Structural, functional and inhibition studies of human histone deacetylase 7

Danielle Claude Desravines

UNIVERSITÉ DE GRENOBLE

EMBL

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité Biologie Structurale et Nanobiologie

Arrêté ministériel : 7 août 2006

Présentée et soutenue publiquement par

DESRAVINES Danielle Claude

le **13 décembre 2010**

Etudes Structurales, Fonctionnelles et d'Inhibition de l'Histone Déacétylase 7 Humaine

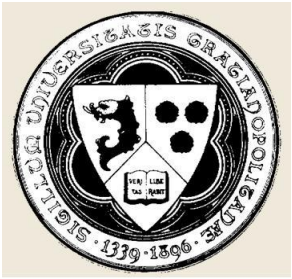
---

Thèse dirigée par **Dr HART Darren**

JURY

Dr Khochbin Saadi	**IAB**	**Président**
Dr Haser Richard	**IBCP**	**Rapporteur**
Dr Bottomley Matthew	**Novartis Vaccines and Diagnostics**	**Rapporteur**
Dr Longhi Sonia	**AFMB**	**Examinateur**
Dr Pillai Ramesh	**EMBL**	**Examinateur**

Thèse préparée au sein de l'EMBL Grenoble dans **l'Ecole Doctorale Chimie et Sciences du Vivant**



UNIVERSITÉ DE GRENOBLE

EMBL

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité Biologie Structurale et Nanobiologie

Arrêté ministériel : 7 août 2006

Présentée et soutenue publiquement par

DESRAVINES Danielle Claude

le **13 décembre 2010**

Analyses structurales, fonctionnelles et d'inhibition de l'histone déacétylase 7 humaine.

Thèse dirigée par **Dr HART Darren**

JURY

Dr Khochbin Saadi	**IAB**	**Président**
Dr Haser Richard	**IBCP**	**Rapporteur**
Dr Bottomley Matthew	**Novartis Vaccines and Diagnostics**	**Rapporteur**
Dr Longhi Sonia	**AFMB**	**Examinateur**
Dr Pillai Ramesh	**EMBL**	**Examinateur**

Thèse préparée au sein de l'EMBL Grenoble dans **l'Ecole Doctorale Chimie et Sciences du Vivant**

THESIS

To obtain the title of

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

In Structural Biology and Nanobiology

Arrêté ministériel : 7 août 2006

Defended in public by

Danielle Claude DESRAVINES

the **13th of december 2010**

Structural, Functional and Inhibition Studies of Human Histone Deacetylase 7.

Thesis director **Dr Darren HART**

JURY

Dr Khochbin Saadi	**IAB**	**Chair**
Dr Haser Richard	**IBCP**	**Reviewer**
Dr Bottomley Matthew	**Novartis Vaccines and Diagnostics**	**Reviewer**
Dr Longhi Sonia	**AFMB**	**Examiner**
Dr Pillai Ramesh	**EMBL**	**Examiner**

Thesis prepared within EMBL Grenoble in the **Ecole Doctorale Chimie et Sciences du Vivant**

« Voici à quoi je pense chaque jour fort souvent: ma vie extérieure et intérieure dépend du travail de mes contemporains et de celui de mes ancêtres et je dois m'efforcer de leur fournir la même proportion de ce que j'ai reçu et que je reçois encore » de Albert Einstein

« Si un bureau en désordre dénote un esprit brouillon, que dire d'un bureau vide ? » *de Albert Einstein*

"A hundred times everyday I remind myself that my inner and outer life depend upon the labors of other men [and women], living and dead, and that I must exert myself in order to give in the measure as I have received and I am still receiving." from Albert Einstein

"If a cluttered desk is a sign of a cluttered mind, then what are we to think of an empty desk?" from Albert Einstein

ACKNOWLEDGEMENTS

I would like to thank Darren Hart for allowing me to work in his lab over four years on such an interesting and exciting project and for his guidance.

I would like to thank the European Community who funded my project (LSHM-CT-2006-037498).

I thank the members of my Thesis Advisory Committee, Ramesh Pillai, Klaus Scheffzek and Thierry Vernet for their advice and guidance throughout my Ph.D. project.

I am grateful and thank the members of my thesis jury Saadi Khochbin, Richard Haser, Matthew Bottomley, Sonia Longhi and Ramesh Pillai for devoting part of their time to participate in my thesis defense.

I would like to thank people who participated in my project: Alessandro Angelini for helping with the early protein engineering phase, Nataliia Aleksandrova for performing fluorescence anisotropy and surface plasmon resonance analyses, Philippe Mas and Eva Maria Geenen who helped me with SMRT library generation and NMR experiments. I thank as well all the collaborators who contributed to my project, Malene Ringkjobing Jensen, Robert Schneider and Martin Blackledge for NMR experiments, Mats Ökvist, Sean McSweeney, Gordon Leonard for providing me advices for structural analyses. Finally I thank Frédérik Oger, Gérard Simmonneaux, Gilles Salbert and Daniele Carettoni for SAHA-derivatives design. Thank you all, without your help my project would not have progressed so much.

I would like of course to thank all the past and present members of my laboratory, Franck Tarendeau, Philippe Mas, Alban Chesneau, Patrick Meresse, Hayretin Yumerefendi, Stéphane Boivin, Bertrand Simon, Yingfeng An, Eva Maria Geenen "La Petite", Damien Maurin, Rasa Sukackaite and Denis Bouvier, for all the nice scientific discussions and pleasant times we shared.

I thank the members of the EMBL International PhD Program, especially Helke Hillebrand and Milanka Stojkovic, for their support for allowing EMBL PhD students including myself to complete their work in the best conditions.

I thank as well all my friends, especially Hervé, Maud, Karla, Julianna, Marie, Sarah, Thibaud who always supported me in the best as well as in worst times of my Ph.D.

Last but not least I would like to thank the members of my family, my mother Monique and my father Jean-Vincent who taught me patience, perseverance, humility, open-mindedness and optimism, for their permanent encouragements and love. I thank as well my brother for his refreshing ideas and views of life which helped me to "breathe fresh air". Finally I thank my little sister "Cathy" who was always there for me especially in the last days of thesis manuscript submission, she was of a great support along with Karla. Thank to you all, my success is as well yours.

Thanks to all of you, and my apologies to anyone I might have forgotten, I thank you as well.

ABSTRACT

Cardiovascular diseases, in particular atherosclerosis, are a worldwide burden. Atherosclerosis occurs due to accumulation of cholesterol in the arteries leading to rigid plaque formation; this can eventually lead to strokes and heart attacks. Several efficient drugs exist to decrease blood cholesterol levels, but they are inefficient for treating familial hypercholesterolemia. Therefore new drugs must be developed. It was observed previously that inhibiting HDAC7 increased conversion of cholesterol into excreted bile acids by enhancing the activity of 7α-cholesterol hydroxylase, the rate-limiting enzyme in cholesterol catabolism. This represents a potential new strategy for decreasing cholesterol levels. However HDAC7 shares high homology with ten other human HDACs and this makes designing selective HDAC7 inhibitors challenging due to highly similar active sites. To elucidate mechanisms underlying HDAC inhibitor isoform-specificity, crystal structures are needed. After several protein engineering steps, using the ESPRIT technology, a well-behaving HDAC7 form was designed. Although not yet crystallised, it was relatively easy to purify, highly active and allowed characterisation of a new inhibitor synthesised by collaborators. This compound was derived from a pan-HDAC inhibitor, SAHA, with methyl groups added in meta positions to its cap moiety. Kinetic, biophysical and cell-based experiments revealed that this compound was more potent than SAHA, and more isoform specific, probably due to extra interactions at the rim of the HDAC7 catalytic site. A second completely new strategy to inhibit HDAC7 function is to disrupt its binding to the SMRT corepressor complex in which is active *in vivo*. However, SMRT being a 2525 amino acid intrinsically disordered protein is difficult to approach structurally. We therefore focused on the 500 amino acid repression domain III of SMRT, mapped previously as the interaction region with HDAC7. Protein engineering using ESPRIT generated 51 well expressed soluble SMRT constructs ranging from 15 to 34 kDa and spanning almost the entire region. Six of them were selected based upon their expression yields and were tested for interaction with HDAC7. One construct (69D20) was shown to interact with HDAC7 by ^{15}N-HSQC NMR. The binding was confirmed by complementary biophysical methods revealing an interaction of low micromolar affinity. The interaction region is being characterised in detail by NMR to assign the residues involved in binding, as first step towards developing small-molecules targeting this interface.

RESUME

Les maladies cardiovasculaires, en particulier, l'athérosclérose, constituent un problème majeur de santé publique. L'athérosclérose consiste en l'accumulation de cholestérol dans les artères causant la formation de plaques rigides ; ces plaques pouvent ensuite se rompre causant attaques et malaises cardiaques. Il existe de nombreux médicaments réduisant les taux de cholestérol dans le sang, ils sont cependant inefficaces pour le traitement de patients souffrant d'hypercholestérolémie familiale. Il est donc important de développer de nouvelles molécules. Il a été observé précédemment que l'inhibition de HDAC7 augmentait la conversion du cholestérol en acides biliaires en augmentant l'activité de la 7α-cholestérol hydroxylase, l'enzyme limitante du catabolisme du cholestérol. Ainsi, cibler HDAC7 pourrait constituer une nouvelle stratégie pour réduire les taux de cholestérol. Cependant HDAC7 présente une haute homologie de son site catalytique avec dix autres HDAC humaines rendant difficile la conception d'inhibiteurs spécifiques de HDAC7. Des informations structurales sont requises afin de pouvoir décrypter les mécanismes responsables de la spécificité d'inhibition en fonction des isoformes d'HDACs considérés. Après plusieurs étapes d'ingénierie de HDAC7,à l'aide, notamment, de la technologie ESPRIT, une construction optimale de HDAC7 a été conçue. Cette protéine a, jusqu'à maintenant, résisté à la cristallisation, mais est néanmoins purifiable, active et a permis de caractériser un nouvel inhibiteur synthétisé par des collaborateurs à Rennes. Ce nouvel inhibiteur a été synthétisé à partir d'un inhibiteur pan-HDACs, SAHA, substitué par des groupements méthyle ajoutés en positions méta du groupe fonctionnel phényle aussi appelé cap. Des expériences de cinétique, biophysiques et en cellules de mammifères ont été menées et ont révélé que ce nouveau composé présentait une inhibition de HDAC7 plus importante que SAHA. En outre ce nouveau composé était sans doute plus spécifique, probablement grâce à des interactions supplémentaires avec des résidus de l'entrée de la poche catalytique de HDAC7. Une deuxième stratégie innovante a été mise au point pour inhiber l'activité de HDAC7 en empêchant son interaction avec le complexe corépresseur contenant SMRT. Cependant, SMRT étant une protéine nativement désordonnée de 2525 acides aminés, il était difficile d'envisager des études structurales. Nous nous sommes donc concentrés sur les 500 acides aminés du domaine répresseur III de SMRT, identifié préalablement comme la région d'interaction avec les HDACs de class Iia dont HDAC7. La technologie ESPRIT a également été appliquée à SMRT et a permis de générer 51 constructions surexprimées et solubles, de 15 a 35 kDa, se répartissant sur toute la région. Six d'entre elles ont été sélectionnées pour leur bonne expression et leurs rendements après purification. Leur interaction avec HDAC7 a été testée. Une construction (69D20) interagissait avec HDAC7 par ^{15}N-HSQC NMR. Cette interaction a été confirmée par des méthodes biophysiques complémentaires indiquant une interaction raisonnable de l'ordre du micromolaire. Actuellement la construction est soumise à des analyses de RMN approfondies afin de d'identifier les résidus impliqués dans l'interaction, dans la perspective de créer des molécules ciblant l'interface HDAC7-SMRT.

ABBREVIATIONS

AKR1D1: Aldo-Keto Reductase family 1, member D1

BA: Bile Acid

BA-CoA: Bile Acid-Coenzyme A

BACS: BA-CoA synthase

BAP: Biotin Acceptor Peptide

BAP: Biotin Acceptor Peptide

BARE: Bile Acid Responsive Element

BAT: BA-CoA:Amino acid N-acylTransferase

BCCP: Biotin Carboxyl Carrier Protein

BCCP: biotin carboxyl carrier protein

BCL6: B-cell lymphoma 6

BCoR: BCL6 CoRepressor

CA: Cholic Acid

CaMK : calcium/CalModulin-dependent protein Kinase

CAT: Chloramphenicol AcetylTransferase

CBHA: M-CarBoxycinnamic acid bisHydroxAmide

CBP: CREB Binding Protein

CDCA: ChenoDeoxyCholic Acid

CFTR : Cystic Fibrosis Transmembrane conductance Regulator

CREB: cAMP-response element-binding protein

CRM1: ChRomosome Maintenance region 1

CtBP: E1A C-terminal Binding Protein

CVD: Cardiovascular Disease

CYP27A1: Cytochrome P450, family 27, subfamily A, polypeptide 1 = sterol 27α-hydroxylase

CYP7A1 : Cytochrome P450, family 7, subfamily A, polypeptide 1 = cholesterol 7α-hydroxylase

CYP8B1: Cytochrome P450, family 8, subfamily B, polypeptide 1 = sterol 12α-hydroxylase

DAD: Deacetylase-Activating Domain

DAD: Deacetylase-Activating Domain

ESPRIT: Expression of Soluble Protein by Random Incremental Truncation

FDA : Food and Drug Administration

FTF: α-Fetoprotein Transcription Factor

FXR = BAR: Farnesoid X Receptor = Bile Acid Receptor

FXRE: Farnesoid X *Receptor* Responsive Element

GATA1: GATA-binding factor 1
GFP: Green Fluorescent Protein
GNAT: Gcn5-related N-terminal acetyltransferase
H3K4me: Monomethylation of Lysine 4 of histone 3
HAT: Histone Acetyl Transferase
HDAC: Histone DeACetylase
HDAC: Histone DeACetylase
HDLP : Histone Deacetylase-like Protein
HIF-1α: Hypoxia Inducible Factor 1α
HMG-CoA: HydroxyMethylGlutaryl-Coenzyme A
HNF4α : Hepatocyte Nuclear Factor 4α
HP1: Heterochromatin Protein 1
HSD3B7: Hydroxy-δ-5-Steroid Dehydrogenase, 3β- and steroid δ-isomerase 7
HSP90: Heat Shock Protein 90kDa
IDP: Intrinsically Disordered Protein
LDL: Low Density Lipoprotein
LDL-R: LDL-Receptor
LHR1: Liver Homologue Receptor 1
LSD1: Lysine Specific Demethylase 1
MAP2: Microtubule Associated Protein 2
MCA: 4-MethylCoumarin-7-Amide
MDM2: Murine Double Minute 2
MEF2: myocyte enhancer factor-2
MMP10: Matrix MetalloProteinase-10
MYST: MOZ, Ybf2/Sas3, Sas2, and Tip60
NAD: Nicotinamide Adenine Dinucleotide
NBS1: Nijmegen breakage syndrome 1
NCoR: Nuclear receptor CoRepressor
ND: Nuclear receptor Domain
NES: Nuclear Export Signal
NFκB: Nuclear Factor κB
NFκB: nuclear factor-kappa B
NLS: Nuclear Localisation Signal
p300: E1A binding protein p300
PB: PhenylButyrate
PDB: Protein Data Bank
PEG: PolyEthylene Glycol.
PGC1: PPARγ coactivator 1
PIC: PreInitiation Complex

PKC: Protein Linase C

PKD: Protein Kinase D

PPAR: *Peroxisome Proliferator-Activated Receptor*

pRB: RetinoBlastoma protein

PSI-BLAST: Position-Specific Iterative Basic Local Alignment Search Tool

PTEN: Phosphatase and TENsin homologue

RD: Repression Domain

Runx: Runt box

RXR: Retinoid X Receptor

SAC: SAHA Cap

SAHA: SuberoylAnilide Hydroxamic Acid

SANT: SWI3/ADA2/NCoR/TFIIIB-like

SAR : Structure Activity Relationship

SENP: **SEN**trin/SUMO-specific **P**rotease

SHP: Short Heterodimer Partner

SIRT: SIRTuin

SMRT: Silencing Mediator of Retinoic Acid and Thyroid Hormone Receptor

SRC1: Steroid Receptor Coactivator1

STAT: Signal Transducer and Activator of Transcription

STD: Saturation Transfer Difference

SUMO: Small Ubiquitin-related MOdifier

SUV39H1: SUppressor of Variegation 3-9 Homolog 1 = Histone-lysine N-methyltransferase

TAFII: TATA binding protein-associated factor in the transcription factor IID

TEV: Tobacco Etch Virus

TIMP1: Tissue Inhibitor of MetalloProteinase 1

TSA: TrichoStatin A

VA or VPA: Valproic Acid

VLDL: Very-Low-Density Lipoprotein

WaterLOGSY: Water-Ligand Observation with Gradient SpectroscopY

CONTENTS

CHAPTER 1

Introduction

RESUME

Les maladies cardiovasculaires, en particulier, l'athérosclérose, constituent un problème majeur de santé publique. L'athérosclérose consiste en l'accumulation de cholestérol dans les artères causant la formation de plaques rigides ; ces plaques pouvant ensuite se rompent causant attaques et malaises cardiaques. Il existe de nombreux médicaments réduisant les taux de cholestérol dans le sang, ils sont cependant inefficaces pour le traitement de patients souffrant d'hypercholestérolémie familiale. Il est donc important de développer de nouvelles molécules. Il a été observé précédemment que l'inhibition de l'histone deacetylase 7 (HDAC7) augmentait la conversion du cholestérol en acides biliaires en augmentant l'activité de la 7α-cholestérol hydroxylase, l'enzyme limitante du catabolisme du cholestérol. Ainsi cibler HDAC7 pourrait constituer une nouvelle stratégie pour réduire les taux de cholestérol. Cependant HDAC7 présente une haute homologie de son site catalytique avec dix autres HDAC humaines rendant la conception d'inhibiteur spécifique de HDAC7 difficile. Des informations structurales sont requises afin de pouvoir décrypter les mécanismes responsables de la spécificité d'inhibition en fonction des isoformes d'HDACs considérés.

Les histones deacetylase humaines sont considérées comme des corépresseurs de transcription cruciaux dans de nombreux systèmes physiologiques. Il existe 18 HDACs humaines divisées en 4 classes. La classe I comprend les HDACs 1, 2, 3 et 8, la classe IIa les HDACs 4, 5, 7, 9 et la class IIb les HDACs 6 et 10, la classe III les HDACs appelées Sirtuins SIRT1, 2, 3, 4, 5, 6, and 7, et la classe IV qui comprend HDAC11. Les HDACs de classe IIa, classe à laquelle appartient HDAC7, présentent quelques différences avec les histones de classe I. En effet elles sont exprimées de façon spécifique dans des types de tissus déterminés et exercent leur fonction de répression de transcription dans les muscles squelettiques, cardiaques et lisses, dans les os, dans le système immunitaire, dans le système vasculaire et dans le cerveau alors que les histones de classe I sont exprimées de façon ubiquitaire dans tous les tissus. De plus les HDACs de class IIa possèdent un domaine de régulation N-terminal (absent chez les histones de classe I) qui permet d'établir des interactions avec des facteurs de transcription et des corépresseurs spécifiques de chaque type de tissus. Enfin les HDACs de class IIa sont des corépresseurs de transcription dépendant de différentes signalisations cellulaires. Elles sont ainsi phosphorylées au niveau de deux ou trois serines conservées au sein de la class IIa et

situées au sein du domaine N-terminal de régulation. La phosphorylation des HDACs de classe IIa constitue une modification post-traductionnelle importante influençant leur localisation cellulaire soit dans le noyau soit dans le cytoplasme et influençant ainsi leur fonction de corépresseur de transcription dans le noyau. En combinant l'activité de kinase ou phosphatase avec des signaux extracellulaires, différentes voies de signalisation du développement peuvent être induites.

Dans le noyau, les HDACs de class IIa interagissent avec les corépresseurs SMRT/N-CoR et sont recrutées au niveau des promoteurs de leurs cibles à l'aide de facteurs de transcription tissu-spécifique, comme par exemple les facteurs de transcription myocyte enhancer factor 2 (MEF2) et Runx 2. Les HDACs de class IIa interagissent aussi avec d'autres HDACs, HDAC3 par exemple. L'interaction de HDAC7 avec le complexe corépresseur SMRT/N-CoR n'est pas direct mais s'effectue par le biais de HDAC3. Il a été montré que SMRT activait l'activité de deacetylase de HDAC3 par un domaine activateur de deacetylation (DAD), en revanche un tel effet n'a pas été observé pour les HDACs de classe IIa. En outre les HDACs de class IIa, bien que présentant une forte homologie avec les HDACs de classe I, voient un de leurs résidus catalytiques importants une tyrosine, remplacé par une histidine. Cette différence pourrait expliquer pourquoi les HDACs de classe IIa sont inactives sur des histones acétylées. Par ailleurs les substrats de ces HDACs ne sont pas connus : le substrat de ces HDACs pourrait ne pas être une histone mais un tout autre substrat, idée renforcée par la présence de ces HDACs à la fois dans le cytoplasme et dans le noyau.

La protéine SMRT est une grosse protéine de 2525 acides aminés qui recrute les HDACs et réprime la transcription. SMRT peut interagir avec une centaine de protéines. Les HDACs de classe IIa interagissent avec le domaine de répression de SMRT. SMRT est prédite comme étant une protéine nativement désordonnées rendant toute étude structurale difficile, même si quelques structures de peptides de SMRT ont été publiées. L'absence de repliement des protéines est importante pour de nombreux processus cellulaires. Certaines protéines désordonnées sont ainsi impliquées dans des maladies comme par exemple dans certains cancers, maladies cardiovasculaires et neurodegeneratives, maladies auto-immunes et diabètes.

Au début du projet, aucune structure d'HDACs de classe IIa n'était connue. Leur activité enzymatique en tant que protéine recombinante purifiée n'avait pas bien été caractérisée, une activité avait été détectée sur des échantillons préparés à partir d'extraits cellulaires probablement contaminés par d'autres molécules actives telles que d'autres HDACs. De

plus, en dépit de quelques inhibiteurs d'HDACs disponibles (pan-HDAC ou spécifique), quelques uns approuvés par la FDA et disponibles sur le marché, leur mode d'action reste peu compris. Il n'existe pas non plus d'inhibiteurs spécifiques pouvant être utilis2s en tant que médicaments ou outils pour la biologie cellulaire. De nombreuses questions soulevées n'avaient pas encore de réponses : quelle est la structure de HDAC7 ? La structure de HDAC7 est-elle semblable à celles déjà publiées (comme pour HDAC8 et un homologue bactérien) ? Quelles sont les caractéristiques structurales sous-jacentes de la spécificité des inhibiteurs de HDACs, et est-il possible de créer des inhibiteurs spécifiques de HDAC7 ? Est ce que HDAC7 présente une activité de deacetylation sur des histones acétylées ou sur des peptides acétylés dérivés d'histones ?

L'objectif du projet de la thèse consistait à déterminer s'il était possible de créer des inhibiteurs spécifiques de HDAC7. Dans ce but, deux strategies ont été explorées : cibler le domaine catalytique de HDAC7, et cibler l'interface d'intercation entre HDAC7 et SMRT.

1 Cardiovascular diseases in the World

Cardiovascular diseases (CVDs) caused 17.5 million deaths globally in 2005, representing 30% of all deaths (Fig. 1). The incidence is set to increase: by 2030, it is predicted that 23.6 million people will die annually from CVDs defining it as one of the leading causes of death (World Health Organisation, www.who.int/topics/ cardiovascular_diseases/en/). CVDs are not only a concern for high income countries but are also a serious problem for low- and middle-income countries where over 80 % of the world's deaths from CVDs occur.

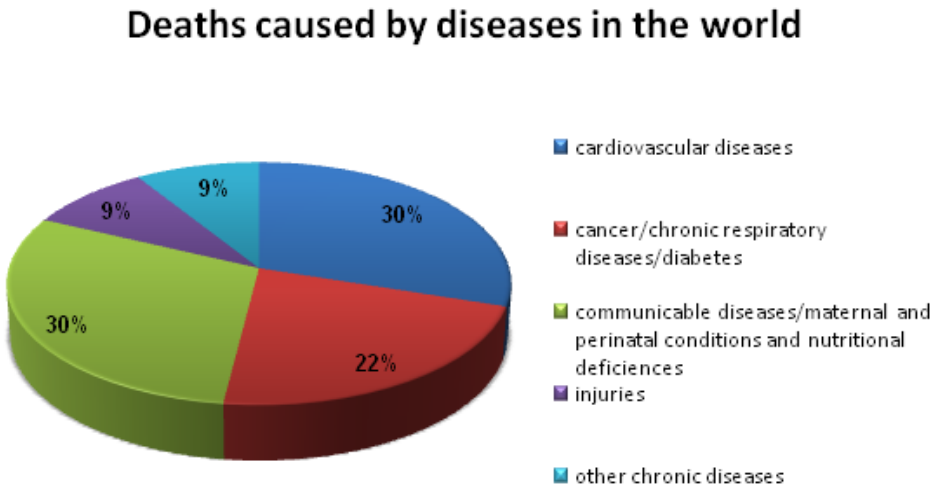


Figure 1. Global death rates in the world. In 2005, about 17.5 million deaths were due to cardiovascular diseases representing about 30% of all global deaths.

Atherosclerosis belongs to the group of CVDs and is an inflammatory disease characterised by the accumulation of lipids and fibrous elements in the large arteries (Ross 1999; Lusis 2000; Glass and Witztum 2001). It results from interactions between modified lipoproteins (modified Low Density Lipoproteins LDL), monocyte-derived macrophages, T cells and the normal cellular elements of the arterial wall. This process leads to the formation of plaques containing lipid-rich necrotic debris and smooth muscle cells in the arterial lumen that can cause complications such as acute occlusion due to the formation of a thrombus or blood clot, leading to myocardial infarction or stroke. Among the many genetic and environmental risk factors identified by epidemiologic studies, elevated level of serum cholesterol is sufficient for the development of atherosclerosis even in absence of the other risk factors (Fig. 2).

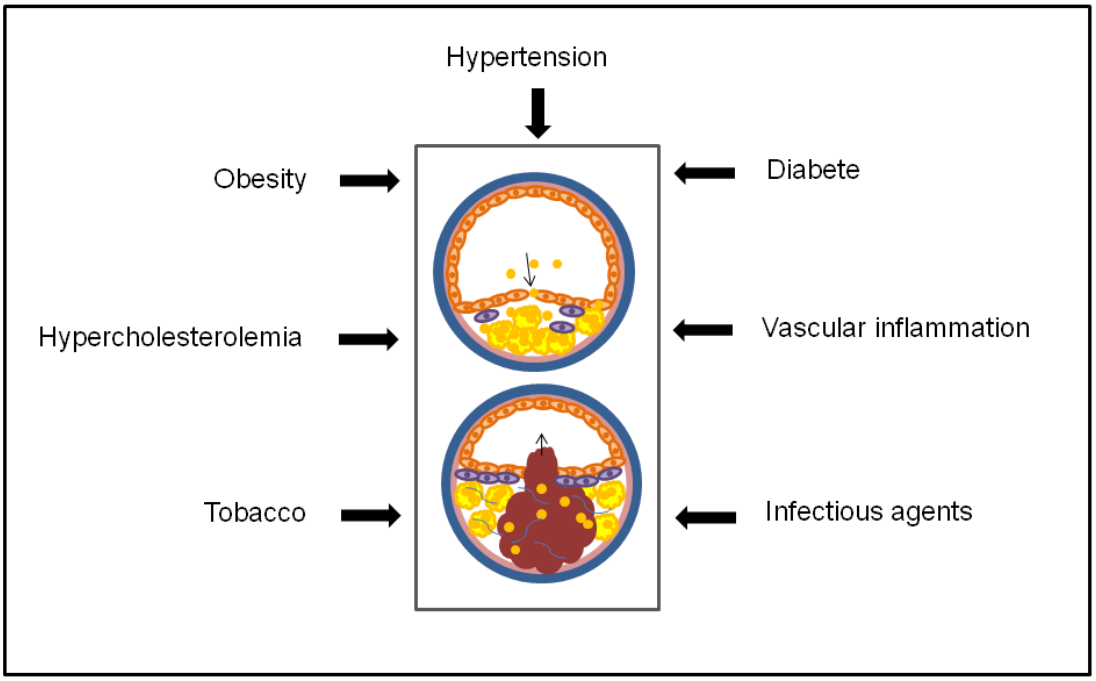


Figure 2. Atherosclerosis and its risk factors. An excess of LDL infiltrates the arteries and accumulates in foam cells (monocytes and macrophages). Those foam cells build up on the wall of arteries forming a lipid core and then a plaque in the later stages of the disease. This plaque is mainly made of fibrous and necrotic elements. It can evolve in the worst case as a thrombus occluding or even disrupting the artery leading to severe infarction. Atherosclerosis is caused by different risk factors such as hypertension, obesity, hypercholesterolemia, diabetes, vascular inflammation, tobacco and infectious agents.

Thanks to a better understanding of cholesterol biogenesis and serum-cholesterol controls, a number of different drugs and therapies have been developed. The bile-acid resins which bind bile acids in the intestine interrupt the enterohepatic circulation of bile acids and increase the conversion of cholesterol into bile acids in the liver. Nicotinic acids have a cholesterol-lowering effect by inhibiting the mobilisation of free fatty acids from peripheral tissues, thereby reducing hepatic synthesis of triglycerides and secretion of VLDL (Very Low Density Lipoproteins). Fibrates, another class of drugs, lower triglycerides levels. Nowadays two families of medication are widely used for treating hypercholesterolemia: the statins and ezetimibes. Statins are structurally similar to hydroxymethylglutaryl-Coenzyme A (HMG-CoA), a precursor of cholesterol and are competitive inhibitors of the HMG-CoA reductase, the regulating enzyme involved in the last regulated step of cholesterol synthesis. The statins lower serum LDL cholesterol concentrations by upregulating LDL-receptor (LDL-R) activity as well as reducing the entry of LDL into circulation. Ezetimibes decrease cholesterol absorption in the intestines. Statins and ezetimibes are often used in combination to benefit from their synergistic and efficient effects. However, the drugs that are available on the market so

far are not sufficient for fully preventing the progression of atherosclerosis in many individuals. For instance such treatments are ineffective for patients with very high risks of CVDs including patients suffering from genetic defects in lipid and lipoprotein metabolisms (e.g. familial hypercholesterolemia); these patients do not respond to hypolipidemic agents. Moreover, these treatments, like most other drugs result in some side effects for patients, most commonly nausea and headache. In the most serious cases, patients can suffer from memory loss, muscle weakness, muscle pain, and even rhabdomyolisis (rapid lysis of skeletal muscles). This is then a serious concern especially because these treatment regimes are long-term (they may be administered over the whole life of a patient). Hence, novel therapies applied as alternatives or in combination with existing treatments are desirable.

Despite the efficacy of existing treatments, atherosclerosis remains a burden for most countries of the world. Scientists and medics from many countries are making great efforts to better understand atherosclerosis with the aim of identifying new drug targets and therapies. Therefore the European commission financed the Framework programme 6, SOUTH consortium for "*applied-oriented studies on regulatory networks involved in lipid homeostasis and atherosclerosis*". This thesis, focusing on one of the risk factors, hypercholesterolemia was financed by this program. It follows a previous project (Framework programme 5, NORTH) where it was demonstrated that the generic HDAC inhibitors trichostatin A (TSA) and valproic acid (VA) were somehow targeting HDAC7 that is recruited to the promoter of the *Cyp7a1* gene that encodes the p450 enzyme 7α-cholesterol hydroxylase. This had the effect of increasing *Cyp7a1* expression leading to a significant decrease in both plasma total cholesterol and LDL-cholesterol, due to its increased conversion into bile acids (Mitro et al. 2007). From this work, it was hypothesised that targeting HDAC7 with inhibitors might be new way to stimulate cholesterol catabolism by upregulating the expression of a gene coding for the p450 enzyme responsible for the catabolism of cholesterol into bile acids. This highlights how eukaryotic gene transcription can be controlled by small molecule intervention. Before going into further detail on the acetylation of histones and HDAC biology, an overview of one of the key components of gene transcription, chromatin, is provided.

2 The chromatin

DNA is the key molecule for storing and propagating genetic information. In mammals, about 30 000 genes exist (Lander et al. 2001; Waterston et al. 2002a; Waterston et al. 2002b) for controlling all cellular processes (such as cellular division, metabolism, etc). For regulating these processes, with such a huge amount of genetic information, other molecules are required. In eukaryotes, the DNA is not naked, but is packaged with an equal mass of proteins forming a large macromolecular assembly called chromatin. As a consequence the chromatin is considered as being the physiological substrate of all cellular processes involving DNA (Felsenfeld and Groudine 2003) forming either transcriptionally active euchromatin or inactive heterochromatin (Clayton et al. 2006; Mellor 2006).

2.1 The nucleosome

Chromatin is composed of a repeating basic unit called the nucleosome. The role of chromatin is to accomplish essentially two roles: (i) allowing the packaging of gigantic DNA molecules into the nucleus, and (ii) allowing the access of DNA at an appropriate time for the genetic information to be decoded.

The nucleosome consists of a protein core, the histone octamer, around which 146 base pairs of DNA are wrapped to form a tight superhelix (Luger et al. 1997). The protein core is composed of 2 copies of 4 histones: H2A, H2B, H3 and H4. The core histones possess long N-terminal tails which protrude outside the nucleosome. These tails represent key elements in the chromatin dynamic process since they can be covalently modified. The sum of these covalent modifications of the chromatin template forms the epigenome.

2.2 Histone post-translational modifications

2.2.1 The histone code

Most of the cells (somatic cells especially) have the same genetic information. Nevertheless, gene expression patterns are markedly different depending on cell types and these differences are strongly influenced by the chromatin structure. Transcriptionally active or inactive regions of chromatin are created by specific modifications of histones, most commonly acetylation, phosphorylation, methylation (Jenuwein and Allis 2001;

Grewal and Elgin 2002; Turner 2002; Felsenfeld and Groudine 2003). Those modifications generally take place on the disordered tails of histones. The chromatin then plays the role of a platform for the integration and recording of signalling events (Cheung et al. 2000). These modifications are induced by extracellular and intracellular regulatory signals. The ensemble of these epigenetic modifications (the epigenome) is inheritable through the cell cycle and meiosis from one generation to the next. This epigenetic information is thought to be indexed as a code called the "histone code", where it works as a fundamental regulatory mechanism in addition to DNA and the genetic information itself (Strahl and Allis 2000; Turner 2000; Jenuwein and Allis 2001). Strahl and Allis proposed a definition of the histone code: "distinct covalent histone modifications acting alone, sequentially or in combination, form a 'histone code' that is then read by effector proteins to bring about different downstream events" (Strahl and Allis 2000). Thus this code might create binding sites for proteins to allow or inhibit gene transcription (Strahl and Allis 2000; Jenuwein and Allis 2001; Fischle et al. 2003a). As with any code, rules have to be defined, but with the knowledge we have so far it is difficult to define definitive ones even though we start to have strong experimental support for acetylation for instance (Marmorstein 2001a; Agalioti et al. 2002; Turner 2002; Kanno et al. 2004); this will be described later on in the introduction.

2.2.2 Histone modifications

Early biochemical studies indicated a variety of post-translational modifications on histones (Allfrey et al. 1964). These post-translational modifications are covalent, and occur mostly on histone tails though modifications have also been reported for the histone core domains (Cosgrove 2007; Godde and Ura 2008). These modifications have the ability to form transcriptionally active regions in otherwise inactive chromatin by inducing changes in the structure of the nucleosomes, impairing their interactions with other protein factors, modifying their propensity to fold into higher order structures or combinations of these effects (Luger and Richmond 1998). Three main roles can be attributed to these histone tail modifications: they are thought to modulate the accessibility of nucleosomal DNA for base-specific recognition; they are interaction platforms for cellular factors (Strahl and Allis 2000; Jenuwein and Allis 2001; Berger 2002); they are responsible for the formation of chromatin higher order structures by

creating protein-protein interactions with neighbouring nucleosomes in order to induce the condensation of the chromatin fibre (Dorigo et al. 2003).

The N-terminal tails of histones are a means for interaction with many nuclear factors that read the signals encoded by post-translational modifications of, commonly, lysine, arginine, serine and threonine residues. The main modifications are methylation, acetylation, phosphorylation sumoylation, ubiquitylation, ADP-ribosylation, deimination and isomerisation (Berger 2002; Kouzarides 2007). Methylation consists of the addition of mono-, di- or trimethyl groups on lysines and arginines. For acetylation, the addition of acetyl groups occurs on lysines. The addition of phosphoryl groups occurs on serines, threonines and tyrosines. Sumoylation and ubiquitylation take place on lysines. Glutamates are subjected to ADP-ribosylation. Finally deimination occurs on arginines and isomerisation on prolines.

These modifications are responsible for different chromosomal and transcriptional states. Acetylation of histone tails is generally associated with activating gene transcription and a number of residues have been identified as being acetylated on all major histones (Sterner and Berger 2000; Cosgrove et al. 2004; Freitas et al. 2004). The proteins involved in the addition of acetyl groups are the histone acetyl transferases including Gcn5 (Brownell et al. 1996) in yeast and CBP/p300 in humans. Acetylation is reversible and the enzymes that deacetylate are the histone deacetylases (HDACs); HDAC1 was the first HDAC to be discovered (Taunton et al. 1996). Phosphorylation is a common, well-studied modification taking place on serines and threonines. In common with acetylation, phosphorylation of histones is involved in various processes such as transcription and DNA repair and is mediated by protein kinases. Phosphorylation is also reversible: marks are removed by serine/threonine phosphatases. Methylation was the first modification to be observed. In contrast to acetylation and phosphorylation which exist as mono modifications, methylated residues can be modified by multiple moieties (mono-, di-, trimethylation, up to two methylations can be observed on arginines and up to three methyl groups on lysines). Methylation is a stable modification (especially on lysines) however demethylating enzymes have been described (Shi et al. 2004). In contrast to the modifications previously described, ubiquitylation consists of the covalent addition of a 76-residue conserved polypeptide. Several marks of ubiquitylation of lysine residues (formation of a polyubiquitination chain) lead to a proteasome-mediated degradation of the target proteins.

2.2.3 Cross-talk of histone modifications

There are two main hypotheses on the way that various histone modifications can affect DNA-templated processes. The first is based upon what is observed for acetylation. Acetylation changes the overall charge of histones leading to disruption of nucleosome and/or chromatin structure allowing protein factors, like transcription factors for example, to reach their DNA targets (Ren and Gorovsky 2001). The second hypothesis relies on the observation that covalently modified residues can specifically bind certain domains, resulting in the recruitment of effector proteins to the chromatin for specific tasks. For instance bromodomains can specifically bind acetylated lysines (Bottomley 2004; Boyer et al. 2004; de la Cruz et al. 2005). Many of the proteins binding covalently modified histones are chromatin-modifying enzymes or part of such complexes. This can lead to cross-talk between histone modifications, where the presence or absence of a histone modification can facilitate or inhibit another modification (Fischle et al. 2003b). For example the Set1-dependant methylation of H3 tail at K4 triggers the association of the Iswi1 ATPase during gene activation in budding yeast (Santos-Rosa et al. 2004) whilst the same modification by Set9 in human cells prevents the binding of the HDAC-containing corepressor NuRD to chromatin (Nishioka et al. 2002). Another example is that H2B ubiquitylation in budding yeast is closely linked to H3K4me and H3K79me representing a trans-tail cross-talk between histone modification where H3K4me and H3K79me requires ubiquitylation of H2B on lysine 123 but not the reverse (Briggs et al. 2002; Dover et al. 2002; Sun and Allis 2002).

Chromatin is the dynamic and physiological template of our genetic information. It is subjected to an array of post-translational modifications on their N-terminal tails which provide a means of regulating the access to the underlying DNA. These modifications are combinatorial and their sum is viewed as a code, the histone code, which still has to be deciphered. Almost all the known histone modifications can either be activating or repressive marks (Berger 2002). Amongst those activating histone modifications, acetylation is recognised as playing a critical role in gene transcription. In addition acetylation has been detected in non-histone proteins as well, suggesting that acetylation might play a role outside of transcription. The next paragraph will be dedicated to this crucial post-translational modification.

3 Histone acetylation and deacetylation

As reviewed above, post-translational modifications (and the enzymes responsible for performing them) of the N-terminal tails of the histone proteins play a key role in gene regulation (Grant 2001; Marmorstein 2001a; Zhang and Reinberg 2001). How those modifications regulate gene expression is not well understood but they seem to work in a combinatorial and coordinated fashion to elicit different biological processes (Strahl and Allis 2000; Schreiber and Bernstein 2002; Turner 2002). Acetylation of histones is a key regulatory modification. It is governed by a dynamic equilibrium between two opposite activities: acetylation by the histone acetyltransferases HATs and deacetylation by histone deacetylases HDACs (Vogelauer et al. 2000; Eberharter and Becker 2002)(Fig. 3).

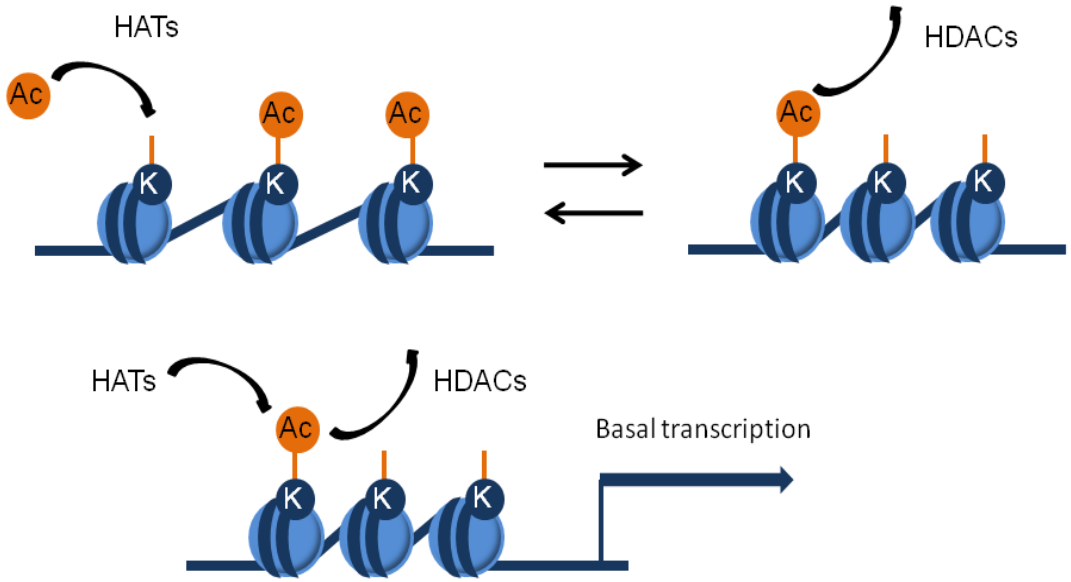


Figure 3. Activities of HATs and HDACs. HATs add acetyl moieties onto histone lysines, whilst HDACs remove them. In cells HAT and HDAC activities are in equilibrium allowing fine adjustment of gene basal transcription.

3.1 Chemical mechanism of acetylation and its role

Acetylation is carried out by the HAT families. There exist about 20 HATs in eukaryotes ranging from yeast to humans. They are classified into different families that exhibit high sequence similarities but little or no sequence similarity exists between these families (Kuo and Allis 1998; Marmorstein 2001b). The different HAT families have specific but distinct histone substrates. Despite substrate differences they have a common feature which is their association with large multi-protein complexes *in vivo* (Grant et al. 1998). Moreover acetylation of nucleosomes occurs only *in vivo* when HAT complexes

are formed (Sterner and Berger 2000) and the substrate specificity is modulated in the presence of these complexes (Grant et al. 1999). HAT families comprise: (i) the GNAT-MYST family, (ii) the CBP/p300 family, (iii) the $TAF_{II}250$ family.

There may be another role for histone acetylation apart from transcription which is nucleosome assembly. This is possible thanks to chaperones that target specifically newly synthesised histones. After their translation, histones are rapidly acetylated allowing their recognition by those chaperones. These are hypothesised to protect acetylated residues from the action of HDACs until the nucleosome has been assembled. Then, after nucleosome packaging to form chromatin, histones can be deacetylated (Shahbazian and Grunstein 2007). Both HAT and HDAC activities are required for correct gene transcription (Carrozza et al. 2005; Clayton et al. 2006; Shahbazian and Grunstein 2007). Acetylation is mostly associated with transcriptional activation and deacetylation with transcriptional repression. However this correlation is not always clear. Recent studies carried out on transcriptionally active genes showed that HDACs, and thus hypoacetylation, play an important role in their transcriptional activation (Smith 2008). These genes are characterised by a fast turnover of acetylation/deacetylation which implies the recruitment of HDACs for immediate repression. The net result is that although acetylation is necessary for gene function, it is not detected.

3.2 Acetylation of non-histone proteins

Early studies demonstrated that many lysine residues in histones are acetylated and that acetylation regulates gene transcription (Allfrey et al. 1964; Vidali et al. 1968). These acetylation sites are conserved amongst vertebrates implying a strong conserving selective pressure. Acetylation activities are also found in *Escherichia coli* suggesting that this function is evolutionarily conserved and has ancient origins (Zhang et al. 2009). Clearly *E. coli* does not have histones; thus it is clear that acetylation occurs on non-histone proteins, as shown recently (Polevoda and Sherman 2002; Glozak et al. 2005; Norris et al. 2009; Spange et al. 2009). Phylogenetic analyses have demonstrated that HDACs are more ancient than histones indicating that the early activity of HDACs was for non-histone substrates (Gregoretti et al. 2004). These non-histone protein targets include the cytoskeleton components α-tubulin and actin, transcription factors (e.g. p53, p65 and pRB), chaperone proteins (HSP90 and HSP70), kinases and phosphatases (e.g. C-Abl tyrosine kinase and PTEN phosphatase), apoptotic factors (Ku70), viral proteins

(HIV Tat) and enzymes involved in DNA repair (DNA glycosylase and NBS1). Acetylation of these non-histone proteins leads to different outcomes, either positive or negative. For instance, acetylation can regulate protein stability (*e.g.* increased stability for p53 or decreased stability for GATA1), DNA binding, gene transcription, protein interactions, cellular localisation, mRNA stability.

So far, acetylation analysis has lacked a global view. However, the acetylome field recently benefitted from technical developments where proteome wide studies have become possible using newly developed anti-acetyllysine antibodies and mass spectrometry. Kim and colleagues studied acetylation at the whole proteome level (Kim et al. 2006) revealing around 400 lysine acetylation sites in almost 200 proteins. In 2009, in a separate study, the dynamics of acetylation were quantified by measuring the change in specific acetylation events (Choudhary et al. 2009). A further 3500 acetylation sites in about 1700 acetylated proteins were discovered. These studies have shown that acetylation events might be "a regulatory modification to rival phosphorylation" in scale (Kouzarides 2000).

4 HDAC enzymes

4.1 Description of HDACs

4.1.1 Classification

The human HDAC family comprises 18 members divided into 4 classes based upon sequence homology with their yeast homologues. HDACs of class I share sequence homology with the yeast Rpd3. This class includes HDAC1, 2, 3 and 8 that contain a catalytic domain of 377-488 amino acids (42-55 kDa) (de Ruijter et al. 2003; Thiagalingam et al. 2003). The class II HDACs are related to the yeast Hda1 deacetylase and are subdivided into two subclasses: HDAC4, 5, 7 and 9 for class IIa and HDAC6 and 10 for class IIb (Grozinger and Schreiber 2002; de Ruijter et al. 2003). Class IIb HDACs, in contrast to other classes, possess two catalytic domains with the second domain being probably inactive as has been shown for HDAC10 (Gallinari et al. 2007). Class II HDACs comprise 855 to 1122 amino acids and have an N-terminal regulatory domain that is absent in class I HDACs (de Ruijter et al. 2003; Thiagalingam et al. 2003; Verdin et al. 2003). The class IV comprises HDAC11 that shares homologies with both class I and II HDACs. Classes I, II and IV all require Zn^{2+} as a cofactor for their enzymatic activity.

Class III HDACs are SIRT1 to 7 that are homologous to yeast Sir2; for this reason they are called sirtuins. They have one catalytic domain ranging from about 300 to 750 amino acids and deacetylate lysines with an activity that is dependent on NAD^+ (and not zinc). This class of protein also has a mono-ADP-ribosyltransferase activity (de Ruijter et al. 2003; Thiagalingam et al. 2003; Saunders and Verdin 2007). None of the HDACs show an intrinsic DNA binding activity; access to DNA is facilitated through a large number of transcription and chromatin-related factors that interact with and recruit class I and class II HDACs to specific chromosomal regions.

4.1.2 Cellular localisation

HDACs from class I are ubiquitously expressed in all cells. They are mostly located in the nucleus although HDAC3 is also present in the cytoplasm (de Ruijter et al. 2003; Thiagalingam et al. 2003). HDAC3 can be sequestered in the cytoplasm by IKBα, an inhibitor of the translocation of NFκB (Yang et al. 2002; Viatour et al. 2003). Class II HDACs shuttle between the nucleus and cytoplasm in response to certain cellular signals (Grozinger and Schreiber 2000; McKinsey et al. 2000a; Wang et al. 2000; Kao et al. 2001; Miska et al. 2001; Zhao et al. 2001). Class IIa HDACs are more strongly expressed in a limited subset of cell types. For instance HDAC4 is abundant in brain, colon, heart, ovary, skeletal muscle, small intestine and thymus. HDAC5 is more common in brain, heart, liver and skeletal muscle whilst HDAC7 is observed in heart, lung and thymus. HDAC9 is most abundant in the brain and heart (Verdin et al. 2003).

4.2 Catalytic mechanism of HDACs

HDACs share features with both metallo- and serine proteases. A catalytic mechanism was proposed based upon the crystal structure of the histone deacetylase catalytic core of a homologue from the hyperthermophilic bacterium *Aquifex aeolicus* which shares 35,2 % identity with human HDAC1 (Finnin et al. 1999; Schultz et al. 2004; Hodawadekar and Marmorstein 2007; Ficner 2009). The active site consists of a tubular pocket, a zinc-binding site and two aspartate-histidine charge relay systems (Fig. 4). The important catalytic residues are conserved amongst HDACs, however the mechanistically important tyrosine (Tyr306) is replaced by a histidine in class IIa HDACs. When a histidine is present, a water molecule seems to replace the hydroxyl group of the missing tyrosine. This difference with class I HDACs might explain why class IIa have a low or

"cryptic" enzymatic activity as shown for HDAC7 (Schuetz et al. 2008) and HDAC4 (Lahm et al. 2007; Bottomley et al. 2008).

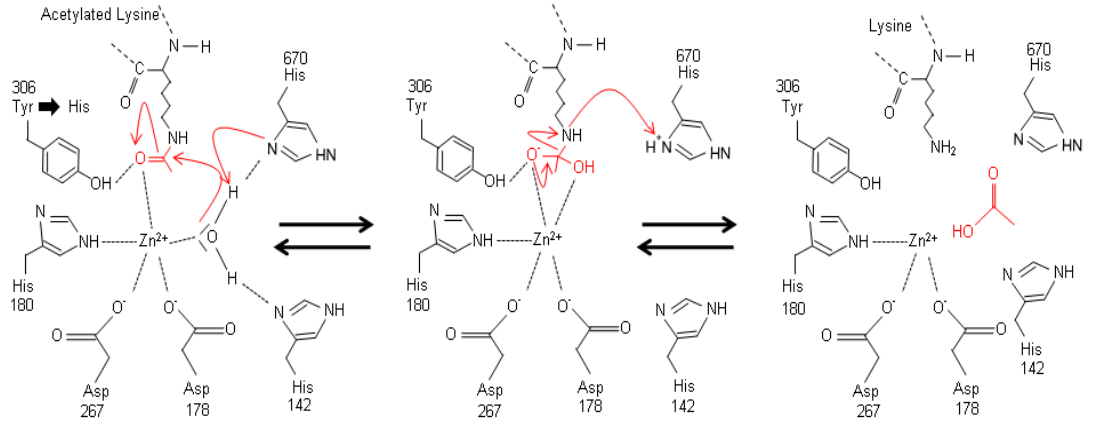


Figure 4. Deacetylation chemical mechanism. The carbonyl oxygen of the N-acetyl amide bond of the lysine binds the catalytic Zn^{2+} ion. The Zn^{2+} polarises the carbonyl group to increase the electrophilicity of the carbon. A water molecule is bound to the Zn^{2+} and its nucleophilicity is increased by hydrogen bonding to two histidine side chains (His143 and 142) and by the charged relay system of Asp178 and His142. The activated water molecule performs a nucleophilic attack on the carbonyl bond of the acetylated lysine giving rise to a tetrahedral oxyanion intermediate. It is stabilised by both the Zn^{2+} and the tyrosine residue 306 (which is replaced by a histidine in class IIa HDACs). In the last step, the carbon nitrogen bond breaks and a proton is transferred from His143 to the scissile nitrogen leading to the acetate and lysine products. (Numbering is as used for class I HDAC8).

4.3 Regulation of HDAC activity

4.3.1 HDAC-containing protein complexes

HDACs are not found in isolation *in vivo* but are always present as components of complexes. The following account will focus on the main interactions of class IIa HDACs (including HDAC7).

Class IIa HDACs contain two main regions: a regulatory N-terminal domain and a C-terminal catalytic domain (Fig. 5). Both of these regions are involved in interactions with others partners. Class IIa HDACs interact with myocyte enhancer factor-2 (MEF2) transcription factors (Sparrow et al. 1999; Lemercier et al. 2000). These transcriptions factors play a major role in the activity of class IIa HDACs in myogenesis, cardiac hypertrophy and T-cell development. With the recruitment of MEF2, class IIa HDACs repress transcription at the MEF2-regulated promoters of genes such as Nur77, myogenin and c-jun.

Class IIa HDACs interact as well with the runt box family (Runx) of transcription factors (Otto et al. 2003). The Runx family regulates the transcription of a variety of

genes and either activates or represses their transcription. Some of the Runx family members play a major role in the developmental regulation of CD4 and CD8 T-cell receptors. Class IIa HDACs are thought to be co-repressors of Runx (especially Runx2) by inhibiting its binding to DNA.

Interaction with B-cell lymphoma 6 (BCL6) has been described (Lemercier et al. 2002). BCL6 is an oncogene involved in the pathogenesis of non-Hodgkin's B-cell lymphomas and is a transcriptional repressor that regulates genes involved in B-cell activation and differentiation, inflammation and cell cycle regulation. The interaction with BCL6 can be direct or via the recruitment of the silencing mediator of retinoic acid and thyroid hormone receptor (SMRT also known as NCoR2) or the BCL6 CoRepressor (BCoR). HDACs interact directly with two closely related co-repressors: SMRT and NCoR (Downes et al. 2000; Huang et al. 2000; Kao et al. 2000). SMRT and NCoR repress gene transcription via three autonomous repression domains called RD1, RD2 and RD3. SMRT and NCoR interact with the C-terminal catalytic domain of class IIa HDACs via RD3 (the interaction with class I HDACs occurs via the SANT domain of SMRT). SMRT and NCoR are found very often in complexes with TBL1, GPS2 and HDAC3. Class IIa HDACs are not active alone but appear to be active in complexes always containing SMRT/NCoR and HDAC3 (Fischle et al. 2001; Guenther et al. 2001; Fischle et al. 2002).

Additional interactions of class IIa HDACs

- Members of the family 14-3-3 proteins bind to class IIa HDACs and regulate their nucleocytoplasmic shuttling (Fu et al. 2000; Grozinger and Schreiber 2000; Wang et al. 2000). This cytoplasmic localisation of class IIa HDACs represents an important and critical aspect of their regulation.
- Class IIa HDACs bind to Ca^{2+}/calmodulin and this disrupts their interaction with MEF2
- Heterochromatin protein 1 (HP1) contains a chromodomain involved in the specific recognition of methylated lysine 9 of histone H3. HP1 interacts with the histone methyltransferases SUV39H1 which coimmunoprecipitate with class IIa HDACs. This establishes a link between HDAC and histone methyltransferase activity: the class IIa/HP1/SUV39H1 complex may mediate a coupled deacetylation/methylation activity important for heterochromatin formation and maintenance (Zhang et al. 2002a).

- Class IIa HDACs interact with other proteins such as the corepressor E1A C-terminal binding protein (CtBP),and can interact with other HDACs in homotypic interactions *i.e.* HDAC7 can bind to HDAC1, 2, 4 and 5 and itself (Dressel et al. 2001).

These interactions are summarised in Fig. 5.

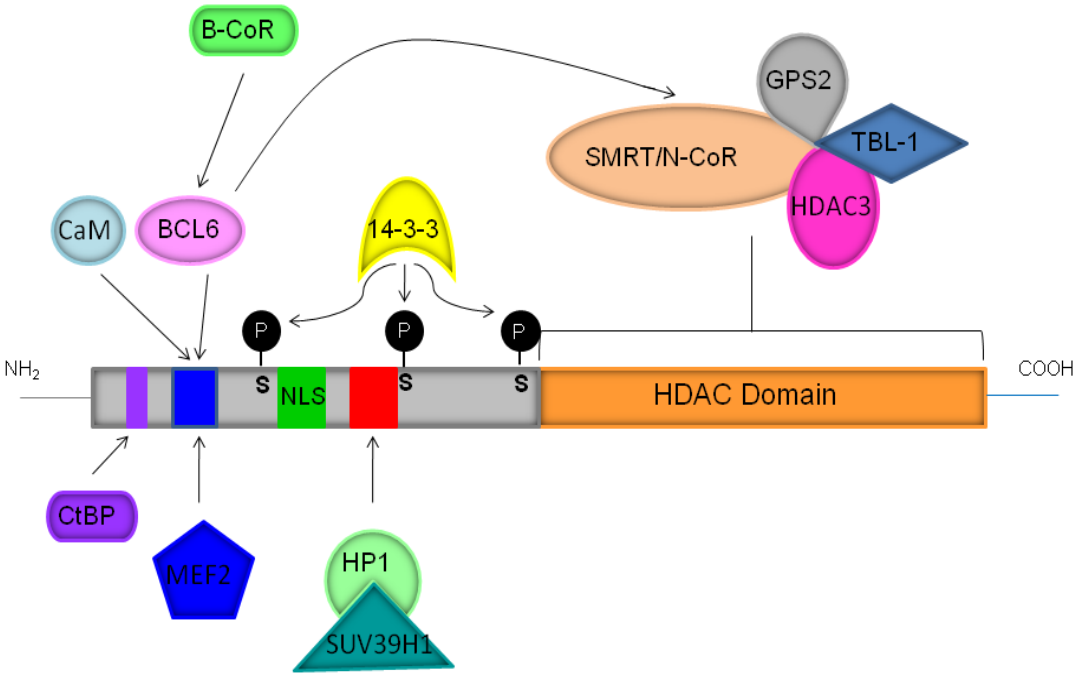


Figure 5. Class IIa HDAC main interaction partners. Class IIa HDACs interact with different partners via the N-terminal regulatory region and C-terminal catalytic domain. The regulatory domain is targeted by different transcription factors. HP1, CtBP, SMRT/N-CoR binding to the catalytic domain and B-CoR are responsible for the repressive activity of class IIa HDACs.

4.3.2 Regulation of HDAC activity

Since HDACs are major players in the regulation of cellular processes such as gene expression, cell growth, differentiation, etc (Kouzarides 1999; Lagger et al. 2002; Lehrmann et al. 2002; Glozak and Seto 2007), it is crucial that their activity is tightly regulated. The regulatory mechanisms are carried out through their cellular localisation, post-translational modifications, gene expression, presence of cofactors, and proteolytic cleavage (Sengupta and Seto 2004; Gallinari et al. 2007). Two main means of regulation will be described hereafter: regulation via subcellular localisation and via post-translational modifications.

4.3.2.1 Regulation by subcellular localisation: the nucleocytoplasmic shuttling

Class IIa HDACs shuttle between the nucleus and the cytoplasm by binding 14-3-3 proteins (a highly conserved acidic protein family). The binding of 14-3-3 with class IIa HDACs requires the phosphorylation of two or three conserved serines in the N-terminal regulatory domain and mediates their cytoplasmic sequestration. The 14-3-3 proteins modify the subcellular localisation of their targets by modulating the activity of nuclear localisation signals (NLS) and nuclear export signals (NES). When the serines in the N-terminal domain are phosphorylated the protein 14-3-3 binds to the HDAC masking the NLS therefore preventing trafficking of HDAC into the nucleus. This interaction exposes the NES for CRM1 binding that triggers its export to the cytoplasm. This shuttling may be a means for regulating the activity of class IIa HDACS by removing them either from the nucleus or the cytoplasm (Kao et al. 2001; Gao et al. 2006).

4.3.2.2 Phosphorylation of class IIa HDACs

Phosphorylation of the 14-3-3 binding sites of HDACs is the main mechanism regulating nucleocytoplasmic shuttling and the signal-dependent repression of target genes of the class IIa HDACs. The phosphorylation sites of these HDACs are closely related to the consensus phosphorylation sites for the Ca^{2+}/calmodulin-dependent protein kinases (CaMKs). For instance CaMKI and CaMKIV can phosphorylate all four class IIa HDACs promoting their association with 14-3-3 proteins and stimulating their export to the cytoplasm (McKinsey et al. 2000b; Wang et al. 2000; Kao et al. 2001; McKinsey et al. 2001). PKD, a protein kinase related to both PKC and CaMK has been shown to phosphorylate HDAC5 and HDAC7 (Vega et al. 2004; Dequiedt et al. 2005). The latter occurs in T-cells leading to the nuclear export of HDAC7 and may play a role in early T-cell development, the formation of heart or nervous systems.

4.3.2.3 Other post-translational modifications

HDAC4 and 9 are subjected to sumoylation (the target lysine is also conserved in HDAC5) (Kirsh et al. 2002). It seems likely that SUMO could alter the interaction properties of its targets thereby affecting cell localisation; how this functions is not clear. It has also been proposed that sumoylation might be coordinated with phosphorylation by an unknown mechanism where two proteins are involved: SENP “sumo remover” and an

E3 ligase, both acting to promote HDAC5 14-3-3 binding site phosphorylation (Chang et al. 2005).

HDAC4 and HDAC7 are both cleaved by caspases releasing the N-terminal NLS from the NES containing C-terminal domain (Liu et al. 2004; Paroni et al. 2004). The released N-terminal fragment of HDAC4 keeps its repressor ability, but the cleavage of HDAC7 abolishes its transcription repressor activity.

Class IIa HDAC activity is controlled tightly by different processes. Nevertheless despite these regulation mechanisms, abnormal HDAC activity can occur leading to diseases such as cancers (Johnstone 2002; Dokmanovic and Marks 2005; Glozak and Seto 2007) resulting from an imbalance between acetylation and deacetylation of histones and non-histone targets. Many research laboratories and pharmaceutical companies are putting efforts to develop HDAC inhibitors as a means for treating those diseases.

5 HDAC inhibitors

5.1 HDAC inhibitor classes

HDAC inhibitors (HDACi) are divided into 5 main structural classes (Fig. 6). Trichostatin A (TSA), originally used as antifungal agent, is the first natural hydroxamate discovered to inhibit HDACs directly (Yoshida et al. 1990). After this discovery other promising compounds were synthesised based upon the hydroxamate active group. One of these, suberoyl anilide hydroxamic acid (SAHA) is less complex structurally than TSA and inhibits partially purified HDACs at nanomolar concentrations (Richon et al. 1998). Other hydroxamate compounds have been described as potent HDAC inhibitors: M-carboxycinnamic acid bishydroxamide (CBHA) and derivatives (Curtin et al. 2002; Bouchain et al. 2003). TSA and SAHA inhibit both class I and class II (there is only one reference demonstrating specificity of SAHA for HDAC7 by (Dokmanovic et al. 2007)).

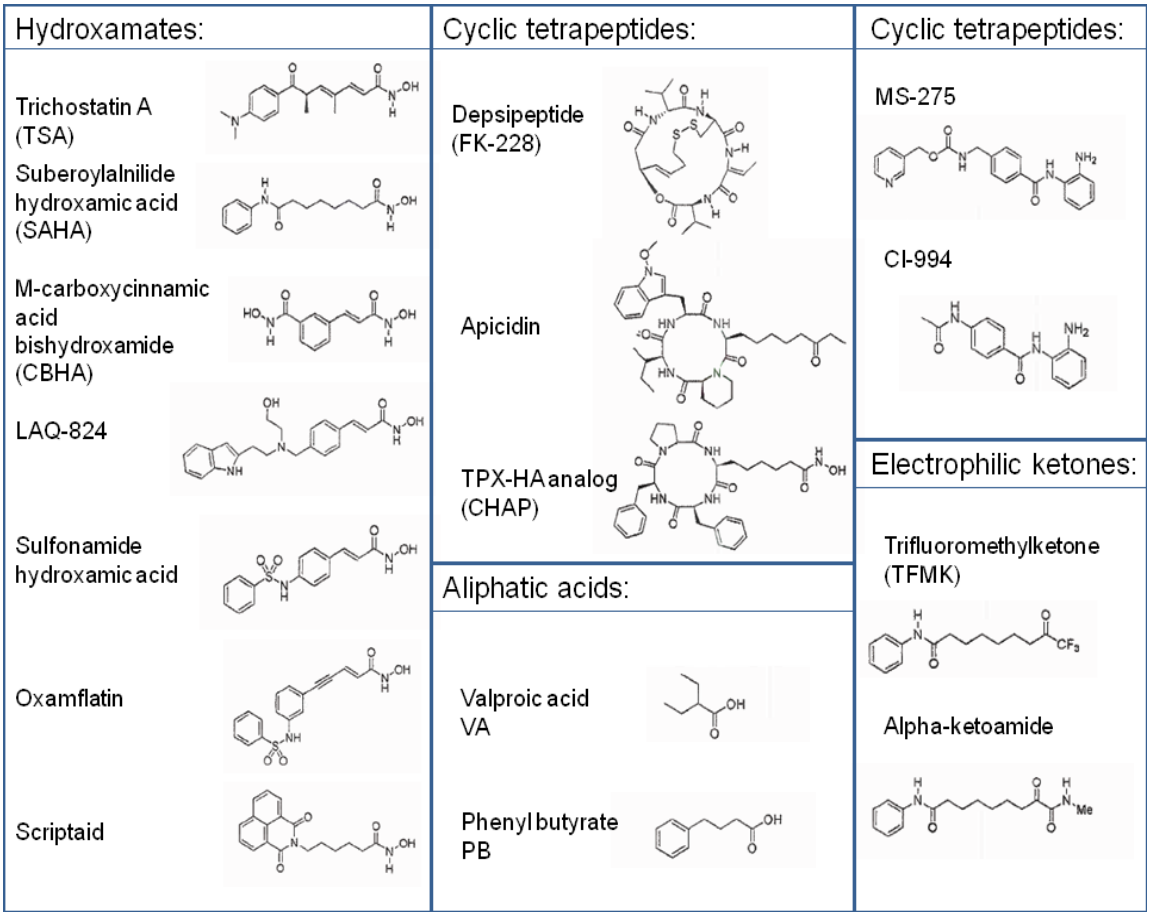


Figure 6. Histone deacetylase inhibitors. HDAC inhibitors are divided into five different classes: (i) the hydroxamates, (ii) the cyclic peptides, (iii) the aliphatic acids, (iv) the benzamides and (v) the electrophilic ketones).

Other inhibitors include:

- The cyclic tetrapeptides are the most structurally complex class of HDAC inhibitors. They include depsipeptide (inhibitor of HDAC1 and 2) and CHAPs, both of which are active at nanomolar levels (Furumai et al. 2001; Furumai et al. 2002; Singh et al. 2002). Their efficiency for disease treatment remains poorly described.
- The aliphatic acids are the least potent class of HDAC inhibitor and exhibit effects at millimolar concentrations; these include valproic acid (VA) and phenylbutyrate (Gottlicher et al. 2001; Phiel et al. 2001; Boivin et al. 2002). VA is a well-tolerated anti-epileptic agent which has been recently shown to be a potent HDAC inhibitor (Gottlicher et al. 2001; Phiel et al. 2001).
- The benzamide class is less potent than hydroxamates and cyclic tetrapeptides. They include MS-275 and CI-994 (Saito et al. 1999; Prakash et al. 2001). These agents typically possess HDAC inhibiting activity at micromolar concentrations.

- The aliphatic ketones are a new recent class of HDAC inhibitors. They include the trifluoromethylketones and the α-ketoamides. They inhibit HDACs in the micromolar range (Frey et al. 2002).

The crystal structure of HDLP (a bacterial homologue of mammalian HDAC) solved in complex with TSA and SAHA revealed the interaction between HDLP and its inhibitors (Finnin et al. 1999) and HDAC8 with a hydroxamic acid (Somoza et al. 2004). The structure-activity relationship (SAR) of the HDAC inhibitor classes shed light on key inhibitor features. It appears that the direct interaction of the hydroxamate group with the zinc at the bottom of the catalytic pocket is a prerequisite for inhibitory activity through blocking of the catalytic site. The structural characteristics of most active HDAC inhibitors are: (i) a hydroxamic acid group, (ii) a spacer aliphatic chain of 5-6 carbons, (iii) a second carbonyl polar group and (iv) a protein surface recognition domain such as a phenyl amino group. All these features are summarised in Fig. 7.

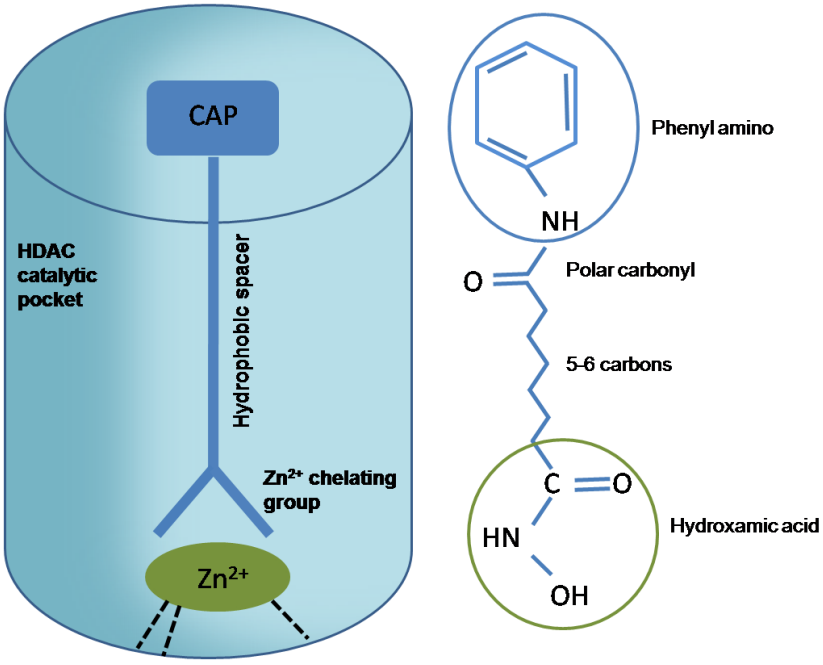


Figure 7. Schematic HDAC catalytic site with a typical inhibitor. The catalytic pocket is tubular. The inhibitor is composed of a cap pointing out of the catalytic pocket, a hydrophobic spacer and a hydroxamic group chelating the zinc sitting in the catalytic pocket.

5.2 Activity of HDACi

HDACi have different actions. They cause induction of differentiation, growth arrest and apoptosis in a broad spectrum of transformed cells in culture and in animal tumours (Marks et al. 2001; Kelly et al. 2002). This inhibitory effect is due in part to accumulation

of acetylated proteins which play a major role in gene transcription regulation (Fig. 8). HDACi alter the acetylated state of histones and other proteins such as regulators of gene transcription (*e.g.* NFKB, GATA1, CREB, MyoD), regulators of cell cycle progression (pRB and p53), hormone receptors (such as glucocorticoids and thyroid hormone receptors) and chaperone proteins (such as HSP90). Normal cells seem to be less sensitive to growth inhibition and apoptotic effects of the HDACi than abnormal cells (Qiu et al. 1999). HDACi share an important feature which is their selectivity in altering gene expression in transformed cells.

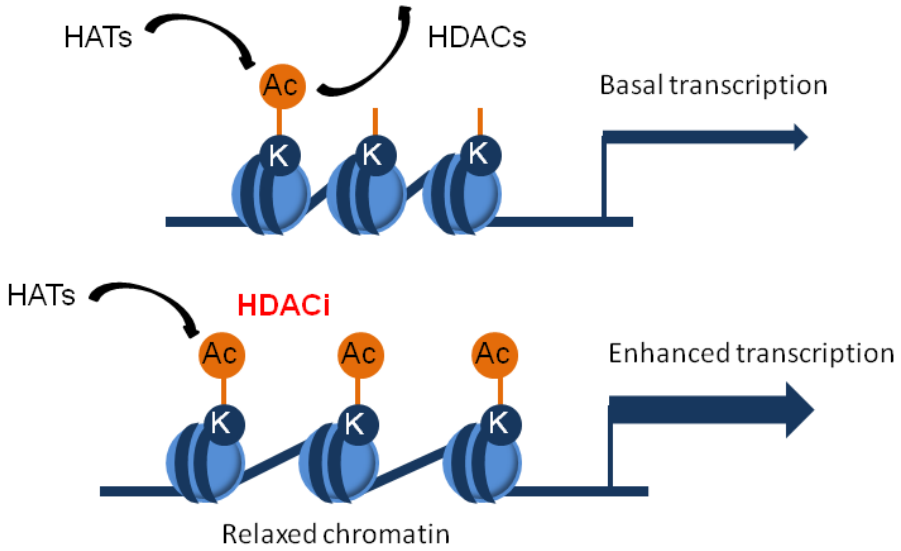


Figure 8. Action of HDAC inhibitors on gene transcription. In normal condition HAT and HDAC activities are in equilibrium leading to basal gene transcription. By targeting HDACs by inhibitors, the equilibrium is disrupted leading to a hyperacetylation. When the chromatin is hyperacetylated it is in a relaxed form leading to enhanced gene transcription. Acetylated histones are a common feature for transcriptionally active chromatin and deacetylated histones for transcriptionally silent chromatin.

Fewer than 2-5 % of expressed genes are up- or down-regulated by more than 2-fold within 1-6 h of culture treatment with TSA or SAHA. It has been shown that treatment with TSA or SAHA increased p21 expression (Richon et al. 2000) and decreased the expression of the thymidylate synthase leading to cell cycle arrest (Glaser et al. 2003). There is as well the example of genes dependent of STAT5 that are repressed by TSA; transcriptional activation of those genes requires a deacetylase activity (Rascle et al. 2003).

It was shown as well that in the SAHA- and MS-275-sensitive acetylome, only 10% of all acetylation sites showed increases in acetylation levels following HDACi treatment (Norris et al. 2009). From these data, it appears that inhibitors have more specific and limited *in vivo* effects than previously thought. Different inhibitors result in differences in acetylation events. For example, HSP90 acetylation (inhibits its activity) was induced by SAHA but not MS-275, whereas acetylation of p53 was induced by MS-275 and not SAHA (Norris et al. 2009).

Many HDACi have been developed within the past few years, with several being approved by the FDA or in clinical trials. They seem to act by complicated mechanisms that are difficult to explain. Research in the field is ongoing both from a structural and biological point of view.

6 HDAC7 and its role in disease mechanisms

6.1 HDAC7 and T-cells

HDAC7 is a regulator of apoptosis in developing thymocytes (Dequiedt et al. 2003). In particular HDAC7 plays a critical role in determining the threshold level at which a developing T-cell undergoes positive versus negative selection. HDAC7 inhibits negative selection, promoting positive selection of T-cells by inhibiting Nur77-mediated apoptosis (Fig. 9).

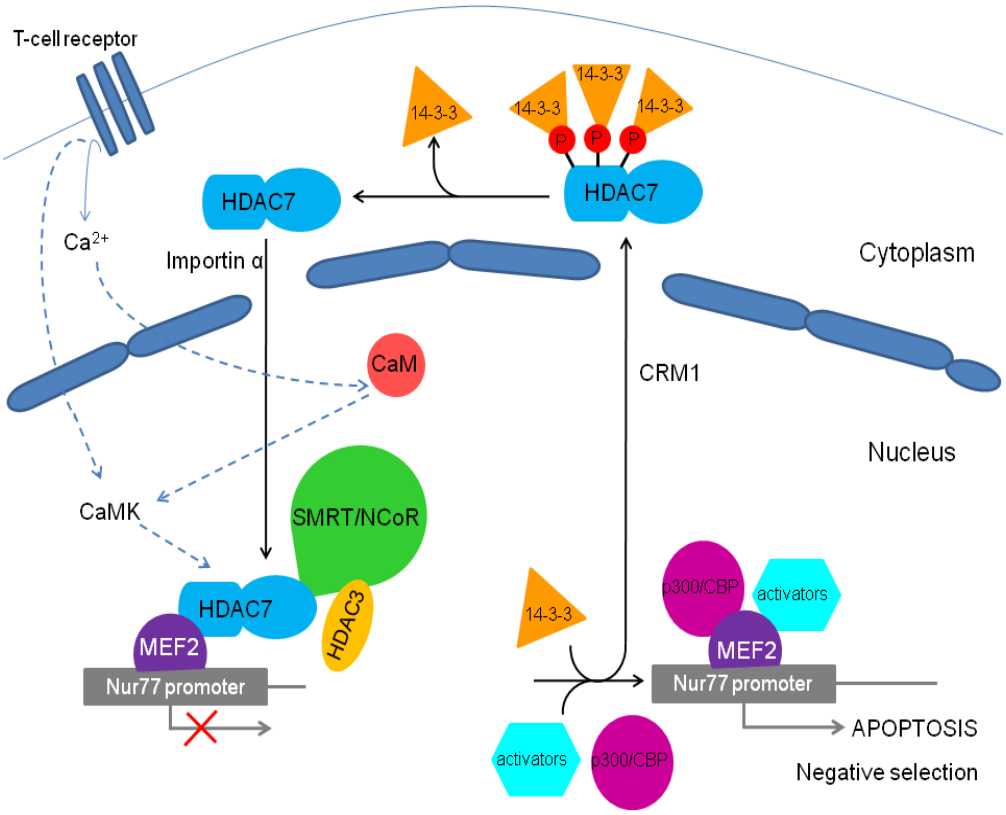


Figure 9. Apoptosis regulation by HDAC7 in developing thymocytes. The Ca^{2+} sensor calmodulin (CaM) is activated upon binding with Ca^{2+}. The activated CaM then activates Ca^{2+}/CaM-dependent protein kinases (CaMKs) which phosphorylate the N-terminal domain of HDAC7. This phosphorylation, maybe involving conformational change of HDAC7, disrupts the binding of HDAC7 with MEF2. HDAC7 is then recruited by 14-3-3 protein and CRM1 contributing to HDAC7 nuclear export. Coactivators like p300/CBP and activators can freely bind to MEF2 activating the transcription of Nur77 leading to apoptosis and therefore to negative selection of T-cells. With HDAC7 bound to MEF2, a repression of transcription of Nur77 prevents apoptosis.

6.2 HDAC7 and blood vessels

HDAC7 regulates blood vessel integrity mediated by the interaction with the myocyte enhancer factor-2 (MEF2). Upon interaction with MEF2, HDAC7 down-regulates the expression of the matrix metalloproteinase-10 (MMP10) and up-regulates the expression of the tissue inhibitor of metalloproteinase 1 (TIMP1) (Chang et al. 2006). Both phenomena contribute to maintenance of vascular integrity. When HDAC7 is silenced, MMP10 is overexpressed and TIMP1 is underexpressed leading to extracellular matrix disruption and blood vessel rupture (Fig. 10).

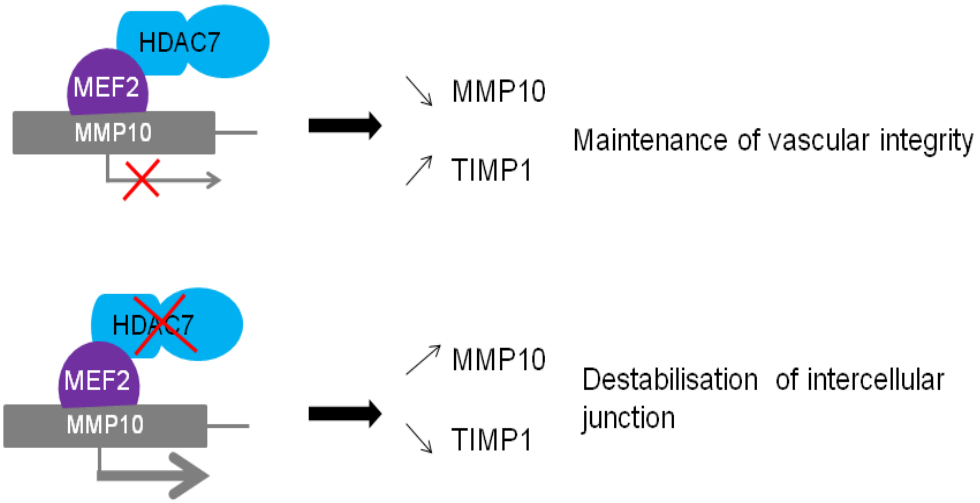


Figure 10. HDAC7 supports vascular integrity. HDAC7 represses the expression of MMP10 leading to vascular integrity. In contrast, when HDAC7 is silenced, MMP10 is overexpressed whilst TIMP1 expression is decreased. This leads to vascular destabilisation and disruption.

6.3 HDAC7 and cystic fibrosis

Recently HDAC7 was shown to be involved in cystic fibrosis (CF) (Hutt et al. 2009). This is a hereditary disease responsible for defects in the cystic fibrosis transmembrane conductance regulator (CFTR), an ion pump responsible for keeping lung epithelium hydrated and free of bacterial infections. About 90% of people suffering from CF have a deletion of phenylalanine 508 (ΔF508) on at least one allele. This deletion produces a misfolded version of the protein which is not transported to the cell surface and which is instead flagged by the endoplasmic reticulum for degradation by the proteasome. Research efforts have focused on rescuing the defective protein from degradation and redirect it to the cell surface. To prevent the degradation of ΔF508, Balch and colleagues proposed to block HDAC activities that regulate chaperones responsible for CFTR folding. By combining the HDACi SAHA with siRNA analysis, they showed that HDAC7 was involved in this process: siRNA knockdown of HDAC7 corrected the defect of ΔF508, the same was observed with SAHA.

It is still not known how HDAC7 inhibition improves the folding of the mutant CFTR. The chaperone protein HSP90 might be involved, but further investigations are needed to confirm the involvement of HSP90 or other chaperones. Alternatives to HSP90 may be preferable as HSP90 is involved in the folding of many other proteins important for human health so side effects would be likely.

6.4 HDAC7 and cholesterol catabolism

This project is based upon work that showed HDAC7 to be involved in cholesterol catabolism (Gilardi et al. 2007). HDAC7 was thus thought to represent a possible target for preventing atherosclerosis and CVDs caused by abnormally high levels of cholesterol. One of the main regulation pathways is the catabolism of cholesterol into bile acids. The synthesis of bile acid is taking place through two different pathways: the classical path starting with the enzyme CYP7A1 and the alternative pathway with the enzyme CYP27A1 being involved (Endo 1992; Javitt 1994; Javitt 2002). It was shown that the synthesis of bile acids is mainly regulated at the level of the gene encoding cholesterol 7α-hydroxylase (CYP7A1), the rate-limiting enzyme of this metabolic pathway. The CYP7A1 enzyme is a cytochrome P450 enzyme which catalyses the hydroxylation of cholesterol at the 7α position, the first reaction in the classical pathway of bile acid synthesis (Fig. 11). Cholesterol catabolism was has been of interest since 1970s but research in the field to unravel the mechanisms improved greatly from the 1990s when it was discovered that CYP7A1 was regulated at the transcriptional level influenced by two effectors: bile acids and cholesterol (Russell and Setchell 1992; Russell 2003). Bile acid synthesis is regulated by many mechanisms, the FXR/SHP and HNF4α pathways.

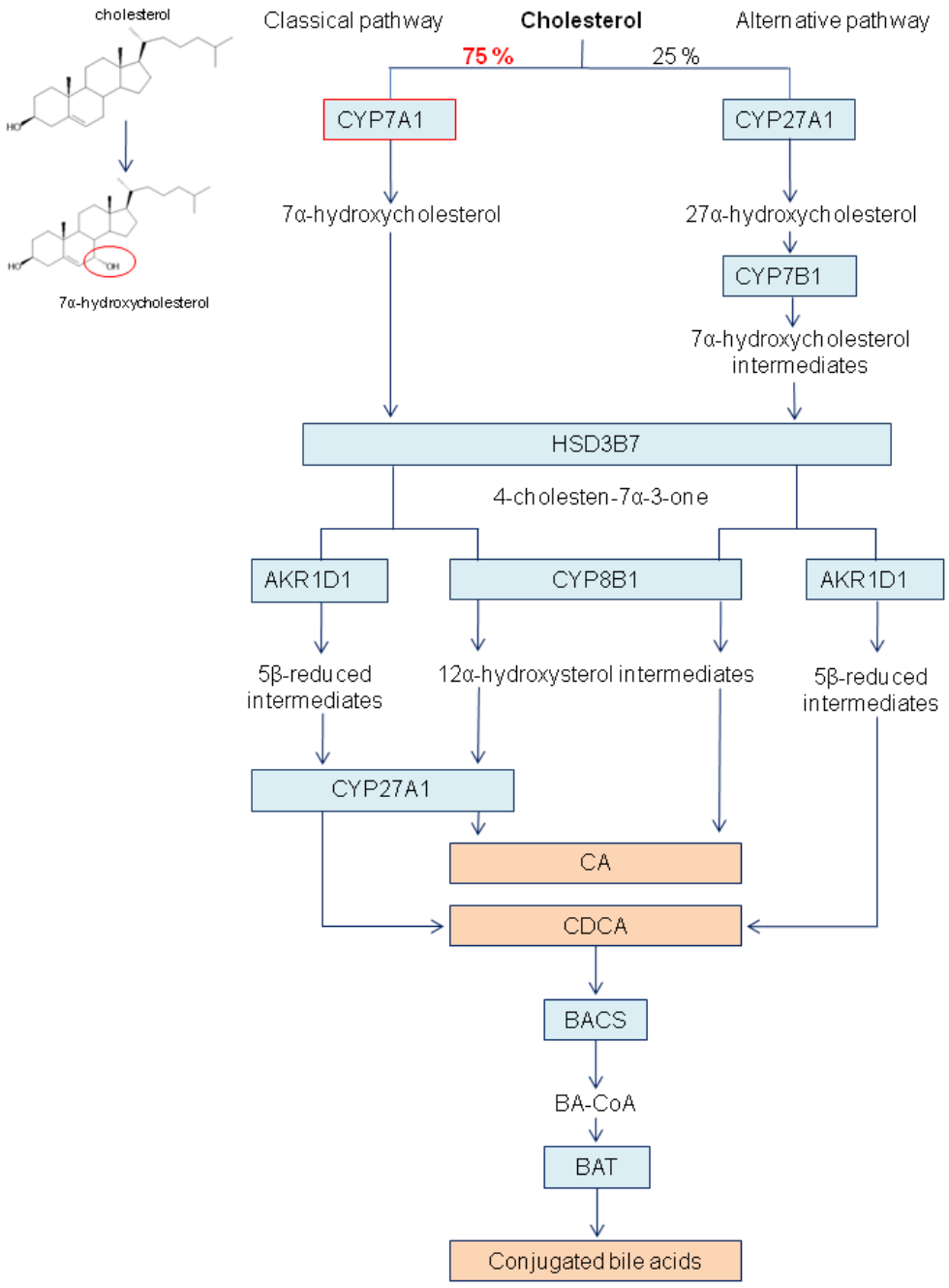


Figure 11. Bile acid synthesis pathways. Cholesterol catabolism into bile acids (BA) involves two different pathways: the classical (or neutral) pathway and the alternative (or acidic) pathway. The classical pathway is responsible for about 75% of total bile acid synthesis. The cholesterol 7α-hydroxylase or CYP7A1 is the rate-limiting enzyme of the classical pathway, the sterol 27α-hydroxylase or CYP27A1 is the rate-limiting enzyme in the alternative pathway. The steps leading to bile acid synthesis include initiation by hydroxylation of cholesterol in position 7 by CYP7A1 (and CYP7B1 in the alternative pathway), modification of the sterol ring (performed by HSD3B7, AKR1D1 and CYP8B1 enzymes in both pathways), oxidation and shortening of the side chain by CYP27A1 (in both pathways). Those steps generate primary bile acids: cholic acids (CA) and chenodeoxycholic acid (CDCA). Conjugation is the last step, it is carried out by BA-CoA synthase (BACs) and BA-CoA:amino acid N-acyltransferase (BAT).

6.4.1 The FXR/SHP pathway

Bile acids were shown to be the most active ligand of FXR (Wang et al. 1999). On the promoter region of the *cyp7a1* gene sits a bile acid responsive element BARE (Stroup et al. 1997). Under basal conditions the nuclear factor α-fetoprotein transcription factor (FTF)/Liver homologue receptor 1 (LHR1) or *cyp7a1* promoter binding factor binds to BARE (Nitta et al. 1999) for the activation of the transcription of *cyp7a1* gene inducing the production of bile acids. Those bile acids activate the transcription of the *shp* gene. Thanks to the binding of a heterodimer of RXR/FXR on the FXRE of the promoter region of *shp*, the produced SHP represses both the transcription of *cyp7a1* by binding to the FTF/LRH1 on the *cyp7a1* promoter and its own transcription. The FXR activation is due to its heterodimerisation with RXR leading to rather low transcription rates and to the recruitment of different coactivators like SRC1 (Steroid Receptor Coactivator1) and PPARγ coactivator 1 (PGC1) (Fig. 12).

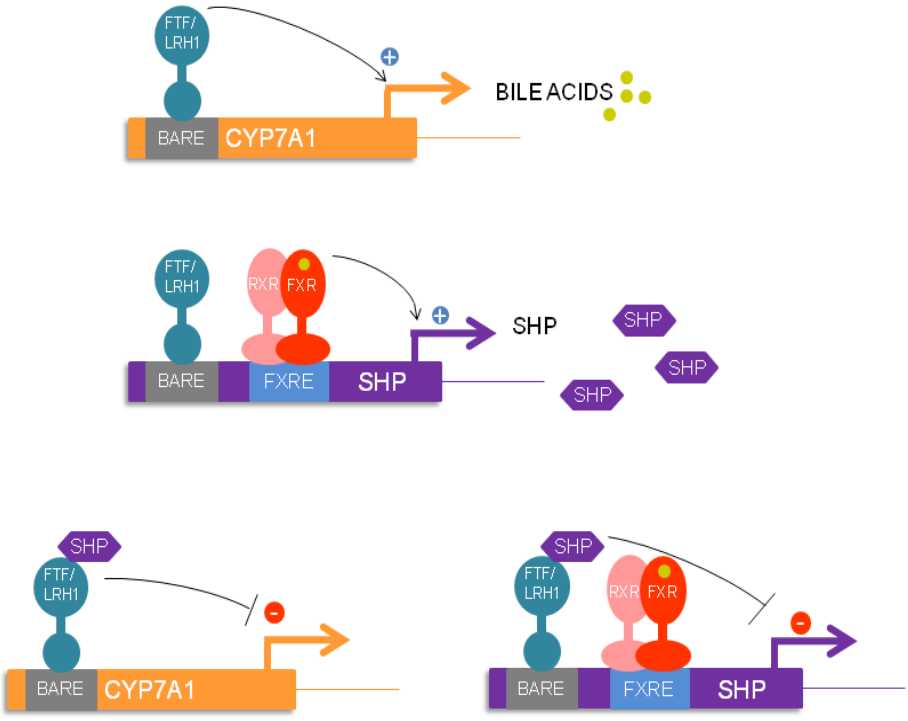


Figure 12. Regulation of Cyp7a1 gene by bile acids involving the nuclear receptors FXR, SHP and FTF/LRH1. FTF/LRH1 binds to the bile acid responsive element BARE of the *cyp7a1* gene activating its transcription. Moreover bile acids can induce *shp* transcription, when FXR responsive element of the promoter of *shp* is occupied by the bile acid-activated heterodimer FXR:RXR. This leads to expression of SHP proteins that act as repressors when bound to the FTF/LRH1. SHP is as well able to repress its own transcription.

6.4.2 The HNF4α pathway

HNF4α (Hepatocyte Nuclear Factor 4α) is one of the activators of the *cyp7a1* promoter (Crestani et al. 1998). Under basal conditions HNF4α binds to the BARE of the *cyp7a1* gene and induces its transcription by recruiting the coactivators PGC1α and CBP. The complex formed recruits the preinitiation complex (PIC) on the TATA box and activates the RNA polymerase II leading to the transcription activation of *cyp7a1*. The bile acids produced perform a negative feedback regulation by dissociating CBP and PGC1, probably by conformational change due to a post-translational modification of the HNF4α decreasing its affinity for the coactivators (Fig. 13).

Figure 13. Regulation of *cyp7a1* gene transcription by the HNF4 pathway. HNF-4 binds to the BARE of the *cyp7a1* gene and recruits coactivators like PGC-1 and CBP. This complex then recruits the preinitiation complex PIC leading to activation of the RNA polymerase II allowing transcription. Upon bile acid production the coactivator complex dissociates from HNF-4 which prevents the recruitment of PIC triggering the *cyp7a1* transcription.

6.4.3 Chromatin-modifying enzyme complexes

In addition to the above regulation mechanism, there might be another form of regulation by chromatin remodelling. Kemper and colleagues showed that bile acids triggered the recruitment of the ATP-dependent mSin3A-Swi/Snf chromatin remodelling complex to the *cyp7a1* promoter through direct interaction with SHP (associated with HDAC1) (Kemper et al. 2004). Moreover it was shown that SHP interacts with the G9a methyltransferase which associates with hypoacetylated and lysine 9-methylated H3 (an epigenetic mark for transcriptionally inactive chromatin) (Boulias and Talianidis 2004; Kemper et al. 2004; Fang et al. 2007).

It was shown that FXR/SHP rather mediates a long-term regulation while there might exist other mechanisms with rapid effects (Mitro et al. 2007). Notably, the corepressor complex made of SMRT/NCoR, HDAC3 and HDAC7 associates rapidly to the promoter region of *cyp7a1*. This leads to the deacetylation of H3 and H4 and to decreased

elongation activity of RNA polymerase II. Bile acids might decrease rapidly the phosphorylation of the large subunit of RNA polymerase II on serine 2 (associated with transcriptional elongation). It might be possible that this decrease in phosphorylation triggers the recruitment of the corepressor complex on the *cyp7a1* promoter.

Moreover bile acids stimulate the shuttling of HDAC7 from the cytoplasm to the nucleus (Mitro et al. 2007). They showed that siRNA against HDAC7 and phosphatase inhibitors (inhibiting nuclear translocation) antagonised the repression of bile acids on *cyp7a1* mRNA levels indicating that the translocation of HDAC7 to the nucleus might be the important event for transcription repression of *cyp7a1*. One possible explanation of the action of HDAC7 on *cyp7a1* transcription might come from the interaction of HDAC7 with HNF4α and its deacetylation inactivating receptor activity (Soutoglou et al. 2000) (Fig. 14). Those observations made it possible to explain the negative feedback of bile acids on *cyp7a1* gene transcription by HDACi. This strategy should increase the catabolic conversion of cholesterol to bile acids and result in a decrease in plasma cholesterol. Two HDAC inhibitors VPA and TSA increased mRNA levels of *cyp7a1* in LDL receptor deficient mice and decreased plasma total and LDL cholesterol (Mitro et al. 2007). These results suggest a new strategy for treating people suffering from familial hypercholesterolemia (due to a mutation in LDL-R) that frequently leads to atherosclerosis and early myocardial infarctions (Kwiterovich et al. 1974).

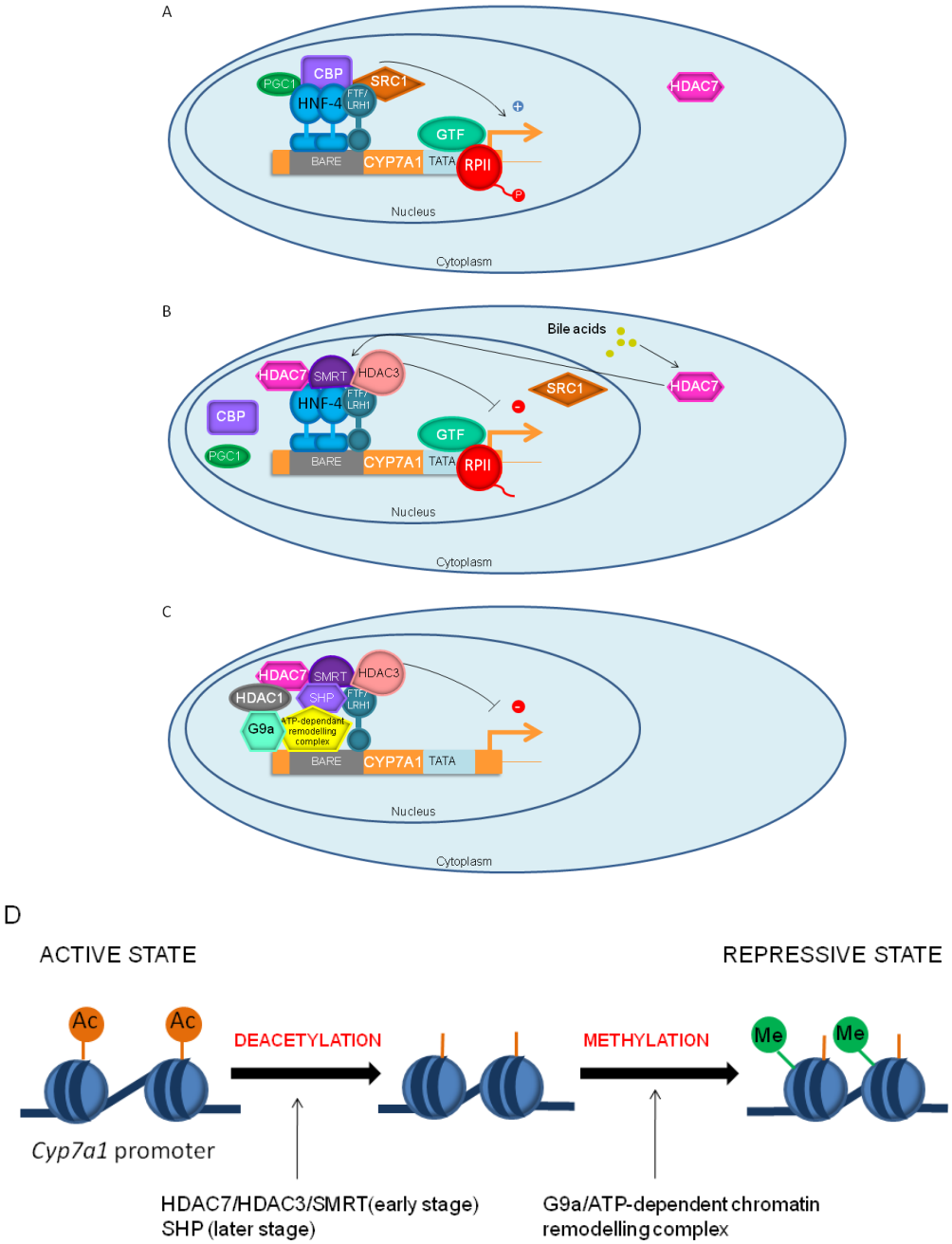


Figure 14. Regulation of *Cyp7a1* gene transcription by chromatin-remodelling enzymes. (A) When the level of bile acids is low, HDAC7 is localised in the cytoplasm, in this condition the activation of *cyp7a1* gene transcription is induced by the binding of HNF-4 and FTF/LRH1 to the BARE of *cyp7a1* promoter. This binding to BARE allows the recruitment of coactivators such as PGC1, CBP and SRC1, general transcription factors (GTF) and phosphorylated RNA polymerases are finally recruited to start transcription. (B) When the concentration of bile acid increases, HDAC7 shuttles to the nucleus and recruits a repressor complex composed of SMRT and HDAC3 onto the HNF4-bound to BARE. At the same time the coactivators are released and the RNA polymerase II is dephosphorylated stopping transcription elongation. (C and D) In the last stages the RNA polymerase dissociates from the promoter, SHP associates with FTF/LRH1, and the corepressor complex remains associated. Finally HDAC1 and an ATP-dependent chromatin remodelling complex are recruited to remodel the chromatin. A last step involves the histone methyltransferase G9a which methylates lysine 9 of histone H3, this methylation being a transcriptional inactive state mark.

6.4.3.1 Towards specific HDAC inhibitors

With the example of regulation of CYP7A1, an important target for lowering-cholesterol drugs, bile acids were shown as important signalling molecules. HDACs were shown to play possibly an important role for the repression of CYP7A1. Inhibiting their activity might provide a means towards lowering plasma cholesterol by increasing its conversion to bile acids. The inhibitors available so far are pan-HDAC or class-specific inhibitors and selective inhibitors for a specific HDAC do not exist. The design of specific HDACi will be of a great value to validate targets in cholesterol catabolism and to develop drugs showing less side effects, as HDACs are involved in many important and vital cellular processes.

For elucidating HDACi specificity it was necessary to obtain the structure of different HDACs with different inhibitors. To this end, it was originally the aim of this thesis to express, then solve the structure of the apo HDAC7 catalytic domain, then use this to study inhibitors designed by collaborators. But soon after the start of this work, two groups published structures of HDAC7 and HDAC4 bound to different inhibitors. In the next part these results will be reviewed.

6.4.3.2 Structural and functional analysis of HDAC7 and HDAC4

Before the release of the structures of HDAC4 and HDAC7 in the PDB (Bottomley et al. 2008; Schuetz et al. 2008). there were few HDAC structures available: that of human HDAC8 (Somoza et al. 2004; Vannini et al. 2004) and two bacterial HDAC-like enzymes HDLP from *Aquifex aeolicus* class I-related (Finnin et al. 1999) and the class IIb-related HDAH from *Bordotella alcaligenes* (Nielsen et al. 2005). The structures available indicate that the HDAC catalytic domain is conserved consisting of a single globular α/β domain with a central parallel β-sheet flanked by α-helices. The catalytic domain of HDAC7 consists of 21 α-helices and 10 β-strands in a single domain with the active site formed by a tube with the catalytic zinc at its bottom. A second zinc, observed only in class IIa, is also present in a flexible surface exposed region and might be involved in interaction with partners (Bottomley et al. 2008; Schuetz et al. 2008) (Fig. 15).

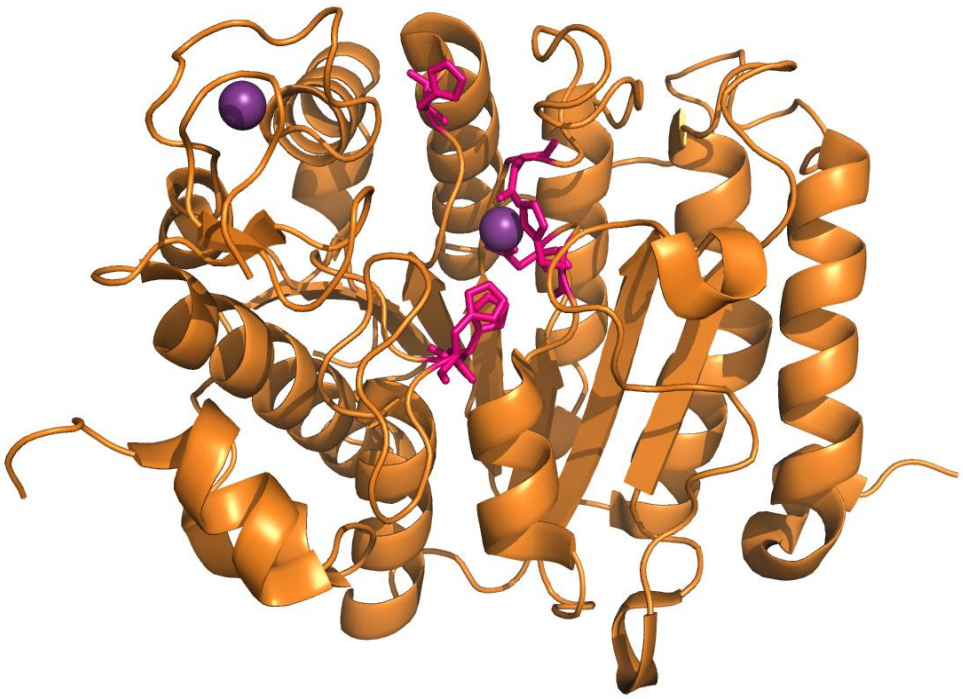


Figure 15. The HDAC7 catalytic domain. Zinc ions are shown as purple balls. The residues in pink represent the catalytic residues.

The substrates of HDAC4 and HDAC7 are not known but it seems that they are inactive on canonical acetyl lysine substrates (Lahm et al. 2007; Bottomley et al. 2008; Schuetz et al. 2008). A common feature of class IIa HDACs is that an important tyrosine involved in the catalysis of the deacetylation is replaced by a histidine which is pointing away from the catalytic pocket. Mutating this class IIa histidine to a tyrosine reintroduces the activity on canonical acetyl lysine peptides (Lahm et al. 2007; Bottomley et al. 2008; Schuetz et al. 2008). In this context one can ask whether class IIa HDACs even possess an enzymatic activity.

7 The interaction of HDACs with the SMRT corepressor

7.1 HDAC3-SMRT interaction

HDAC3 has no intrinsic DNA binding activity; this is mediated by other proteins that target HDAC3 to DNA or to histones. It was discovered that SMRT and NCoR could bind and stimulate HDAC3 enzymatic activity (Wen et al. 2000; Guenther et al. 2001; Zhang et al. 2002b). Interacting fragments of SMRT were identified by deletion mutagenesis and were tested for increased deacetylase activity. A fragment, from residues 267-549 showed high affinity to HDAC3 *in vitro* and significantly increased deacetylase

activity (other fragments that did not bind HDAC3 did not have this effect on HDAC3 activity). In a more precise analysis, a deacetylase-activating domain (DAD) was mapped from residue 395-489 of SMRT (aa 403 to 497 of NCoR) (Guenther et al. 2001) with an essential eight amino acid region (aa 420-427 of NCoR) identified as critical for the interaction (Zhang et al. 2002b). This DAD is sufficient for HDAC3 enzymatic activation and both the C and N termini of HDAC3 interact with the SMRT DAD. Both NCoR and SMRT, by activating HDAC3 activity, are dynamic regulating cofactors. The NMR structure of SMRT DAD (residue 412 to 480) revealed that the DAD comprised 4 α-helices (H0 to H3) (Codina et al. 2005).

7.2 HDAC7-SMRT interaction

HDAC7 does not bind directly to HDAC3 but recruits it via SMRT and NCoR (Fischle et al. 2001). Whereas SMRT activates HDAC3 deacetylase activity, such enhancement upon binding with SMRT has not been observed for class IIa HDACs (Kao et al. 2000; Fischle et al. 2001; Fischle et al. 2002), reappraising the hypothetic enzymatic activity of HDAC7.

SMRT (Silencing Mediator of Retinoid and Thyroid hormone receptors) recruits HDACs and represses transcription. SMRT is a very large protein comprising 2525 amino acids which binds hundreds of different proteins. SMRT can be divided into a N-terminal region having three repression domains RD I, II, and III; two conserved nuclear receptor interaction domains in the C-terminal region NDI and II and 2 SANT domains (Ordentlich et al. 1999; Jepsen and Rosenfeld 2002) (Fig. 16).

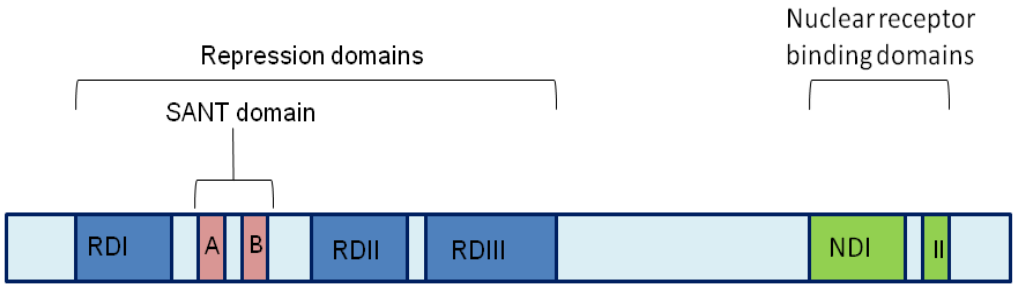


Figure 16. SMRT domains. SMRT is composed of three repression domains RDI, II and III and SANT domains A and B. In its C-terminal part there are two nuclear receptor binding domains. Class IIa HDACS were reported to interact with RDIII and HDAC3 (class I HDAC) interact with the SANT domain.

Class IIa HDACs interact via their catalytic domains with RDIII (Kao et al. 2000; Fischle et al. 2002). The repression domains also recruit additional proteins, therefore

SMRT can be viewed as a giant platform upon which partners assemble forming the corepressor complex assembly.

SMRT, with its 2525 amino acids, is a difficult protein to study from a structural point of view, not least because SMRT is predicted to be largely unstructured (Fig. 17). Nevertheless some SMRT fragment structures have been determined, such as for SMRT DAD and several complexes of proteins containing short SMRT-derived peptides (Codina et al. 2005).

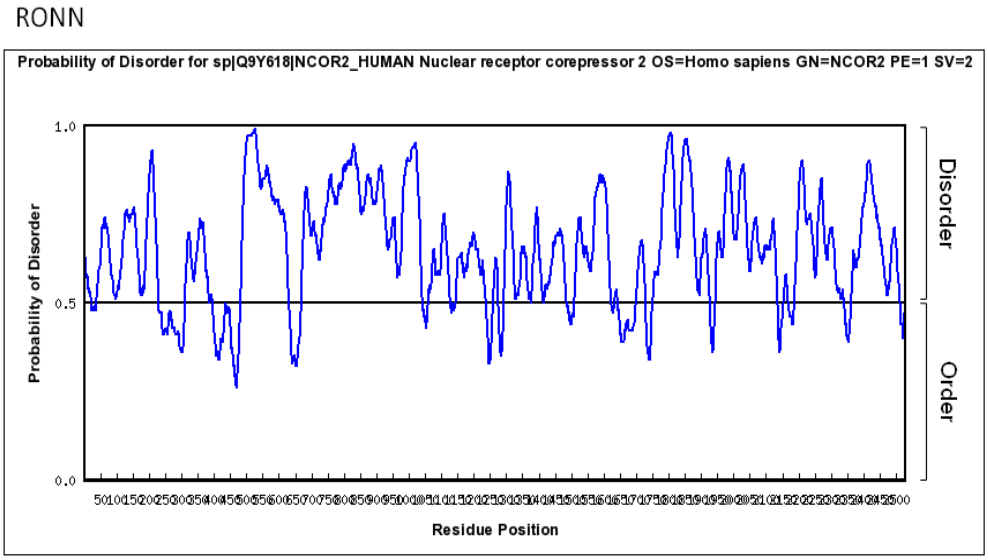


Figure 17. Predicted disorder in SMRT (also called NCoR2). RONN score was calculated on the full-length protein. A score above the threshold of 0.5 is considered disordered. In the case of SMRT the majority of probabilities are above 0.5 indicating that SMRT is mainly unstructured. Other disorder predictors were used and they all agreed for disorder propensity of SMRT (not shown).

Intrinsic disorder is mechanistically important in many cellular processes and found in all three kingdoms of life and will be briefly reviewed below due to its importance in this project regarding the SMRT corepressor-HDAC7 interaction.

7.3 Intrinsically disordered proteins and their biological relevance

7.3.1 A brief history of intrinsically disordered protein research

The structure-function paradigm was a dominant view of protein science for a long time, however experimental observations contradicted this concept when structural disorder was observed. The field of intrinsically disordered protein research accelerated from around the 2000s. Before that, people were reluctant to consider the possible functional importance of intrinsically disordered proteins since structure was so tightly

associated with function (Dunker et al. 2001). The structure-function paradigm started with the observation of stereospecificity in enzyme catalysis. This was explained by the "lock-and-key" model of Fisher and colleagues in 1894 for explaining the specificity of beer yeast which could hydrolyse α- but not β-glucosides. This was the first suggestion of a strict geometrical complementarity of enzyme and substrate, where the enzyme had to be structured correctly. This was subsequently confirmed by the observation that enzymes would lose their activity upon thermal, acidic or urea denaturation. Moreover Anfinsen observed that denatured ribonuclease (RNase) could be renatured *in vitro*, rescuing its enzymatic activity (Anfinsen 1973). Then determination by X-ray crystallography, with the first protein structures - myoglobin (Kendrew et al. 1958), haemoglobin (Perutz 1960; Perutz et al. 1960) and lysozyme (Blake et al. 1965) - reinforced the structure-function paradigm and was so powerful that alternative views withered.

However some exceptions to the lock-and-key model appeared that could not be explained entirely by this structure-function paradigm. Indeed Karush in 1950 showed that serum albumin (even though not an enzyme) had a universal capacity for binding with high-affinity to small hydrophobic molecules. This implied that the binding site of serum albumin had a large number of configurations in equilibrium, and that the best configuration was stabilised upon interaction with a given small molecule. Karush termed that phenomenon "conformational adaptability". Subsequently the paradigm was modified by Koshland. He stated that a substrate did not bind to a rigid active site but rather one that is structurally adapted to the shape of the substrate, this then enabling its catalytic activity (Koshland 1958). Pauling added that the active site had the flexibility to accommodate the highest-energy transition state intermediate (Pauling 1948).

A notable advancement occurred with studies on the microtubule-associated protein 2 (MAP2), a homologue of tau implicated in Alzheimer's disease. It stabilises the unstable neuronal microtubule polymer. MAP2 resisted crystallisation and was classified as intrinsically disordered protein (IDP) (Hernandez et al. 1986). MAP2 was revealed to be resistant to heat treatment, and sedimentation equilibrium and gel filtration analyses revealed that MAP2 was not globular but had an elongated shape and exhibited very little secondary structure by circular dichroism. Then Sigler generalised the concept of protein disorder: "whereas crystal structures at atomic resolution are crucial to our understanding of […] specific molecular interactions, we can imagine many assemblies […] whose function requires strong but less precisely defined arrangements than the ones we have seen crystallographically" (Sigler 1988)

A further breakthrough in the field was made through demonstrating the functionality of IDPs by Wright and colleagues. They studied the cyclin dependent kinase (Cdk) inhibitor $p21^{Cip1}$, a protein important for the p53-dependent control of the cell cycle (Kriwacki et al. 1996). They showed that it had few secondary structure elements in the unbound state and that it had a stable conformational state upon binding to its partner Cdk2. This induced folding allowed $p21^{Cip1}$ to bind and inhibit a diverse family of cyclin-Cdk complexes.

With the increased amount of evidence of IDP functional importance, Romero and colleagues applied disorder predictors (based upon amino acid propensity) to the Swiss-Prot database and showed that more than 15000 proteins contained disordered regions of at least 40 amino acids, emphasising that structural disorder is a general phenomenon (Romero et al. 1998). It was as well suggested that disordered regions might be involved in binding to other molecules (Dunker et al. 1998). Finally amino acid composition analyses of IDPs provided additional information of their features (Uversky et al. 2000), demonstrating in particular that IDPs are localised in a unique region of the net charge-mean hydrophobicity phase space combining low overall hydrophobicity and large net charge responsible for their inability to fold into well-defined structures.

7.3.2 IDPs and biological processes

A number of studies have been performed to analyse biological processes and molecular functions associated with protein disorder. Ward et al used the DISOPRED algorithm to estimate the frequency of proteins with intrinsically disordered regions with more than 30 consecutive residues in 24 genomes from the three kingdoms of life (Ward et al. 2004). Tompa classified the proteome (using IUPred and PONDR VLS1) of *E. coli* and yeast based upon the percentage of disordered residues, the percentage of IDPs and of proteins with intrinsically disordered regions of more than 30 consecutive residues (Tompa et al. 2006). Finally Dunker classified disorder In SwissProt entries focusing on statistically significant functional annotation keywords (238 out of 710 functional keywords) (Xie et al. 2007). Taken together, these studies allowed classification of IDPs into five main functional classes:

- Transcription and its regulation, notably molecular functionalities such as DNA and nucleotide binding
- Signal transduction and regulation of cell cycle

- Biogenesis and functioning of cellular entities containing nucleic acids such as the ribosome and chromatin
- Processing of mRNA (including splicing)
- Organisation and biogenesis of the cytoskeleton

Ultimately it may be possible to link patterns of disorder in a protein sequence with protein function (Lobley et al. 2007). It should be noted however that the actual datasets reflect general trends rather than real correlations. To establish real correlations more predictive knowledge will be required. Nevertheless, from what is known so far it seems that disorder in the interior part of a protein is often associated with DNA-binding transcription factors. C-terminal disorder is predominant in transcription factor activators and repressors, whereas N-terminal disorder seems to be a common feature in ion-channels.

7.3.3 IDPs and diseases

Some proteins involved in diseases naturally contain a high level of disorder. This is the case for example in cancer-related proteins for which a high level of disorder is predicted, demonstrated by the comparison of four datasets: (i) human-cancer associated proteins (HCAP), (ii) signalling proteins collected by the Alliance for Cellular Signalling (AfCS), (iii) eukaryotic proteins from SwissProt and (iv) non-homologous protein segments with well-defined three-dimensional structures (Iakoucheva et al. 2002). Predictions also revealed a high frequency of structural disorder in proteins involved in CVDs. CVD-related proteins have a high proportion of disorder with 57 ± 4% containing at least one intrinsically disordered region of more than 30 consecutive residues, showing similar levels to signalling proteins (66 ± 6%) and higher levels than for eukaryotic proteins. This high proportion of disorder is observed as well for diabetes and autoimmune diseases (Cheng et al. 2006a). Several neurodegenerative diseases are caused by the deposition of insoluble protein aggregates known as amyloids. These diseases involve large conformational (misfolding) changes of some key proteins (Chiti and Dobson 2006).

Some IDPs can be considered as drug targets given their role in many biological processes and diseases. Indeed protein-protein interactions involving one structured (ordered) partner and one disordered partner have many features that make them potentially druggable: (i) they interact weakly which allows tighter competitive binding

by small molecule drugs to the structured partner, (ii) the protein-protein interface is not large and flat as observed for ordered protein-protein interfaces, but rather curved, (iii) the protein-protein interface is small, compact and most often involves helical structures which may facilitate mimicry with small molecules. For example, several successful peptide inhibitors derived from p53 that bind MDM2 have been designed and shown to be functional in assays (Wasylyk et al. 1999; Chene et al. 2000).

Thus by studying and identifying the interaction between HDAC7 and a fragment of SMRT it might be possible to decipher the structural binding interface and to design small molecules preventing its association with SMRT, as a means for finding new drugs for treating HDAC7-associated diseases and Class IIa HDAC-associated diseases in general.

8 Objectives of the project

At the start of this project, there were no structures of class IIa HDACs including HDAC7. Their enzymatic deacetylase activity had not been well assessed on purified proteins, only samples prepared from extracts that were potentially contaminated with additional activities. Moreover, despite reasonable numbers of available HDAC inhibitors (pan-HDAC or class-specific inhibitors), some of them approved by the FDA and already on the market, their mode of action was not well understood. Neither were there HDAC-specific inhibitors that could be used as drugs or tools for cell biology. Many questions were unanswered: what is the structure of HDAC7? Is it similar to the structures already published (HDAC8, and bacterial HDAC-like deacetylases)? What are the structural features underlying HDAC inhibitor specificity, and is it possible to design HDAC7-specific inhibitors? Does HDAC7 show a deacetylase activity on acetylated histones and histone-derived peptides?

To begin to address those questions we focused on the HDAC7 catalytic domain expression, purification and crystallisation. However the structures of HDAC7 apoenzyme and with two different inhibitors bound were soon deposited in the PDB by the Structural Genomics Consortium with an accompanying paper nearly one year later (Schuetz et al. 2008). The project then moved towards reproduction of HDAC7 crystals and determination of the structure of HDAC7 with inhibitors designed and synthesised by consortium collaborators. Chapter 2 will present the work of HDAC7 expression, purification and crystallisation trials. It shows how, despite the published information, HDAC7 is a difficult target to work on and significant further efforts were necessary to

improve expression and purification of HDAC7 to generate material for crystallisation, biochemistry and biophysical analyses. Chapter 3 describes compound profiling experiments in which the deacetylase activity of purified HDAC7 was profiled with two different synthetic substrates that became available on the market. The binding and the activity of one promising inhibitor molecule designed by our collaborators is described together with crystallisation attempts of the protein-small molecule complex. Separately, we investigated an alternative strategy for targeting HDAC7 specifically. Chapter 4 describes this strategy where the interest was on the binding partners of HDAC7 and especially the interaction between HDAC7 and SMRT (SMRT being a crucial partner for HDAC7 biological function). The aim was to identify a HDAC7-binding SMRT fragment, and to characterise the binding interaction as a possible approach for targeting HDAC7 activity.

In summary this thesis aims to show the work performed in order to understand and target HDAC7 activity.

The project was thus divided into three main objectives:

- Improving the HDAC7 construct (expression and solubility)

- To understand and improve the specificity of HDAC7 inhibitors by assessing new compounds synthesised within the SOUTH consortium
    - by reproducing crystals obtained by Schuetz in 2008
    - by characterising the binding of HDAC7 with inhibitors synthesised within the consortium using structural, biochemical and biophysical methods.

- To investigate a new strategy for HDAC inhibition by exploring the SMRT/HDAC7 interaction
    - by identifying the SMRT fragment interacting with HDAC7
    - by characterising structurally the interaction between the identified SMRT fragments and HDAC7
    - by functional investigation of this interaction and its effect on HDAC7 activity.

CHAPTER 2

HDAC7: Construct Engineering, Protein Expression and Purification

RESUME

Pour étudier de nouveaux inhibiteurs de HDAC7, il était d'abord nécessaire d'obtenir HDAC7 en quantité et qualité suffisantes et reproductibles, compatibles avec la mise en place d'un système de cristallisation permettant une analyse de nouveaux inhibiteurs en haut-débit. Des tests préliminaires d'expression et de purification ont été effectués en utilisant différentes constructions de HDAC7, différentes souches d'expression bactérienne, différentes températures d'induction, et en utilisant un gène synthétique de HDAC7 pour augmenter les rendements. A la suite de ces tests, HDAC7 s'est révélée être une protéine faiblement exprimée. Entre temps le Consortium de Génomique Structurale de Toronto a résolu la structure de HDAC7. Le projet de thèse s'est donc réorienté pour reproduire les conditions de purification et de cristallisation publiées. En utilisant les conditions publiées il s'est avéré que les rendements obtenus étaient trop faibles pour envisager le système de cristallisation. Les auteurs de la publication de la structure de HDAC7 ont confirmé qu'ils rencontraient également les mêmes difficultés et que tout comme nous ils pouvaient espérer, dans le meilleur des cas, quelques milligrammes de HDAC7 après purification provenant de culture de 50 l. Dès lors l'accent a été mis sur l'amélioration des conditions de purification. Ainsi l'étiquette hexahistidine en N-terminale a été échangée par une étiquette Strep II et une étape de précipitation du lysat par du sulfate d'ammonium a été ajoutée, permettant d'obtenir 5-8 mg de HDAC7 dans les meilleures semaines. Cependant la protéine présentait un profile d'élution hétérogène par chromatographie échangeuse d'ions et ne cristallisait pas mais était active. Une autre étape d'ingénierie a été effectuée sur HDAC7 afin de trouver des constructions alternatives. Pour cela la technologie ESPRIT a été utilisée. Elle se caractérise par trois étapes principales : la création d'une librairie non biaisée de constructions d'ADN de HDAC7 par troncation en 5' par l'exonuclease III, l'expression et la sélection de clones exprimant des constructions soluble révélées par biotinylation *in vivo* du peptide accepteur de biotine, positionné à l'extrémité C-terminal des constructions. De façon surprenante, à la suite de la sélection des clones solubles, la construction utilisée pour résoudre la structure de HDAC7 (faiblement exprimée) avait également été sélectionnée parmi les meilleurs clones. Des délétions C-terminales du peptide accepteur de biotine ont révélé que 6 acides aminés suffisaient à rétablir de meilleures conditions d'expressions et de solubilité. Cette construction a été utilisée pour

des essais de purification qui ont révélé que HDAC7 était à présent homogène (par chromatographie échangeuse d'ions). Malheureusement cette protéine résistait encore à la cristallisation mais était active, c'est donc avec cette construction que les travaux présentés dans le reste du manuscrit ont été effectués. En parallèle une dernière étape d'ingénierie a été effectuées : différentes délétions des 32 acides aminés invisibles dans la structure de HDAC7 (mais présents *a priori* lors des étapes de purification) ont été réalisées et testées pour leur expression et rendement après purification. Une construction a d'abord été choisie pour sa bonne expression et purification. Des essais de cristallisation sont en cours avec cette construction et était toujours en cours au moment de rédaction du manuscrit de thèse.

1 Introduction

1.1 The challenges of protein expression for structural biology

The field of structural genomics has highlighted the bottlenecks of the structure solution process. Probably the most challenging is obtaining highly expressed and soluble proteins in a form that crystallises. Expression and solubility levels of a recombinant protein are dependent on their amino acid sequences, these amino acids influencing the intrinsic physical properties of the proteins and how the cell processes them, yet predicting the solubility and crystallisability of a protein from its amino acid sequence is currently impossible. Even when conditions are published, reproducibility, especially of the crystallisation steps, can problematic as will be seen with HDAC7 (results and discussion).

In structural biology, the most common method of obtaining soluble protein is to design a small number of constructs rationally based upon knowledge of the target protein (literature, biochemistry, comparison of homologues). However in many cases, these strategies are highly time-consuming due to multiple iterations of cloning and testing, but still fail. For novel proteins with no homologues and little published functional information, the failure rate is very high. Commonly targets are abandoned if expression and/or crystallisation trials are unsuccessful.

The SMRT corepressor is not only novel by the above criteria, but has the additional challenge of being largely unstructured. Such a protein would not normally be considered for structural studies (chapter 4). In contrast, HDAC7 would appear to be very well characterised since the catalytic domain is now understood at atomic resolution by X-ray crystallography. But problems in expressing material in useful quantities and unreliable crystallisation required a large amount of construct engineering using novel methods to be performed and this chapter describes this work. Notably, a directed evolution type method was used to define better expressing constructs of both HDAC7 and SMRT.

1.2 Random library methods in protein engineering

It is often extremely difficult to design well-behaving constructs of a target protein. To circumvent this, several groups have developed methods to generate all possible genetic constructs of a target protein and screen them for the desired phenotype, *e.g.* solubility (Magliery and Regan 2004; Hart and Tarendeau 2006; Prodromou et al. 2007;

Savva et al. 2007; Dyson et al. 2008). This strategy is based on the protein engineering approach of directed evolution (Tobin et al. 1999), mimicking natural evolution in a test tube by producing large numbers of variants by random mutation and subsequently selecting the best adapted variants. There are several strategies for creating random diversity by mutagenesis methods and these are generic and applicable to any gene without prior knowledge of its biological characteristics.

For enzyme engineering, the common application of directed evolution, a genetic diversity of a target is usually generated by random point mutagenesis. In error-prone PCR the target gene is amplified by a DNA polymerase in conditions where the replication fidelity is low, introducing mutations in the new copies. DNA-shuffling can be performed on a pool of PCR products (containing mutations) digested by DNAse I that cuts DNA non-specifically. The generated fragments are then reassembled by PCR simulating the natural process of recombination and allowing different lineages of mutations to mix. DNA shuffling resulted in many successes e.g. improving the fluorescence of GFP (Crameri et al. 1996; Crameri et al. 1998) or adaptation of mammalian paraoxonases PON1 and PON3 for expression in *E. coli* (Aharoni et al. 2004).

Although well suited for changing the properties of single enzymes, point mutagenesis is not appropriate for defining subconstructs in larger genes. For this, gene fragmentation libraries can be generated using DNAse I (Anderson 1981), physical DNA breakage by sonication (Deininger 1983) and DNA shearing (Oefner et al. 1996; Thorstenson et al. 1998). Additionally, and as used in this work, libraries can be generated by enzymatic digest of the full-length construct. Unidirectional or bidirectional random truncations can be performed using exonuclease III in combination with mung bean nuclease or nuclease S1 to remove generated single strands DNA (Henikoff 1984; Ostermeier and Lutz 2003).

With the aim of generating random truncation libraries and screening them for soluble proteins, the ESPRIT technology was designed. This technology creates unidirectional or bidirectional truncation libraries using exonuclease III and has led to a number of successes on difficult targets including CagA from *Helicobacter pylori* (Angelini et al. 2009), influenza polymerase PB2 subunit (Tarendeau et al. 2007; Guilligay et al. 2008; Tarendeau et al. 2008), NFκB (Yumerefendi et al. 2010), herpesvirus terminase (Nadal et al. 2010).

1.3 ESPRIT: An overview of the technology used for HDAC7 and SMRT expression

The ESPRIT technology employs a strategy for generating DNA fragment libraries by exonuclease III truncation: this digest creates nested deletion of genes (Henikoff 1984; Ostermeier and Lutz 2003). These are then screened for soluble protein expression (Fig. 1).

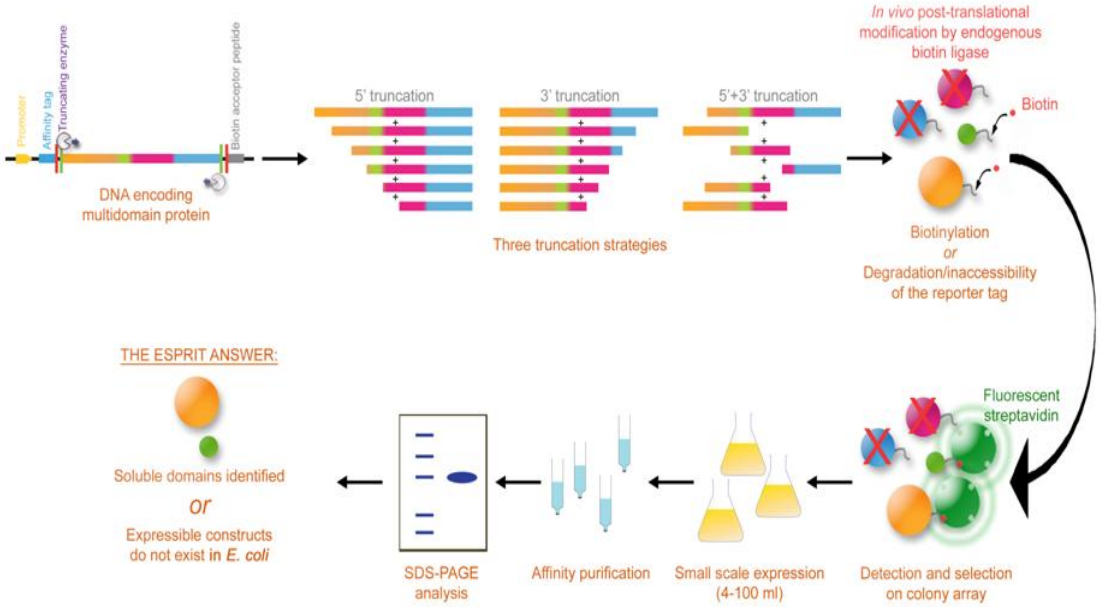


Figure 1. Overview of ESPRIT technology. A random truncation library of a target protein is generated. The constructs are screened for solubility through *in vivo* biotinylation (reporter of solubility). In a final steps, selected hits are confirmed by purification and SDS-PAGE analysis. (Reproduced with permission of Dr F. Tarendeau)

The target gene construct is cloned into a vector designed for truncation where restriction sites create 5' exonuclease-sensitive and 3' exonuclease-resistant overhangs. In addition, the vector contains an N-terminal hexahistidine tag for purification purpose and a C-terminal biotin acceptor peptide (BAP) for solubility assessment. The target gene is digested in a time-course exonuclease III reaction where small aliquots of reaction solution are removed each minute and quenched in a highly concentrated salt solution (2M NaCl) at 70°C. This digestion generates single strand overhangs which are removed by incubating the DNA with mung bean nuclease and are further polished by a proofreading DNA polymerase (e.g. *Pfu*). The vector is then recircularised with T4 DNA ligase and *E.coli* transformed with the library (Fig. 2).

The vector used for truncation allows three types of truncations: 5' or 3' unidirectional and bidirectional truncations (Fig. 1). In unidirectional truncations one end

of the construct is fixed and in frame with the C-terminal BAP or the N-terminal hexahistidine tag, the other end being truncated. Thanks to this technique, all the possible genetic variants are generated. When using unidirectional truncation, only one-third of the constructs are in frame after ligation with a library diversity of N (N being the gene length in base pairs). Bidirectional truncation focuses on seeking for internal gene fragments expressing solubly. This bidirectional truncation is completed using 2 steps of unidirectional truncations: one end is randomly truncated and recircularised, this vector product is then subjected to another unidirectional truncation from its other end and recircularised (Fig. 2). This two-step exonuclease protocol maintains the proper orientation of the DNA insert compared to gene fragmentation strategies for example where only 50 % of inserts are in the right orientation. The theoretical generated diversity is much higher than in unidirectional truncation libraries with a large theoretical diversity of N(N+1)/2 (Prodromou et al. 2007). This level of diversity cannot be screened in its entirety in the ESPRIT screening format which is limited to 28 000 constructs. Therefore, to both reduce the number of clones to be screened and to improve sampling efficiency, the library is first size-fractionated on agarose gel according to the size of protein domain being targeted.

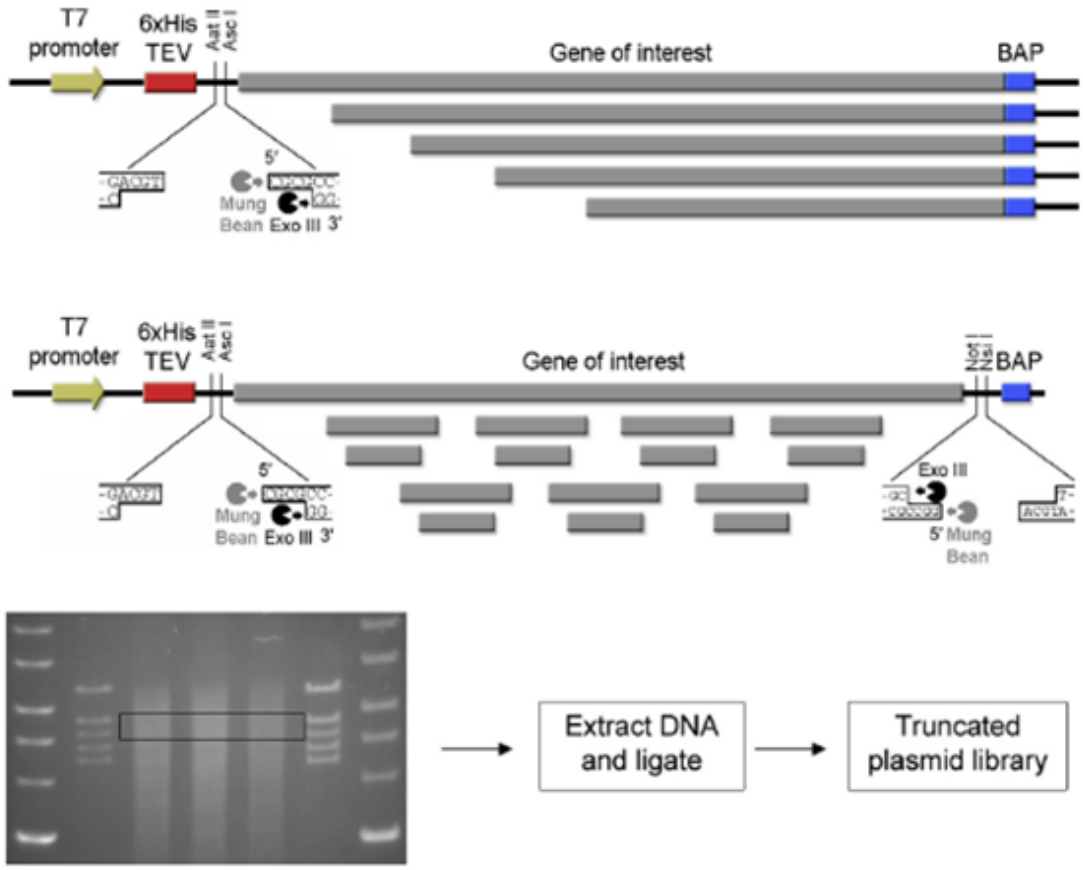


Figure 2. Different ESPRIT strategies. A) Unidirectional truncation used when it is suspected that one of the ends is compatible with solubility. B) Bidirectional truncation to access internal soluble domains. The library is size-fractionated by agarose gel to extract the band of interest and generate a library made of recircularised truncated plasmids. (Adapted from Yumerefendi *et al* 2010)

Assessing all the diversity generated in ESPRIT libraries by direct solubility screen using centrifugation or filtration is impractical. To address this difficulty of handling huge libraries, an indirect solubility assay was developed using *in vivo* post-translational modification of a short BAP. This *in vivo* post-translational modification occurs commonly in cells were cytoplasmic biotin ligase BirA enzyme adds the biotin cofactor covalently to the biotin carboxyl carrier protein (BCCP) (Chapman-Smith and Cronan 1999). Even though BCCP has been used as a fusion reporter, the fact it is a globular protein may influence the solubility of the target by a passenger solubilisation effect. For this reason BCCP was replaced by a synthetic 14 amino acid peptide, named BAP, mimicking the BCCP recognition site (Schatz 1993; Beckett et al. 1999). The BAP is used widely for protein immobilisation on surfaces and is efficiently labelled by endogenous biotin ligase in both prokaryotes and eukaryotes (Duffy et al. 1998).

Moreover this peptide can be removed when required as in crystallisation experiments with little effect on the target solubility.

In the first step of the ESPRIT screen clones are printed as high density macroarrays on nitrocellulose membranes resulting in up to 56 000 geometrically arrayed colonies (28 000 colonies in duplicates). The colonies are lysed *in situ* under denaturing conditions (Bussow et al 1998) to perform colony blots that are then probed with fluorophores. Fluorescent streptavidin is used to detect biotinylation of the C-terminal BAP, reporting on the potential solubility of the fragment. In parallel, the N-terminal hexahistidine tag is probed by a fluorescent monoclonal antibody (Fig. 3). Finally constructs showing both signals (i.e. intact and non-proteolysed), are ranked based upon their biotinylation signal intensity. Typically the first 96 best clones are selected for expression trials in a small-scale format in 24 deep-well plates, where they are then lysed and purified through their histidine tag on Ni-NTA affinity resin.

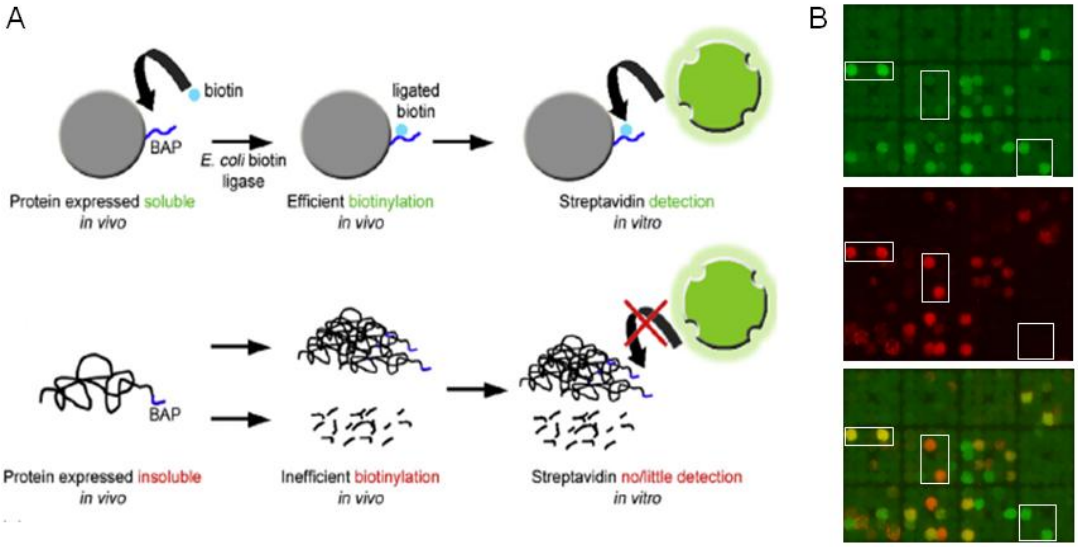


Figure 3. Principle of ESPRIT solubility screen. A) When a protein is expressed solubly, the endogenous *E. coli* biotin ligase biotinylates the biotin acceptor peptide of the protein. In case of degradation or aggregation, biotinylation does not occur or is less efficient. B) Expression of biotinylated proteins is assessed on colony macroarrays using fluorescent conjugates. (Adapted from Yumerefendi *et al* 2010).

The presence of purified (i.e. soluble) protein from the selected clones is checked by Coomassie stained SDS-PAGE. This is also a means to check the behaviour of the selected clones (expression level, presence of degradation, etc). In many cases the best clones identified are only partially soluble, however they often yield sufficient amount of material to allow further investigations such as crystallisation assay (Nadal et al. 2010), and biophysical characterisation (Rawlings et al. 2010). Aggregation should result in

reduced biotinylation signal due to hiding of the tag in the inclusion bodies. Nevertheless in some cases, strong insoluble aggregates (inclusion bodies) still show some biotinylation signal probably due to slow aggregation rates, or due to biotinylation of surface exposed BAPs. Therefore the colony array analysis is really an enrichment step and a second step of purification is key for distinguishing fully or partially soluble constructs from insoluble ones.

The third aspect of the technology, in addition to the library synthesis method and use of a non-perturbing solubility reporter, are two different types of robotics: (i) a colony picking robot handles clones individually and allows clonal separation of library members. The arraying function of the same instrument grids them out on membranes permitting easy tracking, and the rearraying function permits the final 96 selected clones to be extracted from the master library. Finally the replication function makes copies of the whole library (allowing repetition of experiments in different conditions of expression for example), (ii) a liquid handling robot is used for automated protein purification. Without these robotics it would be impractical to handle thousands of clones (Fig. 4).

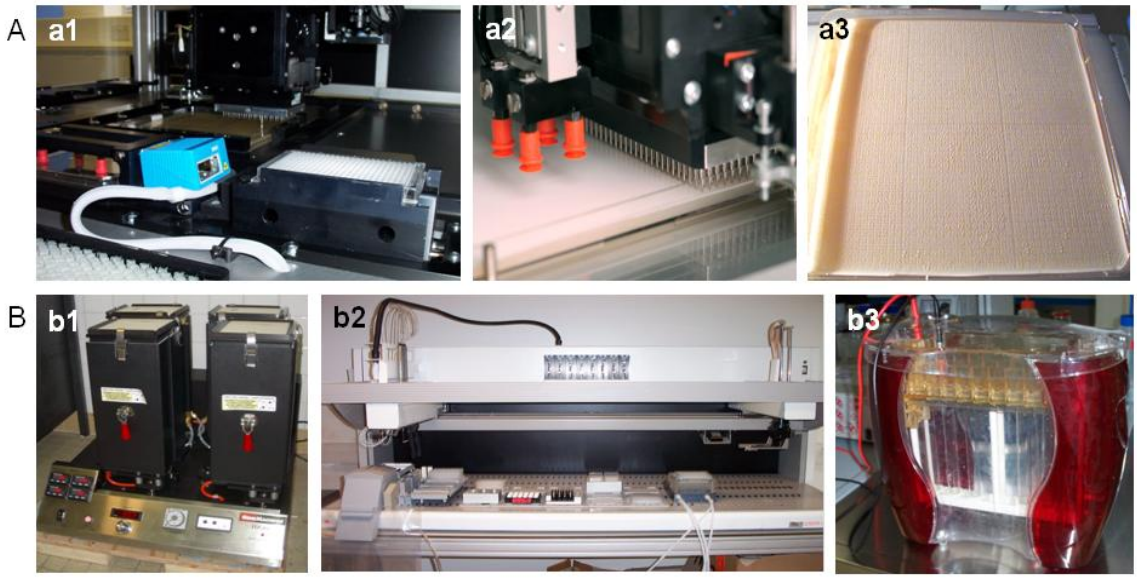


Figure 4. High-throughput automation. A) picker gridder robot (a1) is used for picking and isolating colonies (a2) to print them on macroarrays (a2, a3). B) Cells are grown using plate incubators (b1). Protein constructs are purified on a liquid handling robot (b2) and then analysed by SDS-PAGE thanks to an electrophoresis system allowing high-throughput processing of 128 samples in parallel.

2 Results and discussion

2.1 Initial expression trials

The first aim of the project was to screen inhibitors provided by members of the SOUTH consortium, to better understand and elucidate HDAC inhibitor specificity. For that purpose we wanted to create a crystallisation system to increase the speed of crystal screening and structure determination, where we could produce HDAC7 in large amounts either as crystals for soaking with new inhibitors, or as purified protein allowing co-crystallisation. Unlike functional studies, structural projects are protein-intensive; several milligrams of proteins are required to carry out one experiment. The first stage of the project was to get sufficient amounts of purified HDAC7 and reproduce recently published crystal conditions.

For this purpose small-scale expression trials were performed to define the best conditions for obtaining HDAC7 at sufficient amounts. Four alternative catalytic domain constructs were designed: (i) R483-S903, (ii) T515-S903, (iii) R483-L952 and (iv) T515-L952. These were cloned into a pET9 derived vector containing a N-terminal hexahistidine tag cleavable with hexahistidine tagged Tobacco Etch Virus (TEV) protease. They were expressed in three different *E. coli* expression strains: (i) BL21 (DE3) RIL, (ii) BL21 AI RIL and (iii) Rosetta 2 (DE3) pLysS, at four different temperatures (15°C, 20°C, 25°C and 30°C) of induction. Unfortunately no well expressing construct was apparent (data not shown) with levels being very low at 0.2 mg/l. It became clear that construct engineering would be necessary to improve the expression levels.

2.2 Expression and purification of HDAC7

Near the beginning of the project the structure of the HDAC7 catalytic domain (residues 483-903) was released in the PDB by the Structural Genomics Consortium (SGC). There was no accompanying publication; this came several months after deposition, but it was clear that HDAC7 might be expressed and purified in reasonable amounts compatible with structural studies. Our focus was then directed towards this construct, first using the conditions published on the SGC website, then using the later published conditions comprising affinity purification using Ni-NTA resin, anion-exchange, TEV cleavage, hexahistidine tag and TEV removal and gel filtration. Yields

were low at <0.3 mg/l of culture (confirmed to us by the SGC although not detailed in their publication). The low quantity of protein was apparent from SDS-PAGE where it only detectable by western blot. A minimum of 50 l of culture were grown and induced to get a few mgs of protein and after several attempts, it was concluded that the purification as published was too inefficient without large fermentation facilities.

It was then decided to try and improve the purification step by improving the first affinity purification that was not stringent enough to get rid of the large amount of *E. coli* contaminants (even after bead resin volume, washing steps optimisations etc). Consequentially, additional column steps were necessary leading to large loss of material. The N-terminal hexahistidine tag was replaced by a Strep tag II subcloned into the same pET9-derived vector. This tag is composed of 8 amino acids (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) and binds to the biotin binding pocket of streptavidin. The purification is based on the strength of the biotin-streptavidin interaction. The streptavidin resin has been optimised as streptactin resin allowing a good recovery of purified proteins following elution with free biotin (Schmidt and Skerra 2007). Due to the specificity and strength of the binding of the Strep tag II, purifications lack the non specific binding of cellular contaminants observed when the hexahistidine tag is used for purification.

The Strep-HDAC7cd was expressed in a Rosetta 2 pLysS strain in 50 l of TB medium. After lysis of the cells (about 1,2 l) the clarified lysate was precipitated with ammonium sulphate to concentrate it to 300 ml. The protein was purified to near homogeneity in 4 steps: the protein was subjected to an initial affinity purification step using streptactin resin. This step enriched and purified HDAC7 in reasonable amounts. The protein was further purified by Q-sepharose anion-exchange. The protein eluted in multiple peaks upon salt gradient (Fig. 5A).

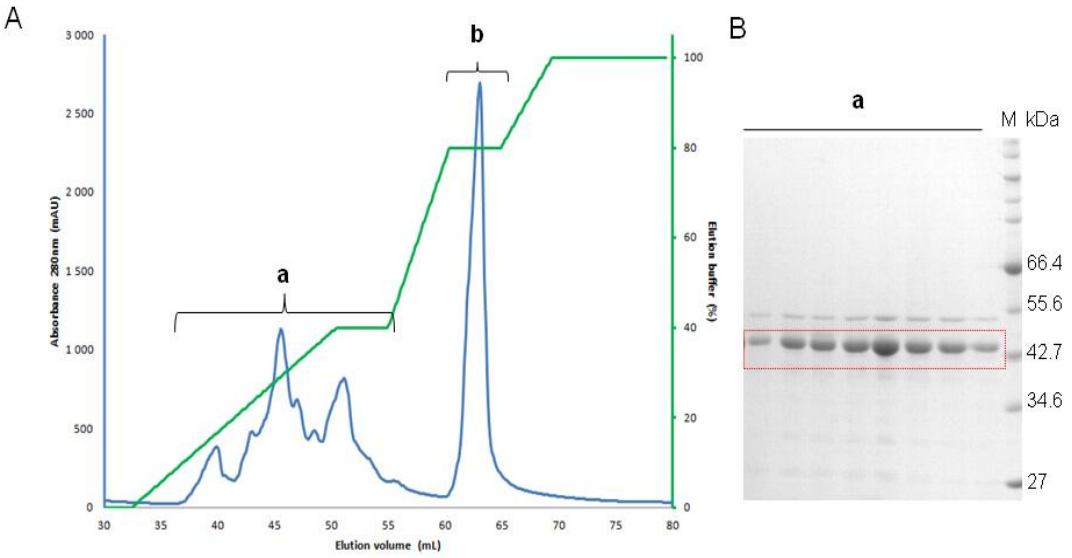


Figure 5. Purification of HDAC7 catalytic domain. A) the profile of HDAC7 elution after Q-sepharose anion-exchange chromatography. HDAC7 eluted in multiple peaks shown in region a. No protein was detected in peak b. B) SDS-PAGE with HDAC7 eluted from gel filtration chromatography as a monomer at about 43 kDa. A minor contaminant with a slightly higher molecular weight coeluted with HDAC7.

This elution profile suggested that multiple conformational species of HDAC7 were present. Nevertheless the fractions containing most of HDAC7 were pooled and incubated with TEV protease. The tag was removed by a second step of affinity purification. Finally, in the last step, size exclusion chromatography Superdex 75 16/60 indicated that HDAC7 was monomeric (Fig. 5B), but revealed as well the presence of a higher molecular weight contaminant (not removable). The yields of protein ranged from 5 to 8 mg of purified proteins from 50 l of culture. Despite the presence of this contaminant the protein was sent for crystallisation at the high-throughput crystallisation facility of EMBL Grenoble. Several trials were performed either robotically or manually reproducing published data (10-15% PEG 3350, 0,1M HEPES-NaOH pH7,5, 10% isopropanol and 10 mM DTT), but no crystals were obtained.

Despite obviously having been performed successfully by the SGC, reproduction of crystals from their construct and protocols proved impossible in our hands. Moreover, the purification protocol as optimised with the Strep tag II was long and difficult to perform with the basic fermentation equipment available in the institute, thus preventing the project from reaching high-yield scale required for screening all the inhibitors synthesised by collaborators. It was therefore decided to try and improve expression and purification of HDAC7 using the ESPRIT (Expression of Soluble Protein by Random Incremental Truncation) technology applied to HDAC7. This technology was developed by Dr Darren

Hart and Dr Franck Tarendeau (former post-doctoral fellow in the lab) and the work on HDAC7 was performed with the help of Alessandro Angelini a post-doctoral fellow. This technology allows testing of protein constructs for solubility in a high-throughput manner.

2.3 ESPRIT technology applied to HDAC7

HDAC7 had proven to be a problematic protein exhibiting weak expression, low yields, possible conformational heterogeneity and difficulties in reproducing crystallisation conditions. For these reasons, improving HDAC7 expression yield and solubility was crucial. For this purpose better behaving soluble fragments were sought using the ESPRIT technology.

2.3.1 Random truncation library

Two different constructs were used to perform the truncation libraries. They comprised the core HDAC7 catalytic domain starting with the same N-terminal residue but ending with 2 different C-terminal extensions (these constructs being longer than the construct used to solve the structure of HDAC7 by Schuetz (Schuetz et al. 2008). One construct started from Leu449 until Leu952 (construct LLWE/ PMNL), the second construct comprised residues from Leu449 until Ser903 (construct LLWE/RLAS) (Fig. 6). The aim was thus to truncate randomly the N terminus (disordered in the published crystal structure) whilst keeping The C-terminus both as published (Ser903) and full-length (Leu952). These constructs were obtained performing PCRs from a synthetic *E. coli* codon-optimised gene of HDAC7 catalytic domain (from Leu449 until Leu952) and cloned into a pET9 derived vector containing a TEV cleavable N-terminal hexahistidine tag and a C-terminal biotin acceptor peptide BAP. A random unidirectional 5' truncation was performed on the linearised vector containing the gene (generating all N-terminal deletions of HDAC7 protein) with the 3' end remaining fixed. The library was size fractionated on agarose to keep DNA constructs within a range of truncation of about 240 base pairs (about 80 amino acids deletion range at the N terminus) ensuring that the construct would yield protein in the range of 35 to 45 kDa. All the generated constructs were in frame with the C-terminal BAP whilst 33% would be expected to be in frame with the N-terminal hexahistidine tag after religation of the linearised vector. A colony PCR screen was performed on some clones selected out of the library to assess the efficiency of the truncation and the size distribution of the DNA constructs. Colony PCR

indicated an even distribution of constructs with no obvious bias; this was then confirmed by sequencing (Fig. 7).

Figure 6. Different HDAC7 constructs used for constructing the ESPRIT library. Two different constructs were designed both starting with the same N-terminal residue (Leu449) with two different ends, Ser903 and Leu952.

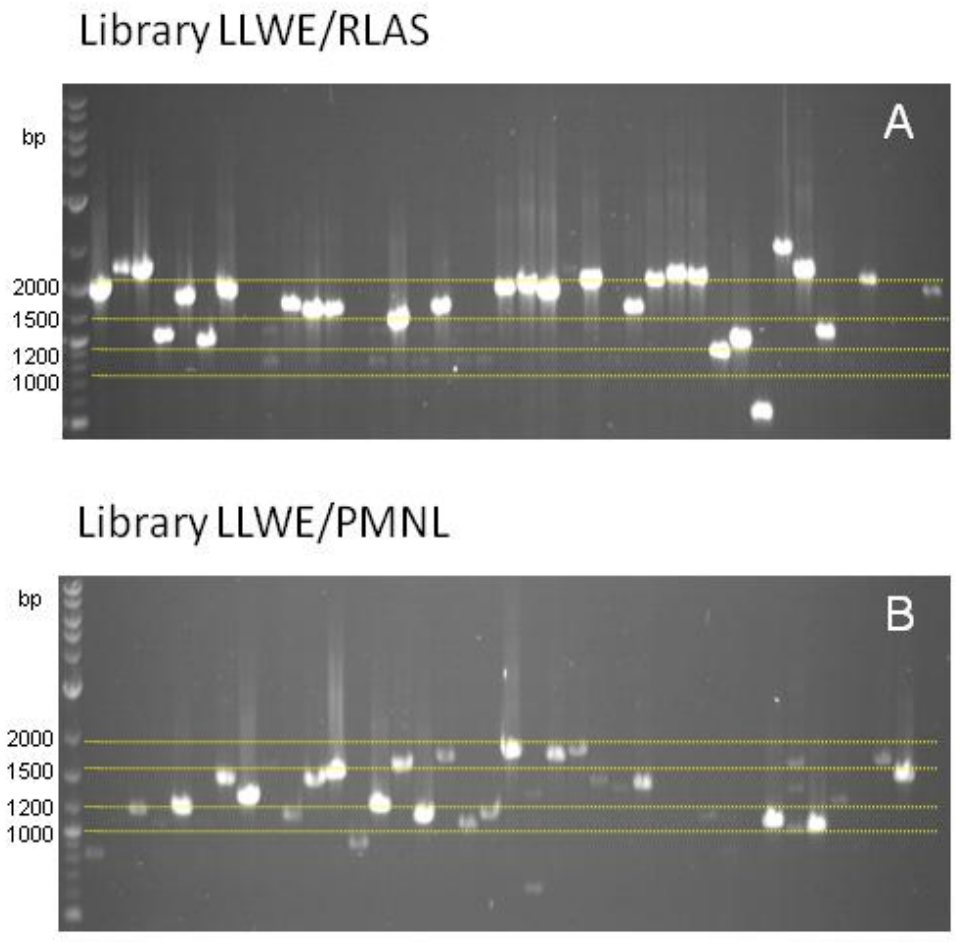


Figure 7. Assessment of HDAC7 5' truncation library quality.. Most of the constructs rank between 1000 and 2000 bp corresponding to protein size comprised between 300 and 600 amino acids. A. 23/40 tested clones were in this range representing 57,5 % of efficiency, B. 22/40 clones were comprised in the same range size with 55% of efficiency.

58% of DNA inserts for library LLWE/RLAS were located in the desired size range 1000 bp to 2000 bp corresponding to translation products between about 300 amino acids and 600 amino acids. 55% of DNA inserts of library LLWE/PMNL were located in this range. These efficiencies allowed calculation of the construct oversample factor when 9216 clones from each library were tested: 5.3 and 5.1 for library LLWE/RLAS and library LLWWE/PMNL respectively. The libraries were transformed into BL21 AI RIL cells, with 9216 clones picked into plates for each library.

2.3.2 Identification of HDAC7 soluble constructs

In total, 18432 constructs (2 libraries printed on the same membrane) were assayed for potential solubility. The solubility screen was processed using 3 main stages: (i) the clones were arrayed on a nitrocellulose membrane; (ii) they were grown on the membrane sitting on LB agar plate; (iii) they were then induce on the membrane by moving it across onto another LB agar plate containing the inducer L-arabinose. The clones were lysed *in situ* to generate colony blots, where putatively soluble proteins were detected with a fluorescent Alexa 488 streptavidin conjugate and the presence of the N-terminal hexahistidine tag was detected using a monoclonal anti-hexahistidine tag antibody with fluorescent Alexa 532 conjugated anti mouse secondary antibody (Fig. 8).

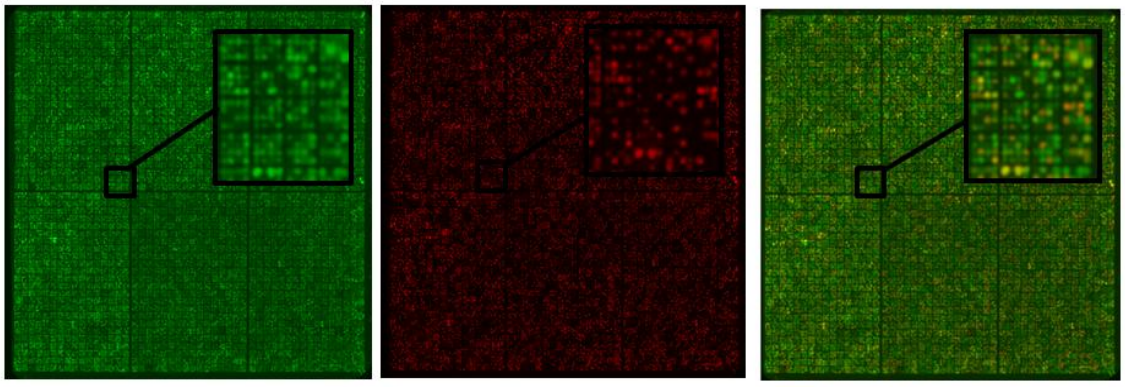


Figure 8. Screening for expression and putative solubility. The colonies are printed onto a nitrocellulose membrane and lysed *in situ*. In this colony-based solubility screening the C-terminal biotinylation is represented in green in the left panel, the N-terminal hexahistidine tag is represented in red in the middle panel and the right panel shows the merged images, where yellow clones indicate the presence of both signals simultaneously.

The values of signal intensities were extracted using VisualGrid array analysis software (GPC-Biotech). Clones with no detectable hexahistidine signal were eliminated and only clones showing both signals were considered further, thereby removing clones

exhibiting internal translation initiation, premature translation termination or proteolysis. The clones were then ranked according to their biotin signal intensities. The 96 clones with highest biotinylation signals were selected for a further level of screening by purification from 4 ml expression cultures using a liquid handling robot (Tecan Genesis workstation) and were analysed by SDS-PAGE (Fig. 9).

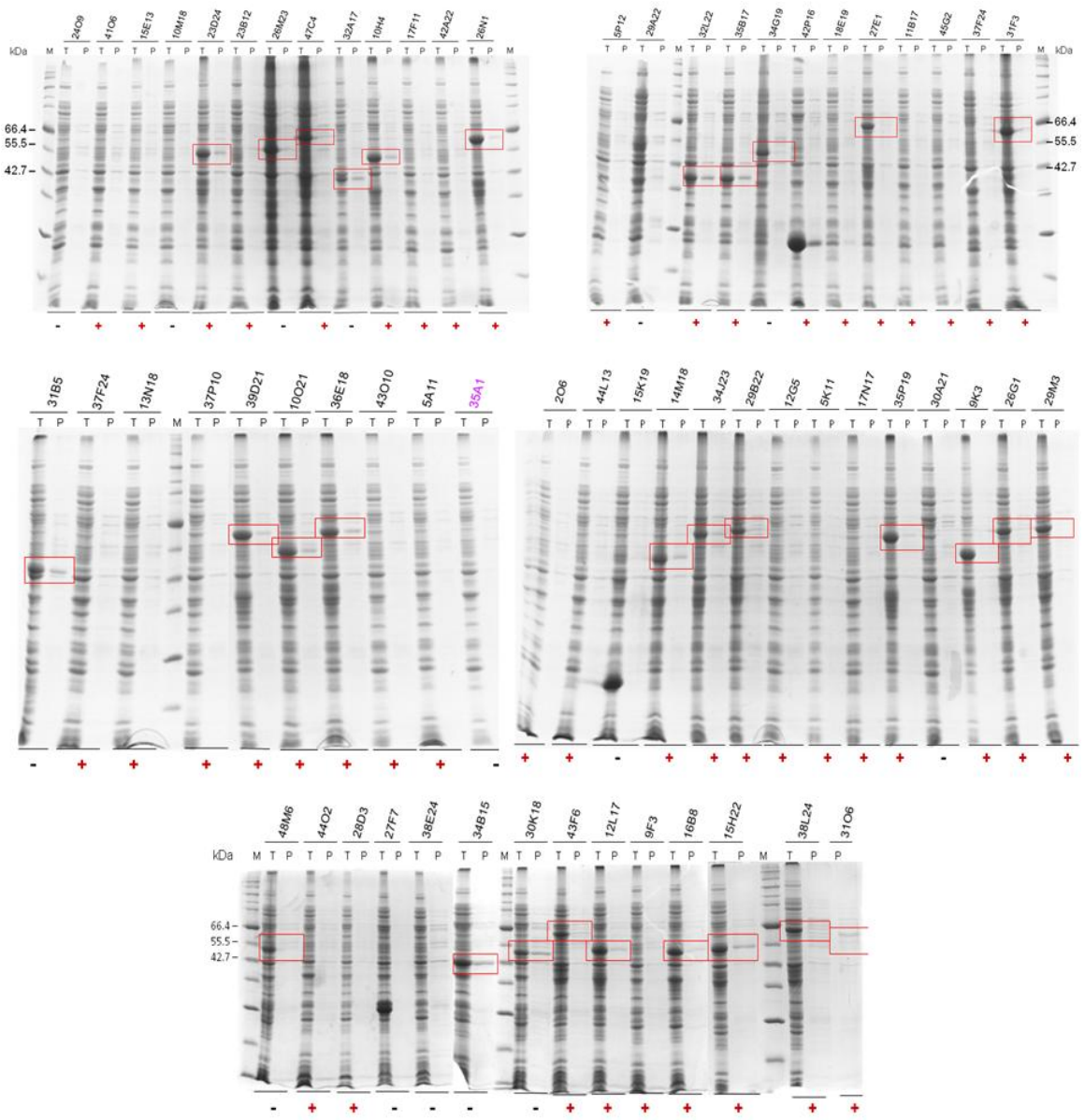


Figure 9. Identification of purifiable constructs of HDAC7. The 96 selected clones were analysed by Coomassie stained SDS-PAGE (T: total fraction, P: purified fractions) with 62 out 96 loaded on gel. The clone 35A1 in purple was loaded as a control (HDAC construct used by Schuetz) and was not screened (does not have a C-Terminal BAP). A few clones showed reasonable overexpression visible in total fractions, they are highlighted in red squares. Moreover a deacetylase activity assay was performed on the eluted proteins indicating the presence (+) or absence (-) of activity.

Several clones were reasonably overexpressed according to their total fraction: 31 out of 96 clones with molecular weights from 35 to 66 kDa. All of them were active on the class IIa-specific trifluoroacetyl substrate (Fig. 9). The presence of lower molecular weight constructs was also observed for 3 clones but they showed no activity indicating that the truncation may have deleted important catalytic residues. It is sometimes observed that overexpression is not successful when scaling up expression conditions from 4 ml to 50 ml or 1 l. To exclude this possibility, the 96 clones were expressed in 50 ml cultures and the same results as for 4 ml cultures were observed. The figure below summarises purifiable HDAC7 fragments alignment (Fig. 10). In total 24 constructs showed reasonable overexpresssion, solubility and deacetylase activity and it was observed that up to 40 amino acids were deleted by truncation yielding soluble, purifiable and active proteins. In the next step, large-scale expression of 1 l of 12 constructs were performed, to assess solubility and TEV cleavage efficiency (Fig. 11).

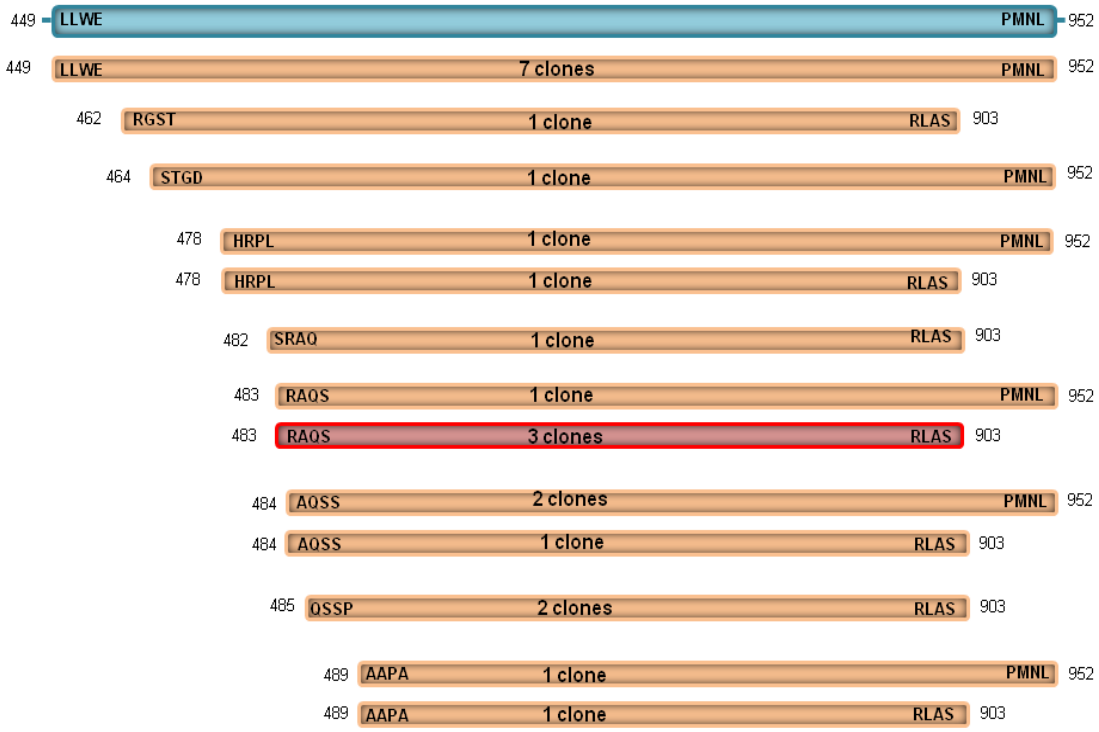


Figure 10. Sequence alignment of the selected clones. The truncation was performed on the parental construct in blue (only one parental construct shown). In light orange are the sequences of the selected constructs. The sequence in red is identical to the construct used for crystallisation by Schuetz *et al*.

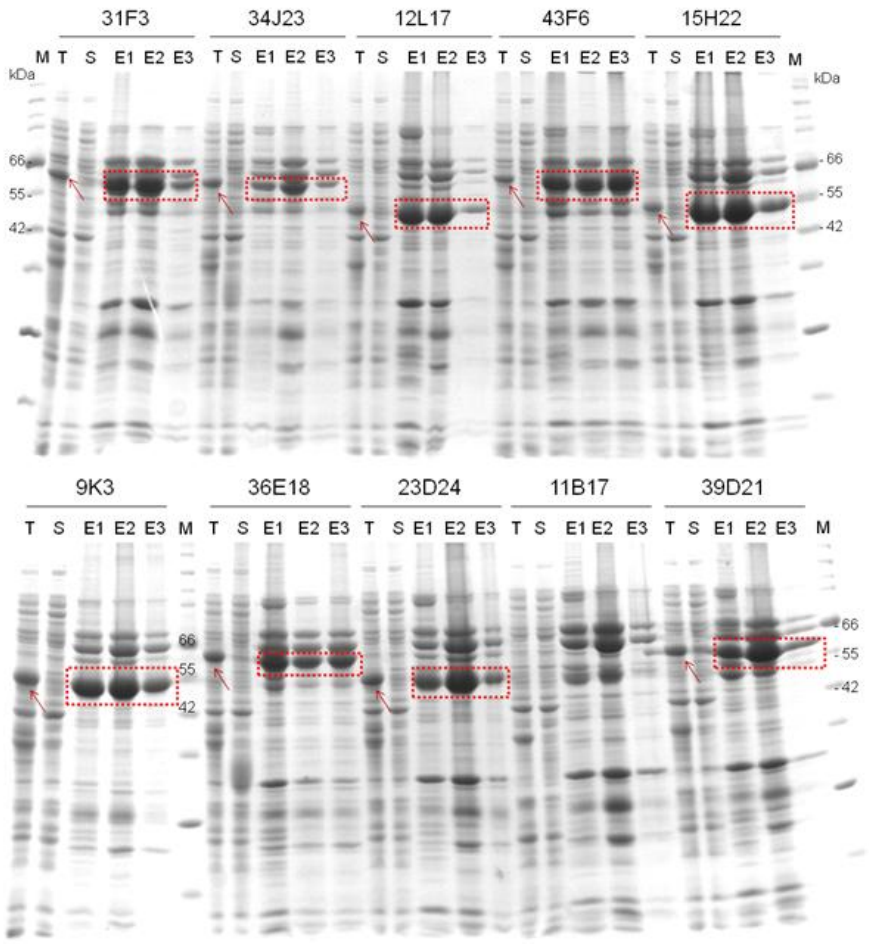


Figure 11. Large-scale expression and purification of selected HDAC7 fragments. Total fractions (T), soluble fractions (S) and elution fractions (E1, E2, E3) were loaded on SDS-PAGE. The proteins were overexpressed and visible in the total fraction, but the proteins were faintly or not detected in the soluble fractions. Besides, the proteins were purified successfully except for clone 11B17 that sequencing revealed as out of frame.

Despite reasonable expression, the selected fragments were only partially soluble, however the yields were compatible with further steps of processing. Moreover, it was possible to remove the hexahistidine tag by TEV protease digestion. Finally, the C-terminal BAP was suppressed by site directed mutagenesis.

Surprisingly, amongst the best selected hits, some constructs (Arg483-Ser903 or RAQS/RLAS; Fig. 10) were similar to the original poorly-expressed published construct. We hypothesised that this might be explained by the presence of the C-terminal BAP that was somehow enhancing the levels of soluble expression of this construct. Experiments were then carried out to investigate this effect.

2.3.3 Deletion analyses performed on the C-terminal BAP

Three different constructs were designed from the RAQS/RLAS construct by performing deletion of the linker and BAP (Fig. 12). The constructs were expressed in 1L and purified by affinity purification. The expression yields were better for the hit construct (selected out of the library with the BAP tag in place) and for the designed ones containing amino acids from the linker of the BAP tag. Purification yields followed the same tendency with equivalent yields for the screened construct and the construct extended by 6 amino acids. Thus this construct was selected for further use in crystallisation trials and enzymology assays.

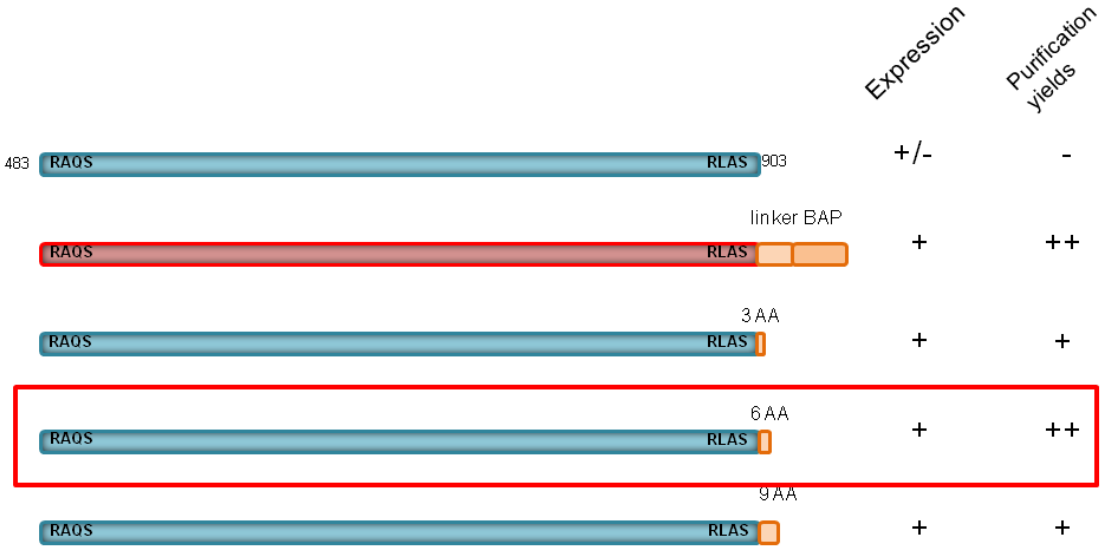


Figure 12. C-terminal deletions of the BAP tag and its linker. Deletions of the linker of the BAP tag were performed. Three constructs were generated containing either 3 (residues NAS), 6 (residues NASNNG) or 9 (residues NASNNGSGG) non-native amino acids from the linker of BAP tag. These amino acids were sufficient to get good expression and protein yields.

2.4 The optimised HDAC7 construct

2.4.1 Expression and purification of HDAC7

The expression of HDAC7 was carried out in 12 l of culture. This represents an improvement of about 5-fold compared to previous conditions where 50 l cultures were required to get sufficient amount of purified protein. HDAC7 was purified by a 4 step purification process. HDAC7 was eluted in a single peak at a salt concentration of 200 mM. In contrast to the original construct, the elution profile of the optimised construct strongly suggested that HDAC7 was purified as homogenous species (Fig. 13; compare

with Fig 1). Moreover the protein was further purified by gel filtration and was characterised as monodisperse, eluting at the expected weight (Fig. 14).

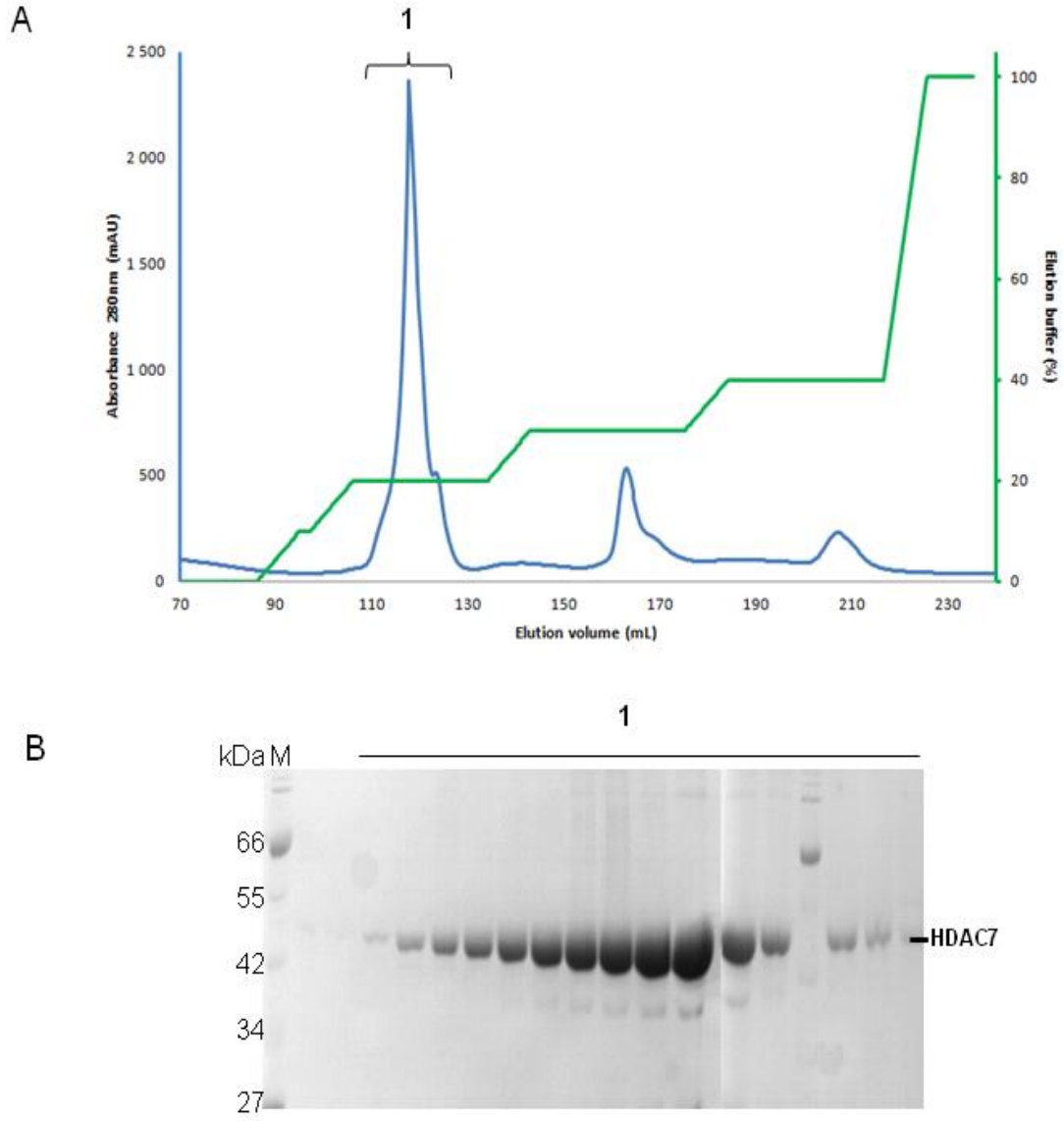


Figure 13. Anion-exchange chromatography of HDAC7. HDAC7 was purified on a Q sepharose anion-exchange column. HDAC7 eluted in a single major peak (peak 1 in part A). After this purification step, SDS-PAGE analysis (B) showed that HDAC7 was highly pure.

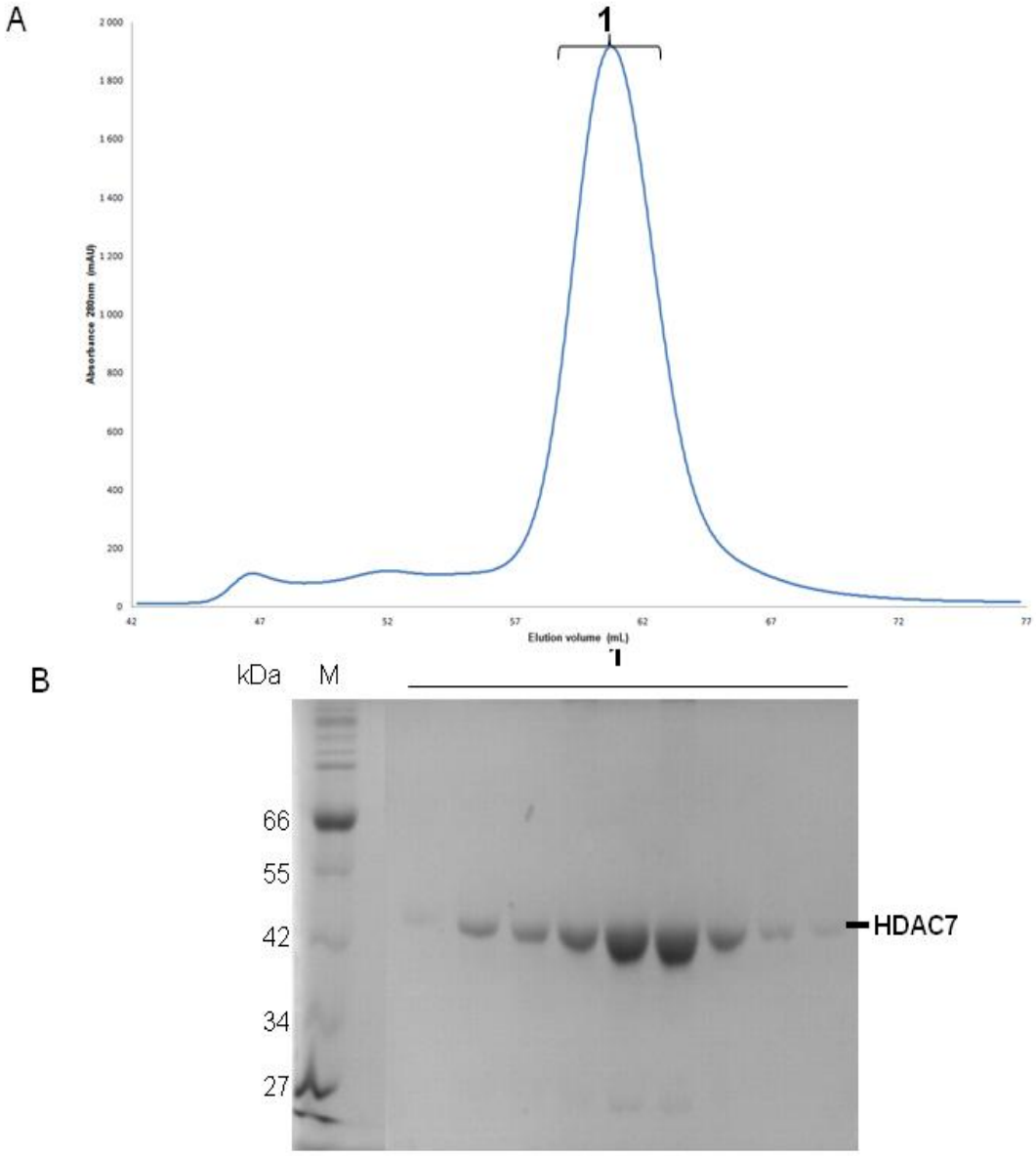


Figure 14. Size-exclusion chromatography. A. HDAC7 eluted in a single peak (elution volume of 60 ml). SDS-PAGE analysis revealed that it was pure (B).

The optimised construct showed greatly improved expression, solubility and purification. The 6 non native C-terminal amino acids somehow stabilised the HDAC7 catalytic domain and we hypothesise that the SGC construct may be slightly overtruncated leading to destablised protein, as evidenced by the ion exchange profiles (Figs. 1 & 14), and perhaps misfolding and protein turnover accounting for the low yield. Looking at the C-terminal region of the HDAC7 crystal structure, there is a short helix (aa 896-898) right at the terminus and we speculate that this may be stabilised, or even extended in the presence of the extra amino acids (Fig. 15).

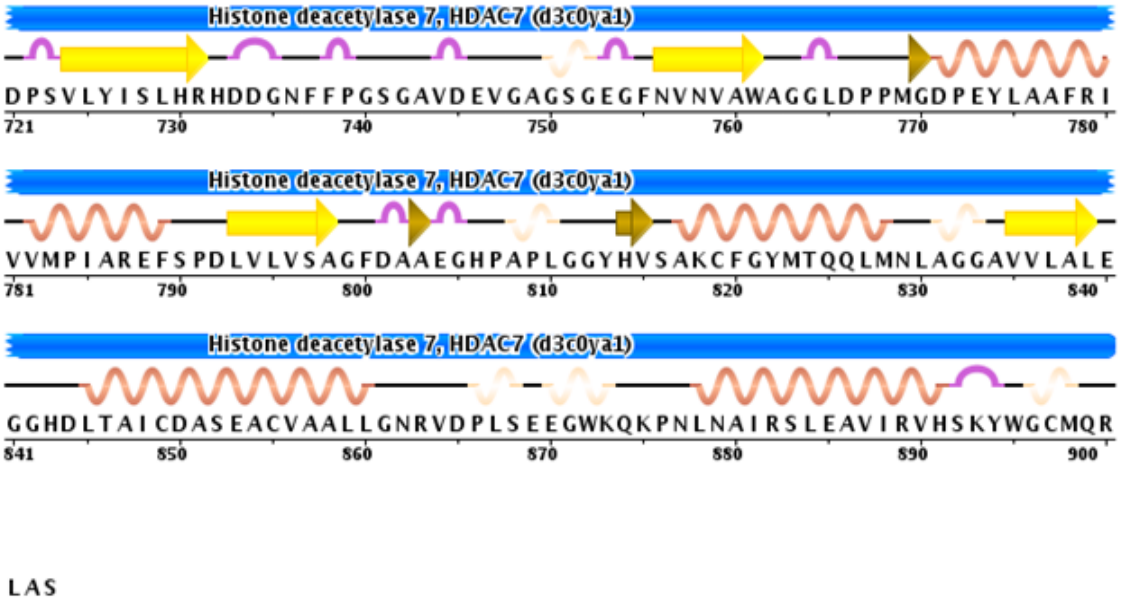


Figure 15. C-terminal sequence of HDAC7 catalytic domain structure (3C0Y). Structural information of HDAC7 extends until R901, LAS residues being invisible in the structure.

2.4.2 Crystallisation trials of HDAC7

The improved construct was then used for crystallisation trials. About 30 trials (using the high-throughput crystallisation facility of EMBL Grenoble, and manually) were performed both on the old and optimised constructs. Unfortunately no crystals of HDAC7 were obtained. Surface lysine methylation was attempted (Walter et al. 2006; Kim et al. 2008) employing reductive alkylation of lysine residues (Rypniewski et al. 1993; Rayment 1997). It proved to be unsuccessful for HDAC7 triggering its precipitation and preventing crystallisation experiments.

HDAC7 was engineered to improve its behaviour in expression and purification. This revealed to be fruitful, but HDAC7 remained recalcitrant to crystallisation. It is not rare, not to say quite common, that crystallisation may give crystals once and then impossible to reproduce. We then considered whether this construct could be further modified to allow crystallisation.

2.5 Further engineering to remove the disordered N terminus of HDAC7

In the structure of HDAC7 Arg483-Ser903 (Schuetz et al. 2008) the N-terminal region of the HDAC7 catalytic domain construct was invisible, probably flexible and disordered. Disorder prediction analyses were performed on this region (Fig. 16) revealing that the first 32 amino acids were unfolded which might hinder crystallisation

success. It seems possible that under the SGC crystallisation conditions these residues may have been removed slowly by contaminating proteases in the crystallisation drops (no analysis was presented on the protein crystals so it is unclear whether the 32 predicted disordered residues were present in the crystal). We therefore decided to see if these could be deleted from the new C-terminally engineered construct to improve crystal packing. Three amino acid incremental truncations of the N-terminal disordered residues were generated using the C-terminally optimised HDAC7 construct generating constructs starting from Leu506, Ser509, Pro512 and Thr515 (the first visible residue in the HDAC7 structure) (Fig. 17). Expression tests were performed on these constructs in 1l of culture with affinity purification and SDS-PAGE analysis (Fig. 17).

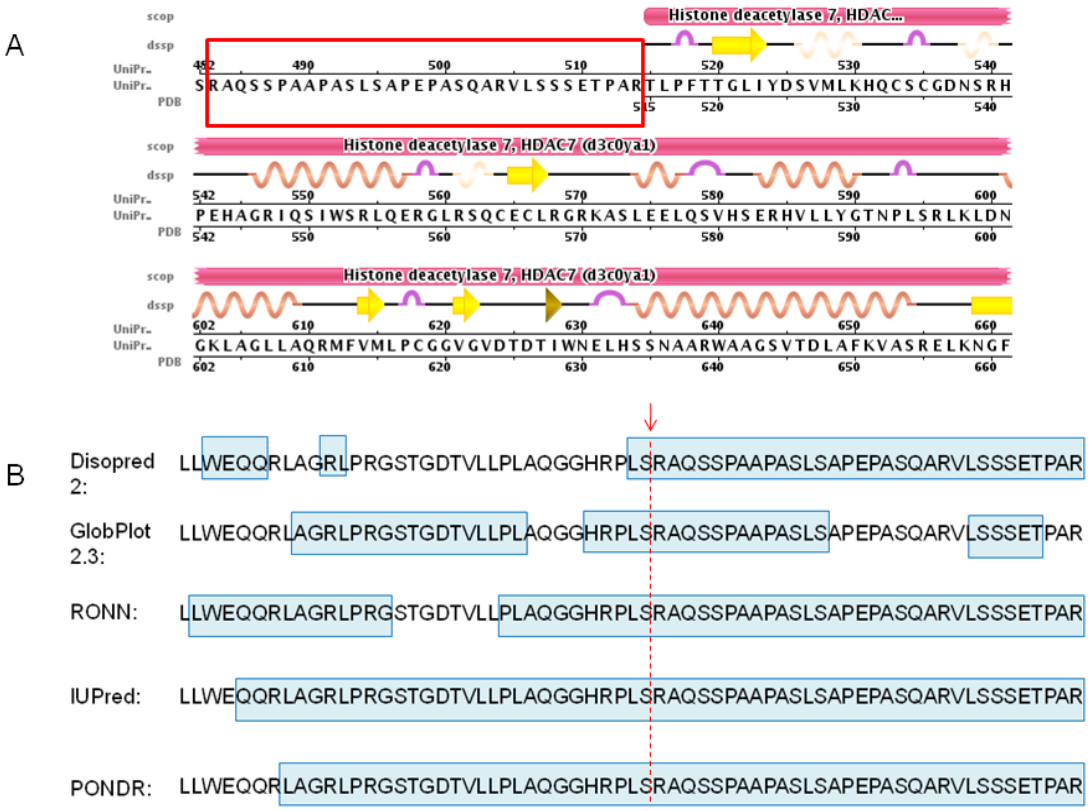


Figure 16. Analysis of the N-terminal region of the HDAC7 catalytic domain. A) The secondary structures of the first residues of the N-terminal catalytic domain are shown. The visible structure starts from Thr515. The region comprising Arg403 to Arg514 is invisible in the crystal structure. B) Disorder predictions were performed on this region by 5 different disorder predictors with disordered region highlighted in light blue. They all predict that region to be substantially disordered. The red arrow indicates the start of the construct from which a structure was obtained.

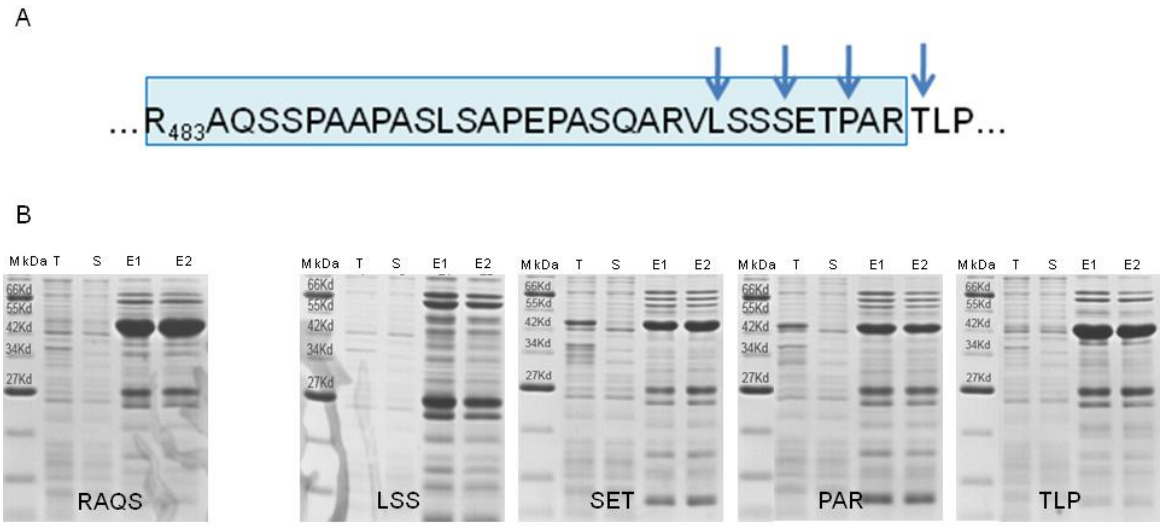


Figure 17. Alternative constructs to delete the N-terminal disordered region of HDAC7. The sequence starts at Arg483 being the same as the crystallised construct. The predicted disordered region is highlighted in light blue. The crystallographic ordered region starts from residue Thr515 (TLP). 4 constructs were designed starting from residues marked by a blue arrow. The constructs were purified and analysed on SDS-PAGE (T: total fraction, S: soluble fraction, E1 and E2: elution fractions). The construct starting with Arg483 was expressed and purified as a control. The constructs starting at Ser509, Pro512 and Thr515 were soluble and well purified.

Two out of the 4 designed constructs showed good behaviour upon affinity purification (expression and purification failed for the construct LSS). The shorter construct was chosen for a first attempt for crystallisation. This construct was purified large-scale and showed same profile as for the soluble optimised construct. This work is ongoing at the time of thesis submission.

3 Conclusions

During the early attempts of expression and purification of HDAC7 it became clear that there was a problem. The protein was very poorly expressed, being visible in lysates only by western blot. Initially large culture volumes (50 l in shake flasks) were attempted, but large amounts of contaminating *E. coli* proteins copurified so a more specific affinity tag (Strep II tag) was used. The resulting material was highly pure but the multiple peaks observed during ion exchange steps suggested it was not conformationally homogenous, perhaps explaining the lack of crystallisation. This was confusing since the Structural Genomics Consortium had succeeded in solving the structure of HDAC7 from this construct, but correspondence with them confirmed that very large volumes and difficult preparation steps had made their work very difficult also. We cannot explain the difference between our data and theirs, other than that crystallisation is often unreliable and difficult to reproduce, however we attempted it repeatedly. One hypothesis is that non-specific proteolysis was leading to crystallisation by slowly removing ends of the construct that are invisible in their structure, however they did not analyse the protein from the crystals so this cannot be confirmed.

As a consequence it was decided to carry out protein engineering on HDAC7 to identify better behaving constructs with improved expression and homogeneity. For this purpose a new technology developed in our laboratory, ESPRIT, was applied to HDAC7. This technology is based on random genetic truncation of a target sequence followed by a screen for soluble expression on the library using a small fused solubility reporter (biotin acceptor peptide). Better expressing constructs were identified both longer and shorter than the original form (suggesting at this point that the disordered N terminus might be removed). During analysis of these hits it was discovered unexpectedly that the addition of 6 non native amino acids at the C-terminus from the linker of the solubility reporter provided a strong effect on stabilising HDAC7 leading to improved expression, solubility, homogeneity and yields of protein. Therefore this construct was used in crystallisation trials, initially with the same disordered N terminus used in the original structure. However, despite its improved characteristics, HDAC7 remained recalcitrant to crystallisation. However this material provided the basis for all subsequent biochemistry and biophysical studies (chapter 3 & 4).

A second step of construct optimisation was carried out by the deletion of 32 N-terminal amino acids that are invisible in the structure. This was possible only once the C

terminus had been modified (above). This final optimised construct remains highly active and shows the same behavior as the optimised HDAC7 above (well expressed, homogenous, highly active) but is potentially more compact, a property useful in crystallisation. Crystallisation trials with this construct are in progress.

CHAPTER 3

Characterisation of a new SAHA-derived HDAC7 inhibitor

RESUME

Afin de pouvoir tester des inhibiteurs synthétisés par des collaborateurs à Rennes, des tests d'activité de déacétylation ont été effectués à l'aide de deux substrats commerciaux : lysine acétylée et lysine trifluoroacétylée. HDAC7 était active sur la lysine trifluoroactylée et non active sur la lysine acétylée en accord avec les controles positif et négatif. Des substrats portant différentes substitutions sur le groupement phenyl de SAHA ont été testés *ex vivo* et *in vitro* (tests sur des HDACs recombinantes commerciales et RMN avec HDAC7 optimisée envoyée aux collaborateurs). Ces tests ont révélé que des substitutions par des groupements methyle en position meta du groupement phényle contribuaient à augmenter le pouvoir inhibiteur de SAHA et sa spécificité. Un composé a été retenu, le composé 13 spécifique des HDACs de classe I et HDAC7. Les résultats de ces études ont été publiés (Oger *et al* 2010). Des tests de liaison du composé avec HDAC7 ont été effectués et suivies par dénaturation thermique suivie par dichroïsme circulaire. Ils ont montré que le composé 13 stabilise HDAC7 indiquant que celui-ci effectuerait des contacts supplémentaires avec HDAC7 via les groupement methyles en position méta la proline 809 serait impliquée dans la stabilisation du composé au sein de poche catalytique de HDAC7 (montré par modélisation moléculaire). Cependant des expériences complémentaires telles que mutagénèse et cristallisation sont requises afin de comprendre le mécanisme d'inhibition et de sélectivité de ce composé. Des essais de cristallisation en présence de composé 13 ont été effectués et n'ont donné aucun cristaux exploitables à ce jour, les essais sont toujours en cours.

1 Introduction

HDACs play an important role as regulators of cell functions and are involved in many cellular processes (chapter 1). In particular, they regulate gene expression by acetylating histones around which DNA is ordered. HDACs have been associated with many cellular events and diseases, like cell growth, cell differentiation, cancer formation (Minucci and Pelicci 2006), but they are also involved in other processes; for example, HDAC7 is involved in blood vessel integrity and cholesterol catabolism (see chapter I). Therefore HDAC inhibitors (HDACi) are extremely attractive targets for many therapies.

Intense research activities are ongoing to improve the pharmacokinetic and therapeutic indices of current HDACi. The characteristic features for hydroxamic acids inhibitors are represented by three distinct structural groups: the zinc-binding group, a hydroxamic linker and a cap group (Chapter1, Fig. 7). The hydroxamate moiety is the common zinc-binding group of many HDACi. Modifications of this group were moderately successful (they comprised benzamide, α-ketoesters, electrophilic ketones, mercapto amide and phosphonates) (Miller et al. 2003; Monneret 2005). The hydrophobic linker orientates the zinc-binding hydroxamic acid to the active site, and allows positioning of the cap group at the entrance of the active site. Thus the cap group can interact with amino acids located at the rim of the channel. Nevertheless in the crystal structure of HDAC7 bound to SAHA (PDB: 3C0Y), the cap group was not visible. It was thus hypothesised that substitutions of the cap might improve binding of this part of the molecule and the efficiency of the HDACi. Indeed, targeting the cap groups has been proposed for development of more potent and/or more selective HDACi, based upon molecular dynamics simulations (Estiu et al. 2008).

Selective inhibitors, affecting either a single HDAC isoform (isoform-selective HDACi) or several isoforms from a single class (class-selective HDACi) will provide a more effective drug therapy with less side effects compared to the current pan-inhibitors. They would also constitute an ideal research tool for investigating the specific functions of each HDAC isoform. However, despite the potential therapeutic advantages of selective HDACi, designing such selective inhibitors is difficult and challenging. The encountered obstacles are mainly due the high sequence and structural similarities of HDAC active sites. In addition, few structural data are available preventing characterisation of the subtle differences between HDACs to address the problem of specificity. Only three structures are available: HDAC8, HDAC7 and HDAC4. However,

some advances in isoform selective HDACi against class IIa HDACs have been described: 2-trifluoroacetylthiophenes showing potency and selectivity for class II HDACs with a 10-fold selectivity and diphenylmethylene hydroxamic acids showing selectivity for class IIa HDACs (Jones et al. 2008; Tessier et al. 2009)

Other limitations may come from the HDAC activity assays. Previously radioactivity was used widely to measure deacetylation using radiolabeled substrates. However even though successful, it required time consuming additional steps (producing the labelled substrates, separation of the substrate from the product, necessity to work with additional special safety conditions, potential contamination problems on high throughput automation), all limiting the assay throughput for drug discovery projects. The first non-isotopic methods for assaying HDAC activity used MAL (N-(4-methyl-7-coumarinyl)-N-α-(tert.-butyloxy-carbonyl)-N-Ω-acetyllysinamide) as a substrate (Hoffmann et al. 1999). However the formation of deacetylated product was measured by HPLC and fluorescence detection after extraction with ethyl acetate which decreased considerably the throughput. Recently assays were improved with the use of fluorescently-labelled octapeptide substrates (Hoffmann et al. 2001) and with the use of internal standards for the quantification of fluorescence substrate by HPLC. But this method was poorly suited to high-throughput screens as it required a step of substrate and product separation. More recently, fluorescent HDAC activity assays have been developed to ease high-throughput screening of pharmaceutical compound collections (commercialised by BIOMOL/ENZO and BPS Biosciences). These fluorogenic assays generally rely on a two-step reaction. A peptide (or peptide mimic) containing a ε-acetyl lysine residue is fused to 4-methylcoumarin-7-amide (MCA) at its C terminus. The acetyl group is first removed by the deacetylase activity of the HDAC releasing acetate from the lysine. Secondly the deacetylated peptide can then be recognised by trypsin which cleaves the peptide only in this deacetylated form. The fluorescent MCA is released and can be monitored in a fluorescent plate reader. Currently there are two main classes of substrates: (i) acetyllysine substrates (Boc-L-Lys (ε-acetyl)-MCA), class I, IIb and IV-specific (the only substrate available in the early stages of this project) and (ii) trifluoroacetyllysine substrate Boc-L-Lys (ε-trifluoroacetyl)-MCA), a class IIa HDAC-specific substrate recently designed that exhibits higher sensitivity due to the highly labile trifluoroacetyl leaving group (Lahm et al. 2007) (Fig.1). These 2 classes of substrates were used to assess the deacetylase activity of the recombinant HDAC7.

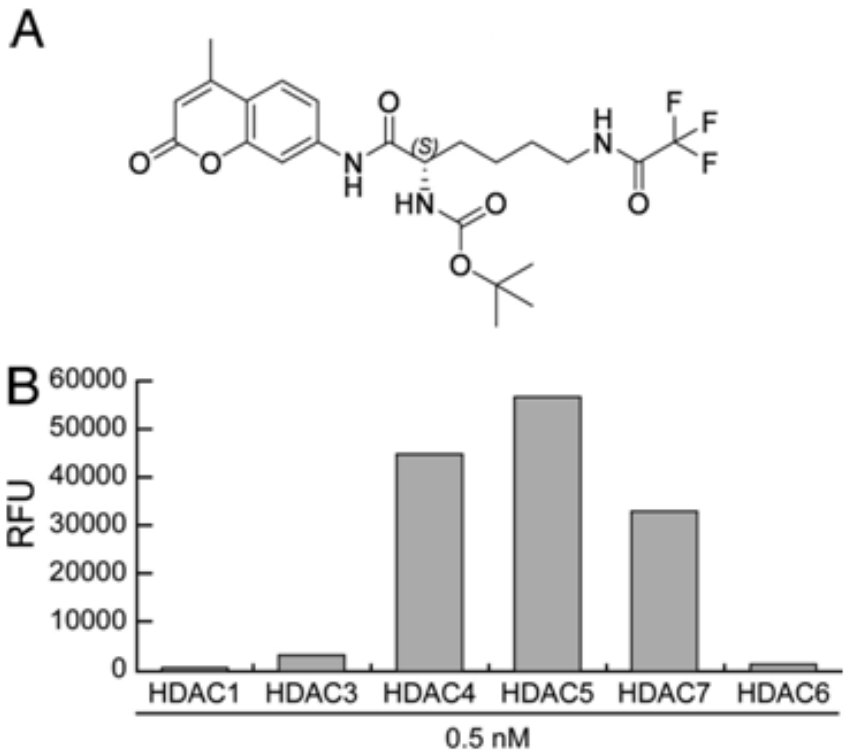


Figure 1. Trifluoroacetyllysine substrate for high throughput HDACi screening. A) The trifluoroacetyllysine substrate molecule. B) Catalytic activities of class I and class II HDACs were assessed using this substrate. This substrate is specific for HDAC4, 5 and 7 (class IIa HDACs), no activity was observed for class I HDACs (HDAC1 and 3) and class IIb HDACs (HDAC6). Figure adapted from (Lahm et al. 2007)

2 Results and discussion

2.1 Preliminary deacetylation assay of optimised HDAC7

As described in chapter II, the HDAC7 construct was optimised for improved solubility and yields. To assess its enzymatic quality, deacetylase assays on both available substrates were performed (Fig. 2). The optimised HDAC7 was active on trifluoroacetyllysine substrate, but there was no detectable activity on the acetyllysine substrate. This was to be expected, but indicated clearly that the HDAC7 construct optimisation and purification protocols had not disrupted the HDAC7 enzymatic activity, and that the protein was natively folded.

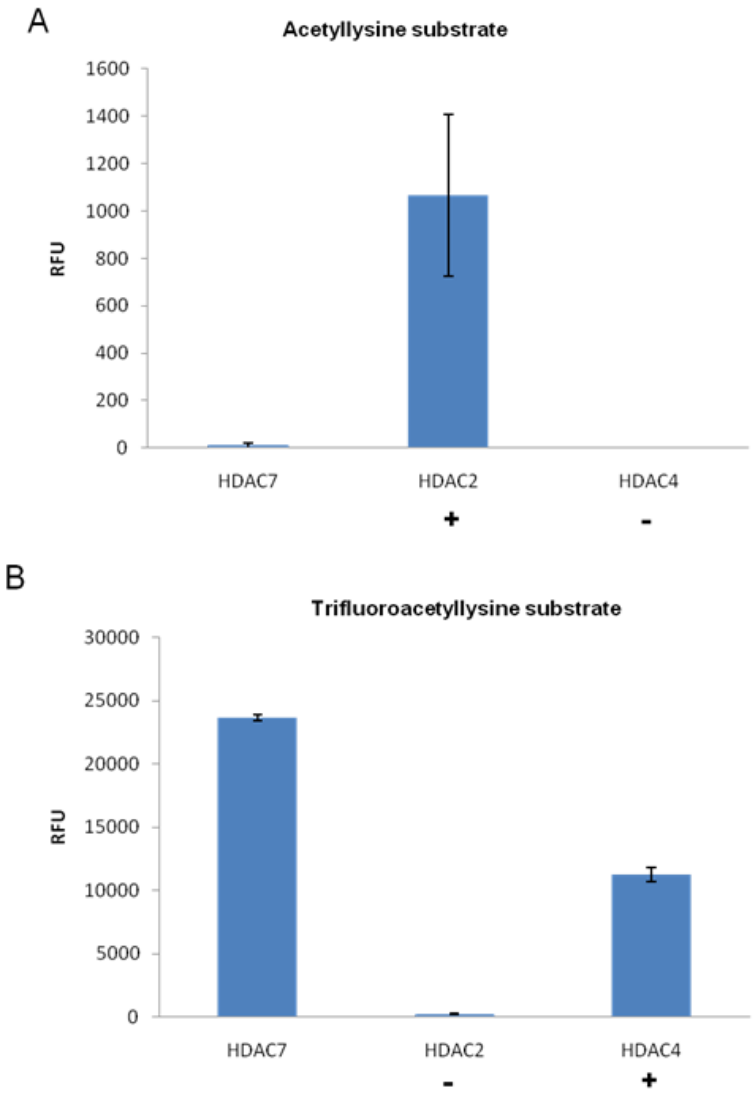


Figure 2. Enzymatic activity of the optimised HDAC7 constuct. Comparative enzymatic activity of HDAC7 on both acetyllysine (A) and trifluoroacetyllysine (B) substrates at equivalent concentrations (10 µM). Data represent mean values of three measurements ± standard deviation. HDAC7 is active on trifluoroacetyllysine substrate, but inactive on acetyllysine substrate. RFU, relative fluorescence units.

2.2 Presentation of Manuscript (Oger *et al.* 2010): Identification of a SAHA derivative with improved inhibition characteristics

The following section describes work published in Oger et al. J. Med. Chem. 2010 where we are co-authors and which is attached at the end of the chapter. From the construct engineering and protein production (Chapter 2), we were able to provide all of the purified material used in the NMR studies. Other parts of the work were performed by coauthors at Rennes, Bratislava, Milan, and L'Aquila. A commentary of the paper is provided below with further details in the manuscript itself. The work in the subsequent sections 2.3, 2.4 and 2.5 (not detailed in the paper) was performed entirely in our laboratory.

One of the driving principles in the field of HDACi discovery is that selective inhibitors can be designed exploiting the subtle differences in the active site of each HDAC isoform. To this end a number of attempts have been made to modify the cap group of the pan-inhibitor SAHA, the metal-binding moiety and the linker.

Many class IIa-selective inhibitors were discovered by screening against synthetic acetylated substrates (whilst class I HDACi were screened on acetyl-histone derived proteins or peptides). For example the HDACi tubacin was identified by monitoring deacetylation of α-tubulin, α-tubulin being a substrate for HDAC6 (Haggarty et al. 2003). It was shown that the capping group has a significant influence on HDACi selectivity for both class I and class II HDACs (Bieliauskas and Pflum 2008). It was hypothesised that the capping group moieties interact with the amino acids at the rim or near the rim of the active site (Bieliauskas and Pflum 2008; Andrianov et al. 2009). Nevertheless nothing is known about the specific groups required for discriminating class I and class II HDAC isoforms.

Within the SOUTH consortium it was decided to perform modifications on the cap group of SAHA (Fig. 3). SAHA cap derivatives were synthesised by monoethyl, mono- and dimethyl substitutions. The activity of these SAHA-derived small molecules were then assayed for inhibitory activity in cells and *in vitro* (Oger et al. 2010).

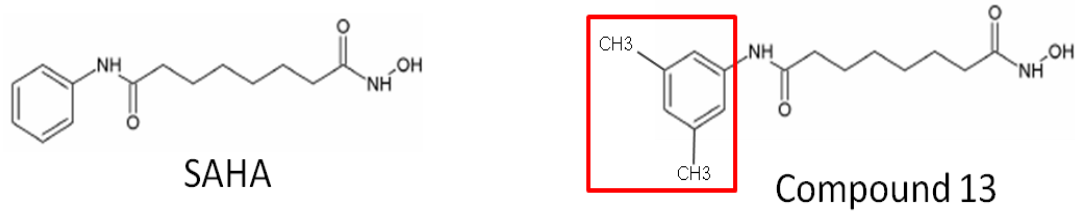


Figure 3. Structure of SAHA and one of the derived compound. SAHA was substituted in its meta position by 2 methyl groups.

2.2.1 Inhibition of SAHA Cap derivatives (SACs) compound *ex vivo*

The SACs were synthesised as (i) para-methoxy, (ii) ortho-, meta-, para- methyl, (iii) ortho-, meta-, para- ethyl, and (iv) ortho-, meta-, para- dimethyl substitutions. They were tested on Caco-2 colon cancer cells (Fig. 4).

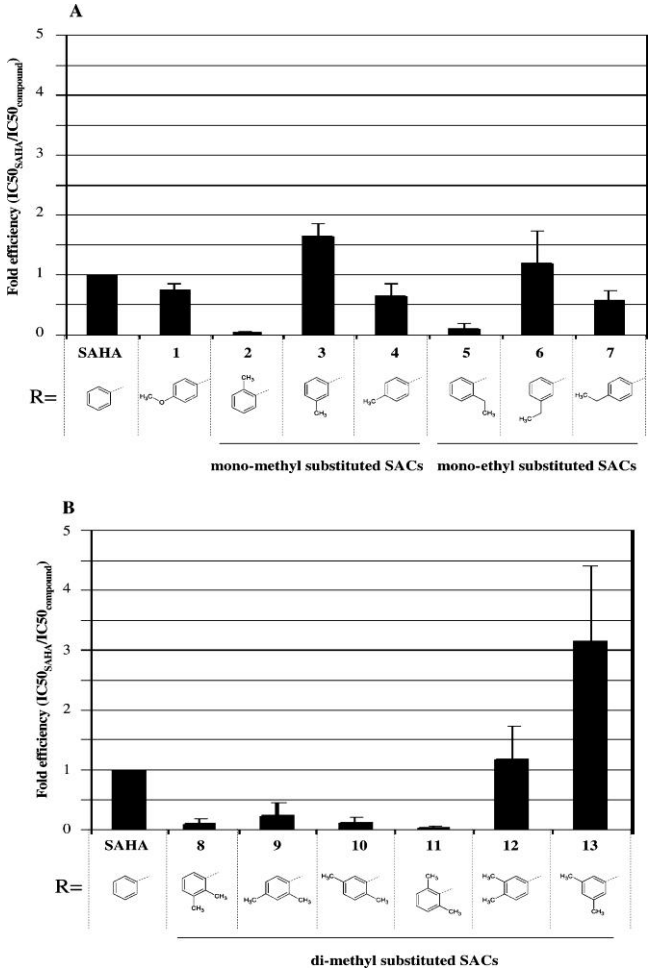


Figure 4. *Ex vivo* HDAC inhibition in Caco-2 cells in response to treatment with (A) monomethyl- and monoethyl-substituted SACs or (B) dimethyl-substituted SACs. Caco-2 cells were treated in triplicate using increasing concentrations from 0.1 to 50 µM (six concentrations) of the indicated compounds for 4 h. Data are shown as ratio of IC_{50} values (IC_{50}SAHA/IC_{50}compound). Results are the mean of at least three independent experiments except for compound 2. Error bars are standard errors of the mean (SEM) except for compound two (standard deviation, SD).

For monosubstitutions it was observed that para substitution (p-substitution) were 2-fold less active than SAHA, ortho substitution (o-substitution) 5-10 fold less active than SAHA, and meta substitutions (m-substitution) 1.5 fold more active than SAHA. The o-disubstituted compounds were inactive indicating that the ortho position affects SAHA efficiency. An o-substitution combined with either a m- or p-substitution restored inhibition activity. In addition m- and p-dimethyl substituted compounds were 1.2 to 3-fold more active than SAHA. This indicated that the position of substitution has an influence on effectiveness of the compounds.

2.2.2 Position of alkylation influence protein acetylation in Caco-2 cells

ELISA (Enzyme-Linked Immunosorbent Assays) were performed to monitor the effect of SACs on total protein acetylation levels in Caco-2 cells. The m- and p-monomethyl substitutions showed similar effects to SAHA. And o-substitutions (either mono or dimethyl) did not induce hyperacetylation (a marker for HDACi activity) except in combination with p-substitution, where hyperacetylation was observed. This indicated that there was a correlation between regulation of protein acetylation levels in these cells and the ability to inhibit HDACs, i.e. a positive correlation between the effects observed *ex vivo* and *in vivo*.

2.2.3 SACs profiling on HDAC1, HDAC3, HDAC5, HDAC6 and HDAC7

In order to determine possible selectivity or specificity of SACs on individual HDACs *in vitro* deacetylation assays were performed on HDAC1 and 3 from class I, HDAC5 and HDAC7 from class IIa and HDAC6 from class IIb. All o-substituted compounds were less active than SAHA at inhibiting all HDACs. This indicated that o-substitution alone is sufficient to decrease SAHA efficiency. Methyl and ethyl in m- and p-position influenced potency, those compounds showed 1.3 to 3-fold more potency than SAHA. Ethyl p-substituted compounds were as active as SAHA on all HDAC except on HDAC7 (4-fold more efficient). This indicated that short alkyl chains in m-and p-positions may selectively potentiate SAHA inhibition *in vitro*. Dimethyl m- and p-substituted compounds especially compound 13 (Fig. 1) were more potent on class I HDACs and on HDAC7, and o-dimethyl substituted compounds were the least active (compound 11). Therefore compounds 11 and 13 were selected for further analyses.

2.2.4 Molecular modelling of HDAC7 with compounds 11 and 13

Molecular modelling was performed on compounds 11 and 13 in order to determine structural features underlying the efficiency of o-dimethyl substituted (compound 11) and m-dimethyl substituted compounds (compound 13). SAHA was used as a reference. SAHA docked efficiently in the HDAC7 active site allowing both interaction with the catalytic Zn^{2+} at the bottom of the catalytic pocket and, via one of its NH groups, a hydrogen bond with the carbonyl group of Pro809. Compound 13 made the same hydrogen bonds between its NH amide and carbonyl of P809. But this was not observed for compound 11 which tilted away from Pro809, the absence of this hydrogen bond may affect the interaction of the compound with the active site of HDAC7. The root mean squared deviation (r.m.s.d) of SAHA, compound 13 and 11 were calculated and showed that the binding of compound 11 was less stable than in the interactions with SAHA and compound 13. For HDAC4 the same hydrogen bond was observed but with the equivalent Pro942 residue. These two prolines are likely to be involved in the binding of compounds and are only present in class II HDACs, suggesting a means for designing class-specific inhibitors.

2.2.5 Analysis of compound 11 and 13 affinity for recombinant HDAC7.

It was shown by activity assays on recombinant HDACs that o-substituted compounds were 5-100-fold less efficient than SAHA (compound 11) and that m- and/or p-substituted compounds (compound 13) exhibited higher efficiency than SAHA and showed as well partial enzymatic selectivity. Molecular modelling suggested a structural feature explaining the lack of activity of compound 11 but did not explain structurally the differences in activity between compound 13 and compound 11. For this purpose structure-activity relationship (SAR) analyses were performed for compounds 11 and 13. This was followed by NMR: saturation transfer difference (STD) and water-ligand observation with gradient spectroscopy (WaterLOGSY). The NMR experiments were carried out with batches of our recombinant optimised HDAC7 (chapter 2). The different biological activities of compound 13, compound 11 and SAHA may be related to differences in binding HDAC active sites (Fig. 5). It was thus determined that compound 11 had a weak affinity for HDAC7 compared to compound 13, this was followed through the methyl proton signals of both compounds.

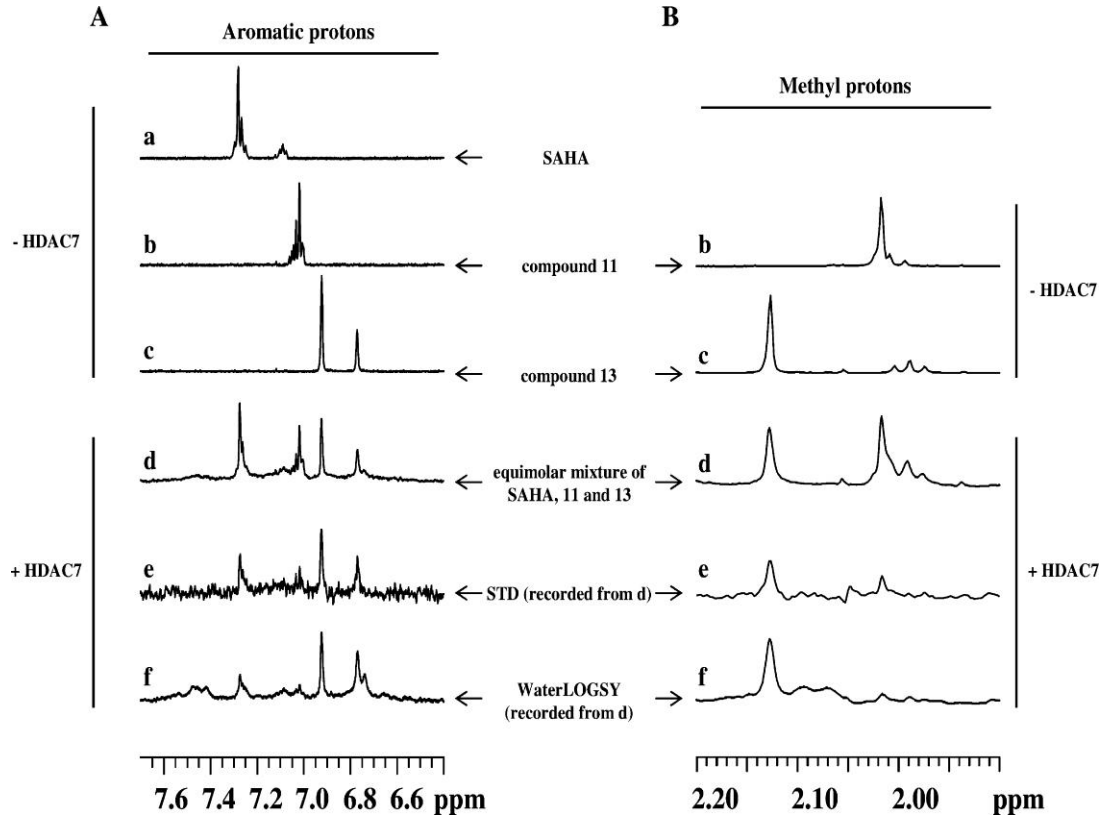


Figure 5. STD and WaterLOGSY experiments of compound 11 and 13. Expanded regions of one-dimensional spectra of (A) the aromatic protons of SAHA, compounds 11 and 13, and (B) the methyl protons. Experiments were performed in 20 mM phosphate buffer, pH7.4, 100 mM NaCl, and 10% DMSO and were recorded at 298 K. (a) SAHA, (b) compound 11, (c) compound 13, (d) equimolar mixture (60 µM) of the three ligands in the presence of 15 µM HDAC7, (e) STD spectrum recorded on sample d, and (f) WaterLOGSY spectrum recorded on sample d.

It was then determined in comparisons with the SAHA standard that the affinity for HDAC7 was higher for compound 13 over SAHA and compound 11 lower. It was thus concluded that the combination of 2 alkyl substituents in the m-position of the phenyl ring increased the potency of SAHA possibly by stabilising interactions. This might represent a route towards the development of new HDACi and this compound was studied further as part of this thesis.

The complete work can be found in the article attached in the end of this chapter.

2.3 Thermal stability assay to assess binding of Compound 13

The SAHA inhibitor sits in the catalytic pocket of HDACs. Its cap group is not well-defined in the three dimensional structure of HDAC7 (3C0Y), indicating that this region might be flexible. Therefore it was hypothesised that substitutions in the cap region might stabilise the binding to the inhibitor and as a consequence improve its specificity for HDAC7. To assess whether compound 13 bound in a different manner to SAHA, thermal

denaturation followed by circular dichroism were performed both with SAHA and compound 13 (Fig.6).

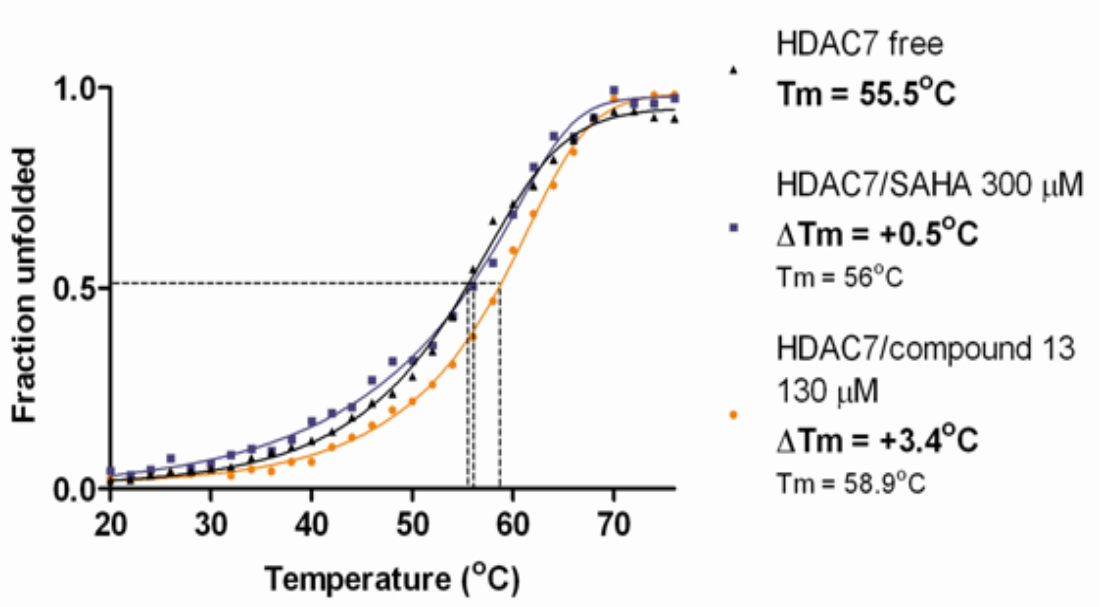


Figure 6. Binding assay of SAHA and compound 13. Thermal denaturation temperature (T_m) of HDAC7 alone was 55.5°C. The T_m increase upon binding SAHA was +0.5°C, whilst with compound 13 it increased by +3.4°C.

The T_m of HDAC7 increased by +3.4°C when incubated with compound 13. This suggests that compound 13 makes more molecular contacts with HDAC7 than SAHA and is in accordance with the NMR results of Oger *et al*. 2010. Indeed we hypothesise that the substituents may interact with amino acids located at the entrance of the catalytic pocket (Fig.7). The distance between SAHA and residues at the entrance of the catalytic pocket indicated that the distance with Pro809 was of 7.1 Å whereas His843 was a bit further. This may indicate a possible interaction with this proline when SAHA cap is substituted, as suggested by molecular modelling experiments in paragraph 2.2.4. However the three dimensional structure of this compound bound to HDAC7 is required to confirm this hypothesis. Targeting this cap be one way to improve HDACi selectivity, but it may be possible to target other residues in other loops according to differences between HDAC isofoms. However this will require further structural information as only three HDACs structures are available.

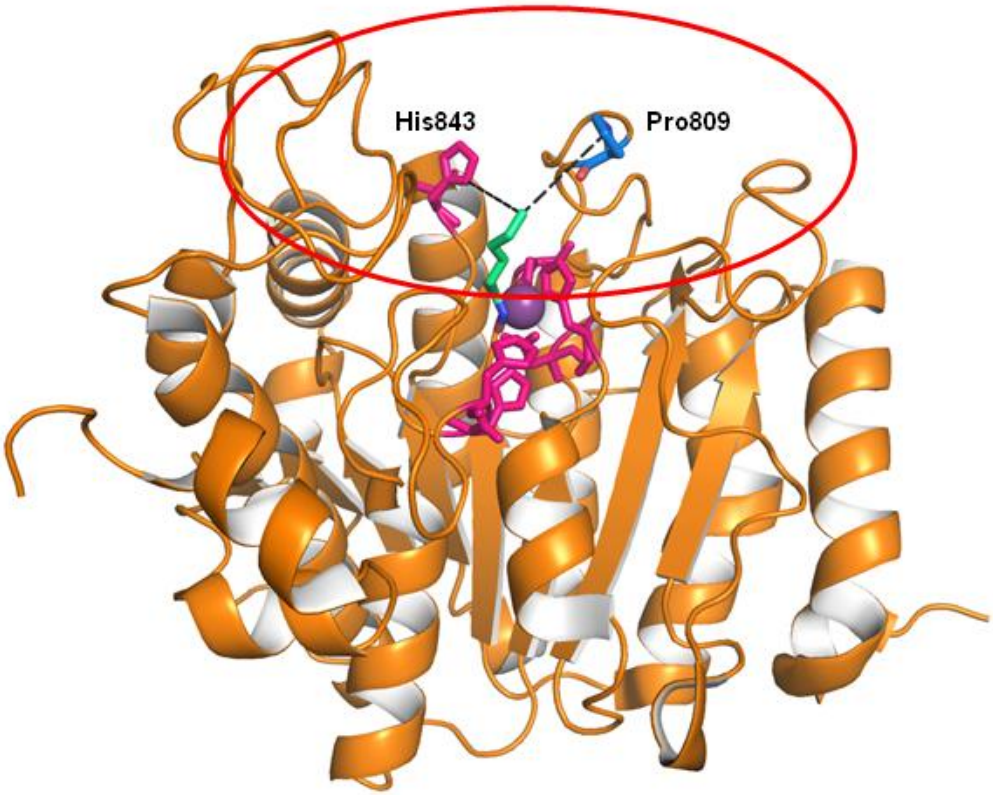


Figure 7. Structure of HDAC7 (3COZ). Catalytic residues are shown in fuchsia. The SAHA compound is indicated in green although the cap phenyl ring is not visible in the structure. The catalytic Zn^{2+} is in purple. The rim of the catalytic pocket is indicated by a red circle. The substituted phenyl ring of compound 13 may interact within this region, perhaps with proline 809. Distance between SAHA and histidine 843 was calculated as of 8.5 Å and proline 809 as of 7.1 Å.

2.4 Inhibition of HDAC7 by compound 13

Compound 13 seems to bind HDAC7 with more contacts than SAHA. To assess if this correlates with an increase in potency, IC_{50} values were calculated in a comparative analysis. This experiment was divided into two parts: (i) determination of enzymatic kinetic constants of HDAC7 on both substrates, (ii) determination of IC_{50} based upon kinetic constants obtained.

2.4.1 Michaelis-Menten kinetic analysis of HDAC7

No activity of HDAC7 was observed on acetyllysine substrate (Fig. 2A), this is a common feature of class IIa HDACs, as observed previously for HDAC4 and 7 (Lahm et al. 2007; Bottomley et al. 2008; Schuetz et al. 2008). This raises the question of enzymatic activity of HDAC7, which will be further discussed in chapter 4. Kinetic constants have been determined previously for the acetyllysine substrate (Schuetz et al. 2008). This can be explained by the inaccurate conditions they used, employing a massive

excess of enzyme (almost equal to the substrate concentration): this is incompatible with the Michaelis-Menten requirements (i.e. enzyme concentration should be negligible relative to the substrate concentration) for getting accurate and relevant measurements.

Assays were therefore performed with the more labile trifluoroacetyl substrate (Fig. 1) using recombinant purified HDAC7. It showed that HDAC7 followed Michaelis-Menten kinetics and allowed measurement of the kinetic constants of the reaction (Fig. 8). Two parameters were determined: K_m, the Michaelis constant, and V_{max}, the maximal initial rate of the catalysed reaction, when the enzyme is saturated with substrate. The K_m of HDAC7 was of 5,5 ± 2,1 µM and V_{max} was determined as 1089 s^{-1}.

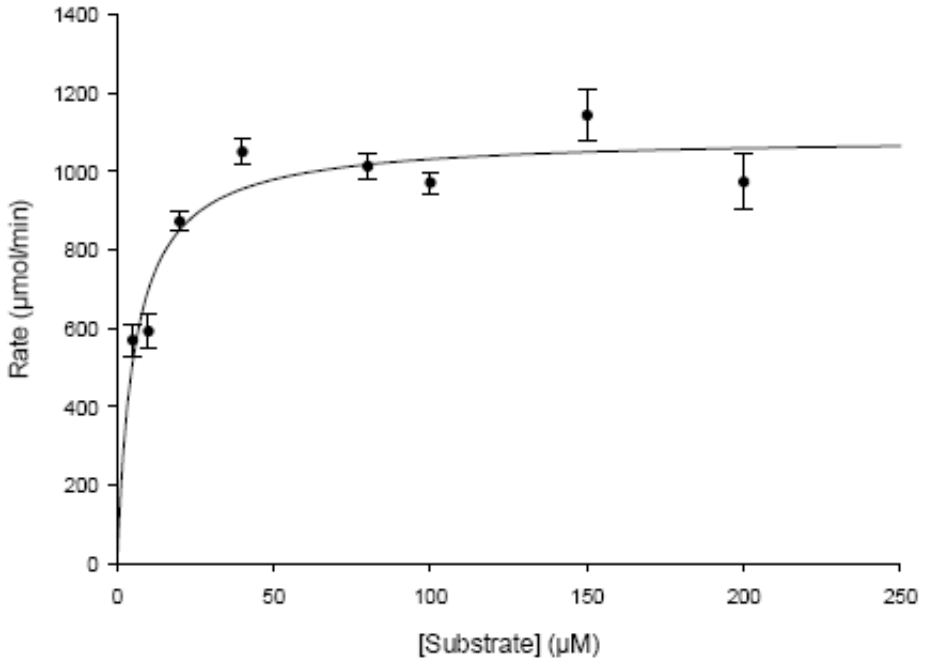


Figure 8. HDAC7 catalytic activity analysis. Deacetylase activity of recombinant optimised-HDAC7 was assayed using the trifluoroacetyl substrate (BPS Biosciences). The curve showed a characteristic hyperbolic shape characteristic of Michaelis-Menten kinetics. HDAC7 at 1 nM was assayed using substrate concentration from 5 µM to 200 µM. The K_m for the substrate was 5,5 µM and V_{max} 1089,4 s^{-1}.

2.4.2 IC_{50} determination

Dose-response studies were performed with SAHA and compound 13 to compare their inhibitory activity against HDAC7; half-maximal inhibitory concentrations (IC_{50}) were thus calculated (Fig. 9). The experiments revealed that compound 13 was 3-time more potent than SAHA on HDAC7 *in vitro*.

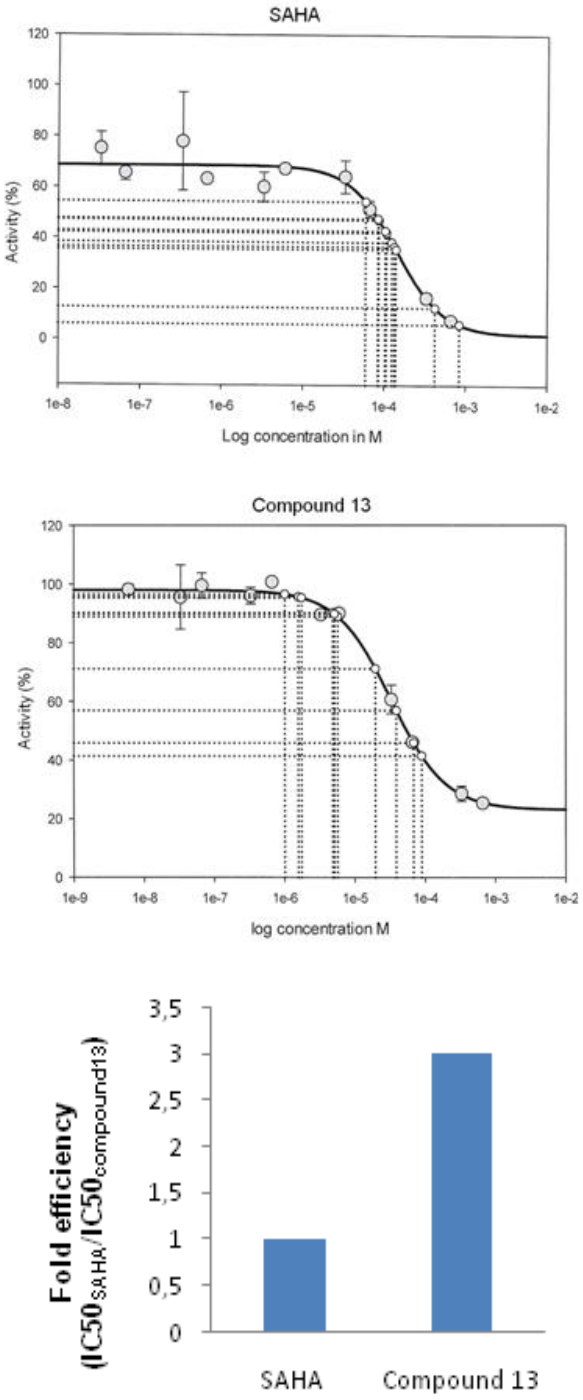


Figure 9. Inhibition of HDAC7 by SAHA and compound 13. The IC_{50} of SAHA was 100 ± 38 µM and IC_{50} of compound 13 was calculated 33 ± 4 µM. Data represent mean values of triplicate measurements ± standard deviation. The histogram indicates fold efficiency of compound 13 over SAHA: compound 13 is 3-fold more efficient than SAHA.

2.5 Crystallisation of HDAC7 with compound 13

To understand the molecular mechanism underlying the thermal stabilisation effect upon compound 13 binding, together with improved IC_{50} values over SAHA, cocrystallisation trials of HDAC7 and the inhibitor were performed. Crystallisation was

unsuccessful under these conditions, yielding heavy precipitation regardless of concentration optimisation and purification buffer used (two different Tris and Hepes derived buffers were tested). Under some conditions compound 13 itself crystallised (Fig. 10). All these crystallisation experiments were performed with the enhanced soluble HDAC7 (Chapter 2, section 2.4.2). This form of HDAC7 was heavily recalcitrant to crystallisation, as described in chapter 2 in the case of apo optimised HDAC7. This might be due to the short disordered N-terminal region that might prevent crystal contact formation. Crystallisation trials with the lately optimised HDAC7 (i.e. HDAC7 with 6 non-native amino acids at the C-terminal end and deletion of disordered region at the N-terminal end; (Chapter 2, section 2.5) are currently in progress (one crystallisation trial had been performed at thesis submission).

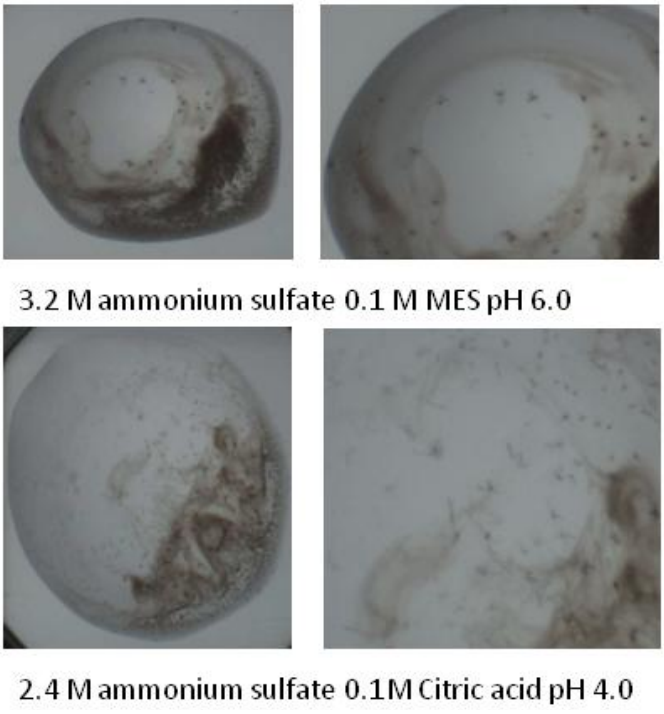


Figure 10. Sample of crystallisation drops of HDAC7 incubated with compound 13. These drops comprised the optimised HDAC7 construct co-crystallised with compound 13. Micro crystals with an urchin form grew in 2 days but little diffraction was observed at ESRF X-ray beamline ID23. However after reproduction of crystallisation conditions with the compound alone, diffraction results suggested that the compound 13 itself was crystallising. Heavy precipitation was observed, most likely HDAC7.

3 Conclusions

In this chapter were presented the results with a newly synthesised inhibitor (compound 13 obtained by cap modification of SAHA). With this compound several experiments were carried out and, additionally, studies were performed in Caco2 cells and by NMR in the groups of our collaborators. Some of these data were published in the Journal of Medicinal Chemistry.

First of all the optimised HDAC7 was tested for activity on trifluoroacetyllysine and acetyllysine fluorescent substrates. This revealed that HDAC7 is active on the much more labile trifluoroacetyllysine substrate therefore indicating that this HDAC7 construct was properly folded.

It was hypothesised that substitution of the cap phenyl group of SAHA may influence the selectivity and potency of SAHA so different SAHA-derived inhibitors were synthesised as alkyl mono or di-substitutions in ortho, meta and para positions. *In vitro* and *ex vitro* experiments by us and our collaborators confirmed that these substitutions did influence the activity of SAHA. In particular dimethyl substitution in the meta position increased the inhibitory effect on HDAC7. This may be explained by a possible interaction involving proline 809, located at the rim of the catalytic pocket, present in class II HDAC but not in class I HDACs. Therefore the rim of the catalytic pocket could represent a possible pharmacophore for designing class specific and/or isoform specific inhibitors provided more structural information becomes available for comparing small differences between HDAC isoforms. Then we compared binding of HDAC7 to both compound 13 and SAHA by thermal denaturation and CD. This experiment revealed that compound 13 bound more tightly to HDAC7, suggesting it made more molecular contacts than SAHA. Crystallisation experiments of HDAC7 and compound 13 have been unsuccessful with the earlier optimised HDAC7 construct (containing 6 stabilising non-native C-terminal amino acids), and even in some conditions the compound itself crystallised. Currently crystallisation trials are being performed with the further optimised HDAC7 (without the 32 aa N-terminal disordered region and including the 6 non-native C-terminal amino acids). Getting this structure will allow us to understand the interaction between HDAC7 and compound 13 and to identify moieties important for increased potency and isoform selectivity. Finally an enzymatic inhibition assay was performed to quantify the increase in inhibitory effect of compound 13 over SAHA. It demonstrated that compound 13 was 3-fold more potent than SAHA.

Journal of Medicinal Chemistry Article

J. Med. Chem. **2010,** *53,* 1937–1950 **1937**
DOI: 10.1021/jm901561u

Biological and Biophysical Properties of the Histone Deacetylase Inhibitor Suberoylanilide Hydroxamic Acid Are Affected by the Presence of Short Alkyl Groups on the Phenyl Ring

Frédérik Oger,[†] Aurélien Lecorgne,[‡] Elisa Sala,[§] Vanessa Nardese,[§] Florence Demay,[†] Soizic Chevance,[‡] Danielle C. Desravines,[∥] Nataliia Aleksandrova,[∥] Rémy Le Guével,[†,¶] Simone Lorenzi,[⊥] Andrea R. Beccari,[⊥] Peter Barath,[#] Darren J. Hart,[∥] Arnaud Bondon,[▽] Daniele Carettoni,[§] Gérard Simonneaux,[‡] and Gilles Salbert*[,†]

[†]*Equipe SPARTE, UMR CNRS 6026-Université Rennes 1, Campus Beaulieu, Bâtiment 13, 35042 Rennes, France, Cedex,* [‡]*Equipe ICMV, UMR CNRS 6226-Université Rennes 1, Campus Beaulieu, Bâtiment 10C, 35042 Rennes, France, Cedex,* [§]*Axxam, San Raffaele Biomedical Science Park via Olgettina 58, 20132 Milan, Italy,* [∥]*Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, UMI 3265, 6 rue Jules Horowitz, BP181, 38042 Grenoble, France, Cedex 9,* [⊥]*Dompé SpA Via Campo di Pile, I-67100 L'Aquila, Italy,* [#]*Cancer Research Institute, Slovak Academy of Sciences, Vlarska 7, SK-833 91 Bratislava, Slovak Republic, and* [▽]*Equipe RMN-ILP, Université de Rennes 1, UMR6026 CNRS, Campus de Villejean, Bat 05, 35033 Rennes, France, Cedex.* [¶]*Present address: U522 INSERM, Faculté de Médecine, Campus de Villejean, 35033 Rennes, France*

Received June 2, 2009

Inhibition of histone deacetylases (HDACs) leads to growth arrest, differentiation, or apoptosis of tumor cell lines, suggesting HDACs as promising targets for cancer therapy. At present, only one HDAC inhibitor (HDACi) is used in therapy: suberoylanilide hydroxamic acid (SAHA). Here, we describe the synthesis and biological evaluation of a new series of compounds derived from SAHA by substituting short alkyl chains at various positions of the phenyl ring. Such modifications induced variable effects ranging from partial loss of activity to increased potency. Through molecular modeling, we describe a possible interaction between HDAC7 proline 809, a residue that is strictly conserved within class 2 enzymes only, and the amide group of HDACi, while nuclear magnetic resonance experiments indicated that dimethyl *m*-substitution may stabilize the inhibitor in the active site. Our data provide novel information on the structure−activity relationship of HDACi and suggest new ways for developing second generation SAHA-like molecules.

Introduction

Histone deacetylases (HDACs)[a] are enzymes that catalyze lysine deacetylation in various proteinaceous substrates such as histones. HDACs have been divided into four classes on the basis of their protein sequence homology with yeast (*Saccharomyces cerevisiae*) deacetylase enzymes.[1] Briefly, class 1 HDACs (composed of HDAC1, -2, -3, and -8) are closely related to the yeast transcriptional regulator Rpd3 (reduced potassium dependency 3). Class 2 HDACs, including HDAC4, -5, -7, and -9 (class 2a) and HDAC6 and -10 (class 2b), share domain similarity with yeast histone deacetylase I. The class 3 HDACs form the nicotinamide adenine dinucleotide-dependent Sir2 (silent information regulator or sirtuin) family. In humans, seven members of the Sir2 family have been characterized and ordered in four subclasses.[2] Recently, a separate class of HDACs (class 4), comprising only HDAC11 in mammals, was described.[1]

Because aberrant histone acetylation has been linked to malignant diseases,[3] HDAC inhibitors (HDACi) bear great potential as new antitumor drugs. Indeed, they can induce differentiation, growth arrest, and apoptosis of transformed cells in culture. Many of these agents are effective in inhibiting tumor growth in vivo, and some have entered clinical trials as possible antitumor agents.[4] Class 1 and class 2 HDACs are sensitive to similar inhibitors (HDACi) derived from both natural sources and chemical synthesis. HDACi can be classified into six main groups: short chain fatty acids (e.g., butyrate and valproic acid), hydroxamic acids [e.g., trichostatin A (TSA), suberoylanilide acid (SAHA), and oxamflatin], epoxyketones (e.g., trapoxin and HC-toxin), cyclic peptides (e.g., apicidin and depsipeptide), epoxides (e.g., depudecin), and benzamidines (e.g., MS-27-275).[5] Class 3 HDACs, however, are not sensitive to these HDACi, and specific inhibitors such as sirtinol (Sir2 inhibitor naphtol) have been developed for the specific inactivation of these particular HDACs.[6]

HDAC inhibition has recently been clinically validated as a new therapeutic strategy for cancer treatment with the Food and Drug Administration approval of suberoylanilide hydroxamic acid (SAHA) for the treatment of cutaneous T cell lymphoma.[7] Intense research activities are ongoing in pharmaceutical and academic laboratories toward improving the pharmacokinetic and therapeutic indices of current HDACi. The classic pharmacophores for hydroxamic acid inhibitors consist of three distinct structural motifs: the zinc-binding group, a hydrophobic linker, and a cap group.[8] The X-ray structures of a bacterial HDAC homologue, histone deacetylase-like protein (HDLP),[9] bound to SAHA or TSA, and, more recently, human HDAC8,[10] confirmed that the zinc-binding group interacts with a Zn^{2+} ion at the bottom of a

*To whom correspondence should be addressed. Tel: 33(0)2 23 23 66 25. Fax: 33(0)2 23 23 67 94. E-mail: gilles.salbert@univ-rennes1.fr.

[a]Abbreviations: HDAC, histone deacetylase; HDACi, HDAC inhibitors; SAHA, suberoylanilide acid; SACs, SAHA cap derivatives; STD, saturation transfer difference; TSA, trichostatin A; WaterLOGSY, water−ligand observation with gradient spectroscopy.

Published on Web 02/09/2010

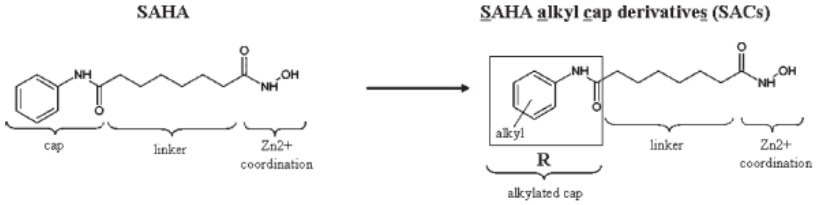


Figure 1. Structure of SAHA and general structure of the synthesized compounds. Compounds were derived from SAHA by substitution with alkyl on the phenyl ring and were named SACs for "SAHA alkyl cap derivatives". SACs conserve the hydroxamic acid function, the length of the aliphatic chain, and the amide function of SAHA and are only modified on the phenyl ring.

channel-like active site. The common zinc-binding group of many HDACi is the hydroxamate moiety, and modifications of this group have been modestly successful, yielding isosteres such as benzamide, α-ketoesters, electrophilic ketones, mercaptoamide, and phosphonates.[8,11] The hydrophobic linker presents the zinc-binding group to the active site, fills the hydrophobic channel, and positions the cap group at the entrance of the active site. The cap group could thus establish interactions with amino acid residues at the rim of the channel.[12] However, in the crystal structure of SAHA in complex with the HDAC7 catalytic domain (PDB ID: 3C0Z), the cap moiety is not well-defined, suggesting that addition of substituents in this part may improve binding and thus efficiency of the HDACi. Hence, targeting of the cap group represents an alternative approach to discovering more potent and/or more selective HDACi such as suggested by Wiest and co-workers using molecular dynamics simulation assays.[13]

Here, we describe the synthesis of a new series of SAHA derivatives harboring alkyl substituents at different positions of the phenyl ring. Our data demonstrate that the biological activity and the biochemical properties of SAHA can be finely modulated by the differential positioning of these substituents.

Results

Chemistry. SAHA cap derivatives (SACs, Figure 1) synthesized and used in this study are depicted in Table 1. All SACs and SAHA were synthesized using the method described by Mai et al.[14] The synthetic route to SAHA analogues is shown in Figure 2. First, suberic anhydride **14** was prepared from a mixture of suberic acid and acetic anhydride in good yield (98%). Then, suberic anhydride was condensed with various aniline derivatives to introduce diversity. In all of the syntheses, we noticed the formation of the expected amide together with the dianilide byproduct (about 10−40%). The latter was removed from solution by precipitation and filtration. The suberanilic acid (**1a**−**13a**) was subsequently treated with hydroxylamine in methanol to give the expected SAHA derivatives (**1**−**13**). Depending on the nature of the aromatic ring, the compounds were obtained in 14−95% yields. All of the compounds were characterized by ^{1}H NMR, ^{13}C NMR, and mass spectrometry. As expected, ^{1}H NMR spectroscopy mainly showed the existence of the *Z*-isomer.[15]

Biological Evaluation. *o*-,*m*-,*p*-Substitutions of the Phenyl Ring Differentially Modulate SAHA's Ability To Inhibit HDACs in Caco-2 Cells. As HDACi are promising drugs to treat colorectal cancer[16,17] (www.clinicaltrials.gov), we analyzed SACs activity in Caco-2 colon cancer cells (Figure 3A,B). Cells were treated with SACs, and the total HDAC activity was then measured. As shown in Figure 3A, the position of substitution impacted the effectiveness of compounds. Indeed, although *p*-substituted compounds

Table 1. List of Synthesized Compounds[a]

		SAHA cap substituent			
		nature	position		
Compound	R		ortho	meta	para
SAHA		/			
1		methoxy			✓
2		methyl	✓		
3		methyl		✓	
4		methyl			✓
5		ethyl	✓		
6		ethyl		✓	
7		ethyl			✓
8		dimethyl	✓ **(2)**	✓ **(3)**	
9		dimethyl	✓		✓
10		dimethyl	✓ **(2)**	✓ **(5)**	
11		dimethyl	✓ **(2,6)**		
12		dimethyl		✓	✓
13		dimethyl		✓ **(3,5)**	

[a] The position of alkyl substituents on the SAHA phenyl cap is indicated.

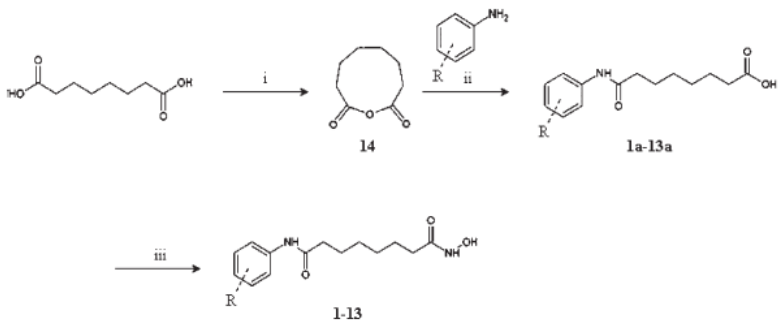


Figure 2. Reagents: (i) Acetic anhydride, 140 °C, 1 h.[31] (ii) THF, room temperature, 30 min. (iii) Ethyl chloroformiate, triethylamine (THF, 0 °C, and 10 min) and then hydroxylamine hydrochloride and potassium hydroxide (MeOH, 0 °C, and 15 min).

(**4** and **7**) were 2-fold less active than SAHA, *m*-substituted compounds (**3** and **6**) were approximately 1.5-fold more active than SAHA. On the contrary, *o*-substituted compounds (**2** and **5**) were 5−10 times less active than SAHA.

We next produced disubstituted compounds. Given that monomethyl and monoethyl substitution of SAHA showed similar effects, only dimethylated SACs were subsequently synthesized. As shown in Figure 3B, the *o*-substituted compound (**11**) was totally inactive toward HDAC inhibition, confirming that *o*-substitution strongly affects SAHA efficiency. HDAC inhibition was partially restored when one *o*-substitution was combined with one *m*- or *p*-substitution (compounds **8**−**10**). Interestingly, the dimethyl-substituted compounds without *o*-substitution (**12** and **13**) were 1.2−3-fold more active than SAHA, respectively. Taken together, these results indicate that alkyl substitution of the SAHA phenyl ring has a profound influence on the effectiveness of the inhibitors.

Position of Alkyl Substituents in SACs Both Quantitatively and Qualitatively Influences Protein Acetylation Levels in Caco-2 Cells. Next, indirect enzyme-linked immunosorbent assays (ELISAs) were performed to monitor the effects of SACs on total protein acetylation levels in Caco-2 cells (Figure 4). The ability of SACs to regulate protein acetylation levels in these cells was in good agreement with their ability to inhibit HDACs (compare Figures 3 and 4). Because different proteins could account for the acetylation signal detected in this assay, the protein acetylation profiles of Caco-2 cells treated with SACs were analyzed by Western blots using antiacetyl lysine antibodies. As shown in Figure 5A, SAHA induced hyperacetylation of proteins previously identified by Sehested and co-workers as α-tubulin and histones.[18] As expected, previously identified active compounds (**1**, **3**, **4**, **6**, and **7**) induced acetylation of these proteins to the same extent as SAHA. Conversely, no *o*-substituted compounds, except **9** and to a lower extent **10**, were able to induce protein hyperacetylation (Figure 5A,B). Compounds **12** and **13** both induced α-tubulin and histone acetylation, although **13** induced only a weak α-tubulin acetylation as compared to SAHA (Figure 5B), suggesting that this particular compound could have a different selectivity than SAHA. These results were directly confirmed by Western blot analysis using antiacetylated α-tubulin and antiacetylated histone H3/H4 antibodies (Supporting Information, S1, S2). Similar results were obtained in the hepatoma cell line HepG2 (Supporting Information, S3).

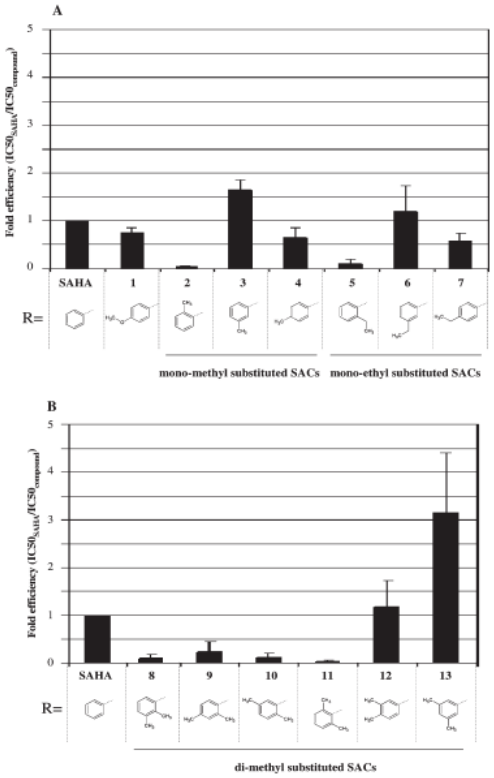


Figure 3. Ex vivo HDAC inhibition in Caco-2 cells in response to treatment with (A) monomethyl- and monoethyl-substituted SACs or (B) dimethyl-substituted SACs. Caco-2 cells were treated in triplicate using increasing concentrations from 0.1 to 50 μM (six concentrations) of the indicated compounds during 4 h. Data are shown as a ratio of IC_{50} values ($IC_{50SAHA}/IC_{50compound}$). Results are the mean of at least three independent experiments except for compound **2** (two independent experiments). Error bars are standard errors of the mean (SEM) except for compound **2** (standard deviation, SD).

Profiling of SACs Efficiency on HDAC1, HDAC3, HDAC5, HDAC6, and HDAC7. To determine the potential

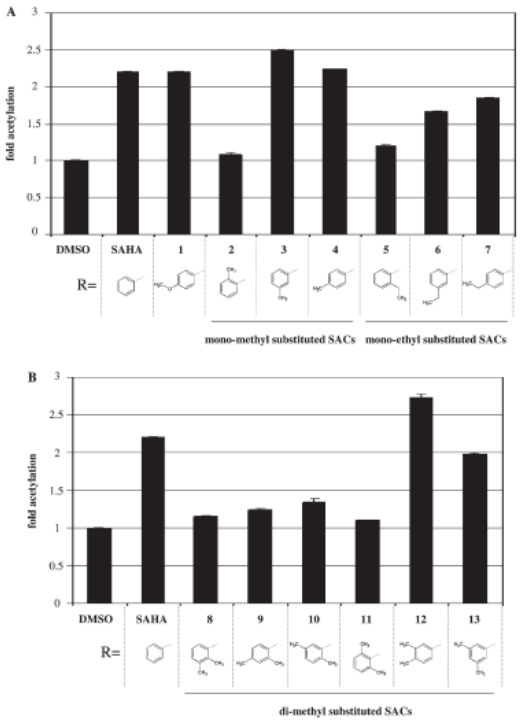


Figure 4. Quantitative analysis of protein acetylation levels in Caco-2 cells in response (A) to monomethyl- and monoethyl-substituted SACs treatments and (B) to dimethyl-substituted SACs treatments. Caco-2 cells were treated with a 5 μM concentration of the indicated compounds during 6 h, and indirect ELISA was performed in triplicate using 1 μg of whole cell protein extracts and anti acetyl-lysine as the primary antibody. Results are expressed as fold acetylation corresponding to protein acetylation level measured in response to indicated compounds as compared to the protein acetylation level measured using vehicle (DMSO). Error bars are standard deviations (SDs).

selectivity or specificity of SACs toward individual HDACs, we performed in vitro experiments with various recombinant HDACs. Two class 1 (HDAC1 and -3), two class 2a (HDAC5 and -7), and one class 2b (HDAC6) HDACs were used for these assays (dose−response curves are detailed in the Supporting Information, S4−S6). As shown in Table 2, all *o*-substituted compounds were less effective than SAHA at inhibiting class 1, 2a, and 2b HDACs. These data indicate that a single *o*-substitution is necessary and sufficient to dramatically reduce SAHA efficiency toward class 1 and class 2 HDACs. Conversely, methyl *m*- and *p*-substituted compounds (**3** and **4**) and ethyl *m*-substituted compound (**6**) were 1.3−3-fold more potent than SAHA (Table 2). Ethyl *p*-substituted compound (**7**) was as active as SAHA on all HDACs, except on HDAC7 (four times more efficient), suggesting some selectivity of this compound. Taken together, these data indicate that short alkyl chains, *m*- or *p*-substituted, may selectively potentiate SAHA-mediated inhibition of HDAC activity in vitro. Interestingly, crystallographic structures have revealed that the HDAC active site is surrounded by hydrophobic residues, which could provide targets for hydrophobic interactions involving alkyl chains substituted on the SAHA phenyl ring.[10,19−21] Remarkably, the dimethyl *m*- and *p*-substituted compound (**12**) and particulary the dimethyl *m*-substituted compound (**13**) were more potent on class 1 HDACs and on HDAC7, whereas their *o*-dimethyl counterpart (**11**) was the least active compound of the series on all tested HDACs. Because of the very high differential activity between **11** and **13**, we decided to explore further these compounds in structure−activity relationship (SAR) studies.

Molecular Modeling Reveals Proline 809 as a Potential Pharmacophore for the Design of Class 2 HDAC Selective Inhibitors. From the series of recombinant HDACs that we used for in vitro assay of SACs activity, HDAC7 is the only enzyme for which crystallographic data are available. Hence, to address structural features related to the contrasting biological efficiency of *o*- and *m*-dimethyl-substituted compounds (**11** and **13**, respectively), molecular modeling analyses were performed using SAHA, **11**, and **13** with HDAC7 (PDB ID: 3C0Z) as a template. As depicted in Figure 6A, SAHA was efficiently docked in the HDAC7 active site in a position allowing both interaction of the hydroxamic acid group with a Zn^{2+} atom and engagement of the NH from the amide group in a hydrogen bond with the carbonyl group of P809. Such a positioning of the HDACi is likely to be favored by interaction of the phenyl ring with F737. Interestingly, **13** was predicted to establish a similar hydrogen bond with P809, whereas **11**, most likely due to conformational constraints, had its NH group tilted away from P809 (Figure 6A). Through the analysis of the molecular dynamics runs of the three complexes, it appears that the lack of interaction between the NH group of **11** and the carbonyl group of the P809 may affect the interactions of the entire molecule with the protein active site. In particular, the hydroxamic moiety of **11** was predicted to adopt a different conformation in the zinc-binding pocket as compared to that of SAHA and **13** (Figure 6B). The root mean squared deviation (rmsd) of all three compounds calculated on the heavy atoms of truncated SAHA suggests that the binding of **11** is less stable than the other two ligands (Figure 6A). SAHA and **13** have a lower rmsd (calculated on the common structure atoms) than **11** and place their hydroxamic acid moiety in a very similar way as compared with TSA (not shown).

Similarly, molecular modeling of SAHA, **11**, and **13** in the active site of the other class 2a enzyme HDAC4 (PDB ID: 2VQW) suggested hydrogen bonding between the amide NH of SAHA and **13** and the carbonyl group of P942 (Supporting Information, S7). Accordingly, P942 of HDAC4 as well as P809 of HDAC7 are likely to be strong determinants of ligand binding. Examination of the solved HDAC4 and HDAC7 crystal structures in complex with HDACi revealed that these proline residues are indeed involved in the binding of compounds, although through bridging water molecules (Supporting Information, S8). Interestingly, proline at this position is only present in class 2 HDACs and thus provides a possible class-specific pharmacophore for drug targeting.

Substitution in *m*-Position Increases Compound Affinity for Recombinant HDAC7. Activity assays on recombinant HDAC enzymes with SACs indicated that *o*-substituted compounds (e.g., **11**) were 5−100-fold less efficient than SAHA, whereas dimethyl *m*- and/or *p*-substituted compounds (e.g., **13**) harbor not only a higher efficiency when compared to SAHA but also a partial enzymatic selectivity.

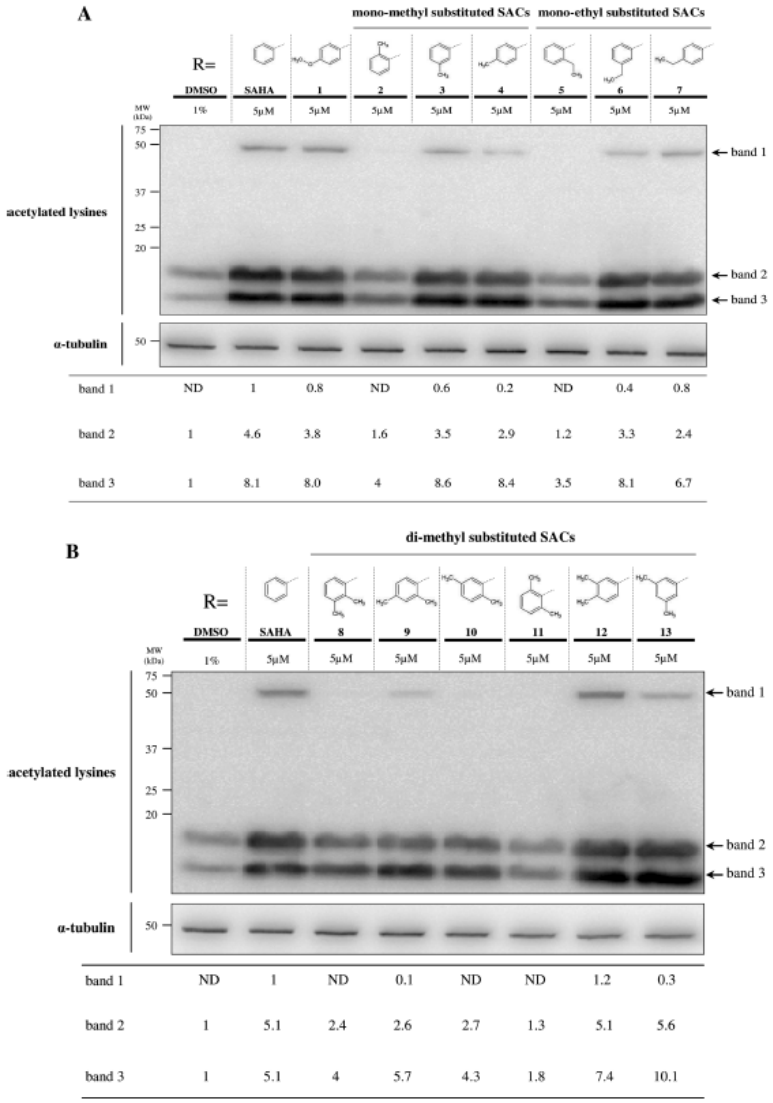


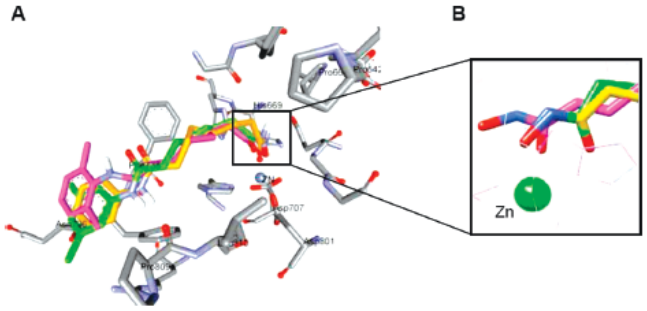
Figure 5. Protein acetylation profile in the soluble fraction of Caco-2 cells in response to (A) monomethyl- and monoethyl-substituted SACs treatments and (B) dimethyl-substituted SACs treatments. Caco-2 cells were treated with a 5 μM concentration of the indicated compounds during 6 h. α-Tubulin was used as a loading control. This result is representative of two independent experiments. When cells were treated with DMSO (1%) as a negative control, only two bands lower than 20 kDa were detectable (bands 2 and 3). The intensity of these bands was arbitrarily fixed to 1 to perform a densitometric analysis in comparison to SACs treatment. SAHA increased acetylation of at least one protein close to 50 kDa. As this signal was not detectable in the negative control, the intensity of this band was arbitrarily fixed to 1.

Although molecular modeling suggested structural clues for the explanation of the lack of activity of **11**, these analyses did not differentiate **13** from SAHA in terms of interaction. Thus, in an attempt to assess the SAR of compounds **11** and **13** in vitro, we performed saturation transfer difference (STD)[22] and water−ligand observation with gradient spectroscopy (WaterLOGSY)[23] NMR experiments using purified recombinant HDAC7 (aa 483−903). STD and WaterLOGSY sequences are known to be both particularly efficient for the screening of new molecules[24,25] and can be successfully applied to the analysis of mixtures of different ligands. Both methods rely on the existence of a strong negative nuclear Overhauser effect (NOE) associated with large molecules, in contrast with the weak positive NOE observed for free small molecules in solution. Hence, each ligand bound to a protein behaves as its large biomolecule partner and suffers long correlation time and negative NOE. STD corresponds to a direct transfer of magnetization

Table 2. SACs IC_{50} Values on Representative Class 1 (HDAC1 and HDAC3), Class 2a (HDAC5 and HDAC7), and Class 2b (HDAC6) HDACs[a]

	$IC_{50} \pm$ SEM (μM)				
	class 1		class 2a		class 2b
compounds	HDAC1	HDAC3	HDAC5	HDAC7	HDAC6
SAHA	0.119 ± 0.031	0.147 ± 0.053	12.60 ± 2.10	41.00 ± 10.20	0.037 ± 0.003
1	0.071 ± 0.011	0.137 ± 0.034	16.00 ± 3.30	21.60 ± 1.30	0.032 ± 0.002
2	0.576 ± 0.286	1.348 ± 0.002	42.70 ± 4.50	> 100	0.251 ± 0.069
3	0.057 ± 0.023	0.058 ± 0.032	11.30 ± 0.50	16.00 ± 1.80	0.030 ± 0.010
4	0.076 ± 0.026	0.114 ± 0.034	18.90 ± 2.70	19.00 ± 5.60	0.030 ± 0.001
5	1.101 ± 0.149	2.263 ± 0.427	> 100	> 100	0.215 ± 0.005
6	0.027 ± 0.002	0.035 ± 0.005	15.70 ± 4.10	16.10 ± 2.60	0.021 ± 0.001
7	0.149 ± 0.061	0.240 ± 0.070	9.60 ± 0.30	10.50 ± 1.60	0.026 ± 0.004
8	1.110 ± 0.240	1.750 ± 0.450	51.20 ± 12.20	> 100	0.206 ± 0.026
9	0.638 ± 0.013	1.100 ± 0.202	37.60 ± 6.10	45.40 ± 11.60	0.171 ± 0.002
10	1.142 ± 0.437	1.346 ± 0.164	54.30 ± 18.30	> 100	0.141 ± 0.041
11	11.614 ± 8.836	4.989 ± 1.051	> 100	> 100	0.615 ± 0.024
12	0.030 ± 0.010	0.056 ± 0.004	12.40 ± 0.50	15.90 ± 1.00	0.020 ± 0.001
13	0.017 ± 0.007	0.030 ± 0.001	14.30 ± 7.60	18.30 ± 7.30	0.035 ± 0.005

[a] IC_{50} values were determined from a minimum of two independent experiments. In each experiment, dose−response curves of compounds were tested in quadruplicate. The average IC_{50} values and the SEMs are reported in micromolar units.



Ligand	RMSD (Å)	StdDev (Å)
SAHA	0.747	0.049
11	1.463	0.058
13	0.722	0.040

Figure 6. Molecular modeling of interaction between SAHA, compounds **11** and **13**, and the HDAC7 catalytic site. (A) SAHA (yellow) and compounds **11** (magenta) and **13** (green) are depicted in the HDAC7 active site (gray). Amino acids of the HDAC7 active site are indicated. The putative H-bond interaction between the proton of NH belonging to the amide function and the carbonyl group of proline 809 (P809) of HDAC7 is depicted as a green dotted line. (B) Close-up view of hydroxamic acid function. The table indicates the rmsd for each tested compound. The minimized average structure of SAHA (yellow) in complex with HDAC7 protein (gray) has been reported and is shown with the superimposition of **11** (magenta) and **13** (green) and belonging to the respective minimized average complex structures.

through the protein protons, whereas WaterLOGSY uses magnetization transfer through water molecules present at the protein−ligand interface. The results were visualized as a difference of spectra with and without magnetization transfer and are displayed in Figure 7 for SAHA, **11**, and **13**. Interestingly, the relative intensities of the aromatic resonances were no longer representative of an equimolar mixture in STD and WaterLOGSY assays (Figure 7A, e and f spectra, respectively). The signals corresponding to the *o*- (6.92 ppm) and *p*- (6.77 ppm) protons of **13** were stronger than the multiplet at 7.26 ppm related to the four *o*- and *m*-protons of SAHA. The multiplet centered at 7.03 ppm, which corresponds to three protons (*m*- and *p*-) of **11**, was only weakly detectable. Hence, these data indicate that the differential biological activities of **13**, SAHA, and **11** are likely related to their difference in binding to HDAC active sites. The weak affinity of **11** for HDAC7 catalytic domain when compared to **13** was confirmed through analysis of the methyl protons of both compounds (resonances at 2.13 and 2.01 ppm for **13** and **11**, respectively; Figure 7B). Again, despite being in equimolar concentration with **11** and SAHA, **13** generated the strongest signal. Hence, on the basis of all of the NMR data, we can rank affinity of these three ligands for HDAC7 in vitro as **13** > SAHA ≫ **11**.

These results indicate that a combination of two alkyl substituents in the *m*-position of the phenyl ring increases the potency of SAHA, likely through a stabilized interaction within the active site, and could lead to valuable new

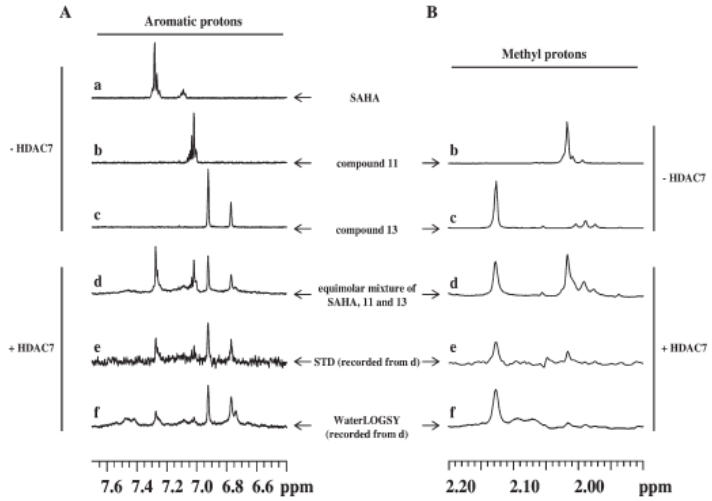


Figure 7. Expanded regions of one-dimensional spectra of (A) the aromatic protons of SAHA, compounds **11** and **13** (A), and the methyl protons (B). Experiments were performed in 20 mM phosphate buffer, pH 7.4, 100 mM NaCl, and 10% DMSO-d_6 and were recorded at 298 K. (a) SAHA, (b) compound **11**, (c) compound **13**, (d) equimolar mixture (60 μM) of the three ligands in the presence of 15 μM HDAC7, (e) STD spectrum recorded on sample d, and (f) WaterLOGSY spectrum recorded on sample d. All of the signals are presented arbitrary positive for convenience.

HDACi. In this context, the optimal combination and length of alkyl chains still remain to be determined. Accordingly, molecular modeling was performed using **13** docked in HDAC4 (PDB ID: 2VQW) as a reference and various compounds harboring from two to six carbon-length *m*-substituents (Supporting Information, S9). Results indicated that the number of different solutions increased with lengthening of the alkyl chain, revealing unconsistent positioning of the molecules within the active site. Consistent with the occurrence of variability in ligand position, the number of solutions relying on hydrogen bonding between the P942 and the amide NH decreased with chain length (Supporting Information, S9). A similar trend was observed when analyzing the position of the hydroxamic moiety of the compounds in proximity to the zinc ion (Supporting Information, S9). Hence, short chain derivatives are likely to group in the most active set of compounds.

Cell Growth Inhibitory Effects of SACs in Caco-2 Cancer Cells and in Nontumoral Cells. Because differential positioning of alkyl substituents on the SAHA phenyl ring differentially impacted HDACi effectiveness in vitro, we performed cell growth assays (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide, MTT) to correlate these in vitro data to the growth inhibition profile of the various compounds in tumor cells. In parallel, we monitored $p21^{waf1/cip1}$ gene expression, a known SAHA-induced gene regulating cell cycle and apoptosis.[26] Caco-2 cells were treated during 48 h with increasing concentrations of SACs, SAHA, or vehicle only as a negative control. Mono *m*-substituted or *p*-substituted SACs (**3**, **4**, **6**, and **7**, respectively) showed IC_{50} values within the range of SAHA (Figure 8A). As monitored by Western blot, compounds **3**, **4**, **6**, and **7** also induced $p21^{waf1/cip1}$ expression in a dose-dependent manner similarly to SAHA (Figure 8A). Mono *o*-substituted SACs (**2** and **5**) were less efficient than SAHA with IC_{50} values approximately 5-fold higher than SAHA (78.5 and 40.5 μM, respectively). In agreement with these observations, $p21^{waf1/cip1}$ expression was barely induced by a 5 μM treatment with these compounds (Figure 8A). Nonetheless, at a concentration of 50 μM, *o*-substituted SACs **2**, **5**, **8**, and **11** were able to induce $p21^{waf1/cip1}$ expression (Supporting Information, S10). Taken together, these results suggest that mono *o*-substituted SACs are poor inhibitors of tumoral cell growth.

As shown in Figure 8A, dimethyl-substituted SACs harboring at least one *o*-substitution (**8**−**11**) exhibited higher IC_{50} values than SAHA in MTT assay. These high IC_{50} values were correlated to a lack of $p21^{waf/cip1}$ induction. Conversely, *m*-, *p*-dimethyl SACs (**12** and **13**) were as potent as SAHA in inhibiting cell growth and induced $p21^{waf1/cip1}$ to a similar extent (Figure 8A). Similar results were obtained with HepG2 cells (Supporting Information, S11), albeit **5** and **10** were more potent in inducing p21 expression in hepatoma cells than in Caco-2 cells. These observations indicate a strong correlation between the ability of these compounds to inhibit HDACs in vitro and in vivo and their ability to inhibit cell growth in cancer cells. Although IC_{50} values from **13** and SAHA were identical, dose−response curves were different and suggested that **13** is more potent than SAHA at concentrations ranging from 12.5 to 50 μM (Figure 8B). Indeed, Hill slopes were markedly different (−2.03 and −7.16 for SAHA and **13**, respectively), suggesting different modes of interaction with HDACs. The steepest slope observed with **13** could reflect a concentration-dependent stabilization of conformational transitions favoring binding of the inhibitor at high concentrations. Alternatively, binding of **13** to HDACs could induce phase transitions leading to enzyme aggregation, a phenomenon that could be interpreted as increased inhibition. However, slopes observed in in vitro inhibition assays with recombinant enzymes did not differ signicantly between SAHA and **13** (see the Supporting Information, S6), suggesting that the steepest slope observed in the presence of **13** in MTT assay might reflect cell-associated mechanisms.

A

Compound	R	Caco-2 cells IC_{50} (μM)	Caco-2 cells $p21^{waf1/cip1}$ induction 1 μM	5 μM
SAHA		8.1		
1		11.6		
2		78.5		
3		10.7		
4		9.7		
5		40.5		
6		8.3		
7		4.5		
8		57.9		
9		23.8		
10		29.5		
11		>100		
12		5.4		
13		8		

B

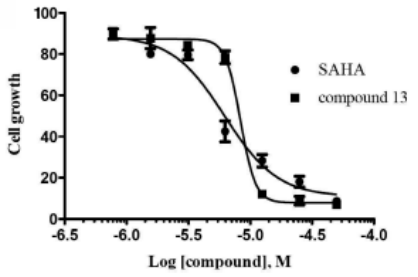


Figure 8. Growth inhibition assays (MTT) and $p21^{waf1/cip1}$ induction in response to SACs treatments in Caco-2 cells. (A) Cytotoxicity assays were performed using SACs concentrations ranging from 0.05 to 100 μM during 48 h of treatment. Results shown are representative of at least two independent experiments. $p21^{waf1/cip1}$ induction was analyzed after 1 or 5 μM SACs treatment during 24 h. IC_{50} = compound concentration leading to 50% of viable cells. $p21^{waf1/cip1}$ expression was monitored using Western blot after 24 h of treatment with 1 and 5 μM compounds. Equal loading between lanes was verified using antiactin antibodies (data not shown). (B) MTT dose−response curves for compound 13 and SAHA-treated Caco-2 cells.

Because **13** was more efficient than SAHA in several in vitro experiments and was partially more efficient in tumoral cell growth inhibition, we decided to compare **13** and SAHA activity toward growth of BJ and MUF nontumoral cells. Indeed, HDACi are known to be poor inducers of apoptosis in normal cells.[27] As expected, both HDACi had a moderate effect on nontumoral cell growth (Figure 9). Indeed, SAHA and **13** IC_{50} values could not be determined (> 100 μM). Cell growth was similarly affected by SAHA and **13** in BJ and MUF cells as judged from the almost superimposable dose−response curves (Figure 9). These data indicate that **13**, beside being more active than SAHA on several biological parameters, is not more detrimental than SAHA to normal cell growth, suggesting that **13** could be a potential candidate for further investigation.

Discussion

Distinguishing characteristics of HDACi include a metal binding moiety, a carbon linker, and a capping group. Considering SAHA, the molecule is characterized as having an anilide as a capping group, a $(CH_2)_6$ chain as a linker, and a hydroxamic group as a metal-chelating agent. On the basis of crystallographic analyses, the capping group is solvent-exposed and could potentially interact with amino acids near the entrance of the active site.[28] Most of the previous HDACi design have emphasized variations of the capping group of SAHA derivatives but with large modifications sampling the surface of the enzyme, such as changing the phenyl group to a benzyl group and adding functionalities in the α-position (α-aminosuberic acid).[29] In contrast, there is no systematic study involving discrete modifications of the phenyl ring while leaving the remainder of the molecule unchanged.

Our data demonstrate the importance of the position at which substituents are placed on the phenyl ring of SAHA. Indeed, *o*-substituted compounds were mostly inactive when compared to SAHA, whereas *m*- and *p*-substituted compounds were often more active than SAHA. The lack of potency of *o*-substituted compounds could be due to their low affinity for HDAC active sites. Indeed, although HDAC active sites might harbor a certain structural plasticity, as suggested by the recent resolution of HDAC4 crystallographic structure,[19] their surface might not be flexible enough to accommodate *o*-substitution of the SAHA phenyl ring.[10] Hence, the lack of potency of *o*-substituted compounds might not be due to steric constraints imposed by the protein moiety. Accordingly, molecular modeling did not reveal clashes between **11** and HDAC7 catalytic site. Interestingly, the collected data rather indicate that *o*-substitution might favor a particular compound conformation, altering both Zn^{2+} coordination and H-bond formation with the carbonyl group of P809. This amino acid, which is strictly conserved among class 2 HDACs only,[19] might be essential to stabilize the binding of HDACi, especially hydroxamic acids such as TSA and SAHA, in the active site of class 2 enzymes. Indeed, modeling data for SAHA and **13** indicate that Zn^{2+} is well-coordinated by the hydoxamic acid function and that the amide NH of SAHA and **13** could establish a H-bond with the carbonyl group of P809. Although molecular modeling allowed us to delineate structural features, which could explain the weak efficiency of *o*-substituted compounds, it did not explain the gain of potency of **13** on various biological parameters when compared to SAHA. However, STD and

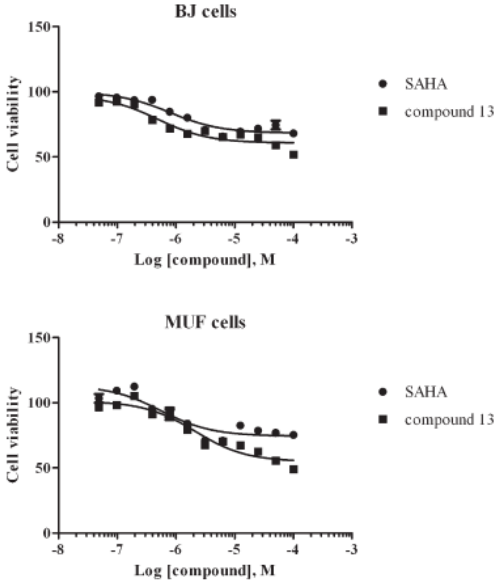


Figure 9. Growth of BJ and MUF nontumoral cells in the presence of SAHA and compound **13**. Cell growth assays were performed using SAHA and **13** concentrations ranging from 0.05 to 100 micromolar during 48 h of treatment.

WaterLOGSY NMR experiments indicated that **13** has a higher affinity for HDAC7 catalytic domain than SAHA. Indeed, analysis of NMR spectra suggested that the K_{off} (first-order dissociation rate constant) of **13** might be lower than the K_{off} of SAHA. Interestingly, although the cap group of SAHA is unresolved in the crystal structure of HDAC7, the protons do show interaction with the protein in STD experiments. This interaction seemed to be stronger when methyl groups were substituted in the *m*-position (**13**). The proximity of F737 to the phenyl ring of **13** could be involved in the strengthening of this interaction. Cocrystallization of HDAC7 with **13** would help to define the structural requirements for its binding to the catalytic site of the enzyme.

The key structural features of amides, especially short C−N bond lengths and hindered rotation about the C−N bond, have been extensively studied.[30] It is now well-established that the *E*/*Z* isomerism arises from delocalization of the lone pair of nitrogen into the π^* orbital of the carbonyl group with some double bond character and slow rotation. The barriers of rotation of amides have been studied mainly by nuclear magnetic resonance, but there are few studies of monosubstituted amides, due in part to the predominance of the *trans* form over the *cis* form.[31] In contrast, data for acetanilide derivatives are quite abundant, and the amount of *cis* isomer increases for acetanilides as the bulk of the *ortho*-substitution increases. For example, it has been reported that the *trans*/*cis* isomer ratio is > 99 for acetanilide, 94/6 for 2-methylacetanilide, and 74/26 for 2,6-dimethylacetanilide. This increase in the *cis* isomer is paralleled by an increase in the barrier to rotation. Although a detailed explanation for the effects of varying the *o*-substituent is still missing, it was proposed that the *o*-substituted ring is out of the amide plane in the ground state and that this effect is amplified as the bulk of the *o*-substituent increases. In some cases, such as with 2,4,6-tri-*t*-butylacetanilide, a 55/45 (*trans*/*cis*) mixture of isomers is obtained, and the isomers can be separated. It is thus likely that *o*-substitution leads to SACs conformations that preclude a stable interaction with the enzymes and possibly to a decrease in the ability of the amide function to establish critical bonds within its immediate environment.

Conclusion

Our data provide clear evidence that minor modifications of the phenyl ring of SAHA can significantly change the properties of the inhibitor. Indeed, we report that the presence of alkyl groups in *o*-position (methyl or ethyl) decreases activity, whereas in *m*- and *p*-position, the presence of alkyl groups leads to a selective increase in activity, as revealed for **12** and **13**. In addition, we provide biophysical evidence that the increase in biological activity of **13** is correlated to an enhanced affinity of the compound toward the class 2 HDAC7 catalytic domain. Nevertheless, our understanding of the molecular mechanisms involved awaits structural information describing interaction between SACs and HDACs at the atomic level. Our data indicate that, when designing novel HDACi derived from SAHA, targeting the cap moiety may provide opportunities to develop more selective and/or more potent molecules, which could be used in physiopathological events such as hypercholesterolemia[32] or in various cancers.[33]

Experimental Section

Chemistry. All reactions were performed under argon. Solvents were distilled from the appropriate drying agent prior to use: tetrahydrofuran (THF) from sodium and benzophenone and MeOH from magnesium. Commercially available reagents were used without further purification unless otherwise stated. All reactions were monitored by TLC with Merck precoated aluminum foil sheets (Silica gel 60 with fluorescent indicator UV_{254}). Compounds were visualized with UV light at 254 and 365 nm. ^{1}H NMR and ^{13}C NMR were recorded using Bruker (Advance 300dpx and 200dpx) spectrometers at 300 (or 200 MHz) and 75 MHz (or 50 MHz), respectively. High-resolution mass spectra were recorded on a ZabSpec TOF Micromass spectrometer in ESI positive mode (compounds **1**−**13**), on a Varian MAT 311 Micromass spectrometer in EI mode (compounds **1a** and **3a**−**13a**) at the CRMPO (Centre Régional de Mesures Physiques de l'Ouest, University of Rennes), and on a Waters Q-TOF2 Micromass spectrometer in ESI positive mode (compound **2a**). Melting points were determined on a Kofler melting apparatus. All tested compounds are >95% pure by elemental analysis (Supporting Information, S12).

All of the syntheses of the SAHA derivatives were adapted from a method previously reported by Mai et al.[14] in three steps. The preparation of *N*-hydroxy-*N*′-(4-methylphenyl)octanediamide (AC22) (**4**) is described as a general example.

Suberic Anhydride (14). A solution of suberic acid (5.0 g, 28.7 mmol) in acetic anhydride (10 mL) was heated at reflux while it was stirred for 1 h. After it was cooled to room temperature, the solvent was evaporated under reduced pressure to give 4.40 g (98%) of a white solid (mp 52 °C). ^{1}H NMR (DMSO-d_6), white solid, mp 52 °C, δ (ppm): 1.34 (m, 4H, $H_{4,5}$), 1.62 (m, 4H, $H_{3,6}$), 2.41 (m, 4H, $H_{2,7}$).

8-[(4-Methylphenyl)amino]-8-oxooctanoic Acid (4a). 4-Methylaniline (343 mg, 3.21 mmol) was added to a stirred solution of suberic anhydride (500 mg, 3.21 mmol) in anhydrous THF. After it was stirred at room temperature for 30 min, the resulting mixture was filtered to give a solid as the bis-amide

(150 mg, 13%). The resulting filtrate was then diluted with water, yielding a white precipitate that was also filtered to give 370 mg (44%) of a white solid; mp 145 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.29 (m, 4H, $2H_4$, and $2H_5$), 1.53 (m, 4H, $2H_3$, and $2H_6$), 2.24 (m, 7H, CH_3−Ar, $2H_2$, and $2H_7$), 7.06 (d, 2H, 2H−Ar), 7.45 (d, 2H, 2H−Ar), 9.78 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 21.3, 25.3, 25.9, 29.3, 34.5, 37.2, 39.1, 119.9, 129.9, 132.6, 137.7, 171.9, 175.4. MS: M^+ calcd for $C_{15}H_{21}NO_3$, 263.1521; found, 263.1510

***N*-Hydroxy-*N*′-(4-methylphenyl)octanediamide (4).** To a 0 °C cooled solution of 8-[(4-methylphenyl)amino]-8-oxooctanoic acid (**4a**) (370 mg, 1.41 mmol) in anhydrous THF, ethyl chloroformiate (213 mg, 1.97 mmol) and triethylamine (213 mg, 2.1 mmol) were added, and the mixture was stirred for 10 min. The solid was filtered off under argon, and the filtrate was added to freshly prepared solution of hydroxylamine in methanol. To prepare the hydroxylamine, a solution of hydroxylamine hydrochloride (293 mg, 4.22 mmol) in methanol (5 mL) was added to a stirred solution of potassium hydroxide (236 mg, 4.22 mmol) in methanol (5 mL) at 0 °C. After it was stirred for 15 min, the precipitate was removed, and the filtrate was used as such. The resulting mixture was stirred at room temperature for 15 min and then was evaporated. After the addition of 10 mL of acetonitrile to the residue, the solution was filtered to give a white solid of (**4**) (390 mg, 44%); mp 152 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.27 (m, 4H, $2H_4$, and $2H_5$), 1.52 (m, 4H, $2H_3$, and $2H_6$), 1.94 (m, 2H, $2H_7$), 2.27 (m, 5H, $2H_2$, and CH_3), 7.06 (d, 2H, 2Ar−H), 7.46 (d, 2H, 2Ar−H), 8.69 (s, 1H, OH), 9.80 (s, 1H, NH), 10.37 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 21.3, 25.9, 29.3, 33.1, 37.2, 78.6, 119.9, 129.9, 132.6, 137.7, 169.9, 171.8. MS: $[M + Na]^+$ calcd for $C_{15}H_{22}N_2O_3Na$, 301.1528; found, 301.1526.

SAHA. White solid, mp 160 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.27 (m, 4H, $2H_4$, and $2H_5$), 1.52 (m, 4H, $2H_3$, and $2H_6$), 1.94 (m, 2H, $2H_7$), 2.29 (m, 2H, $2H_2$), 7.00 (m, 1H, Ar−H), 7.27 (m, 2H, 2Ar−H), 7.59 (d, 2H, 2Ar−H), 8.66 (s, 1H, OH), 9.92 (s, 1H, NH), 10.29 (s, 1H, NH).

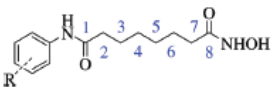
***N*-Hydroxy-*N*′-(4-methoxyphenyl)octanediamide (1).** White solid (51%), mp 151 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.26 (m, 4H, $2H_4$, and $2H_5$), 1.50 (m, 4H, $2H_3$, and $2H_6$), 1.93 (m, 2H, $2H_7$), 2.25 (m, 2H, $2H_2$), 3.70 (m, 3H, OMe), 6.83 (d, 2H, 2Ar−H), 7.48 (d, 2H, 2Ar−H), 8.66 (s, 1H, OH), 9.78 (s, 1H, NH), 10.37 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 25.9, 26.0, 29.3, 33.1, 37.1, 56.0, 114.6, 121.4, 133.5, 155.8, 170.0, 171.6. MS: $[M + Na]^+$ calcd for $C_{15}H_{22}N_2O_4Na$, 317.1477; found, 317.1477.

***N*-Hydroxy-*N*′-(2-methylphenyl)octanediamide (2).** White solid (38%), mp 135 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.31 (m, 4H, $2H_4$, and $2H_5$), 1.54 (m, 4H, $2H_3$, and $2H_6$), 1.95 (t, 2H, $2H_7$), 2.18 (s, 3H, CH_3), 2.32 (m, 5H, $2H_2$, and CH_3), 7.00−7.30 (m, 4H, 4Ar−H), 8.67 (s, 1H, OH), 9.23 (s, 1H, NH), 10.35 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 18.7, 25.9, 26.1, 29.3, 33.1, 33.6, 36.6, 125.9, 126.1, 126.7, 131.1, 132.7, 137.3, 170.0, 172.0. MS: $[M + Na]^+$ calcd for $C_{15}H_{22}N_2O_3Na$, 301.1528; found, 301.1525.

***N*-Hydroxy-*N*′-(3-methylphenyl)octanediamide (3).** White solid (52%), mp 118 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.26 (m, 4H, $2H_4$, and $2H_5$), 1.48 (m, 2H, $2H_6$), 1.56 (m, 2H, $2H_3$), 1.93 (t, 2H, $2H_7$), 2.25 (m, 5H, $2H_2$, and CH_3), 6.83 (d, 1H, Ar−H), 7.14 (t, 1H, Ar−H), 7.36 (d, 1H, Ar−H), 7.43 (s, 1H, Ar−H), 8.66 (s, 1H, OH), 9.79 (s, 1H, NH), 10.33 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 22.1, 25.9, 29.2, 33.1, 37.2, 117.1, 120.4, 124.5, 129.3, 138.6, 140.1, 169.9, 172.1. MS: $[M + Na]^+$ calcd for $C_{15}H_{22}N_2O_3Na$, 301.1528; found, 301.1523

***N*-(2-Ethylphenyl)-*N*′-hydroxyoctanediamide (5).** White solid (71%), mp 116 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.10 (t, 3H, CH_2CH_3), 1.29 (m, 4H, $2H_4$, and $2H_5$), 1.53 (m, 4H, $2H_3$, and $2H_6$), 1.95 (t, 2H, $2H_7$), 2.31 (t, 2H, $2H_2$), 2.51 (q, 2H, CH_2CH_3), 7.21 (m, 4H, 4Ar−H), 8.67 (s, 1H, OH), 9.30 (s, 1H, NH), 10.39 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 15.2, 24.7, 25.9, 26.0, 26.2, 29.3, 33.2, 36.6, 122.5, 126.7, 127.2, 129.3, 136.7, 139.0, 170.0, 172.3. MS: $[M + Na]^+$ calcd for $C_{16}H_{24}N_2O_3Na$, 315.1685; found, 315.1680

Scheme 1. ^{1}H Atom-Labeling Used for SACs



***N*-(3-Ethylphenyl)-*N*′-hydroxyoctanediamide (6).** White solid (62%), mp 120 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.15 (t, 3H, CH_2CH_3), 1.26 (m, 4H, $2H_4$, and $2H_5$), 1.48 (m, 2H, $1H_3$, and $1H_6$), 1.55 (m, 2H, $1H_3$, and $1H_6$), 1.93 (m, 2H, $2H_7$), 2.27 (t, 2H, $2H_2$), 2.51 (q, 2H, CH_2CH_3), 6.86 (d, 1H, Ar−H), 7.16 (t, 1H, Ar−H), 7.40 (d, 1H, Ar−H), 7.45 (s, 1H, Ar−H), 8.65 (s, 1H, OH), 9.86 (s, 1H, NH), 10.35 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 16.5, 26.0, 29.2, 29.3, 33.1, 37.2, 117.4, 119.3, 123.3, 129.4, 140.3, 145.0, 170.0, 172.1. MS: $[M + Na]^+$ calcd for $C_{16}H_{24}N_2O_3Na$, 315.1685; found, 315.1673

***N*-(4-Ethylphenyl)-*N*′-hydroxyoctanediamide (7).** White solid (97%), mp 146 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.14 (t, 3H, CH_2CH_3), 1.26 (m, 4H, $2H_4$, and $2H_5$), 1.49 (m, 4H, $2H_3$, and $2H_6$), 1.93 (m, 2H, $2H_7$), 2.27 (t, 2H, $2H_2$), 2,51 (q, 2H, CH_2CH_3), 7.10 (d, 2H, 2Ar−H), 7.49 (d, 2H, 2Ar−H), 8.65 (s, 1H, OH), 9.84 (s, 1H, NH), 10.35 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 21.3, 25.9, 29.3, 33.1, 37.2, 78.6, 119.9, 129.9, 132.6, 137.7, 169.9, 171.8. MS: $[M + Na]^+$ calcd for $C_{16}H_{24}N_2O_3Na$, 315.1685; found, 315.1692

***N*-(2,3-Dimethylphenyl)-*N*′-hydroxyoctanediamide (8).** White solid (50%), mp 129 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.30 (m, 4H, $2H_4$, and $2H_5$), 1.54 (m, 4H, $2H_3$, and $2H_6$), 1.95 (t, 2H, $2H_7$), 2.05 (s, 3H, CH_3), 2.24 (s, 3H, CH_3), 2.30 (t, 2H, $2H_2$), 7.03 (m, 3H, 3Ar−H), 8.68 (s, 1H, OH), 9.36 (s, 1H, NH), 10.40 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 15.0, 21.1, 26.0, 26.2, 29.3, 33.2, 36.6, 124.6, 126.6, 127.6, 132.1, 137.1, 137.7, 170.2, 172.1. MS: $[M + Na]^+$ calcd for $C_{16}H_{24}N_2O_3Na$, 315.1685; found, 315.1685

***N*-(2,4-Dimethylphenyl)-*N*′-hydroxyoctanediamide (9).** White solid (57%), mp 117 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.30 (m, 4H, $2H_4$, and $2H_5$), 1.53 (m, 4H, $2H_3$, and $2H_6$), 1.95 (t, 2H, $2H_7$), 2.13 (s, 3H, CH_3), 2.24 (s, 3H, CH_3), 2.30 (t, 2H, $2H_2$), 6.94 (d, 1H, 1Ar−H), 6.99 (s, 1H, 1Ar−H), 7.19 (d, 1H, 1Ar−H), 8.67 (s, 1H, OH), 9.18 (s, 1H, NH), 10.37 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 18.7, 21.6, 26.0, 26.2, 29.3, 33.1, 32.6, 39.1, 126.2, 127.2, 131.6, 132.7, 134.7, 134.9, 135.1, 170.0, 172.0. MS: $[M + Na]^+$ calcd for $C_{16}H_{24}N_2O_3Na$, 315.1685; found, 315.1688

***N*-(2,5-Dimethylphenyl)-*N*′-hydroxyoctanediamide (10).** White solid (52%), mp 134 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.30 (m, 4H, $2H_4$, and $2H_5$), 1.54 (m, 4H, $2H_3$, and $2H_6$), 1.95 (t, 2H, $2H_7$), 2.12 (s, 3H, CH_3), 2.24 (s, 3H, CH_3), 2.30 (t, 2H, $2H_2$), 6.87 (d, 1H, 1Ar−H), 7.05 (d, 1H, 1Ar−H), 7.17 (s, 1H, 1Ar−H), 8.67 (s, 1H, OH), 9.20 (s, 1H, NH), 10.37 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 18.4, 21.5, 26.0, 26.2, 29.3, 33.1, 33.7, 122.5, 126.6, 129.5, 130.9, 135.7, 137.2, 170.0, 172.0. MS: $[M + Na]^+$ calcd for $C_{16}H_{24}N_2O_3Na$, 315.1685; found, 315.1678

***N*-(2,6-Dimethylphenyl)-*N*′-hydroxyoctanediamide (11).** White solid (95%), mp 133 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.31 (m, 4H, $2H_4$, and $2H_5$), 1.50 (m, 4H, $2H_3$, and $2H_6$), 1.95 (t, 2H, $2H_7$), 2.12 (s, 6H, CH_3), 2.30 (t, 2H, $2H_2$), 7.05 (s, 3H, 3Ar−H), 8.81 (s, 1H, OH), 9.23 (s, 1H, NH), 10.39 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 19.0, 26.0, 26.3, 29.3, 29.4, 33.1, 36.2, 127.1, 128.5, 136.0, 136.3, 170.0, 171.7. MS: $[M + Na]^+$ calcd for $C_{16}H_{24}N_2O_3Na$, 315.1685; found, 315.1677

***N*-(3,4-Dimethylphenyl)-*N*′-hydroxyoctanediamide (12).** White solid (42%), mp 134 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.26 (m, 4H, $2H_4$, and t $2H_5$), 1.53 (m, 4H, $2H_3$, and $2H_6$), 2.16 (m, 10H, $2CH_3$, $2H_2$, and $2H_7$), 7.00 (d, 1H, Ar−H), 7.28 (d, 1H, Ar−H), 7.35(s, 1H, Ar−H), 8.66 (s, 1H, OH), 9.71 (s, 1H, NH), 10.35 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 19.6, 20.5, 23.9, 26.0,

29.3, 33.1, 117.5, 121.2, 130.3, 131.4, 137.0, 138.0, 170.0, 171.9. MS: $[M + Na]^+$ calcd for $C_{16}H_{24}N_2O_3Na$, 315.1685; found, 315.1673

***N*-(3,5-Dimethylphenyl)-*N*-hydroxyoctanediamide (13).** White solid (53%), mp 171 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.27 (m, 4H, $2H_4$, and $2H_5$), 1.51 (m, 4H, $2H_3$, and $2H_6$), 1.94 (t, 2H, $2H_7$), 2.21 (s, 6H, CH_3), 2.27 (t, 2H, $2H_2$), 6.65 (s, 1H, 1Ar−H), 7.22 (s, 2H, 2 Ar−H), 8.67 (s, 1H, OH), 9.76 (s, 1H, NH), 10.35 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 22.0, 25.9, 26.0, 29.3, 33.1, 37.2, 117.7, 125.3, 138.4, 140.1, 170.0, 172.0. MS: $[M + Na]^+$ calcd for $C_{16}H_{24}N_2O_3Na$, 315.1685; found, 315.1677

8-[(4-Methoxyphenyl)amino]-8-oxooctanoic Acid (1a). White solid (21%), mp 144 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.29 (m, 4H, $2H_4$, and $2H_5$), 1.51 (m, 4H, $2H_3$, and $2H_6$), 2.22 (m, 4H, $2H_2$, and $2H_7$), 3.70 (s, 3H, OCH_3), 6.84 (m, 2H, 2Ar−H), 7.47 (m, 2H, 2Ar−H), 9.70 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 25.3, 25.9, 29.2, 29.3, 34.5, 37.1, 56.0, 114.6, 121.4, 133.4, 155.8, 171.5, 175.4. MS: $[M + Na]^+$ calcd for $C_{15}H_{21}NO_4$, 279.1471; found, 279.1446

8-[(2-Methylphenyl)amino]-8-oxooctanoic Acid (2a). White solid (41%), mp 126 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.31 (m, 4H, $2H_4$, and $2H_5$), 1.53 (m, 4H, $2H_3$, and $2H_6$), 2.00−2.35 (m, 7H, $2H_2$, $2H_7$, and CH_3), 6.95−7.40 (m, 4H, 4Ar−H), 9.24 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 18.7, 25.2, 26.1, 29.2, 34.5, 36.6, 126.0, 126.2, 126.7, 131.1, 132.7, 137.3, 172.1, 175.4. MS: M^+ calcd for $C_{15}H_{21}NO_3Na$, 286.1419; found, 286.1418

8-[(3-Methylphenyl)amino]-8-oxooctanoic Acid (3a). White solid (26%), mp 102 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.29 (m, 4H, $2H_4$, and $2H_5$), 1.53 (m, 4H, $2H_3$, and $2H_6$), 2.25 (m, 7H, $2H_2$, $2H_7$, and CH_3), 6.83 (d, 1H, Ar−H), 7.15 (t, 1H, Ar−H), 7.36 (d, 1H, Ar−H), 7.43 (s, 1H, Ar−H), 9.75 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 22.1, 25.3, 25.9, 29.2, 29.3, 34.5, 37.2, 117.1, 120.4, 124.5, 129.3, 138.7, 140.1, 172.1, 175.4. MS: M^+ calcd for $C_{15}H_{21}NO_3$, 263.1521; found, 263.1510

8-[(2-Ethylphenyl)amino]-8-oxooctanoic Acid (5a). White solid (26%), mp 102 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.11 (t, 3H, CH_2CH_3), 1.31 (m, 4H, $2H_4$, and $2H_5$), 1.55 (m, 4H, $2H_3$, and $2H_6$), 2.21 (t, 2H, $2H_7$), 2.31 (t, 2H, $2H_2$), 2.50 (q, 2H, CH_2CH_3), 7.22 (m, 4H, 1Ar−H), 9.24 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 16.5, 26.1, 26.7, 27.5, 30.6, 35.9, 37.9, 127.8, 128.1, 128.5, 130.7, 138.0, 140.3, 173.7, 176.8. MS: M^+ calcd for $C_{16}H_{23}NO_3$, 277.1678; found, 277.1686

8-[(3-Ethylphenyl)amino]-8-oxooctanoic Acid (6a). White solid (38%), mp 102 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.16 (t, 3H, CH_2CH_3), 1.19 (m, 4H, $2H_4$, and $2H_5$), 1.54 (m, 4H, $2H_3$, and $2H_6$), 2.23 (m, 4H, $2H_2$, and $2H_7$), 2.54 (q, 2H, CH_2CH_3), 6.86 (d, 1H, Ar−H), 7.18 (t, 1H, Ar−H), 7.41 (d, 1H, Ar−H), 7.44 (s, 1H, Ar−H), 9.79 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 16.4, 25.3, 25.9, 29.2, 29.3, 34.5, 37.2, 117.3, 119.3, 123.3, 124.5, 129.4, 140.2, 145.0, 172.1, 175.4. MS: M^+ calcd for $C_{16}H_{23}NO_3$, 277.1678; found, 277.1686

8-[(4-Ethylphenyl)amino]-8-oxooctanoic Acid (7a). White solid (19%), mp 138 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.15 (t, 3H, CH_2CH_3), 1.29 (m, 4H, $2H_4$, and $2H_5$), 1.53 (m, 4H, $2H_3$, and $2H_6$), 2.30 (m, 4H, $2H_2$, and $2H_7$), 2.53 (q, 2H, CH_2CH_3), 7.11 (d, 2H, 2Ar−H), 7.48 (d, 2H, 2Ar−H), 9.77 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 16.7, 25.3, 25.9, 29.2, 34.5, 37.2, 120.0, 128.7, 137.9, 139.1, 171.8, 175.4. MS: M^+ calcd for $C_{16}H_{23}NO_3$, 277.1678; found, 277.1686

8-[(2,3-Dimethylphenyl)amino]-8-oxooctanoic Acid (8a). White solid (42%), mp 142 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.32 (m, 4H, $2H_4$, and $2H_5$), 1.55 (m, 4H, $2H_3$, and $2H_6$), 2.05 (t, 3H, CH_3), 2.24 (m, 7H, $2H_2$, $2H_7$, and CH_3), 7.03 (m, 3H, 3Ar−H), 9.28 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 14.9, 21.1, 25.3, 26.1, 29.2, 29.3, 34.6, 36.6, 124.6, 126.0, 127.7, 132.1, 137.1, 137.7, 172.0, 175.4. MS: M^+ calcd for $C_{16}H_{23}NO_3$, 277.1678; found, 277.1658

8-[(2,4-Dimethylphenyl)amino]-8-oxooctanoic Acid (9a). White solid (32%), mp 122 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.31 (m, 4H, $2H_4$, and $2H_5$), 1.51 (m, 4H, $2H_3$, and $2H_6$), 2.13 (s, 3H, CH_3), 2.17 (t, 2H, $2H_7$), 2.24 (s, 3H, CH_3), 2.28 (t, 2H, $2H_2$), 6.94 (d, 1H, 1Ar−H), 7.00 (s, 1H, 1Ar−H), 7.20 (d, 1H, Ar−H), 9.16 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 18.7, 21.3, 25.3, 26.1, 26.2, 29.2, 29.3 29.4, 34.5, 36.5, 126.2, 127.2, 131.5, 132.6, 134.7, 135.0, 172.0, 175.4. MS: M^+ calcd for $C_{16}H_{23}NO_3$, 277.1678; found, 277.1686

8-[(2,5-Dimethylphenyl)amino]-8-oxooctanoic Acid (10a). White solid (48%), mp 128 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.31 (m, 4H, $2H_4$, and $2H_5$), 1.54 (m, 4H, $2H_3$, and $2H_6$), 2.13 (m, 5H, $2H_7$, and CH_3), 2.22 (m, 5H, $2H_2$, and CH_3), 6.88 (d, 1H, Ar−H), 7.06 (d, 1H, Ar−H), 7.17 (s, 1H, 1Ar−H), 9.17 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 18.3, 21.5, 25.3, 26.1, 29.2, 29.3, 34.6, 36.6, 126.6, 129.5, 130.9, 135.7, 137.1, 172.0, 175.4. MS: M^+ calcd for $C_{16}H_{23}NO_3$, 277.1678; found, 277.1686

8-[(2,6-Dimethylphenyl)amino]-8-oxooctanoic Acid (11a). White solid (22%), mp 129 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.32 (m, 4H, $2H_4$, and $2H_5$), 1.55 (m, 4H, $2H_3$, and $2H_6$), 2.12 (s, 6H, CH_3), 2.21 (t, 2H, $2H_7$), 2.30 (t, 2H, $2H_2$), 7.05 (s, 3H, 3Ar−H), 9.18 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 19.1, 25.3, 26.2, 29.2, 29.4, 34.5, 36.2, 127.1, 128.5, 136.0, 136.3, 171.6, 175.4. MS: M^+ calcd for $C_{16}H_{23}NO_3$, 277.1678; found, 277.1658

8-[(3,4-Dimethylphenyl)amino]-8-oxooctanoic Acid (12a). White solid (56%), mp 120 °C, ^{1}H NMR (DMSO-d_6) δ (ppm): 1.29 (m, 4H, $2H_4$, and $2H_5$), 1.53 (m, 4H, $2H_3$, and $2H_6$), 2.22 (m, 10H, 2 CH_3−Ar, $2H_2$, and $2H_7$), 7.00 (d, 1H, Ar−H), 7.32 (d, 1H, Ar−H), 7.36 (s, 2H, 2Ar−H), 9.69 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 19.6, 20.5, 25.3, 25.9, 29.3, 34.5, 37.2, 117.5, 121.2, 130.3, 131.4, 137.0, 138.0, 171.8, 175.4. MS: M^+ calcd for $C_{16}H_{23}NO_3$, 277.1678; found, 277.1686

8-[(3,5-Dimethylphenyl)amino]-8-oxooctanoic Acid (13a). White solid (30%), mp 123 °C. ^{1}H NMR (DMSO-d_6) δ (ppm): 1.28 (m, 4H, $2H_4$, and $2H_5$), 1.53 (m, 4H, $2H_3$, and $2H_6$), 2.22 (m, 10H, $2H_2$, $2H_7$, and CH_3), 6.66 (s, 1H, 1Ar−H), 7.18 (s, 2H, 2Ar−H), 9.69 (s, 1H, NH). ^{13}C NMR (DMSO-d_6) δ (ppm): 22.0, 25.3, 26.0, 29.3, 34.5, 37.2, 117.7, 125.3, 138.4, 140.1, 172.0, 175.4. MS: M^+ calcd for $C_{16}H_{23}NO_3$, 277.1678; found, 277.1686

Cell Lines and Drug Treatment. All compounds (**1**−**13**, including SAHA) were dissolved in DMSO (dimethyl sulfoxide, Sigma) at 10 mM as a stock solution and stored at −20 °C. The Caco-2 cell line (ATCC number HTB-37), derived from a human colorectal adenocarcinoma, as well as HepG2 cell line (ATCC number HB-8065) were maintained under the conditions recommended by the suppliers at 37 °C and 5% CO_2. MUF is a primary cell line of human foreskin diploid fibroblasts created by Dr. Piršel (Cancer Research Institute, Slovak Academy of Sciences, Slovak Republic). Human fibroblasts BJ (ATCC number CRL-2522) were established from skin taken from normal foreskin. MUF and BJ cell lines were normal and nontransformed and, therefore, suitable as controls to tumor cell lines. MUF and BJ cell lines were cultured in DMEM supplemented with antibiotics and 10% fetal calf serum. For protein extract preparation and Western blotting analysis, cells were plated 1 day before drug treatment in six-well plates at a density of 3×10^5 cells per well. Cells were treated with desired compound at concentrations and times indicated in the figure captions.

Ex Vivo HDAC Activity. All compounds were evaluated for their ability to inhibit total HDAC activity ex vivo in Caco-2 cells using the HDAC Fluorimetric Cellular Activity Assay Kit (Biomol International) and employing TSA as a positive control. This assay is based on the use of Fluor de Lys fluorescent substrate, which is cell-permeable and a substrate for most class 1 and 2 HDACs. Caco-2 cells were seeded in triplicate in 96-well plates at a density of 3×10^4 cells per well and allowed to attach overnight (70% confluency). Media were replaced with 50 μL per well of media containing 200 μM Fluor de Lys substrate and 1 μM TSA as a positive control. Plates were incubated at 37 °C and 5% CO_2 for 4 h. To terminate HDAC activity and begin development of the fluorescence signal, 50 μL per well of the 1× Developer (Biomol International) diluted in cold lysis

buffer [50 mM Tris/HCl (pH 8.0), 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$, and 1% NP-40] + 2 μM TSA was added and mixed by pipetting. After developer was added, plates were incubated for an additional 15 min at 37 °C, and the fluorescence was read (excitation 365 nm and emission 450 nm) in a microplate-reading fluorimeter (Fluorolite 1000, Dynatech Laboratories). Comparison of inhibited vs control relative fluorescence was employed to determine the percent HDAC activity remaining.

Indirect ELISA. Whole cell extracts were prepared using a Nuclear Extract Kit (Active Motif) according the manufacturer's instructions using TSA (1 μM) in each step to avoid protein deacetylation. Protein extracts (antigen) were diluted to a final concentration of 20 μg/mL in PBS, and wells of a PVC microtiter plate were coated with 1 μg of antigen. Plates were covered and incubated for 2 h at room temperature. Then, after the coating solution was removed, plates were washed three times by filling the wells with 200 μL of PBS. The solutions or washes were removed by flicking the plate over a sink. The remaining protein binding sites in the coated wells were blocked by adding 200 μL of blocking buffer [5% bovine serum albumin (BSA) in PBS] per well. Plates were covered and incubated for 2 h at room temperature. Plates were then washed twice with PBS, and 100 μL of antiacetyl lysines antibody (from rabbit, diluted 1/1000, Abcam) diluted in blocking buffer was added to each well. Plates were covered and incubated overnight at 4 °C. Subsequently, plates were washed four times with PBS, and 100 μL of conjugated secondary antibody (antirabbit, diluted 1/2500, Abcam) diluted in blocking buffer was added to each well. Then, plates were covered and incubated 2 h at room temperature. After four washes, the horseradish peroxidase (HRP) activity was measured using the colorimetric kit (Substrate Reagent Pack, R&D Systems) following the supplier's recommendations.

Western Blot Analysis. After the media were removed from six-well plates, cells were washed with PBS and harvested in 150 μL of PBS using a scraper. Protein concentrations were directly quantified in cell suspension using DC Protein Assay Kit (Biorad) according to the manufacturer's instructions. After centrifugation (5 min, 200g, 4 °C), cell pellets were resuspended in Laëmmli loading buffer to obtain the desired protein concentration. Then, samples were denatured by sonication (bioruptor, Diagenode, 7 min, 4 °C, max power) and boiling (10 min, 100 °C). In all Western blotting analyses, equal protein quantities were loaded on 12% SDS-PAGE, and electrophoresis separation was followed by proteins transfer on nitrocellulose membranes (Hybond-C Extra, Amersham Biosciences). The quality of protein transfer was systematically verified using red Ponceau staining (data not shown). Membranes were probed with appropriate primary antibodies in milk saturation buffer (PBS/0.1% Tween, 5% milk) for anti-$p21^{waf1/cip1}$ (from mouse, diluted 1/1000, Millipore), antiactin (from mouse, diluted 1/5000, Abcam), antiacetylated H3 (from rabbit, diluted 1/25000, Abcam), antihistone H3 (from rabbit, diluted 1/2000, Abcam), antiacetylated α-tubulin (from mouse, diluted 1/1000, Abcam), and anti-α-tubulin (from mouse, diluted 1/1000, Sigma) or in BSA saturation buffer (PBS/0.1% Tween, 5% BSA) for antiacetyl lysine (from rabbit, diluted 1/1000, Abcam). The secondary antibodies corresponding to antimouse (diluted 1/10000, Santa Cruz) or antirabbit (diluted 1/2500, Abcam) were coupled to HRP. The signal was enhanced by chemiluminescence using Immobilon (Millipore) as a substrate for HRP and detected on film (Biomax Film, Kodak) or camera (Biorad).

Compound Profiling on HDAC1, HDAC3, HDAC5, HDAC6, and HDAC7. HDAC1, HDAC3, and HDAC6 enzymes, HDAC substrates, HDAC buffer, and developer were from BPS Bioscience (San Diego, CA). Human HDAC5 and HDAC7 were recombinantly expressed in Sf9+ insect cells as chimeric fusions with N-terminal GST and C-terminal HIS tag and then purified by glutathione-sepharose and Ni-NTA Superflow chromatography. HDAC enzymes were diluted in HDAC buffer in the presence of 0.01% (w/v) BSA (Sigma) at the following working concentrations (1×): 75 nM HDAC1, 1 nM HDAC3, 1 nM HDAC5, 3 nM HDAC6, and 0.5 nM HDAC7. The respective enzyme concentrations were optimized to ensure the detection of the HDAC activities in the linear range of the reaction for at least 60 min, with a consumption of less than 10% of the initial substrate. Specific fluorogenic substrates were used at working concentrations equivalent to their respective $2 \times K_m$ values. In particular, the class 1 HDAC substrate (#50032) was used at 12 μM (HDAC1 and HDAC3), the class 2a HDAC substrate (#50040) was used at 4 μM (HDAC5 and HDAC7), while the specific HDAC6 substrate (#50037) was used at 30 μM. Compounds dissolved in DMSO were assayed in quadruplicate at 11 concentrations ranging from 0.316 nM to 31.6 μM for HDAC1, HDAC3, and HDAC6 and from 6.32 nM to 632 μM for HDAC5 and HDAC7. All of the dilution steps were robotically performed with a Matrix Control Mate (Thermo Scientific, Hudson, NH). The reaction was incubated for 45 min at 32 °C and stopped with developer 2× in the presence of 8 μM TSA (Sigma). Plates were further incubated at 32 °C for 30 min. The fluorescence intensity was detected with Safire2 (Tecan, Mannedorf, Switzerland) at gain 40 and with excitation−emission wavelengths of 353−450 nm, respectively. End point relative fluorescence units (RFU) data were transformed as a percentage of activity to compare the relative potency of the compounds on the different HDAC enzymes. Curve fitting was performed by GraphPad Prism 4.0, using a sigmoidal dose−response (variable slope) model. The average IC_{50} values were obtained as a minimum from two independent experiments (for details, see the Supporting Information, S4−S6).

Protein Expression and Purification of HDAC7. The catalytic domain of human HDAC7 (Uniprot: Q8WUI4; amino acids 483−903 with an expression-enhancing NASNNG C-terminal sequence) was expressed from a pET9-derived vector encoding an N-terminal TEV cleavable hexahistidine tag. *Escherichia coli* BL21 AI (Invitrogen) cotransformed with RIL plasmid (Stratagene) was used for protein expression. Protein was purified as described[19] with minor modifications (Supporting Information, S13).

Acquisition of NMR Spectra. All NMR spectra were recorded at 298 K with a spectral width of 14 ppm on Bruker Avance 500 MHz spectrometer, equipped with a 5 mm TXI inverse triple-resonance cryoprobe head. (PRISM platform, University of Rennes 1, France). Spectra processing was performed using TOPSPIN 1.3 software (Bruker). The in vitro experiments were performed as already described.[34]

STD Experiments. Selective saturation of the protein was achieved by a train of Gauss-shaped pulses of 50 ms length each, truncated at 1%, and separated by a 1 ms delay. Forty selective pulses were applied, leading to a total length of the saturation train of 2 s. The on-resonance irradiation of the protein was performed at a chemical shift of 1.3 ppm. Off-resonance irradiation was set at 40 ppm, where no protein signals were present. Unwanted magnetization water was suppressed using a WATERGATE 3-9-19 pulse sequence. A 30 ms spin lock pulse was applied to remove protein background signals. The total scan number in the STD experiments was 2000, acquired as 16 blocks of 128 scans each. The spectra on- and off-resonances were subtracted after multiplication by an exponential line-broadening function of 0.8 Hz prior to Fourier transformation.

WaterLOGSY Experiments. WaterLOGSY experiments were performed using a selective excitation with a 180° shaped pulse at H_2O position. The first water selective 180° pulse was applied during 5 ms. The first two pulsed field gradients (PFGs) had a typical duration of 2 ms. A weak rectangular PFG is applied during the entire length of the mixing time (1 s). A short gradient recovery time of 1 ms is applied at the end of the mixing time before the detection pulse. The water suppression in

both experiments was achieved with the excitation sculpting sequence. The data were collected with a sweep width of 14 ppm, an acquisition time of 0.585 s, and a relaxation delay of 2 s. Prior to Fourier transformation, the data were multiplied with an exponential function with a line-broadening of 0.8 Hz.

Molecular Modeling. Molecular modeling studies were performed with Macromodel 8[35] on a Intel Xeon dual core with Fedora Core 7 as the operating system. The OPLSAA-2001 force field[35,36] was applied for energy minimization and molecular dynamics simulation. Energy minimization protein−ligand complexes were performed with the conjugate gradient method[37] to an energy gradient of 0.001 kJ/(mol Å) with distant-dependent electrostatic treatment. The three ligands SAHA, **11**, and **13** were built by the Build tool and energy minimized with previous force field and parameters.

Protein−Ligand Complex Simulations. The crystals complexes HDAC7−SAHA (PDB ID: 3C0Z) and HDAC7−TSA (PDB ID: 3C10) have been downloaded from the public RCSB Protein Data Bank (www.rcdb.org).[38] The chains A of all complexes were used for all of the simulations. In all three complexes of HDAC7−SAHA, the ligands were not completely resolved, and particularly, the *N*-phenylacetamide was not present. This precludes evaluation of the interactions of the *N*-phenylacetamide moiety with the amino acids of the protein, and using the complex HDAC7−SAHA with a truncate ligand in the preliminary minimization phases could bring partial occlusion of the entrance of the cavity by the closest amino acids side chains. After these considerations from both crystal cells, the two A chains have been selected and superimposed (rmsd on the heavy atoms of the backbone was 0.134 Å) to evaluate the peculiar differences in the amino acid side chain orientations or in the tertiary structure near the binding cavity and in the receptor surface close to the entrance of the cavity. After this checking phase, where no important differences were found, the chain A of the complex HDAC7−TSA was selected for the successive simulations. All hydrogens were inserted into the structure of chain A of 3C10 complex, all water molecules were removed, the $N^{\delta}H$ tautomer of the H709 was chosen, and a partial charge of +2 was assigned to the zinc ions. The chain A, with TSA, was energetically minimized with OPLSAA-2001 force field keeping the entire backbone fixed. The TSA ligand was removed from the minimized complex, and the truncated SAHA has been introduced as a guideline for the successive insertion of the complete SAHA, **11**, and **13**. The three complexes HDAC7−SAHA, HDAC7−**11**, and HDAC7−**13** were generated, and their energies were minimized. All of the minimum energy conformations of the complexes were submitted to dynamic simulations with the OPLSAA-2001 force field at 300 K, with an 10 ps of equilibration, and 100 ps of production time with a time step of 1 fs collecting 100 snapshots every 1 ps. Only the amino acid side chains inside a sphere of 5 Å from the ligand were free to move, while the rest of the amino acid side chains and the entire backbone were kept fixed. The average structures obtained from 10 representative snapshots of MD trajectory were energy minimized with the OPLSAA-2001 force field to a gradient of 0.001 kJ/(mol Å), with a fixed protein backbone.

Growth Inhibition Assays. One day prior to treatment, Caco-2, HepG2, and normal human fibroblast cells (MUF and BJ cells) were plated on 96-well plates in appropriate media, at densities of 2.104 cells/well for Caco-2 and HepG2 cells and 4.103 cells/well for MUF and BJ cells. Cells were exposed to the tested compounds for 48 h at concentrations ranging from 100 to 0.05 μM with 2-fold dilutions. DMSO was used as a negative control with the same dilutions starting from 1%. After treatment for 48 h, 10 μL of MTT (thiazolyl blue, 5 mg/mL) was added to each well and cultured for 2 h. Then, after the media were removed, formazan crystals were dissolved in 100 μL of DMSO. The absorbance was measured at 570 nm, and the background was measured at 630 nm using a microplate reader (Power Wave X5; Biotek). Data were acquired using KC4 version 3.4 software (Biotek Instruments Inc.). Results (% of cell growth) were calculated as follows: $[(A_{570} - A_{630})_{compound}/(A_{570} - A_{630})_{DMSO}] \times 100$. Results were averaged from three independent experiments for Caco-2 cells and from at least six independent experiments for BJ and MUF cells. Curve fitting was performed by GraphPad Prism 4.0.

Acknowledgment. We thank Maud Bizot and Gaëlle Palierne for technical assistance. We thank Jean-François Geffroy, Anne-Sophie Julien, Alicia Cassant, and Elodie Guiheux for their technical support. We thank Dr. Jérome Eeckhoute for his precious help with the writing of the manuscript. MUF cells were kindly provided by Dr. Miroslav Piršel. This work was funded by the European Community (LSHM-CT-2006-037498), the Région Bretagne (SIE 211-B3-11), the Centre National pour la Recherche Scientifique, and the Ministère de l'Enseignement Supérieur et de la Recherche.

Supporting Information Available: Acetylation level of histones H3/H4 and α-tubulin in response to SACs treatment (S1), materials and methods for histone preparation (S2), Western blot analysis of protein acetylation in HepG2 cells treated with SACs (S3), compound handling and pipetting scheme for compound profiling (S4), *Z′* factor evaluation[39] (S5), dose−reponse curves on HDAC1, -3, -5, -6, and -7 using SACs compounds (S6), modeling experiments suggesting a conserved protein−ligand interaction mode between the cap domain of SAHA and class 2 HDACs, method for HDAC4−SACs complex modeling (S7), importance of HDAC7 proline 809 and HDAC4 proline 274 for ligand interaction (S8), lengthening of the carbon chain on the phenyl ring of SAHA leading to eratic docking (S9), $p21^{waf1/cip1}$ induction in response to increasing concentrations (5, 50, and 100 μM) of representative *o*-substituted SACs (S10), MTT assay and $p21^{waf1/cip1}$ induction in HepG2 cells treated with SACs (S11), table of RP-HPLC purity and microanalytical data (S12), and complete material and methods for recombinant HDAC7 preparation and purification (S13). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Yang, X. J.; Seto, E. The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 206–218.
(2) Michan, S.; Sinclair, D. Sirtuins in mammals: Insights into their biological function. *Biochem. J.* **2007**, *404*, 1–13.
(3) Lu, Q.; Qiu, X.; Hu, N.; Wen, H.; Su, Y.; Richardson, B. C. Epigenetics, disease, and therapeutic interventions. *Ageing Res. Rev.* **2006**, *5*, 449–467.
(4) Paris, M.; Porcelloni, M.; Binaschi, M.; Fattori, D. Histone deacetylase inhibitors: from bench to clinic. *J. Med. Chem.* **2008**, *51*, 1505–1529.
(5) Bolden, J. E.; Peart, M. J.; Johnstone, R. W. Anticancer activities of histone deacetylase inhibitors. *Nat. Rev. Drug. Discovery* **2006**, *5*, 769–784.
(6) Mai, A.; Massa, S.; Lavu, S.; Pezzi, R.; Simeoni, S.; Ragno, R.; Mariotti, F. R.; Chiani, F.; Camilloni, G.; Sinclair, D. A. Design, Synthesis, And Biological Evaluation of Sirtinol Analogues as Class III Histone/Protein Deacetylase (Sirtuin) Inhibitors. *J. Med. Chem.* **2005**, *48*, 7789–7895.
(7) Marks, P. A.; Breslow, R. Dimethyl sulfoxide to vorinostat: Development of this histone deacetylase inhibitor as an anticancer drug. *Nat. Biotechnol.* **2007**, *25*, 84–90.
(8) Miller, T. A.; Witter, D. J.; Belvedere, S. Histone Deacetylase Inhibitors. *J. Med. Chem.* **2003**, *46*, 5097–5116.
(9) Finnin, M. S.; Donigian, J. R.; Cohen, A.; Richon, V. M.; Rifkind, R. A.; Marks, P. A.; Breslow, R.; Pavletich, N. P. Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* **1999**, *401*, 188–193.
(10) Vannini, A.; Volpari, C.; Filocamo, G.; Casavola, E. C.; Brunetti, M.; Renzoni, D.; Chakravarty, P.; Paolini, C.; De Francesco, R.;

Gallinari, P.; Steinkuhler, C.; Di Marco, S. Crystal structure of a eukaryotic zinc-dependent histone deacetylase, human HDAC8, complexed with a hydroxamic acid inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 15064–15069.
(11) Monneret, C. Histone deacetylase inhibitors. *Eur. J. Med. Chem.* **2005**, *40*, 1–13.
(12) Andrianov, V.; Gailite, V.; Lola, D.; Loza, E.; Semenikhina, V.; Kalvinsh, I.; Finn, P.; Petersen, K. D.; Ritchie, J. W.; Khan, N.; Tumber, A.; Collins, L. S.; Vadlamudi, S. M.; Bjorkling, F.; Sehested, M. Novel amide derivatives as inhibitors of histone deacetylase: design, synthesis and SAR. *Eur. J. Med. Chem.* **2009**, *44*, 1067–1085.
(13) Estiu, G.; Greenberg, E.; Harrison, C. B.; Kwiatkowski, N. P.; Mazitschek, R.; Bradner, J. E.; Wiest, O. Structural Origin of Selectivity in Class II-Selective Histone Deacetylase Inhibitors. *J. Med. Chem.* **2008**, *51*, 2898–2906.
(14) Mai, A.; Esposito, M.; Sbardella, G.; Massa, S. A new facile and expeditious synthesis of N-hydroxy-N′-phenyloctanediamide, a potent inducer of terminal cytodifferentiation. *OPPI Briefs* **2001**, *33*, 391–394.
(15) Codd, R. Coordination chemistry of hydroxamic acid. *Coord. Chem. Rev.* **2008**, *252*, 1387–1408.
(16) Lea, M. A.; Ibeh, C.; Shah, N.; Moyer, M. P. Induction of differentiation of colon cancer cells by combined inhibition of kinases and histone deacetylase. *Anticancer Res.* **2007**, *27*, 741–748.
(17) Mariadason, J. M. HDACs and HDAC inhibitors in colon cancer. *Epigenetics* **2008**, *3*, 28–37.
(18) Khan, N.; Jeffers, M.; Kumar, S.; Hackett, C.; Boldog, F.; Khramtsov, N.; Qian, X.; Mills, E.; Berghs, S. C.; Carey, N.; Finn, P. W.; Collins, L. S.; Tumber, A.; Ritchie, J. W.; Jensen, P. B.; Lichenstein, H. S.; Sehested, M. Determination of the class and isoform selectivity of small-molecule histone deacetylase inhibitors. *Biochem. J.* **2008**, *409*, 581–589.
(19) Bottomley, M. J.; Lo Surdo, P.; Di Giovine, P.; Cirillo, A.; Scarpelli, R.; Ferrigno, F.; Jones, P.; Neddermann, P.; De Francesco, R.; Steinkuhler, C.; Gallinari, P.; Carfi, A. Structural and functional analysis of the human HDAC4 catalytic domain reveals a regulatory structural zinc-binding domain. *J. Biol. Chem.* **2008**, *283*, 26694–26704.
(20) Schuetz, A.; Min, J.; Allali-Hassani, A.; Schapira, M.; Shuen, M.; Loppnau, P.; Mazitschek, R.; Kwiatkowski, N. P.; Lewis, T. A.; Maglathin, R. L.; McLean, T. H.; Bochkarev, A.; Plotnikov, A. N.; Vedadi, M.; Arrowsmith, C. H. Human HDAC7 harbors a class IIa histone deacetylase-specific zinc binding motif and cryptic deacetylase activity. *J. Biol. Chem.* **2008**, *283*, 11355–11363.
(21) Somoza, J. R.; Skene, R. J.; Katz, B. A.; Mol, C.; Ho, J. D.; Jennings, A. J.; Luong, C.; Arvai, A.; Buggy, J. J.; Chi, E.; Tang, J.; Sang, B. C.; Verner, E.; Wynands, R.; Leahy, E. M.; Dougan, D. R.; Snell, G.; Navre, M.; Knuth, M. W.; Swanson, R. V.; McRee, D. E.; Tari, L. W. Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases. *Structure* **2004**, *12*, 1325–1334.
(22) Meyer, B.; Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew. Chem., Int. Ed. Engl.* **2003**, *42*, 864–890.
(23) Dalvit, C.; Pevarello, P.; Tato, M.; Veronesi, M.; Vulpetti, A.; Sundstrom, M. Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. *J. Biomol. NMR* **2000**, *18*, 65–68.
(24) Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51*, 219–242.
(25) Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T. L.; Homans, S. W.; Kessler, H.; Luchinat, C.; Meyer, B.; Oschkinat, H.; Peng, J.; Schwalbe, H.; Siegal, G. Perspectives on NMR in drug discovery: A technique comes of age. *Nat. Rev. Drug Discovery* **2008**, *7*, 738–745.
(26) Ocker, M.; Schneider-Stock, R. Histone deacetylase inhibitors: signalling towards p21cip1/waf1. *Int. J. Biochem. Cell Biol.* **2007**, *39*, 1367–1374.
(27) Kelly, W. K.; Marks, P. A. Drug insight: Histone deacetylase inhibitors-development of the new targeted anticancer agent suberoylanilide hydroxamic acid. *Nat. Clin. Pract. Oncol.* **2005**, *2*, 150–157.
(28) Dowling, D. P.; Gantt, S. L.; Gattis, S. G.; Fierke, C. A.; Christianson, D. W. Structural Studies of Human Histone Deacetylase 8 and Its Site-Specific Variants Complexed with Substrate and Inhibitors. *Biochemistry* **2008**, *47*, 13554–13563.
(29) Kahnberg, P.; Lucke, A. J.; Glenn, M. P.; Boyle, G. M.; Tyndall, J. D. A.; Parsons, P. G.; Fairlie, D. P. SAHA from α-Aminosuberic Acid. *J. Med. Chem.* **2006**, *49*, 7611–7622.
(30) Avalos, M.; Babiano, R.; Barneto, J. L.; Bravo, J. L.; Cintas, P.; Jimenez, J. L.; Palacios, J. C. Can We Predict the Conformational Preference of Amides. *J. Org. Chem.* **2001**, *66*, 7275–7282.
(31) Stewart, W. E.; Siddall, T. H., III. Nuclear Magnetic Resonance Studies of Amides. *Chem. Rev.* **1970**, *70*, 517–551.
(32) Mitro, N.; Godio, C.; De Fabiani, E.; Scotti, E.; Galmozzi, A.; Gilardi, F.; Caruso, D.; Vigil Chacon, A. B.; Crestani, M. Insights in the regulation of cholesterol 7alpha-hydroxylase gene reveal a target for modulating bile acid synthesis. *Hepatology* **2007**, *46*, 885–897.
(33) Witt, O.; Deubzer, H. E.; Milde, T.; Oehme, I. HDAC family: What are the cancer relevant targets? *Cancer Lett.* **2009**, *277*, 8–21.
(34) Le Guevel, R.; Oger, F.; Lecorgne, A.; Dudasova, Z.; Chevance, S.; Bondon, A.; Barath, P.; Simonneaux, G.; Salbert, G. Identification of small molecule regulators of the nuclear receptor HNF4alpha based on naphthofuran scaffolds. *Bioorg. Med. Chem.* **2009**, *17*, 7021–7030.
(35) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. An integrated software for modeling organic and bioorganic molecules using molecular mechanics. *J. Comput. Chem.* **1990**, *11*, 440–467.
(36) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
(37) Polak, E.; Ribière, G. Note sur la convergence de directions conjuguées. *Rev. Fr. Informat. Rech. Operationelle* **1969**, *16*, 35–43.
(38) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–42.
(39) Zhang, J. H.; Chung, T. D.; Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screening* **1999**, *4*, 67–73.

Supporting Information

Title

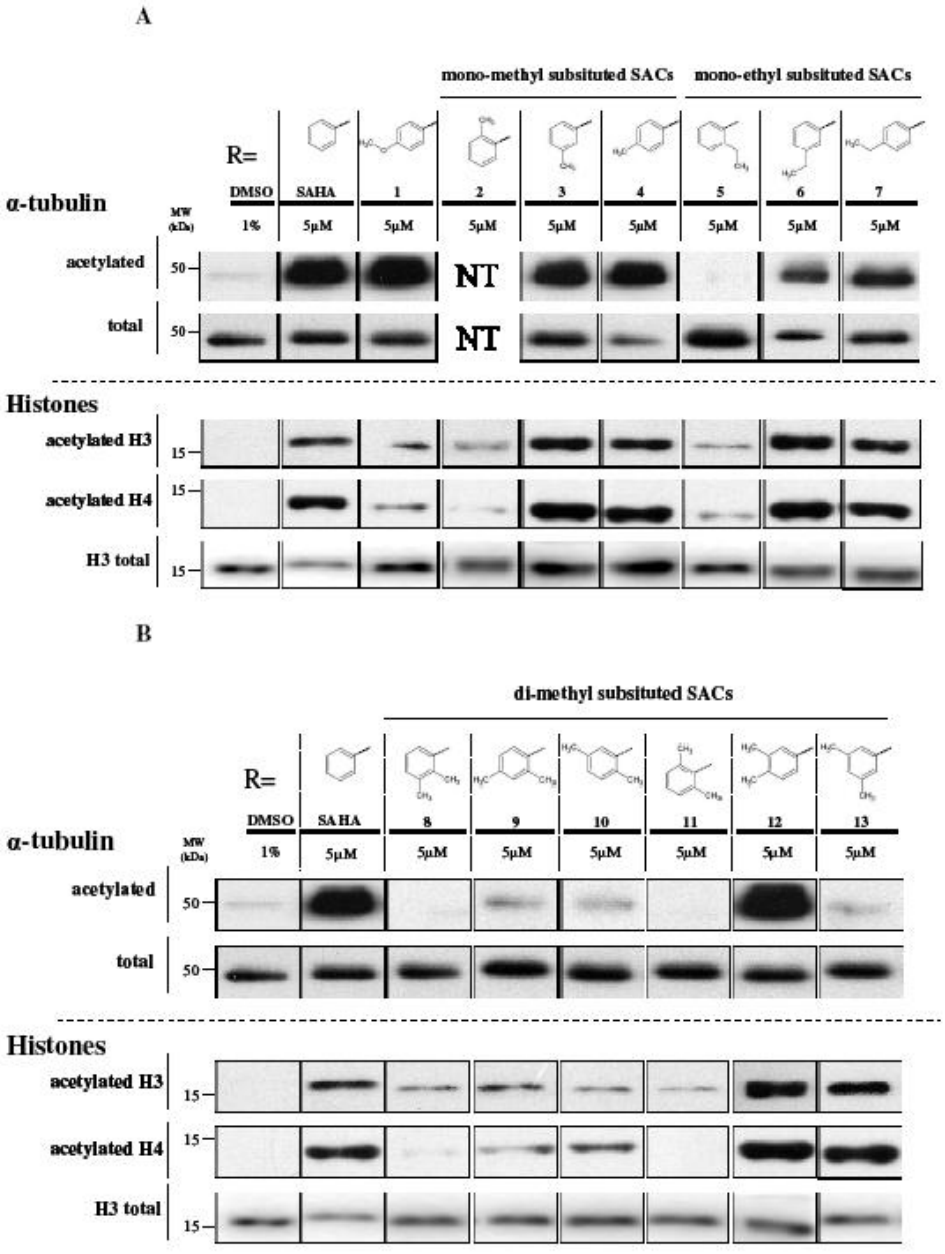
Biological and biophysical properties of the histone deacetylase inhibitor suberoylanilide hydroxamic acid are affected by the presence of short alkyl groups on the phenyl ring.

Authors
Frédérik Oger, Aurélien Lecorgne, Elisa Sala, Vanessa Nardese, Florence Demay, Soizic Chevance, Danielle C. Desravines, Nataliia Aleksandrova, Rémy Le Guével, Simone Lorenzi, Andrea R. Beccari, Peter Barath, Darren J. Hart, Arnaud Bondon, Daniele Carettoni, Gérard Simonneaux and Gilles Salbert

Table of contents

Supporting Information

Supporting Information - S1 : Acetylation level of histones H3/H4 and α-tubulin in response to SACs treatment in Caco-2 cells.



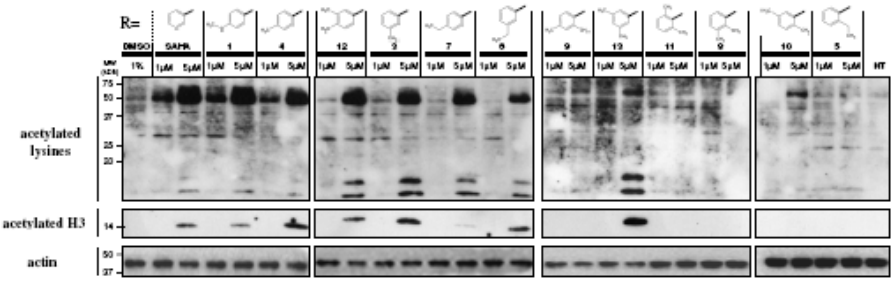
Supporting Information

Supporting Information - S2: Material and methods for histone preparation.

2.10^6 of Caco-2 cells were lysed in 500 µL of Extraction Buffer (EB: 10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol) with 0.2% NP40. After centrifugation at 6,500 g at 4°C, nuclei were washed in 500 µL of EB. Then they were lysed in 300 µL of No Salt Buffer (3 mM EDTA, 0.2 mM EGTA) for 30 min at 4°C. After 5 min centrifugation at 6,500 g, chromatin pellets were resuspended in 300 µL High Salt Buffer (50 mM Tris-Cl pH 8, 2.5 M NaCl, 0.05% NP40) for 30 min at 4°C. Histones were recovered after 10 min centrifugation at 16000 g. Protein concentrations were quantified using DC Protein Assay Kit (Biorad®) according to the manufacturer's instructions.

Supporting Information

Supporting information – S3: Western blot analysis of protein acetylation in HepG2 cells treated with SACs.



Supporting Information – S4: Compound handling and pipetting scheme for compound profiling.

Stock solution of each compound was prepared at 40 mM in dimethyl sulfoxide (DMSO) and stored at -20 °C. The following dilution scheme was adopted for the preparation of dose-response curves for HDAC1, HDAC3 and HDAC6 assays (compound range: 0.316 nM-31.6 µM, see Material and Methods): "mother" plates (96 well-plate, Matrix #4917), containing 200X concentrated dose-response curves, were assembled by performing ten 3.16-fold serial dilutions in 100% DMSO. Compounds were 10-fold further diluted in HDAC Buffer to obtain 20X concentrated "daughter" plates (96 well-plate, Matrix #4917). The following dilution scheme was applied for the preparation of dose-response curves for the HDAC5 and HDAC7 assays (compound range: 6.32 nM-632 µM, see Material and Methods): 20X concentrated dose-response curves were prepared in 96 well-plate (Matrix #4917) by performing ten 3.16-fold serial dilutions in HDAC Buffer.
Substrate mixes (2X) assembled as described in Material and Methods were prepared in HDAC buffer and pre-dispensed in 384 well-plates (Corning #3655). Compounds (20X) were diluted 10-fold in enzyme mix 2.2X (see Material and Methods) in 96-well plate format and the reaction was immediately started by dispensing 15 µL of 2X compound-enzyme mix in quadruplicate into the 384 well-plate containing 15 µL of 2X substrate mix.

Data analysis of compound profiling
To analyze the data of compound profiling (see Material and Methods), the "No Inhibitor" and the "Trichostatin A 100 µM" samples were used as maximum (*MAX*) and minimum (*MIN*) controls, respectively.
In order to assess the assay performance during the compound profiling the *Z'* factor was calculated as follows:

$$Z' = 1 - \frac{3*(\sigma_{MAX} + \sigma_{MIN})}{Abs(\overline{MAX} - \overline{MIN})}$$

where σ_{MAX} and σ_{MIN} are the standard deviation of the *MAX* and *MIN* controls, respectively, and $Abs(\overline{MAX} - \overline{MIN})$ indicates the absolute value of the difference between the mean of the *MAX* and *MIN* controls. Calculation and graphical representation of the *Z'* factor were performed using Spotfire DecisionSite version 9.1 (supporting information S8).
The average IC_{50} values (see Table 2) obtained from a minimum from two independent experiments were calculated using GraphPad Prism 4.0. The standard error mean (*SEM*) was calculated as follows:

$$SEM = \frac{\sigma}{\sqrt{n}}$$

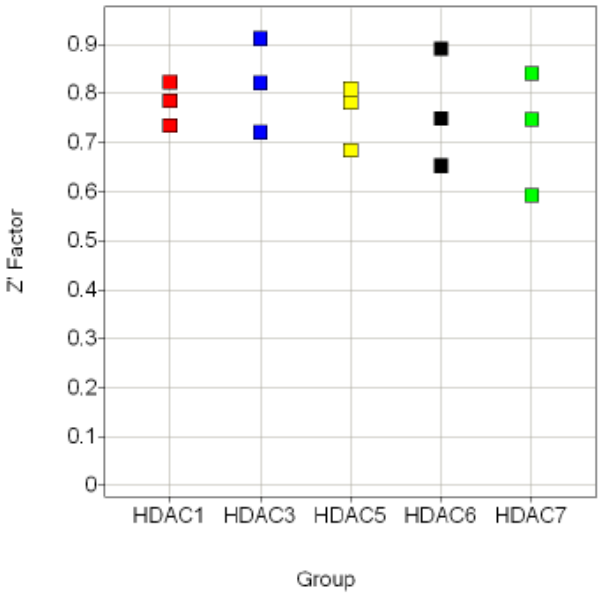
where σ is the standard deviation and *n* is the number of experiments.
Dose-response curve fitting was performed by GraphPad Prism 4.0, using sigmoidal dose-response (variable slope) model with equation:

$$Y = Bottom + (Top - Bottom)/(1 + 10^{[(LogIC_{50} - X)*HillSlope]})$$

where *X* is the logarithm of concentration and *Y* is the percent activity; no bounds were imposed (supporting information S10).

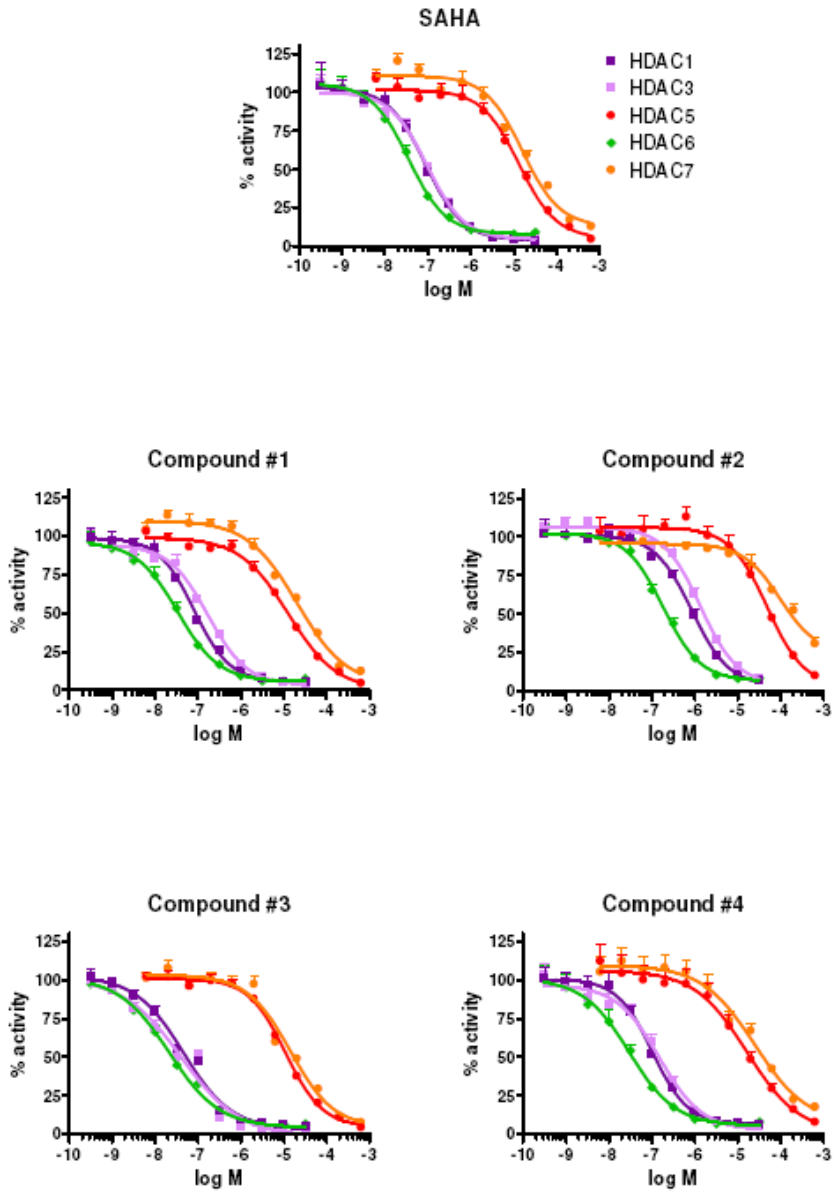
Supporting Information

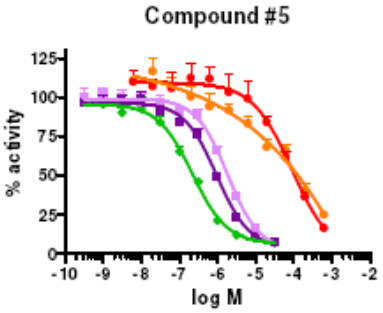
Supporting Information – S5: *Z'* factor evaluation.

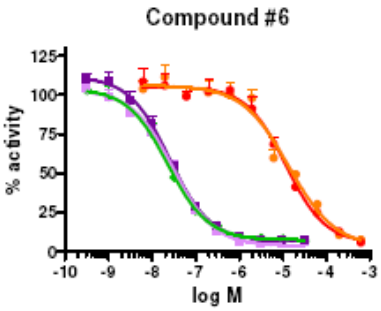


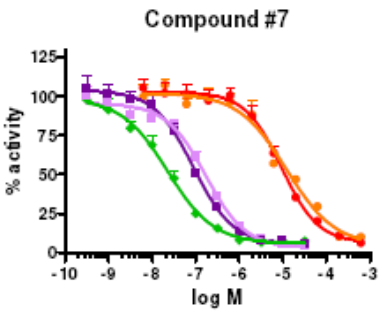
Z' factor was determined based upon all *MAX* and *MIN* test wells that were present on each plate of the compound profiling. The graph represents an overview of the performance of one experiment. Color boxes indicate the *Z'* value for each plate of the enzyme group.
Assays with $1 > Z' \geq 0.5$ are considered suitable for generating high-quality data in screening applications.

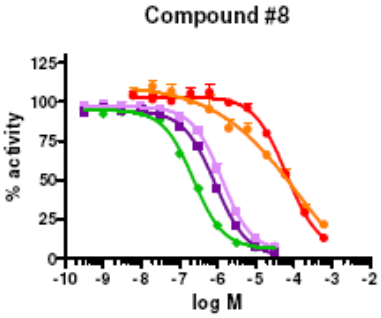
Supporting Information – S6: Dose-reponse curves on HDAC1, 3, 5, 6 and 7 using SACs compounds.

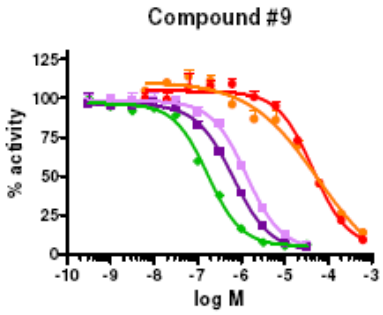


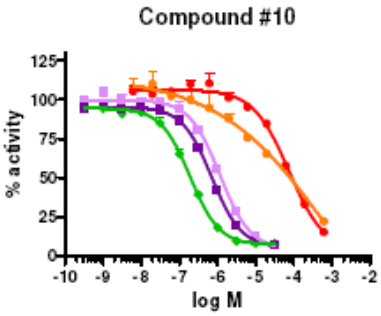

Compound #5
% activity
log M

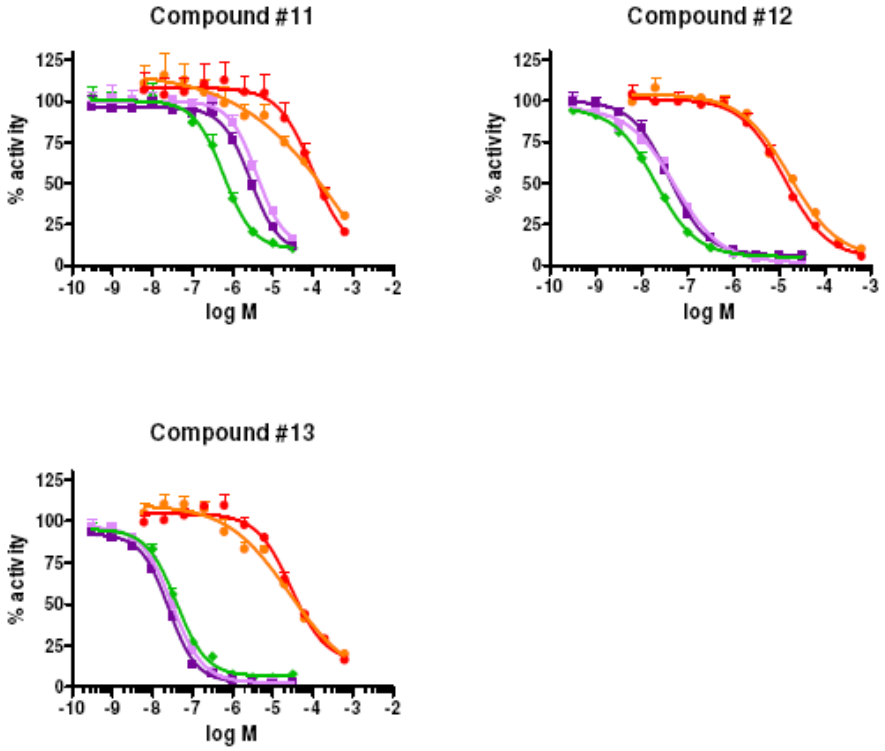

Compound #6
% activity
log M


Compound #7
% activity
log M


Compound #8
% activity
log M

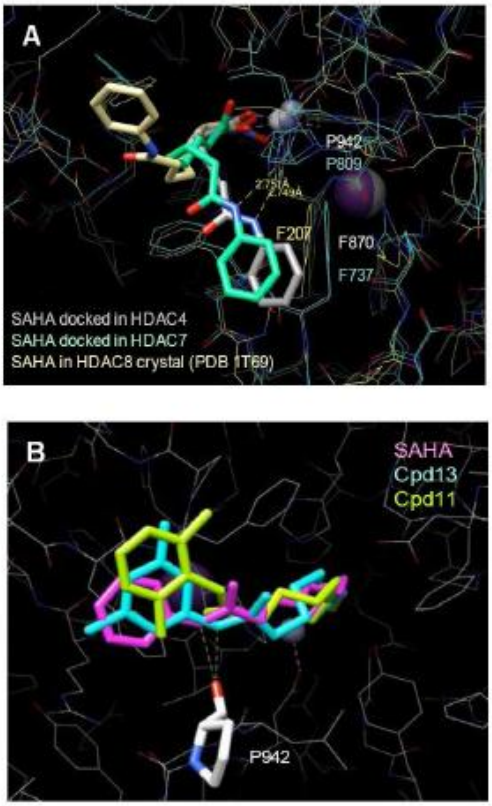

Compound #9
% activity
log M


Compound #10
% activity
log M



Dose-response curves of SAHA and their derivatives were obtained for HDAC1 (■), HDAC3 (■), HDAC5 (●), HDAC6 (◆) and HDAC7 (●). End-point relative fluorescence units (RFU) data were transformed as percentage of activity. Points of the curves represent the mean of the four replicates. Bars represent standard deviations.

Supporting Information – S7: Modeling experiments suggest a conserved protein-ligand interaction mode between the cap domain of SAHA and class 2 HDACs. Method for HDAC4-SACs complex modeling.



(**A**) Superimposed HDAC4 (grey), HDAC7 (blue) and HDAC8 (yellow) structures in complex with SAHA (modelised for HDAC4 and HDAC7, PDB 1T69 for HDAC8). Docking experiments suggest that the NH of the amide group of SAHA establishes a hydrogen bond with the carbonyl group of P942 in HDAC4 and P809 in HDAC7, whereas the phenyl ring of SAHA interacts with F870 in HDAC4 and F737 in HDAC7. Such a positioning of the HDACi in HDAC8 is unfavored by the position of F207 and the absence of proline residue. Note that all class 2 HDACs harbour a proline at the position of HDAC7 P809 whereas class 1 HDACs show a gap in their sequence at the corresponding position [18].
(**B**) Modelisation of SAHA, **11** and **13** interaction with HDAC4. P942 is predicted to establish a hydrogen bond with the NH of the amide group of both SAHA and **13**, whereas the NH of **11** is tilted away due to conformational constraints imposed by the dual *ortho* substitution.

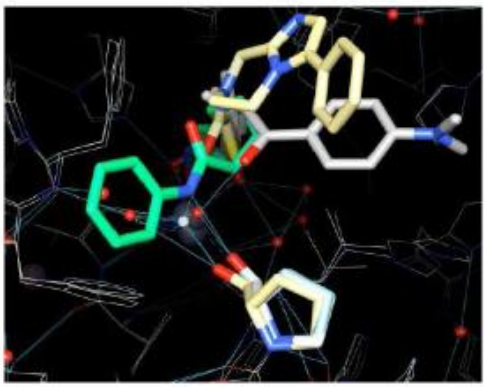
Supporting Information

Protein input file preparation:
A clean HDAC4 (PDB id: 2VQW) input file was generated by using the protein preparation wizard of Maestro software (Maestro 8.5, academic campaign, http://www.schrodinger.com). The bond orders were assigned and Hydrogen atoms were added. The resulting protein was saved to a mol2 file.

Ligands input file preparation:
The ligands input structures were generated and 3D optimized with MarvinSketch Academic Package (MarvinSketch 5.1.4, 2008, ChemAxon http://www.chemaxon.com). The ligands structures were saved to mol2 files.

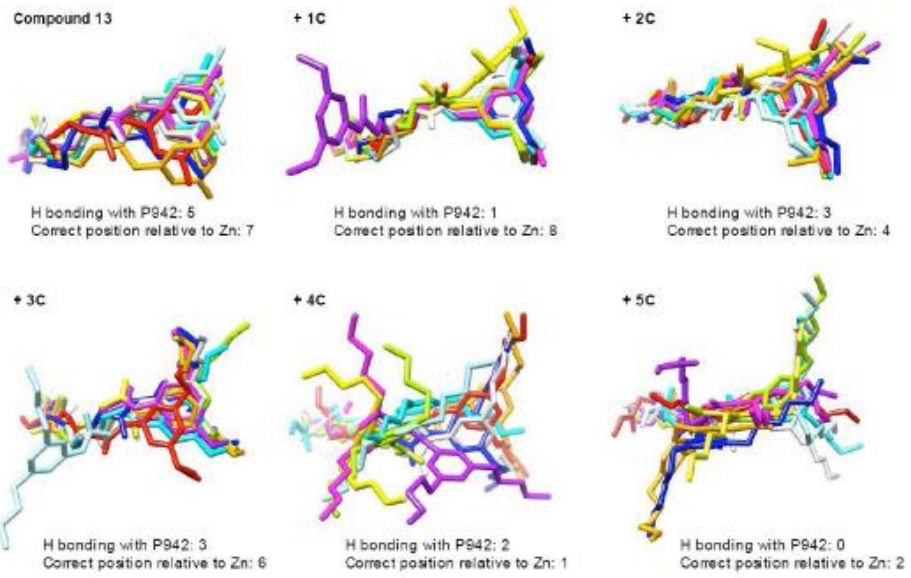
GOLD 3.0 docking procedure:
The binding pocket of HAC4 catalytic domain was defined from the crystallographic coordinates of Zn 1409 of the original pdb. Dockings were performed under "Standard default settings" of GOLD software: Van der Waals: 4, hydrogen bonding: 2.5, search efficiency: 200%, population size of 100, number of islands was 5, number of operations was 100,000, a niche size of 2, and a selection pressure of 1.1.

Supporting Information – S8: Importance of HDAC7 Proline 809 and HDAC4 Proline 942 for ligand interaction.



A water molecule (red dot) is engaged in hydrogen bonding with both Proline 809 of HDAC7 (white) and the carbonyl group of TSA (grey) in PDB 3C10. The same proline residue (light blue) is predicted to engage directly with the NH of the amide group of SAHA (spring green) in docking experiments. In HDAC4 (khaki), Proline 942 has been shown to interact with a water molecule (white dot) engaged in a hydrogen bond with the carbonyl group of a hydroxamate HDACi (khaki – PDB id: 2VQM). Structures were matched with Chimera software.

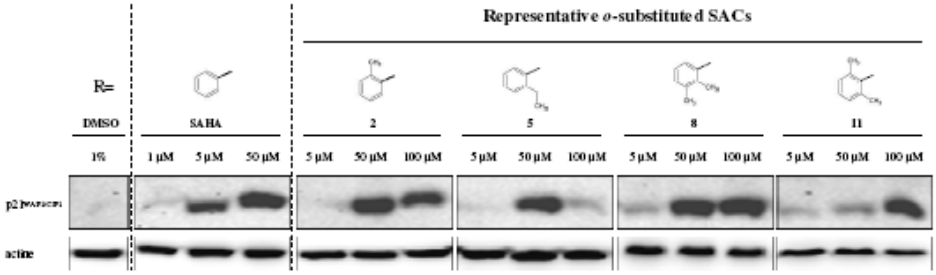
Supporting Information – S9: Lengthening the carbon chain on the phenyl ring of SAHA leads to eratic docking.



Ligands bearing different lengths of alkyl chain in *meta* of the phenyl ring of SAHA were modelised in HDAC4 (PDB id: 2VQW). Images show superimposition of 10 docking solutions (Gold software) for every compound in HDAC4. Note the increased disorder in cap positioning in compounds with long chains. The number of solutions allowing hydrogen bonding between the NH of the amide group and the carbonyl group of P942 is indicated for each compound. The number of solutions showing a correct positioning of the hydroxamic acid function relative to the zinc atom is also indicated for each compound.

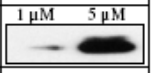
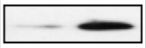
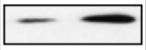
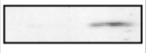
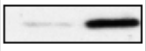
Supporting Information

Supporting Information – S10: $p21^{waf1/cip1}$ induction in response to increasing concentrations (5 µM, 50 µM and 100 µM) of representative *o*-substituted SACs.



Supporting Information

Supporting information – S11: MTT assay and $p21^{waf1/cip1}$ induction in HepG2 cells treated with SACs.

		HepG2 cells	
Compound	R	IC50 (μM)	$p21^{waf1/cip1}$ induction
			1 μM 5 μM
SAHA		9.3	
1		9.9	
2		83.7	NT
3		5.2	
4		7.1	
5		51	
6		4.8	
7		5	
8		44.6	
9		28.7	
10		45.2	
11		>100	
12		6.8	
13		7.4	

Supporting information – S12: Table of RP-HPLC Purity and Microanalytical Data

Analytical HPLC was run on a Waters HPLC system. A first RP-HPLC analysis[1] was performed on a C18 XTerra® (4.6x250 mm, 5µm) column using a water/acetonitrile linear gradient (5-100% B in 40 min, 1 mL/min, 30°C, 225nm). The following buffers were used: (eluent A) water containing 0.08% TFA by volume; (eluent B) acetonitrile containing 1% TFA by volume.
A second RP-HPLC analysis[2] was performed on a C18 XTerra® (4.6x250 mm, 5µm) column using 50 % of eluent A and 50 % of eluent B in isocratic conditions.

compound	RP-HPLC purity	RP-HPLC retention time[1] (minutes)	RP-HPLC retention time[2] (minutes)	formula	calculated			found		
					C	H	N	C	H	N
1	98 %	15.90	3.43	$C_{15}H_{22}N_2O_4$	61.21	7.53	9.52	60.90	7.58	9.45
2	95 %	20.05	4.04	$C_{15}H_{22}N_2O_3$	64.73	7.97	10.06	64.41	8.16	9.61
3	95 %	21.65	4.55	$C_{15}H_{22}N_2O_3$	64.73	7.97	10.06	64.28	7.96	9.51
4	95 %	19.48	3.99	$C_{15}H_{22}N_2O_3$	64.73	7.97	10.06	64.93	8.11	9.49
5	97 %	21.40	3.94	$C_{16}H_{24}N_2O_3$	65.73	8.27	9.58	65.75	8.44	9.30
6	95 %	27.09	10.44	$C_{16}H_{24}N_2O_3$	65.73	8.27	9.58	65.04	8.20	9.34
7	98 %	24.08	4.70	$C_{16}H_{24}N_2O_3$	65.73	8.27	9.58	65.26	8.26	9.13
8	99 %	20.87	4.22	$C_{16}H_{24}N_2O_3$	65.73	8.27	9.58	64.79	8.30	8.89
9	97 %	22.16	3.94	$C_{16}H_{24}N_2O_3$	65.73	8.27	9.58	65.34	8.20	9.26
10	95 %	20.94	4.09	$C_{16}H_{24}N_2O_3$	65.73	8.27	9.58	64.98	8.29	8.91
11	95 %	21.37	4.00	$C_{16}H_{24}N_2O_3$	65.73	8.27	9.58	66.56	8.68	7.87
12	98 %	25.28	4.68	$C_{16}H_{24}N_2O_3$	65.73	8.27	9.58	65.93	8.42	9.24
13	99 %	25.72	4.53	$C_{16}H_{24}N_2O_3$	65.73	8.27	9.58	65.98	8.48	8.75
14	99 %	21.79	4.05	$C_{15}H_{19}F_3N_2O_3$	54.21	5.76	8.43	54.01	5.70	8.35
15	95 %	25.88	5.18	$C_{15}H_{19}F_3N_2O_3$	54.21	5.76	8.43	54.15	5.88	8.46
16	98 %	25.81	6.25	$C_{15}H_{19}F_3N_2O_3$	54.21	5.76	8.43	54.23	5.71	8.36

Supporting Information - S13: Complete material and methods for recombinant HDAC7 preparation and purification.

The DNA encoding the catalytic domain of HDAC7 (residues 483-903 with additional NASNNG C-terminal sequence) was cloned into a pET9-derived vector adding an N-terminal TEV cleavable hexahistidine tag. The HDAC7 was produced in *E. coli* BL21 AI RIL in Terrific Broth medium (50 μg/ml kanamycin, 50 μg/ml chloramphenicol). Expression was induced with 13 mM L-arabinose at OD_{600nm} 0.5-0.8. Cells were harvested by centrifugation and resuspended in lysis buffer (25 mM Hepes pH7.5, 200 mM KCl, 30 % v/v glycerol, 0.5% Igepal (Sigma Aldrich, St Louis, MO), 1 mM DTT and EDTA-free complete protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IN)). Cells were lysed with a microfluidizer (M110-L; Microfluidics Corporation, Newton, MA, USA) and the lysate cleared by centrifugation (48000 g, 40 min, 4ºC).
The supernatant was incubated overnight with Ni^{2+}-NTA agarose resin (Qiagen) at 4ºC, loaded in to a plastic column. Protein was eluted in elution buffer (25 mM Hepes pH7.5, 200 mM KCl, 10% glycerol, 0.1% octyl-β-D-glucopyranoside (Sigma Aldrich), 150 mM imidazole). The eluted fractions were diluted 4 times in buffer (25 mM Hepes pH7.5, 10% glycerol, 0.1% octyl-β-D-glucopyranoside, 1mM DTT) and purified by anion exchange (Q sepharose 16/10 column, GE Healthcare) with elution at 200 mM KCl. The hexahistidine tag was removed by incubation with hexahistidine tagged tobacco etch virus protease overnight at 4ºC. Finally the protein was purified by gel filtration (Superdex 75 column 16/60, GE Healthcare) in buffer (25 mM Hepes pH7.5, 150 mM KCl, 0.1% octyl-β-D-glucopyranoside and 1mM DTT), dialysed overnight at 4ºC in dialysis buffer (50 mM KH_2PO_4/K_2HPO_4, 150 mM KCl, 0.1% octyl-β-D-glucopyranoside, 1 mM DTT), and concentrated to 6 mg/ml with an Amicon Ultra concentrator (Millipore).

CHAPTER 4

The interaction between HDAC7 and the large unstructured SMRT corepressor

RESUME

HDAC7 interagit avec de nombreuses protéines, en particulier avec le corépresseur SMRT. SMRT est une protéine de haut poids moléculaire contenant 2525 acides aminés. Le domaine de répression III de SMRT interagit avec le domaine catalytique des HDACs de classe IIa dont HDAC7 fait partie (mis en évidence par expériences de double hybride en levure). HDAC7 est uniquement trouvée en présence de SMRT dans le noyau, l'inhibition de l'interaction entre HDAC7 et SMRT devrait empêcher l'activité de HDAC7 dans le noyau.

SMRT est une protéine intrinsèquement désordonnée et ne contient ainsi aucun cœur hydrophobe caractéristique des protéines globulaires. Ainsi obtenir des protéines solubles à partir de SMRT de façon rationnelle semble difficile, rendant par conséquent aussi difficile de cibler l'interaction HDAC7-SMRT. Cependant des essais d'inhibition d'interaction protéine-protéine ont été un succès pour quelques interfaces, que ce soit des interfaces impliquant deux protéines globulaires ou impliquant une protéine globulaire et une protéine intrinsèquement désordonnée. L'interface entre une protéine intrinsèquement désordonnée et une protéine globulaire constitue une cible intéressante en drug discovery grâce à 4 principales caractéristiques : leur interface est légèrement plus petite que les autres interfaces, leur interface est concave (plate pour les autres), un segment unique est impliqué dans l'interaction (plusieurs pour les autres), et les interfaces impliquent es interactions hydrophobes (interactions à caractère polaire pour les autres). Ainsi 4 familles d'inhibiteurs ont été créés pour inhiber avec succès des interactions impliquant une protéine intrinsèquement désordonnée et une protéine globulaire. Ce fut le cas notamment pour l'interaction entre le suppresseur de tumeur p53 qui est intrinsèquement désordonné et l'ubiquitine ligase MDM2 globulaire. La famille d'inhibiteur nutlin se sont révélés particulièrement efficaces pour cibler cette interface avec des inhibitions allant jusqu'à 90 nM, provoquant l'activation de p53 et induisant l'arrêt du cycle cellulaire et apoptose. De telles stratégies se sont donc révélées faisable. Nous avons donc décidé de poursuivre cette stratégie. La technologie ESPRIT a été appliquée à SMRT afin d'obtenir des constructions solubles. 51 clones ont ainsi pu être sélectionnés, exprimant des protéines allant de 15 kDa a 34 kDa, proteines hautement solubles. 6 de ces constructions ont été choisies pour leur bonne expression et leurs bons rendements après purification. L'effet de ces fragments sur l'activité de déacétylase de HDAC7 a été testé : aucun effet

n’a été détecté sur aucun des substrats commerciaux. Des expériences de RMN ont été effectuées : les spectres HSQC des fragments présentent les caractéristiques des protéines désordonnées, à savoir des corrélations dans la dimension des protons regroupées entre 8-8,5 ppm. L’interaction de ces fragments avec HDAC7 a aussi été suivie par RMN et a identifié un fragment interagissant avec HDAC7 : 69D20. Des expériences d’anisotropie de fluorescence ont confirmé ce résultat. Des constantes d’affinité ont pu être déterminées par anisotropie de fluorescence et résonnance plasmonique de surface de l’ordre du micromolaire. La résonance plasmonique de surface a parmi de révéler la dynamique de l’interaction. En effet le sensorgramme obtenu a révélé une interaction en deux phases : une association rapide suivie d’une association plus lente de HDAC7 avec 69D20 et une dissociation rapide suivie d’une dissociation plus lente. Cela semble indiquer un lent changement de conformation par l’interaction. Des expériences complémentaires de RMN sont actuellement en cours afin d’identifier les résidus impliqués dans l’interaction entre SMRT et HDAC7. Une fois l’interaction caractérisée, des analyses d’inhibition pourront être effectuées.

1 Introduction

1.1 The cryptic HDAC7 activity

As was detailed earlier, HDAC7 does not deacetylate canonical acetylated histone peptides (Lahm et al. 2007; Bottomley et al. 2008; Schuetz et al. 2008). This absence of deacetylase activity has been explained by a catalytically important tyrosine in the active site (Y312 in HDAH and Y306 in HDAC8), which stabilises the oxyanion intermediate produced during catalysis (see chapter 1, paragraph 4.2, figure 4), that is replaced by a histidine in HDAC7 (H843). Interestingly, when this histidine is mutated into a tyrosine, the deacetylase activity is rescued (Lahm et al. 2007; Bottomley et al. 2008; Schuetz et al. 2008). This is puzzling since it demonstrates that the active site is otherwise totally able to catalyse the deacetylation reaction in that it binds the substrate and possesses all the residues necessary for the chemistry. A truly dead enzyme would be expected to have lost most of these aspects very quickly once the natural selective pressure for catalytic activity had been lost. So a key question is *does HDAC7 possess an enzymatic activity in vivo? And if so, what is the substrate?*

To try and understand the biological relevance and elucidate the possible enzymatic activity of HDAC7, a parallel is often drawn with HDAC3 and its partners, especially SMRT (also called NCoR2), and the highly homologous NCoR. The HDAC3 catalytic domain directly interacts with SMRT/NCoR and this interaction induces the HDAC3 enzymatic activity (Wen et al. 2000; Guenther et al. 2001; Zhang et al. 2002b). The minimal interacting fragment was identified as a 68 amino acid region folding into 4 α-helical deacetylase-activating domain (DAD) (Codina et al. 2005). For HDAC7, we hypothesised that the interaction with SMRT may occur via some type of overlooked DAD and promote HDAC7 deacetylase activity in a similar manner, perhaps through i) modifying the structure of the active site to compensate for inactivating Y308H substitution, or ii) by changing the nature of the protein surface around the active site and thereby enhancing the affinity with low affinity substrates (maybe histones). In support of this is the observation that HDAC7 immunoprecipitated from the cytoplasm is inactive, but from the nucleus where SMRT coimmunoprecipates it is active on acetylated histone substrates (Kao et al. 2000; Fischle et al. 2001). Similarly, HDAC4 (from the same type IIa class as HDAC7) also showed no deacetylase activity in the absence of SMRT (Fischle et al. 2002).

SMRT is clearly an essential interacting partner for HDAC7 *in vivo*, and the interaction has been mapped by yeast-2-hybrid deletion analysis to the repression domain III of SMRT (Kao et al. 2000; Fischle et al. 2001). The complex of HDAC7 and SMRT also contains HDAC3, but the role of HDAC3/SMRT for HDAC7 is unclear: *does active HDAC3 account for the deacetylation activity of the complex? Or does HDAC3 somehow switch on the deacetylase activity of HDAC7?*

To address those questions, it was necessary to explore in more detail the HDAC7/SMRT complex. Importantly, understanding the HDAC7/SMRT interaction and binding interface could lead to a new strategy for targeting HDAC7 function; blocking HDAC7 recruitment into the corepressor complex would be expected to disrupt completely its function. This could provide an effective alternative to (relatively unspecific) HDAC7 active site inhibitors, and therefore addresses the HDAC7 inhibitor theme of this thesis.

1.2 The study of Instrinsically Disordered Proteins

There are several methods that characterise disorder in proteins. Whilst globular proteins precipitate upon boiling, IDPs resist to boiling at high temperatures. This was observed for several IDPs such as microtubule associated protein 2 (MAP2) (Hernandez et al. 1986). Heating was even suggested as a general method for purifying recombinant IDPs (Kalthoff 2003). IDPs are resistant to heat thanks to their unusual amino acid content, i.e. the highly charged and low hydrophobic content of amino acids means they do not expose buried hydrophobic residues at high temperatures (Uversky et al. 2000; Dunker et al. 2001).

IDPs are resistant to chemical denaturation due to their amino acid composition. For globular proteins , negatively charged side chains are protonated at low pH leading to charge unbalance disrupting salt bridges and causing aggregation (Dill and Shortle 1991). In contrast lowering the pH of IDPs has a different effect. IDPs are commonly highly charged at neutral pH, and decreasing the pH lowers their net charge, therefore this reduces electrostatic repulsion and sometimes triggers a change of conformation into a more compact state (Uversky 2002).

Another common feature of IDPs is their unusual migration on SDS-PAGE. This is due to a decreased binding of SDS and is explained in more detail in section 2.2.3.

The behaviour of IDPs can be followed by size exclusion chromatography. It is often observed that IDPs elute at higher volumes than that expected for a globular protein of a given molecular weight. This indicates that that the proteins are in an extended structural state.

Notwithstanding that the different predictors available are based on different principles, they all rely on the biased sequence features of IDPs. There are about 20 predictors and meta-predictors. The list of predictors is available at the DisProt website (http://www.disprot.org) and a summary is provided here.

Propensity-based predictors. These rely on statistical analyses of the physical/chemical features of amino acids. They include the low-complexity of sequence, biased amino acid composition and the lack of secondary structure elements. For example, FoldIndex simply relies on amino acid propensity (Prilusky et al. 2005); GlobPlot is based upon amino acid propensity combined with a preference for ordered secondary structure (Linding et al. 2003b); Prelink relies also on amino acid propensity but adds hydrophobic cluster analysis (Coeytaux and Poupon 2005); and NORSp is based upon secondary structure propensity and incorporates the Position-Specific Iterative Basic Local Alignment Search Tool (PSI-BLAST) (Liu and Rost 2003).

Machine-learning algorithms (MLA) are more sophisticated in the prediction of disorder. They are trained to distinguish between sequences that encode for disorder versus order. MLAs are implemented by two different methods: neural networks (NN), an adaptive system that changes based upon information gained during learning phases, and support-vector machines (SVM) that involve supervised learning methods. Amongst the MLA predictors based on NN there are PONDR (Li et al. 1999), RONN (Yang et al. 2005), and DisEMBL (Linding et al. 2003a). The DISOPRED2 predictor combines SVM and NN learning routines with PSI-BLAST analysis (Ward et al. 2004).

Predictions can exploit the idea that IDPs are disordered because they cannot make sufficient inter-residue contacts. These predictor tool variants include FoldUnfold which treats data based upon single amino acid propensity (Garbuzynskiy et al. 2004), IUPred relying on estimated pairwise interaction energies (Dosztanyi et al. 2005).

Finally as any one predictor has both strengths and weaknesses, combinations of predictors can be used in the form of meta-servers which carry out parallel predictions and produce a consensus output. MeDor and metaPrDOS represent such meta-servers. They submit the query sequence to 12 servers simultaneously (MeDor and metaPrDOS)

and adds to this predictions of secondary structure (metaPrDOS) and hydrophobic clusters (Lieutaud et al. 2008, Ishida and Kinoshita 2008).

1.3 Drug design and disordered proteins

Most drugs available on the market target the active site of enzymes (Sebolt-Leopold and English 2006) or ligand binding sites of receptors (Lagerstrom and Schioth 2008). Blocking protein-protein interactions with drug molecules is difficult and has been largely unsuccessful (Drews 2000). Only 8 small molecule drugs have been successful in targeting such interactions (Arkin and Wells 2004; Arkin 2005) (see as well chapter 1, section 7.3.3).

Despite difficulties in targeting protein interactions between globular proteins, their inhibition remains very attractive since protein interaction surfaces are much less conserved than catalytic sites and ligand binding pockets offering the potential for highly specific inhibition. Disruption of interactions between IDPs and globular proteins seems particularly feasible because of their different mode of interaction. IDPs often bind their partner through short linear recognition elements which bind grooves or clefts of the partner. In this way, the interface resembles the type of interaction observed in receptor-ligand or enzyme-substrate binding, which can often be targeted by small molecules. Even more feasible is that the interaction can be inhibited by peptides, offering a means of validating the strategy via targeting peptides in cell based assays.

In more detail, interfaces of IDPs are slightly smaller than those of ordered complexes, their average surface is of 1141± 110 $Å^2$ (Vacic et al. 2007) whilst for globular proteins the surface is larger with an average area of 1600 ± 400 $Å^2$ (Lo Conte et al. 1999). Moreover IDPs use a larger portion of their surface for interaction: 50% in contrast to most globular proteins that use only 5-15% (Gunasekaran et al. 2004; Meszaros et al. 2007). Another characteristic feature lies in the number of structural regions involved in the interaction. In the case of the folding of globular proteins distinct segments are brought together to participate in the creation of binding sites, thus the primary sequence regions contacting a partner protein are often discontinuous (Meszaros et al. 2007). In contrast, IDP interfaces often only involve a single sequentially continuous segment. A further difference is that while globular proteins contribute most of their hydrophobic resides to the protein core, IDPs expose their few hydrophobic residues to the surface to allow interaction with binding partners. 40-90% of IDP

hydrophobic residues are exposed to the surface versus only 5-15% for ordered proteins (Meszaros et al. 2007). Therefore IDP interfaces make more hydrophobic-hydrophobic interaction contacts (33% for IDPs and 22% for ordered proteins) whereas ordered proteins make more polar-polar interactions (Gunasekaran et al. 2004; Meszaros et al. 2007).

Of the 8 designed small molecule drugs targeting protein-protein interfaces, 4 of them block protein-protein interactions between disordered and ordered partners demonstrating the feasibility of this strategy (Uversky et al. 2008). One of the successes involved the targeting of p53/MDM2 interaction. The activity of the tumour suppressor p53 is controlled by the E3 ubiquitin ligase MDM2 through a negative feedback mechanism. In the crystal structure of the complex between MDM2 and a p53 N-terminal fragment, MDM2 exhibits a well-defined hydrophobic pocket filled by 3 side-chains from p53. Potent and selective inhibitors like nutlins have been designed (Klein and Vassilev 2004; Vassilev et al. 2004). Nutlins bind MDM2 in the p53 binding pocket and activate the p53 pathway in cancer cells, leading to cell cycle arrest and apoptosis *in vitro*. The number of such similar targets is estimated to be in the range of thousands of protein-protein interfaces involving disordered proteins (Cheng et al. 2006b).

Chapter 3 described work performed for validation of small molecule HDAC7-specific inhibitors. In this chapter an alternative strategy was explored, where the interaction between HDAC7 and SMRT is characterised, with the ultimate goal of inhibiting this binding interface.

2 Results and discussion

2.1 Bioinformatic characterisation of the SMRT corepressor

SMRT is a gigantic intrinsically disordered protein (IDP) of 2525 amino acids. Thus the interaction between HDAC7 and SMRT is characterised by the presence of a disordered protein binding to an ordered partner and presumably involves a folding-upon-binding event. To understand better the structural composition and possible domain content of SMRT, the sequence was analysed using series of tools (Figs. 1 and 2). The SMRT protein and especially SMRT repression domain is predicted to be mainly unstructured with few secondary structure elements being present. The major challenge was therefore to determine whether it was possible to express soluble proteolysis-resistant forms of SMRT.

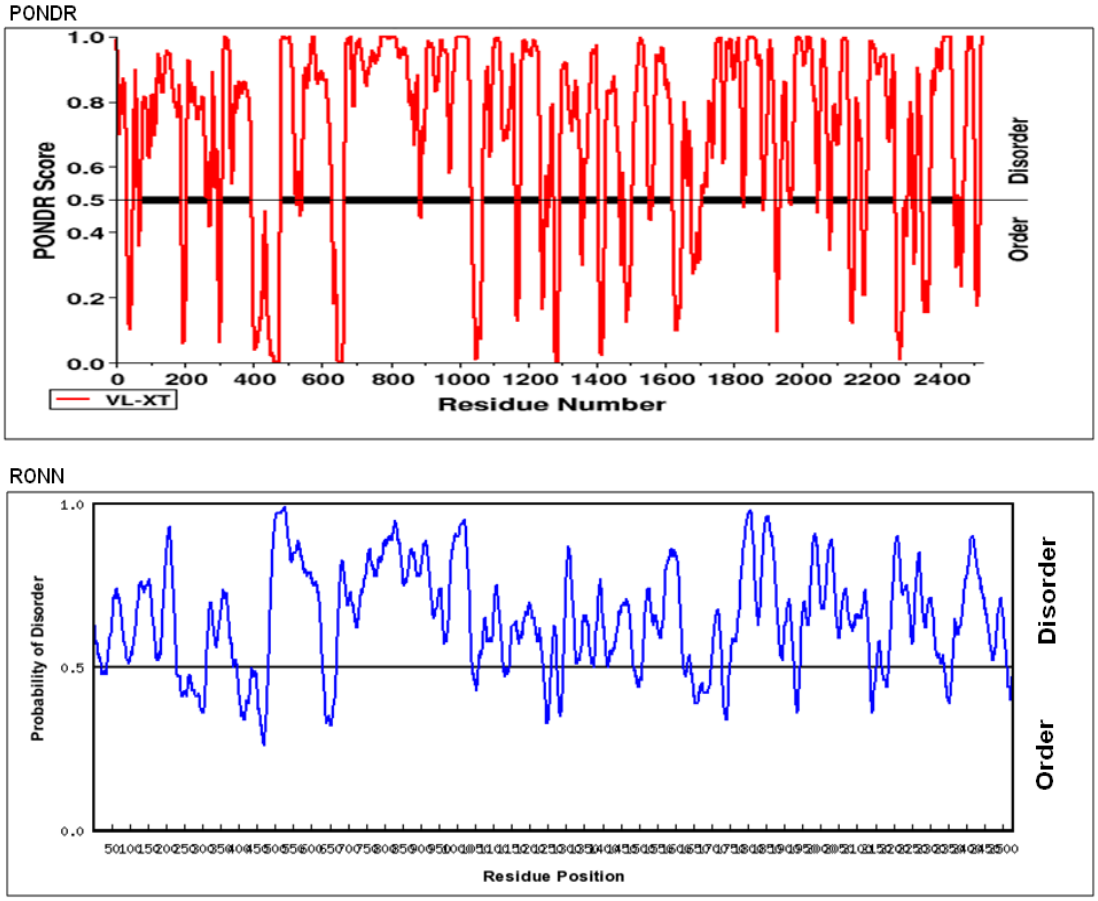


Figure 1. Disorder predictions of full-length SMRT. Two individual predictors were used to predict the disorder propensity of SMRT. Both algorithms predicted that SMRT is highly unstructured (probabilities >0.5).

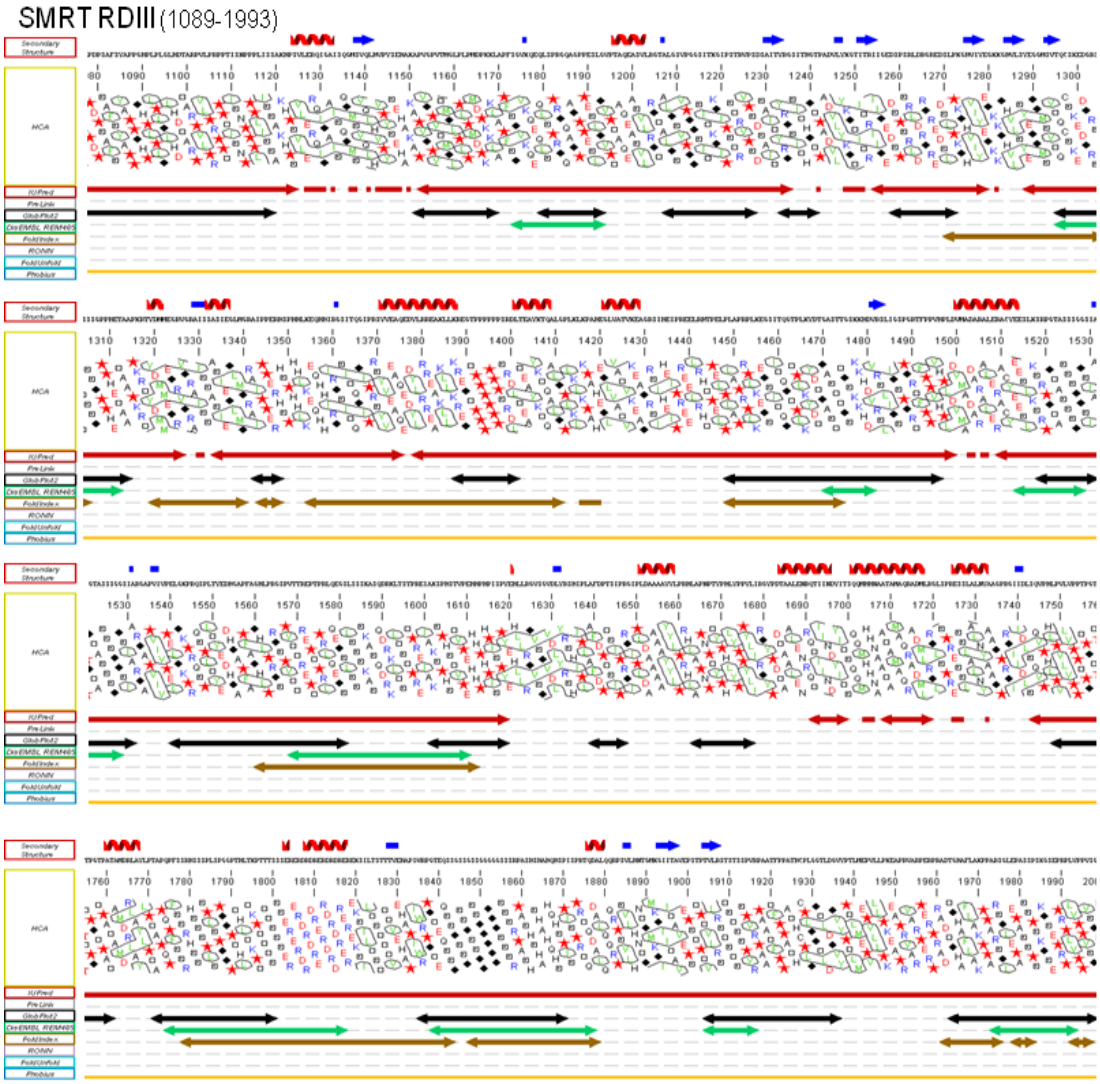


Figure 2. MeDOR analysis of the SMRT repression domain III. The Meta-server predictor was used to analyse SMRT RDIII. It predicted that SMRT RDIII is mainly unstructured (narrows) and contains few predicted secondary structures. Besides, there are no hydrophobic clusters but rather small patches distributed along the sequence.

2.2 ESPRIT technology applied to SMRT to obtain soluble proteins

2.2.1 SMRT construct design

The SMRT region interacting with class IIa HDACs, localised in RDIII of SMRT (Fischle et al. 2001; Guenther et al. 2001; Fischle et al. 2002) was considered for investigation. In particular the minimal interacting region from deletion analyses (Kao et al. 2000; Fischle et al. 2001) starting from E1289 to L1793 was selected and extended by 200 amino acids at both N- and C-termini (Fig. 3). DNA encoding this enlarged region (G1089 to V1993) was produced synthetically with *E. coli* optimised codons, and was

cloned into a pET9 derived vector allowing truncation with exonuclease III according to the ESPRIT protocols.

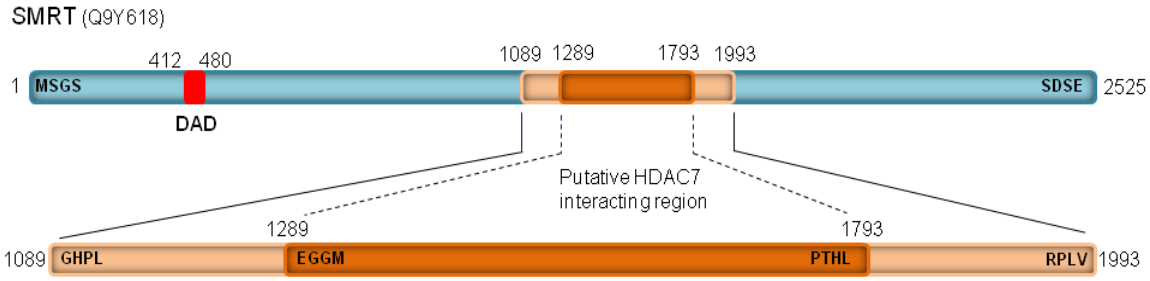


Figure 3. SMRT construct. At the top in blue is the full length SMRT protein. The DAD domain (interacting with HDAC3) is highlighted in red. In dark orange is the region shown to interact with class IIa HDACs, in light orange are the 200 amino acid extensions added at the N-terminus and C-terminus of the interacting region.

2.2.2 Bidirectional truncation

A random bidirectional truncation strategy was chosen to generate an internal fragment library containing constructs ranging from 300 bp to 3000 bp in size (Fig. 4). The overall efficiency of the library was estimated at 95%, defined as clones containing inserts in the desired size range. The library was then used to transform *E. coli* BL21 AI RIL. In total, 27642 colonies were isolated resulted in a calculated sampling of the total theoretical diversity of 3%.

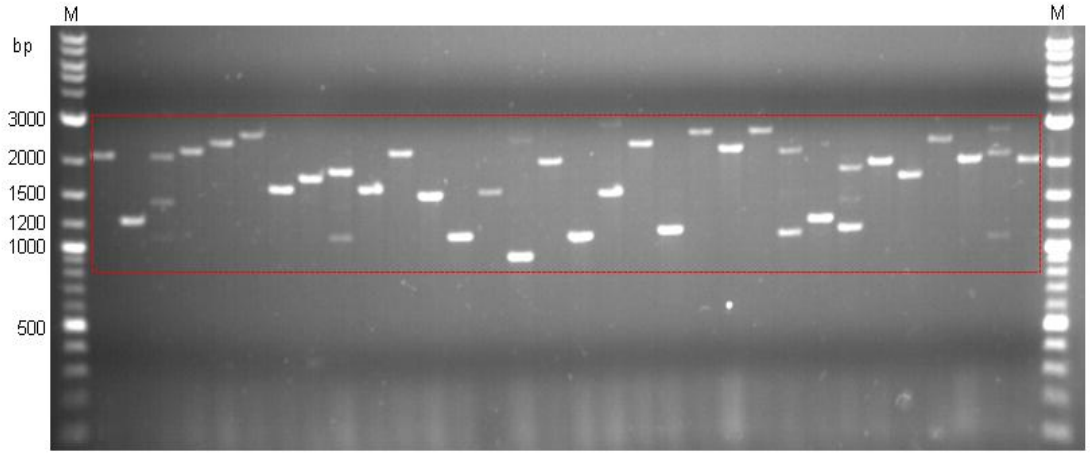


Figure 4. Evaluation of truncation library quality. Colony PCR screen using insert flanking primers showing 32 out of 96 tested clones. Inserts fall between 1000 bp and 3000 bp in size.

2.2.3 Selection of soluble constructs.

The isolated clones were arrayed on nitrocellulose membranes and tested for both N-terminal hexahistidine tag and C-terminal biotinylation (as described in chapter 2 for

HDAC7). The first 96 clones showing strongest biotinylation signals were selected for further analysis by purification screen. This second level of screen was performed from 4 ml expression cultures by affinity purification with Ni^{2+}-NTA resin using a liquid handling robot. The purified fractions were analysed on SDS-PAGE gels (Fig. 5). Many constructs were clearly visible on SDS-PAGE after purification demonstrating that they were highly soluble, in particular 51 clones showed clean purification profiles. They clustered between 15 and 34 kDa (including 5 kDa of peptide tags). After sequencing, it was apparent that the selected hits spanned most of the length of parental construct with the exception of a 57 amino acid region (from residue 1619 to 1676, numbering as full-length SMRT) where no construct was selected (Fig 6 and Fig 7A). Some of the constructs were tested for large scale-expression and purification based upon analysis of 4 ml purifications on SDS-PAGE. Six constructs were chosen: 69D20 (V1276/Y1405), 39L23 (I1255/S1458), 69M23 (H1122/I1254), 27M12 (P1784/V1993), 68F1 (T1798/A1966), 42D23 (R1882/V1993) (Fig 6B). They were expressed in 1 l culture and purified by affinity chromatography with Ni-NTA resin (Fig 8). The selected constructs showed high levels of expression and solubility. It can be estimated according to SDS-PAGE (from total and soluble fractions) that the proteins are highly soluble. They were purified at high yields, estimated in the range of 10mg to 20 mg of protein per litre of culture.

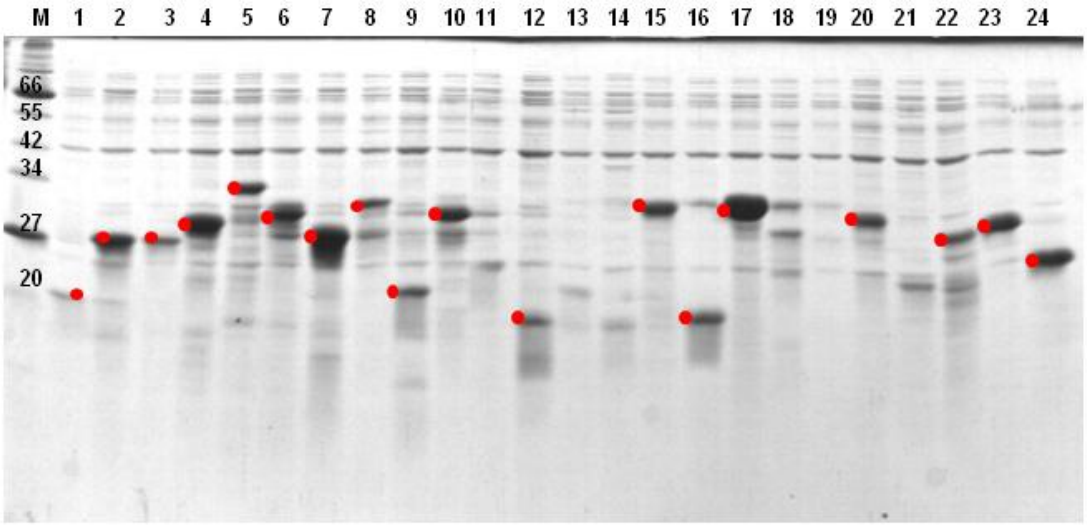


Figure 5. Identification of soluble and purifiable SMRT fragments. 24 of 96 selected constructs are presented, analysed by Coomassie-stained SDS-PAGE. Well purifying solbule constructs are highlighted with a red dot. (M: molecular weight marker).

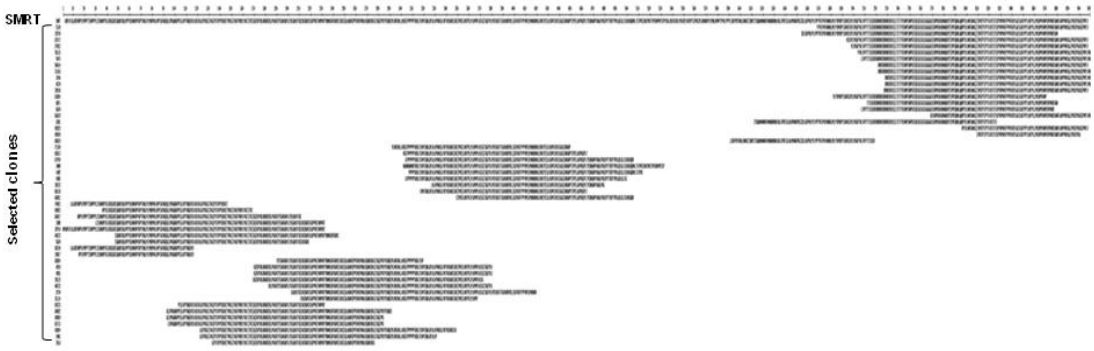


Figure 6. Sequence alignment of the SMRT selected hits. The 51 selected clones span the length of the SMRT fragment, with only a small region not being covered.

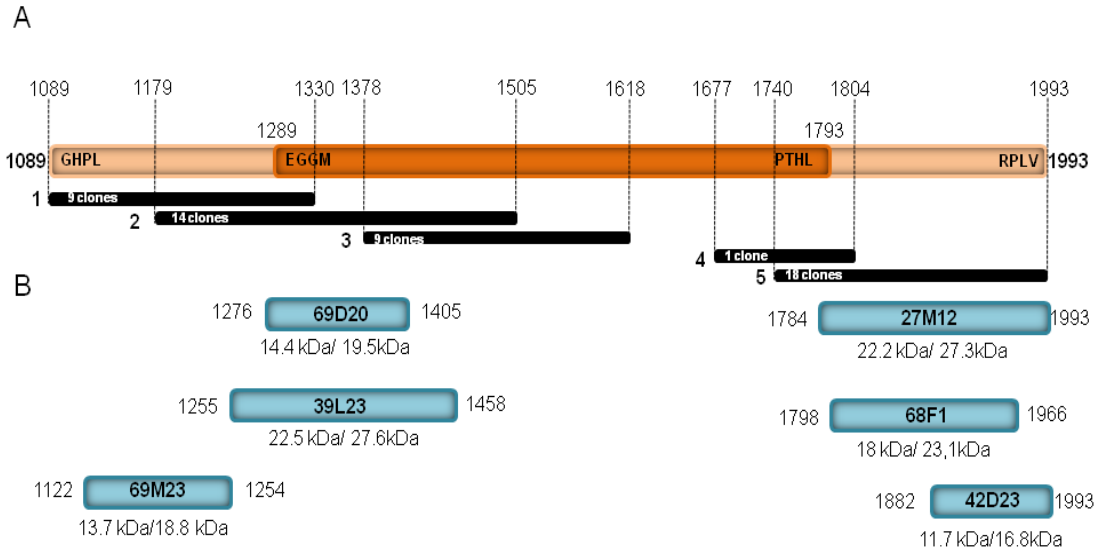


Figure 7. Schematics of the some selected hits. A) The SMRT construct used for truncation is in dark/light orange. The clusters (regions) of clone coverage are shown in black with the number of clones found in these regions indicated. B) 6 constructs were selected based upon their high yields by SDS-PAGE analysis after purification distributed within regions 1, 2 and 5: 69D20, 39L23, 69M23, 27M12, 68F1, 42D23 (molecular weights are indicated under each construct both without and with peptide tags).

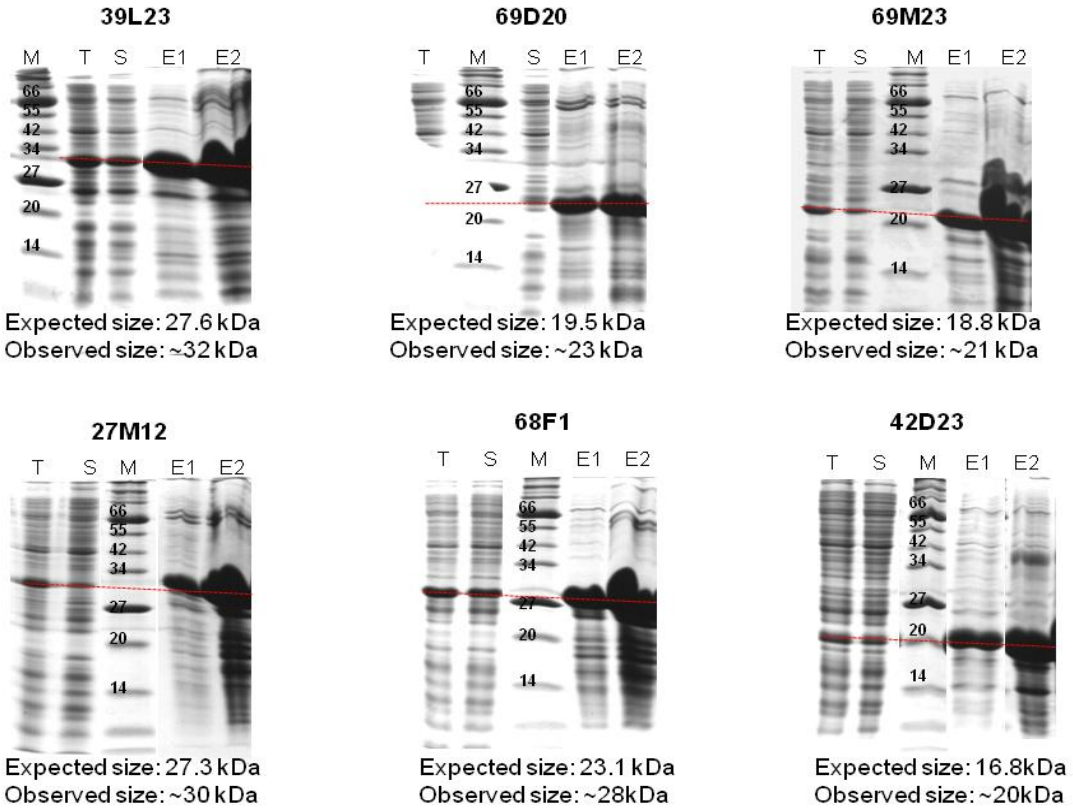


Figure 8. Expression and purification of 6 SMRT constructs. All the constructs are well-expressed: T, total fraction; S, soluble fraction. The constructs were well purified as apparent from elution fractions (E1 and E2). The proteins (highlighted with a red dashed line) did not migrate at appropriate size, in general they migrate between 3 and 5 kDa higher than expected (see bottom of SDS-PAGE).

All proteins exhibited aberrant mobility on SDS-PAGE i.e. they migrated at higher apparent molecular weights. Generally globular proteins bind almost the same amount of SDS per amino acid, thus the proteins are negatively charged and their charge/mass ratio is uniform. Therefore their migration in the electric field is homogenous, the migration of the proteins depends only on their molecular weight. In contrast, IDPs contain a highly biased charged amino acid content that prevents SDS binding through charge repulsion i.e. IDPs bind less SDS than globular proteins. As a consequence they migrate slower than expected and appear larger. This phenomenon has been observed for many proteins and it appears that molecular weights of IDPs are generally overestimated by SDS-PAGE by a factor of 1.2 to 1.8 (Tompa 2002), as observed for these SMRT constructs. In addition it was observed that the selected proteins were resistant to boiling for >10 minutes without triggering their precipitation (Results: Eva Geenan, data not shown).

2.3 Interaction studies of SMRT fragments with HDAC7

2.3.1 NMR

Six SMRT fragments were analysed by N-H correlation NMR, this being the method of choice to obtain insights into both structure and dynamics of proteins such as IDPs. Heteronuclear single quantum coherence (HSQC) spectra were recorded on the ^{15}N-labeled SMRT fragments. Correlations of the backbone amide nitrogens with the directly attached protons were measured. In a ^{1}H-^{15}N HSQC spectrum, the amide bond of each amino acid (except prolines) and the amides of side-chains give a signal. The spectra for SMRT fragments are shown in Fig. 9. The SMRT fragments showed peaks in the range of 7.8 ppm to 8.5 ppm in the ^{1}H dimension and between 108 ppm to 129 ppm in the ^{15}N correlation dimension. These frequencies are characteristic of IDPs with the peaks in the proton dimension being poorly dispersed (most peaks tend to be gathered between 8.0 ppm and 8.5 ppm). In contrast, in the nitrogen dimension, correlation peaks are well-spread from 105 ppm to 130 ppm for backbone and side-chain amide groups. This typical dispersion in the two dimensions is a reliable indicator of disordered proteins and demonstrates the accuracy of HSQC for identifying IDPs. Thus it was confirmed that the 6 SMRT fragments were intrinsically disordered in isolation.

HSQC experiments are also ideal for detecting interactions with proteins. Chemical shifts (i.e. the position of the signals in the NMR spectrum) are highly sensitive to the local electronic environment of the nuclei. Upon protein-protein interaction, changes in the local environment of residues directly involved in the interaction occur leading either to shifts or disappearance of these residues from the spectrum. Therefore to analyse whether any of the SMRT fragments interacted with HDAC7, HSQC experiments were performed in the presence of unlabeled HDAC7 (Fig. 10). Only one of the SMRT fragments, 69D20, showed a clear interaction with HDAC7. Despite this strong interaction, the longer fragment 39L23 that contained the 69D20 sequence (Fig. 7), no binding was observed in this experiment. We reason that this might be explained by a protein conformation reducing interaction between both proteins, possibly supported by fluorescence anisotropy data (see below).

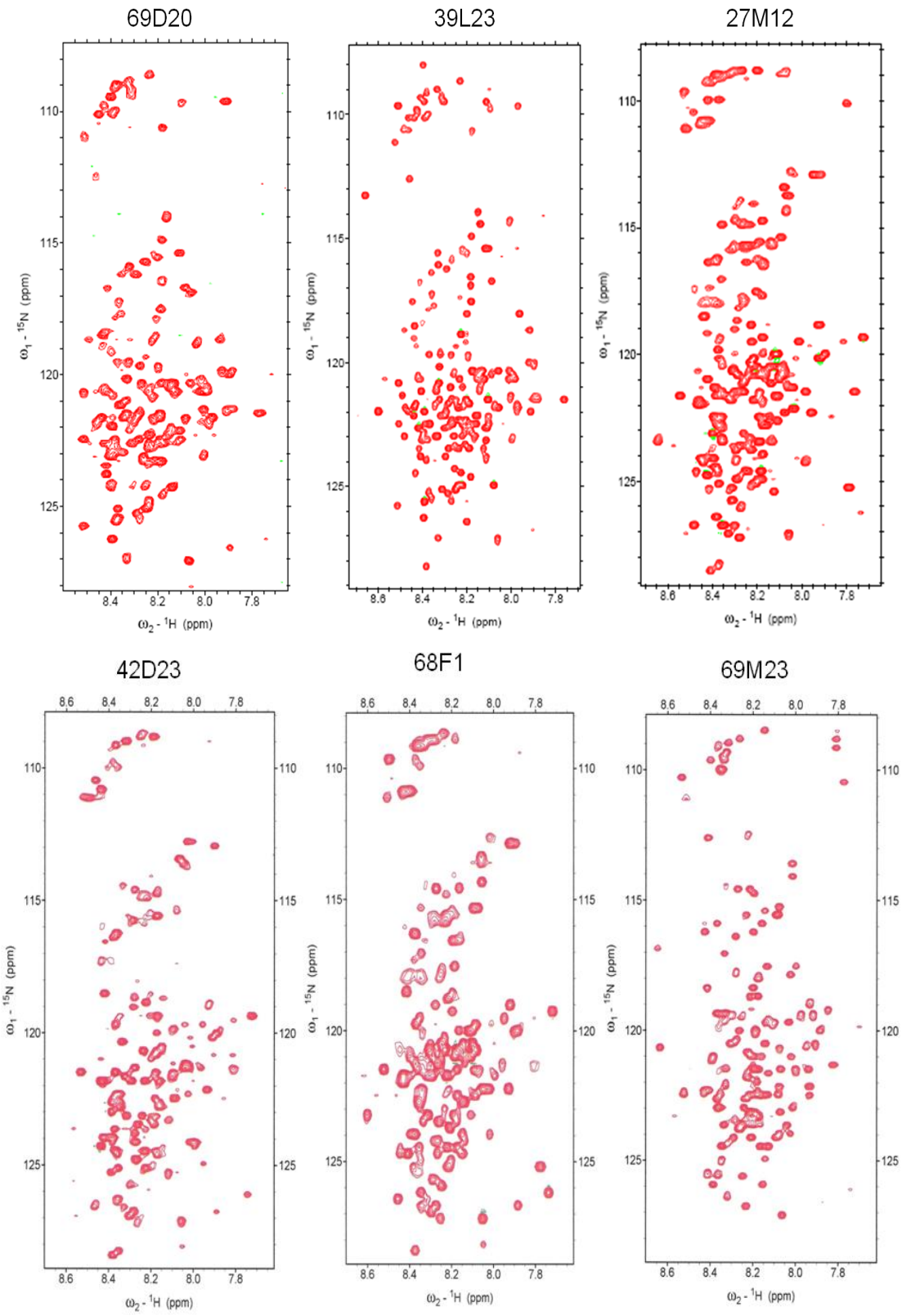


Figure 9. HSQC spectra of SMRT fragments. For all purified SMRT fragments, the resonances are observed between 7.8 ppm and 8.5 ppm in the ^{1}H dimension. Those for the amide backbone and sidechain amides (^{15}N correlations) spanned from 108 ppm to 129 ppm.

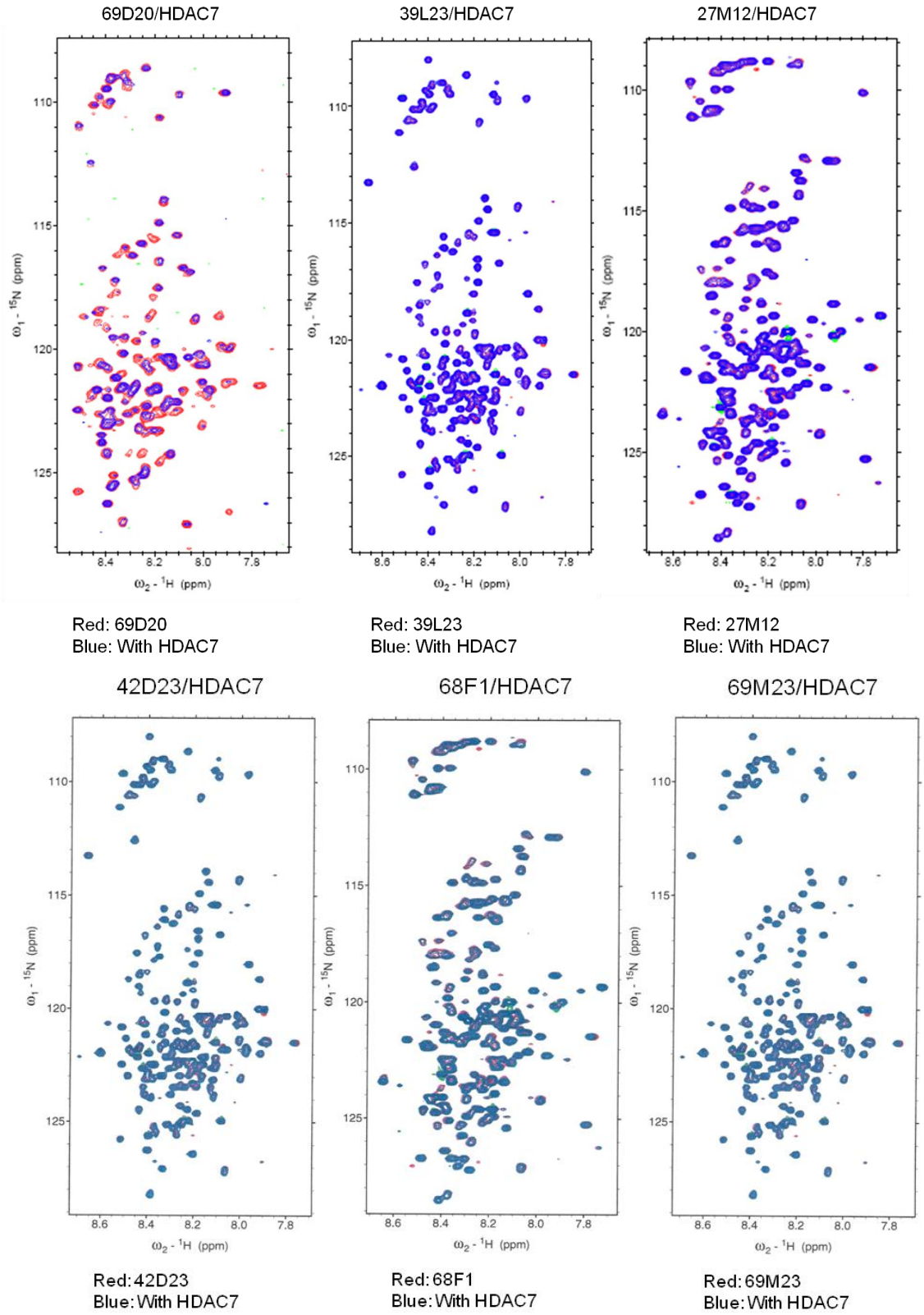


Figure 10. Binding studies by HSQC NMR. Spectra of SMRT fragments alone and SMRT fragments incubated with HDAC7 in the NMR tube are superimposed with the blue spectrum for SMRT fragments and red spectrum for SMRT fragment/HDAC7. For 69D20, peaks disappeared upon binding, whilst for the other constructs no such effect was observed resulting in a perfect overlay of both spectra.

2.3.2 Binding of SMRT fragments to HDAC7 by fluorescence anisotropy

The hydrodynamic volume of a probe attached to a protein will increase when the protein binds to other partners forming a complex. The hydrodynamic properties of labelled proteins can be determined by measuring the fluorescence polarisation (FP) or fluorescence anisotropy (FA) of the emission using steady-state measurements. This method allows measurement of the affinity of a protein-protein interaction in solution and is well-suited for interactions of a small partner with a larger one.

Dr Natasha Aleksandrova (post doc in the laboratory) should be strongly acknowledged for her leading role in the execution of the following experiment. We aimed to determine whether the interactions observed initially by NMR could be reproduced using a second solution-phase method. The SMRT fragments were first fluorescently-labelled. Highly purified HDAC7 was added to fluorescent fragments with a 10-fold molar excess, and interactions were measured at 25°C by measuring relative changes of anisotropy (Fig. 11). Two of the fragments showed a significant change in anisotropy: 2-fold for 69D20 and 1.75-fold for 39L23. The other fragments showed less change. These data are consistent with both 69D20 and 39L23 forming a complex with HDAC7. If the averaged signal from the majority of the constructs is taken as the baseline background signal, or as an indication of non-specific binding (since no indication of specific binding was apparent in NMR), and subtracted from the signals of 69D20 and 39L23, then values of 1.0 and 0.75 above background are obtained for both of these constructs. Thus although this assay is a bit noisy, it seems that both fragments containing the 69D20 region interact with HDAC7, with the shorter construct binding more strongly. These data are therefore consistent with the NMR results above.

Dissociation constants for the 69D20 interaction with HDAC7 were calculated by titration of HDAC7 (8 different concentrations from 0.7 µM to 12 µM). Two similar dissociation constant values were obtained by non-linear regression analysis (curve-fitting) and Scatchard plot: 2.4×10^{-6} M and 6.7×10^{-6} M respectively. These constants indicate that the interaction between HDAC7 and SMRT 69D20 is a medium affinity interaction.

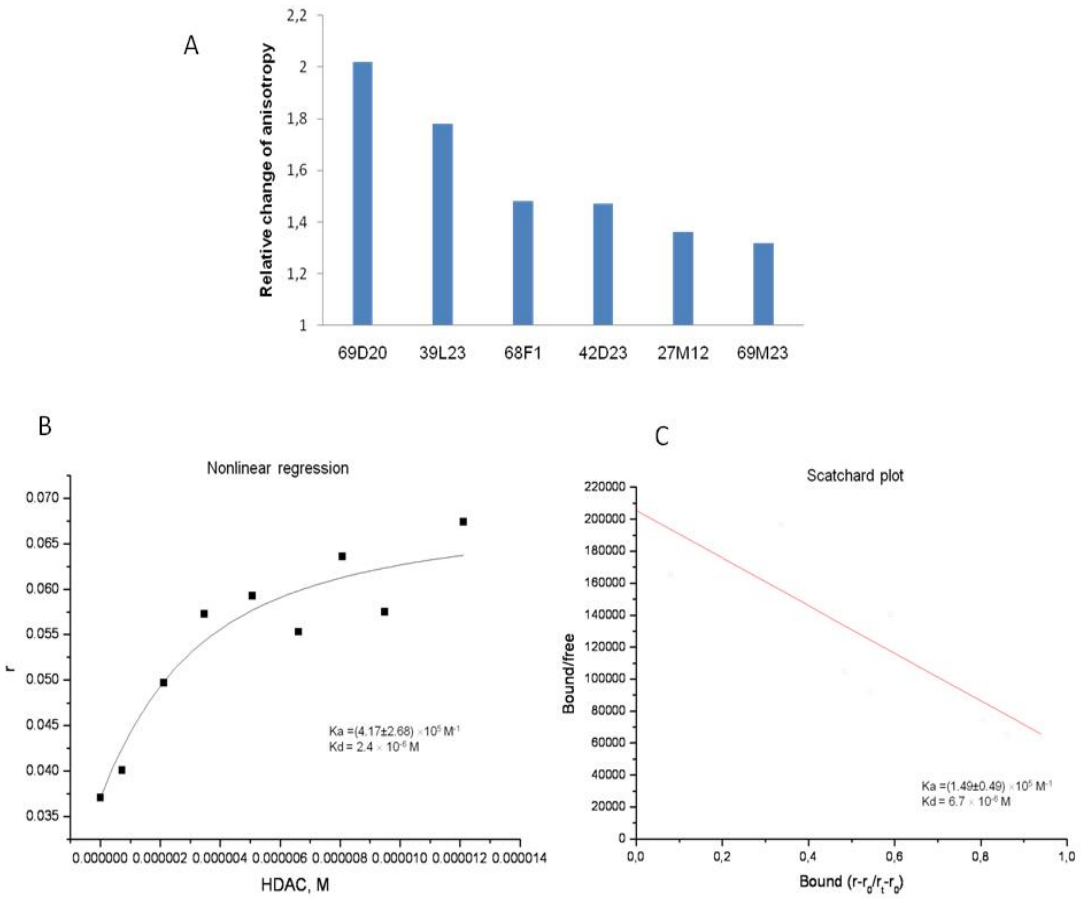


Figure 11. Binding of SMRT fragments to HDAC7 measured by fluorescence anisotropy. A. Fluorescently labeled SMRT fragments were incubated with HDAC7 at 25°C and relative changes of anisotropy were measured permitting comparison. B & C. binding experiments upon titration of HDAC7 to calculate affinities (K_d). Nonlinear regression of the data (B) predicted a Kd of 2.4 x 10^{-6} M whilst Scatchard plot (C) predicted Kd as 6.7 x 10^{-6} M.

2.3.3 Characterisation of 69D20 binding to HDAC7 by surface plasmon resonance

In order to elucidate mode of interaction of SMRT 69D20 and confirm derived binding constants obtained by fluorescence anisotropy, surface plasmon resonance (SPR) analyses were performed using a Biacore X instrument. SPR allows measurement of the affinity through individual determination of association and dissociation rates. It therefore provides mechanistic insights that are not possible using an equilibrium binding method. Experiments and analyses of data were performed with the significant input and expert guidance of Dr Natasha Aleksandrova.

Measurements were made on immobilised 69D20 (CM5 chip), injecting HDAC7 through the flowcell at different concentrations. Data were obtained and analysed using BiaEvaluation software for curve alignment and fitting. This revealed that the 69D20

association with HDAC7 occurred in two phases comprising a fast initial phase, followed by a slower phase (Fig. 12). Likewise the dissociation phase was also biphasic involving a first rapid dissociation, followed by a long, slow dissociation phase of remaining material. This two-phase binding mode suggests that there is probably a slow conformational change occurring after an initial rapid binding event; this has been reported previously by Sevcik and colleagues, also using SPR (Sevcik et al. 2007) when studying the binding of the IDP tau. The slower phase would seem to represent accumulation of stable complex; this fraction of tightly bound material then dissociates slowly and is visible once the weakly bound material has been washed away. When the half-life of the complex is calculated from the latter dissociation phase, a $T_{1/2}$ value of 216 s is obtained ($T_{1/2}$ (s) is calculated as ln2/k_d). Therefore the second slow phase might indicate that a conformational rearrangement occurs during binding and is what would be expected given the "fly-casting" mechanism of IDP binding (Levy et al. 2007) where and initial reversible low affinity interaction sometimes leads to a rearrangement and tighter binding mode in a fraction of the molecules.

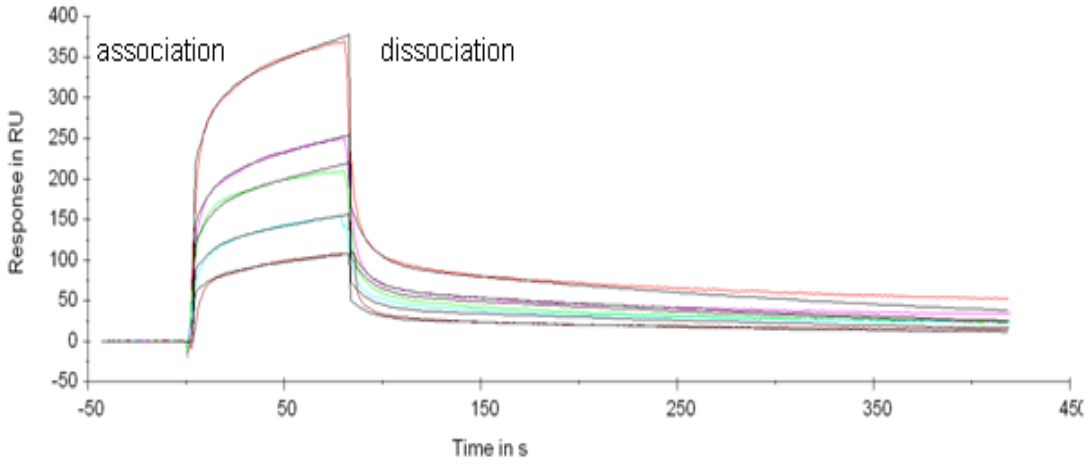


Figure 12. Two-phase binding of 69D20 to HDAC7. A two-phase binding interaction model was used for K_D determination of 69D20/HDAC7 interaction. The association and dissociation rate constants (k_a and k_d) respectively were calculated: first phase k_a = 1.87 x 10^4 Ms^{-1} and k_d = 9.03 x 10-2 s^{-1}; second phase k_a = 1.38 x 10-2 Ms^{-1} and k_d = 3.2 x 10-3 s^{-1}; K_D = 1.1 µM. Coloured curves represent a concentration series (250 nM,375 nM, 562 nM, 625 nM and 1 µM) and black curves are fits to the experimental data used for extraction of kinetic data. (RU: Response Unit)

This phenomenon is observed where disordered regions of proteins are involved in the interaction mechanism and we speculate that this SMRT fragment makes primary contacts via short motifs. The fast-association phase might suggest certain region of SMRT might be responsible for initiating the interaction with HDAC7. Such a mechanism has been described for endocytosis for example where the assembly of large

membrane-bound complexes of highly repetitive and disordered membrane-associated proteins occurs, such as AP180, epsin1, and auxilin (Kalthoff et al. 2002; Scheele et al. 2003) and where fast assembly of the complexes is required for the proper execution of endocytosis. The large capture radius of specific, exposed recognition elements of these disordered proteins provides the principal "hook" for this mechanism, also called "protein fishing" (Evans and Owen 2002; Dafforn and Smith 2004). The protein fishing mechanism is based upon the idea of enhancement of association rates by structural disorder. This molecular mechanism has several interesting features, such as the involvement of relatively non-specific initial association events (Pontius and Berg 1991; Pontius 1993), the exposure of primary contact sites for the initial contact (Csizmok et al. 2005) and the increased capture radius of the disordered fragment providing an efficient spatial search for a partner (i.e. protein fishing). The effective induced folding initiated by these initial contacts is known as "fly-casting" (Shoemaker et al. 2000; Levy et al. 2007). This model of fly-casting suggests that an IDP can have an enhanced capture radius for a specific binding site to which it binds weakly at a relatively large distance. Therefore the induced folding is nucleated by the contact between the two molecules.

We speculate that this phenomenon might occur in the case of SMRT bound to HDAC7, but further dynamic NMR studies are needed to confirm this hypothesis. One can as well note that the conformational change induced in SMRT fragment might be accompanied by one in HDAC7 itself. Indeed it was observed that the region containing a second surface exposed structural Zn^{2+} interacting with SMRT is flexible (Fig. 13). This conformational change in HDAC7 might occur as well when SMRT binds, but some further biophysical methods (like SAXS) will be required to address this hypothesis.

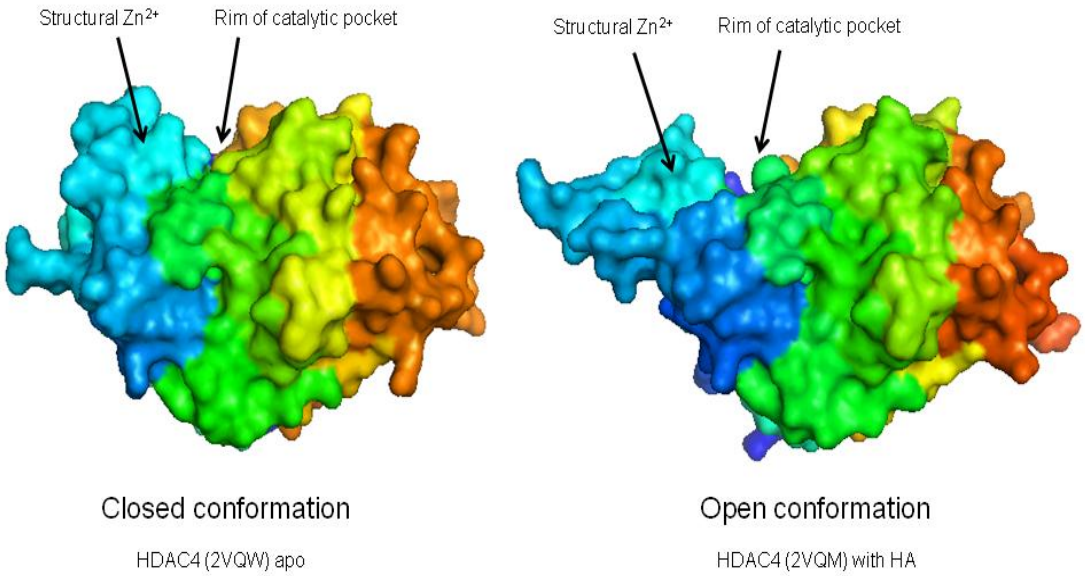


Figure 13. HDAC4 surfaces in apo and liganded forms. In the left panel is represented the surface of HDAC4 (from N terminus in blue until C terminus in orange) with the structural Zn^{2+} region in blue and the rim of the catalytic pocket shown with an arrow. In the left panel the entrance of the catalytic pocket is wider, due to a movement of the structural Zn^{2+} region.

2.3.4 Bioinformatic analysis of SMRT fragment 69D20

Since NMR, fluorescence anisotropy and SPR all suggested SMRT 69D20 bound HDAC7 with micromolar affinity, secondary prediction analyses were performed on this fragment to understand its possible composition (Fig. 12) When 69D20 secondary structure predictions are compared with the DAD structure (which binds HDAC3 and activates it (Codina et al. 2005)), there is no obvious similarity. Nevertheless, these are only predictions and could be inaccurate. Alternatively, this region of SMRT that binds HDAC7 may be structured differently to the previously described DAD. In progress currently at the time of thesis submission is the spectral assignment of the residues of this region by ^{15}N-^{13}C-labled proteins thus it will soon be possible to map the region of SMRT fragment that interacts with HDAC7.

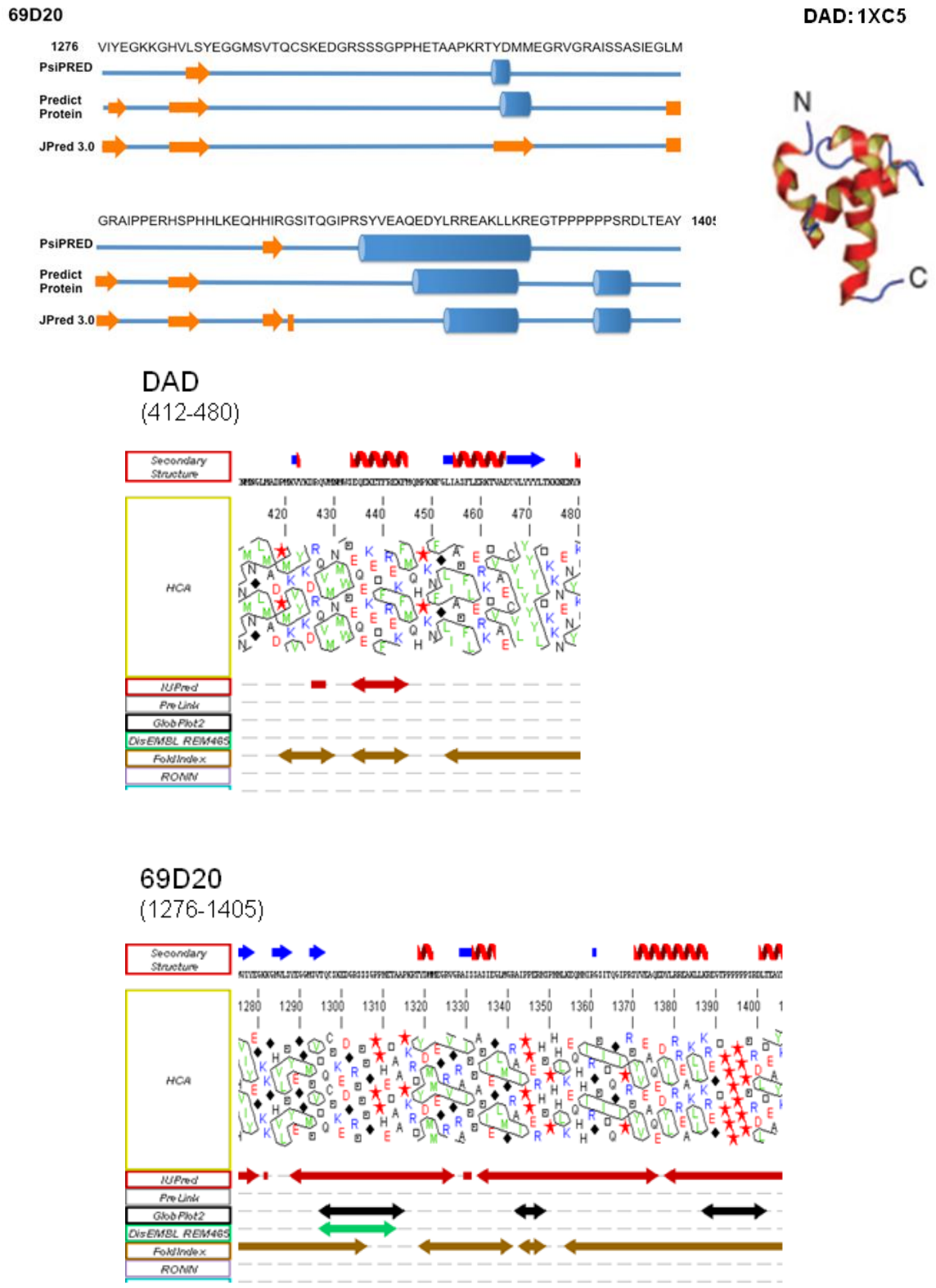


Figure 12. Secondary structure prediction analyses of 69D20. Three different secondary structure predictors were used (PsiPRED, PredictProtein, and JPred 3.0) to predict secondary structures in 69D20. They predicted the presence of β-sheets and α-helices in the same regions. The predicted secondary structures are different to those predicted for the SMRT DAD that interacts with HDAC3 i.e. β-sheets are not present in the DAD structure. A similar trend was observed using the metaserver predictor, MeDOR.

2.4 Influence of SMRT fragments on HDAC7 activity

It was described in chapter 3 how our recombinant "improved" HDAC7 was devoid of detectable enzymatic activity on acetyllysine-containing peptides, despite deacetylating the much more labile trifluoroacetyllysine substrate. It has been shown that HDAC3/SMRT complex purified from cell extracts is highly enzymatically active (Wen et al. 2000; Guenther et al. 2001; Zhang et al. 2002b) although the cryptic activity of the type IIa family member, HDAC4, was not rescued by the presence of SMRT in cellular extracts (Guenther et al. 2001). Nevertheless this has not been investigated for HDAC7. Additionally, it is often not clear whether detected activities in extracts from eukaryotic cells are due to contaminating enzymes that copurify, so reconstitution of this complex from purified recombinant proteins provides a way to test this.

We therefore analysed whether any of the soluble SMRT fragments studied by NMR (including 69D20) reconstituted HDAC7 enzymatic activity in *in vitro* deacetylase assays. HDAC7 activity was measured using both trifluoroacetyllysine and acetyllysine substrates. Addition of SMRT fragments did not lead to significant enhancement of HDAC7 deacetylase activity, neither on the canonical acetyllysine substrate nor on the synthetic trifluoroacetyllysine substrate (Figure 13). Thus, although the fragment 69D20 was shown to interact with HDAC7, this did not rescue the cryptic activity suggesting that while SMRT can bind and activate the type I HDAC3, a similar mechanism does not occur for the type IIa HDAC7.

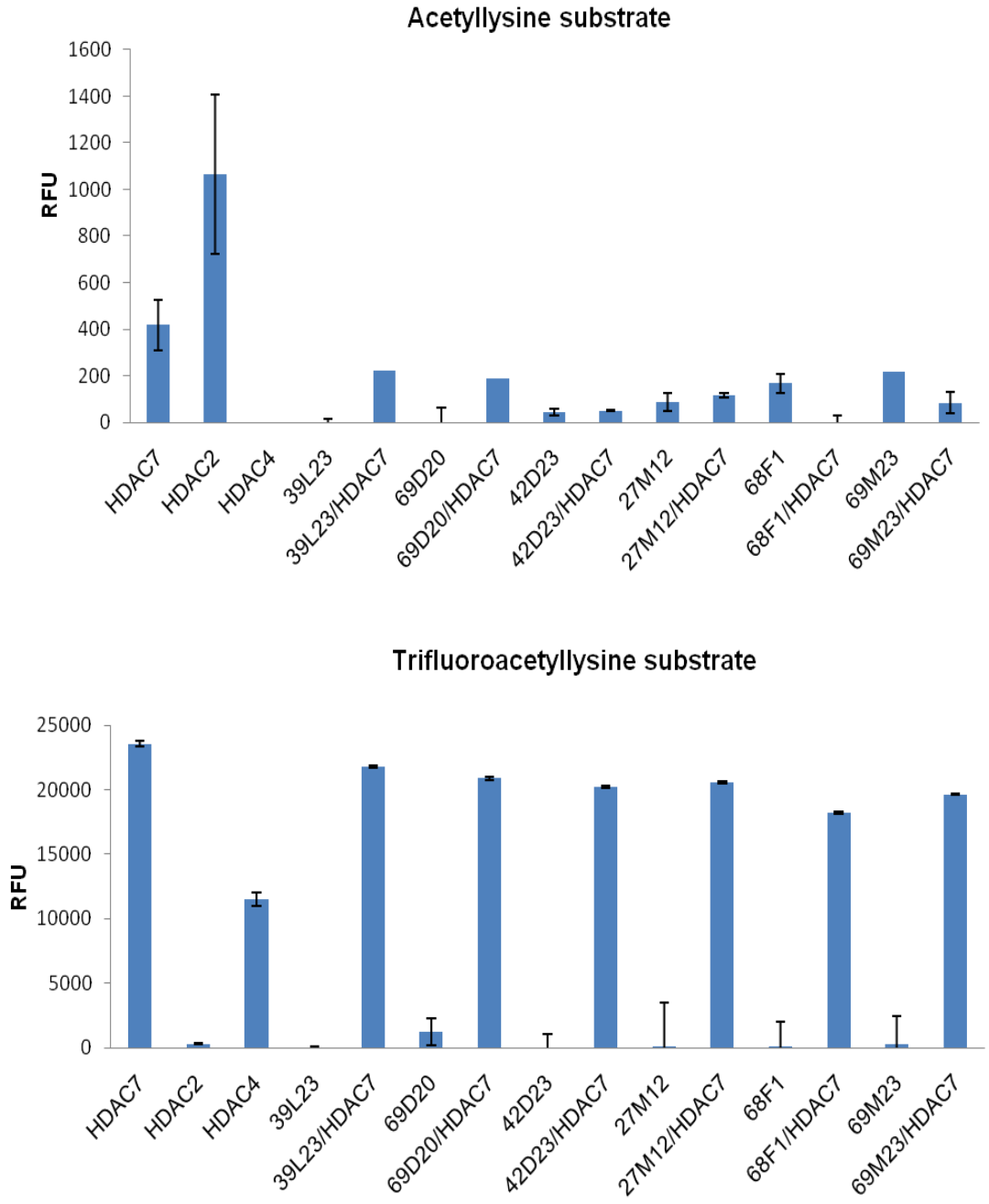


Figure 13. SMRT fragments do not rescue the cryptic HDAC7 deacetylase activity. In the upper histogram, the possible effect of HDAC activation on an acetyllysine peptide substrate was tested. The activity of HDAC7 was measured and showed insignificant levels, as with the HDAC4 control. The HDAC2 positive control provided in the assay kit showed strong activity. SMRT fragments alone did not exhibit activity on the acetyllysine substrate (not shown). In the lower histogram, deacetylase activity was measured on the trifluoroacetyllysine substrate. Purified HDAC7 (positive control) showed high activity on the substrate, even higher than the positive control HDAC4 provided with the deacetylase assay kit. HDAC2 (negative control) was not active on this substrate. No activity enhancement was observed when SMRT fragments were incubated with HDAC7.

Several hypotheses might explain this: HDAC7 might require other interacting proteins for activity, or other regions of SMRT to the region characterised here as

binding. Moreover as the substrate of HDAC7 is not known yet it might be possible that activation exists only in the presence of the proper substrate. Indeed HDAC7 shuttles between the cytoplasm and the nucleus in contrast to class I HDACs, therefore HDAC7 substrate might be a cytoplasmic protein rather than histone. It is the case for example that HDAC6 deacetylates tubulin (Hubbert et al. 2002; Serrador et al. 2004) and not histones. Alternatively, HDAC7 might not be an active enzyme but rather a scaffold protein for bringing interacting partners to the same place at the same moment for an activity to be processed by other HDACs in the corepressor complex. Then HDAC7 would rather behave like an acetyl recognition domain similar to bromodomains (Bottomley 2004; Moriniere et al. 2009). However it is puzzling that a fully functional active site is still maintained (except for one residue at 843 (His in class I replaced by a Tyr in class II) that has a suboptimal identity and appears to inactivate an otherwise perfectly functional enzyme).

Finally it seems entirely possible that studying isolated catalytic domains in dilute solution using Michaelis-Menten kinetics is misleading. At the natural (probable) site of action of HDAC7, chromatin, the giant flexible corepressor complex is probably physically associated with its substrate via a multitude of transcription factors and other DNA binding proteins. So there is probably not a high turnover of acetylated substrate proteins diffusing in and out of the HDAC active site at high velocity, as would be the case with a simple monomeric metabolic enzyme. Thus there would be little natural selective pressure to maximise the enzyme velocity (V_{max}) or substrate affinity (K_m). Indeed, perhaps the problem that has been addressed by evolution is that the enzyme is too active in this substrate-tethered form and the apparently inactivating Y308H substitution is necessary to slow it down.

Considering the role of SMRT, it is a classical hub protein interacting with many partners to bring them together in a coordinated manner, either at the same time and place, or via a specific cascade of events, binding their partners at different times and locations (Han et al. 2004). Indeed, about 120 SMRT-interacting partners are described in interaction databases (e.g. BioGrid (Breitkreutz et al. 2003; Stark et al. 2006)). Most of the time such scaffold IDPs have a modular organisation with several ordered domains involved in protein-protein interactions connected by long disordered regions. For example, CREB-Binding Protein (CBP), a transcription co-activator of CREB, is also such a scaffold protein (Goodman and Smolik 2000). At 2442 amino acids it is a similar length to SMRT and contains several well-folded domains in contrast to the one real

domain (HDAC3 DAD) discovered so far in SMRT. More than 50% of the CBP sequence including some functional domains are located in intrinsically disordered regions which has made it a model IDP for study (Dyson and Wright 2005). SMRT would appear to be an equally interesting system with much remaining to discover.

3 Conclusions

The ESPRIT random library construct screening technology was applied to the SMRT protein to obtain many soluble purifiable fragments that have never been studied previously. They were tested for binding to HDAC7 (also expressed using ESPRIT) using several complementary biophysical methods. One fragment of 130 amino acids interacted with SMRT at micromolar affinity and, from the NMR analysis, with high specificity. The availability of the two purified proteins at milligram levels from recombinant hosts permitted us to test, in the absence of any possible contaminants, whether binding of SMRT stimulated the HDAC7cryptic activity. At concentrations of protein where HDAC7 would be saturated with SMRT no enhancement of HDAC7 deacetylase was detected, demonstrating that this region of SMRT is not an HDAC7 activator. The so-called cryptic activity of type IIa HDACs remains unexplained.

From this work, we propose that the SMRT corepressor represents a fascinating IDP model system with much interesting biology to discover. The production of milligram quantities of purified, isotopically labelled material opens the way to studying many different interactions of this hub protein and the availability of these fragments in stable frozen aliquots ready to use in biophysical and biochemical studies represents a powerful new tool in this field. Moreover, the ESPRIT technology provides long constructs of IDPs (ranging from 92 amino acids to 240 amino acids in this study). This is far longer than the few fragments for which structures have been published of non-HDAC proteins binding SMRT-derived synthetic peptides (Progesterone Receptor with Bound Asoprisnil and a Peptide from SMRT: 17 residues in total with 10 engaged in one α-helix (Madauss et al. 2007), Estrogen Related Receptor-gamma ligand binding domain complexed with 4-hydroxy-tamoxifen and a SMRT peptide: 22 residues with 9 in one α-helix (Wang et al. 2006), BCL6 complexed with a SMRT co-repressor peptide: 19 residues in a β-sheet made of 5 residues (Ahmad et al. 2003) and human PPAR-alpha ligand-binding domain in complex with an antagonist GW6471 and a SMRT corepressor motif: 19 residues with 6 residues in one α-helix (Xu et al. 2002)). This makes possible the study of IDPs not only at the level of a peptide, but as long constructs that are perhaps more physiologically significant and provide multiple contact regions, maybe permitting measurements of protein-fishing and fly-casting effects. Already another IDP (a nucleoporin protein) has been processed in our laboratory with similar results, suggesting that this is a generic strategy.

CHAPTER 5

Summary & Future Perspectives

Les maladies cardiovasculaires constituent un problème majeur de santé publique pour tous les pays du monde. Ces maladies sont dues à l'accumulation de cholestérol transporté par les LDL dans les vaisseaux sanguins. Une stratégie courante consiste à cibler la voie de métabolisme du cholestérol en empêchant sa synthèse ou son absorption par l'organisme. Bien qu'il existe de nombreuses molécules telles que les statines ou ezetimibes, elles sont généralement peu efficaces pour les patients souffrant d'hypercholestérolémie familiale. En outre, ses traitements sont généralement prescrits pour toute la vie du patient, et peuvent causer des effets secondaires à long-terme. Pour toutes ces raisons, des molécules alternatives doivent être créées, de nombreux laboratoires et compagnies pharmaceutiques sont engagés dans cette voie.

Il a récemment été découvert que HDAC7 pourrait jouer un rôle important dans le catabolisme du cholestérol. HDAC7 régule l'activité d'une enzyme clé, l'enzyme 7α-cholestérol hydroxylase (ou cytochrome p450, ou Cyp7a1), impliquée dans la dégradation du cholestérol en acides biliaires, diminuant par conséquent les taux de cholestérol dans le sang. Cependant les HDACs sont des régulateurs clés pour de nombreux processus cellulaires, ainsi cibler les HDACs doit se faire de façon très spécifique pour éviter d'altérer d'autres fonctions. Actuellement des inhibiteurs spécifiques de chaque isoforme d'HDAC ne sont pas disponibles et les inhibiteurs que l'on possède présentent quelques effets secondaires. Ceci peut être acceptable pour le traitement des cancers car les traitements sont administrés sur de courtes périodes, en revanche, ils ne sont pas préconisés pour les traitements au long court comme dans le traitement de maladies cardiovasculaires. Mon projet consistait en l'étude et la compréhension de la spécificité d'inhibiteurs synthétisés par mes collaborateurs.

Tout d'abord, HDAC7 a été surexprimée et purifiée pour avoir le matériel nécessaire pour réaliser des études structurales, de biochimie et de biophysique. Malgré la publication d'un protocole pour le clonage, la purification et la cristallisation de HDAC7 d'un groupe de génomique structurale ayant donné lieu au dépôt de la structure de HDAC7 dans la PDB, la procédure s'est avérée difficile dans la mesure où HDAC7 était faiblement exprimée et hétérogène. La difficulté du protocole a été confirmée par ce laboratoire.

Apres le dépôt de la structure dans la PDB et publication des résultats de ce groupe, nous avons concentré nos efforts à la reproduction des conditions de cristallisation publiées dans le but de créer un système de cristallisation pour tester de multiples inhibiteurs. Cela s'est avéré sans succès, principalement due à de faibles rendements en protéine et une impossibilité d'obtenir des cristaux. Ainsi une autre étape d'ingénierie de protéine a été effectuée pour obtenir de façon empirique de meilleures constructions de HDAC7. La technologie ESPRIT (développée dans le laboratoire) a été utilisée à cet effet. Une bibliothèque de constructions génétiques obtenue par troncation N-terminale (délétion d'ADN de l'extrémité 5') a été synthétisée. La solubilité de chacune des constructions exprimées a été testée. Certains clones ont été sélectionnés, mais après suppression du rapporteur de solubilité et de son linker situés à l'extrémité C-terminale, les niveaux d'expression et de solubilité sont revenus à ceux de la construction d'origine. Des délétions de la région C-terminale ont été réalisées et ont montré que l'addition de six acides aminés non-natifs du linker rétablissait une bonne expression de protéine, cette protéine étant homogène (après chromatographie échangeuse d'anions et tamis moléculaire). Néanmoins la cristallisation a échoué avec cette construction optimisée soit seule ou en complexe avec le compose 13 (un inhibiteur prometteur synthétisé par des collaborateurs). La construction a donc subi une dernière étape d'ingénierie en enlevant 32 acides aminés invisibles dans la structure publiée et supposés non repliés. Des essais de cristallisation sont en cours avec cette construction.

En dépit des difficultés rencontrées pour obtenir des informations structurales sur HDAC7 liée au composé 13, son activité a pu être mesurée. Deux substrats ont été utilisés : les substrats lysine trifluoroacetylée et lysine acétylée. L'activité de déacétylase a été démontrée pour la lysine trifluoroacetylée, substrat spécifique des HDACs de classe IIa à laquelle HDAC7 appartient. En revanche, HDAC7 s'est révélée inactive sur la lysine acétylée (ce qui avait été montré pour les HDACs de classe IIa). Une fois l'activité de HDAC7 mesurée, des analyses comparatives d'inhibition ont été réalisées pour SAHA et le composé 13. Le composé 13 inhibait 3 fois plus HDAC7 à la fois *in vivo* et *in vitro* suggérant que le composé 13 établirait plus de contacts d'interaction avec HDAC7 que SAHA. Il a été démontré que les substitutions du groupement fonctionnel phényle du cap de SAHA constituaient une modification importante améliorant l'activité et la sélectivité des inhibiteurs d'HDACs. Le cap du composé 13 a été modifié par substitution avec des groupements methyle en positions meta. Nous avons supposé que cette substitution permettait d'établir plus de contacts avec HDAC7 à l'entrée de son site actif, augmentant

son affinité pour HDAC7 et ainsi sa spécificité. Des essais de cristallisation ont été effectués pour expliquer cet effet mais aucun cristal de protéine n'a été obtenu jusqu'à maintenant. Des essais sont encore en cours avec la dernière construction améliorée de HDAC7.

Enfin, une stratégie alternative a été choisie : cibler l'interface d'interaction entre HDAC7 et un de ses partenaires, le corépresseur SMRT. SMRT interagit avec HDAC3 et induit l'activation de sa fonction de deacetylation sur des substrats acétylés. HDAC7 interagit avec le domaine de répression III de SMRT. Etudier cette interaction permettrait de déterminer si l'activité de HDAC7 sur des substrats acétylé est rétablie et s'il est possible d'empêcher le recrutement de HDAC7 par SMRT. Mais SMRT est une énorme protéine nativement désordonnée, il n'était donc pas certain de pouvoir obtenir de la protéine soluble considérant en outre l'absence de domaine fonctionnel. Il était donc impossible de créer, par des méthodes rationnelles, des constructions s'exprimant de façon soluble. Pour cette raison la méthode ESPRIT a été utilisée, avec un accent sur la région de SMRT contenant le domaine de répression III. Cette méthode a permis de générer des protéines alternatives de SMRT qui ont été testées pour leur expression et solubilité. Leur éventuelle interaction avec HDAC7 ainsi que leur capacité à activer l'activité de HDAC7 ont été testées. Au total, 51 constructions solubles et purifiables ont été obtenues, se répartissant sur toute la région de SMRT considérée. L'activation de déacétylation de HDAC7 a été testée avec six protéines hautement surexprimées et purifiées. Aucune d'elles n'a eu d'effet sur l'activité de déacétylation de HDAC7. Des tests d'interaction ont été menés par RMN, résonance plasmonique de surface et anisotropie de fluorescence. L'anisotropie de fluorescence a révélé que deux fragments (39L23 et 69D20, avec 39L23 légèrement plus longue que 69D20) de la même région interagissaient avec HDAC7. Des expériences ^{15}N-HSQC ont montré le même résultat pour le fragment 69D20, bien qu'aucune interaction n'ait été détectée pour la construction plus longue 39L23. Aucune explication claire n'a permis d'expliquer cette contradiction à l'heure actuelle : il est possible que la construction la plus longue soit dans une conformation empêchant l'interaction. De très bons spectres ont été obtenus pour la construction 39L23, c'est donc cette construction qui a été utilisée pour identifier les résidus impliqués dans l'interaction. Cette expérience est actuellement en cours et semble indiquer que les données collectées sont de bonne qualité pour permettre une attribution des résidus (collaboration : Martin Blackledge, IBS). L'utilisation de bibliothèque de constructions génétiques aléatoires telle que la technologie ESPRIT, constitue une

approche innovante pour l'étude de différent types d'interaction comme le repliement d'une protéine induit par l'interaction avec son partenaire. En effet, sans cette technologie la seconde partie du projet concernant l'interaction SMRT-HDAC7 n'aurait pas pu être réalisée. Un avantage majeur de cette méthode réside dans la possibilité d'obtenir plusieurs protéines solubles, au lieu d'une unique protéine, pouvant être testées à la demande pour différentes interactions, comme ce fut le cas avec 51 protéines de SMRT. Cette méthode permet d'envisager des études d'interaction de grande envergure avec la possibilité de conserver des fragments de protéines purifiées éventuellement isotopiquement marqués et de les tester à la demande pour différentes interactions. Il y a environ 120 partenaires validés expérimentalement interagissant avec SMRT.

Des essais de cristallisation de HDAC7 et de HDAC7 liée au composé 13 sont actuellement en cours, il est donc encore possible d'élucider les mécanismes de sélectivité du composé 13. De plus, la structure d'un des fragments de SMRT sera déterminée par RMN, des expériences sont actuellement en cours, et permettront l'identification des résidus et/ou domaines impliqués dans l'interaction, ce travail sera publié dans les prochains mois.

Cardiovascular diseases (CVDs) represent a serious health burden for countries all over the world. Several of these conditions have their origin in the accumulation of cholesterol carried by LDL in blood vessels. A common strategy is to target cholesterol metabolism by preventing its synthesis and/or adsorption. Although existing drugs like statins and eztimibes are generally quite efficient, they are ineffective for treating familial hypercholesterolemia. Moreover, as these treatments are prescribed during the whole life of the patient, long-term side effects are likely and unacceptable. For these reasons alternative more effective drugs must be designed and pharmaceutical companies and academics are actively engaged towards this objective.

It was discovered recently that HDAC7 may play an important role in cholesterol catabolism. HDAC7 regulates expression of a key cytochrome p450 enzyme, Cyp7a1, involved in the degradation of cholesterol into bile acids, thereby decreasing cholesterol levels in blood. HDAC7 has therefore been proposed as a potential drug target to design new molecules. However as HDACs are key regulators in cells, targeting them should ideally be done in a highly specific manner so as not to alter any other cellular functions. Currently such HDAC isoform-specific inhibitors are not available and the present pan-specific HDACi compounds exhibit some side effects. These may be acceptable for cancer therapy, but are not for long term CVD treatment. My project was designed so to study inhibitors synthesised by my collaborators and to understand key features of their specificity.

First of all HDAC7 had to be expressed and purified to produce material for biochemical, biophysical and structural experiments. Despite published protocols for cloning, purification and crystallisation from a structural genomics initiative who deposited the HDAC7 structure in the PDB, this was not trivial since the protein produced according to these protocols (construct, purification protocol) was poorly expressed and structurally heterogeneous. These issues were confirmed by that laboratory as significant obstacles in their work also.

Following the structure determination and publication of an accompanying paper nearly one year later, we refocused our efforts into reproducing the published crystallisation conditions with the aim of generating a crystallisation system for testing multiple small molecule compounds. However this was unsuccessful, mainly due to very

low protein yields and absence of crystals. Therefore protein engineering was performed to identify empirically better behaving HDAC7 constructs. For this purpose the ESPRIT technology (developed in our laboratory) was applied to HDAC7. An N-terminal (5′ DNA deletion) truncation library was synthesised and tested for soluble expression. Some clones were selected with greatly improved characteristics but after removal of the C-terminal solubility reporter and its linker, expression and solubility decreased back to that of the original construct. Therefore deletion analyses were performed on this C-terminal region and revealed that the addition of six non-native C-terminal amino acids rescued expression and resulted in protein preparations that were apparently structurally homogeneous (from ion exchange and gel filtration data). However crystallisation was again unsuccessful either in the case of apo-HDAC7 or HDAC7 with compound 13 (a promising compound designed by collaborators). The construct was then further optimised by removal of 32 N-terminal disordered residues (invisible in the structure). Crystallisation trials are ongoing with this construct.

Notwithstanding the difficulties in obtaining structural information on HDAC7 bound to compound 13, its enzymatic activity could be assayed. Two substrates were used: the trifluoroacetyllysine and acetyllysine substrates. HDAC7 showed activity on the trifluoroacetyllysine substrate, designed specifically for class IIa HDACs (to which HDAC7 belongs) and was inactive on the acetyllysine substrate (as described previously for class IIa HDACs). Once the activity was assayed, comparative inhibition analyses were performed between SAHA and compound 13. Compound 13 revealed to be 3-fold more efficient on HDAC7 both *in vivo* and *ex vivo*. Biophysical analyses using thermal denaturation followed by circular dichroism indicated that HDAC7 binding compound 13 was more stable than HDAC7 alone or when bound to SAHA. This suggested that compound 13 may make more interaction contacts with HDAC7 than SAHA. This difference in binding affinity was explained by the chemical differences between compound 13 and SAHA. It was shown that modifying substitutions in the cap group (phenyl ring) of SAHA (HDAC pan-inhibitor) was an important modification that enhanced potency and selectivity of HDAC inhibitors. Compound 13 was modified in its cap moiety by an addition of two methyl groups in the meta positions. We speculate that this substitution allows compound 13 to make more contacts with HDAC7 at the rim of the active site, thereby enhancing affinity and, as a result, specificity. Crystallisation trials were performed to explain this effect but no crystals were obtained so far; trials are still ongoing with the last improved HDAC7 construct.

Lastly, another strategy was investigated to target the binding interface between HDAC7 and an essential cellular partner, the SMRT corepressor. SMRT interacts with HDAC3 and induces its activity upon acetylated substrates. Since it has been shown that HDAC7 interacts with SMRT in its repression domain III, it was decided to study this interaction to a) determine whether binding SMRT relieved the low activity of HDAC7 on acetyllysine substrates, and b) see whether inhibition of this interface blocks recruitment of HDAC7 into the larger complex in which it is active. However SMRT is a giant intrinsically disordered protein (IDP), and it was unclear how to obtain soluble protein since the absence of predicted soluble domains meant that it was impossible to design rationally constructs. It was decided to perform an ESPRIT screen on a region of SMRT containing the repression domain III to identify soluble proteins, and to test them for interaction and deacetylase activation. In total, 51 soluble purifiable constructs were obtained spanning most areas of this SMRT region. The six most highly expressed and purified proteins were tested for deacetylation activation of HDAC7. None of them rescued HDAC7 activity on the substrates. Binding tests were performed by NMR, surface plasmon resonance and fluorescence anisotropy. Fluorescence anisotropy revealed that two fragments (39L23 and 69D20, with 39L23 being slightly longer than 69D20) of the same region interacted with HDAC7. ^{15}N HSQC showed the same preference for 69D20, although the longer construct 39L23 did not interact in that experiment. At this point the explanation for this is unclear: it is possible that the longer construct is in a conformation that reduces the interaction. The 39L23 fragment gave the best resolved spectrum so was ^{15}N- and ^{13}C-labeled to assign by NMR the residues involved in the interaction with HDAC7. This experiment is currently ongoing and preliminary results indicate that the acquired dataset is of a sufficient quality to complete the assignment (collaboration: M. Blackledge, IBS). Applying random library construct screening (ESPRIT) to an IDP is a completely novel approach to obtaining material for study of interactions e.g. folding-upon-binding effects upon complex formation. Indeed, without this technology the SMRT-HDAC7 project would not have been possible. A significant and novel advantage of this technology is that not one soluble protein, but more than 50 were obtained, allowing the creation of a system in which a library of frozen purified isotope-labelled fragments can be tested on-demand for each new interactor. There are over 120 experimentally validated interactors of SMRT.

Crystallisation of HDAC7 and compound 13 are ongoing and will still hopefully provide structural insights into the increased selectivity observed for HDAC7.

Additionally the structure of one binding SMRT fragment will be determined by NMR and will allow identify residues and/or domains involved in this interaction; this work will be published in the next few months.

CHAPTER 6

Materials and Methods

1 Materials and methods from chapter 2

Genes

A synthetic gene encoding human HDAC7 (UNIPROT: Q8WUI4; Leu449-Leu952) was purchased from Geneart with codon optimised sequence for *E. coli* expression. This optimisation addresses codon bias, internal ribosome binding sites and mRNA secondary structure.

HDAC7 synthetic gene sequence:

```
5'ctgctgtgggaacagcagcgtctggccggtcgtctgccgcgtggtagcacgggtgatacggtgctgctgc
cgctggcccagggtggtcatcgcccgctgtctcgtgcacagagcagcccggcagcaccggccagcctgtctg
caccggaaccggcgagccaggcgcgtgttctgagcagcagcgaaaccccggcacgtaccctgccgtttacca
ccggcctgatttatgatagcgtgatgctgaaacatcagtgcagctgcggcgataacagccgtcatccggaac
atgcgggccgtattcagagcatttggagccgtctgcaggaacgtggcctgcgtagccagtgcgaatgcctgc
gtggccgtaaagcgagcctggaagaactgcagagcgtgcatagcgaacgtcatgtgctgctgtatggcacca
atccgctgagtcgcctgaaactggataacggcaaactggccggcctgctggcccagcgtatgtttgtgatgc
tgccgtgcggcggtgtgggtgtggataccgataccatttggaacgaactgcatagcagcaatgcggcgcgtt
gggcggcaggtagcgttaccgatctggcctttaaagtggcgagccgtgaactgaaaaacggctttgcggtgg
tgcgtccgccgggtcatcatgcggatcatagcaccgcgatgggcttttgctttttaacagcgtggcgattg
cgtgccgtcagctgcagcagcagagcaaagcgagcaaaattctgattgtggattgggatgtgcatcatggca
acggcacccagcagacctttttatcaggatccgagcgtgctgtatattagcctgcatcgtcatgatgatggca
actttttttccgggcagcggtgcggtggatgaagtgggtgcgggcagcggcgaaggctttaacgtgaacgtgg
cgtgggcgggtggtctggatccgccgatgggcgatccggaatatctggccgcgtttcgtattgtggtgatgc
cgattgcgcgtgaatttagcccggatctggtgctggtgagcgcgggttttgatgcggcggaaggccatccgg
caccgctgggtggttatcatgtgagcgcgaaatgctttggctatatgacccagcagctgatgaacctggccg
gtggtgcggttgttctggccctggaaggcggccatgatctgaccgcgatttgcgatgcgagcgaagcgtgtg
ttgcggcgctgctgggcaatcgtgtggatccgctgtctgaagaaggctggaaacagaaaccgaacctgaacg
cgattcgtagcctggaagcggtgattcgtgtgcatagcaaatattggggctgcatgcaacgcctggcctctt
gtccggactcttgggttccgcgtgttccgggtgcggataaagaagaagtggaagcggttaccgcgctggcta
gcctgtctgtgggcattctggccgaagatcgtccgagcgaacagctggttgaagaggaagaaccgatgaacc
tg 3'
```

Translation:

```
LLWEQQRLAGRLPRGSTGDTVLLPLAQGGHRPLSRAQSSPAAPASLSAPEPASQARVLSSSETPARTLPFTT
GLIYDSVMLKHQCSCGDNSRHPEHAGRIQSIWSRLQERGLRSQCECLRGRKASLEELQSVHSERHVLLYGTN
PLSRLKLDNGKLAGLLAQRMFVMLPCGGVGVDTDTIWNELHSSNAARWAAGSVTDLAFKVASRELKNGFAVV
RPPGHHADHSTAMGFCFFNSVAIACRQLQQQSKASKILIVDWDVHHGNGTQQTFYQDPSVLYISLHRHDDGN
FFPGSGAVDEVGAGSGEGFNVNVAWAGGLDPPMGDPEYLAAFRIVVMPIAREFSPDLVLVSAGFDAAEGHPA
PLGGYHVSAKCFGYMTQQLMNLAGGAVVLALEGGHDLTAICDASEACVAALLGNRVDPLSEEGWKQKPNLNA
IRSLEAVIRVHSKYWGCMQRLASCPDSWVPRVPGADKEEVEAVTALASLSVGILAEDRPSEQLVEEEEPMNL
```

Cloning of DNA inserts

Initial expression trials. Constructs for initial expression tests were constructed using standard PCR cloning methods with both wild-type (Genbank accession: 51564) and synthetic gene as template.

ESPRIT experiments. The synthetic HDAC7 gene was cloned into a modified pET9a vector (Novagen) , pESPRIT1 (Fig. 1), containing N-terminal hexahistidine tag and TEV protease cleavage site (MGHHHHHHDYDIPTTENLYFQG) and C-terminal biotin

acceptor peptide with linker (SNNGSGGGLNDIFEAQKIEWHE). The plasmid was designed with two restriction sites flanking the target gene to express. They are located on the 5' end of the gene, *Aat*II/*Asc*I respectively; and on the 3' end *Nsi*I/*Not*I. These pairs of restriction sites define the truncation direction of the exonuclease. Here, two different constructions were to be processed encoding HDAC7 Leu449-Leu952 (HDAC7EspFOR1: 5' gatcct agggc gcgcc gctgc tgtgg gaaca gcagc gtctg g 3'; HDAC7EspRevPMNL: 5' catgg atcat gcatt caggt tcatc ggttc ttcct cttca acc 3', and Leu449-Ser903 (HDAC7EspFOR1: 5' gatcc taggg cgcgc cgctg ctgtg ggaac agcag cgtct gg 3'; HDAC7EspRevRLAS: 5' catgg atcat gcatt agagg ccagg cgttg catgc agc 3').

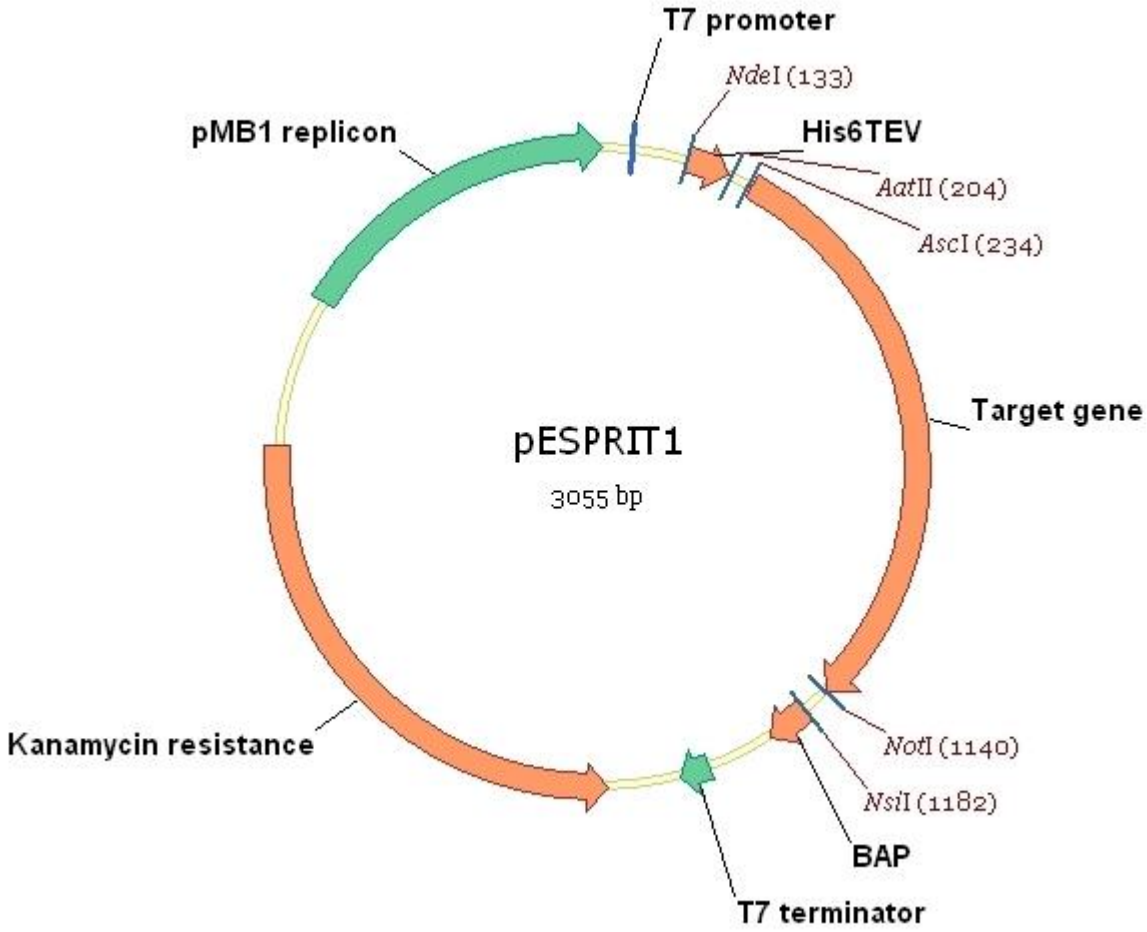


Figure 1: pESPRIT1 used for library construction. This vector was used for both HDAC7 and SMRT (chapter 5) analyses.

ESPRIT Protocol

Library generation. Plasmid was yielded from 200 ml overnight culture *E. coli* by classical alkaline lysis and phenol chloroform extraction and ammonium acetate precipitation protocols. Plasmid was further purified with a plasmid miniprep kit (Qiagen) to remove remaining RNA. 10 µg of supercoiled plasmid were cut to completion with

library specific restriction enzymes (*Aat*II and *Asc*I) in 200 µl total reaction volume. 4 µg of linearised plasmid was truncated with 400 U of ExoIII (NEB) at 22°C in 120 µl total volume. Truncation rate was adjusted to the DNA size using NaCl concentration calculated: $N = (Iog(\delta/47.9))/-0.0064$ (Ostermeier and Lutz 2003), where N=NaCl concentration (mM; here 100 mM), and δ is enzymatic rate expressed in bp/min. Each 60 s during 1 h, 2 µl of the reaction was quenched into 200 µl of 3 M NaCl. At the end of the truncation experiment, exonuclease is inactivated at 70°C for 20 min. Then, DNA is purified with Nucleospin Extract (Macherey Nagel) and eluted in 35 µl. The single stranded DNA previously generated is digested with mung bean nuclease (NEB) using 5 U of enzyme, at 30°C, 30 min. Finally, blunt ends were polished with *Pfu* (Stratagene) treatment at 72°C for 20 min with 25 mM dNTPs mix. Libraries were electrophoresed at 5 V/cm on EtBr-containing 0.5% w/v agarose gels at 4°C. The HDAC7 gene was approximately 1500 bp, and the smear was size fractionated in three slices of about 500 bp (0-500 bp, 500-1000 bp and 1000-1500 bp, with linearised vector 3000-3500 bp, 3500-4000 bp and 4000-4500 bp). DNA from gel slices was extracted using QiaexII (Qiagen). Purified DNA was recircularised by ligation with T4 DNA ligase and MACH1 (Invitrogen) chemically transformed. The total population of clones was estimated, and colonies were scraped from plates and pooled, then plasmid extracted using a midiprep kit (Qiagen). For all libraries, electrocompetent BL-21 AI (DE3) cells (Invitrogen), additionally transformed with the RIL plasmid (Stratagene), were transformed by electroporation.

Preparation of agar plates for screening. For colony picking, Qtrays (X6023, Genetix) were prepared with 300 ml of LB-agar supplemented with kanamycin (50 µg/ml) and chloramphenicol (50 µg/ml) poured on a horizontal surface under a microbiological hood. Induction plates were prepared in the same way, except that they were supplemented with 50 µM biotin and inducer (0.2% L-arabinose).

Colony picking with picker gridder. Transformants were spread on Qtrays in a dilution series so as to have 4000 colonies per plate. Then, a picker-gridder robot (KBiosystems) isolated clones into 384 well plates (X7007, Genetix) containing 80 µl of TB with 10% final glycerol. Cultures were grown overnight at 37°C in a HiGro incubator (GeneMachines).

Gridding and macroarray preparation. The library was replicated in 384 well plates in fresh TB medium with antibiotics. Clones were printed robotically on nitrocellulose

membranes (Amersham) laid on Qtrays containing 300 ml LB agar with antibiotics. The gridding pattern was chosen according to total number of 384 well plates of the library. Colonies were grown overnight at room temperature and nitrocellulose membranes were transferred to prewarmed Qtrays containing the inducer at 30°C. The induction was performed for 4 h at 30°C. *In situ* lysis was performed by transferring membranes onto filter paper soaked with denaturing solution (500 mM NaOH, 1.5 M NaCl) for 10 min. Two cycles of neutralisation on filter paper soaked with neutralisation solution pH7.5 (1 M Tris HCl, 1.5 M NaCl) were performed for 5 min each. Finally, the nitrocellulose arrays were incubated in 2× SSC solution pH7 (300 mM NaCl, 35 mM Na Citrate) for 15 min and cell debris removed gently.

Immunofluorescent detection. The activated membrane was placed in a roller (Techne) and washed with water and 0.1% Tween-PBS for 5 min at 4°C, then blocked overnight with 50 ml of SuperBlock (Pierce). The membrane was washed after blocking 3 × 5 min with 0.1% Tween-PBS, then incubated with 1/3125 dilution in 0.1% Tween-PBS of mouse anti-his-tag primary antibody (Amersham) for 1 h at 4°C. The membrane was washed 3 × 5 min with 0.1% Tween-PBS. Both streptavidin ALEXA488 (Molecular probes) diluted 1/5000 and secondary antibody anti-mouse ALEXA532 diluted 1/1000 were incubated for 1 h at 4°C. Membranes were then washed 3 × 5 min with 0.1% Tween-PBS, and finally rinsed with distilled water.

Fluorescence scanning and data analysis. A Typhoon scanner (GE Healthcare) was used for scanning the membranes. The settings used were conserved for the two detections used for one membrane. Data was saved for subsequent analysis using VisualGrid software (GPC Biotech) to extract pixel intensity to Excel (Microsoft). Clones with no visible hexahistidine signal were eliminated and the remainder ranked upon biotin signal to generate a list of clones according to their signal intensities.

High throughput E. coli expression testing. The 96 first hits were grown overnight at 37°C at 220 rpm in 96 deep-well plates (Qiagen) in the HiGro. Protein expression trials were performed from 4 ml TB medium in 4 × 24 deep-well plates (Qiagen) and inoculated with a 1/100 dilution of overnight inocula. Cultures were grown at 37°C to OD_{600} 0.6-0.8 and induced by addition of L-arabinose at 25°C overnight and then harvested by centrifugation at 3700 rpm for 10 min in a swinging rotor. Cell pellets were resuspended with 4 ml of buffer containing 20 mM Tris pH 8, 250 mM NaCl, 20%

sucrose and 1 mg/ml lysosyme and incubated 15 min in cold room on a rocking platform. The resulting sphaeroplasts were pelleted at 3700 rpm, 15 min. The supernatant was removed and pellets kept at -80°C for at least 30 min prior to resupension in 700 µl of buffer containing 10 mM Tris pH 8, 0,5% Brij58, 1/1000 dilution of benzonase, 5 mM β-ME and protease inhibitor cocktail without EDTA. An aliquot is saved for total fraction analysis. 96 protein purifications were carried out in parallel using a TECAN Liquid handling robot (Tecan). This protocol was a modification from a published version (Scheich et al. 2003). 30 ml of Ni^{2+}-NTA agarose resin (Qiagen) was dispensed in a 96 well receiver plate (Macherey Nagel). Lysates were distributed into each well of the receiver plate containing washed resin for in batch binding for 30 min in the cold room. Wells were washed with 10 volume of washing buffer (300 mM NaCl, 50 mM Na_2HPO_4/NaH_2PO_4 pH 7.5 mM imidazole, 5 mM β-ME) prior to elution performed with 70 ml of elution buffer (300 mM NaCl, 50 mM Na_2HPO_4/NaH_2PO_4 pH 7.5, 300 mM imidazole, 5 mM β-ME). Total and purified protein fractions were mixed with Laemmli buffer and boiled for 10 min.

Large-scale expression and purification. HDAC7 was produced in *E. coli* BL21 AI RIL in Terrific Broth medium (50 µg/ml kanamycin, 50 µg/ml chloramphenicol). The medium was supplemented with 50 µM of $ZnCl_2$ 30 min to 1 h before induction. Expression was induced with 13 mM L-arabinose at OD_{600nm} 0.5-0.8. Cells were harvested by centrifugation and resuspended in lysis buffer (25 mM Hepes pH 7.5, 200 mM KCl, 30% v/v glycerol, 0.5% Igepal (Sigma Aldrich, St Louis, MO), 1 mM DTT and EDTA-free complete protease inhibitor cocktail (Roche Diagnostics, Indianapolis, lN). Cells were lysed with a rnicrofluidiser (M11O-L; Microfluidics Corporation, Newton, MA, USA) and the lysate cleared by centrifugation (48000 g, 40 min, 4°C).

The supernatant was incubated overnight with Ni^{2+}-NTA agarose resin (Qiagen) at 4°C, loaded in to a plastic column. Protein was eluted in elution buffer (25 mM Hepes pH 7.5, 200 mM KCl, 10% glycerol, 0.1% octyl-P-D-glucopyranoside (Sigma Aldrich), 150 mM imidazole). The eluted fractions were diluted 4 times in buffer (25 mM Hepes pH 7.5, 10% glycerol, 0.1% octyl-P-D-glucopyranoside, 1mM DTT) and purified by anion exchange (Q sepharose 16/10 column, GE Healthcare) with elution at 200 mM KCl. The hexahistidine tag was removed by incubation with hexahistidine tagged tobacco etch virus protease overnight at 4°C. Finally the protein was purified by gel filtration (Superdex 75 column 16/60, GE Healthcare) in buffer (25 mM Hepes pH 7.5, 150 mM

KCl, 0.1% octyl-P-D-glucopyranoside and 1mM DTT), dialysed overnight at 4°C in dialysis buffer (50 mM KH_2PO_4/K_2HP0_4, 150 mM KCl, 0.1% octyl-P-D-glucopyranoside, 1 mM DTT), and concentrated to 6 mg/ml with an Amicon Ultra concentrator (Millipore).

Crystallisation trials

Purified HDAC7 was concentrated between 6 to 12 mg/ml with an Amicon Ultra concentrator (Millipore). Concentrated proteins were then submitted to the high-throughput nanodrop crystallisation facility at EMBL Grenoble.

Surface lysine methylation was performed following this protocol (JBS Methylation kit, Jena Bioscience, Germany). The purified protein (in Hepes buffer) was incubated twice with 40 µl of dimethylamine borane complex and 20 µl of 1 M formaldehyde at 4°C for 2 h. Then 10 µl of dimethylamine borane complex was added, the mixture incubated overnight at 4°C. The reaction was quenched by adding 125 µl of 1 M Tris pH 7.5. The protein was finally separated from the mixture by size exclusion chromatography (superdex 75 10/300 GE Healthcare) and then incubated with DTT at a final concentration of 5 mM.

2 Materials and methods from chapter 3

Enzyme kinetics and inhibition experiments

Single point analysis. 1 ng (8 nM) of our purified HDAC7, commercial HDAC4 and commercial HDAC2 were incubated with 20 µM of substrate (trifluoroacetyllysine and acetyllysine substrates) in HDAC assay buffer provided with the assay kit (BPS Biosciences). The reaction was incubated 1.5 h at 37°C, then stopped by a developer solution provided (containing 50 µM TSA and trypsin which develops the fluorescence signal) and further incubated for 25 min at room temperature (about 22 °C). The fluorescence intensity was detected with a spectrophotometer Perkin Elmer Victor^2. The fluorescence intensity was detected with excitation-emission wavelengths of 355 nm and 460 nm respectively and plotted as relative fluorescence units (RFU).

Compound Profiling. The concentration of HDAC7 was optimised to ensure the detection of deacetylase activity in the linear range of the reaction for 60 min with less than 10% of substrate being depleted, ie 1 nM (8 nM). The K_m of HDAC7 was determined by trifluoroacetyllysine substrate titration (substrate concentrations: 5 µM, 10 µM, 20 µM, 40 µM, 80 µM, 100 µM, 150 µM, 200 µM). The reaction were carried out at 37°C over 20 min and quenched by the developer solution for 25 min at room temperature. The fluorescence signal was measured as described above. Data were analysed by SigmaPlot for kinetic constants determination by curve-fitting.

For compound profiling trifluoroacetyllysine substrate was used at 10 µM. Compound 13 and SAHA dissolved in DMSO were assayed in triplicates at 11 concentrations: 6 nM, 33 nM, 66 nM, 330 nM, 660 nM, 3.3 µM, 6 µM, 33 µM, 66 µM, 330 µM and 660 µM. The reactions were performed in white NUNC 96-well plates. The reaction was incubated for 1 h at 37°C and stopped with developer solution (containing 50 µM TSA). The plates were then further incubated with the solution developer 25 min at room temperature. The fluorescence intensity was detected as above. The end point relative fluorescence (RFU) data were transformed as percentage of activity to compare the relative potency of compound 13 and SAHA on HDAC7. Curve-fitting was performed by SigmaPlot using a sigmoidal dose-response model to determine IC_{50}.

Thermal denaturation followed by circular dichroism

Thermal denaturation analyses followed by circular dichroism were carried out on a Jasco-810 spectropolarimeter (Jasco Ltd, Tokyo, Japan) equipped with a temperature control. Thermal unfolding of HDAC7 alone or incubated with compound 13 or SAHA was monitored by recording the ellipticity at 220 nM as a function of temperature. Ellipticity measurements were converted to percentage of folded/unfolded protein. Data were plotted with PRISM and melting temperatures where 50% of protein was folded and 50% unfolded, were calculated.

Crystallisation experiments

Crystallisation experients of HDAC7 and compound 13 were carried out according to chapter 2 material and methods (section 1).

3 Materials and methods from chapter 4

Gene

A synthetic gene encoding human SMRT (UNIPROT: Q9Y618; Gly1089_Val1993) was used. The synthetic gene was purchased from Geneart with optimised sequence for *E. coli* expression.

SMRT synthetic gene sequence:

```
5'ggtcacccactgccactgggcctgcatgataccgcacgtccggttctgccacgtccaccgaccattagca
atccacctccactgattagcagcgcgaaacatccgagcgtgctggaacgtcagattggcgcaattagccagg
gcatgagcgtgcagctgcatgtgccgtatagcgaacatgcgaaagcaccggtgggtccggttaccatgggtc
tgccgctgccgatggatccgaaaaaactggcaccgttttctggcgtgaaacaggaacaactgtctccacgcg
gtcaggcaggtccaccggaaagcctgggtgttccgaccgcacaggaagcgagcgttttacgcggtacggcac
tgggtagcgttccaggcggtagcattaccaaaggcattccgagcacccgtgtgccgagcgatagcgcgatta
cctatcgtggcagcattacccatggcaccccagcggatgtgctgtataaaggcaccattacccgtattattg
gcgaagatagcccgagccgtctggatcgtggccgtgaagatagcctgccgaaaggccatgtgatttatgaag
gcaaaaaaggccatgttctgagctatgaaggcggtatgagcgtgacccagtgcagcaaagaagatggccgta
gcagcagcggtccaccgcatgaaaccgcagcaccgaaacgtacctatgatatgatggaaggccgtgtgggtc
gtgcgatcagcagcgcgagcattgaaggcctgatgggtcgcgcaattccaccggaacgtcattctccgcatc
atctgaaagaacagcatcatattcgcggtagcatcacccagggcattccgcgtagctatgtggaagcgcagg
aagattatctgcgtcgcgaagcgaaactgctgaaacgtgaaggcaccccaccgcctccgccaccgagccgtg
atctgaccgaagcgtataaaacccaggcgctgggtccgctgaaactgaaaccggcacatgaaggcctggtgg
cgaccgtgaaagaagcgggtcgtagcattcatgaaattccgcgtgaagaactgcgtcatactccggaactgc
cgctggcaccgcgtccgctgaaagaaggttctatcacgcagggtacgccactgaaatatgacaccggtgcga
gcaccaccggcagcaaaaaacatgatgtgcgtagcctgattggctctccgggtcgcacctttccaccagttc
acccactggacgtgatggcggatgcgcgtgcgctggaacgtgcgtgctatgaagaaagcctgaaaagccgtc
cgggtacggcctcttcttccggtggcagcattgcgcgtggtgcaccggtgattgtgccggaactgggcaaac
cgcgtcagtctccgctgacctatgaagatcatggtgcaccgtttgcgggtcatctgccacgtggttctccgg
tgaccatgcgtgaaccgactccgcgtctgcaggaaggtagcctgtctagctctaaagcgagccaggatcgta
aactgaccagcacaccacgcgagattgcgaaatctccgcatagcaccgtgccggaacatcatccgcatccga
ttagcccgtatgaacatctgctgcgtggcgtgtctggcgtggatctgtatcgtagccatattccgctggcct
ttgatccgaccagcattccgcgtggcattccgctggatgcggcagcagcgtattatctgccacgtcatctgg
caccgaatccgacctatccgcatctgtatccaccgtatctgattcgtggctatccggataccgcagcgctgg
aaaaccgtcagaccatcatcaacgattatattaccagccagcagatgcaccataacaccgcgacggcaatgg
cccagcgtgcggatatgctgcgtggtctgtctccgcgtgaaagcagcctggccctgaactatgcagcgggtc
cgcgtggtattattgatctgagccaggtgccgcatctgccggttctggttccaccgactccgggtactccgg
caaccgccatggatcgtctggcctatctgccgaccgcaccgcagccgtttagcagccgtcatagcagctctc
cgctgtctccgggtggtccgacccatctgaccaaaccgaccaccacgagcagcagcgaacgtgaacgtgatc
gcgatcgtgaacgtgaccgcgaccgcgagcgcgaaaaaagcattctgacttctacgacgacggtggaacatg
caccgatttggcgtccgggtacggaacagagcagcggtagcagcggcagcagcggtggtggtggtggttctt
ctagtcgcccagcaagccatagccatgcgcatcagcatagcccgatctctccgcgtacccaggatgccctgc
aacaacgtccgagcgttctgcataacaccggcatgaaaggcattattaccgcagtggaaccgagcaaaccga
ccgtgctgcgtagcaccagcaccagctctccggttcgtccggcagcgaccttcccaccggcaacccattgtc
cgctgggtggcacgctggatggtgtgtatccgaccctgatggaaccggttctgctgccaaaagaagccccctc
gcgtcgctcgtccagagcgcccacgcgccgataccggtcatgcgtttctggccaaaccaccggcacgtagcg
gtctggaaccggcaagcagcccgagcaaaggctctgagccacgcccactggtt 3'
```

Translation:

```
GHPLPLGLHDTARPVLPRPPTISNPPPLISSAKHPSVLERQIGAISQGMSVQLHVPYSEHAKAPVGPVTMGL
PLPMDPKKLAPFSGVKQEQLSPRGQAGPPESLGVPTAQEASVLRGTALGSVPGGSITKGIPSTRVPSDSAIT
YRGSITHGTPADVLYKGTITRIIGEDSPSRLDRGREDSLPKGHVIYEGKKGHVLSYEGGMSVTQCSKEDGRS
SSGPPHETAAPKRTYDMMEGRVGRAISSASIEGLMGRAIPPERHSPHHLKEQHHIRGSITQGIPRSYVEAQE
DYLRREAKLLKREGTPPPPPPSRDLTEAYKTQALGPLKLKPAHEGLVATVKEAGRSIHEIPREELRHTPELP
LAPRPLKEGSITQGTPLKYDTGASTTGSKKHDVRSLIGSPGRTFPPVHPLDVMADARALERACYEESLKSRP
GTASSSGGSIARGAPVIVPELGKPRQSPLTYEDHGAPFAGHLPRGSPVTMREPTPRLQEGSLSSSKASQDRK
```

LTSTPREIAKSPHSTVPEHHPHPISPYEHLLRGVSGVDLYRSHIPLAFDPTSIPRGIPLDAAAAYYLPRHLA
PNPTYPHLYPPYLIRGYPDTAALENRQTIINDYITSQQMHHNTATAMAQRADMLRGLSPRESSLALNYAAGP
RGIIDLSQVPHLPVLVPPTPGTPATAMDRLAYLPTAPQPFSSRHSSSPLSPGGPTHLTKPTTTSSSERERDR
DRERDRDREREKSILTSTTTVEHAPIWRPGTEQSSGSSGSSGGGGGSSSRPASHSHAHQHSPISPRTQDALQ
QRPSVLHNTGMKGIITAVEPSKPTVLRSTSTSSPVRPAATFPPATHCPLGGTLDGVYPTLMEPVLLPKEAPR
VARPERPRADTGHAFLAKPPARSGLEPASSPSKGSEPRPLV

Cloning

The gene was cloned into a modified pET9a vector (Novagen) digested by AscI and NotI, pESPRIT1 (Fig. 1) as an *Asc*I/*Not*I fragment.

ESPRIT Protocol

A bidirectional truncation was performed for generating an internal gene fragments library by sequential exonuclease III truncation. The protocol is essentially the same as in chapter 2, but after pooling clones from the first 5′ truncation, DNA was extracted by midiprep (Qiagen) and submitted to a second truncation reaction from 3′ end after first digesting with *Nsi*I/*Not*I.

The linearised plasmids with 150-750 nucleotide SMRT inserts were excised from 1% agarose gel prior to a second ligation step and recovered by transformation of MACH 1 cells (Invitrogen). *E. coli* strain BL21 AI (Invitrogen) cells containing the RIL plasmid (Stratagene) were transformed with the purified library. The colonies were printed onto nitrocellulose membranes. Colony blots of 27642 clones (in duplicates) were hybridised with Alexa 488 conjugated streptavidin (Invitrogen) and screened for expression and solubility by a fluoroimager. 96 best clones were then selected.

Expression and purification of 6 SMRT constructs

SMRT constructs were expressed in *E. coli* BL21 AI RIL in Terrific Broth medium (50 µg/ml kanamycin, 50 µg/ml chloramphenicol). Expression was induced with 13 mM L-arabinose at OD_{600nm} 0.4-0.6. Cells were harvested by centrifugation and resuspended in lysis buffer (50 mM Tris pH8.0, 300 mM KC1, 5mM β-ME, 100 µM $MgCl_2$, EDTA-free complete protease inhibitor cocktail (Roche Diagnostics, Indianapolis) and benzonase). Cells were lysed with a microfluidiser (M11O-L; Microfluidics Corporation, Newton, MA, USA) and the lysate was cleared by centrifugation (48000 g, 40 min, 4°C).

The supernatant was incubated overnight with Ni^{2+}-NTA agarose resin (Qiagen) at 4°C, loaded into a plastic column. Protein was eluted in elution buffer (50 mM Tris pH8.0, 300 mM KC1, 5mM β-ME, and 250 mM imidazole). The hexahistidine tag was removed by incubation with hexahistidine tagged tobacco etch virus protease (TEV) overnight at 4°C.

Finally the protein was purified by gel filtration (Superdex 75 column 16/60, GE Healthcare) in buffer (50 mM Tris pH8.0, 150 mM KCl, 1 mM DTT), dialysed (with HDAC7) overnight at 4°C in dialysis buffer (50 mM KH_2PO_4/K_2HP0_4 pH6.5, 150 mM KCl, 1 mM DTT) for NMR experiments, and concentrated by Amicon Ultra concentrator (Millipore).

NMR experiments

SMRT fragments were isotopically labelled by expression in minimal medium with $^{15}NH_4Cl$. All NMR spectra were recorded at the IBS, Grenoble by members of the Blackledge Group, at 25°C using 600 MHz or 800 MHz Varian spectrometers equipped with cryogenically cooled probes and analyzed using NMRPipe and Sparky software. The proteins, at 100 µM alone or 50 µM when incubated with HDAC7, were buffer-exchanged in a 50 mM potassium phosphate and 150 mM KCl pH 6.5 solution. Resonances of the amide groups were obtained by ^{1}H, ^{15}N-HSQC spectra on samples dissolved in 10% D_20 (v/v).

Fluorescence anisotropy

SMRT was labelled with fluorescein isothiocyanate (FITC, Sigma) to approximately 1:1 protein/dye molar ratio according to the manufacturer's protocol. Excess dye was removed by gel-filtration. Steady-state anisotropy of fluorescein-labelled SMRT alone and in the presence of increasing concentrations of HDAC7 was measured on QuantaMaster 40 spectrofluorimeter (PTI, USA) in L-configuration, with excitation at 490 nm and emission of >520 nm, in the cell at 25°C. The concentration of fluorescein-labelled SMRT in the reaction mixture was 1.5 µM. SMRT fragments were titrated against HDAC7 (RAQS-NNG construct) with concentrations between 1-20 µM in a first screen. The best construct 69D20 was precisely titrated with HDAC7 concentrations of 0 µM, 0.72 µM, 2.12 µM, 3.45 µM, 5.07 µM, 6.60 µM, 8.08 µM, 9.48 µM and 12.1 µM.
Affinity constant was determined from both Scatchard plot and non-linear fitting of steady-state anisotropy data according to HDAC7 concentrations, according to the modified Scatchard equation (Vuilleumier et al. 1999; Avilov et al. 2009):

$S=S_o+(S_t-S_o)x(1+Kx(P_t+n \times N_t)-((1+K \times (P_t+n \times N_t))\text{^}2^{-4} \times n \times N_t \times P_t \times K\text{^}2)\text{^}0.5)/(2 \times n \times K \times N_t)$

with P_t: total HDAC concentration added; S: steady-state anisotropy; N_t: total concentration of labelled SMRT (constant); n: number of HDAC molecules that can be

bound to one SMRT molecule (assumed that n=1); K: affinity constant, M^{-1}; S_o: initial anisotropy when no protein added and S_t: anisotropy at infinite HDAC concentration.

Surface plasmon resonance

SPR measurements were conducted at 760 nm in a Biacore X instrument (Biacore AB, Uppsala, Sweden) equipped with two flow cells. The flow cell temperature was set at 25°C in all experiments. The sample surfaces used were carboxy-methylated dextran CM5 chips (Biacore AB, Uppsala, Sweden). Coupling of ligands (SMRT fragments) to the carboxylic acid groups of the dextran hydrogel was carried out by conventional carbodiimide chemistry using 200 mM EDC (N-ethyl-N'-(3-diethylaminopropyl) carbodiimide) and 50 mM NHS (N-hydroxysuccinimide). The chip was activated for 7 min, followed by a 2 - 7 min ligand injection. Deactivation of the remaining active esters was performed by a 7 min injection of 1M ethanolamine/hydrochloride at pH 8.5. A flow rate of 10 µl/min was used during immobilisation. All ligands were diluted in 10 mM sodium acetate buffer pH 5.5, i.e., below the protein isoelectric point, thus enhancing the electrostatic interactions between the dextran matrix and the ligands. The contact time varied between 2 and 7 min resulting in levels of immobilisation between 5000 and 7000 RU (response units). One of the flow cells was used as a reference to monitor the response due to buffer changes and non-specific interactions. The reference cell was treated in the same way as the sample cell during the immobilisation procedure, but omitting the ligand immobilisation step.

Analyte (HDAC7 RAQS-NNG) was diluted in phosphate buffer (the same one as used for NMR, ie. 50 mM KH_2PO_4/K_2HPO_4 pH6.5, 150 mM KCl, 1 mM DTT) and injected at 250 nM, 375 nM, 562 nM, 625 nM and 1 µM at a flow rate of 20 µl/min. Obtained SPR curves were analyzed by BiaEVAL V4.1 (Biacore AB, Uppsala, Sweden).

Enzymology experiments.

The influence of SMRT fragments on HDAC7 deacetylase activity on two substrates, trifluoroacetyllysine and acetyllysine substrates (BPS Biosciences) were tested over 1.5 h of incubation at 37°C (one end point measured). 1 ng (8 nM) of HDAC7 were incubated with 20 µM of substrate in HDAC assay buffer provided with the assay kit. The reaction was stopped by a developer solution provided and further incubated for 25 min at room temperature (about 22 °C). The fluorescence intensity was detected with the spectrophotometer Perkin Elmer Victor^2. The fluorescence intensity was detected with

excitation-emission wavelengths of 355 nm and 460 nm, respectively and plotted as relative fluorescence units (RFU).

BIBLIOGRAPHY

Agalioti, T., Chen, G., and Thanos, D. 2002. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* **111**(3): 381-392.

Aharoni, A., Gaidukov, L., Yagur, S., Toker, L., Silman, I., and Tawfik, D.S. 2004. Directed evolution of mammalian paraoxonases PON1 and PON3 for bacterial expression and catalytic specialization. *Proc Natl Acad Sci U S A* **101**(2): 482-487.

Ahmad, K.F., Melnick, A., Lax, S., Bouchard, D., Liu, J., Kiang, C.L., Mayer, S., Takahashi, S., Licht, J.D., and Prive, G.G. 2003. Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain. *Mol Cell* **12**(6): 1551-1564.

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. 1964. Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A* **51**: 786-794.

Anderson, S. 1981. Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res* **9**(13): 3015-3027.

Andrianov, V., Gailite, V., Lola, D., Loza, E., Semenikhina, V., Kalvinsh, I., Finn, P., Petersen, K.D., Ritchie, J.W., Khan, N. et al. 2009. Novel amide derivatives as inhibitors of histone deacetylase: design, synthesis and SAR. *Eur J Med Chem* **44**(3): 1067-1085.

Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science* **181**(96): 223-230.

Angelini, A., Tosi, T., Mas, P., Acajjaoui, S., Zanotti, G., Terradot, L., and Hart, D.J. 2009. Expression of Helicobacter pylori CagA domains by library-based construct screening. *FEBS J* **276**(3): 816-824.

Arkin, M. 2005. Protein-protein interactions and cancer: small molecules going in for the kill. *Curr Opin Chem Biol* **9**(3): 317-324.

Arkin, M.R. and Wells, J.A. 2004. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* **3**(4): 301-317.

Avilov, S.V., Godet, J., Piemont, E., and Mely, Y. 2009. Site-specific characterization of HIV-1 nucleocapsid protein binding to oligonucleotides with two binding sites. *Biochemistry* **48**(11): 2422-2430.

Beckett, D., Kovaleva, E., and Schatz, P.J. 1999. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci* **8**(4): 921-929.

Berger, S.L. 2002. Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* **12**(2): 142-148.

Bieliauskas, A.V. and Pflum, M.K. 2008. Isoform-selective histone deacetylase inhibitors. *Chem Soc Rev* **37**(7): 1402-1413.

Blake, C.C., Koenig, D.F., Mair, G.A., North, A.C., Phillips, D.C., and Sarma, V.R. 1965. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* **206**(986): 757-761.

Boivin, A.J., Momparler, L.F., Hurtubise, A., and Momparler, R.L. 2002. Antineoplastic action of 5-aza-2'-deoxycytidine and phenylbutyrate on human lung carcinoma cells. *Anticancer Drugs* **13**(8): 869-874.

Bottomley, M.J. 2004. Structures of protein domains that create or recognize histone modifications. *EMBO Rep* **5**(5): 464-469.

Bottomley, M.J., Lo Surdo, P., Di Giovine, P., Cirillo, A., Scarpelli, R., Ferrigno, F., Jones, P., Neddermann, P., De Francesco, R., Steinkuhler, C. et al. 2008. Structural and functional analysis of the human HDAC4 catalytic domain reveals a regulatory structural zinc-binding domain. *J Biol Chem* **283**(39): 26694-26704.

Bouchain, G., Leit, S., Frechette, S., Khalil, E.A., Lavoie, R., Moradei, O., Woo, S.H., Fournel, M., Yan, P.T., Kalita, A. et al. 2003. Development of potential antitumor agents. Synthesis and biological evaluation of a new set of sulfonamide derivatives as histone deacetylase inhibitors. *J Med Chem* **46**(5): 820-830.

Boulias, K. and Talianidis, I. 2004. Functional role of G9a-induced histone methylation in small heterodimer partner-mediated transcriptional repression. *Nucleic Acids Res* **32**(20): 6096-6103.

Boyer, L.A., Latek, R.R., and Peterson, C.L. 2004. The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol* **5**(2): 158-163.

Breitkreutz, B.J., Stark, C., and Tyers, M. 2003. The GRID: the General Repository for Interaction Datasets. *Genome Biol* **4**(3): R23.

Briggs, S.D., Xiao, T., Sun, Z.W., Caldwell, J.A., Shabanowitz, J., Hunt, D.F., Allis, C.D., and Strahl, B.D. 2002. Gene silencing: trans-histone regulatory pathway in chromatin. *Nature* **418**(6897): 498.

Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. 1996. Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**(6): 843-851.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P. et al. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**(4): 581-592.

Chang, S., Bezprozvannaya, S., Li, S., and Olson, E.N. 2005. An expression screen reveals modulators of class II histone deacetylase phosphorylation. *Proc Natl Acad Sci U S A* **102**(23): 8120-8125.

Chang, S., Young, B.D., Li, S., Qi, X., Richardson, J.A., and Olson, E.N. 2006. Histone deacetylase 7 maintains vascular integrity by repressing matrix metalloproteinase 10. *Cell* **126**(2): 321-334.

Chapman-Smith, A. and Cronan, J.E., Jr. 1999. In vivo enzymatic protein biotinylation. *Biomol Eng* **16**(1-4): 119-125.

Chene, P., Fuchs, J., Bohn, J., Garcia-Echeverria, C., Furet, P., and Fabbro, D. 2000. A small synthetic peptide, which inhibits the p53-hdm2 interaction, stimulates the p53 pathway in tumour cell lines. *J Mol Biol* **299**(1): 245-253.

Cheng, Y., LeGall, T., Oldfield, C.J., Dunker, A.K., and Uversky, V.N. 2006a. Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry* **45**(35): 10448-10460.

Cheng, Y., LeGall, T., Oldfield, C.J., Mueller, J.P., Van, Y.Y., Romero, P., Cortese, M.S., Uversky, V.N., and Dunker, A.K. 2006b. Rational drug design via intrinsically disordered protein. *Trends Biotechnol* **24**(10): 435-442.

Cheung, P., Allis, C.D., and Sassone-Corsi, P. 2000. Signaling to chromatin through histone modifications. *Cell* **103**(2): 263-271.

Chiti, F. and Dobson, C.M. 2006. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**: 333-366.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. 2009. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**(5942): 834-840.

Clayton, A.L., Hazzalin, C.A., and Mahadevan, L.C. 2006. Enhanced histone acetylation and transcription: a dynamic perspective. *Mol Cell* **23**(3): 289-296.

Codina, A., Love, J.D., Li, Y., Lazar, M.A., Neuhaus, D., and Schwabe, J.W. 2005. Structural insights into the interaction and activation of histone deacetylase 3 by nuclear receptor corepressors. *Proc Natl Acad Sci U S A* **102**(17): 6009-6014.

Coeytaux, K. and Poupon, A. 2005. Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics* **21**(9): 1891-1900.

Cosgrove, M.S. 2007. Histone proteomics and the epigenetic regulation of nucleosome mobility. *Expert Rev Proteomics* **4**(4): 465-478.

Cosgrove, M.S., Boeke, J.D., and Wolberger, C. 2004. Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol* **11**(11): 1037-1043.

Crameri, A., Raillard, S.A., Bermudez, E., and Stemmer, W.P. 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**(6664): 288-291.

Crameri, A., Whitehorn, E.A., Tate, E., and Stemmer, W.P. 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* **14**(3): 315-319.

Crestani, M., Sadeghpour, A., Stroup, D., Galli, G., and Chiang, J.Y. 1998. Transcriptional activation of the cholesterol 7alpha-hydroxylase gene (CYP7A) by nuclear hormone receptors. *J Lipid Res* **39**(11): 2192-2200.

Csizmok, V., Bokor, M., Banki, P., Klement, E., Medzihradszky, K.F., Friedrich, P., Tompa, K., and Tompa, P. 2005. Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. *Biochemistry* **44**(10): 3955-3964.

Curtin, M.L., Garland, R.B., Heyman, H.R., Frey, R.R., Michaelides, M.R., Li, J., Pease, L.J., Glaser, K.B., Marcotte, P.A., and Davidsen, S.K. 2002. Succinimide hydroxamic acids as potent inhibitors of histone deacetylase (HDAC). *Bioorg Med Chem Lett* **12**(20): 2919-2923.

Dafforn, T.R. and Smith, C.J. 2004. Natively unfolded domains in endocytosis: hooks, lines and linkers. *EMBO Rep* **5**(11): 1046-1052.

de la Cruz, X., Lois, S., Sanchez-Molina, S., and Martinez-Balbas, M.A. 2005. Do protein motifs read the histone code? *Bioessays* **27**(2): 164-175.

de Ruijter, A.J., van Gennip, A.H., Caron, H.N., Kemp, S., and van Kuilenburg, A.B. 2003. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem J* **370**(Pt 3): 737-749.

Deininger, P.L. 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal Biochem* **129**(1): 216-223.

Dequiedt, F., Kasler, H., Fischle, W., Kiermer, V., Weinstein, M., Herndier, B.G., and Verdin, E. 2003. HDAC7, a thymus-specific class II histone deacetylase, regulates Nur77 transcription and TCR-mediated apoptosis. *Immunity* **18**(5): 687-698.

Dequiedt, F., Van Lint, J., Lecomte, E., Van Duppen, V., Seufferlein, T., Vandenheede, J.R., Wattiez, R., and Kettmann, R. 2005. Phosphorylation of histone deacetylase 7 by protein kinase D mediates T cell receptor-induced Nur77 expression and apoptosis. *J Exp Med* **201**(5): 793-804.

Dill, K.A. and Shortle, D. 1991. Denatured states of proteins. *Annu Rev Biochem* **60**: 795-825.

Dokmanovic, M. and Marks, P.A. 2005. Prospects: histone deacetylase inhibitors. *J Cell Biochem* **96**(2): 293-304.

Dokmanovic, M., Perez, G., Xu, W., Ngo, L., Clarke, C., Parmigiani, R.B., and Marks, P.A. 2007. Histone deacetylase inhibitors selectively suppress expression of HDAC7. *Mol Cancer Ther* **6**(9): 2525-2534.

Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T.J. 2003. Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J Mol Biol* **327**(1): 85-96.

Dosztanyi, Z., Csizmok, V., Tompa, P., and Simon, I. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**(16): 3433-3434.

Dover, J., Schneider, J., Tawiah-Boateng, M.A., Wood, A., Dean, K., Johnston, M., and Shilatifard, A. 2002. Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. *J Biol Chem* **277**(32): 28368-28371.

Downes, M., Ordentlich, P., Kao, H.Y., Alvarez, J.G., and Evans, R.M. 2000. Identification of a nuclear domain with deacetylase activity. *Proc Natl Acad Sci U S A* **97**(19): 10330-10335.

Dressel, U., Bailey, P.J., Wang, S.C., Downes, M., Evans, R.M., and Muscat, G.E. 2001. A dynamic role for HDAC7 in MEF2-mediated muscle differentiation. *J Biol Chem* **276**(20): 17007-17013.

Drews, J. 2000. Drug discovery: a historical perspective. *Science* **287**(5460): 1960-1964.

Duffy, S., Tsao, K.L., and Waugh, D.S. 1998. Site-specific, enzymatic biotinylation of recombinant proteins in Spodoptera frugiperda cells using biotin acceptor peptides. *Anal Biochem* **262**(2): 122-128.

Dunker, A.K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., and Villafranca, J.E. 1998. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput*: 473-484.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W. et al. 2001. Intrinsically disordered protein. *J Mol Graph Model* **19**(1): 26-59.

Dyson, H.J. and Wright, P.E. 2005. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**(3): 197-208.

Dyson, M.R., Perera, R.L., Shadbolt, S.P., Biderman, L., Bromek, K., Murzina, N.V., and McCafferty, J. 2008. Identification of soluble protein fragments by gene fragmentation and genetic selection. *Nucleic Acids Res* **36**(9): e51.

Eberharter, A. and Becker, P.B. 2002. Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep* **3**(3): 224-229.

Endo, A. 1992. The discovery and development of HMG-CoA reductase inhibitors. *J Lipid Res* **33**(11): 1569-1582.

Estiu, G., Greenberg, E., Harrison, C.B., Kwiatkowski, N.P., Mazitschek, R., Bradner, J.E., and Wiest, O. 2008. Structural origin of selectivity in class II-selective histone deacetylase inhibitors. *J Med Chem* **51**(10): 2898-2906.

Evans, P.R. and Owen, D.J. 2002. Endocytosis and vesicle trafficking. *Curr Opin Struct Biol* **12**(6): 814-821.

Fang, S., Miao, J., Xiang, L., Ponugoti, B., Treuter, E., and Kemper, J.K. 2007. Coordinated recruitment of histone methyltransferase G9a and other chromatin-modifying enzymes in SHP-mediated regulation of hepatic bile acid metabolism. *Mol Cell Biol* **27**(4): 1407-1424.

Felsenfeld, G. and Groudine, M. 2003. Controlling the double helix. *Nature* **421**(6921): 448-453.

Ficner, R. 2009. Novel structural insights into class I and II histone deacetylases. *Curr Top Med Chem* **9**(3): 235-240.

Finnin, M.S., Donigian, J.R., Cohen, A., Richon, V.M., Rifkind, R.A., Marks, P.A., Breslow, R., and Pavletich, N.P. 1999. Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* **401**(6749): 188-193.

Fischle, W., Dequiedt, F., Fillion, M., Hendzel, M.J., Voelter, W., and Verdin, E. 2001. Human HDAC7 histone deacetylase activity is associated with HDAC3 in vivo. *J Biol Chem* **276**(38): 35826-35835.

Fischle, W., Dequiedt, F., Hendzel, M.J., Guenther, M.G., Lazar, M.A., Voelter, W., and Verdin, E. 2002. Enzymatic activity associated with class II HDACs is dependent on a multiprotein complex containing HDAC3 and SMRT/N-CoR. *Mol Cell* **9**(1): 45-57.

Fischle, W., Wang, Y., and Allis, C.D. 2003a. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**(6957): 475-479.

Fischle, W., Wang, Y., and Allis, C.D. 2003b. Histone and chromatin cross-talk. *Curr Opin Cell Biol* **15**(2): 172-183.

Freitas, M.A., Sklenar, A.R., and Parthun, M.R. 2004. Application of mass spectrometry to the identification and quantification of histone post-translational modifications. *J Cell Biochem* **92**(4): 691-700.

Frey, R.R., Wada, C.K., Garland, R.B., Curtin, M.L., Michaelides, M.R., Li, J., Pease, L.J., Glaser, K.B., Marcotte, P.A., Bouska, J.J. et al. 2002. Trifluoromethyl ketones as inhibitors of histone deacetylase. *Bioorg Med Chem Lett* **12**(23): 3443-3447.

Fu, H., Subramanian, R.R., and Masters, S.C. 2000. 14-3-3 proteins: structure, function, and regulation. *Annu Rev Pharmacol Toxicol* **40**: 617-647.

Furumai, R., Komatsu, Y., Nishino, N., Khochbin, S., Yoshida, M., and Horinouchi, S. 2001. Potent histone deacetylase inhibitors built from trichostatin A and cyclic tetrapeptide antibiotics including trapoxin. *Proc Natl Acad Sci U S A* **98**(1): 87-92.

Furumai, R., Matsuyama, A., Kobashi, N., Lee, K.H., Nishiyama, M., Nakajima, H., Tanaka, A., Komatsu, Y., Nishino, N., Yoshida, M. et al. 2002. FK228 (depsipeptide) as a natural prodrug that inhibits class I histone deacetylases. *Cancer Res* **62**(17): 4916-4921.

Gallinari, P., Di Marco, S., Jones, P., Pallaoro, M., and Steinkuhler, C. 2007. HDACs, histone deacetylation and gene transcription: from molecular biology to cancer therapeutics. *Cell Res* **17**(3): 195-211.

Gao, C., Li, X., Lam, M., Liu, Y., Chakraborty, S., and Kao, H.Y. 2006. CRM1 mediates nuclear export of HDAC7 independently of HDAC7 phosphorylation and association with 14-3-3s. *FEBS Lett* **580**(21): 5096-5104.

Garbuzynskiy, S.O., Lobanov, M.Y., and Galzitskaya, O.V. 2004. To be folded or to be unfolded? *Protein Sci* **13**(11): 2871-2877.

Gilardi, F., Mitro, N., Godio, C., Scotti, E., Caruso, D., Crestani, M., and De Fabiani, E. 2007. The pharmacological exploitation of cholesterol 7alpha-hydroxylase, the key enzyme in bile acid synthesis: from binding resins to chromatin remodelling to reduce plasma cholesterol. *Pharmacol Ther* **116**(3): 449-472.

Glaser, K.B., Staver, M.J., Waring, J.F., Stender, J., Ulrich, R.G., and Davidsen, S.K. 2003. Gene expression profiling of multiple histone deacetylase (HDAC)

inhibitors: defining a common gene set produced by HDAC inhibition in T24 and MDA carcinoma cell lines. *Mol Cancer Ther* **2**(2): 151-163.

Glass, C.K. and Witztum, J.L. 2001. Atherosclerosis. the road ahead. *Cell* **104**(4): 503-516.

Glozak, M.A., Sengupta, N., Zhang, X., and Seto, E. 2005. Acetylation and deacetylation of non-histone proteins. *Gene* **363**: 15-23.

Glozak, M.A. and Seto, E. 2007. Histone deacetylases and cancer. *Oncogene* **26**(37): 5420-5432.

Godde, J.S. and Ura, K. 2008. Cracking the enigmatic linker histone code. *J Biochem* **143**(3): 287-293.

Goodman, R.H. and Smolik, S. 2000. CBP/p300 in cell growth, transformation, and development. *Genes Dev* **14**(13): 1553-1577.

Gottlicher, M., Minucci, S., Zhu, P., Kramer, O.H., Schimpf, A., Giavara, S., Sleeman, J.P., Lo Coco, F., Nervi, C., Pelicci, P.G. et al. 2001. Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *EMBO J* **20**(24): 6969-6978.

Grant, P.A. 2001. A tale of histone modifications. *Genome Biol* **2**(4): REVIEWS0003.

Grant, P.A., Eberharter, A., John, S., Cook, R.G., Turner, B.M., and Workman, J.L. 1999. Expanded lysine acetylation specificity of Gcn5 in native complexes. *J Biol Chem* **274**(9): 5895-5900.

Grant, P.A., Sterner, D.E., Duggan, L.J., Workman, J.L., and Berger, S.L. 1998. The SAGA unfolds: convergence of transcription regulators in chromatin-modifying complexes. *Trends Cell Biol* **8**(5): 193-197.

Gregoretti, I.V., Lee, Y.M., and Goodson, H.V. 2004. Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis. *J Mol Biol* **338**(1): 17-31.

Grewal, S.I. and Elgin, S.C. 2002. Heterochromatin: new possibilities for the inheritance of structure. *Curr Opin Genet Dev* **12**(2): 178-187.

Grozinger, C.M. and Schreiber, S.L. 2000. Regulation of histone deacetylase 4 and 5 and transcriptional activity by 14-3-3-dependent cellular localization. *Proc Natl Acad Sci U S A* **97**(14): 7835-7840.

Grozinger, C.M. and Schreiber, S.L. 2002. Deacetylase enzymes: biological functions and the use of small-molecule inhibitors. *Chem Biol* **9**(1): 3-16.

Guenther, M.G., Barak, O., and Lazar, M.A. 2001. The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Mol Cell Biol* **21**(18): 6091-6101.

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W., Ortin, J., Hart, D.J. et al. 2008. The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* **15**(5): 500-506.

Gunasekaran, K., Tsai, C.J., and Nussinov, R. 2004. Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. *J Mol Biol* **341**(5): 1327-1341.

Haggarty, S.J., Koeller, K.M., Wong, J.C., Butcher, R.A., and Schreiber, S.L. 2003. Multidimensional chemical genetic analysis of diversity-oriented synthesis-derived deacetylase inhibitors using cell-based assays. *Chem Biol* **10**(5): 383-396.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P. et al. 2004. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**(6995): 88-93.

Hart, D.J. and Tarendeau, F. 2006. Combinatorial library approaches for improving soluble protein expression in Escherichia coli. *Acta Crystallogr D Biol Crystallogr* **62**(Pt 1): 19-26.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**(3): 351-359.

Hernandez, M.A., Avila, J., and Andreu, J.M. 1986. Physicochemical characterization of the heat-stable microtubule-associated protein MAP2. *Eur J Biochem* **154**(1): 41-48.

Hodawadekar, S.C. and Marmorstein, R. 2007. Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene* **26**(37): 5528-5540.

Hoffmann, K., Brosch, G., Loidl, P., and Jung, M. 1999. A non-isotopic assay for histone deacetylase activity. *Nucleic Acids Res* **27**(9): 2057-2058.

Hoffmann, K., Soll, R.M., Beck-Sickinger, A.G., and Jung, M. 2001. Fluorescence-labeled octapeptides as substrates for histone deacetylase. *Bioconjug Chem* **12**(1): 51-55.

Huang, E.Y., Zhang, J., Miska, E.A., Guenther, M.G., Kouzarides, T., and Lazar, M.A. 2000. Nuclear receptor corepressors partner with class II histone deacetylases in a Sin3-independent repression pathway. *Genes Dev* **14**(1): 45-54.

Hubbert, C., Guardiola, A., Shao, R., Kawaguchi, Y., Ito, A., Nixon, A., Yoshida, M., Wang, X.F., and Yao, T.P. 2002. HDAC6 is a microtubule-associated deacetylase. *Nature* **417**(6887): 455-458.

Hutt, D.M., Herman, D., Rodrigues, A.P., Noel, S., Pilewski, J.M., Matteson, J., Hoch, B., Kellner, W., Kelly, J.W., Schmidt, A. et al. 2009. Reduced histone deacetylase 7 activity restores function to misfolded CFTR in cystic fibrosis. *Nat Chem Biol* **6**(1): 25-33.

Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradovic, Z., and Dunker, A.K. 2002. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* **323**(3): 573-584.

Ishida, T. and Kinoshita, K. 2008. Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* **24**(11): 1344-1348.

Javitt, N.B. 1994. Bile acid synthesis from cholesterol: regulatory and auxiliary pathways. *FASEB J* **8**(15): 1308-1311.

Javitt, N.B. 2002. Cholesterol, hydroxycholesterols, and bile acids. *Biochem Biophys Res Commun* **292**(5): 1147-1153.

Jenuwein, T. and Allis, C.D. 2001. Translating the histone code. *Science* **293**(5532): 1074-1080.

Jepsen, K. and Rosenfeld, M.G. 2002. Biological roles and mechanistic actions of co-repressor complexes. *J Cell Sci* **115**(Pt 4): 689-698.

Johnstone, R.W. 2002. Histone-deacetylase inhibitors: novel drugs for the treatment of cancer. *Nat Rev Drug Discov* **1**(4): 287-299.

Jones, P., Bottomley, M.J., Carfi, A., Cecchetti, O., Ferrigno, F., Lo Surdo, P., Ontoria, J.M., Rowley, M., Scarpelli, R., Schultz-Fademrecht, C. et al. 2008. 2-Trifluoroacetylthiophenes, a novel series of potent and selective class II histone deacetylase inhibitors. *Bioorg Med Chem Lett* **18**(11): 3456-3461.

Kalthoff, C. 2003. A novel strategy for the purification of recombinantly expressed unstructured protein domains. *J Chromatogr B Analyt Technol Biomed Life Sci* **786**(1-2): 247-254.

Kalthoff, C., Alves, J., Urbanke, C., Knorr, R., and Ungewickell, E.J. 2002. Unusual structural organization of the endocytic proteins AP180 and epsin 1. *J Biol Chem* **277**(10): 8209-8216.

Kanno, T., Kanno, Y., Siegel, R.M., Jang, M.K., Lenardo, M.J., and Ozato, K. 2004. Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. *Mol Cell* **13**(1): 33-43.

Kao, H.Y., Downes, M., Ordentlich, P., and Evans, R.M. 2000. Isolation of a novel histone deacetylase reveals that class I and class II deacetylases promote SMRT-mediated repression. *Genes Dev* **14**(1): 55-66.

Kao, H.Y., Verdel, A., Tsai, C.C., Simon, C., Juguilon, H., and Khochbin, S. 2001. Mechanism for nucleocytoplasmic shuttling of histone deacetylase 7. *J Biol Chem* **276**(50): 47496-47507.

Kelly, W.K., O'Connor, O.A., and Marks, P.A. 2002. Histone deacetylase inhibitors: from target to clinical trials. *Expert Opin Investig Drugs* **11**(12): 1695-1713.

Kemper, J.K., Kim, H., Miao, J., Bhalla, S., and Bae, Y. 2004. Role of an mSin3A-Swi/Snf chromatin remodeling complex in the feedback repression of bile acid biosynthesis by SHP. *Mol Cell Biol* **24**(17): 7707-7719.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**(4610): 662-666.

Kim, S.C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L. et al. 2006. Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* **23**(4): 607-618.

Kim, Y., Quartey, P., Li, H., Volkart, L., Hatzos, C., Chang, C., Nocek, B., Cuff, M., Osipiuk, J., Tan, K. et al. 2008. Large-scale evaluation of protein reductive methylation for improving protein crystallization. *Nat Methods* **5**(10): 853-854.

Kirsh, O., Seeler, J.S., Pichler, A., Gast, A., Muller, S., Miska, E., Mathieu, M., Harel-Bellan, A., Kouzarides, T., Melchior, F. et al. 2002. The SUMO E3 ligase RanBP2 promotes modification of the HDAC4 deacetylase. *EMBO J* **21**(11): 2682-2691.

Klein, C. and Vassilev, L.T. 2004. Targeting the p53-MDM2 interaction to treat cancer. *Br J Cancer* **91**(8): 1415-1419.

Koshland, D.E. 1958. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* **44**(2): 98-104.

Kouzarides, T. 1999. Histone acetylases and deacetylases in cell proliferation. *Curr Opin Genet Dev* **9**(1): 40-48.

Kouzarides, T. 2000. Acetylation: a regulatory modification to rival phosphorylation? *EMBO J* **19**(6): 1176-1179.

Kouzarides, T. 2007. Chromatin modifications and their function. *Cell* **128**(4): 693-705.

Kriwacki, R.W., Hengst, L., Tennant, L., Reed, S.I., and Wright, P.E. 1996. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A* **93**(21): 11504-11509.

Kuo, M.H. and Allis, C.D. 1998. Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* **20**(8): 615-626.

Kwiterovich, P.O., Jr., Fredrickson, D.S., and Levy, R.I. 1974. Familial hypercholesterolemia (one form of familial type II hyperlipoproteinemia). A study of its biochemical, genetic and clinical presentation in childhood. *J Clin Invest* **53**(5): 1237-1249.

Lagerstrom, M.C. and Schioth, H.B. 2008. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat Rev Drug Discov* **7**(4): 339-357.

Lagger, G., O'Carroll, D., Rembold, M., Khier, H., Tischler, J., Weitzer, G., Schuettengruber, B., Hauser, C., Brunmeir, R., Jenuwein, T. et al. 2002. Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *EMBO J* **21**(11): 2672-2681.

Lahm, A., Paolini, C., Pallaoro, M., Nardi, M.C., Jones, P., Neddermann, P., Sambucini, S., Bottomley, M.J., Lo Surdo, P., Carfi, A. et al. 2007. Unraveling the hidden catalytic activity of vertebrate class IIa histone deacetylases. *Proc Natl Acad Sci U S A* **104**(44): 17335-17340.

Lander, E.S. Linton, L.M. Birren, B. Nusbaum, C. Zody, M.C. Baldwin, J. Devon, K. Dewar, K. Doyle, M. FitzHugh, W. et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**(6822): 860-921.

Lehrmann, H., Pritchard, L.L., and Harel-Bellan, A. 2002. Histone acetyltransferases and deacetylases in the control of cell proliferation and differentiation. *Adv Cancer Res* **86**: 41-65.

Lemercier, C., Brocard, M.P., Puvion-Dutilleul, F., Kao, H.Y., Albagli, O., and Khochbin, S. 2002. Class II histone deacetylases are directly recruited by BCL6 transcriptional repressor. *J Biol Chem* **277**(24): 22045-22052.

Lemercier, C., Verdel, A., Galloo, B., Curtet, S., Brocard, M.P., and Khochbin, S. 2000. mHDA1/HDAC5 histone deacetylase interacts with and represses MEF2A transcriptional activity. *J Biol Chem* **275**(20): 15594-15599.

Levy, Y., Onuchic, J.N., and Wolynes, P.G. 2007. Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J Am Chem Soc* **129**(4): 738-739.

Li, X., Romero, P., Rani, M., Dunker, A.K., and Obradovic, Z. 1999. Predicting Protein Disorder for N-, C-, and Internal Regions. *Genome Inform Ser Workshop Genome Inform* **10**: 30-40.

Lieutaud, P., Canard, B., and Longhi, S. 2008. MeDor: a metaserver for predicting protein disorder. *BMC Genomics* **9 Suppl 2**: S25.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., and Russell, R.B. 2003a. Protein disorder prediction: implications for structural proteomics. *Structure* **11**(11): 1453-1459.

Linding, R., Russell, R.B., Neduva, V., and Gibson, T.J. 2003b. GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res* **31**(13): 3701-3708.

Liu, F., Dowling, M., Yang, X.J., and Kao, G.D. 2004. Caspase-mediated specific cleavage of human histone deacetylase 4. *J Biol Chem* **279**(33): 34537-34546.

Liu, J. and Rost, B. 2003. NORSp: Predictions of long regions without regular secondary structure. *Nucleic Acids Res* **31**(13): 3833-3835.

Lo Conte, L., Chothia, C., and Janin, J. 1999. The atomic structure of protein-protein recognition sites. *J Mol Biol* **285**(5): 2177-2198.

Lobley, A., Swindells, M.B., Orengo, C.A., and Jones, D.T. 2007. Inferring function using patterns of native disorder in proteins. *PLoS Comput Biol* **3**(8): e162.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**(6648): 251-260.

Luger, K. and Richmond, T.J. 1998. The histone tails of the nucleosome. *Curr Opin Genet Dev* **8**(2): 140-146.

Lusis, A.J. 2000. Atherosclerosis. *Nature* **407**(6801): 233-241.

Madauss, K.P., Grygielko, E.T., Deng, S.J., Sulpizio, A.C., Stanley, T.B., Wu, C., Short, S.A., Thompson, S.K., Stewart, E.L., Laping, N.J. et al. 2007. A structural and in

vitro characterization of asoprisnil: a selective progesterone receptor modulator. *Mol Endocrinol* **21**(5): 1066-1081.

Magliery, T.J. and Regan, L. 2004. Library approaches to biophysical problems. *Eur J Biochem* **271**(9): 1593-1594.

Marks, P., Rifkind, R.A., Richon, V.M., Breslow, R., Miller, T., and Kelly, W.K. 2001. Histone deacetylases and cancer: causes and therapies. *Nat Rev Cancer* **1**(3): 194-202.

Marmorstein, R. 2001a. Protein modules that manipulate histone tails for chromatin regulation. *Nat Rev Mol Cell Biol* **2**(6): 422-432.

Marmorstein, R. 2001b. Structure and function of histone acetyltransferases. *Cell Mol Life Sci* **58**(5-6): 693-703.

McKinsey, T.A., Zhang, C.L., Lu, J., and Olson, E.N. 2000a. Signal-dependent nuclear export of a histone deacetylase regulates muscle differentiation. *Nature* **408**(6808): 106-111.

McKinsey, T.A., Zhang, C.L., and Olson, E.N. 2000b. Activation of the myocyte enhancer factor-2 transcription factor by calcium/calmodulin-dependent protein kinase-stimulated binding of 14-3-3 to histone deacetylase 5. *Proc Natl Acad Sci U S A* **97**(26): 14400-14405.

McKinsey, T.A., Zhang, C.L., and Olson, E.N. 2001. Identification of a signal-responsive nuclear export sequence in class II histone deacetylases. *Mol Cell Biol* **21**(18): 6312-6321.

Mellor, J. 2006. Dynamic nucleosomes and gene transcription. *Trends Genet* **22**(6): 320-329.

Meszaros, B., Tompa, P., Simon, I., and Dosztanyi, Z. 2007. Molecular principles of the interactions of disordered proteins. *J Mol Biol* **372**(2): 549-561.

Miller, T.A., Witter, D.J., and Belvedere, S. 2003. Histone deacetylase inhibitors. *J Med Chem* **46**(24): 5097-5116.

Minucci, S. and Pelicci, P.G. 2006. Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat Rev Cancer* **6**(1): 38-51.

Miska, E.A., Langley, E., Wolf, D., Karlsson, C., Pines, J., and Kouzarides, T. 2001. Differential localization of HDAC4 orchestrates muscle differentiation. *Nucleic Acids Res* **29**(16): 3439-3447.

Mitro, N., Godio, C., De Fabiani, E., Scotti, E., Galmozzi, A., Gilardi, F., Caruso, D., Vigil Chacon, A.B., and Crestani, M. 2007. Insights in the regulation of cholesterol 7alpha-hydroxylase gene reveal a target for modulating bile acid synthesis. *Hepatology* **46**(3): 885-897.

Monneret, C. 2005. Histone deacetylase inhibitors. *Eur J Med Chem* **40**(1): 1-13.

Moriniere, J., Rousseaux, S., Steuerwald, U., Soler-Lopez, M., Curtet, S., Vitte, A.L., Govin, J., Gaucher, J., Sadoul, K., Hart, D.J. et al. 2009. Cooperative binding of

two acetylation marks on a histone tail by a single bromodomain. *Nature* **461**(7264): 664-668.

Nadal, M., Mas, P.J., Blanco, A.G., Arnan, C., Sola, M., Hart, D.J., and Coll, M. 2010. Structure and inhibition of herpesvirus DNA packaging terminase nuclease domain. *Proc Natl Acad Sci U S A* **107**(37): 16078-16083.

Nielsen, T.K., Hildmann, C., Dickmanns, A., Schwienhorst, A., and Ficner, R. 2005. Crystal structure of a bacterial class 2 histone deacetylase homologue. *J Mol Biol* **354**(1): 107-120.

Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C.D., Tempst, P., and Reinberg, D. 2002. Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes Dev* **16**(4): 479-489.

Nitta, M., Ku, S., Brown, C., Okamoto, A.Y., and Shan, B. 1999. CPF: an orphan nuclear receptor that regulates liver-specific expression of the human cholesterol 7alpha-hydroxylase gene. *Proc Natl Acad Sci U S A* **96**(12): 6660-6665.

Norris, K.L., Lee, J.Y., and Yao, T.P. 2009. Acetylation goes global: the emergence of acetylation biology. *Sci Signal* **2**(97): pe76.

Oefner, P.J., Hunicke-Smith, S.P., Chiang, L., Dietrich, F., Mulligan, J., and Davis, R.W. 1996. Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucleic Acids Res* **24**(20): 3879-3886.

Oger, F., Lecorgne, A., Sala, E., Nardese, V., Demay, F., Chevance, S., Desravines, D.C., Aleksandrova, N., Le Guevel, R., Lorenzi, S. et al. 2010. Biological and biophysical properties of the histone deacetylase inhibitor suberoylanilide hydroxamic acid are affected by the presence of short alkyl groups on the phenyl ring. *J Med Chem* **53**(5): 1937-1950.

Ordentlich, P., Downes, M., Xie, W., Genin, A., Spinner, N.B., and Evans, R.M. 1999. Unique forms of human and mouse nuclear receptor corepressor SMRT. *Proc Natl Acad Sci U S A* **96**(6): 2639-2644.

Ostermeier, M. and Lutz, S. 2003. The creation of ITCHY hybrid protein libraries. *Methods Mol Biol* **231**: 129-141.

Otto, F., Lubbert, M., and Stock, M. 2003. Upstream and downstream targets of RUNX proteins. *J Cell Biochem* **89**(1): 9-18.

Paroni, G., Mizzau, M., Henderson, C., Del Sal, G., Schneider, C., and Brancolini, C. 2004. Caspase-dependent regulation of histone deacetylase 4 nuclear-cytoplasmic shuttling promotes apoptosis. *Mol Biol Cell* **15**(6): 2804-2818.

Pauling, L. 1948. Nature of forces between large molecules of biological interest. *Nature* **161**(4097): 707-709.

Perutz, M.F. 1960. Structure of hemoglobin. *Brookhaven Symp Biol* **13**: 165-183.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C. 1960. Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature* **185**(4711): 416-422.

Phiel, C.J., Zhang, F., Huang, E.Y., Guenther, M.G., Lazar, M.A., and Klein, P.S. 2001. Histone deacetylase is a direct target of valproic acid, a potent anticonvulsant, mood stabilizer, and teratogen. *J Biol Chem* **276**(39): 36734-36741.

Polevoda, B. and Sherman, F. 2002. The diversity of acetylated proteins. *Genome Biol* **3**(5): reviews0006.

Pontius, B.W. 1993. Close encounters: why unstructured, polymeric domains can increase rates of specific macromolecular association. *Trends Biochem Sci* **18**(5): 181-186.

Pontius, B.W. and Berg, P. 1991. Rapid renaturation of complementary DNA strands mediated by cationic detergents: a role for high-probability binding domains in enhancing the kinetics of molecular assembly processes. *Proc Natl Acad Sci U S A* **88**(18): 8237-8241.

Prakash, S., Foster, B.J., Meyer, M., Wozniak, A., Heilbrun, L.K., Flaherty, L., Zalupski, M., Radulovic, L., Valdivieso, M., and LoRusso, P.M. 2001. Chronic oral administration of CI-994: a phase 1 study. *Invest New Drugs* **19**(1): 1-11.

Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I., and Sussman, J.L. 2005. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21**(16): 3435-3438.

Prodromou, C., Savva, R., and Driscoll, P.C. 2007. DNA fragmentation-based combinatorial approaches to soluble protein expression Part I. Generating DNA fragment libraries. *Drug Discov Today* **12**(21-22): 931-938.

Qiu, L., Kelso, M.J., Hansen, C., West, M.L., Fairlie, D.P., and Parsons, P.G. 1999. Anti-tumour activity in vitro and in vivo of selective differentiating agents containing hydroxamate. *Br J Cancer* **80**(8): 1252-1258.

Rascle, A., Johnston, J.A., and Amati, B. 2003. Deacetylase activity is required for recruitment of the basal transcription machinery and transactivation by STAT5. *Mol Cell Biol* **23**(12): 4162-4173.

Rawlings, A.E., Levdikov, V.M., Blagova, E., Colledge, V.L., Mas, P.J., Tunaley, J., Vavrova, L., Wilson, K.S., Barak, I., Hart, D.J. et al. 2010. Expression of soluble, active fragments of the morphogenetic protein SpoIIE from Bacillus subtilis using a library-based construct screen. *Protein Eng Des Sel* **23**(11): 817-825.

Rayment, I. 1997. Reductive alkylation of lysine residues to alter crystallization properties of proteins. *Methods Enzymol* **276**: 171-179.

Ren, Q. and Gorovsky, M.A. 2001. Histone H2A.Z acetylation modulates an essential charge patch. *Mol Cell* **7**(6): 1329-1335.

Richon, V.M., Emiliani, S., Verdin, E., Webb, Y., Breslow, R., Rifkind, R.A., and Marks, P.A. 1998. A class of hybrid polar inducers of transformed cell differentiation inhibits histone deacetylases. *Proc Natl Acad Sci U S A* **95**(6): 3003-3007.

Richon, V.M., Sandhoff, T.W., Rifkind, R.A., and Marks, P.A. 2000. Histone deacetylase inhibitor selectively induces p21WAF1 expression and gene-associated histone acetylation. *Proc Natl Acad Sci U S A* **97**(18): 10014-10019.

Romero, P., Obradovic, Z., Kissinger, C.R., Villafranca, J.E., Garner, E., Guilliot, S., and Dunker, A.K. 1998. Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomput*: 437-448.

Ross, R. 1999. Atherosclerosis--an inflammatory disease. *N Engl J Med* **340**(2): 115-126.

Russell, D.W. 2003. The enzymes, regulation, and genetics of bile acid synthesis. *Annu Rev Biochem* **72**: 137-174.

Russell, D.W. and Setchell, K.D. 1992. Bile acid biosynthesis. *Biochemistry* **31**(20): 4737-4749.

Rypniewski, W.R., Holden, H.M., and Rayment, I. 1993. Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: an X-ray analysis at 1.8-A resolution. *Biochemistry* **32**(37): 9851-9858.

Saito, A., Yamashita, T., Mariko, Y., Nosaka, Y., Tsuchiya, K., Ando, T., Suzuki, T., Tsuruo, T., and Nakanishi, O. 1999. A synthetic inhibitor of histone deacetylase, MS-27-275, with marked in vivo antitumor activity against human tumors. *Proc Natl Acad Sci U S A* **96**(8): 4592-4597.

Santos-Rosa, H., Bannister, A.J., Dehe, P.M., Geli, V., and Kouzarides, T. 2004. Methylation of H3 lysine 4 at euchromatin promotes Sir3p association with heterochromatin. *J Biol Chem* **279**(46): 47506-47512.

Saunders, L.R. and Verdin, E. 2007. Sirtuins: critical regulators at the crossroads between cancer and aging. *Oncogene* **26**(37): 5489-5504.

Savva, R., Prodromou, C., and Driscoll, P.C. 2007. DNA fragmentation based combinatorial approaches to soluble protein expression Part II: library expression, screening and scale-up. *Drug Discov Today* **12**(21-22): 939-947.

Schatz, P.J. 1993. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Biotechnology (N Y)* **11**(10): 1138-1143.

Scheele, U., Alves, J., Frank, R., Duwel, M., Kalthoff, C., and Ungewickell, E. 2003. Molecular and functional characterization of clathrin- and AP-2-binding determinants within a disordered domain of auxilin. *J Biol Chem* **278**(28): 25357-25368.

Scheich, C., Sievert, V., and Bussow, K. 2003. An automated method for high-throughput protein purification applied to a comparison of His-tag and GST-tag affinity chromatography. *BMC Biotechnol* **3**: 12.

Schmidt, T.G. and Skerra, A. 2007. The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat Protoc* **2**(6): 1528-1535.

Schreiber, S.L. and Bernstein, B.E. 2002. Signaling network model of chromatin. *Cell* **111**(6): 771-778.

Schuetz, A., Min, J., Allali-Hassani, A., Schapira, M., Shuen, M., Loppnau, P., Mazitschek, R., Kwiatkowski, N.P., Lewis, T.A., Maglathin, R.L. et al. 2008. Human HDAC7 harbors a class IIa histone deacetylase-specific zinc binding motif and cryptic deacetylase activity. *J Biol Chem* **283**(17): 11355-11363.

Schultz, B.E., Misialek, S., Wu, J., Tang, J., Conn, M.T., Tahilramani, R., and Wong, L. 2004. Kinetics and comparative reactivity of human class I and class IIb histone deacetylases. *Biochemistry* **43**(34): 11083-11091.

Sebolt-Leopold, J.S. and English, J.M. 2006. Mechanisms of drug inhibition of signalling molecules. *Nature* **441**(7092): 457-462.

Sengupta, N. and Seto, E. 2004. Regulation of histone deacetylase activities. *J Cell Biochem* **93**(1): 57-67.

Serrador, J.M., Cabrero, J.R., Sancho, D., Mittelbrunn, M., Urzainqui, A., and Sanchez-Madrid, F. 2004. HDAC6 deacetylase activity links the tubulin cytoskeleton with immune synapse organization. *Immunity* **20**(4): 417-428.

Sevcik, J., Skrabana, R., Dvorsky, R., Csokova, N., Iqbal, K., and Novak, M. 2007. X-ray structure of the PHF core C-terminus: insight into the folding of the intrinsically disordered protein tau in Alzheimer's disease. *FEBS Lett* **581**(30): 5872-5878.

Shahbazian, M.D. and Grunstein, M. 2007. Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem* **76**: 75-100.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**(7): 941-953.

Shoemaker, B.A., Portman, J.J., and Wolynes, P.G. 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* **97**(16): 8868-8873.

Sigler, P.B. 1988. Transcriptional activation. Acid blobs and negative noodles. *Nature* **333**(6170): 210-212.

Singh, S.B., Zink, D.L., Liesch, J.M., Mosley, R.T., Dombrowski, A.W., Bills, G.F., Darkin-Rattray, S.J., Schmatz, D.M., and Goetz, M.A. 2002. Structure and chemistry of apicidins, a class of novel cyclic tetrapeptides without a terminal alpha-keto epoxide as inhibitors of histone deacetylase with potent antiprotozoal activities. *J Org Chem* **67**(3): 815-825.

Smith, C.L. 2008. A shifting paradigm: histone deacetylases and transcriptional activation. *Bioessays* **30**(1): 15-24.

Somoza, J.R., Skene, R.J., Katz, B.A., Mol, C., Ho, J.D., Jennings, A.J., Luong, C., Arvai, A., Buggy, J.J., Chi, E. et al. 2004. Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases. *Structure* **12**(7): 1325-1334.

Soutoglou, E., Katrakili, N., and Talianidis, I. 2000. Acetylation regulates transcription factor activity at multiple levels. *Mol Cell* **5**(4): 745-751.

Spange, S., Wagner, T., Heinzel, T., and Kramer, O.H. 2009. Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *Int J Biochem Cell Biol* **41**(1): 185-198.

Sparrow, D.B., Miska, E.A., Langley, E., Reynaud-Deonauth, S., Kotecha, S., Towers, N., Spohr, G., Kouzarides, T., and Mohun, T.J. 1999. MEF-2 function is modified by a novel co-repressor, MITR. *EMBO J* **18**(18): 5085-5098.

Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**(Database issue): D535-539.

Sterner, D.E. and Berger, S.L. 2000. Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev* **64**(2): 435-459.

Strahl, B.D. and Allis, C.D. 2000. The language of covalent histone modifications. *Nature* **403**(6765): 41-45.

Stroup, D., Crestani, M., and Chiang, J.Y. 1997. Identification of a bile acid response element in the cholesterol 7 alpha-hydroxylase gene CYP7A. *Am J Physiol* **273**(2 Pt 1): G508-517.

Sun, Z.W. and Allis, C.D. 2002. Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* **418**(6893): 104-108.

Tarendeau, F., Boudet, J., Guilligay, D., Mas, P.J., Bougault, C.M., Boulo, S., Baudin, F., Ruigrok, R.W., Daigle, N., Ellenberg, J. et al. 2007. Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat Struct Mol Biol* **14**(3): 229-233.

Tarendeau, F., Crepin, T., Guilligay, D., Ruigrok, R.W., Cusack, S., and Hart, D.J. 2008. Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog* **4**(8): e1000136.

Taunton, J., Hassig, C.A., and Schreiber, S.L. 1996. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**(5260): 408-411.

Tessier, P., Smil, D.V., Wahhab, A., Leit, S., Rahil, J., Li, Z., Deziel, R., and Besterman, J.M. 2009. Diphenylmethylene hydroxamic acids as selective class IIa histone deacetylase inhibitors. *Bioorg Med Chem Lett* **19**(19): 5684-5688.

Thiagalingam, S., Cheng, K.H., Lee, H.J., Mineva, N., Thiagalingam, A., and Ponte, J.F. 2003. Histone deacetylases: unique players in shaping the epigenetic histone code. *Ann N Y Acad Sci* **983**: 84-100.

Thorstenson, Y.R., Hunicke-Smith, S.P., Oefner, P.J., and Davis, R.W. 1998. An automated hydrodynamic process for controlled, unbiased DNA shearing. *Genome Res* **8**(8): 848-855.

Tobin, M., Affholter, J.A., Stemmer, W.P., and Minshull, J. 1999. Colorless green ideas. *Nat Biotechnol* **17**(4): 333-334.

Tompa, P. 2002. Intrinsically unstructured proteins. *Trends Biochem Sci* **27**(10): 527-533.

Tompa, P., Dosztanyi, Z., and Simon, I. 2006. Prevalent structural disorder in E. coli and S. cerevisiae proteomes. *J Proteome Res* **5**(8): 1996-2000.

Turner, B.M. 2000. Histone acetylation and an epigenetic code. *Bioessays* **22**(9): 836-845.

Turner, B.M. 2002. Cellular memory and the histone code. *Cell* **111**(3): 285-291.

Uversky, V.N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **11**(4): 739-756.

Uversky, V.N., Gillespie, J.R., and Fink, A.L. 2000. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41**(3): 415-427.

Uversky, V.N., Oldfield, C.J., and Dunker, A.K. 2008. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* **37**: 215-246.

Vacic, V., Oldfield, C.J., Mohan, A., Radivojac, P., Cortese, M.S., Uversky, V.N., and Dunker, A.K. 2007. Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res* **6**(6): 2351-2366.

Vannini, A., Volpari, C., Filocamo, G., Casavola, E.C., Brunetti, M., Renzoni, D., Chakravarty, P., Paolini, C., De Francesco, R., Gallinari, P. et al. 2004. Crystal structure of a eukaryotic zinc-dependent histone deacetylase, human HDAC8, complexed with a hydroxamic acid inhibitor. *Proc Natl Acad Sci U S A* **101**(42): 15064-15069.

Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C. et al. 2004. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**(5659): 844-848.

Vega, R.B., Harrison, B.C., Meadows, E., Roberts, C.R., Papst, P.J., Olson, E.N., and McKinsey, T.A. 2004. Protein kinases C and D mediate agonist-dependent cardiac hypertrophy through nuclear export of histone deacetylase 5. *Mol Cell Biol* **24**(19): 8374-8385.

Verdin, E., Dequiedt, F., and Kasler, H.G. 2003. Class II histone deacetylases: versatile regulators. *Trends Genet* **19**(5): 286-293.

Viatour, P., Legrand-Poels, S., van Lint, C., Warnier, M., Merville, M.P., Gielen, J., Piette, J., Bours, V., and Chariot, A. 2003. Cytoplasmic IkappaBalpha increases NF-kappaB-independent transcription through binding to histone deacetylase (HDAC) 1 and HDAC3. *J Biol Chem* **278**(47): 46541-46548.

Vidali, G., Gershey, E.L., and Allfrey, V.G. 1968. Chemical studies of histone acetylation. The distribution of epsilon-N-acetyllysine in calf thymus histones. *J Biol Chem* **243**(24): 6361-6366.

Vogelauer, M., Wu, J., Suka, N., and Grunstein, M. 2000. Global histone acetylation and deacetylation in yeast. *Nature* **408**(6811): 495-498.

Vuilleumier, C., Bombarda, E., Morellet, N., Gerard, D., Roques, B.P., and Mely, Y. 1999. Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study. *Biochemistry* **38**(51): 16816-16825.

Walter, T.S., Meier, C., Assenberg, R., Au, K.F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., and Grimes, J.M. 2006. Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**(11): 1617-1622.

Wang, A.H., Kruhlak, M.J., Wu, J., Bertos, N.R., Vezmar, M., Posner, B.I., Bazett-Jones, D.P., and Yang, X.J. 2000. Regulation of histone deacetylase 4 by binding of 14-3-3 proteins. *Mol Cell Biol* **20**(18): 6904-6912.

Wang, H., Chen, J., Hollister, K., Sowers, L.C., and Forman, B.M. 1999. Endogenous bile acids are ligands for the nuclear receptor FXR/BAR. *Mol Cell* **3**(5): 543-553.

Wang, L., Zuercher, W.J., Consler, T.G., Lambert, M.H., Miller, A.B., Orband-Miller, L.A., McKee, D.D., Willson, T.M., and Nolte, R.T. 2006. X-ray crystal structures of the estrogen-related receptor-gamma ligand binding domain in three functional states reveal the molecular basis of small molecule regulation. *J Biol Chem* **281**(49): 37773-37781.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**(3): 635-645.

Wasylyk, C., Salvi, R., Argentini, M., Dureuil, C., Delumeau, I., Abecassis, J., Debussche, L., and Wasylyk, B. 1999. p53 mediated death of cells overexpressing MDM2 by an inhibitor of MDM2 interaction with p53. *Oncogene* **18**(11): 1921-1934.

Waterston, R.H., Lander, E.S., and Sulston, J.E. 2002a. On the sequencing of the human genome. *Proc Natl Acad Sci U S A* **99**(6): 3712-3716.

Waterston, R.H. Lindblad-Toh, K. Birney, E. Rogers, J. Abril, J.F. Agarwal, P. Agarwala, R. Ainscough, R. Alexandersson, M. An, P. et al. 2002b. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**(6915): 520-562.

Wen, Y.D., Perissi, V., Staszewski, L.M., Yang, W.M., Krones, A., Glass, C.K., Rosenfeld, M.G., and Seto, E. 2000. The histone deacetylase-3 complex contains nuclear receptor corepressors. *Proc Natl Acad Sci U S A* **97**(13): 7202-7207.

Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Uversky, V.N., and Obradovic, Z. 2007. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J Proteome Res* **6**(5): 1882-1898.

Xu, H.E., Stanley, T.B., Montana, V.G., Lambert, M.H., Shearer, B.G., Cobb, J.E., McKee, D.D., Galardi, C.M., Plunket, K.D., Nolte, R.T. et al. 2002. Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha. *Nature* **415**(6873): 813-817.

Yang, W.M., Tsai, S.C., Wen, Y.D., Fejer, G., and Seto, E. 2002. Functional domains of histone deacetylase-3. *J Biol Chem* **277**(11): 9447-9454.

Yang, Z.R., Thomson, R., McNeil, P., and Esnouf, R.M. 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**(16): 3369-3376.

Yoshida, M., Kijima, M., Akita, M., and Beppu, T. 1990. Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *J Biol Chem* **265**(28): 17174-17179.

Yumerefendi, H., Tarendeau, F., Mas, P.J., and Hart, D.J. 2010. ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets. *J Struct Biol* **172**(1): 66-74.

Zhang, C.L., McKinsey, T.A., and Olson, E.N. 2002a. Association of class II histone deacetylases with heterochromatin protein 1: potential role for histone methylation in control of muscle differentiation. *Mol Cell Biol* **22**(20): 7302-7312.

Zhang, J., Kalkum, M., Chait, B.T., and Roeder, R.G. 2002b. The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2. *Mol Cell* **9**(3): 611-623.

Zhang, J., Sprung, R., Pei, J., Tan, X., Kim, S., Zhu, H., Liu, C.F., Grishin, N.V., and Zhao, Y. 2009. Lysine acetylation is a highly abundant and evolutionarily conserved modification in Escherichia coli. *Mol Cell Proteomics* **8**(2): 215-225.

Zhang, Y. and Reinberg, D. 2001. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev* **15**(18): 2343-2360.

Zhao, X., Ito, A., Kane, C.D., Liao, T.S., Bolger, T.A., Lemrow, S.M., Means, A.R., and Yao, T.P. 2001. The modular nature of histone deacetylase HDAC4 confers phosphorylation-dependent intracellular trafficking. *J Biol Chem* **276**(37): 35042-35048.

RESUME :

L'athérosclérose consiste en l'accumulation de cholestérol dans les artères causant la formation de plaques rigides pouvant causer attaques et malaises cardiaques. Il existe des molécules réduisant les taux de cholestérol dans le sang, mais sont inefficaces pour les patients souffrant d'hypercholestérolémie familiale. Il est donc crucial de développer de nouvelles molécules. L'inhibition de HDAC7 influence la conversion du cholestérol en acides biliaires en augmentant l'activité de la 7α-cholestérol hydroxylase, enzyme limitante du catabolisme du cholestérol. Ainsi cibler HDAC7 pourrait constituer une nouvelle stratégie pour réduire les taux de cholestérol. Le projet s'est articulé autour de deux principaux axes : comprendre et améliorer la spécificité des inhibiteurs de HDAC7 et explorer l'interaction SMRT-HDAC7 comme stratégie alternative d'inhibition de HDAC7. Des étapes successives d'ingénierie de HDAC7 ont été menées à l'aide de la technologie ESPRIT. Des essais de cristallisation et des tests d'inhibition avec un composé prometteur ont été effectués. Ce composé est un dérivé de SAHA, pan-inhibiteur des HDACs, modifié par substitution en position méta de son groupe phényle. Des essais de cristallisation sont en cours. ESPRIT a également été appliquée à la protéine SMRT. Les fragments solubles de SMRT ont été testés pour interaction avec HDAC7 par ^{15}N-HSQC RMN. Un fragment de SMRT a été identifié comme interagissant avec HDAC7. Cette interaction a été confirmée par anisotropie de fluorescence et résonance plasmonique de surface. Des expériences de RMN complémentaires sont en cours pour attribuer les résidus impliqués dans l'interaction.

ABSTRACT :

Atherosclerosis occurs due to accumulation of cholesterol in the arteries leading to rigid plaque formation leading eventually to strokes and heart attacks. Several efficient drugs exist to decrease blood cholesterol levels, but are inefficient for treating familial hypercholesterolemia. Therefore new drugs must be developed. It was observed that inhibiting HDAC7 increased conversion of cholesterol into excreted bile acids by enhancing the activity of 7α-cholesterol hydroxylase, the rate-limiting enzyme in cholesterol catabolism. This represents a potential new strategy for decreasing cholesterol levels. The project was divided into two main parts: to understand and elucidate mechanisms underlying HDAC inhibitor isoform-specificity and to explore HDAC7-SMRT interaction as an alternative strategy of HDAC7 inhibition. Several protein engineering steps were carried out on HDAC7, using the ESPRIT technology. Crystallisation trials and inhibition tests were performed with a promising compound. This compound was derived from the pan-HDAC inhibitor SAHA, which was substituted in the meta positions of its phenyl cap. Crystallisation trails are ongoing with this compound. ESPRIT was applied as well on SMRT. The generated soluble fragments were tested for their interaction with HDAC7 by ^{15}N-HSQC NMR. A interacting fragment was identified. This interaction was confirmed by fluorescence anisotropy and surface Plasmon resonance. NMR experiments are ongoing to assign residues involved in the interaction.