

**ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ**

**Centre de Biologie Intégrative, IGBMC, UMR7104, Illkirch**

**THÈSE** présentée par :

**Tajith Baba SHAIK**

soutenue le : **22 septembre 2017**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Biophysique et biologie structurale

(Biophysics and Structural biology)

**ETUDE BIOCHIMIQUE, BIOPHYSIQUE  
ET STRUCTURALE DU MECANISME  
D'ACTION ET DE L'INHIBITION  
SELECTIVE DE L'HISTONE  
DESACETYLASE HDAC8**

**THÈSE dirigée par :**

**Dr. ROMIER Christophe**

Directeur de Recherche CNRS, IGBMC, Illkirch

**RAPPORTEURS :**

**Prof. EINSLE Oliver**

Professeur, Albert-Ludwigs-Universität Freiburg,  
Freiburg, Allemagne

**Dr. FRIBOURG Sébastien**

Directeur de Recherche INSERM, IECB, Bordeaux

**AUTRES MEMBRES DU JURY :**

**Dr. SCHUSTER Catherine**

Directeur de Recherche INSERM, INSERM Unité 1110, Strasbourg

---





## Acknowledgements

First, I would like to express my sincere gratitude to my supervisor Dr. Christophe Romier for giving me this opportunity to pursue PhD in his lab. Thank you, Christophe, for giving me this fascinating subject and for being supportive from the day one. I specially thank you for being so patient with me for the past four years which have been a great experience for a great learning experience.

I take this opportunity to thank all the jury members, Dr. Catherine Schuster, Prof. Dr. Oliver Einsle, and Dr. Sébastien Fribourg who enthusiastically agreed to evaluate my work.

I would like to thank all our A-ParaDDisE collaborators, specially Dr. Raymond Pierce, Prof. Dr. Wolfgang Sippl and Prof. Dr. Manfred Jung for providing great scientific platform.

I would like to thank all the former and present members of the lab. A special thanks to Dr. Martin Marek for being a wonderful co-worker and for active discussions. I would like to thank Dr. Regis Back for such a wonderful company. I would like to thank Dr. Pierre Antony for giving me the valuable suggestions during my thesis writing.

A special thanks to Nata and Pernelle for the wonderful company in the lab. A single sentence is not enough to thank you guys, we can discuss about this in the lab. Really thank you for the nice working environment in the lab. Ah ha! A special thanks to Elizabeth for introducing adventurous Peru style of work.

A very special gratitude goes out to all the platform services, Structural biology department and administrative staff of IGBMC for helping me with lot of administrative stuff. With a special mention to Alastair, who has assisted me in collecting X-ray data at synchrotron. I am also grateful to all my friends at IGBMC, Kareem, Anna, Pau, Christophe, Moamen, and all the colleagues at CBI for providing a nice working environment.

And a special mention to Naima and Rana. Thank you for your support and for the nice food. I am very glad about all the Indians at the institute for the good company specially, Nisha, Vasugi, AK, ashique, and all others. Thank you, guys.

This list never ends without mentioning all the cricketing community of Illkirch. Especially Waseem, it's been wonderful to play cricket with you guys.

Last but by no means least, I would like to thank my family members. My dad, mom, brother for their love and encouragement. Thanks mom for living my passion as your dream.

# Résumé de thèse

## Introduction

L'ADN eucaryote est compacté sous forme d'une structure dynamique connue sous le nom de chromatine et dont la sous-unité minimale est le nucléosome (147 paires de bases d'ADN enroulées par les protéines histone). La structure de la chromatine module l'accessibilité à l'information génétique et régule ainsi les autres processus nucléaires. Fait important, la structure de la chromatine peut être modifiée par les mécanismes épigénétiques, permettant la régulation fine des processus nucléaires. En conséquence, les mécanismes épigénétiques ont un impact majeur sur l'homéostasie cellulaire, le type cellulaire et le développement. La dérégulation des mécanismes épigénétiques est de ce fait impliquée dans de nombreuses maladies telles que le cancer, mais aussi les maladies neurologiques et inflammatoires. Il est donc essentiel de comprendre les bases fondamentales des mécanismes épigénétiques. Mais les effecteurs épigénétiques représentent également des cibles importantes pour les interventions thérapeutiques dans de nombreuses maladies.

Parmi les mécanismes épigénétiques, « l'écriture » et « l'effacement », par des enzymes épigénétiques dédiées, des marques épigénétiques covalentes sur les histones et sur d'autres effecteurs nucléaires ont un effet majeur sur la structure de la chromatine et sa modification en réponse à divers stimuli. Plus précisément, l'addition covalente réversible de petits groupes chimiques (par exemple acétyle, méthyle, phosphate) ou macromolécules (par exemple ubiquitine, SUMO) à des résidus spécifiques des histones peut faciliter ou empêcher l'accessibilité de l'ADN par les mécanismes nucléaires.

L'acétylation est l'une des principales marques épigénétiques et est souvent associée à une structure de chromatine permissive aux processus nucléaires. L'acétylation des histones et d'autres effecteurs nucléaires est contrôlée par l'action opposée des HAT (transférases d'acétyle d'histone – histone acetyl transferases) et HDAC (desacétylases d'histone – histone deacetylases). Fait intéressant, les enzymes HDAC représentent les principales cibles des médicaments épigénétiques (épimédicaments) actuellement approuvés. Ces épimédicaments sont cependant restreints au traitement de quelques cancers car ils inhibent presque tous les membres de la famille HDAC qui ont des fonctions très différentes chez l'homme. Par conséquent, un axe de recherche majeur dans la découverte d'épimédicaments est de développer de nouvelles molécules qui montrent une sélectivité élevée pour un membre précis de la famille HDAC.

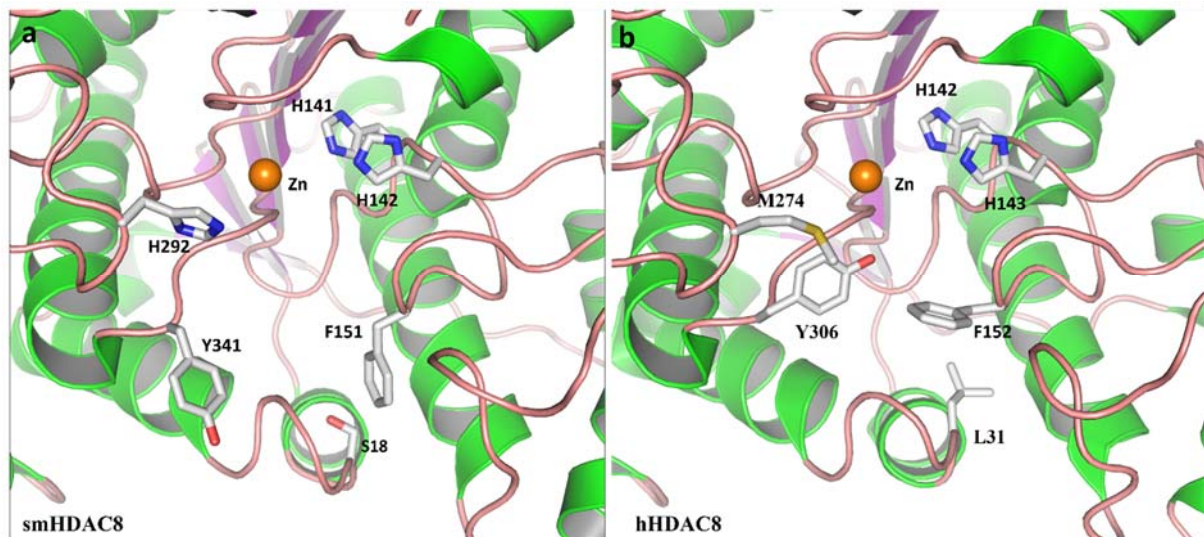
Il est important de noter que, malgré des années d'études des HDAC, la façon dont les différents membres de cette famille reconnaissent spécifiquement leurs cibles et comment ils

agissent sur ces cibles reste très obscurs. En outre, les HDAC font souvent partie de complexes macromoléculaires, et d'autres sous-unités de ces complexes peuvent affecter leur activité, la reconnaissance de leurs cibles, mais aussi l'inhibition par les épimédicaments. Il est donc essentiel de combiner à la fois des recherches fondamentales et applicatives sur les HDACs pour développer des épimédicaments plus puissants et plus sélectifs qui puissent être utilisés plus largement pour traiter un plus grand panel de maladies.

L'importance des mécanismes épigénétiques n'est pas limitée à l'homme, mais aussi à tous les organismes eucaryotes. Cela ouvre la porte à la lutte contre les parasites eucaryotes qui provoquent des millions de décès chaque année dans le monde entier. De tels parasites (par exemple le plasmodium, les trypanosomes, les schistosomes) ont souvent des cycles de vie très complexes où les mécanismes épigénétiques sont censés jouer des rôles essentiels. D'ailleurs, le traitement de parasites avec des épimédicaments approuvés ciblant les HDAC a révélé une forte sensibilité des parasites à ces épimédicaments.

Cela ouvre clairement la voie au développement de nouveaux médicaments antiparasitaires pour lutter contre des maladies pour lesquelles il n'existe que très peu de médicaments contre lesquels les phénomènes de résistance augmentent. Étant donné que le processus de découverte de médicaments nécessite beaucoup de temps et coûte cher, une stratégie de «portage» consistant à modifier les médicaments actuellement approuvés pour les rendre plus puissants contre les parasites devrait accélérer la recherche de nouveaux médicaments antiparasitaires. Encore une fois, un important goulet d'étranglement est de modifier les médicaments initiaux afin qu'ils puissent être sélectifs pour les enzymes parasitaires, mais plus pour les enzymes humaines.

Au cours du projet SETReND (*Schistosoma* Epigenetics: Targets, Regulation, New Drugs - 2010-2012) financé par l'Union Européenne, mon laboratoire a fourni la preuve de concept de cette stratégie de portage épigénétique en ciblant notamment HDAC8 du parasite *Schistosoma mansoni* (smHDAC8). En résolvant la structure de smHDAC8 sous forme apo et en complexe avec des inhibiteurs non sélectifs de HDACs, y compris le médicament déjà approuvé Vorinostat, l'équipe a fourni des informations inestimables et inattendues qui ont été utilisées pour trouver des premiers inhibiteurs présentant une bonne sélectivité mais une puissance moyenne. Les structures de smHDAC8 avec ces molécules ont par la suite été utilisées pour développer une nouvelle série d'inhibiteurs à forte puissance (dans la gamme du nM) et de sélectivité forte à excellente (*Fig i*).



**Fig i: Comparaison entre les poches du site actif smHDAC8 et hHDAC8 :**

Représentation des structures cristallographiques de smHDAC8 (a) et hHDAC8 (b). Les acides aminés importants de la poche catalytique sont représentés sous forme de bâtons et sont numérotés respectivement. Le zinc catalytique est représenté par une sphère orange. Deux différences structurales sont notées au niveau de la poche catalytique, la présence de H292 contre M274 et la conformation de la phénylalanine dans les structures flipped-in vs flipped-out dans smHDAC8 et hHDAC8 respectivement. Ces différences de poche catalytique sont importantes dans la conception des inhibiteurs sélectifs de SmHDAC8.

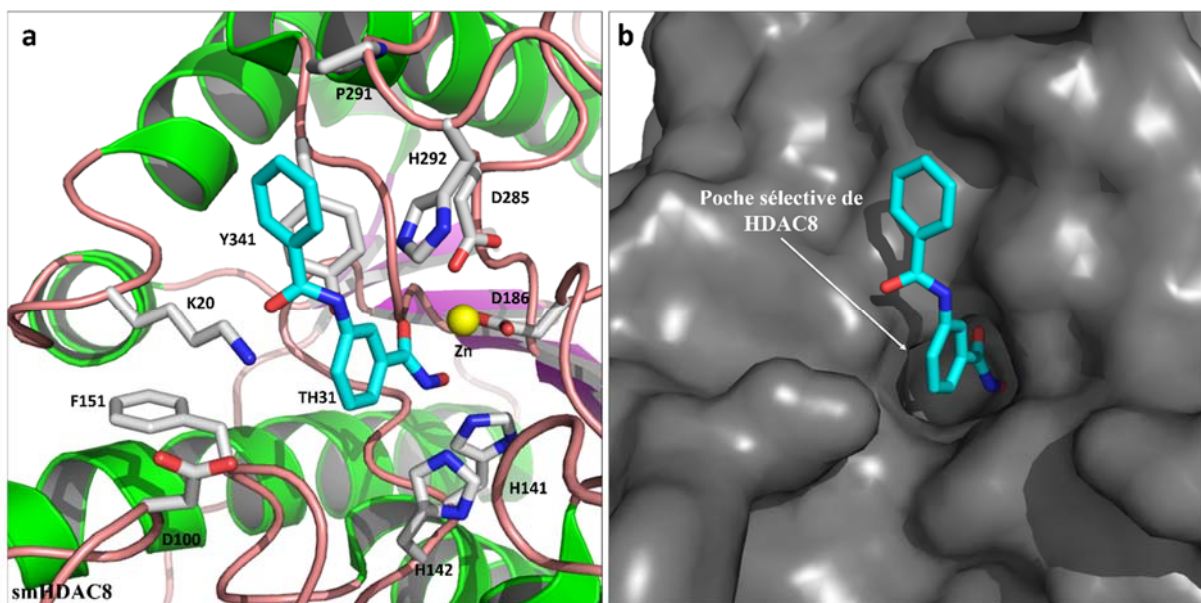
Au cours de ma thèse et dans le cadre du plus grand projet financé par l'Union Européenne, A-ParaDDisE (Anti-Parasitic Drug Discovery in Epigenetics - 2014-2017), j'ai travaillé sur la caractérisation structurale des complexes entre smHDAC8 et cette nouvelle série d'inhibiteurs pour mieux comprendre les bases moléculaires de leur plus grande puissance et de leur plus grande sélectivité. J'ai complété ce travail par l'analyse biochimique, biophysique et structurale des complexes entre smHDAC8, mais aussi la HDAC8 humaine (hHDAC8), avec des inhibiteurs hautement sélectifs de hHDAC8. Ce travail m'a amené à étudier par analyse mutationnelle l'importance de l'architecture du site actif de HDAC8 pour l'activité et l'inhibition de cette enzyme. Enfin, dans une dernière partie de ma thèse, j'ai étudié des aspects plus fondamentaux de la biologie de HDAC8, en considérant notamment son interaction avec l'une de ses principales cibles, le complexe Cohésine.

## Résultats

Au cours de ma thèse, j'ai résolu des dizaines de structures de smHDAC8 complexée avec divers inhibiteurs de nos collaborateurs. La structure de smHDAC8 avec la molécule la plus simple de cette série, TH31, a révélé un nouveau mécanisme de fixation de ces inhibiteurs sur smHDAC8 (*Fig ii*). Plus précisément, en plus de la coordination connue du

zinc catalytique de HDAC8 par un groupement hydroxamate, la liaison de TH31 dans la poche du site actif est stabilisée par l'interaction de son amide interne avec deux résidus de smHDAC8, K20 et H292. La liaison est encore plus stabilisée par des contacts hydrophobes établis entre la coiffe de TH31 et la boucle L6 de smHDAC8.

Cette liaison permet d'obtenir une sélectivité plus grande pour smHDAC8 par rapport aux autres HDAC humaines. Seule hHDAC8 est encore fortement inhibé par TH31. Par conséquent, nous avons sélectionné plusieurs autres inhibiteurs de cette série qui ont montré une puissance accrue pour smHDAC8, mais aussi une sélectivité accrue pour le smHDAC8 par rapport à hHDAC8. Ces inhibiteurs possèdent des groupes substituants supplémentaires ainsi que différentes coiffes par rapport à TH31. La structure de smHDAC8 avec ces différents inhibiteurs a fourni une foule d'informations sur la façon dont la puissance et la sélectivité peuvent être obtenues en améliorant et en étendant les interactions observées initialement entre TH31 et smHDAC8.



**Fig ii Structure cristallographique de smHDAC8-TH31 :**

a) Représentation en ruban de la structure cristallographique de smHDAC8-TH31. TH31 (représenté par des bâtons de couleur cyan) est en coordination avec le zinc catalytique (sphère jaune) et d'autres résidus (représentés sous forme de bâtons et numérotés respectivement). K20 et H292 contiennent l'inhibiteur TH31 dans la poche du site actif de smHDAC8. b) Représentation structurale cristallographique de smHDAC8-TH31 montrant la poche sélective HDAC8.

La HDAC8 humaine est l'une des HDACs pour lesquelles des inhibiteurs hautement sélectifs (par exemple PCI-34051) ont été trouvés. Mais les bases moléculaires de cette

sélectivité restent inconnues. J'ai étudié la base moléculaire de la sélectivité de HDAC8 par ces inhibiteurs par détermination de la structure de hHDAC8 et de smHDAC8 avec plusieurs de ces inhibiteurs. Ces structures ont montré que ces inhibiteurs se lient de manière similaire aux inhibiteurs sélectifs de smHDAC8 dans une poche spécifique de HDAC8 que nous avons appelée poche sélective de HDAC8 (*Fig ii*). Une fois de plus, toutes nos données structurales fournissent des informations très précises sur la base moléculaire de l'inhibition sélective de HDAC8, ouvrant la voie à la conception d'inhibiteurs plus puissants et plus sélectifs.

Ce travail sur l'inhibition sélective de HDAC8 a révélé l'importance de la taille et de la conformation de certaines boucles impliquées dans la formation du site actif de HDAC8. J'ai donc complété mon analyse d'inhibition par une analyse mutationnelle des boucles du site actif de HDAC8. Mes résultats montrent que la conformation de ces boucles est fortement contrainte, ce qui explique comment l'inhibition sélective peut être obtenue. Ces résultats ouvrent également la question de l'importance de ces boucles dans l'activité de HDAC8 et de sa reconnaissance de ses cibles, notamment en raison de la faible spécificité communément supposée des HDACs pour leurs cibles.

La cible la plus connue de HDAC8 est la Cohésine, un complexe qui joue un rôle important dans la cohésion des chromatides soeurs, la régulation transcriptionnelle et la réparation de l'ADN par recombinaison homologe. HDAC8 et la Cohésine sont impliquées dans de nombreux cancers, et des mutations dans HDAC8 et différentes sous-unités de la Cohésine conduisent à la même maladie, le syndrome de Cornelia de Lange, caractérisé par un nanisme et un handicap intellectuel.

Dans la dernière partie de ma thèse, j'ai commencé à étudier l'interaction entre la Cohésine et HDAC8 qui a été montrée comme désacétylant spécifiquement la sous-unité Smc3 de la Cohésine. La Cohésine est un complexe large et très flexible qu'il est difficile à caractériser structuralement. La plupart des structures actuellement publiées sont celles de petits domaines, généralement d'eucaryotes inférieurs. En utilisant la technique de co-expression développée dans l'équipe, j'ai pu reconstituer différents sous-complexes de la Cohésine humaine et, pour certains d'entre eux, j'ai déjà pu obtenir des cristaux. Ce travail ouvre la voie à la caractérisation structurale de la Cohésine humaine, mais aussi à son interaction avec HDAC8, pour mieux comprendre comment ces deux effecteurs nucléaires interagissent, avec des implications pour la maladie et l'inhibition sélective.

## **Conclusions**

Une grande quantité de travail a été réalisée et continue d'être entreprise sur les HDACs qui représentent des cibles thérapeutiques importantes pour le développement



d'épimédicaments. En dépit de tout ce travail, des questions majeures subsistent sur les HDACs qui concernent des aspects fondamentaux tels que la reconnaissance spécifique de leurs cibles et leur mode d'action au sein de complexes, mais aussi des aspects plus applicatifs comme le développement d'inhibiteurs sélectifs pour étendre l'utilisation d'épimédicaments ciblant les HDACs.

Au cours de mon doctorat, j'ai abordé la plupart de ces problèmes en utilisant la HDAC8 humaine et du parasite *Schistosoma mansoni*. Plus précisément, mon travail sur l'inhibition sélective de HDAC8 a fourni des informations détaillées sur les bases moléculaires de cette inhibition sélective, ouvrant ainsi la voie au développement d'inhibiteurs plus puissants et plus sélectifs. Il est important de souligner que, à un moment où la recherche de nouveaux médicaments par des stratégies de screening à haut débit de chimiothèques et des méthodes de screening in silico sont favorisées par de nombreux chimistes médicaux, mon travail montre comment les données structurales peuvent apporter des informations hautement complémentaires et essentielles dans le processus de découverte de nouveaux médicaments. De plus, ce travail m'a amené à étudier plus précisément l'importance de l'architecture globale du site actif des HDACs. Ce travail remet en question le dogme actuel qui considère que les enzymes HDACs sont peu sélectives en termes de cibles. Pour adresser ce problème, j'ai commencé à caractériser l'interaction entre HDAC8 et sa cible principale, le complexe Cohésine. Ce travail a déjà apporté des résultats essentiels qui permettront d'étudier finement cette interaction. Ainsi, le travail effectué lors de mon doctorat combine des recherches fondamentales et applicatives avec des implications pour des interventions thérapeutiques vers le cancer, et les maladies neurologiques et parasitaires.

# Summary

The eukaryotic genomic DNA is packaged into a compact dynamic structure known as chromatin whose minimal subunit is the nucleosome (147 base pairs of DNA wrapped by histone proteins). Chromatin structure modulates the accessibility to the genomic information and thereby regulates the other nuclear processes. Importantly, chromatin structure can be modified by epigenetic mechanisms, enabling the fine tuning of nuclear processes. Accordingly, epigenetic mechanisms have a major impact on cell homeostasis, cell type and development, and deregulation of epigenetic mechanisms has been shown to be implicated in many diseases such as cancer, but also neurological and inflammatory diseases. It is therefore essential to understand the fundamental basis of epigenetic mechanisms. But epigenetic effectors also represent important targets for therapeutic interventions in many diseases.

Among the epigenetic mechanisms, the “writing” and “erasing” by dedicated epigenetic enzymes of covalent epigenetic marks on histones and other nuclear effectors has a major effect on chromatin structure and its modification in response to various stimuli. Specifically, the reversible covalent addition of either small chemical groups (e.g. acetyl, methyl, phosphate) or macromolecules (e.g. ubiquitin, SUMO) to specific residues of histone can facilitate or prevent DNA accessibility by the nuclear machineries.

Acetylation is one of the major epigenetic mark and is often associated with a chromatin structure that is permissive to nuclear processes. Acetylation of histones and other nuclear effectors is controlled by the opposing action of HATs (Histone acetyl transferases) and HDACs (histone deacetylases). Interestingly, HDAC enzymes represent the major targets of currently approved epigenetic drugs (epidrugs). These epidrugs are however restricted to the treatment of the few cancers since they inhibit almost all members of the HDAC family that have very different functions. Therefore, a major research axis in epidrug discovery is to develop new molecules that will show high selectivity for given members of the HDAC family.

Importantly, despite years of study of HDACs, it remains very obscure how the different members recognize specifically their targets and how they act on these targets. In addition, HDACs are quite often part of macromolecular complexes and other subunits of these complexes can affect activity, substrate recognition, but also inhibition by epidrugs. It is therefore essential to combine both basic and applicative research on HDACs to develop future more potent and more selective epidrugs that can be used more extensively to treat a large panel of diseases.

Interestingly, the importance of epigenetic mechanisms is not restricted to human but also to all eukaryotic organisms. This opens the door to the fight against eukaryotic parasites that causes yearly millions of deaths worldwide. Such parasites (e.g. plasmodium, trypanosomes, schistosomes) often have very intricate life cycles where epigenetic mechanisms are expected to play essential roles. Accordingly, treatment of parasites with approved epidrugs targeting HDACs have revealed a strong sensitivity of the parasites to these epidrugs.

This clearly opens the way for the development of new anti-parasitic drugs to fight diseases for which there are no to very few drugs available, with increasing report of resistance. Owing to the fact that the drug discovery process is extremely time-consuming and expensive, a “piggyback” strategy consisting in modifying currently approved drugs to make them highly potent against parasites is expected to speed up the search for new anti-parasitic drugs. Here again, a major bottleneck is to modify the drugs so that they can be selective for the parasitic enzymes but not anymore for the human ones.

During the course of the European-funded project SEtTREND (*Schistosoma* Epigenetics: Targets, Regulation, New Drugs – 2010-2012), my laboratory has provided the proof of concept of this epigenetic piggyback strategy by targeting notably HDAC8 from the parasitic flatworm *Schistosoma mansoni* (smHDAC8). By solving the structure of smHDAC8 in apo form and in complex with pan-HDAC inhibitors, including the approved Vorinostat drug, the team provided invaluable specific and unexpected information that has been used for finding initial inhibitors showing good selectivity but medium potency. Structures of smHDAC8 with these molecules were further used to develop a new series of inhibitors with high potency (nM range) and good to excellent selectivity.

During my PhD thesis and within the frame of the larger European-funded project A-ParaDDisE (Anti-Parasitic Drug Discovery in Epigenetics – 2014-2017), I have worked on the structural characterization of the complexes between smHDAC8 and the new series on inhibitors to better understand the molecular basis of high potency and selectivity. I have complemented this work by the biochemical, biophysical and structural analysis of the complex between smHDAC8, and also human HDAC8 (hHDAC8), with known highly selective hHDAC8 inhibitors. This work led me to look by mutational analysis at the importance of HDAC8 active site architecture for activity and inhibition. Finally, in a last part of my PhD thesis, I have investigated more

fundamental aspects of HDAC8 biology, by notably looking at its interaction with one of its main target, the Cohesin complex.

During my PhD thesis I solved tens of structures of smHDAC8 complexed with various inhibitors made by our collaborators based on our initial structures. The crystal structure of smHDAC8 with the simplest molecule of this series, TH31, revealed a novel mechanism of inhibition of these inhibitors for smHDAC8. Specifically, in addition to the common coordination of the HDAC catalytic zinc by a hydroxamate moiety, binding of TH31 in smHDAC8 active site pocket is stabilized by the interaction of its internal amide with two smHDAC8 residues, K20 and H292. Inhibitor binding is further stabilized by hydrophobic contacts made between its capping group and the L6 loop of smHDAC8.

This binding enables to gain selectivity for smHDAC8 over other human HDACs by several folds. Only hHDAC8 was still strongly inhibited by TH31. Therefore, we selected several other inhibitors of this series that showed increased potency for smHDAC8 but also increased selectivity for smHDAC8 over hHDAC8. These inhibitors had additional substitutions and different capping groups compared to TH31. The structure of smHDAC8 with these different inhibitors has provided a wealth of information how potency and selectivity can be gained by improving and extending the interactions observed initially between TH31 and smHDAC8.

Interestingly, human HDAC8 is one of the HDACs for which highly selective inhibitors (e.g. PCI-34051) have been found but the structural basis for this selectivity has remained unknown. I have investigated the molecular basis of HDAC8 selectivity with these inhibitors by determination of the structure of hHDAC8 and smHDAC8 with several of these inhibitors. Importantly, these structures showed that these inhibitors bind similarly to the smHDAC8-selective inhibitors to these enzymes in a HDAC8-specific pocket that we have termed HDAC8-selective pocket. Again, all our structural data provide precise information on the structural basis for selective inhibition, paving the way for the design of more potent and more selective inhibitors.

This work on the selective inhibition of HDAC8 has revealed the importance of the size and conformation of some of the loops building HDAC8 active site. I have therefore complemented our inhibition analysis by a mutational analysis of the active site loops of HDAC8. Our results show that the conformation of these loops is highly restricted, which further explains how selective inhibition can be achieved. This also opens the question of the importance of these

loops for activity and substrate recognition, notably arguing against the commonly accepted poor substrate specificity of HDACs.

HDAC8 best characterized target is Cohesin, a complex that plays an important role in sister chromatid cohesion, transcriptional regulation and DNA repair via homologous recombination. Specifically, HDAC8 and Cohesin are involved in a wide range of cancers, and mutations in both HDAC8 and different subunits of Cohesin lead to the same disease, the Cornelia de Lange syndrome, characterized by dwarfism and intellectual disability.

In the last part of my thesis, I have started to study the interaction between Cohesin and HDAC8 that has been shown to deacetylate specifically the Smc3 subunit of Cohesin. Cohesin is a large, highly flexible complex that is difficult to characterize structurally. Most currently published structures are of small domains, generally from lower eukaryotes. By using the technique of co-expression developed in the team, I have been able to reconstitute different sub-complexes of the human Cohesin and, for some of them, I have already obtained crystals. This work opens the way to the structural characterization of human Cohesin but also of its interaction with HDAC8 to better understand how these two nuclear effectors interact, with implication in disease and selective inhibition.

A large amount of work has been performed and continues to be carried out on HDACs that represent important therapeutic targets in epidrug discovery. Despite all this work, major issues remain that concern fundamental aspects such as the specific recognition of their targets by HDACs and their mode of action within complexes, but also the development of selective inhibitors against the different HDACs to extend the use of epidrugs targeting these enzymes.

During my PhD, I have addressed most of these issues using HDAC8 from human and the from parasite *Schistosoma mansoni*. Specifically, my work on the selective inhibition of HDAC8 has provided detailed information on the molecular basis of this selective inhibition and is paving the way for the development of more potent and more selective inhibitors. Importantly, at a time where high-throughput screening and in silico screening methods appear favoured by many medicinal chemists, my work is showing how structural data can bring highly complementary and essential data in the drug discovery process.

Importantly, this work has brought me to investigate more specifically the importance of the overall architecture of the active site of HDACs. This work questions the actual dogma that

implies that HDAC enzymes are poorly selective in term of substrate. To address this issue, I have started the characterization of the interaction between HDAC8 and its main target, the Cohesin complex. This work has already brought essential results that will enable from now on to investigate finely this interaction. Thus, the work done during my PhD is combining fundamental and applicative research with implication for therapeutic interventions towards cancer, neurological and parasitic diseases.

# Table of contents

Figures.....	17
Tables List.....	19
1. Introduction .....	24
1.1. Chromatin and epigenetic mechanisms: impact on nuclear processes in health and diseases .	24
1.1.1. The chromatin.....	26
1.1.1.1. The nucleosome: basic unit of the chromatin .....	26
1.1.1.2. Higher order chromatin organization .....	29
1.1.2. Epigenetic mechanisms.....	31
1.1.2.1. Chromatin structure modulation through epigenetic mechanisms .....	31
1.1.2.2. The histone epigenetic marks .....	32
1.1.2.2.1. Post-translational modifications.....	32
1.1.2.2.2. Modification of non-histone proteins.....	36
1.1.2.2.3. Writers, readers and erasers of the epigenetic marks .....	36
1.1.2.2.4. Histone acetyltransferases and histone deacetylases.....	37
1.1.2.2.5. Histone methyltransferases and histone demethylases.....	37
1.1.2.2.6. Kinases and phosphatases .....	39
1.1.2.2.7. Readers domains of the epigenetic marks .....	40
1.1.2.3. ATP-dependent chromatin remodellers .....	40
1.1.2.4. Histone variants and histone chaperones .....	43
1.1.2.5. Long non-coding RNAs .....	45
1.1.3. Epigenetic mechanisms in diseases and therapy.....	46
1.1.3.1. Epigenetic mechanisms in autoimmune diseases .....	47
1.1.3.2. Epigenetic mechanisms in neurological diseases .....	48
1.1.3.3. Epigenetic mechanisms in cancer .....	49
1.1.3.3.1. DNA Methylation in cancer.....	49
1.1.3.3.2. Post translational modifications of histones in cancer.....	50
1.1.3.3.3. Other epigenetic effectors in cancer .....	51
1.1.3.4. Targeting epigenetic players.....	51
1.1.3.4.1. Inhibitors targeting DNA methyltransferases.....	51
1.1.3.4.2. Inhibitors targeting bromodomain acetylation readers .....	52
1.1.3.4.3. Inhibitors targeting HATs and HDACs .....	52



1.1.3.4.4.	Inhibitors targeting protein methyltransferases .....	52
1.1.4.	A structural view of epigenetic targets .....	53
1.2.	The acetylation mark .....	69
1.2.1.	Role of the acetyl epigenetic mark .....	69
1.2.1.1.	Acetylation and mechanisms of acetylation .....	69
1.2.1.2.	A brief history of acetylation.....	70
1.2.1.3.	A handshake of acetylation with other PTMs .....	72
1.2.1.4.	Histone acetyl transferases.....	73
1.2.1.4.1.	HAT enzymes.....	73
1.2.1.4.2.	HAT inhibitors and activators .....	76
1.3.	Histone deacetylases: sirtuins and HDACs.....	78
1.3.1.	Histone deacetylases classification.....	78
1.3.2.	Sub cellular localization of histone deacetylases.....	82
1.3.3.	Structure of histone deacetylases.....	83
1.3.3.1.	HDACs.....	83
1.3.3.2.	Sirtuins .....	87
1.3.4.	Complexes of histone deacetylases .....	90
1.3.4.1.	Complexes of class I HDACs .....	91
1.3.4.2.	Complexes of Class II HDAC: .....	94
1.3.4.3.	Sirtuin complexes:.....	95
1.3.5.	Substrates of histone deacetylases.....	96
1.3.6.	Mechanism of action.....	98
1.3.6.1.	Mechanism of class I HDACs .....	98
1.3.6.2.	Mechanism of Class II HDACs.....	101
1.3.6.3.	Mechanism of action of sirtuins.....	102
1.3.7.	Histone deacetylases in diseases and therapy.....	103
1.3.7.1.	Inhibition of HDACs in human diseases .....	103
1.3.7.2.	An introduction to HDAC inhibitors .....	106
1.3.7.3.	Selective inhibition of HDACs.....	110
1.3.8.	HDAC8 .....	112
1.3.8.1.	Introduction to HDAC8.....	112
1.3.8.2.	Role of HDAC8 on the Cohesin complex .....	114
1.3.8.3.	The Cornelia de Lange Syndrome .....	118

1.3.9.	On the interest of studying HATs and HDACs .....	119
1.4.	Human neglected diseases .....	120
1.4.1.	Neglected Tropical Diseases .....	120
1.4.2.	Actual treatment of NTDs .....	122
1.4.3.	Importance of epigenetics and HDACs in NTDs .....	122
1.4.4.	Proof of concept: smHDAC8 a valid drug target to fight schistosomiasis by a piggyback strategy .....	124
1.4.4.1.	Schistosomiasis .....	124
1.4.4.2.	smHDAC8 .....	126
1.5.	PhD thesis rationale: selective inhibition of parasitic enzymes and beyond.....	129
2.	Material and methods .....	130
2.1.	Cloning strategies.....	131
2.1.1.	Expression vectors .....	131
2.1.2.	Restriction endonuclease dependent cloning .....	132
2.1.3.	Sequence and Ligase Independent Cloning (SLIC) .....	133
2.1.4.	Gibson cloning.....	134
2.1.5.	Site directed mutagenesis.....	135
2.1.6.	Cloning of SMC-HD with linker.....	137
2.2.	Expression methods.....	137
2.2.1.	Principle of expression in case of the PET system of expression.....	138
2.2.2.	Expression protocols.....	139
2.2.2.1.	Mini expression test.....	140
2.2.2.2.	Large scale production .....	140
2.3.	Purification techniques .....	140
2.3.1.	Affinity chromatography .....	141
2.3.2.	Ion exchange chromatography .....	142
2.3.3.	Gel filtration chromatography .....	143
2.3.4.	Purification protocols.....	145
2.3.5.	smHDAC8 purification.....	146
2.3.5.1.	Large-Scale Overproduction and Purification of Recombinant Histone Deacetylase 8 (HDAC8) from the Human-Pathogenic Flatworm Schistosoma mansoni .....	146
2.3.5.2.	Modified protocol for smHDC8.....	147
2.4.	Protein characterization .....	148
2.4.1.	SDS-PAGE gel electrophoresis.....	148

2.4.2.	Dynamic Light Scattering .....	150
2.4.3.	Differential Scanning Fluorimetry .....	151
2.4.4.	<i>In vitro</i> HDAC assay .....	151
2.4.5.	Crystallization vapor diffusion Technique.....	153
2.5.	X-ray crystallography .....	154
2.5.1.1.	Crystal systems.....	154
2.5.1.2.	Bragg's law .....	155
2.5.1.3.	Ewald's sphere .....	156
2.5.1.4.	Theory of diffraction .....	157
2.5.2.1.	Molecular replacement.....	160
2.5.2.2.	Refinement.....	162
3.	Results.....	164
3.1.	Purification of acetyl modifying proteins from eukaryotic parasites .....	165
3.1.1.	Histone Acetyltransferase 1 from <i>Trypanosoma cruzi</i> (TcHAT1).....	165
3.1.1.1.	Initial expression tests of TcHAT1 .....	165
3.1.1.2.	TcHAT1-N2C2 Purification.....	167
3.1.1.3.	Optimization of production .....	167
3.1.1.4.	Purification of dl-TcHAT1-N2C2 .....	168
3.1.1.5.	Purification of lh-TcHAT1-N2C2 .....	169
3.1.2.	Deacetylase 3 from <i>Leishmania braziliensis</i> (LbDAC3) .....	170
3.1.2.1.	Purification of LbDAC3:.....	171
3.1.2.2.	Optimization of expression and purification: .....	173
3.1.3.	Conclusion:.....	175
3.2.	Elucidating smHDAC8 selective inhibition mechanism.....	175
3.2.1.	Production of HDACs.....	175
3.2.1.1.	Purification of smHDAC8 and hHDAC8 .....	176
3.2.1.2.	Crystallization and soaking with inhibitors of smHDAC8.....	178
3.2.2.	Publication 1 .....	181
3.2.3.	Publication 2 .....	182
3.2.4.	Crystal structure of smHDAC8 with J1075 derived TB compounds .....	183
3.2.5.	Crystal structures of smHDAC8 with miscellaneous inhibitors.....	185
3.2.5.1.	Triazole derivatives .....	185
3.2.5.1.	Crystal structures of smHDAC8 with Uracil based compounds .....	189

3.3.	Cohesin complex purification .....	191
3.3.1.	Expression of cohesin complex subunits .....	192
3.3.2.	Purification of hsSMC1-HD.....	195
3.3.3.	Purification and initial crystallization attempts of the hsSMC1-HD /RAD21-CTD1 complex	196
3.3.4.	Optimization of the hsSMC1HD-RAD21 complex production.....	198
3.3.5.	Expression of the hsSMC3HD/N-terminal region of RAD21 complex.....	200
3.3.6.	Purification of hsSMC3-HD/RAD21-NTD2 .....	201
3.3.7.	Characterization of the purified complexes using DLS .....	202
3.3.8.	Crystallization attempts of the SMC-HD/RAD21 complexes .....	203
3.3.9.	Data collection analysis.....	204
3.3.10.	Conclusions and future perspectives on the HDAC8/Cohesin project .....	206
4.	Discussion .....	207
5.	Future Perspectives .....	213
6.	List of poster presentations and oral communications .....	214
7.	List of publications .....	215
8.	References: .....	216

# Figures

Figure 1: <i>Chromatin</i> organization: .....	24
Figure 2: <i>Structure of the nucleosome</i> : .....	27
Figure 3: Influence of H1 binding on nucleosomal arrays. ....	29
Figure 4: Models of higher order nucleosome organizations. ....	30
Figure 5: Condensin in chromosome condensation - The beads on a string model:.....	31
Figure 6: Dynamic properties of nucleosomes. ....	32
Figure 7: Histone chaperones and histone deposition mechanism.....	45
Figure 8: One-step HAT mechanism of action. ....	75
Figure 9: HAT mechanism of action:.....	76
<i>Figure 10</i> : Phylogenetic tree of eukaryotic HDACs:.....	80
Figure 11: Domain organization of HDACs and sirtuins:.....	82
Figure 12: HDAC fold: .....	84
Figure 13: Structure of class IIa HDACs.....	86
Figure 14: Structure of HDAC10 .....	87
Figure 15: Structure of SIRT1 with resveratrol .....	89
Figure 16: Sirtuin structures: .....	90
Figure 17: Sequence alignment of HDAC loops .....	92
Figure 18: Crystal structure of HDAC3-SMRT complex.....	93
Figure 19: Surface representation of HDAC1-MTA1 complex crystal structure: .....	94
Figure 20: Sirtuin complexes .....	96
Figure 21: Class I HDACs mechanism of action.....	100
Figure 22: Class IIa HDACs mechanism of action.....	101
Figure 23: HDAC6 mechanism of action .....	102
Figure 24: Sirtuin mechanism of action.....	103
Figure 25: Consequences of HDAC inhibition: .....	105
Figure 26: Chemical constituents of a HDAC inhibitor .....	106
Figure 27: Different classes HDAC inhibitors .....	107
Figure 28: HDAC8 substrates .....	113
Figure 29: Cohesin complex architecture. ....	115
Figure 30: Cohesion - cell cycle regulation. ....	116
Figure 31: Global Overlap of common NTDs. ....	121
Figure 32: NTDs caused by different type of infections.....	121
Figure 33: Schistosoma life cycle. ....	125
Figure 34: Comparison of catalytic subunit of smHDAC8 with human HDACs.....	128

Figure 35: Experimental approach: .....	130
Figure 36: pnEA-tH vector: .....	132
Figure 37: Steps in Traditional cloning .....	133
Figure 38: Sequence and ligase independent cloning .....	134
Figure 39: Overview of Gibson assembly.....	135
Figure 40: Mechanism of pET expression system.....	139
Figure 41: Affinity chromatography: .....	142
Figure 42: Ion exchange chromatography:.....	143
Figure 43: Gel filtration chromatography.....	145
Figure 44: Polymerization of acrylamide .....	148
Figure 45:DLS outlook.....	150
Figure 46: Fluor de Lys assay: A:.....	152
Figure 47: IC50 determination.....	153
Figure 48: Illustration of vapor diffusion using Sitting drop and handing drop setup.....	153
Figure 49: Bravais lattices .....	155
Figure 50: Bragg's law .....	156
Figure 51: Construction of Ewald's sphere .....	157
Figure 52: Argand diagram: .....	159
Figure 53: Calculation of Patterson peaks from electron density: .....	161
Figure 54: Sequence alignment of TcHAT1:.....	166
Figure 55: Gel filtration purification of TcHAT1-N2C2.....	167
Figure 56: TcHAT1 – mini test-expression optimization:.....	168
Figure 57: Purification profile of dl-TcHAT1-N2C2: .....	169
Figure 58: Purification profile of lhTcHAT1N2C2:.....	169
Figure 59: DLS profile of lh-TcHAT1-N2C2 .....	170
Figure 60: LbDAC3 expression optimization:.....	171
Figure 61: Purification of LbDAC3.....	172
Figure 62: Sequence alignment of LbDAC3 with HDAC7 and 4. ....	173
Figure 63: Ion exchange profile of LbDAC3 .....	174
Figure 64: HDAC8 expression optimization: .....	176
Figure 65: smHDAC8 purification profile.....	177
Figure 66: hHDAC8 purification profile.....	177
Figure 67: J1075 derived compounds .....	179
Figure 68: J1038 derived compounds .....	179
Figure 69: Pan-HDAC and HDAC8 selective inhibitors .....	180

Figure 70: Triazole and Uracil based compounds.....	180
Figure 71: smHDAC8-J1075 derivatives crystal structures: .....	185
Figure 72: Crystal structures of smHDAC8 with triazole derivatives and Uracil based compound .....	188
Figure 73: Strategy to SMC complex production:.....	192
Figure 74: Sequence alignment of hsSMC1A with ABC ATPases: .....	193
Figure 75: Expression of SMC1-HD .....	194
Figure 76: Purification profile of hsSMC1HD: .....	195
Figure 77: Characterization of hsSMC1-HD using DLS: .....	196
Figure 78: Purification profile of SMC1-HD/RAD21-CTD1. ....	196
Figure 79: Thrombin cleavage of SMC1-HD/RAD21-CTD1 complex .....	197
Figure 80: Expression profiles of hsSMC1HD and RAD21CTD optimization: .....	199
Figure 81: Purification profile of SMC-HD/RAD21-CTD2: .....	199
Figure 82: SMC3-RAD21NTD co-expression: .....	200
Figure 83: Purification profile of hsSMC3-HD/RAD21-NTD2: .....	201
Figure 84: Optimization of SMC3HD-RAD21NTD2 complex purification:.....	202
Figure 85: DLS profiles of SMC-RAD21 complexes: .....	203
Figure 86: Crystals of hsSMC1HD-RAD21CTD2 complex. ....	204

## Tables List

Table 1: Table showing important histone modifications and their function .....	34
Table 2: Important phosphorylation sites and known kinases (Rossetto, Avvakumov <i>et al.</i> 2012). .....	39
Table 3: Chromatin remodelling complexes in humans. ....	42
Table 4: List of histone variants and their functions (Buschbeck and Hake 2017). ....	44
Table 5: Table of HAT complexes.....	74
Table 6: Inhibitors and activators of HATs.....	77
Table 7: Classification of HDACs: .....	79
Table 8: HDAC inhibitors.....	109
Table 9: Vectors with different options of tags and antibiotic resistance used in this thesis .....	131
Table 10: Rolling circle plasmid synthesis protocol .....	136
Table 11: Thermocycler protocol for rolling circle plasmid synthesis .....	136
Table 12: SDS-PAGE gel composition.....	149
Table 13: SDS-PAGE staining solution.....	149
Table 14: IC50 values of inhibitors.....	184

Table 15: Crystallographic table II ..... 190

Table 16: Data statistics of hsSMC1HD-RAD21CTD2 ..... 205

### List of Abbreviations

Alphabet	List of Abbreviations
	% : percentage µL : microliter µg : microgram ψDAC : pseudo deacetylase domain
<b>A</b>	AAT : (7-[(3-aminopropyl) amino]-1,1,1-trifluoroheptan-2-one) ADP : adenosine diphosphate ALS : Amyotrophic lateral sclerosis Amp : Ampicilline A-ParaDDisE : Anti-Parasitic Drug Discovery in Epigenetics; 2014-2017 APS : ammonium persulfate ARP : actin related proteins ATP : Adenosine triphosphate AURKB : aurora kinase B
<b>B</b>	BET : <i>Bromodomain</i> and Extra-Terminal motif
<b>C</b>	CBP : CREB binding protein CD : catalytic domain CDC : Centers for Disease control and prevention CDK1 : Cyclin Dependent Kinase 1 CdLS : Cornelia de Lange Syndrome cDNA: complementary DNA CHD : chromodomains helicase DNA binding Chl : Chloramphenicol cm: centimeters CNRS : Centre National de la Recherche Scientifique CoA : coenzyme A CoREST : Corepressor of RE1-Silencing Transcription factor cryoEM: Cryo-Electron Microscopy CSB : Cockayne Syndrome group B CTD : C-terminal domain CTD2 : RAD21 C-terminal region
<b>D</b>	Da : dalton DAD : deacetylase activation domain DAXX: death domain-associated protein DFT : <i>density functional theory</i> DHS : deoxyhypusine synthase DLS : Dynamic Light Scattering DNA: DeoxyriboNucleic Acid DNMTs : DNA methyltransferases dNTPs: deoxyNucleotides Tri-Phosphate dsDNA : double strand DNA



	<p>DTT : Dithiothreitol  DUBm : SAGA deubiquitination module</p>
<b>E</b>	<p>EDTA: Ethylenediaminetetraacetic acid  EM : electron microscopy  Epidrugs : epigenetic drugs  ESFRI : European Strategy Forum on Research Infrastructures</p>
<b>F</b>	<p>FDA : American Food and Drug Administration  FRISBI : French Infrastructure for Integrated Structural Biology</p>
<b>G</b>	<p>GST : Glutathione S-transferase</p>
<b>H</b>	<p>H3K4me2 : dimethylation of lysine 4 from histone H3  HATs : Histone Acetyltransferases  HCl: Hydrochloric acid  HDACs : Histone deacetylases  HDACis : HDAC inhibitors  HDLP : HDAC-like protein  HEK293 cells : human epithelial kidney cell line  His: Histidine  HIV : human immunodeficiency virus  hnRNP : heterogeneous nuclear ribonucleoproteins  HPLC: High-performance liquid chromatography  HSA : helicase SANT-associated  HSP90 : heat shock protein 90</p>
<b>I</b>	<p>IC50 : half maximal inhibitory concentration  ICF : immunodeficiency centromeric instability and facial anomalies  IEX : Ion exchange  IFN<math>\alpha</math> : interferon <math>\alpha</math>  IFN-<math>\gamma</math> : interferon- <math>\gamma</math>  IFN<math>\alpha</math>R2 : IFN<math>\alpha</math> receptor 2  IGBMC : Institut de Génétique et de Biologie Moléculaire et Cellulaire  INSERM : Institut National de la Santé et de la Recherche Médicale  IPTG: Isopropyl <math>\beta</math>-D-1-thiogalactopyranoside  IRF9 : interferon regulatory factor 9  ISGF3 : Interferon-stimulated gene factor 3  ISWI : imitation switch</p>
<b>K</b>	<p>Kan : Kanamycin</p> <p>KAT : lysine acetyltransferases  KCl: Potassium chloride  KDAC : lysine deacetylases  Kpr, Kbu and Kcr : lysine propionylation, butyrylation and crotonylation</p>
<b>L</b>	<p>LB : Luria Bertani  LbDAC3 : Deacetylase 3 from <i>Leishmania braziliensis</i>  LINE : Long interspersed nuclear elements  lncRNAs : long non-coding RNAs  LTR : Long Terminal Repeat</p>
<b>M</b>	<p>M : mol/L</p>

	<p>MAGE : melanoma-associated gene  MBD : methyl CpG binding domain  MCS: Multiple Cloning Site  MeCP2 : methyl CpG binding protein 2  MgCl<sub>2</sub>: Magnesium chloride  MiDAC : Mitotic DeAcetylase Complex  miRNA : micro-RNA  MLL1 : Mixed lineage leukaemia protein 1  mRNA: messenger RNA  MSL: male specific lethal  MTA1 : Metastasis-associated protein 1</p>
<b>N</b>	<p>NaCl: sodium chloride  N-CoR : Nuclear receptor Co-Repressor  NCP : Nucleosome Core Particle  ncRNAs : non-coding RNAs  NES : Nuclear export signal  NFkB : nuclear factor κB  NIBL : nipped B like protein  NLS: nuclear localization signal  ng : nanograms  nm : nanometers  NMR : Nuclear magnetic resonance  NSCLC : non-small-cell lung cancer  nt : nucleotide  NTD : N terminal domain  NTDs : Neglected tropical diseases  NuRD : Nucleosome Remodeling Deacetylase</p>
<b>O</b>	<p>OD : Optical Density  O/N: over night</p>
<b>P</b>	<p>PCR: Polymerase Chain Reaction  PDAC : HDAC10 catalytic domain  PDB: Protein Data Bank  PEG : Polyethylene glycol  pH: potential of Hydrogen  PLK1 : Polo-like protein  PMTs : Protein Methyl Transferases  PP2A : protein phosphate 2A  PRMTs : Protein Arginine Methyltransferases  PTMs : Post-Translational Modifications</p>
<b>Q</b>	<p>QM/MM : DFT quantum mechanics / molecular mechanics</p>
<b>R</b>	<p>RENT : regulator of nucleolar silencing and telophase exit  RNA: Ribonucleic acid  ROS : Reactive oxygen species  Rpm: Revolutions per minute  rRNA: Ribosomal RNA  RT: Reverse Transcriptase  RTase: DGR-associated Reverse Transcriptases  RT-PCR: Reverse Transcription-Polymerase Chain Reaction</p>

<b>S</b>	<p>SAXS : Small Angle X-rays Scattering  SAH: S-adenosyl homocysteine  SAHA : suberanolohydroxamic acid  SAM : S-adenosyl methionine  SDS: Sodium Dodecyl Sulfate  SEC : size-exclusion chromatography  SEtTReND (Schistosoma Epigenetics: Targets, Regulation, New Drugs; 2010-2013)  SGO1 : Shugoshin  SHL2 : super-helical location 2  Sir2 : Silent information regulator 2  SIRT1-7 : sirtuins members  SLE : Systemic Lupus Erythematosus  SLIC : Sequence and Ligase Independent Cloning  SMC : Structural Maintenance of Chromosome  SMC3HD : SMC3 head domain  SMRT : Silencing Mediator of Retinoic acid and Thyroid hormone receptors  Sp : Spectinomycin  STAT2 : signal transducer and activator of transcription 2  SUMO : Small Ubiquitin-like Modifier  SWI/SNF : SWItch/Sucrose Non-Fermentable</p>
<b>T</b>	<p>TAF : TATA-binding protein associated factor  Tat : HIV transcriptional activator protein  TB : Terrific Broth  TBE: Tris Borate EDTA  TcHAT1 : Histone Acetyltransferase 1 from <i>Trypanosoma cruzi</i>  TEMED: N,N,N',N' Tetramethylethylene-1,2-diamine  TGF<math>\beta</math> : transforming growth factor <math>\beta</math>  tH4 : testis specific H4 variant  TFIID : Transcription factor II D  TFMO : Trifluoromethyloxadiazolyl  Tris: 2-amino-2-(hydroxymethyl)-1, 3-propanediol  tRNA : transfert RNA  TRX : thioredoxin  TSA : Trichostatin A</p>
<b>U</b>	UDR : ubiquitin-dependent recruitment motif
<b>V</b>	V:Volume
<b>W</b>	<p>WAPL : wings apart-like protein homologue  WHO : World Health Organization</p>

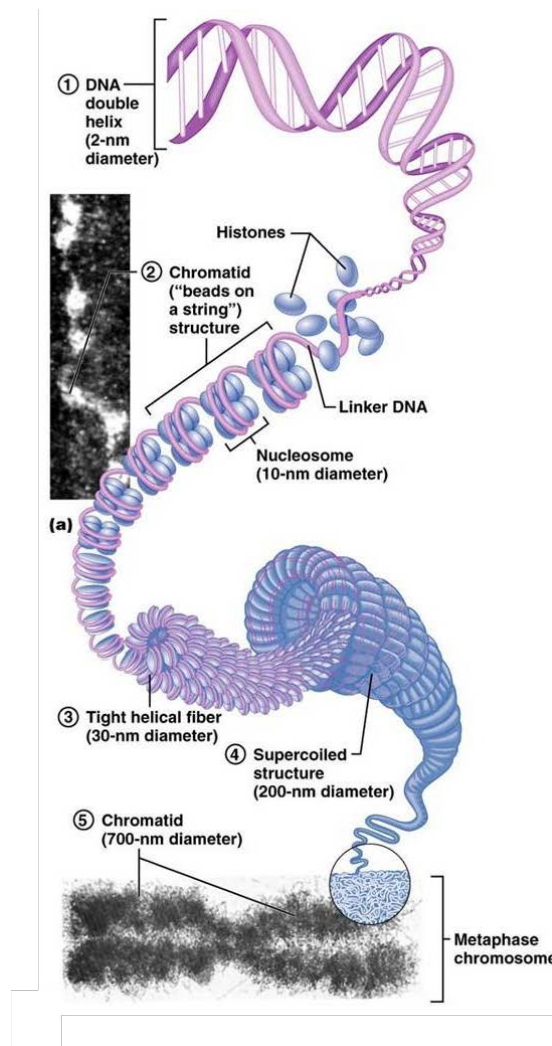
# Introduction

---

# 1. Introduction

## 1.1. Chromatin and epigenetic mechanisms: impact on nuclear processes in health and diseases

In order to fit within the 6 microns large nucleus, the 2 meters long eukaryotic genomic DNA is packaged into a highly ordered compacted structure called chromatin. The chromatin can have several levels of compaction, from a weakly condensed state to a highly condensed state as observed in the chromosomes (Figure 1). Compaction of the eukaryotic DNA into chromatin has a major effect on the nuclear processes (e.g. replication, transcription, DNA repair) as it restricts access to the underlying genetic material.



**Figure 1: Chromatin organization:**  
Picture adapted from [apsbiology.org](http://apsbiology.org)

The current view is that, depending on the chromatin compaction state, the genetic material is more or less accessible to the nuclear effectors. Actually, several major forms of chromatin are distinguished. The euchromatin is generally weakly compacted and is permissive to nuclear processes. In contrast, heterochromatin is generally densely packed and is associated with reduced nuclear activity. Finally, facultative heterochromatin is a less densely compacted chromatin than heterochromatin that displays little nuclear activity but is poised for transformation into euchromatin.

Depending on the gene expression patterns required, the chromatin state of a particular region of the genome can be different between different cell types and can be modified during development and in response to internal and external stimuli. This modulation of the chromatin structure is carried out by the so-called epigenetic mechanisms that enable the chromatin structure to be finely tuned and adapted to the cell's requirements.

Importantly, epigenetic mechanisms, by acting on the chromatin structure, have a strong influence on the various nuclear processes and are directly involved in their regulation. Thus, chromatin structure and epigenetic mechanisms are key to regulate genomic imprinting, X-chromosome inactivation, transcriptional regulation, DNA replication, DNA damage repair and DNA recombination (Sadakierska-Chudy, Kostrzewa *et al.* 2015). In addition, epigenetic mechanisms ensure that heritable changes are carried through mitosis and meiosis without altering the DNA sequence to maintain the tissue specific pattern of gene expression (Sadakierska-Chudy, Kostrzewa *et al.* 2015).

The consequence of the importance of chromatin structure and epigenetic mechanisms on the regulation of nuclear processes is of course that deregulation of these mechanisms has a broad impact on the cellular activity. It comes therefore as no surprise that an increasing number of epigenetic effectors are shown to be involved in the onset and progression of many different diseases. Importantly, the flexibility of epigenetic mechanisms and their capacity to modulate the chromatin structure open the way to epigenetic therapeutic interventions.

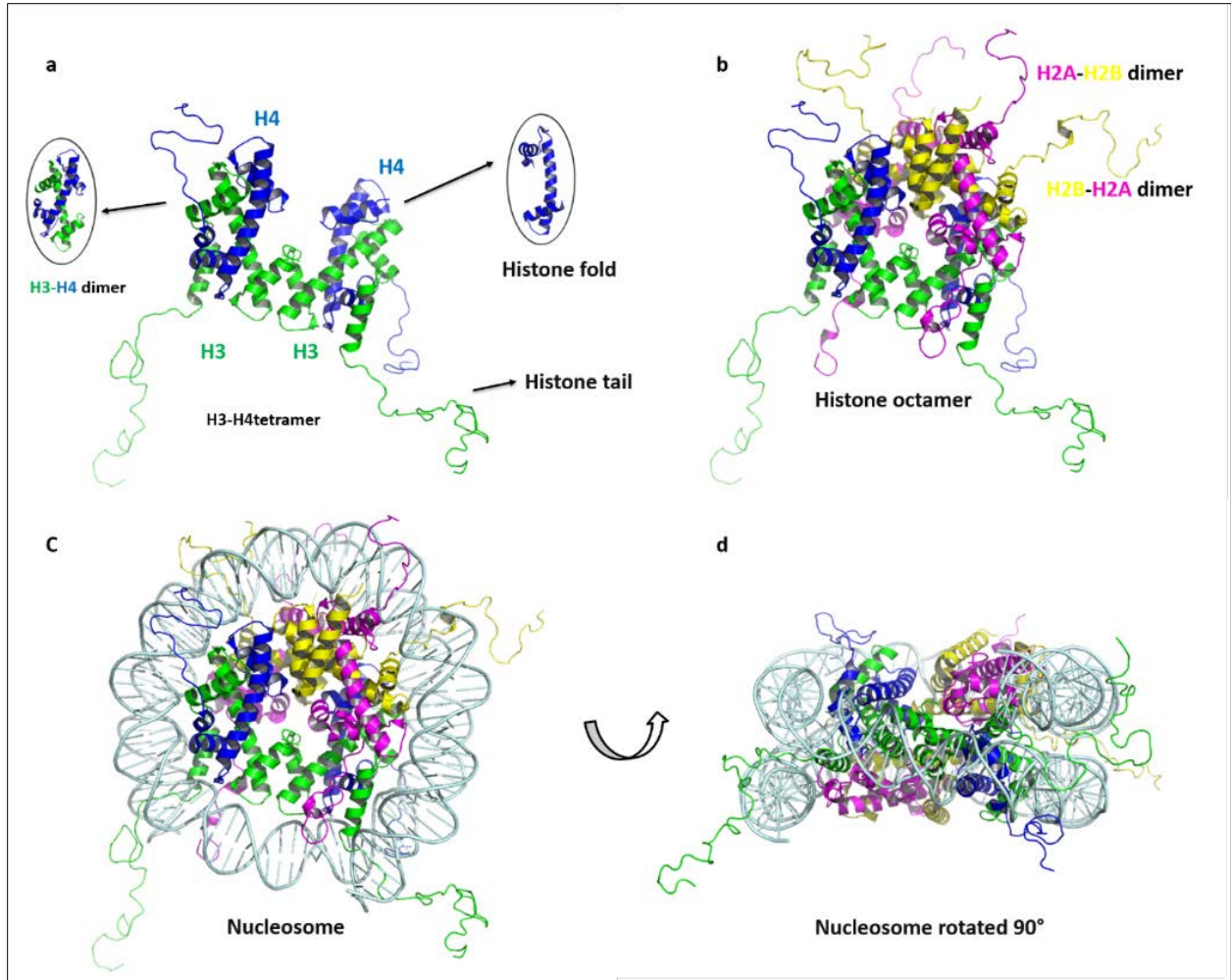
It is therefore essential that epigenetic mechanisms and their interplay with other cellular processes are finely characterized towards the development of therapeutic strategies. This process has already begun, but much more knowledge is required to progress further towards this goal.

### 1.1.1. The chromatin

#### 1.1.1.1. The nucleosome: basic unit of the chromatin

The nucleosome is the smallest structural and functional unit of the chromatin that provides a platform on which epigenetic effectors can act (Cutter and Hayes 2015). Each nucleosome consists of a nucleosome core particle (NCP), linker DNA and a linker histone. The NCP consists of 145-147 bp of DNA wrapped in a 1.7 super helical turn around the histone octamer. The histone octamer disk is made up of two copies of the four histones H2A, H2B, H3 and H4. Histones possess a conserved histone fold and an N-terminal tail domain (Figure 2). H2A contains an additional C-terminal tail that protrudes from the nucleosome towards linker DNA and interacts with histone H1 which is the linker histone (Davey, Sargent et al. 2002, Vogler, Huber et al. 2010). Tail domains are sensitive to proteases and protrude through the nucleosome minor grooves and are solvent exposed (Iwasaki, Miya et al. 2013). Most importantly tail domains participate in inter nucleosomal interactions which are important for maintaining higher order chromatin structure.

H2A specifically dimerize with H2B and H3 with H4. Two H3-H4 dimers are self-associated via H3-H3 interface to form a tetramer. Two H2A-H2B dimers bind on either side of the H4-H3-H3-H4 tetramer. The arrangement generates an octameric disk-like structure on which 121 bp of DNA wraps around histone folds and the sides of the octameric disk are aligned to major and minor grooves of DNA. The remaining 13bp of DNA from both sides are covered by the extensions of histone folds which completes the 147 bp nucleosome (Figure 2).



**Figure 2: Structure of the nucleosome:**

**a.** Tetramer formation of H3:H4 dimers, showing histone fold in inset and histone tails by arrow marks (pdb: 1KX5). **b.** Histone octamer formation with four histones. **c,d.** ribbon representations of the nucleosome in two perpendicular views. The disk form of the nucleosome is shown in d.

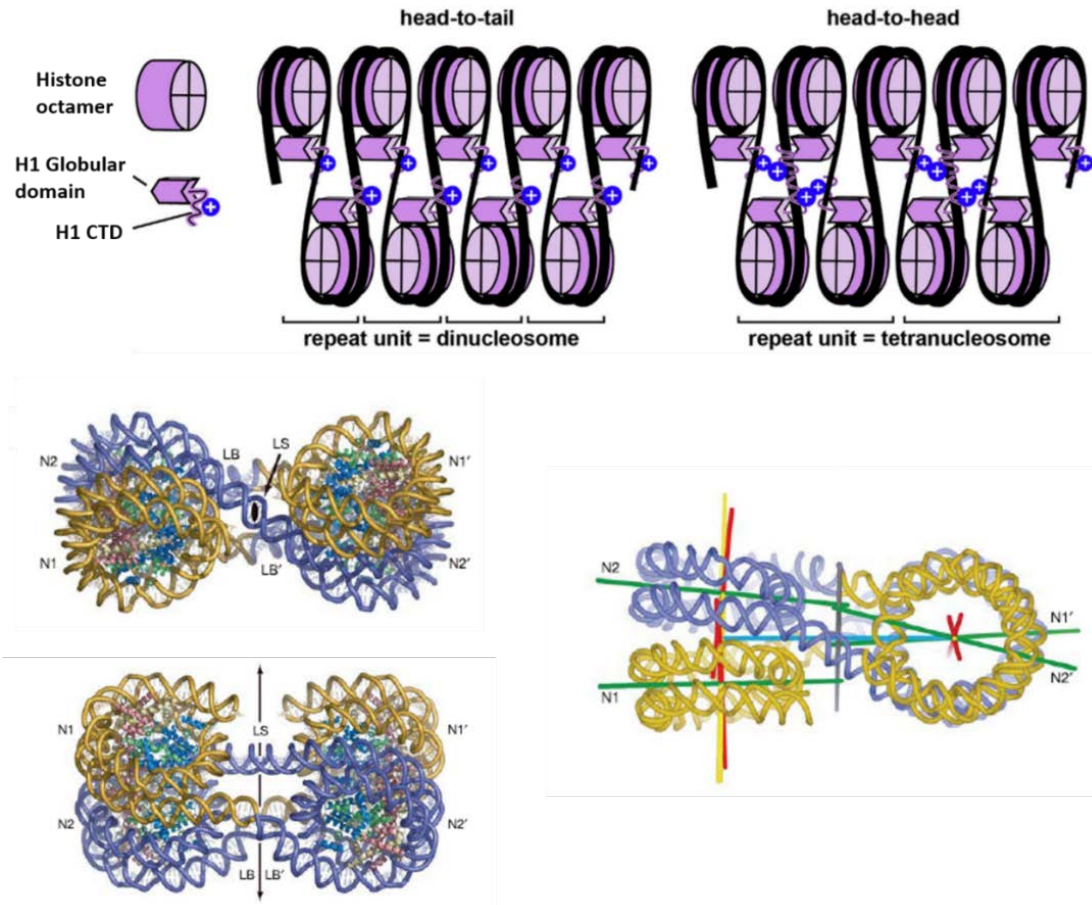
The interactions between positively charged histones and negatively charged DNA are keys for NCP assembly, and for nucleosome compaction. DNA makes 14 contacts with the histone octameric disk in the nucleosome where highly conserved arginine residue at each site interacts with the minor groove to help precisely position the DNA around the octameric disk and to facilitate the overall super helical shape (Gottesfeld and Luger 2001, Davey, Sargent *et al.* 2002, Cairns 2007, Wang, Ulyanov *et al.* 2010).

The linker histone, or H1, binds to the nucleosome with a weaker tendency than of core histones and it is also less conserved compared to core histones. Unlike core histones, the



organization of linker histone is completely different. In mammals 11 isoforms of H1 have been identified, which maintain a general domain organization with a N-terminal protease sensitive tail domain, a central folded globular domain, and highly basic, protease sensitive C-terminal domain (CTD). H1 CTD is essentially basic to neutralize the negative charge of the linker DNA. The globular domain binds at the interface near entry and exit of DNA in the nucleosome where it draws the two linker DNA together and reduce their flexibility (Bednar, Garcia-Saez *et al.* 2017). The globular domain of H1 is enough to increase the compaction of nucleosome.

Adjacent nucleosomes are connected by a segment of linker DNA which determines the space between nucleosomes. This arrangement generates a nucleosomal array in a diameter of 11nm which represents the known “beads on a string” organization of nucleosomes (Olins and Olins 1974). The flexibility of linker DNA is reduced by the H1 binding. The H1 binds the nucleosome in an asymmetric manner. The CTD of H1 binds one linker DNA initially and the globular domain binds asymmetrically on the dyad axis of nucleosome, this asymmetric H1 binding causes asymmetry in electrostatic and mass distribution. This asymmetry helps in the formation of higher order chromatin structure in case of nucleosomal arrays. The head to head and head to tail orientations of H1 proteins in the nucleosomal arrays result in the two different types of repeating structural units, dinucleosomes and tetra nucleosomes (Figure 3) (Bednar, Garcia-Saez *et al.* 2017).



**Figure 3: Influence of H1 binding on nucleosomal arrays.**

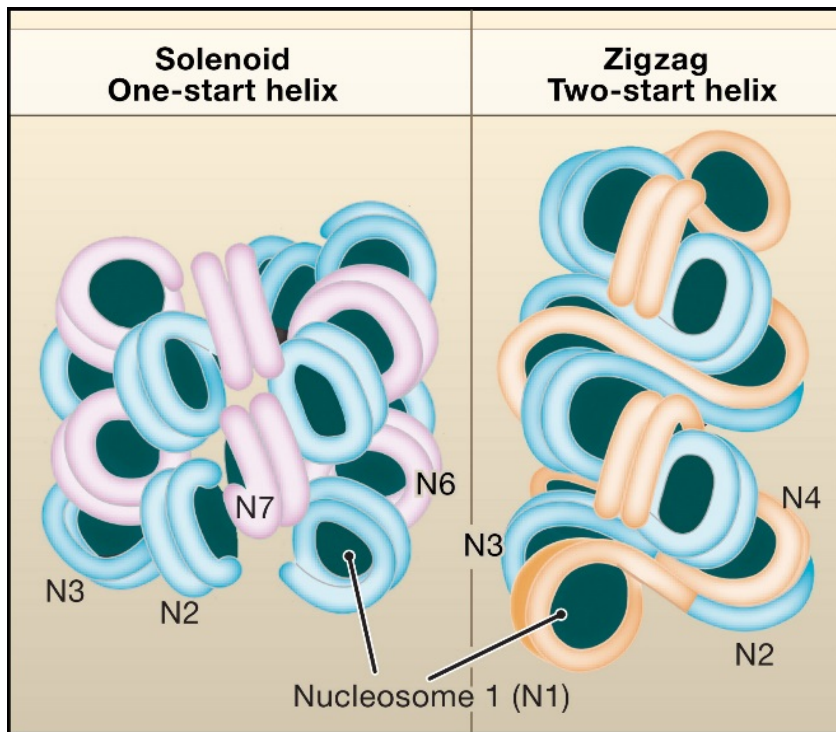
Pictures are modified from (Schalch, Duda *et al.* 2005, Bednar, Garcia-Saez *et al.* 2017).

A crystal structure of an oligonucleosome with four nucleosomes has revealed the arrangement of a tetranucleosome in which nucleosomes are arranged in two stacks connected by three linker DNAs (LB, LB' and LS) (Figure 3). The nucleosome pairs N1, N2 and N1', N2' are rotated with respect to each other along two-fold axis and the two nucleosomes of each stack are interacting with each other through their octameric histone surface. (Schalch, Duda *et al.* 2005).

#### 1.1.1.2. Higher order chromatin organization

The compact 11 nm nucleosomal arrays are further organized into more compact 30 nm chromatin fibers which are considered as the secondary structure of chromatin. The 30 nm fiber shows different conformations such as solenoid helix, twisted ribbon and cross linker

conformations, and is observed in two types of arrangements in the chromatin structure (McGhee, Nickol *et al.* 1983, Woodcock, Frado *et al.* 1984, Zhu and Li 2016). In the one start solenoid structure, consecutive nucleosomes are next to each other and the linker DNA is bent towards the fiber center. In the two start solenoid structures, the consecutive nucleosomes are arranged in a zig zag manner to form two rows of nucleosomes, and the linker DNA is linear (Figure 4). In this latter case, the alternate nucleosomes are the interacting partners. This model further produce coiling and twisting to generate variants of zig zag model (Tremethick 2007).

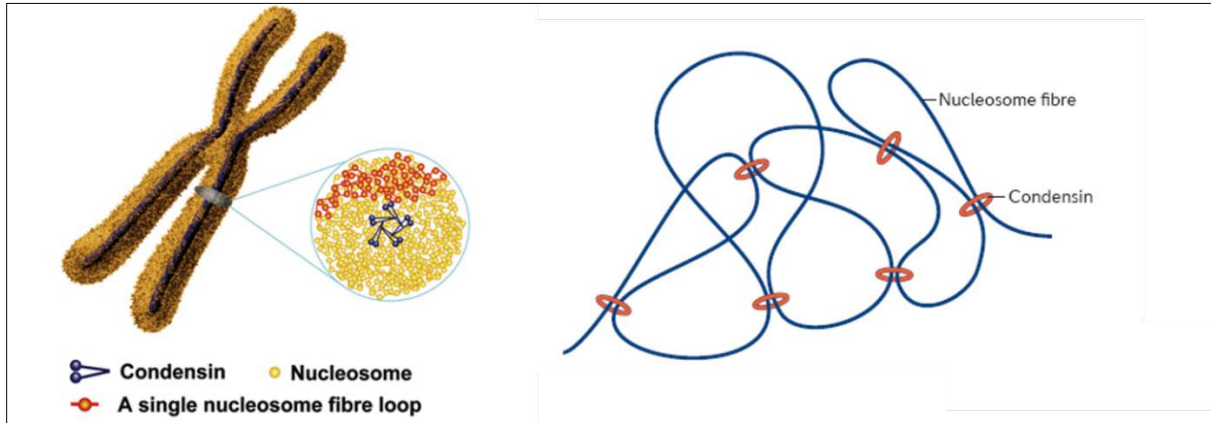


**Figure 4: Models of higher order nucleosome organizations.**  
Adapted from (Tremethick 2007).

Recent advancements in chromatin structure research suggests that the existence of 30 nm chromatin fibers is however controversial. Biophysical and structural (cryo EM, SAXS and ultra SAXS) approaches did not find evidence for the classical 30 nm chromatin scaffolds, and no structural repeating unit that was larger than the 11 nm nucleosome fibers was observed (Nishino, Eltsov *et al.* 2012).

Yet, at the intersections of 11 nm structures, the condensin complex was observed that has been known to help in chromatin compaction. Condensin is a multi-protein complex that plays an important role in the chromatin assembly and segregation during cell cycle (Hirano 2016). This

led to the alternative model that condensin organize chromatin into an arrangement of 11 nm fibres. Condensin is majorly concentrated around centromeres and along the arms of chromosomes. Condensin provides stability to the centromeric chromatin; condensin inactivation leads to loss of sister chromatid separation during anaphase and loss of chromatin compaction (Figure 5).



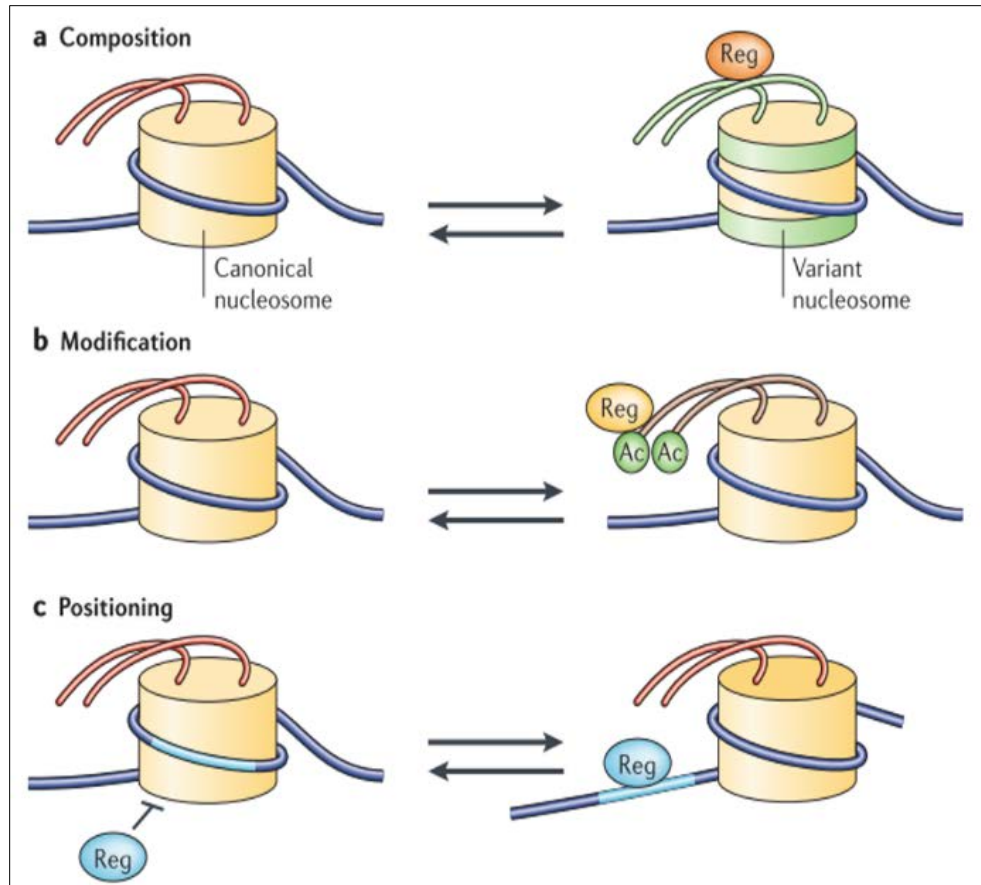
**Figure 5: Condensin in chromosome condensation - The beads on a string model:**

Irregularly folded nucleosome fibres (Hu, Chen *et al.*), held by condensin (blue) resembles ‘beads on a string’. On right representation of chromosome, held by condensin. Pictures adapted from (Nishino, Eltsov *et al.* 2012, Uhlmann 2016).

## 1.1.2. Epigenetic mechanisms

### 1.1.2.1. Chromatin structure modulation through epigenetic mechanisms

The nucleosome and its higher order organizations are highly dynamic structures that can be manipulated by different means (Figure 6). Epigenetic mechanisms are responsible for modifying the chromatin structure in a coordinated manner to help the cell fulfil its nuclear functions. Chromatin modulation mechanisms can be grouped into five major classes: (i) reversible post-translational histone modifications (Suzuki, Muto *et al.*), (ii) ATP-dependent chromatin remodeling, (iii) replacement of canonical histones by histone variants, (iv) long non-coding RNAs (lncRNAs) and (v) DNA methylation.



**Figure 6: Dynamic properties of nucleosomes.**

a) exchange of core histones with histone variants; b) chemical modifications of nucleosome – PTMs of histone; c) repositioning of nucleosomes – sliding over DNA to uncover DNA; Reg – any regulatory proteins; Ac – acetyl group attached to histone tails. Picture adapted from (Saha, Wittmeyer *et al.* 2006).

## 1.1.2.2. The histone epigenetic marks

### 1.1.2.2.1. Post-translational modifications

Histones harbor specific residues that are the target of different reversible post-translational modifications (PTMs). So far, these include methylation, acetylation, phosphorylation, glycosylation, carbonylation, ubiquitylation, biotinylation, sumoylation, citrullination, ADP-ribosylation, N-formylation, crotonylation, propionylation, butyrylation and proline isomerization. According to the Brno nomenclature of European laboratories, the histone modifications are represented with the histone name, followed by the amino acid, followed by the position of the amino acid, followed by the modification type, and finally followed by the degree of modification

(for example H3K4me2 represents a dimethylation of lysine 4 from histone H3) (Sadakierska-Chudy and Filip 2015).

Interestingly, a few sites can undergo different modifications such as H3K4 that can be acetylated or methylated, which enables the use of methyl-acetyl switches (Guillemette, Drogaris *et al.* 2011). In addition, not only single PTMs but also combinations of different PTMs can exert different regulatory functions, as described in the histone code hypothesis (Turner 2000, Jenuwein and Allis 2001). Some of the best characterized histone PTMs that have been identified so far are listed in (Table 1).

The PTMs can change the interaction affinity of histones for DNA and of nucleosomes for other nucleosomes, hence helping reversibly change the chromatin organization between compact and open conformations. But PTMs can also be recognized by epigenetic and nuclear effectors, helping to recruit these effectors at specific loci. This has of course a direct effect on the kind of nuclear activity that will be performed at the PTM-marked genomic locations.

**Table 1: Table showing important histone modifications and their function**

<b>Histone</b>	<b>Amino acid position</b>	<b>PTM</b>	<b>Function</b>
H1	S27	Phosphorylation	Transcriptional activation
	K26	Methylation	Transcriptional silencing
H2A	S1	Phosphorylation	Mitosis, chromatin assembly, transcriptional repression
	K4, K5, K7, K36, K119	Acetylation	Transcriptional activation
	K119	Phosphorylation	Spermatogenesis
	K119	Ubiquitination	Transcriptional repression
	K99	Methylation	Chromatin remodeling
	Q105	Methylation	Chromatin remodeling
H2B	S14, S33	Phosphorylation	Apoptosis, Transcription activation
	K5, K11, K12, K15, K16, K20, K82, K105, K113, K117	Acetylation	Transcriptional activation
	K120	Ubiquitination	Spermatogenesis, meiosis, Transcriptional activation
	K40	Methylation	Unknown
	R96	Methylation	Unknown
H3	K79	Mono methylation	Telomeric silencing, cellular development, cell-cycle checkpoint, DNA repair, and regulation of transcription
	K4	Di methylation	Euchromatin formation
	K4, K9, K17, K27, K36, K56, K64	Tri methylation	Transcription regulation, active euchromatin, X-chromosome inactivation
	R17, R42, R53	Methylation	Transcriptional activation
	K4, K9, K14, K18, K23, K27, K56, K64, K115, K122	Acetylation	DNA repair, transcriptional activation, histone deposition
	T3, T11, T45, T118	Phosphorylation	Mitosis
	S10, S28	Phosphorylation	Mitosis, meiosis, transcriptional activation
	Y41	Phosphorylation	Transcription activation
H4	R3, R92	Mono methylation	Transcription activation
	K20, K59	Mono methylation	Transcription silencing
	K20	Tri methylation	heterochromatin
	K5, K8, K12, K16, 31, K77, K79, K91	Acetylation	Histone deposition, DNA repair, telomere silencing, transcriptional activation
	S1, S47	Phosphorylation	Mitosis

Importantly, the effect of PTMs on histones often varies depending on the type of chemical group added and on the residue modified. For instance, some modifications are associated with transcription repression, whereas others are associated with transcription activation. In the case of acetylation, all four histones can be modified. Acetylation removes the charge from the modified lysine side chain. This has for effect to decrease histone/DNA and histone/histone interactions between neighboring nucleosomes, leading in general to a decompaction of the chromatin into active euchromatin (Zentner and Henikoff 2013, Ma and Zhang 2016). Yet, the acetylation mark can also be specifically recognized. Acetylation of histones is typically carried out by Histone Acetyltransferases (HATs) whereas deacetylation is performed by Histone Deacetylases.

Methylation takes place on lysine and arginine side chains. In the case of lysine, these can be mono-, di-, or tri-methylated, whereas arginines can be mono- and symmetrically or asymmetrically di-methylated. Unlike acetylation, methylation does not change the charge of the side chain but increases its hydrophobic character. Methylation is therefore more associated with the recruitments of nuclear effectors, which explains that methylation, depending on the side chain modified and the level of methylation, can be associated with either repression or activation of different nuclear processes (Ng, Yue *et al.* 2009). Methylation of histones is carried out by histone methyltransferases which use S-adenosyl methionine as a methyl donor, and demethylation is performed by histone demethylases.

Histone phosphorylation is also a major PTM and is set by kinases and removed by phosphatases. Phosphate groups are generally added to the hydroxyl group of amino acids like serines, threonines and tyrosines. Phosphorylation alters the charge status of histones and hence modifies protein-DNA and protein-protein contacts between nucleosomes (North, Javaid *et al.* 2011). But phosphorylation also helps recruit different nuclear effectors.

Many other PTM have been characterized, including small proteins such as ubiquitin and SUMO. These PTMs also affect the chromatin structure and the recruitment of nuclear effectors. Actually, the list of histone modifications and histone modifying enzymes is continuously increasing, highlighting the complexity of chromatin structure and nuclear regulations.



#### 1.1.2.2.2. Modification of non-histone proteins

Post-translational modifications can be deposited on histones not only in the nucleus but also in the cytoplasm, where they for instance serve to mark specifically new synthesized histones with specific marks. PTMs are in fact not only restricted to histones but are also deposited on various nuclear effectors, even epigenetic players. These modifications can also be deposited in the various cellular compartments. Modification of these different effectors can also alter their function, their recruitment, with direct or indirect effect on chromatin structure modulation.

Therefore, modifications of non-histone proteins are probably as important as the modifications of histone and actually participate to the same regulatory process of chromatin modulation. Yet these modifications and their effect on chromatin remain less understood, but pinpoint once again the levels of complexity that underlie chromatin structure regulation (Wu, Connolly *et al.* 2017).

Importantly, homologous enzymes that are responsible for setting or removing PTMs can act either on histones or on non-histone proteins. In some cases, these enzymes may target both histones and non-histone proteins (Glozak, Sengupta *et al.* 2005). As discussed later, this might be a problem for the design of drugs targeting these enzymes since these drugs can non-selectively inhibit a whole range of enzymes that not only act on histones but may also act on other cellular processes.

#### 1.1.2.2.3. Writers, readers and erasers of the epigenetic marks

Epigenetic enzymes that add and remove post-translational modifications are designed as writers and erasers, respectively. Specific domains that recognize the post-translational modifications are designed as readers. As mentioned above, writers, readers and eraser have a whole range of targets. In addition, the list of histone modifying enzymes and of readers is very large. Therefore, in the following sections, only important and well characterized classes of enzymes are discussed: histone acetyltransferases and deacetylases, histone methyltransferases and demethylases, and kinases and phosphatases. Readers recognizing the acetylation and methylations marks will also be described.

#### 1.1.2.2.4. Histone acetyltransferases and histone deacetylases

Histone acetyltransferases (HATs) are important epigenetic enzymes that catalyze the addition of acetyl group to the lysine residues of histones (and non-histone proteins) using acetyl-CoA as cofactor. HATs are found in almost all organisms and they are usually part of large protein complexes. And in some cases, catalytic domains of HATs are shared among different protein complexes. HATs can be broadly classified into GCN5, MYST, SRC, p300/CBP families. Acetylation of histone tails decreases compaction of nucleosomes by decreasing the interaction between lysine and DNA, and result in the increased expression levels. The opposite mechanism of HATs is carried out by histone deacetylases i.e. removal of acetyl group. A detail explanation of HATs and HADCs is given in the section (1.2 & 1.3).

#### 1.1.2.2.5. Histone methyltransferases and histone demethylases

A large number of epigenetic enzymes participate in either addition or removal of methyl group from DNA, RNA and from proteins. The huge number of targets of protein methyl transferases (PMTs), including non-histone proteins, makes them the largest class of epigenetic modifying enzymes. Lysine and Arginine, two basic amino acids that can form hydrogen bonds with the DNA backbone and bases, can be methylated. In arginine, the guanidino group can contribute five hydrogen bonds and the lysine  $\epsilon$  amino group can contribute two hydrogen bonds. Each added methyl group can replace a hydrogen bond and increases the residue hydrophobicity, which can alter the nucleosome structure (Bedford and Clarke 2009, Jahan and Davie 2015).

Based upon target amino acid, PMTs can be divided as lysine methyl transferases and arginine methyltransferases. Lysine methyltransferases can add mono, di and trimethyl groups to lysine side chains, whereas arginine methyltransferases add mono and symmetric or asymmetric dimethyl groups to arginine side chains. All methyltransferases transfer a methyl group from the S-adenosyl methionine (Lahm, Paolini *et al.*) cofactor, and release S-adenosyl homocysteine (SAH) as by-product.

A group of lysine methyl transferases contain a catalytic SET [SU(VAR)3-9, E(Z) and Trithorax] domain, and additionally Immunoglobulin SET (I-SET) domain and post-SET domains are present which contribute to substrate binding. Another group are non-SET domain methyl transferases which use a DOT1L domain for the methyl transferase activity. SET domain-

containing enzymes participate in the methylation reactions of histone tails whereas DOT1L domains methylates core histones (Wood and Shilatifard 2004, Boriack-Sjodin and Swinger 2016).

So far eleven PRMTs have been identified which are grouped into four different classes. Type I PRMTs (1,2,3,4,6,8) participate in  $\omega$ -NG-mono- and asymmetric di-methylation reactions, the type II (PRMT5) enzyme participate in mono- and symmetric di-methylation, and the type III (PRMT7) enzyme only participates in the mono methylation. Recently a type IV PRMT has been found in yeast, which has a functional role in the catalysis of  $\delta$ -N-methyl arginine. PRMTs possess a conserved methyltransferase catalytic domain, which is accompanied by additional modulatory domains like SH3, zinc finger and TRP2 domains. PRMTs contain a conserved MTase domain which is a Rossmann fold and helps in the SAM binding, followed by a PRMT unique  $\beta$ -barrel domain and a dimerization domain (Boriack-Sjodin and Swinger 2016). PRMTs are found in association with protein partners like Blimp1, RioK1, pICLn, MEP50 and MBD/NuRD.

Protein demethylation occurs in two different pathways. In one set, the enzymes use FAD dependent amine oxidation and are known as FAD amine oxidases. The second group of enzymes depends on the Fe(Suzuki, Muto *et al.*) and  $\alpha$ -ketoglutarate and those are known as dioxygenases.

Histone methylation is a thermodynamically very stable post-translational modification which is removed by different enzymatic mechanisms viz.: oxidative cleavage, dioxygenase activity and demethylination of methyl arginine etc. Lysine demethylases have been classified into seven groups KDM1-7. Except for KDM1, all other KDM2-7 contains a Jumonji-C (JmjC) domain. KDM1 contains an amine oxidase domain that requires FAD for the enzymatic activity. KDM1 can only demethylate mono- and di-methyl lysines, because it requires a protonated nitrogen. KDM1 associates with Co-REST transcription co-repressor complex. JmjC proteins can demethylate trimethylated lysine residues. All JmjC proteins requires iron and  $\alpha$ -ketoglutarate for activity. Peptidyl arginine deiminase enzymes depend on calcium for their enzymatic activity that convert methylated arginine to citrulline. JMJD6 is a dioxygenase that shares sequence homology with JmjC and asparaginyl hydroxylase and has been identified as arginine demethylase.

### 1.1.2.2.6. Kinases and phosphatases

Histone phosphorylation is a modification which changes the affinity of other epigenetic regulators for the DNA, and is induced by the extra cellular signals like DNA repair or onset of mitosis. In eukaryotes, several protein kinases and phosphatases are known and classified based upon their catalytic domain and amino acid substrate (Table 2). In H3 four phosphorylation sites have been identified which participate in chromatin compaction. T3, S10, T11 and T28 are the phosphorylation sites in H3, where S10 is the most studied mark which has a role in the mitosis and meiosis. H3S10 is phosphorylated by several enzymes which includes aurora B kinase, MSK1, RSK2, PRK1 etc. Aurora B kinase is an important enzyme that participates in the regulation of mitosis and also in chromatin condensation. H3S10ph is recognized by 14-3-3 proteins which participate in transcriptional activation (Rossetto, Avvakumov *et al.* 2012, Sawicka and Seiser 2014). Phosphorylation of H2A(X) variant is important for DNA damage repair, which is carried out by yTel1 and yMec1 two kinases. Dephosphorylation of H2AX carried out by HTP-C phosphatase is required to exchange.

**Table 2: Important phosphorylation sites and known kinases (Rossetto, Avvakumov *et al.* 2012).**

Histone	Amino acid number	Kinase
H1	S/T	CDK2
H2A	T120	Bub1
	S139 (H2AX)	ATR, Mec1, DNA-PK
	Y142 (H2AX)	WSTF
	S16	RSK2
H2B	S32	RSK2
	S36	AMPK
	Y37	WEE1
	T3, T6	Haspin, PKC $\beta$
	S10	PRK1, MSK1, RSK2
H3	T11, 45	Dlk, PRK1, PK-C $\delta$
	S28,	AuroraB, ERK1, MSK1
	Y41	JAK2
H4	S1	Sps1
	S47	PAK2

#### 1.1.2.2.7. Readers domains of the epigenetic marks

Epigenetic readers are proteins or protein domains that can recognize the post transitional mark on the histones and the DNA. Bromodomains recognize acetyl modifications on histones. Several HATs such as P300, CBP and GCN5 contain bromodomains to facilitate chromatin remodelling and transcription. P300 acetylates the Tat protein which is essential for HIV-1 genome transcription. Acetylated Tat is recognized by the bromodomains of PCAF and promotes HIV gene transcription by recruiting RNA pol II. H4K8 and H3K9,K14 are acetylated by GCN5 which enables to recruit several co-activator complexes like SWI/SNF and TFIID for transcription regulation.

The list of methyl identifying readers is big, this group includes Tudor domain, Agenet, Chromo, PWWP, Malignant Brain Tumor, Plant Homeo Domains and WD40 repeat containing proteins (Zhang, Cooper *et al.* 2015, Greschik, Schule *et al.* 2017). All methyl identifying readers contains a methyl lysine recognition pocket to facilitate the recognition of methylated lysine through different interactions (Herold, Wigle *et al.* 2011). Few epigenetic complexes contain more than a single reader domain which raises the complexity in the recognition mechanism (Ruthenburg, Li *et al.* 2007).

Epigenetic readers contain specific architecture that can recognize histone modifications, and are categorized into four classes. Chromatin architectural proteins, chromatin remodellers, chromatin modifiers and adaptors. Chromatin architectural proteins inhibits transcription by blocking DNA for polymerase binding and induce chromatin compaction. Chromatin remodellers also possess reader domains which have several functions in regulating chromatin structure and gene expression. Many co-repressor complexes contain reader domains that result in the PTM of histones, for example Sin3 a co-repressor that recognizes histone acetylation through its tandem bromodomain and finally deacetylase by HDACs. Adaptors help in the recruitment of factors that help in DNA metabolism and functions transcription replication DNA damage repair etc.

#### 1.1.2.3. ATP-dependent chromatin remodellers

The second type of enzymes that regulate chromatin structure is ATP-dependent chromatin remodellers. Remodellers use the energy of ATP to act on nucleosomes by evicting or sliding

them, or by destabilizing the nucleosomes to help with histone pair exchange or other nuclear processes. Remodellers play important roles in the organization of DNA, transcription regulation, chromosome segregation, DNA recombination, replication and DNA repair. As for many other epigenetic effectors, remodellers maintain the balance between genome packaging and genome access.

Specifically, some remodellers help with the compaction or the proper spacing of nucleosomes, while others act in transcription by exposing DNA sequence to epigenetic regulators. During replication and transcription, the advancing polymerases are followed by nucleosome reconstruction machineries. The DNA unwraps and binds simultaneously in a rapid rate and undergoes conformational changes to facilitate transacting elements binding. Remodellers help in positioning the nucleosomes at functionally required locations.

So far, four major classes of remodellers have been identified. All of them possess an ATPase domain formed of two RecA-like structural domains to perform ATP hydrolysis. Each family also possesses unique auxiliary domains to perform specialized functions (e.g. epigenetic mark recognition, interactions with transcription factors, etc...).

The four families of remodellers were classified on the basis of sequence similarities in their catalytic ATPase domain and are: the SWI/SNF (switch/sucrose non-fermentable) family, the ISWI (imitation switch) family, the CHD (chromodomains helicase DNA binding) family, and the INO80 family. In addition, ATRX and Cockayne Syndrome group B (CSB) complexes are considered as orphan remodellers.

Remodellers often act as multi-protein complexes. SWI/SNF family remodellers possess 8 to 14 subunits which includes ATPase, actin related proteins (ARP), hBAF55/170 and multiple bromodomains. SWI/SNF family remodellers slide and eject nucleosomes during remodeling but lack a role in chromatin assembly. ISWI family members possess 2 to 4 subunits, with one or two catalytic subunits. In the ISWI family, the SANT/SLIDE domains form a module that recognizes free DNA emerging from nucleosomes. Additionally, plant homeodomain, bromodomains and DNA binding motifs are present in this family. ISWI complexes optimize nucleosome spacing, promote chromatin assembly and repress transcription. CHD (chromodomains helicase and DNA binding) family remodellers possess 1 to 10 subunits with two tandemly arranged chromodomains at the N-terminus of catalytic subunit. Few CHD complexes promote transcription while others

repress it. The vertebrate Mi-2/NURD complex from the CHD family contains histone deacetylases HDAC1 and 2 and harbor a methyl CpG binding domain. The INO80 family contain more than 10 subunits, and includes well characterized members SWR1, SRCAP and p400. INO80 participates in transcription activation and DNA repair. SWR1 replaces H2A-H2B dimer with the H2A.Z-H2B variant. Table 3 recapitulates the various remodellers families.

**Table 3: Chromatin remodelling complexes in humans.**

Remodellers	Complex	ATPase domain	Non-catalytic subunits
<b>SWI/SNF</b>  Tumour suppressor, differentiation, development, elongation, signalling, splicing	BAF	hBRM	BAF250, BAF155, BAF170, BAF60a, hSNF5, BAF57, BAF53a/b, $\beta$ -actin
	PBAF	BRG1	BAF180, BAF200, BAF155, BAF170, BAF60a, hSNF5, BAF57, BAF53a/b, $\beta$ -actin
	NURF	SNF2L	BPTF, RbAp46/48
<b>ISWI</b>  Elongation, Pol II repression, replication, X-chromosome regulation, cohesion, embryonic development, differentiation	CHRAC	SNF2H	hACF1, hCHRAC17, hCHRAC15
	ACF	SNF2H	hACF1,
	INO80	hIno80	RuVBL1,2, BAF53a, Arp5,8, hles2,6
	SRCAP	SRCAP	RuVBL1,2, BAF53a, Arp6, GAS41, DMAP1, YL-1, H2AZ-H2B, ZnF-HIT1
<b>INO80</b>  DNA repair, homologous recombination	TRRAP/Tip60	p400	RuVBL1,2, BAF53a, Actin, GAS41, DMAP1, YL-1, Brd8, TRRAP, Tip60, MRGX, MRG15, MRGBP, FLJ11730, EPC1, EPC-like, ING3
	SWR1 (yeast)	Swr1	Rvb1,2; Arp6; Yaf9, Swc4/Eaf2; Swc2/Vps72, Bdf1, H2AZ-H2B, Swc6/Vps71, Swc3,5,7
<b>CHD</b>  Transcriptional repression and silencing	CHD1	CHD1	-
	NURD	Mi-2 $\alpha$ /CHD3, Mi-2 $\beta$ /CHD4	-

#### 1.1.2.4. Histone variants and histone chaperones

Histones are very slowly evolving proteins. Apart from the canonical histones (H1, H2A, H2B, H3 and H4), several variants have been identified with distinct functions. Variant histones can differ from the canonical histones by only a few residues or even large structural domains. Canonical histones are synthesized during S phase and assembled during DNA synthesis behind the replication fork and during DNA repair. Canonical histone genes are in multiple copies and the transcripts lack polyA tails. In contrast, histone variants are synthesized throughout the cell cycle and are assembled independently of DNA synthesis. Histone variants are non-allelic single gene expressions where the transcripts possess introns and tails are poly adenylated (Talbert and Henikoff 2017).

Up to now, seven H1 variants have been identified (H5, H1.10, ooH1.8, scH1, TSH1.6, TSH1.7 and TSH1.9). For H2A, eight variants (H2A.1, H2A.B, H2A.L H2A.P, H2A.W, H2A.X, H2A.Z and macroH2A) are known, for H2B five (H2B.1, H2B.W, H2B.Z, sperm H2B and subH2B), and for H3 six (CENP-A, H3.3, H3.5, H3.X and TSH3.4). So far, no variants were identified for H4 in mammals. Recently a testis specific H4 variant (tH4) has been identified in trypanosomes and urochordates (Moosmann, Campsteijn *et al.* 2011, Buschbeck and Hake 2017). Each variant has distinct functional property from its canonical histone.

Incorporation of histone variants changes the properties of nucleosomes and alters its interactions with remodellers and histone modifiers. Nucleosomes with histone variants such as H2A.Z and H3.3 are important to maintain a less compact region around active gene promoters, which facilitate the binding of transcription factors (Jin, Zang *et al.* 2009). Other examples include, H2A.X that participates in DNA repair, CENP-A which acts in kinetochore assembly, and testis specific H2B (TSH2B) that helps convert chromatin into nucleoprotein during spermatogenesis.

Interestingly, due to the fact that each histone is present twice in the nucleosome, in some cases canonical and variant histones can be present in the same nucleosome (referred as heterotypic nucleosome in contrast to the homotypic nucleosome that contains only one kind of histone from a same family). These nucleosomes are often less compact than the homotypic nucleosomes and can therefore also have specific functional roles.

Finally, most of the histone variants can also undergo post translational modifications (PTMs) either similar or different to the canonical histones depending on their degree of similarity.



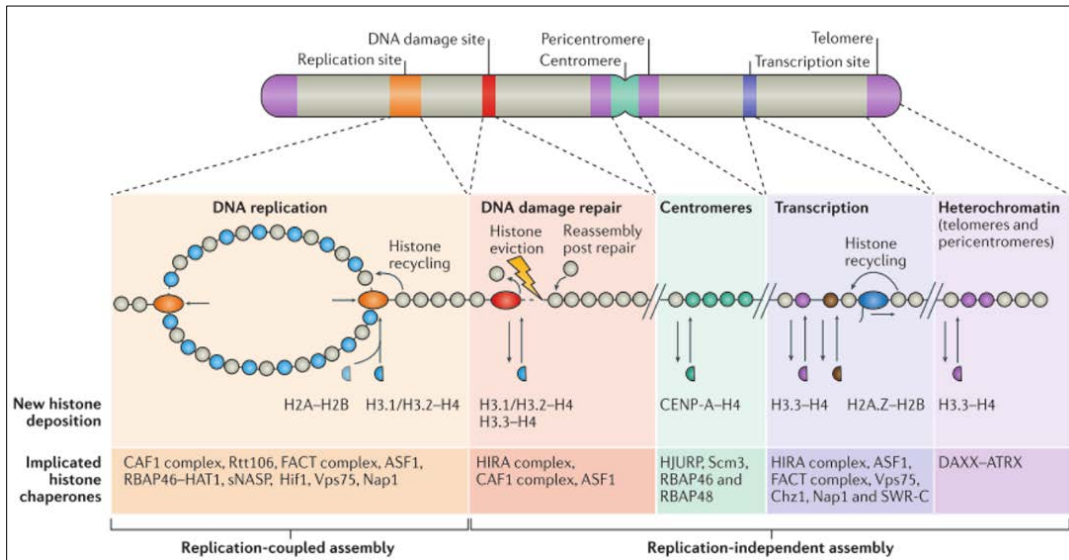
A notable case is H2A.X which possesses a specific phosphorylation site in its C-terminal tail which is used for marking DNA regions requiring DNA repair. Table 4 lists histone variants with their function and the epigenetic effectors they are interacting with.

**Table 4: List of histone variants and their functions (Buschbeck and Hake 2017).**

Histone	Chaperone or remodellers	Function
H2A-H2B	FACT	Deposition and exchange of histones
	NAP1	Nuclear import and deposition
	Nucleolin	Deposition and exchange of histones
H2A.X	FACT	Deposition and exchange
H2A.Z.1	P400, SRCAP	H2A.Z deposition
	ANP32E, INO80	H2A.Z removing
H2A.Z.2.1 and H2A.Z.2.2	P400, SRCAP	H2A.Z deposition
MacroH2A1.1 and Macro H2A2	Unknown	Unknown
MacroH2A1.2	ATRX	Negative regulation
H2A.Bbd	SWI-SNF	Remodeling
TH2B	Unknown	Unknown
H3.1-H4	FACT,	Deposition and exchange
	NAP1	Nuclear import and deposition
	ASF1	Histone transport
	NASP1	Histone protection
	CAF1	Histone deposition and tetramer formation
CENP-A	HJURP	Deposition at centromeres
H3.Y	unknown	Regulation of cell cycle genes
H3.X, H3.1t, H3.5	unknown	unknown

Since the histones (canonical and variant) are shaping the chromatin and play active functional roles, it is essential for the cell that these histones are deposited at precise locations where they will fulfil their roles. For this reason, a large number of epigenetic effectors known as histone chaperones are accompanying histones, from their synthesis in the cytoplasm to their deposition onto the chromatin. Histone chaperones facilitate histone deposition/removal for

nucleosome assembly/disassembly, histone post-translational modifications, histone folding and oligomerization, as well as histone nuclear import, stability and genomic localization.



**Figure 7: Histone chaperones and histone deposition mechanism.**  
 Picture adapted from (Hammond, Stromme *et al.* 2017).

### 1.1.2.5. Long non-coding RNAs

Recent sequencing tools have identified several RNA molecules that do not participate in translation which are termed as non-coding RNAs (ncRNAs). From the whole genomic DNA of a cell, only one percent of DNA is transcribed into RNA in which 75% are ncRNAs. Three main types of RNAs exist: mRNA participates in translation, and tRNA and rRNA possess housekeeping functions. Apart from these, several ncRNAs have been identified with functional role but lacks translational ability.

ncRNAs are divided into two major groups. Small RNAs of nearly 20-30 nt in length which includes miRNA, piRNA, and endogenous siRNA. The second group is long non-coding RNAs (lncRNAs) which are more than 200 nt in length. miRNAs and siRNAs are not directly involved in chromatin regulation, however there are few reports which indicate that the expression of small ncRNAs can alter the chromatin state by restricting activities of few a remodellers (Denis *at al* 2011). lncRNAs can be associated with remodellers and recruit them on genomic loci to alter DNA methylation or histone modifications.

More than 200,000 lncRNAs have been identified so far. lncRNAs are associated with many biological processes such as DNA damage, immune response, induced pluripotency etc. (Guttman 2009). H19, Xist, AIR, HOTAIR and ANRIL are few lncRNAs which were identified to have functional roles in heterochromatin formation and imprinting. H19 is the first identified lncRNA which is essential for cell proliferation and tumorigenesis and has a major role in imprinting. Xist is another lncRNA which participate in X-chromosome inactivation.

lncRNAs are classified into four classes based on their expression direction, (i) antisense lncRNAs which are transcribed in opposite direction to protein coding genes, (Suzuki, Muto *et al.*) intronic lncRNAs which are present in introns of protein coding genes (iii) bidirectional lncRNAs which starts at promoter region in a reverse direction and (iv) intergenic lncRNAs which are present in intergenic regions. lncRNAs are present in both nucleus and cytosol where they participate in chromatin regulation and protein stability, respectively.

lncRNAs can act in both cis and trans. lincRNA-p21 is activated by p53 during DNA damage. Upon activation lincRNA-p21 functions as transcriptional repressor by recruiting hnRNPK to the targeted genes which is a RNA binding protein that belongs to heterogeneous nuclear ribonucleoproteins (hnRNP). The mechanism suggested behind this repressive mechanism is that lincRNA-p21 recruits SETDB1-H3K9 methyl transferase and DNMT1 to the promoters and the methylation of the target genes result in transcription repression. In this manner lincRNA-p21 participate in both DNA methylation alteration and histone modification of chromatin.

HOTAIR is a well characterized lncRNA that binds histone modifying enzyme complexes PRC2 and KDM1, thereby participating in H3K27 methylation and H3K4 demethylation, and also participates in the silencing of the HOX D locus. In breast cancer HOTAIR directs PRC2 to promoters of metastasis suppressor genes and silence the transcription. These are only few examples of several lncRNAs that regulates chromatin stability and gene expression.

### 1.1.3. Epigenetic mechanisms in diseases and therapy

Epigenetic mechanisms are highly coordinated to regulate chromatin structure and maintain cell homeostasis. Dysregulation of these mechanisms can result in cell death, but also in aberrant activation or repression of many different nuclear mechanisms. This is why an ever-

increasing number of diseases are being linked with deregulation of epigenetic mechanisms. Cancer is one of them, but the regulatory role of epigenetic effectors in almost any nuclear mechanisms impacts most of the cellular processes, resulting in a wide variety of other diseases like autoimmune disorders, neurological disorders, metabolic disorders or cardiovascular diseases.

### 1.1.3.1. Epigenetic mechanisms in autoimmune diseases

Several autoimmune diseases have been shown to be due to deregulation of epigenetic mechanisms. ICF (immunodeficiency centromeric instability and facial anomalies) syndrome is one of the best known auto immune disease. It is caused by heterozygous mutations in DNA methyltransferase DNMT3B. This leads to hypomethylation of several genes such as pericentromeric satellite 2 and 3 repeats, alpha satellite sequences, *Alu* sequences and the D4Z4 and NBL2 repeats (Jin, Tao *et al.* 2008, Portela and Esteller 2010). DNA methylation defects also causes systemic lupus erythematosus (Grant, Easley *et al.*) and rheumatoid arthritis. Hypomethylation of PRF1, CD70, CD154, and also 18S and 28S rRNA gene promoters leads to SLE (Javierre, Fernandez *et al.* 2010).

Acetylation of histones regulates the expression of CD154, IL10 and IFN $\gamma$  which are disturbed during rheumatoid arthritis and SLE. HDAC inhibitor TSA was shown to stimulate the expression pattern of these genes during SLE and rheumatoid arthritis. HDAC activity regulates transcription factor NF- $\kappa$ B mediated gene expression. NF- $\kappa$ B binding to DNA is enhanced by several modifications of histones such as acetylation of H3K9 and phosphorylation of H3S10, leading to reduction in H3K9 methylation and increase in H3/H4 acetylation (Vanden Berghe, Ndlovu *et al.* 2006).

Increased H3K9me2 in lymphocytes but not in monocytes is associated with inflammatory pathways and has a role in type I diabetes (Miao, Smith *et al.* 2008). Due to apoptosis, nucleosomes are found in the blood circulation and act as antigens, which is observed in SLE. Several PTMs in these nucleosomes, such as phosphorylation of H2BS14 and H3T45, H3K4me3, acetylation of H4 K8, K12, K16 and H2B K12, would result in auto-antibody production (van Bavel, Dieker *et al.* 2011).

### 1.1.3.2. Epigenetic mechanisms in neurological diseases

In neural cells, mitotic exit is an important transition and signifies the loss of multipotency and the start of neural development which is regulated by epigenetic enzymes. Defect in epigenetic mechanisms leads to neurodevelopmental and neurological diseases. MeCP2 (methyl CpG binding protein 2) is a MBD (methyl CpG binding domain) protein that participates in cellular functions like chromatin architecture, mRNA splicing and transcription of several genes. Aberrant expression of MeCP2 is associated with neurodevelopmental defects. MeCP2 recruits HDACs to methylated DNA and acts as gene silencer. Point mutations in MeCP2 causes the Rett syndrome, an X-linked neurological disorder (Chahrour, Jung *et al.* 2008).

Mutations of acetyl transferases CBP and EP300, leads to the autosomal dominant disease Rubinstein-Taybi syndrome, a disease characterized by short stature, learning difficulties and distinctive facial features (Urduingio, Sanchez-Mut *et al.* 2009). RSK2 is a serine/threonine protein kinase, which is an important member of MAP kinase pathway. RSK2 phosphorylates H3S10 and destabilize the compaction of chromatin and enables the recruitment of CBP. Loss of functional mutations of RSK2 leads to the Coffin-Lowry syndrome which is a rare X-linked disorder (Clayton, Rose *et al.* 2000, Portela and Esteller 2010).

ATRX is the catalytic subunit of a multi-protein ATP-dependent chromatin remodeling complex. ATRX participates in many cellular functions such as X-chromosome inactivation, chromosome cohesion, chromosome alignment on meiotic spindle and heterochromatin formation. Mutations in ATRX gene causes the ATRX (Alpha Thalassemia X-linked mental Retardation) syndrome which is an X-linked disorder characterized by intellectual disabilities, microcephaly (small head size), and distinctive facial features (Portela and Esteller 2010).

In addition to neurodevelopmental disorders, several neurological disorders have been identified in humans which are mainly associated with post-translational modification of histones such as methylation and acetylation. Hyper- or hypo-methylation states are associated with neurological diseases. CGG trinucleotide repeat expansion in 5'-UTR sequence of the FMR1 gene promotes methylation of FMR1 and subsequent transcriptional silencing. This hypermethylation of FMR1 promoter causes the Fragile X syndrome.

Similarly, hypermethylation of neprilysin, FXN, and SMN2 promoters results in Alzheimer's disease, Friedreich's ataxia and spinal muscular atrophy respectively (Gheldof,

Tabuchi *et al.* 2006, Urdinguio, Sanchez-Mut *et al.* 2009). In patients suffering from Parkinson's disease, due to hypomethylation substantia nigra, TNF $\alpha$  is overexpressed which lead to the apoptosis of neuronal cells (Pieper, Evert *et al.* 2008). In multiple sclerosis patients, hypomethylation is observed in PADI2 promoter (Murgatroyd, Patchev *et al.* 2009). Aberrant methylation profiles in gene region 15a11-q13 which is a imprinting region, causes Prader-Willi syndrome (Robertson 2005).

Histone hypo acetylation is also associated in neurological diseases. Amyotrophic lateral sclerosis (ALS) is caused by hypoacetylation because CBP HAT activity is inhibited by the FUS protein which is a part of misfolded cytoplasmic aggregate. Hypoacetylation is also observed in Parkinson's disease and Huntington's disease and Friedreich's ataxia. Hypoacetylation of H3/H4 is accompanied by hyper trimethylation of H3K9 in FXN alleles, associated in Friedreich's ataxia (Herman, Jenssen *et al.* 2006).

### 1.1.3.3. Epigenetic mechanisms in cancer

Cancer is very often associated with dysregulation of epigenetic mechanisms. Histone modifications, DNA methylation at global level and chromatin modifying enzyme profiles contribute to tumorigenesis and tumor progression.

#### 1.1.3.3.1. DNA Methylation in cancer

In cancer cells hypomethylation and hypermethylation are frequently observed. Repetitive sequences which promotes chromosomal instability, translocations, gene disruption and reactivation of endoparasitic sequences, all contribute to the hypomethylation. Hypomethylation of L1 which belongs to the LINE (Long interspersed nuclear elements) family, is found to be hypomethylated in a wide range of cancers including breast, lung, bladder and liver cancers (Wilson, Power *et al.* 2007). S100P, SNCG and melanoma-associated gene (MAGE) and dipeptidylpeptidase6 genes are respectively hypomethylated in pancreatic cancer, breast/ovarian cancers, and melanomas (Wilson, Power *et al.* 2007).

Methylation profiles at distinct promoters can behave differently. MASPIN a tumour suppressor gene is hypermethylated in breast cancer but hypomethylated in other type of cancers

(Futscher, O'Meara et al. 2004). A well-studied hypomethylation and loss of imprinting is due to insulin-like growth factor 2 which is associated with different types of cancers such as breast, liver, lung and colon cancers. Aberrant methylation patterns can be due to the abnormalities in DNA methyltransferases (DNMTs) expressions or recruitment of DNMTs and HDACs to the specific target genes.

In addition, genes that are involved in many cellular pathways such as DNA repair, Ras signalling, cell cycle control, p53 and apoptosis are subjected to hypermethylation in certain cancer types. Apart from CpG islands, aberrant methylation of DNA in CpG island shores have been associated in cancer studies. For example, for the HOXA2 and GATA2 genes (Doi, Park et al. 2009). Hypermethylation at miRNA promoters is observed in many cancer types and also associated with the development of metastasis. Silencing of miR-148, miR34b/c and miR-9 genes by hypermethylation results in tumour dissemination from original locations (Lujambio, Calin et al. 2008).

#### 1.1.3.3.2. Post translational modifications of histones in cancer

Deacetylation of H4K16 is a prominent marker in cancer generation. Overexpression or mutations of HDACs and translocation or deletions of HAT genes are main reasons for impaired acetylation profiles. SirT1 upregulation is observed in several cancers and correlates with DNA methylation profiles as it interacts with DNMT1.

Impaired regulation of methylation marks contributes to several types of cancers. Loss of H3K4me3 and H4K20me3 and gain of H3K9me and H3K27me3 are associated with cancer. Inactivation of methyl transferase SETD2 and demethylase UTX result in renal carcinomas. EZH2 a methyl transferase and a subunit of the PRC2 complex is overexpressed in several cancers. EZH2 also interacts with DNMTs and participates in DNA methylation regulation (Varambally, Cao et al. 2008).

Mixed lineage leukaemia protein 1 (MLL1) is a H3K4 methyl transferase that plays an important role in the regulation of genes involved in haematopoiesis and development. In leukemias, MLL proteins cause abnormal patterns of H3K79 and H3K4 methylation which results in altered gene expression patterns of MLL targets (Wang, Lin et al. 2009).

KDM1, JmjC and UTRX methyltransferases are upregulated in prostate cancers. JAK2 a non-receptor tyrosine kinase which induce cytoplasmic signaling cascades as it phosphorylates H3Y41 during cytokine signaling. Due to chromosomal translocations JAK2 is activated in hematological malignancies (Dawson, Bannister et al. 2009).

#### 1.1.3.3.3. Other epigenetic effectors in cancer

Apart from DNA methylation and histone modifications, nucleosome positioning also plays an important role in cancer. All chromatin remodelling complexes are associated with cancer. BRG1 and BRM are subunits of the SWI/SNF family remodellers and acts as tumour suppressors. During small-cell lung cancer BRG1 and BRM gene are silenced (Hang, Yang et al. 2010). Mutations within the SWI/SNF subunit SNF5 have been associated with several cancer types such as sporadic renal rhabdoid tumours and medulloblastomas. Remodellers also correlates with methylation pattern which are associated in cancer. CHD5 is a chromatin remodeling complex and its gene hypermethylation have been associated with cancer (Mulero-Navarro and Esteller 2008).

#### 1.1.3.4. Targeting epigenetic players

Epigenetic effectors represent important drug discovery targets. First, many epigenetic effectors are enzymes whose activity can be modulated. Second, the reversibility of many epigenetic mechanisms can be used to develop therapeutic strategies that counterbalance the effect of epigenetic dysregulations that are causing diseases. This explains why many studies are conducted to develop epigenetic drugs or epidrugs. Actually, small molecule inhibitors against two classes of epigenetic enzymes, DNA methyltransferases and zinc-dependent histone deacetylases, have been approved by the American Food and Drug Administration (FDA). However, many more inhibitors that target a whole range of epigenetic enzymes are currently in clinical trials.

##### 1.1.3.4.1. Inhibitors targeting DNA methyltransferases

Inhibitors of DNA methyltransferases are generally analogues of 5' methyl-cytosine. The two FDA-approved drugs 5-azacytidine and decitabine are two analogues of cytidine and are available for the treatment of the myelodysplastic syndrome. These two inhibitors bind DNMTs and eventually decreases the methylation of CpG islands. 5-azacytidine is an analogue of cytidine and has a nitrogen atom in the place of carbon 5. Upon entry into cells, this drug phosphorylates



and is incorporated into DNA during replication. Hence, rapidly dividing cancer cells are more prone to uptake this drug. During methyltransferase activity the nitrogen at fifth position forms an irreversible DNMT1-aza linkage, which causes the degradation of the DNMT1 enzyme and hence reduces methylation. Decitabine also has a similar action mechanism like azacitidine as it incorporates in DNA and inhibits methyl transferases. Zebularine, also an analogue of cytidine, is less toxic and more soluble in aqueous solutions than previous inhibitors (Orta, Pastor *et al.* 2017). MG98 is an anti-sense oligonucleotide, binds 3'untranslated region of DNMT1 and prevents DNMT1 transcription.

#### 1.1.3.4.2. Inhibitors targeting bromodomain acetylation readers

JQ1 is a selective inhibitor of bromodomain, it inhibits binding of BET (Bromodomain and extra terminal domain) to acetylated lysine residues (Greschik, Schule *et al.* 2017). Inhibition of BET binding to acetylated lysine residues leads to down-regulation of c-Myc, which is implicated in cancers. Another bromodomain inhibitor BET726 binds acetylated recognition pocket in bromodomain which is implicated in neuroblastoma tumor growth.

#### 1.1.3.4.3. Inhibitors targeting HATs and HDACs

Majority of HAT inhibitors are non-selective and they can inhibit a broad spectrum of HATs. Few examples include curcumin a natural product, garcinol, isothiazolones, among others. HDAC inhibitors (HDACis) are the most successful class of epigenetic inhibitors so far and four of them are already FDA-approved and one by the Chinese government. A more detailed explanation about inhibitors of HATs and HDACs will be given in section (1.2.1.4.2).

#### 1.1.3.4.4. Inhibitors targeting protein methyltransferases

Several protein methyl transferase inhibitors have been proposed. BIX-01294, is a inhibitor of lysine methyl transferase (Portela and Esteller 2010). E72, UNC321, UNC0646, UNC0638 are other inhibitors they are selective towards methyl transferases but shows toxic effect in cellular assays (Tian, Zhang *et al.* 2013). These drugs should be improved further.

#### 1.1.4. A structural view of epigenetic targets

##### ***Foreword***

The different epigenetic mechanisms act in a synergistic manner to modulate the structure of the chromatin. One way of integrating the different epigenetic mechanisms is to have several epigenetic functions born by a single effector. This is for instance the case when epigenetic effectors not only bear an enzymatic activity but will also harbour reader domains responsible for the specific targeting of the enzymatic activity to specific epigenetic marks. Another way is to combine multiple epigenetic activities in protein complexes, the various subunits bearing different epigenetic functions.

Actually, both way of combining epigenetic mechanisms are fully compatible and the functional epigenetic subunits in the cell are often large multi-protein complexes with various subunits bearing different epigenetic functions. A major emphasis in epigenetic research is to decipher how these complexes are assembled and how the different epigenetic functions collaborate within these complexes.

Structural biology brings a lot of information towards this goal and seminal progresses in the structural analysis of epigenetic complexes have been made in the last decade. The following book chapter provides an up-to-date overview of this work.

# Structural Biology of Epigenetic Targets: Exploiting Complexity

**Martin Marek, Tajith Baba Shaik and Christophe Romier**

Département de Biologie Structurale Intégrative, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Université de Strasbourg, CNRS, INSERM, 1 rue Laurent Fries, 67404 Illkirch Cedex, France.

Correspondence to be addressed to: [acmnpv@igbmc.fr](mailto:acmnpv@igbmc.fr) and [romier@igbmc.fr](mailto:romier@igbmc.fr)

## 2.1. Introduction

In the last two decades, epigenetic effectors have increasingly been shown to be major regulators of nuclear processes, with direct implications for cell homeostasis, response to external stimuli, development, but also onset and progression of many diseases(1). As a consequence, both fundamental research in epigenetics and the development of epigenetic drugs (epidrugs) for therapy have become major fields of investigation.

Initial studies focused on epigenetic enzymes involved in the deposition and removal of epigenetic marks and on the reader domains responsible for the specific recognition of these marks(1-4). Yet, the discovery that other epigenetic effectors such as histone variants, histone chaperones and ATP-dependent chromatin remodelers are also implicated in diseases further broadens the number of targets for epidrug design(1,5-9).

A few epidrugs are already approved for the treatment of diseases, notably cancer(4,7). Their clinical use is often accompanied by serious undesirable side effects due to the fact that many epigenetic effectors belong to families whose members are often functionally different but structurally similar. This makes of selective inhibition a major issue for the design of next generation epidrugs. In this respect, structural information is invaluable in helping deciphering precisely in molecular terms the mechanisms governing epigenetic processes, and in aiding next-generation epidrug design.

Another important reason for the reduced usage of epidrugs is the strong interplay between epigenetic effectors. Notably, many epigenetic effectors act within large macromolecular complexes that represent the bona fide functional epigenetic units and that bear different epigenetic activities. This organization has two major consequences for the design and the use of epidrugs. First, these complexes are physically and functionally linking epigenetic activities. Thus, modulating one activity with small molecules is likely to affect the other activities. Second, regulatory subunits can change partner/substrate recognition, enzymatic activity/kinetics, as well as inhibitor binding. Here again, deciphering the structures of these large molecular assemblies, or at least those of their active subcomplexes, is of paramount importance for understanding epigenetic mechanisms and for aiding epidrugs design.

A wealth of structural data has already been obtained on epigenetic effectors and their interactions with inhibitors, substrates and protein partners, unraveling the diversity and complexity of these interactions. The huge amount of published structural data prevents an exhaustive description of all these results. Our previous chapter on epigenetic enzymes (3) and specific chapters of this book are providing precise structural information on epigenetic enzymes and readers. In this chapter, we have chosen to focus primarily on epigenetic macromolecular complexes from the various classes of epigenetic effectors and whose structures have enlarged our understanding of epigenetic mechanisms. Specifically, macromolecular interactions as well as mechanisms leading to structural rearrangements are described, highlighting ways of modulating the activity of epigenetic effectors.

## **2.2. DNA methylases: the DNMT3A/DNMT3L/H3 and DNMT1/USP7 complexes**

In human, DNA methylation occurs predominantly on cytosines (5-methylcytosine) in CpG motifs that often forms clusters known as CpG islands(10). The initial view that DNA methylation is a rather stable epigenetic mark has been completely revisited in the last decade as new demethylation pathways have been characterized(11). It is now commonly accepted that DNA methylation is a highly dynamic mark that is important in developmental processes. Specifically, methylation patterns are strongly perturbed in diseases, notably in cancers(10,11).

De novo DNA methylation is carried out by the DNA methyltransferases DNMT3A and DNMT3B, whereas DNMT1 is required for the maintenance of the methyl mark by methylating hemimethylated DNA. DNMT3L, an inactive paralog of DNMT3A/B, binds to and stimulates the activity of DNMT3A. DNMT3A activity is also stimulated in a DNMT3L-independent manner by histone H3 when its lysine 4 is not methylated. DNMT3A and DNMT3L both have an ADD (ATRX-DNMT3-DNMT3L) domain followed by a methyl transferase domain. Yet, DNMT3A catalytic domain (CD) is active whereas the one of DNMT3L (CD-like) is inactive.

The ADD domains of DNMT3A and DNMT3L are able to bind to the N-terminus of histone H3 when it is not methylated on lysine 4. Methylation prevents binding due to steric hindrance(12,13). The 3.8 Å resolution crystal structure of DNMT3A (ADD-CD) bound to DNMT3L (CD-like) shows that DNMT3L (CD-like) forms an extensive interaction with DNMT3A catalytic domain. This suggests that stimulation of DNMT3A activity by DNMT3L comes from a stabilization of the DNMT3A catalytic domain (Fig.1a)(14). However, DNMT3A ADD domain and the linker region that connects it to DNMT3A (CD) pack against DNMT3A catalytic domain at a position where substrate DNA would be expected to bind, indicating that this structure represents an inhibitory form of the DNMT3A-DNMT3L complex.

Upon binding of histone H3 N-terminus unmethylated on lysine K4 to DNMT3A ADD domain, the ADD domain makes a large movement, interacting with another surface of DNMT3A (CD), thus freeing the DNA-binding surface of this catalytic domain (Fig. 1b)(14). Specifically, H3K4 binds to DNMT3A ADD domain aspartate residues that are otherwise involved in the formation of the inactive DNMT3A conformation.

DNMT1 is also an essential DNA methylase and the target of FDA approved epidrugs(4,7). In contrast to DNMT3A/B enzymes, DNMT1 can only methylate hemimethylated DNA. The 3.0 and 2.6 Å crystal structures of DNMT1 bound to non-methylated DNA and to hemimethylated DNA suggest a mechanism by which this enzyme carries out this discrimination(15,16). Specifically, unmethylated DNA is recognized by a zinc finger of DNMT1(15). This recognition positions the linker that connects the zinc finger to the first bromo-adjacent homology (BAH1) domain of DNMT1 between the DNA and the active site of the enzyme (Fig. 1c).

In the structure of the productive complex(16), the DNA is found inserted into the active site of the catalytic domain (Fig. 1d). Actually, in this structure the major conformational change observed concerns the catalytic loop that adopts a conformation compatible with catalysis. Yet, this structure was obtained with a shorter construct of DNMT1 that does not encompass the zinc finger and the following linker that are playing a major role in DNMT1 autoinhibition in presence of unmethylated DNA. It remains therefore to be understood whether the presence of hemimethylated DNA prevents zinc-finger binding and autoinhibition, or whether the removal of the inhibition is due to an active mechanism.

DNMT1 has been shown to be regulated through various pathways and partner proteins. One of them is the Ubiquitin Specific Protease 7 (USP7) that stabilizes DNMT1. The 2.9 Å crystal structure of USP7 C-terminus (USP7C) in complex with DNMT1 has been solved(17). The overall structure of DNMT1 in this complex is highly similar to the one in the autoinhibited form, including the positioning of the DNMT1 inhibitory N-terminal linker in DNMT1 DNA-binding site (Fig. 1e).

The DNMT1-USP7C complex structure reveals that USP7C, which is composed of several ubiquitin-like domains, binds to DNMT1 on the side opposite to the methylase active site. A critical interaction is made with DNMT1 KG-linker that contains several Lysine-Glycine repeats. Specifically, the lysines of this linker are forming multiple interactions with residues of USP7C and acetylation of these lysines preclude interaction between USP7 and DNMT1, favoring the in vivo degradation of DNMT1(17).

### **2.3. Histone Arginine methyltransferases: the PRMT5-MEP50 complex**

Protein Arginine MethylTransferases (PRMTs) are monomethylating and symmetrically or asymmetrically dimethylating arginine residues in histones and other cellular effectors(18). The role and mode of action of PRMTs have long remained poorly understood. This picture is however changing as more data is obtained on this class of enzymes, showing that they are also involved in a wide range of diseases. Specifically, to develop therapeutic strategies targeting these enzymes, the deciphering in molecular terms of the specific recognition by PRMTs of their substrates and of the influence of partner proteins on PRMTs activity and substrate recognition has to be addressed.

The 2.0 and 3.0 Å crystal structures of human and *X. laevis* PRMT5 in complex with one of its partners, MEP50, have provided novel information on these issues(19,20). PRMT5 mono-methylates and symmetrically di-methylates different substrates(18). PRMT5 is composed of two domains: a N-terminal TIM-barrel and a C-terminal catalytic domain that adopts a canonical arginine methyl transferase fold. The structures of the PRMT5-MEP50 complex reveal the formation of a tetramer of PRMT5-MEP50 dimers where PRMT5 forms the core of the octamer and MEP50 is located on the outside of the complex.

The 2.0 Å crystal structure of the human PRMT5-MEP50 complex in presence of a AdoMet analog and a H4 N-terminal tail peptide shows how the substrate is recognized in the active site of PRMT5 and suggests how active site residues participates to the methylation process (19) (Fig. 2a). Interestingly, the crystal structures of PRMT5-MEP50 bound to selective PRMT5 inhibitors show how these inhibitors can bind directly to these active site residues, leading to selective inhibition(21,22). In addition, these different structures also reveals the molecular basis by which phosphorylation of tyrosine residues in the substrate binding groove can diminish catalytic activity by opposing to substrate binding.

Yet, these structures do not reveal the role of MEP50 in the complex. This information is provided by a lower resolution electron microscopy (EM) structure of the PRMT5-MEP50 complex bound to one of its substrate, nucleoplasmin. This structure reveals that nucleoplasmin interacts predominantly with MEP50 that serves as a docking platform for the substrate to be presented to PRMT5(20).

#### **2.4. Histone lysine methyltransferases: the MLL3–RBBP5–ASH2L and the PRC2 complexes**

Proteins of the MLL family play major roles in development and are mainly responsible for the methylation of lysine 4 of H3 (H3K4), an epigenetic mark associated with activation of transcription(23). MLL1 has been most studied due to its implication in leukemia, but other MLL proteins have also been shown to be involved in cancers(24). Proteins of the MLL family are part of large complexes that all share the ASH2L and RBBP5 subunits. These two subunits are sufficient to stimulate the methyltransferase activity of the MLL family members that otherwise display poor activity. Only MLL1 requires a third partner, WDR5, for full activation.

The molecular basis for MLL protein activity stimulation has long remained poorly understood. The recent crystallographic structure at 2.4 Å resolution of the human complex formed between the SET methyltransferase domain of MLL3, the C-terminal domain of ASH2L and a long peptide of RBBP5 in presence of S-adenosyl-L-homo cysteine (SAH; product of the demethylation reaction of the S-adenosyl-L-methionine (SAM) MLL3 cofactor) and of a H3 peptide substrate has shed light on this issue (Fig. 2b)(25).

In this structure, the RBBP5 peptide is shown forming a link between the MLL3 SET domain and the ASH2L C-terminal domain. Specifically, RBBP5 N-terminus interacts with MLL3 whereas RBBP5 C-terminus binds to ASH2L. Importantly, all three proteins interact at one precise location with an arginine residue of MLL3 binding to different residues from ASH2L and RBBP5. Although this interface is limited, it appears essential for the stability of the complex, its conformation, and for the stimulation of the methyltransferase activity. The stability of the MLL3-ASH2L-RBBP5 complex is further reinforced by residues neighboring the MLL3 arginine. These neighboring residues, in contrast to the arginine, are not conserved in MLL1. This potentially explains the requirement for WDR5 for the stabilization and the full stimulation of the MLL1-ASH2L-RBBP5 complex activity(25).

Surprisingly, the structure of MLL3 SET domain alone is not much different from the one of MLL3 bound to ASH2L-RBBP5, raising the question of the requirement of the ASH2L-RBBP5 complex for MLL3 activity stimulation. Binding of a H3 peptide to this latter complex only induces local conformational changes, but these changes cannot explain the poor activity of free MLL3. Measurements of the structural dynamics of MLL3 alone and in complex with ASH2L-RBBP5 by NMR and molecular simulation techniques suggest that some sub-domains of MLL3 are intrinsically dynamic and that the binding of ASH2L-RBBP5 confers the sufficient stability to the SET domain to be able to bind stably its histone H3 target and to perform its methyltransferase activity.

Probably, one of the most awaited achievements of the last years has been the structural characterization of the Polycomb Repressive Complex 2, PRC2(26,27). PRC2 trimethylates histone H3 at lysine 27, and H3K27me3 is a major epigenetic mark of facultative heterochromatin that is associated with gene silencing(28-31). Perturbation of PRC2 activity has been linked to multiple diseases, notably cancers, which explains that this complex currently represent a major target in epidrugs development, epidrugs targeting PRC2 being currently in pre-clinical and clinical trials(32,33).

PRC2 is composed of a core complex formed by subunits EZH2 (catalytic subunit), EED (Embryonic Ectoderm Development), SUZ12 (Suppressor of Zeste 12) and RBAP46/48(29). Specifically, a subcomplex formed by EZH2, EED and SUZ12 is sufficient for activity, EZH2 being inactive by itself. Interestingly, binding of H3K27me3 to PRC2 has been shown to allosterically stimulate the activity of the complex whereas a H3K27M mutation yields to the inhibition of PRC2(34). In addition, PRC2 is composed of several facultative subunits that are further responsible for the modulation of the activity of this complex(28,29). Our understanding in molecular terms of the function of PRC2 has long remained obscure. The 1.9 Å crystal structure of the WD40 domain of human EED in complex with a H3K27me3 peptide shed the first light on PRC2 structure/function relationships(34). Specifically, the trimethylated lysine is recognized within a central aromatic cage at the surface of EED WD40 domain. Another essential information came from the 2.0 Å crystal structures of the human EZH2 C-terminal SET domain that is responsible for the H3K27 methylation activity of EZH2(35,36). This structure revealed that the SET domain alone adopts an inactive conformation, with the substrate and SAM (S-adenosyl-L-methionine) cofactor binding sites being occluded by different parts of the SET domain itself.

A first step in our structural understanding of the PRC2 complex initially came from the 21 Å EM structure of a human EZH2/EED/SUZ12/RBAP48/AEBP2 sub-complex(37). This low resolution structure, complemented by labelling, cross-linking and mass spectrometry techniques, enabled the location of the different proteins composing the sub-complex as well as specific domains within the EM map. This structure revealed a 4 lobes (A-D) organization, with SUZ12 and AEBP2 forming a physical link between lobes A-B and lobes C-D. Interestingly, EZH2, EED and a SUZ12 C-terminal domain (VEFS), which are sufficient for activity, appear to form the major components of the A-B lobes.

This initial structural information was complemented and refined by the 2.3-2.9 Å resolution crystal structures of the EZH2/EED/SUZ12-VEFS sub-complex from human and the thermophilic yeast *Chaetomium thermophilum* (Fig. 2c)(26,27). The organization of the complex appears very similar to the one derived from the EM structure. The high resolution data however provides unprecedented detailed molecular information on the PRC2 core subcomplex assembly and its function. The complex can be



divided in two lobes. The regulatory lobe is composed of EED and the N-terminal region of EZH2 (corresponding to lobe A in the EM structure). In this lobe, EZH2 is encircling EED through multiple domains ensuring a very tight interaction between the two subunits. The second catalytic lobe (lobe B in the EM structure) is composed primarily of the C-terminal region of EZH2, including its SET methylation domain. The SUZ12 VEFS domain is found at the interface between these two moieties, interacting with both lobes.

Importantly, in the complex the EZH2 SET domain adopts an active conformation, the SAH product being bound at the SAM binding site, and the peptide binding groove being correctly formed. Actually, a H3M27 mutant (H3K27 lysine is replaced by a methionine) peptide, which is absolutely required for crystallization and is known to be involved in tumorigenesis, is found bound to the SET domain where H3M27 occupies the H3K27 binding pocket. A surprising aspect of the EZH2 catalytic domain is that it is bipartite, the C-terminal SET domain being complemented by a functionally essential SET activation loop (SAL) located in the N-terminal region of EZH2 (regulatory lobe).

The EZH2/EED/SUZ12-VEFS sub-complex could be crystallized in absence (basal state) and in presence (activated state) of a H3K27me3 peptide. In the activated state, the peptide is found bound to EED as observed in the EED/H3K27me3 structure (Fig. 2c). Strikingly, upon H3K27me3 binding to EED the N-terminal EZH2 region, termed Stimulation-Responsive Motif (SRM), which directly follows the SAL, becomes ordered and visible in the electron density, interacting with the H3K27me3 peptide. The SRM also makes direct interactions with the SET domain, suggesting an explanation for the allosteric activation of PRC2 by the H3K27me3 epigenetic mark.

## **2.5. Histone lysine ubiquitinylase: the PRC1 complex**

Polycomb Repressive Complex 1 (PRC1) is also involved in transcriptional repression and, like PRC2, has been linked with various cancers(33). However, PRC1 mono-ubiquitinylates histone H2A at lysine 119. Ubiquitylation requires the activity of three enzymes called E1, E2 and E3. First, an E1 enzyme transfers ubiquitin to an E2-conjugating enzyme. Then, the E2 enzyme transfers the ubiquitin to a lysine side chain of a target protein that is specifically recognized by the E3 enzyme(38).

PRC1 acts as a complex that is minimally composed of three proteins. These three proteins have various homologs yielding different PRC1 complexes that have distinct gene targets(29-31). Two of these proteins are forming the E3 enzyme, whereas the third one is the E2 enzyme. Various unrelated E2/E3 structures have shown how E2 and E3 enzymes are acting in concert, but how the PRC1 specifically recognizes its nucleosome target has long remained poorly understood.

The 3.3 Å resolution crystallographic structure of a human PRC1 (RING1B-BMI1-UBCH5C)-nucleosome complex has enabled the deciphering in molecular terms of this recognition (Fig. 3a)(39). RING1B and BMI1 are the two proteins that compose the E3 enzyme that is supposed to recognize specifically the nucleosome. Accordingly, RING1B makes the most extensive interactions with the nucleosome from the three PRC1 proteins. Specifically, it interacts primarily with the so-called “acidic patch” of the nucleosome, a set of acidic residues of H2A and H2B that are exposed at the surface of the histone octamer(40,41). Actually, many different nuclear effectors that interact with histone pairs and the nucleosome have been shown to target the acidic patch through an “arginine anchor” mechanism(42-45). RING1B forms a particularly large number of interactions with the acidic patch, not only with arginines but also with lysine residues.

BMI1 is also interacting with the histone octamer, albeit mostly with the H3 and H4 histones. This interaction is less extensive than in the case of RING1B but still contributes to the recognition of the nucleosome by PRC1. E2 enzyme UBCH5C interacts mostly with RING1B that is centrally positioned in the RING1B-BMI1-UBCH5C complex and that anchors it on the H2A-H2B acidic patch. This positions UBCH5C active site directly over the H2A C-terminal tail that harbors K119 (Fig. 3a). Interestingly, UBCH5C does not interact with any histone. Rather, it makes contacts with the entry/exit and dyad nucleosomal DNA. Thus, through its interactions with all histones and with DNA, PRC1 ensures that it is engaged with the nucleosome.

## **2.6. Histone lysine deubiquitinylases: the SAGA deubiquitination module**

SAGA (Spt-Ada-GCN5-Acetyl transferase) is a 1.8 MDa transcriptional co-activator complex that acts during RNA polymerase II transcription activation and that also couples transcription elongation with RNA export(46). SAGA bears two enzymatic activities: a histone acetyltransferase activity, through its GCN5 subunit, and a histone deubiquitination activity, through its yeast UBP8/human USP22 subunit. Both enzymes require additional subunits for functional activity on their cognate substrate, the nucleosome.

SAGA deubiquitination module (DUBm) is composed of four proteins that are all required for activity: yeast UBP8, SGF11, SGF73 and SUS1, which deubiquitinylate yeast H2B K123, and human Usp22, ATXN7L3, ATXN7 and ENY2, which deubiquitinylate human H2B K120(46). Only the first hundred residues of SGF73/ATXN7 are required for complex formation and the deubiquitination activity, its C-terminal region being involved in the attachment of the DUBm to the rest of the SAGA complex. Importantly, poly-glutamine extensions in human ATXN7 N-terminal domain are responsible for the SCA7 neurodegenerative disease(47).

The requirement for the four proteins and their role in the deubiquitination activity have remained poorly understood. Initial complex reconstitution experiments with the yeast DUBm have shown that the C-terminal zinc finger of SGF11 and the C-terminal deubiquitination domain of UBP8 are not required for complex assembly(48). The same study showed that SGF73 requires the three other subunits to be incorporated within the complex, but also stabilizes UBP8 in this complex.

The 1.9-2.7 Å crystal structures of the yeast SAGA DUBm in absence and in presence of ubiquitin have revealed the intricacy of the DUBm formation and the role of all four subunits in forming a functionally active deubiquitination complex (Fig. 3b)(49,50). The SAGA DUBm is composed of an assembly lobe and a catalytic lobe. The assembly lobe is formed by a long N-terminal  $\alpha$ -helix of SGF11 around which SUS1 is wrapped. The N-terminal zinc-containing domain of UBP8 docks itself onto this SGF11-N/SUS1 complex, which then serves as a platform for binding of the first half of SGF73 N-terminal domain.

The UBP8 C-terminal deubiquitination catalytic domain makes only little interactions with the assembly lobe, being separated from its N-terminal domain by a linker. In fact, it is the second half of the SGF73 N-terminal domain that forms the interface between the assembly lobe and the UBP8 deubiquitination catalytic domain, locking this domain into a position that makes the DUBm catalytically active (Fig. 3b). Importantly, poly-glutamine extensions in ATXN7, which are responsible of the SCA7 disease, are found in the first half of the ATXN7 N-terminal domain that is expected to be part of the assembly lobe(47,50,51). Although the mechanism by which the disease occurs remains obscure, this highlights the importance of SGF73/ATXN7 N-terminal part in the DUBm assembly and function.

Finally, the SGF11 C-terminal zinc finger, which is not required for DUBm assembly, is found bound to the deubiquitination domain of UBP8, close to the active site, and is separated from SGF11 N-terminal  $\alpha$ -helix by a long linker. This zinc finger has initially been thought to recognize DNA in the nucleosome, but the 3.9 Å crystal structure of the DUBm in complex with an ubiquitinated nucleosome revealed that in fact SGF11 zinc finger recognizes the acidic patch of the nucleosome with several arginine residues using an arginine-anchor mechanism (Fig. 3b)(43).

The DUBm does not make extensive interactions with the rest of the nucleosome. Only a small part of the UBP8 deubiquitination domain appears to contact H2B and the DNA (Fig. 3b). Interestingly, the ubiquitin appears not to make any contacts with the nucleosome, and binds to the UBP8 deubiquitination domain in a canonical manner. Comparison with the nucleosome-free DUBm structure reveals some structural rearrangements upon ubiquitin binding, but the DUBm by itself already seems to be in an active conformation. Since UBP8 alone is very poorly active, it is the formation of the DUBm module that is sufficient to lock UBP8 into this active conformation.

The DUBm is part of the larger SAGA complex that also bears a histone acetyltransferase activity. The low resolution structure of a SAGA-nucleosome complex by electron microscopy suggests that the same nucleosome could be engaged by the two enzymatic activities of SAGA at the same time(52). Structure determination of the GCN5-ADA2-ADA3-SGF29 histone acetyltransferase (HAT module) in complex with the nucleosome is awaited to understand how SAGA HAT activity is carried out. Actually, a structure combining both HAT and DUB modules bound to a mononucleosome would give unprecedented information on how two different epigenetic enzymatic functions can collaborate.

## **2.7. Histone acetyltransferases: the MSL1 and NUA4 complexes**

Lysine acetylation is one of the best characterized covalent modifications of histone tails(53). Histone acetyltransferase (North, Javaid *et al.*) enzymes have been classified into five major subfamilies based on sequence homology and substrate acetylation properties: HAT1, GCN5/PCAF, MYST, p300/CBP and Rtt109(54). All HAT enzymes share a structurally conserved acetyl-CoA binding domain. However, activity of most HATs is regulated by tethering into multisubunit complexes. The enzymatic specificities of different HAT complexes are a matter of intensive research as a majority of isolated HAT catalytic domains exhibits no or little substrate selectivity.

MOF is a histone acetyltransferase of the MYST family that is an essential component of the *Drosophila* dosage compensation male specific lethal (MSL) complex, which also harbours proteins MSL1, MSL2, MSL3, MLE and non-coding RNA roX1 and roX2(55). In human, four orthologs of the *Drosophila* proteins were identified: MOF, MSL1, MSL2 and MSL3. Specifically, while isolated MOF HAT domain is able to acetylate free histones H3 and H4, the entire multisubunit complex can only acetylate histone H4 in a nucleosomal context. Both MSL1 and MSL3 are important for stimulating MOF activity.

MSL1 is predicted to be mostly unstructured but has a scaffolding role within the MSL complex, binding to MOF, MSL2 and MSL3. Specifically, the C-terminal region of MSL1 (PEHE domain) binds consecutively MOF and MSL3, whereas its central region binds to MSL2 N-terminal Ring domain. The crystal structures of mammalian MOF HAT domain in complex with MSL1 PEHE N-terminus, of MSL1 PEHE C-terminus in complex with the MRG domain of MSL3, and of MSL1 central domain in complex with MSL2 Ring domain have provided first insights into how the MSL complex assembles(56-58).

Specifically, the 2.0-2.7 Å crystal structures of MOF(North, Javaid *et al.*)-MSL1(PEHE-N) complex show that MSL1 binding does not induce major conformational changes in MOF, raising the question of the role of MSL1 and MSL3 in stimulating MOF activity (Fig. 4a). Yet, both published structures do not agree on the location of the very N-terminus of the MSL1 construct used. In one structure, this region is close to the

active site where it could influence substrate binding. In addition, MSL3 binding site on MSL1 is close to the MOF binding site and MSL3 could also influence substrate recognition.

Another structural analysis, on the Nu4A acetyltransferase complex, has brought further light on how HAT specificity towards a particular histone tail in a nucleosomal context can be achieved(59). NUA4 is a large HAT complex comprising 13 subunits that preferentially acetylates histone H4 in nucleosomes(60,61). The Nu4A HAT module consists of four subunits: in yeast the catalytic subunit ESA1, EPL1, YNG2 and EAF6 and in mammals TIP60, EPC1, ING3 and EAF6. This core complex can recapitulate most of the enzymatic activities of the holo-enzyme.

The 2.7-3.2 Å crystal structures of the yeast NUA4 HAT core module (ESA1 HAT domain, EPL1 N-terminal and central domains, YNG2 N-terminal domain, and EAF6) alone and in complex with acetyl-CoA or a H2A.Z histone N-terminal tail reveal an organisation in two lobes that are interacting weakly and are conformationally relatively independent from each other (Fig. 4b)(59). EPL1 N-terminal region wraps and interacts extensively with ESA1 HAT domain, forming the catalytic lobe. The central domain of EPL1 forms a long coiled coil structure that bundles with the helices formed by the N-terminal domain of YNG2 and by EAF6 to form the assembly lobe.

Importantly, binding of EPL1 to ESA1 HAT induces conformational changes in the active site of the catalytic domain by reorganizing some important loops. The presence of an autoacetylated lysine in ESA1 is observed that stabilizes the new conformation of an active site loop. The structure of the complex in presence of a histone tail peptide shows that these conformational changes observed are required for the correct accommodation of the peptide in a productive way. Interestingly, the complex by itself reveals very little binding specificity towards the peptide, raising the question on how specificity is achieved.

The 8 Å resolution electron microscopy structures of the NUA4 HAT core complex bound to a nucleosome are providing answers to this question(59,62). Specifically, the crystal structure of the NUA4 core complex and of the nucleosome can be unambiguously fitted in the electron density, revealing that the catalytic HAT subunit is oriented in close vicinity to the histone H4 tail. In addition, ESA1 N-terminal Tudor domain, which was included in the EM analysis, is also found bound to the nucleosome. The complex also appears to contact the acidic patch and DNA elements. This is the first structural evidence on how an otherwise poorly selective HAT can be specifically positioned to catalyse acetylation of its cognate histone substrate.

## **2.8. Histone deacetylases: HDAC1/MTA1 and HDAC3/SMRT complexes and HDAC6**

The family of histone deacetylases has been divided into four classes depending on their folds and their sequence similarities. Classes I, II (IIa and IIb), and IV deacetylases (HDACs) adopt an arginase-deacetylase

$\alpha/\beta$  fold and rely on a zinc ion for activity(63). Class III deacetylases are referred as sirtuins and adopt a Rossmann fold, relying on NAD<sup>+</sup> for activity(64).

Zinc-dependent HDAC enzymes catalyze the removal of the acetyl group from the N $\epsilon$  atom of lysines and are the most pharmacologically investigated epigenetic targets so far. Accordingly, most currently approved epigenetic drugs (Vorinostat (SAHA), Romidepsin, Belinostat, Panobinostat and Tucidinostat (Chidamide)) are inhibitors of HDACs(65,66). The HDAC inhibitors have been proven to be effective anticancer agents, mostly in haematological malignancies, but recent discoveries suggested that they can be also therapeutically beneficial in the treatments of multiple other human diseases including neurodegenerative and immune disorders as well as viral and parasitic infections(66-69).

All HDAC inhibitors approved so far are targeting the active site of these enzymes, binding to the catalytic zinc ion, and thus preventing recognition and deacetylation of cognate cellular substrates. The major issue with clinical usage of these approved HDAC inhibitors is the fact that they have no or little specificity, and target the structurally similar, but functionally different 11 human HDAC isozymes (HDAC1-11). Therefore, the treatments with currently-approved HDAC inhibitors are often accompanied by serious undesirable side effects, which hampers their clinical usage. Several isozyme-selective inhibitors have been developed(66), but the approved inhibitors mostly show limited selectivity. The hydroxamates target mostly class I and HDAC6 (class IIb) while Tucidinostat is selective for class I.

Selective targeting of HDACs-protein interactions is emerging as an alternative for inhibition studies(65,70). Class I HDAC1, HDAC2 and HDAC3 are typically recruited into multisubunit macromolecular complexes. Importantly, the recruitment of these HDACs to co-repressor complexes triggers maximal HDAC activity. The structures of HDAC1 and HDAC3 in complex with cognate co-repressors MTA1 (metastasis-associated protein 1) and SMRT (silencing mediator for retinoid and thyroid hormone receptor) provide the molecular explanation for this enhanced deacetylase activity.

In the 3.0, 3.3 and 2.0 crystal structures of HDAC1-MTA1, HDAC1-MTA1-H4 peptide and HDAC3-SMRT, the co-repressor domains wrap around HDAC catalytic domains and, in case of HDAC1, this domain also mediates dimerization of the complex (Fig. 4c,d)(71-73). Strikingly, the structures of HDAC1-MTA1 and HDAC3-SMRT complexes reveal an unexpected regulation mechanism of HDAC activity mediated by inositol phosphates. In HDAC3-SMRT complex, endogenous Ins(1,4,5,6)P<sub>4</sub> is found in a basic pocket formed between HDAC and the co-repressor domain and that is close to the active site pocket. The 3.3 Å crystal structure of HDAC1-MTA1 with exogenous InsP<sub>6</sub> also revealed an inositol molecule bound in the basic pocket, but it shows certain structural differences when compared to HDAC3/SMRT complex(73).

The binding of an inhibitor in the HDAC active site potentiates inositol polyphosphate binding in the basic pocket, which indicates mutual communication between the active site pocket and the inositol-binding pocket. This cross-talk is likely mediated by HDAC1 residue Arg270 and HDAC3 Arg 265 that interconnect the active site pocket with the inositol binding pocket. Interestingly, while mutation of HDAC3 Arg265 led to the complete loss of deacetylase activity, the mutation of HDAC1 Arg270 is less functionally important. Yet, HDAC1 kinetics is regulated upon inositol polyphosphate binding, leading to the hypothesis of HDAC1 activation by entropic allostery(73).

In contrast to other HDACs, HDAC6 is unique as it harbours two functional catalytic domains (CD1 and CD2). The crystal structures of human HDAC6 CD2 and zebrafish HDAC6 CD1, CD2 and tandem CD1-CD2 brought important structural and mechanistic insights into HDAC6 catalytic mechanism and its inhibitions by small-molecule compounds(74,75). The 2.9 Å crystal structure of zebrafish HDAC6 tandem CD1-CD2 reveals that the tandem forms together with the inter-domain linker an ellipsoid-shaped complex with pseudo-twofold symmetry (Fig. 4e)(75). The two catalytic domains interact strongly with each other, and the linker domain connecting CD1 and CD2 also functions as a domain-domain glue to further stabilize the CD1-CD2 complex. Importantly, biochemical experiments show that maximal HDAC6 activity is achieved when both CD1 and CD2 are physically interacting with each other.

Both HDAC6 CD1 and CD2 are structurally very similar. However, biochemical data shows that the CD2 domain exhibits relatively broad substrate specificity, while CD1 appears specific for substrates bearing a carboxy-terminal acetyl-lysine residue(74). This selectivity seems achieved by the fact that CD1 active site pocket is more constricted due to the presence of a lysine residue (a leucine in CD2) that protrudes into the catalytic pocket. This lysine plays the role of a gatekeeper that confers CD1 a specificity towards carboxy-terminal acetyl-lysine substrates by hydrogen bonding with the  $\alpha$ -carboxylate group of the acetyl-lysine. In addition, the hydroxyl group of a CD1 tyrosine (a phenylalanine in CD2 and in other HDACs) could also contribute to the observed specificity via an additional hydrogen-bonding with the carboxy-terminal acetyl-lysine residue.

The structure of HDAC6 was solved in complex with a HDAC6-selective inhibitor, Nexturastat A, providing specific information on HDAC selective inhibition. Other works on the HDAC8 isozyme have also contributed to address this issue that is essential for developing the next generation of HDAC epidrugs(67,69,76-79).

## **2.9. Histone variants and histone chaperones: a complex and modular interplay**

The sequence differences between histone variants and canonical histones can concern a few residues to full structural domains(8). Beside their common DNA packaging role, canonical and variant histones can have very different functions. Specifically, whereas canonical histones are deposited during replication, histone variants are deposited throughout the cell cycle, where they play specific functional roles at precise genomic loci. Due to their broad implication in many nuclear processes, histone variants are involved in many diseases, notably in cancers.

This importance of histone variants explains that the transfer of histones between the different cellular compartments and their deposition onto and removal from the chromatin is a highly regulated process. This requires the action of a large number of dedicated histone chaperones that are ensuring that the various histones are present at the correct genomic loci(9). Depending on their role, some of these chaperones can recognize any histone pair, whereas others bind only to histone pairs of the H3-H4 or H2A-H2B family. Other chaperones are binding to very specific variant pairs.

Importantly, not only the sequence differences, but also the epigenetic marks born by histones can affect their interaction with histones chaperones. In addition, histone chaperones are often members of large epigenetic complexes, where they collaborate with different epigenetic functions. In the last decade, a large amount of structural data has been gathered on the recognition of histone pairs, canonical and variants, by histone chaperones(9). A few of them are described here that highlight the interplay between histones, histone chaperones, and other epigenetic effectors.

During replication, many different effectors are required to disassemble and then reassemble chromatin. Maintenance of the epigenetic state encoded by specific epigenetic marks is essential during this process. This is rendered complicated by the fact that both old histones and newly synthesized histones are used for reassembly. Specifically, new histones need to be identified and modified according to the previous epigenetic state encoded by the old histones. Anti-silencing function 1 (ASF1) histone chaperone participates to this process by binding H3-H4 pairs. ASF1 was the first chaperone whose structure in complex with a histone pair was solved and which was shown to interact with the H3/H4 pair in a way that prevents this pair to homodimerize and to interact with H2A-H2B(80,81).

The replicative helicase plays an important role in the replication process. Its MCM2 subunit has been shown to bind H3-H4, and the structure of a N-terminal fragment of MCM2 in complex with H3-H4 has revealed that MCM2 binds to a H3-H4 dimer, almost completely shielding the histone tetramer DNA binding surface(82,83). Interestingly, ASF1 and MCM2 were shown to act in concert, and the 2.3 and 3.5



Å crystal structures of a ASF1-MCM2-H3-H4 complex reveals that both chaperones can bind simultaneously to the H3-H4 pair (Fig. 5a)(82,84). In this complex, ASF1 breaks the dimerization interface between the H3-H4 pairs and the complex displays a 1:1:1:1 stoichiometry. Yet, the two chaperones interact minimally with each other, showing the modularity of the chaperone-histone pair interactions.

Prior and after replication, it is essential that newly synthesized histone pairs are recognized for further processing. The complex TONSL-MMS22L has been shown to interact and cooperate with free and chromatin-bound H3-H4, and with ASF1 and MCM2. The 2.4 Å crystal structure of TONSL Ankyrin repeat domain in complex with H3-H4 and MCM2 reveals that TONSL binds minimally to the globular domain of the histones but interacts extensively with H4 N-terminal tail (Fig. 5b)(85). Both TONSL and MCM2 chaperones do not interact directly and modeling shows that ASF1 could also bind to the TONSL-MCM2-H3-H4 complex, reinforcing the concept of modularity. Strikingly, TONSL-H4 interactions are only possible when H4K20 is unmethylated and this epigenetic mark appears specific to newly synthesized H4 histone. Modeling shows that TONSL, which has been shown to be associated with chromatin, could remain bound to the H4 tail in a nucleosomal context, thus marking newly synthesized nucleosomes.

None of the ASF1, MCM2 and TONSL chaperones can distinguish between the different members of the H3 family. Yet, many chaperones bear this capacity in order to deposit histone variants at specific loci. One specific H3 variant is CenpA that is found exclusively at centromeres. The human histone chaperone Hjurp has been shown to recognize specifically the CenpA-H4 pair and the 2.3-2.6 Å crystal structures of the human and yeast complexes shows how Hjurp binds to CenpA-H4, preventing CenpA-H4 dimerization and shielding part of the DNA binding surface of this pair (Fig. 5c)(86-88). Interestingly, the determinants for the specific recognition of CenpA over H3 were found to be minimal.

H3.3 is a variant that is deposited in gene bodies as well as in pericentric and telomeric heterochromatin. H3.3 varies by only five residues from the canonical H3.1 histone. Deposition at heterochromatin loci is carried out by the death domain-associated protein (DAXX) histone chaperone. The 2.8 Å crystal structures of the DAXX-H3.3-H4 complex shows that DAXX interacts extensively with the H3.3-H4 pair (Fig. 5d)(89,90). Despite the large interface between DAXX and H3.3-H4, as for CenpA the specific recognition of H3.3 relies on very few amino acid changes, notably on one glycine residue (G90) that is a methionine in H3.1. Interestingly, the 2.2 Å crystal structure of another, unrelated H3.3 histone chaperone, UBN1, in complex with ASF1 and H3.3-H4 reveals the same specificity mechanism of UBN1 through this glycine(91). In addition, the DAXX-H3.3-H4 complex can be targeted by viral BNRF1 from the Epstein-Barr virus to hijack the deposition machinery, BNRF1 interacting not only with DAXX but also H3.3 and H4(92).

Not only H3, but also H2A and H2B have various members in their families and dedicated histone chaperones. New data has shed light on the recognition and processing of the H2A.Z variant which acts both in transcription and DNA repair. Specifically, two large functionally homologous human ATP-dependent chromatin remodelers were shown to act on H2A.Z: P400/TIP60 and SRCAP(93). The P400/TIP60 complex is not only involved in chromatin remodeling, but also in acetylation through its TIP60 subunit that correspond to yeast ESA1 from the NUA4 complex described above. Histone chaperone ANP32E has been shown to belong to P400/TIP60 and to evict H2A.Z from the chromatin. The 1.5-2.6 Å crystal structure of the ANP32E-H2A.Z-H2B complex shows a minimal interface between the chaperone and the histone pair (Fig. 5e)(93,94). ANP32E binding to H2A.Z-H2B causes the doubling in size of H2A.Z  $\alpha$ C helix, with direct implication for H2A.Z-H2B eviction from the nucleosome. This extension of H2A.Z  $\alpha$ C helix is only enabled by the absence, compared to H2A, of a glycine residue at the end of this helix. Strikingly, this glycine is also the only determinant for H2A.Z specific recognition by ANP32E.

The second remodeling complex involved in H2A.Z biology is SRCAP. This complex has been shown to be involved in H2A.Z deposition and histone chaperone YL1 is involved in this mechanism. The 1.9-2.7 crystal structure of the YL1-H2A.Z-H2B complex reveals that this chaperone also doubles in size H2A.Z  $\alpha$ C helix, but it also interacts more extensively with the histone pair, covering its entire DNA binding surface (Fig. 5f)(44,45). Even if the H2A.Z  $\alpha$ C helix glycine plays a role in the specific H2A.Z recognition by YL1, other determinants are required for full recognition, in contrast with that has been observed with ANP32E. In this case, the ATP-dependent remodeling activity of SRCAP is probably needed to prior evict the H2A-H2B pair from the nucleosome. For ANP32E, nucleosome remodeling and recognition by a single determinant appear much more relevant for accessing H2A.Z  $\alpha$ C helix buried at the heart of the nucleosome, prior to H2A.Z-H2B eviction through H2A.Z  $\alpha$ C helix extension.

#### **2.10. ATP-dependent remodelers: CHD1, ISWI, SNF2 and the SNF2/nucleosome complex**

ATP-dependent chromatin remodelers are large multi-subunit complexes that use the energy of ATP to partially or fully disassemble, slide or change the conformation of nucleosomes(5,95). Their ATPase subunits all have a catalytic domain composed of two RecA-like (DExx and HELICc) subdomains. Based on the sequence similarity of their catalytic domains and of the presence or absence of additional auxiliary domains, the remodelers have been classified in four large families: SWI/SNF, IWSI, CHD, and INO80.

The increasing reports of the implication of remodelers in many different diseases make of these epigenetic effectors important targets for therapeutic developments. Specifically, the remodelers of the SWI/SNF family are found mutated in 20% of all human tumors(6). Yet, inhibition of the ATPase catalytic

domain might cause strong selectivity problems. It is therefore essential to understand in molecular terms how these catalytic domains cooperate with the auxiliary domains of the catalytic subunits and with their partner subunits in their respective complexes, but also how they recognize and act on their nucleosomal substrates. The intrinsic flexibility and the large size of the remodelers has slowed down the structural characterization of these epigenetic machines. Yet, our understanding on how remodelers are acting and are being regulated has made important progress.

The CHD family is characterized by a tandem of chromodomains in the N-terminal region of the ATPase subunit. The 3.7 Å crystal structure of CHD1 core encompassing both chromodomains and the ATPase catalytic domain reveals that these different domains are interacting extensively, the two chromodomains bridging both RecA-like lobes of the ATPase (Fig. 6a)(96). The observed positioning of lobe 1 compared to lobe 2 implies that CHD1 ATPase domain is in an inactive conformation. In addition, part of the linker bridging the two chromodomains and that forms two long antiparallel  $\alpha$ -helices is found binding to lobe 2 at a position where DNA is supposed to bind. Binding of methylated proteins to CHD1 chromodomains, like in the case of KDM1A(97), should possibly relieve inhibition. Interestingly, the histone H4 N-terminal tail has also been shown to stimulate CHD1 activity even in absence of the chromodomains, demonstrating of an intricate regulatory mechanism of CHD1 activity.

In the ISWI family, two domains (AutoN and NegC) flanking the ATPase catalytic domain have been shown to regulate the ATPase activity. The 2.4 Å crystal structure of the core IWSI ATPase subunit encompassing the AutoN, ATPase and NegC domains shows, as for CHD1, a non-productive positioning of the two RecA-like lobes (Fig. 6b)(98). Here, the AutoN domain is shown interacting with both lobes of the ATPase domain, possibly stabilizing, as the two chromodomains of CHD1, the non-productive conformation of the enzyme. Interestingly, the N-terminal tail of H4 is also stimulating the activity of ISWI. The 3.0 Å crystal structure of a N-terminal H4 peptide bound to ISWI lobe 2 shows that H4 binds to one of the anchoring surface of AutoN on lobe 2, suggesting an explanation to the stimulating activity of H4 by preventing the interaction between the ISWI AutoN and catalytic domains(98).

In contrast, the role of ISWI NegC domain is less clear from the structural analysis since this domain projects out of the structure. In the IWSI family, the NegC domain is followed by a Hand-Slant-Slide (HSS) domain that contacts naked DNA. The 3.2-3.6 Å crystal structure of the HSS domain from remodeler subunit ISW1a in complex with another subunit of this complex, LOC3, in absence and in presence of DNA showed that the HSS-LOC3 complex is able to bind two DNA molecules(99). The low resolution EM structure of the HSS-LOC3 complex with a nucleosome having free DNA extensions at the nucleosome

entry and exit points further shows that the HSS-LOC3 complex can interact both with the DNA entering and exiting the nucleosome, suggesting that the substrate of ISWI family members could be a dinucleosome rather than a mononucleosome, and that the HSS domain could define the minimal spacing between two nucleosomes(99). Interestingly, the NegC domain could possibly help with this process, functionally linking the ATPase catalytic domain and the HSS domain.

In the SNF family, the ATPase core domains are flanked on the N-terminus by a HSA (helicase SANT-associated) domain and a post-HSA domain and, on the C-terminus, by a SnAC domain. The HSA domain forms a long  $\alpha$ -helix that binds to other proteins, such as actin or actin-related proteins (ARPs), and regulates the remodeling activity by an unknown mechanism(100). The role of the other auxiliary domains is less well understood. The 2.3 Å crystal structure of a yeast Snf2 encompassing a small part of the HSA, the post-HSA, the ATPase and SnAc domains has shed light on the interactions between these different domains(101).

Although the part of the HSA domain included in the analysis is not seen in density, the post-HSA is found forming a long helix interacting with the RecA-like lobe 1 of the ATPase (Fig. 6c). The SnAc forms extensive interactions with lobe 2, having little contacts with the first lobe. Interestingly, the two core domains of the ATPase interact strongly with each other but their relative positioning, as for CHD1 and ISWI, is not compatible with a productive mode. In contrast with CHD1 and ISWI, this nonproductive conformation is however not stabilized by auxiliary domains.

The electron microscopy structure at 4.0-4.7 Å resolution of a SNF2-nucleosome complex has provided a long awaited structural view at sufficient resolution on how the catalytic subunit of a remodeler engages its nucleosome substrate (Fig.6d)(102). Upon binding, the two RecA-like lobes are reorienting themselves to now form a productive ATPase catalytic domain. This not only requires the movement relative to each other of the different structural elements observed in the unbound structure, but also the structuration of several disordered elements.

SNF2 is found bound at super-helical location 2 (SHL2) of the nucleosome, a position already known to bind preferentially several remodelers and where the H4 N-terminal tail protrudes from the nucleosome. Accordingly, the H4 tail is found bound at the same position on lobe 2 as was observed for the ISWI lobe 2-H4 tail complex(98). Since H4 is not required to stimulate SNF2 activity, this interaction may in this case only position correctly SNF2 on the nucleosome. In the case of ISWI, it may both remove the inhibition by the AutoN domain and position ISWI on the nucleosome.

Beside the limited SNF2-H4 interaction, SNF2 interacts strongly with the nucleosomal DNA at the SHL2. This interaction is mediated within the cleft formed between lobes 1 and 2 of the ATPase domain and involves primarily contacts with the phosphate backbone. In addition, SNF2 also interacts with its lobe 1 at SHL-6, most likely to help position the enzyme in a correct orientation to the nucleosome. Importantly, binding of SNF2 to the nucleosome in absence of any nucleotide is already sufficient to distort the DNA at SHL2, lifting the DNA off from its canonical path on the histone octamer surface, thus probably priming the remodeling reaction to come.

In the last decade, our mechanistic understanding of ATP-dependent chromatin remodelers has made a lot of progress. It remains to be understood how remodelers from the INO80 family, which have a long insertion between their RecA domains, are structurally organized, and what are the changes in the remodelers upon nucleotide binding. Importantly, many features observed in the SNF2-nucleosome structure might be observed with other remodelers due to sequence conservation of mechanistically important residues, suggesting similar modes of action. Yet, existing sequence and structural differences between the remodelers are also likely to convey specific functional outputs. The same applies for the associated subunits forming the full remodeler complexes.

### **2.11. Epigenetic readers: histone crotonylation readers and the 53BP1-nucleosome(H2AK15Ub-H4K20me2) complex**

Epigenetic readers are becoming more and more attractive targets for epidrug development. This is exemplified by the case of the bromodomains that recognize the acetyl-lysine epigenetic marks and other known readers are currently also targets for epidrug development (see chapters 14 and 15 in this book). Interestingly, recent progresses in mass spectrometry instrumentation and high-resolution proteomic approaches are identifying so-far unknown histone modifications and their readers, which should define future targets in epidrug development.

Among the epigenetic marks recently identified are lysine propionylation (Kpr), butyrylation (Kbu) and crotonylation (Kcr)(103). These findings suggest that there are corresponding enzymes that lay and erase these marks, but also reader modules that transduce this information into functional outcomes. Among the first structurally characterized readout modules of non-acetyl modifications belong readers of lysine crotonylation such as PHD fingers and YEAST domains(104-106). Crotonylation differs from other acylations in its rigidity and planar configuration due to the  $\pi$ -electron conjugation. Biophysical measurements reveal that the double PHD finger (DPF) domains of human MOZ and DPF2 bind a range of histone lysine acylations, but display the strongest affinity for crotonylated lysine residues(105).

Specifically, the 1.4-1.6 Å crystal structures of the MOZ DPF domain in complex with histone H3K14cr, H3K14bu and H3K14pr revealed that these non-acetyl acylations are accommodated in a hydrophobic “dead-end” pocket with selectivity for crotonylation (Fig. 7a). The observed selectivity towards crotonyl-lysine is achieved by intimate encapsulation and an amide-sensing hydrogen bonding network. Interestingly, sequence and structural comparison revealed that a glycine residue is critical for the pocket formation owing to its side-chain-free feature. In most classical histone H3K4-binding PHD fingers, this glycine is replaced by a bulkier residue such as tyrosine or phenylalanine, which fills the pocket and prevents Kcr accommodation.

A different mode of specific recognition of crotonyl-lysine was described in AF9 YEATS domain(106,107). The NMR and 2.7 Å crystal structures of AF9 YEATS complexed with histone H3K9cr and H3K18cr peptides provided molecular insight into specific recognition of crotonylated substrates (Fig. 7b). In the structure, Kcr is anchored in a specific pocket and the Kcr-flanking H3 residues are recognized by extensive polar or hydrophobic contacts. Careful inspection of the crystal structure revealed an extended aromatic sandwiching cage with crotonyl specificity arising from  $\pi$ -aromatic and hydrophobic interactions between the planar crotonylamide group and aromatic rings of a phenylalanine and tyrosine residues. Interestingly, these structural features are conserved within the YEATS family members, but not within the bromodomains, providing additional evidence for why bromodomains do not bind crotonyl-lysine substrates.

Readers can bind to small peptides, but in vivo this recognition is often carried out in a nucleosomal context. How this is done remains poorly understood. The 4.5 Å electron microscopy structure of tumor suppressor 53BP1, which acts in DNA repair and binds to nucleosomes simultaneously ubiquitinated on H2AK15 and dimethylated on H4K20, provides an interesting view on modified nucleosome recognition (Fig. 7c)(108). Specifically, a small region of 53BP1 composed of a Tudor domain followed by a small ubiquitin-dependent recruitment motif (UDR) is sufficient to bind to this doubly modified nucleosome. Although the medium resolution of the structure renders the structural interpretation somewhat difficult, it appears that the Tudor domain lays over the H4K20me2 epigenetic mark, whereas the following small UDR motif meanders on the nucleosome surface, interacting with specific motifs, including the acidic patch.

An important aspect of this mode of binding is the positioning of a part of this UDR between the histone octamer and the ubiquitin. This binding of the UDR, enables the bridging between the histone octamer and the ubiquitin, in agreement with the observation made with the SAGA DUBm that nucleosome-bound

ubiquitin does not interact with the nucleosome. This ubiquitin-UDR-nucleosome interaction stabilizes the ubiquitin in a specific position and enables the recognition of the H2A-K15 ubiquitinated nucleosome. Interestingly, the previously described 8 Å resolution structure of the core NUA4 HAT complex bound to a nucleosome through a Tudor domain, an interaction with the acidic patch and DNA elements(59), corroborates the 53BP1/nucleosome structure. Yet, better resolution will be required in the future for the structures of such complexes to enable the design of next-generation epidrugs. Nevertheless, the complexity of the recognition of modified nucleosome by larger epigenetic effectors, as the ones described in this chapter, indicate that selective modulation of epigenetic effectors-nucleosome interactions could be achieved.

## **2.12. Conclusions**

Structural analyses provide an essential information when it comes to design more selective and more potent drugs. The results presented here highlight the intricacy of the assembly, the interactions and the mechanisms of epigenetic effectors to achieve precise epigenetic regulation. Specifically, the embedding of epigenetic enzymes within large macromolecular complexes yields a high degree of modulation of their activity through structural changes and allosteric mechanisms. This has major implications for the design of novel, more selective and more potent drugs targeting the active site of epigenetic enzymes.

This also opens the way to design drugs that will modulate (i) the intramolecular interactions within complexes and (Suzuki, Muto *et al.*) the interactions of these complexes with their substrates. Importantly, regulation of these mechanisms through small molecules could help select between processes that yield different biological outcomes. Another important aspect concerns the fact that a drug targeting the same enzyme can have different effects depending on the macromolecular environment of the enzyme (i.e. free or embedded in various complexes). In addition, some complexes are bearing different epigenetic activities. It will be interesting for these complexes to design dual inhibitors that can modulate these different activities at the same time.

A surprising observation made from the different epigenetic effectors-nucleosome complexes whose structures have been solved so far is that a major interaction of the epigenetic players with the nucleosome involves targeting the H2A-H2B acidic patch. The other interactions appear less strong but serve to position correctly the effectors on the nucleosome. It remains to be investigated whether targeting other interaction regions with the nucleosome could help selectively modulating the chromatin accessibility of specific complexes. Yet, epigenetic complexes are often larger than the sub-complexes

currently being structurally characterized when bound to the nucleosome. This should open many more possibilities for selective modulation of epigenetic effectors-nucleosome interactions.

The structural results on epigenetic players obtained in the last decade represent the beginnings of a new era where epigenetic complexes will be at the heart of integrated structural, chemical and medicinal biology. Notably, the ongoing major advances in the electron microscopy technology that are enabling a “resolution revolution” (109) are expected to further ease the high resolution structure determination of epigenetic complexes, with direct implications in epidrugs discovery. Importantly, the combination of new generation epidrugs with the new revolutionary technologies currently developed in genome editing and targeting (110) should provide unprecedented means to specifically intervene at specific genomic loci to correct genetic and epigenetic aberrations to cure human diseases.

### **Acknowledgments**

This work and the authors of this manuscript have been supported by funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreements nos. 241865 (SEtTReND) and 602080 (A-ParaDDisE), by institutional funds from the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM) and the Université de Strasbourg, by la Fondation ARC pour la Recherche sur le Cancer (contract PJA 20141201960), by the IdEX Attractivité Université de Strasbourg, by the French Infrastructure for Integrated Structural Biology (FRISBI; ANR-10-INSB-05-01), and by Instruct as part of the European Strategy Forum on Research Infrastructures (ESFRI).



## References

1. Allis, C.D. and Jenuwein, T. (2016) The molecular hallmarks of epigenetic control. *Nat Rev Genet*, **17**, 487-500.
2. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693-705.
3. Romier, C., Wurtz, J.M., Renaud, J.P. and Cavarelli, J. (2009) In Sippl, W. and Jung, M. (eds.), *Epigenetic Targets in Drug Discovery*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, Vol. 42, pp. 23-56.
4. Cai, S.F., Chen, C.W. and Armstrong, S.A. (2015) Drugging Chromatin in Cancer: Recent Advances and Novel Approaches. *Mol Cell*, **60**, 561-570.
5. Clapier, C.R. and Cairns, B.R. (2009) The biology of chromatin remodeling complexes. *Annual review of biochemistry*, **78**, 273-304.
6. Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J. and Crabtree, G.R. (2013) Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet*, **45**, 592-601.
7. Brien, G.L., Valerio, D.G. and Armstrong, S.A. (2016) Exploiting the Epigenome to Control Cancer-Promoting Gene-Expression Programs. *Cancer Cell*, **29**, 464-476.
8. Buschbeck, M. and Hake, S.B. (2017) Variants of core histones and their roles in cell fate decisions, development and cancer. *Nat Rev Mol Cell Biol*, **18**, 299-314.
9. Hammond, C.M., Stromme, C.B., Huang, H., Patel, D.J. and Groth, A. (2017) Histone chaperone networks shaping chromatin function. *Nat Rev Mol Cell Biol*, **18**, 141-158.
10. Schubeler, D. (2015) Function and information content of DNA methylation. *Nature*, **517**, 321-326.
11. Jeschke, J., Collignon, E. and Fuks, F. (2016) Portraits of TET-mediated DNA hydroxymethylation in cancer. *Curr Opin Genet Dev*, **36**, 16-26.
12. Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D. *et al.* (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature*, **448**, 714-717.
13. Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M. and Shirakawa, M. (2009) Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep*, **10**, 1235-1241.
14. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G. *et al.* (2015) Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature*, **517**, 640-644.
15. Song, J., Rechkoblit, O., Bestor, T.H. and Patel, D.J. (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science*, **331**, 1036-1040.
16. Song, J., Teplova, M., Ishibe-Murakami, S. and Patel, D.J. (2012) Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science*, **335**, 709-712.
17. Cheng, J., Yang, H., Fang, J., Ma, L., Gong, R., Wang, P., Li, Z. and Xu, Y. (2015) Molecular mechanism for USP7-mediated DNMT1 stabilization by acetylation. *Nat Commun*, **6**, 7023.
18. Blanc, R.S. and Richard, S. (2017) Arginine Methylation: The Coming of Age. *Mol Cell*, **65**, 8-24.
19. Antonysamy, S., Bonday, Z., Campbell, R.M., Doyle, B., Druzina, Z., Gheyi, T., Han, B., Jungheim, L.N., Qian, Y., Rauch, C. *et al.* (2012) Crystal structure of the human PRMT5:MEP50 complex. *Proc Natl Acad Sci U S A*, **109**, 17960-17965.
20. Ho, M.C., Wilczek, C., Bonanno, J.B., Xing, L., Seznec, J., Matsui, T., Carter, L.G., Onikubo, T., Kumar, P.R., Chan, M.K. *et al.* (2013) Structure of the arginine methyltransferase PRMT5-MEP50 reveals a mechanism for substrate specificity. *PLoS One*, **8**, e57008.

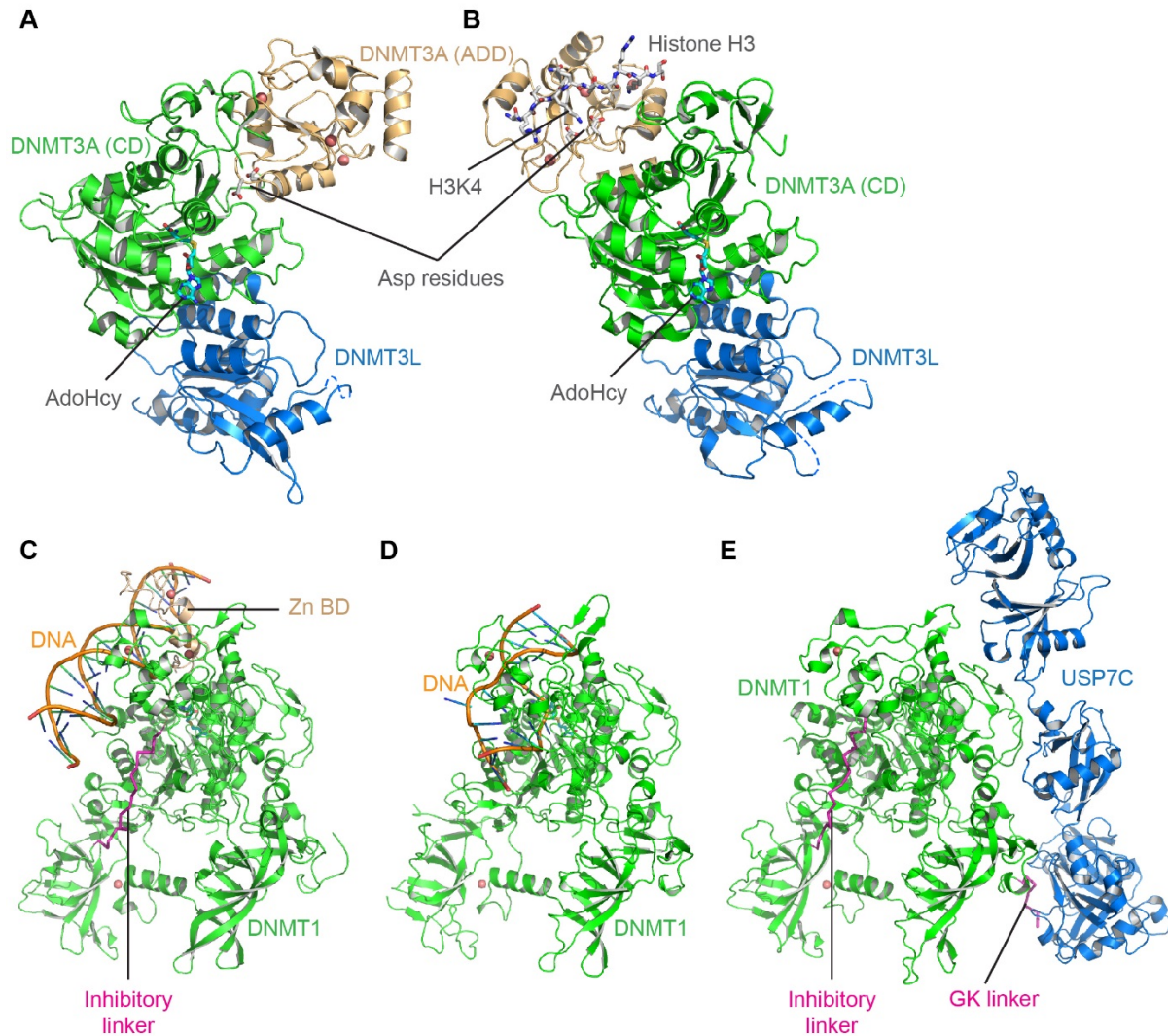
21. Chan-Penebre, E., Kuplast, K.G., Majer, C.R., Boriack-Sjodin, P.A., Wigle, T.J., Johnston, L.D., Rioux, N., Munchhof, M.J., Jin, L., Jacques, S.L. *et al.* (2015) A selective inhibitor of PRMT5 with in vivo and in vitro potency in MCL models. *Nat Chem Biol*, **11**, 432-437.
22. Mavrakis, K.J., McDonald, E.R., 3rd, Schlabach, M.R., Billy, E., Hoffman, G.R., deWeck, A., Ruddy, D.A., Venkatesan, K., Yu, J., McAllister, G. *et al.* (2016) Disordered methionine metabolism in MTAP/CDKN2A-deleted cancers leads to dependence on PRMT5. *Science*, **351**, 1208-1213.
23. Smith, E., Lin, C. and Shilatifard, A. (2011) The super elongation complex (SEC) and MLL in development and disease. *Genes Dev*, **25**, 661-672.
24. Ford, D.J. and Dingwall, A.K. (2015) The cancer COMPASS: navigating the functions of MLL complexes in cancer. *Cancer Genet*, **208**, 178-191.
25. Li, Y., Han, J., Zhang, Y., Cao, F., Liu, Z., Li, S., Wu, J., Hu, C., Wang, Y., Shuai, J. *et al.* (2016) Structural basis for activity regulation of MLL family methyltransferases. *Nature*, **530**, 447-452.
26. Jiao, L. and Liu, X. (2015) Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science*, **350**, aac4383.
27. Justin, N., Zhang, Y., Tarricone, C., Martin, S.R., Chen, S., Underwood, E., De Marco, V., Haire, L.F., Walker, P.A., Reinberg, D. *et al.* (2016) Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nat Commun*, **7**, 11316.
28. Holoch, D. and Margueron, R. (2017) Mechanisms Regulating PRC2 Recruitment and Enzymatic Activity. *Trends Biochem Sci*, **42**, 531-542.
29. Margueron, R. and Reinberg, D. (2011) The Polycomb complex PRC2 and its mark in life. *Nature*, **469**, 343-349.
30. Blackledge, N.P., Rose, N.R. and Klose, R.J. (2015) Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat Rev Mol Cell Biol*, **16**, 643-649.
31. Simon, J.A. and Kingston, R.E. (2013) Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell*, **49**, 808-824.
32. Comet, I., Riising, E.M., Leblanc, B. and Helin, K. (2016) Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. *Nat Rev Cancer*, **16**, 803-810.
33. Koppens, M. and van Lohuizen, M. (2016) Context-dependent actions of Polycomb repressors in cancer. *Oncogene*, **35**, 1341-1352.
34. Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V. *et al.* (2009) Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*, **461**, 762-767.
35. Antonysamy, S., Condon, B., Druzina, Z., Bonanno, J.B., Gheyi, T., Zhang, F., MacEwan, I., Zhang, A., Ashok, S., Rodgers, L. *et al.* (2013) Structural context of disease-associated mutations and putative mechanism of autoinhibition revealed by X-ray crystallographic analysis of the EZH2-SET domain. *PLoS One*, **8**, e84147.
36. Wu, H., Zeng, H., Dong, A., Li, F., He, H., Senisterra, G., Seitova, A., Duan, S., Brown, P.J., Vedadi, M. *et al.* (2013) Structure of the catalytic domain of EZH2 reveals conformational plasticity in cofactor and substrate binding sites and explains oncogenic mutations. *PLoS One*, **8**, e83737.
37. Ciferri, C., Lander, G.C., Maiolica, A., Herzog, F., Aebersold, R. and Nogales, E. (2012) Molecular architecture of human polycomb repressive complex 2. *Elife*, **1**, e00005.
38. Streich, F.C., Jr. and Lima, C.D. (2014) Structural and functional insights to ubiquitin-like protein conjugation. *Annu Rev Biophys*, **43**, 357-379.
39. McGinty, R.K., Henrici, R.C. and Tan, S. (2014) Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature*, **514**, 591-596.
40. 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 Å resolution. *Nature*, **389**, 251-260.

41. Suto, R.K., Clarkson, M.J., Tremethick, D.J. and Luger, K. (2000) Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nat Struct Biol*, **7**, 1121-1124.
42. McGinty, R.K. and Tan, S. (2015) Nucleosome structure and function. *Chem Rev*, **115**, 2255-2273.
43. Morgan, M.T., Haj-Yahya, M., Ringel, A.E., Bandi, P., Brik, A. and Wolberger, C. (2016) Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science*, **351**, 725-728.
44. Latrick, C.M., Marek, M., Ouararhni, K., Papin, C., Stoll, I., Ignatyeva, M., Obri, A., Ennifar, E., Dimitrov, S., Romier, C. *et al.* (2016) Molecular basis and specificity of H2A.Z-H2B recognition and deposition by the histone chaperone YL1. *Nat Struct Mol Biol*, **23**, 309-316.
45. Liang, X., Shan, S., Pan, L., Zhao, J., Ranjan, A., Wang, F., Zhang, Z., Huang, Y., Feng, H., Wei, D. *et al.* (2016) Structural basis of H2A.Z recognition by SRCAP chromatin-remodeling subunit YL1. *Nat Struct Mol Biol*, **23**, 317-323.
46. Spedale, G., Timmers, H.T. and Pijnappel, W.W. (2012) ATAC-king the complexity of SAGA during evolution. *Genes Dev*, **26**, 527-541.
47. Helmlinger, D., Hardy, S., Sasorith, S., Klein, F., Robert, F., Weber, C., Miguet, L., Potier, N., Van-Dorselaer, A., Wurtz, J.M. *et al.* (2004) Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Hum Mol Genet*, **13**, 1257-1265.
48. Diebold, M.L., Fribourg, S., Koch, M., Metzger, T. and Romier, C. (2011) Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. *J Struct Biol*, **175**, 178-188.
49. Kohler, A., Zimmerman, E., Schneider, M., Hurt, E. and Zheng, N. (2010) Structural basis for assembly and activation of the heterotetrameric SAGA histone H2B deubiquitinase module. *Cell*, **141**, 606-617.
50. Samara, N.L., Datta, A.B., Berndsen, C.E., Zhang, X., Yao, T., Cohen, R.E. and Wolberger, C. (2010) Structural insights into the assembly and function of the SAGA deubiquitinating module. *Science*, **328**, 1025-1029.
51. David, G., Abbas, N., Stevanin, G., Durr, A., Yvert, G., Cancel, G., Weber, C., Imbert, G., Saudou, F., Antoniou, E. *et al.* (1997) Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet*, **17**, 65-70.
52. Durand, A., Bonnet, J., Fournier, M., Chavant, V. and Schultz, P. (2014) Mapping the deubiquitination module within the SAGA complex. *Structure*, **22**, 1553-1559.
53. Choudhary, C., Weinert, B.T., Nishida, Y., Verdin, E. and Mann, M. (2014) The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol*, **15**, 536-550.
54. Marmorstein, R. and Zhou, M.M. (2014) Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harb Perspect Biol*, **6**, a018762.
55. Keller, C.I. and Akhtar, A. (2015) The MSL complex: juggling RNA-protein interactions for dosage compensation and beyond. *Curr Opin Genet Dev*, **31**, 1-11.
56. Hallacli, E., Lipp, M., Georgiev, P., Spielman, C., Cusack, S., Akhtar, A. and Kadlec, J. (2012) Msl1-mediated dimerization of the dosage compensation complex is essential for male X-chromosome regulation in *Drosophila*. *Mol Cell*, **48**, 587-600.
57. Huang, J., Wan, B., Wu, L., Yang, Y., Dou, Y. and Lei, M. (2012) Structural insight into the regulation of MOF in the male-specific lethal complex and the non-specific lethal complex. *Cell Res*, **22**, 1078-1081.
58. Kadlec, J., Hallacli, E., Lipp, M., Holz, H., Sanchez-Weatherby, J., Cusack, S. and Akhtar, A. (2011) Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat Struct Mol Biol*, **18**, 142-149.
59. Xu, P., Li, C., Chen, Z., Jiang, S., Fan, S., Wang, J., Dai, J., Zhu, P. and Chen, Z. (2016) The NuA4 Core Complex Acetylates Nucleosomal Histone H4 through a Double Recognition Mechanism. *Mol Cell*, **63**, 965-975.

60. Doyon, Y. and Cote, J. (2004) The highly conserved and multifunctional NuA4 HAT complex. *Curr Opin Genet Dev*, **14**, 147-154.
61. Selleck, W., Fortin, I., Sermwittayawong, D., Cote, J. and Tan, S. (2005) The *Saccharomyces cerevisiae* Piccolo NuA4 histone acetyltransferase complex requires the Enhancer of Polycomb A domain and chromodomain to acetylate nucleosomes. *Mol Cell Biol*, **25**, 5535-5542.
62. Chittuluru, J.R., Chaban, Y., Monnet-Saksouk, J., Carrozza, M.J., Sapountzi, V., Selleck, W., Huang, J., Utley, R.T., Cramet, M., Allard, S. *et al.* (2011) Structure and nucleosome interaction of the yeast NuA4 and Piccolo-NuA4 histone acetyltransferase complexes. *Nat Struct Mol Biol*, **18**, 1196-1203.
63. Lombardi, P.M., Cole, K.E., Dowling, D.P. and Christianson, D.W. (2011) Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes. *Curr Opin Struct Biol*, **21**, 735-743.
64. Moniot, S., Weyand, M. and Steegborn, C. (2012) Structures, substrates, and regulators of Mammalian sirtuins - opportunities and challenges for drug development. *Front Pharmacol*, **3**, 16.
65. Millard, C.J., Watson, P.J., Fairall, L. and Schwabe, J.W. (2017) Targeting Class I Histone Deacetylases in a "Complex" Environment. *Trends Pharmacol Sci*, **38**, 363-377.
66. West, A.C. and Johnstone, R.W. (2014) New and emerging HDAC inhibitors for cancer treatment. *J Clin Invest*, **124**, 30-39.
67. Marek, M., Oliveira, G., Pierce, R.J., Jung, M., Sippl, W. and Romier, C. (2015) Drugging the schistosome zinc-dependent HDACs: current progress and future perspectives. *Future Med Chem*, **7**, 783-800.
68. Shuttleworth, S.J., Bailey, S.G. and Townsend, P.A. (2010) Histone Deacetylase inhibitors: new promise in the treatment of immune and inflammatory diseases. *Curr Drug Targets*, **11**, 1430-1438.
69. Chakrabarti, A., Oehme, I., Witt, O., Oliveira, G., Sippl, W., Romier, C., Pierce, R.J. and Jung, M. (2015) HDAC8: a multifaceted target for therapeutic interventions. *Trends Pharmacol Sci*, **36**, 481-492.
70. Maolanon, A.R., Kristensen, H.M., Leman, L.J., Ghadiri, M.R. and Olsen, C.A. (2017) Natural and Synthetic Macrocyclic Inhibitors of the Histone Deacetylase Enzymes. *Chembiochem*, **18**, 5-49.
71. Millard, C.J., Watson, P.J., Celardo, I., Gordiyenko, Y., Cowley, S.M., Robinson, C.V., Fairall, L. and Schwabe, J.W. (2013) Class I HDACs share a common mechanism of regulation by inositol phosphates. *Mol Cell*, **51**, 57-67.
72. Watson, P.J., Fairall, L., Santos, G.M. and Schwabe, J.W. (2012) Structure of HDAC3 bound to co-repressor and inositol tetrakisphosphate. *Nature*, **481**, 335-340.
73. Watson, P.J., Millard, C.J., Riley, A.M., Robertson, N.S., Wright, L.C., Godage, H.Y., Cowley, S.M., Jamieson, A.G., Potter, B.V. and Schwabe, J.W. (2016) Insights into the activation mechanism of class I HDAC complexes by inositol phosphates. *Nat Commun*, **7**, 11262.
74. Hai, Y. and Christianson, D.W. (2016) Histone deacetylase 6 structure and molecular basis of catalysis and inhibition. *Nat Chem Biol*, **12**, 741-747.
75. Miyake, Y., Keusch, J.J., Wang, L., Saito, M., Hess, D., Wang, X., Melancon, B.J., Helquist, P., Gut, H. and Matthias, P. (2016) Structural insights into HDAC6 tubulin deacetylation and its selective inhibition. *Nat Chem Biol*, **12**, 748-754.
76. Heimbürg, T., Chakrabarti, A., Lancelot, J., Marek, M., Melesina, J., Hauser, A.T., Shaik, T.B., Duclaud, S., Robaa, D., Erdmann, F. *et al.* (2016) Structure-Based Design and Synthesis of Novel Inhibitors Targeting HDAC8 from *Schistosoma mansoni* for the Treatment of Schistosomiasis. *J Med Chem*, **59**, 2423-2435.

77. Kannan, S., Melesina, J., Hauser, A.T., Chakrabarti, A., Heimburg, T., Schmidtkunz, K., Walter, A., Marek, M., Pierce, R.J., Romier, C. *et al.* (2014) Discovery of inhibitors of *Schistosoma mansoni* HDAC8 by combining homology modeling, virtual screening, and in vitro validation. *J Chem Inf Model*, **54**, 3005-3019.
78. Marek, M., Kannan, S., Hauser, A.T., Moraes Mourao, M., Caby, S., Cura, V., Stolfa, D.A., Schmidtkunz, K., Lancelot, J., Andrade, L. *et al.* (2013) Structural basis for the inhibition of histone deacetylase 8 (HDAC8), a key epigenetic player in the blood fluke *Schistosoma mansoni*. *PLoS Pathog*, **9**, e1003645.
79. Stolfa, D.A., Marek, M., Lancelot, J., Hauser, A.T., Walter, A., Leproult, E., Melesina, J., Rumpf, T., Wurtz, J.M., Cavarelli, J. *et al.* (2014) Molecular basis for the antiparasitic activity of a mercaptoacetamide derivative that inhibits histone deacetylase 8 (HDAC8) from the human pathogen *Schistosoma mansoni*. *J Mol Biol*, **426**, 3442-3453.
80. English, C.M., Adkins, M.W., Carson, J.J., Churchill, M.E. and Tyler, J.K. (2006) Structural basis for the histone chaperone activity of Asf1. *Cell*, **127**, 495-508.
81. Natsume, R., Eitoku, M., Akai, Y., Sano, N., Horikoshi, M. and Senda, T. (2007) Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. *Nature*, **446**, 338-341.
82. Huang, H., Stromme, C.B., Saredi, G., Hodl, M., Strandsby, A., Gonzalez-Aguilera, C., Chen, S., Groth, A. and Patel, D.J. (2015) A unique binding mode enables MCM2 to chaperone histones H3-H4 at replication forks. *Nature structural & molecular biology*, **22**, 618-626.
83. Richet, N., Liu, D., Legrand, P., Velours, C., Corpet, A., Gaubert, A., Bakail, M., Moal-Raisin, G., Guerois, R., Compper, C. *et al.* (2015) Structural insight into how the human helicase subunit MCM2 may act as a histone chaperone together with ASF1 at the replication fork. *Nucleic Acids Res*, **43**, 1905-1917.
84. Wang, H., Wang, M., Yang, N. and Xu, R.M. (2015) Structure of the quaternary complex of histone H3-H4 heterodimer with chaperone ASF1 and the replicative helicase subunit MCM2. *Protein Cell*, **6**, 693-697.
85. Saredi, G., Huang, H., Hammond, C.M., Alabert, C., Bekker-Jensen, S., Forne, I., Reveron-Gomez, N., Foster, B.M., Mlejnkova, L., Bartke, T. *et al.* (2016) H4K20me0 marks post-replicative chromatin and recruits the TONSL-MMS22L DNA repair complex. *Nature*, **534**, 714-718.
86. Cho, U.S. and Harrison, S.C. (2011) Recognition of the centromere-specific histone Cse4 by the chaperone Scm3. *Proc Natl Acad Sci U S A*, **108**, 9367-9371.
87. Hu, H., Liu, Y., Wang, M., Fang, J., Huang, H., Yang, N., Li, Y., Wang, J., Yao, X., Shi, Y. *et al.* (2011) Structure of a CENP-A-histone H4 heterodimer in complex with chaperone HJURP. *Genes Dev*, **25**, 901-906.
88. Zhou, Z., Feng, H., Zhou, B.R., Ghirlando, R., Hu, K., Zwolak, A., Miller Jenkins, L.M., Xiao, H., Tjandra, N., Wu, C. *et al.* (2011) Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. *Nature*, **472**, 234-237.
89. Elsasser, S.J., Huang, H., Lewis, P.W., Chin, J.W., Allis, C.D. and Patel, D.J. (2012) DAXX envelops a histone H3.3-H4 dimer for H3.3-specific recognition. *Nature*, **491**, 560-565.
90. Liu, C.P., Xiong, C., Wang, M., Yu, Z., Yang, N., Chen, P., Zhang, Z., Li, G. and Xu, R.M. (2012) Structure of the variant histone H3.3-H4 heterodimer in complex with its chaperone DAXX. *Nature structural & molecular biology*, **19**, 1287-1292.
91. Ricketts, M.D., Frederick, B., Hoff, H., Tang, Y., Schultz, D.C., Singh Rai, T., Grazia Vizioli, M., Adams, P.D. and Marmorstein, R. (2015) Ubinuclein-1 confers histone H3.3-specific-binding by the HIRA histone chaperone complex. *Nat Commun*, **6**, 7711.
92. Huang, H., Deng, Z., Vladimirova, O., Wiedmer, A., Lu, F., Lieberman, P.M. and Patel, D.J. (2016) Structural basis underlying viral hijacking of a histone chaperone complex. *Nat Commun*, **7**, 12707.

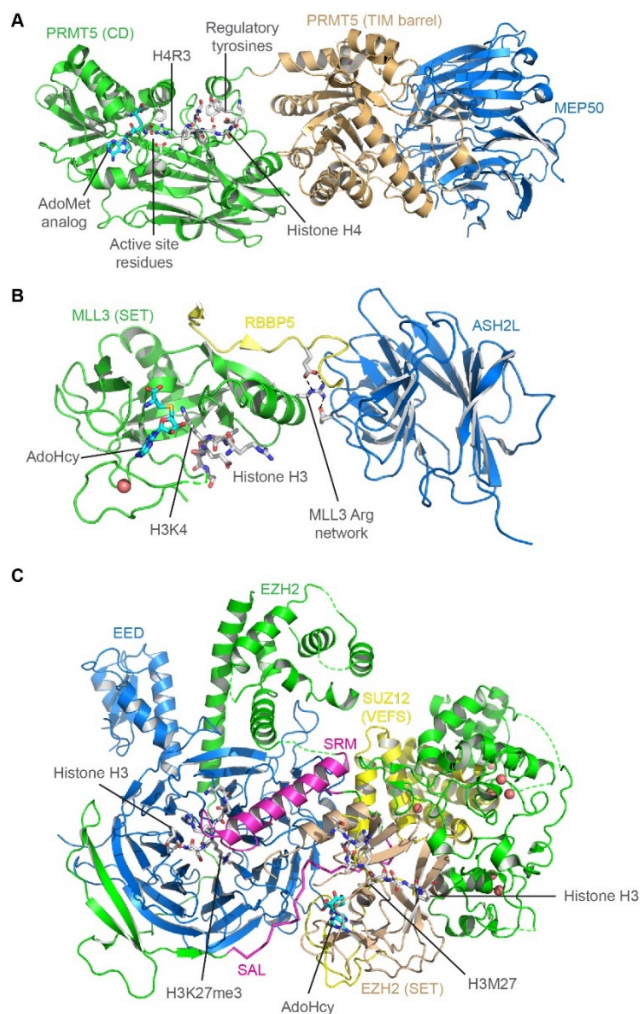
93. Obri, A., Ouararhni, K., Papin, C., Diebold, M.L., Padmanabhan, K., Marek, M., Stoll, I., Roy, L., Reilly, P.T., Mak, T.W. *et al.* (2014) ANP32E is a histone chaperone that removes H2A.Z from chromatin. *Nature*, **505**, 648-653.
94. Mao, Z., Pan, L., Wang, W., Sun, J., Shan, S., Dong, Q., Liang, X., Dai, L., Ding, X., Chen, S. *et al.* (2014) Anp32e, a higher eukaryotic histone chaperone directs preferential recognition for H2A.Z. *Cell Res*, **24**, 389-399.
95. Bartholomew, B. (2014) Regulating the chromatin landscape: structural and mechanistic perspectives. *Annual review of biochemistry*, **83**, 671-696.
96. Hauk, G., McKnight, J.N., Nodelman, I.M. and Bowman, G.D. (2010) The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor. *Mol Cell*, **39**, 711-723.
97. Metzger, E., Willmann, D., McMillan, J., Forne, I., Metzger, P., Gerhardt, S., Petroll, K., von Maessenhausen, A., Urban, S., Schott, A.K. *et al.* (2016) Assembly of methylated KDM1A and CHD1 drives androgen receptor-dependent transcription and translocation. *Nat Struct Mol Biol*, **23**, 132-139.
98. Yan, L., Wang, L., Tian, Y., Xia, X. and Chen, Z. (2016) Structure and regulation of the chromatin remodeller ISWI. *Nature*, **540**, 466-469.
99. Yamada, K., Frouws, T.D., Angst, B., Fitzgerald, D.J., DeLuca, C., Schimmele, K., Sargent, D.F. and Richmond, T.J. (2011) Structure and mechanism of the chromatin remodelling factor ISW1a. *Nature*, **472**, 448-453.
100. Schubert, H.L., Wittmeyer, J., Kasten, M.M., Hinata, K., Rawling, D.C., Heroux, A., Cairns, B.R. and Hill, C.P. (2013) Structure of an actin-related subcomplex of the SWI/SNF chromatin remodeler. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 3345-3350.
101. Xia, X., Liu, X., Li, T., Fang, X. and Chen, Z. (2016) Structure of chromatin remodeler Swi2/Snf2 in the resting state. *Nat Struct Mol Biol*, **23**, 722-729.
102. Liu, X., Li, M., Xia, X., Li, X. and Chen, Z. (2017) Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature*, **544**, 440-445.
103. Sabari, B.R., Zhang, D., Allis, C.D. and Zhao, Y. (2017) Metabolic regulation of gene expression through histone acylations. *Nat Rev Mol Cell Biol*, **18**, 90-101.
104. Andrews, F.H., Shinsky, S.A., Shanle, E.K., Bridgers, J.B., Gest, A., Tsun, I.K., Krajewski, K., Shi, X., Strahl, B.D. and Kutateladze, T.G. (2016) The Taf14 YEATS domain is a reader of histone crotonylation. *Nat Chem Biol*, **12**, 396-398.
105. Xiong, X., Panchenko, T., Yang, S., Zhao, S., Yan, P., Zhang, W., Xie, W., Li, Y., Zhao, Y., Allis, C.D. *et al.* (2016) Selective recognition of histone crotonylation by double PHD fingers of MOZ and DPF2. *Nat Chem Biol*, **12**, 1111-1118.
106. Zhang, Q., Zeng, L., Zhao, C., Ju, Y., Konuma, T. and Zhou, M.M. (2016) Structural Insights into Histone Crotonyl-Lysine Recognition by the AF9 YEATS Domain. *Structure*, **24**, 1606-1612.
107. Li, Y., Sabari, B.R., Panchenko, T., Wen, H., Zhao, D., Guan, H., Wan, L., Huang, H., Tang, Z., Zhao, Y. *et al.* (2016) Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Mol Cell*, **62**, 181-193.
108. Wilson, M.D., Benlekbir, S., Fradet-Turcotte, A., Sherker, A., Julien, J.P., McEwan, A., Noordermeer, S.M., Sicheri, F., Rubinstein, J.L. and Durocher, D. (2016) The structural basis of modified nucleosome recognition by 53BP1. *Nature*, **536**, 100-103.
109. Kuhlbrandt, W. (2014) Biochemistry. The resolution revolution. *Science*, **343**, 1443-1444.
110. Doudna, J.A. and Charpentier, E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, **346**, 1258096.



**Figure 1. Structures of DNA methyltransferases.**

Ribbon representation of DNA methyltransferases **A,B**. Structures of the de novo demethylation complex DNMT3A/DNMT3L in inhibitory (**A**) and activated (**B**) states. DNMT3A catalytic domain (CD) is colored green and its ADD (ATRX-DNMT3-DNMT3L) domain is colored wheat. DNMT3L inactive catalytic domain is colored blue. Zinc ions are shown as red spheres. Histone H3 activating peptide as well as important DNMT3A aspartate residues are represented as sticks with gray carbons. An AdoHcy (S-Adenosyl-L-homocysteine) molecule is represented as sticks with cyan carbons. The coloring is identical in all figures unless stated. **C,D,E**. Structures of methyl mark maintenance DNA methyltransferase DNMT1 in inhibitory (**A**) and active (**B**) states, and in a stabilizing complex with USP7 (**C**). DNMT1 is shown in green, DNA in orange and USP7 in blue. DNMT1 zinc binding domain (Zn BD; wheat) recognizing unmethylated DNA is shown as well as the inhibitory and regulatory GK linkers (magenta ribbons).

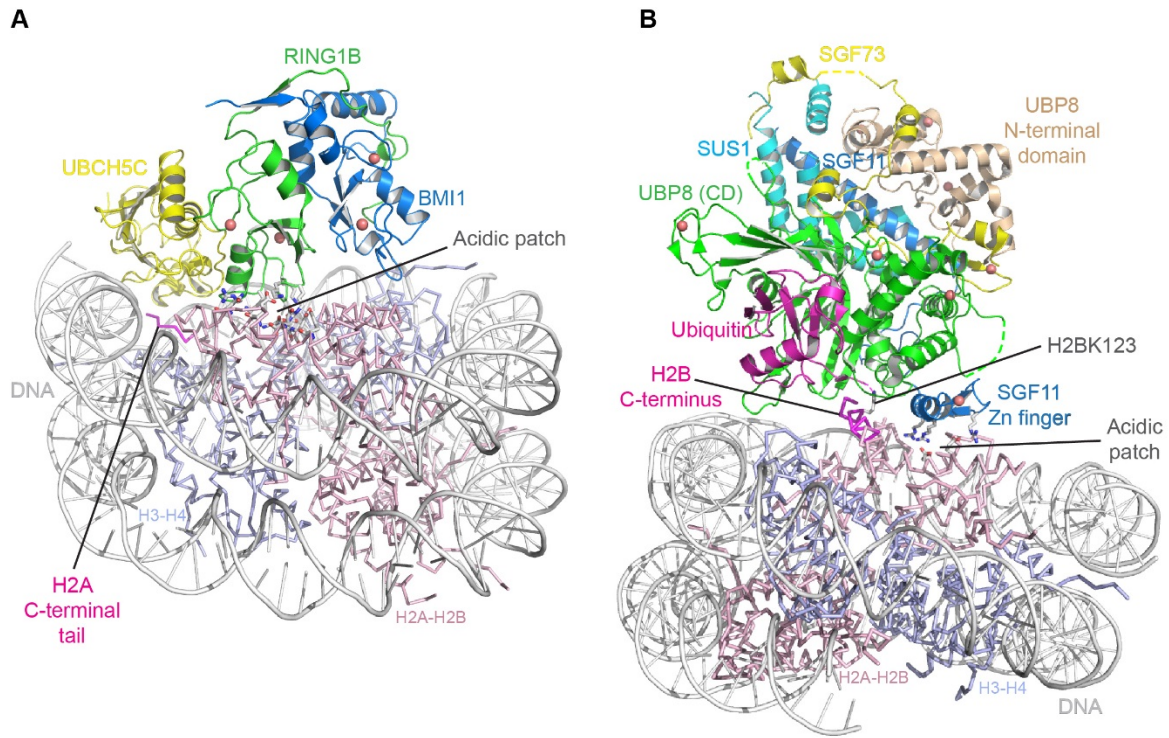




**Figure 2. Structures of arginine and lysine methyltransferases.**

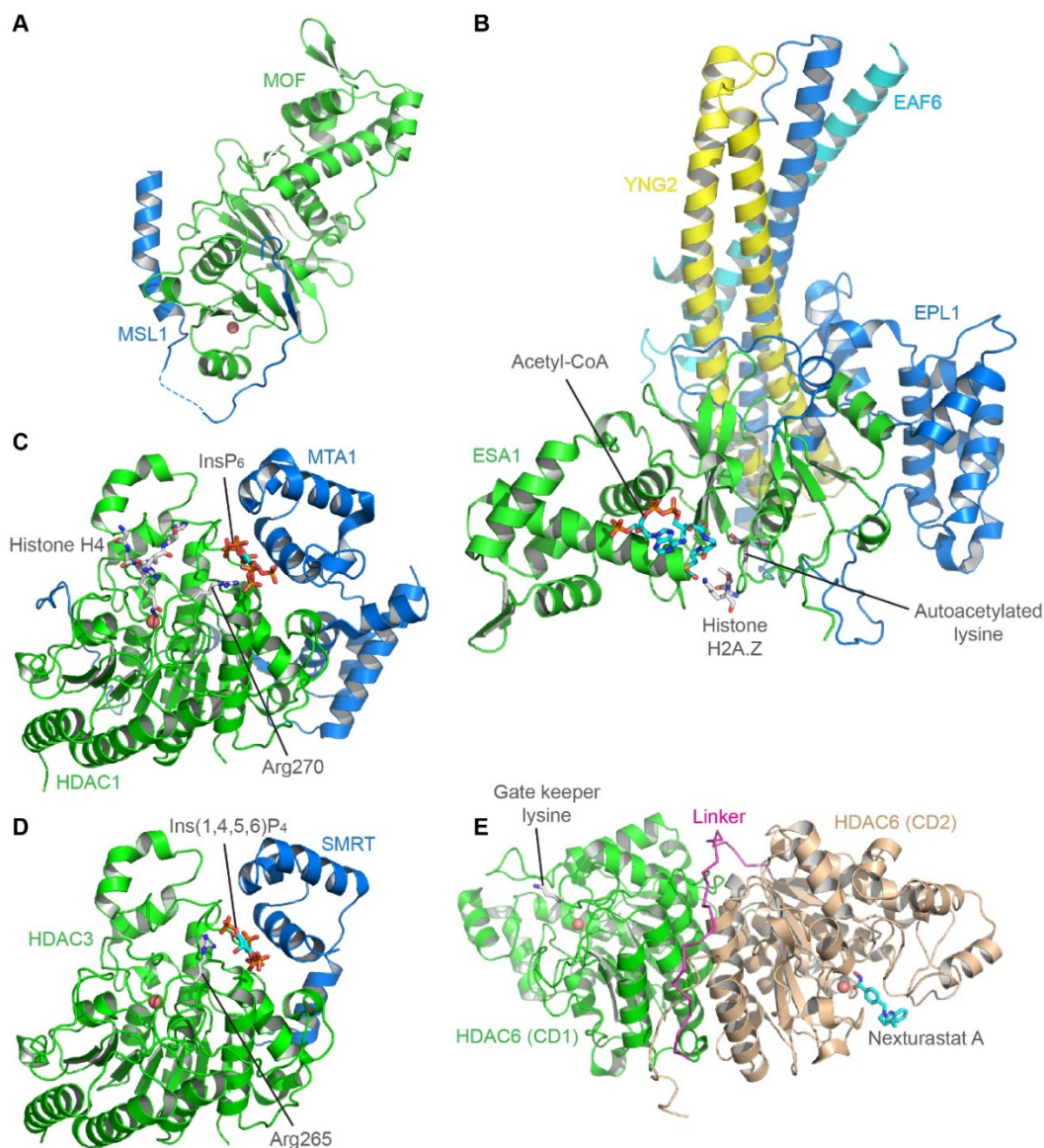
Ribbon representation of arginine and lysine methyltransferases. **A.** Structure of the PRMT5/MEP50 dimer. PRMT5 arginine methyltransferase catalytic domain (CD) is colored green, its TIM-barrel domain is colored wheat, and MEP50 is colored blue. Histone H4 peptide binding to PRMT5 CD is shown as sticks with gray carbon as well as PRMT5 important active site residues and regulatory tyrosines. An AdoMet (S-adenosyl-L-methionine) analog bound to PRMT5 CD is shown as sticks with cyan carbons. **B.** Structure of lysine methyltransferase MLL3 (SET domain; green) in complex with a RBBP5 peptide (yellow) and ASH2L C-terminal domain (blue). Histone H3 N-terminal peptide and an AdoHcy molecule are shown as sticks with gray and cyan carbons, respectively. Residues at the interface of the three proteins that form a hydrogen bond network (MLL3 Arg network) are also shown as sticks. **C.** Structure of the Polycomb Repressive Complex 2 (PRC2; EZH2-EED-SUZ12). Methyltransferase EZH2 is colored green except its catalytic SET domain that is colored wheat. EED is colored blue and the SUZ12 VEF5 domain is colored yellow. An AdoHcy molecule bound to EZH2 SET domain is shown as sticks with cyan ribbons. Two H3 N-terminal peptides are shown as sticks with gray carbon: a H3K27me3 peptide bound to EED and a mutated H3M27 peptide bound to EZH2 SET domain. Two regulatory elements elements are shown in magenta: the SET activation loop (SAL; shown as ribbon) and the Stimulation-responsive motif (SRM). This latter motif is suggested to transmit the signal of H3K27me3 binding to the EED subunit to the catalytic SET domain.





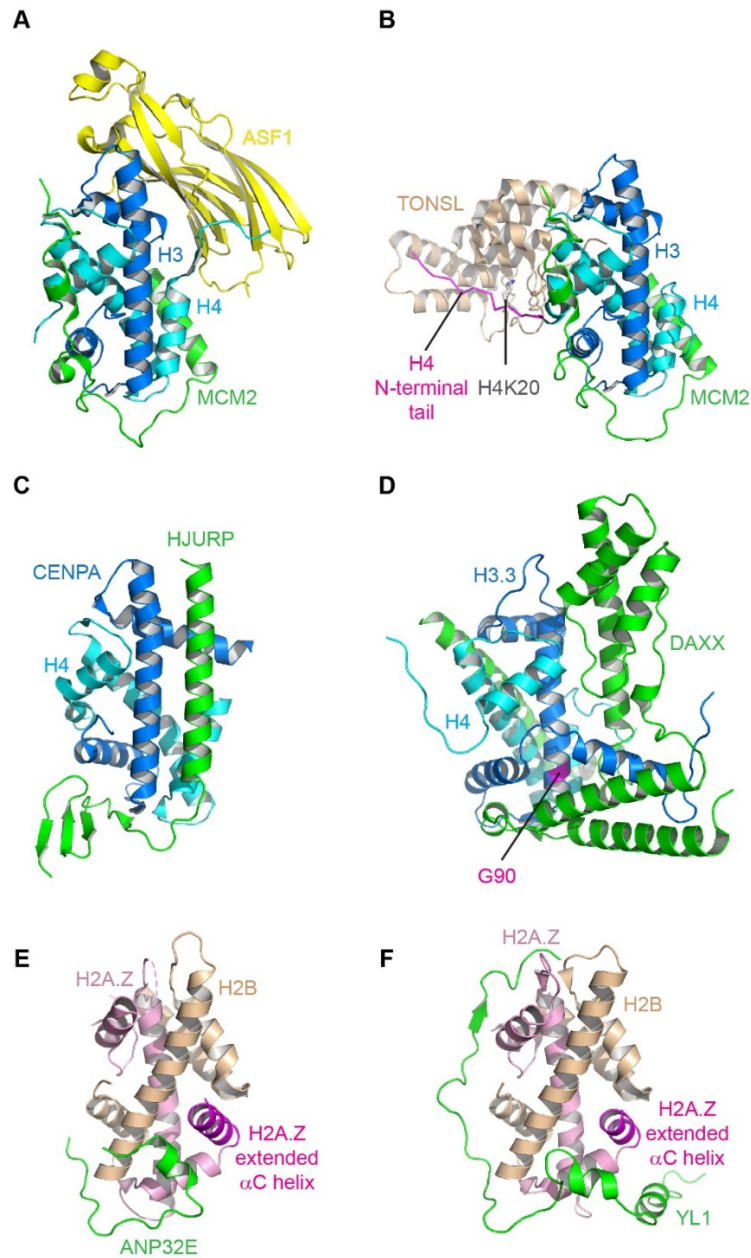
**Figure 3. Structures of (de-)ubiquitinylases.**

Ribbon representation of (de-)ubiquitinylases. **A.** Structure of the Polycomb Repressive Complex 1 (PRC1; RING1B-BMI1-UBCH5C) ubiquitinylase bound to the nucleosome. RING1B is colored green, BMI1 blue and UBCH5C yellow. The DNA is colored white and the H3-H4 and H2A-H2B pairs are colored light blue and light pink, respectively. The nucleosome acidic patch residues as well as RING1B positively charged residues that form a strong interaction network are shown as sticks with gray carbons. The H2A C-terminal tail targeted by UBCH5C is shown as magenta ribbon. **B.** Structure of the deubiquitination module (DUBm; UB8-SGF11-SUS1-SGF73) of the SAGA co-activator bound to the nucleosome. UB8 N-terminal domain is colored wheat, its catalytic domain (CD) is colored green. SGF11 is colored blue for its N-terminal domain and dark blue for its zinc finger C-terminal domain (Zn finger). SUS1 is colored cyan and SGF73 N-terminal region is colored yellow. SGF11 positively charged residues and nucleosome active site residues are shown as stick with gray carbons. H2B C-terminal helix is shown in magenta as well as the ubiquitin molecule bound to H2BK123 residue (sticks with gray carbons). The nucleosome is displayed as in (A).



**Figure 4. Structures of histone acetyltransferases and deacetylases.**

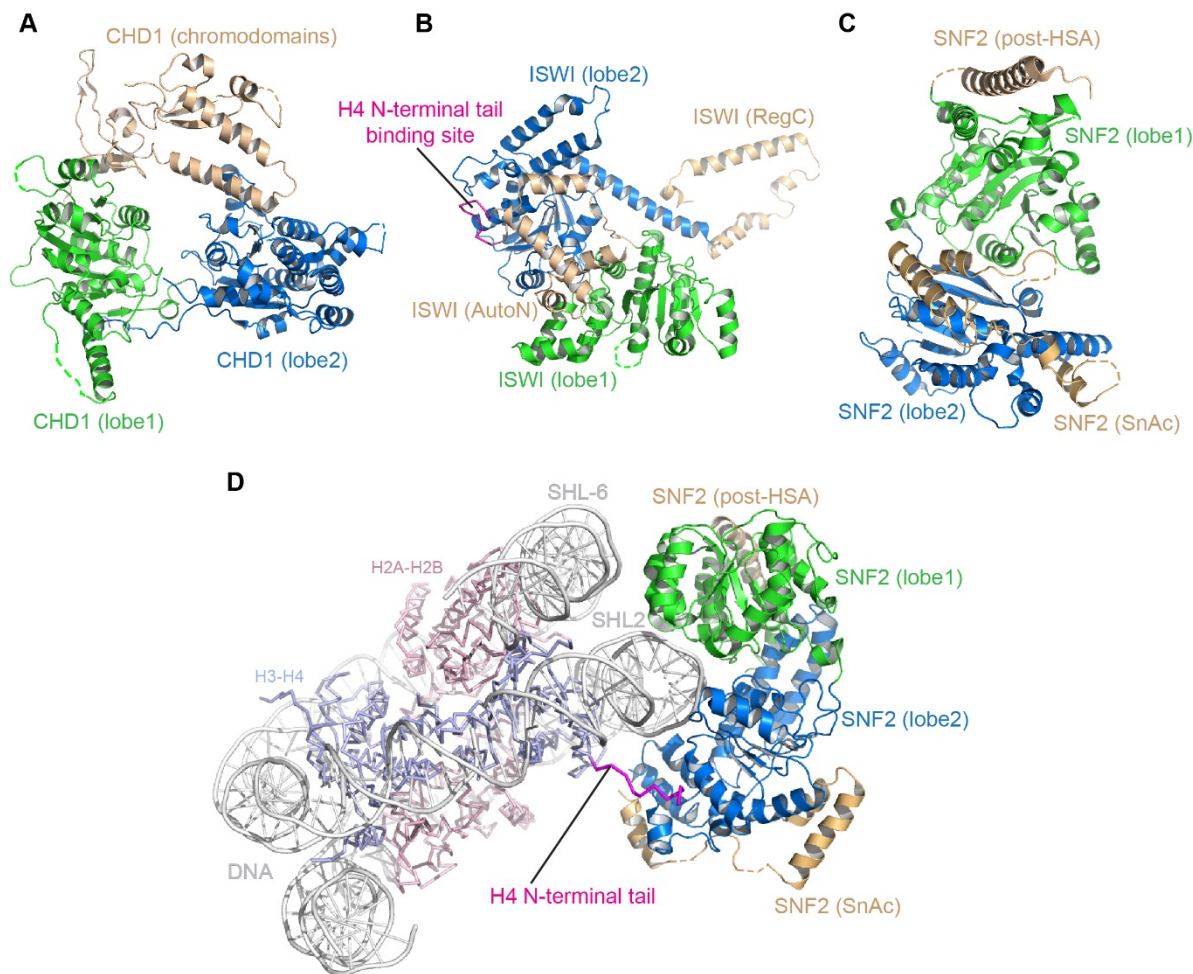
Ribbon representation of histone acetyltransferases (HATs) and deacetylases (HDACs). **A.** Structure of MOF HAT domain (Jaiswal, Turniansky *et al.*) bound to MSL1 (blue). **B.** Structure of the NUA4 HAT complex (ESA1-EPL1-YNG2-EAF6). ESA1 is colored green, EPL1 blue, YNG2 Yellow and EAF6 cyan. Acetyl-CoA is shown as sticks with cyan carbons and histone H2A.Z peptide as well as an autoacetylated lysine as sticks with gray carbons. **C,D.** Structures of HDAC1 (C) and HDAC3 (D) in complex with co-repressors MTA1 and SMRT, respectively, and with inositol phosphate molecules. HDACs are colored green and co-repressors blue, with inositol phosphate molecules shown as sticks with cyan carbons. Histone H4 peptide analog and important arginine residues as shown as sticks with gray carbons. **E.** Structure of HDAC6 tandem CD1-CD2 HDAC domains. CD1 is shown in green and CD2 in wheat. The linker connecting both HDAC domains is shown as magenta ribbon. CD1 gate keeper lysine and HDAC6 selective inhibitor Nexturastat A molecule bound to CD2 are shown as sticks with gray and cyan carbons, respectively.



**Figure 5. Structures of histone chaperones-histone pairs complexes.**

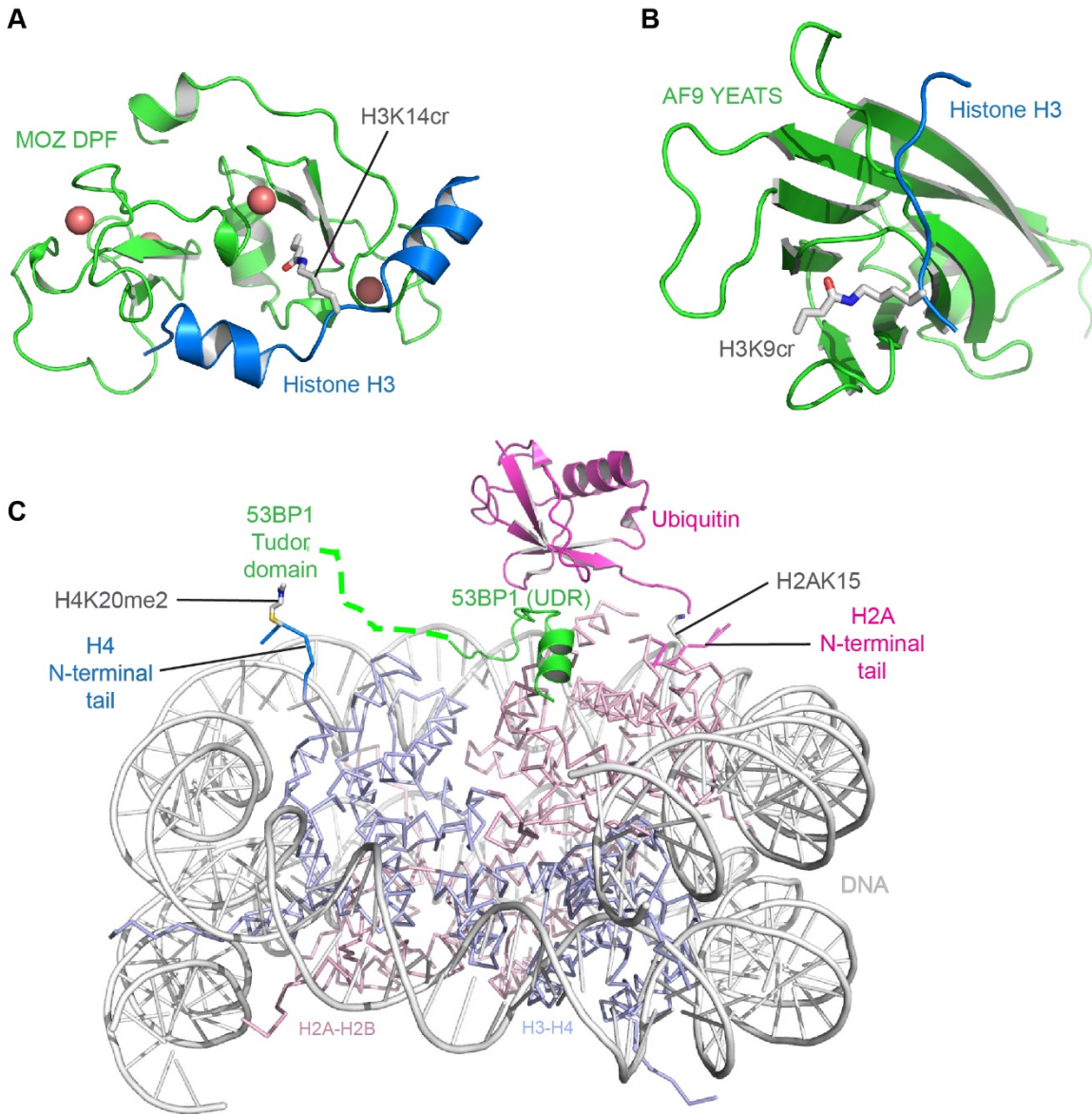
Ribbon representation of histone chaperones-histone pairs complexes. H3 (and related histone variants) and H4 are colored blue and cyan, respectively, H2A (and related histone variants) and H2B are colored light pink and wheat, respectively. **A.** Structure of the MCM2-ASF1-H3-H4 complex. MCM2 is colored green and ASF1 yellow. **B.** Structure of the MCM2-TONSL-H3-H4 complex. MCM2 is colored green and TONSL wheat. H4 N-terminal tail is shown as magenta ribbon and H4K20 as sticks with gray carbons. **C.** Structure of HJURP in complex with the CENPA-H4 variant pair. HJURP is colored green. **D.** Structure of the DAXX-H3.3-H4 complex. DAXX is colored green. G90 that conveys H3.3 specific recognition is colored magenta. **E,F.** Structures of H2A.Z-H2B variant pair in complex with ANP32E removal (E) and YL1 deposition (F) histone chaperones. Both chaperones are colored green. The extended  $\alpha$ C helix from H2A.Z is colored magenta.





**Figure 6. Structures of ATP-dependent chromatin remodeling subunits.**

Ribbon representation of ATP-dependent chromatin remodeling subunits. The RecA-like lobe 1 and lobe 2 of the ATPase domain are colored green and blue, respectively, the auxiliary domains are colored wheat. **A.** Structure of CHD1. The double chromodomains of CHD1 link the two ATPase lobes keeping them in an inactive conformation. **B.** Structure of ISWI. The AutoN domain also links the two ATPase lobes and keep them in an inactive conformation. The part of the AutoN domain that binds to the H4-binding pocket of lobe 2 is shown as magenta ribbon. **C.** Structure of SNF2 in an inactive conformation. **D.** Structure of SNF2 bound to the nucleosome. The same color coding for SNF2 is as in (C). The color coding for the nucleosome is as in Fig. 3. The H4 N-terminal tail that binds to SNF2 lobe 2 is shown as magenta ribbon. SNF2 is bound to the nucleosome in an active conformation and interacts with the nucleosome DNA at SHL2 and SHL-6.



**Figure 7. Structures of epigenetic readers.**

Ribbon representation of epigenetic readers. **A,B.** Structures of crotonylation readers: MOZ double PHD fingers bound to H3K14cr and AF9 YEATS domain bound to H3K9cr. The readers are colored green and histone H3 is colored blue. The crotonylated residues are shown as sticks with gray carbons. **C.** Structure of 53BP1 bound to the H4K20 dimethylated and H2AK15 ubiquitylated nucleosome. The nucleosome is colored as in Fig. 3. 53BP1 is colored green. 53BP1 ubiquitin-dependent recruitment motif (UDR) is shown as ribbon. 53BP1 Tudor domain was not provided in the model deposited and is just indicated. The H4 N-terminal tail is shown as blue ribbon and the H2A N-terminal tail as magenta ribbon. H4K20me2 and H2AK15 are shown as sticks with gray carbons. Ubiquitin is colored magenta.

## 1.2. The acetylation mark

### 1.2.1. Role of the acetyl epigenetic mark

#### 1.2.1.1. Acetylation and mechanisms of acetylation

Acetylation is one of the major post-translational modifications used by the cells for signaling (Allfrey, Faulkner et al. 1964). The opposing action of histone acetyltransferases (HATs) and zinc dependent histone deacetylases (HDACs) and sirtuins, which are responsible for the addition and deletion of acetyl group to the  $\epsilon$  amino group of lysine amino acids, helps maintain the balance of cellular acetylation. Interestingly, there is a direct link between acetylation and the metabolism. Indeed, as for many other post-translational modifications (PTMs), the acetylation mark links small molecules such as acetyl-CoA to epigenetics, which, in turn, influence homeostasis, development and aging (Menzies, Zhang et al. 2016).

Specifically, the number of non-histone substrates of histone acetyltransferases and histone deacetylases has increased dramatically from the time of discovery of these enzymes. Therefore, these enzymes are generally designated as lysine acetyltransferases (KAT) and lysine deacetylases (KDAC). However, for conventional reasons, these two classes will remain described as HATs and HDACs/sirtuins in this thesis.

Acetylation can occur in highly conserved residues in secondary structural elements like  $\alpha$  helices and  $\beta$  strands. In bacteria and mitochondria, acetylation is the major PTM before phosphorylation, which supports the endosymbiotic theory of mitochondria originating from prokaryotic bacteria.

Acetylation masks the positive charge of lysine side chain, and this affects protein interactions with DNA, with partner proteins, as well as subcellular localization. A major function of acetylation is the modulation of histone-DNA and histone-histone interactions, which in turn regulates nuclear processes. Specifically, the neutralized charge on acetylated lysine amino acids weakens these interaction, prevents compaction of the chromatin structure, and allows nuclear players to bind to DNA.

H4K16 acetylation is one of the important PTM that keeps chromatin in a non-compact state by preventing interaction with the H2A-H2B acidic patch. Sas2 and Sir2 maintain acetylation and deacetylation of H4K16 (Cavero, Herruzo et al. 2016). Near the telomeres H4K16 is hypo acetylated and away from telomeres it is hyper acetylated in order to maintain gene regulation.

#### 1.2.1.2. A brief history of acetylation

The discoveries of coenzyme A (CoA) and acetyl-CoA by Nobel laureates Fritz Lipmann (for CoA - 1953), Konard Blach and Feodor Lyman (for acetyl-CoA - 1964) led the foundation for protein acetylation discovery (Verdin and Ott 2015). Protein acetylation was first reported by Vincent Allfrey and his colleagues in the year 1964 but the actual advancement in the field of protein acetylation was only made in the last 20 years (Allfrey, Faulkner *et al.* 1964, Verdin and Ott 2015). In the year 1963, Phillips *et al.* isolated histones from calf thymus with acetyl groups. Soon after, Vincent Allfrey *et al.* showed that radiolabelled acetyl groups were rapidly taken up by histones in the isolated nuclei and were insensitive to translation inhibitor puromycin which suggested the incorporation process was a post translational modification. Further, they showed that the acetylation site is the  $\epsilon$ -amino group of lysines. In the year 1970, the discoveries of Rigg's and his co-workers led Allfrey and Davie to suggest n-butyrate as HDAC inhibitor. This was the first time that a small molecule was suggested to alter cellular functions (Candido, Reeves et al. 1978).

After a long period, in the year 1983, tubulin was the first non-histone protein shown to be acetylated. L'Hernault and Rosenbann showed in a 2D electrophoresis gel a difference in the migration of radiolabelled acetyl tubulin. Further, in 1985, Piperno *et al.* developed monoclonal antibodies against acetyl K40 of tubulin. Later in 1990, lysine residues of histones were shown to be important for gene regulation in yeast cells. Later in 1995, Sternglanz *et al.* identified and cloned the first HAT of yeast, HAT1. A year later Allis *et al.* purified a 55KDa HAT from a protozoan, *Tetrahymena thermophile*, which was an orthologue of Gcn5, a yeast transcriptional regulator (Brownell and Allis 1995).

In the same year, Schreiber *et al.* purified HDAC1 from bovine thymus using tropoxin, an HDAC inhibitor affinity matrix, which was further shown to be an orthologue of Rpd3, a yeast

transcriptional regulator. Additional HATs discoveries followed as well as the characterization of their association with transcriptional regulators such as: the CREB binding protein (CBP), the E1a binding protein (EP300), TAF(Suzuki, Muto *et al.*) 250 of TFIID, and HATs from the MYST family. Further characterization of HATs led to the identification of large multiprotein complexes such as SAGA, Sin3 and NURD. Soon after, Cole and his co-workers developed an EP300 inhibitor, Lys-CoA, where CoA is covalently linked to a lysine residue to mimic the pseudo substrate (Lau, Kundu *et al.* 2000).

Soon after Rpd3 related class I HDACs were discovered, Hda1 related class II HDACs and HDAC10 and HDAC11 were identified (Taunton, Hassig *et al.* 1996). In the year 1999, the first crystal structure of HDAC was solved by Finnin *et al.* from *Aquifex aeolicus*, a hyperthermophile (Finnin, Donigian *et al.* 1999). In the same year, the structure of the HAT domain of PCAF and of the NAD<sup>+</sup>-bound Sir2 were solved. In the same year, the first epigenetic reader that recognized the acetylation PTM, the bromodomain, was discovered (Dhalluin, Carlson *et al.* 1999).

In the year 1997, non-histone protein acetylation was reported for p53. After that, a few more proteins like the HIV transcriptional activator (Shi, Hong *et al.*) protein, E2F, MyoD and the nuclear factor  $\kappa$ B (NF $\kappa$ B) were also showed to be acetylated. In 1999, Zhou *et al.* reported bromodomains which recognize acetylated lysine residues (Dhalluin, Carlson *et al.* 1999). Owens *et al.* crystallized the acetylated H4 peptide in complex with bromodomains of yeast Gcn5 and identified the canonical asparagine residue in the active site pocket which is important for the recognition of acetylated lysine and that is shared by almost all existing 8 classes of bromodomains. Later, YEATS domains were identified as readers of acetylated lysine and shown to possess, to some extent, H3K9ac selectivity (Li, Wen *et al.* 2014).

Guarente *et al.* in 2000 showed that Sir2 proteins are a class of deacetylases and depends on NAD<sup>+</sup> for activity (Imai, Armstrong *et al.* 2000). The authors further derived a link between NAD<sup>+</sup> and nutrients levels, where upon the depletion of nutrients, NAD<sup>+</sup> levels increase which are used by Sir2 to deacetylase histones to repress the transcription. In the year 2006, a first proteomic screen was performed by Zhao *et al.* and identified 388 acetylation sites over 195 proteins (Hallows, Lee *et al.* 2006). The knowledge of acetylation sites has increased dramatically with this experiment when compared to the progress achieved from the beginning of acetylation discovery.



In the same year, SIRT3 was identified to deacetylate mitochondrial acetyl-CoA synthetase, and introduced mitochondria to the acetylation group (Schwer, Bunkenborg et al. 2006).

In 2009, Mann *et al.* used high resolution mass spectrometry to identify 3600 lysine acetylation sites over 1750 proteins, prominently in large protein complexes that are involved in chromatin modulation, cell cycle, splicing, nuclear transport and actin nucleation (Choudhary, Kumar et al. 2009). Further, the acetylation of lysine was proposed to be a part of other acyl modifications such as propionylation, butyrylation, O-acetylation, a concept introduced in 2007. Lysine succinylate and lysine malonate were showed to be acylated by SIRT5 in 2011 (Peng, Lu et al. 2011).

In the year 2003, Sinclair *et al.* discovered polyphenol resveratrol as a SIRT1 activator which was showed to increase the life span of yeasts and metazoans (Howitz, Bitterman et al. 2003). Vorinostat (SAHA) was the first FDA approved HDAC inhibitor in the year 2006 to treat T-cell lymphomas (Mann, Johnson et al. 2007, Zuber, Shi et al. 2011). Further, in 2010 JQ1 and I-BET inhibitors were proposed for bromodomains (Zuber, Shi et al. 2011). Cole *et al.* introduced HAT inhibitors for EP300 which are highly selective (Bowers, Yan et al. 2010).

### 1.2.1.3. A handshake of acetylation with other PTMs

Acetylation functions also in association with other PTMs such as ubiquitination and methylation. In human, nearly one third of acetylation sites may also act as sites of ubiquitination. Competitive regulation of acetylation and ubiquitination has been observed in many proteins. For instance, SMAD7 a negative regulator of TGF $\beta$  (transforming growth factor  $\beta$ ), upon acetylation of K64 and K70, prevents its degradation by restricting ubiquitination. Deacetylation of these two residues promotes ubiquitination and further degradation of SMAD7. Similarly, acetylation of p53 blocks ubiquitination and hence prevents its degradation.

During DNA damage, acetylation and methylation cooperate. Acetylation and methylation of p53 by p300 and SETD7 respectively promotes transcription of p21 and causes cell cycle arrest. Activated interferon  $\alpha$  (IFN $\alpha$ ) translocate CBP to cytoplasm (Bhattacharya, Eckner et al. 1996).

IFN $\alpha$ R2 (IFN $\alpha$  receptor 2) acetylated by CBP, which facilitates the binding of interferon regulatory factor 9 (IRF9). Phosphorylation of IFN $\alpha$ R2 enhances the interaction of IFN $\alpha$ R2 with IRF9. CBP acetylates signal transducer and activator of transcription 2 (STAT2) and promotes hetero dimerization of STAT2 with STAT1 and further forms a complex together with IFN stimulated gene factor 3 (ISGF3) which translocate to the nucleus to promote transcription of several interferon responsive antiviral proteins (Choudhary, Weinert et al. 2014).

#### 1.2.1.4. Histone acetyl transferases

##### 1.2.1.4.1. HAT enzymes

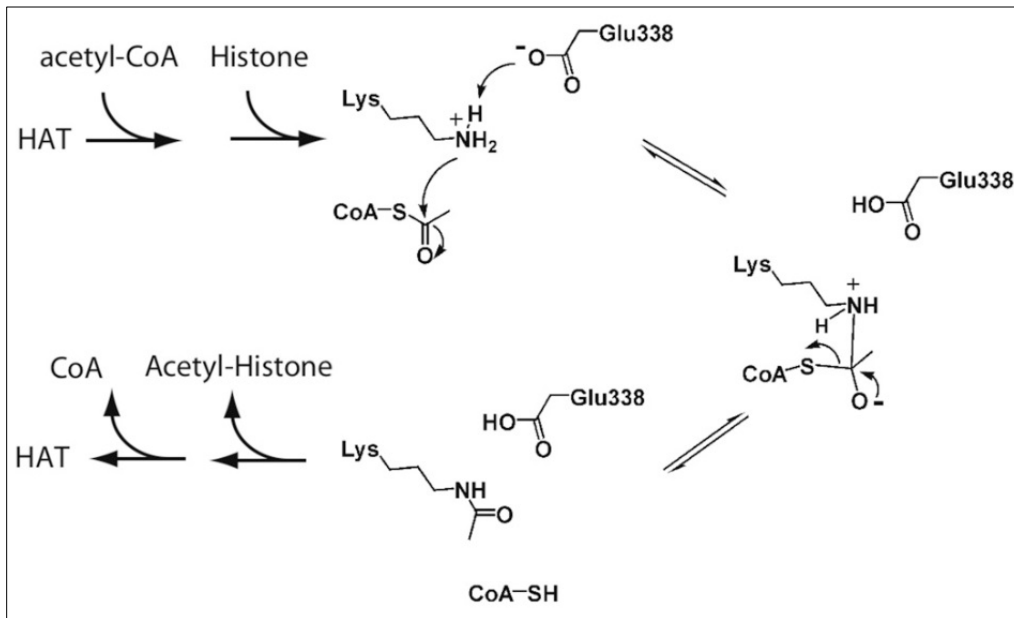
HATs are very important enzymes that regulate chromatin structure and gene transcription. HATs are generally multiprotein complexes (see book chapter in section 1.1.4.), and their overall function depends on other subunits since these partner subunits are providing target specificity (Choudhary, Kumar et al. 2009). HATs are highly divergent proteins and share no similarities between different groups but possess similarity in the same group. Based upon cellular localization HATs were classified as nuclear type A HATs and cytosolic type B HATs. Based upon structural and functional similarities HATs can be grouped into two distinct classes, the GCN5 family and the MYST family. The GCN5 family includes Gcn5, Ada, SAGA and PCAF complexes which majorly acetylate histone H3 at different sites. MYST family members include Sas2, Sas3, Esa1, Nua4, MOF, MOZ, Tip60, MORF and HBO1. The MYST family HATs possess several cellular functions like gene silencing, cell cycle progression, dosage compensation, etc. Few proteins with HAT activity have been identified which lack similarity with other groups. These include CBP/p300, Hat1, TAFII250, ACTR/SRC-1, Elp3, Hpa2, Nut1 and ATF-2 (Berndsen and Denu 2008, Choudhary, Weinert et al. 2014).

**Table 5: Table of HAT complexes**

<b>HAT</b>	<b>Known complexes</b>	<b>Histone targets</b>
<b>GNAT family</b>		
Gcn5	SAGA Gcn5/Ada2/Ada3 ATAC TFTC	H3K9, K14, K36 H3K14
p/CAF	STAGA	
<b>MYST family</b>		
Esa1 (Tip60 in humans)	NuA4, Piccolo NuA4	H4K5, K8, K12, K16, Htz1K14
Sas2 (MOF in humans)	SAS-1	H4K16
Sas3	NuA3	H3K14, K23
Moz	MOZ	H3K14
<b>p300 family</b>		
CBP	Numerous	H2AK5, H2B
p300	Numerous	H2AK5, H2B
Rtt109	Rtt109, Vps75, Rtt109-Asf1	H3K56, K9, K23

A conserved mechanism has been proposed for the acetylation reaction (DesJarlais and Tummino 2016). In the direct attack mechanism of acetyltransferases, acetyl-CoA and substrate binds the HAT in a sequential manner. A glutamate residue in the active site of the HAT deprotonates the histone lysine which result in the formation of a tetrahedral intermediate by the help of nucleophilic attack on the carbonyl carbon of acetyl-coA. Further, the intermediate results

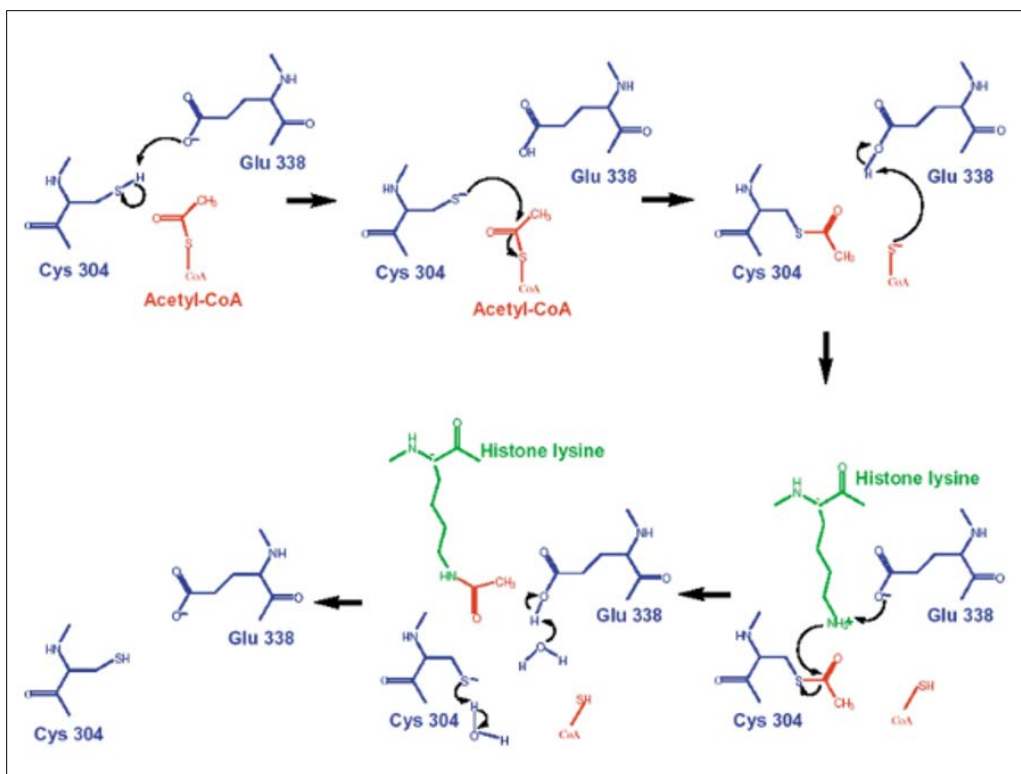
in the formation of acetylated histone and CoA (Figure 8) (Yan, Harper *et al.* 2002, Berndsen and Denu 2008)



**Figure 8: One-step HAT mechanism of action.**

Amino acids numbered according to Esa1. Picture adapted from (Berndsen and Denu 2008).

An alternate mechanism was proposed for the truncated version of Esa1 acetyltransferase domain which follows a two-step mechanism. In the first step, Esa1 binds acetyl-coA, and an active site cysteine residue attacks the acetyl moiety to form an acetylated enzyme intermediate. The subsequent release of CoA allows substrate lysine binding in the active site which is deprotonated by the glutamate and is further attacked by the acetyl-cys to finally yield an acetylated peptide (Figure 9) (Yan, Harper *et al.* 2002). These two mechanisms explain the behavior of HAT domain in truncated and in complex formation.



**Figure 9: HAT mechanism of action:**

Two-step mechanism of Esa1 (blue), acetyl-CoA (Hu, Chen *et al.*) and substrate (Jaiswal, Turniansky *et al.*). Picture adapted from (Yan, Harper *et al.* 2002).

#### 1.2.1.4.2. HAT inhibitors and activators

Curcumin is a natural substrate that has an inhibitory effect on CBP and p300 and inhibits acetylation of H3 and H4 which results in the inhibition of cell proliferation and induction of apoptosis in cancer cells. A similar inhibitory action was observed for garcinol and anacardiac acids against p300 and PCAF with a low affinity. Isothiazolones can also inhibit enzyme activity of acetyl transferases. Lys-CoA is a bisubstrate inhibitor of p300. C646, is a pyrazolone containing inhibitor. It is the only active inhibitor and is selective towards p300 over other HATs. Lys CoA is 20-fold more potent than C646 but it is not cell permeable. Inhibition of p300 by C646 induces caspase dependent apoptosis in prostate cancers (Bowers, Yan *et al.* 2010). A detailed list of inhibitors for HATs is given in the table (Wapenaar and Dekker 2016).

**Table 6: Inhibitors and activators of HATs**

<b>HAT inhibitors</b>	<b>HATs</b>	<b>Proposed target pathology</b>
Bi-substrate inhibitors	Various	-
Garcinol	KAT3B	Breast cancer, colon cancer
Curcumin	KAT3B	Cancer, inflammation, metabolic diseases
Anacardic acid	Non selective	Sensitizing cancer cells
TH1834	KAT5	Breast cancer
Benzylidene barbituric acid	KAT3B	Cell cycle arrest
Isothiazolones	Various	Inhibition of cancer cell proliferation
Thiazinesulfonamide	KAT3B	-
C646	KAT3B	Prostate cancer, melanoma, leukemia
ICG-001	KAT3A	Colon carcinoma
<b>HAT activators</b>		
CTPB	KAT3B	-
TTK21	KAT3A and 3B	Neurogenesis
Pentadecylidenemalonate	KAT2B	Conditioned fear, wound repair

### 1.3. Histone deacetylases: sirtuins and HDACs

Histone deacetylases are acetyl L-lysine deacetylases, often termed lysine deacetylases, which oppose the action of HATs. By removing the acetyl moiety from the  $\epsilon$ -amino group of lysine residues, HDACs promote chromatin compaction and regulates gene expression. HDACs deacetylate a wide range of histone and non-histone substrates, and thus play a pivotal role in numerous cellular functions such as chromatin organization, transcription regulation, cell proliferation, differentiation, immune suppression and angiogenesis (West and Johnstone 2014). However, the biological consequences of non-histone substrate deacetylation for a majority of HDAC members has to be investigated. HDACs are the targets of a growing interest in medicinal chemistry because of their role in several cellular functions. Given their biological significance and their role in drug discovery, an emphasis was put on HDACs in this thesis.

#### 1.3.1. Histone deacetylases classification

Histone deacetylases have been classified into four major classes according to their sequence homology to yeast proteins, subcellular localization and enzymatic activity. Class I, II and IV are  $Zn^{+2}$  dependent enzymes, possess a highly conserved arginase-deacetylase fold (hereafter referred as HDACs), whereas class III enzymes depend on  $NAD^+$  for their deacetylase activity and are known as sirtuins and contain a Rossmann-fold (Gregoretto, Lee *et al.* 2004, Dowling, Di Costanzo *et al.* 2008). So far, 11 human HDACs (HDAC1 to HDAC11) and 7 sirtuins (SIRT1 to SIRT7) have been identified in mammals. Within the HDAC and sirtuin family, the different isozymes differ in size, tissue specific expression patterns, sub cellular localization, enzymatic activity, and substrate specificity. *Table 6* represents a detailed classification of HDACs and Sirtuins, their homology to yeast proteins, their cellular localization and are provided with a few examples of specific substrates.

**Table 7: Classification of HDACs:**

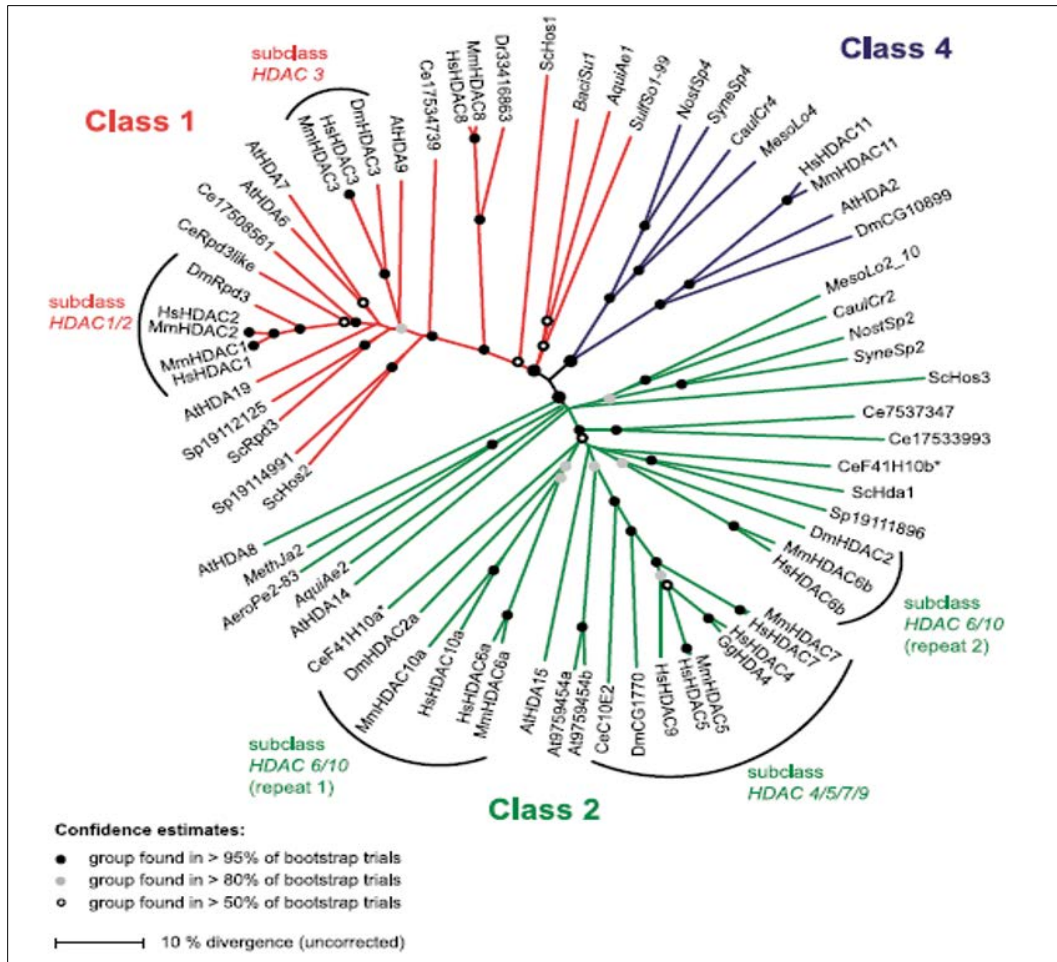
Different classes of HDACs with their member enzymes and yeast homology proteins were mentioned. And also localization, non-histone substrates were mentioned. Note that the substrate list is not complete.

Class	HDACs	Yeast homology	Subcellular localization	Substrates
Class I	HDAC 1	Rpd3	Nucleus	Androgen receptor, SHP, p53, STAT3, MyoD, E2F1
	HDAC 2		Nucleus	Glucocorticoid, YY1, BCL6, STAT3
	HDAC 3		Nucleus	SHO, YY1, GATA1, RELA, STAT3, MEF2D
	HDAC 8		Nucleus, cytoplasm	SMC3, p53
Class IIa	HDAC 4	Hda1	Nucleus, cytoplasm	GCMA, GATA1, HP
	HDAC 5		Nucleus, cytoplasm	GCMA, SMAD7, HP1
	HDAC 7		Nucleus, cytoplasm, mitochondria	PLAG1, PLAG2
	HDAC 9		Nucleus, cytoplasm	-
Class IIb	HDAC 6	-	Cytoplasm	$\alpha$ -tubulin, HSP90, SHP, SMAD7
	HDAC 10		Cytoplasm	Acetyl polyamines
Class IV	HDAC 11	-	Nucleus, cytoplasm	-
Class III (Sirtuins)	SIRT 1-7	Sir2	Cytoplasm, nucleus and mitochondria	various

HDACs and sirtuins are evolutionarily well-established enzymes as they are found in almost all forms of life: archaeobacteria, eubacteria, fungi, plants and animals. An evolutionary relationship of histone deacetylases is provided using a phylogenetic tree (*Figure 10*). All Zn<sup>+2</sup> dependent HDACs display a similar structure with arginases which are metallo-enzymes, which



suggests that HDACs have evolved from a common ancestral metallo-enzyme (Yoshida, Kudo *et al.* 2017). The presence of histone deacetylases in eubacteria also supports the concept of non-histone protein substrates for these enzymes.



**Figure 10: Phylogenetic tree of eukaryotic HDACs:**

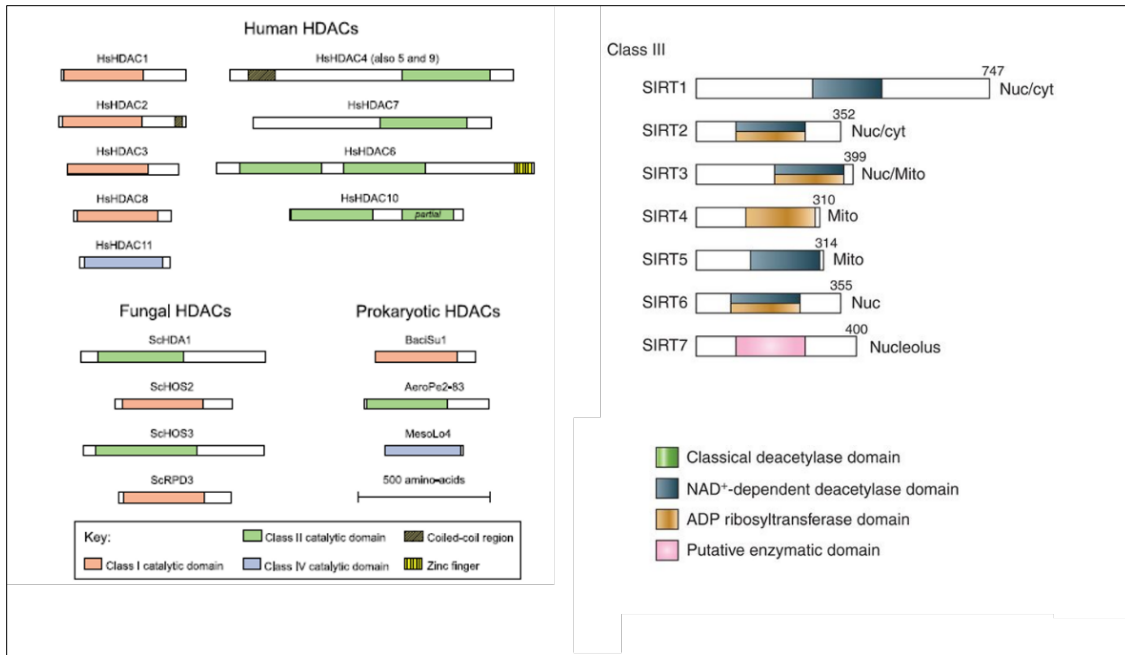
Phylogenetic tree was made using bootstrap neighbor-joining. Nodes or connection points indicate gene divergence, and no assumption was made on which class is close to ancestral genes. Prokaryotes are mentioned in italics. Large black dots indicate nodes where sub classes are divided. Picture adapted from (Gregoretti, Lee *et al.* 2004).

The numbering of histone deacetylases was given in chronological order, based upon their discovery. Class I HDACs are homologues to Rpd3 which is a yeast protein, similarly class II

HDACs have homology with Hda1 and class III Sirtuins with Sir2. Class IV protein contain only a single member HDAC11 that does not have any homology to Hda1 or Rpd3, and was hence placed in a separate group. Class I HDACs, except HDAC8, form dimers, which is an ancestral specificity observed in class I HDACs. HDAC1 can form homo and hetero dimers by self-association or by interaction with HDAC2 while HDAC3 can form a homo dimer by self-association. HDAC8 is a fast-evolving enzyme in class I HDACs which has been lost from invertebrates.

Class II HDACs are divided into two sub classes where class IIa contains HDAC4, HDAC5, HDAC7 and HDAC9 while class IIb contains HDAC6 and HDAC10. Class IIa HDACs interacts with other proteins which includes nuclear receptors, muscle transcription factor and also with HDAC3. The role of class IIa HDACs is much developed in muscle cells. Class IIb HDACs are most interesting in terms of structure as they contain two domains. HDAC6 contains two catalytic domains in tandem, whereas HDAC 10 contains an additional vestigial HDAC domain (Figure 11). HDAC10 has evolved from HDAC6-like protein after vertebrate/invertebrate divergence. Class IV contains a single member that is HDAC11 which is present in all living organisms except fungi, suggests that HDACs have evolved from ancestral organisms and directed towards histones.

Class III deacetylases sirtuins are evolutionarily well conserved and are present from archaea to mammals. Sirtuins have been involved in metabolic and chromatin organization throughout the evolution. The seven members of sirtuins (SIRT1-7) are grouped into four classes and are homologues to yeast protein Sir2 (Silent information regulator 2).



**Figure 11: Domain organization of HDACs and sirtuins:**

Picture is modified from (Gregoretta, Lee *et al.* 2004, Seto and Yoshida 2014).

### 1.3.2. Sub cellular localization of histone deacetylases

Histone deacetylases localization depends upon their functional requirement and structure. After synthesis histone deacetylases are imported to the nucleus by a nuclear localization signal (NLS) or with transporter proteins. Class I HDACs are generally found in the nucleus, and they can deacetylate both histone and non-histone proteins. HDAC1 and HDAC2 are exclusively found in the nucleus while HDAC 3 is also present in the cytoplasm. HDAC8 is distributed in the cytoplasm and the nucleus. Few non-histone substrates and HDACs localization were mentioned (Table 7).

Class II HDACs shuttle in between the nucleus and the cytoplasm to perform their biological functions which includes the deacetylation of structural proteins such as GCMA, GATA, PLAG etc. In order to transport HDACs from the nucleus to the cytoplasm, HDACs contain a nuclear export factor in their sequence. Nuclear export signal (NES) is a short peptide that helps in the export of proteins from the nucleus to the cytoplasm using the nuclear pore

complex. Phosphorylation of HDAC4 by calcium calmodulin protein kinase exports HDAC4 to the cytosol with the help of CRM1 which is a cellular export factor that contains a NES. In the cytosol 14-3-3 proteins bind HDAC4 and retain it in the cytosol. After muscle cell fusion due to reduced phosphorylation, 14-3-3 proteins are dissociated from HDAC4 which shuttles back to the nucleus. HDAC7 also shuttles between the cytoplasm and the nucleus like HDAC5. HDAC7 is similar in sequence to HDAC5 but lacks a NES. During muscle differentiation HDAC5 is transported to the cytoplasm.

Sirtuins majorly participate in senescence and metabolic regulations. SIRT1, 6 and 7 are mostly found in the nucleus. SIRT2 is predominantly found in the cytoplasm and can also be exported to the nucleus, SIRT3, 4 and 5 are mitochondrial sirtuins (Yoshida, Kudo *et al.* 2017). SIRT4-7 were found to be weak deacetylases towards histones. Among sirtuins, SIRT2 and SIRT3 possess both NAD<sup>+</sup> dependent deacetylase activity and ADP-ribosyl transferase activity. SIRT1 and SIRT5 possess deacetylase activity, and SIRT4, SIRT6 have mono ADP-ribosyl transferase activity. SIRT3 is localized in the mitochondria and plays an important role in scavenging reactive oxygen species (ROS) (Kim, Patel *et al.* 2010, Li and Zhu 2014).

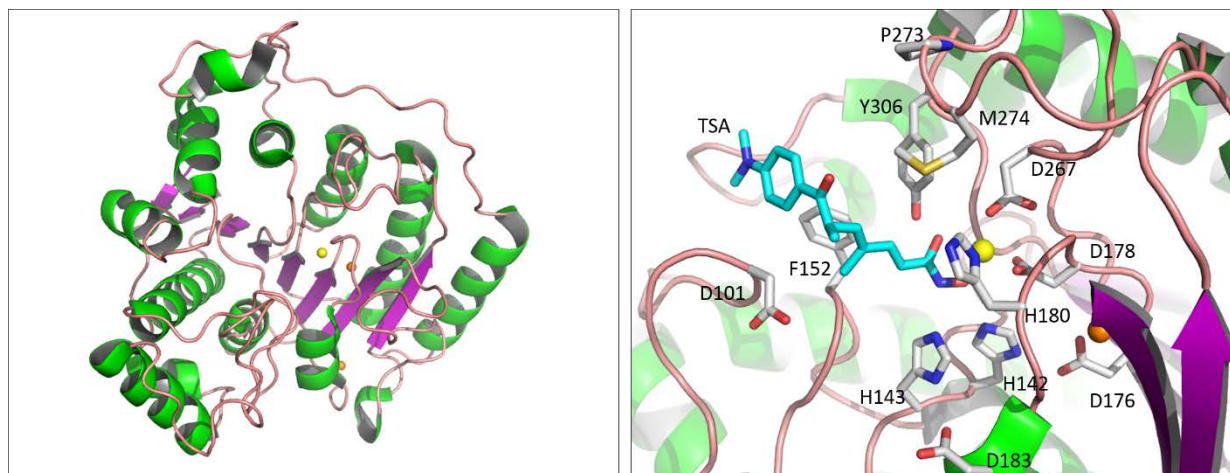
### 1.3.3. Structure of histone deacetylases

#### 1.3.3.1. HDACs

The HDAC-like protein (HDLP) in presence of the TSA inhibitor was the first HDAC-like crystal structure determined from the hyper thermophilic bacterium *Aquifex aeolicus* (Finnin, Donigian *et al.* 1999). HDLP possess 35% sequence identity with HDAC1, which shed light on the structure of class I HDACs. Ever since many crystal structures were solved for HDACs and to date only three HDAC structures are still missing HDAC5, HDAC9 and HDAC11.

All Zn<sup>+2</sup> dependent HDACs contain a common HDAC  $\alpha/\beta$  domain that is similar to arginase fold, where eight parallel  $\beta$  sheets are sandwiched between  $\alpha$  helices (Figure 12). The catalytic pocket contains a narrow hydrophobic pocket in a tube-like shape where Zn<sup>+2</sup> lies at the bottom of the pocket nearly 11Å in depth. The metal ion can be different *in vivo* from Zn<sup>+2</sup>,

depending upon cellular concentrations HDAC can be activated by  $\text{Fe}^{+2}$  (Dowling, Gattis *et al.* 2010). The catalytic pocket is made up of highly hydrophobic amino acids while the catalytic residues that co-ordinate the  $\text{Zn}^{+2}$  are polar in nature and present at the bottom of the active site pocket. The hydrophobic catalytic pocket accommodates aliphatic side chain of lysine during the catalysis. The narrow active site pocket becomes wider at the bottom where the  $\text{Zn}^{+2}$  ion is coordinated by two aspartate residues, a histidine residue and a conserved water molecule.



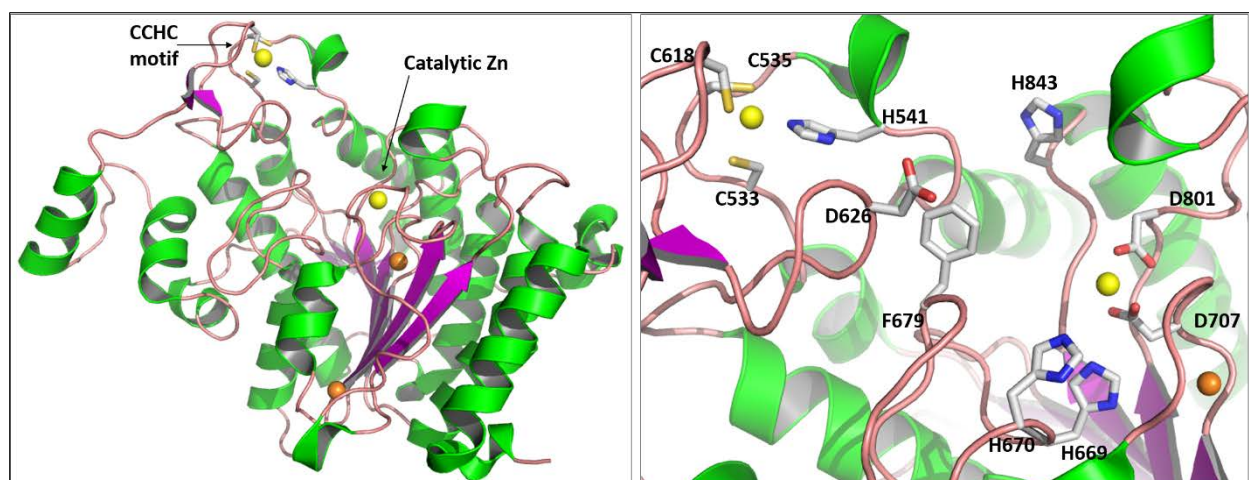
**Figure 12: HDAC fold:**

In the left panel HDAC8 structure (pdb: 1T64) is represented as ribbons where  $\beta$  sheets (magenta) are sandwiched between  $\alpha$  helices (Jaiswal, Turniansky *et al.*), catalytic zinc and potassium (or sodium in present structure) are represented as spheres (yellow and orange colours respectively), all other loops are represented in salmon colour. In the right panel, a close up view of catalytic pocket was represented with all the catalytic amino acids in sticks. In the active site Trichostatin A (Guan, Haggarty *et al.*) was represented as sticks.

In class I HDACs, in addition to the catalytic  $\text{Zn}^{+2}$ , the catalytic pocket contains a tyrosine, Y306 (numbered according to HDAC8), a pair of charge relay system H142-D176 and H143-D183 (numbered according to HDAC8) that participates in the catalytic mechanism which is similar to serine proteases (Fersht and Sperling 1973). H142 is present in the hydrophobic environment while H143 is partially exposed to solvent. Mutations of H142 abolish the deacetylase activity, while mutations of H143 either abolish the activity in HDAC8 or decreases the activity by 12 folds in HDAC 1 and HDAC7 (Hassig, Tong *et al.* 1998, Dowling, Gantt *et al.* 2008). Y306 is conserved in class I and class IIb HDACs. The hydroxyl group of Y306 is oriented towards the catalytic metal ion and stabilizes the tetrahedral oxyanion intermediate through H-bond during the catalysis (Finnin, Donigian *et al.* 1999).

In class I and II, there are two additional metal binding sites apart from the catalytic  $Zn^{+2}$ , which can be occupied by  $Na^+$ ,  $K^+$  or  $Ca^{+2}$  (Vannini, Volpari *et al.* 2004). These two sites are referred as site 1 and 2, which are hexa-coordinated with amino acids and situated 7Å and 21Å away from catalytic pocket, respectively. Amino acids that coordinate site 1 are more conserved than those of site 2, and also site 1 is coordinated by one of the amino acid, D176, which is a part of H142-D176 charge-relay system of the active site. These two metal binding sites influence the catalytic function, depending upon the concentration of salt concentration in the environment. The catalytic function is activated upon low salt concentrations, while in the presence of high salt concentration the activity is inhibited. The mechanism on how these salt concentrations affect catalysis is not well studied, a few reports suggest however that  $K^+$  at site 1 decreases the pKa of H142 which further inhibits the enzyme activity (Gantt, Joseph *et al.* 2010).

The presence of a tyrosine in the catalytic pocket has an important impact on the deacetylation activity. In class IIa HDACs, the catalytic tyrosine is replaced by a histidine residue which is moreover turned away from the catalytic zinc. As a result, the deacetylase activity is reduced when compared to class I HDACs (Bottomley, Lo Surdo *et al.* 2008, Schuetz, Min *et al.* 2008). Mutations of this tyrosine in class I HDACs (Y306F and Y306H in HDAC8) have abolished the HDAC activity, while in class IIa HDACs histidine to tyrosine mutation has restored the canonical activity, which indicates the importance of the catalytic tyrosine in maintaining the proper deacetylase activity (Lahm, Paolini *et al.* 2007, Vannini, Volpari *et al.* 2007, Schuetz, Min *et al.* 2008). In addition to the replacement of the catalytic tyrosine, class IIa HDACs display a specific insertion of CCHC motif which participates in the coordination of zinc (Figure 13). This additional zinc is tetrahedrally coordinated with amino acid residues C533, C535, H541 and C618 (of HDAC7), which was further discussed in section (1.3.7.3).



**Figure 13: Structure of class IIa HDACs**

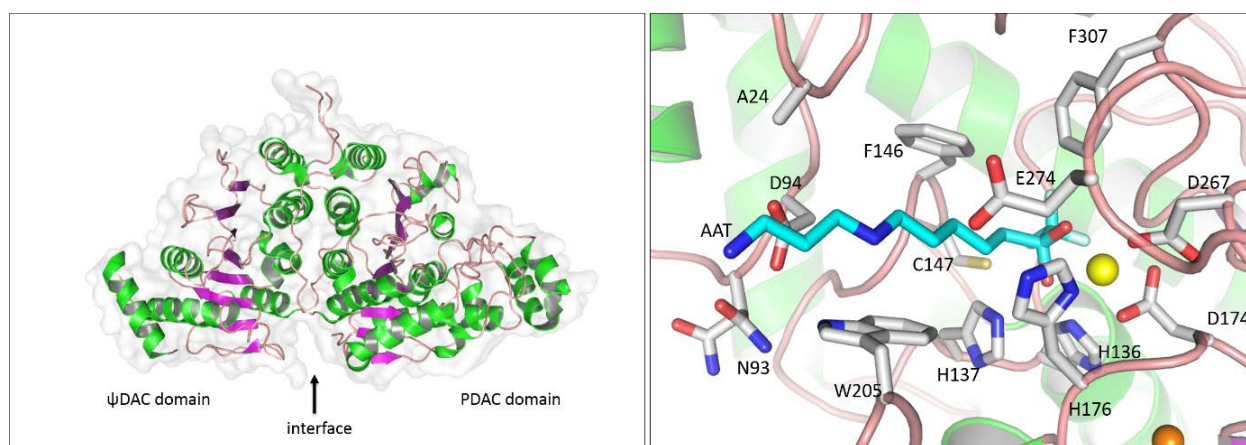
Left panel: Ribbon representation of HDAC7 crystal structure (pdb: 3COY). Right panel: close-up view of CCHC motif and catalytic pocket.

The catalytic pocket of class IIb HDACs is more confined than the one of class I HDACs and is designed to bind cytosolic substrates. Class IIb subclass is composed of two enzymes HDAC6 and HDAC10, where HDAC6 deacetylates K40 of the canonical substrate  $\alpha$ -tubulin and HDAC10 deacetylates polyamines. Among class II HDACs, HDAC6 contains a unique domain organization with two catalytic domains designated as CD1 and CD2, which are followed by a C-terminal ubiquitin binding domain. In class IIb HDACs a charged amino acid (E274 in HDAC10, K330 in HDAC6) is present in the active site pocket whereas in all other HDACs a methionine (M274 in HDAC8) or leucine is present. These charged amino acids in class IIa HDACs serve as gate-keepers and directs the incoming acetyl lysine substrate towards catalytic center. A description about HDAC6 structure was mentioned in the section (1.1.4).

HDAC10 deacetylates N8-acetylspermidine to yield spermidine and acetate. Spermidine is a polyamine and plays an important role in autophagy. Polyamines are preferred targets for HDAC10 than acetyl lysine (hence HDAC10 is also referred as PDAC) which is further supported by the presence of HDAC10 in liver, kidney and spleen. The architecture of HDAC10 is made to prefer polyamines over acetyl lysine (Figure 14). HDAC10 contains a catalytic domain (PDAC) connected to a pseudo deacetylase domain ( $\psi$ DAC) through alpha helices. The  $\psi$ DAC domain of HDAC10 does not have a catalytic function, it may help in the cytoplasmic localization. The



pseudo deacetylase domain lacks active site loops and forms a butterfly like hetero dimer with the catalytic domain. The catalytic domain of HDAC10 is similar to CD1 and CD2 of HDAC6. In HDAC10 the active site pocket is more constricted than other HDAC isozymes where E274 mediates an electrostatic interaction with the substrate and increases specificity for polyamine substrates. Unlike class I HDACs, the amino acids at the active site surface N93 and D94 do not affect substrate binding whereas in class I HDACs the corresponding amino acid D101 plays an important role in substrate binding. Although the overall fold of  $\psi$ DAC domain is similar to PDAC and HDAC6 catalytic domains it does not possess a catalytic  $Zn^{+2}$  or catalytic amino acids (Hai, Shinsky *et al.* 2017).



**Figure 14: Structure of HDAC10**

Left panel: Crystal structure of HDAC10 with  $\psi$ DAC and PDAC domain and their interface. Right panel: The constricted architecture of HDAC10 catalytic site is represented as ribbons (pdb: 5TD7). Active site residues are represented as sticks and labelled accordingly. In the active site HDAC10 specific inhibitor AAT (7-[(3-aminopropyl) amino]-1,1,1-trifluoroheptan-2-one) was shown in cyan colour sticks. Catalytic zinc (yellow) and potassium (orange) are represented as spheres.

### 1.3.3.2. Sirtuins

Sirtuins belongs to deoxyhypusine synthase (DHS)-like NAD/FAD family proteins. The domain organization of sirtuins is represented in (Figure 11). In general, sirtuins contain a N-terminal domain followed by a catalytic domain and a C-terminal region. N- and C-terminal regions varies among sirtuins and are subject to conformational changes during catalysis while the catalytic domain is of nearly 250 amino acids in length. SIRT1 C-terminal region interacts with the catalytic domain and induce conformational changes upon substrate binding.



The catalytic domain of sirtuins comprises two subdomains, a small domain and a large domain. The interface between the two domains forms a tunnel which is conserved among sirtuins where the NAD<sup>+</sup> cofactor and the substrate are accommodated. The large domain is made up of a Rossmann fold with 6  $\beta$ -strands forms a parallel sheet which is packed against 6  $\alpha$ -helices. The small domain can be subdivided into two modules, a zinc binding module and a helical module. The zinc binding module possess a divalent zinc which is tetrahedrally coordinated by four cysteine residues. The helical module is composed of four  $\alpha$ -helices.

In the absence of NAD<sup>+</sup> and substrate, sirtuins adopt an open conformation which is changed to a closed conformation upon substrate binding. The deacetylation function is facilitated by the simultaneous cleavage of nicotinamide and the subsequent transfer of ADP-ribose. The NAD<sup>+</sup> binding pocket is further subdivided into three subpockets. In pocket A, the ADP ribose moiety is bound, whilst NAD<sup>+</sup> is accommodated in the sub pocket B, and during catalysis NAD<sup>+</sup> occupies the sub pocket C, which further facilitates the transfer of the acetyl group from the lysine to the ribose of NAD<sup>+</sup>.

Even though the catalytic domain is fully conserved in sirtuins, surrounding loops are subjected to important conformational changes upon substrate binding. In SIRT2 the substrate and NAD<sup>+</sup> binding induces a 25° rotation of the small domain with respect to the large domain (Figure 16b). As a result L1-L3 loops are rearranged. L1 reorientation results in the unwinding of an  $\alpha$ -helix which facilitates the interaction with ADP-ribose, which further participate to enclose the subpocket C. L2 loop moves closer to L1 which creates a channel in order to accommodate acetylated lysines in the pocket. And the L3 loop which is the part of large domain also undergoes a conformational change due to the reorientation of two conserved residues (V266 and S263), which facilitates the interaction with ADP-ribose. These conformational changes are conserved in SIRT1, SIRT3 and SIRT5. In SIRT6 the helical module is missing, as a result the loop L1 acquires rigidity and the zinc binding module lacks the interaction with the helical module. As a consequence, the zinc binding domain makes a direct interaction with the large domain by a rotation of 45° over the large domain.

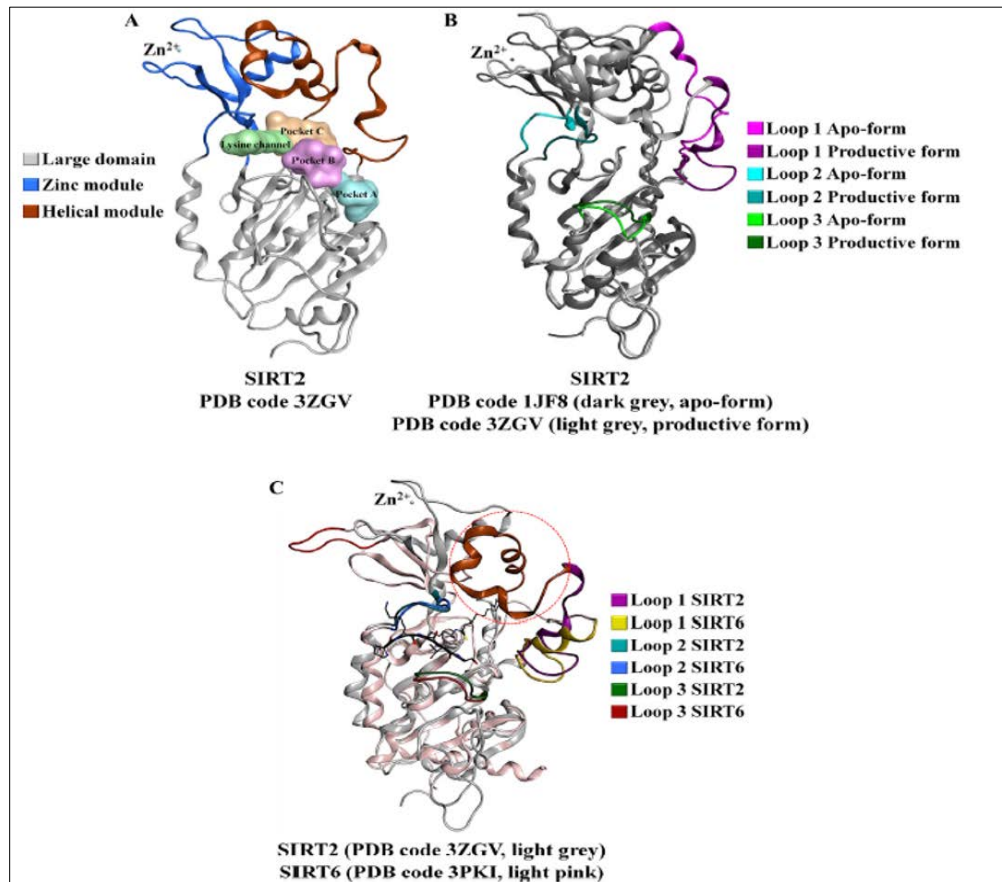
Further, an allosteric modulation of SIRT1 was explained in presence of resveratrol which binds at the substrate entry channel. The domain organization of SIRT1 is unique compared to other sirtuins (Figure 11). It contains an N-terminal region of 50 amino acids followed by a catalytic

domain of 260 amino acids and a C-terminal region ESA of 25 amino acids length. Three molecules of resveratrol were observed in the crystal structure in which two molecules facilitates the tighter binding of the peptide substrate to the SIRT1 which enhances the catalytic activity (Cao, Wang *et al.* 2015). This mechanism explains how modulation of allosteric sites activates the SIRT1 function.



**Figure 15: Structure of SIRT1 with resveratrol**

Ribbon representation of crystal structure (pdb: 5BTR) of SIRT1 in complex with resveratrol (yellow sticks) and p53-acetylated peptide (coloured white sticks). The N-terminal domain of SIRT1 is represented in green colour, catalytic domain in magenta colour and MES in cyan.



**Figure 16: Sirtuin structures:**

Ribbon representation of Sirtuin structures, a) structural organization of sirtuin structure with large domain and zinc binding module and helical module. b) three loops that undergo huge conformational changes during catalysis are represented according to the legend- c) a comparison of SIRT2 with SIRT6. The loop inserted in SIRT6 zinc binding module is coloured in red and the myristoylated peptides is represented in black sticks. Picture adapted from (Sacconay, Carrupt *et al.* 2016).

#### 1.3.4. Complexes of histone deacetylases

A characteristic feature of HDACs is to form multiprotein complexes with corepressors which are directed to specific gene loci to play their repressive function. Another major implication of forming HDAC complexes is to enhance the deacetylation activity of certain HDACs, which is evident in case of class IIa HDACs where the loss of the catalytic tyrosine hinders the enzymatic activity. Complex formation with other HDACs (HDAC3 for example) compensates this loss of

activity. However, HDACs interactions with different complexes and functional characterization is still a growing concept.

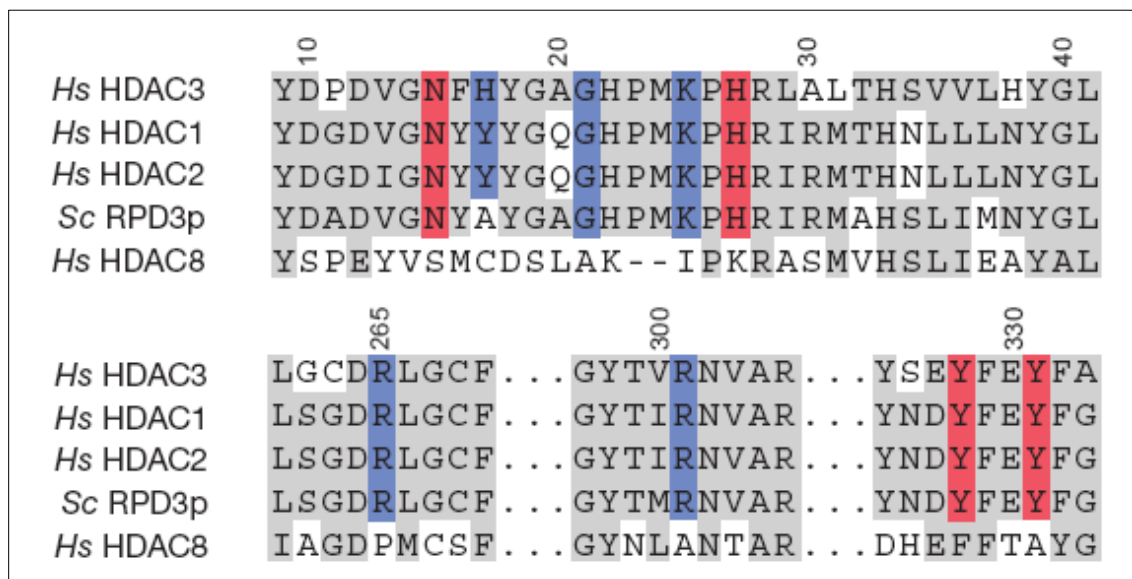
#### 1.3.4.1. Complexes of class I HDACs

Class I HDACs HDAC1, HDAC2 and HDAC3 are known to be a part of multiprotein corepressor complexes (Watson, Fairall *et al.* 2012, Lauffer, Mintzer *et al.* 2013, Millard, Watson *et al.* 2013, Watson, Millard *et al.* 2016, Yoshida, Kudo *et al.* 2017). HDAC1 and HDAC2 are found in several corepressor complexes such as Sin3, Nuclear receptor Co-Repressor (N-CoR), Nucleosome Remodeling Deacetylase (NuRD), Corepressor of RE1-Silencing Transcription factor (CoREST) and Mitotic DeAcetylase Complex (MiDAC) (Grozinger and Schreiber 2002, Itoh, Fairall *et al.* 2015). HDAC3 is exclusively found in association with corepressor complex N-CoR and Silencing Mediator of Retinoic acid and Thyroid hormone receptors (SMRT) (Wen, Perissi *et al.* 2000).

Mammalian Sin3 contains two major isoforms, Sin3A and Sin3B. Sin3A possess several subunits in which ING2 recognizes methylated H3K4 and directs the deacetylation of nucleosome with HDAC1-HDAC2 dimer (Shi, Hong *et al.* 2006). SMRT, N-CoR, NuRD and CoREST complexes possess SANT domains which are important for the interaction of HDACs with the corepressor complexes. Further, SANT domains recognize the PTM on histones and deliver the corepressor complex on the target site. In this way, the SANT domain couples histone binding with deacetylase activity (Boyer, Latek *et al.* 2004). The mechanism of interaction of HDACs with corepressor complexes was first reported by a crystal structure of HDAC3 in complex with SMRT, which has revealed the presence of inositol 1,4,5,6-tetrakisphosphate (Ins(1, 4, 5, 6)P<sub>4</sub> or IP<sub>4</sub>) at the interface of HDAC and SMRT interaction surface (Watson, Fairall *et al.* 2012). The presence IP<sub>4</sub> was surprising because it was co-purified from the mammalian expression cells and it highlights the physiological relevance of IP<sub>4</sub> in the stabilization of HDAC-corepressor complex. IP<sub>4</sub> is produced from Ins (1,4,5)P<sub>3</sub> which is a well-known second messenger after ATP, that releases Ca<sup>2+</sup> upon binding to its receptor INSP<sub>3</sub> receptor.

The crystal structure of the complex formed between HDAC3 and the deacetylase activation domain (DAD) of SMRT has revealed that the IP<sub>4</sub> is sandwiched in between HDAC3 and SMRT, where a basic pocket is formed by the interface residues to facilitate the IP<sub>4</sub> binding

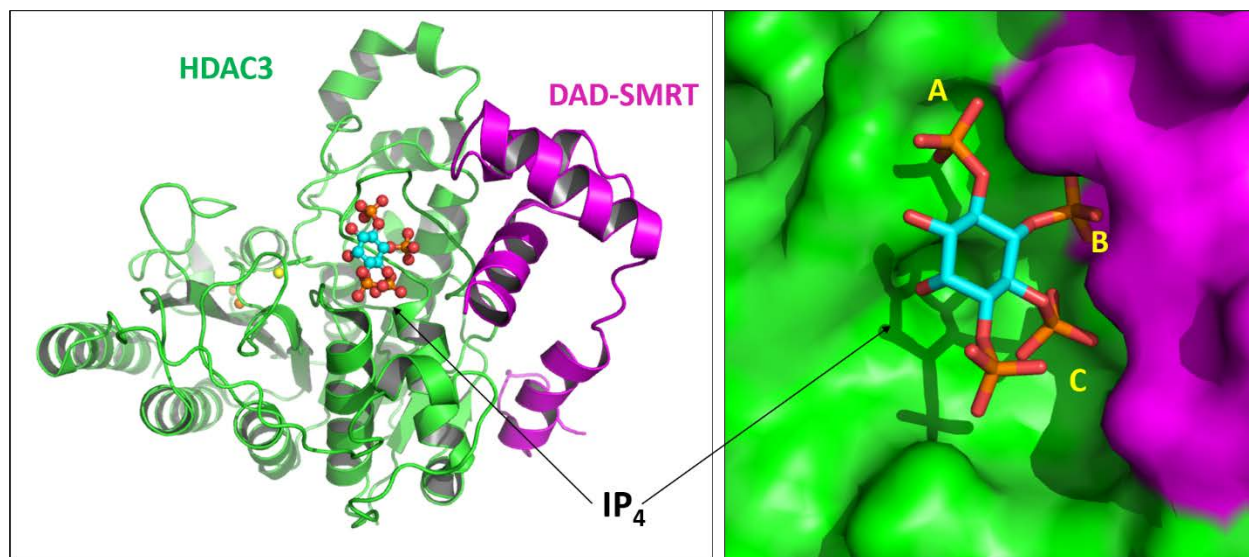
(Figure 18). Upon complex formation, the conformation of the N-terminal helix of the DAD domain rearranges over the surface of HDAC3 covering amino acids from Helices H1, H2, strand S2 and loop L2 and L6 of HDAC3. The interaction surface of HDAC3 and the co-repressor complex is highly conserved in class I HDACs (except HDAC8) as well as in the corepressor complex SANT domains, except Sin3A. These structural parts of HDAC3 are distinct from HDAC8, which makes possible for HDAC8 to act without IP<sub>4</sub> activation (Figure 17). In the absence of IP<sub>4</sub> the basic amino acids would repulse and results in the inactivation of HDAC3 in the complex environment. In this scenario IP<sub>4</sub> acts as molecular glue holding two proteins together.



**Figure 17: Sequence alignment of HDAC loops**

Sequence alignment of class I HDACs with highlighted amino acids important for IP<sub>4</sub> and corepressor binding respectively in blue and red in colour. Picture adapted from (Arrar, Turnham *et al.* 2013).

Further, IP<sub>4</sub> not only plays a structural role but also it participates in the activation process of HDAC3 by modulating HDAC dynamics. R265 of HDAC3 loop L6 forms an important interaction with IP<sub>4</sub> and the consecutive amino acid, L266, is involved in the construction of the active site pocket. Binding of IP<sub>4</sub> may stabilize the dynamic loop L6 which in turn enhances the deacetylase activity (Arrar, Turnham *et al.* 2013).

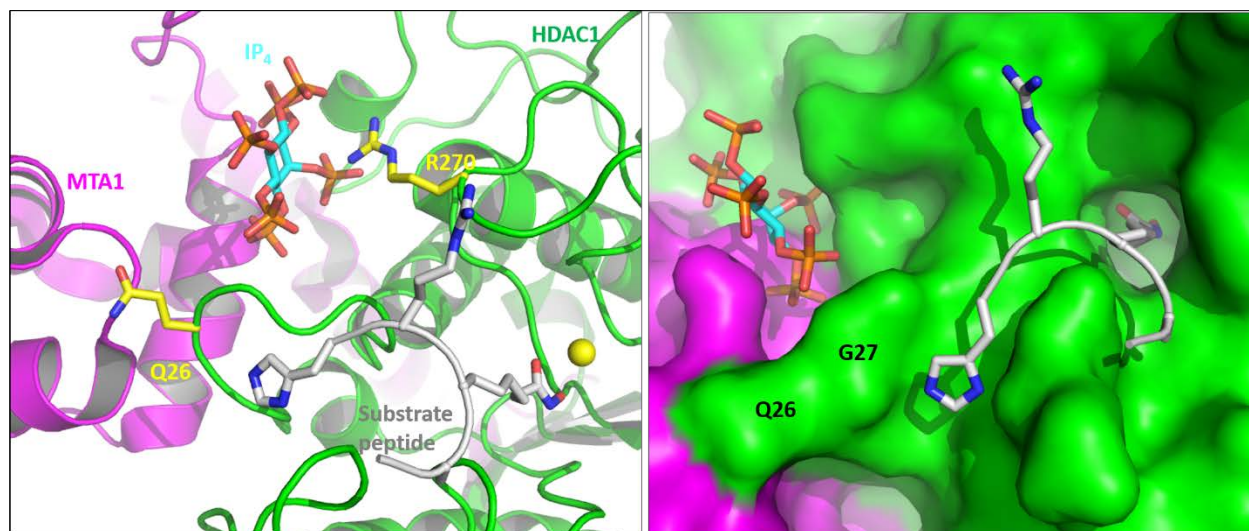


**Figure 18: Crystal structure of HDAC3-SMRT complex**

Left panel: Ribbon representation of HDAC3 (Jaiswal, Turniansky *et al.*) DAD-SMRT (magenta) complex (pdb: 4A69). IP<sub>4</sub> is represented as spheres in cyan and orange colour. Catalytic Zinc and potassium ions are shown as yellow, orange spheres respectively. Right panel: close up view of IP<sub>4</sub> binding interface with sub pocket numbers which are important for activation mechanism. HDAC3 and DAD-SMRT are shown in surface representation while IP<sub>4</sub> is shown as cyan and orange sticks.

The IP<sub>4</sub> mediated interaction mechanism was also reported by the crystal structure of HDAC1-MTA1 (Metastasis-associated protein 1, a subunit of co-repressor complex NuRD) complex in presence of a substrate peptide which is derived from H3 (Figure 19). The binding pocket of IP<sub>4</sub> binding contains three sub pockets A, B and C which plays an important role in the activation mechanism. IP<sub>3</sub> which is the precursor of IP<sub>4</sub>, in spite of its physiological abundance, cannot activate HDACs because it cannot occupy all three sub pockets. Further and most importantly, two amino acids Q26 and G27 in HDAC1 mediates hydrogen bonds with IP<sub>4</sub> and substrate peptide on the either sides of the loop (Watson, Millard *et al.* 2016). These interactions stabilize the active site pocket of class I HDACs (except HDAC8), which results in the enhanced activity.





**Figure 19: Surface representation of HDAC1-MTA1 complex crystal structure:**

Left panel: HDAC1-MTA1 complex ribbon representation (pdb: 5ICN). Important amino acids are numbered. Right panel: surface representation of interaction surface between IP<sub>4</sub> with HDAC1 and MTA1. Important amino acids Q26 and G27 which participates in the interaction with substrate peptide and IP<sub>4</sub> are represented.

#### 1.3.4.2. Complexes of Class II HDAC:

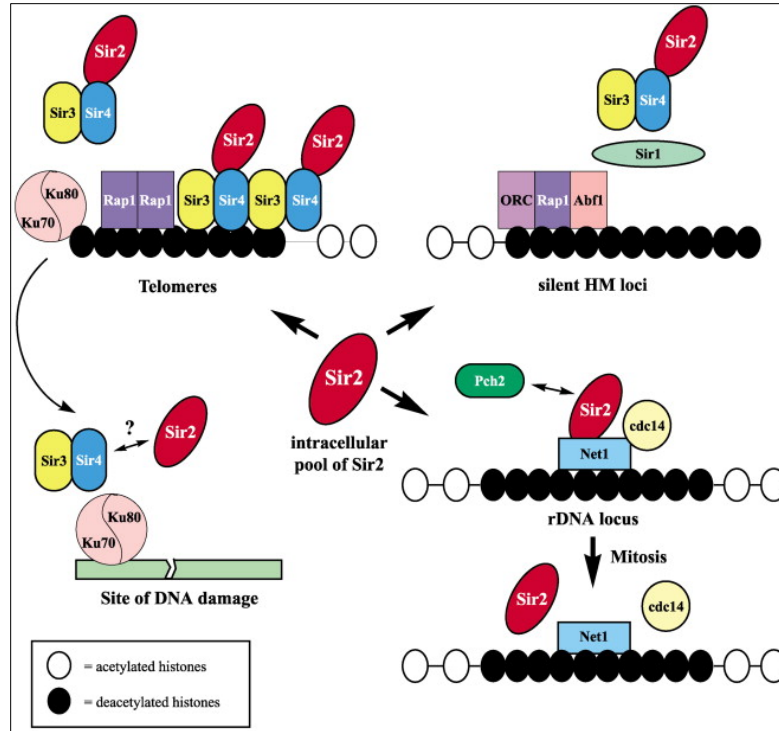
Due to lack of a catalytic tyrosine, class IIa HDACs possess very weak activity. However, the overall catalytic function of class IIa HDACs is dependent on co-repressor complexes. Class IIa HDACs are majorly associated with NCoR/SMRT complexes in presence of HDAC3. Nevertheless, the knowledge of class IIa HDACs mechanism of action in presence of co-repressor complexes is hindered due to the lack of structural information on how these HDACs interact in multiprotein complexes. The complex forming ability of class II HDACs is also evident in lower organisms. Hda1 in yeast is the homologue of class II HDACs, self-associates to form a functional tetrameric complex in presence Hda2 and Hda3 (Wu, Carmen *et al.* 2001). Class IIa HDACs shuttle between the nucleus and the cytoplasm which requires an active interaction with transporter protein such as proteins 14-3-3, kinases and phosphatases. Phosphorylation of HDAC4, HDAC5 and HDAC7 is important for the interaction with the 14-3-3 proteins and enables the nuclear export of class II enzymes. In contrast HDAC6 is found in association with cytoskeletal proteins and chaperons (Zhang, Yuan *et al.* 2007).

#### 1.3.4.3. Sirtuin complexes:

Sirtuin complexes have been less studied. So far, there is no complete characterization that was done for a single sirtuin complex. To date no crystal structure of sirtuin is available in complex with an interacting partner. This hinders the knowledge in understanding sirtuin biology (Sacconnay, Carrupt *et al.* 2016). Nevertheless, yeast sirtuin Sir2 has been identified in two protein complexes (Figure 20), where one complex is formed by Sir2, 3 and 4 proteins and which is responsible for the stabilization of telomeres and is also involved in the silencing the telomeric and *HM* loci. During double strand DNA (dsDNA) break repair, the Ku heterodimer (required for non-homologous end joining DNA break repair) translocate from telomeres to the dsDNA breaks and recruits the Sir2/3/4 complex in order to silence the transcription.

The second complex of Sir2 is formed in complex with Net1 and cdc14 which has the suppressing role in genetic recombination. Sir2 participates in RENT (regulator of nucleolar silencing and telophase exit) complex, which mediates the silencing of ribosomal DNA (rDNA) (Huang and Moazed 2003). At rDNA locus Sir2 recruitment facilitates the condensation of chromatin and thereby down regulates the rRNA expression. As a consequence, rRNA loci is regulated from non-specific recombination. Further, the binding of Pch2, a meiotic check point protein to Sir2 also prevents the recombination at rDNA locus during meiosis. Sir2 also forms a complex with cdc14 which plays an important role in the cell cycle progression.





**Figure 20: Sirtuin complexes**

Yeast Sir2 protein complexes. Sir2 forms different complexes in association with other sirtuins and cellular protein. Picture is adapted from (Grozinger and Schreiber 2002).

### 1.3.5. Substrates of histone deacetylases

Histone deacetylases are among the proteins in the cell that show a wide diversity in their substrates, which includes histone substrates and non-histone substrates. It is difficult to differentiate histone substrates with histone deacetylase isozyme specificity because in *in vitro* assays, all histone deacetylases deacetylate histone substrates without specificity. Several non-histone substrates have been discovered which highlight the vast cellular functions that involve histone deacetylases. The first non-histone histone deacetylase substrate identified was p53 which is acetylated by PCAF and p300 proteins in response to DNA damage, while the same is deacetylated by HDAC1, HDAC2, HDAC3 and SIRT1 (Ito, Kawaguchi *et al.* 2002, Yoshida, Kudo *et al.* 2017). The substrates of histone deacetylases include both nuclear and cytosolic proteins.

The nuclear substrates of histone deacetylases include transcription factors p53, YY1, HMG, STAT3, c-MYC, MyoD, GATA factor, EKLF, E2F/Rb, NF- $\kappa$ B, Smad 7, HIF-1 $\alpha$  and nuclear receptors such as the androgen receptor and the estrogen receptor. Cytosolic substrates of histone deacetylases include  $\alpha$ -tubulin, Importin- $\alpha$ , mitochondrial ku70, Hsp90 before nuclear export, Smc3. Apart from cellular proteins few viral proteins are also subjected to acetylation by p300 and PCAF such as E1A and HDAg whose deacetylation enzymes are not clear (Glozak, Sengupta *et al.* 2005).

### **Class I HDACs:**

HDAC1 in association with the Sin3 corepressor complex participates in the deacetylation of p53. The acetylation sites on p53 are also associated with ubiquitination which suggests that the acetylation of p53 stabilizes p53, whereas deacetylation promotes degradation of p53 via ubiquitination. Class I HDACs also deacetylates YY1 protein (Yin Yang 1) which is a DNA binding protein and has a dual role in transcription activation as well as repression.

One of the most interesting class I HDACs is HDAC8 which has several non-histone protein substrates. *In vivo* evidence for histone deacetylation function for HDAC8 is not determined so far, however, several non-histone proteins are identified such as substrates: inv (16), smc3, ERR- $\alpha$ , CREB, p53, ARID1A (Wolfson, Pitcairn *et al.* 2013).

HMG proteins are involved in DNA bending at distorted regions by which they regulate chromatin structure organization. SRY is one of the HMG family protein which is deacetylated by HDAC3 which results in the loss of nuclear localization of SRY. This is one example where HDACs can regulate chromatin compaction in an indirect manner (Thevenet, Mejean *et al.* 2004). HDAC3 also deacetylates STAT3 which is a cytosolic transcription factor involved in cytokine dependent pathways. Deacetylation of STAT3 affects dimerization ability of STAT3, which is important for its function.

### **Class II HDACs:**

HDAC6 is a major deacetylase in the cytosol. It has been shown to deacetylate different substrates.  $\alpha$ -tubulin is one of the most studied substrate of HDAC6 which is involved in the cell

motility, adhesion, immune synapse and ciliogenesis functions. Cortactin is another substrate of HDAC6 that has a role in the cell motility. Also HSP90 has been identified as HDAC6 substrate, which participates in GR maturation, kinase activation and micropinocytosis. Another substrate of HDAC6 is IFN $\alpha$ R whose function is unknown. Interestingly HDAC6 contains a ubiquitin binding domain, which plays an important role in stress response, autophagy, macropinocytosis and aggresome formation by interacting with ubiquitin (Yang and Seto 2008).

### 1.3.6. Mechanism of action

#### 1.3.6.1. Mechanism of class I HDACs

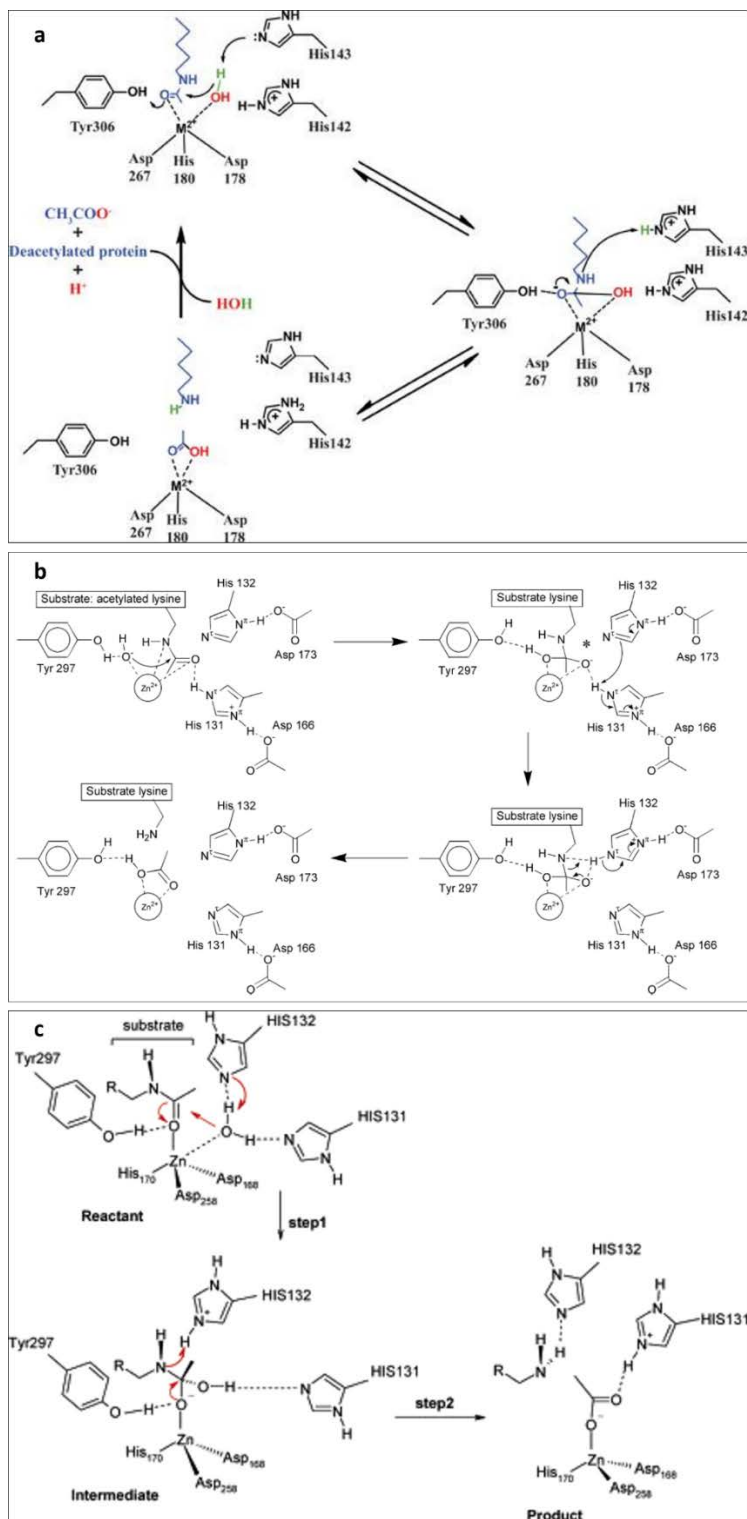
Due to the structural conservation of the HDAC active site residues, it is assumed that all HDACs have a similar mechanism of action. Yet, the HDAC catalytic mechanism remains a matter of debate. So far, three alternate mechanisms were proposed (Figure 21). The first mechanism was proposed based on the crystal structure of the HDAC-like protein HDLP of *Aquifex aeolicus* in presence of SAHA and TSA (Finnin, Donigian *et al.* 1999). According to this mechanism, the divalent Zn<sup>+2</sup> at the bottom of the catalytic pocket coordinates with two aspartate residues D178 and D267 (amino acid numbers corresponding to human HDAC8), and a histidine, H180, residue. Also a pair of histidine-aspartate (H142-D176 and H143-D183) charge-relay hydrogen bond dyad systems present near the bound substrate. This arrangement resembles catalytic triad in serine proteases, and is conserved among all HDACs except class II and class IV HDACs where D183 is replaced by glutamine and asparagine amino acids respectively. A tyrosine (Y306 in HDAC8) residue and also two K<sup>+</sup> ions are important constituents of HDAC structure.

According to the mechanism proposed by Finnin *et al.*, the substrate binds in the active site pocket where Zn<sup>+2</sup> co-ordinates the carbonyl oxygen of the substrate and the catalytic water molecule. The carbonyl carbon is polarized by the Zn<sup>+2</sup> and acts as an electrophile and orients in close proximity of the water molecule. H143 is initially protonated and H142-D176 charge relay system acts as general base and increases the nucleophilicity of water by abstracting a proton. Upon nucleophilic attack of the water on the carbonyl carbon of the substrate, a tetrahedral intermediate is formed which is stabilized by Zn<sup>+2</sup> and also by a hydrogen bond with the hydroxyl group of the catalytic tyrosine. Finally, the carbon-nitrogen bond breaks with the acceptance of a

proton from the histidine-aspartate (H143-D183) charge relay system, and the leaving acetate and lysine side chain remain as final products.

A second mechanism was proposed using the density functional theory (Vanommeslaeghe, De Proft *et al.* 2005). In this mechanism, H143 stays neutral initially, and after substrate binding H142 deprotonates the water molecule and results in the tetrahedral intermediate formation due to the nucleophilic attack of the water molecule. At this step, a proton is transferred from H142 to H143 which is more acidic because of its partial solvent exposure. This leads to deprotonation of H143 by the lysine  $\epsilon$  amino group and subsequent breakage of C-N bond which releases the acetate and a protonated lysine residue.

Recently a third mechanism was proposed using DFT quantum mechanics / molecular mechanics (QM/MM) (Corminboeuf, Hu *et al.* 2006). In this mechanism H142 and H143 both are not protonated initially. Upon binding of the substrate, the water molecule is protonated by H143 which results in tetrahedral intermediate formation. H143 transfers a proton to the amide nitrogen, which leads to the breakage of the amide bond and product release. Finally, a revised new mechanism supports the first proposal that H143 acts as both general base and general acid catalyst, while H142 is protonated and acts as an electrostatic catalyst (Gantt, Decroos *et al.* 2016).

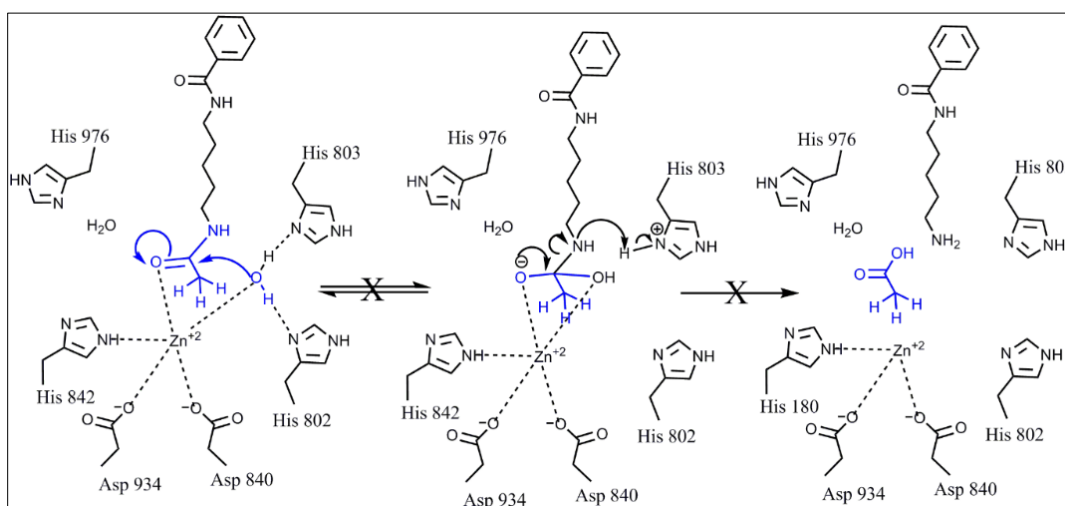


**Figure 21: Class I HDACs mechanism of action.**

a) Class I HDAC mechanism of actions as proposed by Finnin (a), Vanommeslaeghe (b), and Corminboeuf (c). In a) amino acid are numbered according to human HDAC8, in b) and c) amino acid numbers are corresponding to HDLP.

### 1.3.6.2. Mechanism of Class II HDACs.

In class IIa HDACs the absence of a tyrosine at the catalytic site hinders the deacetylation reaction. The substrate binding occurs similarly as in the class I HDACs where it binds to the catalytic zinc and water molecule. The nucleophilic water molecule attacks the carbonyl carbon of the acetyl-lysine which is rendered due to the lack of a hydrogen bond from hydroxyl group of tyrosine. Further, the intermediate undergoes protonation by histidine residue (H803 of HDAC4) to leave the products acetyl group and lysine residue (Figure 22). In general class IIa HDACs participates in interaction with HDAC3 to fulfill their deacetylation activity. Due to lack of tyrosine the active site pocket of class IIa HDACs becomes wider and may bind bigger acyl-lysine compounds.

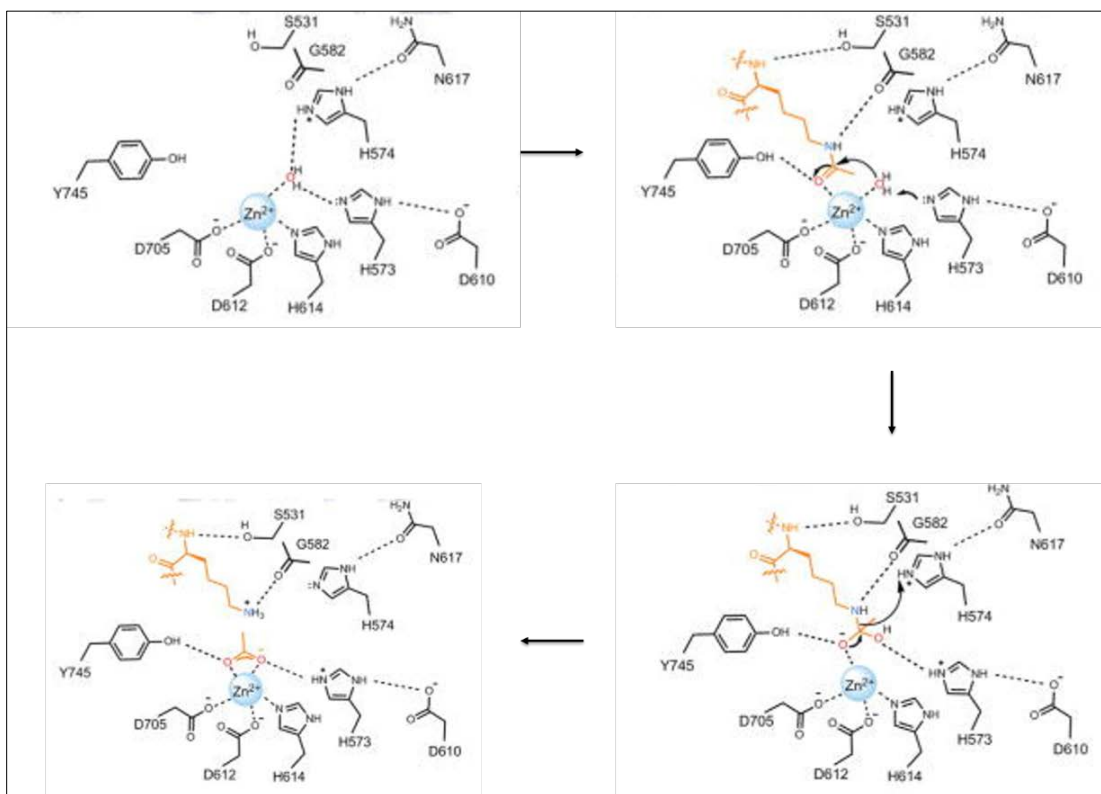


**Figure 22: Class IIa HDACs mechanism of action**

Amino acid numbers correspond to HDAC4. Picture adapted from (Bonomi, Mukhopadhyay *et al.* 2015).

In class IIb HDACs the catalytic mechanism is similar to class I HDACs (Figure 23). The catalytic zinc of HDAC6 coordinates active site residues and a water molecule to form a tetra coordinated state. When the substrate binds to the catalytic zinc, it does not displace the water molecule and results in the penta-coordinated metal ion. The nucleophilic attack of the water molecule leads to the formation of the tetrahedral intermediate which is further stabilized by Y745 (equivalent to Y306 in HDAC8). In HDAC6 the tandem histidines of the charge relay system can act as general acid and general base individually, which is in contrast to the proposed HDAC8 mechanism where only one histidine participates in the protonation, while the other histidine acts

as electrostatic catalyst. Further, the tetrahedral intermediate collapses to form acetyl and lysine products.

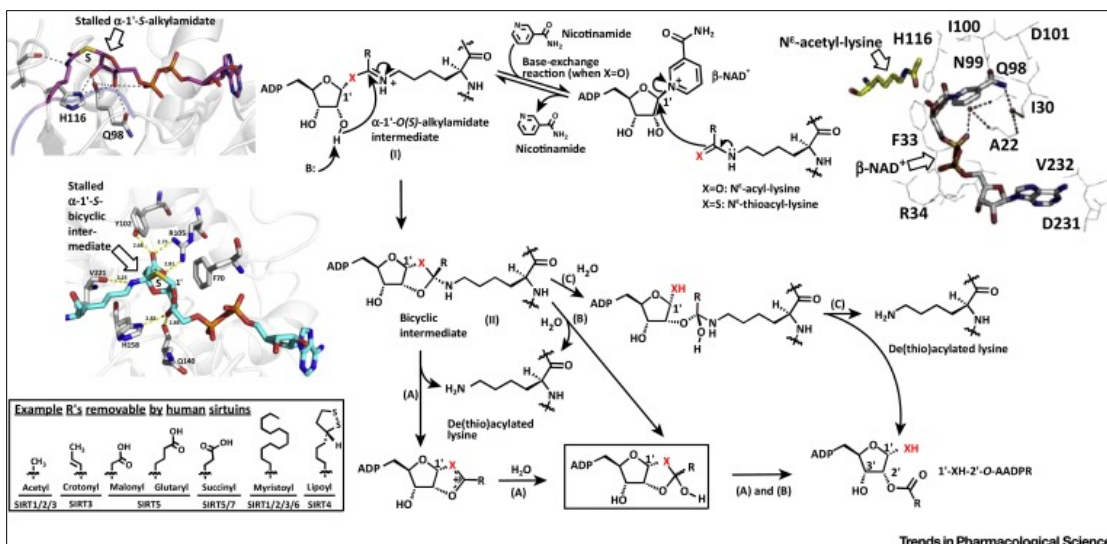


**Figure 23: HDAC6 mechanism of action**

Picture modified from (Hai and Christianson 2016).

### 1.3.6.3. Mechanism of action of sirtuins

Unlike  $Zn^{+2}$  dependent HDACs, sirtuins requires  $NAD^{+}$  for their catalytic reaction (Figure 24). Sirtuins reaction mechanism follows a sequential kinetic mechanism in which a ternary complex is formed by  $NAD^{+}$  and acyl- or acetyl-lysine (Jiang, Liu *et al.* 2017). In the active site pocket, acetyl-lysine binds before the  $NAD^{+}$  and the release of products follows the order as nicotinamide which is followed by deacetylated product and ADP-ribose. In presence of acyl-lysine and  $NAD^{+}$ , sirtuins form a complex which is known as Michaelis complex, which then passes through two intermediate stages alkylamidate and bicyclic intermediate which follows either of the proposed three steps to release the products.



**Figure 24: Sirtuin mechanism of action**

Steps involved in sirtuin mechanism are depicted in sequence. I mechanism is for deacetylation and second mechanism is for deacylation reactions. Intermediate stages were supported with the crystal structures. B: is general base. Different R groups that can be deacylated or depicted in the left bottom corner. A to C path indicates three possible ways to collapse bicyclic intermediate. Picture adapted from (Jiang, Liu *et al.* 2017).

### 1.3.7. Histone deacetylases in diseases and therapy

#### 1.3.7.1. Inhibition of HDACs in human diseases

One of the interest in HDAC studies is to understand the role of deacetylation in human health and diseases. Hypo- and hyper-acetylation are associated with many diseases due to the alterations of different cellular functions like gene regulation, cellular development and differentiation, cell cycle progression and chromosome translocation. HDACs deacetylation activity is predominantly associated with several human cancers, but also non-malignant diseases such as neurodegenerative diseases, inflammatory disorders, metabolic disorders and cardiovascular diseases (Wiech, Fisher *et al.* 2009, Shakespeare, Halili *et al.* 2011, Arrowsmith, Bountra *et al.* 2012). For example, class I HDACs are associated with lung cancer, ovarian cancer and also non-cancerous diseases such as cardiac diseases and gastric defects (Bartling, Hofmann *et al.* 2005, Song, Noh *et al.* 2005, Khabele, Son *et al.* 2007).



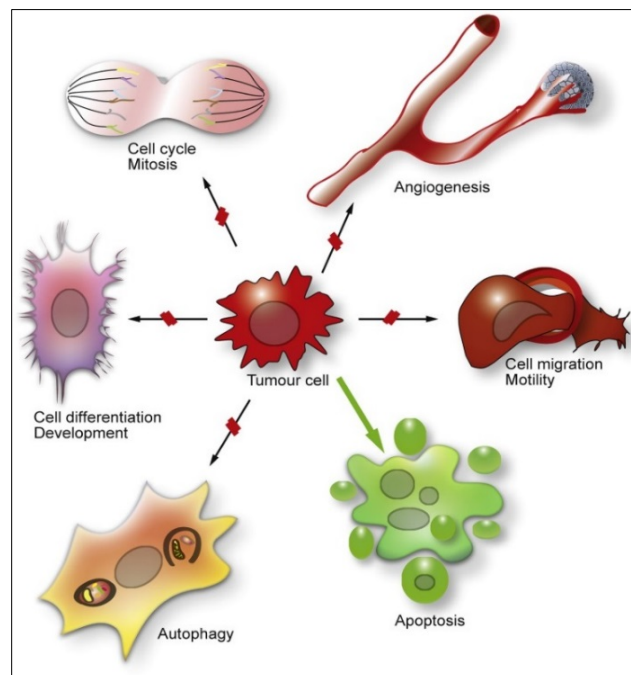
Inv (16) is a fusion protein that plays an important role in the transcription repression which is assisted by HDAC8. Abnormality in this process results in acute myeloid leukemia (Durst, Lutterbach *et al.* 2003, Krennhrubec, Marshall *et al.* 2007). Mutations in HDAC8 are observed in the Cornelia de Lange Syndrome, a developmental disorder (Decroos, Bowman *et al.* 2014). HDAC6 regulation is affected by estrogen signaling which is associated with estrogen receptor-positive breast cancer (Saji, Kawakami *et al.* 2005).

HDACs are involved in many different pathways, which are dysregulated during HDAC pathophysiology. An inhibition of an HDAC alters cellular functions in many aspects, as shown in (Figure 25). HDAC inhibitors (HDACi's) have been shown to play a role in neuroprotection. Valproic acid reduces brain damage in cerebral ischemia models (Ren, Leng *et al.* 2004). The role of HDACi's on memory retention have been investigated in many mice models which has implications in Alzheimer's disease. HDAC2 has a negative role in the memory formation: HDAC2 knock out and HDACi's treated mice showed increased memory formation and retention (Guan, Haggarty *et al.* 2009). HDACi's facilitate angiogenesis and myogenesis and improves recovery from myocardial infarction. HDACi's also plays a role in regeneration and reprogramming of differentiated cells in to induced pluripotent cells which has significance in stem cell therapy (Webster, Yan *et al.* 2013).

A novel promising approach in cancer therapy that is currently investigated consists in combining epigenetic drugs (epidrugs) with other cytotoxic drugs. This approach is proposed to increase the specificity for cancer cells and to reduce the cytotoxic effects. This approach has also showed good results in stabilizing the diseases by inhibiting cancer progression. Cytotoxic drugs such as calpeptin and telomere homolog oligo nucleotide are able to sensitizes breast and ovarian cell lines when used in combination with HDACi's (Sarkar, Horn *et al.* 2013). HDACi's were also shown to down regulate DNMT1 and induce demethylation of tumor suppressor genes. Combination of HDACi and azacitidine showed synergistic results (Sarkar, Goldgar *et al.* 2013). In a recent phase I/II clinical trials, patients with recurrent metastatic NSCLC (non-small-cell lung cancer) were given azacitidine in combination with entinostat. This treatment was well tolerated and the methylation of promoters (APC, RASSF1A, CDH13, CDKN2A) was found decreased in two hypermethylated promoter regions (Sarkar and Faller 2011). Combinations of ACY-125, a HDAC6 inhibitor, and bortezomib, a protease inhibitor, showed positive results in myeloma mouse

models (Santo, Hideshima *et al.* 2012). All these examples are highlighting the potential of combinational approaches in cancer therapy.

In addition, HDACs are found in all major human parasites and they regulate a wide range of functions in parasites (Andrews, Haque *et al.* 2012). Host macrophages are important components of the immune system that engulf pathogens upon activation by interferon- $\gamma$  (INF- $\gamma$ ). *T. gondii* can down regulate the interaction between macrophages and INF- $\gamma$ , hence affecting macrophage activation. This process can be reversed by HDACi's which shows HDACs importance (Lang, Hildebrandt *et al.* 2012). Another interesting strategy that parasites employ is the transformation of the host cell by altering the host cell epigenetic enzymes upon intracellular infection. *Theileria* parasite causes easy coast fever, and induces host cell transformation in the leukocyte epigenetic enzymes for the parasite replication (Marsolier, Perichon *et al.* 2015). HDAC9 is one of the major epigenetic enzyme that is manipulated by *Theileria* (Kinnaird, Weir *et al.* 2013). Likewise, HDACs play different set of functions in parasites, to survive in the host cell, which allow HDACs to stand as an important class of enzymes to develop potential anti-parasitic drugs.

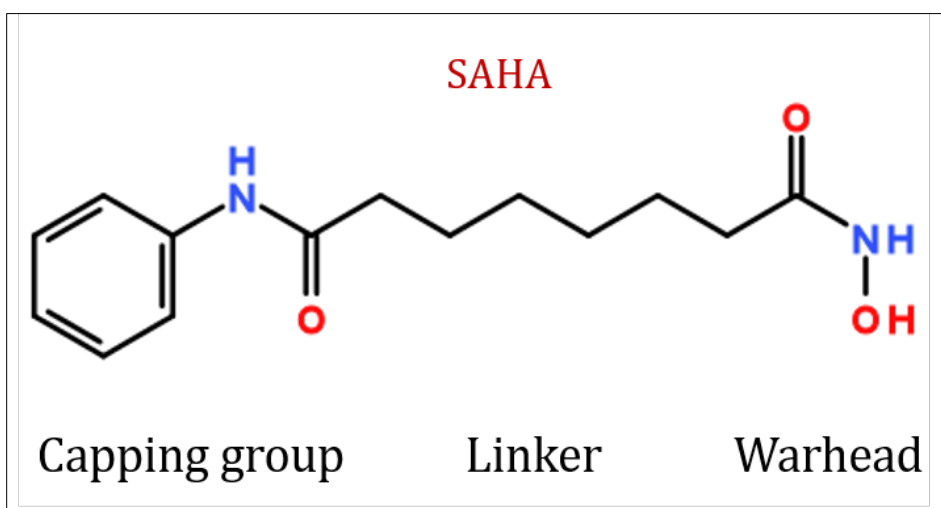


**Figure 25: Consequences of HDAC inhibition:**

HDACi's blocks few cellular pathways which are important for cell survival, and activates few pathways that leads to cell death. Image adapted from (New, Olzscha *et al.* 2012).

### 1.3.7.2. An introduction to HDAC inhibitors

HDAC inhibitors (HDACi's) affect different cellular functions. Some functions are activated upon treatment, whereas others are repressed. Inhibited functions include tumor cell proliferation, angiogenesis, autophagy, cell differentiation and development, whereas activated pathways include, cell cycle arrest, upregulation of crucial genes with anti-cancer effects and apoptosis (Marks, Richon *et al.* 2000, Bolden, Peart *et al.* 2006). To date, four HDAC inhibitors have been approved by the American and Chinese Food and Drug Administration (FDA), Vorinostat (Zolista), Romidepsin (Istodax), Belinostat (Beleodaq), Panobinostat (Farydak), and one by the Chinese FDA, Chidamide (Figure 26). Further, Quisinostat, Entinostat and Mocetinostat are under clinical trials. A table of HDACi's current status is shown in (Table 8).

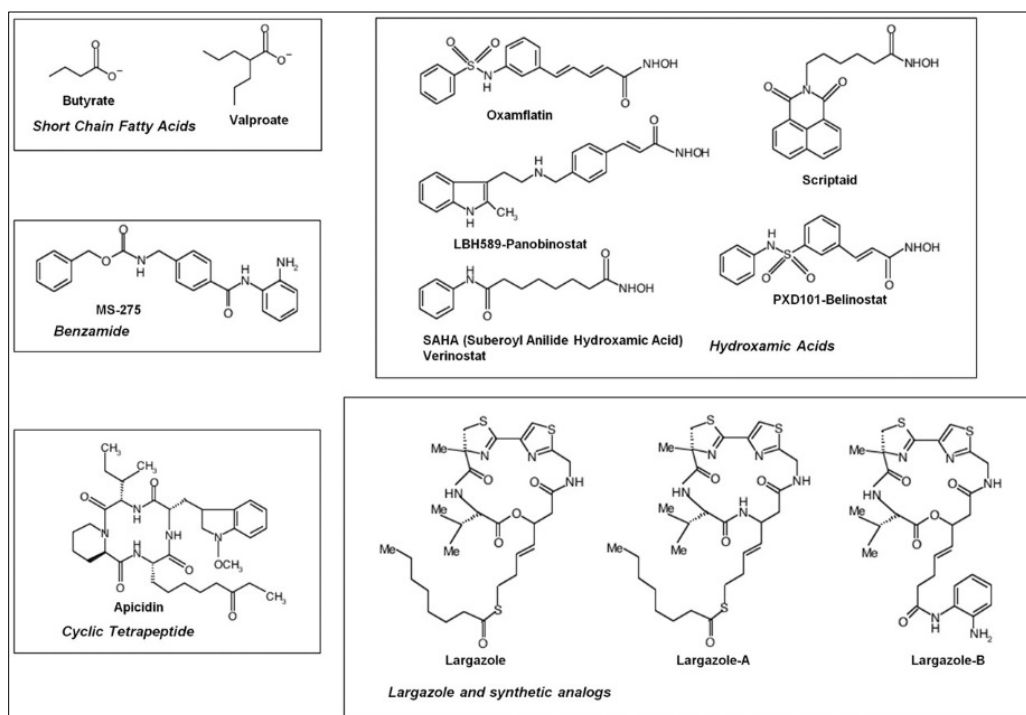


**Figure 26: Chemical constituents of a HDAC inhibitor**

A pan-HDAC inhibitor SAHA showing three different chemical moieties which are importantly found in a HDAC inhibitor.

HDACi's in general, mimic acetylated lysine residues and contains three chemical moieties a metal binding group (warhead) that chelates the catalytic  $Zn^{+2}$ , a linker group that mimics the lysine side chain which occupies the catalytic pocket, and finally a capping group that interacts with surface residues of the active site pocket. Several warheads have been proposed for HDACi

and that include carboxylic acids, hydroxamic acids, cyclic peptides, short chain fatty acids and Benzamides (Figure 27).



**Figure 27: Different classes HDAC inhibitors**

Picture is adapted from (Ghosh, Perrine *et al.* 2012)

SAHA (suberanilohydroxamic acid) is the pioneer of HDAC inhibitors and was approved by the American FDA in 2006 for the treatment of cutaneous T cell lymphoma. SAHA can inhibit HDAC1-9 without any isoform selectivity, and is therefore known as pan-HDAC inhibitor (Grant, Easley *et al.* 2007). SAHA causes hyperacetylation of histones and non-histone proteins such as p53 and heat shock protein 90 (HSP90). SAHA induces apoptosis and cell death in cancer cells, however, several side effects have been noticed in SAHA administration such as diarrhea, fatigue, nausea and anorexia.

Among different classes of HDACi's mentioned in the Figure 27, hydroxamate derivatives have different drawbacks such as poor pharmacokinetic properties, poor absorption in human body, off target effects with other metallo-enzymes, enzymatic hydrolysis in the cell, etc. (Mann, Johnson *et al.* 2007, Botta, Cabri *et al.* 2011). Due to their off target effects, hydroxamate derivatives may cause nausea, anemia, thrombocytopenia etc, Therefore, inhibitor selectivity is a very important aspect when designing HDACi's (Mann, Johnson *et al.* 2007). However, there are

few advantages that also exist for using hydroxamate inhibitors. Due to their pKa of 9-9.5, the hydroxamate moiety remains neutral at physiological pH, which helps with the cell membrane permeability (Wang, Helquist *et al.* 2007). After co-ordination with the catalytic  $Zn^{+2}$ , the pKa of the hydroxamate moiety appears to be reduced and forms an anionic form. Due to this, the hydroxamate warhead more tightly coordinate with catalytic zinc and the tetrahedral state is further stabilized.

Three important constituents of a HDACi which are depicted in Figure 26, plays three different roles in inhibiting an HDAC enzyme. The warhead interacts with catalytic residues H142, H143 and Y306 (corresponding to HDAC8). Specifically, the hydroxyl group of the hydroxamate moiety replaces the catalytic water molecule, thereby the inhibitor is able to form a tetrahedral intermediate as mentioned in section 1.3.6.1. The linker of the inhibitor mimics the aliphatic side chain of the acetyl-lysine and it is located in the active site pocket where it forms Van der Waals interactions with active site residues. The capping group of the inhibitor interacts with amino acids on the surface of the catalytic pocket. By manipulating the chemical constituents of this general form of the inhibitor, one can potentially design more potent and more selective inhibitors.

**Table 8: HDAC inhibitors**

Table represents HDACis that are available to treat cancers and those under clinical trials.

<b>Inhibitor</b>	<b>Trade name</b>	<b>Classification</b>	<b>Target HDAC</b>	<b>Treatment</b>	<b>Status</b>
SAHA (Vorinostat)	Zolina	Hydroxamate	Pan HDAC	Cutaneous T-cell lymphoma	2006
Romidepsin	Istodax	Cyclic peptide	Pan HDAC	Cutaneous T-cell lymphoma	2009
Belinostat	Beleodaq	Hydroxamate	Pan HDAC	Peripheral T-cell lymphoma	2014
Panobinostat	Farydak	Hydroxamate	Pan HDAC	Multiple myeloma	2015
Chidamide	-	Benzamide	Class I and IV and HDAC 10	Peripheral T-cell lymphoma	2015 (approved by Chinese FDA)
Mocetinostat	-	Benzamide	HDAC 1 (minimal HDAC2, 3 and 11)(Fournel, Bonfils <i>et al.</i> 2008)	Various cancers	Clinical trials
Entinostat	-	Benzamide	HDAC 1 and 3	Hodgkin lymphoma, lung and breast cancer	Clinical trials
Givinostat	-	Hydroxamate	Pan HDAC	Refractory myeloma	Clinical trials

### 1.3.7.3. Selective inhibition of HDACs

The majority of HDAC inhibitors are non-selective inhibitors. Due to their specific and their wide range of cellular functions, it is very important to target HDACs with decreased adverse effects, and also to deliver the inhibition of a desired biological function, through isozyme selectivity and specificity. So far, the FDA approved inhibitors are pan-HDAC inhibitors. Recently two HDAC inhibitors NCC149 and PCI-34051 were proposed as HDAC8 selective inhibitors, but the molecular basis of their inhibition mechanism remains poorly understood.

Due to the high structural conservation of HDACs active site pockets, it is difficult to design isozyme-selective inhibitors. Structural information is a much needed information to overcome this problem. Capping groups of HDACi's that interact with distinct surfaces can provide the possibility to design isozyme-selective inhibitors.

Cyclic peptides which mimics substrate peptides displays class I selectivity. The natural compounds trapoxin A and B are examples of cyclic peptides that possess class I selectivity (Furumai, Komatsu *et al.* 2001). Class II HDACs active site pocket surface does not allow the binding of cyclic peptide inhibitors. However, due to its structural similarity with class I HDACs, HDAC6 can bind cyclic peptide inhibitors (Hildmann, Wegener *et al.* 2006). In addition, Tubacin, a structural mimic of HDAC6 canonical substrate  $\alpha$ -tubulin, can inhibit HDAC6 and also HDAC1 (Haggarty, Koeller *et al.* 2003). Tubacin inhibits only one catalytic domain of HDAC6.

The loops of different sizes at the rim of active site pocket leads to different shapes of entry sites for distinct classes of HDACs. These loops are flexible enough to undergo different conformational changes during substrate recognition and binding, which facilitates substrate specificity. In HDAC8 loop L2, at the surface of the active site pocket contains a conserved amino acid, D101, that is present in class I and II, and which has significance for substrate binding. Mutation of this residue decreases enzyme activity (Bottomley, Lo Surdo *et al.* 2008, Weerasinghe, Estiu *et al.* 2008). Loop L1 is more flexible in HDAC8 than in other class I HDACs, and participates to the formation of the acetate release channel in association with F152.

HDAC isozyme-selective inhibition can be achieved by the use of structural information and by optimizing the chemical moieties of inhibitors. Benzamides are more selective for class I HDACs compared to hydroxamate inhibitors (Chou, Herman *et al.* 2008, Khan, Jeffers *et al.*

2008). Benzamides exhibit slow binding properties in HDAC catalytic pocket (Lauffer, Mintzer *et al.* 2013).

In class I HDACs, a foot pocket is formed by hydrophobic amino acids near the catalytic pocket and is absent in class II HDACs. After deacetylation, the protonated lysine side chain leaves through the catalytic channel while the acetate leaves through this acetate channel (Wang, Wiest *et al.* 2004). This foot pocket is 14Å wide and connects the surface of the HDAC to the catalytic pocket. L1 rearranges to facilitate the release of the acetate with the help of two other amino acids R27 and R16, which guide the acetate exit via hydrogen bonding (Wang, Wiest *et al.* 2004).

In HDAC8 this internal channel is distinct compared to other class I HDACs, whereas it shows two subchannels with a 12 Å release channel and a 14Å disposal channel. R37 and W141 guide acetate release by concerted mechanism (Whitehead, Dobler *et al.* 2011). When acetate moves in the disposal channel, the release channel is closed by W141 due to the rearrangement of R37. Acetate releases through the channel only when W141 is in out conformation. Water molecules reach the catalytic pocket either through the catalytic channel (Whitehead, Dobler *et al.* 2011) or by exchange of acetate via the acetate release-channel (Vannini, Volpari *et al.* 2004). Apart from this several surface amino acids that facilitate acetate release. Impairment in acetate release inhibits HDAC8 activity (Lee, Rezai-Zadeh *et al.* 2004).

Pockets in HDACs are distinct and are key determinants for the selective inhibition (Burli, Luckhurst *et al.* 2013). Trifluoromethoxydiazolyl (TFMO) compounds forms a U shape that occupies the lower foot pocket of class IIa enzymes which shows selectivity towards this class of enzymes (Lobera, Madauss *et al.* 2013). Other possibilities are to target the surface of active site pocket, targeting specific complexes and other interacting partners of HDACs which are important for deacetylation activity. Allosteric sites could also be used to design isoform selective inhibitors such as IP4 binding site in HDAC 1-3 and CCHC motif in class II HDACs. Distinct pockets in HDACs provide an additional degree of selectivity opportunities, where the inhibitors need not to be a strong chelators of a catalytic metal atom, but extra protein-ligand interactions can achieve high potency.

One major limitation in developing HDAC therapeutics is the lack of understanding about HDAC and canonical substrate interactions. How HDACs recognize substrates? Especially in case



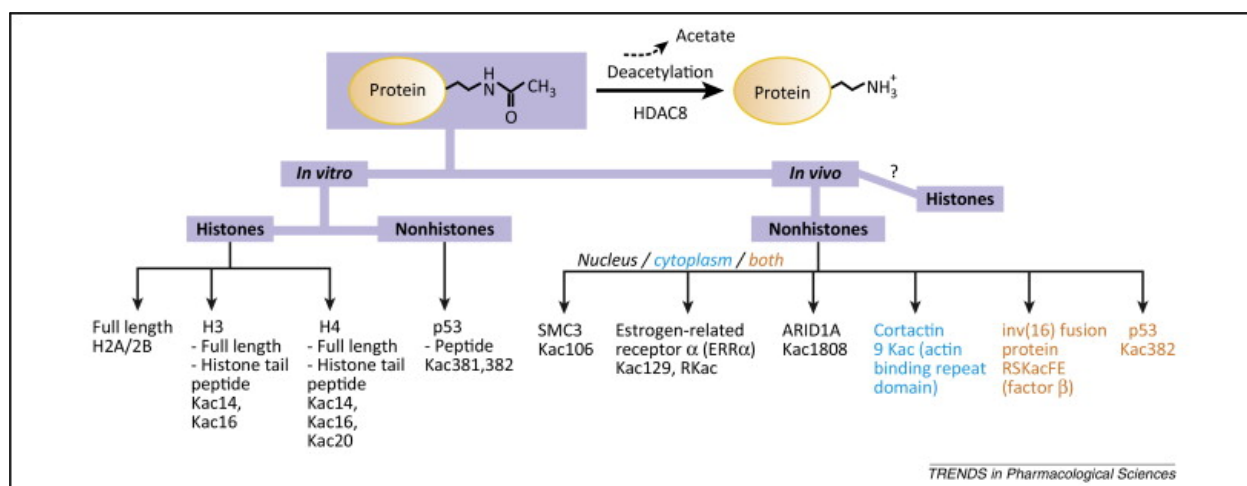
of HDAC8, in spite of having a large number of non-histone substrates, the mechanism of recognition and mechanism of catalysis are poorly understood. This fundamental aspect can be understood by examining one of the HDAC8 major substrate, the Cohesin complex.

### 1.3.8. HDAC8

#### 1.3.8.1. Introduction to HDAC8

Among all the HDACs, HDAC8 is one of the best characterised enzyme which is associated with several cellular functions and pathophysiology. The first crystal structure of HDLP was solved in 1999, which is a homologue to human HDAC8 (Finnin, Donigian *et al.* 1999). After a while, the first crystal structure of human HDAC8 (hHDAC8) was solved in 2004 (Somoza, Skene *et al.* 2004). Ever since a plethora of HDAC8 structures were published encompassing complexes with different inhibitors and mutated versions. To date, 52 HDAC8 crystal structures were deposited in the protein data bank, of them 45 are of human origin. These numbers reflect the importance and the amount of characterization carried out with HDAC8 enzyme. Detailed structural aspects of HDAC8 were described in the section 1.3.3.1).

The debate on HDAC8 substrates is a long-lasting aspect which remains to be addressed. *In vitro* HDAC8 can deacetylase all histones, but *in vivo* histone deacetylation by HDAC8 is unnoticeable. The first non-histone substrate of HDAC8 identified was p53, a transcription factor. The deacetylation activity of p53 derived peptide was much faster than the deacetylation of H4 derived peptide substrates, which suggests that non-histone proteins are preferable substrates of HDAC8 *in vivo* (Wolfson, Pitcairn *et al.* 2013). Identification of p53 as HDAC8 substrate also suggested the cytosolic localization of HDAC8 and further triggered the identification of non-histone substrates of HDAC8. In the (Figure 28) a list of non-histone substrates of HDAC8 was mentioned which included Smc3, Estrogen-related receptor  $\alpha$ , AR1D1A, cortactin, inv (16), p53 etc. The substrates of HDAC8 are distributed in both cytosol and nucleus which also reflects the HDAC8 localization.



**Figure 28: HDAC8 substrates**

Picture adapted from (Chakrabarti, Oehme *et al.* 2015).

HDAC8 is implicated in many cancer and non-malignant diseases. HDAC8 overexpression is observed in different cancers like colon, breast, lung, hepatocellular carcinoma, pancreas tumours, acute myeloid leukaemia, acute lymphocytic leukemias, gastric cancer, child hood tumours of the nervous system like neuroblastoma etc (Wu, Du *et al.* 2013, Song, Wang *et al.* 2015, Wilmott, Colebatch *et al.* 2015). Overexpression of HDAC8 in cancer cells results in cell proliferation and inhibits apoptosis.

Among the FDA approved anti-cancer drugs which target HDACs none are isozyme selective. However, with the increasing structural information of HDAC members, selective inhibition of isozymes is of growing interest. So far two HDAC8 selective inhibitors have been suggested which includes PCI-34051 and NCC149, where PCI-34051 was shown to induce PLC $\gamma$  mediated apoptosis in Jurkat cell lines (Balasubramanian, Ramos *et al.* 2008, Suzuki, Muto *et al.* 2014).

HDAC8 mutations are associated with non-malignant diseases, which includes Cornelia de Lange syndrome which will be discussed in the next section. In another case, HDAC8 mutations have been identified in an X-linked intellectual disorder which is characterised by truncal obesity, gynecomastia, hypogonadism and unusual face, where the symptoms are overlapping with the Wilson Turner syndrome (Harakalova, van den Boogaard *et al.* 2012). The Wilson syndrome is also an X-linked disorder in which HDAC8 gene mutations were proposed as the only reason for this disease.

HDAC8 is a unique enzyme among all other HDAC enzymes as it has few similarities and differences with other HDACs. Among class I HDACs, HDAC8 is the exception from participating in the formation of complexes, where HDAC1-3 are usually the constituents of corepressor complexes, and the possible reason is due to the divergent evolution of HDAC8 from the ancestors which helped it to function in isolation (Gregorette, Lee *et al.* 2004). Yet, several interesting reports have been shown that co-immunoprecipitated or pulled down HDAC8 associates with different protein partners. However the biological relevance and characterization of these complexes are not investigated (Wolfson, Pitcairn *et al.* 2013).

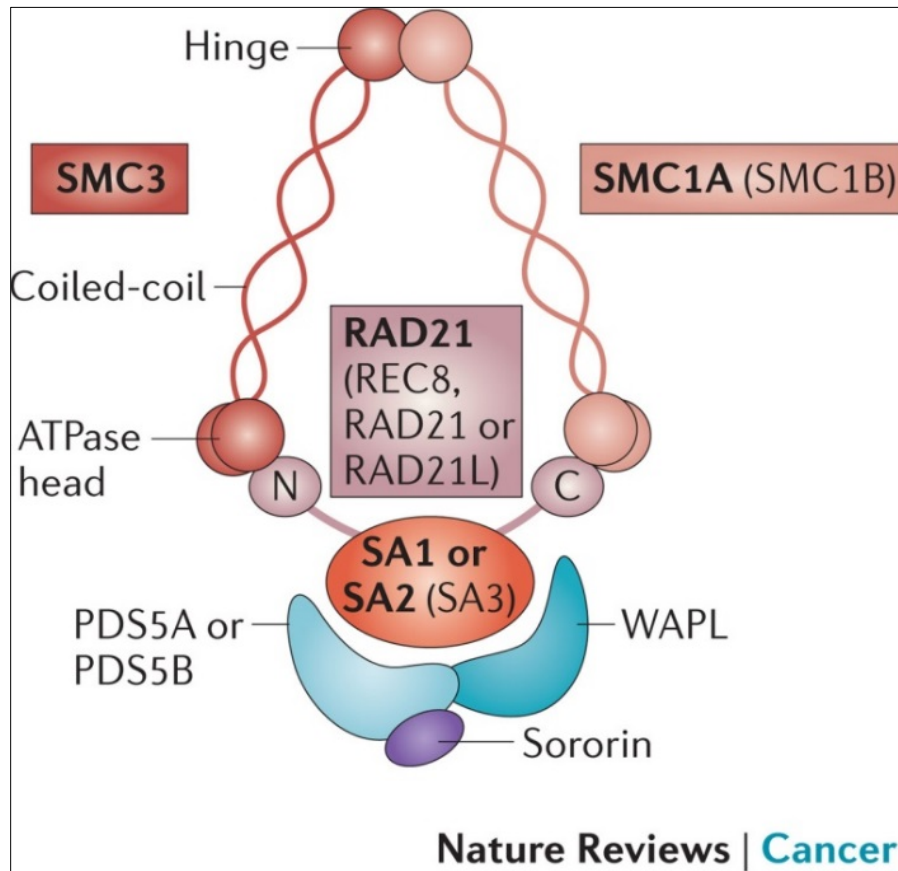
CREB is a cellular transcription factor which is activated upon phosphorylation and results in gene expression while dephosphorylation by PP1 inactivates CREB. HDAC1 plays an important role in CREB inactivation by targeting PP1 to CREB. Recently, HDAC8 has been shown to co-immuno precipitate with CREB and PP1, and *in vivo* studies shown the down regulation of CREB by HDAC8 expression (Gao, Siddoway *et al.* 2009). In another study, immune florescence staining experiments revealed co-localization of HDAC8 with smooth muscle  $\alpha$ -actin in murine fibroblast cells and explained HDAC8 role in the cell contraction (Waltregny, De Leval *et al.* 2004). Further, pull-down experiments demonstrate the interaction of HDAC8 with Hsp20 and cofilin which are actin interacting protein, which strongly suggests HDAC8 role in complexes that are involved in actin dynamics (Karolczak-Bayatti, Sweeney *et al.* 2011). Phosphorylated HDAC8 protects human ever shorter telomerase 1B (hEST1B) from E3 ubiquitin ligase CHIP (*C-terminal heat shock protein interacting protein*) by recruiting Hsp70. The bacterial two-hybrid assay suggested that Hsp70 forms a complex with phosphorylated HDAC8 and hEST1B, in presence of HOP1, which is a chaperone, this complex having a role in stabilization of hEST1B (Lee, Sengupta *et al.* 2006). However, these results are only preliminary and a detailed investigation must be carried out to understand the possible role of HDAC8 in complex formation. A non-histone substrate of HDAC8, the Cohesin complex is detailed in the next section.

#### 1.3.8.2. Role of HDAC8 on the Cohesin complex

From Archaea to higher eukaryotes every cell undergoes chromatin segregation during cell division where the genome has to be split equally into the two new daughter cells (Cobbe and Heck 2004, Gligoris and Lowe 2016). The ring-shaped Cohesin complex mediates chromatin segregation and is one of the important target of HDAC8. Apart from chromosome cohesion, this

complex also participates in chromosome condensation, chromatin organization, transcription regulation and DNA repair (Lopez-Serra, Kelly *et al.* 2014).

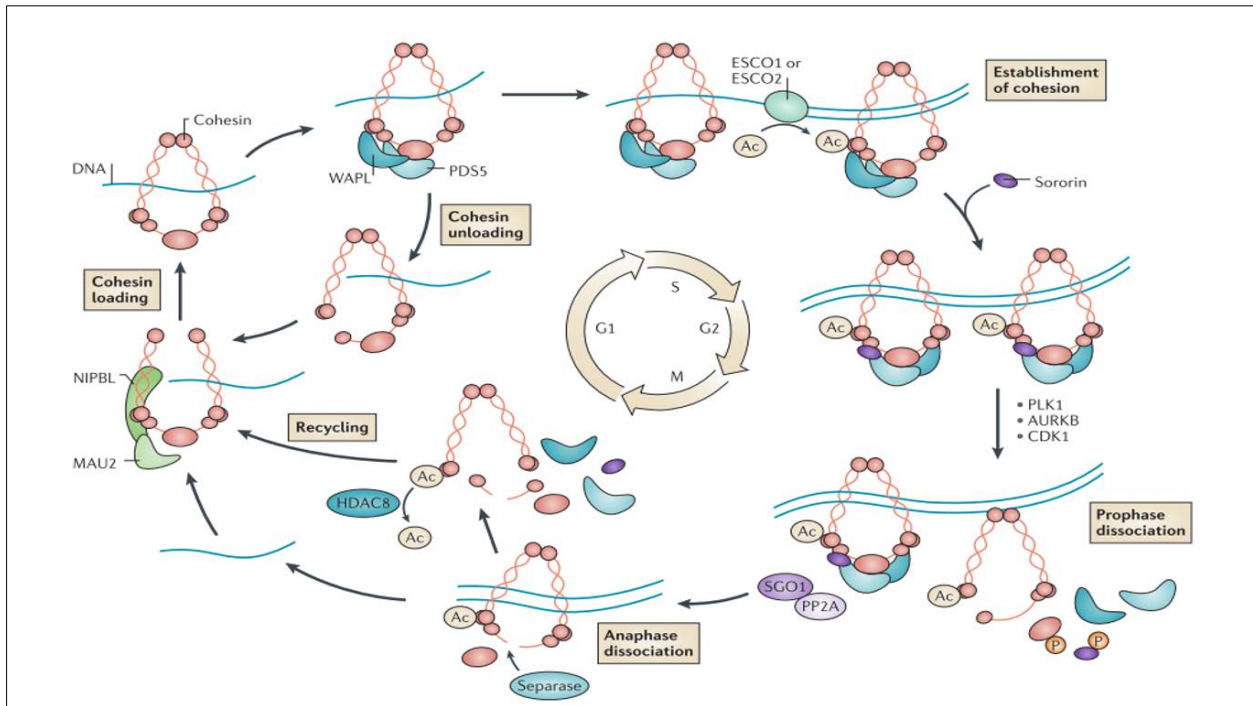
The architecture of the Cohesin complex is depicted in (Figure 29). SMC (Structural Maintenance of Chromosome) proteins contain three domains, the middle hinge domain, which helps with the homo dimerization, the N- and C- terminal ends that fold back at the hinge and interact together to forms a nucleotide binding ATPase domain (Cobbe and Heck 2006). These two regions are separated by a long-coiled coil region. Smc heterodimers with a third protein kleisin (Sister Chromatid Cohesion 1 - SCC1), forms a ring-shaped complex which holds sister chromatids during S phase of cell cycle (Schleiffer, Kaitna *et al.* 2003).



**Figure 29: Cohesin complex architecture.**  
Image adapted from (Losada *et al.* 2014)

In eukaryotes four SMC trimeric complexes have evolved through *smc* gene duplication and speciation (Cobbe and Heck 2006, Gligoris and Lowe 2016). The Cohesin complex is composed of Smc1, Smc3 and Scc1 and participates in sister chromatids cohesion. Condensin I

and condensin II contains Smc2, Smc4 and Brn1 and has functions in gene regulation, DNA repair and recombination. The Dosage compensation complex is a variant of condensin found in *Caenorhabditis elegans*, is composed MIX1, DPY27 and DPY26, and is involved in hetero chromatin formation. The Smc5-6 complex contains Smc5, Smc6 and kleisin delta (Nse4) and has an important role in DNA repair (Gligoris and Lowe 2016).



**Figure 30: Cohesion - cell cycle regulation.**  
See text for description. Picture adapted from (Losada 2014).

A graphical description of cohesion cycle is shown in (Figure 30). HDAC8 has been shown to deacetylate subunit Smc3 from the Cohesin complex. Cohesin is loaded onto sister chromatids in early G1 phase which is assisted by, nipped B like protein (NIBL) - MAU2 heterodimer. This heterodimer transfers ATP hydrolysis energy from the Smc head domains to the hinge domain to transiently dissociate the hinge dimer for the entry of DNA (16). Once the DNA is trapped inside the Cohesin complex, it cannot escape from the Cohesin complex because of the high tension created by the metastatic spindle. This arrangement resembles a ring where V shaped Smc1/3 heterodimers is closed by Scc1. The binding of Scc1 to the Smc1 and Smc3 head domains is not symmetrical. Specifically, the binding interfaces are different in Scc1-Smc1 and Scc1-Smc3. To

be precise, the C-terminal part of Scc1 is involved in the binding to Smc3 at the bottom of the head domain, while the N-terminal part of the Scc1 interacts with Smc3 head domain near the coiled coil region.

During telophase and G1 phase, the Cohesin complex which is loaded on chromatin can be released dynamically from the chromatin. This release is mediated by the dimer of PDS5 and WAPL (wings apart-like protein homologue), where this dimer helps in the opening of the Cohesin ring at the smc3-scc1 interface. However, the exact mechanism of how PDS5-WAPL mediates the unloading is unknown, but a possible hypothesis involved that the PDS5-WAPL dimer is involved in the allosteric regulation of Smc3 head domain and results in the ATP hydrolysis which destabilizes the Cohesin ring at Smc3-Scc1 interface and ultimately causes the unloading of Cohesin complex from chromatin (Hara, Zheng *et al.* 2014).

During replication, acetyltransferases ESCO1 and ESCO2 acetylates K105 and K106 of Cohesin Smc3 subunit. The acetylation of Smc3 promotes the recruitment of soronin to the PDS5 subunit. The binding of soronin to the PDS5 partially dislocates the WAPL and disrupts the interaction between PDS5 and WAPL at the N-terminal site. The interaction between PDS5 and WAPL still lasts at the middle region and at the C-terminal regions with cohesin, so WAPL will not dissociate from the complex. The complex of Cohesin with soronin is able to stabilize the cohesion of sister chromatids during the S phase.

Phosphorylation plays important role in the destabilization of the Cohesin complex. Polo-like protein (PLK1) phosphorylates Scc1, and AURKB (aurora kinase B), CDK1 (Cyclin Dependent Kinase 1) phosphorylates soronin. During mitosis the phosphorylation of soronin and Scc1 results in the destabilization of PDS5 and soronin interface, which helps WAPL to restore its interaction with PDS5 which ultimately results in the disassociation of soronin along with WAPL and PDS5. The heterodimer of Shugoshin (SGO1) and protein phosphatase 2A (PP2A) accumulates on Cohesin complexes at centromere to prevent phosphorylation and subsequent dissociation. Hence cohesin complexes at centromeres are very stable. At anaphase the separase enzyme cleaves Scc1, which breaks the cohesion to release the sister chromatids. Smc3 which is released from the complex can be reused by the next cycle of cohesion, after its deacetylation by HDAC8.

However, a clear molecular mechanism of the Cohesin complex is not fully understood. Crystal structures are only available for the partial domains and are not sufficient to make clear conclusions at the global level. Especially, the role of HDAC8 in the cohesion cycle is very poorly understood. The interacting partners of HDAC8 in the Cohesin complex, the interaction basis of HDAC8 with Smc3, and how it regulates the cohesion cycle is far from clear.

### 1.3.8.3. The Cornelia de Lange Syndrome

Malfunction of Cohesion cycle results in a congenital disorder known as Cornelia de Lange Syndrome (CdLS). Particularly, it is associated with the mutations of genes Smc1, Smc3, Scc1, NIPBL and HDAC8. In CdLS patients, nearly 17 HDAC8 mutations have been identified which represent different parts of HDAC8 structure. HDAC8 inhibition causes accumulation of acetylated Smc3, which delays the cell cycle progression, suppresses the proliferation and finally induces apoptosis (Dasgupta, Antony *et al.* 2016).

Mutations of amino acids which are responsible for HDAC8 activity as well as those on the surface, both can cause CdLS, which suggests that the deacetylase function of HDAC8, and the interaction of HDAC8 with Smc3 are both important factors in CdLS. Another set of residues that are important for catalysis are also observed as mutants in CdLS. For instance, H180, a residue which is important for catalytic Zn<sup>+2</sup> binding, is found mutated in arginine in CdLS. In this case, the mutant arginine residue can protrude into the catalytic pocket and as a result can inhibit the deacetylase function.

A few mutations of residues are found in close proximity to the L1 and L2 loops whose conformational flexibility is important for the optimal activity of the enzyme. Otherwise the functional activity is compromised due to the mutations that affect loop molecular dynamics. This observation stresses the importance of loops in the interaction of HDAC8 with the Smc3 subunit (Decroos, Christianson *et al.* 2015). These mutations explain three different aspects of HDAC8, which include molecular dynamics, catalytic activity and surface interactions that are important for full functionality of HDAC8.

### 1.3.9. On the interest of studying HATs and HDACs

The importance of acetylation in cell signaling and its implication in metabolism and diseases make of this mark and of the proteins that deal with acetylation important players in cell homeostasis and for drug design. Accordingly, HDACs and bromodomains are the focus of extensive studies towards drug development, with already approved drugs in the case of HDACs. In contrast, inhibitors of HATs and sirtuins appear more complicated to develop. For all these players, selective inhibition remains anyway a problem.

Another aspect to consider is the sometimes currently restricted basic knowledge on the protein players involved in acetylation. This lack of knowledge prevents targeting more precisely specific effectors for inhibition. If bromodomains can easily be characterized structurally and functionally, this is more complicated for HATs which are generally parts of large complexes. Surprisingly, even more open questions remain on sirtuins and HDACs concerning their mode of action (e.g. are they all involved in acetylation?) and their targets (what are their real targets and how are these recognized?).

Actually, these issues in basic and applicative research are related since a better understanding of the acetylation players' structure and function is useful to choose targets and to design selective inhibitors, whereas selective inhibitors can help decipher function. During my thesis, I have been addressing these two issues, dealing initially with selective inhibition to come back finally to more basic research.

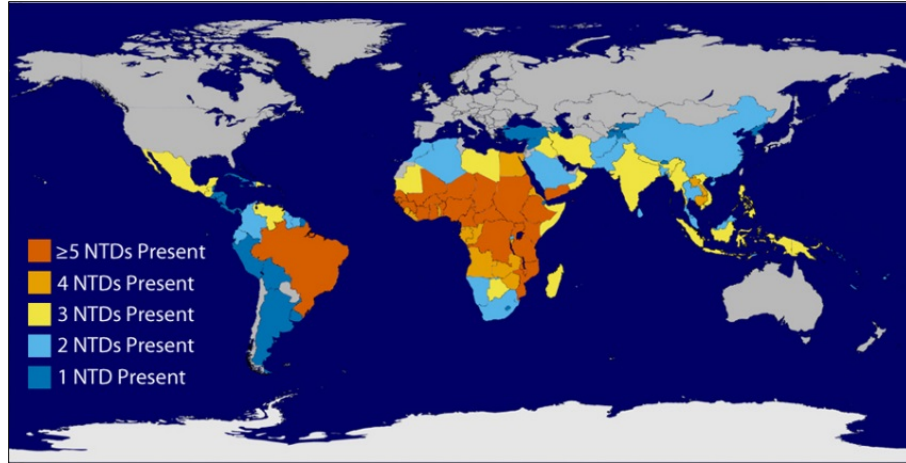
Interestingly, this approach was first directed towards the treatment of neglected diseases, but the knowledge obtained has brought me back to the human enzymes and the treatment of cancer.



## 1.4. Human neglected diseases

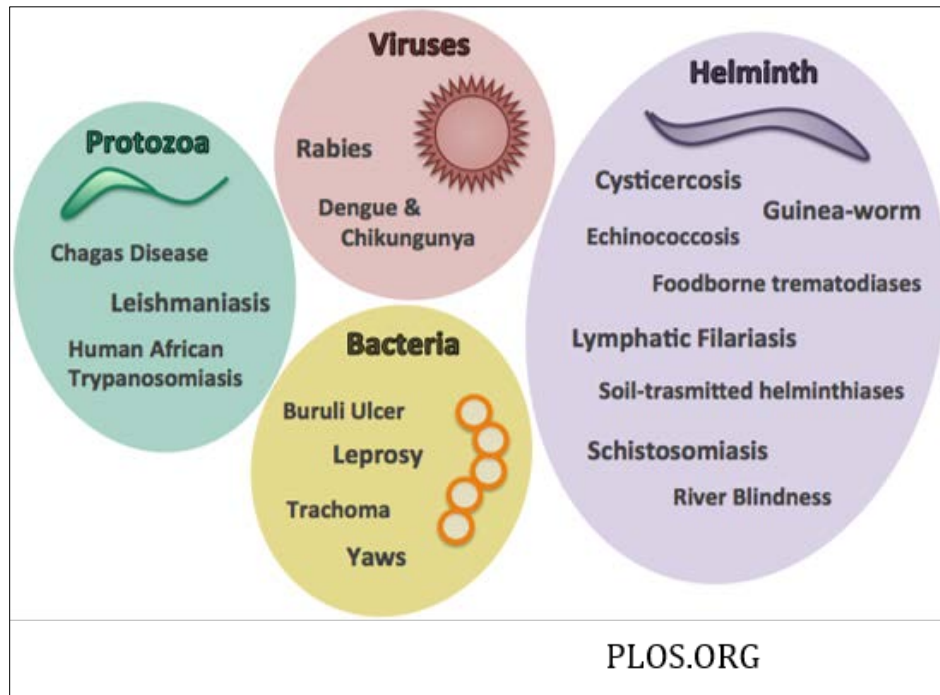
### 1.4.1. Neglected Tropical Diseases

Human neglected diseases or Neglected tropical diseases (NTDs) are a group of diseases caused by the infection of bacteria, eukaryotic parasites and viruses (Figure 32). NTDs are more prevalent in low and middle-income countries where people have little access to clean water and sanitary products. NTDs affect physical and cognitive development, and makes difficult to earn a living and also reduces the productivity. According to the World Health Organization (WHO), NTDs cause 300,000 deaths annually and affect more than one billion people in tropical and subtropical regions, widespread in 149 countries (2015) (Figure 31). According to WHO, 20 diseases have been declared as NTDs, few of them listed in the figure, categorically based upon the type of infectious organism. Due to lack of vaccines for parasitic infections, the treatment solely depends upon the use of drugs., when available. Further, to add complexity, in many cases different parasites infect at the same time the same individual, which renders difficult the treatment. Emerging resistance of parasites against existing non-specific drugs is raising a serious concern and calls urgently for the development of new drugs. However, due to the prevalence of NTDs in poor countries, very little research activities are dedicated to the search of drugs against these diseases that not only affect the health of billions of people but also are a burden for the world economy.



**Figure 31: Global Overlap of common NTDs.**

Specifically, guinea worm disease, lymphatic filariasis, onchocerciasis, schistosomiasis, soil-transmitted helminths, trachoma. Image source CDC (Centers for Disease control and prevention).



**Figure 32: NTDs caused by different type of infections.**

#### 1.4.2. Actual treatment of NTDs

The treatment of NTDs involve the massive amounts of drugs oral administration. So far, no vaccines are available to treat NTDs. For the treatment of schistosomiasis administration of Praziquantel, nearly 40 mg/kg body weight is recommended (Hopkins 2013). In the same way, Azithromycin 20 mg/kg for trachoma, Albendazole 400mg for lymphatic filariasis are recommended. There is no alternate treatment is available to treat NTDs.

#### 1.4.3. Importance of epigenetics and HDACs in NTDs

Due to the eukaryotic parasites' complex life cycles - with different hosts and multiple morphological stages - it is expected that epigenetic enzymes play crucial roles in the homeostasis of these parasites. Specifically, parasites have to respond quickly to sudden environmental changes, which is facilitated by the rapid change of gene expression profiles where epigenetic regulatory mechanisms play important roles. However, there is lack of information on how epigenetic enzymes regulate parasite homeostasis and respond to environmental stress.

Parasites also evolved unique regulatory mechanisms compared to human (Croken, Nardelli *et al.* 2012). For instance, pathogens choose many ways to escape the host immune response. One of them is antigenic variations: the pathogen varies its antigens and try to overcome host cell antibodies. In the case of *Plasmodium*, *this parasite* uses antigenic switching to produce variants in the *var* gene to escape from the host cell acquired immune response, and thus can sustain a long-term infection in the host cell (Merrick and Duraisingh 2010, Robert McMaster, Morrison *et al.* 2016). This antigenic switching solely depends on epigenetic regulations in parasites which restricts the simultaneous expressions of pathogenic genes to vary expression patterns. Therefore, understanding epigenetic regulations in pathogens can help to dysregulate antigenic switching, which could be an effective strategy to target pathogens and to increase host cell immune response (Rivero, Saura *et al.* 2010, Croken, Nardelli *et al.* 2012).

Yet, the major epigenetic modification acetylation also plays an important role in parasite homeostasis and HDACs have been reported as emerging drug targets to treat parasite diseases (Andrews, Haque *et al.* 2012). *Toxoplasma gondii* which causes toxoplasmosis, can switch

between replicative and non-replicative life cycles upon histone acetylation (Gomez-Diaz, Jorda *et al.* 2012).

One major problem with neglected diseases is that the drug discovery process is very time consuming and beyond the cost-effectiveness for the economic resources allocated for NTDs research. Clearly large pharmaceutical companies see no market in the treatment of NTDs and the only funding available is the institutional one. To be able to cope with these restrictions, one strategy that can be applied is the so-called piggyback strategy. This strategy consists in making use of already approved drugs and to modify them to be able to target the parasite's enzymes (Marek, Kannan *et al.* 2013). The major advantage of the piggyback strategy is that it speeds up the initial steps of the drug discovery process and that it should be more cost effective. The major drawback of the piggyback strategy is that the new drugs designed should not target the human enzymes anymore, adding another layer of complexity to the selectivity problem discussed previously.

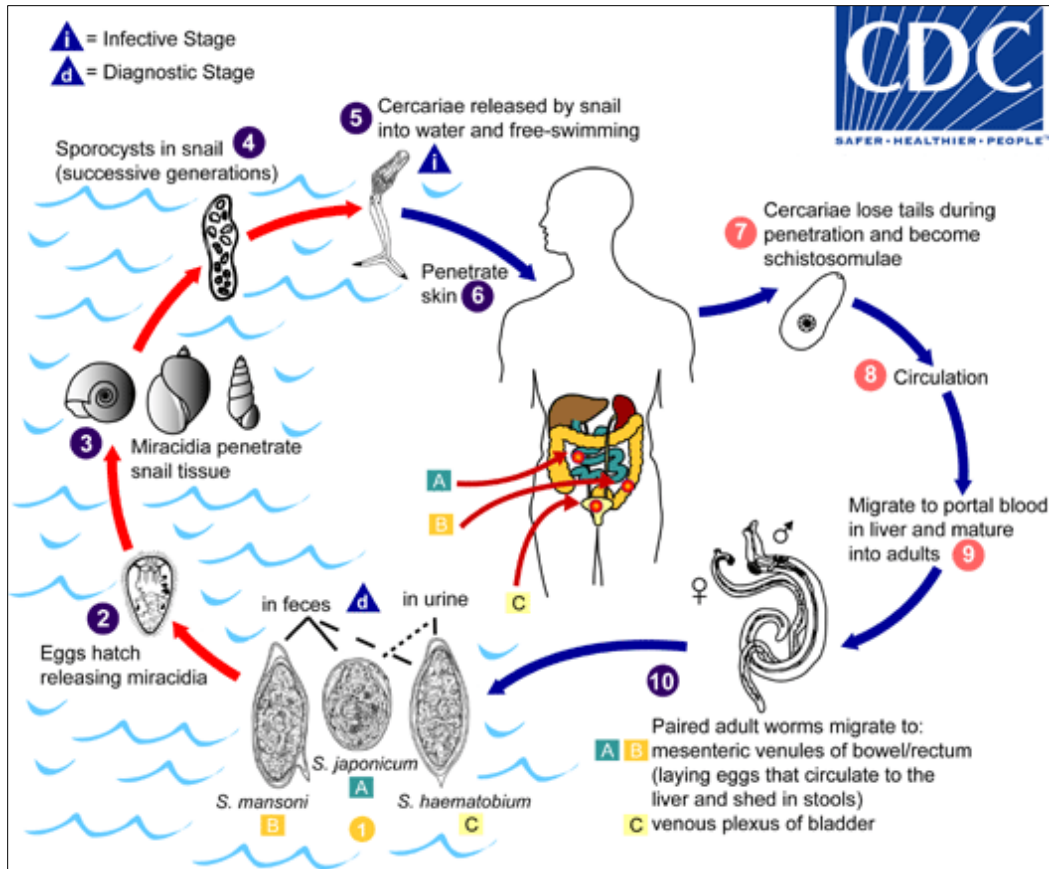
The importance of epigenetics in parasites makes of approved epidrugs good candidates for applying such a strategy. My laboratory has pursued this strategy within two large EU-funded projects, SEtTReND (Schistosoma Epigenetics: Targets, Regulation, New Drugs; 2010-2013) and A-ParaDDisE (Anti-Parasitic Drug Discovery in Epigenetics; 2014-2017). I have been involved in the second project which has built on the success of the first project, notably with the establishment of the proof of concept of the piggyback strategy in the case of the HDAC8 enzyme from *Schistosoma mansoni* (Marek, Kannan *et al.* 2013).

#### 1.4.4. Proof of concept: smHDAC8 a valid drug target to fight schistosomiasis by a piggyback strategy

##### 1.4.4.1. Schistosomiasis

During the course of European funded SEtTReND project, the piggyback strategy was investigated and a proof of concept was provided where schistosomiasis was pursued as case study. Schistosomiasis, also known as bilharzia, is a human neglected disease characterized by the infection of blood fluke parasite *Schistosoma* genus flatworms. After malaria, schistosomiasis is the second deadliest disease, affects nearly 200 million people worldwide (WHO and CDC reports). *Schistosoma* is of two types, intestinal schistosomiasis caused by *S. mansoni*, *S. japonicum* and *S. mekongi* and *S. intercalatum* and urogenital schistosomiasis caused by the infection of *S. haematobium*.

*Schistosoma* completes its life cycle in two hosts, humans and snails. Infection occurs when humans come in contact with fresh water that is contaminated with schistosoma infected snail (Figure 33). Larvae of schistosoma (cercariae) penetrate through the skin, and mature into adult worms inside blood vessels of the human body. The female produces eggs which can be excreted through urinary and digestive systems which can be infectious through contaminated water. Few eggs trapped in body tissues and travel to liver, intestine, lungs and bladder. After years of infection parasites can damage organs severely through the formation of cysts.



**Figure 33: Schistosoma life cycle.**

Picture adapted from CDC. Fresh water contaminated with Schistosoma eggs hatch and release miracidia which can infect snails to complete two generations of sporocysts. Cercariae released from snails, penetrate through human skin. In humans cercariae loses its tail and become schistosomulae which migrates to liver and other organs to mature adult worm. Schistosome adult worms always live in pair which is important for their homeostasis.

Schistosomiasis treatment solely depends on the oral administration of the drug Praziquantel (trade name biltricide, Cesol® 600 - Merck) for people above the age of 6. This drug is also used for the other human trematode diseases like clonorchiasis, cysticercosis, tapeworm infections etc. There is no treatment for children below age 6. The exact mechanism of action of Praziquantel is not clear. However, praziquantel is expected to increase the permeability of the parasite membranes leading to Ca<sup>2+</sup> uptake, which results in the contraction and death of the parasites (Pax, Bennett et al. 1978). Praziquantel is a non-specific drug against schistosoma, includes side effects such as headache, abdominal pain, and disturbance in digestive and nervous

systems. Also recent reports of increased resistance of schistosoma against Praziquantel stresses the urgent need for novel drugs against schistosomiasis (Cioli and Pica-Mattocchia 2003).

In schistosomes acetylation has been better characterized than other PTMs. Knock-down of CBP/300 and GCN5 acetyl transferases decrease the production of egg shell Smp14 protein and results in the damaged reproductive system of adult female worms (de Moraes Maciel, de Silva Dutra *et al.* 2004, Carneiro, de Abreu da Silva *et al.* 2014, Liu 2016). Class I HDACs HDAC 1, 2 and 3, and Sirt2 were identified in schistosoma, and upon inhibition with HDACi's, induced caspase 3/7 activity and result in death (Marek, Kannan *et al.* 2013, Schiedel, Marek *et al.* 2015, Singh and Pandey 2015). These observations indicate that acetylation is an important mechanism in schistosomes and has the potential to serve as valid drug targets.

#### 1.4.4.2. smHDAC8

Since human HDACs are the targets of the majority of FDA-approved drugs, these enzymes represented perfect targets for applying a piggyback strategy in schistosomes. In schistosomes, HDAC8 is the most abundant HDAC and is expressed at all stages of the life cycle (Oger, Dubois *et al.* 2008). This is in contrast to humans where HDAC8 is the least expressed HDAC among four classes of HDACs ( $Zn^{+2}$  dependent and sirtuins) (Hu, Chen *et al.* 2000). The availability of HDAC8 in all forms of schistosome life cycle gives an advantage over other HDACs, to develop an anti-schistosome drug target.

During the course of the SEtTREND project, smHDAC8 (HDAC8 of *S. mansoni*) was showed as a valid drug target to treat schistosomiasis. RNA interference mediated smHDAC8 down regulation studies in schistosomula suggested that smHDAC8 inhibition in schistosomula reduces the worm and egg recovery from infected mice, which further demonstrates smHDAC8 importance in parasite infectivity and homeostasis.

Comparison of our crystal structure of smHDAC8, notably its active site pocket, with the already available human HDAC8 structure (and also those of other human HDACs), revealed important structural differences which provided essential information for structure-guided drug

design (Figure 34). Specifically, one important difference between both active sites is the replacement of M274 in hHDAC8 by H292 in smHDAC8. This single amino acid change introduces a charged amino acid in the highly hydrophobic HDAC8 active site.

The second important difference concerns the conformation of F151 side chain in smHDAC8 that it is observed in flipping-out conformation, turned away from active site pocket, whereas in hHDAC8 (and all the other human HDACs) this phenylalanine is locked into a flipped-in conformation, turned towards the active site. The very reason behind this conformational difference is because L31 of hHDAC8 sterically prevents the phenylalanine to adopt a flipped-out conformation in hHDAC8. The presence of the smaller serine S18 residue at the equivalent position in smHDAC8 makes possible for F151 to adopt a flipped-out conformation. Importantly, S18 is well conserved in all HDAC8 from schistosoma species and in other pathogenic trematodes, providing an important feature for the design of drugs targeting selectively pathogenic nematodes.

These active site pocket key differences were investigated in presence of pan-HDAC inhibitors, SAHA and M344. SAHA binds in smHDAC8 active site pocket in a kinked position with F151 in a flipped-out conformation and the capping group of SAHA stack over the side chain of the catalytic tyrosine Y341 (Y306 in hHDAC8), whereas in hHDAC8 active site pocket SAHA adopts a straight conformation. M344 adopts a kinked orientation in hHDAC8 active site pocket while it is straight in smHDAC8 active site pocket, and is further, stabilized at the surface of active site pocket by tyrosine Y99 of smHDAC8 that makes contact with the capping group of M344.

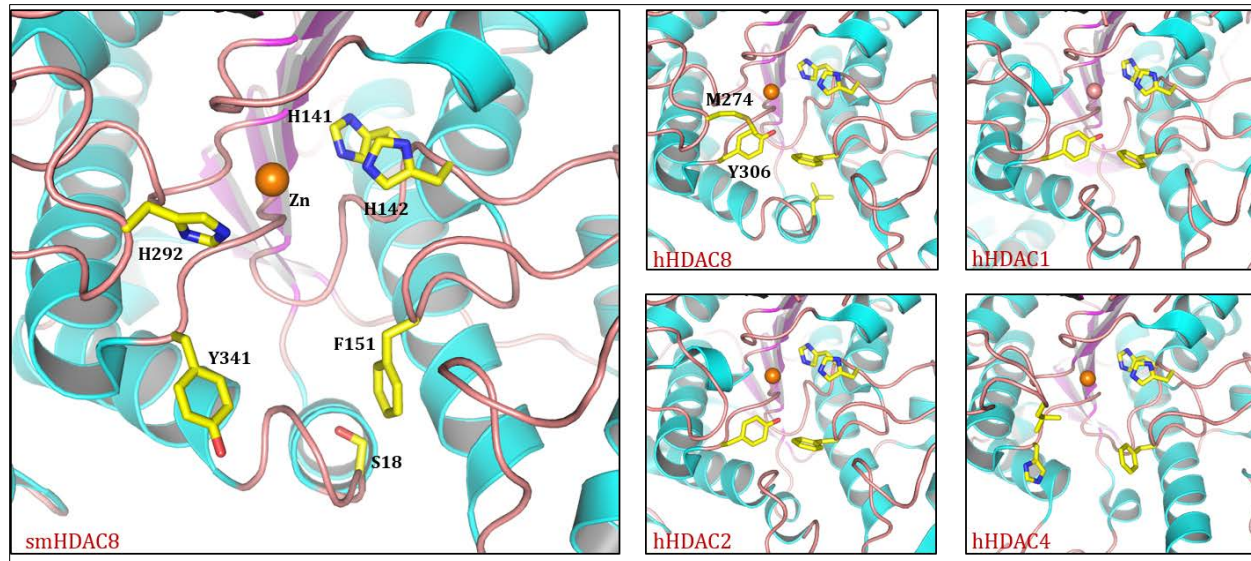
Interestingly, upon M344 binding, F151 from smHDAC8 adopts a flipped-in conformation, revealing the conformational flexibility of F151. Importantly, the schistosoma-specific flipped-out conformation widens smHDAC8 active site pocket, which provides, together with the M274H change, two important physico-chemical/structural changes for the design of bulkier and charged inhibitors that could selectively inhibit smHDAC8.

Virtual screening by our collaborator Wolfgang Sippl enabled to search for inhibitor scaffolds with bulky linker groups and yielded several hits. Inhibition assays by our collaborator Manfred Jung measuring the IC50 (half maximal inhibitory concentration) values for the initial hit compounds highlighted three compounds, J1037, J1038 and J1075, that retained inhibition of



smHDAC8 (as well as hHDAC8 and to a lesser extent hHDAC6) but that had lost their ability to inhibit hHDAC1 and hHDAC3.

The crystal structures of smHDAC8 with J1038 and J1075 revealed that J1038 can form a hydrogen bond with H292, but lacks schistosome-specific F151 flipped-out conformation, while J1075 has no interaction with H292 but favors F151 flipped-out conformation. These two inhibitors used the active site specificities of smHDAC8 in two different ways, and revealing the plasticity of smHDAC8 active site pocket. Further, *in vitro* and *in vivo* studies showed that these inhibitors can cause the separation of male and female schistosomes, reduce the egg production, and result in worm death via apoptosis.



**Figure 34: Comparison of catalytic subunit of smHDAC8 with human HDACs**

Together, these results provided a successful outcome to the SETReND project, and opened the way to the larger A-ParaDDisE project that used the same strategy but with four different parasites: schistosomes, *Plasmodium falciparum*, *Leishmania*, and *Trypanosoma cruzi*.

### 1.5. PhD thesis rationale: selective inhibition of parasitic enzymes and beyond

The SEtTReND project, by providing a proof of concept for the piggyback strategy, has brought essential information on how to set up this strategy and to apply it to other parasitic targets. It has also demonstrated the interest in targeting parasitic epigenetic enzymes. Yet, the selectivity issue had not been fully addressed since some human HDACs were still inhibited by the first hits characterized.

During my thesis, within the A-ParaDDisE project, I have further developed these aspects by looking at the inhibition of other parasitic enzymes. I have also continued the smHDAC8 selective inhibition project in collaboration with the other members of the A-ParaDDisE consortium. Interestingly, this has also brought me to characterize the human HDAC8 in comparison to smHDAC8, but also as a bona fide target, especially since hHDAC8-selective inhibitors have already been developed that could be used in my comparative study. This has led me not only to look at inhibitor binding, but also to analyze precisely the conformation of HDACs' active sites and to decipher the importance of loops in shaping this active site.

In the introductory part of this thesis, I have discussed of the importance of combining basic and applicative research in order to progress in these two complementary research directions. The detailed analysis of the active site conformation of HDAC8 already led the way to more basic research, raising the question of the specific recognition of its substrates by HDAC8. I have chosen to study the interaction of HDAC8 with its best characterized target, the Cohesin complex, to better understand the function of this enzyme and pave the way to the development of more potent, more selective and also more complex epidrugs.

Thus, my PhD thesis has aimed at addressing the following three major objectives:

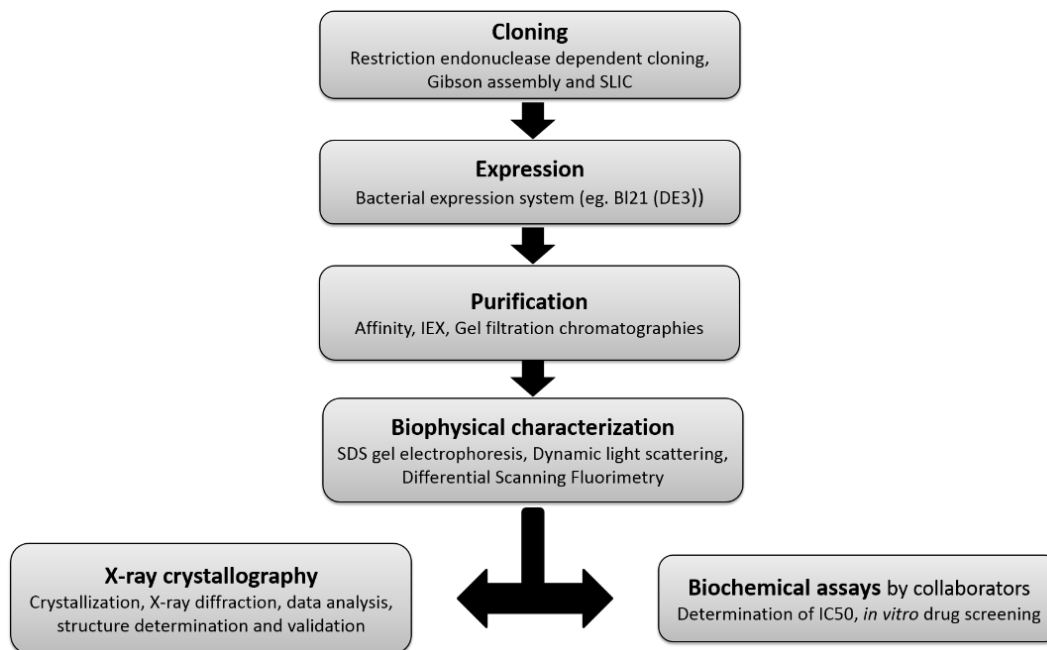
- (i) To apply the piggyback strategy to other parasitic enzymes in the acetylation pathway
- (ii) To further characterize the selective inhibition of HDAC8
- (iii)** To investigate HDAC8 interaction with a multi-protein complex substrate

# Materials and Methods

---

## 2. Material and methods

To understand the function of a protein through biochemical, biophysical and structural methods, the prerequisite is very often a very large amount of purified sample. During this thesis, I have used an integrated approach by combining molecular biology, biochemical, biophysical and structural tools, and I have collaborated with other teams on inhibition assays. A pipeline of experimental setup is provided in Figure 35. All materials used in this thesis are from IGBMC unless specified. In this section, a brief introduction to the different techniques and protocols used during my thesis work are summarized.



**Figure 35: Experimental approach:**  
Flow chart of methods used in this thesis.

## 2.1. Cloning strategies

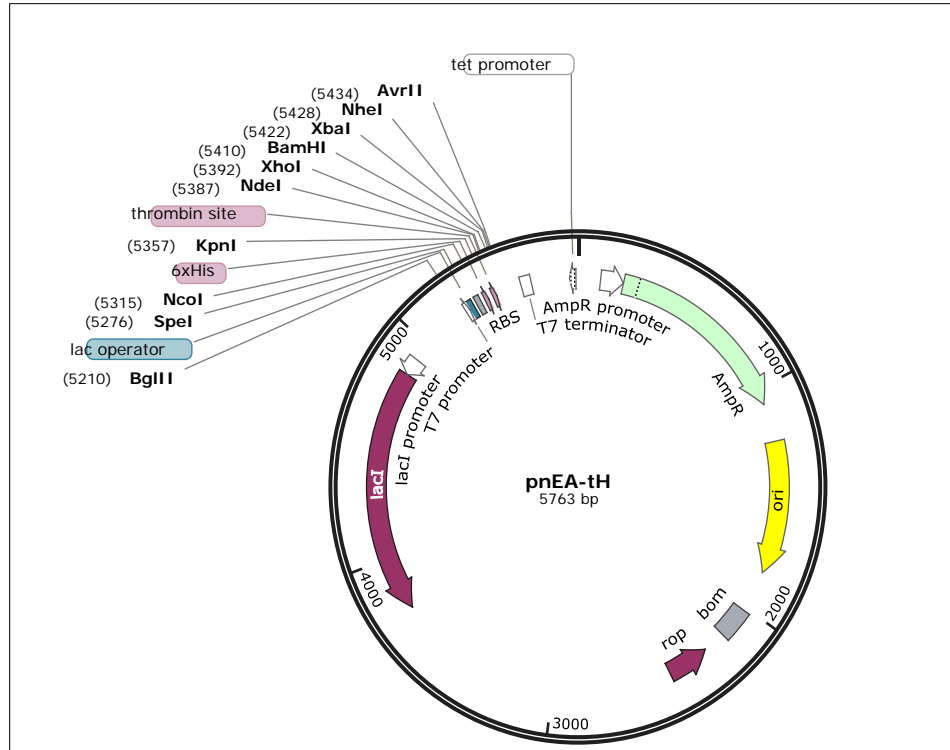
In order to produce proteins recombinantly a gene of interest is inserted into an expression vector which is then transferred into an expression host. Different methods were employed to perform cloning such as traditional cloning strategy (restriction endonuclease dependent ligation), but sequence independent cloning techniques were also used (Gibson method, SLIC). Protocols for cloning are adopted from New England Biolabs and Molecular Cloning: A laboratory Manual by Sambrook *et al.*

### 2.1.1. Expression vectors

An expression vector is a circular DNA that carries the gene of interest (insert) into a host cell for expression. Expression vectors contain all the necessary elements for multiplication and expression inside the host cell. Vectors with different origin of replication and antibiotic resistance can be suitable for co-expression studies and were also used in this thesis. A multi expression tool was developed by our laboratory: the pET-MCN and pET-MCP vectors which were derived from the vectors pET, pCDF (Novagen) and pACYC11b (Fribourg *et al.* 2001) (Table 9). The vectors are available on the platform maintained by our laboratory for the research teams inside and outside of the institute. These vectors contain T7-promoters which are compatible with B121 (DE3) expression bacteria, and also provide a variety of options in multiple cloning sites, antibiotic resistance, N- and C-terminal affinity and fusion tags, protease cleavage sites, etc. Example of a vector from this series is displayed in (Figure 36).

**Table 9: Vectors with different options of tags and antibiotic resistance used in this thesis**

<b>Affinity and solubility tags</b>	<b>Antibiotic options</b>
<ul style="list-style-type: none"><li>• His6 (N-ter)</li><li>• His10 (N-ter)</li><li>• His10-Thioredoxin (N-ter)</li><li>• GST (N-ter)</li><li>• His10 SUMO (N-ter)</li><li>• His10 (C-ter)</li><li>• Native (no tag)</li></ul>	<ul style="list-style-type: none"><li>• Ampicillin (Amp<sup>r</sup>)</li><li>• Chloramphenicol (Chl<sup>r</sup>)</li><li>• Kanamycin (Kan<sup>r</sup>)</li><li>• Spectinomycin (Sp<sup>r</sup>)</li></ul>

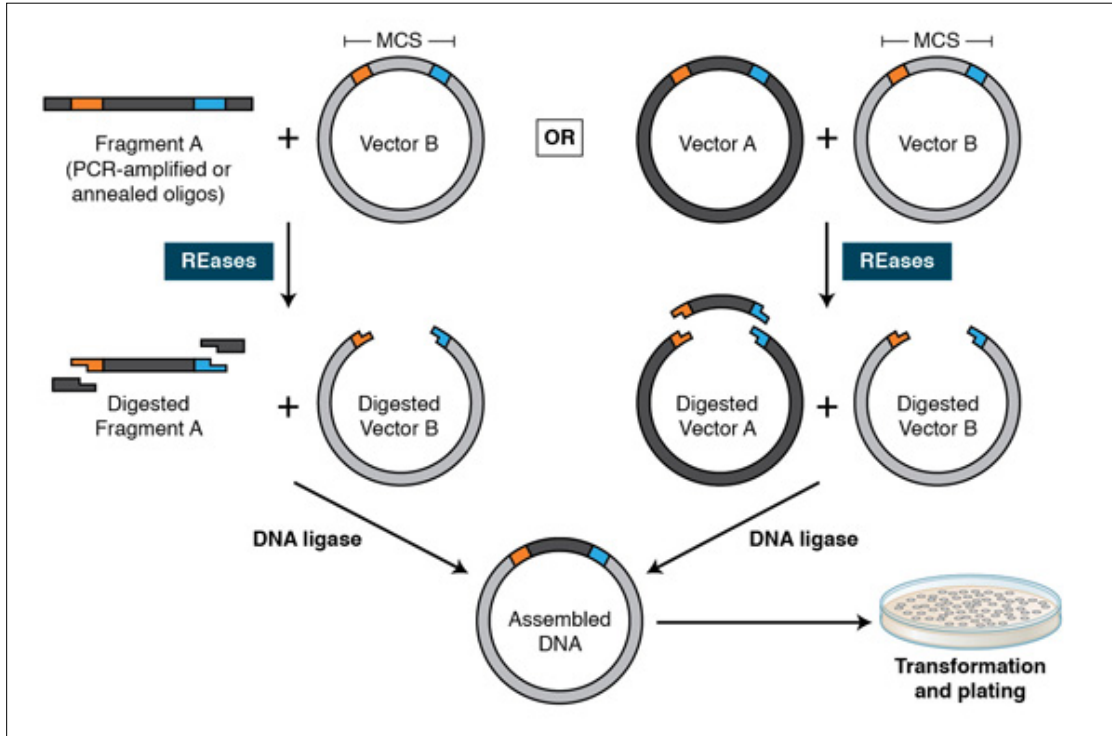


**Figure 36: pNEA-tH vector:**

pNEA-tH vector displaying important features like multiple cloning sites, lacI promoter, origin of replication, and ampicillin resistance gene. pNEA-tH is one of the pET-MCN series vectors developed by our laboratory.

### 2.1.2. Restriction endonuclease dependent cloning

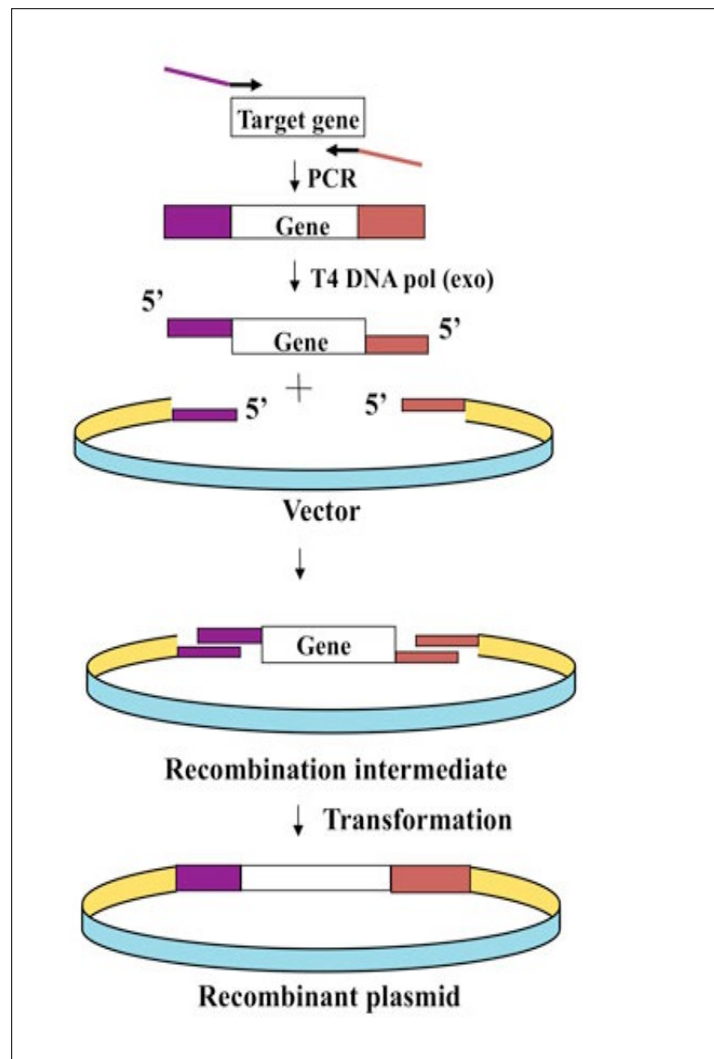
Traditional cloning refers to restriction endonuclease dependent cloning. The gene of interest is PCR amplified with gene specific primers having restriction sites as overhangs at both ends (Figure 37). The amplified product is digested with restriction endonucleases to generate sticky ends. Simultaneously, the expression vector is also prepared by restriction endonuclease treatment to provide complementary sticky ends, followed by alkaline phosphatase treatment which will dephosphorylate the vector to avoid self-ligation. Finally, vector and insert are mixed with T4 DNA ligase to enable ligation, and the mixture is transformed into the E. coli cloning strain DH5 $\alpha$  and plated onto agar medium containing the antibiotic to select for the cells that have incorporated the plasmid. Single colonies are then used to inoculate small cultures. After overnight growth, the cells are used for plasmid preparation which is then sent for sequencing using primers internal to the expression vector that flank the promoter/gene of interest region.



**Figure 37: Steps in Traditional cloning**  
 Picture adapted from *New England Biolabs*

### 2.1.3. Sequence and Ligase Independent Cloning (SLIC)

SLIC does not require restriction enzymes or ligase, instead it depends on the 3' to 5' exonuclease activity of the T4 DNA polymerase (*Figure 38*). The amplified gene contains an overlapping sequence with the vector. When the mixture of vector and insert is mixed with T4 DNA polymerase in the absence of dNTPs, the nucleotides of double stranded DNA are digested by T4 DNA polymerase. The exonuclease activity can be stopped by the addition of dCTP (or any single dNTP), and T4 DNA polymerase acquires polymerase activity and stalled due to lack of all dNTPs. The result of the reaction is vector and insert with compatible sticky ends. The mixture of digested vector and insert is transformed into an *E. coli* cloning strain which will repair the nicks. The procedure for selection and sequencing is the same as described previously.



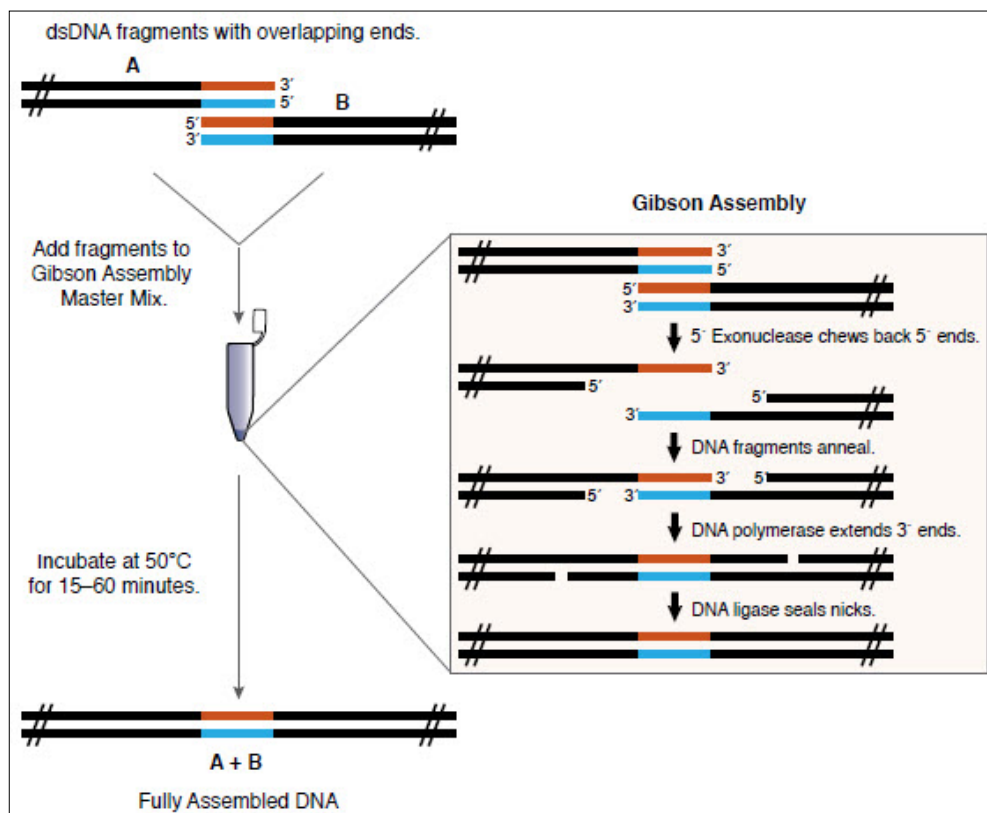
**Figure 38: Sequence and ligase independent cloning**  
 Picture adapted from *Elledge Lab.med.harvard.edu*

#### 2.1.4. Gibson cloning

The Gibson protocol was developed by Daniel Gibson *et al.* and depends on three enzymatic activities (Figure 39). T5 exonuclease creates 3' overhangs which can be annealed at 50°C and simultaneous inactivation of T5 exonuclease happens at high temperature. Further, the polymerase fills the gaps, and the ligase activity enables to generate the recombinant plasmid.



Gibson method is extensively used in this thesis work which has the highest success rate. This technique is highly efficient and has eased enormously cloning.



**Figure 39: Overview of Gibson assembly**  
Picture adapted from *New England Biolabs*

### 2.1.5. Site directed mutagenesis

Two different methods, nested PCR and rolling circle plasmid synthesis were used in this thesis work to obtain mutant clones. In both cases nearly 20 bp long primers were synthesized (from Sigma) where the mutated sequence was kept in the middle, and the  $T_m$  of primers was maintained close to 60°C.

In the first method, 5' and 3' sequences were amplified in two independent reactions. Amplified fragments were purified using spin columns. In the second step purified fragments were used as template and the whole fragment was amplified with N and C terminal primers. Amplified fragments were then used for cloning.

In the second method, the whole plasmid was amplified with the mutated primers. The protocol used for the rolling circle plasmid synthesis was mentioned in the Table 10 and the parameters for thermocycler was mentioned in the Table 11. After PCR amplification, DpnI digestion was performed for 1hr at 37°C according to manufacturers recommended quantities. After DpnI digestion plasmids were purified using spin columns which were further transformed into DH5α competent cells.

**Table 10: Rolling circle plasmid synthesis protocol**

<b>Ingredient</b>	<b>Volume</b>
Template DNA	50ng
2mM dNTP	2 µl
100 µM Forward primer	0.3 µl
100 µM Reverse primer	0.3 µl
GC buffer	4 µl
100% DMSO	0.6 µl
Phusion polymerase (2U/µl)	0.5 µl
H <sub>2</sub> O	to 20 µl

**Table 11: Thermocycler protocol for rolling circle plasmid synthesis**

<b>Temperature (°C)</b>	<b>Time</b>	<b>Number of cycles</b>
98	5 min	1
98	30 sec	30
55	30 sec	
72	30 sec/kb	
72	7 min	1
4	hold	hold

### 2.1.6. Cloning of SMC-HD with linker

In this thesis, all the cloning experiments were based on the above-mentioned methods. Here I have discussed a special case of SMC-HD cloning. The protocol to clone SMC-HD domains (by joining N- and C-terminal domains with a linker) was as follows.

In the first step, N- and C- terminal regions (NTD and CTD respectively) were PCR amplified with four different primers where the middle primers have overlapping sequence by about 20 nucleotides and additionally a linker of thrombin cleavage site was also introduced in this region. The linker extensions contain *NheI* restriction site in overlapping regions. In the second step, a nested PCR (for SMC1-HD) was used where two fragments (NTD and CTD) were amplified with extreme end primers. And then the purified PCR fragment was used for the cloning as one of the above described methods.

In an alternate approach of second step (used for SMC3HD), both PCR products were digested with *NheI* restriction enzyme which was followed by purification using agarose gel. After purification, the sticky end containing products were ligated with the help of T4 DNA ligase. Finally, the ligated product was purified using agarose gel electrophoresis and then the cloning was finished using one of the above-mentioned methods. Thus, the recombinant SMC-HD contains SMC NTD, a linker of thrombin cleavage site and CTD.

## 2.2. Expression methods

For heterologous recombinant expression, several techniques are available such as bacterial expression system, yeast expression system, Baculovirus-insect cell expression systems and eukaryotic expression system. Among these bacterial expression system is the cheapest, simplest, fastest method and generally yields high quantity of the protein of interest. Since the proteins studied in this thesis could be expressed in a bacterial system, only this technique has been used in my thesis work.

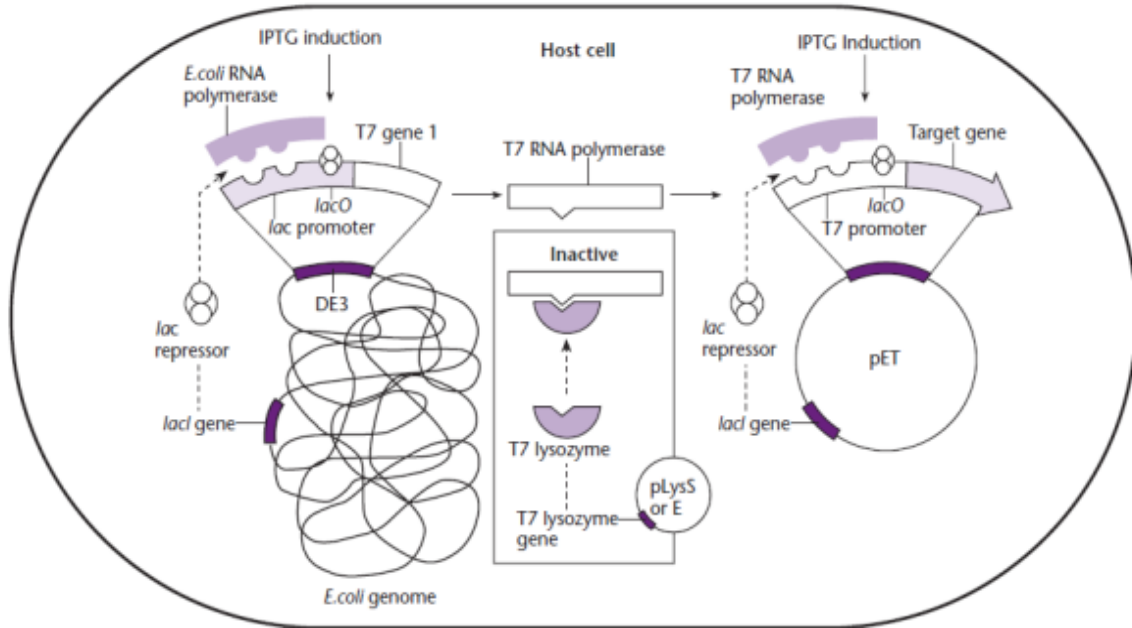
### 2.2.1. Principle of expression in case of the PET system of expression

A genetically modified *E. coli* strain contains the T7 RNA polymerase gene integrated in its chromosome which is under control of a lac promoter. (Studier and Moffatt, 1986; Studier *et al.*, 1990). Bacteriophage T7 RNA polymerase is a highly selective and active enzyme that can elongate five times faster than *E. coli* RNA polymerase. In addition, the promoter sequence of T7 bacteriophage RNA polymerase is rarely present in any other organisms, which makes it selective. When inducer (IPTG) is added to the system, expression of T7 RNA polymerase is triggered. Vectors which carry gene of interest under the control of a T7 promoters are recognized by the RNA polymerase which transcribes the gene of interest (*Figure 40*).

The gene of T7 RNA polymerase is generally integrated in lacUV5 which is activated upon addition of inducer (IPTG). Then the T7 RNA polymerase binds to its selective promoters which provided in the pET expression vectors. As a result, the gene of interest which is cloned downstream to T7 promoter is over expressed. T7 RNA polymerase out competes the cellular protein expression and result in the maximum expression levels.

The strain used in this thesis is B121(DE3), a genetically modified genotype of  $F^-$  *ompT* *hsdSB* ( $r_B^- m_B^-$ ) *gal dcm* (DE3).

- The designation  $F^-$  means that the bacteria does not carry the F plasmid which is required for the conjugation.
- *ompT* stands for the mutation in outer membrane protein protease VII, which reduces the proteolysis of over expressed protein.
- *hsdSB* ( $r_B^- m_B^-$ ) means endogenous restriction enzymes can't digest the exogenous DNA
- *gal* stands for the inability to metabolize galactose as carbon source
- *dcm* is the gene encoding cytosine methylase that methylates second cytosine of CCWGG motif (where W is A or T)
- DE3 means the strain contains  $\lambda$  DE3 lysogen that carries the T7 RNA polymerase gene



**Figure 40: Mechanism of pET expression system.**  
Picture adapted from pET system manual.

### 2.2.2. Expression protocols

A set of general protocols were applied for all the expressions in this thesis which are summarized in (Haffke, Marek *et al.* 2015). First, a mini expression test was performed to check in parallel different expression conditions and purification buffers. Once optimized conditions have been found, large scale expression is set up to purify the protein of interest.

For mini expression tests, in general a specialized media was used which was termed graffinity media which enables to grow cultures at high density, providing enough cells for analysis.

In the case of large scale expressions, beside Graffinity, different media were used: 2XLB (double concentrated Luria Bertani media), Terrific Broth (TB) and auto induction. In all cases, appropriate antibiotics were used. Depending on the expression vector(s) used, the following final concentrations were used: ampicillin 100 µg/ml, kanamycin 50 µg/ml, chloramphenicol 34 µg/ml and spectinomycin 50 µg/ml.

### 2.2.2.1. Mini expression test

Mini expression tests were performed in a high throughput manner to analyze multiple conditions in a parallel manner using 24 well plates. In each condition, a 2ml 2XLB media supplemented with required antibiotics and 0.5% glucose was inoculated with BL21(DE3) bacterial colony which is transformed with the vector(s) encoding the gene(s) of interest. The plates were sealed with a porous membrane and incubated at 37°C in a shaker incubator until the cultures show high density. In a second step, additional 2ml of 2XLB supplemented with required antibiotics, 0.6% lactose, 20 mM HEPES pH 7.0 and 0.5 mM IPTG, was added. Cultures were then grown overnight at 25°C on a shaker incubator. The following day, the cells were harvested by centrifugation and used immediately for purification or stored until next use at -20°C.

### 2.2.2.2. Large scale production

The media used was inoculated with a preculture which was prepared on LB-agar plates of 40 ml in volume from a BL21(DE3) transformed plate. Required antibiotics were supplemented and cultures were grown at 37°C on a shaker incubator until they reached the OD<sub>600</sub> of 1.5. The cultures were then cooled down to 25°C and then expression was induced with IPTG with a final concentration of 0.7 mM. After overnight incubation at 25°C, the cells were harvested by centrifugation and stored at -20°C until next use.

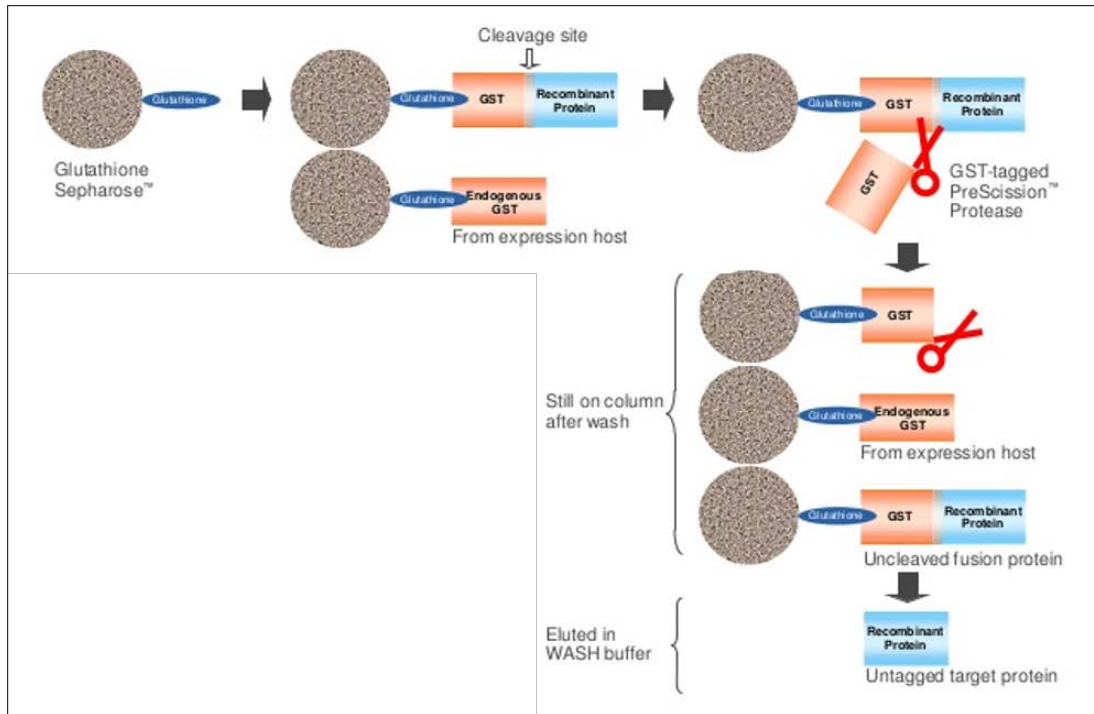
## 2.3. Purification techniques

To be characterized by biophysical and structural means, a protein of interest should be purified to almost homogeneity, i.e. from all other contaminants coming from the expression host. Chromatography techniques are currently widely used for purification. Chromatography was derived from the Greek chroma, which means ‘color’, and graphien, which means ‘to write’, and was first introduced by Mikhail Tswett in 1903, where he purified plant pigments on calcium carbonate column. The basic principle involves a mobile phase in which the solutes are dissolved, which flows over stationary phase and, based upon different properties (e.g. charge, affinity, size, shape, pH etc.) of the solutes, the separation takes place. In this thesis, three main chromatography techniques were employed, affinity, ion exchange and gel filtration chromatographies.

### 2.3.1. Affinity chromatography

Affinity purification is one of the most commonly used chromatographic technique which is often used as a first step purification. Affinity chromatography separates proteins on the basis of an affinity group attached to the protein that interacts with the immobilized matrix (based on agarose or sepharose beads onto which ligands are attached). The interaction between the affinity group and the ligand on the matrix is specific and reversible. A variety of systems like hormone/receptor, antibody/antigen, protein/dye, histidine tag/metal ion, GST tag/glutathione can be used.

In favorable conditions, the target protein which is expressed with an affinity tag interacts with the specific ligand. Unbound material is washed out from the column and the bound protein can then be eluted by applying different conditions such as a non-specific elution by change of pH, ionic strength, polarity, or by a specific elution with a competitive ligand. Two affinity based chromatographic techniques were employed in this thesis, GST (Glutathione S-transferase) tag and metal ion-based affinity purifications. An outline of GST tagged purification is showed in (Figure 41). A GST-tagged protein possess affinity towards glutathione which is immobilized on sepharose (Glutathione sepharose GE®) beads. In metal ion-based affinity chromatography a metal ion ( $\text{Cu}^{+2}$ ,  $\text{Ni}^{+2}$ ,  $\text{Co}^{+2}$ ) is immobilized on a chelator matrix (agarose, sepharose), which shows affinity towards histidine residues (his-tags) in the protein. All the his-tagged affinity purifications were carried out with  $\text{Co}^{+2}$  based Talon® resin from Clontech laboratories Inc. The elution was carried out with reduced glutathione and imidazole, from GST and Talon beads, respectively. In an alternate way, the tag was digested on the column with a sequence specific protease (thrombin, TEV protease, 3C protease) whose recognition sequence is placed between the affinity tag and the protein.



**Figure 41: Affinity chromatography:**

Representation of GST tagged protein purification by affinity chromatography. Picture adapted from GE healthcare.

### 2.3.2. Ion exchange chromatography

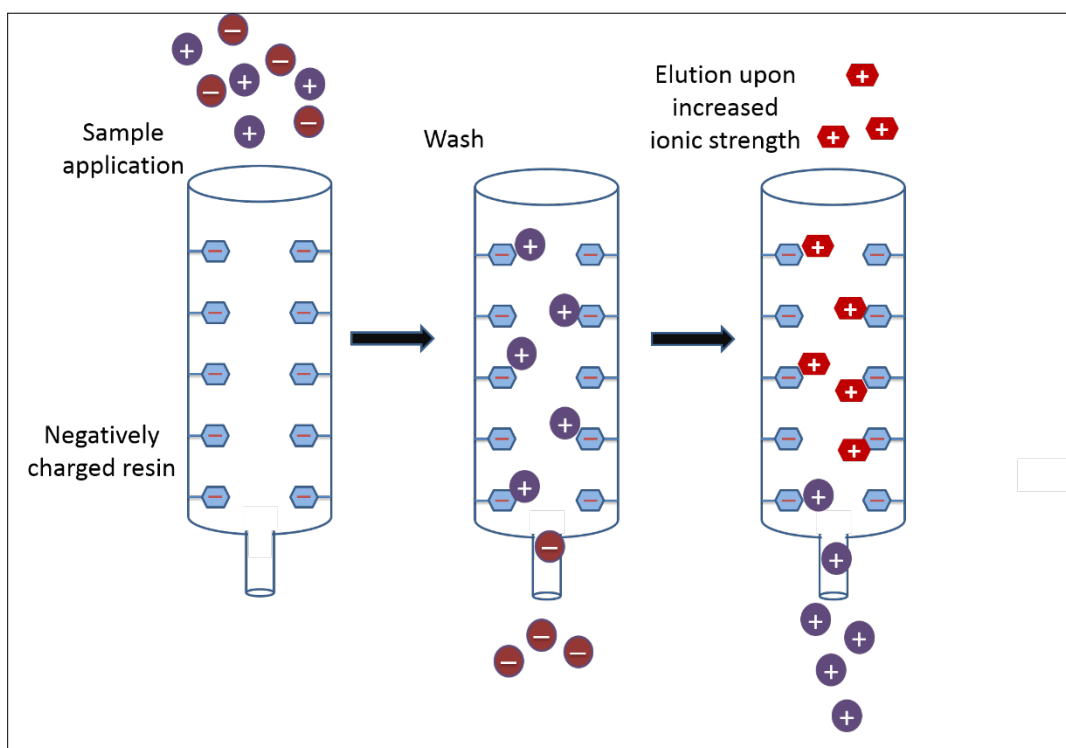
Ion exchange (IEX) chromatography involves the separation of molecules based on their net surface charge. It is a powerful tool capable of separating samples with a single charged amino acid difference. Proteins are amphoteric molecules whose net surface charge changes as the pH of the environment changes. An oppositely charged matrix is used to retain the sample while the non-specific, uncharged or same (charge with the matrix) charged species are washed out of the column. The retained protein is then eluted by changing the ionic strength or the pH. A protein is positively charged when the pH of the environment is below its pI, where the cation exchanger is used. Similarly, a protein is negatively charged when the pH of the environment is above its pI, and an anion exchanger is used.

Further, weak ion exchangers are susceptible to small changes of pH and loses its charge beyond small limits (DEAE, ANX weak anion exchangers and carboxymethyl – weak cation exchanger) while the strong ion exchangers possess a wide range of working pH (sulphopropyl (SP), sulfonic acid-strong cation exchangers and quaternary ammonium (Q)-strong anion



exchanger). Cation-exchanger chromatography is depicted in the (Figure 42). Further, in ion exchange chromatography the sample is being concentrated on the matrix which is eluted in the small fraction volume and higher the charge, more ionic strength is needed to elute the protein.

In this thesis work, Q-sepharose (strong anion exchanger) was used to purify proteins where a simple standard protocol was used. Q-sepharose column was equilibrated with a buffer containing 50 mM Tris pH 8.0, 50 mM NaCl and 2mM DTT. The protein sample was applied and followed by wash with the same equilibration buffer. Finally, the protein was eluted in a gradient of increasing salt concentration from 50 to 1000 mM NaCl in the initial buffer.



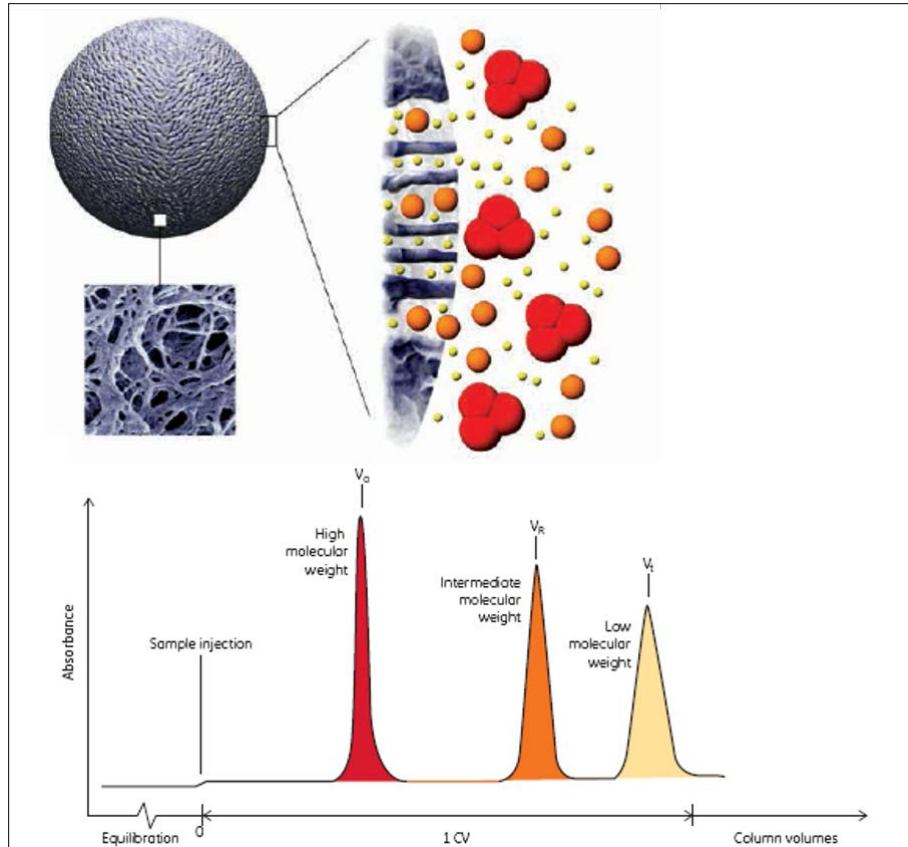
**Figure 42: Ion exchange chromatography:**  
Representation of a cation exchanger

### 2.3.3. Gel filtration chromatography

Gel filtration chromatography or size-exclusion chromatography (SEC) or molecular sieve chromatography is often used as the final step of purification, as it uses the principle of molecular size differences to separate the sample (Figure 43). Unlike IEX or affinity chromatography, gel filtration chromatography doesn't require an interaction of the sample with the matrix. Hence it is recommended that the sample should be concentrated before application. SEC provides several

advantages such as wide range of buffer compatibility, different temperatures, and can also be used in case of globular proteins to determine the exact size of a molecule by deriving a standard gel filtration curve.

An inert porous matrix which lacks reactivity or adsorptive capacity is filled in the columns which acts as a stationary phase. The mobile phase or buffer fills the matrix and the space between the particles, and the elution occurs over the isocratic flow of a single buffer. Molecules with a bigger size than the matrix pore diameter are excluded from the resin (hence the term size exclusion) and elute before the smaller molecules which enter the pores and eventually elute later. The group separation ability of SEC makes it also suitable for rapid buffer exchange since the small salt particles elute at the end of the purification step. Different SEC resins are available with different pore size options for different separation capacity. Sephadex (suitable for desalting), Superdex, Superpose and Sephacryl (dextran + bisacrylamide polymer) are suitable for high resolution separation. Gel filtration was used extensively in my thesis as a final step purification.



**Figure 43: Gel filtration chromatography.**

Top panel electron micrograph of a bead with a graphical illustration of small molecules that enters the pores while large molecules are excluded. Bottom panel is the representation of a typical gel filtration chromatogram with different marks noted.  $V_0$ : Void volume where the large molecules are eluted that doesn't enter the pores,  $V_R$ : retention volume of intermediate molecular size particles and  $V_t$ : total volume where the smallest molecules have eluted. Picture modified from GE Healthcare.

### 2.3.4. Purification protocols

In the first attempt, a general protocol was used to purify the proteins, from which slight modifications (buffer compositions, protease cleavage) were adopted during optimization.

Cells were thawed from  $-20^{\circ}\text{C}$  and resuspended in lysis buffer containing 50 mM Tris pH 8.0, 150 mM NaCl. A 3 Liter culture pellet was dissolved in nearly 70 ml of buffer, and sonication was used to lyse the bacteria, at 40 mAmp, with pulse time 0.5 sec on and 0.5 sec off for 180 secs for three times and mixed after each cycle. Lysate was clarified by centrifugation at  $4^{\circ}\text{C}$ . the supernatant was applied to pre-equilibrated Talon or GST beads and incubated at  $4^{\circ}\text{C}$  for at least

2 hrs. on a roller shaker at slow speed. After incubation resin was washed with lysis buffer a minimum of three times and proceeded for the next step either elution or affinity tag cleavage by protease 3C or thrombin (1mg/ml) overnight. The eluted or cleaved protein was concentrated if necessary in an amicon ultra centrifugal device and applied on 16/60 Superdex 75 or 200 gel filtration column which was pre-equilibrated with gel filtration buffer (50 mM Tris pH 8.0, 150 mM NaCl, 2mM DTT). After each purification step proteins were analyzed using SDS gel electrophoresis.

### 2.3.5. smHDAC8 purification

smHDAC8 is highly sensitive enzyme to small changes in the expression/purification conditions such as histidine tag position, expression conditions, buffer composition, etc. A protocol was optimized prior to my thesis to produce smHDAC8 and which is different from the above-mentioned protocols. This protocol is described in the following book chapter.

#### 2.3.5.1. Large-Scale Overproduction and Purification of Recombinant Histone Deacetylase 8 (HDAC8) from the Human-Pathogenic Flatworm *Schistosoma mansoni*

## Large-Scale Overproduction and Purification of Recombinant Histone Deacetylase 8 (HDAC8) from the Human-Pathogenic Flatworm *Schistosoma mansoni*

Martin Marek, Tajith B. Shaik, Sylvie Duclaud, Raymond J. Pierce, and Christophe Romier

### Abstract

Epigenetic mechanisms underlie the morphological transformations and shifts in virulence of eukaryotic pathogens. The targeting of epigenetics-driven cellular programs thus represents an Achilles' heel of human parasites. Today, zinc-dependent histone deacetylases (HDACs) belong to the most explored epigenetic drug targets in eukaryotic parasites. Here, we describe an optimized protocol for the large-scale overproduction and purification of recombinant smHDAC8, an emerging epigenetic drug target in the multicellular human-pathogenic flatworm *Schistosoma mansoni*. The strategy employs the robustness of recombinant expression in *Escherichia coli* together with initial purification through a poly-histidine affinity tag that can be removed by the thrombin protease. This protocol is divided into two steps: (1) large-scale production of smHDAC8 in *E. coli*, and (2) purification of the target smHDAC8 protein through multiple purification steps.

**Key words** Histone deacetylase, Enzyme, Recombinant expression, Purification, *Schistosoma*

---

### 1 Introduction

Schistosomiasis, or bilharzia, is a parasitic disease caused by trematode flatworms of the genus *Schistosoma* (*S. mansoni*, *S. japonicum*, and *S. haematobium* are the main species of medical relevance) [1, 2]. According to the World Health Organization (WHO) statistics, schistosomes infect around 230 million people worldwide and cause at least 300,000 deaths yearly, with about 800 million people further at risk of infection [3]. The control of schistosomiasis is dependent on mass treatment with a single drug, praziquantel [4], and the consequent risk of the appearance of resistant strains raises the spectrum of widespread drug resistance. Ultimately, praziquantel-resistant schistosome strains have already been reported [5, 6], and these findings rendered the development of new anti-schistosomal drugs a strategic priority.

Schistosomes, like many eukaryotic pathogens, typically display various morphologically distinct stages during their complex life cycles. Epigenetic mechanisms fundamentally underlie the pathogens' morphological transformations, and the targeting of epigenetics-driven cellular programs therefore represents an Achilles' heel of human parasites. Today, zinc-dependent histone deacetylases (HDACs) belong to the most explored epigenetic targets, notably for anticancer therapies [7–9]. This fact significantly speeds up the search for new antiparasitic agents since drugs validated against cancers can be effectively tailored into antiparasitic therapeutics. Nevertheless, one of the key bottlenecks in antiparasitic drug discovery is recombinant production in large quantities of parasites' epigenetic targets for structure-based and pharmacological studies.

In this chapter, we describe an optimized protocol for the large-scale overproduction and purification of *Schistosoma mansoni* HDAC8 (smHDAC8), an emerging epigenetic drug target in this pathogenic organism [10–12]. Our protocol employs the robustness of the *Escherichia coli* expression system that enables cost-effective production, as well as scale up to industrial-scale fermentation production. Specifically, our protocol details the various parameters required during growth, induction and purification that enable the production of soluble smHDAC8 in milligram quantities, since standard parameters only lead to the production of this enzyme in insoluble inclusion bodies.

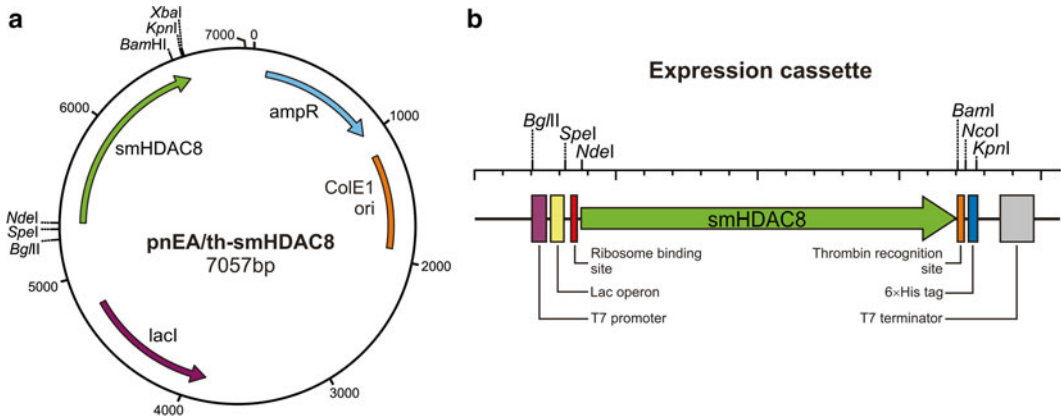
---

## 2 Materials

Prepare all solutions and media using ultrapure deionized water and analytical grade chemicals.

### 2.1 Cell Transformation

1. Expression plasmid vector pNEA/tH-smHDAC8 [10] (Fig. 1), where the full-length *smHDAC8* gene is inserted between the *NdeI* and *BamHI* restriction sites of the pNEA-tH expression vector [13] and is in frame with a sequence encoding a C-terminal thrombin cleavage site followed by a poly-histidine affinity purification tag (*see Note 1*).
2. Chemically competent cells of *Escherichia coli* BL21(DE3) strain.
3. Ice bucket.
4. 42 °C water bath.
5. 2 × Luria broth (2 × LB) medium: for 1 L, weigh 20 g tryptone, 10 g yeast extract, and 20 g NaCl. Add distilled water to reach 1 L. Sterilize by autoclaving.
6. 37 °C shaking incubator.
7. LB agar.



**Fig. 1** Schematic representation of the plasmid used for the expression of smHDAC8 in *E. coli*. **(a)** Map of the pNEA/th-smHDAC8 plasmid. The *smHDAC8* gene is inserted between *NdeI* and *BamHI* restriction sites. Selection in *E. coli* is performed by the beta-lactamase ampicillin resistance gene (*ampR*). Origin of replication sequence (*ColE1 ori*) is available for maintenance in *E. coli* cells, and the *lacI* gene is present for expression of the Lac repressor protein. **(b)** Details of the expression cassette. The *smHDAC8* gene is controlled by the T7 promoter and T7 terminator. The *smHDAC8* gene is cloned in frame with a sequence coding for a C-terminal thrombin cleavage site followed by a poly-histidine tag

8. Ampicillin 1000× stock solution; 100 mg/mL (in H<sub>2</sub>O).
9. Standard petri dishes (diameter of 9 cm) and large petri dishes (diameter of 15 cm).
10. 37 °C incubator.

## 2.2 Cell Cultures

1. Salt medium (0.17 M KH<sub>2</sub>PO<sub>4</sub>, 0.72 M K<sub>2</sub>HPO<sub>4</sub>): for 100 mL, dissolve 2.31 g of KH<sub>2</sub>PO<sub>4</sub> and 12.54 g of K<sub>2</sub>HPO<sub>4</sub> in 90 mL of distilled water. Stir till the salts have dissolved, then adjust the volume of the solution to 100 mL with distilled water and sterilize by filtering.
2. Terrific-Broth (TB) rich medium: for 1 L, add 12 g tryptone, 24 g yeast extract and 4 mL glycerol to 900 mL distilled water and sterilize by autoclaving. Prior to use, add 100 mL of salt medium.
3. Ampicillin 1000× stock solution; 100 mg/mL (in H<sub>2</sub>O).
4. 5-L flasks.
5. Thermostatic shaker Certomat BS-1.
6. Single-beam spectrophotometer model BioPhotometer Plus.
7. Isopropyl β-d-1-thiogalactopyranoside (IPTG; 1 M stock solution).
8. Zinc chloride (100 mM stock solution).
9. 1-L centrifugal bottles for Type JS4.2 rotor.
10. Centrifuge (J6MI floor model centrifuge equipped with Type JS4.2 rotor).

11. Resuspension buffer (10 mM Tris-HCl pH=8.0; 50 mM KCl). For 1 L, dissolve 1.21 g of Trizma base and 3.73 g of KCl in 900 mL distilled water. After the buffer and salt have dissolved, adjust pH with HCl to reach a pH value of 8.0. Finally, adjust volume of the solution to 1 L with distilled water and sterilize by filtering.

### **2.3 Purification Steps**

1. Lysis buffer (10 mM Tris-HCl pH=8.0; 50 mM KCl). Same as the resuspension buffer used to resuspend the cell pellets at the end of the production step.
2. High-pressure homogenizer Microfluidizer Processor M-110EH.
3. Ultracentrifuge Beckman Coulter Optima L90K equipped with Type Ti-45 fixed-angle rotor.
4. Thick-wall ultracentrifuge tubes for Type Ti-45 fixed-angle rotor.
5. Talon Metal affinity resin.
6. 10-mL glass column with adaptor for use with a peristaltic pump.
7. Peristaltic pump model EP-1 Econo Pump.
8. Thrombin stock solution (1 U/ $\mu$ l in 25 mM Tris pH=8.0 and 50% glycerol; kept at  $-20^{\circ}\text{C}$ ).
9. Rolling mixer model RM-5.
10. Ion-exchange chromatography buffers. Low-salt buffer (10 mM Tris-HCl at pH=8.0; 50 mM KCl) and high-salt buffer (10 mM Tris-HCl at pH=8.0 and 1 M KCl). The low-salt buffer corresponds to the lysis buffer. The high-salt buffer is prepared as described for the lysis buffer, with the exception that 74.55 g of KCl are used to reach 1 M final concentration.
11. Polypropylene gravity-flow Econo-Pac column.
12. Bio-Rad Protein Assay reagent.
13. FPLC protein purification system.
14. 1-mL HiTrap Q FF column.
15. Gel filtration buffer: 10 mM Tris-HCl pH=8.0; 50 mM KCl; 2 mM DTT. The buffer is prepared as the lysis buffer but is supplemented with 0.31 g of DTT prior to dissolution of the chemicals in water.
16. Column for gel filtration (16/60 Superdex 200).
17. Amicon Ultra centrifugal filter units with 30-kDa cutoff.
18. Apparatus for SDS-PAGE model Mini-PROTEAN Tetra System.
19. Single-beam spectrophotometer.



### 3 Methods

#### 3.1 Cell Transformation

1. Mix 50  $\mu$ L chemically competent BL21(DE3) *E. coli* cells with 50–100 ng of p<sub>n</sub>EA/tH-smHDAC8 plasmid expression vector and incubate the mix on ice for 20 min. Heat-shock the mix at 42 °C for 45 s and then incubate on ice for 2 min. Add 300  $\mu$ L of 2 $\times$ LB medium and incubate in a 37 °C shaker for 1 h. After incubation, spread 300  $\mu$ L of bacterial suspension on a regular agar plate (9 cm) containing ampicillin (100  $\mu$ g/mL). Incubate the plate overnight at 37 °C.
2. Inoculate large agar plates (15 cm) containing ampicillin (100  $\mu$ g/mL) with several ampicillin-resistant colonies from the transformation in **step 1** (*see Note 2*). To do so, collect several colonies from the small agar plate with a kinked plastic tip and spread this inoculum homogeneously on the large agar plates. Incubate the plates overnight at 37 °C.
3. The next morning, resuspend the film-forming *E. coli* cells from the large agar plates. For resuspension, add approximately 10 mL of fresh sterile 2 $\times$ LB medium per large agar plate (15 cm). Scratch the surface of the plate with a kinked Pasteur glass pipette to release the cells into the 2 $\times$ LB medium. Once resuspension is done, transfer the liquid fraction into a 50-mL Falcon tube. Measure the OD<sub>600</sub> value of this harvested inoculum.

#### 3.2 Large-Scale Cultures

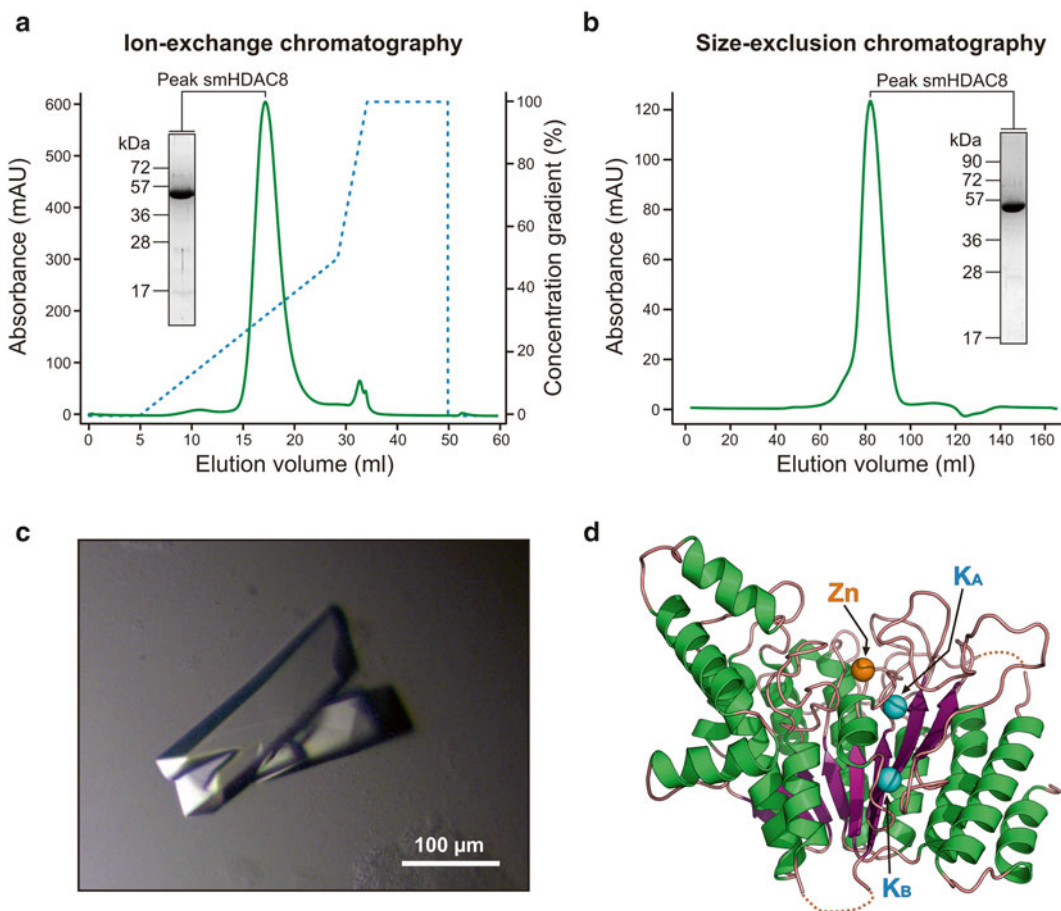
1. Transfer 1 L of Terrific-Broth (TB) medium in a 5-L flat-bottomed Erlenmeyer flask. Add ampicillin to reach a final concentration of 100  $\mu$ g/mL. Use the harvested inoculum to inoculate the culture to start with an OD<sub>600</sub> of approximately 0.2.
2. Grow the cells at 37 °C with shaking to high-density culture (OD<sub>600</sub> approximately 4.0–6.0).
3. Add IPTG (0.5 mM final concentration) and zinc chloride (100  $\mu$ M final concentration) and continue to incubate the cultures at temperature 37 °C.
4. After 1 h, recover the cells by centrifuging the cultures for 25 min at 4000 $\times g$  at 4 °C (*see Note 3*) using 1 L centrifugation jars.
5. Resuspend the cell pellets in an ice-cold resuspension buffer (10 mM Tris-HCl pH=8.0; 50 mM KCl). Use approximately 15 mL of this buffer to resuspended the cell pellet from 1 L culture. The cell suspension can be directly used for purification procedure or stored at -80 °C until further use (*see Note 4*).

#### 3.3 Purification Procedure

1. The following protocol is provided considering the use of cell pellets from 3 L of cultures. First, if required, thaw the resuspended cell pellets. Adjust the volume of the cell resuspension to 40 mL

per liter of culture (i.e., final volume of 120 mL for 3 L of culture) using the lysis buffer (identical to the resuspension buffer).

2. Lyse the cell suspension using a Microfluidizer Processor at high pressure (18,000 psi) using a single round of lysis (*see Note 5*). After the lysis, centrifuge the disrupted cell suspension at  $210,000 \times g$  for 1 h and collect the supernatant in a ice-cold bottle.
3. Apply the supernatant to a column with 2 mL of Talon Metal affinity resin pre-equilibrated in lysis buffer. Briefly, connect the column with pre-equilibrated Talon resin to a peristaltic pump and pump the supernatant from **step 1** through the column with Talon resin at a flow rate of 4.0–5.0 mL/min. Every 30 min, disconnect the column from the peristaltic pump and mix the resin to release the excess pressure. After the loading, wash the column extensively with approximately 100 mL of the lysis buffer to remove nonspecifically bound proteins (*see Note 6*).
4. Release the smHDAC8 enzyme from the Talon resin by thrombin treatment. Briefly, resuspend the Talon resin with bound smHDAC8-His fusion protein with the lysis buffer and transfer it to a new sterile 15-mL Falcon tube. The volume of the resin suspension should be approximately 5 mL. Add 60  $\mu$ L of thrombin (1U/ $\mu$ L) and place the tube on a rolling mixer overnight at 4 °C (*see Note 7*).
5. Next morning, separate the released smHDAC8 protein from the Talon resin particles by applying the resin suspension onto an Econo-Pac column and collect the unbound flow-through fraction into a fresh sterile 15-mL Falcon tube. Wash the resin with additional 3 mL of the lysis buffer to harvest all thrombin-released smHDAC8 enzyme. Check the presence and concentration of smHDAC8 in the flow-through fraction by the Bio-Rad Protein Assay.
6. Load the flow-through containing smHDAC8 enzyme from the **step 4** onto a 1-mL HiTrap Q FF column pre-equilibrated with low-salt ion-exchange chromatography buffer (10 mM Tris-HCl pH=8.0; 50 mM KCl). Elute the bound protein with a gradient of KCl (50 mM to 1 M KCl): *see Fig. 2a* for a typical ion-exchange purification of smHDAC8. Identify fractions containing the smHDAC8 protein by SDS-PAGE.
7. Pool the peak fractions from the ion-exchange chromatography from **step 5** and load this sample onto a gel filtration column (16/60 Superdex 200) equilibrated with gel filtration buffer (*see Note 8*). Identify fractions containing the target protein by SDS-PAGE. *See Fig. 2b* for a typical gel filtration purification of smHDAC8.
8. Pool the peak fractions from gel filtration from **step 6**, and concentrate the smHDAC8 protein with an Amicon Ultra



**Fig. 2** Purification and crystallization of smHDAC8. **(a)** Chromatogram of ion-exchange purification of smHDAC8. The gradient used for this purification step is displayed. **(b)** Chromatogram of gel filtration purification of smHDAC8. **(c)** Picture of an smHDAC8 crystal. **(d)** Atomic structure of smHDAC8 represented as ribbons. *Orange sphere* catalytic zinc ion, *blue spheres* potassium ions ( $K_A$  and  $K_B$ )

centrifugal filter unit to reach a final concentration of 2.5 mg/mL (*see Note 9*). Check purity of the purified smHDAC8 enzyme by SDS-PAGE and determine protein concentration by the Bio-Rad Protein Assay reagent (*see Note 10*).

- Flash-freeze the final product with liquid nitrogen and store at  $-80\text{ }^\circ\text{C}$  (*see Note 11*).

## 4 Notes

- Initial affinity purification of smHDAC8 is suboptimal in presence of an N-terminal poly-histidine tag. Therefore, a C-terminal poly-histidine tag is used. The presence of a *Bam*HI cloning site and a sequence encoding a thrombin protease cleavage site

before the sequence encoding the poly-histidine tag leaves several residues at the C-terminus of smHDAC8 after thrombin cleavage of the poly-histidine tag: GSLVPR. These residues do not affect the enzyme's activity [10].

2. The expression yields for a protein are often better when using as starter for large cultures colonies that have grown on petri dishes rather than a liquid preculture. This is particularly true when using ampicillin resistance. Since smHDAC8 is not produced in large quantities in *E. coli*, the use of colonies grown on petri dishes as starters is preferred. To have sufficient colonies for inoculating the large cultures, streaking of the initial transformed colonies on larger petri dishes is carried out. In general, one large petri dish is sufficient to inoculate one liter of large liquid culture.
3. Do not exceed 1 h incubation in the presence of IPTG: when too much smHDAC8 enzyme is produced in a cell, it goes into insoluble inclusion bodies. This is essential since during our pilot experiments longer incubation times led to an almost complete loss of soluble protein. Reducing the expression time decreased the quantity of smHDAC8 proteins produced per cell and the formation of inclusion bodies. Use of Terrific Broth medium with induction at high density compensates for the small quantity of smHDAC8 present in each cell.
4. The cell paste is immediately resuspended in a buffer that corresponds to the lysis buffer. This avoids resuspension of the frozen cell paste in the lysis buffer at the beginning of the purification procedure. Indeed, this resuspension appears more deleterious to the solubility of smHDAC8 than when the cells are resuspended in the lysis buffer prior to freezing. A possible explanation for this behavior is that thawing the already resuspended cells is less stringent. Importantly, we and others [14] noted that zinc-dependent HDACs lose their enzymatic activities very fast when inappropriately handled. We therefore recommend the resuspension of the cell pellet immediately after centrifugation, and storing the cell suspension at  $-80\text{ }^{\circ}\text{C}$  promptly to minimize loss of enzyme activity. Note also that long storage of the cell pellets at  $-80\text{ }^{\circ}\text{C}$  negatively impacts the quality of the smHDAC8 protein: old cell pellets lead to purified protein that appears well-behaved by SDS-PAGE analysis, but that has reduced enzymatic activity and poorer behavior during crystallization.
5. The use of a microfluidizer for cell lysis is essential for keeping smHDAC8 soluble during lysis. Use of sonication leads to the loss of half of the protein yield compared to the use of the microfluidizer.
6. The binding buffer (10 mM Tris-HCl pH=8.0; 50 mM KCl) may be enriched with 5 mM imidazole to increase the stringency of washing and to eliminate protein contaminants. This

procedure causes however some loss of smHDAC8 due to partial elution from the affinity resin.

7. In this step, thrombin specifically cuts the fusion protein at the thrombin recognition site located between smHDAC8 and the poly-histidine tag. This cleavage results in the release of smHDAC8 enzyme from the affinity resin, whilst the poly-histidine tag and the remaining nonspecifically bound contaminants stay bound to the Talon resin. Using this procedure a good level of purity of the smHDAC8 enzyme is already obtained at the first purification step.
8. We noted that the presence of DTT in the gel filtration buffer may cause a problem (decreased enzymatic activity) in some high-throughput deacetylase screening assays. To avoid this problem, it is possible to replace DTT (2 mM) with TCEP (1 mM) in the gel filtration buffer.
9. We usually concentrate the smHDAC8 protein to 2.5 mg/mL with no observable protein precipitation/aggregation. This purification procedure yields an smHDAC8 enzyme exhibiting deacetylase activity similar, if not identical, to its human counterpart, human HDAC8 [10]. In addition, crystallization of the smHDAC8 at this concentration yielded diffraction-quality crystals that enabled structure determination of this enzyme (Fig. 2c, d) [10].
10. Our protocol typically yields 1.0–1.5 mg of highly pure smHDAC8 per 1 L of bacterial culture.
11. The purified smHDAC8 enzyme shows rather quickly a decrease of its activity and of its propensity to crystallize when kept for 3–4 days at 4 °C. Storage at –80 °C slows down but does not stop this process. Therefore, it is advisable to use the enzyme relatively rapidly even when stored at –80 °C.

---

## Acknowledgements

This work and the authors of this manuscript have been supported by funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreements nos. 241865 (SEtTReND) and 602080 (A-ParaDDisE). The authors are supported by institutional funds from the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Université de Strasbourg and the Université de Lille 2, the French Infrastructure for Integrated Structural Biology (FRISBI; ANR-10-INSB-05-01), and by Instruct as part of the European Strategy Forum on Research Infrastructures (ESFRI).

## References

1. Brown M (2011) Schistosomiasis. *Clin Med* 11(5):479–482
2. Ross A, Bartley P, Sleight A et al (2002) Schistosomiasis. *N Engl J Med* 346(16):1212–1220
3. Gray D, Ross A, Li Y, Mcmanus D (2011) Diagnosis and management of schistosomiasis. *BMJ* 342:2651
4. Dömling A, Khoury K (2010) Praziquantel and schistosomiasis. *ChemMedChem* 5(9):1420–1434
5. Doenhoff M, Cioli D, Utzinger J (2008) Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Curr Opin Infect Dis* 21:659–667
6. Doenhoff M, Kusel J, Coles G, Cioli D (2002) Resistance of *Schistosoma mansoni* to praziquantel: is there a problem? *Trans R Soc Trop Med Hyg* 96(5):465–469
7. Li Z, Zhu W (2014) Targeting histone deacetylases for cancer therapy: from molecular mechanisms to clinical implications. *Int J Biol Sci* 10(7):757–770
8. Campbell R, Tummino P (2014) Cancer epigenetics drug discovery and development: the challenge of hitting the mark. *J Clin Investig* 124(1):64–69
9. West A, Johnstone R (2014) New and emerging HDAC inhibitors for cancer treatment. *J Clin Investig* 124(1):30–39
10. Marek M, Kannan S, Hauser A et al (2013) Structural basis for the inhibition of histone deacetylase 8 (HDAC8), a key epigenetic player in the blood fluke *Schistosoma mansoni*. *PLoS Pathog* 9(9), e100364
11. Stolfá D, Marek M, Lancelot J et al (2014) Molecular basis for the antiparasitic activity of a mercaptoacetamide derivative that inhibits histone deacetylase 8 (HDAC8) from the human pathogen *Schistosoma mansoni*. *J Mol Biol* 426(20):3442–3453
12. Kannan S, Melesina J, Hauser A et al (2014) Discovery of inhibitors of *Schistosoma mansoni* HDAC8 by combining homology modeling, virtual screening, and in vitro validation. *J Chem Inf Model* 54(10):3005–3019
13. Diebold M-L, Fribourg S, Koch M, Metzger T, Romier C (2011) Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. *J Struct Biol* 175(2):178–188
14. Olson D, Udeshi N, Wolfson N et al (2014) An unbiased approach to identify endogenous substrates of “histone” deacetylase 8. *ACS Chem Biol* 9(10):2210–2216

### 2.3.5.2. Modified protocol for smHDAC8

The above-mentioned protocol for smHDAC8 production was very tedious and yielded a low amount of protein (1 mg per liter of culture), which is somewhat incompatible with repetitive crystallization experiments. During my thesis, I have optimized the production of smHDAC8, to meet the high demand of crystallization experiments and to supply our collaborators for *in vitro* activity tests. Two constructs were prepared one with C-terminal his tag and the second with N-terminal histidine-thioredoxin fusion tag, which were used for mini expression tests to compare the expression level with different expression conditions.

The construct of pnEA/tH vector which has smHDAC8 coding sequence followed by a thrombin cleavage site and C-terminal histidine tag was transformed in to BI21(DE3) bacteria and used for large scale expression. Three liters of 2XLB was inoculated with preculture, with appropriate antibiotic and expressed at 37°C until it reached the OD above 1.5. The culture was induced with 0.7 mM IPTG and over expressed at 37°C overnight. After overnight expression, cells were harvested and frozen at -20°C until next use.

For crystallization experiments a 3 liter culture pellet was used to purify. Cells were resuspended in lysis buffer containing 50 mM Tris, pH 8.0 and 150mM NaCl. Sonication was used to break the cells, at 40 mAMP and 0.5 sec on and 0.5 sec off for 90 seconds. After mixing the bacterial suspension, sonication was repeated three times with same setup. Supernatant was separated by centrifugation at 4°C, 21,000g and loaded on Talon® affinity beads which were pre-equilibrated with lysis buffer. After two hours of incubation on a roller shaker at 4°C, beads were washed with lysis buffer thoroughly. At this stage, an additional wash was conducted with a buffer containing 10 mM Tris, pH 8.0 and 50mM KCl, and then beads were collected in a falcon tube and thrombin was added to cleave the tag. After overnight incubation on a roller shaker at 4°C, flow through was collected from the beads and loaded on to gel filtration column Superdex S200, which was pre-equilibrated with 10mM Tris pH 8.0, 50 mM KCl and 2 mM DTT. After SDS gel analysis, peak fractions were pooled together and concentrated in an amicon ultra centrifugal device for further use of crystallization or *in vitro* testing.

## 2.4. Protein characterization

After protein purification, a number of biophysical methods were employed to check the quality of the samples. In the following sections principles of these techniques are discussed.

### 2.4.1. SDS-PAGE gel electrophoresis

SDS polyacrylamide gel electrophoresis serves as a size selective sieve. In the presence of catalyst ammonium persulfate (APS), acrylamide polymerizes in a radical process which are further cross linked in presence of N,N'-methylenebisacrylamide. TEMED (N,N,N',N'-Tetramethylethylene-1,2-diamine) splits APS to form free radicals which is then initiates the polymerization of acrylamide (Figure 44). The electrophoretic mobility of proteins solely depends on the size in denatured conditions when sodium dodecyl sulfate (SDS) is used as it masks the charge on the protein surface. The procedure was obtained from Molecular Cloning: A Laboratory Manual, Sambrook *et al.* to run a SDS gel electrophoresis.

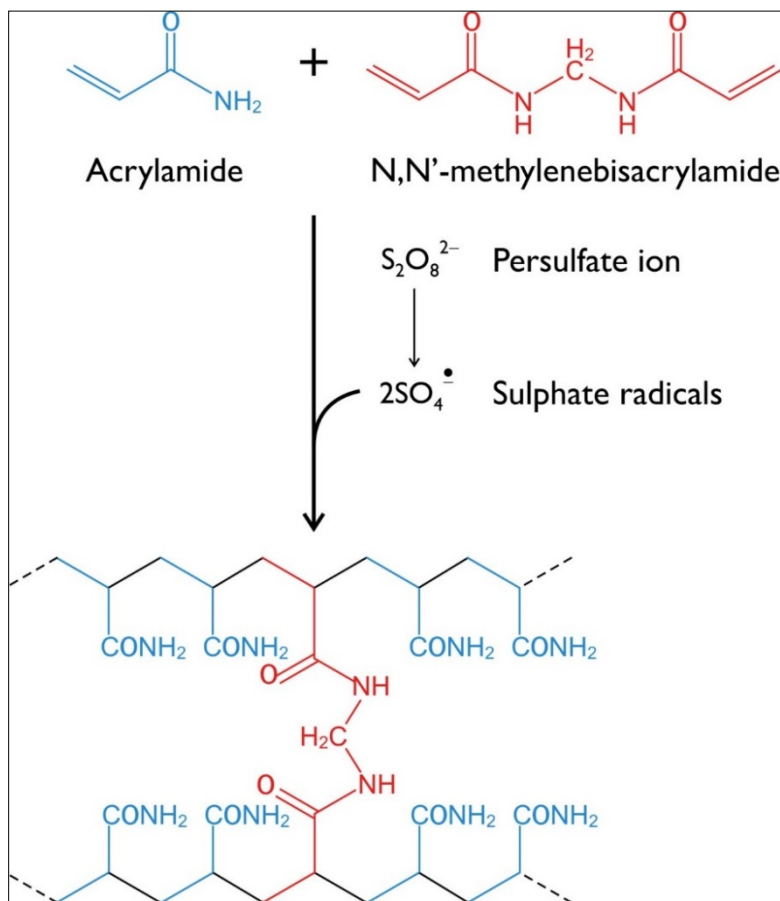


Figure 44: Polymerization of acrylamide



The composition of SDS-PAGE gels and staining solutions were mentioned in the Table 12 and Table 13 respectively. After electrophoresis gels were washed with solution 1 for five minutes which was then followed by solution 2 (Table 13) for staining.

**Table 12: SDS-PAGE gel composition**

Resolving gel composition (50 ml)	Acrylamide percentage				Stacking gel composition (50 ml)	Volume (ml)
	10%	12%	15%	20%		
	Volume (ml)					
40% Acrylamide/bis acrylamide (29:1)	12.6	15.2	18.73	25	40% Acrylamide/bis acrylamide (29:1)	6.25
3M Tris HCl pH 8.8	12.5				1M Tris HCl pH 6.8	6.25
10% SDS	0.5				10% SDS	0.5
10% APS	0.5				10% APS	0.5
TEMED	0.01				TEMED	0.01
Water	23.8	21.26	17.76	11.5	Water	36.5

**Table 13: SDS-PAGE staining solution**

	Solution 1	Solution 2*
Ethanol	500 ml	50 ml
Acetic acid	100 ml	75 ml
Water	400 ml	875 ml
*1 ml of 0.25% Coomassie brilliant blue in ethanol is added to the 50ml of solution 2		

## 2.4.2. Dynamic Light Scattering

Dynamic Light Scattering (DLS) is a widely used technique to determine the size and the size distribution profiles in solution or suspension (Figure 45). In structural biology DLS is very useful to determine the quality of a protein prior to crystallization experiments, to check aggregation, stability, purity and to some extent the molecular weight of the protein studied. The laser beam in DLS instrument takes an advantage of the Brownian motion exhibited by the molecules in the solution. The scattered light from the moving particles reach the detector and contribute to the scattering signal. Moreover, the changing phase of a molecule may add a constructive signal or a destructive signal which is directly related to the diffusion coefficient. Small molecules possess high diffusion coefficient and vice versa. The detectors that measure fluctuations will convert it to a measure of size which is known as hydrodynamic radius. The instrument DynaPro NanoStare (Wyatt technology) (LASER wavelength 830 nm) was used for all the measurements.

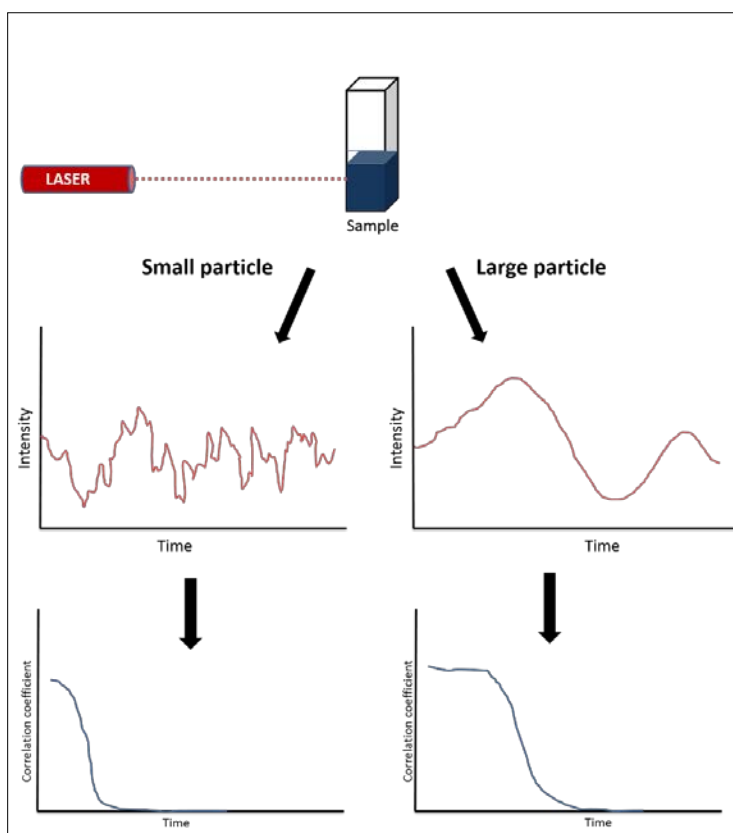


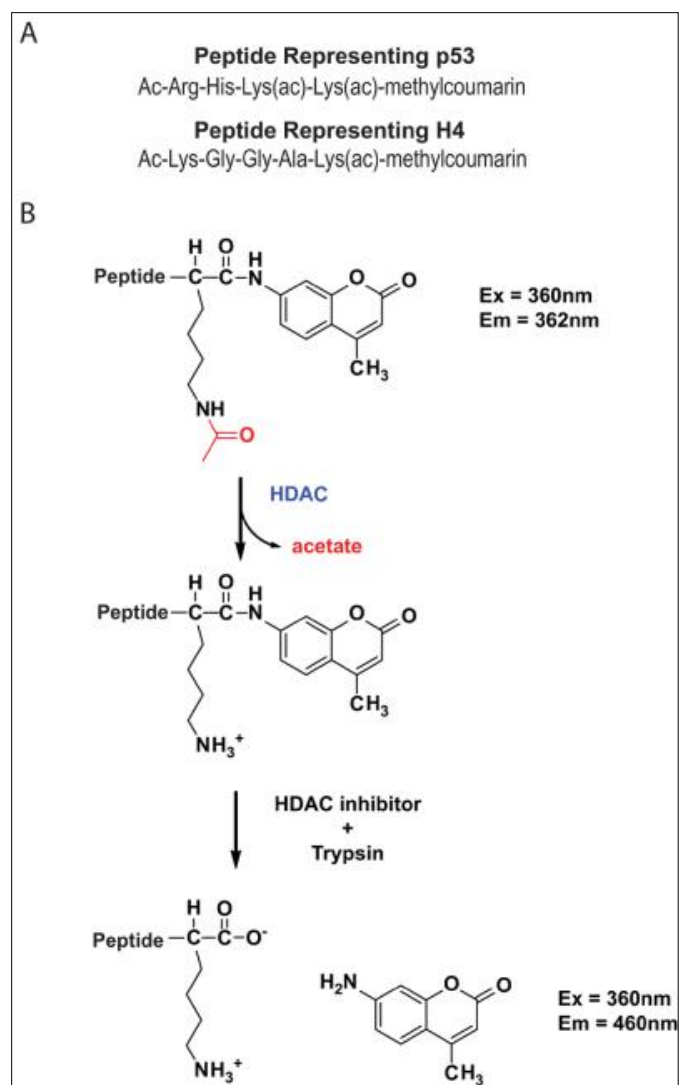
Figure 45:DLS outlook

### 2.4.3. Differential Scanning Fluorimetry

nanoDSF is a differential scanning Fluorimetry which measures tryptophan fluorescence and correlates with melting curves of the protein. Tryptophan has a fluorescence emission at wavelengths 330 nm in a non-polar environment and 350 nm in a polar environment. With an increase in temperature the  $T_m$  or melting temperature (the temperature at which half of the protein is unfolded) can be calculated by measuring the fluorescence of tryptophan at single wavelength or by the ratio of 330/350 nm. A single wavelength measurement sometimes may not produce a well-defined transition whereas the ratio of 330/350 nm can do. The  $T_m$  of the proteins studied was measured by using a Prometheus NT.48 in different buffer conditions or in presence of inhibitors to compare the differences in stability (between different proteins or mutants) or stabilization (e.g. by small molecules).

### 2.4.4. *In vitro* HDAC assay

All *in vitro* HDAC assays were done by our collaborators using a Fluorimetry drug discovery kit (commercially available as Fluor de Lys(R)- HDAC8, BML-KI178 from Enzo Life sciences). The principle of this assay is a peptide substrate (derived from p53 and H4) containing methylcoumarin which is attached to the terminal acetylated lysine residue (Figure 46). In the first step, the substrate peptide is deacetylated by the HDAC, then in the second step trypsin cleaves the fluorophore off the lysine residue, which increases the fluorescence at 460nm.

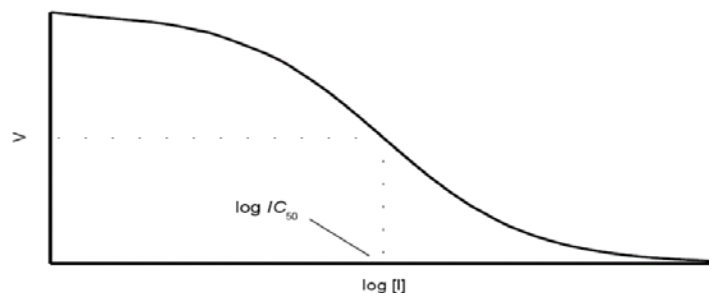


**Figure 46: Fluor de Lys assay: A:**

Substrate peptides derived from p53 and H4 that are conjugated with methylcoumarin fluorophore.

B: Scheme of Fluor de Lys assay reactions. Picture adapted from (Wolfson, Pitcairn *et al.* 2013).

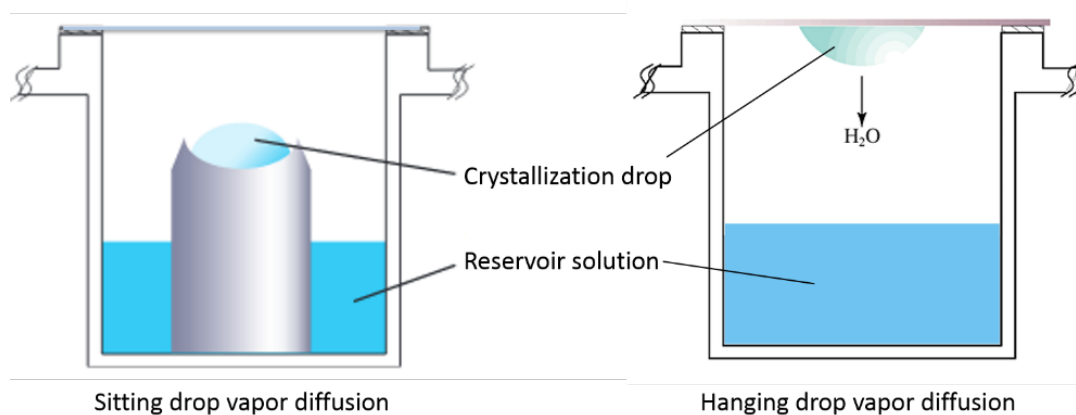
Furthermore, IC<sub>50</sub> values were determined by our collaborators to compare inhibitor potency. IC<sub>50</sub> is the concentration of the inhibitor required to decrease the rate of enzymatic reaction by half (Figure 47). IC<sub>50</sub> values are determined by a series of experiments with a constant enzyme and substrate concentrations and with varying inhibitor concentration. The rate of reaction V decreases upon increasing inhibitor's concentration log [I]. After a set of experiments, V is plotted against log [I] to generate a sigmoidal curve where the point of inflection corresponds to the log of inhibitor concentration that decreases V by half.



**Figure 47: IC50 determination.**  
At half V, inflection point is log IC50.

#### 2.4.5. Crystallization vapor diffusion Technique

Several crystallization methods are available among which vapour diffusion method was used during the thesis work, which is one of the easiest and most commonly employed crystallization method. It separates protein molecules from the solution and simultaneously self-associates protein molecules to facilitate the formation crystal lattice. A drop composed of protein and reagent mixture is allowed to equilibrate with reservoir solution in a sealed chamber. Due to initial low reagent concentration in the drop, water evaporates to attain equilibrium, as a result protein (and reagent) concentration of the drop increases and attains supersaturation. During the process phase separation and nucleation takes place which favours the crystal formation. Vapour diffusion technique can be used in sitting drop and hanging drop methods which is illustrated in (Figure 48).



**Figure 48: Illustration of vapor diffusion using Sitting drop and handing drop setup.**  
Image source: modified from Hampton research.

## 2.5. X-ray crystallography

A three-dimensional structure of a protein defines its function. To determine macromolecular protein structures, three main techniques are available. Nuclear magnetic resonance (NMR) is generally limited to small proteins, electron microscopy (EM) is currently suitable for large proteins and complexes, and X-ray crystallography is widely used, irrespective of protein size. Other techniques to obtain lower resolution data such as SAXS (small angle X-ray scattering) are also complementary. To date, 132,428 structures are available in the protein data bank, which have been solved mostly by crystallography and NMR techniques. The number of EM structures is however growing at the fastest pace. In this thesis, I have used X-ray crystallography in order to investigate the protein-ligand and protein-protein interactions. Different software and methods are available to process the X-ray diffraction data, in the following section I have summarized few that I have used during my thesis.

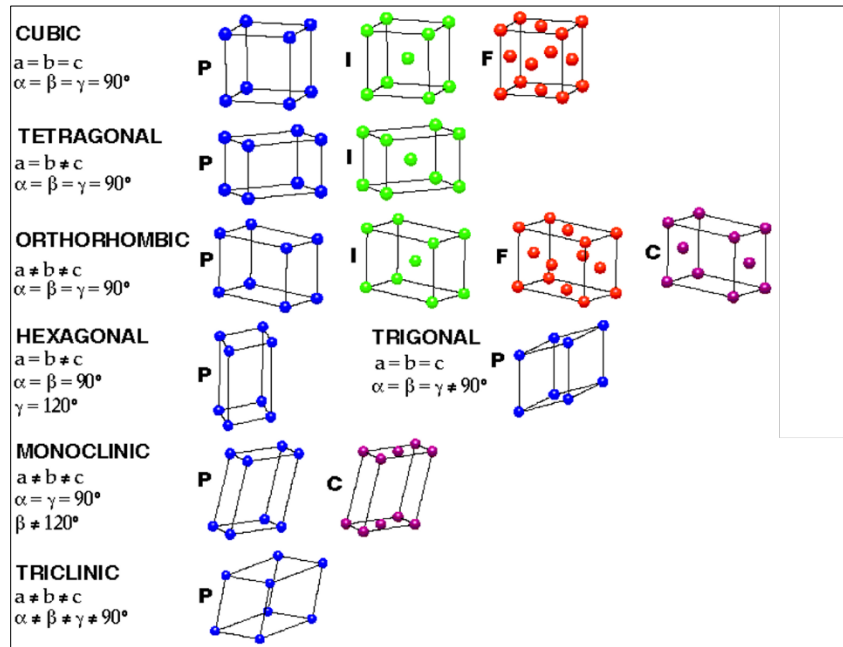
### 2.5.1. Principle of X-ray diffraction

X-ray crystallography is the ultimate tool in the biological macromolecule structure determination. The major hurdle in solving a protein structure through X-ray crystallography is to obtain crystals which is time consuming experiment. Crystal is the periodic arrangement of molecules in three-dimensional space. In order to solve a protein 3D structure from crystal obtaining to structure determination several steps are involved, and it is an extensive and time-consuming process. In this section, I have discussed few basics which will help to understand the process of protein structure determination using X-ray crystallography.

#### 2.5.1.1. Crystal systems

The ordered crystal lattice can be defined by the simplest repeating unit which is called as unit cell. By definition unit cell has a translational symmetry which can build the whole crystal lattice. The lattice parameters that define a unit cell includes three axes  $a$ ,  $b$ ,  $c$  and three angles  $\alpha$ ,

$\beta, \gamma$ . The figure (Figure 49) shows classification of lattice systems in which seven combinations of crystal systems and four lattice types that makes a total of fourteen possible Bravais lattices.



**Figure 49: Bravais lattices**

P: simple cubic, I: body centered, F: face centered cubic, C: side centered

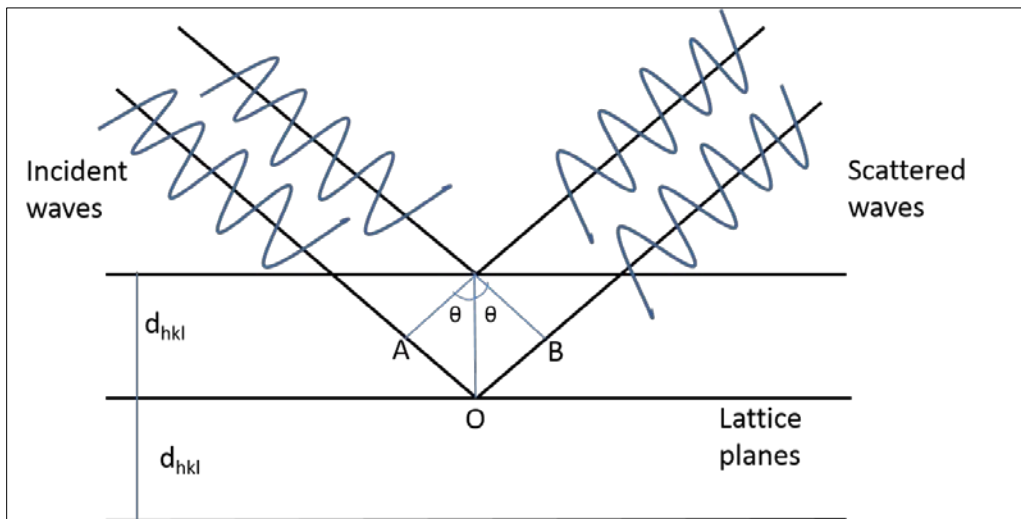
### 2.5.1.2. Bragg's law

Bragg's equation was introduced by Sir William Lawrence Bragg and his father William Henry Bragg in the year 1913. They have proposed to interpret the scattering as reflection from lattice planes (Figure 50). According to Bragg's law when the path difference between diffracted

planes is equal to the integers of wavelength then constructive interference happens which results in the reflection on the detector. The destructive interference leads to no reflection.

$$n\lambda = 2d \sin\theta$$

Where  $n$  is the integer,  $\lambda$  is the wavelength of X-rays,  $d$  is the distance between two planes and  $\theta$  is the angle of the incident light.



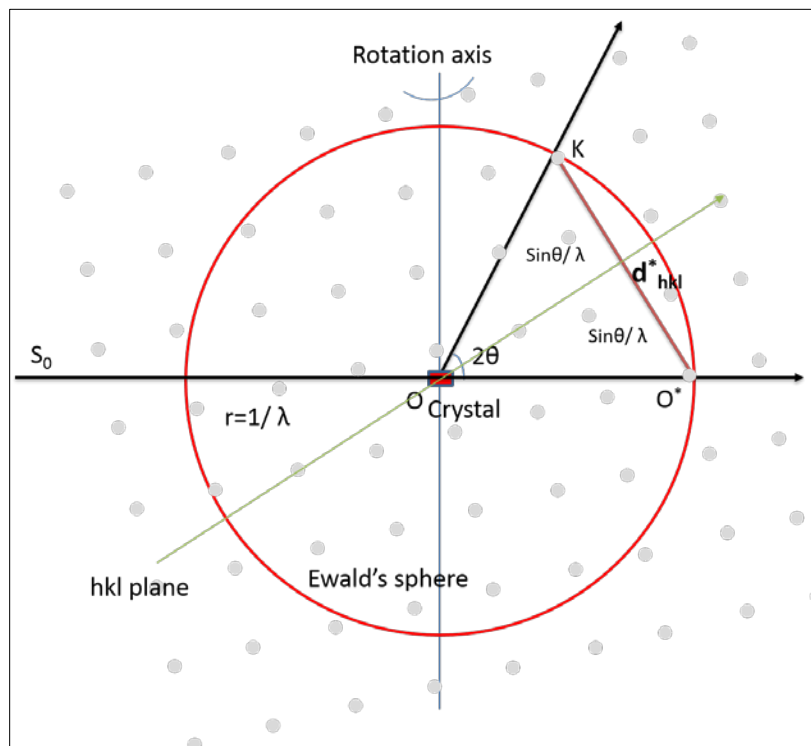
**Figure 50: Bragg's law**

The extra path travelled by the incident X-rays is equal to  $AO+OB$  which is equal to  $2d \sin\theta$ .  $d_{hkl}$  is the interplanar distance.

### 2.5.1.3. Ewald's sphere

Ewald's sphere is a geometrical construction that represents the visualization of diffraction patterns which was demonstrated by Paul Ewald in 1921. The superimposition of reciprocal lattice with Ewald's sphere, shows the diffraction pattern where the reciprocal lattice point that lies on Ewald's sphere will fulfill the diffraction condition (Figure 51). Ewald's sphere is constructed with a radius of  $1/\lambda$  from the origin of crystal center. By changing wavelength and or rotating crystal one can manipulate the diffraction patterns. By rotating the crystal, new lattice points fall on the Ewald's sphere that gives rise to the diffraction. Depending upon the symmetry of the crystal few lattice points will never touch the Ewald's sphere which is called as blind zone, in such case tilting of the crystal can help to get the diffraction of (previous) blind zone.





**Figure 51: Construction of Ewald's sphere**

Ewald's sphere constructed with the radius  $1/\lambda$ , the origin O, reciprocal lattice origin  $O^*$ , reciprocal lattice point represented as grey spheres. Incident X-ray  $S_0$  passes through the origin O and reciprocal lattice origin  $O^*$ . the scattering vector  $OK$  can only make the diffraction pattern when it fulfils Bragg's law.  $n\lambda = 2d \sin \theta$ . The difference between incident X-rays and diffracted X-rays is equal to scattering vector which is  $1/d_{hkl}^*$ .

#### 2.5.1.4. Theory of diffraction

. The wavelengths of X-rays 0.01 to 10 nm are in the range of interatomic distances of biological macromolecules. For instance, the carbon-carbon distance is  $1.54 \text{ \AA}$  which falls under the wavelength of X-rays. When X-rays bombard the crystals, it can interact in three possible ways. The electrons can be released from the matter as photonization where energy and momentum are transferred to electrons from the X-rays. In the second type of interaction the electrons are excited from their ground state but not released from the matter which is also known as Compton scattering. The first two processes are inelastic scattering where the energy of incident X-rays is decreased. In the third kind of interaction, when X-rays hit the crystal, the electrons change the direction of the X-rays without consuming the energy of X-rays. In elastic scattering the outgoing X-rays should have the same energy as the incident X-rays. In biological macromolecular

crystallography, elastic scattering results in the diffraction patterns where the diffracted waves are in constructive interference according to Bragg's law, and the in-phase diffraction results in the reflection on the detector.

### 2.5.2. Structure determination using X-ray crystallography

The information obtained by X-ray crystallography on a macromolecular structure is the electron density of this macromolecule in the three-dimensional space. Determination of the electron density from a diffraction experiment involves several mathematical calculations. A Fourier series is the sum of wave equations that describes a periodic function. Constructing the Fourier series of a wave can describe all of its properties: amplitude, frequency and phase. Since the electron density in 3D space is a periodic function, it can also be represented in the following Fourier expression.

$$\rho_{(x,y,z)} = 1/v \sum_h \sum_k \sum_l F_{(hkl)} e^{-2\pi i(hx+ky+lz)}$$

Where,  $\rho_{(x,y,z)}$  defines the electron density in 3D space with x,y,z fractional co-ordinates,

$1/v$  is the volume of the unit cell which expresses electron density in the units of  $e/\text{\AA}^3$

$F_{(hkl)}$  is the structure factor at miller indices h, k and l

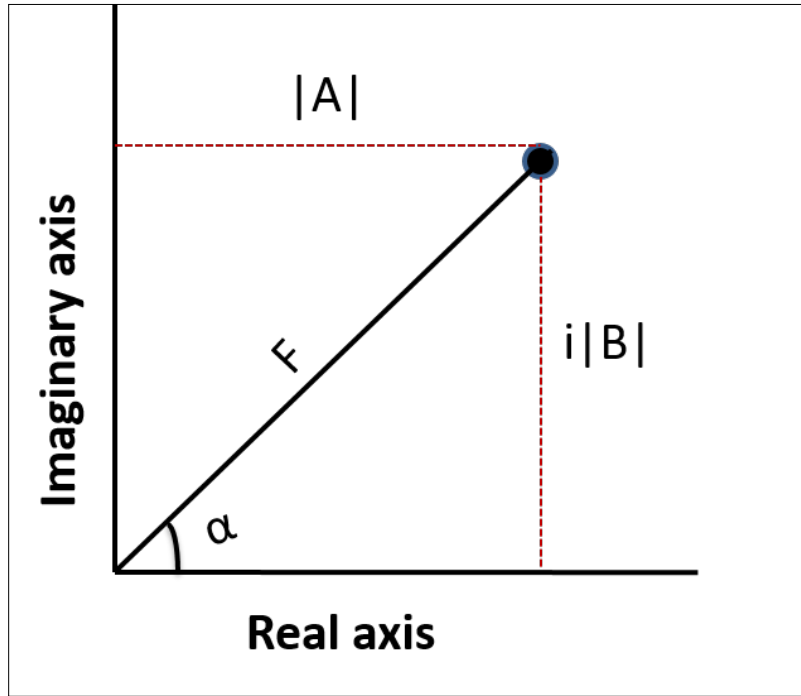
$e^{-2\pi i(hx+ky+lz)}$  are the coordinates in the real and reciprocal space that is coming from Fourier synthesis

Since the structure factor  $F_{(hkl)}$  is a Fourier series of wave function, it defines its three components: frequency, amplitude and phase. The frequency is known from the X-ray source, and the amplitude is the square root of intensities  $I_{(hkl)}$  which are measured during the diffraction experiment. However, the phase information is being lost during data collection. This is known as the "phase problem".

In order to be able to obtain the electron density of a macromolecule from its diffraction experiment, it is required that the phase problem is solved. Its mathematical representation is described briefly in the following section.

The two important components of a structure factor from the above equation are the structure factor amplitude and the phase. These can be represented as complex number in a two-

dimensional representation (Figure 52). Any unknown number ( $iB$ ) with a known real number ( $A$ ) can be represented as a complex number ( $A+iB$ ) in a two-dimensional space using the Argand diagram. A complex number is the one that contains both real and imaginary numbers where  $i = (-1)^{1/2}$ .



**Figure 52: Argand diagram:**

The structure factor is represented as a vector with complex numbers. The length or amplitude of structure factor  $F$  is the square root of intensities (measured in the diffraction experiment) which makes an angle of alpha (phase) with the real axis (amplitude).

The electron density can be derived with known intensities and unknown phases, from the above Argand representation of a structure factor. Two expressions can be derived from Figure 52.

$$\cos \alpha = |A|/|F| \text{ and } \sin \alpha = i|B|/|F|$$

Therefore,

$$|A| = |F| \cos \alpha \text{ and } i|B| = |F| \sin \alpha$$

The complex vector

$$F = |A| + i|B| \text{ is equal to } |F|(\cos \alpha + i \sin \alpha)$$

According to complex number theory  $\cos \theta + i \sin \theta = e^{i\theta}$ , the complex vector becomes:

$$F = |F|e^{i\theta}$$

Substituting this expression in the electron density equation, and using phase angle  $\alpha = 2\pi i\alpha'$  gives:

$$\rho_{(x,y,z)} = 1/v \sum_h \sum_k \sum_l |F_{hkl}| e^{2\pi i\alpha'_{hkl}} e^{-2\pi i(hx+ky+lz)}$$

Further simplifying the equation by adding exponents gives:

$$\rho_{(x,y,z)} = 1/v \sum_h \sum_k \sum_l |F_{hkl}| e^{-2\pi i(hx+ky+lz) - \alpha'_{hkl}}$$

This electron density equation is complete by representing as a function of known amplitudes and unknown phases of each structure factor. This equation represents all three components of wave frequency  $h$  along axis  $x$ ,  $y$ ,  $z$ , amplitude  $|F_{hkl}|$  and phase  $\alpha_{hkl}$ .

Different techniques are available to solve phase problem

- Direct methods (limited to small molecules)
- Molecular replacement
- Isomorphous replacement SIR, MIR
- Anomalous dispersion (SAD, MAD)
- Combination methods (SIRAS, MIRAS)
- Radiation damage induced phasing with anomalous scattering RIP, RIPAS

During my thesis, I have used the molecular replacement method which is described in the next section.

### 2.5.2.1. Molecular replacement

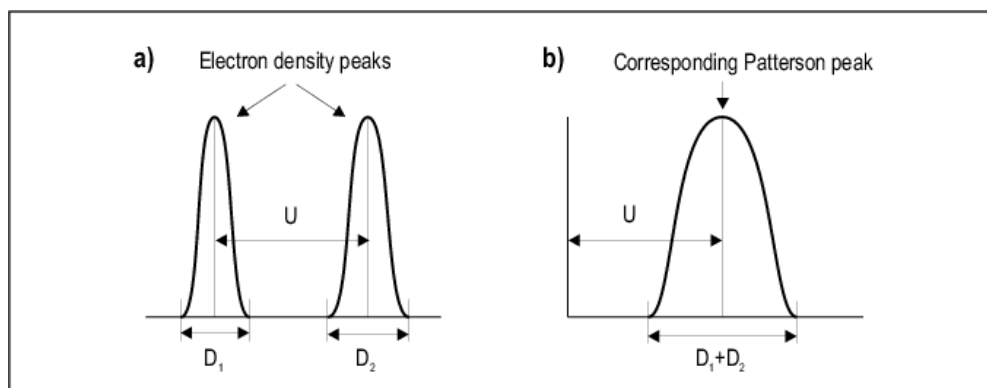
The molecular replacement method was first developed by Michael Rossmann. With this method the phases are estimated from the structure factors calculated from a known homologous model structure. The calculated structure factor  $|F_c|$  is compared to the observed structure factor  $|F_o|$  by applying rotation and translational operations to the initial model to get the final model. To be successful, molecular replacement requires a total of six parameters ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $x$ ,  $y$ ,  $z$ ) for each structure factor of all the atoms in the unit cell has to be calculated which is computationally exhaustive.

In order to simplify the calculation, molecular replacement softwares split these six parameters into two three-dimensional parameters, i.e. three rotational parameters ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) and three translational parameters ( $x$ ,  $y$ ,  $z$ ). After finding a correct rotational fit, the translational search is implemented. However, after the first step of rotational search, the  $|F_c|$  cannot be calculated when the translation parameters are still unknown. This prevents the comparison of  $|F_c|$  with  $|F_o|$ . In order to overcome this barrier, two methods are available for the rotation search (i) a Patterson search function (i.e. a Fourier transform of intensities), or (Suzuki, Muto *et al.*) a maximum likelihood method (i.e. a statistical approach in reciprocal space to average all the possible translations).

The Patterson method is the Fourier synthesis of squared structure factor amplitudes without phases. Since the Patterson function uses intensities in place of phases as Fourier coefficients, the expression of electron density changes to Patterson function of:

$$P_{(uvw)} = 1/v \sum_h \sum_k \sum_l |F_{(hkl)}|^2 \cos 2\pi [hu + kv + lz]$$

The Patterson function of observed data contains intramolecular self-vectors as well as intermolecular vectors due to crystallographic and non-crystallographic symmetry. The Patterson function of the model is calculated in a large P1 unit cell with the model's intramolecular self-vectors present around its origin. The crystal or observed Patterson is calculated in its actual unit cell which contains self-vectors and cross vectors of intermolecular and intramolecular vectors respectively (Figure 53).



**Figure 53: Calculation of Patterson peaks from electron density:**

a) individual electron density peaks with a distance of  $U$  and the width of  $D_1$  and  $D_2$ . b) Patterson peak derived from the two electron density peaks in a.

Once the Patterson functions are calculated for both the model and the target, the Patterson of the large P1 model is placed inside the unit cell of the target and the rotation function is applied. The overlap between the observed and calculated Pattersons can be measured as a product function or correlation coefficient or as Patterson product in reciprocal space, which is at its maximum when the two Pattersons overlap.

$$RF(R) = \int_{r_{min}}^{r_{max}} P_{observed}(u)P_{model}(R, u)du$$

RF is the Patterson function of rotation R and is equal to the product of the observed Patterson  $P_{observed}(u)$  and of the rotated model Patterson  $P_{model}(R, u)$  integrated over all points of u in the Patterson space.  $r_{max}$  and  $r_{min}$  are the limits of the sphere of radius which is centered on the origin.

The translation search is related to the crystallographic symmetry. In case of the P1 space group, the translation result in the same. The translation search for a Patterson is defined as the product of observed and model Pattersons over the whole unit cell. Translation and rotation result in the movement of orthogonal coordinates of Patterson into a new frame, where the translation is straightforward, but the rotation is expressed in different ways such as Polar angles, Eulerian angles and Lattman angles. A correlation coefficient or R factor is calculated for each translation search.

$$R = \frac{\sum |F_{observed}| - |F_{calculated}|}{\sum |F_{observed}|}$$

The model obtained from molecular replacement may contain local errors and systematic displacement. These errors might be corrected in case of rigid body refinement. In rigid body refinement, the atoms are fixed relatively to each other.

#### 2.5.2.2. Refinement

Models obtained from the structure determination using experimental phasing and molecular replacement, are not complete but in general contain enough phase information to build the missing residues and identify the regions which are not correct in the initial molecular

replacement model. In order to obtain a better fit between the model and the experimental data, repeated cycles of map calculations and model building are implemented. Refinement programs use least square refinement and maximum likelihood methods to modify the model to get the best agreement between the model and experimental data. In this thesis, I have used programs from Phenix and CCP4 for the refinement and COOT for the model building.

# Results

---



### 3. Results

During my PhD thesis, I have been developing three research axes that all contributed to address the scientific questions discussed in the introduction of this thesis on epigenetic enzymes involved in the acetylation pathways, notably on histone deacetylases. Accordingly, this Results section has been divided into three parts. In the first part, I describe the work that I did on epigenetic enzymes of the acetylation pathways of eukaryotic pathogens and that have been validated as drug targets by the A-ParaDDisE consortium. Although I have been studying many different enzymes, I am describing here mostly the production work done on two enzymes: a histone acetyltransferase from the pathogen *Trypanosoma cruzi*, and a deacetylase from the pathogen *Leishmania braziliensis*.

In the second part, I describe the work I did on the understanding of the selective inhibition of HDAC8. Here, I have continued the initial work done on smHDAC8 and I have complemented it with the work on hHDAC8. This second part of my results is divided into five subparts. First one is the expression, purification and crystallization of HDAC8 (both smHDAC8 and hHDAC8). Second and third subparts are the structural characterization of HDAC8 selective inhibition, which are accompanied by two articles, one which is published and a second one which should be soon submitted for publication. In subparts four and five, I describe additional work done on the selective inhibition of HDAC8 that should be published later.

Finally, in the last two subparts, I describe the on-going work I am doing on the Cohesin complex that is a target of HDAC8. This work describes cloning, expression, purification and crystallization studies on subcomplexes of Cohesin that should be used for looking at the interaction between Cohesin and HDAC8.

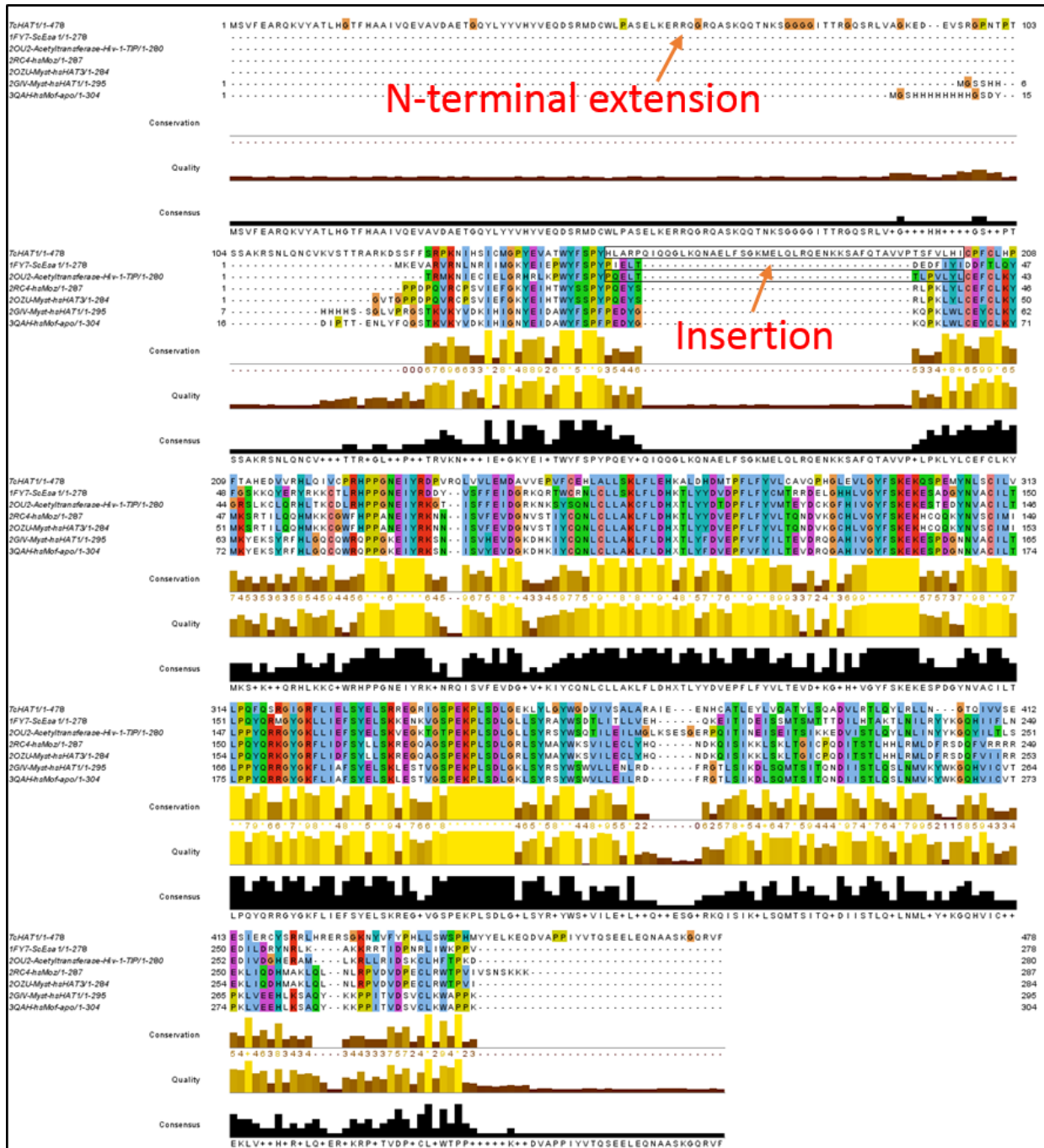
### 3.1. Purification of acetyl modifying proteins from eukaryotic parasites

Several epigenetic enzymes from eukaryotic parasites have been identified by the partner laboratories of the A-ParaDDisE consortium and, for some of them, were validated as drug targets. Among them, potential candidates were selected for the *in vitro* screening assays for candidate inhibitors. The aim of my work was to produce some of these proteins recombinantly to analyse them by structural means and to deliver to our partner laboratories for *in vitro* inhibition assays. I describe here the work I did on two such validated epigenetic drug targets: a HAT from *Trypanosoma cruzi* (TcHAT1) and a deacetylase from *Leishmania braziliensis* (LbDAC3).

#### 3.1.1. Histone Acetyltransferase 1 from *Trypanosoma cruzi* (TcHAT1)

##### 3.1.1.1. Initial expression tests of TcHAT1

Bioinformatic analysis suggests that TcHAT1 belongs to the MOZ/SAS family of acetyltransferases, which has been produced recombinantly. Initially, TcHAT1 full-length protein was cloned and expressed in bacteria, as mentioned in the materials and methods section (2.2.2 & 2.3.4). The full-length protein expression was poor and I failed to scale up its production. Bioinformatics analysis was conducted and the sequence alignment studies revealed that TcHAT1 contains an N-terminal additional sequence as well as an insertion in the middle of the protein (Figure 54). In the first attempt, the N-terminal additional sequence was removed and re-cloned (TcHAT1-N2C2). Different combinations of expression tags were used. In minitest-expression assays, an N-terminal histidine-thioredoxin fusion tag showed the most promising results in terms of yields, notably for the construct TcHAT1-N2C2.

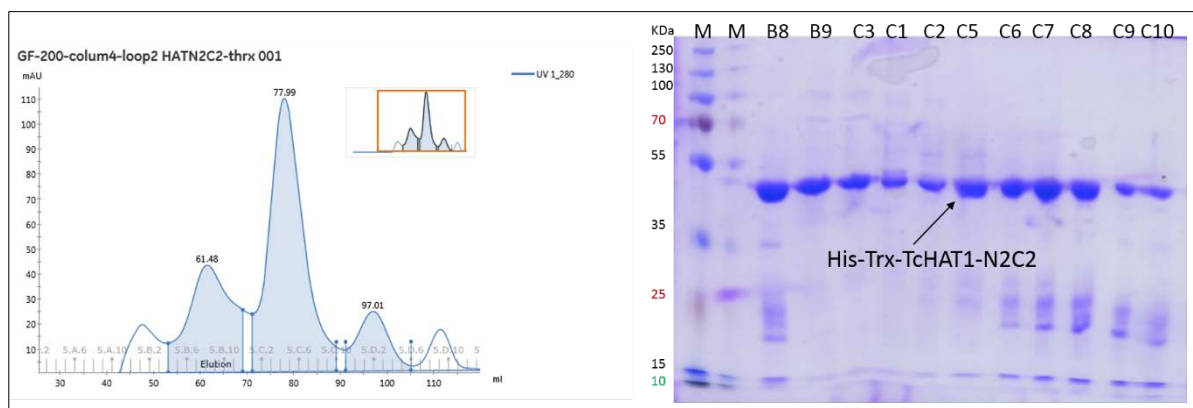


**Figure 54: Sequence alignment of TCHAT1:**

BLAST search was conducted for TCHAT1 against protein sequences available in pdb. Insertion was represented in black border. Image was prepared using Jalview.

### 3.1.1.2. TcHAT1-N2C2 Purification

This better behaviour of the TcHAT1-N2C2 construct in fusion with an N-terminal His-Thioredoxin let me choose this construct for large scale production studies. After cell lysis with sonication, Talon affinity binding was done followed by washing and imidazole elution. A final step of gel filtration chromatography was then performed (*Figure 55*). The gel filtration buffer was 50 mM Tris, 150mM NaCl and 2mM DTT. I observed however that this protein was highly degrading, and protein precipitation was observed all over the purification. The degradation was reduced to some extent by the thioredoxin fusion tag. However, since acetyl-CoA is a cofactor of HATs, we also supplemented it during purification to look whether it can stabilize the protein. However, this protein is not well behaved in DLS, as it shows high molecular weight content which suggests soluble aggregation. Nevertheless, the purified protein was delivered to our partner laboratories for activity tests.



**Figure 55: Gel filtration purification of TcHAT1-N2C2.**

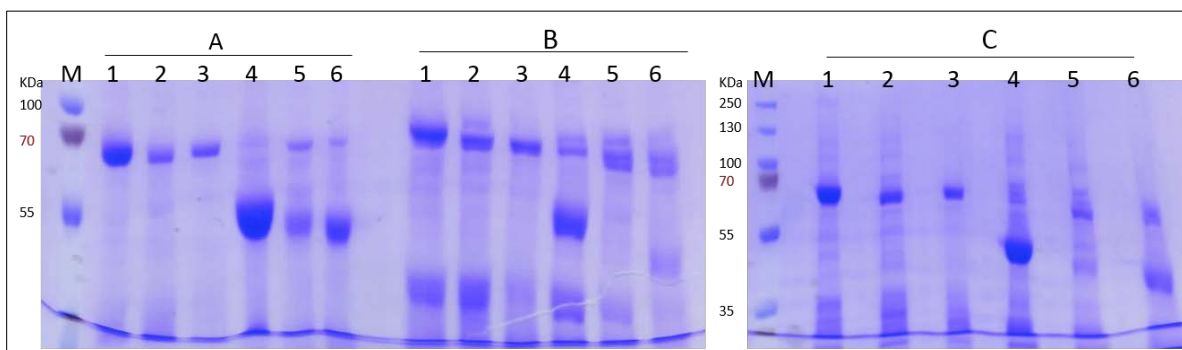
Top panel: gel filtration profile; bottom panel: SDS-PAGE analysis of gel filtration fractions. The protein (MW of 48 kDa with the thioredoxin fusion tag) is found in two peaks on gel filtration chromatography.

### 3.1.1.3. Optimization of production

In order to try to improve the quality of protein, notably by reducing its aggregation state, I further engineered TcHAT1-N2C2 and modified its purification conditions. The presence of an insertion in the middle of sequence may be the cause of improper quality of protein. And also, the high cysteine content is another issue, where TcHAT1-N2C2 has nine cysteine residues, which

can be addressed in partial by increasing reducing agent concentration. In another attempt, the loop of TcHAT1 was deleted and two mutants were prepared, one with deleted loop (insertion - Figure 54) (dl-TcHAT1) and in the second case the loop at insertion was replaced with human sequence (lh-TcHAT1, lh for loop human).

Two loop mutants were prepared by cloning TcHAT1 (full length) and TcHAT1-N2C2 one with deleted loop (dl), other one replaced with human sequence (lh). All constructs were maintained N-his-thioredoxin fusion tag. Mini tests were performed as mentioned in the section (2.2.2.1). In expression profile, TcHAT1-N2C2 looks more soluble, after lh-TcHAT1-N2C2. And the least soluble is the deleted loop mutant lh-TcHAT1 and lh-TcHAT1-N2C2 (Figure 56). Further, these proteins were used for large scale production.

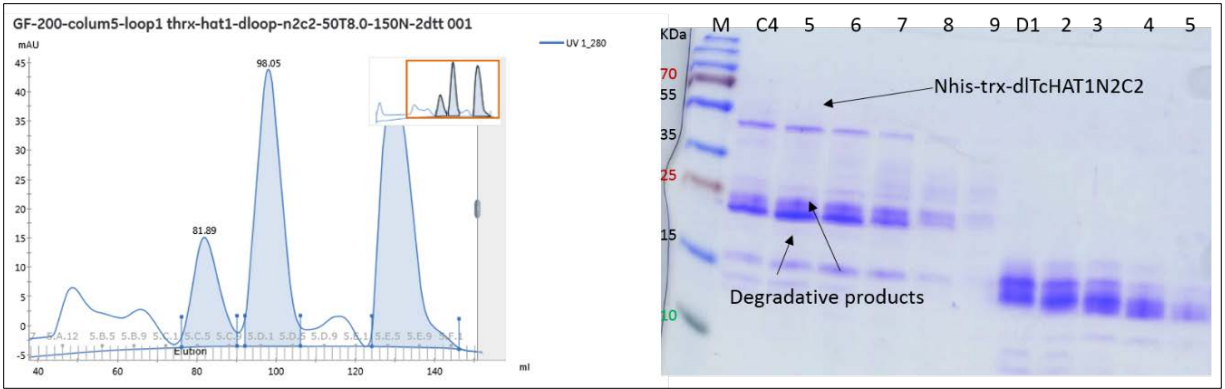


**Figure 56: TcHAT1 – mini test-expression optimization:**

1-6, TcHAT1, dl-TcHAT1, lh-TcHAT1, TcHAT1-N2C2, dl-TcHAT1-N2C2 and lh-TcHAT1-N2C2 respectively. All proteins are tagged with N-terminal-his-thioredoxin fusion tag. A-C: 50 mM Tris, pH 8.0, where A – 50 mM NaCl, B - 150 mM NaCl and C – 500 mM NaCl.

#### 3.1.1.4. Purification of dl-TcHAT1-N2C2

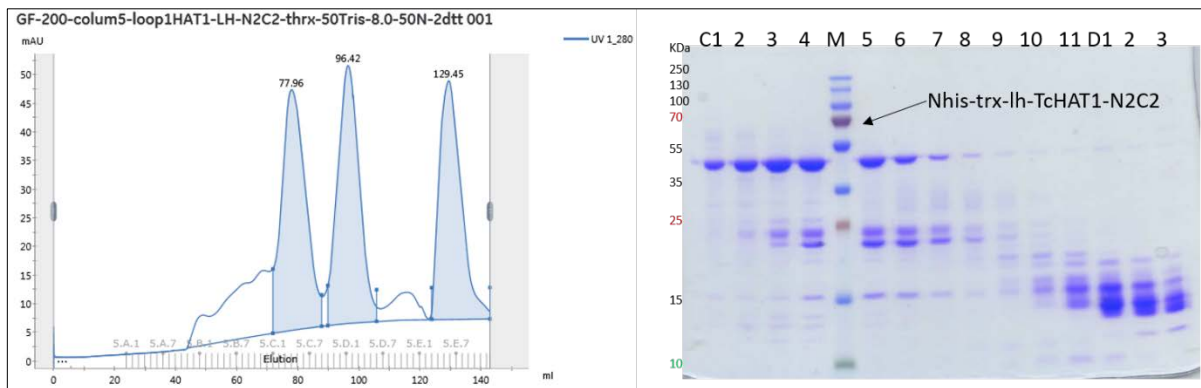
dl-TcHAT1-N2C2 was purified from three liters of culture. Purification was done by affinity tag purification followed by gel filtration chromatography (Figure 57). The protocol followed was the same as for the wild-type enzyme. The purification of dl-TcHAT1-N2C2 was not successful as hinted in the mini test (Figure 56). The solubility and proteolytic stability of the protein were highly affected by the loop deletion. These results suggest that this loop is critical for the TcHAT1 structure and this construct was not further characterized.



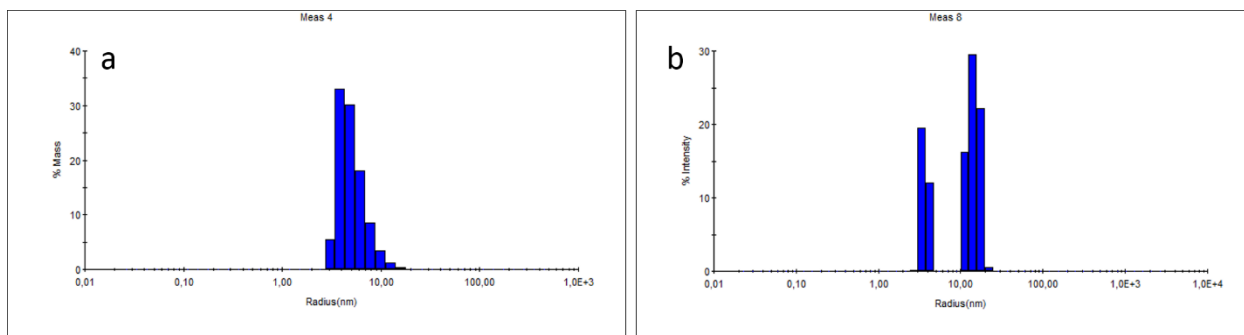
**Figure 57: Purification profile of dl-TcHAT1-N2C2:**  
Gel filtration chromatogram and SDS gel indicating unstable protein.

### 3.1.1.5. Purification of lh-TcHAT1-N2C2

The mutant lhTcHAT1N2C2 was purified in a similar way to above mentioned wild type enzyme (Figure 58). Surprisingly, this mutant has better solubility and stability than deleted-loop mutant and wild type enzymes (Figure 59). There was very little degradation observed, and overall precipitation during the purification was reduced completely. This observation again supports the importance of this insertion in the structure of TcHAT1. The purified protein was delivered for activity and *in vitro* drug screening tests. This mutant purification was a starting point towards crystallization trials.



**Figure 58: Purification profile of lhTcHAT1N2C2:**  
Gel filtration chromatogram and SDS gel indicates good improvement in the protein yield and quality. The degradation and precipitation while purification much reduced with this new construct.



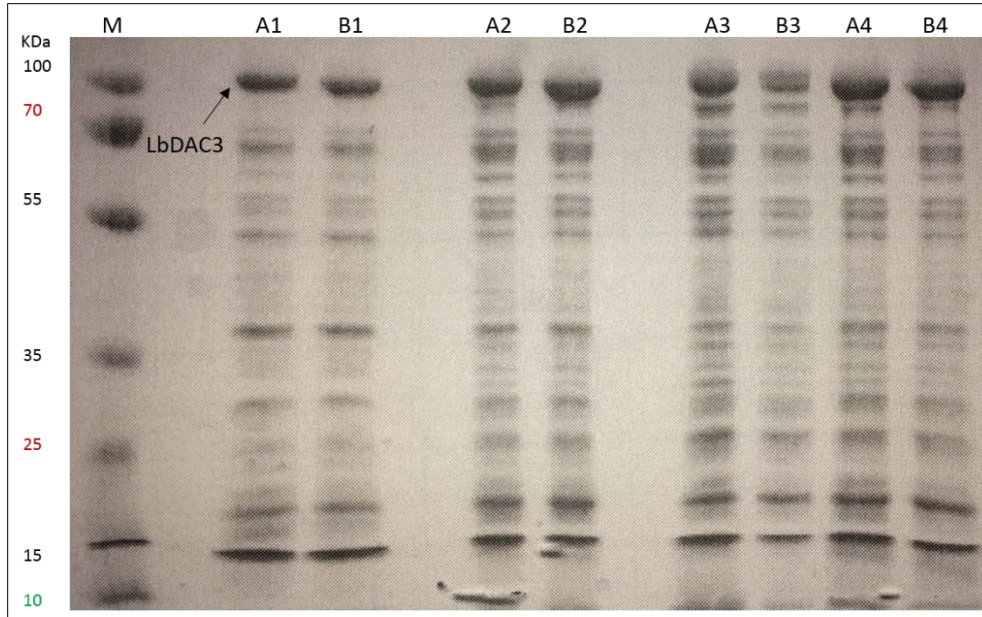
**Figure 59: DLS profile of lh-TcHAT1-N2C2**

a) Purified lh-TcHAT1-N2C2 protein in presence of thioredoxin fusion tag which is stabilized when compared to wild type enzyme. Upon addition of acetyl CoA the stability is improved which reflects in the DLS data b)

### 3.1.2. Deacetylase 3 from *Leishmania braziliensis* (LbDAC3)

LbDAC3 is another parasitic epigenetic enzyme that has been validated as a target by the A-ParaDDisE consortium. It is a homologue of class II HDACs (Figure 62). LbDAC3 was cloned and expressed as mentioned in the materials and methods (2.2.2 & 2.3.4). Among different fusion tags and expression conditions thioredoxin fusion tag was selected with a better solubility. In the mini expression test, different constructs and buffers were used to optimize protein production (Figure 60 ). More or less all conditions yield similar solubility profiles.





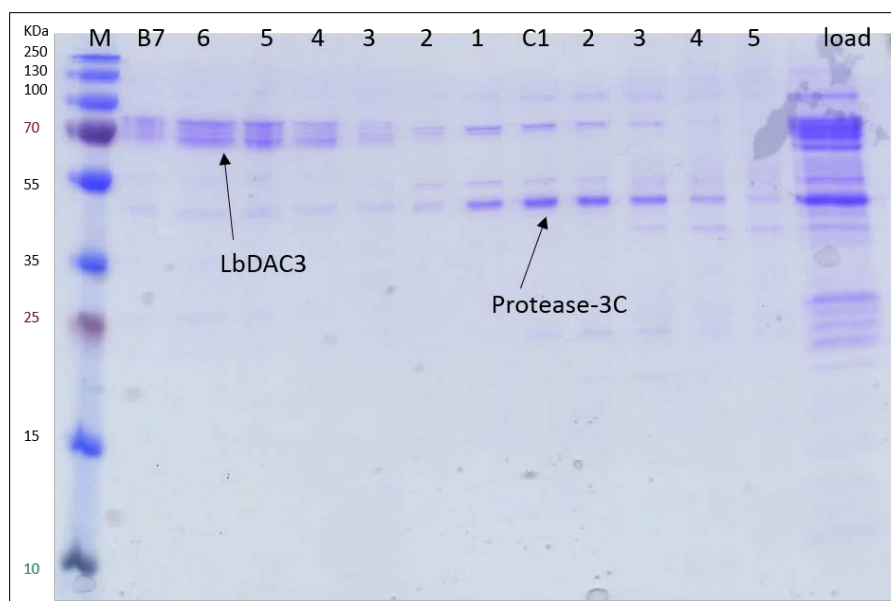
**Figure 60: LbDAC3 expression optimization:**

A: N-his-thrx-3C-LBDAC3, B: N-his-thrx-thrombin-LbDAC3. 1-4 are different buffer conditions, 1: 50 mM KCl, 10 mM Tris, pH8.0, 2: 150 mM KCl, 10 mM Tris, pH8.0, 3: 300 mM KCl, 10 mM Tris, pH8.0 and 4: 150 mM KCl, 10 mM Tris, pH 7.5.

### 3.1.2.1. Purification of LbDAC3:

PnEA-thX-LbDAC3 with 3C protease site (N-his-thrx-3C-LBDAC3) was expressed and purified as mentioned in materials and methods section (2.2.2 & 2.3.4). After expression, cell lysis was done using sonication. The soluble protein was purified using affinity purification followed by protease 3C treatment and finally gel filtration chromatography was used (Figure 61).





**Figure 61: Purification of LbDAC3**  
 Profile of LbDAC3 purification using gel filtration chromatography

However, the yield was too low to pursue further purification and the protein is highly degrading. In order to understand the protein behavior bioinformatics analysis were pursued which revealed a huge insertion in LbDAC3 sequence (Figure 62). The sequence alignment of LbDAC3 with class II human HDACs has indicated that LbDAC3 is different from human enzymes. It lacks an N-terminal region but contain insertions in the catalytic domain. In order to optimize the protein, several constructs were made by deleting insertion at different lengths. Constructs were prepared with different fusion tag options. In the mini expression tests, protein expression was observed but failed to scale up the volume. No construct has yielded a stable protein, all were degrading and precipitating during the purification.

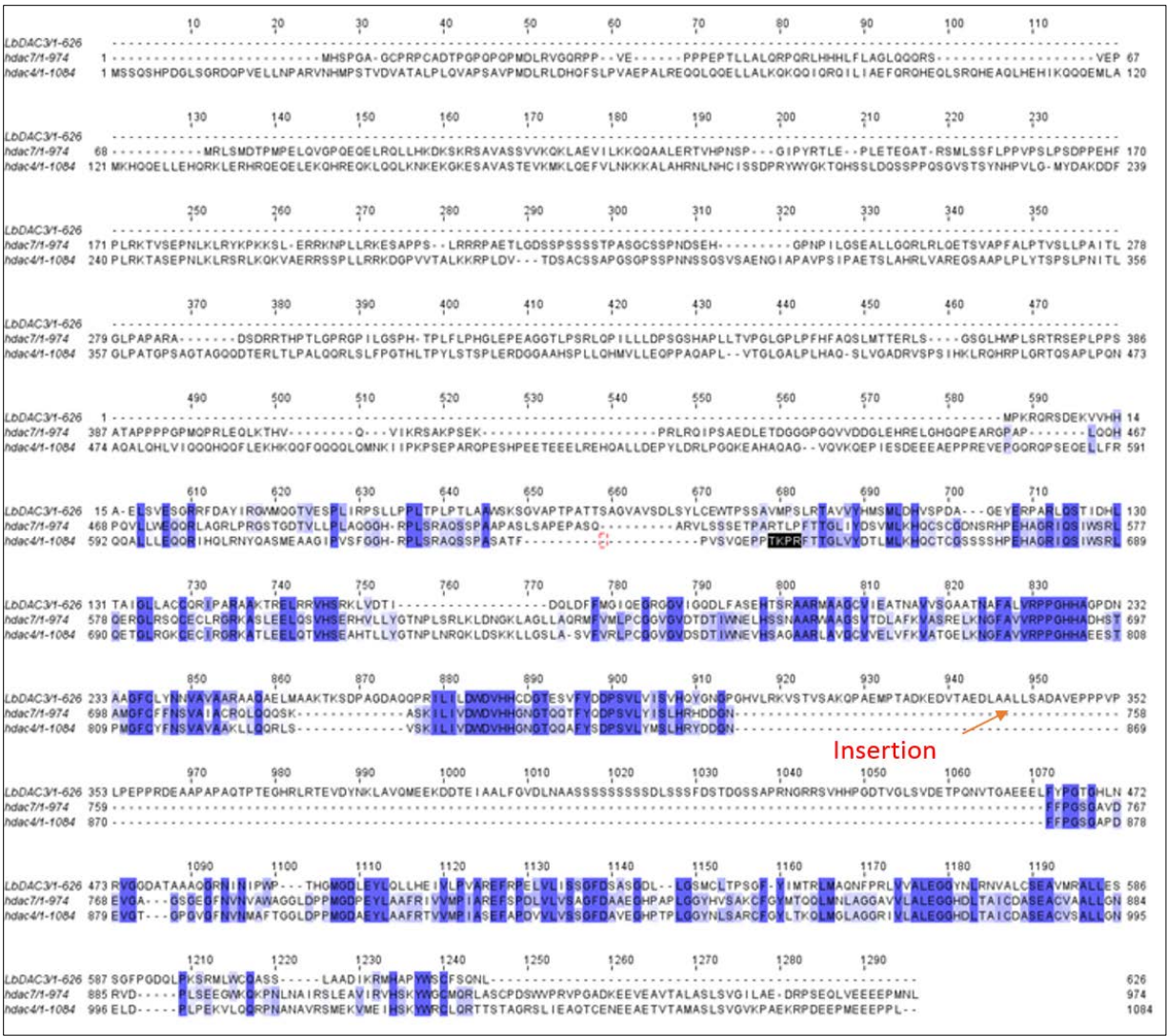
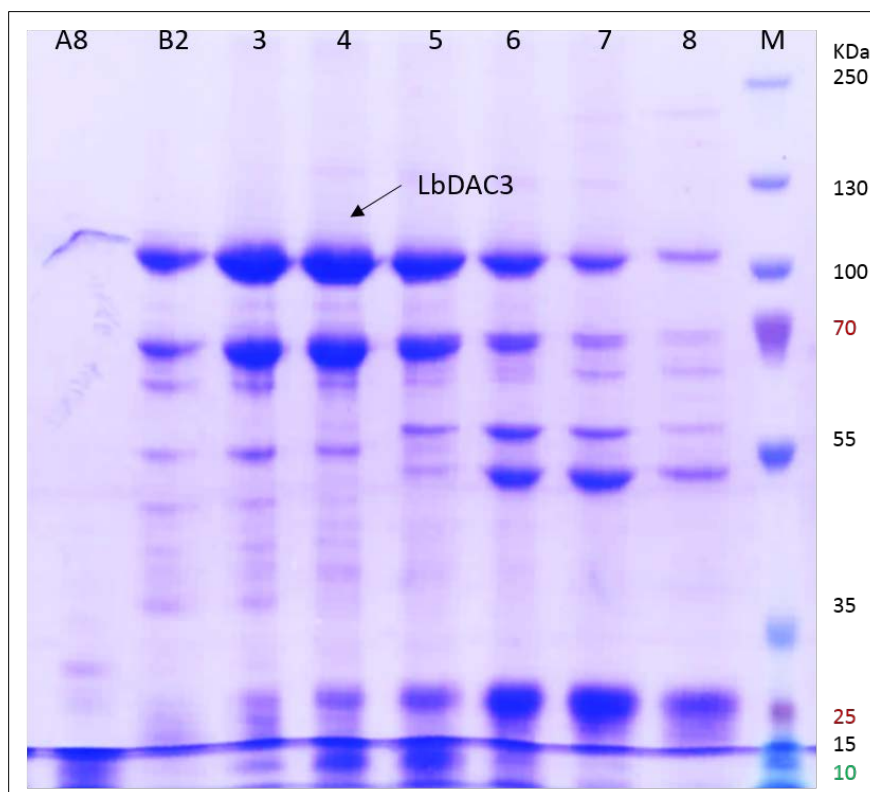


Figure 62: Sequence alignment of LbDAC3 with HDAC7 and 4.

### 3.1.2.2. Optimization of expression and purification:

After several trials with cloning different constructs of LbDAC3, which didn't give good results, an optimization with expression conditions were tried. The expression conditions were changed from 2XLB media to auto induction media, which has improved the solubility of the protein several folds. LbDAC3 (N-his-thrx-3C-LBDAC3) was expressed in auto induction media and followed by its purification with affinity and ion exchange chromatography. For ion exchange chromatography, the procedure mentioned in section (2.3.2) was followed.

After sonication affinity purification was performed and the protein was eluted with 200 mM imidazole in lysis buffer (50 mM Tris, pH 8.0, 150 mM NaCl and 1mM TCEP). Before ion-exchange chromatography, the salt concentration was reduced to 50 mM. An ion-exchange kit from GE life sciences was used to check ion-exchanger compatibility with mini tests which enabled to select Q-sepharose column for the large-scale protein purification. Protein was then loaded on a Q sepharose column and eluted with a gradient of increasing salt concentration. Purified protein has many contaminants, and most affectively it was degrading rapidly over time (Figure 63). Because of degradation the protein was not able to process for additional purification steps. However, the production wise it has improved comparatively from the previous expression conditions which can be used for optimization.



**Figure 63: Ion exchange profile of LbDAC3**

N-his-thrx-3C-LBDAC3 was purified in two steps, affinity chromatography followed by Q-sepharose chromatography. The resulted protein was degrading and precipitating which didn't allow to pursue further purification studies.

### 3.1.3. Conclusion:

Acetyl modifying enzymes in eukaryotic parasites are often important for parasite survival and for maintaining homeostasis, making them valid drug targets. Several epigenetic enzymes of eukaryotic parasites have been identified by the A-ParaDDisE consortium. The production of these enzymes recombinantly can be useful to pursue *in vitro* assays to screen for drug candidates. However, the presence of long insertions in parasitic enzymes make difficult to purify these enzymes. LbDAC3 and TcHAT1 are two examples characterized in this thesis that are highly unstable enzymes, possibly, but not only, due to the presence of additional insertions in comparison with that of human enzymes.

However, I showed that protein engineering by modifying or deleting insertion sequences can yield more stable enzymes. In case of TcHAT1 a fusion tag thioredoxin and the replacement of loop with human enzyme has given interesting results. Additionally, the use of acetyl CoA also contributed to the stabilization, even if it did not prevent completely oligomerization. The efforts are less successful with LbDAC3, and suggest that an interacting partner is missing that could stabilize this protein.

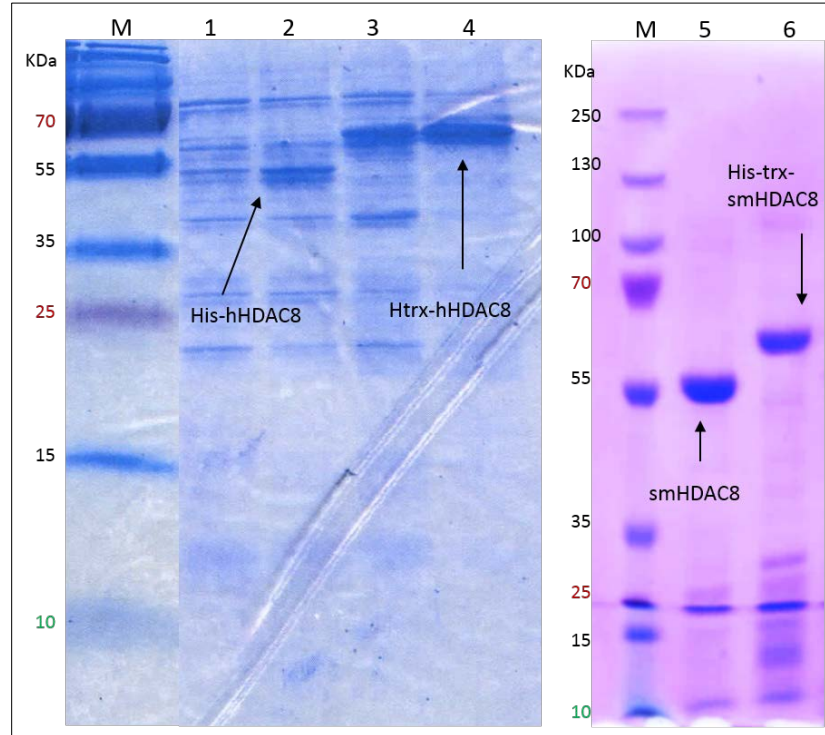
## 3.2. Elucidating smHDAC8 selective inhibition mechanism

### 3.2.1. Production of HDACs

When I joined the lab, *S.mansoni* HDAC8 was the only HDAC for which a protocol of purification existed (Marek, Shaik *et al.* 2016). Human HDAC8 had been shown by other research groups to be purified fused to a C-terminal his-tag (Decroos, Clausen *et al.* 2015). To continue the project on smHDAC8 and start the one on hHDAC8, I dedicated time to setup purification protocols that were easier to use and led higher amounts of proteins.

For that purpose, I cloned *S. mansoni* and human HDAC8 into different expression vectors with different fusion tag combinations. For wild-type smHDAC8 the production was optimized in a pNEA/tH vector where a thrombin protease site followed by a His tag was fused C-terminal to the protein of interest. For hHDAC8 the expression was optimized in pNEA-tX vector in which a

His tag followed by thioredoxin and a thrombin cleavage site was fused N-terminally to the protein of interest.



**Figure 64: HDAC8 expression optimization:**

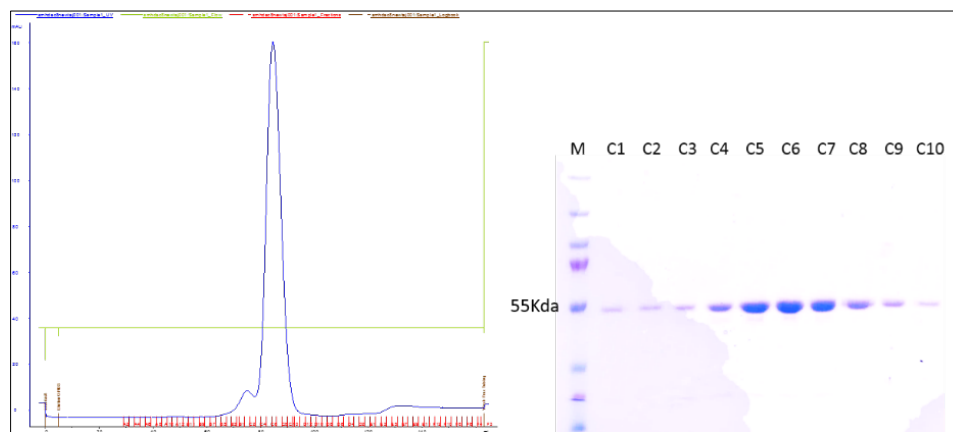
1 and 2 are N-his-hHDAC8 expressed at 25°C and 37°C respectively. 3 and 4 are N-his-trx-hHDAC8 expressed at 25°C and 37°C respectively. 5 and 6 are smHDAC8-Chis and N-his-trx-smHDAC8 respectively expressed at 37°C. SDS gels are edited to match marker.

### 3.2.1.1. Purification of smHDAC8 and hHDAC8

From the above mentioned constructs, a large scale protein production and subsequent purification were performed. hHDAC8 was purified with a simple protocol as mentioned in sections (2.2.2 & 2.3.4) whereas two different protocols were employed to purify smHDAC8 proteins. The wild type smHDAC8 was purified using the new protocol which was mentioned in the section 2.3.5.2). This new protocol has yielded three times more protein when compared to old protocol (3 mg vs 1 mg per 1 ltr culture). However mutant smHDAC8 could not be purified using this new protocol. The reason behind this difference is possibly due to the nature of the smHDAC8 enzyme. The new protocol has rough expression and purification conditions like overnight

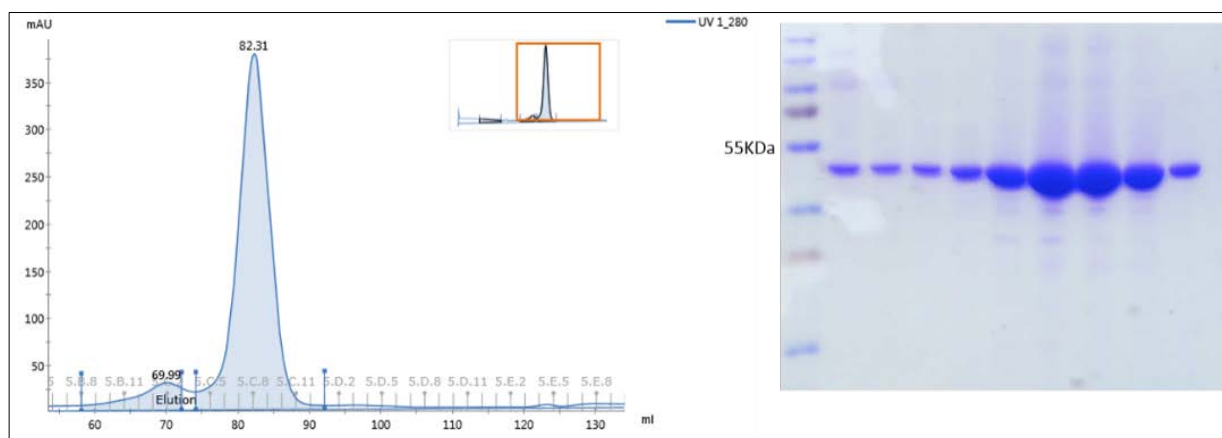
expression at 37°C, using high salt at 150 mM for cell lysis etc. The very sensitive smHDAC8 mutant proteins were not compatible with these harsh conditions but yielded moderate protein with old protocol which was mentioned in the book chapter and section at (2.3.5.1) (Marek, Kannan *et al.* 2013, Marek, Shaik *et al.* 2016).

Final buffers used for gel filtration chromatography were 10mM Tris pH 8.0, 150mM KCl, 2mM DTT for hHDAC8 and 10mM Tris pH 8.0, 50mM KCl, 2mM DTT for smHDAC8. Peak fractions of proteins were concentrated to 3 mg/ml for crystallization experiments.



**Figure 65: smHDAC8 purification profile.**

On the left side Superdex S200 gel filtration chromatography of smHDAC8 was shown. The chromatogram showing a single peak and in the right panel SDS gel with all the peak fractions were showing the purity of the protein.



**Figure 66: hHDAC8 purification profile.**

Superdex S200 gel filtration chromatogram of hHDAC8 and in the right SDS gel analysis of peak fraction.

### 3.2.1.2. Crystallization and soaking with inhibitors of smHDAC8

smHDAC8 crystallization was done using hanging drop method. For crystallization, the protocol was adapted from (Marek, Kannan *et al.* 2013). After gel filtration chromatography, peak fractions were pooled and concentrated in a 30 KDa cut off amicon ultra centrifugal device®. Nearly 3 mg/ml concentration was maintained and the sample was centrifuged prior crystallization experiment. A fine gradient of PEG 3350 from 13% to 20% in presence of 0.1 M Na<sup>+</sup> K<sup>+</sup> tartarate was used as reservoir solution. Nearly 0.5 ml reservoir solution was used in a 24 well XRL crystallization plates, and 1 µl of protein was mixed with 1 µl of reservoir solution on a cover slide and sealed with grease. The crystallization plates were stored at 20 °C, for three days. Crystals were about 200 µm in size and the best crystals were picked and transferred into a new cover slide containing 1 µl inhibitor solution again the plates were sealed and stored at 20°C overnight. The inhibitors were obtained from our collaborators, which were used to prepare a 100 mM stock by dissolving into an appropriate volume of DMF or DMSO. For soaking experiments solution was prepared with inhibitors at a final concentration of 5-10 mM with a buffer containing 10mM Tris pH 8.0, 50mM KCl, 18-24% PEG 3350. Structures of all the inhibitors used in this thesis are mentioned in figures (Figure 67, Figure 68, Figure 69 and Figure 70). After overnight incubation crystals were harvested using a cryo solution containing 22% glycerol in the soaking buffer. Crystals were harvested and stored in liquid nitrogen until data collection.

Several inhibitors were used in this thesis work, which are received from different collaborators, hence the compounds belong to different chemical groups. As a continuation of SEtTReND, the initial lead compounds of smHDAC8 specific inhibitors J1038 and J1075 were developed into two different series, named as TH and TB-series respectively (Marek, Kannan *et al.* 2013). Apart from these two major series, few other group of compounds were also used which includes triazole derivatives, uracil based compounds and also pan-HDAC and HDAC8 selective commercially available compounds were used to study different aspects of HDACs.



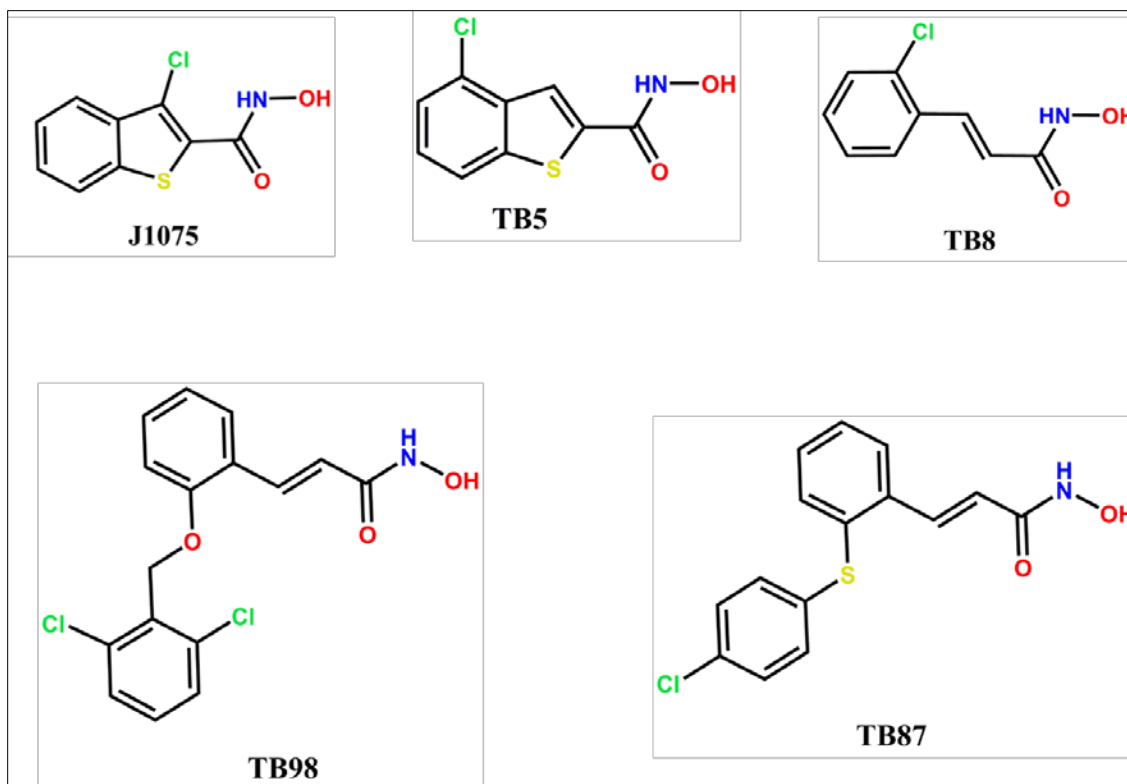


Figure 67: J1075 derived compounds

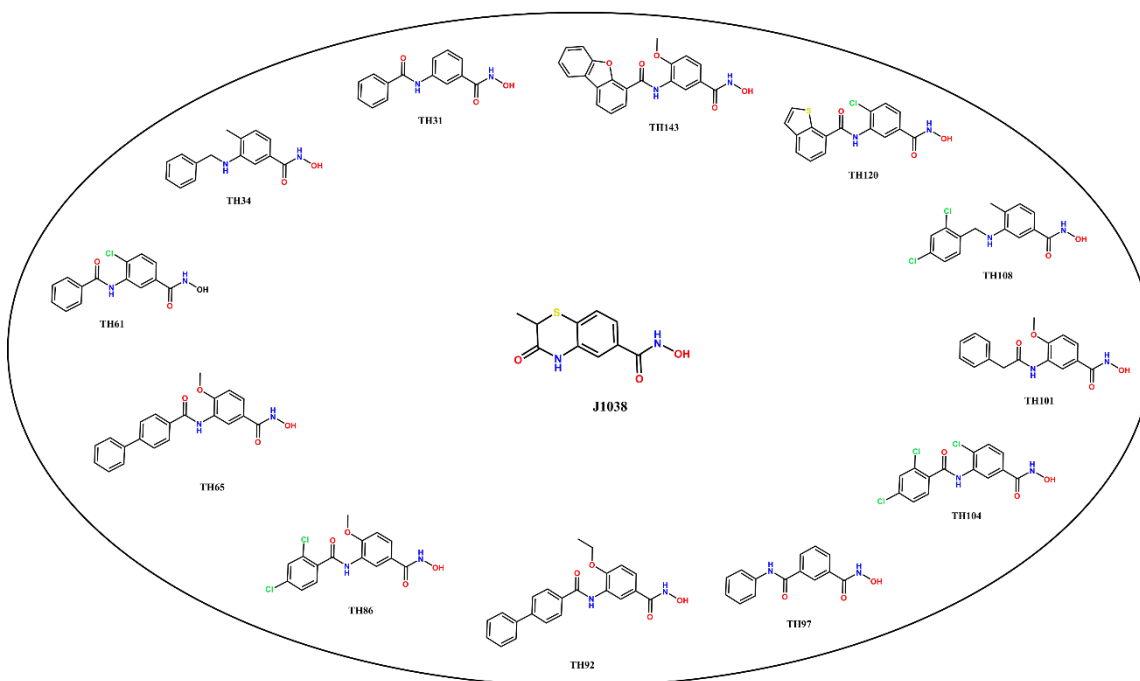


Figure 68: J1038 derived compounds



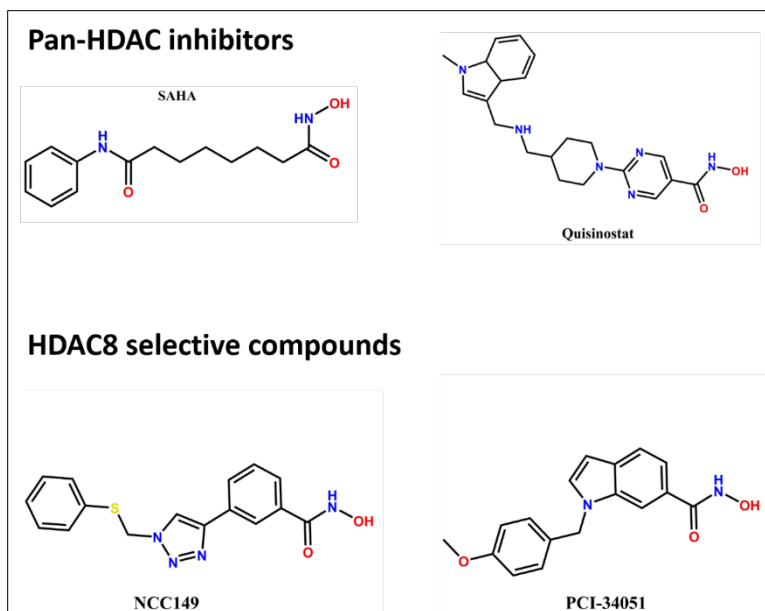


Figure 69: Pan-HDAC and HDAC8 selective inhibitors

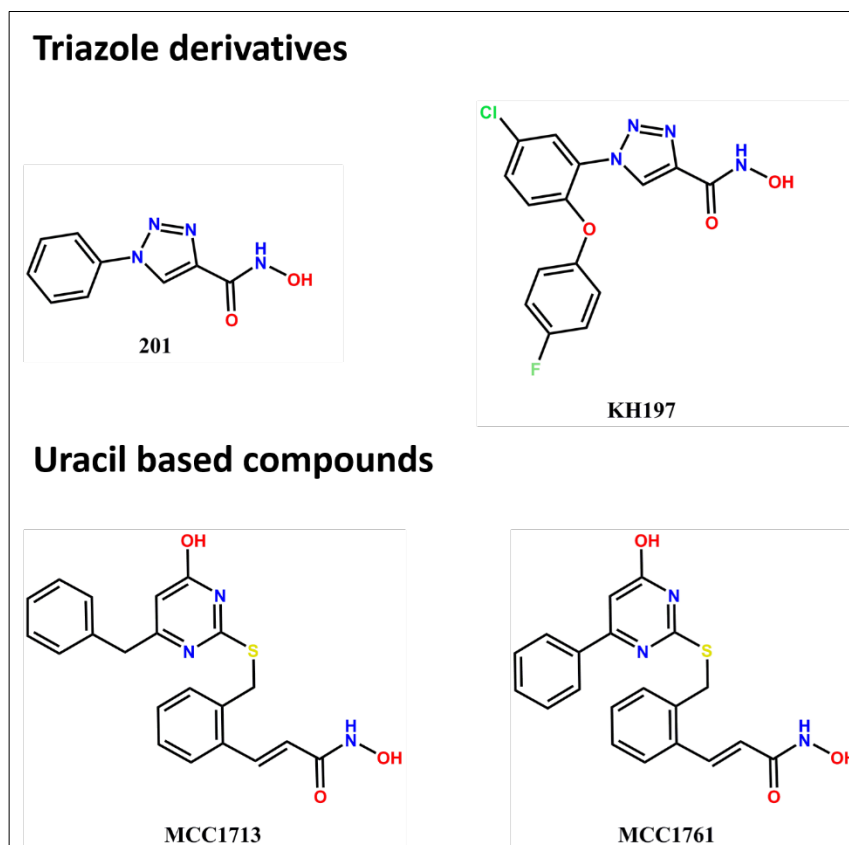


Figure 70: Triazole and Uracil based compounds

### 3.2.2. Publication 1

#### **Summary of publication 1:**

The knowledge of smHDAC8 catalytic pocket structural specificities that we obtained during the SEtTReND project was used to design new smHDAC8 specific inhibitors by our collaborators (Wolfgang Sippl's group at Halle University in Germany). These compounds were tested for inhibition by our other collaborators (Manfred Jung's group at Freiburg University, Germany). In particular, the screening of J1038-derived compounds provided promising results. Particularly, benzhydroxamate compounds having an internal amide group in their linker and with methyl or methoxy substitution at para position were showed to be more potent for smHDAC8 and, for some of them, showed increased selectivity for smHDAC8 over hHDAC8

Our crystal structure of smHDAC8 in complex with the simplest of these derivative TH31 (13a) revealed that these compounds make use of smHDAC8-specific active site specificities for selective inhibition. TH31 was clamped in the active site pocket by two smHDAC8 residues, K20 and H292 which form a 'clamp jaw'. Structural comparison with human HDACs suggest that the clamp jaw formation is smHDAC8 specific. Further, these observations were supported by biochemical data where IC50 values are more favorable towards smHDAC8 over hHDAC1 and 6, and to some extent towards hHDAC8. Cytotoxic studies also suggested that these inhibitors do not have toxic effects on HEK293 cells (human epithelial kidney cell line). These results suggest that the new series of compounds are valid candidates for further optimization.

#### **Contribution:**

For this project, I have produced the smHDAC8 and hHDAC8 enzymes. These enzymes were used by our collaborators for *in vitro* assays to determine IC50 values.

# Structure-Based Design and Synthesis of Novel Inhibitors Targeting HDAC8 from *Schistosoma mansoni* for the Treatment of Schistosomiasis

Tino Heimburg,<sup>†,‡</sup> Alokta Chakrabarti,<sup>‡,‡</sup> Julien Lancelot,<sup>§</sup> Martin Marek,<sup>||</sup> Jelena Melesina,<sup>†</sup> Alexander-Thomas Hauser,<sup>‡</sup> Tajith B. Shaik,<sup>||</sup> Sylvie Duclaud,<sup>||</sup> Dina Robaa,<sup>†</sup> Frank Erdmann,<sup>†</sup> Matthias Schmidt,<sup>†</sup> Christophe Romier,<sup>||</sup> Raymond J. Pierce,<sup>§</sup> Manfred Jung,<sup>‡</sup> and Wolfgang Sippl<sup>\*,†</sup>

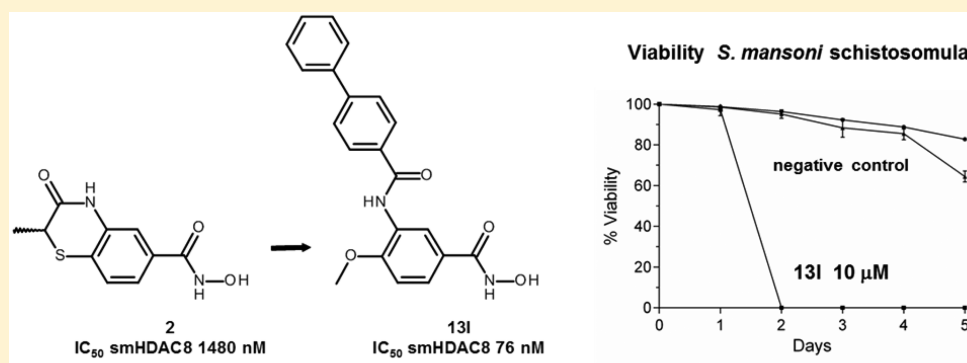
<sup>†</sup>Institute of Pharmacy, Martin-Luther University of Halle-Wittenberg, 06120 Halle/Saale, Germany

<sup>‡</sup>Institute of Pharmaceutical Sciences, University of Freiburg, 79104 Freiburg, Germany

<sup>§</sup>Université de Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 8204 - CIIL - Centre d'Infection et d'Immunité de Lille, 59000 Lille, France

<sup>||</sup>Département de Biologie Structurale Intégrative, Institut de Génétique et Biologie Moléculaire et Cellulaire (IGBMC), Université de Strasbourg (UDS), CNRS, INSERM, 67404 Illkirch Cedex, France

## Supporting Information



**ABSTRACT:** Schistosomiasis is a major neglected parasitic disease that affects more than 265 million people worldwide and for which the control strategy consists of mass treatment with the only available drug, praziquantel. In this study, a series of new benzohydroxamates were prepared as potent inhibitors of *Schistosoma mansoni* histone deacetylase 8 (smHDAC8). Crystallographic analysis provided insights into the inhibition mode of smHDAC8 activity by these 3-amidobenzohydroxamates. The newly designed inhibitors were evaluated in screens for enzyme inhibitory activity against schistosome and human HDACs. Twenty-seven compounds were found to be active in the nanomolar range, and some of them showed selectivity toward smHDAC8 over the major human HDACs (1 and 6). The active benzohydroxamates were additionally screened for lethality against the schistosome larval stage using a fluorescence-based assay. Four of these showed significant dose-dependent killing of the schistosome larvae and markedly impaired egg laying of adult worm pairs maintained in culture.

## INTRODUCTION

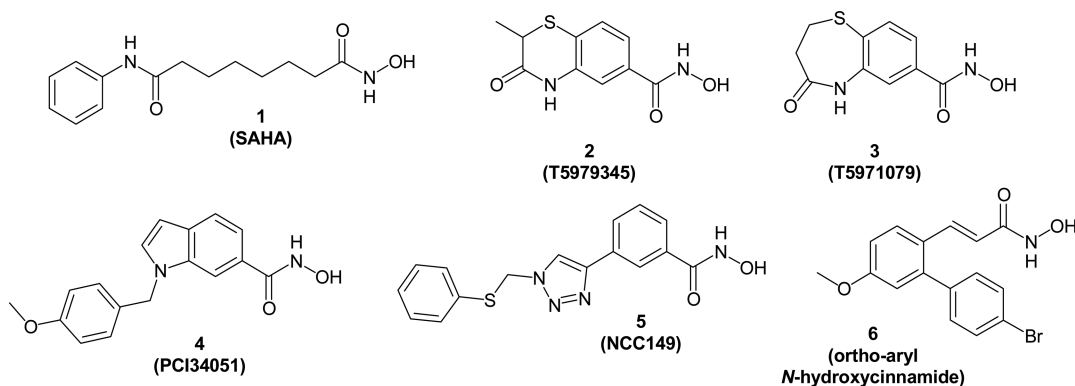
Schistosomiasis is one of the major neglected parasitic diseases,<sup>1</sup> second in importance only to malaria. It is caused by parasites from the genus *Schistosoma*,<sup>2,3</sup> with *Schistosoma mansoni* being the most widely distributed species.<sup>4</sup> Worldwide more than 265 million individuals are infected,<sup>5</sup> of whom 280 000 die annually.<sup>4,6</sup> To date there is no effective vaccine available, and control depends on mass drug administration using the only available treatment, praziquantel, which is effective against all species of human schistosomiasis.<sup>7,8</sup> In 2013 over 30 million people were treated in sub-Saharan Africa,<sup>9</sup> and this raises serious concerns about the selection of drug resistance. The reduced efficiency of praziquantel and the

observed resistance in laboratory strains underline the need to consider alternative therapeutic strategies.<sup>7,8,10,11</sup> Indeed, drug resistance represents an increasing problem for the treatment of a number of parasitic diseases for which only a few drugs are available. Thus, novel potential drug targets and drug candidates against eukaryotic parasites are urgently required.<sup>12</sup>

Histone modifying enzymes (HMEs), which are responsible for post-translational modifications of histone and non-histone substrates, have been reported as drug targets for many diseases such as cancer, inflammation, metabolic diseases, and neuro-

Received: September 23, 2015

Published: March 3, 2016



**Figure 1.** Chemical structures of pan-HDAC inhibitor **1** (SAHA) and reported hHDAC8 inhibitors.

psychiatric disorders as well as in regenerative medicine.<sup>13–16</sup> One of the best-investigated post-translational modifications is acetylation/deacetylation of lysine residues in histone and non-histone proteins. The processes of acetylation and deacetylation are controlled by histone acetyl transferases (HATs) and histone deacetylases (HDACs), respectively. The equilibrium between acetylated and non-acetylated histone proteins must be maintained for proper transcriptional activity and cellular function.<sup>17</sup> Additionally, an increasing number of non-histone proteins have been reported as substrates of HDACs.<sup>18</sup> These proteins may be involved in transcription complexes, which play a pivotal role in the regulation of gene expression as well as cell proliferation, migration, death, and angiogenesis.<sup>18</sup> HDACs are a family of enzymes found in many organisms, including bacteria, fungi, plants, and animals. Eighteen different members of the HDAC family have been annotated in the human genome and have been classified into four categories on the basis of their homology to yeast HDACs.<sup>19</sup> Class I consists of four different subtypes (HDAC1, -2, -3, and -8) and shows homology to the yeast protein RPD3. Class II includes six subtypes, which are divided into two subclasses: class IIa with subtypes HDAC4, -5, -7, and -9 and class IIb with HDAC6 and -10. HDAC11 is placed in class IV. While the activity of the enzymes belonging to class I, II, and IV HDACs depends on a zinc-based catalytic mechanism, class III enzymes, also called sirtuins, use nicotinamide adenine dinucleotide as a cofactor.<sup>20</sup> Hereafter, the term HDAC will refer only to the classical zinc-dependent deacetylases.

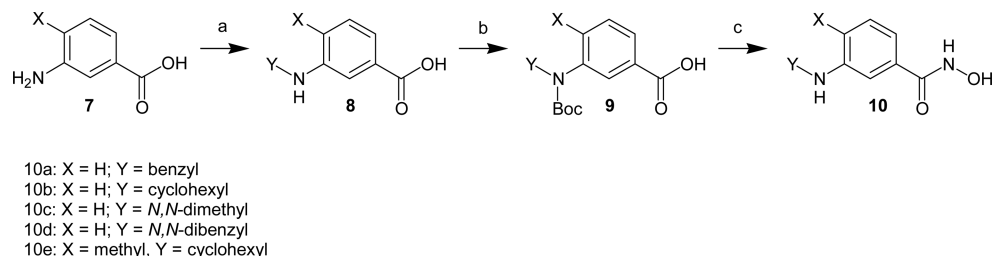
Numerous HDAC inhibitors (HDACi) have been identified in the past decade; several are in clinical trials, and five are currently approved for the treatment of cancer, for example, the aliphatic hydroxamate SAHA (**1**) (Figure 1), marketed under the name Vorinostat.<sup>21–26</sup> The pan-inhibitor **1** is a weak inhibitor of HDAC8 ( $\mu\text{M}$  range). Several aromatic<sup>27–29</sup> and cinnamic acid-based<sup>30</sup> hydroxamates have recently been developed as selective inhibitors of human HDAC8 (**2–6**; Figure 1). The indole derivative **4**<sup>27</sup> is the most selective HDAC8 inhibitor in vitro, with an  $\text{IC}_{50}$  of 10 nM for HDAC8 and selectivity indexes of 290 and 400 for HDAC8 over HDAC6 and HDAC1, respectively. Recent investigations<sup>14,15,31,32</sup> have shown that eukaryotic parasites possess HDAC orthologues and that histone acetylation seems to play a key role in gene-transcription regulation and cell-cycle progression. Also, many human parasites share several characteristics with tumor cells, including high metabolic activity, a dependence on lactate fermentation as an energy source within the human host, uncontrolled cell division, and a

degree of invisibility to the host immune responses.<sup>14</sup> The therapeutic potential of HDAC inhibitors as antiparasitic agents was first shown for the cyclic tetrapeptide apicidin.<sup>31</sup> In addition, several studies using various HDACi demonstrated the antiproliferative and antiparasitic activities of these inhibitors on major human parasitic diseases such as leishmaniasis, malaria, schistosomiasis, toxoplasmosis, and trypanosomiasis.<sup>14,15</sup>

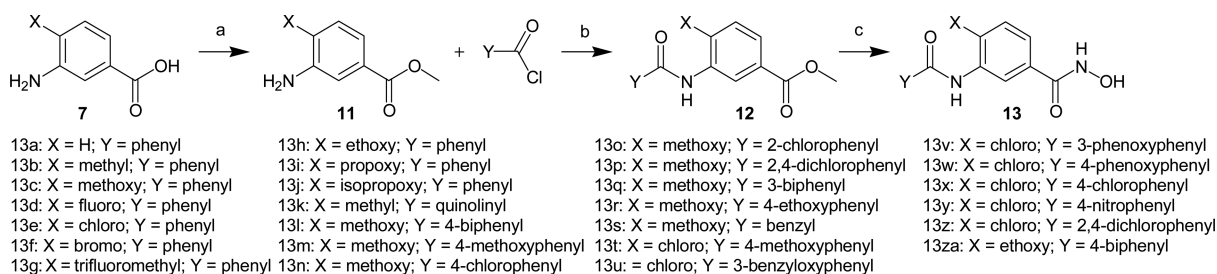
To date, only class I (smHDAC1, -3, and -8) and class III (smSirt1, -2, -5, -6, and -7) HDACs of *S. mansoni* have been cloned and characterized.<sup>33,34</sup> Treatment of schistosomes with generic HDAC inhibitors caused protein acetylation and dose-dependent mortality of schistosome larvae (schistosomula) and adult worms.<sup>33</sup> All three *S. mansoni* class I HDACs (smHDAC1, -3, and -8) are expressed at all life-cycle stages, with HDAC8 transcripts always being the most abundant,<sup>32</sup> indicating that this latter enzyme is a potential target for the design of schistosome-specific inhibitors. This observation was quite surprising because normal levels of HDAC8 transcripts are generally lower than those of HDAC1 and HDAC3 in human cells, with the exception of some cancers, where HDAC8 expression is often strikingly upregulated.<sup>35</sup>

The potential of smHDAC8 as a therapeutic target was supported by biochemical and in vivo assays.<sup>36</sup> RNA interference (RNAi)-mediated downregulation of smHDAC8 expression in schistosomula followed by their intravenous injection into mice and harvesting of the surviving worms 35 days later showed significantly reduced worm recovery compared with that from mice infected with schistosomula treated with control double-stranded RNA.<sup>36</sup> Finally, the crucial roles of zinc-dependent HDACs in schistosome biology were confirmed by the use of small-molecule HDAC inhibitors.<sup>33,36,37</sup> Therefore, a therapy with small-molecule HDACi represents a promising approach for the treatment of schistosomiasis.

In a previous study, we were able to identify the first small-molecule inhibitors of smHDAC8 by a combination of virtual screening and in vitro testing.<sup>38</sup> Two of the identified hits were cocrystallized with smHDAC8, paving the way for structure-based optimization.<sup>36</sup> In the present work, we applied structure-based design on a benzohydroxamate template, taking into consideration appropriate synthetic strategies, to obtain compounds with smHDAC8 inhibitory activity in vitro and antischistosomal activity in cellular assays. A major goal of the current work was to identify compounds that show selectivity for smHDAC8 over major human HDAC isoforms, especially hHDAC1 and hHDAC6.

Scheme 1<sup>a</sup>

<sup>a</sup>Conditions: (a) aldehyde, toluene, Na(AcO)<sub>3</sub>BH, AcOH; (b) Boc<sub>2</sub>O, MeOH, *t*-BuOH; (c) PyBOP, DIPEA, NH<sub>2</sub>OTHP, THF; cat. HCl, THF; TFA, CHCl<sub>3</sub>.

Scheme 2<sup>a</sup>

<sup>a</sup>Conditions: (a) SOCl<sub>2</sub>, MeOH; (b) DIPEA, THF; (c) aq. NaOH sol., MeOH; PyBOP, DIPEA, NH<sub>2</sub>OTHP, THF; cat. HCl, THF.

## RESULTS AND DISCUSSION

**Structural Validation.** From the available X-ray structure of smHDAC8/2<sup>38</sup> (Figure 1) it was known that a hydrogen bond is formed between the amide NH group of the inhibitor and His292.<sup>38</sup> Therefore, open-ring analogues that maintained this hydrogen bond were designed. This resulted in the synthesis of the first series of inhibitors, 3-aminobenzohydroxamates 10a–e (Scheme 1) and 3-amidobenzohydroxamates 13a–c (Scheme 2). In vitro testing showed that all compounds exhibit specificity for the HDAC8 isoforms over the other human HDACs tested (HDAC1 and -6). However, while the 3-aminobenzohydroxamates (10a–e) are micromolar inhibitors of both human and schistosomal HDAC8, they show a significant preference for the human isoform (Table 1). In contrast, the 3-amidobenzohydroxamates (13a–c) are active in the nanomolar range and show very similar inhibitory activities against hHDAC8 and smHDAC8. These results prompted us to focus on the derivatization of 13a–c, since this seemed to be a more promising strategy to obtain selective compounds.

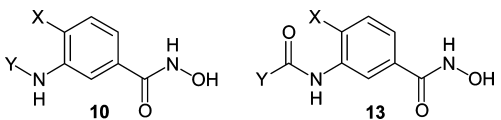
However, prior to the synthesis of derivatives of this series, we looked at the binding mode of this scaffold to smHDAC8. Toward this aim, the crystallographic structure of the complex between smHDAC8 and compound 13a at 2.2 Å was solved and refined (Table S1 in the Supporting Information). The crystal structure of the smHDAC8/13a complex reveals that the inhibitor binds in the smHDAC8 active-site pocket, forming specific interactions with the protein (Figure 2). First and as expected, the hydroxamate warhead of 13a interacts with the catalytic zinc ion and three residues, namely, His141, His142, and Tyr341 (Figure 2). The last of these residues adopts the flipped-in conformation typically observed in most HDAC/hydroxamate complexes.<sup>11</sup> Second, reminiscent of what was observed in the smHDAC8/2 complex,<sup>36</sup> smHDAC8 H292 is able to interact with inhibitor 13a, with its side chain forming a hydrogen bond (3.0 Å) with the amine group of the amide (Figures 2 and 3).

Strikingly, two additional smHDAC8-specific features were observed in this complex. First, and in contrast to what was observed for the smHDAC8/2 complex,<sup>36</sup> smHDAC8 Phe151 is observed to be in its flipped-out position (Figures 2 and 3) despite the fact that binding of 13a would not prevent the Phe151 side chain from adopting a flipped-in conformation. A major consequence of this flipped-out conformation of Phe151 is that Lys20 also adopts a flipped-in conformation, with the aliphatic part of its side chain lying on the Phe151 side chain and its amine forming a hydrogen bond (2.8 Å) with the carbonyl oxygen of the 13a amide group.

Interestingly, this is the first time that Lys20 has been observed to interact clearly with the inhibitor in an smHDAC8/inhibitor complex. Such a conformation had only been observed previously in the noninhibited structure of smHDAC8 in which Lys20 also interacted with an *L*-tartrate molecule that was present in the crystallization buffer and was observed to bind to the catalytic zinc (Figure 3). In all of the other smHDAC8/inhibitor complexes solved to date, the Lys20 side chain was either not seen in density or was prevented from reaching into the active site by the flipped-in conformation adopted by Phe151.<sup>39</sup>

In addition to these hydrogen bonds formed between smHDAC8 and 13a, hydrophobic contacts between the two molecules are also observed, building on the overall hydrophobic character of the smHDAC8 active site and further stabilizing the smHDAC8/13a interaction. However, the schistosome-specific clamp formed by smHDAC8 Lys20 and His292, which distinctively interacts with the 13a amide group, helps anchor 13a in the enlarged (because Phe151 is flipped out) smHDAC8 active site, thus suggesting the molecular basis for the improved inhibitory activity of the 3-amidobenzohydroxamates toward smHDAC8.

**Synthesis.** We then continued the optimization of the 3-amidobenzohydroxamate inhibitors. Toward this aim and to guide the optimization process, docking studies using the

Table 1. IC<sub>50</sub> Values for 3-Aminobenzohydroxamate Derivatives


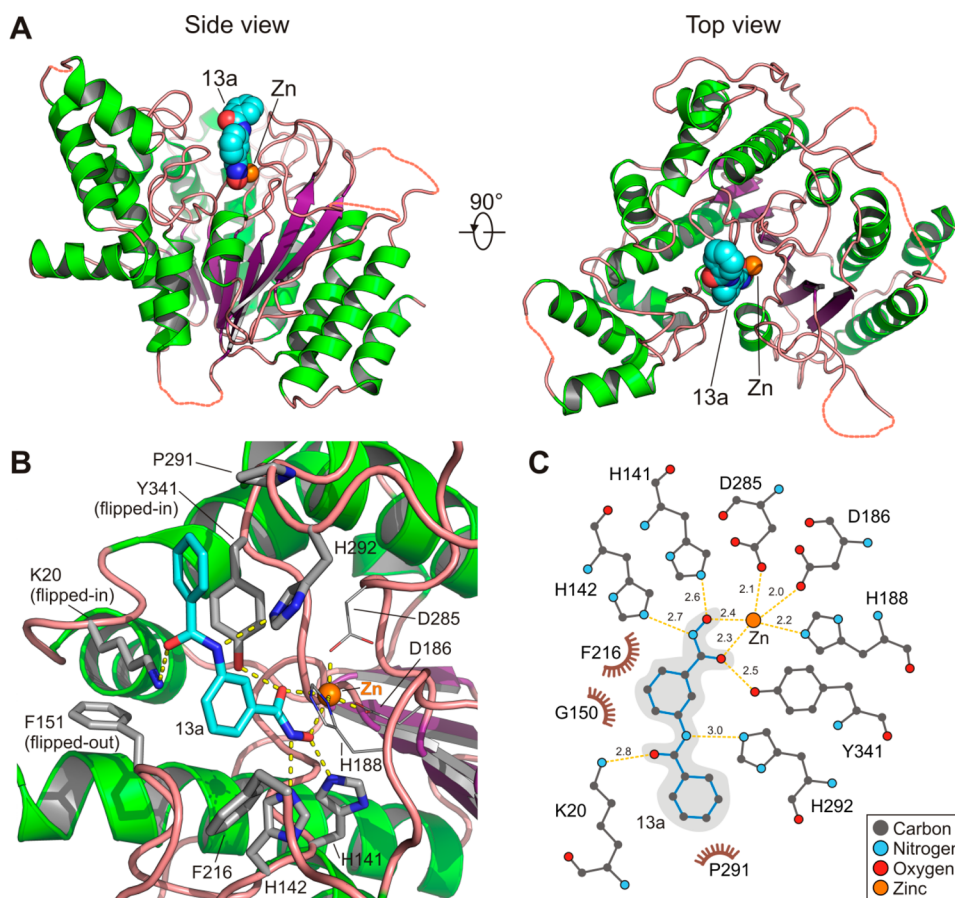
compound	X	Y	smHDAC8 IC <sub>50</sub> (nM)	hHDAC8 IC <sub>50</sub> (nM)	hHDAC1 IC <sub>50</sub> (μM)	hHDAC6 IC <sub>50</sub> (μM)
10a	H	benzyl	1080 ± 250	143.4 ± 7.3	41.4% @ 10 μM	n.d.
10b	H	cyclohexyl	3630 ± 620	830 ± 40	129.9 ± 18.5	n.d.
10c	H	<i>N,N</i> -dimethyl	1576 ± 146.0	70.2 ± 10.8	37.4% @ 10 μM	n.d.
10d	H	<i>N,N</i> -dibenzyl	9290 ± 1500	2190 ± 420	42.3 ± 3.8	8.6 ± 1.9
10e	methyl	cyclohexyl	600.3 ± 195.5	104.3 ± 12.0	49.5% @ 10 μM	n.d.
13a	H	phenyl	468.2 ± 79.0	582.0 ± 48.0	33.6 ± 1.8	3.0 ± 0.3
13b	methyl	phenyl	116.2 ± 38.2	204.0 ± 22.0	8.4 ± 2.0	0.9 ± 0.4
13c	methoxy	phenyl	189.8 ± 54.2	88.3 ± 24.0	2.3 ± 1.2	2.5 ± 1.1
13d	fluoro	phenyl	177.6 ± 8.1	317.8 ± 54.2	22.3 ± 7.7	0.50 ± 0.01
13e	chloro	phenyl	67.0 ± 10.2	120.0 ± 36.7	11.6 ± 3.9	0.12 ± 0.02
13f	bromo	phenyl	150.4 ± 8.5	191.4 ± 26.0	7.4 ± 0.8	0.15 ± 0.01
13g	trifluoromethyl	phenyl	139.6 ± 8.3	342.2 ± 76.1	2.4 ± 0.1	0.14 ± 0.02
13h	ethoxy	phenyl	129.3 ± 7.6	171.5 ± 15.6	4.6 ± 0.3	1.3 ± 0.1
13i	propoxy	phenyl	266.9 ± 49.5	n.d.	n.d.	n.d.
13j	isopropoxy	phenyl	220.1 ± 56.2	29.0 ± 0.2	3.6 ± 0.3	2.6 ± 0.4
13k	methyl	2-quinolinyl	96.1 ± 13.7	30.3 ± 7.3	2.7 ± 0.8	0.09 ± 0.01
13l	methoxy	4-biphenyl	75.4 ± 25.5	26.1 ± 17.6	6.3 ± 2.1	0.390 ± 0.002
13m	methoxy	4-methoxyphenyl	106.0 ± 17.5	77.1 ± 10.6	2.6 ± 0.2	0.4 ± 0.1
13n	methoxy	4-chlorophenyl	146.0 ± 4.3	239.7 ± 96.0	2.9 ± 0.3	0.9 ± 0.1
13o	methoxy	2-chlorophenyl	699.3 ± 27.4	211.16 ± 27.59	9.6 ± 1.0	3.1 ± 0.7
13p	methoxy	2,4-dichlorophenyl	121.6 ± 18.7	548.3 ± 93.9	13.0 ± 1.9	2.3 ± 0.4
13q	methoxy	3-biphenyl	289.7 ± 20.0	n.d.	n.d.	n.d.
13r	methoxy	4-ethoxyphenyl	305.0 ± 35.0	438.4 ± 48.0	4.4 ± 0.6	1.0 ± 0.1
13s	methoxy	benzyl	182.7 ± 39.3	512.2 ± 29.8	28.9 ± 8.6	5.1 ± 0.7
13t	chloro	4-methoxyphenyl	147.1 ± 4.8	235.6 ± 49.5	4.1 ± 0.9	0.13 ± 0.01
13u	chloro	3-benzyloxyphenyl	378.1 ± 44.9	214.4 ± 27.0	9.4 ± 2.8	1.5 ± 0.1
13v	chloro	3-phenoxyphenyl	396.4 ± 43.3	448.6 ± 100.4	6.4 ± 0.7	0.3 ± 0.1
13w	chloro	4-phenoxyphenyl	979.1 ± 1100	1080 ± 300	8.5 ± 2.1	0.15 ± 0.01
13x	chloro	4-chlorophenyl	234.7 ± 10.3	292.0 ± 53.3	3.8 ± 0.2	0.09 ± 0.05
13y	chloro	4-nitrophenyl	393.6 ± 50.5	n.d.	n.d.	n.d.
13z	chloro	2,4-dichlorophenyl	191.4 ± 16.7	1184.0 ± 45.1	31.6 ± 19.8	0.8 ± 0.1
13za	ethoxy	4-biphenyl	92.0 ± 26.0	148.7 ± 22.7	2.08 ± 0.14	0.6 ± 0.1
14a			8205 ± 1300	582.3 ± 88.5	n.d.	n.d.
15a			268.2 ± 21.1	23.9 ± 4.7	12.1 ± 5.7	2.9 ± 0.3
16a			485.0 ± 158.2	19.8 ± 5.9	20.0 ± 5.9	2.4 ± 0.9
17a			n.a.	n.a.	n.a.	n.a.
17b			n.a.	n.a.	n.a.	n.a.
1			1560 ± 200	400 ± 100	0.117 ± 0.006	0.042 ± 0.011
2			1480 ± 460	970 ± 110	27.5 ± 8.3	3.6 ± 0.6
3			1220 ± 280	620 ± 80	9.8 ± 0.7	1.3 ± 0.2
4			435.6 ± 61.0	77.7 ± 28.1	48% @ 100 μM	41% @ 100 μM

available inhibited hHDAC8 and smHDAC8 structures were carried out. Examination of the crystal structures showed that there is the possibility to use more bulky and more lipophilic residues at position 4 of the benzohydroxamate moiety. Recognizing that physicochemical properties might also play an important role in the antischistosomal activity, lipophilic substituents were also included. Taking this into consideration, we synthesized compounds **13d–j** containing different halides and alkoxy groups at position 4 to examine the effects of these substituents on the activity and selectivity for smHDAC8 (Scheme 2). The compounds containing halogen at position 4 of the benzohydroxamate moiety (**13d–f**) were slightly more selective for smHDAC8 compared with the 4-methoxy derivative **13c**. Compound **13h** with an ethoxy group at

position 4 did not show an increase in the activity against smHDAC8 but was less active on hHDAC8 compared with **13c**.

Meanwhile, analogues containing more lipophilic alkoxy residues at position 4, such as **13i** and **13j** (Scheme 2), exhibited decreased activity and selectivity for smHDAC8 compared with **13h**. Further modifications were introduced on the benzamide moiety, including chloro, nitro, and alkoxy groups (**13m–p**, **13r**, **13t**, **13x–z**). In addition, the introduction of aromatic lipophilic residues to address the hydrophobic side pocket of smHDAC8 (**13k**, **13l**, **13q**, **13u–w**, **13za**) was investigated. Compound **13k** bearing a quinolinyl residue and compound **13l** bearing a 4-biphenyl residue showed increased activity against smHDAC8 and hHDAC8.





**Figure 2.** Overall structure of inhibitor **13a** bound to smHDAC8. (A) Structure of smHDAC8 (shown in ribbon representation) with bound **13a** (shown in space-filling representation). The orange sphere represents the catalytic zinc ion. (B) Close-up view of the binding mode of **13a** in the smHDAC8 active-site pocket. Protein residues are shown as gray sticks, and **13a** is shown as cyan sticks. Yellow dashed lines represent salt bridges made by smHDAC8, **13a**, and the catalytic zinc ion. (C) LigPlot-generated two-dimensional schematic overview of molecular interactions between **13a** and smHDAC8 active-site zinc and protein residues. Hydrogen bonds and interactions are indicated by yellow dashed lines, and the corresponding distances between the atoms (in Å) are given. Hydrophobic contacts are shown by brown arcs with spokes radiating toward the atoms involved.

Substitution at the meta position (as in **13q**) or the ortho position (as in **13o**) decreased the activity on the tested enzymes.

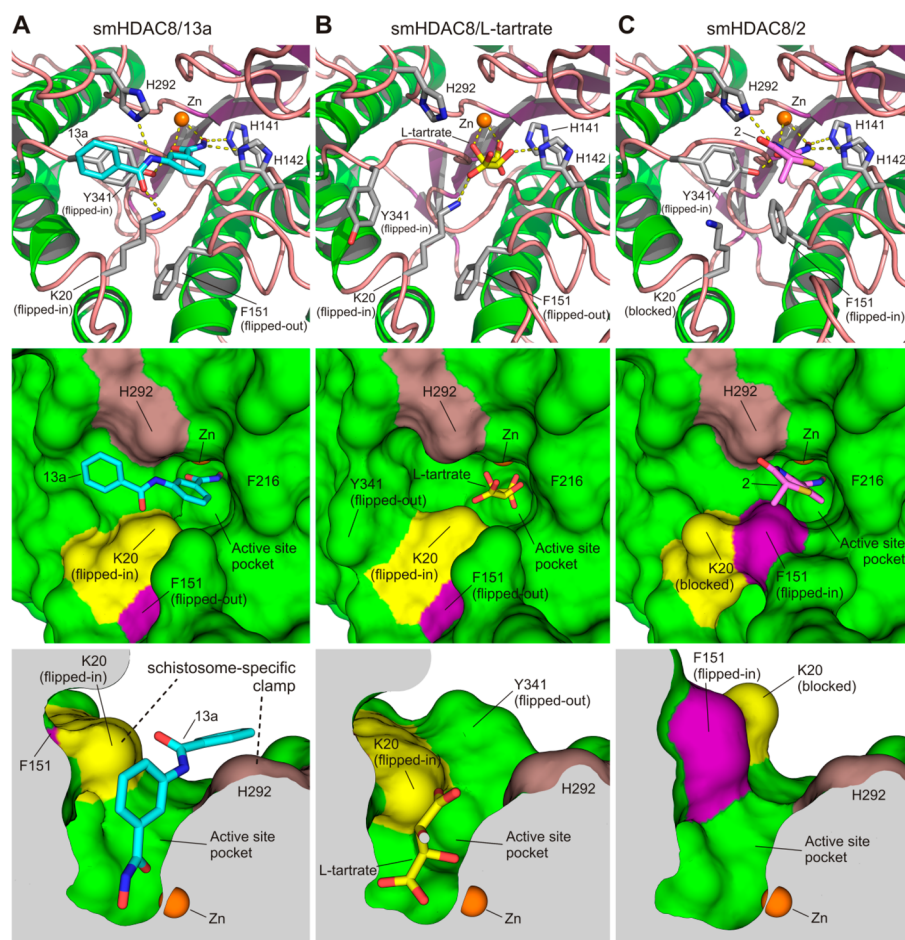
Interestingly, the combination of para and ortho substitution (as in **13p**) restored the activity against smHDAC8 without increasing the activity against the human enzymes. To prove that this substitution pattern is important for the selectivity of this compound, we synthesized compound **13z**, which was indeed found to be more active on the schistosomal enzyme than on the human counterpart. Increasing the distance between the two aromatic rings by a further methylene group (**13s**) led to increased smHDAC8 selectivity.

In addition, we tested other linker groups between the two aromatic rings (Scheme 3). The introduction of an ether (**15a**) or sulfonamide (**16a**) resulted in loss of activity for smHDAC8 and increased activity for hHDAC8 compared with **13c**. Also, using another scaffold such as a condensed aromatic ring system (**14a**) did not result in an improvement of the activity or selectivity compared with **13c**.

To check the impact of the zinc-chelating moiety, we synthesized two analogues of **13l** containing a carboxylate or carboxyl ester group instead of the hydroxamate (**17a** and **17b**, respectively; Scheme 4). As expected neither compound showed any effect in the enzymatic HDAC assay.

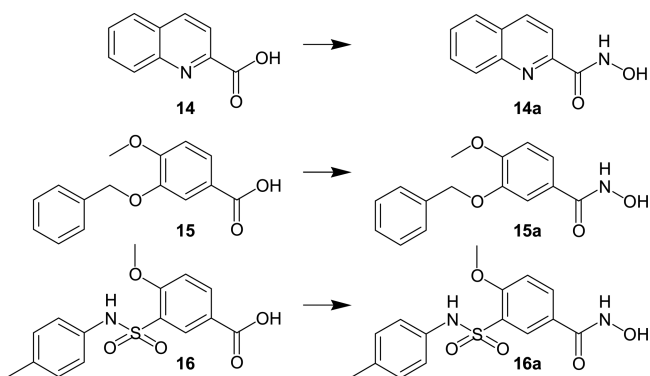
**Docking Studies.** To rationalize the obtained biochemical data, notably to understand the change in specificity between the schistosomal and human enzymes, the synthesized inhibitors were docked to the available crystal structures of smHDAC8, hHDAC8, and hHDAC1 and a homology model of hHDAC6. The applied docking method (for details, see Experimental Methods) was first successfully validated on the X-ray structures of hHDAC8 and smHDAC8. By means of this docking setup, consistent binding models were derived for both hHDAC8 and smHDAC8. In the case of smHDAC8, all of the derivatives having an amide linker between the two aromatic rings showed hydrogen bonds to Lys20 and His292, as observed for **13a** in its crystal structure.

Most importantly, the hydrogen bond between the amide linker and His292 cannot be formed in hHDAC8 since His292 is replaced by a methionine in this latter enzyme. However, in the available crystal structures of hHDAC8 a conserved water molecule bound to the zinc-coordinating histidine (His180 in hHDAC8) is observed, which was found to be a hydrogen-bonding partner with most of the amides in the docking studies (Figure 4). Thus, the observed hydrogen bond in the case of hHDAC8 (as well as in the homology model of HDAC6) could partially explain the same range of activity of some of the



**Figure 3.** Specific structural interactions observed in the smHDAC8 active-site pocket in (A) the smHDAC8/13a complex, (B) the smHDAC8/*L*-tartrate complex, and (C) the smHDAC8/2 complex. Upper panels: close-up views with protein and small molecules shown as ribbon and sticks, respectively. Middle panels: surface representations. Lower panels: cutaway surface representations of the active site. For clarity, in (C) the 2 molecule was removed from the cutaway surface representation. In the smHDAC8/13a complex structure in (A), specific structural arrangements are observed, notably involving two specific residues: Lys20 and Phe151. The flipped-out conformation of Phe151 enables Lys20 to flip in and contact inhibitor 13a. This unique active-site conformation is not observed in other smHDAC8/inhibitor complexes. In addition, the schistosome-specific residue His292 together with the flipped-in Lys20 form a clamp that promotes stabilization and binding of 13a in the smHDAC8 active-site pocket.

### Scheme 3<sup>a</sup>



<sup>a</sup>Conditions: PyBOP, DIPEA, NH<sub>2</sub>OTHP, THF; cat. HCl, THF.

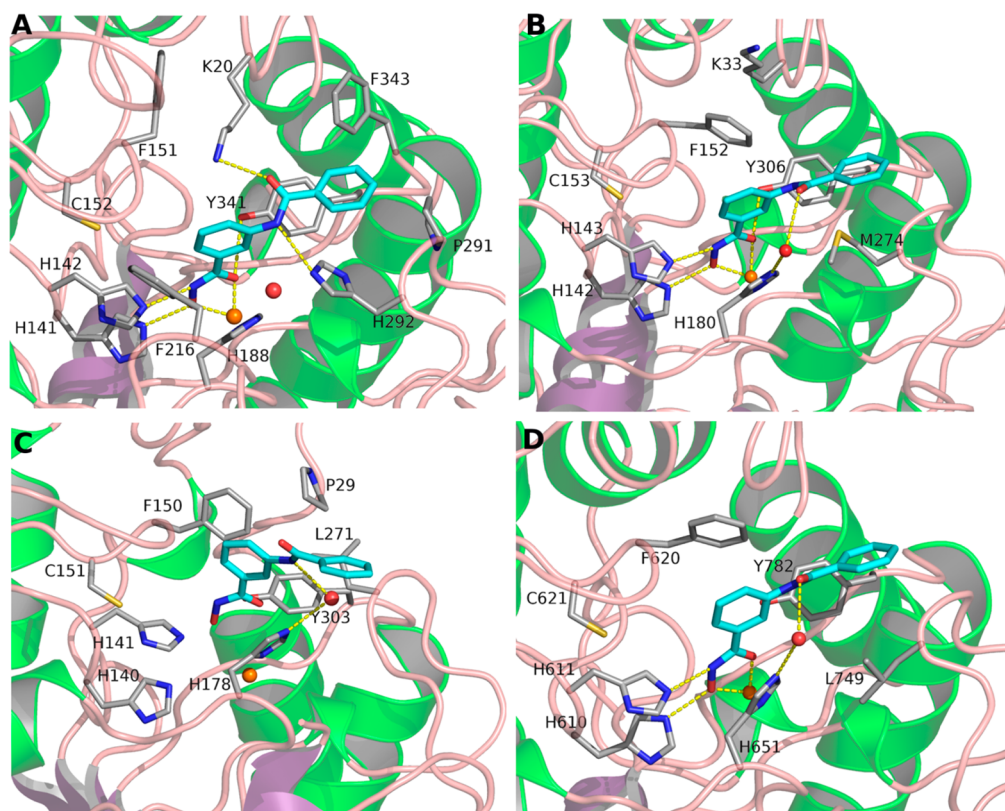
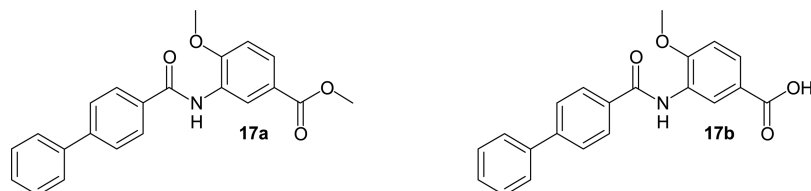
compounds on the human HDAC8 and HDAC6 isoforms as on smHDAC8.

Adding a substituent at the para position of the parent compound 13a (methyl, alkyloxy, or halide) results in a 2- to 7-fold gain in smHDAC8 inhibitory potency and a 2- to 20-fold

gain in hHDAC8 potency (e.g., 13b, 13c, and 13e compared with 13a; Table 1). Increasing the size of the substituent at the para position enables additional interactions at the entrance of the pocket. Furthermore, the nature of the para substituent might influence the biologically active conformation of the compounds, which affects not only the potency but also the selectivity. In the case of compound 13a, the X-ray structure shows an out-of-plane orientation of the amide linker ( $\Phi = -60^\circ$  to  $-102^\circ$ ; Table S2 in the Supporting Information) that is stabilized by two hydrogen bonds to Lys20 and His292 (Figure 2). A substituent at the para position of the benzohydroxamate favors the out-of-plane conformation of the amide linker, which might explain the higher inhibitory potency of the para-substituted compounds. Measuring the dihedral angle  $\Phi$  between the amide linker and the first aromatic ring observed in the docking poses shows a clear preference for the out-of-plane conformation of the amide group ( $-77^\circ$  to  $-90^\circ$ ; Table S2). In the case of hHDAC8, the predicted conformation of the amide linker is close to a coplanar orientation ( $29$ – $51^\circ$ ,  $148^\circ$  for 13g) because of the modified hydrogen-bonding pattern (Figure 4).



Scheme 4



**Figure 4.** Docking poses of compound 13a (cyan carbon atoms) in (a–c) the X-ray structures of (a) smHDAC8 (this study), (b) hHDAC8 (PDB ID 2V5X), and (c) hHDAC1 (PDB ID SBKX) and (d) the homology model of hHDAC6. The protein backbone is depicted in ribbon representation, and the side chains of important residues are shown in stick representation with carbons colored gray. The conserved water molecule that should help stabilize the binding of 13a in the active sites of hHDAC6 and hHDAC8 is shown as a red ball, and the zinc ion is shown as an orange ball. Hydrogen bonds and interactions with the metal ion are depicted as yellow dashed lines.

The decreased HDAC1 activity might be attributed to the narrower pocket of hHDAC1 in comparison to hHDAC8 and -6. In the docking poses, this can be clearly seen in a comparison of the distances from the zinc-binding groups of meta-substituted benzohydroxamates to the zinc ion. In the case of hHDAC1, the distance is higher than in the other investigated HDAC isoforms (Figure 4). The docking results and structure–activity relationship studies suggest that meta substitution of benzohydroxamates is important to gain HDAC8 selectivity. This is also supported by previous publications showing that meta-substituted benzohydroxamic acids are more active on human HDAC8 and -6 in comparison with HDAC1 and other class I HDACs.<sup>28,30,40–45</sup>

Adding hydrophobic substituents such as halides or further aromatic rings to the second aromatic ring (e.g., 13k, 13l, 13p, 13z, and 13za) resulted in additional van der Waals interactions with Phe216, Pro291, and Phe343 of smHDAC8. However, the higher selectivity of 13s and 13z (3–6-fold hHDAC8/smHDAC8 selectivity, 4–28-fold hHDAC6/smHDAC8 selectivity) could not be explained on the basis of the derived

docking solutions. More sophisticated methods that take into account protein flexibility and binding free energy calculations might be helpful to track down the subtle differences in the protein–ligand interactions deep inside the binding pocket.

**Phenotypic Response.** We next analyzed the effect of the developed compounds on the parasites maintained in culture. The compounds were initially tested for their toxicity toward *S. mansoni* schistosomula using an Alamar Blue-based viability assay (see Experimental Methods). Initial testing was done at a concentration of 10  $\mu$ M, and selected compounds were also tested at 20  $\mu$ M in order to determine the dose dependence. Two biological replicates were carried out in triplicate, and the results are shown in Table 2. In addition to the compounds developed during this study, the selective HDAC8 inhibitor 4 and praziquantel, the drug used for treatment of schistosomiasis, were also included in the assay. Of the tested compounds, 13l and 13za provoked the most marked dose-dependent reductions in schistosomula viability. In this assay, 4 showed only very modest activity against schistosomula, and praziquantel was inactive at the concentrations used. This latter

**Table 2. Toxicity Studies on *S. mansoni* Schistosomula (Alamar Blue Assay)**

compound	% viability $\pm$ SEM	
	10 $\mu$ M	20 $\mu$ M
10b	82.7 $\pm$ 2.4	n.d.
10c	92.6 $\pm$ 7.4	n.d.
10d	77.9 $\pm$ 2.6	74.6 $\pm$ 0.4
10e	84.3 $\pm$ 1.7	n.d.
13a	79.2 $\pm$ 0.5	75.6 $\pm$ 5.1
13b	82.9 $\pm$ 2.0	72.4 $\pm$ 2.4
13c	77.4 $\pm$ 3.2	74.8 $\pm$ 2.3
13d	82.0 $\pm$ 8.1	78.6 $\pm$ 9.5
13e	84.7 $\pm$ 1.4	n.d.
13f	70.8 $\pm$ 4.4	63.1 $\pm$ 1.7
13g	99.9 $\pm$ 0.8	77.4 $\pm$ 3.4
13h	89.7 $\pm$ 10.3	n.d.
13j	75.0 $\pm$ 0.8	68.7 $\pm$ 0.7
13k	74.9 $\pm$ 4.1	59.4 $\pm$ 4.2
13l	64.8 $\pm$ 3.5	33.4 $\pm$ 3.5
13m	74.6 $\pm$ 6.7	64.6 $\pm$ 1.0
13o	89.5 $\pm$ 7.2	66.5 $\pm$ 0.8
13p	77.8 $\pm$ 2.4	50.7 $\pm$ 2.7
13q	58.8 $\pm$ 1.6	57.0 $\pm$ 1.2
13s	83.3 $\pm$ 4.2	82.8 $\pm$ 5.8
13t	63.2 $\pm$ 0.5	60.7 $\pm$ 0.3
13u	73.8 $\pm$ 5.4	68.6 $\pm$ 2.4
13v	62.9 $\pm$ 2.7	55.0 $\pm$ 3.9
13x	71.8 $\pm$ 4.1	49.3 $\pm$ 0.6
13y	85.3 $\pm$ 6.8	68.1 $\pm$ 1.2
13z	72.5 $\pm$ 0.3	59.3 $\pm$ 9.9
13za	77.9 $\pm$ 3.9	35.8 $\pm$ 0.8
14a	95.6 $\pm$ 4.4	n.d.
16a	77.7 $\pm$ 5.0	73.8 $\pm$ 4.2
4	70.6 $\pm$ 0.5	62.3 $\pm$ 3.5
praziquantel	92.0 $\pm$ 7.5	89.0 $\pm$ 5.7

result is in line with previous findings<sup>46</sup> and was thought to be due in part to the relatively weak activity of praziquantel on schistosomula and to stimulation of enzyme or ion channel activity, leading to high fluorescence signals in the assay. In view of its selectivity of inhibition with regard to human HDAC1 and -6, inhibitor **13l** was chosen for further testing. We first showed that the EC<sub>50</sub> value for this compound using the Alamar Blue-based assay was 16.1  $\mu$ M (Figure S1 in the Supporting Information). We next showed that compound **13l** is lethal to schistosomula in a microscopy-based assay within 2 days of incubation at 10  $\mu$ M and 1–2 days at 20  $\mu$ M (Figure 5A,B). We further tested **17a** and **17b** (Scheme 4), derivatives of **13l** that were synthesized as negative controls, to check **13l** for off-target effects. In the same assay, compound **17a**, an analogue of **13l** with a carboxyl ester instead of the hydroxamate that showed no inhibitory activity on smHDAC8, had only a very minor effect on schistosomula after 5 days of incubation compared with the dimethyl sulfoxide (DMSO) solvent control (Figure 5A,B). This suggested that the activity of **13l** on schistosomula is indeed related to its capacity to inhibit smHDAC8. We finally tested compound **13l** for its capacity to affect adult schistosomes maintained in culture (Figure 5C,D). At concentrations of 10 and 20  $\mu$ M, **13l** caused a marked separation of adult male and female worm pairs, with 90% of the pairs being separated after 5 days in the presence of 20  $\mu$ M **13l**. A corresponding reduction in egg laying by these

worm pairs was also induced (Figure 5D), reaching 80% for the 20  $\mu$ M dose. Therefore, compound **13l** affects the viability of both larvae and adult worms of *S. mansoni*, most probably through the inhibition of smHDAC8.

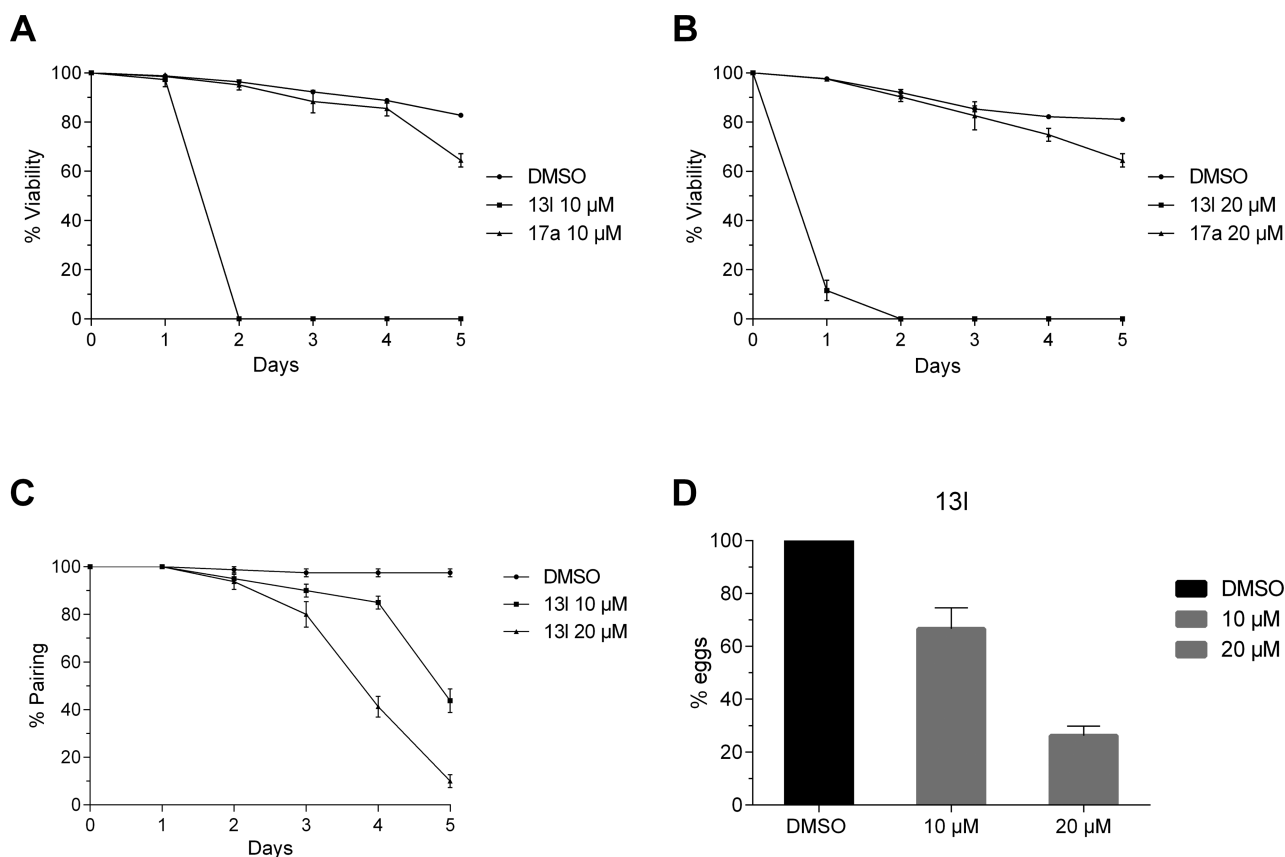
**Cytotoxicity Assay.** It was important to test the selectivity of the compounds against smHDAC8 and to exclude possible toxic effects caused by targeting of human HDACs or other proteins. Therefore, a cytotoxicity assay in a human epithelial kidney cell line (HEK293) was performed. The cells were incubated for 45 h with the indicated compounds at a concentration of 50  $\mu$ M, and the cell viability was determined using the Alamar Blue assay. All of the tested inhibitors exhibited only relatively low cytotoxicity in the human cell system used (Table 3).

## CONCLUSIONS

Our initial work on HDAC8 from *Schistosoma mansoni* provided a proof of concept that HDAC inhibitors of this enzyme could be used to target pathogens.<sup>36</sup> Here a weak screening hit with suboptimal physicochemical properties characterized in that initial study was optimized against HDAC8 from *S. mansoni* using critical structure-guided insights. Central to this optimization was the inclusion of a methyl/methoxy group at the para position and an amide linker at the meta position of the benzohydroxamate. Previous crystallographic studies highlighted the binding of the inhibitors at the acetyllysine tunnel and featured a flexible phenylalanine that is able to shift in response to binding of compound to smHDAC8. The series disclosed here builds on this initial structural feature and represents a novel smHDAC8 inhibition template that provides the possibility to develop potent and selective inhibitors for the therapy of schistosomiasis. The presented compounds demonstrated high selectivity for smHDAC8 over the major human HDAC isoforms HDAC1 and -6, and some compounds even showed a preference for smHDAC8 over its human orthologue hHDAC8. It has been reported that inhibition of human HDAC8 shows limited effects on many cell types<sup>39</sup> and that an HDAC8 inhibitor had the most limited effect on the human acetylome among a panel of inhibitors of HDACs with different selectivities.<sup>47</sup> Cytotoxicity studies of the tested compounds showed that the compounds exhibit a relatively low effect on cell proliferation, indicating that the inhibition of human HDAC8 does not result in intrinsic toxicity. Thus, while we recognize that selectivity over human HDAC8 still needs optimization, there are strong indications that the high selectivity with respect to hHDAC1 and -6 that we have already obtained is more important for a potential therapeutic setting. The most potent derivatives were also shown to impair the viability of schistosomula without affecting cell viability of HEK293 cells. One such compound, **13l**, killed schistosomula in vitro and caused significant separation of adult worm pairs and a significant decrease in egg laying. An analogue of **13l** without inhibitory activity toward smHDAC8 had no effect on the parasite, confirming that inhibition of smHDAC8 is the basis of the antiparasitic effects of these inhibitors and underlining their potential as antischistosomal drug leads.

## EXPERIMENTAL METHODS

**Synthetic Chemistry.** Unsubstituted and 4-substituted 3-amino-benzoic acids were used as starting points for the synthesis of the inhibitors under study. Alkyl and aryl residues were introduced on the aromatic NH<sub>2</sub> group via reductive amination of the imines obtained by



**Figure 5.** (A) Viability assay of *S. mansoni* schistosomula with up to 5 days of incubation with 10 μM 13l or 17a compared to the DMSO solvent control. (B) The same assay with 20 μM 13l or 17a. (C) Separation of adult worm pairs for up to 5 days in culture in the presence of 10 or 20 μM 13l. (D) Cumulative reduction (%) in egg laying by adult worm pairs in culture in the presence of 10 or 20 μM 13l compared with the DMSO solvent control.

**Table 3. Cytotoxicity Studies in HEK293 Cells**

compd	% viability <sup>a</sup> (%)	compd	% viability <sup>a</sup> (%)	compd	% viability <sup>a</sup> (%)
10a	n.d. <sup>b</sup>	13h	70.4	13t	62.5
10b	70.7	13i	97.4	13u	74.4
10c	92.7	13j	128.6	13v	62.0
10d	7.7	13k	66.2	13w	n.d.
10e	70.2	13l	12.3 (199 μM) <sup>c</sup>	13x	81.6
13a	72.0	13m	61.9	13y	79.1
13b	71.2	13n	92.6	13z	70.4
13c	67.3	13o	96.9	13za	80.3
13d	81.8	13p	95.1	14a	73.3
13e	72.2	13q	76.2	15a	69.1
13f	64.3	13r	87.7	16a	80.2
13g	96.0	13s	88.7	17a	47.2

<sup>a</sup>Percent viability of cells in the presence of 50 μM compound in comparison with an untreated sample. Daunorubicin was used as a positive control, and an IC<sub>50</sub> value of 12.55 ± 0.07 μM was determined. <sup>b</sup>n.d. = not determined. <sup>c</sup>IC<sub>50</sub> value.

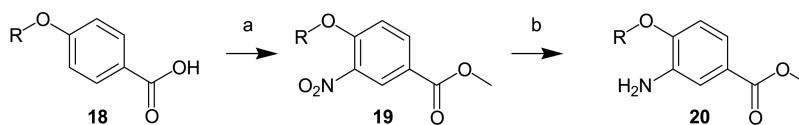
reaction of the amino group of the 3-aminobenzoic acids and an aldehyde using sodium triacetoxyborohydride as a reducing reagent. To avoid byproducts in the following synthetic steps, the secondary amine was protected by a *tert*-butyloxycarbonyl (Boc) group. The corresponding hydroxamates (10a–e) were obtained using PyBOP as an activating reagent and *O*-(tetrahydro-2H-pyran-2-yl)hydroxylamine and the subsequent cleavage of the protecting groups (Scheme 1). The 4-amidobenzohydroxamate derivatives 13a–g and 13k–z were prepared from different 3-aminobenzoic acids with modifications at

position 4 and various benzoic acid derivatives (14a–16a). The conversion of the amino group to the corresponding amide was accomplished by the reaction with activated benzoic acid derivatives and *N,N*-diisopropylethylamine (DIPEA). Different methods for activation of the carboxylic acids were tested. PyBOP, DCC, and chloroethyl formate were insufficient and formed byproducts, so activation with thionyl chloride was selected as the method of choice. Using methyl 3-aminobenzoate derivatives instead of 3-aminobenzoic acid derivatives increased the yields and also facilitated the purification. The corresponding hydroxamates were obtained using PyBOP and *O*-(tetrahydro-2H-pyran-2-yl)hydroxylamine followed by cleavage of the protecting group (Scheme 2). Generally, using tetrahydropyran (THP)-protected hydroxylamine increased the yields of the desired benzohydroxamates compared with other methods using hydroxylamine hydrochloride and KOH or potassium methanolate.

Several inhibitors (14a–16a) were synthesized directly from commercially available carboxylic acids (Scheme 3). In other cases (13h–j, 13za), the inhibitors were synthesized via compounds 20a–c, which were obtained starting from 4-alkoxybenzoic acids via a methyl esterification/nitration/reduction sequence (Scheme 5). For the nitration step, nitric acid, which is a mixture of 1 mL of nitric acid (68%) and 1.2 mL of sulfuric acid (98%), was used. This method is suitable for the reaction of 10 mmol of moderately activated aromatic rings. After the purification step, the nitro group was reduced to the corresponding amine group using Fe<sup>0</sup> and dilute hydrochloric acid.

**Materials.** All of the materials and reagents were purchased from Sigma-Aldrich Co. Ltd. and Carbolution Chemicals. All of the solvents were analytically pure and dried before use. Thin-layer chromatography was carried out on aluminum sheets coated with silica gel 60 F254 (Merck, Darmstadt, Germany). For column chromatography under normal pressure, silica gel 60 (0.036–0.200 mm) was used.



Scheme 5<sup>a</sup>

R: **20a**) CH<sub>3</sub>-CH<sub>2</sub>-; **20b**) CH<sub>3</sub>-(CH<sub>2</sub>)<sub>2</sub>-; **20c**) (CH<sub>3</sub>)<sub>2</sub>-CH-

<sup>a</sup>Conditions: (a) SOCl<sub>2</sub>, MeOH, HNO<sub>3</sub>/H<sub>2</sub>SO<sub>4</sub>; (b) Fe<sup>0</sup>, dil. HCl, MeOH.

Final compounds were confirmed to be of >95% purity based on HPLC. The purity was measured by UV absorbance at 256 nm. The HPLC system consisted of an XTerra RP18 column (3.5 μm 3.9 × 100 mm) from the manufacturer Waters (Milford, MA, USA), two LC-10AD pumps, an SPD-M10A VP PDA detector, and a SIL-HT autosampler, all from the manufacturer Shimadzu (Kyoto, Japan). The mobile phase was in all cases a gradient of methanol/water (starting at 95% water and going to 5% water).

Mass spectrometry analyses were performed with a Finnigan MAT 710C mass spectrometer (Thermo Separation Products, San Jose, CA, USA) for the ESI-MS spectra and with an LTQ (linear ion trap)-Orbitrap XL hybrid mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) for the HRMS-ESI (high-resolution mass spectrometry) spectra. For the HRMS analyses, the signals for the isotopes with the highest prevalence (<sup>35</sup>Cl, <sup>79</sup>Br) were given and calculated.

<sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded on Varian Gemini 2000 and Varian Inova 500 spectrometers using deuterated chloroform (CDCl<sub>3</sub>) and deuterated DMSO ((CD<sub>3</sub>)<sub>2</sub>SO) as solvents. Chemical shifts are referenced to the residual solvent signals. The following abbreviations for solvents and reagents were used: ethyl acetate (EtOAc), methanol (MeOH), tetrahydrofuran (THF), chloroform (CHCl<sub>3</sub>), water (H<sub>2</sub>O).

**Computational Studies.** Homology modeling of hHDAC6 was performed using the program MODELLER<sup>48</sup> as described in a previous publication.<sup>38</sup> Molecular docking of all of the inhibitors to the X-ray structures of smHDAC8, hHDAC8, and hHDAC1 and the homology model of hHDAC6 was carried out with the program Glide<sup>49</sup> (Schrödinger, LLC, New York, NY, USA) using the same protocol as in a previous study.<sup>37</sup> Briefly, the protein structures were prepared using Schrödinger's Protein Preparation Wizard. Hydrogen atoms were added, protonation states were assigned, and a restrained minimization was performed. Inhibitor structures were prepared in MOE 2012.10 (Chemical Computing Group, Montreal, Canada). All of the compounds were docked in neutral form with multiple low-energy starting conformations to produce more unbiased results.

**Enzymes and in Vitro Inhibition Assays.** Recombinant human HDAC1 and -6 were purchased from BPS Biosciences, and recombinant human HDAC8 was produced as described before.<sup>36</sup> Recombinant smHDAC8 enzyme was overproduced in *Escherichia coli* cells and purified by a method described previously.<sup>36</sup> Inhibition assays of smHDAC8 and human HDACs were performed as described earlier.<sup>36,37</sup> Briefly, the commercial Fluor de Lys drug discovery kit (BML-KI178) was used to test the inhibition of smHDAC8 and human HDAC8. The test compound, Fluor de Lys-HDAC8 substrate (50 μM), and enzyme were incubated for 90 min at 37 °C with subsequent addition of 50 μL of Developer II (BML-KI176) and further incubation for 45 min at 30 °C. Fluorescence was measured in a plate reader (BMG Polarstar) with excitation at λ = 390 nm and emission at λ = 460 nm. Inhibition tests of human HDAC1 and -6 were conducted using Cbz-(Ac)Lys-AMC (ZMAL) as the substrate and trypsin as the developer. After incubation of the test compound, ZMAL (10.5 μM), and enzyme for 90 min at 37 °C, 60 μL of trypsin was added, and the mixture was further incubated for 20 min at 37 °C. Trichostatin A (2 μM) was used in both assays to stop the reaction. Fluorescence was measured as mentioned above. IC<sub>50</sub> values were determined with OriginPro version 9.0.0 (OriginLab, Northampton, MA, USA). Values in Table 1 represent mean ± standard error.

**Cytotoxicity Studies.** HEK293 cells (DSMZ Braunschweig, ACC305) were incubated at 37 °C in a humidified incubator with 5% CO<sub>2</sub> in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal calf serum (FCS) and 5 mM glutamine. Cells were seeded out at 1.5 × 10<sup>3</sup> cells per well in a 96-well cell-culture plate (TPP, Trasadingen, Switzerland). The test compound was added immediately to the medium at 50 μM or increasing concentrations to determine IC<sub>50</sub> values. After 24 h, Alamar Blue reagent (Invitrogen, CA) was added according to the manufacturer's instructions, and the mixture was incubated again for 21 h before samples were analyzed. Detection of viable cells, which convert the resazurine compound of the reagent into the highly fluorescent resorufin, was performed using a FLUOstar OPTIMA microplate reader (BMG Labtec, Ortenberg, Germany) with the following filter set: Ex 560 nm/Em 590 nm. All of the measurements were performed in triplicate, and the reported data are means with standard deviations of ≤12%. Daunorubicin was used as a positive control, and an IC<sub>50</sub> value of 12.55 ± 0.07 μM was determined.

**Phenotypic Screening.** The screening assay to determine the effects of novel inhibitors targeting smHDAC8 on the viability of *S. mansoni* schistosomula was carried out as previously described.<sup>50</sup> Briefly, newly transformed schistosomula (NTS) were obtained in vitro by mechanical transformation of *S. mansoni* cercaria as previously described.<sup>51</sup> An NTS suspension was prepared at a concentration of 100 NTS per 100 μL using Medium 199 (Invitrogen) supplemented with 10% FCS (Gibco), penicillin (50 units mL<sup>-1</sup>), streptomycin (50 μg mL<sup>-1</sup>) and rifampicin (60 μg mL<sup>-1</sup>). Schistosomula were kept in culture for 3 h at 37 °C and 5% CO<sub>2</sub> prior to use in screening. Drug stock solutions (20 mM in DMSO) were used. Mid-dilutions were performed in 100% DMSO, and 1 μL was added to 100 μL of M199 medium in the wells of black 96 well plates (Nunc, UK) supplemented with Medium 199 and 100 μL of prepared NTS suspension (100 NTS/well). Live and dead schistosomula (treated with 70% ethanol) were used as positive and negative controls. Experiments were carried out in triplicate wells in two biological replicates, and the compounds were tested at final concentrations of 10 and 20 μM. After 48 h of drug exposure, 20 μL of resazurine solution (AbdSerotec) was added to each well. Finally, after a further 24 h of exposure, the fluorescence intensity of the highly red-fluorescent resorufin product was measured at an excitation wavelength of 530 nm and an emission wavelength of 590 nm in an Infinite M200 Pro microplate reader (TECAN). Background fluorescence of the drug-containing medium was determined for each drug dilution using wells containing only DMSO as a control. The EC<sub>50</sub> for compound 13I was measured using the same assay with different concentrations of the compound.

The effect of selected compounds on the viability of schistosomula was further tested using a microscopy-based assay as described elsewhere.<sup>52</sup> Schistosomula (2000 per well) prepared as described above were maintained in six-well plates in M199 medium kept at pH 7.4 with 10 mM HEPES and supplemented as above at 37 °C in a humid atmosphere containing 5% CO<sub>2</sub>. Inhibitor (at 10 or 20 μM) was added, dissolved in DMSO, and the culture medium was refreshed each day. The assessment of parasite mortality was carried out after microscopic examination on the basis of three criteria: granular appearance, tegumental defects, and absence of motility. At least 300 schistosomula were observed at each time point for each condition, and the results were expressed as a percentage of viable larvae remaining. Three biological replicates (different batches of larvae) were examined in duplicate for each condition.

The stability of adult worm pairs and egg laying was assayed as previously described.<sup>51</sup> Worm pairs were obtained from infected hamsters by perfusion as described previously<sup>53</sup> and washed in M199 medium, and 10 pairs were placed in 2 mL of M199 buffered complete medium (as for schistosomula above) in each well of a six-well culture plate. Worms were maintained in culture for 5 days at 37 °C (humid atmosphere, 5% CO<sub>2</sub>) before the addition of smHDAC8 inhibitors dissolved in DMSO as above. Both the culture medium and the inhibitors were refreshed daily. The number of couples remaining as pairs was determined daily by microscopy, and the medium of each well containing eggs laid by the couples was recovered and centrifuged to allow the eggs to be counted under the microscope. Two biological replicate experiments were performed in triplicate.

**Crystallization and X-ray Data Collection.** Diffraction-quality crystals of native smHDAC8 enzyme were obtained at 17 °C after 3–4 days by mixing of equal volumes of smHDAC8 (2.5 mg/mL) with reservoir solution composed of 21% PEG 3350 (Fluka) and 0.2 M Na<sup>+</sup>/K<sup>+</sup> L-tartrate and crystallization using the hanging-drop vapor diffusion technique. After 3 days, grown crystals were soaked in mother liquor supplemented with inhibitor 13a (10 mM final concentration of the inhibitor) for 20 h. Crystals used for X-ray data collection were briefly transferred in reservoir solution supplemented with 22% glycerol and flash-frozen in liquid nitrogen. Crystallographic data obtained in this project were collected at 100 K on SOLEIL beamline PROXIMA1.

#### Structure Determination, Model Building, and Refinement.

The crystallographic data were processed and scaled using HKL2000.<sup>54</sup> Since the crystals of the smHDAC8/13a complex belonged to the same space group (*P1*) and had the same unit cell as native smHDAC8 crystals,<sup>36</sup> only rigid-body refinement was used to adapt to the slight differences in unit cell constants using Phenix.<sup>55</sup> The initial model was refined through several cycles of manual building using Coot<sup>56</sup> and automated refinement with Phenix<sup>55</sup> and Buster.<sup>57</sup> The final model was validated using tools provided in Coot.<sup>56</sup> Visualization of structural data was done with Pymol,<sup>58</sup> and a two-dimensional diagram summarizing the molecular interactions between inhibitor 13a and smHDAC8 enzyme was prepared using the LigPlot program.<sup>59</sup> Atomic coordinates and structure factors of the smHDAC8/13a complex were deposited in the Protein Data Bank under PDB ID 5FUE.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.5b01478.

X-ray data collection and refinement statistics for the smHDAC8/13a complex, dihedral angle analysis of docking poses of *m*-phenylamidobenzohydroxamates with different para substituents, Alamar Blue-based viability assay, and additional synthetic procedures and analytical data (PDF)

SMILES strings, IC<sub>50</sub> values, and percent viability data (CSV)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Phone: +493455525040. E-mail: wolfgang.sippl@pharmazie.uni-halle.de.

### Author Contributions

<sup>†</sup>T.H. and A.C. contributed equally.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work and the authors of this article received funding from the European Union's Seventh Framework Programme for Research, Technological Development and Demonstration under Grant Agreements 241865 (SEtTReND) and 602080 (A-ParaDDisE). We thank Karin Schmidt-kunz, Simone Knies, Anja Kuberski, and Inka Negwer for technical assistance. The work of C.R. and M.M. was supported by institutional funds from the Centre National de la Recherche Scientifique (CNRS), the Institute National de la Santé et de la Recherche Médicale (INSERM), and the Université de Strasbourg. The work of J.L. and R.J.P. was supported by institutional funds from the CNRS, the Institut Pasteur de Lille, and the Université de Lille. This work was supported by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01 and Instruct as part of the European Strategy Forum on Research Infrastructures (ESFRI). Work on human HDAC8 was supported by the Deutsche Forschungsgemeinschaft (Ju295/13-1 to M.J. and Si868/13-1 to W.S.). M.M., T.B.S., S.D., and C.R. thank members of SOLEIL for the use of their beamline facilities and for help during data collection.

## ■ ABBREVIATIONS USED

aq, aqueous; Boc, *tert*-butyloxycarbonyl; Boc<sub>2</sub>O, di-*tert*-butyl dicarbonate; PyBOP, benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate; CDCl<sub>3</sub>, deuterated chloroform; CEF, chloroethyl formate; CHCl<sub>3</sub>, chloroform; DCC, dicyclohexylcarbodiimide; DIPEA, diisopropylethylamine; dDMSO, deuterated dimethyl sulfoxide; EtOAc, ethyl acetate; Et<sub>3</sub>N, triethylamine; eq, equivalent; MeOH, methanol; Na(AcO)<sub>3</sub>BH, sodium triacetoxyborohydride; n.a., not active; n.d., not determined; NH<sub>2</sub>OTHP, *O*-(tetrahydro-2*H*-pyran-2-yl)-hydroxylamine; smHDAC8, *Schistosoma mansoni* histone deacetylase 8; SOCl<sub>2</sub>, thionyl chloride; sol, solution; *t*-BuOH, *tert*-butyl alcohol; TEA, triethylamine; TFA, trifluoroacetic acid; THF, tetrahydrofuran; TLC, thin-layer chromatography

## ■ REFERENCES

- Hotez, P. J.; Kamath, A. Neglected Tropical Diseases in Sub-Saharan Africa: Review of Their Prevalence, Distribution, and Disease Burden. *PLoS Neglected Trop. Dis.* **2009**, *3*, e412.
- Brown, M. Schistosomiasis. *Clin. Med.* **2011**, *11*, 479–482.
- Ross, A. G.; Bartley, P. B.; Sleigh, A. C.; Olds, G. R.; Li, Y.; Williams, G. M.; McManus, D. P. Schistosomiasis. *N. Engl. J. Med.* **2002**, *346*, 1212–1220.
- Gray, D. J.; Ross, A. G.; Li, Y.-S.; McManus, D. P. Diagnosis and Management of Schistosomiasis. *BMJ.* **2011**, *342*, d2651.
- World Health Organization. *Wkly. Epidemiol. Rec.* **2015**, *90*, 25–32.
- Dömling, A.; Khoury, K. Praziquantel and Schistosomiasis. *ChemMedChem* **2010**, *5*, 1420–1421.
- Ismail, M.; Botros, S.; Metwally, A.; William, S.; Farghally, A.; Tao, L.; Day, T. I. M. A.; Bennett, J. L. Resistance to Praziquantel: Direct Evidence from *Schistosoma mansoni* Isolated from Egyptian Villagers. *Am. J. Trop. Med. Hyg.* **1999**, *60*, 932–935.
- Doenhoff, M. J.; Kusel, J. R.; Coles, G. C.; Cioli, D. Resistance of *Schistosoma mansoni* to Praziquantel: Is There a Problem? *Trans. R. Soc. Trop. Med. Hyg.* **2002**, *96*, 465–469.
- Webster, J. P.; Molyneux, D. H.; Hotez, P. J.; Fenwick, A. The Contribution of Mass Drug Administration to Global Health: Past, Present and Future. *Philos. Trans. R. Soc., B* **2014**, *369*, 20130434.
- Doenhoff, M. J.; Cioli, D.; Utzinger, J. Praziquantel: Mechanisms of Action, Resistance and New Derivatives for Schistosomiasis. *Curr. Opin. Infect. Dis.* **2008**, *21*, 659–667.

- (11) Marek, M.; Oliveira, G.; Pierce, R. J.; Jung, M.; Sippl, W.; Romier, C. Drugging the Schistosome Zinc-Dependent HDACs: Current Progress and Future Perspectives. *Future Med. Chem.* **2015**, *7*, 783–800.
- (12) Hotez, P. J.; Pecoul, B. Manifesto for Advancing the Control and Elimination of Neglected Tropical Diseases. *PLoS Neglected Trop. Dis.* **2010**, *4*, e718.
- (13) Schäfer, S.; Jung, M. Chromatin Modifications as Targets for New Anticancer Drugs. *Arch. Pharm.* **2005**, *338*, 347–357.
- (14) Pierce, R. J.; Dubois-Abdesselem, F.; Lancelot, J.; Andrade, L.; Oliveira, G. Targeting Schistosome Histone Modifying Enzymes for Drug Development. *Curr. Pharm. Des.* **2012**, *18*, 3567–3578.
- (15) Andrews, K. T.; Haque, A.; Jones, M. K. HDAC Inhibitors in Parasitic Diseases. *Immunol. Cell Biol.* **2012**, *90*, 66–77.
- (16) Ouaisi, M.; Ouaisi, A. Histone Deacetylase Enzymes as Potential Drug Targets in Cancer and Parasitic Diseases. *J. Biomed. Biotechnol.* **2006**, *2006*, 13474.
- (17) Grunstein, M. Histone Acetylation in Chromatin Structure and Transcription. *Nature* **1997**, *389*, 349–352.
- (18) Peserico, A.; Simone, C. Physical and Functional HAT/HDAC Interplay Regulates Protein Acetylation Balance. *J. Biomed. Biotechnol.* **2011**, *2011*, 371832.
- (19) 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.
- (20) North, B. J.; Verdin, E. Sirtuins: Sir2-Related NAD-Dependent Protein Deacetylases. *Genome Biol.* **2004**, *5*, 224.
- (21) Cho, Y. S.; Whitehead, L.; Li, J.; Chen, C. H. T.; Jiang, L.; Vögtle, M.; Francotte, E.; Richert, P.; Wagner, T.; Traebert, M.; Lu, Q.; Cao, X.; Dumotier, B.; Fejzo, J.; Rajan, S.; Wang, P.; Yan-Neale, Y.; Shao, W.; Atadja, P.; Shultz, M. Conformational Refinement of Hydroxamate-Based Histone Deacetylase Inhibitors and Exploration of 3-Piperidin-3-Ylindole Analogues of Dacinostat (LAQ824). *J. Med. Chem.* **2010**, *53*, 2952–2963.
- (22) Marks, P.; Rifkind, R. A.; Richon, V. M.; Breslow, R.; Miller, T.; Kelly, W. K. Histone Deacetylases and Cancer: Causes and Therapies. *Nat. Rev. Cancer* **2001**, *1*, 194–202.
- (23) Ito, K.; Ito, M.; Elliott, W. M.; Cosio, B.; Caramori, G.; Kon, O. M.; Barczyk, A.; Hayashi, S.; Adcock, I. M.; Hogg, J. C.; Barnes, P. J. Decreased Histone Deacetylase Activity in Chronic Obstructive Pulmonary Disease. *N. Engl. J. Med.* **2005**, *352*, 1967–1976.
- (24) Mukherjee, P.; Pradhan, A.; Shah, F.; Tekwani, B. L.; Avery, M. A. Structural Insights into the Plasmodium Falciparum Histone Deacetylase 1 (PfHDAC-1): A Novel Target for the Development of Antimalarial Therapy. *Bioorg. Med. Chem.* **2008**, *16*, 5254–5265.
- (25) Giannini, G.; Cabri, W.; Fattorusso, C.; Rodriguez, M. Histone Deacetylase Inhibitors in the Treatment of Cancer: Overview and Perspectives. *Future Med. Chem.* **2012**, *4*, 1439–1460.
- (26) 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.
- (27) Balasubramanian, S.; Ramos, J.; Luo, W.; Sirisawad, M.; Verner, E.; Buggy, J. J. A novel histone deacetylase 8 (HDAC8)-specific inhibitor PCI-34051 induces apoptosis in T-cell lymphomas. *Leukemia* **2008**, *22*, 1026–1034.
- (28) Suzuki, T.; Ota, Y.; Ri, M.; Bando, M.; Gotoh, A.; Itoh, Y.; Tsumoto, H.; Tatum, P. R.; Mizukami, T.; Nakagawa, H.; Iida, S.; Ueda, R.; Shirahige, K.; Miyata, N. Rapid Discovery of Highly Potent and Selective Inhibitors of Histone Deacetylase 8 Using Click Chemistry to Generate Candidate Libraries. *J. Med. Chem.* **2012**, *55*, 9562–9575.
- (29) Huang, W. J.; Wang, Y. C.; Chao, S. W.; Yang, C. J.; Chen, L. C.; Lin, M. H.; Hou, W. C.; Chen, M. Y.; Lee, T. L.; Yang, P.; Chang, C. I. Synthesis and biological evaluation of ortho-aryl N-hydroxycinnamides as potent histone deacetylase (HDAC) 8 isoform-selective inhibitors. *ChemMedChem* **2012**, *7*, 1815–1824.
- (30) Suzuki, T.; Muto, N.; Bando, M.; Itoh, Y.; Masaki, A.; Ri, M.; Ota, Y.; Nakagawa, H.; Iida, S.; Shirahige, K.; Miyata, N. Design, synthesis, and biological activity of NCC149 derivatives as histone deacetylase 8-selective inhibitors. *ChemMedChem* **2014**, *9*, 657–664.
- (31) Andrews, K. T.; Walduck, A.; Kelso, M. J.; Fairlie, D. P.; Saul, A.; Parsons, P. G. Anti-Malarial Effect of Histone Deacetylation Inhibitors and Mammalian Tumour Cytodifferentiating Agents. *Int. J. Parasitol.* **2000**, *30*, 761–768.
- (32) Oger, F.; Dubois, F.; Caby, S.; Noël, C.; Cornette, J.; Bertin, B.; Capron, M.; Pierce, R. J. The Class I Histone Deacetylases of the Platyhelminth Parasite *Schistosoma mansoni*. *Biochem. Biophys. Res. Commun.* **2008**, *377*, 1079–1084.
- (33) Dubois, F.; Caby, S.; Oger, F.; Cosseau, C.; Capron, M.; Grunau, C.; Dissous, C.; Pierce, R. J. Histone Deacetylase Inhibitors Induce Apoptosis, Histone Hyperacetylation and up-Regulation of Gene Transcription in *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* **2009**, *168*, 7–15.
- (34) Lancelot, J.; Caby, S.; Dubois-Abdesselem, F.; Vanderstraete, M.; Trolet, J.; Oliveira, G.; Bracher, F.; Jung, M.; Pierce, R. J. *Schistosoma mansoni* sirtuins: characterization and potential as chemotherapeutic targets. *PLoS Neglected Trop. Dis.* **2013**, *7*, e2428.
- (35) Nakagawa, M.; Oda, Y.; Eguchi, T.; Aishima, S.-I.; Yao, T.; Hosoi, F.; Basaki, Y.; Ono, M.; Kuwano, M.; Tanaka, M.; Tsuneyoshi, M. Expression Profile of Class I Histone Deacetylases in Human Cancer Tissues. *Oncol. Rep.* **2007**, *18*, 769–774.
- (36) Marek, M.; Kannan, S.; Hauser, A. T.; Moraes Mourão, M.; Caby, S.; Cura, V.; Stolfa, D. A.; Schmidtkunz, K.; Lancelot, J.; Andrade, L.; Renaud, J. P.; Oliveira, G.; Sippl, W.; Jung, M.; Cavarelli, J.; Pierce, R. J.; Romier, C. Structural Basis for the Inhibition of Histone Deacetylase 8 (HDAC8), a Key Epigenetic Player in the Blood Fluke *Schistosoma mansoni*. *PLoS Pathog.* **2013**, *9*, e1003645.
- (37) Stolfa, D.; Marek, M.; Lancelot, J.; Hauser, A.-T.; Walter, A.; Leproult, E.; Melesina, J.; Rumpf, T.; Wurtz, J.-M.; Cavarelli, J.; Sippl, W.; Pierce, R. J.; Romier, C.; Jung, M. Molecular Basis for the Antiparasitic Activity of a Mercaptoacetamide Derivative That Inhibits Histone Deacetylase 8 (HDAC8) from the Human Pathogen *Schistosoma mansoni*. *J. Mol. Biol.* **2014**, *426*, 3442–3453.
- (38) Kannan, S.; Melesina, J.; Hauser, A.; Chakrabarti, A.; Heimburg, T.; Schmidtkunz, K.; Walter, A.; Marek, M.; Pierce, R. J.; Romier, C.; Jung, M.; Sippl, W. Discovery of Inhibitors of *Schistosoma mansoni* HDAC8 by Combining Homology Modeling, Virtual Screening, and in Vitro Validation. *J. Chem. Inf. Model.* **2014**, *54*, 3005–3019.
- (39) Chakrabarti, A.; Oehme, I.; Witt, O.; Oliveira, G.; Sippl, W.; Romier, C.; Pierce, R. J.; Jung, M. HDAC8: A Multifaceted Target for Therapeutic Interventions. *Trends Pharmacol. Sci.* **2015**, *36*, 481–92.
- (40) Rodrigues, D. A.; Ferreira-Silva, G. Á.; Ferreira, A. C. S.; Fernandes, R. A.; Kwee, J. K.; Sant’Anna, C. M. R.; Ionta, M.; Fraga, C. A. M. Design, Synthesis, and Pharmacological Evaluation of Novel N-Acylhydrazono Derivatives as Potent Histone Deacetylase 6/8 Dual Inhibitors. *J. Med. Chem.* **2016**, *59*, 655–670.
- (41) Wang, L.; Kofler, M.; Brosch, G.; Melesina, J.; Sippl, W.; Martinez, E. D.; Easmon, J. 2-Benzazoyl-4-Piperazin-1-Ylsulfonylbenzenecarbo-hydroxamic Acids as Novel Selective Histone Deacetylase-6 Inhibitors with Antiproliferative Activity. *PLoS One* **2015**, *10*, e0134556.
- (42) Lobera, M.; Madauss, K. P.; Pohlhaus, D. T.; Wright, Q. G.; Trocha, M.; Schmidt, D. R.; Baloglu, E.; Trump, R. P.; Head, M. S.; Hofmann, G. A.; Murray-Thompson, M.; Schwartz, B.; Chakravorty, S.; Wu, Z.; Mander, P. K.; Kruidenier, L.; Reid, R. A.; Burkhart, W.; Turunen, B. J.; Rong, J. X.; Wagner, C.; Moyer, M. B.; Wells, C.; Hong, X.; Moore, J. T.; Williams, J. D.; Soler, D.; Ghosh, S.; Nolan, M. A. Selective class IIa histone deacetylase inhibition via a nonchelating zinc-binding group. *Nat. Chem. Biol.* **2013**, *9*, 319–325.
- (43) Olson, D. E.; Wagner, F. F.; Kaya, T.; Gale, J. P.; Aidoud, N.; Davoine, E. L.; Lazzaro, F.; Weiwler, M.; Zhang, Y. L.; Holson, E. B. Discovery of the First Histone Deacetylase 6/8 Dual Inhibitors. *J. Med. Chem.* **2013**, *56*, 4816–4820.
- (44) Tang, W.; Luo, T.; Greenberg, E. F.; Bradner, J. E.; Schreiber, S. L. Discovery of histone deacetylase 8 selective inhibitors. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 2601–2605.



(45) KrennHrubec, K.; Marshall, B. L.; Hedglin, M.; Verdin, E.; Ulrich, S. Ma. Design and evaluation of 'Linkerless' hydroxamic acids as selective HDAC8 inhibitors. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2874–2878.

(46) Panic, G.; Flores, D.; Ingram-Sieber, K.; Keiser, J. Fluorescence/luminescence-based markers for the assessment of *Schistosoma mansoni* schistosomula drug assays. *Parasites Vectors* **2015**, *8*, 624.

(47) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science* **2009**, *325*, 834–840.

(48) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling Using Modeller. *Current Protocols in Bioinformatics*; Wiley: New York, **2006**; Chapter 5, Unit 5.6.

(49) Suite 2012: Maestro Version 9.3, Protein Preparation Wizard, Epik Version 2.3, Glide Version 5.8; Schrödinger, LLC: New York, 2012.

(50) Marxer, M.; Ingram, K.; Keiser, J. Development of an in Vitro Drug Screening Assay Using *Schistosoma haematobium* schistosomula. *Parasites Vectors* **2012**, *5*, 165.

(51) Ramalho-Pinto, F. J.; Gazzinelli, G.; Howells, R. E.; Mota-Santos, T.; Figueiredo, E.; Pellegrino, J. *Schistosoma mansoni*: Defined System for Stepwise Transformation of Cercaria to Schistosomule in Vitro. *Exp. Parasitol.* **1974**, *36*, 360–372.

(52) Vanderstraete, M.; Gouignard, N.; Cailliau, K.; Morel, M.; Lancelot, J.; Bodart, J. F.; Dissous, C. Dual Targeting of Insulin and Venus Kinase Receptors of *Schistosoma mansoni* for Novel Anti-Schistosome Therapy. *PLoS Neglected Trop. Dis.* **2013**, *7*, e2226.

(53) Smithers, S. R.; Terry, R. J. The Infection of Laboratory Hosts with Cercariae of *Schistosoma mansoni* and the Recovery of the Adult Worms. *Parasitology* **1965**, *55*, 695–700.

(54) Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, *276*, 307–326.

(55) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 213–221.

(56) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126–2132.

(57) Blanc, E.; Roversi, P.; Vonnrhein, C.; Flensburg, C.; Lea, S. M.; Bricogne, G. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2210–2221.

(58) The PyMOL Molecular Graphics System, version 1.7.4; Schrödinger, LLC: New York.

(59) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng., Des. Sel.* **1995**, *8*, 127–134.

## Supporting Information

### Structure-Based Design and Synthesis of Potent Inhibitors Targeting HDAC8 of *Schistosoma mansoni* for the Treatment of Schistosomiasis

Tino Heimbürg<sup>1</sup>, Alokta Chakrabarti<sup>2</sup>, Julien Lancelot<sup>3</sup>, Martin Marek<sup>4</sup>, Jelena Melesina<sup>1</sup>, Alexander-Thomas Hauser<sup>2</sup>, Tajith B. Shaik<sup>4</sup>, Sylvie Duclaud<sup>4</sup>, Dina Robaa<sup>1</sup>, Frank Erdmann<sup>1</sup>, Matthias Schmidt<sup>1</sup>, Christophe Romier<sup>4</sup>, Raymond J. Pierce<sup>3</sup>, Manfred Jung<sup>2</sup> and Wolfgang Sippl<sup>\*,1</sup>

<sup>1</sup>*Institute of Pharmacy, Martin-Luther University of Halle-Wittenberg, 06120 Halle/Saale, Germany*

<sup>2</sup>*Institute of Pharmaceutical Sciences, University of Freiburg, 79104 Freiburg, Germany*

<sup>3</sup>*University of Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 8204 - CIIL - Centre d'Infection et d'Immunité de Lille, 59000 Lille, France*

<sup>4</sup>*Département de Biologie Structurale Intégrative, Institut de Génétique et Biologie Moléculaire et Cellulaire (IGBMC), Université de Strasbourg (UDS), CNRS, INSERM, 67404 Illkirch Cedex, France*

\* *Institute of Pharmacy, Martin-Luther-Universität Halle-Wittenberg, Wolfgang-Langenbeck-Str. 4, 06120 Halle/Saale, email: [wolfgang.sippl@pharmazie.uni-halle.de](mailto:wolfgang.sippl@pharmazie.uni-halle.de)*

**Contents:** S2: smHDAC8/**13a** X-ray structure. Data collection and refinement statistics.  
S3: Dihedral angle analysis of docking poses of *m*-phenylamido-benzohydroxamates with different para substituents.  
S4: AlamarBlue-based viability assay.  
S4–S18: Additional synthetic procedures and analytical data .

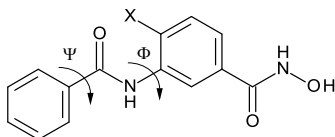


**Table S1.** smHDAC8/13a X-ray structure. Data collection and refinement statistics.

smHDAC8/13a complex	
<b>Data collection</b>	
Space group	P1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	70.96, 70.67, 98.66
$\alpha$ , $\beta$ , $\gamma$ (°)	77.99, 75.41, 85.40
Resolution (Å)	50.0-2.20 (2.24-2.20)*
$R_{\text{sym}}$ or $R_{\text{merge}}$	0.119 (0.289)
$I / \sigma I$	17.88 (4.75)
Completeness (%)	96.4 (94.8)
Redundancy	2.8 (2.7)
<b>Refinement</b>	
Resolution (Å)	44.15-2.20
No. reflections	88494
$R_{\text{work}} / R_{\text{free}}$	0.148/0.196
No. atoms	
Protein	12988
Ligand/Ions	167
Water	517
<i>B</i> -factors	
Protein	28.06
Ligand/Ions	41.52
Water	31.76
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	0.98

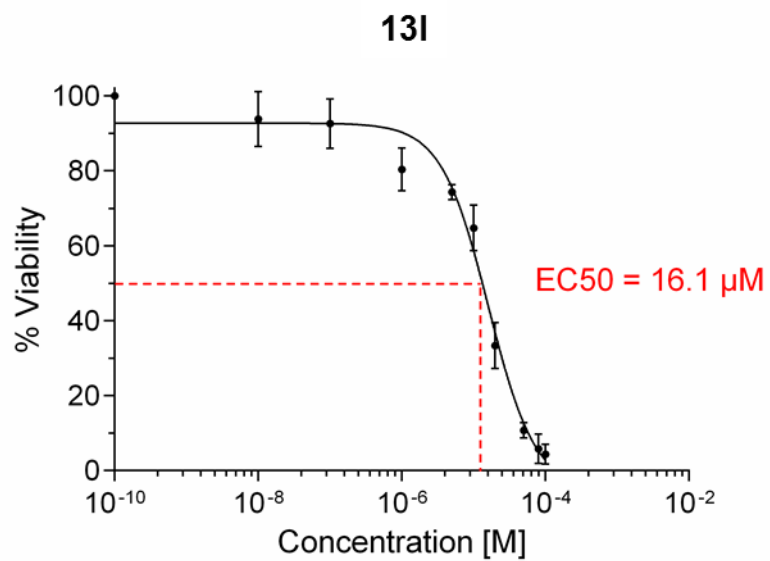
\* Values in parentheses are for highest-resolution shell.

**Table S2.** Dihedral angle ( $\Phi$ ) analysis of docking poses of *meta*-phenylamido-benzohydroxamates with different *para*-substituents.



Cpd.	<i>para</i> -substituent X	Dihedral angle $\Phi$ (C4-C3-N-C, deg.)		Dihedral angle $\Psi$ (O-C-C1'-C2', deg.)		Activity gain compared to <b>13a</b> (smHDAC8)	Activity gain compared to <b>13a</b> (hHDAC8)
		smHDAC8	hHDAC8	smHDAC8	hHDAC8		
<b>13a</b>	H	-60 to -102 (X-ray) -82	29	-15 to -51 (X-ray) -33	30		
<b>13b</b>	methyl	-88	51	-39	30	4.0	2.9
<b>13c</b>	methoxy	-89	43	-36	29	2.5	6.6
<b>13d</b>	fluoro	-83	41	-36	28	2.6	1.8
<b>13e</b>	chloro	-85	49	-36	28	7.0	4.9
<b>13f</b>	bromo	-90	50	-38	28	3.1	3.0
<b>13g</b>	trifluoromethyl	-95	148	-35	31	3.4	1.7
<b>13h</b>	ethoxy	-83	41	-40	27	3.6	3.4
<b>13i</b>	propoxy	-77	44	-27	28	1.8	not determined
<b>13j</b>	isopropoxy	-90	33	-33	50	2.1	20.1

**Figure S1.** The activity on schistosomula shown as EC<sub>50</sub>-value of compound **13I** using the AlamarBlue-based viability-assay.



## SYNTHESIS

### General procedures

#### General procedure for the synthesis of the amines (**10a-e**)

The 3-amino-benzoic acid (2 mmol) derivative was dissolved in toluene (100 ml), the aldehyde (4 mmol) was added and the reaction mixture was heated for 2 h under reflux using a water trap. The reaction was monitored by TLC. Subsequently, the solvent and the excess of the aldehyde were evaporated under reduced pressure. The crude product was dissolved in THF and cooled to 0 °C then sodium triacetoxyborohydride (16 mmol) was added and the reaction mixture was stirred for 15 minutes after which acetic acid was given to the reaction and stirring was continued over night at room temperature<sup>1</sup>. The reaction was stopped by adding water and adjusting the pH to 5 with aq. potassium hydrogen carbonate solution. The mixture was extracted with EtOAc (3 x 25 ml), the organic layers were combined and evaporated under vacuum. The crude amine was dissolved in MeOH (50 ml) and *t*-BuOH (50 ml) then Boc<sub>2</sub>O was added and the reaction was stirred at room temperature over night. The product was purified by column chromatography (chloroform/ methanol, 98:2).

#### Cleavage of the Boc-protecting group

The boc-protected hydroxamate derivative was dissolved in a mixture of CHCl<sub>3</sub> (80 ml) and TFA (20 ml) and stirring was continued for 1 h at room temperature<sup>2</sup>. After that the mixture was cooled to 0 °C and an aq. potassium carbonate solution was added to adjust the pH to 6-7. The reaction mixture was extracted with CHCl<sub>3</sub> (3 x 25 ml), the organic layers were combined and evaporated under vacuum. The product was purified by column chromatography (chloroform/ methanol/ formic acid, 95:4.95:0.05).

General procedure for the synthesis of the amides (**13a–za**)

The 3-amino-benzoic acid derivative (1 eq.) was dissolved in MeOH (50 ml) and the solution was cooled to 0 °C. Thionylchloride (3 eq.)<sup>3</sup> was added dropwise and the reaction mixture was heated under reflux for 1 h. After evaporating the solvent and the excess of thionylchloride under vacuum, the obtained methyl ester was dissolved in dry THF (50 ml) and DIPEA (2.5 eq.).

Method A: Activation of the carboxylic acid with PyBOP (**13b, 13c, 13e, 13k, 13l, 13m, 13p, 13r–x, 13z**)

The solution of methyl 3-amino-benzoate derivatives and DIPEA in THF was added to a mixture of benzoic acid (1.2 eq.) and PyBOP (1,5 eq.)<sup>4,5</sup> and stirred at room temperature overnight. The solvent was subsequently evaporated under vacuum and the mixture was dissolved in EtOAc (50 ml) and washed with aq. potassium hydrogen carbonate solution and brine. The organic layer was evaporated under vacuum and the obtained amide was purified by column chromatography (chloroform/ methanol, 99:1).

Method B: Activation of the carboxylic acid with thionylchloride (**13a, 13l, 13n, 13o, 13q, 13y** and **13za**)

The carboxylic acid was cooled to 0 °C and then thionylchloride (2 eq.) was added dropwise. The mixture was heated under reflux for 30 minutes<sup>37</sup>. After evaporating the excess of thionylchloride under vacuum the acid chloride was dissolved in dry THF (50 ml) and added to a solution of the 3-amino-methyl-benzoate and DIPEA in THF. The reaction was monitored by TLC. Subsequently, the solvent was evaporated under vacuum and the mixture was dissolved in EtOAc (50 ml) and washed with aq. sodium hydroxide solution. The organic layer was evaporated under vacuum and the amide was purified by column chromatography (chloroform/ methanol, 99:1).

Method C: Using benzoyl chloride as activated benzoic acid (**13d**, **13f-j**)

The benzoyl chloride was dissolved in dry THF (50 ml) and added to a solution of the 3-amino-methyl-benzoate and DIPEA in THF. The reaction was monitored by TLC. Subsequently, the solvent was evaporated under vacuum and the mixture was dissolved in EtOAc (50 ml) and washed with aq. sodium hydroxide solution. The organic layer was evaporated under vacuum and the amide was purified by column chromatography (chloroform/ methanol, 99:1).

Hydrolysis of the methyl ester (**12a**, **12b**, **12d**, **12f-za**) to the carboxylic acid derivatives.

The amide was dissolved in MeOH (25 ml) and 1M aq. sodium hydroxide solution (10 ml) and heated to 50 °C for 2 h. The reaction was monitored by TLC. After that the solvent was evaporated under vacuum and the product was dissolved in EtOAc and washed with 1M HCl solution, the organic layer was evaporated under vacuum.

General procedure for the synthesis of the hydroxamate

Method A: (**13a-c**)

The substituted benzoic acid derivative (1 eq.) was dissolved together with TEA (1.3 eq.) in dry THF and cooled to 0 °C. After that CEF was added and stirred for 30 minutes at room temperature and then the reaction mixture was filtered directly into a mixture of a freshly prepared hydroxyl amine solution\* with an excess of hydroxyl amine<sup>6</sup>.

\*A calculated amount of hydroxyl amine hydrochloride was added to a solution of sodium methanolate in methanol. After stirring for 30 minutes at room temperature, the flask was cooled to 0 °C and filtered.

#### Method B: (13d–za)

The substituted benzoic acid derivative (1 eq.) was dissolved in dry THF (50 ml) and PyBOP (1.2 eq.) was added. To the activated acid a mixture of NH<sub>2</sub>OTHP (1.5 eq.) and DIPEA (2.5 eq.) in dry THF (5 ml) was added and the reaction mixture was stirred at room temperature overnight<sup>4,5,7,8</sup>. The solvent was evaporated under vacuum and the mixture was dissolved in EtOAc (50 ml) and washed with aq. potassium hydrogen carbonate solution and brine. The organic layer was evaporated under vacuum and the amide was purified by column chromatography (chloroform/ methanol/ TEA, 99.5:0.45:0.05). The obtained product was dissolved in THF and a catalytic amount of diluted HCl was added and it was stirred at room temperature. The reaction was controlled by TLC. After that the solvent was evaporated under vacuum and the mixture was dissolved in EtOAc (50 ml) and washed with brine. The organic layer was evaporated under vacuum and the amide was purified by column chromatography (chloroform/ methanol/ formic acid, 95:4.95:0.05).

#### General procedure for the synthesis of the 3-amino-4-alkoxy-methylbenzoate derivatives (20a–c)

At first the 4-alkoxybenzoic acid derivative (1 eq.) was dissolved in methanol and cooled to 0 °C. Thionyl chloride was added dropwise and the reaction mixture was heated under reflux for 1 h. After evaporating the solvent and the excess of thionyl chloride under vacuum<sup>3</sup> the obtained methyl ester was cooled to 0 °C again and nitro sulfuric acid (1.5 eq.) was added dropwise while cooling the reaction. To form the nitro sulfuric acid 1 ml nitric acid 68% (1.5 mmol) was cooled to 0 °C and then 1.2 ml concentrated sulfuric acid was added.<sup>9</sup> The reaction mixture stirred for 30 min at 50 °C, crushed ice was subsequently added to the reaction mixture and the precipitated product was filtered and washed with water. The product was recrystallized from ethanol. The

methyl 3-nitro-4-alkoxybenzoate (1 eq.) was added to a mixture of MeOH and 1 M HCl (4:1) and iron powder (10 eq.)<sup>10</sup> and the reaction mixture was heated under reflux for 1 h. After that the solvent was evaporated and the product was dissolved in 50 ml 1 M NaHCO<sub>3</sub> solution and extracted with ethyl acetate (3 x 25 ml). The organic layer was evaporated under vacuum and the product was purified by column chromatography (chloroform/ methanol, 99.5:0.5).



## Characterization data

### 3-N-Benzylamino-benzohydroxamate (**10a**)

MS m/z: 241.29 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 10.95 (s, 1H, -HN-OH), 8.83 (s, 1H, -NH-OH), 7.38 – 7.24 (m, 4H, Ar-H), 7.20 (t, J = 6.9 Hz, 1H, Ar-H), 7.06 (t, J = 7.8 Hz, 1H, Ar-H), 6.95 (s, 1H, Ar-H), 6.84 (d, J = 7.6 Hz, 1H, Ar-H), 6.66 (d, J = 8.0 Hz, 1H, Ar-H), 6.41 (t, J = 5.9 Hz, 1H, Ar-H), 4.27 (d, J = 5.9 Hz, 2H, -CH<sub>2</sub>-).

HR-MS m/z: 243.1130 [M+H]<sup>+</sup>; calculated C<sub>14</sub>H<sub>15</sub>N<sub>2</sub>O<sub>2</sub><sup>+</sup>: 243.1128

HPLC: rt 7.02 min (95.07%)

Yield: 7 mg; 0.03 mmol; 3%

### 3-(N-Cyclohexyl)-amino-benzohydroxamate (**10b**)

MS m/z: 233.36 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 10.96 (s, 1H, -HN-OH), 8.83 (s, 1H, -NH-OH), 7.15 – 7.09 (m, 1H, Ar-H), 7.09 – 7.02 (m, 1H, Ar-H), 6.89 (d, J = 1.8 Hz, 1H, Ar-H), 6.80 (d, J = 7.7 Hz, 1H, Ar-H), 6.68 – 6.62 (m, 1H, -NH-Ar), 5.56 (d, J = 7.7 Hz, 1H, CH<sub>2</sub>-NH-Ar), 1.95 – 1.84 (m, 2H, -CH<sub>2</sub>-), 1.74 – 1.64 (m, 2H, -CH<sub>2</sub>-), 1.62 – 1.54 (m, 1H, -CH<sub>2</sub>-), 1.37 – 1.06 (m, 6H, -CH<sub>2</sub>-).

HR-MS m/z: 235.1442 [M+H]<sup>+</sup>; calculated C<sub>13</sub>H<sub>19</sub>N<sub>2</sub>O<sub>2</sub><sup>+</sup>: 235.1441

HPLC: rt 5.67 min (96.62%)

Yield: 25 mg; 0.11 mmol; 11%

### 3-(N,N-Dimethylamino)-benzohydroxamate (**10c**)

MS m/z: 181.13 [M+H]<sup>+</sup>

<sup>1</sup>H NMR (500 MHz, DMSO) δ 11.07 (s, 1H, -HN-OH), 7.24 – 7.19 (m, 1H, Ar-H), 7.05 (s, 1H, Ar-H), 7.00 (d, J = 7.3 Hz, 1H, Ar-H), 6.84 (d, J = 8.3 Hz, 1H, Ar-H), 2.92 (s, 6H, -N-(CH<sub>3</sub>)<sub>2</sub>).

HR-MS m/z: 181.0973 [M+H]<sup>+</sup>; calculated C<sub>9</sub>H<sub>13</sub>N<sub>2</sub>O<sub>2</sub><sup>+</sup>: 181.0972

HPLC: rt 2.31 min (95.41%)

Yield: 90 mg; 0.5 mmol; 50%

### 3-Dibenzylamino-benzohydroxamate (**10d**)

MS m/z: 331.29 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 10.99 (s, 1H, -HN-OH), 8.84 (d, J = 1.6 Hz, 1H, -NH-OH), 7.37 – 7.27 (m, 4H, Ar-H), 7.27 – 7.16 (m, 6H, Ar-H), 7.14 – 7.06 (m, 2H, Ar-H), 6.92 (d, J = 7.6 Hz, 1H, Ar-H), 6.78 – 6.74 (m, 1H, Ar-H), 4.70 (s, 4H, -N-(CH<sub>2</sub>-Ar)<sub>2</sub>).

HR-MS m/z: 333.1599 [M-H]<sup>-</sup>; calculated: C<sub>21</sub>H<sub>21</sub>N<sub>2</sub>O<sub>2</sub><sup>-</sup>: 333.1598

HPLC: rt 13.35 min (95.68%)

Yield: 110 mg; 0.33 mmol;

### 3-(N-Cyclohexyl)-amino-4-methyl-benzohydroxamate (**10e**)

MS m/z: 249.03 [M+H]<sup>+</sup>

<sup>1</sup>H NMR (500 MHz, DMSO) δ 10.99 (s, 1H, -HN-OH), 8.81 (s, 1H, -NH-OH), 6.99 (d, J = 7.6 Hz, 1H, Ar-H), 6.90 (d, J = 11.9 Hz, 1H, Ar-H), 6.87 (d, J = 6.6 Hz, 1H, Ar-H), 2.08 (s, 3H, Ar-CH<sub>3</sub>), 1.98 – 1.90 (m, 2H, -CH<sub>2</sub>-), 1.77 – 1.69 (m, 2H, -CH<sub>2</sub>-), 1.66 – 1.58 (m, 1H, -CH<sub>2</sub>-), 1.41 – 1.11 (m, 6H, -CH<sub>2</sub>-).

HR-MS m/z: 249.1598 [M+H]<sup>+</sup>; calculated C<sub>14</sub>H<sub>21</sub>N<sub>2</sub>O<sub>2</sub><sup>+</sup>: 249.1598

HPLC: rt 6.70 min (95.18%)  
Yield: 140 mg; 0.56 mmol; 11.2%

3-Benzamido-benzohydroxamate (**13a**)

MS m/z: 255.23 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.11 (s, 1H, -NH-OH), 9.47 (s, 1H, -CONH-Ar), 9.02 (s, 1H, -NH-OH), 8.22 – 8.15 (m, 1H, Ar-H), 8.02 – 7.87 (m, 3H, Ar-H), 7.64 – 7.47 (m, 3H, Ar-H), 7.47 – 7.33 (m, 2H, Ar-H).

HR-MS m/z: 257.0923 [M+H]<sup>+</sup>; calculated C<sub>14</sub>H<sub>13</sub>N<sub>2</sub>O<sub>3</sub><sup>+</sup>: 257.0921

HPLC: rt 9.88 min (86.11%)

Yield: 20 mg; 0.08 mmol; 8%

3-Benzamido-4-methyl-benzohydroxamate (**13b**)

MS m/z: 269.25 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (500 MHz, DMSO) δ 11.17 (s, 1H, -NH-OH), 9.97 (s, 1H, -CONH-Ar), 8.97 (s, 1H, -NH-OH), 8.00 – 7.95 (m, 2H, Ar-H), 7.77 (s, 1H, Ar-H), 7.62 – 7.50 (m, 4H, Ar-H), 7.34 (d, J = 8.2 Hz, 1H, Ar-H), 2.26 (s, 3H, Ar-CH<sub>3</sub>).

HR-MS m/z: 271.1079 [M+H]<sup>+</sup>; calculated C<sub>15</sub>H<sub>15</sub>N<sub>2</sub>O<sub>3</sub><sup>+</sup>: 271.1077

HPLC: rt 9.83 min (96.67%)

Yield: 35 mg; 0.13 mmol; 6.5%

3-Benzamido-4-methoxy-benzohydroxamate (**13c**)

MS m/z: 285.31 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.09 (s, 1H, -NH-OH), 9.53 (s, 1H, -CONH-Ar), 8.90 (d, J = 1.6 Hz, 1H, -NH-OH), 8.17 (d, J = 2.2 Hz, 1H, Ar-H), 7.99 – 7.92 (m, 2H, Ar-H), 7.65 – 7.56 (m, 2H, Ar-H), 7.55 – 7.49 (m, 2H, Ar-H), 7.14 (d, J = 8.7 Hz, 1H, Ar-H), 3.87 (s, 3H, -O-CH<sub>3</sub>).

HR-MS m/z: 287.1028 [M+H]<sup>+</sup>; calculated C<sub>15</sub>H<sub>15</sub>N<sub>2</sub>O<sub>4</sub><sup>+</sup>: 287.1026

HPLC: rt 10.28 min (96.31%)

Yield: 70 mg; 0.25 mmol; 12.5%

3-Benzamido-4-fluoro-benzohydroxamate (**13d**)

MS m/z: 273.18 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.01 (s, 1H, -NH-OH), 9.44 (s, 1H, -CONH-Ar), 8.90 (s, 1H, -NH-OH), 8.22 (d, J = 1.9 Hz, 1H, Ar-H), 7.99–7.91 (m, 2H, Ar-H), 7.62 – 7.45 (m, 4H, Ar-H), 7.12 (d, J = 8.6 Hz, 1H, Ar-H).

HR-MS m/z: 273.0676 [M-H]<sup>-</sup>; calculated: C<sub>14</sub>H<sub>10</sub>FN<sub>2</sub>O<sub>3</sub><sup>-</sup>: 273.0681

HPLC: rt 8.47 min (98.23%)

Yield: 20 mg; 0.07 mmol; 4.7%

3-Benzamido-4-chloro-benzohydroxamate (**13e**)

MS m/z: 289.25 [M-H]<sup>-</sup>/ 291.15 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.32 (s, 1H, -NH-OH), 10.15 (s, 1H, -CONH-Ar), 9.10 (s, 1H, -NH-OH), 8.04 – 7.92 (m, 3H, Ar-H), 7.68 – 7.58 (m, 3H, Ar-H), 7.57 – 7.48 (m, 2H, Ar-H).

HR-MS m/z: 313.0351 [M+Na]<sup>+</sup>; calculated C<sub>14</sub>H<sub>11</sub><sup>35</sup>ClN<sub>2</sub>O<sub>3</sub>Na<sup>+</sup>: 313.0350

HPLC: rt 4.72 min (95.17%)

Yield: 170 mg; 0.58 mmol; 58%

3-Benzamido-4-bromo-benzohydroxamate (**13f**)

MS m/z: 334.95 [M-H]<sup>-</sup>/ 337.01 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.33 (s, 1H, -NH-OH) 10.13 (s, 1H, -CONH-Ar), 9.10 (s, 1H, -NH-OH), 7.99 (d, J = 7.4 Hz, 2H, Ar-H), 7.93 (s, 1H, Ar-H), 7.80 (d, J = 8.4 Hz, 1H, Ar-H), 7.64 – 7.51 (m, 4H, Ar-H).

HR-MS m/z: 332.9875 [M-H]<sup>-</sup>; calculated: C<sub>14</sub>H<sub>10</sub><sup>79</sup>BrN<sub>2</sub>O<sub>3</sub><sup>-</sup>: 332.9880

HPLC: rt 9.75 min (95.95%)

Yield: 120 mg; 0.36 mmol; 33%

3-Benzamido-4-trifluoromethyl-benzohydroxamate (**13g**)

MS m/z: 323.25[M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.49 (s, 1H, -NH-OH), 10.23 (s, 1H, -CONH-Ar), 9.24 (s, 1H, -NH-OH), 7.97 – 7.92 (m, 2H, Ar-H), 7.90 (d, J = 7.9 Hz, 3H, Ar-H), 7.64 – 7.58 (m, 1H, Ar-H), 7.58 – 7.50 (m, 2H, Ar-H).

HR-MS m/z: 323.0638 [M-H]<sup>-</sup>; calculated: C<sub>15</sub>H<sub>10</sub>F<sub>3</sub>N<sub>2</sub>O<sub>3</sub><sup>-</sup>: 323.0649

HPLC: rt 9.04 min (99.75%)

Yield: 75 mg; 0.23 mmol; 11.8%

3-Benzamido-4-ethoxy-benzohydroxamate (**13h**)

MS m/z: 300.73 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.01 (s, 1H, -NH-OH), 9.44 (s, 1H, -CONH-Ar), 8.90 (s, 1H, -NH-OH), 8.22 (d, J = 1.9 Hz, 1H, Ar-H), 7.94 (d, J = 7.3 Hz, 2H, Ar-H), 7.64 – 7.47 (m, 4H, Ar-H), 7.12 (d, J = 8.6 Hz, 1H, Ar-H), 4.14 (q, J = 6.9 Hz, 2H, -O-CH<sub>2</sub>-CH<sub>3</sub>), 1.35 (t, J = 6.9 Hz, 3H, -O-CH<sub>2</sub>-CH<sub>3</sub>).

HR-MS m/z: 299.1030 [M-H]<sup>-</sup>; calculated: C<sub>16</sub>H<sub>15</sub>N<sub>2</sub>O<sub>4</sub><sup>-</sup>: 299.1037

HPLC: rt 10.28 min (99.05%)

Yield: 140 mg; 0.47 mmol; 12.7%

3-Benzamido-4-propoxy-benzohydroxamate (**13i**)

MS m/z: 313.24 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.08 (s, 1H, -NH-OH), 9.46 (s, 1H, -CONH-Ar), 8.90 (s, 1H, -NH-OH), 8.17 (d, J = 1.9 Hz, 1H, Ar-H), 7.93 (d, J = 7.2 Hz, 2H, Ar-H), 7.62 – 7.49 (m, 4H, Ar-H), 7.12 (d, J = 8.6 Hz, 1H, Ar-H), 4.03 (t, J = 6.4 Hz, 2H, -O-CH<sub>2</sub>-CH<sub>2</sub>-), 1.78 – 1.70 (m, 2H, -CH<sub>2</sub>-CH<sub>2</sub>-CH<sub>3</sub>), 0.95 (t, J = 7.4 Hz, 3H, -CH<sub>2</sub>-CH<sub>2</sub>-CH<sub>3</sub>).

HR-MS m/z: 313.1187 [M-H]<sup>-</sup>; calculated: C<sub>17</sub>H<sub>17</sub>N<sub>2</sub>O<sub>4</sub><sup>-</sup>: 313.1194

HPLC: rt 10.59 min (99.23%)

Yield: 30 mg; 0.1 mmol; 7.5%

3-Benzamido-4-isopropoxy-benzohydroxamate (**13j**)

MS m/z: 313.14 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.06 (s, 1H, -NH-OH), 9.34 (s, 1H, -CONH-Ar), 8.88 (s, 1H, -NH-OH), 8.25 (d, J = 2.0 Hz, 1H, Ar-H), 7.92 (d, J = 7.4 Hz, 2H, Ar-H), 7.62 – 7.49 (m, 4H, Ar-H), 7.13 (d, J = 8.7 Hz, 1H, Ar-H), 4.71 (dt, J = 11.9, 5.9 Hz, 1H, -O-CH-(CH<sub>3</sub>)<sub>2</sub>), 1.30 (d, J = 6.0 Hz, 6H, -O-CH-(CH<sub>3</sub>)<sub>2</sub>).

HR-MS m/z: 313.1189 [M-H]<sup>-</sup>; calculated: C<sub>17</sub>H<sub>17</sub>N<sub>2</sub>O<sub>4</sub><sup>-</sup>: 313.1194

HPLC: rt 10.38 min (98.35%)

Yield: 80 mg; 0.25 mmol; 5%

3-(Quinaltin-2-amido)-4-methyl-benzohydroxamate (**13k**)

MS m/z: 320.03 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.18 (s, 1H, -NH-OH), 10.54 (s, 1H, -CONH-Ar), 8.97 (s, 1H, -NH-OH), 8.65 (d, J = 8.5 Hz, 1H, Ar-H), 8.31 – 8.23 (m, 2H, Ar-H), 8.21 (d, J = 8.5 Hz, 1H, Ar-H), 8.13 (d, J = 8.0 Hz, 1H, Ar-H), 7.96 – 7.86 (m, 1H, Ar-H), 7.79 – 7.72 (m, 1H, Ar-H), 7.53 – 7.57 (m, 1H, Ar-H), 7.37 (d, J = 7.9 Hz, 1H, Ar-H), 2.40 (s, 3H, Ar-CH<sub>3</sub>).

HR-MS m/z: 322.1186 [M+H]<sup>+</sup>; calculated C<sub>18</sub>H<sub>16</sub>N<sub>3</sub>O<sub>3</sub><sup>+</sup>: 322.1186

HPLC: rt 13.00 min (97.95%)

Yield: 180 mg; 0.59 mmol; 59%

3-(4-Biphenylamido)-4-methoxy-benzohydroxamate (**13l**)

MS m/z: 361.38 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.09 (s, 1H, -NH-OH), 9.58 (s, 1H, -CONH-Ar), 8.89 (s, 1H, -NH-OH), 8.18 (d, J = 2.1 Hz, 1H, Ar-H), 8.05 (d, J = 8.3 Hz, 2H, Ar-H), 7.82 (d, J = 8.3 Hz, 2H, Ar-H), 7.74 (d, J = 7.3 Hz, 2H, Ar-H), 7.66 – 7.59 (m, 1H, Ar-H), 7.50 (t, J = 7.6 Hz, 2H, Ar-H), 7.41 (t, J = 7.4 Hz, 1H, Ar-H), 7.14 (d, J = 8.6 Hz, 1H, Ar-H), 3.88 (s, 3H, -O-CH<sub>3</sub>).

HR-MS m/z: 363.1341 [M+H]<sup>+</sup>; calculated: C<sub>21</sub>H<sub>19</sub>N<sub>2</sub>O<sub>4</sub><sup>+</sup>: 363.1339

HPLC: rt 13.45 min (95.98%)

Yield: 50 mg; 0.14 mmol; 3.5%

3-(4-Methoxy-benzamido)-4-methoxy-benzohydroxamate (**13m**)

MS m/z: 315.34 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.06 (s, 1H, -NH-OH), 9.36 (s, 1H, -CONH-Ar), 8.88 (s, 1H, -NH-OH), 8.15 (d, J = 1.7 Hz, 1H, Ar-H), 7.94 (d, J = 8.7 Hz, 2H, Ar-H), 7.65 – 7.55 (m, 1H, Ar-H), 7.12 (d, J = 8.6 Hz, 1H, Ar-H), 7.04 (d, J = 8.7 Hz, 2H, Ar-H), 3.86 (s, 3H, Ar-O-CH<sub>3</sub>), 3.80 (s, 3H, Ar-O-CH<sub>3</sub>).

HR-MS m/z: 315.0981 [M-H]<sup>-</sup>; calculated: C<sub>16</sub>H<sub>15</sub>N<sub>2</sub>O<sub>5</sub><sup>-</sup>: 315.0986

HPLC: rt 8.96 min (95.35%)

Yield: 20 mg; 0.06 mmol; 6%

3-(4-Chloro-benzamido)-4-methoxy-benzohydroxamate (**13n**)

MS m/z: 319.30 [M-H]<sup>-</sup>/ 321.31 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.08 (s, 1H, -NH-OH), 9.66 (s, 1H, -CONH-Ar), 8.90 (s, 1H, -NH-OH), 8.10 (d, J = 1.9 Hz, 1H, Ar-H), 7.97 (d, J = 8.4 Hz, 2H, Ar-H), 7.66 – 7.54 (m, 3H, Ar-H), 7.13 (d, J = 8.6 Hz, 1H, Ar-H), 3.85 (s, 3H, -O-CH<sub>3</sub>).

HR-MS m/z: 319.0482 [M-H]<sup>-</sup>; calculated: C<sub>15</sub>H<sub>12</sub><sup>35</sup>ClN<sub>2</sub>O<sub>4</sub><sup>-</sup>: 319.0491

HPLC: rt 8.33 min (97.65%)

Yield: 12 mg; 0.04 mmol; 4%

3-(2-Chloro-benzamido)-4-methoxy-benzohydroxamate (**13o**)

MS m/z: 319.13 [M-H]<sup>-</sup>/ 321.13 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.09 (s, 1H, -NH-OH), 9.70 (s, 1H, -CONH-Ar), 8.88 (s, 1H, -NH-OH), 8.32 (s, 1H, Ar-H), 7.61 – 7.38 (m, 5H, Ar-H), 7.11 (d, J = 8.6 Hz, 1H, Ar-H), 3.84 (s, 3H, -O-CH<sub>3</sub>).

HR-MS m/z: 319.0484 [M-H]<sup>-</sup>; calculated: C<sub>15</sub>H<sub>12</sub><sup>35</sup>ClN<sub>2</sub>O<sub>4</sub><sup>-</sup>: 319.0491

HPLC: rt 9.12 min (96.34%)

Yield: 100 mg; 0.31 mmol; 38%

3-(2,4-Dichloro-benzamido)-4-methoxy-benzohydroxamate (**13p**)

MS m/z: 353.18 [M-H]<sup>-</sup> / 355.13 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.10 (s, 1H, -NH-OH), 9.82 (s, 1H, -CONH-Ar), 8.90 (s, 1H, -NH-OH), 8.33 (s, 1H, Ar-H), 7.71 (s, 1H, Ar-H), 7.63 – 7.45 (m, 3H, Ar-H), 7.11 (d, J = 8.5 Hz, 1H, Ar-H), 3.84 (s, 3H, -O-CH<sub>3</sub>).

HR-MS m/z: 353.0094 [M-H]<sup>-</sup>; calculated: C<sub>15</sub>H<sub>11</sub><sup>35</sup>Cl<sub>2</sub>N<sub>2</sub>O<sub>4</sub><sup>-</sup>: 353.0101

HPLC: rt 10.24 min (97.43%)

Yield: 200 mg; 0.57 mmol; 28.5%

3-(3-Biphenyl)-4-methoxy-benzohydroxamate (**13q**)

MS m/z: 361.26 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.11 (s, 1H, -NH-OH), 9.74 (s, 1H, -CONH-Ar), 8.92 (s, 1H, -NH-OH), 8.24 (s, 1H, Ar-H), 8.12 (d, J = 1.5 Hz, 1H, Ar-H), 7.93 (d, J = 7.6 Hz, 1H, Ar-H), 7.88 (d, J = 7.5 Hz, 1H, Ar-H), 7.76 (d, J = 7.5 Hz, 2H, Ar-H), 7.67 – 7.58 (m, 2H, Ar-H), 7.50 (t, J = 7.5 Hz, 2H, Ar-H), 7.40 (t, J = 7.2 Hz, 1H, Ar-H), 7.15 (d, J = 8.6 Hz, 1H, Ar-H), 3.86 (s, 3H, -O-CH<sub>3</sub>).

HR-MS m/z: 361.1182 [M-H]<sup>-</sup>; calculated: C<sub>21</sub>H<sub>17</sub>N<sub>2</sub>O<sub>4</sub><sup>-</sup>: 361.1194

HPLC: rt 12.08 min (99.00%)

Yield: 60 mg; 0.17 mmol; 8.5%

3-(4-Ethoxy-benzamido)-4-methoxy-benzohydroxamate (**13r**)

MS m/z: 329.22 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.06 (s, 1H, -NH-OH), 9.34 (s, 1H, -CONH-Ar), 8.87 (s, 1H, -NH-OH), 8.16 (d, J = 2.1 Hz, 1H, Ar-H), 7.92 (d, J = 8.8 Hz, 2H, Ar-H), 7.61 – 7.57 (m, 1H, Ar-H), 7.12 (d, J = 8.7 Hz, 1H, Ar-H), 7.02 (d, J = 8.9 Hz, 2H, Ar-H), 4.10 (q, J = 7.0 Hz, 2H, -O-CH<sub>2</sub>-CH<sub>3</sub>), 1.34 (t, J = 7.0 Hz, 3H, -O-CH<sub>2</sub>-CH<sub>3</sub>).

HR-MS m/z: 331.1289 [M+H]<sup>+</sup>; calculated: C<sub>17</sub>H<sub>19</sub>N<sub>2</sub>O<sub>5</sub><sup>+</sup>: 331.1289

HPLC: rt 10.01 min (98.72%)

Yield: 200 mg; 0.6 mmol; 30%

3-[(Phenylacetyl)amino]-4-methoxy-benzohydroxamate (**13s**)

MS m/z: 299.38 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.00 (s, 1H, -NH-OH), 9.33 (s, 1H, -CONH-Ar), 8.82 (s, 1H, -NH-OH), 8.31 (s, 1H, Ar-H), 7.47 (dd, J = 8.6, 2.0 Hz, 1H, Ar-H), 7.36 – 7.28 (m, 4H, Ar-H), 7.26 – 7.20 (m, 1H, Ar-H), 7.06 (d, J = 8.6 Hz, 1H, Ar-H), 3.86 (s, 3H, -O-CH<sub>3</sub>), 3.73 (s, 2H, -CH<sub>2</sub>-).

HR-MS m/z: 301.1187 [M+H]<sup>+</sup>; calculated: C<sub>16</sub>H<sub>17</sub>N<sub>2</sub>O<sub>4</sub><sup>+</sup>: 301.1183

HPLC: rt 8.58 min (99.31%)

Yield: 195 mg; 0.65 mmol; 32.5%

3-(4-Methoxy-benzamido)-4-chloro-benzohydroxamate (**13t**)

MS m/z: 320.99 [M+H]<sup>+</sup> /322.99 [M+H]<sup>+</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.31 (s, 1H, -NH-OH), 9.97 (s, 1H, -CONH-Ar), 9.09 (s, 1H, -NH-OH), 8.00 – 7.93 (m, 3H, Ar-H), 7.66 – 7.60 (m, 2H, Ar-H), 7.06 (d, J = 8.8 Hz, 2H, Ar-H), 3.83 (s, 3H, Ar-O-CH<sub>3</sub>).

HR-MS m/z: 319.0488 [M-H]<sup>-</sup>; calculated: C<sub>15</sub>H<sub>12</sub><sup>35</sup>ClN<sub>2</sub>O<sub>4</sub><sup>-</sup>: 319.0491

HPLC: rt 9.62 min (97.22%)

Yield: 50 mg; 0.16 mmol; 8.2%

3-(3-Benzyloxy-benzamido)-4-chloro-benzohydroxamate (**13u**)

MS m/z: 395.17 [M-H]<sup>-</sup> /397.15 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.32 (s, 1H, -NH-OH), 10.13 (s, 1H, -CONH-Ar), 9.10 (s, 1H, -NH-OH), 7.96 (s, 1H, Ar-H), 7.64 (s, 2H, Ar-H), 7.63 – 7.59 (m, 1H, Ar-H), 7.57 (d, J = 7.8 Hz, 1H, Ar-H), 7.50 – 7.42 (m, 3H, Ar-H), 7.39 (t, J = 7.3 Hz, 2H, Ar-H), 7.36 – 7.30 (m, J = 7.2 Hz, 1H, Ar-H), 7.28 – 7.22 (m, 1H, Ar-H), 5.18 (s, 2H, Ar-CH<sub>2</sub>-O-Ar).

HR-MS m/z: 395.0797 [M-H]<sup>-</sup>; calculated: C<sub>21</sub>H<sub>16</sub><sup>35</sup>ClN<sub>2</sub>O<sub>4</sub><sup>-</sup>: 395.0804

HPLC: rt 12.25 min (99.49%)

Yield: 50 mg; 0.13 mmol; 5.5%

3-(3-Phenoxy-benzamido)-4-chloro-benzohydroxamate (**13v**)

MS m/z: 381.20 [M-H]<sup>-</sup> /383.20 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.31 (s, 1H, -NH-OH), 10.21 (s, 1H, -CONH-Ar), 9.10 (s, 1H, -NH-OH), 7.93 (s, 1H, Ar-H), 7.76 (d, J = 7.8 Hz, 1H, Ar-H), 7.65 – 7.62 (m, 2H, Ar-H), 7.60 – 7.56 (m, 1H, Ar-H), 7.54 (d, J = 7.9 Hz, 1H, Ar-H), 7.42 (t, J = 7.9 Hz, 2H, Ar-H), 7.27 – 7.23 (m, 1H, Ar-H), 7.18 (t, J = 7.4 Hz, 1H, Ar-H), 7.07 (d, J = 7.9 Hz, 2H, Ar-H).

HR-MS m/z: 381.0641 [M-H]<sup>-</sup>; calculated: C<sub>20</sub>H<sub>14</sub><sup>35</sup>ClN<sub>2</sub>O<sub>4</sub><sup>-</sup>: 381.0648

HPLC: rt 12.18 min (99.83%)

Yield: 165 mg; 0.43 mmol; 19.3%

3-(4-Phenoxy-benzamido)-4-chloro-benzohydroxamate (**13w**)

MS m/z: 381.17 [M-H]<sup>-</sup> /383.17 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.32 (s, 1H, -NH-OH), 9.91 (s, 1H, -CONH-Ar), 9.10 (s, 1H, -NH-OH), 8.02 (d, J = 8.8 Hz, 2H, Ar-H), 7.96 (s, 1H, Ar-H), 7.66 – 7.60 (m, 2H, Ar-H), 7.45 (t, J = 7.9 Hz, 2H, Ar-H), 7.22 (t, J = 7.4 Hz, 1H, Ar-H), 7.10 (t, J = 8.0 Hz, 4H, Ar-H).

HR-MS m/z: 381.0638 [M-H]<sup>-</sup>; calculated: C<sub>20</sub>H<sub>14</sub><sup>35</sup>ClN<sub>2</sub>O<sub>4</sub><sup>-</sup>: 381.0648

HPLC: rt 12.35 min (99.00%)

Yield: 10mg; 0.03 mmol; 0.59%

3-(4-Chloro-benzamido)-4-chloro-benzohydroxamate (**13x**)

MS m/z: 323.15 [M-H]<sup>-</sup> /325.26 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.32 (s, 1H, -NH-OH), 10.26 (s, 1H, -CONH-Ar), 9.11 (s, 1H, -NH-OH), 8.05 – 7.92 (m, 3H, Ar-H), 7.67 – 7.59 (m, 4H, Ar-H).

HR-MS m/z: 322.9992 [M-H]<sup>-</sup>; calculated: C<sub>14</sub>H<sub>9</sub><sup>35</sup>Cl<sub>2</sub>N<sub>2</sub>O<sub>3</sub><sup>-</sup>: 322.9996

HPLC: rt 10.46 min (97.07%)

Yield: 140 mg; 0.43 mmol; 21.5%

3-(4-Nitro-benzamido)-4-chloro-benzohydroxamate (**13y**)

MS m/z: 334.27 [M-H]<sup>-</sup> /336.28 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.34 (s, 1H, -NH-OH), 10.55 (s, 1H, -CONH-Ar), 9.12 (s, 1H, -NH-OH), 8.38 (d, J = 8.6 Hz, 2H, Ar-H), 8.20 (d, J = 8.5 Hz, 2H, Ar-H), 8.00 – 7.93 (m, 1H, Ar-H), 7.69 – 7.64 (m, 2H, Ar-H).

HR-MS m/z: 334.0227 [M-H]<sup>-</sup>; calculated: C<sub>14</sub>H<sub>9</sub><sup>35</sup>ClN<sub>3</sub>O<sub>5</sub><sup>-</sup>: 334.0236

HPLC: rt 9.15 min (97.24%)

Yield: 30 mg; 0.09 mmol; 4.5%

3-(2,4-Dichloro-benzamido)-4-chloro-benzohydroxamate (**13z**)

MS m/z: 357.67 [M-H]<sup>-</sup> /359.44 [M-H]<sup>-</sup> /361.37 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.34 (s, 1H, -NH-OH), 10.37 (s, 1H, -CONH-Ar), 9.11 (s, 1H, -NH-OH), 8.06 (s, 1H, Ar-H), 7.76 (d, J = 1.6 Hz, 1H, Ar-H), 7.68 – 7.60 (m, J = 6.0 Hz, 3H, Ar-H), 7.58 – 7.54 (m, 1H, Ar-H).

HR-MS m/z: 358.9753 [M+H]<sup>+</sup>; calculated: C<sub>14</sub>H<sub>10</sub><sup>35</sup>Cl<sub>3</sub>N<sub>2</sub>O<sub>3</sub><sup>+</sup>: 358.9752

HPLC: rt 8.60 min (99.89%)

Yield: 30 mg; 0.09 mmol; 4.5%

3-(4-Biphenylamido)-4-ethoxy-benzohydroxamate (**13za**)

MS m/z: 375.24 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.09 (s, 1H, -NH-OH), 9.51 (s, 1H, -CONH-Ar), 8.90 (s, 1H, -NH-OH), 8.23 (d, J = 1.4 Hz, 1H, Ar-H), 8.04 (d, J = 8.1 Hz, 2H, Ar-H), 7.83 (d, J = 8.1 Hz, 2H, Ar-H), 7.75 (d, J = 7.5 Hz, 2H, Ar-H), 7.62 – 7.56 (m, 1H, Ar-H), 7.50 (t, J = 7.4 Hz, 2H, Ar-H), 7.44 – 7.37 (m, 1H, Ar-H), 7.13 (d, J = 8.6 Hz, 1H, Ar-H), 4.15 (q, J = 6.7 Hz, 2H, -O-CH<sub>2</sub>-CH<sub>3</sub>), 1.36 (t, J = 6.8 Hz, 3H, -O-CH<sub>2</sub>-CH<sub>3</sub>).

HR-MS m/z: 375.1341 [M-H]<sup>-</sup>; calculated: C<sub>22</sub>H<sub>19</sub>N<sub>2</sub>O<sub>4</sub><sup>-</sup>: 375.1350

HPLC: rt 12.66 min (98.10%)

Yield: 60 mg; 0.16 mmol; 8%

Quinaltic-2-hydroxamate (**14a**)

MS m/z: 189.11 [M+H]<sup>+</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ (bs, 1H, -HN-OH), 8.53 (d, J = 8.5 Hz, 1H, Ar-H), 8.10 – 8.02 (m, 3H, Ar-H), 7.88 – 7.80 (m, 1H, Ar-H), 7.73 – 7.65 (m, 1H, Ar-H).

HR-MS m/z: 189.0657 [M+H]<sup>+</sup>; calculated C<sub>10</sub>H<sub>9</sub>N<sub>2</sub>O<sub>2</sub><sup>+</sup>: 189.0659

HPLC: rt 8.70 min (96.06%)

Yield: 65 mg; 0.35 mmol; 35%

3-Benzyloxy-4-methoxy-benzohydroxamate (**15a**)

MS m/z: 274.04 [M+H]<sup>+</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.03 (s, 1H, -HN-OH), 8.87 (s, 1H, -NH-OH), 7.47 – 7.41 (m, 3H, Ar-H), 7.41 – 7.35 (m, 3H, Ar-H), 7.35 – 7.29 (m, 1H, Ar-H), 7.01 (d, J = 8.4 Hz, 1H, Ar-H), 5.09 (s, 2H, O-CH<sub>2</sub>-Ar), 3.79 (s, 3H, Ar-O-CH<sub>3</sub>).

HR-MS m/z: 274.1073 [M+H]<sup>+</sup>; calculated C<sub>15</sub>H<sub>16</sub>NO<sub>4</sub><sup>+</sup>: 274.1074

HPLC: rt 10.85 min (98.46%)

Yield: 110 mg; 0.4 mmol; 40.25%

3-(N-para-touluenyl)-sulfonamido-4-methoxy-benzohydroxamate (**16a**)

MS m/z: 335.32 [M-H]<sup>-</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.25 (s, 1H, -HN-OH), 9.87 (s, 1H, -CONH), 8.98 (s, 1H, -NH-OH), 8.13 (s, 1H, Ar-H), 7.90 (d, J = 7.0 Hz, 1H, Ar-H), 7.20 (d, J = 8.7 Hz, 1H, Ar-H), 7.03 – 6.87 (m, 3H, Ar-H), 4.03 – 3.83 (m, 3H, -O-CH<sub>3</sub>), 2.12 (s, 3H, Ar-CH<sub>3</sub>).

HR-MS m/z: 337.0855 [M+H]<sup>+</sup>; calculated C<sub>15</sub>H<sub>17</sub>N<sub>2</sub>O<sub>5</sub><sup>+</sup>: 337.0853

HPLC: rt 8.24 min (99.51%)

Yield: 25 mg; 0.08 mmol; 25.8%

3-(4-Biphenylamido)-4-methoxy-methylbenzoate (**17a**)

MS m/z: 362.04 [M+H]<sup>+</sup>

<sup>1</sup>H NMR: (400 MHz, DMSO) δ 9.60 (s, 1H, -CONH), 8.44 (d, J = 2.2 Hz, 1H, Ar-H), 8.05 (d, J = 8.4 Hz, 2H, Ar-H), 7.87 – 7.79 (m, 3H, Ar-H), 7.79 – 7.71 (m, 2H, Ar-H), 7.50 (t, J = 7.5 Hz, 2H, Ar-H), 7.41 (t, J = 7.4 Hz, 1H, Ar-H), 7.22 (d, J = 8.7 Hz, 1H, Ar-H), 3.93 (s, 3H, -O-CH<sub>3</sub>), 3.83 (s, 3H, -COOCH<sub>3</sub>).

HPLC: rt 14.08 min. (95.79%)

Yield: 900 mg; 2.49 mmol; 71.15%

3-(4-Biphenylamido)-4-methoxy-benzoic acid (**17b**)

MS m/z: 348.06 [M+H]<sup>+</sup>

<sup>1</sup>H NMR (400 MHz, DMSO) δ 12.67 (s, 1H, -COOH), 9.57 (s, 1H, -CONH), 8.39 (d, J = 2.1 Hz, 1H, Ar-H), 8.05 (d, J = 8.4 Hz, 2H, Ar-H), 7.88 – 7.77 (m, 3H, Ar-H), 7.75 (d, J = 7.0 Hz, 2H, Ar-H), 7.50 (t, J = 7.6 Hz, 2H, Ar-H), 7.42 (d, J = 7.2 Hz, 1H, Ar-H), 7.19 (d, J = 8.8 Hz, 1H, Ar-H), 3.91 (s, 3H, -O-CH<sub>3</sub>).

HPLC: rt 13.52 min (95.47%)

Yield: 835 mg; 2.41 mmol; 96.79%



- (1) Abdel-Magid, A. F.; Mehrman, S. J. A Review on the Use of Sodium Triacetoxyborohydride in the Reductive Amination of Ketones and Aldehydes. *Org. Process Res. Dev.* **2006**, *10*, 971–1031.
- (2) Kocienski, P. J. *Protecting Groups*, 3<sup>rd</sup> ed.: Georg Thieme Verlag: · New York, 2005; p 206.
- (3) Becker, R.; Fanghänel, E.; Habicher, W. D. *Organikum*, 22<sup>nd</sup> ed.: WILEY-VCH Verlag GmbH: Weinheim, 2004; p 498.
- (4) Kim, M. H.; Patel, D. V. “BOP” as a Reagent for Mild and Efficient Preparation of Esters. *Tetrahedron Lett.* **1994**, *35*, 5603–5606.
- (5) Lenguyen, D.; Castro, B. D. LeNguyen. *Tetrahedron Lett.* **1990**, *31*, 205–208.
- (6) Reddy, A. S.; Kumar, M. S.; Reddy, G. R. A Convenient Method for the Preparation of Hydroxamic Acids. *Tetrahedron Lett.* **2000**, *41*, 6285–6288.
- (7) Misra, R. N.; Botti, C. M.; Haslanger, M. F.; Engebrecht, J. R.; Mahoney, E. M.; Ciosek, C. P. Cyclic Aryl Hydroxamic Acids: Synthesis and Inhibition of 5-Lipoxygenase. *Bioorg. Med. Chem. Lett.* **1991**, *1*, 295–298.
- (8) Suzuki, T.; Ota, Y.; Ri, M.; Bando, M.; Gotoh, A.; Itoh, Y.; Tsumoto, H.; Tatum, P. R.; Mizukami, T.; Nakagawa, H.; Iida, S.; Ueda, R.; Shirahige, K.; Miyata, N. Rapid Discovery of Highly Potent and Selective Inhibitors of Histone Deacetylase 8 Using Click Chemistry to Generate Candidate Libraries. *J. Med. Chem.* **2012**, *55*, 9562–9575.

- (9) Becker, R.; Fanghänel, E.; Habicher, W. D. *Organikum*, 22<sup>nd</sup> ed.: WILEY-VCH Verlag GmbH: Weinheim, 2004; pp 358–359.
- (10) Pflästerer, D.; Dolbundalchok, P.; Rafique, S.; Rudolph, M.; Rominger, F.; Hashmi, a. S. K. On the Gold-Catalyzed Intramolecular 7- Exo-Trig Hydroamination of Allenes. *Adv. Synth. Catal.* **2013**, 355, 1383–1393.

### 3.2.3. Publication 2

#### **Summary of publication 2:**

Several smHDAC8 specific inhibitors and HDAC8 selective inhibitors were used to understand the basis of HDAC8 selective inhibition. Crystal packing in hHDAC8 crystals is made in head to head manner, where active site residues and inhibitors are involved in crystal packing and make difficult to study the mechanism of selective inhibition in a non crystal-biased way. In contrast, smHDAC8 provides a tool to pursue these studies since the lack of crystal contacts at the active site make it an ideal system for inhibitor binding studies.

Crystal structures of smHDAC8 with hHDAC8-selective and with TH compounds (benzhydroxamate derivatives) consistently showed that the binding onto smHDAC8 catalytic tyrosine Y341 and additional contacts with active site loop L6 contribute to HDAC8 selective inhibition. Specifically, HDAC8 crystal structure with pan-HDAC inhibitor quisinostat and smHDAC8 crystal structures with TH compounds and hHDAC8-selective inhibitors showed that small changes at the inhibitor level can influence dramatically inhibitor conformation without changing the binding pocket. In addition, mutational analysis of loops composing the active site revealed the importance of the loops conformation, not only for inhibitor binding but also for activity. In conclusion, this article provides comprehensive information on the selective inhibition of HDAC8.

#### **Contribution:**

I have done all the structural work in this article (except TH33, TH39 and TH65), production of enzymes, crystallization, data collection at synchrotrons and data processing, and contributed to analysis of the data. I have also purified the proteins which were used for *in vitro* biochemical assays by our collaborators.

## Structural basis of HDAC8 selective inhibition

Marek, M.<sup>1,2,\*§</sup>, Shaik, T.B.<sup>1,\*</sup>, Heimbürg, T.<sup>3</sup>, Chakrabarti, A.<sup>4</sup>, Lancelot, J.<sup>5</sup>, Ramos-Morales, E.<sup>1</sup>, Da Veiga, C.<sup>6</sup>, Kalinin, D.<sup>7</sup>, Melesina, J.<sup>3</sup>, Robaa, D.<sup>3</sup>, Schmidtkunz, K.<sup>4</sup>, Suzuki, T.<sup>8</sup>, Holl, R.<sup>9</sup>, Ennifar, E.<sup>6</sup>, Pierce, R.J.<sup>5</sup>, Jung, M.<sup>4</sup>, Sippl, W.<sup>3</sup> & Romier, C.<sup>§,1</sup>

<sup>1</sup>Département de Biologie Structurale Intégrative, Institut de Génétique et Biologie Moléculaire et Cellulaire (IGBMC), Université de Strasbourg, CNRS, INSERM, 1 rue Laurent Fries, 67404 Illkirch Cedex, France.

<sup>2</sup>Current address : Loschmidt Laboratories, Department of Experimental Biology & RECETOX, Faculty of Science, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic.

<sup>3</sup>Institute of Pharmacy, Martin-Luther-Universität Halle-Wittenberg, Wolfgang-Langenbeck-Straße 4, 06120 Halle/Saale, Germany.

<sup>4</sup>Institute of Pharmaceutical Sciences, Albert-Ludwigs-Universität Freiburg, Albertstraße 25, 79104 Freiburg, Germany.

<sup>5</sup>Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, U1019 - UMR 8204- CIIL -Centre d'Infection et d'Immunité de Lille, F-59000 Lille, France.

<sup>6</sup>Architecture et Réactivité de l'ARN, Institut de Biologie Moléculaire et Cellulaire (IBMC), UPR 9004 du CNRS, Université de Strasbourg, 15 Rue René Descartes, 67084 Strasbourg Cedex, France.

<sup>7</sup>Institute of Pharmaceutical and Medicinal Chemistry, University of Münster, Corrensstr. 48, 48149 Münster, Germany

<sup>8</sup>Graduate School of Medical Science, Kyoto Prefectural University of Medicine, 1-5 Shimogamohangi-Cho, Sakyo-Ku, 606-0823 Kyoto, Japan

<sup>9</sup>Institute of Organic Chemistry, Department of Chemistry, University of Hamburg, Martin-Luther-King-Platz 6, 20146 Hamburg, Germany

\*These authors have contributed equally to this work.

§Correspondence should be addressed to Martin Marek (martin.marek@recetox.muni.cz) and Christophe Romier (romier@igbmc.fr).

## **Abstract**

Currently approved epigenetic drugs (epidrugs) target mainly zinc-dependent histone deacetylases (HDACs). Yet, cross-reactivity of these drugs for the structurally similar but functionally different HDAC isozymes hampers their broad usage in clinical settings. Selective inhibitors targeting single HDAC isozymes are being developed, but our precise understanding in molecular terms of their selectivity remains sparse. PCI-34051 and NCC-149 were among the first inhibitors displaying isozyme-specificity for their target, HDAC8. Here, we show how HDAC8-selective inhibitors build their selectivity on specific interactions, notably with the HDAC8 active site catalytic tyrosine, but also through contacts with the HDAC8 L6 loop that is forming, together with L1 loop, a HDAC8-specific pocket. These interactions are enabled by the specific size and conformation of HDAC8 L1 and L6 loops, which leave the catalytic tyrosine uncovered, and by the constrained L-shape of HDAC8-selective inhibitors. Collectively, our results highlight the importance of HDAC active site loops and architecture, and pave the way for the design of next-generation selective HDAC inhibitors.

## Introduction

Acetylation of lysine residues in proteins is a major signaling mark that impacts most cellular processes<sup>1-3</sup>. In the cell nucleus, acetylation of histones has been shown to be essential for modulating chromatin structure and for acting in epigenetic signaling that drives and regulates nuclear mechanisms and development<sup>1,3,4</sup>. Protein lysine acetylation is a reversible process relying on the opposing effect of acetyltransferases and deacetylases<sup>3,5,6</sup>. In addition, the acetylation marks on lysines are recognized by epigenetic readers harboring structural modules (e.g. bromodomains) that enable the recruitment of cellular effectors to specific subcellular and genomic loci<sup>5,7</sup>.

Due to the functional importance of acetylation mechanisms, deregulation of these mechanisms has been linked with multiple human diseases, including cancer<sup>4,8-11</sup>. The reversibility of acetylation and the possibility of modulating recognition of acetylated lysines by bromodomains provide a way to act on acetylation pathways. Thus, epigenetic effectors involved in these pathways represent important therapeutics targets<sup>4,8-10,12,13</sup>.

Accordingly, among the currently approved epigenetic drugs, a majority (Vorinostat (SAHA), Romidepsin, Belinostat, Panobinostat and Chidamide) target lysine deacetylases<sup>13-15</sup>. The family of lysine deacetylases has been divided into four classes depending on their folds and their sequence similarities. Classes I, II (IIa and IIb), and IV adopt an arginase-deacetylase  $\alpha/\beta$  fold and rely on a zinc ion for activity (thereafter referred to as histone deacetylases or HDACs)<sup>6</sup>. The class III deacetylases are referred to as sirtuins and adopt a Rossmann fold, relying on NAD<sup>+</sup> for activity<sup>6</sup>. Eleven HDACs and seven sirtuins are found in human.

The currently approved drugs against lysine deacetylases target only proteins from the HDAC family. Yet, these drugs show poor selectivity against a single member of the structurally similar but functionally different human HDAC isozymes, thus hampering their broader therapeutic usage<sup>13,15</sup>.

Several small-molecule inhibitors exhibiting selectivity for specific HDACs have been developed. PCI-34051 and NCC-149 were among the first HDAC isozyme-selective inhibitors discovered<sup>16,17</sup>. These two aromatic hydroxamate derivatives show high selectivity for human HDAC8 (hHDAC8), an HDAC isozyme that has been shown to be overexpressed in several cancers<sup>18-20</sup> and whose mutations can lead to the Cornelia de Lange syndrome<sup>21-23</sup>. Specifically, PCI-34051, which is an indole-based derivative, is currently the most selective HDAC8 inhibitor with a selectivity index of 290 and 400 for HDAC6 and HDAC1, respectively, making it a strong chemical tool for studying the biological role of HDAC8 in vivo<sup>16,24-27</sup>. In addition, our work on HDAC8 from the human-pathogenic flatworm *Schistosoma mansoni* (smHDAC8) has led to the design of new selective HDAC8 inhibitors<sup>28,29</sup>.

To date, the structural and mechanistic basis underlying HDAC-selective inhibition remain poorly understood. To address this issue, we have dissected the molecular basis of HDAC8-selective inhibition by combining biochemical, biophysical and crystallographic studies on hHDAC8 and smHDAC8. Our

results reveal that HDAC8-selective inhibitors bind into a specific HDAC8-selective pocket formed by the active site catalytic tyrosine and by residues from L1 and L6 loops. This specific enzyme-ligand recognition is favoured by the constrained L-shaped conformation of HDAC8-selective inhibitors. This selective binding relies on a specific conformation of HDAC8 L6 loop and a shorter L1 loop that are not observed in any other HDAC isozymes. Collectively, our results highlight the structural/functional similarities and dissimilarities between the various HDAC isozymes and pave the way for the development of new HDAC isozyme-selective inhibitors to treat human diseases.

## Results

### *In vitro* and *in vivo* effects of **PCI-34051** and **NCC-149**

**PCI-34051** and **NCC-149** have been developed to target human HDAC8 (hHDAC8) selectively<sup>16,17</sup>. To investigate whether these inhibitors also target smHDAC8, we have looked at their inhibition and binding to smHDAC8. As a comparison, we have used the highly potent but non-selective Phase II HDAC inhibitor Quisinostat (**QSN**)<sup>30,31</sup>.

Measurements of the maximal-half inhibitory concentration ( $IC_{50}$ ) showed that all three inhibitors possess inhibitory activity in the submicromolar range against hHDAC8 and smHDAC8. **NCC-149** showed the most potent inhibition, followed by **QSN** and **PCI-34051** (Fig. 1). Measurement of the thermodynamic parameters using Isothermal Titration Calorimetry (ITC) confirmed the inhibition results obtained, the equilibrium dissociation constant ( $K_d$ ) values determined being in the same range that the corresponding  $IC_{50}$  values, with the exception of **QSN** that showed a lower  $K_d$  value for smHDAC8 (Fig. 1; Suppl. Fig. 1).

The biological effects, especially anti-cancer properties of **PCI-34051** and **NCC-149** in various cell types have been well characterized<sup>16,17,24-27</sup>, and we have previously shown that pan-HDAC inhibitors affect schistosome pathogens<sup>29,32</sup>. We therefore asked whether **PCI-34051** and **NCC-149** could also have anti-parasitic effect on schistosomes. Our various biological assays confirmed that **PCI-34051** and **NCC-149** affect the pathogens, triggering their apoptosis (Suppl. Fig. 2). These results demonstrated that both hHDAC8 and smHDAC8 can be used for studying HDAC8 inhibition by **PCI-34051** and **NCC-149**.

So far, few structures of HDACs in complex with selective inhibitors have been solved. Moreover, in many HDAC/inhibitor structures, the active site of the HDAC and the bound inhibitor are involved in extensive crystal packing contacts. This complicates the delineation between biologically-relevant and crystal packing-driven HDAC-inhibitor interactions. Therefore, in addition to the co-crystallization attempts of hHDAC8 with inhibitors, we have used the possibility offered by apo smHDAC8 crystals to look at HDAC8/inhibitor interactions in a crystal lattice-open environment<sup>29</sup>.

Despite intensive efforts, we were not able to obtain well-diffracting crystals of hHDAC8 in complex with **PCI-34051**, **NCC-149** and **QSN**. In contrast, soaking experiments of apo smHDAC8 crystals with all

three inhibitors were successful and yielded high resolution structures in complex with smHDAC8 (Fig. 2; Suppl. Table 1).

#### Binding mode of QSN to HDAC8

Analysis of the smHDAC8/**QSN** structure revealed that **QSN** adopts a straight conformation as its piperidine-pyrimidine linker allows limited conformational flexibility (Fig. 2c). **QSN** hydroxamate warhead coordinates the catalytic zinc and simultaneously interacts via hydrogen bonding with the histidine dyad, H141 and H142 (hHDAC8 H142 and H143) and with the catalytic tyrosine Y341 (hHDAC8 Y306) hydroxyl, as commonly observed for most other hydroxamate-containing HDAC inhibitors. Furthermore, **QSN** piperidine-pyrimidine linker is sandwiched between the side chains of smHDAC8 F151 (hHDAC8 F152) and F216 (hHDAC8 F208), where it forms planar  $\pi$ - $\pi$  stacking and non-polar contacts.

Specifically, **QSN** piperidine ring adopts a chair conformation, which allows **QSN** methyl-amino-methyl linker to form a hydrogen bond (2.4 Å) with the carboxyl group of smHDAC8 D100 (hHDAC8 D101), a conserved class I HDAC residue that has been shown to interact with the backbone of incoming acetylated peptides<sup>33,34</sup>. Finally, **QSN** capping methyl-indole group is solvent exposed, making minimal non-polar contacts with Y99 (hHDAC8 Y100). Interestingly, we previously observed a very similar binding mode to smHDAC8 for the other pan-HDAC inhibitor M344, including an interaction between D100 and the M344 internal amide group<sup>29</sup>. The M344 conformation is less constrained by its linker, which suggests that this binding mode is common to and favoured by many pan-HDAC inhibitors. In agreement, such a binding was also observed upon SAHA binding to hHDAC2 in a crystal-lattice open environment<sup>35</sup>.

#### Binding mode of PCI-34051 to HDAC8

**PCI-34051** hydroxamate warhead interacts with the catalytic zinc and active site residues as observed for **QSN** (Fig. 2a). However, in contrast to **QSN**, the hinge connecting the central indole-based spacer and the methoxyphenyl group of **PCI-34051** favours binding of its capping group onto the side chain of smHDAC8 Y341 (hHDAC8 Y306). This tyrosine, together with the catalytic zinc, has been shown in hHDAC8 to be involved in catalysis by polarizing the leaving acetyl group of the incoming acetylated lysine<sup>33</sup>. Here, the methoxyphenyl capping group is perpendicularly (86°) oriented over the aromatic ring of this tyrosine, which favours T-shaped  $\pi$ - $\pi$  stacking (4.9 Å). Thus, the binding of **PCI-34051** onto Y341 is favoured by the L-shape of this inhibitor.

Further, the methoxyphenyl capping group of **PCI-34051** is positioned in close vicinity to the smHDAC8 L6 loop, being inserted in a small pocket shaped by the side chains of P291 and H292 (hHDAC8 P273 and M274). While the methoxy group forms non-polar contacts with the pyrrolidine ring of P291, the phenyl ring of the inhibitor interacts (4.3 Å) via either  $\pi$ - $\pi$  or cation- $\pi$  interaction with H292, depending



upon the protonation state of the histidine residue (Fig. 2a). These interactions complement the aromatic interaction made with Y341.

smHDAC8 and hHDAC8 differ by one residue in their active sites, where hHDAC8 M274 is replaced by smHDAC8 H292. Since this latter residue is involved in inhibitor binding, we asked whether the smHDAC8-H292M mutant binds **PCI-34051** in the same way as the wild-type (WT) enzyme. The crystal structure of **PCI-34051** bound to the “humanized” smHDAC8-H292M mutant reveals that **PCI-34051** still adopts an L-shaped conformation when bound to the smHDAC8-H292M mutant, but this conformation is slightly different from the one adopted with the WT enzyme (Suppl. Fig. 3a-c; Suppl. Table 1). Specifically, **PCI-34051** still lies over Y341 catalytic tyrosine but appears more centred in the pocket created by smHDAC8 Y341, F151 and the L6 loop. In contrast to the WT enzyme, the central indole group of **PCI-34051** is axially rotated by an angle of  $\sim 20^\circ$ , which favours the positioning of the capping methoxyphenyl group over the aromatic ring of Y341 (4.8 Å), effecting nearly parallel ( $8.9^\circ$ )  $\pi$ - $\pi$  stacking. As a consequence, **PCI-34051** capping group is interacting differently with the L6 loop than observed with the WT enzyme, still making close non-polar contacts with the aliphatic ring of P291 and the side chain of M292 in this loop (Suppl. Fig. 3a).

#### Binding mode of **NCC-149** to HDAC8

**NCC-149** hydroxamate also binds in a canonical way to the catalytic zinc and active site residues, and the rest of the L-shaped inhibitor is turned towards and interacts with Y341 and the smHDAC8 L6 loop (Fig. 2b). Specifically, the 1,2,3-triazole ring of the linker is oriented in a position (4.9 Å) that is slightly off perpendicular ( $\sim 83^\circ$ ) to the aromatic ring of Y341, indicating their  $\pi$ - $\pi$  contacts. At the same time, the 1,2,3-triazole ring packs against L6 loop H292, which allows their mutual T-shaped ( $\sim 67^\circ$ ) aromatic interactions. In addition, and as observed for **PCI-34051**, the phenylthiomethyl capping group of **NCC-149** is inserted in the small subpocket of HDAC8 L6 loop, where it makes both upright ( $\sim 76^\circ$ )  $\pi$ - $\pi$  stacking with H292 and hydrophobic contacts (3.6 Å) with P291.

We also solved the structure of **NCC-149** bound to the smHDAC8-H292M mutant. Here, the hydroxamate and linker of **NCC-149** bind very similarly to the smHDAC8-H292M mutant and to the WT enzyme, and show fewer conformational changes than observed with **PCI-34051**. Interestingly, the 1,2,3-triazole ring is closer to the L6 loop, where it interacts with M292 via a sulphur-aromatic interaction (3.7 Å), suggesting a similar interaction with hHDAC8 (Suppl. Fig. 3d-f). This binding mode still favours T-shaped ( $82^\circ$ )  $\pi$ - $\pi$  stacking between the 1,2,3-triazole and Y341 (4.7 Å), as well as hydrophobic contacts between the internal benzene ring and the two phenylalanines, F151 and F216. However, the terminal phenylthiomethyl capping group of the inhibitor changes its position and is turned away from the L6 loop, lying in another binding subpocket formed by smHDAC8 K20 and F21

(L1 loop) and Y341 and F343 (L7 Loop) whose hydrophobic character is conserved in hHDAC8 (Suppl. Fig. 3d).

The conformational adaptation of **NCC-149** to the smHDAC8-H292M selective pocket is eased by the intrinsically higher conformational flexibility of this inhibitor that allows the repositioning of its capping group. In the case of **PCI-34051**, which is more rigid as it contains only a one-atom hinge, a major part of the inhibitor had to be repositioned. Yet, these changes do not affect the major interaction of HDAC8-selective inhibitors with the uncovered aromatic ring of the catalytic tyrosine and with residues of the L6 loop showing the importance of these elements as key binding surfaces for these inhibitors.

#### Selective inhibition of smHDAC8 over other human HDACs

Previous work on the selective inhibition of smHDAC8 has yielded the development of an inhibitor series of 3-benzamido-benzohydroxamates that show strong selectivity for smHDAC8 and hHDAC8 over other human HDACs<sup>28</sup>. The structure of smHDAC8 with the simplest inhibitor of this series (**1**) reveals that the capping benzamido moiety binds to HDAC8 in a similar way to **PCI-34051** and **NCC-149**, laying over Y341 (3.9 Å), its benzene ring capping group further making non-polar contacts with smHDAC8 L6 loop, notably with P291 (3.6 Å) (Fig. 3a). This inhibitor also exploits smHDAC8-specific interactions with residues K20 and H292 (hHDAC8 K33 and M274)<sup>28</sup>.

Many of the 3-benzamido-benzohydroxamate inhibitors that were subsequently developed displayed higher potency than **1** in inhibiting smHDAC8<sup>28</sup>. To understand the molecular basis of these observations, we have further solved the structures of smHDAC8 bound to several of these inhibitors (compounds **2-14**) (Fig. 3; Suppl. Fig. 4; Suppl. Table 2; Suppl. Table 3).

All these compounds showed a similar mode of binding to smHDAC8 than **1**. Yet, slight differences in chemical composition impacted specific interactions, potentially relating to the differences in the IC<sub>50</sub> values observed. Compounds **2-4** only have different substitutions at the para position of the benzohydroxamate moiety compared to **1**. These compounds bind very similarly to smHDAC8 as **1** (Suppl. Fig. 4). Their lower IC<sub>50</sub> values most likely stem from the additional contacts of their substituent groups with F216, as well as the possible stabilization by these groups of the non-canonical geometry of the amide group of these inhibitors.

In the case of **5** and **6**, which have respectively larger biphenyl and benzothiophene capping groups, these latter form more extensive hydrophobic contacts (3.5 Å) with P291 (Fig. 3b; Suppl. Fig. 4). In the case of **7**, which only has an inverted internal amide compared to **1**, the orientation and the length of the hydrogen bonds between **7** and smHDAC8 K20 and H292 appear more favourable for interaction (Suppl. Fig. 4).

Compounds **8-10** are particularly interesting since they also show a higher selectivity for smHDAC8 over hHDAC8 (3-, 4.5-, and 6-fold; respectively) (Fig. 3c; Suppl. Fig. 4; Suppl. Table 2). **8** has one additional methylene group between the internal amide and the phenyl capping group compared to **1**. **8** appears

to bind less deeply in the pocket to maximize its interactions with Y341 and the L6 loop. This change is compatible with the presence of smHDAC8 H292, but would be sterically unfavourable with hHDAC8 M274, which possibly explains the weaker inhibition observed for the human enzyme.

Compounds **9** and **10** both have a dichlorophenyl rather than a phenyl capping group, and bind perfectly into the pocket formed by smHDAC8 Y341 and the L1 and L6 loops (Fig. 3c; Suppl. Fig. 4). The presence of the halogen atoms in the capping group of **9/10** forces the inhibitor to be slightly tilted towards the L6 loop, where it forms cation- $\pi$  interaction (4.3 Å) with smHDAC8 H292. The bulkier character of the dichlorophenyl capping group may complicate the adaptation of **9/10** to the active site pocket of hHDAC8. Modelling studies show that the dichlorophenyl capping group of **9** and **10** is turned around 180° in comparison to its position in smHDAC8 and is making van der Waals interactions with the hydrophobic residues of the L6 loop P273 and M274. In this case, both chloro-substituents are surface-exposed, which presents a possible explanation for their decreased affinities towards hHDAC8 (Suppl. Fig. 5).

We have used another 3-benzamido-benzohydroxamate inhibitor (**11**) which has an internal amine rather than an internal amide in its linker and shows a low nM IC<sub>50</sub> for hHDAC8 but only a low  $\mu$ M IC<sub>50</sub> for smHDAC8 (Suppl. Table 2). The smHDAC8/**11** structure reveals that this inhibitor does not form any strong interaction with K20 and H292 and its capping group interacts less extensively with the L6 loop (Suppl. Fig. 6). The inhibitor binds centrally into the pocket, where it forms hydrophobic contacts with F216 (3.4 Å) and Y341 (3.7 Å), in a conformation identical to that observed for **PCI-34051** when bound to the smHDAC8-H292M mutant (Suppl. Fig. 6). This suggests that HDAC8-selective inhibitors bind more centrally in this pocket in hHDAC8 than in smHDAC8 due to the slight different physico-chemical properties of the active sites of these two proteins.

Finally, an unrelated compound, **12**, which has a triazole linker, binds similarly to smHDAC8, interacting with Y341 and the L6 loop (Suppl. Fig. 4). This compound also adopts an L-shaped conformation to bind to the enzyme. **12** does not make direct contact to K20 and H292, which might explain its higher IC<sub>50</sub> value for smHDAC8 (Suppl. Table 2). This compound display however a 4-fold higher potency for smHDAC8 over hHDAC8, and is also unable to inhibit strongly HDAC1 and HDAC6. Collectively, our results highlight how small chemical variations may be used to influence inhibition potency.

#### Structural specificity of HDAC8-selective pocket

Our findings show an HDAC8-selective inhibition relying on the common binding of the HDAC8 isozyme-selective inhibitors to a HDAC8 pocket that forms a shallow groove and that we have termed HDAC8-selective pocket. This pocket is defined by the catalytic tyrosine side-chain (L7 loop), which forms the pocket bottom, and residues from L6 loop and, to a lesser extent, of L1 loop of HDAC8, that both form the sides of the pocket. This pocket is highly specific to HDAC8.

In class IIa HDACs, the catalytic tyrosine is replaced by a histidine whose side chain is turned away from the active site and cannot provide the same interaction surface as HDAC8 catalytic tyrosine (Fig. 4). The situation is different for HDAC isozymes 1, 2, 3, 6 and 10 (thereafter called HDAC1-3,6,10) that all have retained a catalytic tyrosine at the same position. In addition, these isozymes have a L6 loop that displays a similar conformation to that observed in HDAC8 (Fig. 4). L6 loops in HDAC1-3,6,10 however protrude slightly more over the catalytic tyrosine side chain than in the case of HDAC8 and could sterically clash with and perturb the binding of HDAC8-selective inhibitors (Fig. 4), as supported by docking studies of **PCI-34051** and **NCC-149** with other HDAC isozymes (Suppl. Fig. 7).

Yet, our results with the smHDAC8-H292M mutant show that HDAC8-selective inhibitors can adapt to small changes within the HDAC8-selective pocket and could potentially overcome a more protrusive L6 loop. Actually, another specific and conserved feature of HDAC1-3,6,10 is a larger L1 loop that extends towards the L6 loop. Notably, at the tip of HDAC1-3,6,10 L1 loop, a proline (or an isoleucine residue in HDAC10) is present that forms hydrophobic interactions with L6 loop residues and the catalytic tyrosine, thus forming a lock over the catalytic tyrosine and preventing the formation of a pocket similar to the HDAC8-selective pocket (Fig. 4; Suppl. Fig. 7).

In HDAC8, the L1 loop, together with L2 loop, have been shown to display flexibility, being able to change their conformation to adapt to larger inhibitors, such as largazole analogues<sup>36</sup> (Suppl. Fig. 8). This raises the question whether the same L1 loop flexibility might exist in HDAC1-3,6,10. Current structural data on HDAC1-3,6,10, however, show that their L1 loops make much more extensive contacts with the rest of the enzyme. This likely explains their more constrained conformation.

#### Essential roles of L1 and L6 loops in HDAC8 catalysis and inhibition

To further question HDAC8 L1 and L6 loops conformation and functional importance in catalysis and inhibitor binding, we have performed a mutational analysis of these loops in HDAC8 and studied the effect of these mutations by biochemical, biophysical and structural means. Despite an identical number of residues, HDAC8 L6 loop has a conformation that is slightly different from the one adopted by the corresponding loops in HDAC1-3,6,10. This slight conformational change prevents L6 loop residues to protrude over the catalytic tyrosine (Fig. 4).

Sequence and structural comparison highlighted two residues that could, in first place, be responsible for this specific conformation of HDAC8 L6 loop: hHDAC8 P273/smHDAC8 P291 and hHDAC8 C275/smHDAC8 R293. However, we could not exclude that larger rearrangement are required, and we therefore created several different mutants for both hHDAC8 and smHDAC8 (Suppl. Table 4).

First, specific point mutants (hHDAC8 P273R and P273R/C275G and smHDAC8 P291R and P291R/R293G) were made, where the residues were replaced by their HDAC1 counterparts. Second, we created mutants where we exchanged the HDAC8 L6 loop completely with the one of HDAC1. We also constructed mutants where not only the L6 loop but also the L1 loop had been exchanged. Finally,

triple mutants were also generated where a leucine (hHDAC8 L179/smHDAC8 L187) was replaced by an isoleucine, as observed in HDAC1, since this mutation could possibly facilitate L6 loop conformational change (Fig. 4; Suppl. Table 4).

Except for the point mutants, all smHDAC8 mutants turned out to be insoluble. In contrast, all hHDAC8 mutants were soluble. Surprisingly, activity assays showed however that all mutants had drastically reduced activity (Suppl. Table 4). While smHDAC8 single point mutants lost around one third of their activity, hHDAC8 point mutants or L6 loop replacement mutants showed around 10-fold activity loss. Mutants of hHDAC8 combining L6 loop replacement with either L1 loop replacement or the L179I point mutation showed a 50-fold loss of activity, and the triple mutant displayed almost no more activity.

We used Thermal Shift Assay experiments to assess whether the mutations had an effect on the stability of the various mutants. All mutants only showed a decreased  $T_m$  of only about 5°C compared to the WT enzymes but did not indicate partial or complete unfolding of the proteins (Suppl. Table 4). Due to the residual activity of the mutants, we next measured the  $IC_{50}$  values for **PCI-34051**, **NCC-149** and **QSN** for all mutants. All inhibitors showed significantly higher  $IC_{50}$  values indicative of a poorer inhibition capacity (Suppl. Table 5). This was most pronounced for inhibitor **PCI-34051**, with more than 64-fold increase of the  $IC_{50}$  value for the triple HDAC8 mutant. The  $IC_{50}$  values for **NCC-149** were also significantly increased, but not as much as for **PCI-34051**. **QSN** also displayed decreased inhibition, albeit to a much lesser extent, possibly mirroring only the slight stability decrease of the mutants, but supporting our conclusions on the crucial role of L1 and L6 loops for HDAC8 selective inhibition.

To investigate the molecular basis of L1 and L6 loops conformational changes upon mutations, we attempted to solve the structures of the different HDAC8 mutants in complex with **PCI-34051**, **NCC-149** and **QSN**. Although different mutants gave crystals in presence of some of these inhibitors, only the crystals obtained with the hHDAC8-mL6/**QSN** complex led to exploitable structural data (Suppl. Table 6). To our knowledge this is the first structure of compound **QSN** bound to human HDAC8. Here again the capping group of the inhibitor is extensively involved in crystal packing and it is impossible to understand whether the binding conformation of **QSN**, which is different to that observed when bound to smHDAC8, represents a favoured binding conformation in solution to hHDAC8 (Suppl. Fig. 9). Strikingly, the structure of this complex (Fig. 5) revealed that, despite the complete exchange of the L6 loop, the conformation of this loop remains as observed in hHDAC8 WT. The hHDAC8 P273R replacement however leads to the partial occupation of the HDAC8-selective pocket by the side chain of the arginine replacing the proline (Fig. 5). Actually, the only mutation common to all mutants is the hHDAC8 P273R/smHDAC8 P291R change. This corresponds to the arginine in HDAC1 and HDAC3 that is important for activity and is involved in inositol phosphate binding<sup>34,37,38</sup>. Our results therefore further highlight the importance of the L6 loop for class I HDACs activity and inhibitor binding, and

pinpoint differences between HDAC8 and the other members of this class that can be used for selective inhibition of the former enzyme.

### Discussion

The results described here provide a comprehensive and detailed molecular view of HDAC8 selective inhibition, highlighting a specific HDAC8-selective pocket where selective inhibitors form preferential interactions with HDAC8 catalytic tyrosine and L6 loop residues. They particularly emphasize the balance between chemical structure and inherent conformational flexibility of the inhibitors, on one hand, and small changes at the protein level, on the other hand, with important implications for selective inhibition. Specifically, the restricted conformation of HDAC8 L6 loop, compared to the more flexible character of the L1 loop point out the importance of the L6 loop for selective inhibition. Actually, the constrained conformation of the L7 loop that bears the catalytic tyrosine also appears as an essential feature of HDAC8 selective inhibition.

Interestingly, in class IIa HDACs the replacement of the catalytic tyrosine by a histidine creates a specific pocket at the position of the tyrosine side chain. This feature prevents HDAC8-selective inhibitors from binding to class IIa HDACs and has been exploited to design class IIa-selective inhibitors that cannot distinguish, however, between the different class IIa isozymes<sup>39</sup>.

In case of HDAC1-3,6,10, although the catalytic tyrosine is conserved and the L6 loop is similar in these isozymes, the HDAC8-selective pocket is not present in these isozymes due to a larger L1 loop in HDAC1-3,6,10 that covers the catalytic tyrosine and forms a lock over this selectivity pocket. The recent structures of HDAC6 obtained with HDAC6-selective inhibitors show that the capping group of these inhibitors actually interacts with the proline at the tip of the L1 loop as well as with the preceding histidine residue<sup>40</sup>. This mirrors somehow the interaction observed between the HDAC8-selective inhibitors and smHDAC8 L6 loop P291-H292 motif. On a structural basis, it remains however difficult to appreciate why the HDAC6-selective inhibitors do not bind to the same conserved motif in HDAC1-3.

Interestingly, the HDAC8-selective pocket could adapt bulkier inhibitors that could also interact with the back of the pocket, as partially observed for compound **NCC-149** when bound to the smHDAC8-H292M mutant, thus potentially paving the way for the design of more potent HDAC8-selective inhibitors. Yet, the reduced conformational flexibility of compound **PCI-34051**, if it prevents this inhibitor to adapt easily to small changes within HDAC8 active site, might also explain why this inhibitor is so selective for this enzyme.

The existence of a unique selective specific pocket in HDAC8 and the effects observed on activity when this pocket is partially occupied indicate that it might also be essential for the recognition of specific targets. Collectively, our results unravel for the first time the structural determinants underlying HDAC8 selective inhibition and pave the way to the design of more potent HDAC selective inhibitors,

towards the development of novel epidrugs and the delineation of HDACs specific biological role through chemical biology approaches.

## **Acknowledgments**

This work and the authors of this manuscript have been supported by funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreements nos. 241865 (SEtTReND) and 602080 (A-ParaDDisE). CR, MM, and RJP are supported by institutional funds from the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM), the Université de Strasbourg, the Institut Pasteur de Lille and the Université de Lille. The authors acknowledge the support and the use of resources of the French Infrastructure for Integrated Structural Biology FRISBI ANR-10-INBS-05 and of Instruct-ERIC. We wish to thank members of the ESRF-EMBL joint structural biology groups, the SOLEIL, and the Swiss Light Source (SLS) synchrotrons for the use of their beamline facilities and for help during data collection. We are grateful to Alastair McEwen (IGBMC) for his kind assistance during X-ray data collections.

## **Author contributions**

M.M. and T.B.S. contributed equally to this work. M.M., T.B.S., E.R.M. and C.R. solved the HDAC8-inhibitor crystal structures, constructed and biophysically characterized the HDAC8 mutants. J.L. and J.R.P. performed anti-parasitic studies. C.D.V., M.M. and E.E. performed and interpreted the ITC experiments. T.S. synthesized the NCC-149 inhibitor. D.K. and R.H synthesized the KH197 (**12**) inhibitor. T.H., J.M. and D.R. carried out docking studies, synthesized and characterized smHDAC8 inhibitors. A.C., K.S. and M.J. performed enzymatic activity and inhibition assays. M.M., J.R.P, M.J., W.S. and C.R. designed experiments, analysed data and wrote the paper.

## **Competing financial interests**

The authors declare no competing financial interests.



## Material and Methods

### Small-molecule inhibitors

The inhibitors **PCI-34051** and **NCC-149** were purchased from Selleck Chemicals (USA). The smHDAC8-selective inhibitors were synthesized and characterized as described previously<sup>28,41</sup>.

Compound **6** (3-(Benzthiophene-7-carboxamido)-4-chlorobenzohydroxamate) was synthesized as follows, in a manner analogous to the smHDAC8-selective inhibitors mentioned above (Suppl. Fig. 10a). Procedure. (a) Benzthiophene-7-carboxylic acid (1.4 mmol) was cooled to 0 °C and then thionyl chloride (3 mmol) was added dropwise. The mixture was heated under reflux for 30 minutes. After evaporating the excess of thionylchloride under vacuum, the obtained acid chloride was dissolved in dry THF (50 ml) and added to a solution of methyl 3-amino-4-chlorobenzoate (1.4 mmol) and DIPEA (3 mmol) in THF. The reaction was monitored by TLC. Subsequently, the solvent was evaporated under vacuum and the mixture was dissolved in EtOAc (50 ml) and washed with aq. sodium hydroxide solution. The organic layer was evaporated under reduced pressure and the product was purified by column chromatography (chloroform/ methanol, 99:1). (Yield: 88.8 %). (b) The obtained amide was dissolved in MeOH (25 ml) and 1M aq. sodium hydroxide solution (10 ml) and heated to 50 °C for 2 h. The reaction was monitored by TLC. Subsequently, the solvent was evaporated under reduced pressure and the product was dissolved in EtOAc and washed with 1M HCl solution, the organic layer was finally evaporated under reduced pressure. (Yield: 85.0 %). (c) The substituted benzoic acid derivative (1 mmol) was dissolved in dry THF (50 ml) and PyBOP (1.2 mmol) was added. To the activated acid, a mixture of NH<sub>2</sub>OTHP (1.5 mmol) and DIPEA (2.5 mmol) in dry THF (5 ml) was added, and the reaction mixture was stirred at room temperature overnight. The solvent was evaporated under reduced pressure and the mixture was dissolved in EtOAc (50 ml) and washed with aq. potassium hydrogen carbonate solution and brine. The organic layer was evaporated under reduced pressure and the product was purified by column chromatography (chloroform/ methanol/ TEA, 99.5:0.45:0.05). The obtained product was dissolved in THF and a catalytic amount of diluted HCl was added and it was stirred at room temperature. The reaction was controlled by TLC. After that, the solvent was evaporated under reduced pressure and the mixture was dissolved in EtOAc (50 ml) and washed with brine. The organic layer was evaporated under reduced pressure and the product was purified by column chromatography (chloroform/methanol/formic acid, 95:4.95:0.05). (Yield: 35 %).

Analytical data. MS m/z: 345.23 (Cl<sup>35</sup>) | 347.24 (Cl<sup>36</sup>)[M-H]<sup>-</sup>.

<sup>1</sup>H NMR (400 MHz, DMSO) δ 11.37 (s, 1H), 10.42 (s, 1H), 9.14 (s, 1H), 8.22 (d, J = 7.3 Hz, 1H), 8.15 (d, J = 7.7 Hz, 1H), 8.02 (s, 1H), 7.85 (d, J = 5.2 Hz, 1H), 7.74 – 7.63 (m, 2H), 7.59 (t, J = 7.5 Hz, 1H), 7.53 (d, J = 5.2 Hz, 1H).

HRMS m/z: 369.0072 [M+Na]<sup>+</sup>; calculated: C<sub>16</sub>H<sub>11</sub>N<sub>2</sub>O<sub>3</sub>ClSNa<sup>+</sup> 369.0071. HPLC: rt 10.68 min (99.65%). Yield: 120 mg; 0.35 mmol; 25 %.

Compound **12** (1-[5-chloro-2-(4-fluorophenoxy)phenyl]-*N*-hydroxy-1*H*-1,2,3-triazole-4-carboxamide) was synthesized as follows (Suppl. Fig. 10b). Procedure: (a) 2-Amino-4-chlorophenol (7.0 mmol) was dissolved in 1 M HCl (20 mL). At -5 °C, a solution of sodium nitrite (8.4 mmol) in water (2 mL) was added dropwise over a period of 5 min. After stirring additional 5 min, urea (50 mg) was added. Then the mixture was added to a cold solution of sodium azide (14 mmol) and sodium acetate (0.06 mmol) in water (10 mL). The resulting mixture was stirred for 2 h at -5 °C. Then it was extracted with diethyl ether (3×). The combined organic layers were dried (Na<sub>2</sub>SO<sub>4</sub>), filtered, and the solvent was removed *in vacuo*. The residue was purified by flash column chromatography (cyclohexane/ethyl acetate = 10/1). (Yield: 86 %). (b) The obtained azide (3.0 mmol) was dissolved in a 1:1 mixture of water and *tert*-butyl alcohol (15 mL). Methyl propiolate (4.0 mmol), sodium ascorbate (0.20 mmol), and copper(II) sulfate pentahydrate (0.04 mmol) were added and the mixture was stirred for 18 h at room temperature. Then water was added and the mixture was extracted with ethyl acetate (3×). The combined organic layers were dried (Na<sub>2</sub>SO<sub>4</sub>), filtered, and the solvent was removed *in vacuo*. The residue was purified by flash column chromatography (cyclohexane/ethyl acetate = 2/1). (Yield: 89 %). (c) A 100 mL round-bottom flask was charged with the obtained triazole (0.59 mmol), Cu(OAc)<sub>2</sub> (0.59 mmol), 4-fluorophenylboronic acid (0.71 mmol), and powdered 4 Å molecular sieves. Then dichloromethane (4.5 mL) was added. After the addition of triethylamine (2.9 mmol), the reaction mixture was stirred at ambient temperature overnight. Then the suspension was filtered. The filtrate diluted with water and extracted with ethyl acetate (3×). The combined organic layers were dried (Na<sub>2</sub>SO<sub>4</sub>), filtered, and the solvent was removed *in vacuo*. The residue was purified by flash column chromatography (cyclohexane/ethyl acetate = 9/1 → 3/1). (Yield: 13 %). (d) A 5.4 M solution of sodium methoxide in methanol (1.1 mmol) was added to a solution of the obtained diphenyl ether (0.11 mmol) and hydroxylamine hydrochloride (0.55 mmol) in dry methanol (3 mL). The mixture was stirred at ambient temperature overnight. Then the solvent was removed *in vacuo* and the residue was purified by automatic flash column chromatography using a Biotage purification apparatus (5% → 50% ACN in H<sub>2</sub>O, Biotage® SNAP KP-C18-HS 12 g). Fractions containing the desired product were combined, dried from acetonitrile under reduced pressure and then subjected to lyophilization. (Yield: 95 %).

Analytical data: <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>): δ [ppm] = 7.02 – 7.06 (m, 1H), 7.15 – 7.20 (m, 2H), 7.22 – 7.28 (m, 2H), 7.53 – 7.57 (m, 1H), 7.87 – 7.90 (m, 1H), 8.31 (s, 1H); <sup>13</sup>C NMR (DMSO-*d*<sub>6</sub>): δ [ppm] = 116.9 (d, *J* = 23.7 Hz, 2C), 120.4 (1C), 121.1 (d, *J* = 8.7 Hz, 2C), 123.1 (1C), 125.7 (1C), 127.4 (1C), 128.8 (1C), 130.2 (1C), 147.1 (1C), 148.4 (1C), 151.3 (d, *J* = 2.4 Hz, 1C), 158.3 (1C), 158.8 (d, *J* = 241 Hz, 1C); HRMS (*m/z*): [M+H]<sup>+</sup> calcd for C<sub>15</sub>H<sub>11</sub>ClFN<sub>4</sub>O<sub>3</sub>: 349.0498, found: 349.0522; HPLC: *t*<sub>R</sub> = 16.3 min, purity 92.4 %.

#### **Cloning, expression and purification of HDAC8 proteins**

The full-length cDNA constructs (WT and mutants) for hHDAC8 and smHDAC8 were amplified by polymerase-chain reaction (PCR) and cloned into bacterial expression vectors. The hHDAC8 (WT and

mutant) gene was inserted between the NdeI and BamHI restriction sites of the pNEA-3HT expression vector<sup>42</sup>, where it is in frame with a sequence coding for a N-terminal poly-histidine affinity purification tag followed by thioredoxin and a protease 3C cleavage site. The smHDAC8 (WT and mutant) gene was cloned between the NdeI and BamHI restriction sites of the pNEA-tH expression vector<sup>42</sup> and is in frame with a sequence coding for a C-terminal thrombin cleavage site followed by a poly-histidine affinity purification tag.

For wildtype smHDAC8 overproduction, a modified protocol was used than the one published previously. Expression was carried out in BL21 (DE3) cells in 2xLB medium. Cultures were induced at 37°C with 0.7mM IPTG, in presence of 100 µM ZnCl<sub>2</sub>. After overnight incubation at 37°C, cells were harvested and resuspended in lysis buffer A (150mM NaCl, 50mM Tris pH 8.0). Lysis was done by sonication, lysate was clarified by centrifugation. The supernatant was loaded onto Talon Superflow Metal Affinity Resin (Clontech) pre-equilibrated with the lysis buffer A. The his-tagged protein was released from the Talon resin by thrombin protease treatment in buffer B (50mM KCl, 10mM Tris pH 8.0) and subsequently loaded onto 16/60 Superdex 200 gel filtration column (GE Healthcare) pre-equilibrated with buffer C (50 mM KCl, 10 mM Tris-HCl pH 8.0 and 2 mM DTT). Peak fractions were concentrated with an Amicon Ultra centrifugal filter unit. This protocol yielded three times more protein for the wild type enzyme. smHDAC8 mutants could not be purified using this protocol and their overproduction was carried out as described initially<sup>29,43</sup>.

hHDAC8 overproduction was carried out in *E. coli* BL21 (DE3) cells in 2xLB medium. Culture induction was done at 23°C by adding 0.5 mM final isopropyl-1-thio-β-D-galactopyranoside (IPTG, Euromedex), in presence of 100 µM ZnCl<sub>2</sub>. Harvested bacteria were re-suspended in lysis buffer (50 mM KCl, 10 mM Tris-HCl pH 8.0) and lysed by sonication. The lysate was clarified by centrifugation (17,500 rpm, 50 min, 4°C, Sorvall Lynx 6000 Thermo Scientific). The supernatant was loaded onto Talon Superflow Metal Affinity Resin (Clontech) pre-equilibrated with lysis buffer. The his-thioredoxin-tagged protein was released from the Talon resin by 3C protease treatment and subsequently loaded onto 16/60 Superdex 200 gel filtration column (GE Healthcare) pre-equilibrated with the buffer (50 mM KCl, 10 mM Tris-HCl pH 8.0, and 0.5 mM TCEP). The recombinant proteins were concentrated with an Amicon Ultra centrifugal filter unit.

### **Mutagenesis experiments**

The mutant constructs were generated using standard PCR-based nested protocols and inserted into the corresponding expression vectors. The L1 and L6 loops exchange mutants were designed based on structural comparison. The hHDAC1 L6 sequence S<sub>265</sub>LSGDRLGC was introduced instead of the T<sub>268</sub>LAGDPMCS sequence in hHDAC8, to create the hHDAC8-mL6 mutant. In the second step, the hHDAC1 L1 sequence Y<sub>23</sub>YGQGHMPK was introduced instead of the L<sub>31</sub>AKI sequence in hHDAC8-mL6, to create the double hHDAC8-mL1/mL6 mutant. Finally, the triple mutant hHDAC8-mL1/mL6/L179I

was generated by an introduction of isoleucine residue in the L179 position, using the double hHDAC8-mL1/mL6 mutant as a template. smHDAC8 mutants were produced as described for hHDAC8 mutants. Sequences replaced in smHDAC8 were the same as the ones replaced in hHDAC8.

### **Crystallization and X-ray data collection**

Diffraction-quality crystals of native smHDAC8 enzyme were obtained at 17°C after 3 days by mixing equal volumes of smHDAC8 (2.5 mg/ml) with reservoir solution composed of 21% PEG 3350 (Fluka) and 0.05 M Na<sup>+</sup>/K<sup>+</sup> L-tartrate, and crystallized using the hanging-drop vapor diffusion technique. After 3 days, grown crystals were soaked for 20 hours in mother liquor supplemented with the corresponding inhibitor (10 mM final concentration of the inhibitor, preparation from a 100mM stock in N,N-dimethylformamide (DMF) or DMSO). Crystals used for X-ray data collection were briefly transferred in reservoir solution supplemented with 22% glycerol and flash-frozen in liquid nitrogen. Co-crystallization of hHDAC8-mL6 together with **QSN** (Quisinostat) inhibitor was performed using the hanging-drop vapor diffusion technique. The hHDAC8-mL6/**QSN** complex was formed by incubating the hHDAC8-mL6 mutant protein (5 mg/mL) with **QSN** (5 mM resuspended in DMF) at 4°C for 1 h. Diffraction-quality crystals were obtained at 20°C after 3–4 days by mixing equal volumes of the hHDAC8-mL6/**QSN** complex with reservoir solution composed of 20% polyethylene glycol 3350 (Fluka), 0.2 M KNO<sub>3</sub> and 0.1 M Bis-tris propane (pH=7.5). Crystals used for X-ray data collection were briefly transferred in reservoir solution supplemented with 22% glycerol and flash-frozen in liquid nitrogen. Crystallographic data obtained in this project were collected at 100 K on SOLEIL beamline PROXIMA1, ESRF beamlines ID30b, ID29 and ID23 and SLS PX beamlines.

### **Structure determination, model building and refinement**

Crystallographic data were processed and scaled using HKL2000<sup>44</sup> or XDS<sup>45</sup>. Phases for smHDAC8/inhibitor complexes were obtained by molecular replacement followed by rigid body refinement against smHDAC8 native structure as a model (4BZ5). The initial models were refined through several cycles of manual building using Coot<sup>46</sup> and automated refinement with Phenix<sup>47</sup>. The structure of hHDAC8-mL6 complexed with **QSN** was solved by molecular replacement with Phenix<sup>47</sup> using the hHDAC8 structure (PDB 1T67) as a search model. The final models were validated using tools provided in Coot<sup>46</sup> and Molprobit<sup>48</sup>. Visualization of structural data was done with Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.), and two-dimensional diagrams summarizing molecular interaction between inhibitors and HDAC8 enzymes were prepared with the help of the LigPlot program<sup>49</sup>. Atomic coordinates and structure factors of the smHDAC8 and hHDAC8-mL6 complexes were deposited in the Protein Data Bank under the PDB codes xxx.

### **HDAC activity and inhibition assays**

Catalytic activity and inhibition assays of smHDAC8 and hHDAC8 were performed as described earlier<sup>29</sup>. Briefly, the hHDAC8 and smHDAC8 activity testing was carried out with the HDAC8 Fluorimetric Drug

Discovery Kit (Fluor de Lys(R)- HDAC8, BML-KI178) from Enzo Life Sciences, according to the manufacturer's instructions with a substrate concentration of 50  $\mu$ M. Fluorescence was measured in a plate reader (BMG Polarstar) with excitation at  $\lambda$  =390 nm and emission at  $\lambda$  = 460 nm. IC<sub>50</sub> values were determined with OriginPro (version 9.0.0, Northampton, Massachusetts).

#### **In vivo studies on *Schistosoma mansoni***

The in vivo effects of **PCI-34051** and **NCC-149** compounds on the viability of *S. mansoni* schistosomula (Puerto-Rican strain) were tested using a microscopy-based assay, as described previously<sup>50</sup>. In brief, schistosomula (2,000 per well), prepared by standard mechanical transformation from cercaria<sup>51</sup>, were maintained in 6-well plates in M199 medium kept at pH 7.4 with 10 mM HEPES and supplemented at 37°C in a humid atmosphere containing 5% CO<sub>2</sub>. Three different concentrations of inhibitors (25, 50 and 100  $\mu$ M) were tested, the inhibitors were dissolved in DMSO, and the culture medium was refreshed each day. The assessment of parasite mortality was carried out after microscopic examination, based on three criteria: a granular appearance, tegumental defects and the absence of motility. At least 300 schistosomula were observed at each time point for each condition and results were expressed as a percentage of viable larvae remaining. Three biological replicates (different batches of larvae) were examined in duplicate for each condition.

The stability of adult worm pairs and egg production were assayed as previously described<sup>51</sup>. Adult worm pairs were obtained from infected golden hamsters (*Mesocricetus auratus*) by perfusion, washed in M199 medium and ten pairs placed in 2 mL of M199 buffered complete medium (as for schistosomula above) in each well of a 6-well culture plate. Worms were maintained in culture for 5 days at 37°C (humid atmosphere, 5% CO<sub>2</sub>) before the addition of HDAC8-selective inhibitors, the application of DMSO alone served as a negative control experiment. Both the culture medium and the inhibitors were refreshed daily. The number of couples remaining as pairs was determined daily by microscopy and the medium of each well containing eggs laid by the couples was recovered and centrifuged to allow eggs to be counted under the microscope. Three biological replicate experiments were performed in triplicate.

#### **TUNEL assay**

Detection of DNA strand breaks in inhibitor-treated *S. mansoni* schistosomula was done using the terminal deoxynucleotidyl transferase dUTP nick-end labelling (TUNEL) method using the *In Situ* Cell Death Detection Kit TMR Red (Roche). The method designed for cell suspensions was followed as described in the manufacturer's instructions with few modifications. Briefly, 2,000 schistosomula were treated with 100  $\mu$ M **PCI-34051** or **NCC-149** for 96 h, in 6-well plates containing 2 mL of complete medium. The treatment with DMSO alone served as negative control experiment. After 96-h incubation, culture media were removed and the schistosomula were centrifuged (1,000 rpm, 2 min), washed three times in PBS buffer, and then fixed in 2% formaldehyde for 60 min. Schistosomula were

afterwards washed once more in PBS and permeabilization solution (0.1% Triton X-100, 0.1% sodium citrate) was added for 10 min on ice. Labeling of schistosomula with 4',6-diamidino-2-phenylindole (DAPI) and TMR Red dUTP was performed according to the manufacturer's instructions and TUNEL-positive parasites were observed by fluorescence using an AxioImager Z1-Apotome microscope (Zeiss).

### **Isothermal titration calorimetry (ITC)**

ITC experiments were done at 25°C using a PEAQ microcalorimeter (Malvern Instruments). All protein samples were purified in the same ITC buffer (50 mM KCl, 10 mM Tris pH=8.0 and 0.5 mM TCEP). In a typical experiment, aliquots of 2.0 µl of HDAC8 protein sample at 200 µM were injected into a 300 µl inhibitor solution at 20 µM. Blank experiments were used to retrieve signal due to solvent (DMSO or DMF) dilution into ITC buffer. Data were analyzed with Microcal PEAQ-ITC Analysis Software and with Affinimeter.

### **Differential scanning fluorimetry**

Thermal stability of HDAC8 proteins was analysed by a label-free differential scanning fluorimetry (DSF) approach using a Prometheus NT.48 instrument (NanoTemper Technologies). Briefly, the shift of intrinsic tryptophan fluorescence of HDAC8 proteins upon gradual temperature-triggered unfolding (temperature gradient 20–95°C) was monitored by detecting the emission fluorescence at 330 and 350 nm. The measurements were carried out nanoDSF-grade high sensitivity glass capillaries (NanoTemper Technologies) at a heating rate of 1°C/min. Protein melting points ( $T_m$ ) were inferred from the first derivative of the ratio of tryptophan emission intensities at 330 and 350 nm. Finally, the  $\Delta T_m$  value of an HDAC8 protein for a particular inhibitor was calculated as the difference between the  $T_m$  values of the inhibitor-bound and inhibitor-free proteins. All the assays were done in triplicate.

### **Computational and docking experiments**

Crystal structures of hHDAC1-3, hHDAC8 and hHDAC6-CD2 for docking as well as zHDAC10 for homology modelling of hHDAC10 were downloaded from Protein Data Bank (corresponding PDB IDs 4BKX, 4LXZ, 4A69, 2V5X, 5EDU, 5TD7). Also the zHDAC6 CD1-2 structure with Nexturastat A (PDB ID 5G0J) was downloaded from PDB in order to retrieve conserved water molecules for docking to hHDAC6-CD2. The homology model of human HDAC10 was built on zebrafish HDAC10 structure using MODELLER version 9.11<sup>52</sup>. Structures of inhibitors were generated in MOE version 2014.09 (Molecular Operating Environment (MOE), 2014.09; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2014).

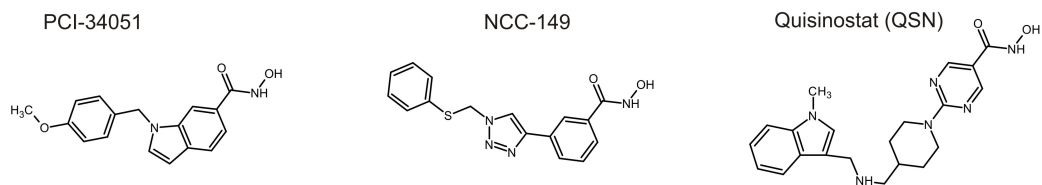
Protein and inhibitor structures were further prepared for docking in Schrödinger Suite (Schrödinger Suite 2014-2: Maestro version 9.8, Protein Preparation Wizard; Epik version 2.8, Glide version 9.8, Schrödinger, LLC, New York, NY, 2014). Human HDAC1-3,6,10 were prepared using Protein Preparation Wizard tool. Hydrogen atoms and missing amino acid residues side chains were added. Solvent molecules were removed except two conserved water molecules: one near the catalytic zinc ion and

another above the zinc coordinating histidine residue (H180, hHDAC8 numbering). Next, hydrogen bonds network and amino acid residues protonation states and tautomers were optimized. Finally, protein models were subjected to energy minimization using OPLS-2005 force field with default settings. Inhibitor structures were prepared with LigPrep and ConfGen tools. Namely, the tautomeric forms and stereoisomers were created and energy minimization was performed with OPLS force field using Ligprep. Conformers were generated with default settings (fast) and energy minimization of the output conformations was performed using ConfGen.

Molecular docking was performed using Glide from the Schrödinger Suite. Receptor grids were generated using default settings. The Standard Precision docking protocol with default settings without any constraints was used, except the number of docking poses for post-docking minimization per ligand was increased to 20 and the maximal number of output poses per ligand was increased to 2.

# Figure 1

**a**

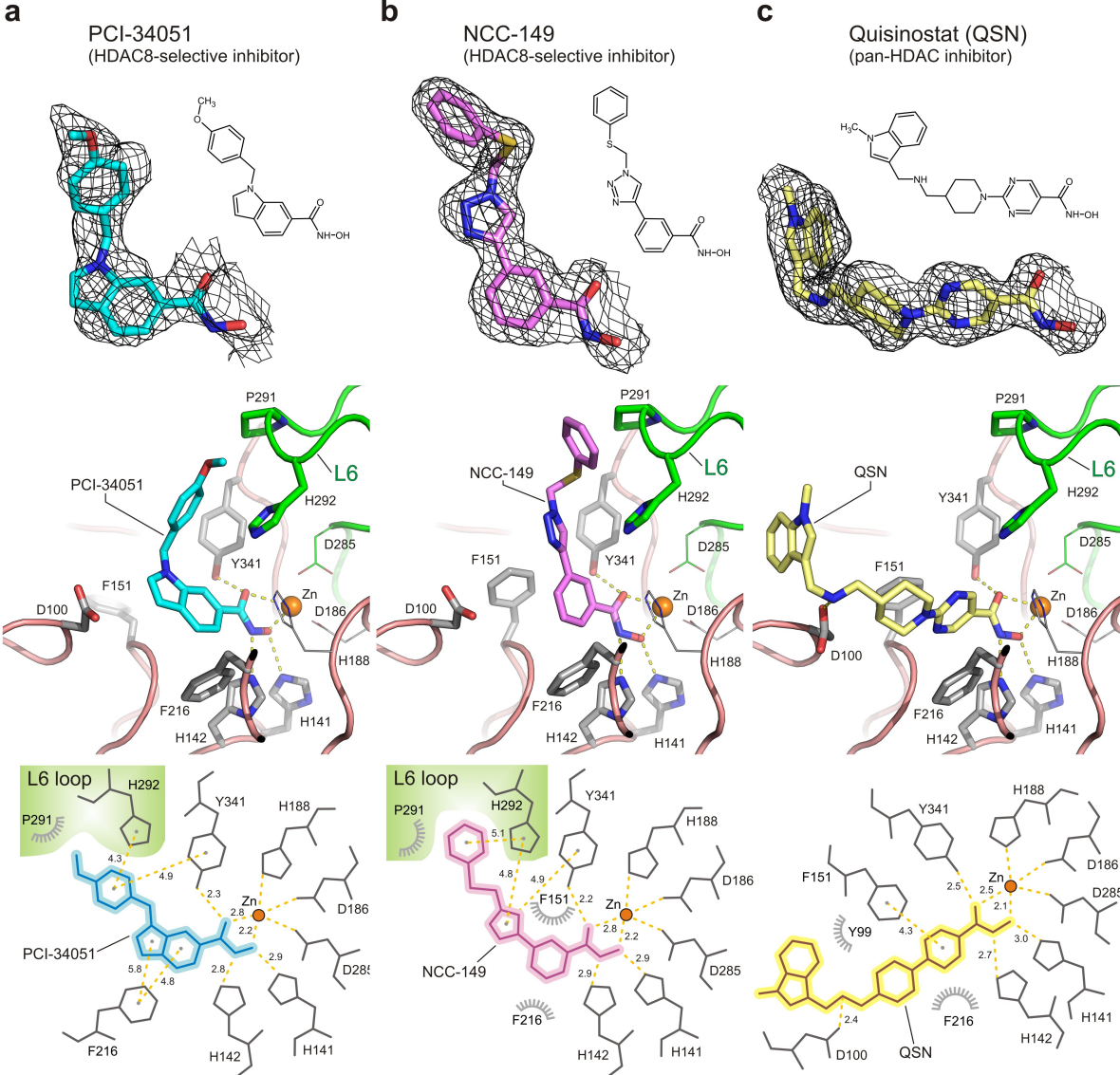


**b**

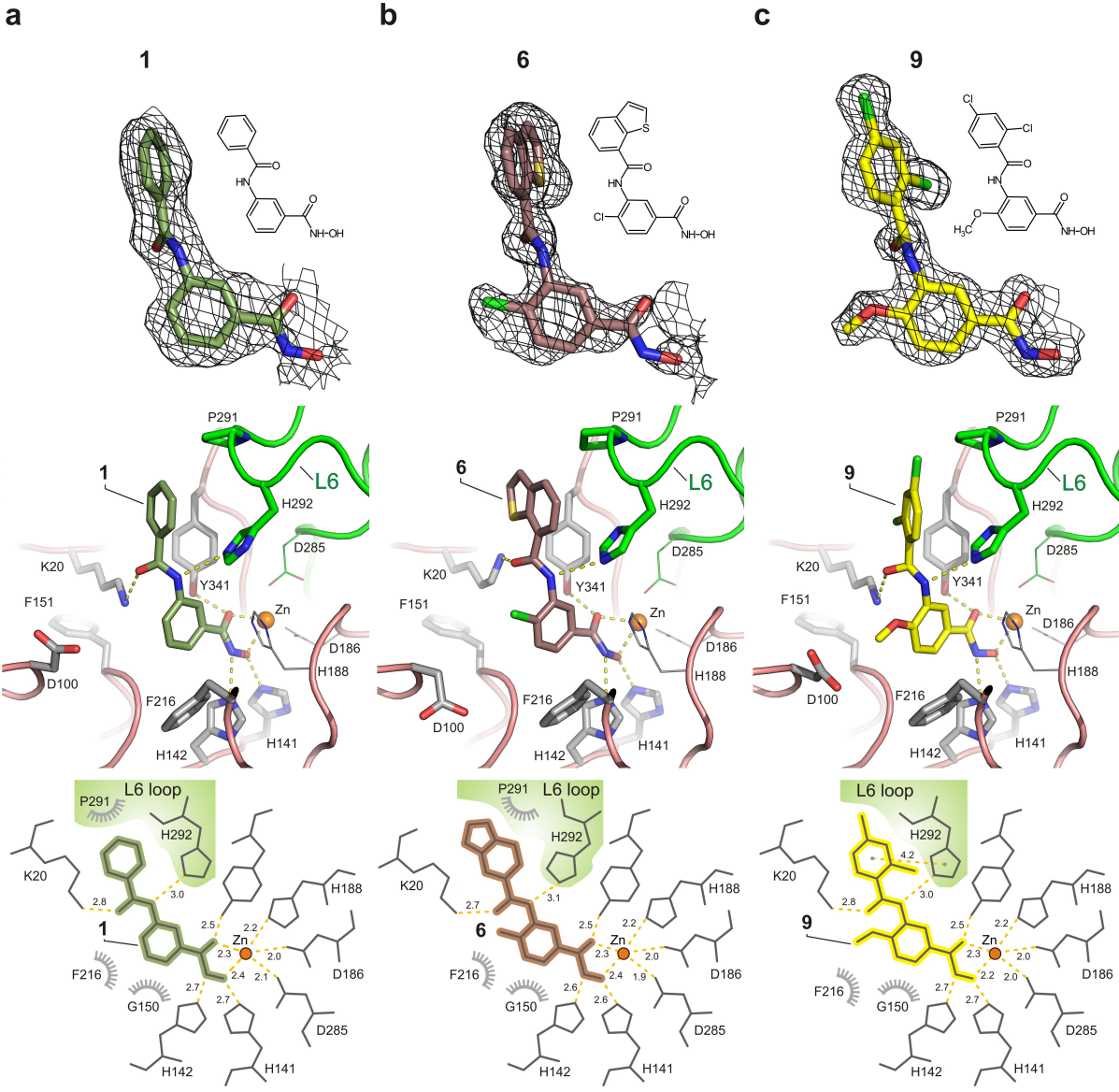
Enzyme	Inhibitor	IC <sub>50</sub> (nM)	K <sub>d</sub> (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	-TΔS (kcal/mol)
hHDAC8	PCI-34051	77.7 ± 18.1	73.9 ± 8.6	-9.7	-11.2	1.5
hHDAC8	NCC-149	44 ± 4.9	91.8 ± 21.5	-9.6	-11.0	1.4
hHDAC8	QSN	64.4 ± 3.4	19.7 ± 4.1	-10.5	-13.0	2.5
smHDAC8	PCI-34051	435.8 ± 61	363 ± 60.8	-8.8	-8.6	-0.2
smHDAC8	NCC-149	95.4 ± 9.3	92.5 ± 25.2	-9.6	-9.8	0.2
smHDAC8	QSN	303.4 ± 39	15.8 ± 7.4	-10.6	-8.4	-2.2



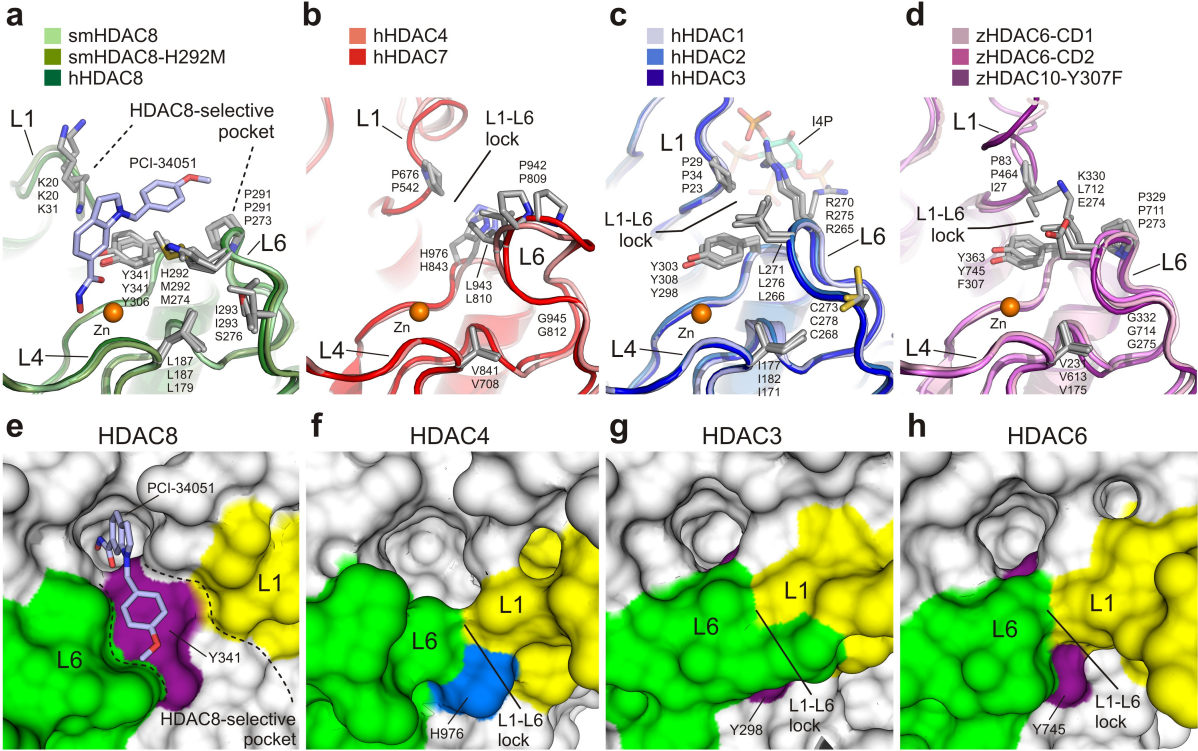
**Figure 2**



**Figure 3**



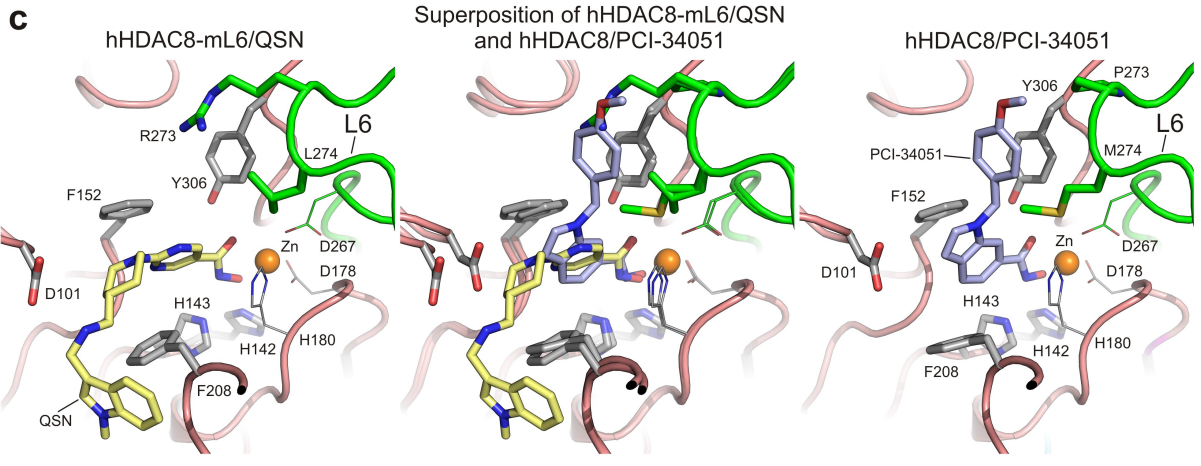
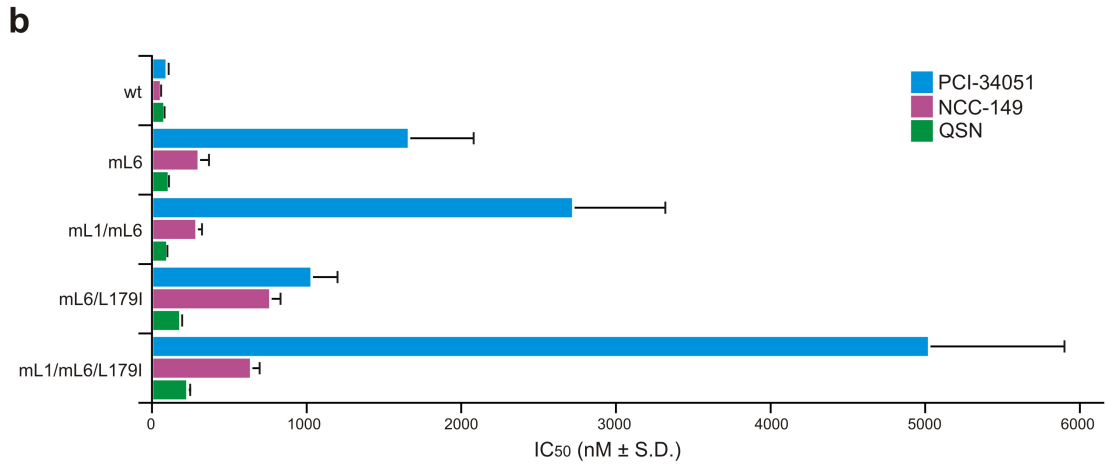
**Figure 4**



**Figure 5**

**a**

	L1 loop			L4 loop			L6 loop										
smHDAC8	14	LCCSS	-----PKFGDR	24	182	YVDLDL	HHGDGV	193	279	VVQC	ADCLAT	--DPHRI	FRLT	(17)	LSGYL	320	
hHDAC8	27	MCDSL	-----AKIPKR	37	174	YVDLDL	HHGDGV	185	261	VLQL	GADT	IAG	--DPMCS	FNMT	PVGI	G	285
hHDAC1	19	VGNYYYG	---QGHMPKPHR	34	172	YIDIDI	HHGDGV	183	258	VLQCGS	DSL	SG	--DRLG	CFNLT	IKGHA	282	
hHDAC2	20	IGNYYYG	---QGHMPKPHR	35	173	YIDIDI	HHGDGV	184	259	VLQCGS	ADSL	SG	--DRLG	CFNLT	VKGHA	283	
hHDAC3	13	VGNFHYG	---AGHPMKPHR	28	166	YIDIDI	HHGDGV	177	253	VLQCGS	ADSL	GC	--DRLG	CFNLS	IRGHG	277	
hHDAC4	662	MLKHQCT	CGSSSSHPEHAGR	681	836	IVDWDV	HHGNGT	847	928	LVSSGF	DAVE	GHTPL	GGYNLS	ARCFG	954		
hHDAC5	691	MLKHQCM	CGNTHVHPEHAGR	710	866	IVDWDI	HHGNGT	877	958	LVSAGF	DAVE	GHLSP	GGYSVT	ARCFG	984		
hHDAC7	530	MLKHQCS	CGDNSRHPEHAGR	549	705	IVDWDV	HHGNGT	716	797	LVSAGF	DAEAG	HPAPL	GGYHVS	AKCFG	823		
hHDAC9	641	MLKHQC	VCNSTTHPEHAGR	660	816	IVDL	DVHHGNGT	827	908	LVSAGF	DALE	GHTPPL	GGYKVT	AKCFG	934		
hHDAC6-CD1	94	LNEFHCLW	--DDSFPEGPER	111	249	IVDWDV	HHGQGT	260	340	LVAAGF	DALQ	--DPK	GEMAAT	PAGFA	364		
hHDAC6-CD2	489	MMNHCNLW	--DSHPEVPER	506	645	IVDWDV	HHGNGT	656	736	LVSAGF	DAARG	--DPL	GGCQVS	PEGYA	760		
hHDAC10	11	MTATRL	LWDDPECEIERPER	30	168	VVDWDV	HHGQGI	179	259	LVSAGF	D	SAIG	--DPE	GQM	QAT	PECFA	283
hHDAC11	24	YNITFMGL	--EKLHPFDAGK	41	177	IIDLD	DAHQNGH	188	255	VYNAGT	DILEG	--DRL	GGLS	IS	PAGIV	279	



## Figure legends

**Figure 1. HDAC8 inhibition and binding by inhibitors PCI-34051, NCC-149 and QSN.** (a) Chemical structures of **PCI-34051**, **NCC-149** and **Quisinostat (QSN)**. (b)  $IC_{50}$  values and binding affinities/thermodynamic parameters of **PCI-34051**, **NCC-149** and **QSN** for human HDAC8 (hHDAC8) and *Schistosoma mansoni* HDAC8 (smHDAC8).

**Figure 2. Structural characterization of PCI-34051, NCC-149 and QSN binding to smHDAC8.**

(a) Simulated annealing omit electron density map contoured at  $2\sigma$  for **PCI-34051**, **NCC-149** and **QSN** when bound to smHDAC8. (b) Binding mode of **PCI-34051**, **NCC-149** and **QSN** in the active site of smHDAC8. The inhibitors and important residues are shown as sticks. The L6 loop is coloured green. The catalytic zinc ion is shown as orange sphere. Zinc coordination and hydrogen bonds are shown as dashed lines. Only HDAC8-selective inhibitors **PCI-34051** and **NCC-149** adopt an L-shaped conformation and interact with the catalytic tyrosine and the L6 loop. (c) Two dimensional illustrations of binding. Zinc coordination, hydrogen bonds and aromatic interactions are shown as dashed lines. The corresponding distances between the atoms and/or chemical groups are given in Å. Hydrophobic contacts are shown by grey arcs with spokes radiating toward the atoms involved.

**Figure 3. Structural characterization of 1, 6 and 9 HDAC8-selective inhibitors binding to smHDAC8.**

(a) Simulated annealing omit electron density map contoured at  $2\sigma$  for **1**, **6** and **9** when bound to smHDAC8. (b) Binding mode of **1**, **6** and **9** in the active site of smHDAC8. (c) Two dimensional illustrations of binding. Representations, colouring and display of interactions are as in Fig. 2.

**Figure 4. Structural delineation of HDAC8 selective inhibition.**

(a-d) Close-up views shown as ribbon and sticks of the superposed structures of (a) smHDAC8, smHDAC8-H292M and hHDAC8, (b) hHDAC4 and hHDAC7, (c) hHDAC1, hHDAC2 and hHDAC3, and (d) zebrafish zHDAC6-(catalytic domain1)CD1, zHDAC6-CD2 and zHDAC10. The catalytic zinc is shown as orange sphere. HDAC8-selective inhibitor **PCI-34051** is shown in (a) as light blue sticks when bound in HDAC8-selective pocket. In the other HDACs this pocket is not formed: residues from L1 and L6 loops are protruding and forming a lock over the catalytic tyrosine (other class I and class IIb HDACs) or its replacing histidine in the case of class IIa HDACs. (e) Surface representation of the pocket accommodating the linker and capping groups of the HDAC8-selective inhibitors. The **PCI-34051** inhibitor is represented as sticks and lays on the catalytic tyrosine (purple). The pocket walls are formed by residues from the L1 (yellow) and L6 (green) loops. (f,g,h) Surface representation of the same region

in hHDAC4 (f), hHDAC3 (g), and hHDAC6-CD2 (h) using the same colour code as in (e). In these latter HDACs, L1 and L6 loop residues are interacting and forming a L1-L6 lock over the pocket.

**Figure 5. HDAC8 mutants inhibition by PCI-34051, NCC-149 and QSN, and structure of the human HDAC8-mL6/QSN complex.** (a) Structure-based sequence alignment of L1, L4 and L6 loop sequences from various HDACs. Red boxes show the regions that have been swapped between HDAC1 and HDAC8 in the mutational analysis. (b)  $IC_{50}$  values for **PCI-34051**, **NCC-149** and **QSN** on human HDAC8 WT and mutants. Assays were done in triplicate. Error bars represent the SD. (c) Close-up view of the hHDAC8-selective pocket in the human HDAC8-mL6/QSN complex (left panel) compared to WT hHDAC8 bound to **PCI-34051** (adapted from the smHDAC8-H292M structure) (right panel) and the superposition of the two structures (middle panel). The side chain of the arginine R273 (from the hHDAC8-mL6/QSN complex) replacing P273 (from WT hHDAC8) binds into the HDAC8-selective pocket where **PCI-34051** (and the other HDAC8-selective inhibitor) were shown to bind. This mutation, which is the only common mutation to all mutants used, is most likely responsible for the general decrease of activity observed for these mutants.



## References

1. Choudhary, C., Weinert, B.T., Nishida, Y., Verdin, E. & Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol* **15**, 536-50 (2014).
2. Drazic, A., Myklebust, L.M., Ree, R. & Arnesen, T. The world of protein acetylation. *Biochim Biophys Acta* **1864**, 1372-401 (2016).
3. Verdin, E. & Ott, M. 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nat Rev Mol Cell Biol* **16**, 258-64 (2015).
4. Haberland, M., Montgomery, R.L. & Olson, E.N. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet* **10**, 32-42 (2009).
5. Marmorstein, R. & Zhou, M.M. Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harb Perspect Biol* **6**, a018762 (2014).
6. Seto, E. & Yoshida, M. Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb Perspect Biol* **6**, a018713 (2014).
7. Musselman, C.A., Lalonde, M.E., Cote, J. & Kutateladze, T.G. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19**, 1218-27 (2012).
8. Arrowsmith, C.H., Bountra, C., Fish, P.V., Lee, K. & Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat Rev Drug Discov* **11**, 384-400 (2012).
9. Campbell, R.M. & Tummino, P.J. Cancer epigenetics drug discovery and development: the challenge of hitting the mark. *J Clin Invest* **124**, 64-9 (2014).
10. Dawson, M.A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12-27 (2012).
11. Falkenberg, K.J. & Johnstone, R.W. Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nat Rev Drug Discov* **13**, 673-91 (2014).
12. Jones, P.A., Issa, J.P. & Baylin, S. Targeting the cancer epigenome for therapy. *Nat Rev Genet* **17**, 630-41 (2016).
13. West, A.C. & Johnstone, R.W. New and emerging HDAC inhibitors for cancer treatment. *J Clin Invest* **124**, 30-9 (2014).
14. Chan, T.S., Tse, E. & Kwong, Y.L. Chidamide in the treatment of peripheral T-cell lymphoma. *Onco Targets Ther* **10**, 347-352 (2017).
15. Maolanon, A.R., Madsen, A.S. & Olsen, C.A. Innovative Strategies for Selective Inhibition of Histone Deacetylases. *Cell Chem Biol* **23**, 759-68 (2016).
16. Balasubramanian, S. et al. A novel histone deacetylase 8 (HDAC8)-specific inhibitor PCI-34051 induces apoptosis in T-cell lymphomas. *Leukemia* **22**, 1026-34 (2008).
17. Suzuki, T. et al. Design, synthesis, and biological activity of NCC149 derivatives as histone deacetylase 8-selective inhibitors. *ChemMedChem* **9**, 657-64 (2014).
18. Chakrabarti, A. et al. Targeting histone deacetylase 8 as a therapeutic approach to cancer and neurodegenerative diseases. *Future Med Chem* **8**, 1609-34 (2016).
19. Chakrabarti, A. et al. HDAC8: a multifaceted target for therapeutic interventions. *Trends Pharmacol Sci* **36**, 481-92 (2015).
20. Ramos, T.L. et al. HDAC8 overexpression in mesenchymal stromal cells from JAK2+ myeloproliferative neoplasms: a new therapeutic target? *Oncotarget* **8**, 28187-28202 (2017).
21. Deardorff, M.A. et al. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature* **489**, 313-7 (2012).
22. Deardorff, M.A., Porter, N.J. & Christianson, D.W. Structural aspects of HDAC8 mechanism and dysfunction in Cornelia de Lange syndrome spectrum disorders. *Protein Sci* **25**, 1965-1976 (2016).
23. Kaiser, F.J. et al. Loss-of-function HDAC8 mutations cause a phenotypic spectrum of Cornelia de Lange syndrome-like features, ocular hypertelorism, large fontanelle and X-linked inheritance. *Hum Mol Genet* **23**, 2888-900 (2014).

24. Dasgupta, T., Antony, J., Braithwaite, A.W. & Horsfield, J.A. HDAC8 Inhibition Blocks SMC3 Deacetylation and Delays Cell Cycle Progression without Affecting Cohesin-dependent Transcription in MCF7 Cancer Cells. *J Biol Chem* **291**, 12761-70 (2016).
25. Hsieh, C.L. et al. Alterations in histone deacetylase 8 lead to cell migration and poor prognosis in breast cancer. *Life Sci* **151**, 7-14 (2016).
26. Lopez, G. et al. HDAC8, A Potential Therapeutic Target for the Treatment of Malignant Peripheral Nerve Sheath Tumors (MPNST). *PLoS One* **10**, e0133302 (2015).
27. Zhang, K. et al. HDAC8 functions in spindle assembly during mouse oocyte meiosis. *Oncotarget* **8**, 20092-20102 (2017).
28. Heimbürg, T. et al. Structure-Based Design and Synthesis of Novel Inhibitors Targeting HDAC8 from *Schistosoma mansoni* for the Treatment of Schistosomiasis. *J Med Chem* **59**, 2423-35 (2016).
29. Marek, M. et al. Structural basis for the inhibition of histone deacetylase 8 (HDAC8), a key epigenetic player in the blood fluke *Schistosoma mansoni*. *PLoS Pathog* **9**, e1003645 (2013).
30. Arts, J. et al. JNJ-26481585, a novel "second-generation" oral histone deacetylase inhibitor, shows broad-spectrum preclinical antitumoral activity. *Clin Cancer Res* **15**, 6841-51 (2009).
31. Venugopal, B. et al. A phase I study of quisinostat (JNJ-26481585), an oral hydroxamate histone deacetylase inhibitor with evidence of target modulation and antitumor activity, in patients with advanced solid tumors. *Clin Cancer Res* **19**, 4262-72 (2013).
32. Dubois, F. et al. Histone deacetylase inhibitors induce apoptosis, histone hyperacetylation and up-regulation of gene transcription in *Schistosoma mansoni*. *Mol Biochem Parasitol* **168**, 7-15 (2009).
33. Somoza, J.R. et al. Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases. *Structure* **12**, 1325-34 (2004).
34. Watson, P.J. et al. Insights into the activation mechanism of class I HDAC complexes by inositol phosphates. *Nat Commun* **7**, 11262 (2016).
35. Lauffer, B.E. et al. Histone deacetylase (HDAC) inhibitor kinetic rate constants correlate with cellular histone acetylation but not transcription and cell viability. *J Biol Chem* **288**, 26926-43 (2013).
36. Decroos, C. et al. Variable active site loop conformations accommodate the binding of macrocyclic largazole analogues to HDAC8. *Biochemistry* **54**, 2126-35 (2015).
37. Millard, C.J. et al. Class I HDACs share a common mechanism of regulation by inositol phosphates. *Mol Cell* **51**, 57-67 (2013).
38. Watson, P.J., Fairall, L., Santos, G.M. & Schwabe, J.W.R. Structure of HDAC3 bound to co-repressor and inositol tetrakisphosphate. *Nature* **481**, 335-340 (2012).
39. Burli, R.W. et al. Design, synthesis, and biological evaluation of potent and selective class IIa histone deacetylase (HDAC) inhibitors as a potential therapy for Huntington's disease. *J Med Chem* **56**, 9934-54 (2013).
40. Hai, Y. & Christianson, D.W. Histone deacetylase 6 structure and molecular basis of catalysis and inhibition. *Nat Chem Biol* **12**, 741-7 (2016).
41. Heimbürg, T. et al. Structure-based design and biological characterization of selective histone deacetylase 8 (HDAC8) inhibitors with anti-neuroblastoma activity. *J Med Chem* (2017).
42. Diebold, M.L., Fribourg, S., Koch, M., Metzger, T. & Romier, C. Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. *J Struct Biol* **175**, 178-88 (2011).
43. Marek, M., Shaik, T.B., Duclaud, S., Pierce, R.J. & Romier, C. Large-Scale Overproduction and Purification of Recombinant Histone Deacetylase 8 (HDAC8) from the Human-Pathogenic Flatworm *Schistosoma mansoni*. *Methods Mol Biol* **1436**, 109-18 (2016).
44. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-26 (1997).
45. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).

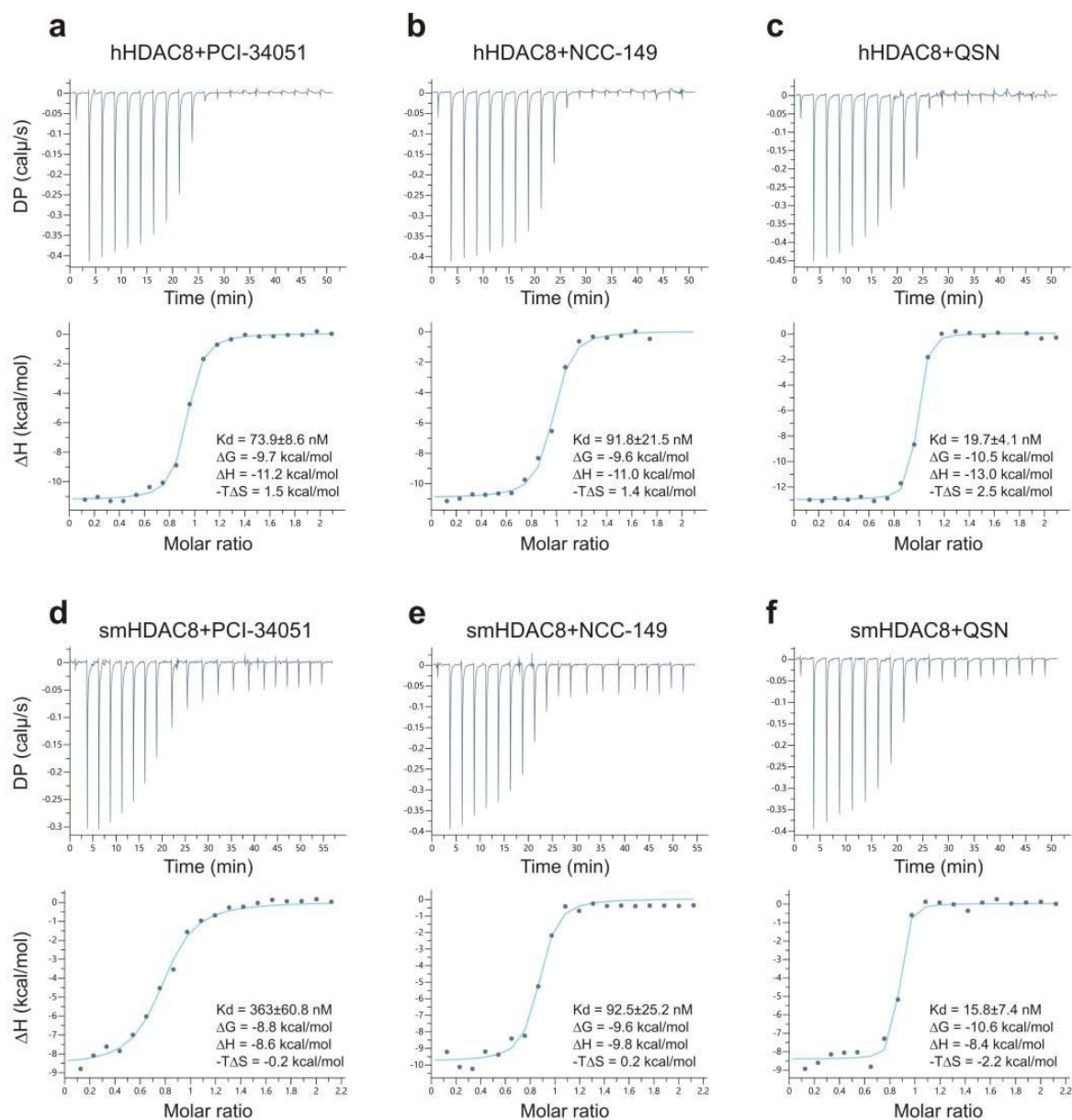


46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
47. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-21 (2010).
48. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
49. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **8**, 127-34 (1995).
50. Vanderstraete, M. et al. Dual targeting of insulin and venus kinase Receptors of *Schistosoma mansoni* for novel anti-schistosome therapy. *PLoS Negl Trop Dis* **7**, e2226 (2013).
51. Ramalho-Pinto, F.J. et al. *Schistosoma mansoni*: defined system for stepwise transformation of cercaria to schistosomule in vitro. *Exp Parasitol* **36**, 360-72 (1974).
52. Sali, A. & Blundell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815 (1993).

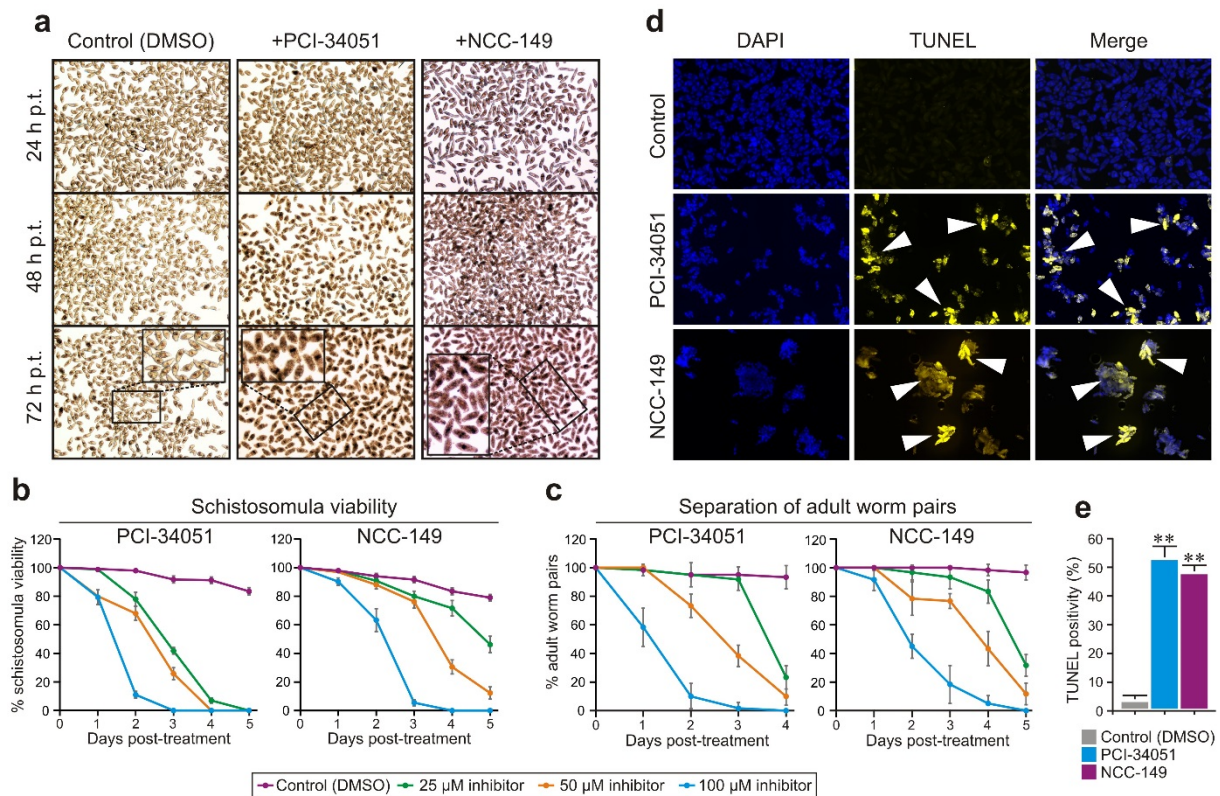
## **Structural basis of HDAC8 selective inhibition**

Marek, M.<sup>1,2,\*,\$</sup>, Shaik, T.B.<sup>1\*</sup>, Heimbürg, T.<sup>3</sup>, Chakrabarti, A.<sup>4</sup>, Lancelot, J.<sup>5</sup>, Ramos-Morales, E.<sup>1</sup>, Da Veiga, C.<sup>6</sup>, Kalinin, D.<sup>7</sup>, Melesina, J.<sup>3</sup>, Robaa, D.<sup>3</sup>, Schmidtkunz, K.<sup>4</sup>, Suzuki, T.<sup>8</sup>, Holl, R.<sup>7</sup>, Ennifar, E.<sup>6</sup>, Pierce, R.J.<sup>5</sup>, Jung, M.<sup>4</sup>, Sippl, W.<sup>3</sup> & Romier, C.<sup>5,1</sup>

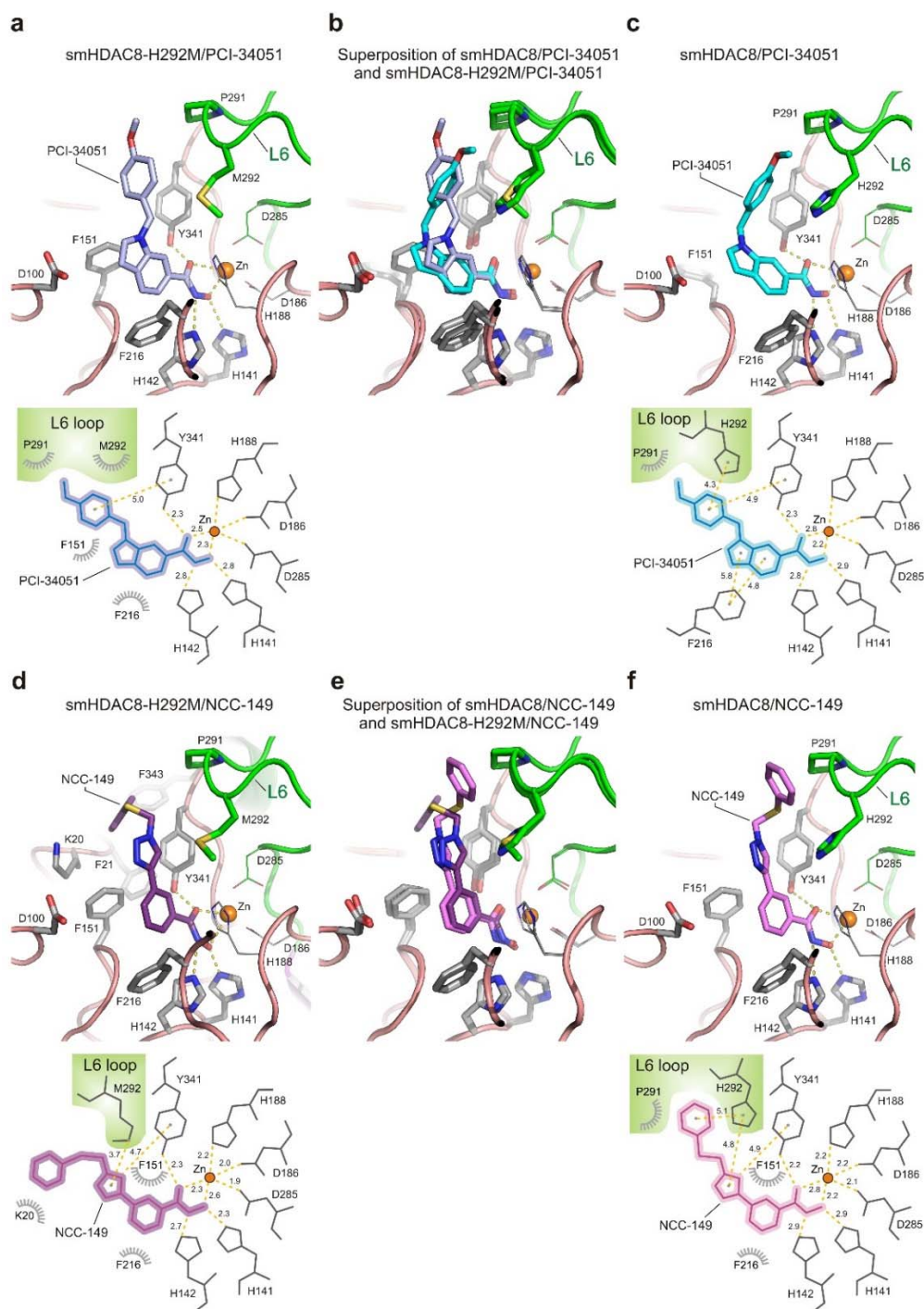
## **Supplementary Information**



**Supplementary Figure 1. ITC measured affinities and thermodynamic parameters of PCI-34051, NCC-149 and Quisinostat (QSN) binding to hHDAC8 and smHDAC8.** ITC profiles of the titration of hHDAC8 (a–c) and smHDAC8 (d–f) with PCI-34051 (a, d), NCC-149 (b, e) and QSN (c, f). Top panels, titration data of enzyme into corresponding inhibitor solution. DP, Differential power. Bottom panels, integrated heat measurements for the titration enzyme with the corresponding inhibitor.

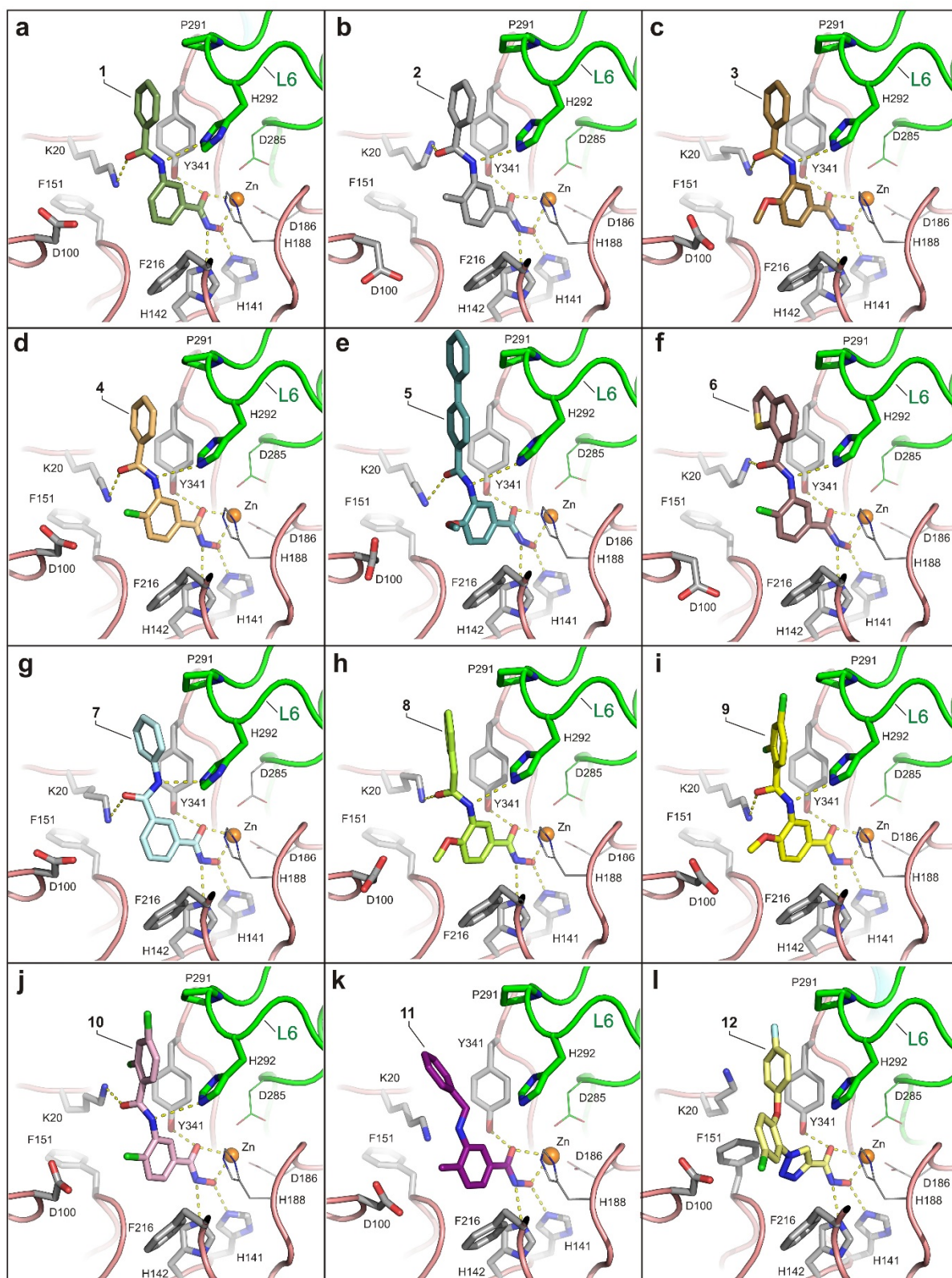


**Supplementary Figure 2. Fitness analysis of schistosomes treated with inhibitors PCI-34051 and NCC-149.** (a) Time-course analysis of schistosomula viability. *S. mansoni* schistosomula were cultivated in the presence of PCI-34051 or NCC-149, and their viability was assessed by microscopic observations at the indicated time points post treatment (p.t.). Schistosomula cultivated in the DMSO alone served as a control. (b) Dose- and time-dependent mortality of schistosomula induced by PCI-34051 and NCC-149. *S. mansoni* schistosomula (1000 per well) were incubated in 1-mL of culture medium with varying quantities of inhibitors or the solvent (DMSO). Both inhibitors have a dose-dependent effect on larval viability, with PCI-34051 inducing a somewhat stronger phenotypic response than NCC-149. Assays were done in triplicate; error bars represent the SD. (c) Induction of separation of *S. mansoni* adult worm pairs. The paired status of male and female adult worms was assessed daily in the presence of varying quantities of PCI-34051 and NCC-149. Both inhibitors have a dose-dependent effect on adult worm pairing. Assays were done in triplicate; error bars represent the SD. (d) TUNEL assays. Fluorescent microscopy of *S. mansoni* schistosomula incubated with DMSO alone (top panels), PCI-34051 (middle panels), or NCC-149 (bottom panels) for 96 h. DAPI (blue), TUNEL (yellow) and merged figures are presented. (e) Quantification of TUNEL positivity of schistosomula incubated for 96 h with PCI-34051 or NCC-149 (both at 100  $\mu$ M), or with DMSO alone. Mortality of *in vitro* cultivated schistosomes treated with PCI-34051 and NCC-149 is due to the triggering of apoptosis. Assays were done in triplicate; error bars represent the SD.

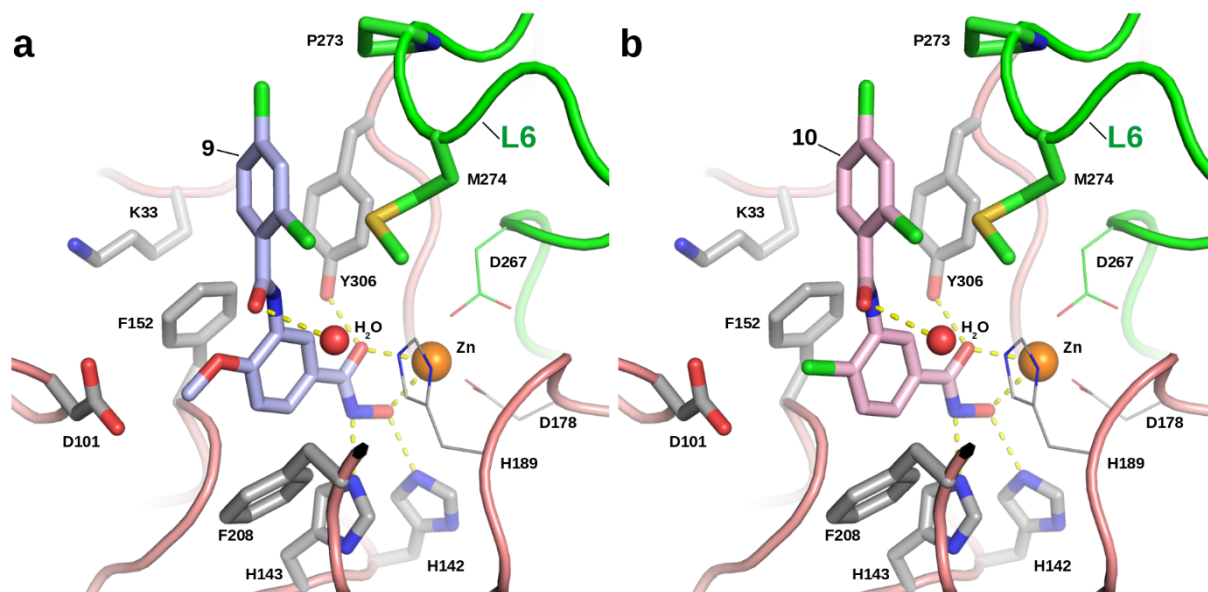


**Supplementary Figure 3. Structures of PCI-34051 and NCC-149 bound to the smHDAC8-H292M mutant.** (a-c) Close-up view of HDAC8-selective inhibitor **PCI-34051** bound to smHDAC8 WT (a) and the smHDAC8 H292M mutant (c), and the superposition of the two structures (b). The upper panels show the structural data, whereas the lower panels provide a schematic view of the interaction. In the structural views, the inhibitors and important residues are shown as sticks. The L6 loop is colored green. The catalytic zinc ion is shown as orange sphere. Zinc coordination and hydrogen bonds are shown as dashed lines. In the schematic views, Zinc coordination, hydrogen bonds and aromatic interactions are shown as dashed lines, and the corresponding distances between the atoms and/or chemical groups are given in Å. Hydrophobic contacts are shown by grey arcs with spokes radiating toward the atoms involved. (d-f) Same as for (a-c) with the **NCC-149** HDAC8-selective inhibitor.

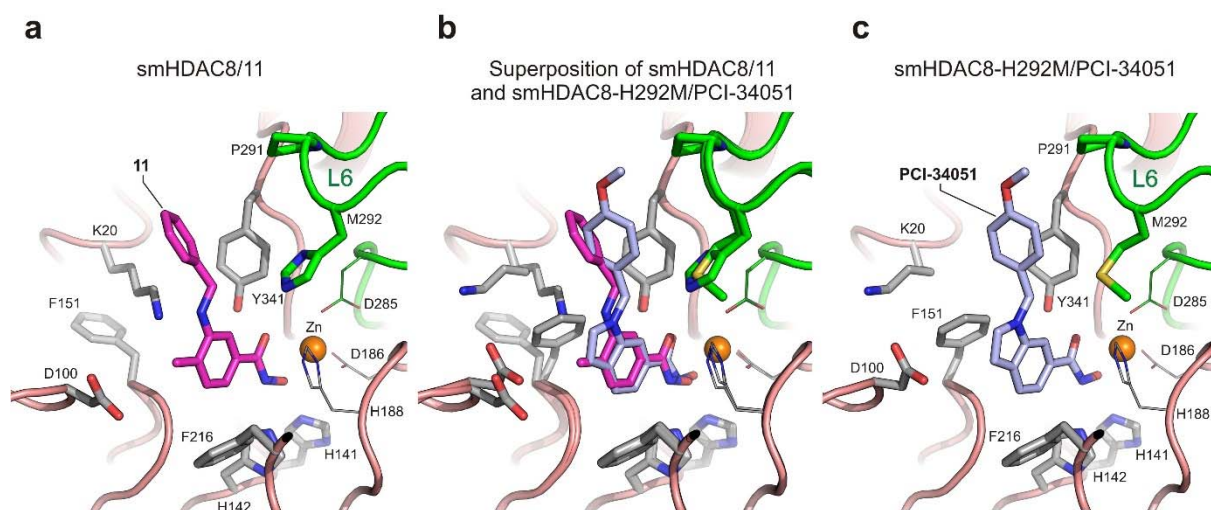




**Supplementary Figure 4. Close-up view of HDAC8-selective inhibitors bound to smHDAC8.** (a-h) Close-up views of 3-benzamido-benzohydroxamates HDAC8-selective inhibitors **1-11** bound to smHDAC8. (i) Close-up view of the unrelated 1-[5-chloro-2-(4-fluorophenoxy)phenyl]-N-hydroxy-1H-1,2,3-triazole-4-carboxamide inhibitor **12** that also shows HDAC8-selective inhibition. All inhibitors bind into a HDAC8-specific pocket formed by catalytic tyrosine Y341 and residues from L1 and L6 loops. The inhibitors and important residues are shown as sticks. The L6 loop is colored green. The catalytic zinc ion is shown as orange sphere. Zinc coordination and hydrogen bonds are shown as dashed lines.

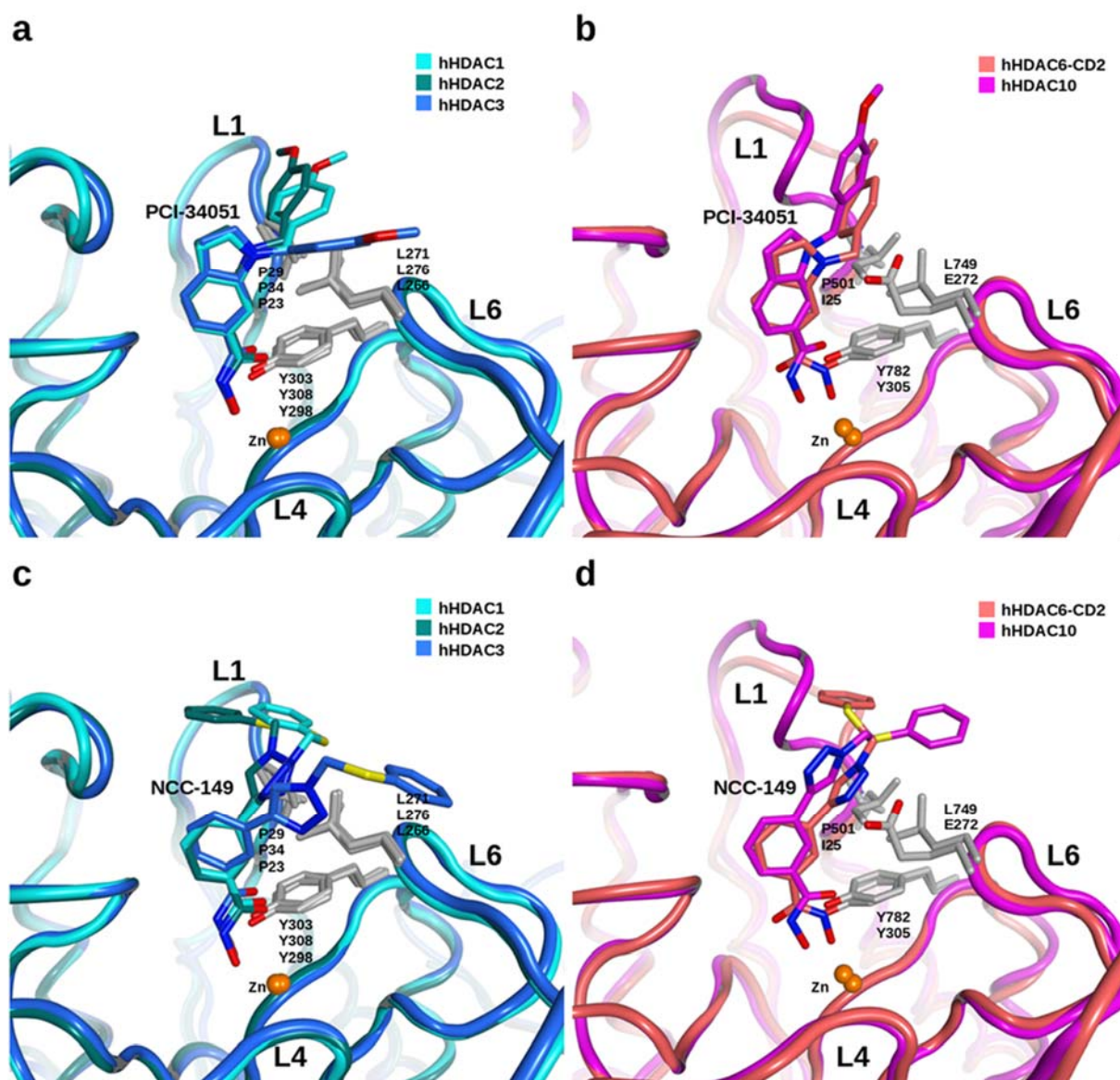


**Supplementary Figure 5. Docking results of compounds 9 and 10 bound to human HDAC8.** Inhibitor 9 (a), pale blue colored sticks, and 10 (b), pale pink colored sticks, are shown at the human HDAC8 binding pocket. Compared to smHDAC8, the capping group of inhibitors 9 and 10 are flipped when bound to hHDAC8, which exposes the chloro-substituents to the solvent, potentially reducing binding affinity. Protein backbone is represented as pale salmon colored ribbon. Protein residues are shown as grey colored sticks and the zinc ion as orange sphere. The ribbon and residues of the L6 loop are colored in green. Yellow dashed lines represent zinc coordination and hydrogen bonds.

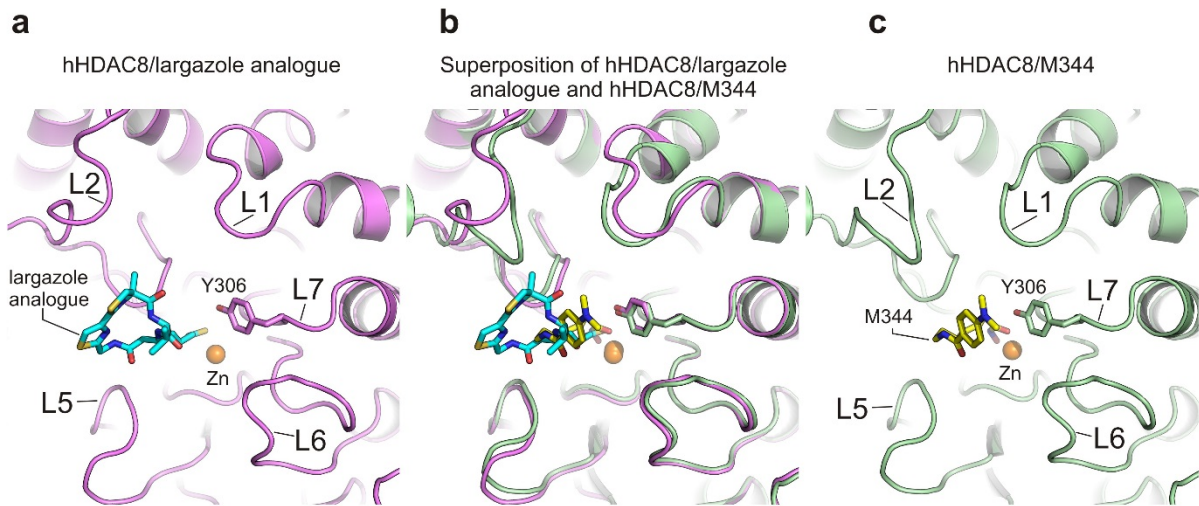


**Supplementary Figure 6. Comparison of the binding modes of compound 11 and PCI-34051 to smHDAC8.** Structures of compound **11** bound to smHDAC8 (**a**) and **PCI-34051** bound to the smHDAC8-H292M mutant (**c**), and the superposition of these two structures (**b**). Both inhibitors binds with a similar conformation to smHDAC8 that most likely mimics the binding of HDAC8-selective inhibitors to hHDAC8. Protein residues are shown as sticks, the catalytic zinc ion is shown as orange sphere, and L6 loop residues are colored green.

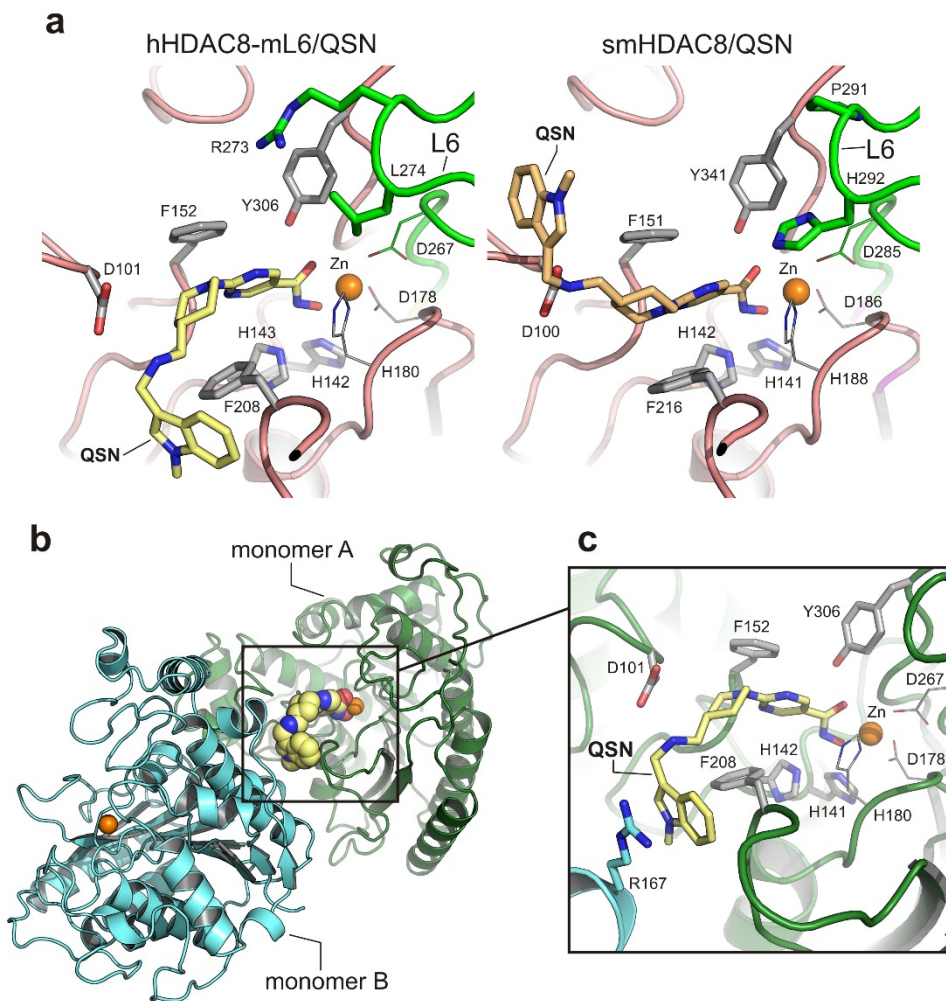




**Supplementary Figure 7. Docking poses of PCI-34051 and NCC-149 in different HDAC isozymes.** Modeling of binding of HDAC8-selective inhibitors **PCI-34051** (a,b) and **NCC-149** (c,d) to hHDAC1-3 (a,c) and hHDAC6-CD2 and hHDAC10 (b,d). The hydroxamic acid group of both L-shaped HDAC8 inhibitors is not able to reach the catalytic zinc ion in HDAC1-3 and HDAC10 and to chelate it in a bidentate fashion. Meanwhile, in HDAC6, the hydroxamate group also shows a monodentate chelation of the zinc ion similar to that observed in the crystal structures of zHDAC6 with the benzohydroxamate derivatives Nexturastat A and HPOB (PDB ID: 5G0J and 5EF7, respectively). Yet, the capping groups of both inhibitors are resting on the surface of the protein and cannot interact strongly with the catalytic tyrosine due to the lock imposed by residues from L1 and L6 loops of these different HDACs. As a result, inhibitors are more solvent-exposed, which probably negatively influences their inhibitory activity. Protein backbone is represented as ribbon. Color codes for protein backbone and corresponding inhibitor docking pose are depicted in the legend. Protein residues (grey color) and inhibitors (see color codes in the legend) are shown as sticks, the catalytic zinc ion as orange sphere.

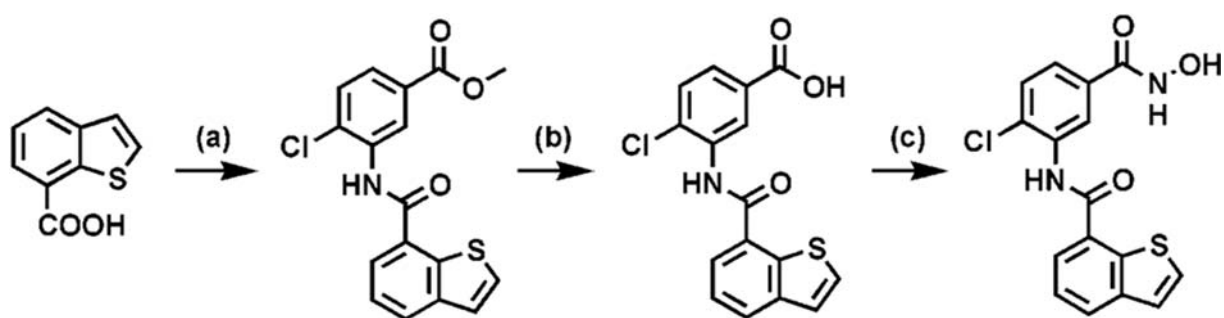


**Supplementary Figure 8. Structural plasticity and flexibility of human HDAC8 L1 and L2 loops.** Structure of a largazole analogue (a) (PDB ID 4RNO) and inhibitor M344 (b) (PDB ID 1T67) bound to human HDAC8, and the superposition of both structures (c). HDAC8 L1 and L2 loops show strong plasticity and flexibility to accommodate the inhibitors in their active sites. In contrast, L6 and L7 (containing the catalytic tyrosine 306) loop conformations are constrained, providing a stable unique surface for the interaction with HDAC8-selective inhibitors. The proteins are shown as ribbons and the inhibitors as sticks, the catalytic zinc ion is shown as orange sphere.

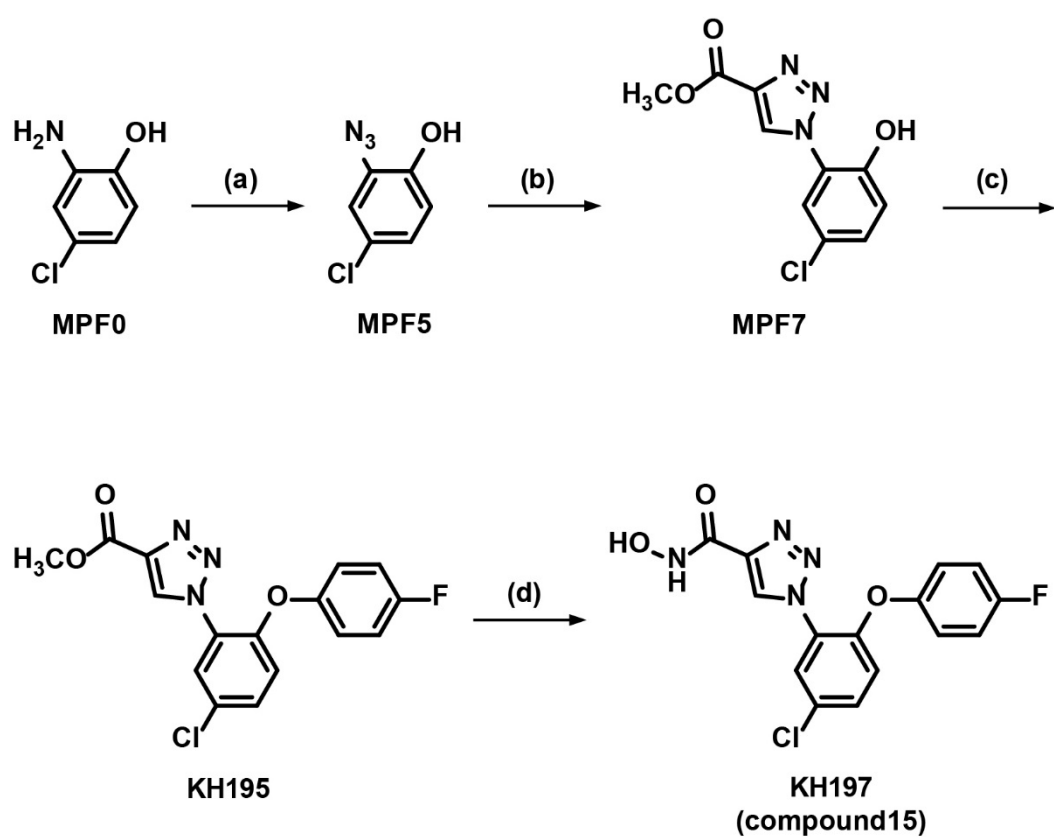


**Supplementary Figure 9. Binding mode of Quisinostat (QSN) in the structure of the human HDAC8-mL6/ QSN complex.** (a) Conformation of Quisinostat (QSN) bound to the human HDAC8-mL6 mutant (left panel) and to smHDAC8 (right panel). QSN does not interact with the L6 loop in both structures. The conformation difference observed therefore most likely stems from the involvement of QSN in crystal packing in the hHDAC8-mL6/QSN complex, an artefact observed in many of the HDAC/inhibitor structures published so far. (b,c) Overall (b) and closeup (c) views of QSN involvement in crystal packing. QSN is absolutely required for crystal appearance.

a



b



**Supplementary Figure 10.** Synthesis pathways for (a) the 3-(Benzthiophene-7-carboxamido)-4-chlorobenzohydroxamate inhibitor (**6**) and (b) the 1-[5-chloro-2-(4-fluorophenoxy)phenyl]-N-hydroxy-1H-1,2,3-triazole-4-carboxamide inhibitor (**12**).

**Supplementary Table 1. Data collection and refinement statistics for the structures of smHDAC8 WT and H292M mutant bound to PCI-34051, NCC-149 and Quisinostat.**

Data collection*	smHDAC8/ PCI-34051	smHDAC8/ NCC-149	smHDAC8/ Quisinostat	smHDAC8-H292M/ PCI-34051	smHDAC8-H292M/ NCC-149
Space group	P1	P1	P1	P1	P1
Cell dimensions					
a, b, c (Å)	70.55 70.62 97.89	71.4, 71.46, 99.25	70.78, 70.78, 97.9	70.6, 70.67, 98.18	70.8, 70.84, 98.41
$\alpha$ , $\beta$ , $\gamma$ (°)	75.52 78.01 85.51	78.13, 75.53, 85.38	77.97, 75.66, 85.74	75.82, 78.24, 86.05	78.13, 75.97, 85.98
Resolution (Å)	50-2.09 (2.22-2.09)	50-1.84 (1.96-1.84)	50-1.55 (1.64-1.55)	50-1.90 (2.01-1.9)	50 -1.85 (1.96 -1.85)
Rsym or Rmerge	11.5 (82.5)	5.4 (58.7)	5 (66.4)	15.4 (99.4)	15.5 (75.6)
I / $\sigma$	8.36 (1.47)	15.4 (2.2)	8.31 (1.02)	6.6 (1.19)	5.5 (1.45)
Completeness (%)	97 (92.1)	94.9 (89.2)	82.7 (48.4)	97.5 (94.7)	97.1 (92.2)
Redundancy	3.4 (3.4)	3.4 (3.2)	1.86 (1.85)	3.5 (3.6)	3.5 (3.5)
CC(1/2)	99.4 (58.0)	99.9 (74.4)	99.8 (22.9)	99.3 (49.4)	98.9 (50.8)
<b>Refinement</b>					
Resolution (Å)	43-2.50	47-1.85	47-1.55	44-1.90	44-1.85
No. reflections	60906	152126	218321	138739	151470
Rwork / Rfree	0.180 / 0.253	0.150 / 0.183	0.168 / 0.202	0.211 / 0.265	0.207 / 0.253
Number of atoms					
Protein	12897	13119	13124	12877	12989
Ligand/ion	136	322	369	222	231
Water	197	1160	1102	797	910
B-factors					
Protein	29.83	29.67	28.16	29.32	27.69
Ligand/ion	52.10	45.82	43.95	55.95	45.49
Water	26.46	40.81	38.61	36.45	34.64
R.m.s deviations					
Bond lengths (Å)	0.008	0.008	0.007	0.009	0.008
Bond angles (°)	1.032	0.832	0.891	1.001	0.887
* Values in parentheses are for the highest-resolution shell.					

**Supplementary Table 2. IC<sub>50</sub> values measured for HDAC8-selective inhibitors on different HDACs.**

Compound	Chemical structure	smHDAC8 IC <sub>50</sub> (nM)	hHDAC8 IC <sub>50</sub> (nM)	hHDAC1 IC <sub>50</sub> (μM)	hHDAC6 IC <sub>50</sub> (μM)
PCI-34051		436 ± 61	78 ± 18	28.3 ± 2.0	48.2 ± 6.2
NCC-149		95 ± 9	44 ± 5	nd	nd
1 (TH31)*		468 ± 79	582 ± 48	33.6 ± 1.8	3.0 ± 0.3
2 (TH33)*		116 ± 38	204 ± 22	8.4 ± 2.0	0.9 ± 0.4
3 (TH39)*		190 ± 54	88 ± 24	2.3 ± 1.2	2.5 ± 1.1
4 (TH61)*		67 ± 10	120 ± 37	11.6 ± 3.9	0.12 ± 0.02
5 (TH65)*		75 ±	26 ± 18	6.3 ± 2.1	0.39 ± 0.002
6 (TH120)		97 ± 16	155 ± 52	21.4 ± 2.2	0.053 ± 0.015
7 (TH97)		220 ± 4	120 ± 37	39.1 ± 8.9	3.1 ± 0.1
8 (TH101)*		183 ± 39	512 ± 30	28.9 ± 8.6	5.1 ± 0.7
9 (TH86)*		122 ± 19	548 ± 94	13.0 ± 1.9	2.3 ± 0.4
10 (TH104)*		191 ± 17	1184 ± 45	31.6 ± 19.8	0.8 ± 0.1
11 (TH34)*		1260 ± 170	260 ± 36	21.8 ± 2.1	5.1 ± 0.3
12 (KH197)		504 ± 46	2200 ± 680	nd	nd

\* Taken from <sup>1</sup> and <sup>2</sup>; nd, not determined.

**Supplementary Table 3. Data collection and refinement statistics for the structures of smHDAC8 bound to compounds 1-12.**

Data collection*	smHDAC8 / 2	smHDAC8 / 3	smHDAC8 / 4	smHDAC8 / 5	smHDAC8 / 6	smHDAC8 / 7
Space group	P1	P1	P1	P1	P1	P1
Cell dimensions						
a, b, c (Å)	70.92 71.22 99.88	70.71 70.75 98.29	70.59, 70.61, 98.19	70.62 71.13 98.48	70.7, 70.71, 98.32	70.36, 70.49, 98.27
$\alpha, \beta, \gamma$ (°)	77.43 75.61 85.45	77.95 75.54 85.38	77.72, 75.87, 85.57	77.84 76.19 85.81	75.66, 78.01, 85.7	75.84, 77.74, 85.59
Resolution (Å)	50-2.04 (2.16-2.04)	50 - 2.03 (2.16-2.03)	50–1.94 (2.06–1.94)	50-1.94 (2.06-1.94)	50–1.69 (1.8–1.69)	50–1.75 (1.85-1.75)
Rsym or Rmerge	8.2 (43)	12.4 (57.2)	25.9 (42.5)	16.8 (127.1)	10.7 (79.4)	4.2 (27.5)
I / $\sigma$	10.62 (2.55)	7.35 (1.87)	3.44 (1.57)	4.65 (0.83)	8.28 (1.31)	11.8 (2.52)
Completeness (%)	97 (93.4)	96.7 (90.5)	95.0 (90.9)	95.7 (89.2)	94.3 (90.7)	93.7 (91.5)
Redundancy	3.5 (3.4)	3.5 (3.4)	3.6 (3.5)	2.87 (2.83)	3.3 (3.45)	1.7 (1.7)
CC(1/2)	99.6 (80.2)	99.8 (80.9)	92.0 (76.4)	98.4 (28.5)	99.4 (55.7)	99.8 (83.7)
<b>Refinement</b>						
Resolution (Å)	35-2.37	35-2.50	49-1.94	47-1.95	48-1.69	48-1.75
No. reflections	69306	61487	127009	125345	189740	169960
Rwork / Rfree	0.159 / 0.220	0.158 / 0.236	0.169 / 0.215	0.210 / 0.257	0.197 / 0.229	0.150 / 0.183
Number of atoms						
Protein	12971	12973	13108	12760	12999	13021
Ligand/ion	133	184	247	293	128	343
Water	325	474	758	480	487	1365
B-factors						
Protein	35.52	28.74	27.35	22.65	28.08	23.35
Ligand/ion	43.44	43.36	45.18	44.66	41.82	39.41
Water	35.68	30.94	34.13	27.66	31.17	35.18
R.m.s. deviations						
Bond lengths (Å)	0.009	0.011	0.010	0.016	0.007	0.006
Bond angles (°)	0.985	1.166	1.000	1.364	0.908	0.798
* Values in parentheses are for the highest-resolution shell.						

Data collection*	smHDAC8 / 8	smHDAC8 / 9	smHDAC8 / 10	smHDAC8 / 11	smHDAC8 / 12
Space group	P1	P1	P1	P1	P1
Cell dimensions					
a, b, c (Å)	71.26, 71.27, 99.14	70.65, 70.73, 98.12	71.2, 71.22, 99.02	71.2, 71.29, 99.07	70., 70.7, 98.3
$\alpha, \beta, \gamma$ (°)	77.92, 75.54, 85.5	75.46, 78.05, 85.54	78.28, 75.64, 85.72	75.69, 78.14, 85.65	75.9, 78.3, 85.6
Resolution (Å)	50–1.84 (1.95–1.84)	50–1.75 (1.85–1.75)	50–2.0 (2.12–2.0)	50–1.99 (2.11–1.99)	50–1.65 (1.76–1.65)
Rsym or Rmerge	15.5 (81.8)	12.6 (55.5)	10.8 (66.6)	7.4 (58.2)	4.3 (78.5)
I / $\sigma$	8.4 (1.75)	13.94 (2.39)	8.67 (1.84)	11.91 (2.05)	14.93 (1.43)
Completeness (%)	93.2 (88.4)	95.7 (91.3)	95.6 (91.9)	95.7 (92.3)	92.1 (87.3)
Redundancy	3.3 (3.1)	3.2 (3.1)	3.5 (3.4)	3.5 (3.4)	3.6 (3.6)
CC(1/2)	98.5 (61.1)	98.9 (84.5)	99.3 (73.6)	99.7 (72.9)	99.9 (58.7)
<b>Refinement</b>					
Resolution (Å)	48-1.84	48-1.75	47-2.00	48-1.99	49-1.66
No. reflections	149024	175117	120102	122257	196617
Rwork / Rfree	0.177 / 0.218	0.189 / 0.226	0.160 / 0.205	0.156 / 0.197	0.122 / 0.194
Number of atoms					
Protein	13076	12726	12952	13131	13168
Ligand/ion	132	406	287	213	243
Water	999	1596	787	794	850
B-factors					
Protein	23.10	18.36	28.84	32.50	34.54
Ligand/ion	35.82	38.37	49.26	47.83	55.27
Water	31.11	33.40	37.45	39.79	42.88
R.m.s. deviations					
Bond lengths (Å)	0.008	0.007	0.011	0.008	0.008
Bond angles (°)	0.842	0.846	0.967	0.834	0.898
* Values in parentheses are for the highest-resolution shell.					



**Supplementary Table 4. HDAC8 mutants and activity measurements.**

HDAC8 variant	Relative deacetylase activity (%)	T <sub>m</sub> (°C)
hHDAC8 WT	100	45 ± 0.5
hHDAC8-P273R	6.4 ± 0.3	40.2 ± 0.7
hHDAC8-P273R/C275G	11.4 ± 0.1	40 ± 0.5
hHDAC8-mL6	10.6 ± 0.1	40.6 ± 0.2
hHDAC8-mL1/mL6	3.3 ± 0.05	39.6 ± 0.4
hHDAC8-mL6/L179I	1.8 ± 0.06	38.8 ± 0.5
hHDAC8-mL1/mL6/L179I	0.5 ± 0.04	37.1 ± 0.3
smHDAC8 WT	100	54.1 ± 0.9
smHDAC8-P291R	34.3 ± 1	49.5 ± 1
smHDAC8-P291R/R293G	18.3 ± 1	49.5 ± 1.5

**Supplementary Table 5. IC<sub>50</sub> values for hHDAC8 mutants with PCI-34051, NCC-149 and Quisinostat (QSN).**

Enzyme variant	Inhibitor	IC50 (nM)	Fold increase compared to WT
hHDAC8 WT	PCI-34051	77.7 ± 18.1	
hHDAC8-mL6	PCI-34051	1630 ± 450	21
hHDAC8-mL1/mL6	PCI-34051	2700 ± 620	34.8
hHDAC8-mL6/L179I	PCI-34051	1000 ± 200	12.9
hHDAC8-mL1/mL6/L179I	PCI-34051	5000 ± 900	64.4
hHDAC8 WT	NCC-149	44 ± 4.9	
hHDAC8-mL6	NCC-149	283 ± 85	6.4
hHDAC8-mL1/mL6	NCC-149	274 ± 50	6.2
hHDAC8-mL6/L179I	NCC-149	746 ± 85	17
hHDAC8-mL1/mL6/L179I	NCC-149	628 ± 69	14.3
hHDAC8 WT	QSN	64.4 ± 3.4	
hHDAC8-mL6	QSN	93 ± 8	1.4
hHDAC8-mL1/mL6	QSN	94 ± 5	1.5
hHDAC8-mL6/L179I	QSN	169 ± 27	2.6
hHDAC8-mL1/mL6/L179I	QSN	217 ± 31	3.4

**Supplementary Table 6. Data collection and refinement statistics for the structure of human HDAC8-mL6/Quisinostat (QSN) complex.**

<b>Data collection*</b>	<b>hHDAC8-mL6 / QSN</b>
Space group	P3 <sub>2</sub>
Cell dimensions	
a, b, c (Å)	106.36, 106.36, 82.05
α, β, γ (°)	90, 90, 120
Resolution (Å)	50 – 2.10 (2.22 – 2.1)
Rsym or Rmerge	16.0 (111.9)
I / σI	7.27 (1.2)
Completeness (%)	99.7 (98.2)
Redundancy	5.7 (5.5)
CC(1/2)	99.1 (53.5)
<b>Refinement</b>	
Resolution (Å)	37-2.10
No. reflections	60780
Rwork / Rfree	0.152 / 0.189
Number of atoms	
Protein	5695
Ligand/ion	64
Water	461
B-factors	
Protein	38.99
Ligand/ion	37.66
Water	46.94
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.868
* Values in parentheses are for the highest-resolution shell.	

## References

1. Heimburg, T. et al. Structure-Based Design and Synthesis of Novel Inhibitors Targeting HDAC8 from *Schistosoma mansoni* for the Treatment of Schistosomiasis. *J Med Chem* **59**, 2423-35 (2016).
2. Heimburg, T. et al. Structure-based design and biological characterization of selective histone deacetylase 8 (HDAC8) inhibitors with anti-neuroblastoma activity. *J Med Chem* (2017).

### 3.2.4. Crystal structure of smHDAC8 with J1075 derived TB compounds

Apart from J1038-derived TH compounds another class of inhibitors were also developed which target smHDAC8 selectively and are derived from the J1075 initial hit (Marek, Kannan *et al.* 2013). In this purpose, the initial hit compound J1075 was optimized further by our collaborators in Halle, Germany to design novel smHDAC8 selective inhibitors. Structure based drug design yielded many compounds, among which few inhibitors showed promising specificity towards smHDAC8. Among them, few inhibitors with potent inhibition of smHDAC8 were selected for structural studies. Among these, the TB series of compounds TB5, TB8, TB87 and TB98 were used for crystallization studies. All the data statistics were in the acceptable range and shown in the table (**Error! Reference source not found.**).

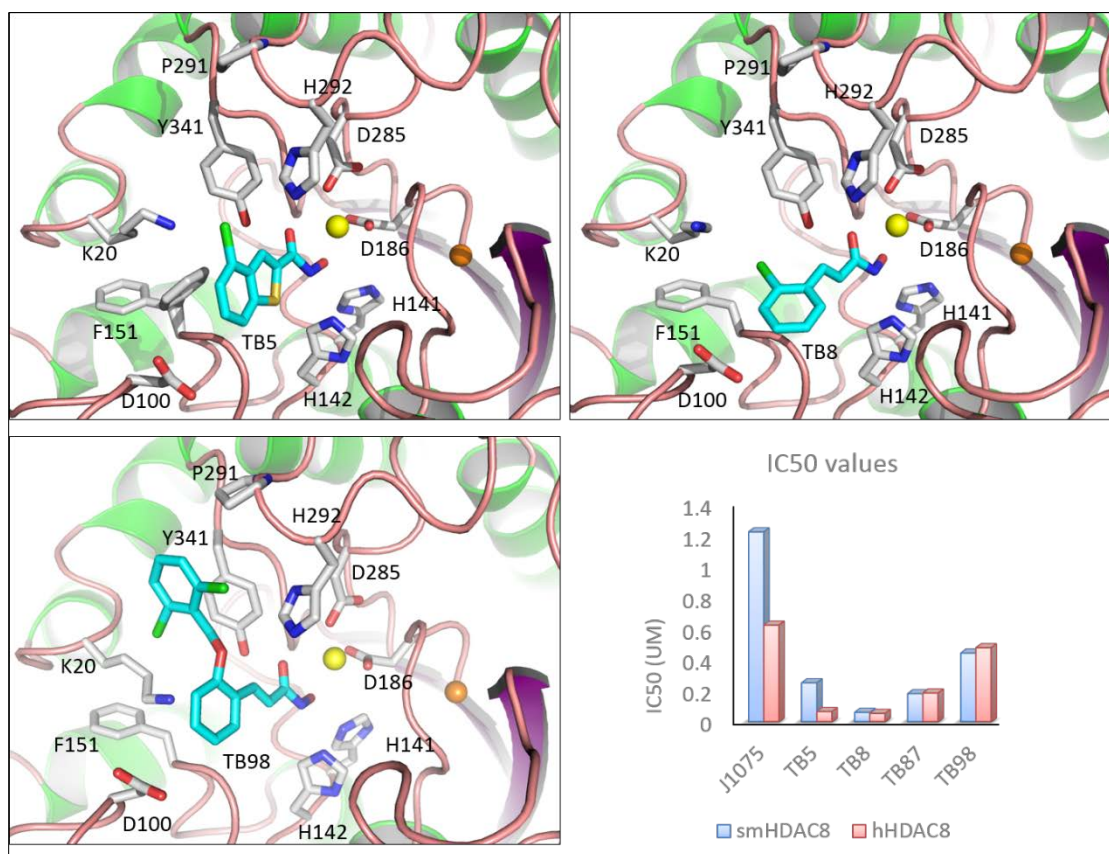
TB5 and TB8 are simple initial compounds of this inhibitor series. These two inhibitors bind in the smHDAC8 active site pocket in two different orientations (Figure 71). Both inhibitors coordinate the catalytic zinc with their hydroxamate war-head and interact with catalytic residues. According to *in vitro* assays, TB8 has the more potency towards HDAC8 when compared to other TB compounds. The direct comparison of TB8 compound is TB5 where both compounds are small linker inhibitors in this series. Structural inspection of these compounds in complex with smHDAC8, has revealed that TB8 binds in a schistosome-specific conformation where it forces F151 to adopt flipping-out conformation where as TB5 structure showed partial occupancy for both flipping-out and flipping-in conformations. Because of schistosome favorable conformation TB8 is more selective when compared to TB5. On the other hand, TB5 is more selective towards hHDAC8 than smHDAC8 because TB5 favors F151 in flipping-in conformation which is a hHDAC8 nature.

In the J1075 series, apart from TB5 and TB8, two compounds TB98 and TB87 were observed with more selectivity towards smHDAC8 over hHDAC8 during *in vitro* assays. To understand the selectivity differences through structural point of view I have solved the structure of TB98 with smHDAC8 while the second compound TB87 did not yield a good diffraction quality data. The crystal structure of smHDAC8 in complex with TB98 has revealed few interesting observations. First, TB98 bound in the active site of smHDAC8 by coordinating catalytic zinc and other catalytic residues. Next, TB98 binding in the active site pocket of smHDAC8 has forced F151 to adopt schistosome specific flipping-out conformation. In addition to that the presence of

chloride ions on the capping group and the L-shape of TB98 in active site pocket allowed to make interactions with loop L6 sub pocket which was made by P291 and H292. This mode of binding was also observed in the case of TH compounds where HDAC8-selective pocket was explained in detail. A similar kind of binding in TB98-smHDAC8 structure is reinforcing the fact that Loop L6 is the driving force in schistosome specific inhibition.

**Table 14: IC50 values of inhibitors**

Compound	smHDAC8 (nM)	hHDAC8 (nM)
TB5	250	65
TB8	60	53
TB87	180	184
TB98	441	475
MCC1761	1380	2720
MCC1713	274	365
KH197	503	2200



**Figure 71: smHDAC8-J1075 derivatives crystal structures:**  
smHDAC8 crystal structures in complex with J1075 derived inhibitors and also showing IC<sub>50</sub> values compared to J1075. Zinc and potassium ions are represented as yellow and orange spheres respectively.

### 3.2.5. Crystal structures of smHDAC8 with miscellaneous inhibitors

During my thesis, different hydroxamate inhibitors were provided by other chemists of the A-ParaDDisE consortium for structural determination in complex with smHDAC8. These inhibitors belong to different classes of inhibitors compared to the initial studies with the TH and TB studies, and were used to understand their mechanisms of inhibition.

#### 3.2.5.1. Triazole derivatives

Triazole derivatives are an important class of molecules in medicinal chemistry with lot of biological significance. Especially, 1,2,3-triazoles are heterocyclic compounds with high solubility and they are more resistant towards metabolic degradation. Triazole derivatives are also proposed as the alternants of capping groups in HDACi's (Pirali, Pagliai *et al.* 2008). In fact, triazole group

restricts the bond rotation and enables the war-head to sit in an appropriate position inside the catalytic pocket which increases the inhibition potency of the drug (Singh, Nazarova *et al.* 2010).

In this series, two triazole derivatives 201 and KH197, were used to study the inhibition mechanism of smHDAC8. The crystallization experiments were successful and all the data statistics are represented in the Table 15. Both inhibitors are bound in the active site pocket by coordinating catalytic zinc and catalytic residues as in the case of previous inhibitors.

In smHDAC8/201 crystal structure, F151 was observed to have adopted both flipping-in and flipping-out conformations. In addition, the inhibitor was bound in a straight orientation in the active site pocket, and it could not reach the surface of the active site pocket because of its small linker (Figure 72-e). This structural inhibition mechanism was further used for the optimization.

The second compound KH197 is a long linker molecule compared to compound-201. KH197 contains three aromatic ring systems in the structure where a triazole group is attached to hydroxamate war-head and to a diphenyl ether group which is attached with chloride and fluoride moieties. Crystal structure of smHDAC8-KH197 complex has revealed an interesting binding mode, where the inhibitor adopted an L-shape in the active site pocket and three aromatic ring systems are buried in the active site pocket.

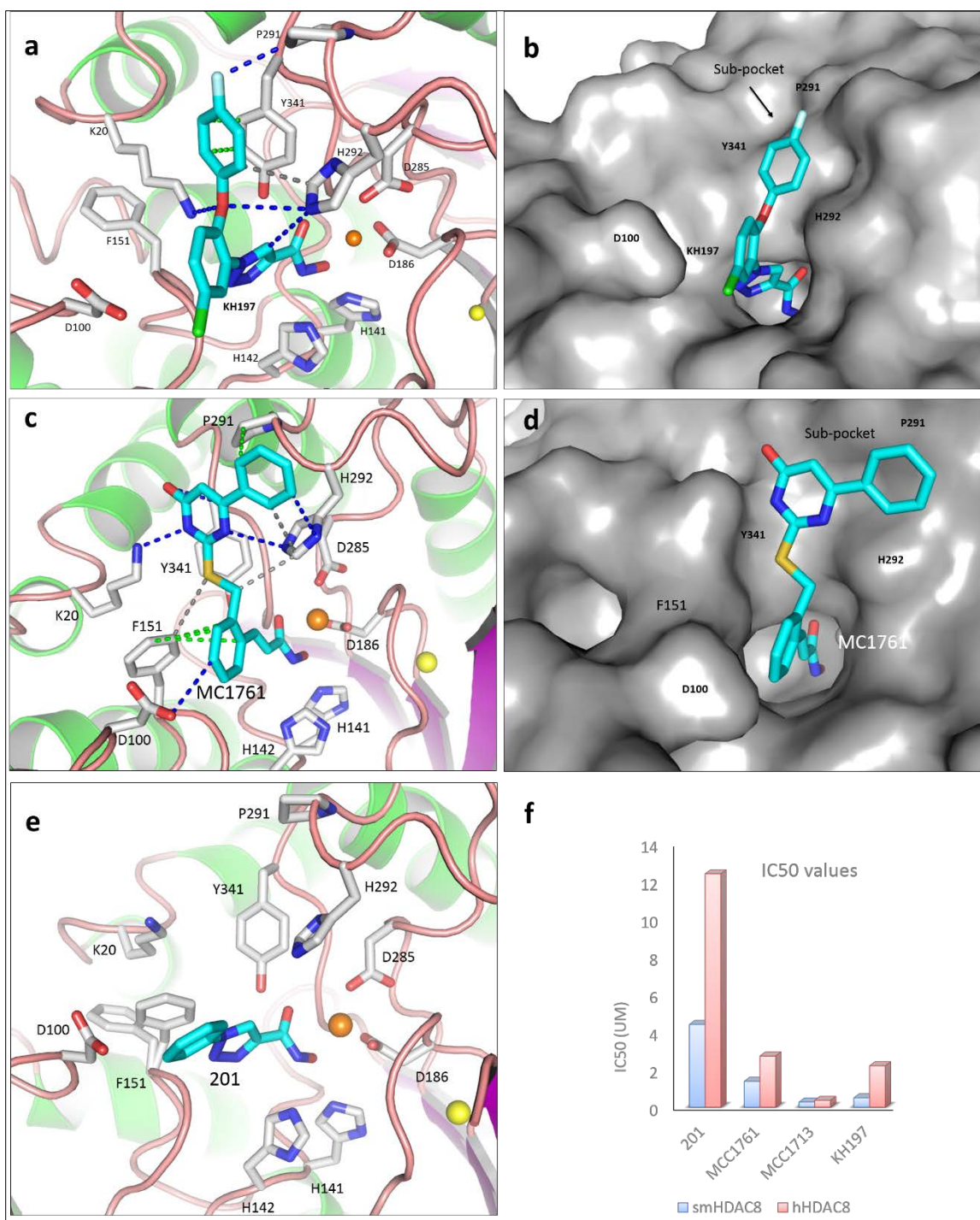
The L-shape of the inhibitor enabled the inhibitor to reach the loop L6 sub pocket which is formed by P291, H292 and Y341. The fluoro-phenyl capping group of the inhibitor is nicely stacked over the Y341 and makes  $\pi$ - $\pi$  interactions (Figure 72-b). The loop L6 sub pocket is specific to smHDAC8 and it is detailed in the article2 (3.2.3). The interactions of this capping phenyl group with P291, H292 and Y341 sub pocket are highly favourable to gain schistosome specificity. Further, the presence of an oxygen atom in the biphenyl ether makes a kink in between the two phenyl groups and the oxygen atom could participate in the “schistosome clamp-jaw” kind of inhibition, which was mentioned in the article1 (3.2.2). K20 and H292 forms hydrogen bonds with the ether group which is holding the inhibitor in the active site pocket (Figure 72-a).

The linker phenyl group (or the second aromatic system) of the inhibitor was situated perpendicular to the D100. And this phenyl is attached to a chloride group which is closely situated near D100, and this phenyl group is pointing outside of the active site pocket towards the loop L2.



All pan-HDAC inhibitors are observed to target D100, suggests that this phenyl group adopts a pan-HDAC like inhibition.

The triazole group forms the base of three aromatic ring systems and it has formed hydrogen bonds with H292. Three aromatic ring systems are placed in three different planes in the active site pocket viz., the triazole ring is buried in the active site pocket the second phenyl group occupies the place of active site channel connecting towards loop L2 and the third phenyl group is placed over schistosome specific subpocket formed by amino acids of loop L6 and Y341. Further, two halogen atoms, F<sup>-</sup> and Cl<sup>-</sup> contribute additional contacts where the fluoride is in close proximity to D100 (4.4 Å) which is known to bind the main chain of incoming acetylated peptides. The chloride atom is in close proximity to P291 (3.3 Å). These residues belong to two different active site loops, L2 for the former and L6 for the latter. Interestingly, the first loop is generally interacting with pan-HDAC inhibitors, whereas the second is part of the HDAC8-selective pocket. The bidentate nature, with two capping groups, of KH197 enables it to bind with selective and non-selective features to HDAC8 enzymes. These specific features are also reflected in the *in vitro* assays where KH197 has IC<sub>50</sub> values 0.5 μM and 2.2 μM for smHDAC8 and hHDAC8 respectively (Table 14: IC<sub>50</sub> values of inhibitors).



**Figure 72: Crystal structures of smHDAC8 with triazole derivatives and Uracil based compound**  
 Crystal structure of smHDAC8 with KH197 (a, b), MCC1761 (c, d), 201 (e) and comparison of IC<sub>50</sub> values (f). Catalytic residues that interacting with inhibitors are labelled (a, c and e) and the interactions between catalytic residues and inhibitor are represented as dashed lines (H-bonds-blue;  $\pi$ - $\pi$  interaction-green; hydrophobic interaction-grey). Zinc and potassium atoms are represented as orange and yellow spheres respectively. The capping group of inhibitors KH197 and MCC1761 are placed over a subpocket that is formed by P291, H292 and Y341 (b, d).

### 3.2.5.1. Crystal structures of smHDAC8 with Uracil based compounds

Another set of compounds which are uracil based inhibitors MC1713 and MC1761, were also used for the crystallization studies. The compound MC1713 was unable to yield diffracting grade crystals because of its insoluble nature in the aqueous buffer while the crystal structure of smHDAC8 in complex with MC1761 shows an L-shaped inhibitor which is bound in the HDAC8-selective pocket (Figure 72- c &d). smHDAC8-MC1761 structure has few similarities and few differences with smHDAC8-KH197 structure. MC1761 was bound in the active site pocket while the hydroxamate war-head coordinate with catalytic zinc and other catalytic residues. The uracil based capping group was nicely fixed over the schistosome specific subpocket which is similar to KH197 (Figure 72-d). Interestingly, in contrast to KH197, F151 is observed in a flipped-in conformation where it forms Van der Waals interactions with the phenyl group of the inhibitor. Another interesting aspect is that K20 has stacked over the F151 and reaching towards pyrimidine ring of the inhibitor to make H-bonding.

Apart from TH and TB series, the inhibitors of miscellaneous groups are interesting because they support the major conclusions drawn from the analysis made by the TH and TB inhibitors. And an additional set of information was provided by these inhibitors should be used to optimize the drug design in order to design schistosome specific inhibitors. Compared to the TH inhibitors, their potency towards smHDAC8 and hHDAC8 is lower which could be because of solubility or thermodynamic properties of the inhibitor. However, these structures are complementary as those described in the articles published and would help further to design more potent and more selective compounds for smHDAC8.

**Table 15: Crystallographic table II**

Data collection	smHDAC8/201	smHDAC8/ KH197	smHDAC8/MCC1 761	smHDAC8-TB5	smHDAC8-TB8	smHDAC8-TB98
Space group	P1	P1	P1	P1	P1	P1
Cell dimensions						
a, b, c (Å)	70.49, 70.7, 97.98	70., 70.7, 98.3	71.49, 71.53, 98.78	70.65, 70.73, 98.29	70.85, 70.9, 98.3	70.6, 70.7, 98.2
$\alpha, \beta, \gamma$ (°)	78.08, 75.79, 85.82	75.9, 78.3, 85.6	75.97, 78.56, 85.71	75.9, 78.32, 85.59	78.1, 75.4, 85.5	75.9, 78.3, 85.6
Resolution (Å)	50 (1.71– 1.61)	50 (1.76– 1.65)	50 (1.89– 1.78)	50 (2.27– 2.14)	50 (2.21– 2.08)	50 (2.20– 2.07)
Rsym or Rmerge	6.4 (108.9)	4.3 (78.5)	5.3 (36.2)	10.2 (64.7)	11.7 (65.8)	16 (65.6)
I / $\sigma$ I	11.38 (1.05)	14.93 (1.43)	8.58 (1.61)	4.79 (1.05)	4.38 (1.05)	6.03 (1.56)
Completeness (%)	92.2 (85.5)	92.1 (87.3)	91.1 (89.1)	93.4 (90)	94.0 (88.5)	94.3 (89.9)
Redundancy	3.6 (3.5)	3.6 (3.6)	1.74 (1.7)	1.76 (1.74)	1.4 (1.4)	3.5 (3.4)
CC(1/2)	99.9 (51.9)	99.9 (58.7)	99.6 (73.7)	98.8 (46.2)	98.3 (37.7)	98.4 (71.6)
<b>Refinement</b>						
Resolution (Å)	48-1.62	49-1.66	49-1.78	48-2.10	43-2.10	43-2.10
No. reflections	213108	196617	160890	102410	100814	100814
Rwork / Rfree	0.166 / 0.195	0.122 / 0.194	0.189 / 0.225	0.189 / 0.240	0.175 / 0.226	0.175 / 0.226
No. atoms						
Protein	13170	13168	13043	12968	13047	13047
Ligand/ion	228	243	254	170	111	111
Water	960	850	885	432	570	570
B-factors						
Protein	31.27	34.54	26.48	39.59	35.11	35.11
Ligand/ion	51.78	55.27	36.78	58.53	53.98	53.98
Water	40.38	42.88	34.96	43.88	39.01	39.01
R.m.s deviations						
Bond lengths (Å)	0.007	0.008	0.008	0.008	0.007	0.007
Bond angles (°)	0.855	0.898	0.909	1.004	0.998	0.998
*Values in parentheses are for highest-resolution shell.						
#The number of protein atoms varies slightly for each structure, depending on the quality of the electron density for poorly folded regions.						

### 3.3. Cohesin complex purification

My work on the selective inhibition of HDAC8, notably on the role of the loops at the rim of the active site, has shown that the active site loops play an important role not only on inhibition, but also on activity. A similar conclusion was drawn recently from the structural analysis of HDAC6, where changes in the active sites of this enzyme influence substrate recognition and, in fact, provide different substrates specificities for the two catalytic domains (Miyake, Keusch et al. 2016). This is also the case for HDAC1 and HDAC3 that need inositol phosphate molecules for activity, these molecules being in direct contact with active site loops (Watson, Fairall et al. 2012).

Yet, the mechanisms by which the active site loops integrate signals coming from substrates and small molecules and transform them in functional output still remains poorly understood. One major reason for this is that structures of HDAC/substrate complexes are restricted so far to small peptide substrates that do not integrate the complexity of full substrates binding to HDACs. It has been suggested that HDACs are poorly specific in terms of substrates, but this view is contradicted by the fact that HDACs have specific substrates that need to be precisely recognized. If this can be due by partner subunits for HDAC1-3, the question remains open for HDAC6 and HDAC8. Specifically, in the case of HDAC8, since the enzyme is supposed to act as a monomer, specific recognition of substrates is supposed to be carried out by the enzyme itself, and it is tempting to think that the specific loops that surround the active site as well as the HDAC8-selective pocket we have characterized participate (among others) to this recognition.

In yeast and human, several lines of evidence demonstrate that the Cohesin complex is a major target for HDAC8 (and its yeast counterpart Hos1). In somatic cells, the components of cohesin complex are SMC1 (structural maintenance of chromosome protein 1), SMC3, RAD21 and stromal antigen SA1 or SA2 (Figure 29). Among these subunits SMC3 is the substrate of HDAC8 that deacetylates two consecutive residues in the head domain of this protein. Our A-ParaDDisE collaborators have however obtained data that would imply direct interactions between HDAC8 and SMC1 and RAD21. A description of cohesin complex architecture was given in the section (1.3.8.2). So far, the knowledge of cohesin complex is limited to domain structures and a clear depiction of overall architecture, and structural knowledge of the interaction between cohesin complex and HDAC8 is missing. In order to understand these aspects, I have initiated the

biochemical and structural investigation of these different complexes by recombinant production and crystallization.

### 3.3.1. Expression of cohesin complex subunits

Different constructs were prepared by cloning N- and C-terminal domains of SMC1, SMC3, and RAD21 individually into separate expression vectors with different antibiotic resistance markers. For SMC proteins only head domains (SMC-HD) were cloned based upon a sequence alignment which was prepared with SMC1 and SMC3 sequences against available ATPase proteins that belongs to ABC (ATP binding cassette) family in the protein data bank (Figure 74). For hsRAD21 (*Homo sapiens* RAD21) N- and C-terminal domains (NTD and CTD respectively) constructs were designed based upon sequence alignment with yeast Scc1 (homologue of hsRAD21). Since the similarity between the two proteins is low, constructs of different lengths were made for RAD21 (Figure 73).



**Figure 73: Strategy to SMC complex production:**

hsRAD21 N-terminal region cloned in three different constructs, NTD1, NTD2 and NTD3. And C-terminal region cloned into CTD1 and CTD2 regions. SMC3HD and SMC1HD domains interact with N and C-terminal regions respectively which are showed by arrows.

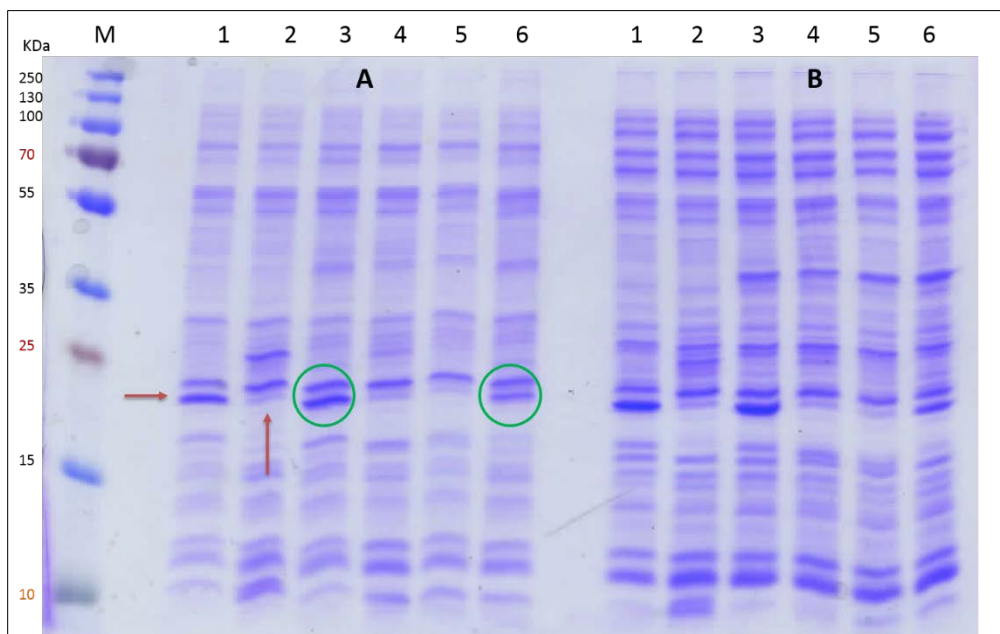


**Figure 74: Sequence alignment of hSMC1A with ABC ATPases:**

N and C-terminal regions separated by a coiled coil domain which is removed for the clarity and the same is labelled with black arrows. Thrombin cleavage site of linker is mentioned. And several conserved motifs which are important for ATP binding are highlighted.



Initially, SMC1 N- and C-terminal domains were co-expressed as separate entities in two different vectors. And additionally, RAD21 (NTD and CTD) was used as a third protein for co-expression experiments. However, co-expression of SMC-HD from with two individually cloned vectors (for NTD and CTD) did not yield good results. The C-terminal domain of SMC1 was not well expressed or found in inclusion bodies (Figure 75). The C-terminal domain was highly unstable and it should be joined with N-terminal domain to form a stable head domain. Co-expression with RAD21-CTD could not rescue this problem. Further, different fusion tags were also used to co-express NTD and CTD of SMC1, the expression levels were increased comparatively, but the solubility and stability during purification was not good and no protein was able to purify in this method.



**Figure 75: Expression of SMC1-HD**

1: his-SMC1-NTD, 2: his-SMC1-CTD, 3: his-SMC1-NTD with SMC1-CTD, 4: SMC1-NTD with his-SMC1-CTD, 5: SMC1-NTD + his-SMC1-CTD + hsRAD21-CTD, 6: his-SMC1-NTD + SMC1-CTD + hsRAD21-CTD; A: 50 mM Tris pH 8.0, 50 mM NaCl, B: 50 mM Tris pH 8.0, 150 mM NaCl. Red arrows indicate SMC1-NTD and CTD proteins green arrows indicate complex of SMC1 NTD+CTD.

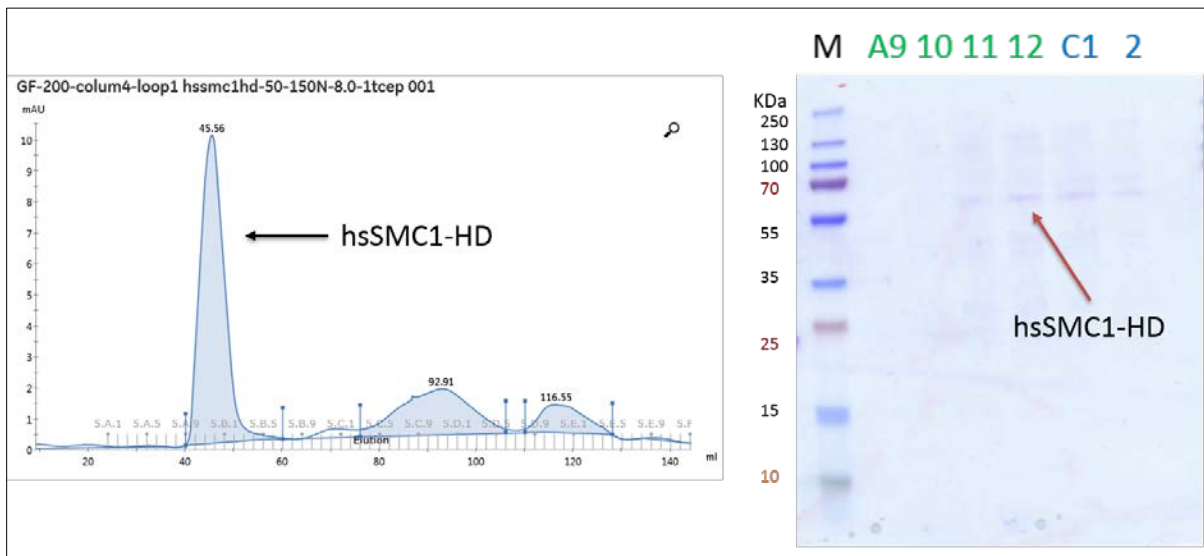
In previous reports, head domains were expressed as soluble entities by joining the N- and C-terminal domains with a thrombin cleavable linker (Haering, Schoffnegger et al. 2004). Since co-expression of the independent N- and C-terminal domains was not successful with the human proteins, I have followed the same strategy and linked these two domains by a small linker,



resulting in a linked fusion protein (SMC1-HD). The protocol for cloning was mentioned in the section (2.1.5).

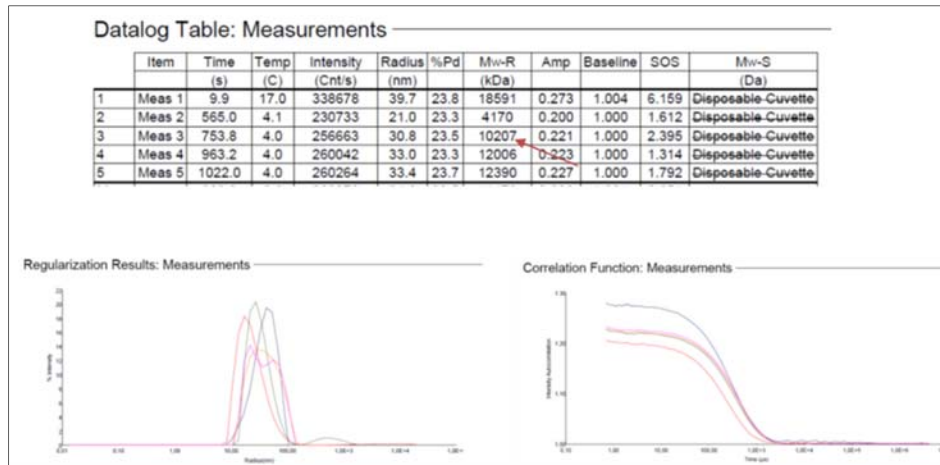
### 3.3.2. Purification of hsSMC1-HD

Expression of SMC1-HD as soluble entity was successful. However, yields were low and, upon gel filtration chromatography, the protein eluted with a higher molecular weight than expected (Figure 76). I reasoned that the head domain could be unstable in the buffer used or lacking an interacting partner. Analysis by DLS (Figure 77) clearly indicated that the protein is probably forming soluble aggregates.



**Figure 76: Purification profile of hsSMC1HD:**

Gel filtration Chromatogram shows the purified protein in very low concentration (10 mAU) as indicated by the black arrow which was analysed on SDS-PAGE.

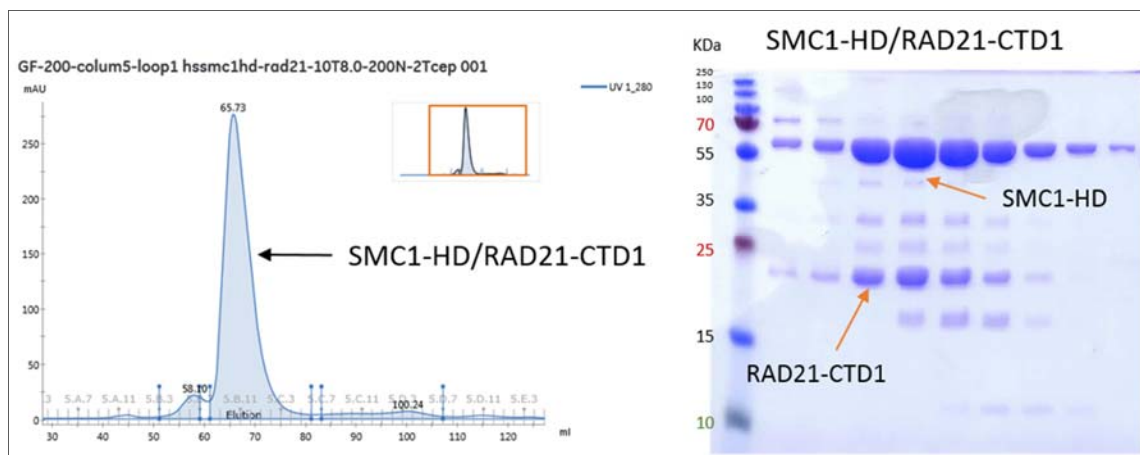


**Figure 77: Characterization of hsSMC1-HD using DLS:**

DLS profile (log table top, regularization and correlation functions plots below) is showing that the purified hsSMC1-HD is a soluble aggregate. The molecular weight is higher than the theoretical molecular weight which is indicated by red arrow.

### 3.3.3. Purification and initial crystallization attempts of the hsSMC1-HD /RAD21-CTD1 complex

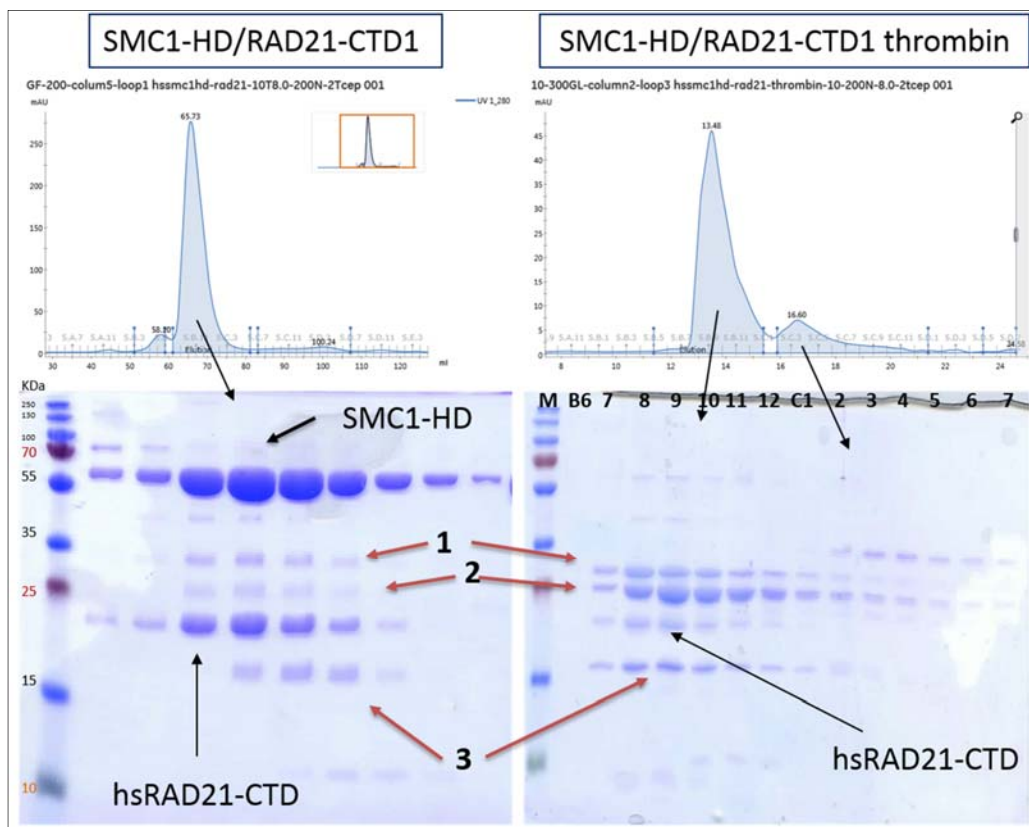
Since SMC1-HD was unstable, I co-expressed it with its natural partner RAD21 (construct RAD21-CTD1 see Figure 73). As expected, RAD21-CTD1 was able to stabilize the head domain of SMC1. The complex eluted at the expected molecular weight in gel filtration chromatography and it behaved very nicely in DLS analysis (Figure 78 & Figure 85). However, few degradative products were observed in the SDS-PAGE analysis.



**Figure 78: Purification profile of SMC1-HD/RAD21-CTD1.**

Gel filtration chromatogram of SMC1-HD/RAD21-CTD1 complex. The complex was represented with black arrow which was analysed using SDS-PAGE.

Further, the degradation as observed in the (Figure 78), was analysed using different strategies which could help in defining better boundaries to clone new constructs. Since the SMC1-HD has a thrombin cleavage site, I have performed a thrombin site cleavage analysis. According to the thrombin cleavage analysis and mass spectrometry analysis the two degrading products below SMC1-HD (represented as 1 and 2 in red arrows in Figure 79) are actually two the cleaved products of SMC1-HD i.e. individual domains SMC1-NTD and SMC1-CTD. In the gel filtration chromatography which was done after thrombin cleavage the complex has eluted at same volume as the fresh complex, indicating that the cleaved products are still part of the complex. However, the product below RAD21-CTD1 was confirmed as a degradative product of RAD21-CTD1 by mass spectrometry analysis.



**Figure 79: Thrombin cleavage of SMC1-HD/RAD21-CTD1 complex**

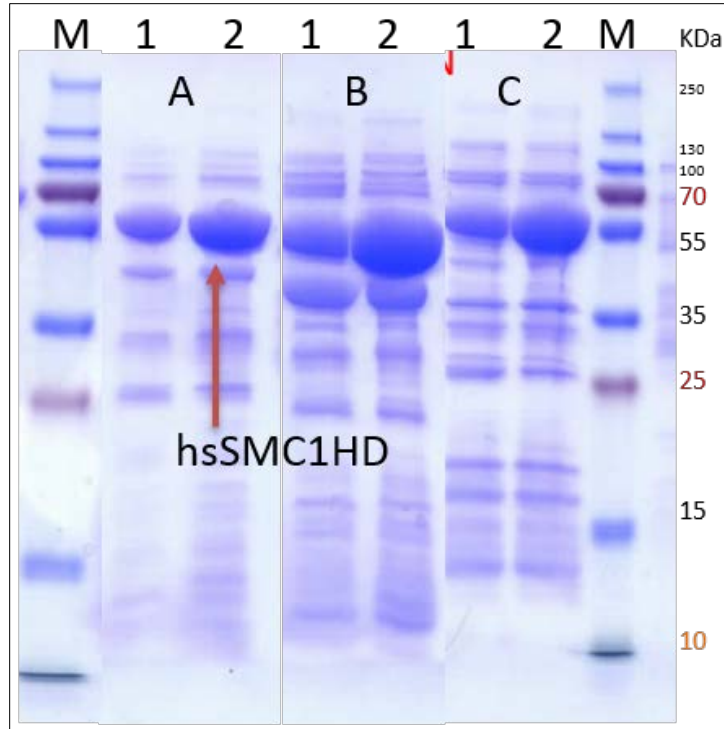
SMC1-HD/RAD21-CTD1 complex was analysed with thrombin cleavage, untreated and thrombin treated purifications are represented left and right respectively. The peak fraction (black arrows) were analysed using SDS-PAGE. The components of complex, SMC1-HD and hsRAD21-CTD are represented with black arrows. The degradation products 1, 2 and 3 are represented with red arrows.

Since the complex appeared sufficiently stable except for the linker in SMC1-HD, I performed crystallization trials with this complex. Unfortunately, these assays were unsuccessful. One of the problem encountered from the beginning was that the poor sequence similarity between RAD21 and its homologue Scc1 (yeast protein) made difficult to choose for RAD21 construct boundaries based on the structure of the homologue region of Scc1 (pdb: 1W1W). The mass spectrometry results, bioinformatics analysis, and model building suggested that the initial RAD21 clone was too long and that a shorter construct should be made.

#### 3.3.4. Optimization of the hsSMC1HD-RAD21 complex production

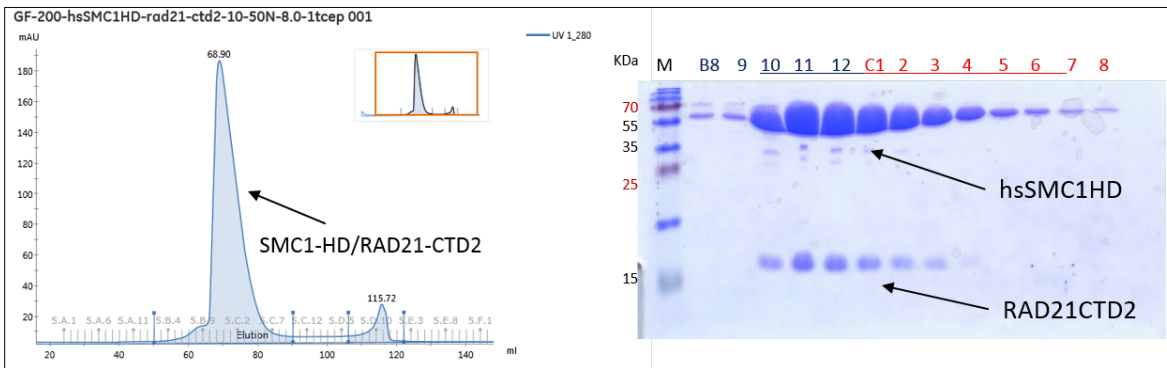
Therefore, a new, shorter construct was prepared for RAD21 C-terminal region (termed CTD2), and was co-expressed with hsSMC1-HD. In co-expression tests, the SMC1-HD/RAD21-CTD2 complex appeared more soluble than the SMC1-HD/RAD21-CTD1 complex (Figure 80). Further, this new construct was used for the optimization studies and subsequent purification.

The complex between hsSMC1-HD and RAD21-CTD2 was expressed and purified in large scale as described above. In the gel filtration chromatography, the complex eluted at the expected molecular weight (Figure 81). A single peak was observed with the complex as confirmed by SDS-PAGE. Surprisingly, this complex more stable and all the degradation products were not repeated in this construct.



**Figure 80: Expression profiles of hsSMC1HD and RAD21CTD optimization:**

1: hsSMC1-HD/RAD21-CTD1, 2: hsSMC1-HD/RAD21-CTD2. A, B and C are three different salt concentrations used to extract the protein from bacteria. Buffer: 50mM Tris pH 8.0 and 50, 150 and 500 mM NaCl respectively in A, B and C. In all cases SMC1HD (51.3 KDa) is expressed without tag and C-terminal his tags were used for RAD21 proteins and pulled down using Co<sup>2+</sup> affinity beads. CTD2 has enhanced the production of SMC1HD. SDS-PAGE gel was edited to match marker.

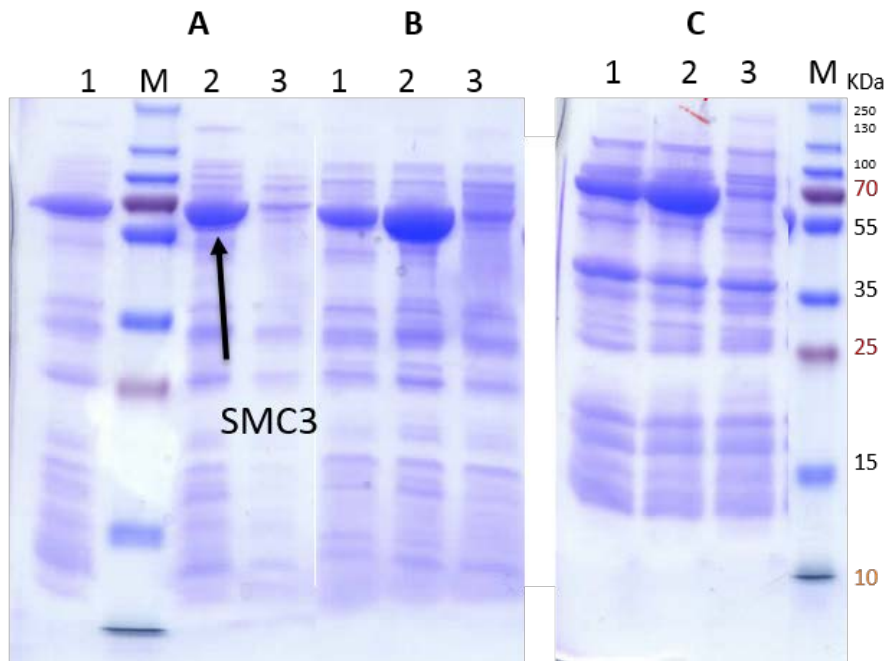


**Figure 81: Purification profile of SMC-HD/RAD21-CTD2:**

Gel filtration chromatogram showing a single peak and the presence of complex was showed on SDS-PAGE. The purified complex is pure and free of contaminants and degradative products.

### 3.3.5. Expression of the hsSMC3HD/N-terminal region of RAD21 complex

Similar efforts were made to prepare another sub-complex of the cohesin complex, namely the complex of SMC3 head domain (SMC3-HD) and RAD21 N-terminal region. The head domain of SMC3 was prepared same as in case of SMC1-HD, with a linker of thrombin cleavage site. Three different constructs were prepared for RAD21 (NTD1-3) (Figure 73) and I mapped the interaction between these different proteins in which RAD21-NTD2 was able to form the most stable complex of the three RAD21 constructs (Figure 82). This complex was further used for large scale production and purification.

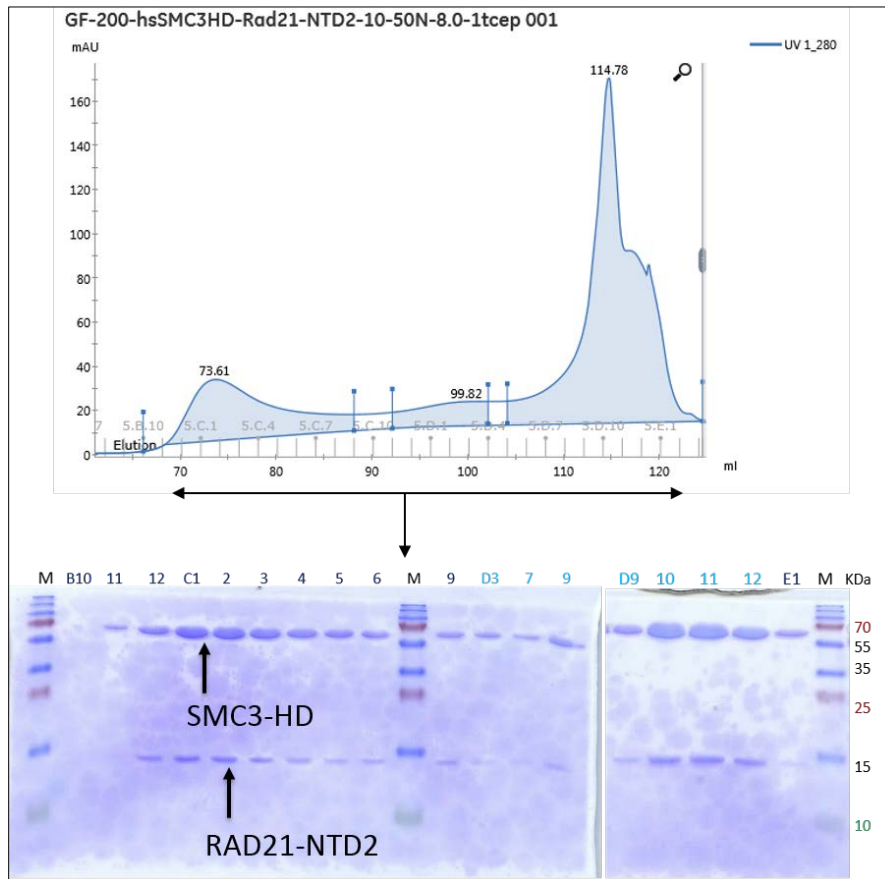


**Figure 82: SMC3-RAD21NTD co-expression:**

A, B and C are three different salt concentrations used to extract the protein from bacteria. Buffer: 50mM Tris pH 8.0 and 50, 150 and 500 mM NaCl respectively in A, B and C. 1: RAD21-NTD1 (24KDa), 2: RAD21-NTD2 (10KDa) and 3: RAD21-NTD3 (8KDa). (SMC3-HD 60.6KDa). In all cases SMC3-HD (60.6KDa) is expressed without tag and C-terminal his tags were used for RAD21 proteins, and pulled down using  $\text{Co}^{+2}$  affinity beads. Among three constructs of RAD21-NTD2 is best since it enhanced the production of SMC3-HD. RAD21 was not visible in the gel, though it was confirmed in the following experiments.

### 3.3.6. Purification of hsSMC3-HD/RAD21-NTD2

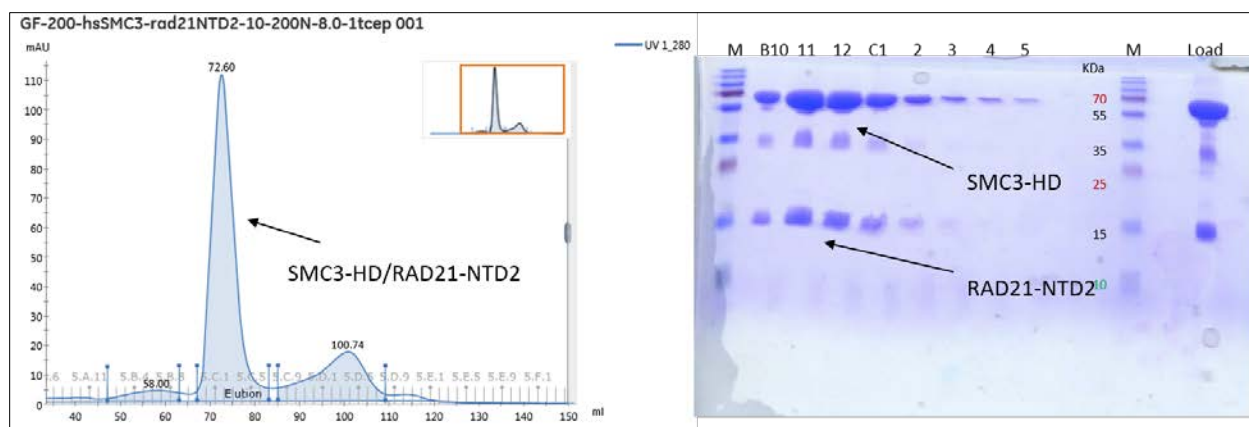
The complex between SMC3-HD and RAD21-NTD2 was expressed and purified as mentioned in the sections (2.2.2 & 2.3.4). During the gel filtration chromatography, the complex eluted however in many different fractions, indicative of a poor biochemical behaviour (Figure 83). The buffer used for gel filtration purification was 10mM Tris pH 8.0, 50mM NaCl, 2mM TCEP. I reasoned that the salt concentration used might have been too low since the complex appeared more soluble at higher salt concentration in the initial minitests. In presence of high salt concentration (200mM NaCl), the complex eluted at the expected molecular weight and behaved stably (Figure 84).



**Figure 83: Purification profile of hsSMC3-HD/RAD21-NTD2:**

Chromatogram showing the complex of SMC3-HD/RAD21-NTD2 has dispersed in the gel filtration from volume 70ml to till the very end of the column volume. Indicating unfavourable conditions for the complex.



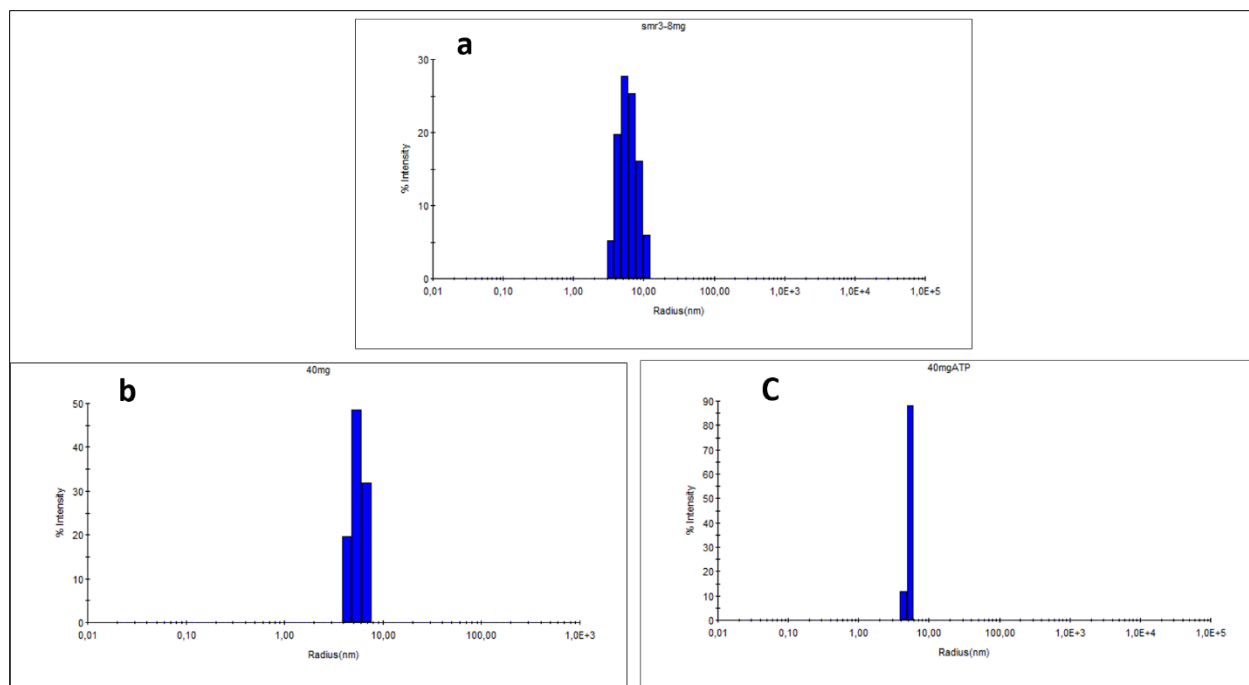


**Figure 84: Optimization of SMC3HD-RAD21NTD2 complex purification:**  
 Chromatogram showing a single peak which was analysed on SDS-PAGE. Upon change in salt concentration the stability of the complex is improved.

### 3.3.7. Characterization of the purified complexes using DLS

Two SMC1-HD/RAD21-CTD2 and SMC3-HD/RAD21-NTD2 complexes have been optimized and purified in stable form. Before pursuing with structural studies, I have assessed the quality of the samples by dynamic light scattering (DLS). In this analysis, both complexes were showed to be monodisperse (Figure 85). Specifically, both complexes were used at different concentrations ranging from 7 to 40 mg/ml and were showed to be monodisperse and to display the expected molecular weight whatever the concentration used.





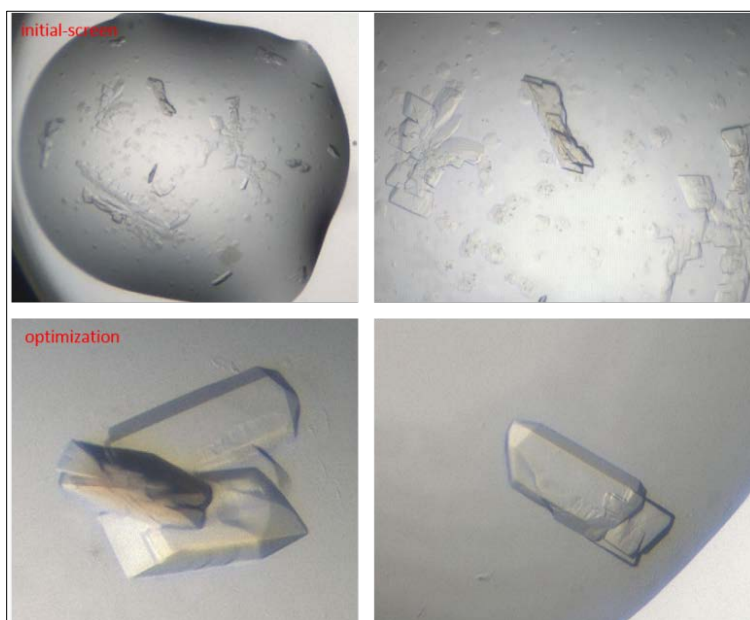
**Figure 85: DLS profiles of SMC-RAD21 complexes:**  
a: hsSMC3-HD/RAD21-NTD2. b and c: hsSMC1-HD/RAD21-CTD2. In c) using ATP $\gamma$ S has improved the sample quality. All Samples are at their respective molecular sizes. For SMC3-HD/RAD21-NTD2 upon ATP $\gamma$ S addition no change was observed.

### 3.3.8. Crystallization attempts of the SMC-HD/RAD21 complexes

I next started the crystallization attempts with both complexes which were used at different concentrations. Since SMC-HD complexes are ATP binding domains, the crystallization attempts were made in absence and in presence of the ATP analogue ATP $\gamma$ S. Several sparse matrix crystallization screens that are available on the Structural Genomics platform of IGBMC were used for these initial attempts. Crystallization plates were setup using a mosquito robot, using different drop volumes (200-400 nl) and stored at 20°C.

Initial hits were obtained only for the SMC1-HD/RAD21-CTD2 complex without ATP $\gamma$ S in PEG conditions. From the initial condition, a novel screen was generated with a finer gradient of pH and precipitant. New crystallization plates were setup with fresh protein using different volumes (200nl to 400nl) with different protein concentrations 2.5 – 30 mg/ml, and also with and without ATP $\gamma$ S. Crystals appeared after 8 days in almost all conditions without ATP $\gamma$ S, with

different shape and size (from 50 to 300  $\mu\text{m}$ ) (Figure 86). Good quality crystals were harvested in different cryo options and flash frozen in liquid nitrogen.



**Figure 86: Crystals of hsSMC1HD-RAD21CTD2 complex.**

Top panel: crystals obtained in the initial sparse matrix screen. Bottom panel: crystals obtained from optimized crystallization screen.

### 3.3.9. Data collection analysis

Data was collected for the hsSMC1-HD/RAD21-CTD2 crystals on IGBMC X-ray home source (Rigaku FR-X microfocuss rotating anode generator, detector: EIGER 4M from Dectris, AFC Partial-k goniometer, VariMax-HF Arc optic). Initial indexing was done after collection of two frames, and suggested a centred monoclinic space group (C2). Data was then collected using  $0.25^\circ$  rotations for  $240^\circ$  in total. The detector distance was set to 90mm.

Data was processed using XDS software, and showed that data was of good quality up to 2.56  $\text{\AA}$  resolution (Table 16). Structure determination was done by molecular replacement with phaser using the yeast Smc1HD-Scc1 structure (PDB code 1w1w) as model. The molecular replacement has been rendered difficult due to the poor sequence similarity between the human and yeast proteins. We therefore generated a model of the human complex by replacing the residues and removing regions that are poorly conserved. This model provided a solution by

molecular replacement that enabled me to locate secondary structure elements and start model building. The current model is however not complete although the R-factor and R-free values are in the low 40%. This model suggests a different mode of interaction between the N- and C-terminal regions of SMC1 but this information will require further model building and refinement steps to be fully confirmed.

**Table 16: Data statistics of hsSMC1HD-RAD21CTD2**

<b>Data collection</b>	<b>SMC1HD-RAD21CTD2</b>
Space group	C2
Cell dimensions	
<i>a, b, c</i> (Å)	188.85, 64.53, 47.79
$\alpha, \beta, \gamma$ (°)	90, 102.66, 90
Resolution (Å)	50 – 2.56
$R_{\text{merge}}$	5.3 (85.4)
$I/\sigma I$	11.9 (1.11)
Completeness (%)	94.4 (94.1)
Redundancy	2.4 (2.3)
CC(1/2)	99.9 (59.3)
Values in parentheses are for highest-resolution shell	

### 3.3.10. Conclusions and future perspectives on the HDAC8/Cohesin project

Understanding the interaction between HDAC8 and cohesin first requires the production of the cohesin subcomplexes of interest, namely those suggested to bind directly to HDAC8: those formed by SMC1, SMC3 and RAD21. SMC subunits of the cohesin complex were cloned and expressed in different combinations to obtain soluble, well-behaved sub-complexes. This could however only be achieved by co-expression with their natural partner RAD21. Interestingly, the SMC1HD-RAD21 complex is very stable in different conditions compared to the SMC3HD-RAD21 complex. This observation is somewhat reflected in the crystallization experiments, although many more crystallization attempts are required in the future.

This work has already led to the almost complete structure determination of a human SMC1HD/RAD21 complex that has escaped structural analysis by many different groups so far. Our current model reveals interesting aspects that should help better characterize the human cohesin complex structure and function. It will be interesting to solve the structure of this complex in presence of an ATP analogue to see whether structural changes are occurring upon ATP binding. Determination of the structure of the human SMC3/RAD21 sub-complex should most likely bring important information on human cohesin.

Clearly, another major interest is to characterize the interactions between HDAC8 and these cohesin subcomplexes. Preliminary and very limited attempts have been unsuccessful and this work needs to be further continued and extended. Specifically, mimics of acetylation states need to be introduced and biophysical methods for looking at interactions needs to be used beside the biochemical ones used so far. Another aspect that I have already started but that I did not describe in this manuscript is the use of the yeast system for addressing the same scientific question. Since we apparently see differences between the yeast and human systems, the question which is raised is whether these two systems have evolved differently, including the interaction between HDAC8 and the cohesin. Working on both the yeast and human system should bring a wealth of information in the future.

# Discussion

---

## 4. Discussion

Since the past decade, epigenetic enzymes have been extensively targeted to treat human cancers. Most of the FDA approved anti-cancer drugs target HDACs, but none of them are isoform selective. Despite of the fact that HDACs possess a highly conserved structural fold, the substrates and cellular functions are diverse among HDAC isoforms. Moreover, isoform selective targeting is of paramount importance to deliver the inhibition to a specific cellular function but not to the whole HDAC family. The strategy of anti-epigenetic drug targeting to treat cancer, has implications in several dimensions, such as the development of anti-parasitic therapeutics. The piggy-back strategy is to develop anti-cancer drug candidates into anti-parasitic drugs to reduce the cost effectiveness and time management which are serious concerns of pharma industries. Eukaryotic parasites which cause neglected tropical diseases, possess complex life cycles with several morphological stages and different hosts in their life cycle, where epigenetic enzymes are expected to be the driving force. Moreover, the similarities between cancer cells and parasitic cells such as uncontrolled cell division, high metabolic rate, depending on lactate fermentation for energy source and a degree of invisibility towards host immune system, indicates to use anti-cancer drugs as anti-parasitic drugs. However, a major concern of cross reactivity and off target effects are raised due to high structural conservation of epigenetic enzymes among homologues. In order to overcome this bottleneck challenge structural studies are the much-needed tools.

The above-mentioned piggy-back strategy was tested and validated during the project SETReND by studying schistosomiasis as a case study. Schistosomiasis is a neglected disease caused by the infection of *Schistosoma* genus flatworms which causes more than 200,000 deaths per year. Schistosomes completes their life cycle in snails as a secondary host with multiple morphological stages. Initial studies by our collaborators have revealed that HDAC8 importance in schistosome homeostasis. In fact, the abundant expression of smHDAC8 during all the stages of life cycle, making it an ideal drug target. Further, RNAi mediated smHDAC8 knockdown experiments have shown the decreased ovulation and pairing capacity of schistosoma, in infected mice. And inhibition of smHDAC8 has resulted in the apoptosis mediated cell death. These experiments serve as proof of concept where epigenetic enzymes in parasites can be potential drug targets.

A close inspection at the active site of smHDAC8 in apo and inhibited states revealed striking differences from that of hHDAC8 structure. The major amino acid substitution (M274 of hHDAC8 to H292 of smHDAC8) at active site pocket in smHDAC8, and the schistosome specific conformation of F151 (the flipping-out conformation) of smHDAC8, draws a line between human and schistosome HDAC8s, which provides significance in the inhibitor design. Further, these differences not only widen the active site pocket but also provides plasticity to the parasite enzyme, hence can accommodate bulkier linker inhibitors. Subsequent drug design and structural studies revealed schistosome specific inhibitors with a lesser affinity towards humanHDAC8 but not to other human HDACs. This was the initial work done prior to my arrival on which my thesis work has expanded towards specific and more fundamental aspects of HDAC8 inhibition in human enzymes as well as parasitic enzymes.

\*\*\*

The above mentioned SEtTReND project outcome led me to ask fundamental questions about HDAC8 selective inhibition mechanism. First, on an applicative purpose how to expand the piggy back strategy towards epigenetic targets of other parasitic enzymes. In second, how to improve the specificity towards smHDAC8 since the inhibitors presented in the SEtTReND, still retain affinity towards hHDAC8. And a final question raised during this thesis was, what is the molecular mechanism of HDAC8 and complex-substrate recognition and how does the deacetylation effects in a complex environment.

As a part of A-ParaDDisE consortium, several epigenetic enzymes were identified as drug targets in different eukaryotic parasites. TcHAT1 and LbDAC3 are two examples that were discussed in this thesis, which explains the complexity of epigenetic enzymes of eukaryotic parasites. Most of the epigenetic parasite enzymes possess insertions in their sequences which creates difficulty in the recombinant production. However, the evolutionary aspects behind these insertions in the parasitic enzymes is not well understood. One possible explanation could be these insertions are used in the interactions with other partner proteins. In order to produce parasitic enzymes, the knowledge of these partner interactions is also much needed information which is lacking in most of the cases. As an alternate protein engineering can serve as a powerful tool that help to increase the solubility and stability of the protein. In this thesis, I have modified the proteins TcHAT1 and LbDAC3 and the approach has stabilized the proteins to some extent.

TcHAT1 an acetyl transferase from *Trypanosoma cruzi* that causes trypanosomiasis or sleeping sickness. The production of TcHAT1 was not successful at the beginning and the sequence alignment showed an insertion and an N-terminal extension in the TcHAT1 sequence. Deletion of the insertion has resulted in the destabilization of TcHAT1 which suggests that the insertion is important for the *in vivo* protein stability. However, the insertions are major hurdles for *in vitro* protein purification because flexible parts of the protein are highly prone to proteosomal degradation and aggregation which is an unfavorable condition for crystallization studies. The same was observed in case of LbDAC3, a histone deacetylase from *Leishmania braziliensis* which causes leishmaniasis. LbDAC3 is a homologue of class II human HDACs. Sequence alignment with class II human HDACs revealed insertions in the LbDAC3 sequence and upon deletion of these insertions the stability of LbDAC3 was hindered. These observations suggest that it is difficult to produce these enzymes in the absence of partner proteins.

However, the replacement of insertion sequence in TcHAT1 with human sequence result in the improvement of protein production and stability. This subtle change in the sequence showed huge impact on the protein stability which reflects the complexity of parasitic enzymes vs human enzymes. Further, the use of co-factor molecules like acetyl CoA in case of TcHAT1 and also the use of fusion tags such as thioredoxin, have shown the positive effect on protein stability. Additional optimization studies are required to proceed for the crystallization studies. These epigenetic enzymes were used by our collaborators of A-ParaDDisE consortium for *in vitro* activity tests which are under progress. To test TcHAT1 activity our collaborator Prof. Dr. Manfred Jung and his colleagues have developed an assay based upon on H3 and H4 peptide lysine acetylation using a europium-N1-labeled secondary antibody. Also, acetyl-CoA was synthesized which an analogue of acetyl CoA, for the inhibition assays. However, the presence of acetyl CoA in the purified protein was incompatible for this assay hence the protein purification trials in the absence of acetyl CoA are being tested.

The major part of my thesis involved in addressing the second question i.e. to improve the schistosome specific inhibitors. In this process, I have solved smHDAC8 crystal structure with several inhibitors which led us to elucidate selective inhibition of HDAC8 and to define the 'HDAC8-selective pocket' at atomic level. Virtual screening from the initial lead compounds let our collaborators to propose more smHDAC8 selective inhibitors. The many inhibitors used in this



thesis can be represented as different series, TH series – the derivatives of J1038 (initial hit from SEtTReND), TB series (derivatives of J1075), triazole derivatives, uracil based compounds, commercially available inhibitors. All these inhibitors contain hydroxamate war-head that tends to chelate catalytic zinc in the active site pocket. Apart from the hydroxamates I have also used carboxylate compounds which didn't yield a diffraction quality data.

Crystal structure of smHDAC8 with one of the simplest compound of TH series (benz-hydroxamate derivatives), TH31 (benz-hydroxamate with benzamide capping group) has revealed a new inhibition mechanism. Two active site residues K20 and H292 of smHDAC8 holds the inhibitor in the active site pocket and the arrangement resembles a 'clamp-jaw' which is only possible in smHDAC8 but not in human HDACs including hHDAC8. Schistosome specific flipped-out conformation of F151 allows K20 to place its side chain towards the active site channel and that allows to form a hydrogen bond with inhibitor. In human HDACs corresponding phenylalanine always found in flipped-in conformation and hence equivalent lysine side chain moves away from the inhibitor. And also, H292 a polar amino acid that can form another hydrogen bond with the inhibitor which is not possible in hHDAC8 due to the presence of M274 at equivalent position. This clamp-jaw binding mode was further used in the optimization studies and all the TH series compounds showed this kind of inhibition in the crystal structures. By changing the orientation of amide group (TH97) after benz-hydroxamate has changed the hydrogen bond distances with K20 and H292 which has favored the specificity towards smHDAC8. These observations indicate how small changes in the inhibitors can influence the inhibition potency. The presence of benzamide capping group in TH31 and subsequent derivatives have shown a sub pocket which is formed by the amino acids P291 and H292 of loop L6 in smHDAC8. The phenyl capping group of TH-series compounds are directed towards this sub pocket to form schistosome specific interactions. Further, introduction of halogens (TH86 and TH104) in the capping group have forced the phenyl capping groups to be tilted towards loop L6 sub pocket. This has enabled the capping groups to make cation- $\pi$  interactions with H292 and thereby increased the affinity towards smHDAC8. results highlight how small modifications can influence inhibition potency.

Further these observations were also supported by the crystal structures of smHDAC8 with inhibitors from other series such as TB98, MC1761 and so on. MC1761 and KH197 are two unique compounds that are different from TH and TB series inhibitors. Crystal structures of smHDAC8

with these two compounds have shown that the capping group is reaching towards loop L6 and also towards loop L2. The D100 (smHDAC8) is an important amino acid in HDAC8 that participate in the interaction with the incoming peptide. All pan-HDAC inhibitors attain a straight conformation in the active site pocket and hence can reach D100. KH197 and MC1761 can interact with residues from loop L6 and loop L2 suggests that these inhibitors showing features like pan-HDAC inhibitors and schistosome specific inhibitors.

These structural studies have shown the importance of loops in the selective inhibition of HDACs. The catalytic pocket of HDAC8 is formed by the amino acids of loops L6, L1 and the catalytic tyrosine, which we have termed as HDAC8-selective pocket. Amino acids of loops L6, L1 contributes to the side walls of the selective pocket and the catalytic tyrosine is involved in the formation of bottom of the pocket. The HDAC8 selective pocket is not possible in all other HDACs. In class IIa HDACs the catalytic tyrosine has replaced by histidine and the side chain is also turned away from the catalytic zinc. This orientation creates a sub pocket in the class IIa HDACs and thus it can't provide the interaction surface same as HDAC8. All other HDACs contains a similar loop L6 in length but protrudes over catalytic tyrosine and hence it sterically clashes with the HDAC8 selective inhibitors. In class I HDACs (except HDAC8) the presence of arginine in loop L6 which is important for the IP<sub>4</sub> interaction hinders the formation of HDAC8 selective pocket. And the loop L1 is larger in all other HDACs compared to HDAC8. The presence of proline or isoleucine (HDAC10) in the loop L1 will allow its interaction with loop L6 and catalytic tyrosine which is incompatible with the formation of an HDAC8 like pocket.

Compared to all other HDACs, the HDAC8 selective pocket is wider and hence it can accommodate larger inhibitors such as largazole. This also explains the possible role of HDAC8 in deacylation reactions which are recent discoveries in HDAC biology. In addition to the loops the geometry of inhibitors also contributes to the inhibition mechanism. The L-shape of the inhibitor adapts towards the geometry of the active site pocket.

Understanding the role of loops in the catalytic function has opened another interesting question of fundamental aspect of HDAC8 complex-substrate recognition. This basic question was studied using the HDAC8 complex-substrate cohesin. The Cohesin complex subunit SMC3 is one of the important target for HDAC8. The Cohesin complex is formed by SMC1 (Structural Maintenance of Chromosome 1), SMC3, RAD21 and SA1/2. SMC1 and SMC3 proteins contain

three domains: the ATPase binding domain formed by N- and C-terminal regions, a long-coiled coil domain and a hinge region. SMC proteins folds back at hinge region to from the Head domains and the two hinge domains dimerize to show a V shaped structure which is further completed by RAD21 to form a ring like structure that holds sister chromatid during cohesion establishment. Acetylation of SMC3 subunit is essential for sister chromatid cohesion, which is performed by ESCOI acetyl transferase. However, HDAC8 plays an important role in deacetylation of SMC3 protein so it will be reutilized for the next cycle of cohesion establishment.

The head domains of SMC complex are highly stable when the N and C terminal regions are expressed in tandem and the presence of RAD21 stabilizes the head domains of SMC proteins. The complex structure of SMC and RAD21 may bring more information that can help in building platform for HDAC8 interaction studies. The molecular basis of HDAC8 interaction with SMC complex may add a new dimension in the HDAC biology and epigenetic drug targeting.

Thus, my thesis work has dealt with the HDAC8 selective inhibition and helps to investigate the fundamental aspects of HDAC8 complex-substrate recognition mechanisms. This work also provides the knowledge to develop isoform selective inhibitors for human and pathogenic diseases.

## 5. Future Perspectives

This thesis work has set future research directions in different aspects of the study of proteins involved in acetylation pathways, notably deacetylation.

First, HDAC8 specific inhibition studies the lead compounds will be selected for the *in vivo* drug studies. Further, optimization in schistosome specificity will be carried out in collaboration with our partner laboratories. And most importantly the knowledge gained in this project will be applied in handling other epigenetic enzymes from different parasites that are screened as valid targets. In this direction, our laboratory has already progressing in positive manner where we are able to solve the structure of a parasitic target that is different from schistosoma. *In vitro* drug screening from a large library of compounds are already started. Similarly, other epigenetic targets such as TcHAT1 and LbDAC3, will be used for further experiments to identify inhibitors.

A second aspect is to continue the work on cohesin complex. The progressive crystallization trials of one of the cohesin sub-complex will be used to understand the biology of cohesin. And most importantly the interactions cohesin complex with HDAC8 will be pursued. Further, our collaborators have identified few interacting partners of HDAC8, which we will use them to understand the structural relation.

Consequently, this thesis work will be continued to understand the molecular interactions of epigenetic enzymes and use the knowledge in an applicative field like epigenetic drug discovery and neglected tropical disease therapeutics.

## 6. List of poster presentations and oral communications

### Présentations orales:

1. **Tajith B. Shaik**, Martin Marek, Sylvie Duclaud and Christophe Romier. Towards specific inhibition of histone deacetylases: the case of HDAC8 from *Schistosoma mansoni* (2016) Les Journées du campus d'Illkirch. 21<sup>st</sup> – 22<sup>nd</sup> April, ESBS –Pole API –Illkirch.
2. **Tajith B. Shaik**, Martin Marek, Sylvie Duclaud and Christophe Romier. Selective inhibition of HDAC8 from *Schistosoma mansoni*: a structural perspective. (2016) Protein Inhibitor Interaction and Protein Function Meeting of the Binational PhD college “Enzyme reactivities and their applications” CDFA 0407. 18<sup>th</sup> November, Strasbourg – Freiburg.
3. **Tajith B. Shaik**, Martin Marek, Sylvie Duclaud and Christophe Romier. Understanding HDAC specificity: implication for epigenetic drug discovery in the treatment of human neglected diseases. (2015) IGBMC internal seminar 26<sup>th</sup> November, IGBMC, Strasbourg.

### Présentations par affiche:

1. **Tajith B. Shaik**, Martin Marek, Sylvie Duclaud and Christophe Romier. Study of *Schistosoma mansoni* HDAC8, a key epigenetic player in the treatment of Schistosomiasis. (2015) EMBO Conference: Chromatin and Epigenetics. 6-10 May, EMBL Heidelberg, Germany.
2. **Tajith B. Shaik**, *et al.*, Towards specific inhibition of histone deacetylases: the case of HDAC8 from *Schistosoma mansoni*. (2016) 3rd Freiburg Epigenetic Spring Meeting: Chemical Biology of Epigenetics. 10<sup>th</sup> – 13<sup>th</sup> April, Freiburg Institute for Advanced Studies (FRIAS), Freiburg, Germany.
3. **Tajith B. Shaik**, Martin Marek, Sylvie Duclaud and Christophe Romier. Towards specific inhibition of histone deacetylases: the case of HDAC8 from *Schistosoma mansoni*. (2016) Forum BioChem – Rencontre Academie-Industrie. 2<sup>nd</sup>-3<sup>rd</sup> June, Ecole Supérieure de Biotechnologie de Strasbourg (ESBS), Parc d'innovation, Illkirch, Strasbourg. **Best poster prize Société de chimie thérapeutique.**
4. **Tajith B. Shaik**, *et al.*, Towards Specific Inhibition of Histone Deacetylases: the case of HDAC8 from *Schistosoma mansoni*. (2016) Drug Innovation in Academia - A conference hosted by Helmholtz Drug Research 8<sup>th</sup>-9<sup>th</sup> December, German Cancer Research Center DKFZ, Heidelberg, Germany.
5. **Tajith B. Shaik**, *et al.*, Selective inhibition of HDAC8: a novel approach to treat human neglected diseases. 53rd International Conference on Medicinal Chemistry RICT 2017 Drug Discovery & Selection July 5-7, 2017 - Toulouse, Occitanie (formerly Languedoc Roussillon Midi-Pyrénées), France.

## 7. List of publications

1. Tino Heimburg, Alokta Chakrabarti, Julien Lancelot, Martin Marek, Jelena Melesina, Alexander-Thomas Hauser, **Tajith B. Shaik**, Sylvie Duclaud, Dina Robaa, Frank Erdmann, Matthias Schmidt, Christophe Romier, Raymond J. Pierce, M. J. and W. S. (2015). Structure-Based Design and Synthesis of Novel Inhibitors targeting HDAC8 from *Schistosoma mansoni* for the Treatment of Schistosomiasis. *Journal of Medicinal Chemistry*. <http://doi.org/10.1021/acs.jmedchem.5b01478>
2. Marek, M., **Shaik, T. B.**, Jung, M., Sippl, W., Pierce, R. J., & Romier, C. (2016). Combattre les maladies négligées en ciblant sélectivement leurs enzymes épigénétiques : le cas de la désacétylase 8 (HDAC8) de *Schistosoma mansoni*. *Biologie Aujourd'hui*, 210(4), 311–320.
3. Martin Marek, **Tajith B. Shaik**, Sylvie Duclaud, Raymond J. Pierce, and C. Romier. (2016). Large-Scale Overproduction and Purification of Recombinant Histone Deacetylase 8 (HDAC8) from the Human-Pathogenic Flatworm *Schistosoma mansoni*. In S. Sarkar (Ed.), *Methods in Molecular Biology* (Vol. 1436, pp. 109–118). *Springer New York*. **Book Section**.

## 8. References:

1. (2015). "Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013." Lancet **385**(9963): 117-171.
2. Allfrey, V. G., R. Faulkner and A. E. Mirsky (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.
3. Andrews, K. T., A. Haque and M. K. Jones (2012). "HDAC inhibitors in parasitic diseases." Immunol Cell Biol **90**(1): 66-77.
4. Arrar, M., R. Turnham, L. Pierce, C. A. de Oliveira and J. A. McCammon (2013). "Structural insight into the separate roles of inositol tetrakisphosphate and deacetylase-activating domain in activation of histone deacetylase 3." Protein Sci **22**(1): 83-92.
5. Arrowsmith, C. H., C. Bountra, P. V. Fish, K. Lee and M. Schapira (2012). "Epigenetic protein families: a new frontier for drug discovery." Nat Rev Drug Discov **11**(5): 384-400.
6. Balasubramanian, S., J. Ramos, W. Luo, M. Sirisawad, E. Verner and J. J. Buggy (2008). "A novel histone deacetylase 8 (HDAC8)-specific inhibitor PCI-34051 induces apoptosis in T-cell lymphomas." Leukemia **22**(5): 1026-1034.
7. Bartling, B., H. S. Hofmann, T. Boettger, G. Hansen, S. Burdach, R. E. Silber and A. Simm (2005). "Comparative application of antibody and gene array for expression profiling in human squamous cell lung carcinoma." Lung Cancer **49**(2): 145-154.
8. Bedford, M. T. and S. G. Clarke (2009). "Protein arginine methylation in mammals: who, what, and why." Mol Cell **33**(1): 1-13.
9. Bednar, J., I. Garcia-Saez, R. Boopathi, A. R. Cutter, G. Papai, A. Reymer, S. H. Syed, I. N. Lone, O. Tonchev, C. Crucifix, H. Menoni, C. Papin, D. A. Skoufias, H. Kurumizaka, R. Lavery, A. Hamiche, J. J. Hayes, P. Schultz, D. Angelov, C. Petosa and S. Dimitrov (2017). "Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1." Mol Cell **66**(3): 384-397.e388.
10. Berndsen, C. E. and J. M. Denu (2008). "Catalysis and substrate selection by histone/protein lysine acetyltransferases." Curr Opin Struct Biol **18**(6): 682-689.
11. Bhattacharya, S., R. Eckner, S. Grossman, E. Oldread, Z. Arany, A. D'Andrea and D. M. Livingston (1996). "Cooperation of Stat2 and p300/CBP in signalling induced by interferon-alpha." Nature **383**(6598): 344-347.
12. Bolden, J. E., M. J. Peart and R. W. Johnstone (2006). "Anticancer activities of histone deacetylase inhibitors." Nat Rev Drug Discov **5**(9): 769-784.
13. Bonomi, R., U. Mukhopadhyay, A. Shavrin, H. H. Yeh, A. Majhi, S. W. Dewage, A. Najjar, X. Lu, G. A. Cisneros, W. P. Tong, M. M. Alauddin, R. S. Liu, T. J. Mangner, N. Turkman and J. G. Gelovani (2015). "Novel Histone Deacetylase Class IIa Selective Substrate Radiotracers for PET Imaging of Epigenetic Regulation in the Brain." PLoS One **10**(8): e0133512.
14. Boriack-Sjodin, P. A. and K. K. Swinger (2016). "Protein Methyltransferases: A Distinct, Diverse, and Dynamic Family of Enzymes." Biochemistry **55**(11): 1557-1569.
15. Botta, C. B., W. Cabri, E. Cini, L. De Cesare, C. Fattorusso, G. Giannini, M. Persico, A. Petrella, F. Rondinelli, M. Rodriguez, A. Russo and M. Taddei (2011). "Oxime amides as a novel zinc binding group in histone deacetylase inhibitors: synthesis, biological activity, and computational evaluation." J Med Chem **54**(7): 2165-2182.
16. Bottomley, M. J., P. Lo Surdo, P. Di Giovine, A. Cirillo, R. Scarpelli, F. Ferrigno, P. Jones, P. Neddermann, R. De Francesco, C. Steinkuhler, P. Gallinari and A. Carfi (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.

17. Bowers, E. M., G. Yan, C. Mukherjee, A. Orry, L. Wang, M. A. Holbert, N. T. Crump, C. A. Hazzalin, G. Liszczak, H. Yuan, C. Larocca, S. A. Saldanha, R. Abagyan, Y. Sun, D. J. Meyers, R. Marmorstein, L. C. Mahadevan, R. M. Alani and P. A. Cole (2010). "Virtual ligand screening of the p300/CBP histone acetyltransferase: identification of a selective small molecule inhibitor." *Chem Biol* **17**(5): 471-482.
18. Boyer, L. A., R. R. Latek and C. L. Peterson (2004). "The SANT domain: a unique histone-tail-binding module?" *Nat Rev Mol Cell Biol* **5**(2): 158-163.
19. Brownell, J. E. and C. D. Allis (1995). "An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in *Tetrahymena macronuclei*." *Proc Natl Acad Sci U S A* **92**(14): 6364-6368.
20. Burli, R. W., C. A. Luckhurst, O. Aziz, K. L. Matthews, D. Yates, K. A. Lyons, M. Beconi, G. McAllister, P. Breccia, A. J. Stott, S. D. Penrose, M. Wall, M. Lamers, P. Leonard, I. Muller, C. M. Richardson, R. Jarvis, L. Stones, S. Hughes, G. Wishart, A. F. Haughan, C. O'Connell, T. Mead, H. McNeil, J. Vann, J. Mangette, M. Maillard, V. Beaumont, I. Munoz-Sanjuan and C. Dominguez (2013). "Design, synthesis, and biological evaluation of potent and selective class IIa histone deacetylase (HDAC) inhibitors as a potential therapy for Huntington's disease." *J Med Chem* **56**(24): 9934-9954.
21. Buschbeck, M. and S. B. Hake (2017). "Variants of core histones and their roles in cell fate decisions, development and cancer." *Nat Rev Mol Cell Biol* **18**(5): 299-314.
22. Cairns, B. R. (2007). "Chromatin remodeling: insights and intrigue from single-molecule studies." *Nat Struct Mol Biol* **14**(11): 989-996.
23. Candido, E. P., R. Reeves and J. R. Davie (1978). "Sodium butyrate inhibits histone deacetylation in cultured cells." *Cell* **14**(1): 105-113.
24. Cao, D., M. Wang, X. Qiu, D. Liu, H. Jiang, N. Yang and R. M. Xu (2015). "Structural basis for allosteric, substrate-dependent stimulation of SIRT1 activity by resveratrol." *Genes Dev* **29**(12): 1316-1325.
25. Carneiro, V. C., I. C. de Abreu da Silva, E. J. Torres, S. Caby, J. Lancelot, M. Vanderstraete, S. D. Furdas, M. Jung, R. J. Pierce and M. R. Fantappie (2014). "Epigenetic changes modulate schistosome egg formation and are a novel target for reducing transmission of schistosomiasis." *PLoS Pathog* **10**(5): e1004116.
26. Cavero, S., E. Herruzo, D. Ontoso and P. A. San-Segundo (2016). "Impact of histone H4K16 acetylation on the meiotic recombination checkpoint in *Saccharomyces cerevisiae*." *Microb Cell* **3**(12): 606-620.
27. Chahrour, M., S. Y. Jung, C. Shaw, X. Zhou, S. T. Wong, J. Qin and H. Y. Zoghbi (2008). "MeCP2, a key contributor to neurological disease, activates and represses transcription." *Science* **320**(5880): 1224-1229.
28. Chakrabarti, A., I. Oehme, O. Witt, G. Oliveira, W. Sippl, C. Romier, R. J. Pierce and M. Jung (2015). "HDAC8: a multifaceted target for therapeutic interventions." *Trends Pharmacol Sci* **36**(7): 481-492.
29. Chou, C. J., D. Herman and J. M. Gottesfeld (2008). "Pimelic diphenylamide 106 is a slow, tight-binding inhibitor of class I histone deacetylases." *J Biol Chem* **283**(51): 35402-35409.
30. Choudhary, C., C. Kumar, F. Gnad, M. L. Nielsen, M. Rehman, T. C. Walther, J. V. Olsen and M. Mann (2009). "Lysine acetylation targets protein complexes and co-regulates major cellular functions." *Science* **325**(5942): 834-840.
31. Choudhary, C., B. T. Weinert, Y. Nishida, E. Verdin and M. Mann (2014). "The growing landscape of lysine acetylation links metabolism and cell signalling." *Nat Rev Mol Cell Biol* **15**(8): 536-550.
32. Cioli, D. and L. Pica-Mattoccia (2003). "Praziquantel." *Parasitol Res* **90 Supp 1**: S3-9.



33. Clayton, A. L., S. Rose, M. J. Barratt and L. C. Mahadevan (2000). "Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation." *Embo j* **19**(14): 3714-3726.
34. Cobbe, N. and M. M. Heck (2004). "The evolution of SMC proteins: phylogenetic analysis and structural implications." *Mol Biol Evol* **21**(2): 332-347.
35. Cobbe, N. and M. M. Heck (2006). "The evolution of ATPase activity in SMC proteins." *Proteins* **63**(3): 685-696.
36. Corminboeuf, C., P. Hu, M. E. Tuckerman and Y. Zhang (2006). "Unexpected deacetylation mechanism suggested by a density functional theory QM/MM study of histone-deacetylase-like protein." *J Am Chem Soc* **128**(14): 4530-4531.
37. Croken, M. M., S. C. Nardelli and K. Kim (2012). "Chromatin modifications, epigenetics, and how protozoan parasites regulate their lives." *Trends Parasitol* **28**(5): 202-213.
38. Cutter, A. R. and J. J. Hayes (2015). "A brief review of nucleosome structure." *FEBS Lett* **589**(20 Pt A): 2914-2922.
39. Dasgupta, T., J. Antony, A. W. Braithwaite and J. A. Horsfield (2016). "HDAC8 Inhibition Blocks SMC3 Deacetylation and Delays Cell Cycle Progression without Affecting Cohesin-dependent Transcription in MCF7 Cancer Cells." *J Biol Chem* **291**(24): 12761-12770.
40. Davey, C. A., D. F. Sargent, K. Luger, A. W. Maeder and T. J. Richmond (2002). "Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution." *J Mol Biol* **319**(5): 1097-1113.
41. Dawson, M. A., A. J. Bannister, B. Gottgens, S. D. Foster, T. Bartke, A. R. Green and T. Kouzarides (2009). "JAK2 phosphorylates histone H3Y41 and excludes HP1alpha from chromatin." *Nature* **461**(7265): 819-822.
42. de Moraes Maciel, R., D. L. de Silva Dutra, F. D. Rumjanek, L. Juliano, M. A. Juliano and M. R. Fantappie (2004). "Schistosoma mansoni histone acetyltransferase GCN5: linking histone acetylation to gene activation." *Mol Biochem Parasitol* **133**(1): 131-135.
43. Decroos, C., C. M. Bowman, J. A. Moser, K. E. Christianson, M. A. Deardorff and D. W. Christianson (2014). "Compromised structure and function of HDAC8 mutants identified in Cornelia de Lange Syndrome spectrum disorders." *ACS Chem Biol* **9**(9): 2157-2164.
44. Decroos, C., N. H. Christianson, L. E. Gullett, C. M. Bowman, K. E. Christianson, M. A. Deardorff and D. W. Christianson (2015). "Biochemical and structural characterization of HDAC8 mutants associated with Cornelia de Lange syndrome spectrum disorders." *Biochemistry* **54**(42): 6501-6513.
45. Decroos, C., D. J. Clausen, B. E. Haines, O. Wiest, R. M. Williams and D. W. Christianson (2015). "Variable active site loop conformations accommodate the binding of macrocyclic largazole analogues to HDAC8." *Biochemistry* **54**(12): 2126-2135.
46. DesJarlais, R. and P. J. Tummino (2016). "Role of Histone-Modifying Enzymes and Their Complexes in Regulation of Chromatin Biology." *Biochemistry* **55**(11): 1584-1599.
47. Dhalluin, C., J. E. Carlson, L. Zeng, C. He, A. K. Aggarwal and M. M. Zhou (1999). "Structure and ligand of a histone acetyltransferase bromodomain." *Nature* **399**(6735): 491-496.
48. Doi, A., I. H. Park, B. Wen, P. Murakami, M. J. Aryee, R. Irizarry, B. Herb, C. Ladd-Acosta, J. Rho, S. Loewer, J. Miller, T. Schlaeger, G. Q. Daley and A. P. Feinberg (2009). "Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts." *Nat Genet* **41**(12): 1350-1353.
49. Dowling, D. P., L. Di Costanzo, H. A. Gennadios and D. W. Christianson (2008). "Evolution of the arginase fold and functional diversity." *Cell Mol Life Sci* **65**(13): 2039-2055.
50. Dowling, D. P., S. L. Gantt, S. G. Gattis, C. A. Fierke and D. W. Christianson (2008). "Structural studies of human histone deacetylase 8 and its site-specific variants complexed with substrate and inhibitors." *Biochemistry* **47**(51): 13554-13563.

51. Dowling, D. P., S. G. Gattis, C. A. Fierke and D. W. Christianson (2010). "Structures of metal-substituted human histone deacetylase 8 provide mechanistic inferences on biological function." Biochemistry **49**(24): 5048-5056.
52. Durst, K. L., B. Lutterbach, T. Kummalu, A. D. Friedman and S. W. Hiebert (2003). "The inv(16) fusion protein associates with corepressors via a smooth muscle myosin heavy-chain domain." Mol Cell Biol **23**(2): 607-619.
53. Fersht, A. R. and J. Sperling (1973). "The charge relay system in chymotrypsin and chymotrypsinogen." J Mol Biol **74**(2): 137-149.
54. Fennin, M. S., J. R. Donigian, A. Cohen, V. M. Richon, R. A. Rifkind, P. A. Marks, R. Breslow and N. P. Pavletich (1999). "Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors." Nature **401**(6749): 188-193.
55. Fournel, M., C. Bonfils, Y. Hou, P. T. Yan, M. C. Trachy-Bourget, A. Kalita, J. Liu, A. H. Lu, N. Z. Zhou, M. F. Robert, J. Gillespie, J. J. Wang, H. Ste-Croix, J. Rahil, S. Lefebvre, O. Moradei, D. Delorme, A. R. Macleod, J. M. Besterman and Z. Li (2008). "MGCD0103, a novel isotype-selective histone deacetylase inhibitor, has broad spectrum antitumor activity in vitro and in vivo." Mol Cancer Ther **7**(4): 759-768.
56. Furumai, R., Y. Komatsu, N. Nishino, S. Khochbin, M. Yoshida and S. Horinouchi (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.
57. Futscher, B. W., M. M. O'Meara, C. J. Kim, M. A. Rennels, D. Lu, L. M. Gruman, R. E. Seftor, M. J. Hendrix and F. E. Domann (2004). "Aberrant methylation of the maspin promoter is an early event in human breast cancer." Neoplasia **6**(4): 380-389.
58. Gantt, S. L., C. G. Joseph and C. A. Fierke (2010). "Activation and inhibition of histone deacetylase 8 by monovalent cations." J Biol Chem **285**(9): 6036-6043.
59. Gantt, S. M., C. Decroos, M. S. Lee, L. E. Gullett, C. M. Bowman, D. W. Christianson and C. A. Fierke (2016). "General Base-General Acid Catalysis in Human Histone Deacetylase 8." Biochemistry **55**(5): 820-832.
60. Gao, J., B. Siddoway, Q. Huang and H. Xia (2009). "Inactivation of CREB mediated gene transcription by HDAC8 bound protein phosphatase." Biochem Biophys Res Commun **379**(1): 1-5.
61. Gheldof, N., T. M. Tabuchi and J. Dekker (2006). "The active FMR1 promoter is associated with a large domain of altered chromatin conformation with embedded local histone modifications." Proc Natl Acad Sci U S A **103**(33): 12463-12468.
62. Ghosh, S. K., S. P. Perrine, R. M. Williams and D. V. Faller (2012). "Histone deacetylase inhibitors are potent inducers of gene expression in latent EBV and sensitize lymphoma cells to nucleoside antiviral agents." Blood **119**(4): 1008-1017.
63. Gligoris, T. and J. Lowe (2016). "Structural Insights into Ring Formation of Cohesin and Related Smc Complexes." Trends Cell Biol **26**(9): 680-693.
64. Glozak, M. A., N. Sengupta, X. Zhang and E. Seto (2005). "Acetylation and deacetylation of non-histone proteins." Gene **363**: 15-23.
65. Gomez-Diaz, E., M. Jorda, M. A. Peinado and A. Rivero (2012). "Epigenetics of host-pathogen interactions: the road ahead and the road behind." PLoS Pathog **8**(11): e1003007.
66. Gottesfeld, J. M. and K. Luger (2001). "Energetics and affinity of the histone octamer for defined DNA sequences." Biochemistry **40**(37): 10927-10933.
67. Grant, S., C. Easley and P. Kirkpatrick (2007). "Vorinostat." Nat Rev Drug Discov **6**(1): 21-22.
68. Gregoret, I. V., Y. M. Lee and H. V. Goodson (2004). "Molecular evolution of the histone deacetylase family: functional implications of phylogenetic analysis." J Mol Biol **338**(1): 17-31.
69. Greschik, H., R. Schule and T. Gunther (2017). "Selective targeting of epigenetic reader domains." Expert Opin Drug Discov **12**(5): 449-463.

70. Grozinger, C. M. and S. L. Schreiber (2002). "Deacetylase enzymes: biological functions and the use of small-molecule inhibitors." *Chem Biol* **9**(1): 3-16.
71. Guan, J. S., S. J. Haggarty, E. Giacometti, J. H. Dannenberg, N. Joseph, J. Gao, T. J. Nieland, Y. Zhou, X. Wang, R. Mazitschek, J. E. Bradner, R. A. DePinho, R. Jaenisch and L. H. Tsai (2009). "HDAC2 negatively regulates memory formation and synaptic plasticity." *Nature* **459**(7243): 55-60.
72. Guillemette, B., P. Drogaris, H. H. Lin, H. Armstrong, K. Hiragami-Hamada, A. Imhof, E. Bonneil, P. Thibault, A. Verreault and R. J. Festenstein (2011). "H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation." *PLoS Genet* **7**(3): e1001354.
73. Haering, C. H., D. Schoffnegger, T. Nishino, W. Helmhart, K. Nasmyth and J. Lowe (2004). "Structure and stability of cohesin's Smc1-kleisin interaction." *Mol Cell* **15**(6): 951-964.
74. Haffke, M., M. Marek, M. Pelosse, M. L. Diebold, U. Schlattner, I. Berger and C. Romier (2015). "Characterization and production of protein complexes by co-expression in *Escherichia coli*." *Methods Mol Biol* **1261**: 63-89.
75. Haggarty, S. J., K. M. Koeller, J. C. Wong, R. A. Butcher and S. L. Schreiber (2003). "Multidimensional chemical genetic analysis of diversity-oriented synthesis-derived deacetylase inhibitors using cell-based assays." *Chem Biol* **10**(5): 383-396.
76. Hai, Y. and D. W. Christianson (2016). "Histone deacetylase 6 structure and molecular basis of catalysis and inhibition." *Nat Chem Biol* **12**(9): 741-747.
77. Hai, Y., S. A. Shinsky, N. J. Porter and D. W. Christianson (2017). "Histone deacetylase 10 structure and molecular function as a polyamine deacetylase." *Nat Commun* **8**: 15368.
78. Hallows, W. C., S. Lee and J. M. Denu (2006). "Sirtuins deacetylate and activate mammalian acetyl-CoA synthetases." *Proc Natl Acad Sci U S A* **103**(27): 10230-10235.
79. Hammond, C. M., C. B. Stromme, H. Huang, D. J. Patel and A. Groth (2017). "Histone chaperone networks shaping chromatin function." *Nat Rev Mol Cell Biol* **18**(3): 141-158.
80. Hang, C. T., J. Yang, P. Han, H. L. Cheng, C. Shang, E. Ashley, B. Zhou and C. P. Chang (2010). "Chromatin regulation by Brg1 underlies heart muscle development and disease." *Nature* **466**(7302): 62-67.
81. Hara, K., G. Zheng, Q. Qu, H. Liu, Z. Ouyang, Z. Chen, D. R. Tomchick and H. Yu (2014). "Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion." *Nat Struct Mol Biol* **21**(10): 864-870.
82. Harakalova, M., M. J. van den Boogaard, R. Sinke, S. van Lieshout, M. C. van Tuil, K. Duran, I. Renkens, P. A. Terhal, C. de Kovel, I. J. Nijman, M. van Haelst, N. V. Knoers, G. van Haften, W. Kloosterman, R. C. Hennekam, E. Cuppen and H. K. Ploos van Amstel (2012). "X-exome sequencing identifies a HDAC8 variant in a large pedigree with X-linked intellectual disability, truncal obesity, gynaecomastia, hypogonadism and unusual face." *J Med Genet* **49**(8): 539-543.
83. Hassig, C. A., J. K. Tong, T. C. Fleischer, T. Owa, P. G. Grable, D. E. Ayer and S. L. Schreiber (1998). "A role for histone deacetylase activity in HDAC1-mediated transcriptional repression." *Proceedings of the National Academy of Sciences of the United States of America* **95**(7): 3519-3524.
84. Herman, D., K. Janssen, R. Burnett, E. Soragni, S. L. Perlman and J. M. Gottesfeld (2006). "Histone deacetylase inhibitors reverse gene silencing in Friedreich's ataxia." *Nat Chem Biol* **2**(10): 551-558.
85. Herold, J. M., T. J. Wigle, J. L. Norris, R. Lam, V. K. Korboukh, C. Gao, L. A. Ingerman, D. B. Kireev, G. Senisterra, M. Vedadi, A. Tripathy, P. J. Brown, C. H. Arrowsmith, J. Jin, W. P. Janzen and S. V. Frye (2011). "Small-molecule ligands of methyl-lysine binding proteins." *J Med Chem* **54**(7): 2504-2511.
86. Hildmann, C., D. Wegener, D. Riester, R. Hempel, A. Schober, J. Merana, L. Giurato, S. Guccione, T. K. Nielsen, R. Ficner and A. Schwienhorst (2006). "Substrate and inhibitor specificity of class 1 and class 2 histone deacetylases." *J Biotechnol* **124**(1): 258-270.

87. Hirano, T. (2016). "Condensin-Based Chromosome Organization from Bacteria to Vertebrates." *Cell* **164**(5): 847-857.
88. Hopkins, A. (2013). "Treating neglected tropical diseases." *Community Eye Health* **26**(82): 26-27.
89. Howitz, K. T., K. J. Bitterman, H. Y. Cohen, D. W. Lamming, S. Lavu, J. G. Wood, R. E. Zipkin, P. Chung, A. Kisielewski, L. L. Zhang, B. Scherer and D. A. Sinclair (2003). "Small molecule activators of sirtuins extend *Saccharomyces cerevisiae* lifespan." *Nature* **425**(6954): 191-196.
90. Hu, E., Z. Chen, T. Fredrickson, Y. Zhu, R. Kirkpatrick, G. F. Zhang, K. Johanson, C. M. Sung, R. Liu and J. Winkler (2000). "Cloning and characterization of a novel human class I histone deacetylase that functions as a transcription repressor." *J Biol Chem* **275**(20): 15254-15264.
91. Huang, J. and D. Moazed (2003). "Association of the RENT complex with nontranscribed and coding regions of rDNA and a regional requirement for the replication fork block protein Fob1 in rDNA silencing." *Genes Dev* **17**(17): 2162-2176.
92. Imai, S., C. M. Armstrong, M. Kaeberlein and L. Guarente (2000). "Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase." *Nature* **403**(6771): 795-800.
93. Ito, A., Y. Kawaguchi, C. H. Lai, J. J. Kovacs, Y. Higashimoto, E. Appella and T. P. Yao (2002). "MDM2-HDAC1-mediated deacetylation of p53 is required for its degradation." *Embo j* **21**(22): 6236-6245.
94. Itoh, T., L. Fairall, F. W. Muskett, C. P. Milano, P. J. Watson, N. Arnaudo, A. Saleh, C. J. Millard, M. El-Mezgueldi, F. Martino and J. W. Schwabe (2015). "Structural and functional characterization of a cell cycle associated HDAC1/2 complex reveals the structural basis for complex assembly and nucleosome targeting." *Nucleic Acids Res* **43**(4): 2033-2044.
95. Iwasaki, W., Y. Miya, N. Horikoshi, A. Osakabe, H. Taguchi, H. Tachiwana, T. Shibata, W. Kagawa and H. Kurumizaka (2013). "Contribution of histone N-terminal tails to the structure and stability of nucleosomes." *FEBS Open Bio* **3**: 363-369.
96. Jahan, S. and J. R. Davie (2015). "Protein arginine methyltransferases (PRMTs): role in chromatin organization." *Adv Biol Regul* **57**: 173-184.
97. Jaiswal, D., R. Turniansky and E. M. Green (2017). "Choose Your Own Adventure: The Role of Histone Modifications in Yeast Cell Fate." *J Mol Biol* **429**(13): 1946-1957.
98. Javierre, B. M., A. F. Fernandez, J. Richter, F. Al-Shahrour, J. I. Martin-Subero, J. Rodriguez-Ubreva, M. Berdasco, M. F. Fraga, T. P. O'Hanlon, L. G. Rider, F. V. Jacinto, F. J. Lopez-Longo, J. Dopazo, M. Forn, M. A. Peinado, L. Carreno, A. H. Sawalha, J. B. Harley, R. Siebert, M. Esteller, F. W. Miller and E. Ballestar (2010). "Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus." *Genome Res* **20**(2): 170-179.
99. Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." *Science* **293**(5532): 1074-1080.
100. Jiang, Y., J. Liu, D. Chen, L. Yan and W. Zheng (2017). "Sirtuin Inhibition: Strategies, Inhibitors, and Therapeutic Potential." *Trends Pharmacol Sci* **38**(5): 459-472.
101. Jin, B., Q. Tao, J. Peng, H. M. Soo, W. Wu, J. Ying, C. R. Fields, A. L. Delmas, X. Liu, J. Qiu and K. D. Robertson (2008). "DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function." *Hum Mol Genet* **17**(5): 690-709.
102. Jin, C., C. Zang, G. Wei, K. Cui, W. Peng, K. Zhao and G. Felsenfeld (2009). "H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions." *Nat Genet* **41**(8): 941-945.
103. Karolczak-Bayatti, M., M. Sweeney, J. Cheng, L. Edey, S. C. Robson, S. M. Ulrich, A. Treumann, M. J. Taggart and G. N. Europe-Finner (2011). "Acetylation of heat shock protein 20 (Hsp20) regulates human myometrial activity." *J Biol Chem* **286**(39): 34346-34355.

104. Khabele, D., D. S. Son, A. K. Parl, G. L. Goldberg, L. H. Augenlicht, J. M. Mariadason and V. M. Rice (2007). "Drug-induced inactivation or gene silencing of class I histone deacetylases suppresses ovarian cancer cell growth: implications for therapy." *Cancer Biol Ther* **6**(5): 795-801.
105. Khan, N., M. Jeffers, S. Kumar, C. Hackett, F. Boldog, N. Khramtsov, X. Qian, E. Mills, S. C. Berghs, N. Carey, P. W. Finn, L. S. Collins, A. Tumber, J. W. Ritchie, P. B. Jensen, H. S. Lichenstein and M. Sehested (2008). "Determination of the class and isoform selectivity of small-molecule histone deacetylase inhibitors." *Biochem J* **409**(2): 581-589.
106. Kim, H. S., K. Patel, K. Muldoon-Jacobs, K. S. Bisht, N. Aykin-Burns, J. D. Pennington, R. van der Meer, P. Nguyen, J. Savage, K. M. Owens, A. Vassilopoulos, O. Ozden, S. H. Park, K. K. Singh, S. A. Abdulkadir, D. R. Spitz, C. X. Deng and D. Gius (2010). "SIRT3 is a mitochondria-localized tumor suppressor required for maintenance of mitochondrial integrity and metabolism during stress." *Cancer Cell* **17**(1): 41-52.
107. Kinnaird, J. H., W. Weir, Z. Durrani, S. S. Pillai, M. Baird and B. R. Shiels (2013). "A Bovine Lymphosarcoma Cell Line Infected with Theileria annulata Exhibits an Irreversible Reconfiguration of Host Cell Gene Expression." *PLoS One* **8**(6): e66833.
108. Krennhrubec, K., B. L. Marshall, M. Hedglin, E. Verdin and S. M. Ulrich (2007). "Design and evaluation of 'Linkerless' hydroxamic acids as selective HDAC8 inhibitors." *Bioorg Med Chem Lett* **17**(10): 2874-2878.
109. Lahm, A., C. Paolini, M. Pallaoro, M. C. Nardi, P. Jones, P. Neddermann, S. Sambucini, M. J. Bottomley, P. Lo Surdo, A. Carfi, U. Koch, R. De Francesco, C. Steinkuhler and P. Gallinari (2007). "Unraveling the hidden catalytic activity of vertebrate class IIa histone deacetylases." *Proc Natl Acad Sci U S A* **104**(44): 17335-17340.
110. Lang, C., A. Hildebrandt, F. Brand, L. Opitz, H. Dihazi and C. G. Luder (2012). "Impaired chromatin remodelling at STAT1-regulated promoters leads to global unresponsiveness of Toxoplasma gondii-infected macrophages to IFN-gamma." *PLoS Pathog* **8**(1): e1002483.
111. Lau, O. D., T. K. Kundu, R. E. Soccio, S. Ait-Si-Ali, E. M. Khalil, A. Vassilev, A. P. Wolffe, Y. Nakatani, R. G. Roeder and P. A. Cole (2000). "HATs off: selective synthetic inhibitors of the histone acetyltransferases p300 and PCAF." *Mol Cell* **5**(3): 589-595.
112. Lauffer, B. E., R. Mintzer, R. Fong, S. Mukund, C. Tam, I. Zilberleyb, B. Flicke, A. Ritscher, G. Fedorowicz, R. Vallerio, D. F. Ortwine, J. Gunzner, Z. Modrusan, L. Neumann, C. M. Koth, P. J. Lupardus, J. S. Kaminker, C. E. Heise and P. Steiner (2013). "Histone deacetylase (HDAC) inhibitor kinetic rate constants correlate with cellular histone acetylation but not transcription and cell viability." *J Biol Chem* **288**(37): 26926-26943.
113. Lee, H., N. Rezai-Zadeh and E. Seto (2004). "Negative regulation of histone deacetylase 8 activity by cyclic AMP-dependent protein kinase A." *Mol Cell Biol* **24**(2): 765-773.
114. Lee, H., N. Sengupta, A. Villagra, N. Rezai-Zadeh and E. Seto (2006). "Histone deacetylase 8 safeguards the human ever-shorter telomeres 1B (hEST1B) protein from ubiquitin-mediated degradation." *Mol Cell Biol* **26**(14): 5259-5269.
115. Li, Y., H. Wen, Y. Xi, K. Tanaka, H. Wang, D. Peng, Y. Ren, Q. Jin, S. Y. Dent, W. Li, H. Li and X. Shi (2014). "AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation." *Cell* **159**(3): 558-571.
116. Li, Z. and W. G. Zhu (2014). "Targeting histone deacetylases for cancer therapy: from molecular mechanisms to clinical implications." *Int J Biol Sci* **10**(7): 757-770.
117. Liu, W. (2016). "Epigenetics in Schistosomes: What We Know and What We Need Know." *Front Cell Infect Microbiol* **6**: 149.
118. Lobera, M., K. P. Madauss, D. T. Pohlhaus, Q. G. Wright, M. Trocha, D. R. Schmidt, E. Baloglu, R. P. Trump, M. S. Head, G. A. Hofmann, M. Murray-Thompson, B. Schwartz, S. Chakravorty, Z. Wu, P. K.

- Mander, L. Kruidenier, R. A. Reid, W. Burkhart, B. J. Turunen, J. X. Rong, C. Wagner, M. B. Moyer, C. Wells, X. Hong, J. T. Moore, J. D. Williams, D. Soler, S. Ghosh and M. A. Nolan (2013). "Selective class IIa histone deacetylase inhibition via a nonchelating zinc-binding group." *Nat Chem Biol* **9**(5): 319-325.
119. Lopez-Serra, L., G. Kelly, H. Patel, A. Stewart and F. Uhlmann (2014). "The Scc2-Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions." *Nat Genet* **46**(10): 1147-1151.
120. Losada, A. (2014). "Cohesin in cancer: chromosome segregation and beyond." *Nat Rev Cancer* **14**(6): 389-393.
121. Lujambio, A., G. A. Calin, A. Villanueva, S. Roperio, M. Sanchez-Cespedes, D. Blanco, L. M. Montuenga, S. Rossi, M. S. Nicoloso, W. J. Faller, W. M. Gallagher, S. A. Eccles, C. M. Croce and M. Esteller (2008). "A microRNA DNA methylation signature for human cancer metastasis." *Proc Natl Acad Sci U S A* **105**(36): 13556-13561.
122. Ma, F. and C. Y. Zhang (2016). "Histone modifying enzymes: novel disease biomarkers and assay development." *Expert Rev Mol Diagn* **16**(3): 297-306.
123. Mann, B. S., J. R. Johnson, K. He, R. Sridhara, S. Abraham, B. P. Booth, L. Verbois, D. E. Morse, J. M. Jee, S. Pope, R. S. Harapanhalli, R. Dagher, A. Farrell, R. Justice and R. Pazdur (2007). "Vorinostat for treatment of cutaneous manifestations of advanced primary cutaneous T-cell lymphoma." *Clin Cancer Res* **13**(8): 2318-2322.
124. Marek, M., S. Kannan, A. T. Hauser, M. Moraes Mourao, S. Caby, V. Cura, D. A. Stolfa, K. Schmidtkunz, J. Lancelot, L. Andrade, J. P. Renaud, G. Oliveira, W. Sippl, M. Jung, J. Cavarelli, R. J. Pierce and C. Romier (2013). "Structural basis for the inhibition of histone deacetylase 8 (HDAC8), a key epigenetic player in the blood fluke *Schistosoma mansoni*." *PLoS Pathog* **9**(9): e1003645.
125. Marek, M., T. B. Shaik, S. Duclaud, R. J. Pierce and C. Romier (2016). "Large-Scale Overproduction and Purification of Recombinant Histone Deacetylase 8 (HDAC8) from the Human-Pathogenic Flatworm *Schistosoma mansoni*." *Methods Mol Biol* **1436**: 109-118.
126. Marks, P. A., V. M. Richon and R. A. Rifkind (2000). "Histone deacetylase inhibitors: inducers of differentiation or apoptosis of transformed cells." *J Natl Cancer Inst* **92**(15): 1210-1216.
127. Marsolier, J., M. Perichon, J. D. DeBarry, B. O. Villoutreix, J. Chluba, T. Lopez, C. Garrido, X. Z. Zhou, K. P. Lu, L. Fritsch, S. Ait-Si-Ali, M. Mhadhbi, S. Medjkane and J. B. Weitzman (2015). "Theileria parasites secrete a prolyl isomerase to maintain host leukocyte transformation." *Nature* **520**(7547): 378-382.
128. McGhee, J. D., J. M. Nickol, G. Felsenfeld and D. C. Rau (1983). "Higher order structure of chromatin: orientation of nucleosomes within the 30 nm chromatin solenoid is independent of species and spacer length." *Cell* **33**(3): 831-841.
129. Menzies, K. J., H. Zhang, E. Katsyuba and J. Auwerx (2016). "Protein acetylation in metabolism - metabolites and cofactors." *Nat Rev Endocrinol* **12**(1): 43-60.
130. Merrick, C. J. and M. T. Duraisingh (2010). "Epigenetics in Plasmodium: what do we really know?" *Eukaryot Cell* **9**(8): 1150-1158.
131. Miao, F., D. D. Smith, L. Zhang, A. Min, W. Feng and R. Natarajan (2008). "Lymphocytes from patients with type 1 diabetes display a distinct profile of chromatin histone H3 lysine 9 dimethylation: an epigenetic study in diabetes." *Diabetes* **57**(12): 3189-3198.
132. Millard, C. J., P. J. Watson, I. Celardo, Y. Gordiyenko, S. M. Cowley, C. V. Robinson, L. Fairall and J. W. Schwabe (2013). "Class I HDACs share a common mechanism of regulation by inositol phosphates." *Mol Cell* **51**(1): 57-67.
133. Miyake, Y., J. J. Keusch, L. Wang, M. Saito, D. Hess, X. Wang, B. J. Melancon, P. Helquist, H. Gut and P. Matthias (2016). "Structural insights into HDAC6 tubulin deacetylation and its selective inhibition." *Nat Chem Biol* **12**(9): 748-754.

134. Moosmann, A., C. Campsteijn, P. W. Jansen, C. Nasrallah, M. Raasholm, H. G. Stunnenberg and E. M. Thompson (2011). "Histone variant innovation in a rapidly evolving chordate lineage." BMC Evol Biol **11**: 208.
135. Mulero-Navarro, S. and M. Esteller (2008). "Chromatin remodeling factor CHD5 is silenced by promoter CpG island hypermethylation in human cancer." Epigenetics **3**(4): 210-215.
136. Murgatroyd, C., A. V. Patchev, Y. Wu, V. Micale, Y. Bockmuhl, D. Fischer, F. Holsboer, C. T. Wotjak, O. F. Almeida and D. Spengler (2009). "Dynamic DNA methylation programs persistent adverse effects of early-life stress." Nat Neurosci **12**(12): 1559-1566.
137. New, M., H. Olzscha and N. B. La Thangue (2012). "HDAC inhibitor-based therapies: can we interpret the code?" Mol Oncol **6**(6): 637-656.
138. Ng, S. S., W. W. Yue, U. Oppermann and R. J. Klose (2009). "Dynamic protein methylation in chromatin biology." Cell Mol Life Sci **66**(3): 407-422.
139. Nishino, Y., M. Eltsov, Y. Joti, K. Ito, H. Takata, Y. Takahashi, S. Hihara, A. S. Frangakis, N. Imamoto, T. Ishikawa and K. Maeshima (2012). "Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure." Embo j **31**(7): 1644-1653.
140. North, J. A., S. Javaid, M. B. Ferdinand, N. Chatterjee, J. W. Picking, M. Shoffner, R. J. Nakkula, B. Bartholomew, J. J. Ottesen, R. Fishel and M. G. Poirier (2011). "Phosphorylation of histone H3(T118) alters nucleosome dynamics and remodeling." Nucleic Acids Res **39**(15): 6465-6474.
141. Oger, F., F. Dubois, S. Caby, C. Noel, J. Cornette, B. Bertin, M. Capron and R. J. Pierce (2008). "The class I histone deacetylases of the platyhelminth parasite *Schistosoma mansoni*." Biochem Biophys Res Commun **377**(4): 1079-1084.
142. Olins, A. L. and D. E. Olins (1974). "Spheroid chromatin units (v bodies)." Science **183**(4122): 330-332.
143. Orta, M. L., N. Pastor, E. Burgos-Moron, I. Dominguez, J. M. Calderon-Montano, C. Huertas Castano, M. Lopez-Lazaro, T. Helleday and S. Mateos (2017). "Zebularine induces replication-dependent double-strand breaks which are preferentially repaired by homologous recombination." DNA Repair (Amst) **57**: 116-124.
144. Pax, R., J. L. Bennett and R. Fetterer (1978). "A benzodiazepine derivative and praziquantel: effects on musculature of *Schistosoma mansoni* and *Schistosoma japonicum*." Naunyn Schmiedebergs Arch Pharmacol **304**(3): 309-315.
145. Peng, C., Z. Lu, Z. Xie, Z. Cheng, Y. Chen, M. Tan, H. Luo, Y. Zhang, W. He, K. Yang, B. M. Zwaans, D. Tishkoff, L. Ho, D. Lombard, T. C. He, J. Dai, E. Verdin, Y. Ye and Y. Zhao (2011). "The first identification of lysine malonylation substrates and its regulatory enzyme." Mol Cell Proteomics **10**(12): M111.012658.
146. Pieper, H. C., B. O. Evert, O. Kaut, P. F. Riederer, A. Waha and U. Wullner (2008). "Different methylation of the TNF-alpha promoter in cortex and substantia nigra: Implications for selective neuronal vulnerability." Neurobiol Dis **32**(3): 521-527.
147. Pirali, T., F. Pagliai, C. Mercurio, R. Boggio, P. L. Canonico, G. Sorba, G. C. Tron and A. A. Genazzani (2008). "Triazole-modified histone deacetylase inhibitors as a rapid route to drug discovery." J Comb Chem **10**(5): 624-627.
148. Portela, A. and M. Esteller (2010). "Epigenetic modifications and human disease." Nat Biotech **28**(10): 1057-1068.
149. Portela, A. and M. Esteller (2010). "Epigenetic modifications and human disease." Nat Biotechnol **28**(10): 1057-1068.
150. Ren, M., Y. Leng, M. Jeong, P. R. Leeds and D. M. Chuang (2004). "Valproic acid reduces brain damage induced by transient focal cerebral ischemia in rats: potential roles of histone deacetylase inhibition and heat shock protein induction." J Neurochem **89**(6): 1358-1367.

151. Rivero, F. D., A. Saura, C. G. Prucca, P. G. Carranza, A. Torri and H. D. Lujan (2010). "Disruption of antigenic variation is crucial for effective parasite vaccine." *Nat Med* **16**(5): 551-557, 551p following 557.
152. Robert McMaster, W., C. J. Morrison and M. S. Kobor (2016). "Epigenetics: A New Model for Intracellular Parasite-Host Cell Regulation." *Trends Parasitol* **32**(7): 515-521.
153. Robertson, K. D. (2005). "DNA methylation and human disease." *Nat Rev Genet* **6**(8): 597-610.
154. Rossetto, D., N. Avvakumov and J. Cote (2012). "Histone phosphorylation: a chromatin modification involved in diverse nuclear events." *Epigenetics* **7**(10): 1098-1108.
155. Ruthenburg, A. J., H. Li, D. J. Patel and C. D. Allis (2007). "Multivalent engagement of chromatin modifications by linked binding modules." *Nat Rev Mol Cell Biol* **8**(12): 983-994.
156. Sacconay, L., P. A. Carrupt and A. Nurisso (2016). "Human sirtuins: Structures and flexibility." *J Struct Biol* **196**(3): 534-542.
157. Sadakierska-Chudy, A. and M. Filip (2015). "A comprehensive view of the epigenetic landscape. Part II: Histone post-translational modification, nucleosome level, and chromatin regulation by ncRNAs." *Neurotox Res* **27**(2): 172-197.
158. Sadakierska-Chudy, A., R. M. Kostrzewa and M. Filip (2015). "A comprehensive view of the epigenetic landscape part I: DNA methylation, passive and active DNA demethylation pathways and histone variants." *Neurotox Res* **27**(1): 84-97.
159. Saha, A., J. Wittmeyer and B. R. Cairns (2006). "Chromatin remodelling: the industrial revolution of DNA around histones." *Nat Rev Mol Cell Biol* **7**(6): 437-447.
160. Saji, S., M. Kawakami, S. Hayashi, N. Yoshida, M. Hirose, S. Horiguchi, A. Itoh, N. Funata, S. L. Schreiber, M. Yoshida and M. Toi (2005). "Significance of HDAC6 regulation via estrogen signaling for cell motility and prognosis in estrogen receptor-positive breast cancer." *Oncogene* **24**(28): 4531-4539.
161. Santo, L., T. Hideshima, A. L. Kung, J. C. Tseng, D. Tamang, M. Yang, M. Jarpe, J. H. van Duzer, R. Mazitschek, W. C. Ogier, D. Cirstea, S. Rodig, H. Eda, T. Scullen, M. Canavese, J. Bradner, K. C. Anderson, S. S. Jones and N. Raje (2012). "Preclinical activity, pharmacodynamic, and pharmacokinetic properties of a selective HDAC6 inhibitor, ACY-1215, in combination with bortezomib in multiple myeloma." *Blood* **119**(11): 2579-2589.
162. Sarkar, S. and D. V. Faller (2011). "T-oligos inhibit growth and induce apoptosis in human ovarian cancer cells." *Oligonucleotides* **21**(1): 47-53.
163. Sarkar, S., S. Goldgar, S. Byler, S. Rosenthal and S. Heerboth (2013). "Demethylation and re-expression of epigenetically silenced tumor suppressor genes: sensitization of cancer cells by combination therapy." *Epigenomics* **5**(1): 87-94.
164. Sarkar, S., G. Horn, K. Moulton, A. Oza, S. Byler, S. Kokolus and M. Longacre (2013). "Cancer development, progression, and therapy: an epigenetic overview." *Int J Mol Sci* **14**(10): 21087-21113.
165. Sawicka, A. and C. Seiser (2014). "Sensing core histone phosphorylation - a matter of perfect timing." *Biochim Biophys Acta* **1839**(8): 711-718.
166. Schalch, T., S. Duda, D. F. Sargent and T. J. Richmond (2005). "X-ray structure of a tetranucleosome and its implications for the chromatin fibre." *Nature* **436**(7047): 138-141.
167. Schiedel, M., M. Marek, J. Lancelot, B. Karaman, I. Almlof, J. Schultz, W. Sippl, R. J. Pierce, C. Romier and M. Jung (2015). "Fluorescence-based screening assays for the NAD(+)-dependent histone deacetylase smSirt2 from *Schistosoma mansoni*." *J Biomol Screen* **20**(1): 112-121.
168. Schleiffer, A., S. Kaitna, S. Maurer-Stroh, M. Glotzer, K. Nasmyth and F. Eisenhaber (2003). "Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners." *Mol Cell* **11**(3): 571-575.
169. Schuetz, A., J. Min, A. Allali-Hassani, M. Schapira, M. Shuen, P. Loppnau, R. Mazitschek, N. P. Kwiatkowski, T. A. Lewis, R. L. Maglathin, T. H. McLean, A. Bochkarev, A. N. Plotnikov, M. Vedadi and C. H. Arrowsmith (2008). "Human HDAC7 harbors a class IIa histone deacetylase-specific zinc binding motif and cryptic deacetylase activity." *J Biol Chem* **283**(17): 11355-11363.



170. Schwer, B., J. Bunkenborg, R. O. Verdin, J. S. Andersen and E. Verdin (2006). "Reversible lysine acetylation controls the activity of the mitochondrial enzyme acetyl-CoA synthetase 2." Proc Natl Acad Sci U S A **103**(27): 10224-10229.
171. Seto, E. and M. Yoshida (2014). "Erasers of histone acetylation: the histone deacetylase enzymes." Cold Spring Harb Perspect Biol **6**(4): a018713.
172. Shakespear, M. R., M. A. Halili, K. M. Irvine, D. P. Fairlie and M. J. Sweet (2011). "Histone deacetylases as regulators of inflammation and immunity." Trends Immunol **32**(7): 335-343.
173. Shi, X., T. Hong, K. L. Walter, M. Ewalt, E. Michishita, T. Hung, D. Carney, P. Pena, F. Lan, M. R. Kaadige, N. Lacoste, C. Cayrou, F. Davrazou, A. Saha, B. R. Cairns, D. E. Ayer, T. G. Kutateladze, Y. Shi, J. Cote, K. F. Chua and O. Gozani (2006). "ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression." Nature **442**(7098): 96-99.
174. Singh, E. K., L. A. Nazarova, S. A. Lopera, L. D. Alexander and S. R. McAlpine (2010). "Histone Deacetylase Inhibitors: Synthesis of Cyclic Tetrapeptides and their Triazole Analogues." Tetrahedron Lett **51**(33): 4357-4360.
175. Singh, R. and P. N. Pandey (2015). "Molecular docking and molecular dynamics study on SmHDAC1 to identify potential lead compounds against Schistosomiasis." Mol Biol Rep **42**(3): 689-698.
176. Somoza, J. R., R. J. Skene, B. A. Katz, C. Mol, J. D. Ho, A. J. Jennings, C. Luong, A. Arvai, J. J. Buggy, E. Chi, J. Tang, B. C. Sang, E. Verner, R. Wynands, E. M. Leahy, D. R. Dougan, G. Snell, M. Navre, M. W. Knuth, R. V. Swanson, D. E. McRee and L. W. Tari (2004). "Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases." Structure **12**(7): 1325-1334.
177. Song, J., J. H. Noh, J. H. Lee, J. W. Eun, Y. M. Ahn, S. Y. Kim, S. H. Lee, W. S. Park, N. J. Yoo, J. Y. Lee and S. W. Nam (2005). "Increased expression of histone deacetylase 2 is found in human gastric cancer." Apmis **113**(4): 264-268.
178. Song, S., Y. Wang, P. Xu, R. Yang, Z. Ma, S. Liang and G. Zhang (2015). "The inhibition of histone deacetylase 8 suppresses proliferation and inhibits apoptosis in gastric adenocarcinoma." Int J Oncol **47**(5): 1819-1828.
179. Suzuki, T., N. Muto, M. Bando, Y. Itoh, A. Masaki, M. Ri, Y. Ota, H. Nakagawa, S. Iida, K. Shirahige and N. Miyata (2014). "Design, synthesis, and biological activity of NCC149 derivatives as histone deacetylase 8-selective inhibitors." ChemMedChem **9**(3): 657-664.
180. Talbert, P. B. and S. Henikoff (2017). "Histone variants on the move: substrates for chromatin dynamics." Nat Rev Mol Cell Biol **18**(2): 115-126.
181. Taunton, J., C. A. Hassig and S. L. Schreiber (1996). "A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p." Science **272**(5260): 408-411.
182. Thevenet, L., C. Mejean, B. Moniot, N. Bonneaud, N. Galeotti, G. Aldrian-Herrada, F. Poulat, P. Berta, M. Benkirane and B. Boizet-Bonhoure (2004). "Regulation of human SRY subcellular distribution by its acetylation/deacetylation." Embo j **23**(16): 3336-3345.
183. Tian, X., S. Zhang, H. M. Liu, Y. B. Zhang, C. A. Blair, D. Mercola, P. Sassone-Corsi and X. Zi (2013). "Histone lysine-specific methyltransferases and demethylases in carcinogenesis: new targets for cancer therapy and prevention." Curr Cancer Drug Targets **13**(5): 558-579.
184. Tremethick, D. J. (2007). "Higher-order structures of chromatin: the elusive 30 nm fiber." Cell **128**(4): 651-654.
185. Turner, B. M. (2000). "Histone acetylation and an epigenetic code." Bioessays **22**(9): 836-845.
186. Uhlmann, F. (2016). "SMC complexes: from DNA to chromosomes." Nat Rev Mol Cell Biol **17**(7): 399-412.
187. Urduingio, R. G., J. V. Sanchez-Mut and M. Esteller (2009). "Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies." Lancet Neurol **8**(11): 1056-1072.

188. van Bavel, C. C., J. W. Dieker, Y. Kroeze, W. P. Tamboer, R. Voll, S. Muller, J. H. Berden and J. van der Vlag (2011). "Apoptosis-induced histone H3 methylation is targeted by autoantibodies in systemic lupus erythematosus." *Ann Rheum Dis* **70**(1): 201-207.
189. Vanden Berghe, W., M. N. Ndlovu, R. Hoya-Arias, N. Dijsselbloem, S. Gerlo and G. Haegeman (2006). "Keeping up NF-kappaB appearances: epigenetic control of immunity or inflammation-triggered epigenetics." *Biochem Pharmacol* **72**(9): 1114-1131.
190. Vannini, A., C. Volpari, G. Filocamo, E. C. Casavola, M. Brunetti, D. Renzoni, P. Chakravarty, C. Paolini, R. De Francesco, P. Gallinari, C. Steinkuhler and S. Di Marco (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.
191. Vannini, A., C. Volpari, P. Gallinari, P. Jones, M. Mattu, A. Carfi, R. De Francesco, C. Steinkuhler and S. Di Marco (2007). "Substrate binding to histone deacetylases as shown by the crystal structure of the HDAC8-substrate complex." *EMBO Rep* **8**(9): 879-884.
192. Vanommeslaeghe, K., F. De Proft, S. Loverix, D. Tourwe and P. Geerlings (2005). "Theoretical study revealing the functioning of a novel combination of catalytic motifs in histone deacetylase." *Bioorg Med Chem* **13**(12): 3987-3992.
193. Varambally, S., Q. Cao, R. S. Mani, S. Shankar, X. Wang, B. Ateeq, B. Laxman, X. Cao, X. Jing, K. Ramnarayanan, J. C. Brenner, J. Yu, J. H. Kim, B. Han, P. Tan, C. Kumar-Sinha, R. J. Lonigro, N. Palanisamy, C. A. Maher and A. M. Chinnaiyan (2008). "Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer." *Science* **322**(5908): 1695-1699.
194. Verdin, E. and M. Ott (2015). "50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond." *Nat Rev Mol Cell Biol* **16**(4): 258-264.
195. Vogler, C., C. Huber, T. Waldmann, R. Ettig, L. Braun, A. Izzo, S. Daujat, I. Chassignet, A. J. Lopez-Contreras, O. Fernandez-Capetillo, M. Dundr, K. Rippe, G. Langst and R. Schneider (2010). "Histone H2A C-terminus regulates chromatin dynamics, remodeling, and histone H1 binding." *PLoS Genet* **6**(12): e1001234.
196. Waltregny, D., L. De Leval, W. Glenisson, S. Ly Tran, B. J. North, A. Bellahcene, U. Weidle, E. Verdin and V. Castronovo (2004). "Expression of histone deacetylase 8, a class I histone deacetylase, is restricted to cells showing smooth muscle differentiation in normal human tissues." *Am J Pathol* **165**(2): 553-564.
197. Wang, D., P. Helquist and O. Wiest (2007). "Zinc binding in HDAC inhibitors: a DFT study." *J Org Chem* **72**(14): 5446-5449.
198. Wang, D., N. B. Ulyanov and V. B. Zhurkin (2010). "Sequence-dependent Kink-and-Slide deformations of nucleosomal DNA facilitated by histone arginines bound in the minor groove." *J Biomol Struct Dyn* **27**(6): 843-859.
199. Wang, D. F., O. Wiest, P. Helquist, H. Y. Lan-Hargest and N. L. Wiech (2004). "On the function of the 14 Å long internal cavity of histone deacetylase-like protein: implications for the design of histone deacetylase inhibitors." *J Med Chem* **47**(13): 3409-3417.
200. Wang, P., C. Lin, E. R. Smith, H. Guo, B. W. Sanderson, M. Wu, M. Gogol, T. Alexander, C. Seidel, L. M. Wiedemann, K. Ge, R. Krumlauf and A. Shilatifard (2009). "Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II." *Mol Cell Biol* **29**(22): 6074-6085.
201. Wapenaar, H. and F. J. Dekker (2016). "Histone acetyltransferases: challenges in targeting bi-substrate enzymes." *Clin Epigenetics* **8**: 59.
202. Watson, P. J., L. Fairall, G. M. Santos and J. W. Schwabe (2012). "Structure of HDAC3 bound to co-repressor and inositol tetrakisphosphate." *Nature* **481**(7381): 335-340.

203. Watson, P. J., C. J. Millard, A. M. Riley, N. S. Robertson, L. C. Wright, H. Y. Godage, S. M. Cowley, A. G. Jamieson, B. V. Potter and J. W. Schwabe (2016). "Insights into the activation mechanism of class I HDAC complexes by inositol phosphates." Nat Commun **7**: 11262.
204. Webster, A. L., M. S. Yan and P. A. Marsden (2013). "Epigenetics and cardiovascular disease." Can J Cardiol **29**(1): 46-57.
205. Weerasinghe, S. V., G. Estiu, O. Wiest and M. K. Pflum (2008). "Residues in the 11 A channel of histone deacetylase 1 promote catalytic activity: implications for designing isoform-selective histone deacetylase inhibitors." J Med Chem **51**(18): 5542-5551.
206. Wen, Y. D., V. Perissi, L. M. Staszewski, W. M. Yang, A. Krones, C. K. Glass, M. G. Rosenfeld and E. Seto (2000). "The histone deacetylase-3 complex contains nuclear receptor corepressors." Proc Natl Acad Sci U S A **97**(13): 7202-7207.
207. West, A. C. and R. W. Johnstone (2014). "New and emerging HDAC inhibitors for cancer treatment." J Clin Invest **124**(1): 30-39.
208. Whitehead, L., M. R. Dobler, B. Radetich, Y. Zhu, P. W. Atadja, T. Claiborne, J. E. Grob, A. McRiner, M. R. Pancost, A. Patnaik, W. Shao, M. Shultz, R. Tichkule, R. A. Tommasi, B. Vash, P. Wang and T. Stams (2011). "Human HDAC isoform selectivity achieved via exploitation of the acetate release channel with structurally unique small molecule inhibitors." Bioorg Med Chem **19**(15): 4626-4634.
209. Wiech, N. L., J. F. Fisher, P. Helquist and O. Wiest (2009). "Inhibition of histone deacetylases: a pharmacological approach to the treatment of non-cancer disorders." Curr Top Med Chem **9**(3): 257-271.
210. Wilmott, J. S., A. J. Colebatch, H. Kakavand, P. Shang, M. S. Carlino, J. F. Thompson, G. V. Long, R. A. Scolyer and P. Hersey (2015). "Expression of the class 1 histone deacetylases HDAC8 and 3 are associated with improved survival of patients with metastatic melanoma." Mod Pathol **28**(7): 884-894.
211. Wilson, A. S., B. E. Power and P. L. Molloy (2007). "DNA hypomethylation and human diseases." Biochim Biophys Acta **1775**(1): 138-162.
212. Wolfson, N. A., C. A. Pitcairn and C. A. Fierke (2013). "HDAC8 substrates: Histones and beyond." Biopolymers **99**(2): 112-126.
213. Wood, A. and A. Shilatifard (2004). "Posttranslational modifications of histones by methylation." Adv Protein Chem **67**: 201-222.
214. Woodcock, C. L., L. L. Frado and J. B. Rattner (1984). "The higher-order structure of chromatin: evidence for a helical ribbon arrangement." J Cell Biol **99**(1 Pt 1): 42-52.
215. Wu, J., A. A. Carmen, R. Kobayashi, N. Suka and M. Grunstein (2001). "HDA2 and HDA3 are related proteins that interact with and are essential for the activity of the yeast histone deacetylase HDA1." Proc Natl Acad Sci U S A **98**(8): 4391-4396.
216. Wu, J., C. Du, Z. Lv, C. Ding, J. Cheng, H. Xie, L. Zhou and S. Zheng (2013). "The up-regulation of histone deacetylase 8 promotes proliferation and inhibits apoptosis in hepatocellular carcinoma." Dig Dis Sci **58**(12): 3545-3553.
217. Wu, Z., J. Connolly and K. K. Biggar (2017). "Beyond histones - the expanding roles of protein lysine methylation." Febs j.
218. Yan, Y., S. Harper, D. W. Speicher and R. Marmorstein (2002). "The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate." Nat Struct Biol **9**(11): 862-869.
219. Yang, X. J. and E. Seto (2008). "The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men." Nat Rev Mol Cell Biol **9**(3): 206-218.
220. Yoshida, M., N. Kudo, S. Kosono and A. Ito (2017). "Chemical and structural biology of protein lysine deacetylases." Proc Jpn Acad Ser B Phys Biol Sci **93**(5): 297-321.
221. Zentner, G. E. and S. Henikoff (2013). "Regulation of nucleosome dynamics by histone modifications." Nat Struct Mol Biol **20**(3): 259-266.

222. Zhang, T., S. Cooper and N. Brockdorff (2015). "The interplay of histone modifications - writers that read." *EMBO Rep* **16**(11): 1467-1481.
223. Zhang, X., Z. Yuan, Y. Zhang, S. Yong, A. Salas-Burgos, J. Koomen, N. Olashaw, J. T. Parsons, X. J. Yang, S. R. Dent, T. P. Yao, W. S. Lane and E. Seto (2007). "HDAC6 modulates cell motility by altering the acetylation level of cortactin." *Mol Cell* **27**(2): 197-213.
224. Zhu, P. and G. Li (2016). "Structural insights of nucleosome and the 30-nm chromatin fiber." *Curr Opin Struct Biol* **36**: 106-115.
225. Zuber, J., J. Shi, E. Wang, A. R. Rappaport, H. Herrmann, E. A. Sison, D. Magoon, J. Qi, K. Blatt, M. Wunderlich, M. J. Taylor, C. Johns, A. Chicas, J. C. Mulloy, S. C. Kogan, P. Brown, P. Valent, J. E. Bradner, S. W. Lowe and C. R. Vakoc (2011). "RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia." *Nature* **478**(7370): 524-528.

## Etude biochimique, biophysique et structurale du mécanisme d'action et de l'inhibition sélective de l'histone desacetylase HDAC8

### Résumé en anglais

Histone deacetylases (HDACs) are the major targets of currently FDA-approved anti-cancer epigenetic drugs. HDACs also play an important role in the homeostasis of eukaryotic pathogens. Hence, a strategy to tackle neglected diseases caused by these pathogens is to modify currently approved epigenetic drugs targeting HDACs. HDAC8 from *Schistosoma mansoni* (smHDAC8) was shown to be a valid drug target to treat schistosomiasis, second deadliest tropical disease after malaria. Structural differences between human HDAC8 and smHDAC8 catalytic pocket enabled the design of schistosome-selective inhibitors that bind in a HDAC8 selective pocket, which is unique to HDAC8 among the highly conserved HDAC isozymes. This thesis work shows how to target selectively related isoforms with the help of atomic resolution structures, and opens the door to the investigation of the mode of action of HDAC8 at the fundamental level.

**Key words:** Epigenetics, structural biology, neglected diseases, HDAC8, selective inhibition.

### Résumé en français

Les histones désacétylases (HDACs) sont les principales cibles des médicaments épigénétiques anticancéreux actuellement approuvés par la FDA. Les HDACs jouent aussi un rôle important dans l'homéostasie des pathogènes eucaryotes. Par conséquent, une stratégie pour lutter contre les maladies négligées causées par ces pathogènes est de modifier les médicaments épigénétiques actuellement approuvés qui ciblent les HDACs. HDAC8 de *Schistosoma mansoni* (smHDAC8) est une cible médicamenteuse valable pour traiter la schistosomiase, deuxième maladie négligée mortelle après le paludisme. Les différences structurales entre les poches catalytiques des HDAC8 humaine et smHDAC8 ont permis la conception d'inhibiteurs sélectifs des schistosomes qui se lient dans une poche sélective unique à HDAC8. Ce travail de thèse montre comment cibler sélectivement des isoformes HDAC l'aide de structures à résolution atomique, et ouvre la porte à l'étude du mode d'action de HDAC8 au niveau fondamental.

**Mots-clés:** Epigénétique, biologie structurale, maladies négligées, HDAC8, inhibition sélective.