studies of the Sin3a complex. While obtaining the crystals was not possible, small and large-scale purifications of various forms of the complex have helped to characterise aspects of Sin3a function.

The binary complex formed by HID 531-772 and HDAC1 was successfully
expressed and purified in small and large-scale experiments. However, this form of the complex was not as stable as the ternary complex, which was stabilised by SDS3 and could be produced in significantly greater quantities. The coiled-coil domain of SDS3 was not required for the interaction with the HID or HDAC1 and the hypothesis that the coiled-coil domain could hinder the process of crystal nucleation was potentially confirmed by the fact that the Δ-CC ternary complex was able to form non-diffracting protein crystals unlike the full ternary complex. Size exclusion chromatography profiles of both the full and Δ-CC ternary complexes have shown that Sin3a is not susceptible to salt concentrations up to 300 mM K-acetate, which is remarkably different from the behaviour shown by the HDAC3/DAD complex purified and crystallised by Watson et al. 2012b. While the HID 531-772 construct is probably too big to be crystallised alone with HDAC1 and SDS3, shorter HID constructs spanning from amino acid 651 to 772 failed to express in sufficient quantities to be utilised in crystallisation trials. Nevertheless, the small Sin3a-HID constructs have confirmed that the predicted alpha helices 3-7 (Figure 3.3 of chapter 3) and in particular alpha helices 5 and 6 are particularly important for the interaction with HDAC1.
CHAPTER 5: Activity and binding assays

5.1 Chapter aims
Mapping experiments and large-scale purifications of the Sin3a binary and ternary complexes have provided partial insights to the nature of the interactions existing among the complex subunits. However, how the activity of this complex is regulated and how its activation profile compares to that of other transcriptional co-repressor complexes remains to be elucidated. As discussed in previous chapters, Sin3a differs from SMRT, NuRD, CoREST and MiDAC in that, seemingly, no subunits of the Sin3a complex contain a SANT domain or a region homologous to it. Since the SANT domain can interact with IP₄, it is possible to hypothesise that Sin3a may be IP₄ independent and therefore regulated in a different way. However, mapping experiments reported in chapters 3 and 4 have also shown that the HID has a region formed by 5 alpha helices –important for the interaction with HDAC1- that may fold to form a tertiary structure resembling a SANT domain potentially able to bind IP₄ through a different interaction mechanism.

In this chapter, the activation profile of Sin3a was investigated through HDAC activity and kinetic assays with the aim of finding whether the complex can be stimulated by inositol phosphate. Activity and kinetic assays methodologies are described in paragraphs 2.6 and 2.8 of chapter 2.

5.2 HDAC activity assays
(Watson et al. 2012b) demonstrated that the activity of the HDAC3/SMRT-DAD
complex is strictly dependent on the presence of IP$_4$. As shown in Figure 5.1A (adapted from Millard et al. 2013), it is noticeable that the HDAC3/DAD complex loses most of its activity when exposed to high concentrations of NaCl. This is due to the displacement of the IP$_4$ molecule responsible for activating the complex. However, the activity could be fully recovered when exogenous IP$_4$ was added back to the complex in low salt conditions (Watson et al. 2012b). Similar results were also observed for the MTA1/HDAC1 complex, the activity of which was dramatically increased in the presence of exogenous IP$_4$ (Figure 5.1 B, adapted from Millard et al. 2013). Unpublished data from Yun Song and Charles Milano have shown the same phenomenon with the HDAC1-dependent CoREST and MiDAC complexes.

**Figure 5.1: SMRT and NuRD activation profiles (Adapted from Millard et al., 2013).** A: HDAC3/SMRT-DAD activation profile. HDAC3 loses most of its activity when exposed to high concentrations of NaCl. The activity is fully restored when exogenous IP$_4$ is added back to the complex diluted in low-salt buffer. B: HDAC1/MTA1 (NuRD) activation profile.

HDAC activity assays were therefore set up to understand whether the Sin3a complex can also be activated by IP$_4$ despite the lack of a SANT domain in any of
its subunits.

### 5.2.1 Binary and ternary complex activity assays

For these experiments, the binary complex was purified as described in paragraph 4.2.2, while the ternary complex was purified as described in paragraph 4.3.2 (+RNase A) and passed through a size exclusion chromatography column in 150 mM K-acetate buffer. The purified protein samples were split into two vials. One vial was stored at 4°C, while the other one was diluted in an excess volume of stripping buffer (containing 1 M NaCl) and kept at room temperature for 4 hours. After this incubation, the sample was transferred into a dialysis membrane with a 10 kDa molecular weight cut off and dialysed against 1 L of HDAC assay buffer (containing 50 mM NaCl) overnight at 4°C. The sample was then re-concentrated and BOC-Lysine HDAC activity assays were carried out as described in paragraph 2.6.2. The aim of these experiments was to compare the catalytic responses to IP₄ of the protein complex before and after being treated in high NaCl. As visible from Figure 5.2, both the binary and ternary complexes tested at an approximate concentration of 150 nM did not seem to respond to IP₄. Indeed, the catalytic activity was not increased by IP₄ either before or after the high salt treatment. An accurate assessment of the variation of activity before and after the high salt treatment was not possible due to the technical difficulty of re-calibrating the amount of protein used in the assays. In fact, the difference in activity observable between the “Pre-” and “Post-NaCl” samples shown Figure 5.2A is significantly dissimilar in Figure 5.2B and in Figure 5.2C.
To ensure that the assay would not be saturated by an excessively high catalytic activity, the protein complex was tested at a final concentration of 25 nM, side-by-side with the SMRT complex at the same concentration as a positive control. Lowering the amount and concentration of protein used in the assay would also increase the IP₄/complex ratio, possibly increasing the probability of stimulating the activity of Sin3a. Nevertheless, the results of this experiment (Figure 5.3) were similar to those shown in Figure 5.2 (no visible change in activity upon addition of exogenous IP₄) and, in fact, represented a further confirmation of the
hypothesis that Sin3a is not activated by IP$_4$. In line with the findings reported by Watson et al. (2012b) and Millard et al. (2013), the HDAC3/SMRT complex showed a 19-fold activation upon addition of exogenous IP$_4$. On the contrary, Sin3a activity could not be stimulated.

![Figure 5.3: SMRT and Sin3a: direct comparison of IP$_4$-induced activation.](image)

**Figure 5.3:** SMRT and Sin3a: direct comparison of IP$_4$-induced activation. Sin3a (here expressed as ternary complex) and SMRT were tested side by side and their activation profile against IP$_4$. Proteins were tested at a concentration of 25 nM. Blank reading values were subtracted from the experimental data.

Experiments were carried out to assess whether the activity of Sin3a could be stimulated by incubating the complex with IP$_4$ under different temperatures and for different times. In a set of experiments we pre-incubated the ternary complex with IP$_4$ at 0°C, room temperature and 37°C for 30 minutes, while a second experiment was designed to incubate the ternary complex with IP$_4$ at 37°C for 15 and 30 minutes. Only after these pre-incubation steps the substrate was added to
the reactions and the HDAC activity measured. These experiments (Figure 5.4) showed that IP₄ was not able to activate Sin3a even when pre-incubated with the complex at different temperatures and for different periods of time. Nevertheless, it was evident that low temperatures could preserve the activity of the complex compared to higher temperatures. At the same temperature conditions, incubating the complex for longer times before adding the substrate caused a partial loss of activity. It is possible to argue that these results were caused by the dissociation rate of the complex subunits in solution being larger than their association rate. This was probably due to the low protein concentrations (ranging from 50 to 25 nM) used in the pre-incubations with IP₄. Indeed, at such concentrations, dissociated subunits would have a lower chance to re-encounter their interaction partners in solution, causing the complex to fall apart and consequently losing activity. This accounts for the fact that at lower temperatures the activity was preserved, as the complex subunits dissociate at lower rates. Similarly, longer pre-incubation times resulted in loss of catalytic activity as larger proportions of the complex in solution would dissociate prior adding the substrate to the assay.
Figure 5.4: Effect of pre-incubating the ternary complex with IP₄ at different temperatures and for different amounts of time. A: Temperature experiment. The “0ºC” sample was pre-incubated on ice for 30 mins; the “room temperature” samples were incubated at 20ºC for 30 mins; the 37ºC samples were incubated on a heat block set at 37ºC for 30 mins. The protein was used at a 25 nM concentration B: Time experiment. All the samples were incubated on a heat block set at 37ºC for the times indicated in the figure. The protein was used at a concentration of 50 nM. All samples were transferred to a black 96 well plate and the assay was carried out as described in paragraph 2.6.2 of chapter 2. Blank reading values were subtracted from the experimental data.

The experiments shown in Figures 5.2, 5.3 and 5.4 are consistent with the hypothesis that Sin3a may not be regulated by IP₄. Therefore, we asked ourselves how the Sin3a activity shown in Figure 5.2 compared to the basal activities of the other HDAC-dependent transcriptional co-repressor complexes. Indeed, if the activity of Sin3a resulted to be significantly higher than the basal activity of the other complexes, or even comparable to their activities upon IP₄ stimulation, it is possible that Sin3a might not require further activation by inositol phosphates. On the contrary, if the basal activity of Sin3a were comparable to activities of other HDAC1/2 complexes, it would probably indicate that the complex does not require further activation.

An experiment was therefore designed to verify these hypotheses. While the full ternary Sin3a complex and the HDAC3/DAD complexes were expressed and
purified by myself, the purified NuRD, MiDAC and CoREST complexes were kindly provided by Dr. Christopher Millard, Mr Charles Milano and Ms Yun Song respectively. In order to calibrate the amount of HDAC in each complex for use in the assay, the proteins were run on an SDS-PAGE gel and the relative intensities of HDACs were measured with a BioRad Gel Doc XR system and adjusted so that the concentration of the proteins was equal (Figure 5.5 A). A second round of normalisation was then carried out with the same method and the different intensities among the HDACs were used to recalibrate the activity of each complex in the HDAC assay (Figure 5.5 B). As visible from Figure 5.5 C, the baseline activities of SMRT, MiDAC and CoREST were similar to the activity shown by the Sin3a ternary complex, giving credit to the hypothesis that Sin3a could interact with a small “activator molecule”, but possibly not IP₄. Unfortunately, due to heavy protein degradation it was not possible to compare the basal activity of the NuRD complex to that of the other complexes.
5.3 Sin3a-IP$_5$ binding assays

The HDAC activity assays performed with the Sin3A complex indicated that inositol phosphate does not stimulate its activity. However, it remained formally possible that IP might bind the Sin3A complex and therefore the affinity of the complex for inositol phosphates was measured through kinetic assays.

As shown in Figure 5.6 (adapted from Watson et al. 2012b) IP$_4$ interacts with HDAC3 and SMRT thanks to the 4 contiguous phosphate groups of the inositol
molecule (at positions 1,4,5,6). Since IP₄ is a very expensive compound, IP₅ - which bears the same 4 contiguous phosphates as IP₄ in addition to one extra phosphate located at position 2 - was used for fluorescent polarisation and Bio-layer interferometry assays reported in the following paragraphs.

**Figure 5.6:** Molecular details of the IP₄ interaction with the HDAC3/DAD complex (adapted from Watson et al. 2012b). **A:** IP₄ in the HDAC3/DAD complex. HDAC3 is represented in blue, while the DAD domain of SMRT is represented in grey. Phosphate groups belonging to IP₄ are labelled in red (P₁, P₆, P₅, P₄). **B:** Electron density and backbone of IP₄. Phosphate groups are labelled in red (P₁, P₆, P₅, P₄).

### 5.3.1 Fluorescence polarisation (FP) assay of IP₅ binding to the Sin3a ternary complex

Fluorescence polarisation is a technique that can provide a direct measurement of the binding of small fluorescent molecules to larger biological molecules such as proteins or nucleic acids. It is based on the principle that small molecules in solution tumble faster than larger molecules. Therefore, when a fluorophore is attached to the ligand of interest, it will tumble more quickly when the ligand is free in solution than when it is bound to a larger molecule such as a protein.
Slow-rotating molecules result in an increased emission of polarised light, therefore a strong binding between the ligand and the large molecule results in a higher amount of polarised light.

Flourescein-labelled IP$_5$ (a kind gift from Professor Barry Potter of the University of Bath) was used to measure the affinity of the Sin3a complex to the inositol phosphate. The experiment was carried out as described in paragraph 2.8.2 of chapter 2. The HDAC3/DAD complex was expressed and purified as described by Watson et al. (2012b) and used as a positive control. The Sin3a ternary complex was purified as described in paragraph 4.3.2 of chapter 4 (+RNase A) and eluted from a size exclusion chromatography column in 150 mM K-acetate gel filtration buffer.

IP$_5$ appears to weakly bind the Sin3a ternary complex (Figure 5.7), while it forms a tight interaction with SMRT. Moreover, the weak interaction with Sin3a could only be detected after 10 mins (Figure 5.7 B) and increased slightly after 20 and 30 mins. This result is consistent with the outcome of the HDAC activity assays described in paragraph 5.2.1, suggesting that IP$_5$ does not stimulate catalytic activity of Sin3A. Nevertheless, a similar molecule with an alternative structure could bind the complex with a higher affinity.
5.3.2 Bio-layer interferometry assay of IP₅ binding to the Sin3a ternary complex

The Bio-layer interferometry assay (OCTET assay) was used to investigate a potential association between Sin3a and IP₅. The results were in agreement with the previous data shown in paragraphs 4.2.1 and 4.3.1. This assay is based on the interference pattern of white light that is reflected from a reference layer on an OCTET biosensor tip and a layer of ligand immobilized onto the same tip. When the tip is immersed in a solution containing the protein of interest, any protein molecule binding to the immobilized ligand causes a shift in the wavelength reflected by the ligand layer itself that is proportional to the amount of protein...
Biotinylated IP$_5$ was a kind gift from Professor Barry Potter (University of Bath) and was immobilized on the OCTET biosensor tips by following the manufacturer's instructions. The binding kinetics of Sin3a with IP$_5$ were measured as described in paragraph 2.8.3 of chapter 2 with a FortèBio Octet$^{\text{®}}$QK$^{\text{e}}$ system. The Sin3a ternary complex was purified as described in paragraph 4.3.2 of chapter 4 (+RNase A) and further purified by size exclusion chromatography in 150 mM K-acetate gel filtration buffer.

The results, reported in Figure 5.8, were consistent with the HDAC catalytic assays and the FP experiments as only a very weak binding with IP$_5$ could be observed even by the bio-layer interferometry assay. The tightest binding for the Sin3a complex was observed at a protein concentration of 1 µM, which is approximately 8 times lower than the binding observed for the MiDAC complex diluted at a concentration of 25 nM (information kindly communicated by Charles Milano).
Figure 5.8: *Bio-layer interferometry assay*. The binding of Sin3a with IP$_5$ was measured at different protein concentrations, which are reported on the right side of the chart. The part of the curve to the left side of the red line at 1200 mins shows the association curves of Sin3a with IP$_5$, while the curve on the right of the line show the rate of dissociation. Binding is determined by measuring the wavelength shift of white light reflected by the bio-layer upon binding of Sin3a to IP$_5$.

5.4 Results summary and conclusions

The experiments reported in this chapter aimed at investigating the possibility that Sin3a is regulated by IP$_4$. The results, however, strongly suggested that this would not be the case. Indeed, HDAC catalytic assays did not show any activity stimulation of the complex upon addition of exogenous IP$_4$. Similarly, FP and bio-layer interferometry assays have shown that Sin3a does not form a strong interaction with IP$_5$, which accounts for the fact that IP$_4$ is not an activator of the complex due to lack of binding. These results are supported by the absence of a SANT domain in the Sin3a complex. However, they open to new hypotheses on the regulation of the activity of Sin3a complex, including the possibility that different inositol phosphate molecules might regulate Sin3a. These hypotheses will be discussed in chapter 6.
CHAPTER 6: Discussion

6.1 HDAC1 and its interaction within the HID domain of Sin3a

6.1.1 HDAC1/HID/SDS3 interaction mapping experiments

Since Millard et al. (2013) have solved the structure of HDAC1 in complex with the ELM2-SANT domain of MTA1, the idea that class I HDACs interact with their co-repressors through multiple surfaces is becoming increasingly more accepted. Indeed, Millard et al. have reported that HDAC1 interacts with the NuRD complex through the SANT domain and the ELM2 region of MTA1, which encompasses both a dimeric interface and an extended N-terminal region that completely wraps around the enzyme (as represented in Figures 1.8 and 1.9). Therefore, mapping experiments (Chapter 3) ought to find a “minimal” interacting region between HDAC1 and the HID, and confirmed that this is also likely to be the case in the Sin3a complex. Indeed, the minimal interacting region of Sin3a could not be identified and, on the contrary, a large number of constructs bearing different deletions seemed to be able to interact with HDAC1 (as summarized in Figure 3.7). This suggested the presence of multiple interaction motifs similarly to the HDAC1/MTA1 structure solved by Millard et al. (2013). The full length HID (531-855) and the HID lacking the last 78 amino acid residues (HID 531-772) appeared to be the two strongest Sin3a interacting partners of HDAC1 (Figures 3.4 and 3.7). The fact that the C-terminal region (HID 772-855) did not seem to be essential for the interaction with the enzyme indicated that the critical interacting surfaces of the HID were either located in the central region of the
domain or towards its N-terminus. HID 531-654, which compared to HID 531-772 lacked a much larger portion of the C-terminal region of the HID, also seemed to be able to interact with HDAC1, while a very similar construct (HID 524-680) could not bind to the enzyme, perhaps due to the misfolding of the construct in solution (Figures 3.5 and 3.7). HID 681-855, however, did form an interaction with HDAC1. These results lead to the conclusion that, possibly, the central portion of the HID may be important for the interaction between Sin3a and HDAC1. This region, which we identified as HID 651-772, bears 4-5 predicted alpha helices that may have the capacity to form a tertiary structure somewhat resembling a SANT domain. However, it should be noted that there is no evidence of homology between the SANT and the HID domains. Short constructs (<100 aa) bearing this region of the HID interacted with HDAC1, although their expression levels were low.

In light of what Millard et al. (2013) have observed with the HDAC1/MTA1 and HDAC3/SMRT complexes as well as the above-discussed data, it is possible to hypothesize that the central region of the HID may form a folded structure that interacts with HDAC1 and that extends (either from its N-terminus or C-terminus) to wrap around the enzyme (as the MTA1 structure shown in Figure 6.1).
The deletion of these potentially extended regions (for example in HID 531-654 and HID 651-772) caused a dramatic loss of expression, indicating that they may be involved in the correct folding of the HID other than simply having a role in interacting with HDAC1 (Figures 3.4, 3.6 and 3.7). Nevertheless, we now know that the role of the SANT domain in the HDAC1/MTA1 and HDAC3/SMRT structures is to interact with the HDACs through the interposition of a molecule of IP₄ (Watson et al. 2012b; Millard et al. 2013), therefore, the question on whether the central helical bundle of the HID may play a similar role in the Sin3a complex arises and remains to be elucidated.

Mapping experiments have pointed out that HDAC1 may be interacting with Sin3a thanks to the mediation of the SDS3 protein, a core component of the Sin3a complex. The full length HID (531-855) was found to be able to interact with SDS3 in an HDAC-independent manner. Truncating the HID in the central region (HID 654-680), however, caused this interaction to become HDAC1-dependent (Figures 3.9 and 3.10). These findings suggested that SDS3 could independently interact with Sin3a through to the central region of the HID and
made contact with HDAC1, which rescued the interaction when this region was compromised. This could be evidence that the central helical bundle of the HID is indispensible for the interaction of SDS3. The possibility that SDS3 makes contact with both HDAC1 and the HID simultaneously reflects the observations of Lechner et al. (2000), who found that SDS3 is essential for the HDAC activity of the Sin3/Rpd3 complex in yeast.

6.1.2 Protein complex purifications and crystallization trials

Mapping experiments (chapter 3) were further validated from the complex purification experiments reported in chapter 4. Even though a binary complex consisting of only HDAC1 and the HID could be expressed and purified, the addition of SDS3 significantly improved the expression of the complex, confirming the stabilizing effect of SDS3 for Sin3a. The presence of large amounts of RNA contaminants was a consistent observation with all of the purifications of the ternary complex (chapter 4). Because large amounts of RNA could only be detected when also SDS3 was co-expressed, it was initially hypothesized that small RNA molecules could associate with SDS3 and possibly even be part of a core complex and/or possess a regulatory role. However, later experiments showed that even upon RNase A digestion of the complex, small nucleic acid molecules were still non-specifically interacting with the complex and could simply be washed away with more stringent salt conditions (Figure 4.21). In fact, this was consistent with the observations of Graur et al. (2013), who reported that the activity of Pol II can be up-regulated even up to 1000-fold in transformed cells, and that up to 90% of Pol
II transcripts may be generated as a result of transcriptional noise and possess no particular function. The reason why high levels of RNA could only be detected with the expression of the ternary complex could then be attributed to a particular shape or electric environment assumed by the complex in solution. Crystallization trials with the full ternary complex solely resulted in the formation of salt crystals of no use (Figure 4.18). One of the plausible reasons of this outcome might have been the high level of protein degradation visible after the complex had been set in the crystallization trials for 3 weeks (Figure 4.19). The degree of degradation observed, in turn, could have been caused by the fact that the HID might not have been fully folded and was therefore highly susceptible to proteolytic attack. Indeed, as previously mentioned, it was not possible to find a minimal interacting region between HDAC1 and the HID, implying that portions of HID 531-772 may not be involved in the interaction and be unstructured in solution. Another possibility, however, was that the coiled coil domain of SDS3 promoted the dimerization of the complex (as suggested by Alland et al. 2002) and was either changing the conformation of the complex into a shape that was not suitable for forming crystals or it could have been sticking out of the complex bulk, hindering crystal formation. It was for this reason that an SDS3 construct lacking the coiled coil domain was designed and employed for crystallization trials. Crystals generated with this truncated form of SDS3 failed to generate an X-ray diffraction (Figure 4.31). Nevertheless, the absence of this diffraction could be indicating that the crystals were actually made of protein complex but had an internal structure that was too heterogeneous to diffract X-rays. Even though this was a significant progress in the process of obtaining crystals of the Sin3a complex, it is
evident that still a lot of optimization around the protein sample may be required. Finally, gel filtrating the full ternary complex in a gel filtration column with a softer matrix, helped purifying the complex from nucleic acid impurities (Figure 4.21) but did not result in the formation of protein crystals, strongly suggesting that the full ternary complex may not be suitable for X-ray crystallography studies. A common characteristic of all of the gel filtrations of the ternary complex was the typical shape of the chromatograms, characterized by a steep increase in absorbance and a rather slowly-descending absorbance towards the end (Figures 4.9, 4.9bis, 4.11, 4.13 and 4.28). This may be indicative of a heterogeneous protein sample less than optimal for crystallization trials. Potentially, this could have been caused by three factors: protein degradation, the dissociation and association rates of the complex subunits and the three-dimensional fold of the complex in solution. Although protein degradation could be observed in some protein samples, the addition of extra-protease inhibitors has tackled this problem, which could therefore be excluded. On the contrary, levels of SDS3 seemed to decrease along the gel filtration fractions (Figures 4.9bis and 4.28), indicating that there could be co-existing species of the complex in solution bearing different amounts of bound SDS3. Finally, large regions of the HID may remain unfolded in solution due to lack of a specific interacting partner. These long and unstructured “arms” could stick onto the beads of the gel filtration matrix causing the characteristic chromatograms described above. Future experiments may address the last two scenarios in at least two possible ways. The first approach could be continuing the mapping experiments in greater detail, as an attempt to find a region of the HID that lacks the fragments dispensable for
the interaction with HDAC1 and SDS3 but yet expresses robustly and forms a strong interaction with the two subunits. However, this approach may be time consuming and finding a well-expressing region of the HID that lacks well conserved regions of the protein might not be possible. Indeed, because the HID domain is highly conserved, it is difficult to use sequence conservation to guide the design of domain boundaries. A second approach could be trying to further stabilize the HID by the addition of extra core subunits to the ternary complex. Probably, the best candidates would be SAP30 and/or SAP18. SAP30 is a 30 kDa protein that is mainly associated with the PAH3 domain of Sin3a (Xie et al. 2011). As reviewed by Grzenda et al. (2009), it forms interactions with a series of transcription factors, targeting Sin3a to chromatin substrates. Interestingly, Zhang et al. (1998) have reported that HDAC1 may form a direct interaction with SAP30. Therefore, it is possible to hypothesize that the addition of this protein to the complex may further stabilize it, increasing the possibilities of generating protein crystals. However, to include this protein in the complex, the presence of PAH3 may be required. SAP18, on the other hand, is an 18 kDa protein that also associates with Sin3a and HDAC1 (as reviewed by Grzenda et al. 2009). Zhang et al. (1997) found that it can stabilize HDAC1 enhancing its catalytic activity, although the exact boundaries of the interaction with Sin3a remain unknown. SAP30 and SAP18 could be added separately to the ternary complex to understand if they can stabilize the HID and improve the chances of generating protein crystals. However, a further option could be adding both SAP30 and SAP18 to the ternary complex. Indeed, Portolano et al. (in press), report that it is possible to successfully co-transfect up to 5 plasmids in HEK 293F cells. Other
associated proteins such as RBBP4, RBBP7, SAP130 and SAP180 may be tested for their ability to stabilize the complex. However, the large dimensions of SAP130 and SAP180 would require further interaction-mapping experiments, while RBBP4 and RBBP7 are not directly associated with Sin3a but interact with the complex through HDAC1, HDAC2 and SAP30 (Hassig and Schreiber 1997; Zhang et al. 1999; Zhang et al. 1998).

6.2 Sin3a: IP$_4$-dependent or independent?

Bioinformatics analyses of Sin3a have revealed no apparent homology between the HID and the SANT domains of other transcriptional co-repressor complexes such as SMRT, CoREST, NuRD and MiDAC. This results in two major implications, the first one being that it is difficult to predict the way Sin3a recruits HDAC1 and HDAC2 into the complex, and the second one being that it is not known whether Sin3a can bind IP$_4$ and regulate the catalytic activity of HDAC1 and HDAC2. Catalytic and kinetic experiments were designed and are reported in chapter 5 to get an insight of this issue. HDAC assays carried out with and without the presence of IP$_4$ seemed to confirm the fact that inositol phosphates are not involved in the regulation of the Sin3a complex, consistent with the absence of a predicted SANT domain (Figures 5.2, 5.3 and 5.4). Similar outcomes were observed with the kinetic assays (reported in Figures 5.7 and 5.8), which further validated the results of the HDAC catalytic assays, suggesting that inositol phosphates do not bind to Sin3a. Moreover, HDAC catalytic assays (Figure 5.4) have shown that the baseline activity of Sin3a is very similar to the baseline activity of the CoREST and MiDAC complexes, suggesting that there is
space for further activation of the complex. These outcomes raise questions on how the catalytic activity of the Sin3a is regulated.

One hypothesis could be that Sin3a is purely independent from IP$_4$ and simply activates HDAC1 and HDAC2 through their interaction with the HID or other subunits of the complex. Indeed, as already mentioned in the introduction, Sin3a is a very ancient transcriptional co-repressor complex and it is very well conserved from yeast to humans (Knoepfler and Eisenman 1999). This could imply that because of the raising complexity of eukaryotic cells, evolution may have found more selective ways to regulate transcription. Thus, other transcriptional co-repressor complexes may have originated to be regulated in more specific ways (for example through IP$_4$). In this scenario, it would be easy to imagine that Sin3a did not require to be specifically regulated by inositol phosphates in ancient organisms. However, it remains unknown whether other subunits of the complex have a similar role to that of IP$_4$, recruiting and activating HDACs 1 and 2 in a similar fashion. Both published scientific literature and the data shown in this thesis (chapters 3 and 4) point out that SDS3 has a critical role for the deacetylase activity of Sin3a. However, while SDS3 seems to be largely responsible for the interaction of HDAC1 with the HID, IP$_4$ is dispensable for the interaction of HDAC1 and HDAC3 with their cognate co-repressors (Millard et al. 2013; Watson et al. 2012b). This could imply that SDS3 might have the mere role of recruiting HDAC1 to the complex to stabilize it, which is consistent with the fact that the binary complex is catalytically active (Figure 5.2 A). In contrast, protein subunits such as SAP30 and SAP18 (dispensable for the interaction with HDAC1 with Sin3a) might potentially be able to tune the catalytic activity of the Sin3a
complex. These speculations, however, should be investigated through experiments of complex purifications and HDAC assays. In particular, future researchers may look at the difference of the catalytic activities of the Sin3a complex with and without SAP30 and SAP18. Also, it would be interesting to carefully compare the baseline activities of normalized binary and ternary complexes to rule out the possibility that SDS3 actively increases the HDAC activity of Sin3a. Another hypothesis is that Sin3a may be regulated by a co-factor different from inositol phosphates. Indeed, while the results of chapter 5 clearly point out that the Sin3a complex is IP$_4$ independent, mapping experiments reported in chapter 3 suggest that the structure of the interaction between HDAC1 and the HID may resemble, at least in part, the HDAC1/MTA1 and HDAC3/SMRT interactions. Indeed, as already mentioned, HDAC1 and HDAC3 interact through the SANT helical bundle that extends into a peptide that braces the enzymes stabilizing them (Millard et al. 2013). Since the SANT domains can bind IP$_4$, it could be hypothesized that the central helical bundle of the HID (helices 3-7) may form a structure able to bind a co-factor that can regulate the activity of the complex. Nevertheless, it should be noted that while MTA1 and SMRT are able to form a strong interaction with the enzymes independently on the presence of other subunits, Sin3a requires SDS3 to form a stable complex with HDAC1. Therefore, the interaction and the regulation of HDAC1 in the Sin3a complex may be unique among other transcriptional co-repressor complexes, implying that it is not possible to exclude that a co-factor may be interposed between HDAC1 and a different subunit. It still remains possible to hypothesize that the Sin3a complex is in fact regulated by IP$_4$. However, this may be
interposed between the surface of HDAC1 and another subunit of the Sin3a complex that is yet to be discovered. Indeed, none of the currently known Sin3a subunits show sequence similarity to a SANT domain, making most likely that IP₄ is not be able to interact with any of them.

Further experiments will help understand the regulation mechanism of Sin3a may add to the knowledge of how Class I HDAC-containing complexes recruit their catalytic subunits in function of the specific needs of the cell. Indeed, it would be interesting to understand how the amount of HDAC activity of each transcriptional co-repressor complex affects their function in the cell. Interestingly, these complexes, canonically known as transcriptional co-repressors, have also been found associated with active genes (as reviewed by Dovey et al. 2010a), suggesting that their function may be more complicated than we previously thought and that the regulation of their enzymatic activity may have more profound implications.

6.3 Summary and conclusions: transcriptional co-repressor complexes, chromatin and the regulation of life.

In this thesis I have discussed how DNA regulates its own activity through a series of intricate mechanisms that are studied in the field of epigenetics. It was described how particular proteins, enzymes and protein complexes can work in synergy to pack DNA into a structure known as chromatin; and how this can actively change its conformation in response to the needs of the cell or even to protect us from dangerous sequences spuriously acquired during the course of evolution. Chromatin modifying enzymes include some that add and/or remove
methyl groups from the DNA itself in order to regulate transcription (DNA methyl transferases and demethylases). Other enzymes reposition nucleosomes within the genome in an ATP-dependent manner (chromatin remodeling complexes), while yet another class of enzymes post-translationally modifies histone tails to condense or relax chromatin, indirectly regulating the expression of genes. Among all of these enzymes, this thesis has focused on histone deacetylases and their interacting partners. These form large co-repressor complexes that remove the acetyl group of histone-acetylated lysines to condense chromatin and inhibit transcription. Experiments were carried out to understand how the Sin3a complex recruits HDAC1 to deacetylate chromatin targets. Although the crystal structure of the complex could not be solved, it was still possible to get a deeper insight to Sin3a and the results of these experiments have been reported in chapters 3, 4 and 5 and discussed in the first two paragraphs of this chapter. However, a number of questions still remain unanswered and further experiments are required to elucidate yet many aspects of the Sin3a complex (as discussed in paragraphs 6.1 and 6.2 of this chapter).

One of the general aspects concerning Class I containing transcriptional co-repressor complexes, that this thesis has not yet covered, is the fact that while these complexes are canonically associated with transcriptional repression, recent findings have shown that they can actually possess a dual function as transcriptional repressors as well as transcriptional activators (as reviewed by Reynolds et al. 2013). It is becoming widely accepted that transcriptional regulation does not always function as a simple on-off switch, but the activities of transcriptional co-repressor complexes are finely tuned to tightly regulate the
transcription of genes. The way transcriptional co-repressor complexes accomplish this still remains largely unexplored. However, Watson et al. (2012b) and Millard et al. (2013) discovery of IP₄ involvement in the regulation of transcription may potentially open new horizons in this field.

Following Aristotle’s hypothetical question on “What controls the DNA?” the next question to be answered will be “What controls the regulators of DNA?” or perhaps more appropriately, “How does DNA regulate its own regulators?” This is, in my opinion, the most important question that science will have to answer in the future in order to make a significant progress in the understanding of diseases and the regulation of life in general.


He, Y. and Radhakrishnan, I. 2008. Solution NMR studies of apo-mSin3A and -mSin3B reveal that the PAH1 and PAH2 domains are structurally independent. Protein Science 17(1), pp. 171–175.


