STRUCTURAL AND BIOCHEMICAL STUDIES OF HISTONE ACETYLTRANSFERASES AND SIRTUIN ENZYMES

by
Alison E. Moreno (Ringel)

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
March 2015
Abstract

Lysine acetylation is an abundant and reversible post-translational modification that regulates a diverse range of biological processes. Dedicated enzyme families exist that catalyze and remove acetyl modifications from lysine, and dysregulation of so-called HATs (Histone Acetyltransferases) and HDACs (Histone DeACetylases) has been implicated in cancer and other human diseases. This dissertation focuses on structural and biochemical characterization of two classes of enzymes that regulate lysine acetylation. First, Gcn5 is a highly conserved acetyltransferase associated with the SAGA transcriptional co-activator. While Gcn5 is catalytically active on its own, its specificity and overall activity changes when it associates with three additional SAGA subunits, Ada2, Ada3, and Sgf29, which are collectively referred to as the Histone Acetyltransferase (HAT) module. Sgf29 contains a tandem Tudor domain that specifically recognizes H3K4 trimethylation (H3K4me3), but how this interaction modulates the acetyltransferase activity of Gcn5 is still unknown. Here, I have studied the interplay between H3K4 trimethyl recognition and acetylation by Gcn5, and have developed a new technique for monitoring crosstalk between lysine acetylation and other histone post-translational modifications. Whereas Gcn5 is only slightly faster on H3K4me3 compared to unmodified nucleosomes, H3K4 trimethyl recognition by Sgf29 promotes processive acetylation by the HAT module, explaining how co-localization between hyperacetylation and H3K4 trimethylation is established at gene promoters on a mechanistic level. Sirtuins, which catalyze NAD⁺-dependent
lysine deacylation, represent the second class of enzymes studied in this dissertation. In light of recent studies demonstrating that many sirtuins remove other acyl modifications than acetylation, we revisited the lysine specificity of four bacterial and archaeal enzymes originally identified as deacetylases using biochemistry and x-ray crystallography. Out of the four sirtuins tested, three preferentially removed acyl modifications other than acetylation from lysine. Crystal structures of Sir2Af1 bound to a succinylated peptide and Sir2Af2 bound to a myristoylated peptide revealed structural determinants of acyl chain specificity, including an unexpected conformational change allowing Sir2Af2 to accommodate a myristoyl chain in its active site. These results identify desuccinylation and demyristoylation as activities catalyzed by archaeal sirtuins, and emphasize the importance of structural and biochemical studies to our understanding of sirtuin specificity.

Advisor: Cynthia Wolberger

Reader: James Berger
Acknowledgements

There are many people who supported me throughout graduate school, and I am deeply grateful to have had such a wonderful network of mentors, family, and friends.

First, I would like to thank my advisor and mentor, Cynthia Wolberger, without whom this work would not have been possible. Beyond your excellent scientific guidance, you have mentored me with incredible patience and caring. I am so appreciative that you took me into your lab, that you gave me the freedom to explore some of my own interests, and that you challenged me every day to be a better scientific thinker. I am forever grateful for your unconditional support and advice throughout graduate school.

I would also like to thank my committee members, Phil Cole and Sean Taverna, for their guidance and insight with my projects. Also, my undergraduate advisor, Scott Holmes, deserves a special thank-you since he fueled my interest in science in the first place, and I would not be in graduate school without his mentorship and support.

My fellow lab members and colleagues in the department have also been a really important part of my graduate experience. In particular, I would like to thank former post-docs in the lab Ajit Datta, Chris Berndsen and Reuven Wiener for their advice and guidance.

Finally, I would like to thank my family for their love and encouragement throughout this process. My husband, Andrew Moreno, has been a pillar of support, as have been my parents and sisters.
Table of Contents

Abstract ............................................................................................................................. ii
Acknowledgements ....................................................................................................... iv
List of Figures ................................................................................................................ vii

Chapter 1: Chromatin and the histone code .............................................................. 1
1.1 Introduction ............................................................................................................. 1
1.2 The “histone code” hypothesis ............................................................................... 3
1.3 Transcriptional co-activators and the SAGA complex ......................................... 7
1.4 Enzymatic activities of the SAGA complex .......................................................... 10
1.5 Lysine deacylation by sirtuin enzymes .................................................................. 15
1.6 Objectives ............................................................................................................... 19

Chapter 2: Sgf29, a subunit of the SAGA HAT module, promotes processive acetylation of histone tails ................................................ 22
2.1 Abstract .................................................................................................................. 22
2.2 Introduction ............................................................................................................. 22
2.3 Methods ................................................................................................................. 26
   Cloning, Expression, and Purification of Proteins .................................................... 26
   Enzyme-mediated ligation of methylated peptide to histone H3 .............................. 30
   MALDI-TOF mass spectrometry ......................................................................... 32
   Fluorescent labeling of histones and nucleosomes ............................................. 32
   Immunoprecipitation assays ............................................................................. 34
   Steady-state acetylation kinetics ...................................................................... 34
   Acid urea gel electrophoresis assay of acetylation ............................................ 36
   Differential labeling assay ................................................................................ 38
   Processivity assay ............................................................................................ 39
2.4 Results .................................................................................................................... 40
   H3K4me3 modestly stimulates acetylation by the HAT module ......................... 40
   Differential labeling technique for studying mixtures of post-translationally modified substrates ............................................................. 53
   Gcn5 preferentially acetylates H3K4me3 nucleosomes .................................. 56
   Sgf29 promotes processive acetylation by the HAT module ................................ 60
2.5 Discussion .............................................................................................................. 64

Chapter 3: Contribution of chromatin-binding domains to SAGA HAT module activity ................................................. 70
3.1 Introduction ............................................................................................................. 70
3.2 Methods .................................................................................................................. 71
   Plasmid constructs for expression in E. coli ......................................................... 71
Table 3.1 List of plasmids containing HAT module domain truncations for bacterial expression .................................................. 72
   Protein expression and purification ................................................................. 73
   Immunoprecipitation assays ............................................................................ 77
   Nucleosome binding assays ........................................................................... 77
<table>
<thead>
<tr>
<th>Chapter 4: Structures of the catalytic domain of human Gcn5L2 bound to propionyl-CoA and butyryl-CoA</th>
<th>91</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Abstract</td>
<td>91</td>
</tr>
<tr>
<td>4.2 Introduction</td>
<td>91</td>
</tr>
<tr>
<td>4.2 Methods</td>
<td>93</td>
</tr>
<tr>
<td>Protein expression and purification</td>
<td>93</td>
</tr>
<tr>
<td>Acyltransferase assay</td>
<td>93</td>
</tr>
<tr>
<td>HAT Domain Crystallization</td>
<td>94</td>
</tr>
<tr>
<td>Data collection and processing</td>
<td>95</td>
</tr>
<tr>
<td>4.3 Results and discussion</td>
<td>96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 5: HAT Module Crystallization Trials</th>
<th>105</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>105</td>
</tr>
<tr>
<td>5.2 Summary of Crystallization Trials</td>
<td>105</td>
</tr>
<tr>
<td>Crystallization trials of the S. cerevisiae HAT Module</td>
<td>105</td>
</tr>
<tr>
<td>Crystallization trials of the pombe HAT module</td>
<td>106</td>
</tr>
<tr>
<td>Crystallization trials of the Gcn5/Ada2 complex</td>
<td>108</td>
</tr>
<tr>
<td>Crystallization trials of the Ada3/Sgf29 complex</td>
<td>114</td>
</tr>
<tr>
<td>4.3 Future Directions</td>
<td>117</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 6: Alternate deacylating specificites for the archaeal enzymes Sir2Af1 and Sir2Af2</th>
<th>119</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Abstract</td>
<td>119</td>
</tr>
<tr>
<td>6.2 Introduction</td>
<td>120</td>
</tr>
<tr>
<td>6.3 Materials and Methods</td>
<td>126</td>
</tr>
<tr>
<td>Protein expression and purification</td>
<td>126</td>
</tr>
<tr>
<td>Enzymatic activity assays</td>
<td>128</td>
</tr>
<tr>
<td>6.4 Results</td>
<td>132</td>
</tr>
<tr>
<td>Structure-based prediction of acyl chain specificity</td>
<td>132</td>
</tr>
<tr>
<td>Alternate deacylating activities of bacterial and archaeal sirtuins</td>
<td>136</td>
</tr>
<tr>
<td>Crystal structure of Sir2Af1 bound to a succinylated peptide</td>
<td>140</td>
</tr>
<tr>
<td>Crystal structure of Sir2Af2 bound to a myristoylated peptide</td>
<td>143</td>
</tr>
<tr>
<td>6.5 Discussion</td>
<td>146</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 7: Supplementary Information</th>
<th>153</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section A: Solution studies of the SAGA DUB module</td>
<td>153</td>
</tr>
<tr>
<td>Section B: Purification of the intact SAGA complex</td>
<td>161</td>
</tr>
<tr>
<td>Section C: A New RING Tossed into an Old HAT</td>
<td>169</td>
</tr>
<tr>
<td>Structure Preview</td>
<td>169</td>
</tr>
</tbody>
</table>

References | 176 |
List of Figures

Figure 1.1 Histone PTMs found at gene promoters in yeast ........................................ 3
Figure 1.2 Mechanisms employed by “readers” of the histone code ............................ 6
Figure 1.3 Modular organization of the SAGA complex ............................................ 9
Figure 1.4 Structure of the SAGA DUB module ....................................................... 12
Figure 1.5 Interactions between HAT module subunits .............................................. 15
Figure 2.1 Fluorophores for differential labeling do not exhibit spectral overlap .......... 39
Figure 2.2 Purified HAT module constructs ............................................................. 42
Figure 2.3 The presence of H3K4me3 increases overall acetylation by the HAT module on peptides ............................................................... 43
Figure 2.4 Producing site-specifically modified histone H3 ........................................ 47
Figure 2.5 MALDI-TOF analysis of H3K4me3 histone ............................................ 48
Figure 2.6 The presence of H3K4me3 increases overall acetylation by the HAT module on nucleosomes ................................................................. 52
Figure 2.7 Overview of differential labeling experiment .......................................... 55
Figure 2.8 The HAT module preferentially acetylates H3K4me3 nucleosomes out of a mixture .......................................................... 58
Figure 2.9 H3K4 trimethylation does not stimulate HAT activity in trans ............... 60
Figure 2.10 Sgf29 promotes processive acetylation against H3K4me3 nucleosomes .......... 63
Figure 2.11 The bromodomain of Gcn5 facilitates acetylation on nucleosomal substrates .................................................................................... 64
Figure 2.12 Model depicting PTM-dependent acetylation by SAGA ......................... 69
Figure 3.1 Chromatin-interacting domains within the SAGA HAT module ............... 71
Figure 3.2 Contribution of chromatin-interacting domains to HAT module integrity and activity ........................................................................................................ 79
Figure 3.3 Connectivity between HAT module subunits .......................................... 83
Figure 3.4 The Sgf29 tandem Tudor and Ada2 SWIRM domains contribute to nucleosome binding .................................................................................. 86
Figure 3.5 The Ada2 SWIRM domain does not bind to nucleosomes ...................... 88
Figure 3.6 The interaction between the HAT module and nucleosomes does not require the H3 tail ............................................................ 89
Figure 4.1 Acyltransferase activity of hsGcn5L2 ...................................................... 97
Figure 4.2 Simulated annealing omit maps of the propionyl and butyryl groups .................................................................................................................. 100
Figure 4.3 Close-up view of the hsGcn5L2 active site .............................................. 101
Figure 4.4 Model of ternary complex between hsGcn5L2, different acyl-CoA molecules, and peptide ............................................................. 103
Figure 5.1 SDS-PAGE analysis of S. cerevisiae HAT module .................................. 106
Figure 5.2 Summary of HAT module complexes subjected to crystallization trials ........................................................................................................ 107
Figure 5.3 Summary of Ada2 ZnF-SANT/ Gcn5 complexes for crystallography ........ 110
Figure 5.4 Nickel-NTA pull-down of Ada2 ZnF-SANT/ Gcn5 complexes ............... 111
Figure 6.1 Summary of the sirtuin-mediated deacylation reaction and chemical structures of various lysine acyl modifications found in cells. ................................. 122
Figure 6.2 Structure- and sequence-guided predictions of acyl chain specificity ................................................................................................................... 135
Figure 6.3 Comparison of deacylation activities .................................................. 138
Figure 6.4 Structure of Sir2Af1 bound to a succinylated peptide ......................... 141
Figure 6.5 Structure of Sir2Af2 bound to a myristoylated peptide ...................... 144
Figure 6.6 Sequence conservation between sirtuins that remove long acyl chains ................................................................................................................... 147
Figure 7.1 Steady-state kinetics for DUBm variants ........................................... 154
Figure 7.2 Sedimentation velocity of different DUBm constructs ...................... 157
Figure 7.3 Sedimentation equilibrium analysis of DUBm variants .................... 159
Figure 7.4 Example of TAP-tag purifications for intact SAGA ......................... 168
Figure 7.5 Domain structure of p300 .................................................................. 171
Figure 7.6 Structure of p300 core fragment ....................................................... 173

List of Tables

Table 2.1 Steady-state kinetic parameters using peptide substrates ................. 44
Table 2.2 Steady-state kinetic parameters using acetyl-CoA as a substrate ........ 44
Table 2.3 Steady-state kinetic parameters using nucleosomal substrates ........ 45
Table 3.1 List of plasmids containing HAT module domain truncations for bacterial expression ................................................................................................................... 72
Table 3.2 Steady-state kinetic parameters using peptide substrates and various HAT module domain truncations ................................................................. 80
Table 3.3 Steady-state kinetic parameters for Gcn5 on peptide substrates ....... 84
Table 3.4 Steady-state kinetic parameters for Gcn5 using acetyl-CoA as a substrate ..................................................................................................................... 84
Table 3.5 Steady-state kinetic parameters for ΔSWIRM and ΔTudor HAT module on nucleosomal substrates ................................................................. 87
Table 4.1 Summary of data collection and refinement statistics ....................... 99
Table 5.1 Constructs for crystallizing the Ada3/Sgf29 sub-complex .................. 115
Table 6.1 Summary of refinement statistics for sirtuin structures ................... 142
Table 7.1 Steady-state kinetic constants for DUBm variants ............................ 155
Table 7.2 DUBm Dimer dissociation constants ............................................... 160
Chapter 1: Chromatin and the histone code

1.1 Introduction

Gene expression in eukaryotes is governed by chromatin, which organizes and compacts DNA within the nucleus (Kornberg & Lorch, 1999). Nucleosomes are the fundamental unit of chromatin, and are composed of two copies each of the histone proteins H2A, H2B, H3, and H4, around which are wound 147 base pairs of DNA (Kornberg & Lorch, 1999; Luger et al, 1997). All four histones contain a globular domain responsible for forming the histone octamer, as well as long unstructured N-terminal tails that are extensively post-translationally modified in vivo (Luger et al, 1997). The majority of the eukaryotic genome is packaged into nucleosomes (Segal & Widom, 2009), which compact DNA to fit into a limited nuclear volume by assembling into chromatin fibers (Woodcock & Ghosh, 2010). However, such high levels of compaction obstruct essential biological processes that require access to DNA, such as gene transcription (Li et al, 2007) and repairing DNA damage (Price & D'Andrea, 2013). To deal with the opposing requirements for compaction and accessibility, cells have evolved a number of mechanisms that modulate chromatin condensation across the genome.

Histone post-translational modifications represent one mechanism that cells employ to facilitate dynamic processes on chromatin. Examples of the various kinds of chemical modifications found on histones include lysine acylation, lysine/arginine methylation, serine/threonine phosphorylation, and lysine ubiquitination (Rothbart & Strahl, 2014). These modifications regulate chromatin accessibility by changing the intrinsic thermodynamic properties of...
nucleosomes, such as their overall stability and propensity to self-associate (Bowman & Poirier, 2014; Zentner & Henikoff, 2013), and also serve as docking sites for effector proteins that regulate downstream signaling events (Musselman et al, 2012; Patel & Wang, 2013). Consequently, enzymes that alter histone post-translational modifications play an important role in gene regulation by making nucleosomal DNA more accessible to transcriptional machinery.

Combinations of histone post-translational modifications divide the genome into two general chromatin environments that differ from one another in overall condensation and transcriptional output. Heterochromatin is transcriptionally repressed, hypoacetylated, and enriched for repressive marks such as H3K9 methylation (Kouzarides, 2007). Euchromatic regions are transcriptionally permissive and contain low levels of H3K4 di- and trimethylation (H3K4me2/3), H3/H4 acetylation (H3ac/H4ac), H3K36 trimethylation (H3K36me3), and H3K79 trimethylation (H3K79me3) (Kouzarides, 2007). Local patterns of histone PTMs further distinguish actively transcribed genes within euchromatin from unexpressed regions. For example, H3K4me3 and H3/H4 hyperacetylation are enriched at the promoters of actively transcribed genes in yeast but not in gene bodies (Figure 1.1), and the relative levels of these modifications correlate with higher levels of mRNA synthesis (Bernstein et al, 2002; Kurdistani et al, 2004; Pokholok et al, 2005).

In many cases, the way in which chromatin domains containing particular combinations of histone post-translational modifications are established and the mechanisms that couple the regulation of multiple post-translational modifications
remain poorly understood. The number of chemical groups and modification sites discovered on histones has only continued to grow in recent years (Tan et al, 2011), highlighting how much more there is to learn about the interplay between chromatin environment and nuclear signaling pathways.

**Figure 1.1 Histone PTMs found at gene promoters in yeast**

### 1.2 The “histone code” hypothesis

The “histone code” hypothesis proposes that distinct patterns of histone PTMs form a combinatorial language interpreted by effector proteins that direct specific biological outcomes (Jenuwein & Allis, 2001; Strahl & Allis, 2000). Virtually every nuclear process that involves DNA, such as transcription, replication, recombination, repair, and condensation, is regulated by post-translational modifications on chromatin (Bannister & Kouzarides, 2011; Kouzarides, 2007). The terms “writers”, “erasers”, and “readers” of the histone code have emerged as nomenclature to describe the enzymes that apply and remove post-translational modifications from histones, as well as the domains that recognize them.
Histone crosstalk refers to the coordinate regulation of histone post-translational modifications, and there are many different ways that unique spatial or temporal arrangements of histone marks can be established along chromatin. For example, combinations of post-translational modifications may exist on the same histone tail (Figure 1.2A), on different tails within the same nucleosome (Figure 1.2B), or on different tails within different nucleosomes (Figure 1.2C). Alternatively, existing patterns of histone marks may direct the deposition of new modifications (Figure 1.2D), thus regulating the order in which different modifications appear on chromatin. Effector modules, or “reader” domains, have evolved that recognize combinatorial patterns of histone post-translational modifications across the genome and initiate particular functional outcomes. Crosstalk between different histone modifications, mediated through reader domains, has emerged as an important and widespread principle directing chromatin-templated processes in eukaryotes.

Over the past ten years, many new reader domains have been discovered, and the molecular strategies employed to distinguish different post-translational modifications have been studied extensively (Taverna et al, 2007a). Some reader domains recognize a single modified residue in a particular sequence context (Figure 1.2E). For example Sgf29, a subunit of the SAGA transcriptional co-activator, recognizes H3K4 trimethyl lysine through its tandem Tudor domain (Bian et al, 2011). A crystal structure of the isolated tandem Tudor domain in complex with an H3K4me3 peptide revealed that the trimethyl lysine (H3K4me3) and the N-terminus of the peptide (H3A1) interact with two negatively
charged pockets on Sgf29, explaining why Sgf29 is so selective for the H3K4 position (Bian et al, 2011). Other reader domains are less sequence-specific, with the modified residue forming a large proportion of the interaction surface (Figure 1.2F). For example, the Gcn5 bromodomain binds to acetylated peptides with a slight preference for hydrophobic and positively charged residues at the +2 and +3 positions, respectively, but otherwise makes few contacts except with the acetyl lysine (Hudson et al, 2000; Owen et al, 2000). Examples where single reader domains bind to multiple histone modifications have also been discovered (Figure 1.2G). The bromodomain of the chromatin-associated protein Brdt binds to diacetylated H4K5ac/H4K8ac peptides with high affinity using a broad pocket that can accommodate both acetyl lysine residues (Moriniere et al, 2009).

Instead of binding two modifications by one domain, linked binding modules within the same polypeptide can also provide a combinatorial read-out of histone modifications (Figure 1.2H). The tandem PHD (plant homeodomain) fingers in MOZ, an acetyltransferase found in vertebrates, recognize peptides harboring unmodified H3R2 and acetylated H3K14 (H3R2/H3K14ac) with high affinity, and the strength of this interaction is significantly diminished by either methylating the arginine or deacetylating the lysine (Qiu et al, 2012). Lastly, multi-protein complexes containing several subunits with reader domains can recognize combinations of histone marks (Figure 1.2I). In these ways, individual proteins and protein complexes possessing reader domains mediate crosstalk between different histone modifications, such that multivalent recognition of histone
modifications can provide high levels of specificity for particular chromosomal regions (Musselman et al, 2012).

**Figure 1.2** Mechanisms employed by “readers” of the histone code.

(A) Read-out of two post-translational modifications on the same histone tail.

Multivalent engagement of two modifications by linked binding modules on (B) different tails within the same nucleosome or (C) histone tails on different nucleosomes. (D) Crosstalk where the deposition of one modification requires the presence of another. (E,F) Examples where a single “reader” domains bind to a single modification. (G) Example where a single “reader” domain binds to two modifications on the same histone tail. (H) Example where two “reader” domains exist within the same polypeptide and engage two modifications on the same histone tail. (I) Macromolecular complexes may contain multiple subunits with “reader” domains.
1.3 Transcriptional co-activators and the SAGA complex

Co-activators that remodel nucleosomes and post-translationally modify histone tails are an important component of the transcriptional machinery in eukaryotes (Li et al, 2007). The SAGA (Spt-Ada-Gcn5 Acetyltransferase) complex is a multi-protein transcriptional co-activator that affects the expression of nearly all yeast genes (Bonnet et al, 2014), and is particularly important for activating transcription in response to environmental stress (Huisenga & Pugh, 2004). Since its discovery in yeast (Grant et al, 1997), homologues of most SAGA subunits have been identified in eukaryotes ranging from insects to humans (Nagy & Tora, 2007; Spedale et al, 2012), pointing to the central role that SAGA plays in transcriptional regulation. SAGA is targeted to the upstream activating sequences of transcribed genes through interactions with acidic transcriptional activators like Gal4 and Gcn4 (Bhaumik & Green, 2001; Larschan & Winston, 2001). There, SAGA coordinates a number of processes involved in transcription initiation, such as recruiting TATA-binding protein and chromatin remodelers to gene promoters (Bhaumik & Green, 2002; Dudley et al, 1999; Hassan et al, 2001), facilitating pre-initiation complex assembly (Bhaumik & Green, 2001), acetylating histone tails (Vignali et al, 2000), and deubiquitinating histone H2B (Daniel et al, 2004; Henry et al, 2003). Recent studies have linked SAGA-associated activities to coding regions as well, where acetylation by Gcn5 promotes nucleosome eviction (Govind et al, 2007) and spliceosome assembly (Gunderson & Johnson, 2009), while H2B deubiquitination by Ubp8 allows RNA
polymerase II to transition to productive elongation (Daniel et al, 2004; Henry et al, 2003; Wyce et al, 2007). In addition, SAGA couples transcription to mRNA export from the nucleus through its Sus1 subunit (Kohler et al, 2006; Pascual-García et al, 2008; Rodriguez-Navarro et al, 2004). Genetic studies have also linked Sus1 to cytoplasmic mRNA degradation pathways (Cuenca-Bono et al, 2010). With roles in transcription initiation, transcription elongation, mRNA export, and mRNA degradation, the SAGA complex touches nearly every aspect of gene regulation.

SAGA is composed of 19-21 protein subunits organized into four structurally independent modules (Figure 1.3). Each module is responsible for different subsets of SAGA-associated functions (Koutelou et al, 2010; Lee et al, 2011; Weake & Workman, 2012), with the coordinated action of these modules at gene promoters required for transcription activation (Bhaumik, 2011). The TAF (TBP-Associated Factor) and SPT modules form the non-catalytic core of the SAGA complex; these modules recruit SAGA to gene promoters, interact with TATA-binding protein (TBP), and facilitate pre-initiation complex assembly (Koutelou et al, 2010; Lee et al, 2011; Weake & Workman, 2012). The other two modules perform catalytic functions: the DUB (DeUBiquitinating) module deubiquitinates histone H2B and the HAT (Histone AcetylTransferase) module acetylates the N-terminal tails of histones H3 and H2B (Koutelou et al, 2010). Because of the modular organization of the SAGA complex, deleting individual subunits can have limited effects on overall complex integrity. A systematic mass spectrometry study looking at the composition of purified SAGA from yeast
strains lacking other SAGA subunits showed: (i) that individual modules can be disrupted while preserving the integrity of the rest of the complex and (ii) that deleting core subunits can cause intact modules to dissociate from SAGA (Lee et al, 2011). The modular architecture of the SAGA complex explains why single null mutations compromising the overall structural integrity of SAGA (such as individually mutating \textit{spt20\Delta} or \textit{spt7\Delta}) phenocopy double mutations affecting multiple modules (such as \textit{gcn5\Delta/spt8\Delta} or \textit{gcn5\Delta/ubp8\Delta}), while null mutations affecting single modules (such as individually mutating \textit{gcn5\Delta}, \textit{spt8\Delta}, or \textit{ubp8\Delta}) cause phenotypes that are comparatively less severe (Henry et al, 2003; Sterner et al, 1999).

\textbf{Figure 1.3} \textit{Modular organization of the SAGA complex}
Whether these modules exist separately from SAGA \textit{in vivo} is not known, but it is clear that many SAGA-regulated genes are differentially sensitive to subunit deletions. Distinct gene expression defects are observed in budding yeast strains harboring \textit{spt3}Δ, \textit{spt20}Δ, and \textit{gcn5}Δ null mutations, suggesting that each module regulates separate sets of SAGA-dependent genes (Lee et al, 2000). This result is supported by studies in fission yeast, where microarray analysis for thirteen SAGA deletion mutants revealed five distinct but overlapping classes of gene expression patterns (Helmlinger et al, 2011). In addition, the deletion mutants displayed a range of growth phenotypes when challenged with 37 kinds of environmental stress that require the expression of SAGA-regulated genes for optimal survival (Helmlinger et al, 2011). Since different transcriptional programs must be activated in response to each treatment, these data support module-specific functions for the SAGA complex. Independent recruitment of different SAGA modules has been observed directly at the \textit{GAL1} locus, where members of the HAT and DUB modules (Gcn5 and Ubp8) transit into the coding region while the Spt8 subunit stays at the promoter (Wyce et al, 2007). However, evidence for dissociation of modules \textit{in vivo} is limited. Further studies will be required to settle whether detached SAGA modules exist in a biological context.

\section*{1.4 Enzymatic activities of the SAGA complex}

One common feature of the DUB and HAT modules is that the activity and substrate specificity of the catalytic subunit changes depending on its binding
partners. In yeast, the ubiquitin-specific protease Ubp8 is the catalytic subunit of the DUB module, which deubiquitinates histone H2B (Daniel et al, 2004; Henry et al, 2003). Ubp8 is catalytically inactive on its own, but becomes a robust deubiquitinating enzyme by binding to three other DUB module subunits: Sgf73, Sgf11, and Sus1 (Kohler et al, 2008). These four proteins are sufficient to reconstitute the DUB activity of SAGA, and have been extensively characterized both biochemically and structurally.

Crystallographic studies showed that the four DUB module subunits form a highly intertwined structure held together by eight structural zinc atoms, where each protein depends on the others for its overall fold (Figure 1.4A) (Kohler et al, 2010; Samara et al, 2010). Ubp8 is organized into two globular domains stapled together by the N-terminus of Sgf73 (Figure 1.4B), where the “catalytic lobe” contains a conserved USP domain and the “assembly lobe” contains a Zinc Finger-Ubiquitin Binding Protein (ZnF-UBP) domain (Samara & Wolberger, 2011). With the exception of Ubp8, the other DUB module subunits adopt highly extended non-globular conformations. Along with Sus1, the N-terminus of Sgf11 forms the core of the “assembly lobe”, while the C-terminus of Sgf11 contains an arginine-rich zinc finger that contacts the USP domain of Ubp8 (Figure 1.4C) (Kohler et al, 2010; Samara et al, 2010). The Sgf11 zinc finger maintains the active site of Ubp8 in an active conformation, since complexes lacking this domain are catalytically impaired (Kohler et al, 2010) and form dramatic domain-swapped dimers (Samara et al, 2012). Domain-swapping misaligns the active site residues on one Ubp8 monomer, and mutations that promote the monomeric
form of the enzyme partially restore DUB module activity (Samara et al, 2012). Beyond its role in activating Ubp8, the Sgf11 zinc finger also participates in substrate recognition. Mutating conserved arginine residues in the Sgf11 zinc finger causes overall levels of H2B ubiquitination to increase \textit{in vivo} without affecting catalysis on model substrates \textit{in vitro} (Kohler et al, 2010). Since Ubp8 is constitutively activated as part of the DUB module, recruitment to SAGA-regulated genes likely results in rapid and complete local deubiquitination of H2B.

\textbf{Figure 1.4} Structure of the SAGA DUB module

(A) Cartoon representation of the DUB module. (B) Surface representation of Ubp8. (C) Surface representation of Sgf73, Sgf11, and Sus1. PDB ID – 3MHS.
Gcn5 is the catalytic subunit of the SAGA Histone AcetylTransferase (HAT) module, which is a four-protein complex that acetylates lysine residues on the N-terminal tails of histones H3, H4, and H2B (Grant et al, 1997; Grant et al, 1999; Kuo et al, 1996; Saleh et al, 1997; Suka et al, 2001). Whereas Gcn5 is not essential in yeast, flies and mice lacking Gcn5 die at larval and embryonic stages, respectively (Carre et al, 2005; Xu et al, 2000). In addition to Gcn5, the SAGA HAT module contains two transcriptional adaptor proteins, called Ada2 and Ada3 (Candau & Berger, 1996; Horiuchi et al, 1995; Marcus et al, 1994), as well as Sgf29 (Lee et al, 2011; Sanders et al, 2002). The four HAT module subunits contain a number of chromatin-interacting domains: Gcn5 possesses a bromodomain that binds to acetyl lysine (Dhalluin et al, 1999; Hudson et al, 2000; Owen et al, 2000), Ada2 contains a zinc finger, SANT, and SWIRM domains, and Sgf29 contains a tandem Tudor domain. The presence of reader domains within the HAT module raises the possibility of histone crosstalk, where patterns of either acetylation (mediated by the Gcn5 bromodomain) or methylation (mediated by the Sgf29 tandem Tudor domain) alter the enzymatic activity of Gcn5.

Although high-resolution structural information is only available for the catalytic core of Gcn5 (Poux et al, 2002; Schuetz et al, 2007; Trievel et al, 1999), bromodomain of Gcn5 (Hudson et al, 2000; Owen et al, 2000), and isolated Sgf29 tandem Tudor domain (Bian et al, 2011), biochemical studies have demonstrated that the non-catalytic HAT module subunits modulate the activity and lysine specificity of Gcn5. Recombinant Gcn5 robustly acetylates free histones but not nucleosomes (Grant et al, 1997; Grant et al, 1999). Binding to
Ada2 stimulates overall Gcn5 activity (Balasubramanian et al, 2002; Boyer et al, 2002; Grant et al, 1997), but only the heterotrimeric complex consisting of Gcn5/Ada2/Ada3 efficiently targets nucleosomal substrates (Balasubramanian et al, 2002). Interestingly, the lysine specificity of Gcn5 also changes in the context of other HAT module subunits. Whereas Gcn5 and Gcn5/Ada2 vastly prefer acetylating histone H3 at the H3K14 position (Grant et al, 1999; Kuo & Andrews, 2013), Gcn5/Ada2/Ada3 targets multiple sites on histone H3, with lysine specificity resembling that of intact SAGA (Balasubramanian et al, 2002).

The structural basis for the way in which the non-catalytic subunits of the HAT module regulate Gcn5 activity has yet to be determined. The connectivity between HAT module subunits is linear, meaning that pairwise interactions have been identified between Sgf29/Ada3 (Ito et al, 2001; Kurabe et al, 2007), Ada3/Ada2 (Benecke et al, 2002; Candau & Berger, 1996; Horiuchi et al, 1995; Ito et al, 2001), and Ada2/Gcn5 (Candau & Berger, 1996; Candau et al, 1997; Horiuchi et al, 1995; Marcus et al, 1994), but not between any other combinations of subunits (Figure 1.4). Beyond its structural role bridging Gcn5 and Ada3 together, Ada2 also tethers the HAT module to the rest of the SAGA complex. Cross-linking studies on intact SAGA have shown that Ada2 makes the greatest number of contacts outside the HAT module (Han et al, 2014), and purifying SAGA from an ada2Δ null mutant results in a complex lacking all four HAT module subunits (Lee et al, 2011). Until a high-resolution structure is determined of the HAT module complex, how Ada2, Ada3, and Sgf29 regulate the acetyltransferase of Gcn5 will remain an open question.
1.5 Lysine deacylation by sirtuin enzymes

Sirtuins are a broadly conserved class of enzymes that catalyze the NAD$^+$-dependent removal of acyl modifications from lysine residues (Imai et al., 2000; Landry et al., 2000; Smith et al., 2000; Tanner et al., 2000a; Tanny & Moazed, 2001). Organisms from all three domains of life encode for sirtuins, suggesting that the sirtuin-catalyzed reaction represents a fundamental and ancient cellular activity (Frye, 2000; Greiss & Gartner, 2009). So far, sirtuin homologues have been identified in the majority of archaea and bacteria and in all eukaryotes (Frye, 2000; Greiss & Gartner, 2009), where in many cases they have been duplicated and allowed to diverge in sequence, function, and sub-cellular localization (Michan & Sinclair, 2007). Many critical cellular processes are regulated by sirtuins. For example, yeast Sir2 regulates transcriptional silencing...
by deacetylating histones and promoting the formation of heterochromatin (Braunstein et al, 1993). Human cells contain seven sirtuin paralogues that regulate a wide variety of processes, many tissue-specific, such as managing oxidative stress, regulating fatty acid oxidation, maintaining genome stability, and altering gene transcription by deacetylating transcription factors (Michan & Sinclair, 2007; Vassilopoulos et al, 2011). Although the biological roles played by sirtuins in bacteria are less well understood, examples where lysine deacetylation affects cellular functions continue to emerge. For example, deacetylation by the bacterial sirtuin CobB activates the metabolic enzyme acetyl coenzyme-A synthetase (ACS) in Salmonella enterica (Starai et al, 2002) and affects the activity of N-hydroxyarylamine O-acetyltransferase (NhoA) in E. coli (Zhang et al, 2013b). CobB has also been linked to signal transduction and transcriptional regulation in E. coli, where it regulates chemotaxis by deacetylating the response regulator CheY (Li et al, 2010), modulates transcription from the stress-responsive promoter cpxP by deacetylating the α-subunit of RNA polymerase (Lima et al, 2011), and promotes transcription from the RcsC regulon by deacetylating the transcription factor RcsB (Hu et al, 2013; Thao et al, 2010). As of yet archaeal sirtuins have no known biological targets, but future studies on archaea may reveal cellular pathways that involve these enzymes.

The distribution of lysine acetylation on a proteome-wide scale has been studied in a growing number of organisms ranging in complexity from bacteria and single-celled eukaryotes to human cell lines and rat tissue samples (Choudhary et al, 2009; Crosby et al, 2012; Finkemeier et al, 2011; Henriksen et
al, 2012; Jeffers & Sullivan, 2012; Kim et al, 2013; Kim et al, 2006; Lundby et al, 2012; Miao et al, 2013; Okanishi et al, 2013; Weinert et al, 2011; Wu et al, 2011; Wu et al, 2013; Zhang et al, 2009a; Zhang et al, 2013a). In all cases the following two patterns have emerged: first that lysine acetylation is highly abundant, as each study identified on the order of hundreds or thousands of unique acetylation sites, and second that acetylation is enriched on proteins involved in central metabolic pathways. Acetyl groups, however, are not the only type of acyl chain that modifies lysine residues in vivo. The same advances in mass spectrometry that enabled large-scale characterization of the lysine acetylome, such as specific enrichment for acetylated proteins combined with high-resolution fractionation techniques and sensitive mass spectrometers (Jensen, 2006; Olsen & Mann, 2013), have also facilitated the discovery of a number of new acyl modifiers that differ in chain length and overall charge from acetylation. Studies in eukaryotes have identified modifications like propionylation (Chen et al, 2007), butyrylation (Chen et al, 2007), crotonylation (Tan et al, 2011), malonylation (Peng et al, 2011; Xie et al, 2012), succinylation (Park et al, 2013; Xie et al, 2012; Zhang et al, 2011), and glutarylation (Tan et al, 2014) decorating lysine residues. Lysine succinylation has also been detected in E. coli (Colak et al, 2013; Weinert et al, 2013), suggesting that at least some of these acyl modifications are conserved across evolution. In fact, the bacterial enzyme propionyl-CoA synthetase (PrpE) from S. enterica is propionylated in vivo, which regulates PrpE catalytic activity (Garrity et al, 2007). Since all other lysine post-translational modifications are reversible, it would stand to reason that enzymes
catalyzing lysine deacylation also exist. Accordingly, reports of sirtuins that are more active in vitro against longer or charged acyl chains have paralleled their discovery in cells.

All members of the sirtuin family share a conserved catalytic core of ~250 amino acids, and structures for a number of sirtuins have been determined that reveal the mechanisms sirtuins use for substrate selection and catalysis. The sirtuin catalytic core is organized into two domains. The larger of the two contains a Rossman fold responsible for binding NAD+. The smaller domain is composed of a zinc-binding module and a helical module, which are each inserted into the primary sequence of the Rossman fold (Sanders et al, 2010; Yuan & Marmorstein, 2012). Four loops connect the two domains, forming a hydrophobic cleft where acetylated peptides bind (Avalos et al, 2002; Zhao et al, 2003; Zhao et al, 2004). Apart from the contacts made with the acetylated residue itself, sirtuin enzymes make few specific contacts with the amino acids that surround the modified lysine (Avalos et al, 2002; Zhao et al, 2003; Zhao et al, 2004), although the identity of the -1 and +2 peptide positions matter to an extent (Cosgrove et al, 2006; Garske & Denu, 2005). Instead, the peptide hydrogen bonds with a β-sheet near the active site forming an additional β-strand through backbone interactions (Avalos et al, 2002; Zhao et al, 2003; Zhao et al, 2004). Therefore, the intrinsic flexibility of the peptide factors into substrate selection (Khan & Lewis, 2005), and sequence context is only one of several physical properties that sirtuins exploit for substrate discrimination.
Acyl chain identity is emerging as another determinant of sirtuin specificity, which is supported by both structural and biochemical studies. Several enzymes that historically exhibited weak deacetylation activity in vitro will robustly deacylate lysine residues with a different modification. For example, the demalonylating, desuccinylating, and deglutarylating activities of human SIRT5 is much stronger than its deacetylase activity (Du et al., 2011; Peng et al., 2011; Tan et al., 2014), and a crystal structure of SIRT5 bound to a succinylated peptide has identified sequence elements that confer specificity for negatively charged modifications (Du et al., 2011). Sirtuins that remove long acyl chains from lysine have also been discovered. Human SIRT6 preferentially removes long chain fatty acids from lysine compared to acetyl groups (Feldman et al., 2013; Hong et al., 2013), a specificity mirrored by PfSir2A, a sirtuin from the malaria parasite *P. falciparum* (Zhu et al., 2011). Both enzymes have been crystallized in complex with myristoylated peptides (Hong et al., 2013; Zhu et al., 2011) revealing a hydrophobic tunnel that accommodates the long acyl chain. Sirtuins are central regulators of many different kinds of cellular processes, and biological roles for these new alternate deacylating activities continue to emerge.

1.6 Objectives

The connection between acetylation and transcription was first proposed over fifty years ago (Allfrey et al., 1964), and since then enzymes that affect acetylation have emerged as a central regulators of gene expression (Brown et al., 2000). This thesis focuses on two classes of enzymes, histone
acetyltransferases and sirtuins, which catalyze opposing reactions by depositing and removing lysine acyl modifications, respectively. Prior to this work, it was known that Gcn5, the acetyltransferase associated with the SAGA transcriptional co-activator (Koutelou et al, 2010), forms a complex with three additional subunits: Ada2, Ada3, and Sgf29 (Balasubramanian et al, 2002; Lee et al, 2011), which are collectively referred to as the HAT module. Sgf29 and Gcn5 possess “reader” domains that recognize H3K4 trimethylation (Bian et al, 2011; Vermeulen et al, 2010) and acetylated lysine residues (Dhalluin et al, 1999; Hudson et al, 2000; Owen et al, 2000), raising the possibility that the HAT module participates in histone crosstalk. The first portion of this thesis addresses how the non-catalytic subunits of the SAGA HAT module regulate acetylation by Gcn5. Chapter 2 identifies a mechanism coupling H3K4me3 recognition by Sgf29 to hyperacetylation by Gcn5 and describes a new color-coding assay where acetylation of individual sub-populations of nucleosomes containing different post-translational modifications can be monitored in a mixture. Chapter 3 describes how the various chromatin interacting domains of the SAGA HAT module contribute to the structural integrity of the complex, participate in nucleosome recognition, and regulate the acetyltransferase activity of Gcn5. Chapter 4 reports structures of the catalytic domain of human Gcn5 (hsGcn5L2) bound to propionyl-CoA and butyryl-CoA, and demonstrates how Gcn5 discriminates between different acyl chains. Chapter 5 summarizes the strategies employed to obtain crystals of the HAT module or sub-complexes within the HAT module, which were ultimately unsuccessful. The second portion of this thesis
focuses on the substrate specificity of four archaeal and bacterial sirtuins, which are NAD\(^+\)-dependent lysine deacetylases. Sir2Af1 and Sir2Af1 are enzymes derived from the archaeabacterium *Archaeoglobus fulgidus*, and were originally used as model deacetylases to study the sirtuin-catalyzed reaction. Chapter 6 describes how a combination of x-ray crystallography and biochemistry were used to redefine the acyl chain specificity of Sir2Af1 and Sir2Af2, and has been published in the journal *Protein Science*. 
Chapter 2: Sgf29, a subunit of the SAGA HAT module, promotes processive acetylation of histone tails

2.1 Abstract

Gcn5 is the histone acetyltransferase subunit of the Spt-Ada-Gcn5 Acetyltransferase (SAGA) complex, a multi-subunit transcriptional co-activator that regulates gene expression in eukaryotes. Although Gcn5 is catalytically active on its own, its specificity and overall activity is regulated by three additional SAGA subunits, Ada2, Ada3, and Sgf29, which form a complex with Gcn5 known as the HAT (Histone AcetylTransferase) module. Sgf29 contains a tandem Tudor domain that binds histone H3 trimethylated at K4 (H3K4me3) and is thought to modulate histone acetylation by Gcn5. We have developed a technique using differential fluorescent labeling to monitor acetylation of individual sub-populations of modified and unmodified nucleosomes in a mixture. We find that the HAT module preferentially acetylates nucleosomes harboring H3K4 trimethylation in a mixture containing excess unmodified nucleosomes. This effect requires the Tudor domain of Sgf29, which promotes processive acetylation of H3K4me3 substrates. These data define a new mechanism for crosstalk between H3K4 trimethylation and acetylation by the SAGA complex.

2.2 Introduction

The different patterns of histone post-translational modifications distributed across the genome constitute a “histone code” that direct downstream events by recruiting effector proteins that orchestrate distinct transcriptional programs (Jenuwein & Allis, 2001; Strahl & Allis, 2000). Crosstalk between
histone modifications, where existing patterns of histone PTMs direct the establishment of new modifications or several modifications are recognized in tandem, has emerged as an important and widespread mechanism regulating chromatin-templated processes (Musselman et al, 2012). The multi-functional complexes that activate transcription typically contain multiple “reader” domains that recognize specific chromatin modifications, along with catalytic subunits that deposit or remove histone modifications (Lalonde et al, 2014; Musselman et al, 2012). The SAGA (Spt-Ada-Gcn5 Acetyltransferase) complex is a multi-protein transcriptional co-activator (Koutelou et al, 2010) that mediates transcription of nearly all yeast genes (Bonnet et al, 2014), and is particularly important for activating transcription in response to environmental stress (Huisinga & Pugh, 2004). SAGA operates downstream of H3K4 trimethylation (H3K4me3) and H2B ubiquitination, which are both present at target genes before SAGA is recruited (Weake & Workman, 2010). SAGA coordinates a number of processes involved in transcription initiation and elongation, including acetylating histone H3 and deubiquitinating histone H2B (Weake & Workman, 2010). Interestingly, overall levels of H3 acetylation in vivo are diminished in the absence of H3K4 trimethylation (Noma & Grewal, 2002), which, coupled with the observation that the SAGA complex contains a “reader” domain that specifically recognizes the H3K4me3 mark (Vermeulen et al, 2010), suggests that SAGA mediates crosstalk between histone methylation and acetylation. The underlying mechanism governing crosstalk between H3K4 trimethylation and acetylation by SAGA, however, remains poorly understood.
The histone acetyltransferase activity of SAGA resides in a distinct four-protein subcomplex known as the HAT module (Balasubramanian et al, 2002; Lee et al, 2011). In addition to its catalytic subunit, Gcn5 (Grant et al, 1997; Grant et al, 1999; Saleh et al, 1997), the SAGA HAT module contains Ada2, Ada3 (Candau & Berger, 1996; Horiuchi et al, 1995) and Sgf29 (Lee et al, 2011), which regulate the overall enzymatic activity and substrate specificity of Gcn5. Whereas recombinant Gcn5 acetylates free histones but not nucleosomes (Balasubramanian et al, 2002; Grant et al, 1997; Grant et al, 1999), a heterotrimeric complex comprising Gcn5/Ada2/Ada3 is active on nucleosomal substrates and has the same lysine specificity as intact SAGA (Balasubramanian et al, 2002). The HAT module also contains domains that recognize specific histone post-translational modifications that may target acetylation to particular chromatin environments. Sgf29 contains a tandem Tudor domain that recognizes H3K4 trimethylation (Bian et al, 2011; Vermeulen et al, 2010), and Gcn5 contains a bromodomain that binds to acetyl lysine (Dhalluin et al, 1999; Hudson et al, 2000; Owen et al, 2000), which are both enriched at the promoters of actively transcribed genes (Pokholok et al, 2005).

The presence of distinct “reader” domains in the HAT module that recognize lysine methylation and acetylation suggests that HAT module activity or recruitment is likely impacted by the presence of these modifications in chromatin templates. The contribution of the Gcn5 bromodomain to SAGA activity has, to date, been best studied. In yeast, deleting the Gcn5 bromodomain reduces growth under stress conditions (Hassan et al, 2002) and decreases
transcription from the inducible *HIS3* promoter (Sterner et al, 1999). *In vitro*, the bromodomain promotes the retention of SAGA on acetylated chromatin (Hassan et al, 2002) and enhances acetyltransferase activity on nucleosomes containing H3K14 or H4K16 acetylation (Li & Shogren-Knaak, 2008; Li & Shogren-Knaak, 2009). A recent study showed that the Gcn5 bromodomain likely facilitates crosstalk between H3K14 and H3K18 acetylation as well, since bromodomain mutations disrupting the ability of Gcn5 to bind to acetylated histones also change the lysine specificity of the HAT module, and Gcn5 does not appreciably catalyze H3K18 acetylation on histones (H3K14R) that cannot be acetylated at the K14 position (Cieniewicz et al, 2014). By contrast, the contribution of H3K4me3 recognition by Sgf29 to HAT module acetylation of histones is not well understood. Although recombinant Sgf29 binds to peptides containing H3K4 trimethylation with micromolar affinity (Bian et al, 2011; Vermeulen et al, 2010), biochemical studies have yet to connect H3K4me3 recognition with differences in Gcn5 enzymatic activity, since purified yeast SAGA acetylates histone peptides with and without H3K4 trimethylation at the same rate (Bian et al, 2011). However, deletion of Sgf29 phenocopies the slow-growth phenotype and resistance to VP16 toxicity of other SAGA subunit deletions in yeast (Lee et al, 2011) and causes a profound decrease in histone acetylation at sites targeted by Gcn5 (Bian et al, 2011). Thus, while multiple lines of evidence suggest that Sgf29 regulates acetylation by Gcn5 *in vivo*, the mechanism connecting H3K4me3 recognition directly to differences in nucleosome acetylation by SAGA is unknown.
To elucidate the crosstalk between H3K4me3 recognition by the Sgf29 tandem Tudor domain and acetylation by Gcn5, we developed a color-coding assay that makes it possible to measure acetylation rates on distinct histone populations within mixtures of modified and unmodified nucleosomes. We find that the HAT module preferentially acetylates H3K4 trimethylated nucleosomes in a mixture and that this preference depends on the Sgf29 tandem Tudor domain. Compared to unmodified nucleosomes, the HAT module is more processive when H3K4me3 is present, thereby explaining how a small increase in specificity translates into a large difference in histone H3 acetylation. These data reveal a mechanism by which different chromatin-interacting domains within the SAGA HAT module facilitate crosstalk between its acetyltransferase activity and other histone post-translational modifications.

2.3 Methods

Cloning, Expression, and Purification of Proteins
SAGA HAT module subunits were PCR-amplified from *S. pombe* genomic DNA using KOD polymerase (EMD Millipore) and cloned into vectors for bacterial expression using the In-Fusion Cloning kit (Clontech). Since *pombe* Sgf29 contains an intron, the two expressed sequences were PCR-amplified using primers containing 15 bp internal overlap and used together in a single In-Fusion reaction. The HAT module subunits were cloned into three vectors, which were compatible for co-expression in *E. coli*. Ada3 was cloned into pET32a, Gcn5/
Gcn5ΔBromomdain was cloned into the first multiple cloning site of CDFduet, Ada2 was cloned into the second multiple cloning site of CDFduet, and Sgf29/ Sgf29ΔTudor was cloned into pRSF. Full-length Sgf29 for pull-down assays was cloned into pET32a. Histone H3 from *X. laevis* containing C110A/ Q125C mutations and/or having the first 32 amino acids truncated (H3Δ32) for sortase reactions were generated using QuickChange Site-Directed Mutagenesis (Agilent Technologies).

Plasmids containing wild-type HAT module and deletion constructs lacking the Sgf29 Tudor domain (ΔTudor corresponding to Sgf29 aa1-95) or Gcn5 Bromodomain (ΔBromodomain corresponding to Gcn5 aa1-340) were co-transformed into Rosetta2-(DE3) *E. coli* cells (EMD Millipore) and co-expressed as intact complexes bearing a hexahistidine tag on the Ada3 subunit. Ada3 was expressed as a thioredoxin fusion in pET32a, Gcn5 and Ada2 were provided on CDFduet, and Sgf29 was supplied on pRSF. Starter cultures co-transformed with all three plasmids were grown overnight at 37°C in Terrific broth (TB) containing 50 µg/mL carbenicillin, 31 µg/mL chloramphenicol, 50 µg/mL streptomycin, and 25 µg/mL kanamycin. The saturated cultures were then diluted by 100-fold into 4-12 liters of TB and grown at 37°C until the reached an OD₆₀₀ of 0.6-0.8. The flasks were transferred to an ice bath for 45 minutes and the shaker temperature was lowered to 16°C. The cells were induced with the addition of 0.25 mM IPTG overnight (16-18 hours), harvested by centrifugation, and re-suspended in lysis.
buffer containing 40 mM HEPES, pH 7.6, 500 mM NaCl, 10% glycerol, 20 mM imidazole, pH 8.0, and 5 mM 2-mercaptoethanol (BME).

To purify the HAT module, cell pellets were thawed, lysed using a Microfluidizer (Microfluidics Corp), and clarified by centrifugation at 32,000 g. The soluble fraction was loaded onto a HisTrap HP column (GE LifeSciences) equilibrated in lysis buffer. The HAT module was eluted with a 20 to 400 mM imidazole gradient over 15 column volumes and dialyzed overnight at 4°C into 20 mM HEPES, pH 7.6, 100 mM NaCl, 10% glycerol, and 5 mM 2-mercaptoethanol (BME) in the presence of 1 mg recombinant TEV protease per 10 mg purified protein to cleave the thioredoxin tag. Following dialysis, imidazole was added to the cleavage reaction at a final concentration of 20 mM, and then the mixture was run over the HisTrap HP column (GE LifeSciences), which retained the hexahistidine tagged-thioredoxin while allowing the HAT module to pass through. Solid ammonium sulfate was added to the flow-through over ten minutes while stirring at 4°C, to a final concentration of 0.7 M. Precipitated protein was removed using a 0.4 µm filter, and the supernatant applied to a Phenyl HP column (GE LifeSciences) equilibrated in buffer “A”, containing 20 mM HEPES, pH 7.6, 200 mM NaCl, 0.7 M ammonium sulfate, 10% glycerol, and 5 mM BME. The column was stepped to 30% buffer “B”, containing 20 mM HEPES, pH 7.6, 40 mM NaCl, 10% glycerol, and 5 mM BME. The column was developed with a 30-100% gradient over 5 column volumes. Fractions containing the HAT module were concentrated and loaded onto a HiPrep 26/60 Sephacryl S-300 (GE LifeSciences) gel filtration
column equilibrated in 20 mM HEPES, pH 7.6, 100 mM NaCl, and 200 µM tris(2-carboxyethyl)phosphine (TCEP). Purified HAT module was concentrated to 4-10 mg/mL, flash-frozen in liquid nitrogen and stored at -80°C until use. Extinction coefficients were calculated using the ProtParam tool from the ExPASy Bioinformatics Resource Portal (www.expasy.org).

Cells expressing Sgf29 were grown and lysed as described above for the HAT module except that 50 µg/mL carbenicillin and 31 µg/mL chloramphenicol were used as antibiotics. Clarified lysate was loaded onto a HisTrap HP column (GE LifeSciences) equilibrated in buffer containing 20 mM HEPES, pH 7.6, 500 mM NaCl, 20 mM imidazole, pH 8, and 5 mM BME. After washing with 10 column volumes of buffer, the proteins were eluted with 300 mM imidazole and then dialyzed overnight into 20 mM HEPES, pH 7.5, 100 mM NaCl, and 5 mM BME. The thioredoxin/ hexahistidine tags were removed by adding 1 mg of TEV protease per 10 mg recombinant protein during dialysis. Dialyzed samples were loaded onto the HisTrap HP column and the tag was separated using a gradient of 0 to 200 mM imidazole over 10 column volumes. Fractions containing pure protein were dialyzed overnight at 4°C against 20 mM HEPES pH 7.6, 100 mM NaCl, and 200 µM TCEP. The proteins were concentrated to approximately 5 mg/mL, flash-frozen in liquid nitrogen and stored at -80°C.

Sortase was purified as previously described (Piotukh et al, 2011). Purified protein was dialyzed overnight at 4°C into 50 mM HEPES, pH 7.5, 150 mM NaCl,
and 1 mM dithiothreitol (DTT), concentrated to 35-48 mg/mL, flash-frozen in liquid nitrogen and stored at -80°C.

Recombinant *X. laevis* histones, including the mutants used for differential labeling and the truncated version of H3 for enzyme-mediated ligation where the first 32 amino acids were deleted (H3Δ32), were purified as previously described (Luger et al, 1999). DNA for recombinant mononucleosomes was purified from an EcoRV digest of pST55-16xNCP601, a kind gift from Dr. Song Tan (Penn State, PA) as described previously (Dyer et al, 2004). Nucleosomes were assembled using the salt gradient dialysis method (Luger et al, 1999), dialyzed into low-salt buffer containing 10 mM Tris-HCl, pH 7.5, 5 mM KCl, and 1 mM DTT for storage at 4°C, where they were used within six weeks.

*Enzyme-mediated ligation of methylated peptide to histone H3*

Peptide for enzyme-mediated ligation with the sequence: H₂N-ART(K-me3)QTARKSTGGKAPRKQLATKAARKSAPATGGK-NH₂, was purchased at >85% purity (United Peptide; Herndon, VA). The peptide was solubilized in MilliQ water at a concentration of 20 mM and dialyzed against three changes of MilliQ water at 4°C using 100-500 kDa molecular weight cut-off (MWCO) cellulose ester dialysis tubing (Spectrum Labs; Rancho Dominguez, CA). After dialysis, peptide concentrations were determined with a bicinchoninic acid (BCA) assay using bovine serum albumin (BSA) as a standard (ThermoScientific) read in a POLARstar Omega plate reader (BMG Labtech), and then stored at -20°C. Full-
length histone H3 containing trimethylated H3K4 and either the wild-type H3 sequence or a C110A/Q125C double mutant, were prepared as previously described (Piotukh et al, 2011) with the following modifications: Ligation reactions were run in assay buffer containing 50 mM HEPES, pH 7.6, 150 mM NaCl, 10 mM CaCl₂, and 1 mM DTT. Each reaction contained 500 µM 35-mer H3 peptide, 90 µM H3Δ32, and 300 µM sortase enzyme. Since the globular form of H3 (missing the first 32 amino acids) has poor solubility in water, lyophilized protein was first dissolved at a concentration of 8 to 10 mM in solubilization buffer (6 M guanidine-HCl, 20 mM HEPES pH 7.6, and 10 mM DTT) before diluting it into the reaction. The reactions were incubated at 30°C for 16-18 hours, during which the globular and ligated histones gradually precipitated while the sortase enzyme and histone peptide remained soluble. The soluble and insoluble portions of the sortase reaction were separated by spinning at 21,000 g in a microcentrifuge for ten minutes at room temperature. The pellet was resuspended in five reaction volumes of buffer containing 7 M deionized urea, 20 mM HEPES, pH 7.5, 5 mM BME, and 0.5 mM EDTA disodium salt and loaded onto a MonoS HR 5/5 column (GE LifeSciences) equilibrated in the same buffer. To separate full-length histone from unligated substrate, the column was developed with a 0.1-1 M NaCl gradient over 40 column volumes. Fractions containing the full-length H3K4me3 histone were pooled and dialyzed at 4°C against three changes of MilliQ water containing 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and 5 mM BME. Dialyzed histone was lyophilized and resuspended in a small amount of MilliQ water, and the concentration was determined spectrophotometrically using the
same extinction coefficient (4040 M\(^{-1}\) cm\(^{-1}\)) as for full-length H3. Equal amounts of purified H3K4me3 histone H3 and full-length bacterially expressed H3 ranging from 0.5 to 10 µg were run on an SDS-PAGE gel and stained with Coomassie Brilliant Blue. To correct the extinction coefficient of purified H3K4me3 for the presence of contaminants, band intensities were compared in ImageJ between the samples to calculate a correction factor. Typically, the extinction coefficient for recombinant H3 overestimated the concentration of H3K4me3 by a factor of 1.1-1.5, and this correction factor was applied when calculating the H3K4me3 concentration at the octamer reconstitution step. The protein was lyophilized for long-term storage at -20°C.

**MALDI-TOF mass spectrometry**

MALDI samples were prepared using the sandwich method. One microliter of matrix containing 10 mg/mL Sinapinic acid diluted in 50% acetonitrile + 0.05% trifluoroacetic acid was spotted onto a MALDI plate and allowed to air-dry. Ten-fold serial dilutions of full-length H3 and H3K4me3 prepared with enzyme-mediated ligation were spotted on top of the matrix, allowed to dry, followed by another microliter of matrix on top. Spectra were collected using a Voyager DE-STR (Applied Biosystems) in reflection mode.

**Fluorescent labeling of histones and nucleosomes**

The endogenous cysteine in *Xenopus* H3 (C110) points to the interior of the histone octamer, and is not in a good position to accommodate a fluorescent dye.
The Q125 position was chosen for fluorescent labeling instead, because it sits on a solvent-exposed helix. Therefore, histone H3 C110A/Q125C was used for fluorescent labeling with cysteine-reactive dyes. Lypophilized histones were resuspended and reduced in 6 M guanidine-HCl, 20 mM HEPES, pH 7.5, and 400 µM TCEP for 30-60 minutes at room temperature at a concentration of 4 mg/mL. Meanwhile, 1 mg of Alexa Fluor-488 C₅ maleimide and 1 mg of Alexa Fluor-647 C₂ maleimide (LifeTechnologies) were each dissolved in 100 µL of dimethyl sulfoxide (DMSO) at a final concentration of 13.9 mM and 7.7 mM, respectively. Histones bearing different post-translational modifications were labeled with unique colors; unmodified H3 was labeled with Alexa Fluor-647 whereas H3K4me3 was labeled with Alexa Fluor-488. Unfolded/ reduced histones were diluted to a concentration of 2 mg/mL with labeling buffer and incubated with a fivefold molar excess dye for two hours at room temperature with gentle agitation protected from light. Reactions were quenched with the addition of 1 mM BME and combined immediately with equimolar amounts of the other three histones to reconstitute histone octamers as previously described (Luger et al, 1999). Refolded octamers were purified by gel filtration, which also removed excess dye that remained after dialysis. Nucleosomes were reconstituted with differentially labeled octamers as previously described (Luger et al, 1999). Labeled nucleosomes were dialyzed into nucleosome storage buffer (50 mM KCl, 10 mM Tris-HCl, pH 7.5, and 1 mM DTT) and concentrated to > 5 µM for storage at 4°C, where they were used within six weeks.
**Immunoprecipitation assays**

To study the interaction between Sgf29 and the H3K4me3 modification, 25 µL of M-280 Streptavidin DynaBeads (LifeTechnologies) were washed into histone binding buffer (20 mM HEPES, pH 7.6, 50 mM NaCl, 0.02% Tween-20, and 1% BSA) and then incubated for one hour at room temperature with 1 µg of biotinylated histone H3 21-mer peptide +/- H3K4me3, which was generously provided by Dr. Sean Taverna. The beads were washed twice with histone binding buffer and then twice in complex binding buffer (20 mM HEPES, pH 7.6, 50 mM NaCl, 0.02% Tween-20). Recombinant Sgf29 was added to the beads at a final concentration of 10 µM and allowed to bind at room temperature for one hour. The beads were washed three times with complex binding buffer, eluted by boiling with 2X SDS-PAGE loading buffer, run on an SDS-PAGE gel, and stained with Coomassie brilliant blue.

**Steady-state acetylation kinetics**

Steady-state kinetic measurements using peptide substrates were done with a continuous spectrophotometric assay as previously described (Berndsen & Denu, 2005), with a few minor modifications. Briefly, each 50 µL reaction contained 5 mM MgCl₂, 1 mM DTT, 0.2 mM thiamine pyrophosphate (Sigma), 0.2 mM NAD⁺, 100 mM HEPES, pH 7.6, 50 mM NaCl, 2.5 mM pyruvate, 1 µL of 0.45 U/mg at 13 mg/mL pyruvate dehydrogenase (Sigma), 0.5-100 µM acetyl-CoA, 25-50 nM HAT module, and 0-1400 µM histone H3 peptide containing the sequence H₂N-ARTKQTARKSTGGKAPRKQLA-COOH (purchased at >90%
purity from United Peptide). Peptide concentrations were determined with a BCA assay using BSA as a standard (ThermoScientific). For titrations where the peptide concentration was varied, the acetyl-CoA concentration was held at 100 µM. For titrations where the acetyl-CoA concentration was varied, the peptide concentration was held at 400 µM. All of the reaction components, except for acetyl-CoA, were assembled in 384-well plates (Griener Bio-One) and incubated at 30°C for five minutes. Reactions were initiated by the addition of acetyl-CoA, and the absorbance at 340 nm was monitored continuously using a POLARstar Omega plate reader (BMG Labtech) for 5-60 minutes. Absorbance at 340nm was converted into the molar concentration of NADH using Beer’s Law, assuming \( \varepsilon_{340\text{nm}} = 6220 \text{ M}^{-1} \text{ cm}^{-1} \). To calculate initial rates, NADH production was plotted as a function of time and fit to a line where initial velocity conditions were satisfied, typically within the first three minutes. A blank reaction containing HAT module and acetyl-CoA, but no substrate, was performed for each titration and the rate subtracted as background from the other reactions. Initial rates were measured in triplicate, normalized to the enzyme concentration, plotted as a function of substrate concentration, and the resulting curve fit to the Michaelis-Menten equation using non-linear least squares regression implemented in GraphPad Prism 5.

Steady-state titrations with nucleosomal substrates were performed using a radioactive filter-binding assay (Berndsen & Denu, 2005). Nucleosome concentrations were determined spectrophotometrically at 260nm using the
extinction coefficient of the 147-bp DNA fragment: \( \varepsilon_{260nm} = 2,346,045 \, \text{M}^{-1} \, \text{cm}^{-1} \).

Briefly, samples containing 100 mM HEPES pH 7.6, 1 mM DTT, 0-10 µM unmodified NCP or 0-1 µM H3K4me3 NCP, 50 mM NaCl, and 50 nM HAT module were incubated at 30°C for five minutes. Reactions were initiated with the addition of acetyl-CoA at 25 µM, containing a 3:1 molar ratio of unlabeled: tritiated acetyl-CoA (PerkinElmer), and quenched by spotting 25 µL onto P81 filter paper. Since proteins and peptides stick to P81 filter but free acetyl-CoA does not, acetylation was monitored by scintillitation counting on the filter paper. At first, six time points were collected at 15-second intervals to determine the range where product formation was linear with time. Under all conditions tested, initial rate conditions were satisfied for the first minute of the reaction. Subsequent experiments were performed by collecting three time points at 15, 30, and 45 seconds. An unwashed filter was reserved at each substrate concentration to convert counts per minute to a molar concentration of acetyl-CoA. Initial rates were determined by plotting product concentration as a function of time and fitting a line to the data, using the rate of acetyl-transfer in the absence of substrate as a reference. Each experiment was performed in duplicate, and kinetic parameters were fit as described for the enzyme-coupled assay earlier in this paper.

**Acid urea gel electrophoresis assay of acetylation**

Nucleosomes were acetylated in buffer containing 20 mM HEPES, pH 7.6, 50 mM NaCl, 1 mM DTT, 50 µM acetyl-CoA, 20 µg/mL BSA and 0.2 µM or 1 µM
nucleosome core particle (NCP). Reactions were incubated in buffer at 30°C for five minutes, initiated by the addition of 50 nM HAT module, quenched at different time points by flash-freezing in liquid nitrogen, and then lyophilized for analysis by acid urea gel electrophoresis.

Acid urea gels were assembled and run as previously described (Shechter et al, 2007). Histones were visualized with either SYPRO Ruby protein stain (LifeTechnologies) or by Western blotting. For Western blotting, proteins were transferred to PVDF membrane as previously described (Shechter et al, 2007). Membranes were blocked overnight in 5% nonfat milk at 4°C and washed in TBS. Primary antibodies were diluted in 1% nonfat milk in TBS supplemented with 0.1% Tween (TBST) as follows: anti-H3 (Abcam ab1791, 1/25,000), anti-H3K14ac (07-353, EMD Millipore, 1/5000), anti-H3K18ac (EMD Millipore 07-354, 1/7500), anti-H3K23ac (EMD Millipore 07-355, 1/5000), or anti-H3K4me3 (Abcam ab8580, 1/5000). Each primary antibody was applied for 1 hour at room temperature followed by washing in TBST. Goat anti-rabbit IgG-horseradish peroxidase secondary antibody (Amersham Biosciences) was diluted to 1/5000 in 1% nonfat milk and TBST and applied for 1 hour at room temperature and washed in TBST. Blots were developed using Pierce ECL Western Blotting Substrate (ThermoScientific) and exposed using film.
**Differential labeling assay**

Assays using differentially labeled nucleosomes were run as described in the section *HAT assays for analysis by gel electrophoresis*. Control reactions containing only one kind of labeled nucleosome were performed at substrate concentrations of 0.2 µM and 1 µM. Nucleosome mixtures were prepared by combining 0.2 µM H3K4me3 nucleosome/ Alexa Fluor-488 with 0.8 µM unmodified nucleosome/ Alexa Fluor-647 and running the same time course experiments. Quenched time points were resolved on 28-cm acid urea gels and visualized using a Typhoon Imager (GE LifeSciences). Reactions containing Alexa Fluor-488 were excited using the 488nm laser and imaged using the 520BP40 filter, and reactions containing Alexa Fluor-647 were excited using the 633nm laser and imaged using the 670BP30 filter. Native gels containing mixtures of the labeled nucleosomes were imaged in the same way to ensure that there was not spectral overlap between the two dyes (Supplementary Figures 2.1A & 2.1B). Merged images were prepared using Adobe Photoshop.
Figure 2.1 Fluorophores for differential labeling do not exhibit spectral overlap
Native PAGE gel on fluorescently labeled nucleosomes imaged using excitation wavelength/emission filter sets specific for (A) Alexa Fluor-647 or (B) Alexa Fluor-488.

Processivity assay

Three reactions were prepared for each combination of nucleosome type/HAT module construct, containing 20 mM HEPES, pH 7.6, 50 mM NaCl, 1 mM DTT, 50 µM acetyl-CoA, 20 µg/mL BSA, and 0.05 µM fluorescently labeled nucleosome in a total volume of 25-35 µL. Reactions were incubated for 5 minutes at 30°C, initiated with the addition of enzyme to a final concentration of 50 nM, and mixed vigorously by pipetting up and down. After allowing the reaction to proceed for 10-60 seconds, reactions were either: (a) quenched by the addition of 1/5th volume acetic acid, (b) mock-quenched with the addition of
1/5\textsuperscript{th} volume nucleosome storage buffer, or (c) subjected to competition by the addition of 1/5\textsuperscript{th} volume unlabeled nucleosomes, bringing the concentration of the unlabeled competitor to 1 µM. After incubating the reactions for a total of 1.5-10 minutes at 30°C, the samples were flash-frozen in liquid nitrogen, lyophilized, resuspended in 10 µL of acid urea sample buffer, and run on a 28-cm acid urea gel (Shechter et al, 2007). Gels were scanned using a Typhoon imager (GE LifeSciences) with the same excitation wavelengths/ emission filter combinations as for the differential labeling assay. ImageJ was used for background subtraction and to quantitate band intensities.

2.4 Results

\textit{H3K4me3 modestly stimulates acetylation by the HAT module}

We wished to investigate how H3K4 trimethylation affects HAT module activity on peptides and nucleosomes and how the Sgf29 tandem Tudor domain may contribute to differential activity on methylated versus unmethylated substrates. To facilitate quantitative \textit{in vitro} studies focused on the H3K4me3 modification, we utilized recombinant HAT module from \textit{S. pombe}, rather than \textit{S. cerevisiae}, because the \textit{S. pombe} proteins can be isolated as a stable four-protein complex containing Sgf29, Gcn5, Ada2, and Ada3. Full-length \textit{S. pombe} HAT module and variants lacking either the Sgf29 tandem Tudor domain or Gcn5 bromodomain were expressed as complexes in \textit{E. coli} and purified to homogeneity (Figure 2.2). We first confirmed that \textit{S. pombe} Sgf29 interacts
specifically with H3K4me3 peptides by performing pull-down assays. As shown in Figure 2.2A, streptavidin beads coated with H3K4me3 peptides efficiently retain Sgf29, whereas unmodified H3 peptides do not.

To determine whether the interaction between Sgf29 and H3K4me3 changes the kinetics of acetylation by the HAT module on peptide substrates, we measured initial acetylation rates at increasing concentrations of either unmodified or H3K4me3 peptides in the presence of saturating amounts of acetyl-CoA and fit the data to the Michaelis-Menten equation. Compared to unmodified peptides, the presence of the H3K4me3 modification decreases $K_M$ from 250 ± 40 µM to 77 ± 10 µM, but does not affect $k_{cat}$ for the enzyme (Figure 2.3B, Table 2.1). Our results differ from earlier studies, where the presence of Sgf29 had no effect on SAGA activity with H3K4me3 versus unmodified peptides (Bian et al, 2011). However, the data presented here is fit from a full steady-state titration rather than a single rate measurement, and is therefore more sensitive to changes in kinetic parameters. Furthermore, constructs lacking the Sgf29 tandem Tudor domain (ΔTudor) are kinetically equivalent on H3K4me3 versus unmodified peptides (Figure 2.3C, Table 2.1), suggesting that the 3.1-fold increase in $k_{cat}/K_M$ seen with the wild-type HAT module reflects a true difference in enzyme specificity. We also considered the possibility that the H3K4me3 modification affects $k_{cat}$ or $K_M$ for acetyl-CoA but found no evidence of crosstalk between H3K4 trimethylation and acetyl-CoA binding (Figure 2.3D, Table 2.2). The fact that acetyl-CoA binding is independent of the H3K4me3 modification is consistent with what we know about the Gcn5 reaction mechanism, where the
cofactor binds before the peptide (Tanner et al, 2000b; Tanner et al, 1999). To confirm that acetylation by Gcn5 is not directly affected by truncating the Sgf29 tandem Tudor, we compared $k_{cat}$ and $K_M$ values for unmodified histone peptides and nucleosomes between wild-type and ΔTudor HAT module complexes. Since the kinetic parameters between the two complexes are nearly the same for peptides (Table 2.1) and nucleosomes (Table 2.3), differences observed with the ΔTudor HAT module must be related to H3K4me3 recognition.

Figure 2.2 Purified HAT module constructs

Left – SDS-PAGE gel of purified S. pombe HAT module constructs containing different domain truncations where all four subunits are present in roughly stoichiometric quantities. Right – Location of catalytic domain, tandem Tudor domain, and bromodomain within the HAT module.
Figure 2.3 The presence of H3K4me3 increases overall acetylation by the HAT module on peptides.

(A) Pull-down using biotinylated histone peptides containing H3K4 trimethylation and Sgf29. Steady-state kinetic titrations comparing the activity of (B) wild-type HAT module or (C) ΔTudor HAT module on unmodified versus H3K4me3 peptides using an enzyme-coupled assay. (D) Steady-state kinetic titrations comparing the activity of wild-type versus ΔTudor HAT module using a constant amount of H3K4me3 peptide and varying the acetyl-CoA concentration using an enzyme-coupled assay.
**Table 2.1 Steady-state kinetic parameters using peptide substrates**

<table>
<thead>
<tr>
<th>HAT Module:</th>
<th>H3 Peptide aa 1-21:</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_M$ (µM)</th>
<th>$k_{cat}/K_M$ (µM$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>unmod</td>
<td>2.5 ± 0.1</td>
<td>250 ± 40</td>
<td>0.010 ± 0.002</td>
</tr>
<tr>
<td>ΔTudor</td>
<td>unmod</td>
<td>2.5 ± 0.1</td>
<td>210 ± 30</td>
<td>0.012 ± 0.002</td>
</tr>
<tr>
<td>ΔBromodomain</td>
<td>unmod</td>
<td>1.8 ± 0.07</td>
<td>380 ± 40</td>
<td>0.0047 ± 0.001</td>
</tr>
<tr>
<td>Wild-type</td>
<td>H3K4me3</td>
<td>2.4 ± 0.08</td>
<td>77 ± 10</td>
<td>0.031 ± 0.004</td>
</tr>
<tr>
<td>ΔTudor</td>
<td>H3K4me3</td>
<td>2.5 ± 0.09</td>
<td>190 ± 20</td>
<td>0.013 ± 0.001</td>
</tr>
</tbody>
</table>

Kinetic parameters were fit from titrations where the concentration of histone H3 aa1-21 peptide was varied while the acetyl-CoA concentration was held constant at 100 µM.

**Table 2.2 Steady-state kinetic parameters using acetyl-CoA as a substrate**

<table>
<thead>
<tr>
<th>HAT Module:</th>
<th>Substrate:</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_M$ (µM)</th>
<th>$k_{cat}/K_M$ (µM$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>Ac-CoA</td>
<td>1.7 ± 0.1</td>
<td>3.4 ± 0.7</td>
<td>0.50 ± 0.1</td>
</tr>
<tr>
<td>ΔTudor</td>
<td>Ac-CoA</td>
<td>1.8 ± 0.1</td>
<td>2.8 ± 0.3</td>
<td>0.64 ± 0.08</td>
</tr>
</tbody>
</table>

Kinetic parameters were fit from titrations where the concentration of acetyl-CoA was varied while the concentration of H3K4me3 aa1-21 peptide was held constant at 400 µM.
Table 2.3 Steady-state kinetic parameters using nucleosomal substrates

<table>
<thead>
<tr>
<th>NCP:</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_M$ (µM)</th>
<th>$k_{cat}/K_M$ (µM$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>1.3 ± 0.1</td>
<td>1.9 ± 0.3</td>
<td>0.68 ± 0.1</td>
</tr>
<tr>
<td>ΔTudor</td>
<td>1.9 ± 0.2</td>
<td>2.8 ± 0.5</td>
<td>0.68 ± 0.1</td>
</tr>
<tr>
<td>Wild-type</td>
<td>1.0 ± 0.05</td>
<td>0.41 ± 0.05</td>
<td>2.4 ± 0.3</td>
</tr>
</tbody>
</table>

Both the above experiments and previously reported studies (Bian et al, 2011) utilized peptide substrates, leaving open the question of whether the full effect of H3K4 trimethylation on SAGA HAT module activity might require a nucleosomal context. We therefore generated nucleosomes containing recombinant full-length histone H3 that was homogeneously modified with H3K4me3. To incorporate this modification, we employed enzyme-mediated ligation using an engineered bacterial transpeptidase, Sortase A (SrtA) (Piotukh et al, 2011). The native form of the enzyme recognizes and splices a short peptide epitope onto another protein (Popp & Ploegh, 2011), while engineered SrtA (F40 SrtA) recognizes a sequence that naturally occurs in histone H3 (Figure 2.4A) (Piotukh et al, 2011). We generated full-length histone H3 for nucleosome reconstitutions by using F40 SrtA to splice synthetic peptides corresponding to residues 1-35 of histone H3 containing H3K4me3 onto a truncated form of histone H3 (H3ΔNterm) (Figures 2.4A & 2.4B). We confirmed the presence of the trimethyl modification using Western blotting with an antibody.
specific for the H3K4me3 modification (Figure 2.4C), and matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) (Figure 2.5). The modified histone H3 was reconstituted with unmodified histones H2A, H2B, and H4 into nucleosome core particles, thus generating a pool of nucleosomes uniformly trimethylated at H3K4.
Figure 2.4 Producing site-specifically modified histone H3.

(A) Schematic describing enzyme-mediated ligation with sortase to produce recombinant full-length H3K4me3 histone. (B) SDS-PAGE gel of sortase reaction components and products. (C) Western blot on an SDS-PAGE gel of recombinant H3 and H3K4me3 made with sortase. (D) SDS-PAGE gel of histone octamers reconstituted using different versions of histone H3.
To test whether H3K4 trimethylation affects acetylation by the HAT module in a nucleosomal context, we measured steady-state rates of acetylation on increasing concentrations of recombinant nucleosomes with and without the H3K4me3 modification using a radioactive filter-binding assay. In the presence of the H3K4me3 modification, $K_M$ for nucleosomes drops from $1.9 \pm 0.3 \text{ µM}$ to $0.41 \pm 0.05 \text{ µM}$ (Figure 2.6A, Table 2.3), which accounts for the ~3.5-fold increase in $k_{cat}/K_M$ observed with methylated compared to unmodified nucleosomes (Table 2.3). While the $K_M$ for nucleosomes is almost two orders of magnitude lower than that for peptides, indicating that the HAT module binds far more tightly to

**Figure 2.5 MALDI-TOF analysis of H3K4me3 histone**
nucleosomes than to peptides, the 3.5-fold increase in $k_{cat}/K_M$ caused by the methyl modification in a nucleosomal context (Figure 2.6A, Table 2.3) is close to the 3.1-fold increase in specificity observed with peptides (Figure 2.3B, Table 2.1). Since H3K4 trimethylation stimulates HAT module activity on both peptide and nucleosomal substrates, other features of the nucleosome including the DNA and globular domains of the histones, do not appear to be a factor in discriminating H3K4me3 substrates.

Whether the increase in acetylation rate by the HAT module on H3K4 trimethylated nucleosomes involves all four histones, or specifically targets histone H3, cannot be distinguished by radioactive filter-binding assays. Therefore, we compared histone acetylation patterns on unmodified versus H3K4 trimethylated nucleosomes using acid urea gels, which separate histones based on molecular weight and the number of acetylated lysine residues (Shechter et al, 2007). At early time points, the HAT module acetylates histone H3 in nucleosomes harboring the H3K4me3 modification more quickly than unmodified nucleosomal histone H3 (Figure 2.6B). Thirty seconds after initiating the reaction, bands corresponding to three acetylation sites are evident on the H3K4me3 histone, but only two acetylation sites are populated for unmodified H3 (Figure 2.6B). The HAT module also acetylates histone H4, but the rate is insensitive to H3K4 trimethylation (Figure 2.6B), suggesting that histone tail orientation plays a role in HAT module specificity. The observation that H4 acetylation is independent of H3K4 trimethylation is consistent with a mechanism in cis,
whereby recognition of the methyl mark stimulates acetylation on the same histone tail.

In addition to enhancing the overall rate of acetylation on histone H3, we wondered whether recognition of H3K4me3 by Sgf29 alters the specificity of the HAT module for particular lysines. The N-terminus of histone H3 contains seven conserved lysine residues that are acetylated by Gcn5 \textit{in vitro}: K4, K9, K14, K18, K23, K27, and K36 (Cieniewicz et al, 2014; Grant et al, 1999; Kuo et al, 1996). A recent study showed that the Gcn5 bromodomain regulates the specificity of the enzyme for particular lysines, as mutations in the bromodomain that abolish acetyl lysine recognition also alter the site-specific acetylation pattern catalyzed by Gcn5 (Cieniewicz et al, 2014). We speculated that the Sgf29 tandem Tudor domain might perform an analogous function by governing the lysine specificity of Gcn5 on H3K4 trimethylated nucleosomes. To test this, we used antibodies that recognize specific acetylated lysine residues within histone H3 to probe the products of HAT module reactions resolved on acid urea gels. Since antibodies to H3K14ac recognize the singly acetylated histone, whereas antibodies to H3K18ac and H3K23ac do not (Figure 2.6C), H3K14 is the first position acetylated by the HAT module on both unmodified and H3K4me3 substrates. Such a strong intrinsic preference for H3K14 acetylation agrees with the results of mass spectrometry studies of the site-selectivity of Gcn5 (Cieniewicz et al, 2014; Kuo & Andrews, 2013).

Although the HAT module acetylates nucleosomal H3 containing H3K4me3 more quickly than unmodified H3, not all lysine residues are equally
affected by the presence of the methyl mark. H3K14 and H3K18 acetylation are accelerated in the presence of the H3K4me3 modification (Figure 2.6C), which is most evident at later time points (2.5 minutes). Compared to unmodified H3, the signal for H3K14 and H3K18 acetylation is skewed toward the upper portion of the gel in the presence of the methyl mark (Figures 2.6C), indicating that H3K14 and H3K18 acetylation are disproportionately represented among the most highly acetylated histones. In contrast, the pattern of H3K23 acetylation is similar for both methylated and unmethylated H3 (Figure 2.6C), indicating that the rate of H3K23 acetylation stays roughly the same. These results agree with in vivo data showing that overall levels of H3K14 and H3K18 acetylation are reduced in an Sgf29 knock-out strain, whereas H3K23 acetylation is maintained at wild-type levels (Bian et al, 2011).
Figure 2.6 The presence of H3K4me3 increases overall acetylation by the HAT module on nucleosomes.

(A) Steady-state kinetic titrations comparing wild-type HAT module on unmodified versus H3K4me3 nucleosome core particles (NCPs) using a radioactive filter-binding assay. (B) HAT reaction time course on unmodified versus H3K4me3 NCPs resolved on an acid urea gel and stained for total protein using SyproRuby. (C) Western blot of a HAT reaction resolved on an acid urea gel using antibodies for total H3, H3K14ac, H3K18ac, and H3K23ac.
**Differential labeling technique for studying mixtures of post-translationally modified substrates**

Based on the relatively modest increase in HAT module activity observed on methylated versus unmodified nucleosomes, we speculated that experiments using highly purified and homogeneously modified nucleosomes had failed to capture an important regulatory mechanism governing HAT module specificity. Compared to nucleosomes reconstituted from purified components, chromatin *in vivo* is heterogeneous with respect to the set of post-translational modifications found on each nucleosome. For an enzyme like the HAT module, the ability to selectively acetylate nucleosomes bearing particular combinations of post-translational modifications may be more important than its activity overall. While there are multiple biochemical assays for measuring bulk acetylation rates (Berndsen & Denu, 2005), there are no tools for selectively monitoring individual acetylation kinetics in the context of a mixture. We therefore developed a technique that combines differential fluorescent labeling of histones with acid urea gel electrophoresis to quantitate HAT module activity on both modified and unmodified nucleosomes within the same mixture.

To track histone-specific acetylation, we generated color-coded nucleosomes containing fluorescently labeled histone H3: unmodified nucleosomes were labeled with one color and H3K4me3 nucleosomes were labeled with another (Figure 2.7A). After incubating the SAGA HAT module with color-coded nucleosomes (Figure 2.7B) and resolving the reaction products on acid urea gels, the products corresponding to one or the other color-coded histone can be visualized on a laser scanner using fluorophore-specific excitation.
wavelengths and emission filters (Figure 2.7C). If the HAT module acetylates unmodified and H3K4me3 nucleosomes at similar rates, we expect to see the same signal distribution for both fluorophores. On the other hand, if the HAT module preferentially acetylates one sub-population out of the mixture, then the banding pattern will differ between the two fluorescent signals.
Figure 2.7 Overview of differential labeling experiment

(A) Histones containing particular histone post-translational modifications are labeled with different fluorescent dyes. (B) Substrate composition of HAT reactions for interrogating different kinetic properties. (C) Using a laser scanner to image sub-populations of nucleosomes individually.
**Gcn5 preferentially acetylates H3K4me3 nucleosomes**

To ask whether the HAT module preferentially acetylates nucleosomes containing H3K4 trimethylation, we used the color-coding method to assay histone H3 acetylation in mixtures of nucleosome that are either unmodified or that contain two copies of histone H3 bearing the H3K4me3 mark. Since H3K4 trimethylation increases the specificity constant, $k_{cat}/K_M$ for the HAT module by nearly four-fold (Figure 2.6A, Table 2.3), the HAT module should acetylate H3K4me3 and unmodified nucleosomes at the same rate in mixtures containing a fourfold molar excess of unmodified nucleosomes (Cornish-Bowden, 1984). We therefore assayed SAGA HAT module activity under competitive conditions containing a 4:1 molar ratio of unmodified to H3K4me3 nucleosome, and at a total substrate concentration where the HAT module acetylates the differentially modified color-coded nucleosomes at similar rates (Figure 2.8A & 2.8B). Surprisingly, the HAT module preferentially acetylates the methylated nucleosome under conditions where the acetylation rates for modified and unmodified nucleosomes were expected to be the same (Figure 2.8C). An overlay of the images corresponding to each labeled substrate reveals that nucleosomes with the H3K4me3 modification are acetylated to completion within the first five minutes, whereas unmodified nucleosomes are still not fully acetylated at the last time point (Figure 2.8C). Notably, the HAT module does not acetylate H3K4me3 nucleosomes more quickly in the context of a mixture, but rather acetylates them first. Experiments conducted at the same concentration of H3K4me3 nucleosomes with (Figure 2.8C) and without (Figure 2.9C) a fourfold
molar excess of unmodified nucleosomes reveal that bands corresponding to acetylated forms of histone H3K4me3 accumulate with similar time-dependence. Preferential acetylation of methylated nucleosomes requires the Sgf29 tandem Tudor domain, since the pattern of acetylation for SAGA HAT module lacking the Tudor domain is the same for both methylated and unmethylated nucleosomes (Figure 2.8D). Thus, when nucleosomes harboring different histone marks compete for the HAT module active site, the interaction between the Sgf29 tandem Tudor domain and the H3K4me3 modification facilitates preferential acetylation of methylated nucleosomes.
**Figure 2.8** The HAT module preferentially acetylates H3K4me3 nucleosomes out of a mixture

Fluorescent images of HAT reactions with differentially labeled nucleosomes resolved on acid urea gels using (A) wild-type HAT module and 1 µM unmodified/H3K4me3 nucleosomes or (B) ΔTudor HAT module and 1 µM unmodified/H3K4me3 nucleosomes.
H3K4me3 nucleosomes. (C) Fluorescent images of a HAT reaction containing a four to one molar ratio of differentially labeled unmodified versus H3K4me3 nucleosomes resolved on an acid urea gel. Left – Merged image with unmodified nucleosomes colored green and H3K4me3 nucleosomes colored red. Right – individual images used to generate the merged image. (D) Fluorescent images of a ΔTudor HAT reaction containing a four to one molar ratio of unmodified versus H3K4me3 nucleosomes resolved on an acid urea gel. Left – Merged image with unmodified nucleosomes colored green and H3K4me3 nucleosome colored red. Right – individual images used to generate the merged image.

To rule out the possibility that H3K4me3 recognition acts in trans by stimulating overall HAT module activity, we assayed acetylation of nucleosomes in the presence of added H3 tail peptide containing the H3K4me3 modification. Peptide was added at a concentration tenfold higher than the reported dissociation constant of methylated peptide for Sgf29 (Bian et al, 2011), while nucleosomes were held at a concentration matching the amount of H3K4me3 substrate in our experiments using nucleosome mixtures (Figure 2.8C & 2.8D). As shown in Figure 2.9, the added peptide did not increase HAT module activity on unmodified (Figure 2.9A & 2.9B) or H3K4me3 nucleosomes (Figure 2.9C & 2.9D), ruling out trans-stimulation of Gcn5 by H3K4 trimethylated histone tails.
**Figure 2.9** H3K4 trimethylation does not stimulate HAT activity in trans

HAT reactions using fluorescently labeled nucleosomes and wild-type HAT module resolved on acid urea gels with the following substrates: (A) 200 nM unmodified nucleosomes, (B) 200 nM unmodified nucleosomes supplemented with 50 µM H3K4me3 aa1-8 peptide (C) 200 nM H3K4me3 nucleosomes, or (D) 200 nM H3K4me3 nucleosomes supplemented with 50 µM H3K4me3 aa1-8 peptide.

*Sgf29 promotes processive acetylation by the HAT module*

When provided with competing substrates, the HAT module clearly modifies H3K4me3 nucleosomes first, even under conditions where one expects to see equal rates of acetylation (Table 2.3, Figure 2.8C). We hypothesized that H3K4me3 recognition by Sgf29 might promote processive acetylation of methylated nucleosomes, whereby the HAT module modifies multiple lysines before dissociating from the nucleosome. To test for processive acetylation by Gcn5, we used the reaction scheme diagrammed in Figure 2.10A. We incubated the HAT module with labeled nucleosome for a short period of time and then
added a twenty-fold molar excess of unlabeled competitor nucleosome. As controls, a portion of the initial reaction was either quenched with acetic acid, which halts all enzymatic activity, or had buffer alone added (mock quench), which allows the reaction to proceed (Figure 2.10A). The resulting pattern of acetylation is then analyzed on acid-urea gels probed for labeled H3. For a non-processive enzyme, the acetylation pattern in the presence of competitor should mimic the chemical quench, since the enzyme should dissociate readily from the labeled substrate and acetylate the unlabeled competitor instead. A processive enzyme should remain bound to the labeled substrate, even in the presence of added competitor, thus allowing for further acetylation of the labeled nucleosome. As a result, the acetylation pattern for a processive enzyme in the presence of competitor should mimic the mock-quench.

When the SAGA HAT module was incubated with labeled but otherwise unmodified nucleosomes, addition of unlabeled nucleosomes competed for HAT module activity and prevented labeled H3 from becoming fully acetylated (compare competitor and mock-quenched samples in Figure 2.10B). By contrast, addition of unlabeled competitor nucleosomes had little effect on acetylation of labeled H3K4 trimethylated nucleosomes, as can be seen by the similarity in the intensity profiles of the competitor and mock-quench samples (Figure 2.10C). This behavior indicates that the SAGA HAT module acetylates H3K4me3 nucleosomes much more processively than unmodified nucleosomes.

To probe whether processive acetylation of H3K4me3 nucleosomes depends on the interaction between the Sgf29 tandem Tudor domain and the
H3K4 trimethyl modification, we performed the same experiment using a HAT module construct lacking the Tudor domain (ΔTudor). Unlike the intact HAT module, the ΔTudor construct can be competed off of both unmodified (Figure 2.10D) and H3K4me3 nucleosomes (Figure 2.10E). These results are consistent with a role for the Sgf29 Tudor domain in engaging the methylated histone H3 tail.

Since the Gcn5 bromodomain recognizes acetyl lysine and regulates the lysine specificity of Gcn5 (Cieniewicz et al, 2014), we asked whether the bromodomain contributes to processive acetylation by the HAT module. Truncating the bromodomain of Gcn5 while leaving the Sgf29 tandem Tudor domain intact causes a two-fold decrease in $k_{cat}/K_M$ for peptide substrates compared to wild-type HAT module (Table 2.1). As been demonstrated previously, deletion of the Gcn5 bromodomain changes the overall acetylation pattern catalyzed by the HAT module (Cieniewicz et al, 2014), particularly by delaying accumulation of the most highly acetylated states (Figure 2.11A & 2.11B). However, processive acetylation is clearly impaired on H3K4me3 nucleosomes (Figure 2.10F & 2.10G), although not to the same degree seen with the ΔTudor construct (Figure 2.10D & 2.10E). These results suggest that, while processivity is primarily governed by binding of the Sgf29 tandem Tudor domain to H3K4me3 marks in nucleosomes, there is some contribution by the Gcn5 bromodomain, presumably through binding to acetylated lysine residues.
**Figure 2.10** Sgf29 promotes processive acetylation against H3K4me3 nucleosomes.

(A) Schematic of competition assay to test for processive behavior. Competition assay using wild-type HAT module and (B) unmodified nucleosomes or (C) H3K4me3 nucleosomes. Results from a competition assay using ΔTudor HAT module and (D) unmodified nucleosomes or (E) H3K4me3 nucleosomes. Results from a competition assay using ΔBromodomain HAT module and (F) unmodified nucleosomes or (G) H3K4me3 nucleosomes Top panel – acid urea gel of competition assay visualized using a Typhoon scanner. Background subtraction performed using a sliding paraboloid in ImageJ. Bottom panel – quantitation of acid urea gel in ImageJ.
Figure 2.11 *The bromodomain of Gcn5 facilitates acetylation on nucleosomal substrates.*

(A) HAT reactions using Gcn5 ΔBromodomain and 200 nM fluorescently labeled nucleosome resolved on an acid urea gel. (B) HAT reactions using Gcn5 ΔBromodomain and 200 nM fluorescently labeled H3K4me3 nucleosome resolved on an acid urea gel.

2.5 Discussion

While crosstalk between histone modifications is an established and well-documented feature of the histone code (Musselman et al., 2012), the underlying mechanism has remained obscure. In this study, we uncover a mechanism by which the HAT subcomplex of the SAGA transcriptional coactivator couples histone hyperacetylation to H3K4 trimethylation, a universal mark associated with promoter regions of actively transcribed chromatin in organisms ranging from yeast (Santos-Rosa et al., 2002) to humans (Barski et al., 2007). While H3K4 trimethylation modestly accelerates acetylation by the HAT module on peptides and nucleosomes (Figure 2.3B, Figure 2.6A, Table 2.1 & 2.3), we find that the
HAT module is highly selective for H3K4 trimethylated histone in mixtures of modified and unmodified nucleosomes (Figure 2.8C). In addition, we find that the HAT module is more processive on nucleosomes containing the H3K4me3 modification (Figure 2.10B & 2.10C), acetylating each H3K4 trimethylated nucleosome multiple times before dissociating. Enhanced processivity by the HAT module on H3K4me3 nucleosomes can explain how a modest four-fold increase in specificity translates into a strong preference for acetylating nucleosomes harboring H3K4 trimethylation. This mechanism accounts for the results of mass spectrometry studies showing that the acetylation state of yeast histone H3 \textit{in vivo} is strongly correlated with H3K4 trimethylation on the same histone tail (Jiang et al, 2007; Taverna et al, 2007b).

Our results explain how the HAT module, which is broadly distributed over transcriptionally active chromatin (Bonnet et al, 2014), can establish different acetylation patterns at promoters versus coding regions of transcribed genes. For example, H3K9 and H3K14 acetylation are both catalyzed by Gcn5 \textit{in vivo} (Zhang et al, 1998), yet high-resolution ChIP-seq data show that H3K9 acetylation is narrowly limited to promoter regions (Bonnet et al, 2014) whereas H3K14 acetylation spreads across gene bodies as well (Johnsson et al, 2009). How can the same enzyme generate different acetylation patterns? Our results are consistent with a model whereby H3K4 trimethylation, which is enriched at gene promoters compared to coding regions (Pokholok et al, 2005; Wang et al, 2008b) and is recognized by the Sgf29 subunit of the HAT module (Bian et al, 2011), promotes a different pattern of acetylation in promoter regions than in
gene bodies (Figure 2.12). In the absence of H3K4 trimethylation, the HAT module likely acetylates histones according to the intrinsic lysine specificity of Gcn5, which vastly favors modification at the H3K14 position (Cieniewicz et al, 2014; Kuo et al, 1996; Kuo & Andrews, 2013). At promoters, however, the prevalence of the H3K4me3 mark promotes processive, multi-site acetylation by the HAT module, thus leading to higher levels of acetylation at non-preferred sites such as H3K9. This hyperacetylation is mediated by Sgf29, which binds to H3K4me3 via its tandem Tudor domain. The bromodomain of Gcn5 can bind to these newly acetylated residues, providing a second feedback mechanism that promotes retention of the complex and, consequently, local hyperacetylation (Figure 2.12). This mechanism can explain how the “reader” functions of Sgf29 and Gcn5 might fine-tune the activity of Gcn5 along different genic regions, thus establishing diffuse patterns of H3K14 acetylation punctuated by narrow regions of hyperacetylation that co-localize with H3K4 trimethylation.

Our results highlight the importance of using defined, homogeneous chromatin templates with native modifications to study the mechanistic basis of crosstalk between histone modifications. For many chromatin-modifying complexes, the mechanism by which “reader” domains translate the histone code into functional outcomes remains an open question that is difficult to answer biochemically with recombinant unmodified histones or with heterogeneously modified chromatin isolated from cells. Our color-coding assay, combined with established methods for generating uniformly modified histones (Chatterjee & Muir, 2010; Piotukh et al, 2011), makes it possible to monitor acetylation on
different pools of nucleosomes within the same reaction, thereby better approximating the heterogeneity of chromatin inside the cell. This differential labeling assay also makes it possible to observe substrate competition directly, rather than making inferences from independent measurements of specificity. In the present study, we compared activity on unmodified and H3K4me3 nucleosomes; however, the number of independent sub-populations that can be interrogated simultaneously using color-coding is limited only by the necessity to avoid significant spectral overlap between the fluorophores used. This technique can be adapted to study tail-specific acetylation on asymmetrically modified nucleosomes as well (Li & Shogren-Knaak, 2008), for example where one copy of H3 contains the H3K4me3 mark and the other copy is unmodified. By incorporating more than one label into the same octamer, different histones within the same nucleosome can be monitored independently. Similarly, nucleosome arrays could be reconstituted with specific proportions of differentially labeled octamers in order to study histone modifications in the context of chromatin fibers. Our assay using fluorescent histone labeling, along with other recently developed techniques such as DNA-barcoded nucleosome libraries (Nguyen et al, 2014), are particularly suited to address these types of questions on a mechanistic level.

This work defines a mechanism that may regulate acetylation by the SAGA complex at gene promoters where the H3K4me3 modification is enriched. Our studies not only reveal the mechanistic basis for crosstalk between H3K4 trimethylation and acetylation by the SAGA HAT module, but also highlight the
fact that crosstalk established by local histone modifications may have profound
effects on the activity and specificity of chromatin-modifying enzymes. The
biochemical approaches developed here will likely facilitate future work
investigating how nuclear machinery translates the combinatorial language of
histone post-translational modifications into functional outcomes.
Figure 2.12 Model depicting PTM-dependent acetylation by SAGA
Chapter 3: Contribution of chromatin-binding domains to SAGA HAT module activity

3.1 Introduction

The SAGA HAT module subunits contain many domains that recognize chromatin epitopes or mediate protein-protein interactions (Figure 3.1), such that the term “Chromatin Interaction Module” has been proposed as an alternative name for the four-protein complex (Spedale et al, 2012). Pointing to their functional importance in vivo, these chromatin-interaction domains have been evolutionarily conserved from yeast to humans, both in identity and distribution among HAT module subunits (Spedale et al, 2012). Gcn5, the catalytic subunit of the HAT module, also harbors a bromodomain that binds to acetyl lysine (Figure 3.1) (Dhalluin et al, 1999; Hudson et al, 2000; Owen et al, 2000). The transcriptional adaptor protein, Ada2, which binds directly to Gcn5 and enhances its acetyltransferase activity (Boyer et al, 2002; Sterner et al, 2002), also contains zinc finger, SANT, and SWIRM domains (Figure 3.1). Some evidence suggests that the Ada2 SANT domain may bind to histone tails (Boyer et al, 2002; Boyer et al, 2004) and SWIRM domains are commonly found in complexes that modify chromatin (Aravind & Iyer, 2002). Even though Ada3 does not contain annotated domains known to interact with chromatin, it binds to both Ada2 (Horiuchi et al, 1995) and Sgf29 (Kurabe et al, 2007), and is required for efficient nucleosomal acetylation by the HAT module (Balasubramanian et al, 2002). Finally, Sgf29 possesses a tandem Tudor domain that specifically recognizes H3K4 trimethylation (Bian et al, 2011; Vermeulen et al, 2010), a modification enriched at the promoters of actively transcribed genes (Pokholok et al, 2005). Whether
these chromatin-interaction domains contribute to the structural integrity of the HAT module, regulate catalysis by Gcn5, or participate in nucleosome recognition has never been studied systematically.

![HAT Module Domains](image)

**Figure 3.1 Chromatin-interacting domains within the SAGA HAT module**

### 3.2 Methods

**Plasmid constructs for expression in E. coli**

SAGA HAT module subunits were PCR-amplified from *S. pombe* genomic DNA and cloned into vectors for bacterial expression using the In-Fusion Cloning kit (Clontech). Since *pombe* Sgf29 contains an intron, the two exons were PCR-amplified using primers containing 15 bp internal overlap and used together in one In-Fusion reaction. Plasmids containing *pombe* HAT module subunits that were generated for this study are summarized in Table 3.1.
Table 3.1 List of plasmids containing HAT module domain truncations for bacterial expression

**Plasmids for recombinant expression of S. pombe HAT module subunits**

NOTE: TEV site DNA sequence included in primer is GAAAACCTGTATTTTCAGGGA, which corresponds to the amino acid sequence ENLYFQG

*Linker region added to facilitate cleavage by TEV (amino acid sequence = ASAS).

<table>
<thead>
<tr>
<th>Vector</th>
<th>Gene</th>
<th>Restriction Sites</th>
<th>Gene</th>
<th>Restriction Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET32a</td>
<td>TEV site-Sgf29</td>
<td>Nco I/ Xho I</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>pET32a</td>
<td>TEV site-Sgf29 aa1-95</td>
<td>Nco I/ Xho I</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>pET32a</td>
<td>TEV site-linker*-Sgf29</td>
<td>Nco I/ Xho I</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>pET32a</td>
<td>TEV site-Ada3 aa1-135</td>
<td>Nco I/ Xho I</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>pET32a</td>
<td>TEV site-Gcn5 aa1-455</td>
<td>Nco I/ Xho I</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>pET32a</td>
<td>TEV site-Ada3 aa1-552</td>
<td>Nco I/ Xho I</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>pET32a</td>
<td>TEV site-Ada2 aa1-120</td>
<td>Nco I/ Xho I</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>pET32a</td>
<td>TEV site-linker*-Ada2 aa317-438</td>
<td>Nco I/ Xho I</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>pRSF</td>
<td>Sgf29 aa1-245</td>
<td>Nco I/ Xho I</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>pRSF</td>
<td>Sgf29 aa1-95</td>
<td>Nco I/ Xho I</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa1-455</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa1-438</td>
<td>Bgl II/ Xho I</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa1-340</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa1-438</td>
<td>Bgl II/ Xho I</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa1-455</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa1-293</td>
<td>Bgl II/ Xho I</td>
</tr>
<tr>
<td>Complex</td>
<td>Protein</td>
<td>Tag</td>
<td>Subunit 1</td>
<td>Subunit 2</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>-----</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa1-455</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa52-438</td>
<td>Bgl II/ Xho I</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa1-455</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa52-293</td>
<td>Bgl II/ Xho I</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa1-455</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa52-293</td>
<td>Bgl II/ Xho I</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa49-455</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa1-438</td>
<td>Bgl II/ Xho I</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa49-455</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa1-293</td>
<td>Bgl II/ Xho I</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa49-455</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa52-293</td>
<td>Bgl II/ Xho I</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa49-455</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa52-293</td>
<td>Bgl II/ Xho I</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa1-340</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa52-293</td>
<td>Bgl II/ Xho I</td>
</tr>
<tr>
<td>CDFduet</td>
<td>Gcn5 aa49-340</td>
<td>Nco I/ Not I</td>
<td>Ada2 aa52-293</td>
<td>Bgl II/ Xho I</td>
</tr>
</tbody>
</table>

Protein expression and purification

The HAT module was co-expressed in Rosetta2-(DE3) E. coli cells (EMD Millipore) as an intact complex bearing a hexahistidine tag on the Ada3 subunit. All complexes containing a total of four subunits (Ada3, Ada2, Gcn5, and Sgf29) and any combination of domain truncations were purified using the same protocol. Ada3 was expressed as a thioredoxin fusion in pET32a, Gcn5 and Ada2 were provided on CDFduet, and Sgf29 was supplied on pRSF. Starter cultures co-transformed with all three plasmids were grown overnight at 37°C in Terrific Broth (TB) containing 50 µg/mL carbenicillin, 31 µg/mL chloramphenicol, 50 µg/mL streptomycin, and 25 µg/mL kanamycin. The saturated cultures were then diluted by a factor of 100 into 4-12 liters of TB and grown at 37°C shaking at 250 rpm until they reached an OD₆₀₀ = 0.6-0.8. The flasks were transferred to an ice bath for 45 minutes and the shaker temperature was turned down to 16°C.
The cells were induced with the addition of 0.25 mM IPTG overnight (16-18 hours), harvested by centrifugation, and re-suspended in lysis buffer containing 40 mM HEPES, pH 7.6, 500 mM NaCl, 10% glycerol, 20 mM imidazole, pH 8.0, and 5 mM 2-mercaptoethanol (BME). To purify the HAT module, cell pellets were thawed in a room temperature water bath, lysed using a Microfluidizer (Microfluidics Corp), and clarified by centrifugation at 32,000 g. The soluble fraction was loaded onto a HisTrap HP column (GE LifeSciences) equilibrated in lysis buffer. The HAT module was eluted with a 20 to 400 mM imidazole gradient over 15 column volumes and dialyzed overnight at 4°C into 20 mM HEPES, pH 7.6, 100 mM NaCl, 10% glycerol, and 5 mM BME in the presence of 1 mg recombinant TEV protease per 10 mg purified protein. Imidazole was then added to the cleavage reaction at a final concentration of 20 mM. The reaction mix was run over the HisTrap HP column (GE LifeSciences), where the HAT module eluted in the flow-through. Solid ammonium sulfate was added to the flow-through to a final concentration of 0.7 M, which was performed over ten minutes while stirring at 4°C. Precipitated protein was removed using a 0.4 µm filter and discarded, while the supernatant was applied to a Phenyl HP column (GE LifeSciences) equilibrated in buffer “A”, containing 20 mM HEPES, pH 7.6, 200 mM NaCl, 0.7 M ammonium sulfate, 10% glycerol, and 5 mM BME. The column was stepped to 30% buffer “B” containing 20 mM HEPES, pH 7.6, 40 mM NaCl, 10% glycerol, and 5 mM BME, then developed with a 30-100% gradient over 5 column volumes. Fractions containing the HAT module were concentrated and loaded onto a HiPrep 26/60 Sephacryl S-300 column (GE LifeSciences).
equilibrated in 20 mM HEPES, pH 7.6, 100 mM NaCl, and 200 µM tris(2-carboxyethyl)phosphine (TCEP). Purified HAT module was concentrated to 4-10 mg/mL and flash-frozen in liquid nitrogen for long-term storage at -80°C. Extinction coefficients were calculated using the ProtParam tool from the ExPASy Bioinformatics Resource Portal (www.expasy.org).

The conditions for expressing and lysing individual HAT module subunits and *E. coli* thioredoxin (from pET32a) are identical to those for the four-protein complex, except that 50 µg/mL carbenicillin and 31 µg/mL chloramphenicol were used as antibiotics. Clarified lysates were loaded onto a HisTrap HP column (GE LifeSciences) equilibrated in buffer containing 20 mM HEPES, pH 7.6, 500 mM NaCl, 20 mM imidazole, pH 8, and 5 mM BME. After washing with 10 column volumes of buffer, the proteins were eluted by adding 300 mM imidazole and then dialyzed overnight into 20 mM HEPES, pH 7.5, 100 mM NaCl, and 5 mM BME. The thioredoxin/ hexahistidine tags were cleaved off of all the proteins by adding 1 mg of TEV protease per mg recombinant protein during dialysis, except for Ada3 aa1-135 and *E. coli* thioredoxin, which were dialyzed without adding TEV protease. Cleaved samples were loaded back onto the HisTrap HP column, and the tag was separated from the target protein by running a gradient to 200 mM imidazole over 10 column volumes. Thioredoxin, Ada3 aa1-135, full-length Sgf29, Sgf29 aa1-95, and Sgf29 aa103-245 were dialyzed overnight at 4°C against 20 mM HEPES pH 7.6, 100 mM NaCl, and 200 µM TCEP, concentrated to approximately 5 mg/mL and flash-frozen in liquid nitrogen for long-term storage.
storage at -80°C. Full-length Gcn5 was further purified over a Superdex 200 16/60 column equilibrated in 20 mM HEPES, pH 7.6, 150 mM NaCl, and 200 µM TCEP. The zinc finger-SANT domain of Ada2 (aa1-120) was further purified over a Q HP column (GE LifeSciences) equilibrated in 100 mM NaCl, 20 mM HEPES, pH 7.6, and 5 mM BME and developed with a 100 to 500 mM NaCl gradient over 10 column volumes, followed by gel filtration over a Superdex200 16/60 column equilibrated in 20 mM HEPES, pH 7.6, 100 mM NaCl, and 200 µM TCEP. Gcn5 and Ada2 aa1-120 were concentrated to approximately 5 mg/mL and flash-frozen in liquid nitrogen for long-term storage at -80°C. After the second HisTrap column, the Ada2 SWIRM domain (aa317-438) was further purified over an Sp HP column (GE LifeSciences) equilibrated in 100 mM NaCl, 20 mM HEPES, pH 7.6, and 5 mM BME and developed with a 100 to 500 mM NaCl gradient over 5 column volumes, followed by gel filtration over a Superdex75 16/60 column equilibrated in 150 mM NaCl, 20 mM HEPES, pH 7.6, 1 mM DTT. Fractions containing pure Ada2 SWIRM (aa317-438) were pooled, concentrated to 12.9 mg/mL, and frozen in liquid nitrogen for long-term storage at 80°C.

The complex between Ada2 aa1-120 and Gcn5 was prepared by mixing the proteins at a concentration of 200 µM, incubating the mixture on ice for 30 minutes, and then injecting the sample onto an analytical Superdex200 10/300 GL (GE LifeSciences) equilibrated with 20 mM HEPES, pH 7.6, 100 mM NaCl, and 200 µM TCEP. Fractions containing both proteins were pooled, concentrated, and flash-frozen in liquid nitrogen for long-term storage at -80°C.
Immunoprecipitation assays

To study its interaction with Sgf29, a fragment of Ada3 bearing a hexahistidine-tag was incubated with full-length Sgf29, the N-terminus of Sgf29, or the Sgf29 tandem Tudor domain for 10 minutes at room temperature in 5 mM sodium phosphate, pH 7.4, 100 mM NaCl, and 0.01% Tween-20. As a control, hexahistidine-tagged *E. coli* thioredoxin was incubated with the same fragments of Sgf29 under the same conditions. All proteins were combined at a final concentration of 10 µM each. 25 µL of His-Tag Dynabeads (LifeTechnologies) was added to each sample and allowed to bind for five minutes at room temperature. The beads were washed four times with 50 mM sodium phosphate, pH 7.4, 300 mM NaCl, and 0.01% Tween-20. Complexes that were retained on the beads were eluted with 50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 0.01% Tween-20, and 300 mM imidazole, run on an SDS-PAGE gel, and stained with Coomassie Brilliant Blue.

Nucleosome binding assays

Nucleosome binding assays were performed in the presence of 20 mM HEPES, pH 7.6, 50 mM NaCl, 5% sucrose, 1 mM DTT, and 0.2 mg/mL bovine serum albumin (BSA). HAT module complex at concentrations ranging from 0 to 10 µM or Ada2 SWIRM domain at concentrations ranging from 0 to 50 µM were incubated with 50 nM nucleosome core particle at room temperature for 30 minutes. Native gels were used to measure the extent of binding. 6% Novex TBE gels (LifeTechnologies) were pre-run at 100V at 4°C for an hour in 0.25 X TBE running buffer, or until the voltage stabilized before loading the samples.
Samples were loaded onto gels running at 100V, and then electrophoresed for 90-120 minutes at 150V at 4°C. Gels were stained for 20 minutes with 1X SYBR Gold nucleic acid gel stain diluted in 0.5X TBE buffer (LifeTechnologies) to visualize nucleosomes and HAT module-nucleosome complexes. For fluorescein-labeled nucleosomes, the gels were scanned with a Typhoon Imager (GE LifeSciences) using the 488nm laser and 520BP40 filter.

**Steady-state acetylation measurements**

Steady-state kinetic measurements were performed with a continuous spectrophotometric assay for peptides and a radioactive filter-binding assay for nucleosomes as previously described (Berndsen & Denu, 2005), with the modifications outlined in section 2.3. The resulting curves were fit to the Michaelis-Menten equation using GraphPad Prims 5.

### 3.3 Results and discussion

**Contribution of chromatin-binding domains to HAT module integrity**

We wondered whether truncating any of the chromatin-interaction domains associated with the HAT module, individually or combinatorially, would destabilize the complex or affect the acetyltransferase activity of Gcn5. Therefore, we purified recombinant *S. pombe* HAT module complexes containing various combinations of domain truncations from *E. coli*, where the most minimal complex lacked the Sgf29 tandem Tudor domain, the Ada2 zinc finger, the Ada2 SWIRM domain, the amino terminus of Gcn5, and the bromodomain of Gcn5.
None of the truncations tested disrupted the complex, even when five domains were truncated at once (Figure 3.2A). To test whether the chromatin interaction domains affected the catalytic activity of Gcn5, we measured steady-state acetylation kinetics of the truncated complexes on peptide substrates. Surprisingly, even when four domains were truncated in the same complex, both $k_{\text{cat}}$ and $K_M$ remained within error of one another (Figure 3.2B, Table 3.2). Therefore, the chromatin interaction domains within the HAT module do not modulate its acetyltransferase activity on simple peptide substrates.

**Figure 3.2** *Contribution of chromatin-interacting domains to HAT module integrity and activity*  
(A) SDS-PAGE gel of purified HAT module constructs with various combinations of domain truncations reveals that chromatin-interacting domains are peripheral to complex integrity. (B) Steady-state kinetic analysis using histone peptides with HAT module constructs containing various combinations of domain truncations.
Table 3.2 Steady-state kinetic parameters using peptide substrates and various HAT module domain truncations

<table>
<thead>
<tr>
<th>HAT Module Construct:</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_M$ (µM)</th>
<th>$k_{cat}/K_M$ (µM$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAT module (wild-type)</td>
<td>2.5 ± 0.1</td>
<td>250 ± 40</td>
<td>0.010 ± 0.002</td>
</tr>
<tr>
<td>ΔTudor</td>
<td>2.5 ± 0.1</td>
<td>210 ± 30</td>
<td>0.012 ± 0.002</td>
</tr>
<tr>
<td>ΔTudor ΔSWIRM</td>
<td>3.2 ± 0.1</td>
<td>270 ± 30</td>
<td>0.012 ± 0.001</td>
</tr>
<tr>
<td>ΔTudor ΔSWIRM ΔN-term Gcn5</td>
<td>3.1 ± 0.1</td>
<td>300 ± 30</td>
<td>0.010 ± 0.001</td>
</tr>
<tr>
<td>ΔTudor ΔSWIRM ΔZnF</td>
<td>3.4 ± 0.2</td>
<td>440 ± 60</td>
<td>0.0078 ± 0.001</td>
</tr>
<tr>
<td>ΔTudor ΔSWIRM ΔZnF ΔN-term Gcn5</td>
<td>3.8 ± 0.1</td>
<td>380 ± 30</td>
<td>0.010 ± 0.001</td>
</tr>
<tr>
<td>ΔSWIRM</td>
<td>2.9 ± 0.1</td>
<td>390 ± 40</td>
<td>0.0075 ± 0.001</td>
</tr>
</tbody>
</table>

Kinetic parameters were fit from titrations where the concentration of H3 aa1-21 peptide was varied while the concentration of acetyl-CoA was held constant at 100 µM.

In budding yeast, Ada2 stimulates the catalytic efficiency of Gcn5 through its SANT domain (Boyer et al, 2002; Sterner et al, 2002), which interacts with histone tails (Boyer et al, 2002). Whether the isolated SANT domain is sufficient to stimulate the activity of Gcn5 on peptides is not known. To address this question, we expressed and purified full-length *pombe* Gcn5 and an N-terminal fragment of *pombe* Ada2 containing both zinc finger (ZnF) and SANT domains. Gcn5 and the Ada2 ZnF-SANT domain form a stable complex over gel filtration (Figure 3.3A), demonstrating that this minimal fragment of Ada2 is sufficient for interacting with Gcn5. Steady-state kinetics performed in the presence and absence of the Ada2 ZnF-SANT domain show that the ZnF-SANT domain
increases $k_{\text{cat}}$ for Gcn5 by approximately three-fold, whereas $K_M$ for peptide remains the same (Figure 3.3B, Table 3.3). A similar experiment in which the acetyl-CoA concentration was varied yields a three-fold increase in $k_{\text{cat}}$ and a nine-fold decrease in $K_M$ for the nucleotide cofactor (Figure 3.3C, Table 3.4), suggesting that the \textit{pombe} SANT domain may also regulate acetyl-CoA binding. These findings differ from previous studies using Gcn5 and full-length Ada2 from budding yeast, where the presence of Ada2 simultaneously increased $k_{\text{cat}}$ and decreased $K_M$ for peptide by three-fold, and had no effect on its affinity for acetyl-CoA (Boyer et al, 2002). Aside from using proteins derived from different organisms, our studies differ in several important ways. First, we use a minimal ZnF-SANT domain from the N-terminus of Ada2, whereas the previous study included the full-length protein. It is possible that the full-length protein has a different effect on catalysis by Gcn5. Second, the peptide $K_M$ for \textit{S. pombe} Gcn5 is much larger than the corresponding values for \textit{S. cerevisiae} (Boyer et al, 2002) and therefore is less well determined. Finally, our steady-state titrations with acetyl-CoA were conducted using sub-saturating amounts of peptide, which may exaggerate differences in apparent kinetic constants when varying the concentration of acetyl-CoA. In contrast to the other chromatin-interacting domains in the HAT module, only the Ada2 ZnF-SANT directly modulates catalysis by Gcn5 and affects the kinetics of acetylation on a peptide substrate.

Next, we characterized the interaction between Sgf29 and Ada3. Yeast two-hybrid and co-immunoprecipitation studies have shown that the mammalian homologues of Sgf29 and Ada3 interact through N-terminal domains (Kurabe et
al, 2007), which are predicted to be coiled-coils. To test whether this also applied in *pombe*, we immobilized the N-terminus of Ada3 on beads and tried to pull-down recombinantly prepared constructs of Sgf29. Full-length Sgf29 and a construct containing the first 95 amino acid residues of Sgf29 were retained on the beads, whereas the Sgf29 tandem Tudor domain was not (Figure 3.3D). Therefore, the N-terminus of Sgf29, and not the tandem Tudor domain, tethers Sgf29 to the rest of the HAT module by interacting with the N-terminus of Ada3.

Clearly, the chromatin interaction domains within the HAT module subunits are peripheral to the core complex (Figure 3.3E), since the domains can be truncated without compromising complex integrity. Relatively small motifs are responsible for tethering different HAT module subunits together (Figure 3.3E), as the ZnF-SANT domain of Ada2, which is less than 120 amino acids long, is sufficient to form a complex with Gcn5 (Figure 3.3A & 3.3E), and the N-terminal 135 amino acid residues of Ada3 can bind to the first 95 residues of Sgf29 (Figure 3.3D & 3.3E). Rather than regulating the catalytic activity of Gcn5, these peripheral domains may participate in substrate recruitment, either by binding to general nucleosomal epitopes or by recognizing specific post-translational modifications on histone tails.
Figure 3.3 Connectivity between HAT module subunits

(A) Gel filtration elution profile of Gcn5/Ada2 ZnF-SANT complexes. Steady-state kinetic titrations of Gcn5-containing complexes on (B) histone peptides or (C) acetyl-CoA. (D) SDS-PAGE gel of a pull-down between (Top) – His-tagged fragment of the Ada3 N-terminus or (Bottom) – E. coli thioredoxin and Sgf29 stained with Coomassie brilliant blue. (E) Model of the connectivity between HAT module subunits.
Table 3.3 Steady-state kinetic parameters for Gcn5 on peptide substrates

<table>
<thead>
<tr>
<th>HAT Module Construct:</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_M$ (µM)</th>
<th>$k_{cat}/K_M$ (µM$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gcn5</td>
<td>0.53 ± 0.1</td>
<td>570 ± 280</td>
<td>0.00093 ± 0.001</td>
</tr>
<tr>
<td>Gcn5 + Ada2 ZnF-SANT</td>
<td>1.6 ± 0.1</td>
<td>510 ± 80</td>
<td>0.0032 ± 0.001</td>
</tr>
</tbody>
</table>

Summary of steady-state kinetic parameters fit from titrations where the concentration of peptide was varied while the acetyl-CoA concentration was held constant at 100 µM.

Table 3.4 Steady-state kinetic parameters for Gcn5 using acetyl-CoA as a substrate

<table>
<thead>
<tr>
<th>HAT Module Construct:</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_M$ (µM)</th>
<th>$k_{cat}/K_M$ (µM$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gcn5</td>
<td>0.28 ± 0.01</td>
<td>7.4 ± 1</td>
<td>0.038 ± 0.01</td>
</tr>
<tr>
<td>Gcn5 + Ada2 ZnF-SANT</td>
<td>0.91 ± 0.02</td>
<td>0.84 ± 0.1</td>
<td>1.1 ± 0.1</td>
</tr>
</tbody>
</table>

Summary of steady-state kinetic parameters fit from titrations where the concentration of acetyl-CoA was varied while the peptide concentration was held constant at 400 µM.

Nucleosome binding studies with the HAT module

To ask whether the chromatin-interaction domains associated with the HAT module affect nucleosome recognition, binding assays were performed with recombinant nucleosome core particles. Binding was assessed using electrophoretic mobility shift assays (EMSAs), since nucleosome-HAT module complexes are stable and migrate more slowly than free nucleosomes on native gels. The wild-type HAT module populates two nucleosome-bound states that
have different electrophoretic mobilities on native gels, with an apparent
dissociation constant in the low micromolar range (Figure 3.4A). Whereas
truncating the tandem Tudor domain in Sgf29 subtly delays the accumulation of
the species with the slowest electrophoretic mobility (Figure 3.4B), truncating the
Ada2 SWIRM domain completely eliminates the second bound state (Figure
3.4C) and truncating the Tudor and SWIRM domains strongly reduces binding to
nucleosomes (Figure 3.4D). The two bound states might correspond to
complexes with different stoichiometries or represent a conformational change
that alters the electrophoretic mobility of the complex. Since nucleosomes
contain two copies of each histone and have pseudo two-fold symmetry, it is
tempting to speculate that two HAT module complexes might bind to each
nucleosome. Steady-state titrations reveal that all four HAT module complexes
acetylate nucleosomes with similar kinetic parameters (Figure 3.4E, 3.4F, 3.4G,
& 3.4H, Table 3.5), and that the binding defect observed with the
ΔTudorΔSWIRM construct translates into a threefold increase in $K_M$ (Table 3.5).
Thus, while truncating the Sgf29 tandem Tudor and Ada2 SWIRM domains
causes a marked decrease in nucleosome binding by the HAT module, the effect
on catalysis by Gcn5 is comparatively small.
Figure 3.4 The Sgf29 tandem Tudor and Ada2 SWIRM domains contribute to nucleosome binding

Native gels visualizing binding between fluorescein-tagged nucleosomes and (A) wild-type HAT module, (B) Sgf29ΔTudor HAT module, (C) Ada2ΔSWIRM HAT module, or (D) ΔTudorΔSWIRM HAT module. Steady-state kinetic titrations using nucleosome core particles and (E) wild-type HAT module, (F) Sgf29ΔTudor HAT module, (G) Ada2ΔSWIRM HAT module, or (H) ΔTudorΔSWIRM HAT module.
Table 3.5 Steady-state kinetic parameters for ΔSWIRM and ΔTudor HAT module on nucleosomal substrates

<table>
<thead>
<tr>
<th>HAT Module Construct:</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_M$ (µM)</th>
<th>$k_{cat}/K_M$ (µM$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>1.3 ± 0.1</td>
<td>1.9 ± 0.3</td>
<td>0.68 ± 0.1</td>
</tr>
<tr>
<td>ΔTudor</td>
<td>1.9 ± 0.2</td>
<td>2.8 ± 0.5</td>
<td>0.68 ± 0.1</td>
</tr>
<tr>
<td>ΔSWIRM</td>
<td>2.3 ± 0.08</td>
<td>2.4 ± 0.2</td>
<td>0.96 ± 0.09</td>
</tr>
<tr>
<td>ΔTudorΔSWIRM</td>
<td>2.0 ± 0.20</td>
<td>6.3 ± 1</td>
<td>0.32 ± 0.06</td>
</tr>
</tbody>
</table>

Summary of steady-state kinetic parameters fit from titrations where the concentration of nucleosome was varied while the acetyl-CoA concentration was held constant at 25 µM

Since the ΔSWIRM HAT module only populates one nucleosome-bound state on a native gel (Figure 3.4C), compared to two for the wild-type HAT module (Figure 3.4A), we wondered whether the Ada2 SWIRM domain might bind to nucleosomes. However, the isolated Ada2 SWIRM domain does not form a stable complex with nucleosome core particles on native gels (Figure 3.5). Therefore, the second band populated by HAT module-nucleosome complexes on native gels is not caused by a direct interaction between the Ada2 SWIRM domain and a second nucleosome.
Figure 3.5 The Ada2 SWIRM domain does not bind to nucleosomes

Native PAGE gel monitoring binding between nucleosomes (50 nM) and increasing concentrations of the Ada2 SWIRM domain.

Previous studies have shown that the presence of H3K4 trimethylation enhances the binding affinity between Sgf29 and histone peptides by 1-2 orders of magnitude (Bian et al, 2011). To test whether the presence of the H3K4me3 modification also affects HAT module affinity for nucleosomes, binding assays were performed using nucleosomes containing site-specific H3K4 trimethylation on native gels. The HAT module bound unmodified nucleosomes (Figure 3.6A), H3K4 trimethylated nucleosomes (Figure 3.6B), and even nucleosomes lacking the entire histone H3 N-terminal tail (Figure 3.6C) with the same apparent affinity, indicating that the H3 tail is not a primary binding determinant for the HAT module-nucleosome interaction.
The interaction between the HAT module and nucleosomes does not require the H3 tail

Native gel monitoring the interaction between the HAT module and (A) wild-type nucleosomes, (B) H3K4me3 nucleosomes, or (C) nucleosomes lacking the H3 N-terminal tail.

Figure 3.6

Does H3K4me3 recognition by Sgf29 recruit SAGA to gene promoters? In vitro binding data show that the HAT module interacts with unmodified and H3K4me3 nucleosomes with similar affinities (Figure 3.6A & 3.6B), which argues against H3K4me3 recognition governing SAGA recruitment. The sheer number of chromatin-interacting domains within the SAGA complex leaves open the possibility that other domains, which are not sensitive to the H3K4me3 modification, are just as important for substrate recruitment as the Sgf29 tandem Tudor domain. Mechanisms that recruit the SAGA complex to gene promoters exist in vivo that do not operate through Sgf29; at many genes, acidic activators
bound to upstream activating sequences are responsible for bringing SAGA to
gene promoters (Brown et al, 2001) rather than specific chromatin modifications.
Complicating the situation, chromatin-immunoprecipitation (ChIP) experiments
measuring SAGA recruitment in Sgf29 knockout strains yield different results
depending on which SAGA subunit is tagged (Bian et al, 2011; Shukla et al,
2011). Therefore, the effect of H3K4me3 recognition by Sgf29 on SAGA
recruitment in vivo remains to be determined.
Chapter 4: Structures of the catalytic domain of human Gcn5L2 bound to propionyl-CoA and butyryl-CoA

4.1 Abstract

Gcn5 is a well-studied enzyme that catalyzes lysine acetylation and propionylation \textit{in vitro}. Recent observations that lysines are modified by a number of chemically diverse acyl chains \textit{in vivo} have motivated further studies of the acyl specificity of enzymes like Gcn5. Here, structures were determined of the catalytic domain of human Gcn5L2 bound to propionyl-CoA and butyryl-CoA, demonstrating how Gcn5 discriminates between acyl chain donors and providing a structural explanation for why Gcn5 butyryltransferase activity is so weak.

4.2 Introduction

Lysine acetylation is an abundant and conserved post-translational modification (Choudhary et al, 2009; Weinert et al, 2011) that changes the overall size and charge of the modified residue. Several classes of enzymes are known to catalyze site-specific lysine acetylation (Marmorstein & Trievel, 2009; Roth et al, 2001; Yang, 2004), many of which localize to the nucleus and alter the post-translational modifications present on histones (Lee & Workman, 2007). These enzymes are collectively referred to as either \textit{Histone AcetylTransferases} (HATs) or Lysine (K) \textit{AcetylTransferases} (KATs), the latter to reflect their ability to acetylate non-histone substrates (Glozak et al, 2005). KATs are divided into several main families based on structural homology within their catalytic domains and the presence of short conserved sequence motifs (Marmorstein & Trievel,
2009). Although different KAT families employ distinct kinetic mechanisms to catalyze acetyl transfer, they all share a common dependence on the nucleotide cofactor, acetyl-CoA, as an acyl chain donor (Berndsen & Denu, 2008).

Gcn5 is a member of the GNAT (Gcn5-related-N-Acetyltransferase) family of histone acetyltransferases and is responsible for acetylating the N-terminal tails of histones H3 and H2B at the promoters of inducible genes (Grant et al, 1997). Not only has Gcn5 been broadly implicated in transcriptional regulation (Huisinga & Pugh, 2004), but its kinetic mechanism has been studied extensively (Tanner et al, 2000b; Tanner et al, 1999) and its catalytic domain has been crystallized from several organisms in the presence of various combinations of substrates (Poux et al, 2002; Roth et al, 2001).

A number of high-energy acyl chain donors in the form of acyl-CoAs exist in cells (King & Reiss, 1985), which may be substrates for KATs like Gcn5. Supporting this possibility, propionyl and butyryl chains, which differ in length from acetylation by one or two carbons respectively, are present on histone lysines in vivo (Chen et al, 2007; Zhang et al, 2009b), albeit at lower stoichiometries compared to acetylation (Zhang et al, 2009b). Some of these acyl-CoA molecules exist at sufficiently high concentrations to affect the activity of KATs, which may link lysine acylation to metabolic fluctuations (Albaugh et al, 2011). Members of three different KAT families are already known to catalyze lysine acylation in vitro: p300/CBP, which catalyzes propionylation and butyrylation of histones and p53 (Chen et al, 2007); yeast Esa1, a member of the MYST family of acetyltransferases that catalyzes propionylation of histone H4
peptides (Berndsen et al, 2007); and human P/CAF, a member of the GNAT family that robustly catalyzes propionylation of histone peptides \textit{in vitro} (Leemhuis et al, 2008).

4.2 Methods

\textit{Protein expression and purification}

A plasmid encoding the his-tagged catalytic domain of human Gcn5L2 (hsGcn5L2) under T7-induction was obtained from Addgene (Plasmid No.: 25482). The protein purified as previously described (Schuetz et al, 2007). Purified protein was dialyzed into 20 mM HEPES, pH 7.5, 150 mM NaCl, and 1 mM DTT, concentrated to 9 mg/mL, flash-frozen in liquid nitrogen, and stored at -80°C.

\textit{Acyltransferase assay}

Rates of propionylation and butyrylation were determined using a spectrophotometric assay monitoring free CoA production. Reactions contained 10 µM hsGcn5L2 catalytic domain, 250 µM histone H3 peptide aa1-21 (purchased from United Peptide at >90% purity), 100 mM HEPES, pH 7.6, 50 mM NaCl, and 500 µM acetyl-CoA (Sigma No.: A2181), propionyl-CoA (Sigma No.: P5397) or butyryl-CoA (Sigma No.: B1508). Reactions were incubated for five minutes at 37°C before adding acyl-CoA, and then maintained at 37°C for the remainder of the experiment. Initially, six data points were collected to find a
time frame where acyl-CoA consumption was linear with time. The reaction was quenched at the indicated time points by the addition of two volumes of quenching buffer (3.2 M guanidine-HCl and 100 mM sodium phosphate pH 6.8). After all the samples had been collected, one volume of 4 mM 5,5’-dithiobis-(2-nitrobenzoic acid) (DTNB) (Sigma No.: D218200) dissolved in 100 mM sodium phosphate, pH 6.8 was added. Samples were moved to a 384-well polystyrene clear-bottom plate (Grenier Bio-One) and absorbance at 412 nM was measured in a POLARstar Omega plate reader (BMG Labtech). Absorbances were converted to concentrations using a standard curve generated by reacting increasing concentrations of CoA (Sigma No.: C3019) with DTNB using an extinction coefficient for 3-thio-6-nitrobenzoate (TNB) of $\varepsilon_{412nm} = 13,700 \text{ M}^{-1} \text{ cm}^{-1}$.

Subsequent reactions were conducted by collecting three data points in triplicate within the linear time frame.

**HAT Domain Crystallization**

Propionyl-CoA and butyryl-CoA were diluted in 20 mM HEPES, pH 7.5 and stored at -20°C at a concentration of 20 mM calculated using $\varepsilon_{260nm} = 16,400 \text{ M}^{-1} \text{ cm}^{-1}$. Purified human Gcn5L2 aa497-662 was mixed with each acyl-CoA to a final concentration of 1.6 mM acyl-CoA and 7.9 mg/mL protein. NaCl was added to a final concentration of 125 mM from a 5 M stock, and the resulting mixture was incubated on ice for 30 minutes. Both complexes were crystallized using hanging drop vapor diffusion by mixing 1 µL of protein:acyl-CoA complex with 1 µL well solution. Human Gcn5L2 bound to propionyl-CoA was crystallized in 20% (v/v)
ethanol and 100 mM Tris, pH 9.0. Human Gcn5L2 bound to butyryl-CoA was crystallized in 10% (v/v) 2-propanol, 3% glycerol, 100 mM HEPES, pH 7.8, and 11% (w/v) PEG 4,000. Crystals were cryoprotected by soaking in well solution supplemented with 9% sucrose, 4% glucose, 8% ethylene glycol, and 8% glycerol. Prior to data collection, crystals were flash frozen in a liquid nitrogen stream.

Data collection and processing

Diffraction data were collected using a Rigaku FR-E SuperBright x-ray generator at a wavelength of 1.54 Å and recorded with a Saturn 944+ CCD detector. Data were processed with HKL2000 (Otwinowski & Minor, 1997). The structures were solved using molecular replacement with MOLREP from the CCP4 suite using the coordinates of human Gcn5L2 (PDB ID 1Z4R) as a search model (Vagin & Teplyakov, 1997; Vagin & Teplyakov, 2010). Refinement was carried out using REFMAC5 from the CCP4 suite (Murshudov et al, 1997; Winn et al, 2011) and the graphics program COOT for model-building (Emsley & Cowtan, 2004). Simulated annealing omit maps were generated by removing either propionyl- or butyryl-CoA from the refined structures, fitting acetyl-CoA into the ligand density, and performing three rounds of refinement with Phenix including two cycles of simulated annealing (Adams et al, 2010).
4.3 Results and discussion

To ask whether human Gcn5, whose catalytic domain shares 95% sequence identity and 99% sequence similarity to P/CAF, also functions as a propionyltransferase, its rate of acyl-CoA consumption was measured in the presence of histone H3 peptide, blanked by the rate of acyl-CoA hydrolysis in the absence of peptide. Consistent with earlier measurements on P/CAF (Leemhuis et al, 2008), human Gcn5L2 efficiently acetylates and propionylates peptides, while its butyrylating activity is nearly undetectable (Figure 4.1). Under these experimental conditions, where peptide concentration is held constant at 250 µM, human Gcn5L2 propionylates histone peptides approximately nine-fold more slowly and butyrylates peptides nearly 70-fold more slowly compared to its acetyltransferase activity (Figure 4.1).
Figure 4.1 Acyltransferase activity of hsGcn5L2

Rate of catalysis by hsGcn5L2 (10 µM) using different acyl-CoA cofactors (500 µM) and N-terminal histone H3 peptide (250 µM) containing the sequence: ARTKQTARKSTGGKAPRKQLA.

Although Gcn5L2 clearly discriminates between different acyl-CoA cofactors, structural studies describing how its active site accommodates longer acyl chains or why its activity drops precipitously with increasing acyl chain length have never been done. Therefore the structure of human Gcn5L2, which had been previously crystallized bound to acetyl-CoA (Schuetz et al, 2007), was
solved in complex with propionyl-CoA and butyryl-CoA to 2.0 Å and 2.1 Å, respectively. Refinement statistics for each structure are summarized in Table 4.1. For both ligands, simulated annealing omit maps show clear density corresponding to the extra methyl group for propionyl-CoA (Figure 4.2A) or extra ethyl chain for butyryl-CoA (Figure 4.2B). Compared to the structure of hsGcn5L2 bound to acetyl-CoA (Schuetz et al, 2007), the structures reported here are very similar; the Cα RMSD (calculated using PDBefold from the EMBL-EBI website) is only 0.13 Å for the structure of human Gcn5L2 in complex with propionyl-CoA and 0.24 Å for its structure in complex with butyryl-CoA. This suggests that the catalytic domain of Gcn5L2 is rigid, and that the shape of its active site limits the types of acyl chains it can accommodate.
Table 4.1 Summary of data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection and refinement statistics</th>
<th>hsGcn5L2 bound to a propionylated peptide</th>
<th>hsGcn5L2 bound to a butyrylated peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diffraction Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P6₁</td>
<td>P6₁</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54</td>
<td>1.54</td>
</tr>
<tr>
<td>Unit Cell Dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a, b, c$ (Å)</td>
<td>38.09, 38.09, 187.06</td>
<td>38.25, 38.25, 186.97</td>
</tr>
<tr>
<td>$\alpha, \beta, \gamma$ (°)</td>
<td>90.0, 90.0, 120.0</td>
<td>90.0, 90.0, 120.0</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>29.17-2.00 (2.07-2.00)</td>
<td>24.81-2.08 (2.15 - 2.08)</td>
</tr>
<tr>
<td>Measured Reflections</td>
<td>51218 (3583)</td>
<td>58587 (2201)</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>10102 (915)</td>
<td>9086 (819)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.98 (92.71)</td>
<td>98.06 (87.88)</td>
</tr>
<tr>
<td>Average I/σ (merged reflections)</td>
<td>13.91 (5.68)</td>
<td>16.86 (5.39)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>5.1 (3.9)</td>
<td>6.4 (2.7)</td>
</tr>
<tr>
<td>$R_{merge}$ (%)</td>
<td>8.68 (31.29)</td>
<td>7.70 (18.61)</td>
</tr>
<tr>
<td><strong>Refinement Statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Atoms</td>
<td>1509</td>
<td>1506</td>
</tr>
<tr>
<td>Protein</td>
<td>1328</td>
<td>1356</td>
</tr>
<tr>
<td>Ligands</td>
<td>68</td>
<td>61</td>
</tr>
<tr>
<td>Water</td>
<td>113</td>
<td>89</td>
</tr>
<tr>
<td>CC1/2</td>
<td>0.996 (0.898)</td>
<td>0.997 (0.942)</td>
</tr>
<tr>
<td>CC*</td>
<td>0.999 (0.973)</td>
<td>0.999 (0.985)</td>
</tr>
<tr>
<td>$R_{work}$ (%) / $R_{free}$ (%)</td>
<td>17.1/ 19.7</td>
<td>16.4/ 20.7</td>
</tr>
<tr>
<td><strong>B-factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>19.30</td>
<td>19.90</td>
</tr>
<tr>
<td>Ligands</td>
<td>24.30</td>
<td>22.00</td>
</tr>
<tr>
<td>Solvent</td>
<td>31.20</td>
<td>29.80</td>
</tr>
<tr>
<td><strong>RMSD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.007</td>
<td>0.008</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.08</td>
<td>1.27</td>
</tr>
</tbody>
</table>

Summary of refinement statistics for the structures of hsGcn5L2 bound to propionyl-CoA and butyryl-CoA. Values in parentheses correspond to data from the highest resolution shell.
The active site of Gcn5 contains three features that facilitate transfer of the acyl chain to lysine: an active site glutamate that functions as a general base (Tanner et al, 1999), a structurally conserved water molecule that forms a proton wire between the general base and incoming lysine (Rojas et al, 1999), and residues that stabilize the position of the acyl-CoA (Rojas et al, 1999; Schuetz et al, 2007). This active site geometry is preserved in the two acyl-CoA bound structures reported here, including the orientation of the acyl-CoA thioester, which is coordinated by the backbone amide of cysteine 579, and the position of the water molecule, which is hydrogen bonded to glutamate 575 (Figure 4.3A, 4.3B, & 4.3C). Acetyl- and propionyl-CoA bind in similar conformations, although the propionyl-CoA thioester rotates relative to the enzyme to fit the extra methyl group into the Gcn5 active site (Figure 4.3B). In contrast, butyryl-CoA binds in an
orientation where the terminal methyl group projects toward the solvent, since the catalytic water molecule blocks it from occupying the Gcn5 active site cleft (Figure 4.3C). Although longer acyl chains could also bind in this orientation, molecules like crotonyl-CoA, which contain unsaturated carbon-carbon bonds that do not freely rotate, cannot adopt conformations compatible with this geometry.

**Figure 4.3 Close-up view of the hsGcn5L2 active site**

The catalytic water molecule is shown as a blue sphere and the acyl-CoA molecules are depicted in stick representation. (A) Structure of hsGcn5L2 (purple) bound to acetyl-CoA (green). (B) Structure of hsGcn5L2 (pink) bound to propionyl-CoA. (C) Structure of hsGcn5L2 (gray) bound to butyryl-CoA (orange).
Models of human Gcn5L2 bound to each acyl-CoA molecule and a histone peptide (Figure 4.4A) were generated using the structure of tetrahymena Gcn5 bound to a bisubstrate analog to position the incoming lysine residue (PDB ID 1M1D) (Poux et al, 2002). For the structure of human Gcn5L2 bound to acetyl-CoA, the incoming lysine residue makes an angle of 105° with the acetyl thioester (Figure 4.4B), which is a reasonable angle of attack for a carbonyl group by a nucleophile (Burgi et al, 1974) and supports the validity of this model. Propionyl-CoA also adopts a conformation compatible with this angle of attack, as the position of the terminal methyl is in the same plane as the thioester, which leaves the lysine attack trajectory open (Figure 4.4C). In contrast, the terminal methyl of butyryl-CoA projects into the channel occupied by the lysine (Figure 4.4D). This explains why Gcn5 is a poor butyryltransferase, as the acyl chain, catalytic water molecule (Figure 4.3C), and incoming lysine (Figure 4.4D) cannot all fit into its active site.
Figure 4.4 Model of ternary complex between hsGcn5L2, different acyl-CoA molecules, and peptide

(A) Overall model with hsGn5L2 colored purple, the peptide colored yellow, and acetyl-CoA colored green. The catalytic water molecule is depicted as a blue sphere. Close-up views of the arrangement between the incoming lysine (yellow) and (B) acetyl-CoA shown in green, (C) propionyl-CoA shown in cyan, or (D) butyryl-CoA shown in orange.

Here, we have determined crystal structures that describe how Gcn5 accommodates propionyl-CoA in its active site, and provide a structural mechanism by which it discriminates between different acyl-CoA molecules. This work suggests that some naturally occurring acyl-CoA molecules, like butyryl-CoA, could inhibit Gcn5 activity by binding to the enzyme in a way that prevents lysine from entering its active site. As intracellular ratios of different acyl-CoAs change in response to metabolic activity (Hosokawa et al, 1986; King & Reiss, 1985; Palladino et al, 2012), this study raises the question whether fluctuating
levels of acyl-CoA molecules may also regulate the activity of Gcn5 in vivo. Since free CoA (Tanner et al, 2000b) and CoA-peptide conjugates (Poux et al, 2002) potently inhibit Gcn5 in vitro, the structures presented here provide a clue as to how other CoA-based molecules may be exploited to design future generations of acetyltransferase inhibitors.
Chapter 5: HAT Module Crystallization Trials

5.1 Introduction

The way in which the non-catalytic subunits of the SAGA HAT module, Ada2, Ada3, and Sgf29, activate Gcn5 and regulate its acetyltransferase activity has never been elucidated on a molecular level. I therefore attempted to crystallize the HAT module with the goal of determining its 3-dimensional structure. This chapter summarizes crystallization attempts for various forms of the SAGA HAT module and HAT module sub-complexes.

5.2 Summary of Crystallization Trials

Crystallization trials of the S. cerevisiae HAT Module

HAT module containing Ada3 aa185-703, Gcn5, and Ada2 from S. cerevisiae was recombinantly expressed in E. coli from the polycistronic pST44 vector (Dr. Song Tan, Penn State, PA) and purified as previously described (Balasubramanian et al, 2002; Barrios et al, 2007). Complexes for crystallography were further purified over a Sephacryl S-300 HR 16/60 column equilibrated in 20 mM HEPES, pH 8.0, 100 mM NaCl, 10% Glycerol, and 5 mM BME. Polycistronic vectors containing N-terminal truncations in Ada3 (aa364-703 or 452-703) were produced using QuickChange Site-Directed Mutagenesis (Agilent Technologies) and purified in the same way. Purified complexes (Figure 5.1) were concentrated to approximately 5mg/mL and subjected to sparse matrix screening at room temperature using Crystal Screen HT (Hampton Research), JCSG+ (Qiagen), ProComplex I & II (Qiagen), and Wizard I & II (Rigaku). However, the complex did not crystallize under the conditions tested.
**Figure 5.1** SDS-PAGE analysis of *S. cerevisiae* HAT module

*Crystallization trials of the *pombe* HAT module*

*S. pombe* HAT module was recombinantly expressed and purified as described in Chapter 2.3 and Chapter 3.2. Truncating the Ada2 SWIRM domain dramatically increased the yield, homogeneity, and purity of the purified HAT module (data not shown). Therefore, complexes lacking the Ada2 SWIRM domain and containing various other domain truncations (Figure 5.2) were concentrated to 4-8 mg/mL and screened at room temperature using Crystal Screen HT (Hampton Research) and PEG Rx (Hampton Research). Complexes vi and viii (Figure 5.2) were also screened with Hampton Crystal Screen Lite (Hampton Research) and JCSG+ (Qiagen) with and without equimolar amounts of H3K4me3 aa1-11 peptide. Complex iv (Figure 5.2) was also screened with JCSG+ (Qiagen), MIDAS (Molecular Dimensions), ProComplex (Qiagen), and Wizard I & II (Rigaku).
Attempts were also made to crystallize the HAT module bound to a nucleosome core particle. Crystallographic quantities of nucleosome core particles assembled with a 147 bp DNA fragment containing the Widom-601 strong positioning sequence (Lowary & Widom, 1998) were prepared as previously described (Luger et al, 1999) and purified over a TSKgel DEAE-5PW column (Tosoh Biosciences, LLC). These were mixed with construct viii (Figure 5.2) at 1:1, 1:2, and 2:1 stoichiometries, in the presence and absence of equimolar amounts of H3K4me3 aa1-11 peptide, and screened using Natrix HT (Hampton Research) and Nucleix (Qiagen) at 20°C and 4°C. Crystals containing the HAT module were never obtained, but nucleosomes alone crystallized out of the mixture in several conditions.
Crystallization trials of the Gcn5/Ada2 complex

The first 120 amino acids of *S. pombe* Ada2 contain zinc finger and SANT domains, which are sufficient to form a complex with Gcn5 (Figure 3.3A) and stimulate acetylation by Gcn5 on histone peptides (Figure 3.3B, Table 3.3). To understand how the ZnF-SANT domain of Ada2 regulates acetylation by Gcn5, heterodimeric complexes of the Ada2 ZnF-SANT and Gcn5 were recombinantly prepared and used for crystallization trials. Complexes were cloned from budding yeast genomic DNA (*Saccharomyces cerevisiae*), fission yeast genomic DNA (*Schizosaccharomyces pombe*), human placental cDNA (*Homo sapiens*), soybean cDNA (*Glycine max*), and *Arabidopsis thaliana* cDNA into pETduet, a vector containing two multiple cloning sites that is suitable for recombinant protein expression in *E. coli* (Figure 5.3). To determine which complexes expressed to high levels in *E. coli*, induction trials were performed in Rosetta2-(DE3) cells (EMD Millipore) using 1 L each of M9ZB medium (Studier, 2005) for the *S. pombe*, *S. cerevisiae*, *G. max*, and *A. thaliana* constructs. Cells were grown at 37°C, induced at an OD600 of 2-3 with 0.2 mM IPTG, and incubated overnight at 15°C. Cells were harvested by centrifugation, resuspended in 500 mM NaCl, 40 mM sodium phosphate, pH 7.4, 5 mM BME, 20 mM imidazole, and 10% glycerol, lysed using a Microfluidizer (Microfluidics Corp), and clarified by centrifugation at 32,000 g. The soluble fraction was bound to 0.5 mL of HisPur resin (ThermoScientific) pre-equilibrated in resuspension buffer and incubated at 4°C for an hour with gentle agitation. The beads were eluted according to
standard protocols and samples corresponding to the insoluble fraction, soluble fraction, unbound fraction, and eluted fraction were run on an SDS-PAGE gel. Of the eight complexes tested, *A. thaliana* Ada2 ZnF-SANT/Gcn5Δ97 had the highest level of expression (Figure 5.4). Therefore, this complex was chosen for large-scale purification and crystallization trials, along with the *S. pombe* Ada2 ZnF-SANT/Gcn5 complex.
Figure 5.3 Summary of Ada2 ZnF-SANT/ Gcn5 complexes for crystallography
Figure 5.4 Nickel-NTA pull-down of Ada2 ZnF-SANT/ Gcn5 complexes
Both complexes (A. thaliana Ada2 ZnF-SANT/Gcn5Δ97 and S. pombe Ada2 ZnF-SANT/Gcn5) were purified according to similar protocols. The proteins were expressed in Rosetta2-(DE3) cells (EMD Millipore) using 8 L of Terrific Broth (TB) for the S. pombe complex and 8 L of M9ZB medium (Studier, 2005) for the A. thaliana complex. Cells were grown at 37°C and moved to an ice bath for 40 minutes after reaching an OD600 of 2.5 for M9ZB cultures and 0.5 for TB cultures. After a 45 minute incubation on ice, protein expression was induced with 0.2 mM IPTG overnight at 15°C. Cells were harvested by centrifugation and resuspended in lysis buffer containing 500 mM NaCl, 40 mM HEPES, pH 7.5, 5 mM BME, 20 mM imidazole, pH 8, and 10% glycerol. Lysates were prepared using a Microfluidizer (Microfluidics Corp) and clarified by centrifugation at 32,000 g. The soluble portion was run over a 5 mL HisTrap HP (GE LifeSciences) equilibrated in lysis buffer, and eluted using a 20-270 mM imidazole gradient over 5 column volumes. Fractions containing both proteins were pooled and dialyzed overnight at 4°C into 100 mM NaCl, 20 mM HEPES, pH 7.5, and 5 mM BME in the presence of 2-4 mg TEV protease. Following dialysis, imidazole was added to a final concentration of 20 mM and the sample was re-run over the HisTrap HP to remove the protease and cleaved tag. For the S. pombe complex, the salt concentration of the flow-through was diluted to 100 mM NaCl using 20 mM HEPES, pH 7.5, 5 mM BME and loaded onto a Q HP (GE LifeSciences) equilibrated in 20 mM HEPES, pH 7.5, 100 mM NaCl, and 5 mM BME. The complex was eluted using a 100-500 mM NaCl gradient over 10 column volumes. A. thaliana complex was not purified using ion exchange
chromatography. Finally, both complexes were concentrated and run over a HiPrep Superdex 200 26/60 (GE LifeSciences) equilibrated in 20 mM HEPES, pH 7.6, 100 mM NaCl, and 500 µM TCEP. Fractions containing pure S. pombe complex were concentrated to 25.7 mg/mL based on an extinction coefficient of 67,730 M⁻¹ cm⁻¹ or 1.032 mg⁻¹ ml cm⁻¹ calculated using the ProtParam tool from the ExPASy Bioinformatics Resource Portal (www.expasy.org) and flash-frozen in liquid nitrogen for long-term storage at -80°C. Fractions containing pure A. thaliana complex were concentrated to 48.5 mg/mL based on an extinction coefficient of 70,250 M⁻¹ cm⁻¹ or 0.990 mg⁻¹ ml cm⁻¹ and flash-frozen in liquid nitrogen for long-term storage at -80°C. The final yield of protein was 1.9 mg/L for the S. pombe complex and 8.3 mg/L for the A. thaliana complex.

Crystallization trials of the heterodimeric Ada2 ZnF-SANT/Gcn5 complex from S. pombe were conducted at room temperature using a 1:1 drop ratio with Hampton Crystal Screen HT (Hampton Research), JCSG+ (Qiagen), and ProComplex (Qiagen). Crystallization trials were also conducted using Index HT (Hampton Research) and JCSG+ (Qiagen) in the presence of acetyl-CoA. Excess acetyl-CoA was added to the complex at a 1.6-fold molar ratio and incubated overnight at 4°C before setting the trays. Crystals were never obtained for the S. pombe Ada2 ZnF-SANT/Gcn5 complex.

Crystallization trials of the heterodimeric Ada2 ZnF-SANT/Gcn5Δ97 complex from A. thaliana were conducted at two protein concentrations: 400 µM or 28.4 mg/mL of the complex or at 200 µM or 14.2 mg/mL of the complex. For each protein concentration, four conditions were tested for a total of eight
combinations: the complex alone, the complex in the presence of 1 mM acetyl-CoA trilithium salt, the complex in the presence of 1.5 mM H3 aa1-21 peptide, and the complex in the presence of both 1 mM CoA and 1.5 mM H3 aa1-21 peptide. These combinations were screened at room temperature using Hampton HT (Hampton Research), Wizard HT (Rigaku Reagents), and JCSG+ (Qiagen). Crystals of the *A. thaliana* Ada2 ZnF-SANT/Gcn5Δ97 complex were never obtained.

*Crystallization trials of the Ada3/Sgf29 complex*

To determine the molecular architecture of the Ada3/Sgf29 portion of the SAGA HAT module, which form a separable and stable complex (Figure 3.3D), crystallographic studies were attempted. Although the proteins corresponding to the N-terminus of Ada3 and Sgf29 are both very soluble and express well in *E. coli*, one major obstacle to co-expressing the Ada3/Sgf29 complex was non-stoichiometric expression of the two subunits (data not shown). Therefore, Ada3 and Sgf29 were cloned into many different expression vectors (pET32a, CDFduet, pCDF-1b, pRSF, pMAL, etc.), and various combinations of compatible plasmids were subjected to small-scale co-expression trials followed by nickel-NTA pull-downs to figure out conditions that produced similar amounts of Ada3 and Sgf29 (data not shown). The plasmid combinations that generated stoichiometric complexes are listed in Table 5.1.
Table 5.1 Constructs for crystallizing the Ada3/Sgf29 sub-complex

<table>
<thead>
<tr>
<th>Plasmid Information</th>
<th>Notes about expression and purification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vector</strong></td>
<td><strong>spAda3</strong></td>
</tr>
<tr>
<td>pET32a</td>
<td>aa1-135</td>
</tr>
<tr>
<td>pET32a</td>
<td>aa1-135</td>
</tr>
<tr>
<td>pMAL</td>
<td>MBP-Ada3 aa1-135</td>
</tr>
<tr>
<td>pRSF</td>
<td>aa1-135</td>
</tr>
<tr>
<td>CDFduet</td>
<td>aa1-135</td>
</tr>
<tr>
<td>CDFduet</td>
<td>T4 Lysozyme Ada3 aa1-135</td>
</tr>
<tr>
<td>CDFduet</td>
<td>T4 Lysozyme Ada3 aa7-135</td>
</tr>
<tr>
<td>CDFduet</td>
<td>T4 Lysozyme Ada3 aa1-135</td>
</tr>
<tr>
<td>CDFduet</td>
<td>T4 Lysozyme Ada3 aa7-135</td>
</tr>
</tbody>
</table>

Complexes for crystallographic screening were produced in Rosetta2-DE3 cells (EMD Millipore), which were grown in 2-6 L of M9ZB culture medium (Studier, 2005) and induced for 16-18 hours with 0.2 mM IPTG at 16°C. Plasmids under T7-induction (pET32a, pRSF, and CDFduet) were inducted at an OD600nm of 2-3, but cultures containing pGEX-based vectors (pMAL) were induced at an
OD\textsubscript{600nm} of 0.6. After induction, cells were harvested by centrifugation and resuspended in lysis buffer containing 600 mM NaCl, 50 mM HEPES, pH 7.6, 5 mM BME, 20 mM imidazole, and 10% glycerol. Lysates were prepared using a Microfluidizer (Microfluidics Corp) and clarified by centrifugation at 32,000 \textit{g}. The soluble fraction was passed over a HisTrap HP (GE LifeSciences) equilibrated in lysis buffer, and eluted with a 20-200 mM imidazole gradient over 10 column volumes. Fractions containing the complexes were dialyzed overnight at 4°C into 100 mM NaCl, 20 mM HEPES, pH 7.6, 5 mM BME, and 10% glycerol in the presence of 4-8 mg TEV protease. After dialysis, imidazole was added to a final concentration of 20 mM and the protein was re-passed over the HisTrap HP. The flow-through from the HisTrap HP was collected and purified by gel filtration for the MBP-tagged complexes, or subjected to ion exchange for the other complexes (see Table 5.1 for details). For ion exchange, the flow-through was diluted six-fold with 20 mM Tris, pH 7.6 and 5 mM DTT and loaded onto a Q HP (GE LifeSciences) or Sp HP (GE LifeSciences) equilibrated in 20 mM Tris, pH 7.6, 20 mM NaCl, and 5 mM DTT (see table 5.1 for details). Complexes were eluted using a 20-500 mM NaCl gradient over 15 column volumes. Fractions containing the complexes were pooled and loaded onto a HiLoad Superdex200 26/60 gel filtration column equilibrated in 10 mM HEPES, pH 7.6, 100 mM NaCl, and 5 mM DTT. Purified proteins (excluding the MBP-tagged complexes) were dialyzed at 4°C into 10 mM HEPES, pH 7.6, 50 mM NaCl, and 5 mM DTT. The MBP-tagged complexes were dialyzed into 15 mM HEPES, pH 7.6, 100 mM NaCl, 5 mM DTT, and 5 mM \textit{D}-maltose. After dialysis, the complexes were
concentrated to 25-40 mg/mL and flash-frozen in liquid nitrogen for long-term storage at -80°C. Crystallization trials with the Ada3/Sgf29 complexes were conducted at room temperature using a 1:1 drop ratio with Crystal Screen HT (Hampton Research), JCSG + (Qiagen), Wizard I & II (Rigaku Reagents), Index HT (Hampton Research), ProComplex (Qiagen), PEG Rx (Hampton Research), and Salt Rx (Hampton Research). Crystals were not obtained for any of the Ada3/Sgf29 complexes.

4.3 Future Directions

How the non-catalytic subunits of the SAGA HAT module regulate the activity of Gcn5 remains an open question. Future structural studies of the entire HAT module, or pieces of the HAT module, will likely explain why four subunits are required for the full catalytic activity of Gcn5 and elucidate the molecular mechanisms governing the substrate specificity of the HAT module. Even though the catalytic domains of many HATs have been crystallized (Roth et al, 2001; Vetting et al, 2005), structures of HAT complexes have remained elusive.

Here, several strategies were employed to produce crystals of the SAGA HAT module. Prior to this work, studies from other labs had demonstrated that the HAT module from *S. cerevisiae* could be recombinantly co-expressed in bacteria and purified to form active complex consisting of Gcn5, Ada2, and Ada3 (Balasubramanian et al, 2002; Barrios et al, 2007). Thus, this complex was used as a starting point for crystallization trials. However, attempts to make a stable four-protein complex containing Gcn5, Ada2, Ada3, and Sgf29 in bacteria were not successful (data not shown). Therefore, constructs containing *S. pombe*
homologues were cloned, co-expressed in bacteria, and purified to homogeneity. Further crystallization trials were conducted using the full-length \textit{S. pombe} complex, as well as minimal complexes containing domain truncations. Since truncating the Ada2 SWIRM domain results in a complex that binds to nucleosomes as a single species (Figure 3.4C), crystallization trials containing \textit{ΔSWIRM} HAT module and nucleosome core particles were performed. Smaller portions of the HAT module corresponding to the Ada3/Sgf29 subcomplex and Gcn5/Ada2 subcomplex were also subjected to crystallization trials. After screening Gcn5/Ada2 complexes from a number of organisms, one construct from \textit{A. thaliana} containing Gcn5\textit{Δ97} and Ada2 aa1-158 expressed to high levels and was used for extensive crystallization screening. Although crystals were never obtained for any of these complexes, many of these constructs express to high levels in bacteria and form stable complexes \textit{in vitro}. Additional screening could still be done, particularly with new fragments identified by limited proteolysis or with homologues cloned from other organisms.
Chapter 6: Alternate deacylating specificites for the archaeal enzymes Sir2Af1 and Sir2Af2

6.1 Abstract

Sirtuins were originally shown to regulate a wide array of biological processes such as transcription, genomic stability, and metabolism by catalyzing the NAD$^+$-dependent deacetylation of lysine residues. Recent proteomic studies have revealed a much wider array of lysine acyl modifications in vivo than was previously known, which has prompted a reevaluation of sirtuin substrate specificity. Several sirtuins have now been shown to preferentially remove propionyl, succinyl and long-chain fatty acyl groups from lysines, which has changed our understanding of sirtuin biology. In light of these developments, we revisited the acyl specificity of several well-studied archaeal and bacterial sirtuins. We find that the *Archaeoglobus fulgidus* sirtuins, Sir2Af1 and Sir2Af2, preferentially remove succinyl and myristoyl groups, respectively. Crystal structures of Sir2Af1 bound to a succinylated peptide and Sir2Af2 bound to a myristoylated peptide show how the active site of each enzyme accommodates a non-canonical acyl chain. As compared to its structure in complex with an acetylated peptide, Sir2Af2 undergoes a conformational change that expands the active site to accommodate the myristoyl group. These findings point to both structural and biochemical plasticity in sirtuin active sites and provide further evidence that sirtuins from all three domains of life catalyze non-canonical deacylation.
6.2 Introduction

Lysine acetylation is an abundant and well-studied post-translational modification that regulates a broad array of cellular processes in organisms ranging from bacteria to mammals (Choudhary et al, 2009; Lundby et al, 2012; Zhang et al, 2009a). In addition to identifying specific acetylation sites (Jensen, 2006; Olsen & Mann, 2013), advances in mass spectrometry have also revealed a surprisingly broad array of lysine acyl modifications in vivo whose chemical properties differ from acetylation. Studies in eukaryotes have identified modifications such as propionylation (Chen et al, 2007), butyrylation (Chen et al, 2007), crotonylation (Tan et al, 2011), malonylation (Peng et al, 2011; Xie et al, 2012), succinylation (Zhang et al, 2011), and glutarylation (Tan et al, 2014) decorating lysine residues. Although it remains to be seen how many of these acyl modifications are conserved throughout evolution, widespread lysine succinylation has been detected in E. coli (Colak et al, 2013; Weinert et al, 2013) and lysine propionylation has been discovered in Salmonella enterica (Garrity et al, 2007). Acyl chain identity is increasingly recognized as an additional factor in regulating cellular activities, and there are now examples of lysine residues that can be modified by more than one type of acyl chain in vivo (Du et al, 2011; Rardin et al, 2013; Weinert et al, 2013). Identifying the enzymes that attach and remove different types of lysine acyl modifications is therefore a key aspect in elucidating regulatory pathways.
Sirtuins were initially identified as a class of NAD\textsuperscript{+}-dependent deacetylases conserved across all three domains of life (Frye, 2000; Greiss & Gartner, 2009; Smith et al, 2000), where they regulate a remarkably broad range of cellular processes. Sirtuins catalyze lysine deacetylation by cleaving nicotinamide from NAD\textsuperscript{+} and generating a mixture of 2’ and 3’-O-acetyl-ADP ribose in addition to the unmodified lysine (Figure 6.1). The founding member of the sirtuin family, yeast Sir2, regulates transcriptional silencing by deacetylating histones and promoting the formation of heterochromatin (Braunstein et al, 1993). CobB, a bacterial homolog of yeast Sir2, regulates signal transduction pathways (Li et al, 2010), the activity of metabolic enzymes (Garrity et al, 2007; Starai et al, 2002; Zhang et al, 2013b) and gene transcription (Hu et al, 2013; Lima et al, 2011; Thao et al, 2010). The seven human sirtuins also regulate a wide variety of processes including the response to oxidative stress, fatty acid oxidation, genome stability and transcription (Michan & Sinclair, 2007; Vassilopoulos et al, 2011). Because of the central roles that sirtuins play in so many critical pathways across biology, these enzymes have been the subjects of intense study (Houtkooper et al, 2012; Michan & Sinclair, 2007; Sauve et al, 2006).
Although all biological functions of sirtuins were originally ascribed to their deacetylation activity, recent studies have shown that several sirtuins are more active \textit{in vitro} in removing other types of lysine acyl modifications (Du et al, 2011; Jiang et al, 2013; Tan et al, 2014; Zhu et al, 2011), and that these alternate specificities are important for their biological roles (Du et al, 2011; Jiang et al, ...
In some cases, sirtuins that have little or no deacetylase activity in vitro were shown to be more active against longer or charged acyl chains, pointing to acyl group identity as an important component of sirtuin specificity (Du et al, 2011; Feldman et al, 2013; Jiang et al, 2013; Peng et al, 2011; Zhu et al, 2011). For example, human SIRT5 preferentially removes negatively charged succinyl, malonyl and glutaryl acyl groups from lysines (Du et al, 2011; Peng et al, 2011; Tan et al, 2014). This activity is important for SIRT5 regulation of the enzyme, carbamoyl phosphate synthase (CPS1), which is desuccinylated (Du et al, 2011) and deglutarylated (Tan et al, 2014) by SIRT5 in vivo. Desuccinylation by SIRT5 also regulates the activity of the pyruvate dehydrogenase and succinate dehydrogenase complexes in mouse embryonic fibroblasts (Park et al), as well as the activity of Cn/Zn superoxide dismutase (SOD1) in human cells (Lin et al, 2013). A crystal structure of human SIRT5 bound to a succinylated peptide has revealed the molecular basis for preferential removal of these negatively charged acyl modifications (Du et al, 2011).

Human SIRT6 also exhibits alternate acyl chain specificity, preferentially removing long chain fatty acyl groups from lysine (Feldman et al, 2013; Jiang et al, 2013). SIRT6 deacetylates histone H3K9 (Michishita et al, 2008) and PGC1-α in vivo (Dominy et al, 2012), however that rate at which SIRT6 deacetylates lysines in vitro is markedly lower than the rate for removal of long-chain fatty acyl modifications, including myristoyl groups (Feldman et al, 2013; Jiang et al, 2013). A biological target for SIRT6 demyristoylating activity has been identified in mammalian cells, where SIRT6-demyristoylates the cytokine, TNF-α, promoting
its secretion (Jiang et al, 2013). PfSir2a, a sirtuin from the malaria parasite *P. falciparum*, also has robust demyristoylating and low deacetylating activities (Zhu et al, 2011), although the underlying biology for PfSir2a remains to be elucidated. The bacterial sirtuin, CobB, has multiple deacylating activities that are biologically relevant. In addition to its deacetylation activity, which regulates the activity of acetyl CoA synthetase (Starai et al, 2002), CobB also regulates propionyl-CoA synthetase (PrpE) activity by depropionylating the enzyme *in vivo* (Garrity et al, 2007) and was recently shown to have desuccinylating activity (Colak et al, 2013). These findings emphasize the importance of alternative sirtuin deacylation activities in diverse biological pathways.

Structural studies of sirtuins bound to different substrates have provided clues to their true acyl specificities as well as insights into how their active sites accommodate non-canonical acyl modifications. The structure of human SIRT5 in complex with a thioacetylated peptide revealed a negatively charged CHES buffer molecule in the active site (Du et al, 2011), which provided a clue that SIRT5 might also remove negatively charged acyl modifications (Du et al, 2011). Biochemical assays confirmed this hypothesis, while a structure of SIRT5 in complex with a succinylated peptide revealed that the succinyl carboxylate was contacted by the same residues that bound the negatively charged buffer molecule (Du et al, 2011). Structures of the demyristoylating enzymes, human SIRT6 and *P. falciparum* PfSir2A, bound to myristoylated peptides show how long hydrophobic tunnels near these sirtuins’ active sites accommodate the long aliphatic chain of the myristoyl group and explain why these enzymes
preferentially remove long acyl chains (Jiang et al, 2013; Zhu et al, 2011). These studies highlight the important clues that structures can shed on the true acyl specificities of sirtuins.

In light of the recent findings that some sirtuins have alternative deacylation activities that are important for their function, we revisited the substrate preferences of several bacterial and archaeal enzymes that have been used as model systems to study the structure-based mechanism of the sirtuin deacetylation reaction (Avalos et al, 2005; Avalos et al, 2002; Cosgrove et al, 2006; Min et al, 2001; Zhao et al, 2004). Sir2Af1 and Sir2Af2 from *Archaeoglobus fulgidus*, TmSir2 from *Thermotoga maritima* and CobB from *Escherichia coli* were tested against a panel of peptides containing different acyl modifications. Of the enzymes tested, only TmSir2 preferred acetylated substrates, whereas the other three enzymes exhibited alternate acyl chain specificities. Crystal structures of Sir2Af1 bound to a succinylated peptide and Sir2Af2 bound to a myristoylated peptide reveal features of the substrate-binding pocket that govern the specificity of each sirtuin for its preferred acyl modification. Interestingly, Sir2Af2 undergoes a conformational change that enables it to accommodate the long myristoyl group. These studies underscore the principle that sirtuins are a family of deacylases rather than deacetylases and reveal principles for structure based modeling of sirtuin acyl group specificity.
6.3 Materials and Methods

Protein expression and purification

TmSir2 was expressed and purified as previously described (Smith et al, 2002). Sir2Af2 was expressed and purified as previously described (Smith et al, 2002) with the following modifications. After heating the cell lysate, the supernatant was loaded first on a Q column (HiTrap Q HP 5 mls, GE LifeSciences) and fractions containing Sir2Af2 were subsequently loaded onto a Cibacron Blue column (HiTrap Blue HP 5 mls, GE LifeSciences). After gel filtration, fractions containing Sir2Af2 were dialyzed into 20 mM HEPES pH 7.4 and 1 mM TCEP, concentrated to 20 mg/mL, and flash-frozen in liquid nitrogen. Sir2Af1 was purified as previously described (Avalos et al, 2002) with the following modifications. According to the published protocol, the cell lysate was heated and the soluble portion containing Sir2Af1 was passed over a Q column (HiTrap Q HP 5 mls, GE LifeSciences). The following chromatographic steps differ from the published protocol. The flow-through from the Q column was bound to an S column (HiTrap Sp HP 5 mls, GE LifeSciences) in 20 mM MES pH 6.0, 25 mM NaCl, and 1 mM DTT, and eluted using a linear gradient to 1 M NaCl. Fractions containing Sir2Af1 were concentrated and loaded on a Superdex75 26/60 (GE LifeSciences) in 20 mM HEPES pH 8.0, 150 mM NaCl, 1 mM DTT, and 25 μM ZnCl_2. Sir2Af1 was concentrated to 20 mg/mL in the same buffer and flash-frozen in liquid nitrogen.
*E. coli* CobB aa1-279 was amplified by colony PCR from XL1-Blue cells (Stratagene) and cloned as a thioredoxin-fusion into pET32a (EMD Millipore, Merck KGaA, Darmstadt, Germany) with an N-terminal TEV site and a four amino acid unstructured linker before the start codon (ENLYFQG-ASAS). Plasmid containing CobB was expressed in Rosetta-2(DE3)pLysS cells (EMD Millipore). Starter cultures were grown overnight in MDG media (Studier, 2005), diluted into M9ZB media for protein expression (Studier, 2005), grown at 37°C to an OD$_{600nm}$ of 2.0, incubated in an ice bath for 45 minutes, and induced with 0.5 mM IPTG overnight at 15°C. Cells were lysed in 20 mM HEPES pH 7.6, 500 mM NaCl, 5 mM $\beta$ME, and 20 mM Imidazole, and passed over a HisTrap column (HisTrap HP 5 mL, GE LifeSciences). CobB was eluted with a gradient to 400 mM Imidazole and dialyzed overnight at 4°C into 100 mM NaCl, 20 mM HEPES pH 7.6, and 5 mM $\beta$ME in the presence of 1 mg TEV protease per 20 mg recombinant protein. Cleaved protein was re-passed over the HisTrap and the flow-through was further purified using a Q column (Q HP 5 mLs, GE LifeSciences) in 20 mM HEPES pH 7.6, 50 mM NaCl, and 200 $\mu$M TCEP with a linear gradient to 1 M NaCl. Fractions containing CobB were dialyzed into 20 mM HEPES pH 7.6, 100 mM NaCl, and 200 $\mu$M TCEP, concentrated to 5 mg/mL, and flash-frozen in liquid nitrogen.

All peptides used in this study had the following sequence: KGLGKGGA-$K^*$-RHRKW, corresponding to amino acids 8-20 of histone H4; $K^*$ indicates the position of the acylated lysine. The following acyl modifications were incorporated
at the sixteenth position (K*): acetyl lysine, propionyl lysine, butyryl lysine, succinyl lysine, and myristoyl lysine. Peptides for biochemical assays were purchased at >70% purity from JPT Peptide Technologies GmBH (Berlin, Germany) and used without further purification. Peptides for crystallization were purchased from the JHMI Synthesis and Sequencing Facility (Baltimore, MD), where they were synthesized as previously described (Zhu et al, 2011) at >90% purity and used without further purification. Peptide concentrations were determined spectrophotometrically using the molar extinction coefficient for tryptophan $\varepsilon_{280\text{nm}} = 5690 \text{ M}^{-1} \text{ cm}^{-1}$.

**Enzymatic activity assays**

Enzyme-coupled assays monitoring NAD$^+$-consumption were performed as previously described with the following modifications (Avalos et al, 2002). Assays were run in buffer containing 100 mM HEPES pH 8.0, 50 mM NaCl, 0.5 mM DTT, and 1 mM NAD$^+$ (Sigma N7004-1G) previously neutralized with NaOH. Peptide stock solutions were made at 1mM, which was calculated spectrophotometrically using the molar extinction coefficient for tryptophan to convert absorbance at 280nm measured on a NanoDrop 1000 spectrophotometer (Thermo Scientific) to a concentration, and used at a final concentration of 250 µM in the assay. Reactions were initiated with the addition of enzyme, which were used at a final concentration of 2 µM for TmSir2, CobB, and Sir2Af2, and at a concentration of 15 µM for Sir2Af1. Reactions containing TmSir2, Sir2Af2, and Sir2Af1 were run at 50°C, and reactions containing CobB were run at 37°C; all reactions were run
in a thermal cycler with heated lid to prevent sample evaporation. Glucose-6-phosphate (Sigma G7879-1G) and glucose-6-phosphate-dehydrogenase (Sigma G8529-2KU) were purchased from Sigma. For each measured rate, at least three time points were collected to ensure that the reaction proceeded linearly with time and that measurements captured the initial reaction velocity, and absorbance at 340nm was converted to a concentration of NADH remaining using an extinction coefficient for NADH $\varepsilon_{340\text{nm}} = 6220 \text{ M}^{-1} \text{ cm}^{-1}$.

Reactions for analysis by HPLC were run in the same assay buffer as the enzyme-coupled assay. For Sir2Af2, 2 µM enzyme and 250 µM H4K16-myristoyl peptide were mixed in the presence and absence of 2 mM NAD$^+$ in a total reaction volume of 50 µL. Reactions were allowed to proceed for 20 minutes at 50°C, quenched with an equal volume of 10% TFA, placed on ice for 30 minutes, and spun twice for 10 minutes at room temperature at 18,000 g to precipitate the enzyme. The reaction products were separated on a Kinetex XB-C18 column (100Å, 75mm x 4.60mm, 2.6µm, Phenomenex) heated to 30°C using a 40-100% gradient (0.1% TFA for “Buffer A” and 0.1% TFA + 90% acetonitrile for “Buffer B”) developed at 0.5 mL/minute over 20 minutes and absorbance was recorded at 280nm. For Sir2Af1, 15 µM enzyme and 250 µM H4K16-succinyl peptide were mixed in the presence and absence of 2 mM NAD$^+$ in a total reaction volume of 50 µL. Reactions were allowed to proceed for 60 minutes at 50°C, quenched with an equal volume of 0.5 N HCl in methanol, incubated for >3 hours at -20°C to precipitate the enzyme, and spun twice for 10 minutes at 18,000g. The reaction
products were separated on the same column as above using a 65-75% gradient (0.1% TFA for “Buffer A” and 0.1% TFA + 95% methanol for “Buffer B”) developed at 0.5 mL/minute over 25 minutes and absorbance was recorded at 280nm.

**Crystallization, Data Collection, and Refinement**

For the crystal structure of Sir2Af1 bound to a succinylated peptide, equal volumes Sir2Af1 at 20 mg/mL and peptide (H4K16-succinyl) at 3 mM in 20 mM HEPES pH 7.4 were mixed and incubated for >30 minutes on ice. The resulting mixture contained 10 mg/mL Sir2Af1 and 1.5 mM peptide. Crystals were grown by vapor diffusion at 20°C by mixing 1 µL of protein solution with 1 µL of well solution containing 0.1 M Tris pH 8.5, 200 mM MgCl₂, and 20% (w/v) PEG 8,000. Crystals grew in 3-5 days, were cryo-protected with the addition of glycerol to 30% (v/v), and flash frozen in a liquid nitrogen stream. Diffraction data were collected at the JHMI x-ray facility using a Rigaku FR-E x-ray generator at a wavelength of 1.54 Å and a Saturn 944+ detector. The complex crystallized in the primitive monoclinic space group P2₁ with unit cell dimensions a = 44.61 Å, b = 51.36 Å, c = 57.69 Å, α = 90.0°, β = 101.9°, and γ = 90.0°.

For the crystal structure of Sir2Af2 bound to a myristoylated peptide, a mixture containing 10 mg/mL Sir2Af2, 1.5 mM H4K16-myristyl peptide, 10% (v/v) polystyrene nanospheres (Nanosphere Size Standards, Cat. No.: 3020A, ThermoScientific), 10 mM HEPES pH 7.4, and 1 mM TCEP was incubated on ice
for >30 minutes. The mixture precipitated upon addition of the nanospheres, but was still used to set trays. Crystals were grown by vapor diffusion at 20°C by mixing 1 µL protein solution with 1 µL well solution. The well solution contained 0.1 M Na-acetate pH 4.8, 14-18% (v/v) 2-propanol, and 14-15% (w/v) PEG 6,000. Crystals grew to their final dimensions in 5-8 days and were cryo-protected with the addition of glycerol to 20% (v/v). Diffraction data were collected at the GM/CA-CAT beamline 23-ID-D at the Advanced Photon Source using a wavelength of 1.033 Å and under standard cryogenic conditions. The complex crystallized in the primitive monoclinic space group P2₁ with unit cell dimensions a = 38.48 Å, b = 77.25 Å, c = 45.90 Å, α = 90.0°, β = 110.7°, and γ = 90.0°.

Both data sets were indexed and scaled with HKL2000 (Otwinowski & Minor, 1997) and structures were determined using molecular replacement with MOLREP from the CCP4 suite (Vagin & Teplyakov, 1997; Vagin & Teplyakov, 2010). For the structure of Sir2Af1 bound to a succinylated peptide, the coordinates of Sir2Af1 (PDB ID 1ICI) (Min et al, 2001) were used as a search model. For the structure of Sir2Af2 bound to a myristoylated peptide, the coordinates of Sir2Af2 (PDB ID 1MA3) (Avalos et al, 2002) were used as a search model. Subsequent rounds of model-building and reciprocal space refinement were done with the graphics program COOT (Emsley & Cowtan, 2004) and REFMAC5 from the CCP4 suite (1994). Crystallographic statistics are summarized in Table 6.1. Simulated annealing omit maps were created by
deleting the atoms corresponding to the succinyl- or myristoyl-groups from the refined structures followed by three rounds of refinement with Phenix that included two cycles of simulated annealing (Adams et al, 2010).

6.4 Results

Structure-based prediction of acyl chain specificity

In light of the chemically diverse array of lysine acyl modifications that have been detected in cells and the recently discovered ability of some sirtuins to remove these modifications, we speculated that the archaeal sirtuins, Sir2Af1 and Sir2Af2, might preferentially remove acyl modifications from lysine other than acetyl groups. Structures of these sirtuins bound to a variety of substrates have been previously reported. These include structures of Sir2Af1 bound to NAD\(^+\) (Chang et al, 2002; Min et al, 2001), and of Sir2Af2 bound to an acetylated peptide (Avalos et al, 2002) as well as in complex with NAD\(^+\)-derived metabolites (Avalos et al, 2005; Avalos et al, 2004). Although both enzymes deacetylate peptides and proteins \textit{in vitro} (Avalos et al, 2002; Chang et al, 2002; Min et al, 2001), comparisons with structures of SIRT5 (Du et al, 2011) and SIRT6 (Jiang et al, 2013) suggested that Sir2Af1 and Sir2Af2 might also remove other types of acyl modifications.

Sir2Af1 contains features in common with SIRT5 that are consistent with a preference for removing negatively charged acyl groups from lysine. Recent structural studies of the human demalonylating and desuccinyling enzyme,
SIRT5, identified tyrosine and arginine residues (Y102 and R105) that confer specificity for the negatively charged succinyl group (Du et al, 2011). Y102 donates a hydrogen bond to the modified lysine, while the R105 guanidinium group participates in both electrostatic and hydrogen bonding interactions with the succinyl carboxylate (Du et al, 2011). Interestingly, SIRT5 is the only human sirtuin where these residues are conserved (Du et al, 2011), and biochemical studies have confirmed that the other human sirtuins are weak desuccinylases in vitro (Feldman et al, 2013). Kinetically, SIRT5 is remarkably specific for removing negatively charged acyl modifications, as its $K_M$ decreases by two orders of magnitude when supplied with succinyl or malonyl lysine compared to acetyl lysine (Du et al, 2011), which translates into an increase of nearly three orders of magnitude in the specificity constant, $k_{cat}/K_M$ (Du et al, 2011). Not only are these two residues conserved in Sir2Af1 as well as in the desuccinylating enzyme, E. coli CobB (Figure 6.2A) (Colak et al, 2013), but they are also strongly evolutionarily coupled (Fischer et al, 2012) and a single point mutation of a corresponding residue in Sir2Af2 (M70R) profoundly alters its specificity, conferring desuccinylating activity on the mutant Sir2Af2-M70R while decreasing deacetylation activity (Fischer et al, 2012). We therefore hypothesized that Sir2Af1 is also a desuccinylating enzyme.

A previously reported crystal structure of Sir2Af2 (Avalos et al, 2002) contains a serendipitously bound PEG molecule (Figure 6.2B) that suggests that Sir2Af2 may be able to remove medium to long fatty acyl modifications. Three of the five monomers within the crystal asymmetric unit contain a long PEG
molecule that binds in a hydrophobic tunnel that extends away from the acetyl lysine binding site (Avalos et al, 2004). The PEG molecules align remarkably well with the myristoyl group in structures of both SIRT6 and PfSir2a bound to myristoylated peptides (Figure 6.2B) (Jiang et al, 2013; Zhu et al, 2011), suggesting that Sir2Af2 may also accommodate long fatty acyl chains and thus catalyze lysine demyristoylation.
Figure 6.2 Structure- and sequence-guided predictions of acyl chain specificity

(A) Sequence alignment between sirtuins that are specific for negatively charged acyl chains (CobB, SIRT5, and Sir2Af1) and those that are not (Sir2Af2 and TmSir2). Conserved tyrosine and arginine residues confer specificity for negatively charged acyl modifications to lysine (Colak et al, 2013; Du et al, 2011). (B) Structural alignment between human SIRT6 and PfSir2a, which are both demyristoylating enzymes, and Sir2Af2. All three ligands are shown, but only the Sir2Af2 structure is displayed. Both myristoyl lysine residues overlay with a PEG molecule from the Sir2Af2 structure, suggesting alternate specificity for Sir2Af2.
Alternate deacylating activities of bacterial and archaeal sirtuins

To test our hypothesis that Sir2Af1 removes negatively charged acyl modifications while Sir2Af2 has demyristoylating activity, we assayed the activity of these enzymes on peptides based on the sequence of histone H4 containing acetyl, propionyl, butyryl, succinyl, or myristoyl lysine at residue K16 (Figure 6.1). As a comparison we also tested the bacterial sirtuin, TmSir2, from *Thermotoga maritima*, whose active site does not have room to accommodate larger acyl groups and is thus predicted to be a *bona fide* deacetylase (Bheda et al, 2011; Hawse & Wolberger, 2009), and *E. coli* CobB, which desuccinylates peptides *in vitro* (Colak et al, 2013). Using an assay that monitors NAD⁺ consumption (Avalos et al, 2002), we found that Sir2Af1 activity is approximately fivefold higher against succinyl lysine as compared to acetyl lysine (Figure 6.3A). As previously observed, the overall activity of Sir2Af1 is weak as compared to other archaeal and bacterial sirtuins (Min et al, 2001). In addition, we confirm that NAD⁺-consumption by Sir2Af1 reflects peptide desuccinylation using an HPLC-based assay that monitors the peptide directly (Figure 6.3B), demonstrated by a shift in its retention time in the presence of enzyme and NAD⁺. CobB also exhibits robust desuccinylating activity in our assay, as it desuccinylates peptides at a rate approximately 2.3-fold faster than deacetylation (Figure 6.3A). Interestingly, CobB activity was lower on the uncharged acyl modifications, propionyl and butyryl, which are smaller than succinyl groups and would be expected to be accommodated in the same binding pocket. CobB depropionylates substrates 1.8-fold more slowly and debutyrylates 3-fold more
slowly than deacetylation of the same peptide, and no demyristoylating activity was detected (Figure 6.3A). Consistent with our structure-based hypothesis, the archaeal sirtuin Sir2Af1 preferentially removes negatively charged acyl chains from lysine, which we predicted based on conserved sequence elements that it shares with the known desuccinylating enzymes SIRT5 (Du et al, 2011) and CobB (Colak et al, 2013).
Figure 6.3 Comparison of deacylation activities.

(A) Rates of NAD⁺-consumption were measured for different sirtuin enzymes against 250 µM acylated histone H4 peptide (5). TmSir2, Sir2Af2, and CobB were used at a concentration of 2µM, and Sir2Af1 was used at a concentration of 15µM. Initial rates were normalized to total enzyme concentration for comparison. (B) HPLC chromatograms from deacylation reactions. Left – 15 µM Sir2Af1 was incubated with 250 µM succinylated H4 peptide for 1 hour at 50°C. Right – 2 µM Sir2Af2 was incubated with 250 µM myristoylated H4 peptide for 20 minutes at 50°C. The peptides were deacylated only when 2mM NAD⁺ was added, causing them to elute at an earlier retention time.
We next tested the prediction that the archaeal sirtuin Sir2Af2 would preferentially remove long fatty acyl modifications. Consistent with our hypothesis, Sir2Af2 shows increased reaction velocity with increasing acyl chain length (Figure 6.3A). As compared to its deacetylating activity, Sir2Af2 depropionylated, debutyrylated, and demyristoylated a peptide of the same sequence at 1.5-, 1.9-, and 3.4-fold higher rates, respectively (Figure 6.3A). Even though the carbon backbone for succinyl lysine is substantially shorter in length than myristoyl lysine, Sir2Af2 had no detectable desuccinylating activity, suggesting that its active site can accommodate substrates that are long and hydrophobic but not charged. To confirm that NAD\(^+\)-consumption reflects peptide deacylation, we monitored reaction progress using an HPLC-based assay, where we observed a shift in the retention time for the peptide when Sir2Af2 and NAD\(^+\) were both present (Figure 6.3B).

We also examined the acyl specificity of the bacterial sirtuin, TmSir2, which we expected to preferentially remove acetyl groups compared to other acyl modifications based on previous structural and biochemical data. Structures of TmSir2 bound to both acetylated (Hawse & Wolberger, 2009) and propionylated peptides (Bheda et al, 2011) have been reported, showing that the modified lysine fits snugly into a tight hydrophobic pocket. In addition, TmSir2 was previously shown to catalyze depropionylation more slowly than deacetylation (Bheda et al, 2011). Consistent with previous data, TmSir2 preferentially removed acetyl groups, with relatively lower activity on all other acyl
modifications tested (Figure 6.3A). No activity was detectable against myristoyl lysine or succinyl lysine (Figure 6.3A).

*Crystal structure of Sir2Af1 bound to a succinylated peptide*

To determine the structural basis of Sir2Af1 specificity for negatively charged acyl chains, we determined the structure of Sir2Af1 sirtuin bound to a succinylated H4 peptide at 1.8 Å resolution (PDB ID 4TWI). Crystallographic statistics are summarized in Table 6.1. The overall structure resembles a previously reported structure of Sir2Af1 bound to NAD⁺ (Figure 6.4A) (Min et al, 2001). Electron density 2Fo-Fc maps (Figure 6.4B) and simulated annealing omit maps show unambiguous density for the succinylated lysine (Figure 6.4C). Conserved arginine and tyrosine residues, Y64 and R67 (Figure 6.2A), hydrogen bond and form electrostatic interactions with the succinyl carboxylate (Figure 6.4D) in the same way that SIRT5 engages a succinylated peptide (Figure 6.4E) (Du et al, 2011). Although CobB has not been crystallized bound to a succinylated substrate, it likely uses the same interactions to bind negatively charged acyl chains. If a succinyl group is modeled into its structure in complex with an acetylated peptide (Zhao et al, 2004), a single rotamer change by residue R95 can recapitulate the hydrogen bonding network found in the active sites of Sir2Af1 and SIRT5 (Figure 6.4F).
Figure 6.4 Structure of Sir2Af1 bound to a succinylated peptide

Structure determined at 1.8 Å resolution (PDB ID 4TWI). (A) The hydrogen-bonding network in the active site of Sir2Af1 (pink) bound to a succinylated peptide (yellow) resembles that for human SIRT5 (3RIY – teal). (B) 2Fo-Fc map of the succinylated peptide contoured at 1σ (gray). (C) Simulated annealing omit map of the succinyl group shows density for the modified lysine. 2Fo-Fc map is colored gray and contoured at 1σ. Fo-Fc map is colored green and contoured at 3σ. Hydrogen-bonding and electrostatic interactions in the (D) Sir2Af1, (E) human SIRT5, and (F) E. coli CobB active sites.
<table>
<thead>
<tr>
<th></th>
<th>Sir2-Af1 bound to a succinylated peptide</th>
<th>Sir2-Af2 bound to a myristoylated peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diffraction Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁</td>
<td>P2₁</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54</td>
<td>1.033</td>
</tr>
<tr>
<td>Unit Cell Dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>44.61, 51.36, 57.69</td>
<td>38.48, 77.25, 45.90</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90.0, 101.9, 90.0</td>
<td>90.0, 110.7, 90.0</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>21.94-1.79 (1.85-1.79)</td>
<td>42.93 - 1.65 (1.71 - 1.65)</td>
</tr>
<tr>
<td>Measured Reflections</td>
<td>53220 (2921)</td>
<td>99631 (8711)</td>
</tr>
<tr>
<td>Unique Reflections</td>
<td>23269 (2008)</td>
<td>30123 (2941)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>95.39 (83.73)</td>
<td>99.73 (97.89)</td>
</tr>
<tr>
<td>Average I/σ (merged reflections)</td>
<td>7.36 (1.85)</td>
<td>13.24 (4.15)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>2.3 (1.5)</td>
<td>3.3 (3.0)</td>
</tr>
<tr>
<td>R_merge (%)</td>
<td>11.7 (33.9)</td>
<td>5.01 (18.3)</td>
</tr>
</tbody>
</table>

**Refinement Statistics**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Atoms</td>
<td>2213</td>
<td>2139</td>
</tr>
<tr>
<td>Protein</td>
<td>1968</td>
<td>1906</td>
</tr>
<tr>
<td>Ligands</td>
<td>7</td>
<td>33</td>
</tr>
<tr>
<td>Water</td>
<td>238</td>
<td>200</td>
</tr>
<tr>
<td>CC1/2</td>
<td>0.977 (0.751)</td>
<td>0.997 (0.949)</td>
</tr>
<tr>
<td>CC*</td>
<td>0.994 (0.926)</td>
<td>0.999 (0.987)</td>
</tr>
<tr>
<td>R_work (%) / R_free (%)</td>
<td>19.2 / 22.1</td>
<td>18.0 / 20.6</td>
</tr>
</tbody>
</table>

**B-factors**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>18.1</td>
<td>18.2</td>
</tr>
<tr>
<td>Ligands</td>
<td>39.6</td>
<td>32.9</td>
</tr>
<tr>
<td>Solvent</td>
<td>29.9</td>
<td>33.0</td>
</tr>
</tbody>
</table>

**RMSD**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond lengths (Å)</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.00</td>
<td>1.02</td>
</tr>
</tbody>
</table>

*Values in parentheses correspond to data from the highest resolution shell.*
Crystal structure of Sir2Af2 bound to a myristoylated peptide

To determine how Sir2Af2 accommodates long acyl chains in its active site, we determined the structure of Sir2Af2 bound to a myristoylated peptide at a resolution of 1.65 Å (PDB ID 4TWJ) (Figure 6.5A). Crystallographic statistics are summarized in Table 6.1. The structure resembles previously reported structures of Sir2Af2 bound to an acetylated p53 peptide (Avalos et al, 2002), NAD⁺, or ADP ribose (Avalos et al, 2004). Electron density for the myristoylated peptide is clearly observable in both 2Fo-Fc and in simulated annealing omits maps (Figure 6.5B). As predicted, the myristoyl group occupies a hydrophobic tunnel in the Sir2Af2 active site where a PEG molecule was located in a previous structure (Avalos et al, 2004). An alignment with crystal structures of other sirtuins in complex with myristoylated peptides (human SIRT6 (Jiang et al, 2013) and PfSir2a (Zhu et al, 2011)) shows that the myristoyl groups from all three structures overlay well, suggesting a conserved function for the hydrophobic tunnel (Figure 6.5C).
Figure 6.5 Structure of Sir2Af2 bound to a myristoylated peptide

Structure determined at 1.65 Å resolution (PDB ID 4TWJ). (A) Overall structure with Sir2Af2 shown in cyan and myristoylated H4 peptide shown in yellow. (B) Simulated annealing omit map of the myristoyl group (yellow). $2F_o-F_c$ map (gray) is contoured at 1σ and $F_o-F_c$ map (green) is contoured at 2σ. (C) The myristoyl group (yellow) aligns with the PEG molecule in a previous structure of Sir2Af2 (green). (D) Overlay of the Sir2Af2 (cyan) bound to myristoylated peptide (yellow) with Sir2Af2 bound to an acetylated peptide (1MA3; purple). Inset shows structural rearrangement required to accommodate the myristoyl group. (E) The
hydrophobic tunnel from the structure of Sir2Af2 (purple) bound to an acetylated peptide (1MA3) is occluded by L74 and F75 (magenta). Superimposed myristoyl group (yellow spheres) clashes with L74 and F75. (F) Conformational rearrangement when Sir2Af2 (cyan) is bound to a myristoylated peptide (yellow spheres) opens up the hydrophobic tunnel by moving the positions of L74 and F75 (magenta).

In contrast with Sir2Af1, which does not require a conformational change to bind succinyl lysine, the active site of Sir2Af2 must rearrange in order to accommodate a myristoyl group. If a myristoyl lysine is docked into the structure of Sir2Af2 bound to an acetylated peptide (Avalos et al, 2002), a kinked helix obstructs the path of the acyl chain through the hydrophobic tunnel (Figure 6.5D). Two residues, L74 and F75, are located on the kinked helix and sterically occlude the part of the hydrophobic tunnel that is furthest from the active site in the structure of Sir2Af2 bound to an acetylated peptide (Figure 6.5E) (Avalos et al, 2002). Superposition with the structure reported here shows that the terminal six carbons of the myristoyl chain would clash with these two residues (Figure 6.5E). However, in the structure of Sir2Af2 bound to a myristoylated peptide, the helix straightens out, thereby moving the α-carbons of L74 and F75 by 5.2 Å and 8.3 Å, respectively (Figure 6.5E). This widens the tunnel by more than 6 Å, allowing Sir2Af2 to accommodate the full length of the myristoyl chain (Figure 6.5E).
6.5 Discussion

Renewed interest in sirtuin substrate preference has been triggered by recent studies that have revealed the presence of a large and chemically diverse repertoire of acyl modifications in cells. For a growing number of sirtuins, the ability to remove acyl modifications other than acetyl groups has been implicated in their biological roles. Determining the true specificities of sirtuins for different acyl chains is thus essential to understanding the roles of sirtuins in biology. Here we have revisited the acyl chain specificity of two archaeal enzymes that were previously used as models for understanding the sirtuin deacetylation reaction and showed that both enzymes preferentially remove other types of acyl modifications. Sir2Af1 preferentially desuccinylates peptides and the demyristoylating activity of Sir2Af2 is markedly faster than its deacetylating activity on a similar peptide. This finding underscores the principle that sirtuins are actually a family of deacylases, and suggests that future studies on other sirtuins will likely uncover alternate specificities as well.

Relative to its deacetylation activity, the archaeal sirtuin Sir2Af2 is more active against acyl chains of increasing length. This activity was suggested by previous structural studies revealing the presence of a hydrophobic tunnel adjacent to the active site that was occupied by a PEG molecule in a structure of Sir2Af2 bound to NAD$^+$ (Avalos et al, 2002). However, in contrast with the desuccinylating enzymes Sir2Af1, SIRT5 and CobB, the ability of Sir2Af2, PfSir2a, and SIRT6 to accommodate long acyl chains is not the result of
conserved sequence elements. When the amino acid sequences of these three enzymes are aligned along with other sirtuins with different specificities (Figure 6.6), no pattern of sequence conservation emerges that distinguishes the demyristoylating enzymes. Of all the amino acid residues within 5 Å of the myristoyl acyl chain in the Sir2Af2 structure, the only residues that are conserved among the demyristoylating enzymes are also generally conserved among sirtuins (Figure 6.6). Therefore, structural and biochemical studies remain the only way at present to identify sirtuins specific for removing long acyl chains.

Figure 6.6 Sequence conservation between sirtuins that remove long acyl chains

Sequence alignment among sirtuins with different acyl specificities performed using ClustalW. Residues within 5 Å of the myristoyl acyl chain are boxed in red.
Structural plasticity in the Sir2Af2 active site may facilitate its ability to accommodate and remove a broad array of acyl chains from lysine. In the structure of Sir2Af2 bound to an acetylated peptide (Avalos et al, 2002), the hydrophobic tunnel is too short to accommodate a myristoyl lysine (Figure 6.5E), despite the fact that it efficiently catalyzes demyristoylation \textit{in vitro} (Figure 6.3A). It is important to point out that shorter acyl chains such as butyryl or octanoyl lysine could likely still bind to Sir2Af2 in this conformation. However, its structure bound to a myristoylated peptide reveals a conformational change, which generates a long channel that extends through the entire enzyme and is open to solvent. In this new conformation, Sir2Af2 may be able to accommodate an aliphatic chain of greater length, raising the possibility that it could remove even longer acyl chains such as palmitic acid (Figure 6.5F). However other structurally characterized sirtuins with open hydrophobic tunnels, like human SIRT6 (Jiang et al, 2013) and human SIRT2 (Moniot et al, 2013), are still weak depalmitoylating enzymes (Feldman et al, 2013; Moniot et al, 2013), suggesting that this modification may not be targeted by sirtuins in general.

Interestingly, crystal structures of SIRT6 and PfSir2a, which also remove long acyl chains from lysine, do not exhibit the same degree of structural flexibility as Sir2Af2 in this region. A comparison of the structures of SIRT6 alone (PDB ID 3K35), in complex with ADP-ribose (PDB ID 3PKI), or bound to a myristoylated peptide (PDB ID 3ZG6) (Jiang et al, 2013) does not show any structural differences in the vicinity of the acyl modification. Likewise, crystal structures of PfSir2a have been determined bound to AMP (PDB ID 3JWP), to a
myristoylated peptide (PDB ID 3U3D) (Zhu et al, 2011), and to both a myristoylated peptide and NAD$^+$ (PDB ID 3U31) (Zhu et al, 2011), all of which adopt similar conformations near the myristoyl chain.

The fact that Sir2Af2 must undergo a structural rearrangement to accommodate myristoyl lysine has important implications for studying the acyl specificity of other sirtuins, as it suggests that modeling substrates into preexisting structures will not be sufficient to predict the types of acyl chains targeted by a particular enzyme. Without directly testing the acyl specificity of Sir2Af2, its ability to remove very long chain acyl modifications would have been overlooked. These observations make clear that simply using docking programs may not capture the acyl chain preferences of many sirtuins with unknown specificities.

For other sirtuins like Sir2Af1, *E. coli* CobB, and human SIRT5, the presence of conserved sequence elements is sufficient to predict their specificity for negatively charged acyl chains. Here we make the surprising observation that Sir2Af1 and *E. coli* CobB, which are both desuccinylating enzymes, display incrementally slower kinetics with increasing acyl chain length, whereas the addition of a terminal carboxyl group to the acyl modification results in markedly higher deacylation rates. A similar pattern has been observed with human SIRT5, which has no detectable debutyrylating activity but robustly catalyzes desuccinylation (Feldman et al, 2013). This is counterintuitive, as we would expect that favorable hydrophobic interactions between the acyl chain and enzyme might give rise to the opposite effect. For example, butyryl lysine and
succinyl lysine contain the same number of backbone carbons and differ only by the presence of a terminal carboxylate for succinyl lysine. However, Sir2Af1 deacetylation activity is 12-fold faster on succinylated versus butyrylated substrates, but only five-fold faster on succinylated versus acetylated substrates (Figure 6.3A). Since full kinetic and thermodynamic studies were not done, it remains to be seen whether these differences are also reflected in catalytic efficiency and how binding affinity for different acyl chains correlates with activity for these sirtuins.

TmSir2 was the only sirtuin of the four studied here that exhibited preferential deacetylase activity (Figure 6.3A). Consistent with our findings, a previous study that characterized the kinetics and thermodynamics of TmSir2 against acetylated and propionylated substrates showed tighter binding but slower overall kinetics with the longer acyl chain (Bheda et al, 2011). Moreover, the active site of TmSir2 does not appear to be able to rearrange and accommodate longer acyl chains, demonstrated by a crystal structure of TmSir2 bound to a propionylated peptide (Bheda et al, 2011). Thus, while the ability to remove diverse acyl modification seems increasingly to be a characteristic of many sirtuins, some enzymes in this family appear to be bona fide deacetylases that are unlikely to remove larger acyl groups.

Promiscuity with regard to acyl chain specificity is emerging as a common feature of sirtuin enzymes, many of which are able to remove a number of chemically related acyl groups from lysine (Feldman et al, 2013; Smith & Denu, 2007). This has been explicitly shown for human SIRT1-6, which are each able to
remove several types of acyl chains *in vitro* (Feldman et al, 2013). For example, human SIRT1 is able to efficiently remove acyl chains ranging in length from acetyl (two carbons) to myristoyl groups (fourteen carbons), although it is only weakly active against palmitoyl lysine (sixteen carbons) and ignores acyl chains containing carboxylates or sulfhydryls (Feldman et al, 2013). Human SIRT5, which is very active against succinyl lysine, also removes a number of long acyl chains from lysine that vary in length from octanoylation (eight carbons) to dodecanoylation (twelve carbons) (Feldman et al, 2013). These data support the idea that promiscuity is a common feature of many sirtuin enzymes, and that this property has been conserved throughout evolution. Intriguingly, the deacetylating activity of human SIRT6, which preferentially removes acyl chains ranging in length from six to fourteen carbons (Feldman et al, 2013), is strongly stimulated by free fatty acids (Feldman et al, 2013). This suggests that SIRT6-catalyzed deacetylation may be coupled to the concentration of various intracellular metabolites. It is likely that free fatty acids bind within the hydrophobic tunnel occupied by myristoyl lysine in the structure of SIRT6 (Jiang et al, 2013), since myristic acid competitively inhibits SIRT6-catalyzed demyristoylation of histone peptides while stimulating SIRT6-catalyzed deacetylation of the same substrate (Feldman et al, 2013). Although not explicitly tested here, the archaeal sirtuin Sir2Af2 may also be regulated in this way, as it also contains a hydrophobic tunnel with the potential to bind free fatty acids.

While the biological functions for the archaeal sirtuins are not well characterized, it is clear that these enzymes are not simply deacetylases. There
are likely other sirtuins whose true specificities remain to be discovered. Although the conserved Arg/Tyr sequence element can be used to predict specificity for negatively charged acyl chains (Du et al, 2011), biochemical and structural studies remain the primary tool for predicting whether a sirtuin will accept long acyl chains in its active site. We expect that the identification of sirtuin enzymes with alternate specificities will keep pace with the discovery of new acyl modifications \textit{in vivo}, as lysine acylation emerges as a common element of the signaling landscape in all three domains of life (Choudhary et al, 2014).
Chapter 7: Supplementary Information

Section A: Solution studies of the SAGA DUB module

This appendix summarizes my contributions to structural and solution studies detailing the role of the Sgf11 zinc finger to the stability and catalytic activity of the SAGA deubiquitinating (DUB) module. Dr. Nadine Samara led this project when she was a graduate student in the lab, and the work presented in this section of my thesis has been published in the journal *Structure* (Samara et al, 2012).

The catalytic subunit of the SAGA DUB module, Ubp8, is only enzymatically active when in complex with three other DUB module subunits (Kohler et al, 2008). Truncating a conserved zinc finger found in Sgf11, which is located near the Ubp8 active site, causes a dramatic decrease in activity by the DUB module (Kohler et al, 2010) caused by a reduction in $k_{cat}$ by nearly 100-fold (Table 7.1) (Samara et al, 2012). To identify residues that might compensate for the absence of the Sgf11 zinc finger, sequence alignments were performed between Ubp8 and other USP-family deubiquitinating enzymes that do not require accessory subunits for activation. Two residues in Ubp8, S144 and S149, are asparagines in the majority of other DUBs, leading us to hypothesize that these residues might stabilize the active site in a catalytically competent conformation, or otherwise impact catalysis by Ubp8.

To see whether these residues altered catalysis by Ubp8, steady-state kinetic titrations using ubiquitin-AMC (Ub-AMC) as a substrate were performed.
with complexes containing Sgf11 zinc finger truncations (ΔZnF) and individual or double serine to asparagine mutations. None of the mutations tested had an effect on $K_M$ for Ub-AMC, but mutating either S144N or S149N in a ΔZnF background caused a four-fold increase in $k_{cat}$, (Figure 7.1A & 7.1B, Table 7.1). Mutating both residues in the same complex resulted in a roughly additive increase in $k_{cat}$ (Figure 7.1A, Table 7.1), suggesting that the two mutations perform different functions.

**Figure 7.1 Steady-state kinetics for DUBm variants**

(A) Steady-state kinetic titrations using Ub-AMC with wild-type, ΔZnF, ΔZnF S144N, ΔZnF S149N, and ΔZnF S144N S149N DUBm. (B) Steady-state kinetic titrations comparing wild-type, S144N, and S149N DUBm.
Table 7.1 Steady-state kinetic constants for DUBm variants

<table>
<thead>
<tr>
<th>dUB Module Variant</th>
<th>$k_{\text{cat}}$ (sec$^{-1}$)</th>
<th>$K_M$ (µM)</th>
<th>$k_{\text{cat}}/K_M$ (M$^{-1}$ sec$^{-1}$)</th>
<th>Critical $\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>4.4 ± 0.3</td>
<td>24 ± 2</td>
<td>(1.8 ± 0.2) x 10$^5$</td>
<td>3.1 x 10$^{-2}$</td>
</tr>
<tr>
<td>S144N</td>
<td>6.5 ± 0.3</td>
<td>28 ± 2</td>
<td>(2.3 ± 0.2) x 10$^5$</td>
<td>2.0 x 10$^{-2}$</td>
</tr>
<tr>
<td>S149N</td>
<td>3.1 ± 0.2</td>
<td>20 ± 2</td>
<td>(1.6 ± 0.2) x 10$^5$</td>
<td>2.2 x 10$^{-2}$</td>
</tr>
<tr>
<td>ΔZnF</td>
<td>0.063 ± 0.01</td>
<td>19 ± 1</td>
<td>(3.3 ± 0.4) x 10$^3$</td>
<td>3.1 x 10$^{-5}$</td>
</tr>
<tr>
<td>ΔZnF S144N</td>
<td>0.28 ± 0.02</td>
<td>17 ± 2</td>
<td>(1.6 ± 0.2) x 10$^4$</td>
<td>4.5 x 10$^{-4}$</td>
</tr>
<tr>
<td>ΔZnF S149N</td>
<td>0.26 ± 0.02</td>
<td>11 ± 1</td>
<td>(2.4 ± 0.3) x 10$^4$</td>
<td>5.9 x 10$^{-4}$</td>
</tr>
<tr>
<td>ΔZnF S144N S149N</td>
<td>0.47 ± 0.02</td>
<td>13 ± 1</td>
<td>(3.6 ± 0.3) x 10$^4$</td>
<td>1.1 x 10$^{-3}$</td>
</tr>
</tbody>
</table>

*All data points measured in triplicate.

To understand why the S144N and S149N mutations affect catalysis by Ubp8, structures of ΔZnF and ΔZnF S144N DUB module were determined by x-ray crystallography. The structures will not be described at length here, since I did not contribute to crystallization and structure determination. However, both the ΔZnF and ΔZnF S144N DUB modules formed unexpected domain-swapped dimers in the crystal, which misaligned certain active site residues. To ask whether the domain-swapped dimers existed in solution, sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation was performed. Whereas wild-type DUB module sedimented as a single species (Figure 7.2A & 7.2F), ΔZnF and ΔZnF S144N DUB modules sedimented as two distinct species with sedimentation coefficients consistent with a monomer-dimer equilibrium (Figure 7.2B, 7.2C, & 7.2F). Interestingly, the S149N mutation stabilized the monomeric form of the DUB module (Figure 7.2D, 7.2E, & 7.2F). Since the active site residues of the domain-swapped dimer are misaligned, the S149N mutation
might enhance catalysis by the ΔZnF DUB module by promoting the monomeric form of the enzyme.
Figure 7.2 Sedimentation velocity of different DUBm constructs. (A) WT DUBm, (B) ΔZnF, (C) ΔZnF S144N, (D), ΔZnF S149N, and (E) ΔZnF S144N and S149N. Data are fit to a sum of two Gaussians (gray) corresponding to sedimentation of monomeric (red) and dimeric (blue) DUB module.

Sedimentation equilibrium experiments were performed to confirm that the species sedimenting with higher s-value corresponded to a DUB module dimer. Consistent with our sedimentation velocity experiments, the profiles obtained with wild-type DUB module could be fit to a single-species model (Figure 7.3A). However, fitting the data collected with the ΔZnF and ΔZnF S144N complexes to a single-species model results in non-randomly distributed residuals (Figure 7.3B & 7.3D). Instead, these data were fit to a self-association scheme $2A \rightarrow A_2$ with dissociation constants of 17 and 28 µM for the ΔZnF and ΔZnF S144N complexes, respectively (Figure 7.3C & 7.3E, Table 7.2). Therefore, the domain-swapped dimers observed crystallographically also exist in solution.
Figure 7.3 Sedimentation equilibrium analysis of DUBm variants

Dimerization happens in solution. (A) Wild-type DUBm at a concentration of 6.5 µM can be modeled as a single species. ΔZnF DUBm was analyzed at three rotor speeds and at three concentrations in the analytical ultracentrifuge, and was fit to a (B) single species model or (C) monomer-dimer self-association scheme. ΔZnF S144N DUBm was also analyzed at three rotor speeds/concentrations in the analytical ultracentrifuge, and was fit to a (D) single species model or (E) monomer-dimer self-association scheme. For both ΔZnF and ΔZnF S144N, the residuals for the monomer-dimer self-association scheme (C and E) are less systematically distributed than for the single species fit.
Dimer dissociation constants fit from sedimentation equilibrium data.

These kinetic and solution studies of the SAGA DUB module helped elucidate how the Sgf11 zinc finger regulates Ubp8 activity. Steady-state kinetics revealed that the kinetic defect observed with ΔZnF DUB module was caused by a decrease in $k_{cat}$ and not $K_M$. Analytical ultracentrifugation not only confirmed that crystallographic dimers were also observed in solution, but that the S149N mutation promoted the monomeric form of the enzyme. This study revealed one mechanism by which the non-catalytic subunits of the DUB module regulate Ubp8 activity.
Section B: Purification of the intact SAGA complex

Intact SAGA complex was purified from \textit{S. pombe} strains containing genomically integrated TAP-tags on the Spt7 or Ada1 subunits. TAP-tagged strains were generously provided by Dr. Fred Winston at Harvard University. Samples corresponding to each step of the TAP-tag purification were run on SDS-PAGE gels and blotted with an $\alpha$-HA antibody, which recognizes a portion of the TAP-tag (Figure 7.4A & 7.4B). Based on the Western blot, the tagged proteins (Spt7 or Ada1) were present at all expected steps within the purification process, and the tagged subunit changed molecular weight when treated with TEV protease, corresponding to removal of an IgG moiety (Figure 7.4A & 7.4B). Purified complexes from Spt7-TAP and Ada1-TAP strains were run on SDS-PAGE gels and stained with SyproRuby to visualize co-purifying members of the SAGA complex (Figure 7.4C & 7.4D). The resulting complexes were very pure and displayed similar banding patterns, suggesting that the composition of the complex did not change based on the identity of the tagged subunit. Functional and structural studies remain to be done with the intact SAGA complex.

Media Recipes –

\textit{YES Media:}
- 5 g yeast extract
- 30 g dextrose
- 112 mLs AA mix
- Bring to a final volume of 1 L and autoclave.
AA mix:
2 g adenine
2 g uracil
2 g histidine
2 g leucine
2 g lysine
Bring to a final volume of 1 L and autoclave or filter for storage.

Buffer Recipes –

2X NP-40:
40 mM HEPES pH 8.0
300 mM NaCl
100 mM NaF
600 µM Na$_3$VO$_4$
Store at 4°C. Immediately before use mix 30mLs of 2X NP-40 buffer with 6 mLs 10% NP-40, 24 mLs cold MilliQ H$_2$O, 80 µL Leupeptin (5 mg/mL), 100 µL PMSF (0.5 M in DMSO), and one crushed EDTA-free Roche complete ultra protease inhibitor tablet.

IPP150:
10 mM Tris-base pH 8.0 at 4°C
150 mM NaCl
0.1% NP-40

TEV Cleavage Buffer (TEVCB):
10 mM Tris-base pH 8.0 at 4°C
150 mm NaCl
0.1% NP-40
1 mM DTT

2X Calmodulin Binding Buffer (CBB):
20 mM Tris-base pH 8.0 at 4°C
300 mM NaCl
2 mM Mg-OAc
2 mM Imidazole
4 mM CaCl$_2$

To make CBB/0.1%:
25 mLs 2X CBB
0.5 mL 10% NP-40
35 µL βME
Add cold H$_2$O to 50 mLs.
To make CBB/0.02%:
25 mLs 2X CBB
100 µL 10% NP-40
35 µL βME
Add cold H₂O to 50 mLs.

Calmodulin Elution Buffer (CEB):
10 mM Tris-base pH 8.0 at 4°C
150 mM NaCl
0.02% NP-40
1 mM Ng-OAc
1 mM Imidazole
3 mM EGTA pH 8.0
10 mM βME

Growing yeast for TAP-tag purification –

1. Streak out yeast strains from glycerol stocks 4 days in advance onto YES plates. If your strain has the tag integrated into the genome, you do not have to grow the cells under selection. Grow at 30°C.

2. Make 4 L of YES media with 500 mLs to 1 L per 2 L flask. Make 50 mLs YES media in a 250 mL unbaffled flask. Autoclave on liquid cycle for 45 minutes.

3. In the evening, inoculate a 50mL YES culture with one colony. Grow for ~24 hours at 32°C shaking at 210 rpm.

4. Measure the OD₆₀₀nm of the turbid yeast culture. Calculate the number of cells/mL using a conversion of OD₆₀₀nm = 0.1 for 2 x 10⁶ cells/mL. Calculate the volume you need to add to each large flask to reach an OD₆₀₀ = 1 in the morning (i.e. 16-18 hours later). This corresponds to ~2x10⁷ cells/mL. Wild-type S. pombe has a doubling time of 120-140 minutes. If you overgrow your cells, they will enter stationary phase and be very difficult to lyse. Grow at 32°C shaking at 210 rpm overnight.
5. Pellet the yeast at 4°C for 15 minutes at 5,000 rpm.

6. Decant the supernatant and wash with 20 mLs of ice cold MilliQ water.
   Spin cells for 5 minutes at 4,000 rpm in 50 mL conicals.

7. Resuspend the cells in 10-20 mLs NP-40 buffer + protease inhibitors.
   Freeze in droplets in liquid nitrogen. Transfer the frozen yeast to a 50 mL
   conical with holes poked in the top. Place the tube on dry ice and allow all
   of the nitrogen to drain away. Store at -80°C with the cap loosely on until
   ready to process.

**TAP-tag purification** –

1. Lyse cells using nitrogen mill. Transfer the resulting powder to one or two
   50 mL conical tubes. Add 10-15 mLs of ice-cold NP-40 buffer + PMSF to
   the top of the powder to help them to thaw on ice. Vortex a few times if
   you need them to thaw more quickly.

2. Spin cell lysates for 30 minutes at 4°C in an SS-34 rotor at 16.5krpm.

3. Pour supernatants into a 50 mL conical.

4. Prepare the IgG sepharose slurry:
   a. Add 1 mL of IgG sepharose slurry (500 µL of beads in a 1:1 slurry)
      to a 15 mL conical tube.
   b. Wash 3 x 5 mLs with NP-40 buffer (no protease inhibitors), spinning
      for 2 minutes at 800 g in between to pellet the beads.

5. Add 1 mL cleared lysate directly to the conical with the beads. Place the
   beads into the 50 mL conical containing the cleared lysate. Rinse the tube
containing the washed beads with one more mL of lysate to obtain all of the beads.

6. Rotate for 2 hours in the cold room.

7. Remove 1% of the sample (500 µL of a 50 mL sample), and wash 3 x 1 mL with IPP-150. Spin for 30 seconds at 10krpm in a microcentrifuge in between washes to pellet the beads. Resuspend the beads in 100 µL 1X sample buffer + βME ("IgG Bound" sample).

8. Pour the rest of the samples over chromatography columns in the cold room to collect the beads.

9. Take 100 µL from the flow-through for the IgG unbound sample.

10. Wash the column:
    a. 3 x 10 mLs IPP150
    b. 1 x 10 mLs TEVCB

11. Resuspend the IgG sepharose in 1 mL TEVCB and transfer to a 2 mL eppendorf tube. Add 300-500 units TEV protease (use the conversion 10,000 units/mg), which corresponds to 40 µL of a 1 mg/mL solution.

12. Rotate in the cold room overnight.

13. Spin down samples at 2,000 g for one minute in the cold room and place the supernatant (1 mL) in a cold 15 mL conical.

14. Resuspend the beads in 1 mL ice-cold TEVCB and spin again at 2,000 g. Add the supernatant to the same 15 mL conical (you should have a total of 2 mLs).
15. Spin the 15 mL conical at 4,000rpm for 5 minutes at 4°C to pellet excess beads. Transfer supernatant to a new cold 15 mL conical.

16. Collect a sample of the remaining protein that was not cleaved by TEV:
   a. Add 1 mL TEVCB to the beads. Spin at 10 krpm for 30 seconds.
   b. Wash 3 times in TEVCB as above. Add 1 mL TEVCB to the beads and transfer to a new tube. Spin the beads down again and resuspend in 1 mL TEVCB.
   c. Remove 45 µL of beads (3% total).
   d. Spin the beads at 10 krpm for 30 seconds and resuspend in 100 µL 1X sample buffer + βME.

17. Add 6 mL ice-cold CBB/0.1% buffer and 6 µL of 1M CaCl₂ (3 µL per mL of sample) to the conical containing the sample.

18. Remove a TEV cleaved sample.

19. Prepare the calmodulin affinity resin:
   a. Add 1 mL of calmodulin affinity resin slurry, (500 µL of beads in a 1:1 slurry) to a 15 mL conical tube.
   b. Wash 3 x 5 mLs with CBB/0.1% buffer, spinning for 2 minutes at 800 g in between to pellet the beads.

20. Add the beads to the samples the same way as in step 5.

21. Rotate the samples in the cold room for 2 hours.

22. Pour samples over chromatography column in cold room. Allow the beads to pack.
23. Remove 100 µL of the flow-through as the calmodulin unbound sample and add 1/3rd volume of 4X sample buffer + βME.

24. Wash the column:
   a. Wash 2 x 1 mL CBB/0.1%
   b. Wash 1 x 1 mL CBB/0.02%

25. Add 1 mL CEB and collect eluate in a 1.5 mL microcentrifuge tube.

26. With the bottom of the column capped, add 1 mL CEB buffer and pipette up and down to homogenize the beads into a slurry.

27. Allow this to incubate in the cold room for 10 minutes.

28. Elute into another 1.5 mL microcentrifuge tube.

29. Collect 75 µL of each eluate sample and add 25 µL 4X sample buffer + βME.

30. Concentrate all of the sample to less than 40 µL using a Vivaspin 500 100K MWCO in the microcentrifuge in the cold room.

31. Run 20 µL on a gel and stain with silver stain or SyproRuby.
Figure 7.4 Example of TAP-tag purifications for intact SAGA

Western blots using α-HA antibody to visualize (A) Ada1-TAP or (B) Spt7-TAP throughout the purification. SDS-PAGE gels stained with SyproRuby to visualize the SAGA complex purified from (C) Ada1-TAP or (D) Spt7-TAP strains.
Section C: A New RING Tossed into an Old HAT

p300 and CBP are multi-domain histone acetyltransferases (HATs) that regulate gene expression and are mutated in human diseases including cancer. In the August issue of *NSMB*, Delvecchio and colleagues report the structure of the p300 catalytic core, revealing the presence of a previously unknown RING domain that regulates the enzyme’s activity. This article preview was published in the journal *Structure* (Ringel & Wolberger, 2013).

*Structure Preview*

Reversible acetylation of lysine residues on histones, transcription factors, and transcriptional co-activators plays a central role in activating eukaryotic transcription (Eberharter & Becker, 2002). Histone acetyltransferase (HAT) enzymes, which attach an acetyl group to lysine, are grouped into three families based on sequence homology within their catalytic domains: the Gcn5 N-acetyltransferase (GNAT) family, the Morf, Ybf2, Sas2, and Tip60 (MYST) family, and the p300/CBP family (Lee & Workman, 2007). Although all three classes of HATs use acetyl-CoA as the acetyl donor and catalyze the same reaction, they differ substantially in structure and use different reaction mechanisms (Berndsen & Denu, 2008). A variety of mechanisms have evolved to regulate acetyltransferase activity and prevent inappropriate gene activation. These mechanisms include incorporation of the HAT into larger complexes that potentiate its enzymatic activity and the presence of one or more chromatin
reader domains that each recognizes a specific type of histone modification (Lee & Workman, 2007). By binding to individual histone modifications such as acetylated or methylated lysine or to combinations of modifications, chromatin reader domains restrict HAT activity to the appropriate chromosomal context (Lee & Workman, 2007). Misregulation of enzymatic activity has been associated with cancer for all three HAT families, and many aggressive tumors are characterized by differences in HAT expression levels (Cohen et al, 2011).

CBP and p300 are large, highly similar HATs of over 2,400 amino acid residues with overlapping cellular functions. The two proteins are ~64% identical in primary sequence, with even higher levels of conservation across their catalytic cores (Kalkhoven, 2004). p300/CBP contains at least nine annotated domains, in addition to the HAT domain, including a bromodomain, which binds acetylated lysines (Kalkhoven, 2004), and a predicted PHD domain, which typically binds methylated lysines (Lee & Workman, 2007). In addition to associating with transcription factors such as TATA-binding protein and TFIIB, p300/CBP also interacts with the tumor suppressor proteins, p53 and BRCA1, as well as oncoproteins such as fos and myb (Kalkhoven, 2004). Among HAT enzymes, p300/CBP is particularly interesting because it activates itself by autoacetylating a basic loop in the HAT domain (Thompson et al, 2004), and also acetylates at least 70 non-histone proteins, including p53 (Wang et al, 2008a). Mutations in CBP give rise to the congenital development disorder, Rubinstein-Taybi Syndrome (RTS) (Kalkhoven, 2004), and multiple human cancers arise from mutations and translocations of p300/CBP (Figure 7.5). While an earlier
structure of the isolated p300 HAT domain showed how certain oncogenic mutations disrupted catalytic activity (Liu et al, 2008), most known mutations and deletions either map outside of this domain or have no obvious impact on catalytic function (Cohen et al, 2011; Kalkhoven, 2004). Given our current understanding of p300/CBP-catalyzed acetylation, the mechanism by which these core mutations dysregulate the acetyltransferase activity of p300/CBP and contribute to carcinogenesis has remained elusive.

Figure 7.5 Domain structure of p300
The catalytic core of p300 is a hotspot for cancer mutations. Domain structure overlaid with mutations and truncations found in solid tumors and lymphoid neoplasms (Iyer et al, 2004; Pasqualucci et al, 2011).

The structure reported by Delvecchio and colleagues (Delvecchio et al, 2013) of a larger p300 catalytic core fragment contains an unanticipated feature that is key to understanding p300/CBP HAT regulation. In addition to the HAT domain bound to a lysine-CoA bisubstrate analogue, the fragment also contains a bromodomain and a sequence known as the CH2 region, which contains a PHD domain. The structure quite unexpectedly revealed that the CH2 region also contains a structurally unusual RING domain, which is inserted within the PHD domain. RING domains are found in a subset of ubiquitin E3 ligases, where they mediate interactions with both E2 ubiquitin conjugating enzymes and ubiquitin (Deshaiies & Joazeiro, 2009). In the new p300 structure, the RING domain contacts the active site of the HAT domain, blocking the substrate-binding cleft and thus suggesting that the RING domain might regulate HAT activity. Indeed, the authors showed that mutating key residues that perturb binding of the RING domain to the HAT domain results in a hyperactive form of the enzyme, pointing to an auto-inhibitory role for the RING domain in p300/CBP catalysis. A number of mutations and deletions that give rise to cancer and RTS also map to the RING-HAT interface, illuminating the likely mechanism by which previously uncharacterized mutations misregulate p300/CBP activity and cause disease. For example, the C1204R mutation found in some B-cell lymphomas disrupts the
integrity of the RING domain by removing a zinc coordination site (Figure 7.6).

Other curated mutations affect the RING domain as well: mutations that disturb the interaction between the RING and active site such as R1645E and E1242K are found in malignant melanoma and RTS, respectively (Figure 7.6), while the RING is entirely deleted in some forms of breast cancer. Accordingly, cells transfected with any of these mutant enzymes display hyperacetylation of p53 relative to a wild-type control.

**Figure 7.6 Structure of p300 core fragment**

*Individual domains colored as in Figure 7.5. Certain p300 mutations found in cancer map to the newly discovered auto inhibitory RING domain. C1204R coordinates a zinc atom in the RING domain itself, whereas E1242 and R1645 form a salt bridge between the RING and HAT domains (PDB ID 4BHW).*
Why was the presence of the RING domain in p300/CBP overlooked? First, the RING is embedded in a PHD domain, whose histone-interacting residues have been re-arranged as a result. Second, while canonical RING domains contain two structural zinc atoms in separate coordination sites (Deshaies & Joazeiro, 2009), the p300 RING contains a single bound zinc, with tightly packed hydrophobic interactions replacing the second metal-binding site. There is also an insertion in the p300 RING loop L2 that, in other RINGs, interacts with E2 enzymes (Deshaies & Joazeiro, 2009). Based on its established role in the ubiquitin conjugation pathway, Delvecchio and colleagues tested the p300 RING for E3 ligase activity in vitro but could detect none, at least with the panel of E2 enzymes tested. While an as-yet undiscovered role in ubiquitination cannot be ruled out, the authors’ findings indicate that the RING domain in p300 has been adapted to auto-regulate HAT activity.

The discovery of a previously unknown RING domain that gates the activity of the HAT domain is a key step forward in understanding how p300 is regulated. It remains to be seen to what degree the RING domain may co-regulate HAT activity in concert with the autoregulatory loop in the HAT domain, which was deleted in both the present and previous structural studies of p300 because it interfered with crystallization. The role of the PHD domain also remains to be determined, as it lacks features that enable other PHD domain to bind methylated lysines on histones. Perhaps this PHD domain evolved to perform a structural role, anchoring the inserted RING domain near the catalytic
site and forming a compact core structure that bridges the bromodomain and
catalytic domain. Although the newly identified RING domain does not appear to
be an active member of the ubiquitination machinery, previous work has shown
that the first ~600 residues of p300/CBP has E3 ligase activity and mediates
it be that this portion of p300/CBP also contains a divergent RING that is not
detectable by primary sequence analysis? More broadly, the discovery of an
unanticipated domain that plays a critical role in regulating HAT activity is a
reminder that other such domains likely await discovery. The inventory of
regulatory mechanisms applied to HATs will surely continue to grow as structural
biologists successfully tackle larger and larger protein fragments and complexes.
References


Bhaumik SR (2011) Distinct regulatory mechanisms of eukaryotic transcriptional activation by SAGA and TFIID. *Biochim Biophys Acta* **1809:** 97-108


Garske AL, Denu JM (2005) SIRT1 Top 40 Hits: Use of One-Bead, One-Compound Acetyl-Peptide Libraries and Quantum Dots to Probe Deacetylase Specificity†. Biochemistry 45: 94-101


Referees:

Cynthia Wolberger, Ph.D.
Dissertation Advisory
Primary Dissertation Reader
Professor, Department of Biophysics and Biophysical Chemistry
Johns Hopkins University School of Medicine
cwolberg@jhmi.edu

James Berger, Ph.D.
Secondary Dissertation Reader
Thesis Committee Member
Professor, Department of Biophysics and Biophysical Chemistry
Johns Hopkins University School of Medicine
jmberger@jhmi.edu

Philip Cole, M.D., Ph.D.
Thesis Committee Member
Professor, Department of Pharmacology and Molecular Sciences
Johns Hopkins University School of Medicine
pcole@jhmi.edu

Sean Taverna, Ph.D.
Thesis Committee Member
Assistant Professor, Department of Pharmacology and Molecular Sciences
Johns Hopkins University School of Medicine
staverna@jhmi.edu

Gregory D. Bowman, Ph.D.
Thesis Committee Member
Associate Professor, Department of Biophysics
Johns Hopkins University
gdbowman@jhu.edu

Rachel Green, Ph.D.
Thesis Committee Member
Professor, Department of Molecular Biology and Genetics
Johns Hopkins University School of Medicine
ragreen@jhmi.edu
Curriculum Vitae for Ph.D. Candidates
The Johns Hopkins University School of Medicine

Alison E. Ringel
Place of birth: New York, NY

March 23, 2015

Educational History

Johns Hopkins University School of Medicine, Baltimore, MD
(音乐)
Ph.D. Molecular Biophysics
Cumulative GPA: 4.00/4.00
Advisor: Dr. Cynthia Wolberger
Thesis: Structural and biochemical studies of HATs and HDACs

Wesleyan University, Middletown, CT
May 2009
B.A. Molecular Biology & Biochemistry
B.A. Physics
Cumulative GPA: 3.98/4.00
Advisor: Dr. Scott G. Holmes
Honors Thesis: Tdh3 (a GAPDH isozyme) is a novel regulator of Sir2-mediated transcriptional silencing in yeast

Research Experience

Johns Hopkins University School of Medicine, Baltimore, MD
2009-present
Graduate research with Dr. Cynthia Wolberger
Biochemical studies of histone acetyltransferases
• Purified a four-protein complex (the HAT module) responsible for acetylating the N-terminal tails of histones at actively transcribed genes
• Developed a new biochemical assay to monitor the effects of different modifications on lysine acetylation within a single pool of recombinant nucleosomes
• Identified a mechanism coupling the recognition of specific histone post-translational modifications to acetylation by the HAT module.

Structural and biochemical studies of sirtuins
• Used x-ray crystallography to explore the acyl chain specificity of archaeal sirtuin enzymes

Wesleyan University, Middletown, CT
2007-2009
Undergraduate research with Dr. Scott G. Holmes
Genetic studies of yeast Sir2

197
• Investigated how a glycolytic enzyme regulates transcriptional silencing using yeast genetics

Teaching Experience

Program in Molecular Biophysics, Johns Hopkins University School of Medicine

Teaching Assistant, Methods in Molecular Biophysics
Graduate-level course required for all students in the Program in Molecular Biophysics

Spring 2011

Teaching Assistant, Methods in Molecular Biophysics
Graduate-level course required for all students in the Program in Molecular Biophysics

Spring 2014

Molecular Biology/ Biochemistry Tutor
Tutored graduate students without a background in biological sciences on basic principles of molecular biology and biochemistry

Department of Molecular Biology and Biochemistry, Wesleyan University

Teaching Assistant, Molecular Biology
Upper-level course required of all molecular biology and biochemistry majors

Fall 2007/2008

Teaching Assistant, Advanced Lab in Molecular Biology and Genetics
Ran lab sections, prepared materials, and managed the ordering for a semester-long course in yeast genetics

Spring 2008/2009

Awards/ Recognition

University Honors, Wesleyan University May 2009
Phi Beta Kappa, Wesleyan University May 2009
Barry M. Goldwater Scholarship, Wesleyan University March 2008
St. Jude National Graduate Research Symposium April 2015

Meetings and Presentations

Keystone: The Biology of Sirtuins, Santa FE, NM March 2015

Oral presentation: Alternate deacylating specificities for the archaeal sirtuins Sir2Af1 and Sir2Af2

Epigenetics and Chromatin, Cold Spring Harbor Laboratory, NY September 2014

Oral presentation: The Sgf29 subunit of the SAGA HAT module promotes processive acetylation of histone tails by recognizing H3K4 trimethylation.
Institute for Biophysical Research, Annual Retreat, Baltimore MD
September 2014
Oral presentation: The Sgf29 subunit of the SAGA HAT module promotes processive acetylation of histone tails by recognizing H3K4 trimethylation.

Mid-Atlantic Crystallography Meeting, University of Maryland, Rockville, MD
April 2014
Poster: Redefining sirtuin specificity using structural studies and biochemistry.

Publications


