The Molecular Basis of Lysine Acetylation: Addition, Removal, and Recognition

Thesis by

Andrew M. Davenport

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2016

(Defended July 27, 2015)

© 2015
Andrew M. Davenport
All Rights Reserved

ACKNOWLEDGEMENTS

There are a host of people to whom I am grateful, and whose help was invaluable for the completion of this thesis. I am thankful to more people than I can mention, and everyone brought up here deserves more than a sentence or two, but some highlights are listed below.

First, I would like to thank my advisor, Andre Hoelz, for great guidance over the past seven years, both before and during graduate school. I'm particularly grateful for the freedom I was granted, and for the opportunity I was given to pursue projects (with varying degrees of success) outside the normal purview of the lab.

I would like to thank the scientists and educators from earlier in my life who exposed me to research, inspired me, and set me on my path to graduate school: Jeffrey Petruska, Lorne Mendell, Wolfgang Peti.

I would like to thank my parents for setting an example for me, and for instilling in me a value of education and intellectual pursuits.

I would like to thank all the members of the Hoelz laboratory for contributing to an extremely friendly, fun, and intellectually stimulating environment.

I would like to thank Ferdinand Huber, my closest scientific collaborator during my time at Caltech, without whose drive and intellect the work presented here would not have been possible.

Lastly, and mostly importantly, I would like to thank my wife, Yunji, for constant support, guidance, patience, and love. I can't imagine what my life would be without you.

ABSTRACT

Acetyltransferases and deacetylases catalyze the addition and removal, respectively, of acetyl groups to the ε-amino group of protein lysine residues. This modification can affect the function of a protein through several means, including the recruitment of specific binding partners called acetyl-lysine readers. Acetyltransferases, deacetylases, and acetyl-lysine readers have emerged as crucial regulators of biological processes and prominent targets for the treatment of human disease. This work describes a combination of structural, biochemical, biophysical, cell-biological, and organismal studies undertaken on a set of proteins that cumulatively include all steps of the acetylation process: the acetyltransferase MEC-17, the deacetylase SIRT1, and the acetyl-lysine reader DPF2. Tubulin acetylation by MEC-17 is associated with stable, long-lived microtubule structures. We determined the crystal structure of the catalytic domain of human MEC-17 in complex with the cofactor acetyl-CoA. The structure in combination with an extensive enzymatic analysis of MEC-17 mutants identified residues for cofactor and substrate recognition and activity. A large, evolutionarily conserved hydrophobic surface patch distal to the active site was shown to be necessary for catalysis, suggesting that specificity is achieved by interactions with the α-tubulin substrate that extend outside of the modified surface loop. Experiments in *C. elegans* showed that while MEC-17 is required for touch sensitivity, MEC-17 enzymatic activity is dispensible for this behavior. SIRT1 deacetylates a wide range of substrates, including p53, NF-κB, FOXO transcription factors, and PGC-1α, with roles in cellular processes ranging from energy metabolism to cell survival. SIRT1 activity is uniquely controlled by a C-terminal regulatory segment (CTR). Here we present crystal structures of the catalytic domain of human SIRT1 in complex with the CTR in an *apo* form and in

complex with a cofactor and a pseudo-substrate peptide. The catalytic domain adopts the canonical sirtuin fold. The CTR forms a β-hairpin structure that complements the β-sheet of the NAD^+-binding domain, covering an essentially invariant, hydrophobic surface. A comparison of the *apo* and cofactor bound structures revealed conformational changes throughout catalysis, including a rotation of a smaller subdomain with respect to the larger NAD^+-binding subdomain. A biochemical analysis identified key residues in the active site, an inhibitory role for the CTR, and distinct structural features of the CTR that mediate binding and inhibition of the SIRT1 catalytic domain. DPF2 represses myeloid differentiation in acute myelogenous leukemia. Finally, we solved the crystal structure of the tandem PHD domain of human DPF2. We showed that DPF2 preferentially binds H3 tail peptides acetylated at Lys14, and binds H4 tail peptides with no preference for acetylation state. Through a structural and mutational analysis we identify the molecular basis of histone recognition. We propose a model for the role of DPF2 in AML and identify the DPF2 tandem PHD finger domain as a promising novel target for anti-leukemia therapeutics.

TABLE OF CONTENTS

Acknowledgements iii
Abstract iv
Table of Contents vi

Chapter I: Introduction 1
Lysine Acetlyation 2
Acetylation and acetyltransferases 5
Deacetylation and deacetylases 7
Acetyl-lysine recognition 9
References 11

Chapter II: Structural and Functional Characterization of the α-tubulin acetyltransferase MEC-17 21
Abstract 22
Introduction 23
Results 25
Discussion 31
Methods 34
Acknowledgements 38
Figure Legends 39
References 42
Figures 48

Chapter III: Structural and Functional Analysis of the Putative Proto-Oncoprotein DPF2 57
Abstract 58
Introduction 59
Results 61
Discussion 67
Methods 69
Acknowledgements 72
Figure Legends 73
References 78
Figures 84

Chapter IV: Structural and Functional Analysis of Human SIRT1 97
Abstract 98
Introduction 99
Results 102
Discussion 113
Methods 115
Acknowledgements 121

Figure Legends 122
References 128
Figures 138

Chapter I

Introduction

LYSINE ACETYLATION

All domains of life carry out chemical modifications of proteins known as posttranslational modifications (PTMs) following initial synthesis by the ribosome. PTMs both increase the functional diversity of proteins, and the speed by which this functional diversity can be employed. Well-studied PTMs include phosphorylation, acetylation, methylation, glycosylation, and ubiquitination, although hundreds of others have been observed.[1]

Protein acetylation was first observed in 1963 via the incorporation of radiolabeled acetate into histone proteins.[2] The modification was shown to occur on the ε-amino group of Lysine residues (Lys) and to be independent of translation.[3] In the decades following its discovery, further examples of acetylated proteins were sparse, although significant, including the microtubule component α-tubulin,[4,5] the transcriptional regulator p53,[6] and the HIV-1 protein TAT.[7,8] This slow pace mainly resulted from the lack of adequate experimental reagents to identify acetylation marks. As such, the acetylation field remained focused on histones, and the enzymes that mediate acetylation were termed histone acetyltransferases (HATs) and histone deacetylases (HDACs). However, recent advances in mass spectrometry have led to the identification of thousands of acetylated proteins in various organisms, establishing acetylation as a widespread, conserved regulatory mechanism.[9] As the scope of acetylated proteins became apparent, the responsible enzymes have been variously called lysine acetyltransferases and deacetylases (KATs or KDACs) or protein acetyltransferases and deacetylases (PATs or PDACs), in addition to the original HAT and HDAC nomenclature.

Lysine acetylation can affect protein-protein interactions, protein-nucleic acid interactions, protein localization/transport, protein stability, and enzymatic activity. These

mechanisms have been shown to regulate a wide range of cellular processes, with particularly prominent roles in transcription and metabolism. The effect of acetylation on transcriptional activity was first suggested by the localization of acetylated histones to specific chromatin domains: histones tails in euchromatin are hyperacetylated while those in heterochromatin are hypoacetylated.[10-12] This connection became clearer upon the discovery that several previously characterized transcriptional regulators carried histone acetyltransferase or deacetylase activity.[13-16] Further genome-wide studies showed that histone tail acetylation is associated with actively transcribed genes.[17-19]

The correlation between histone acetylation and active transcription is believed to at least partly result from a loosening of the chromatin structure due to the neutralization of the positive charge on lysine. Such an opening of chromatin would rationalize observations that acetylated histones allow increased access to transcription factors and passage of the RNA polymerase.[20-23] Histones can also serve as recruitment sites for proteins containing domains that specifically recognize acetylated or unmodified histone tails. These proteins, named acetyl-lysine readers, are common components of transcription factor and chromatin remodeling complexes.

Beyond effects on histone proteins and chromatin architecture, acetylation can effect changes in gene transcription by acting on transcriptional regulators themselves. p53 is acetylated on a several lysine residues, particularly in a C-terminal regulatory domain, which promotes DNA binding and transcriptional activity[6,24] and blocks ubiquitination and subsequent degradation.[25] Cross-talk between acetylation and ubiquitination is thought to be a general mechanism for protein stabilization, with significant overlap in the lysine

residues subject to each modification.[26] Other notable acetylated transcription factors include c-MYC, NF-κB, c-Jun, MyoD, E2F1, and FOXO1.

In addition to transcriptional regulation, acetylation has been strongly linked to cellular metabolism. Required cofactors for all identified acetyltransferases (acetyl-CoA) and one family of deacetylases, the sirtuins (NAD^+), are key metabolic intermediates, and they have been proposed to link the activity of these respective enzymes to the metabolic state of the cell.[1] Large numbers of enzymes in metabolic pathways, including glycolysis, gluconeogenesis, the TCA cycle, fatty acid synthesis, and degradation, are acetylated. In many instances, acetylation is dependent upon the nature of the carbon source, and regulates the catalytic activity of the enzyme or even the direction of carbon flux through the pathway.[27,28]

Given the functions of acetylation in transcriptional and metabolic regulation, it is unsurprising that acetyltransferases, deacetylases, and acetyl-lysine readers are implicated in a number of cancers and metabolic disorders. However, human diseases linked to acetylation are not limited to these areas, and include cardiac diseases, neurodegenerative diseases, viral and fungal infections, psychiatric disorders, and even drug addiction.[29-31] Proteins mediating acetylation have proven validated targets for therapeutics, with particular success in the development of deacetylase inhibitors. Such inhibitors have been approved by the FDA for treatment of cutaneous T cell lymphoma (vorinostat and romidepsin) and multiple myeloma (panobinostat), with a number more in trials for additional cancers.[32,33]

ACETYLATION AND ACETYLTRANSFERASES

There are three major families of lysine acetyltransferases, all named for their founding proteins: Gcn5, CBP/p300, and MYST (whose founding members include Morf, Ybf2, Sas2, and Tip60). Other families of acetyltransferases exist, but they have not been as extensively studied. Moreover, due to relatively poor sequence identity, the total number of acetyltranferases is not well known, and it possible that more acetyltransferases or acetyltransferase families exist.[34]

All families of acetyltransferases use acetyl-CoA as the acetyl group donor. Representative crystal structures of each family have shown that they all share a structurally conserved core region comprising a three-stranded β-sheet flanked on one side by an α-helix that participates in acetyl-CoA co-factor biding.[34-38] Each family has unique regions surrounding this core. Despite this structural similarity, the acetyltransferases use different mechanisms to effect acetyl group transfer. Gcn5 family acetyltransferases bind substrate and cofactor to form a ternary complex, then deprotonate the substrate lysine via a conserved active site general base (Glu173 in the case of Gcn5), allowing for nucleophilic attack on the carbonyl carbon of acetyl-CoA.[36] MYST family acetyltransferases feature a similar chemical strategy of lysine deprotonation via a general base and subsequent nucleophilic attack. However, in different contexts they have been shown to proceed either through a ternary complex like Gcn5, or through a ping-pong catalytic mechanism whereby the enzyme first forms an acetylated intermediate prior to substrate binding.[39,40] A combination of structural and biochemical studies on p300 revealed that CBP/p300 acetyltransferases do not employ a general base, with mutation of only two active site residues (Tyr1467 and Trp 1438) having a significant affect on catalysis.[37] p300 was

proposed to utilize a so-called "hit-and-run" mechanism through a short-lived ternary complex, but further mechanistic insight is lacking.

Lysine acetyltransferases generally exist in large protein complexes with multiple subunits.[41] A common theme in such complexes is the presence of PTM binding domains that allow for recognition of, and interplay between, various PTMs. SAGA, the first identified multi-subunit acetyltransferase complex, includes the acetyltranserase Gcn5 and contains 15 putative PTM interacting domains.[41] Subunits of these complexes can influence the substrate selection or enzymatic activity of the resident acetyltransferase. Such organization may explain an observed lack of substrate specificity or consensus acetylation motif for many individually isolated acetyltransferases, as substrate selection may be provided by these other subunits.[42] In addition to regulation by participation in large protein complexes, acetyltransferases are also often controlled by auto-acetylation. For example, acetylation of several lysines within a loop of p300 relieves a blockage of substrate binding, and stimulates activity.[43]

While cells encode dedicated acetyltransferases, it is also possible for acetyl-CoA to enzymatically modify lysine residues at high concentrations and high pH. The chemical environment in the mitochondria may be sufficient for such reactions,[44] but the full extent to which non-enzymatic acetylation occurs *in vivo* is unknown.

DEACETYLATION AND DEACETYLASES

Deacetylases can be divided into the Zn^{2+}-dependent histone deacetylases (HDACs) and NAD^{+}-dependent sirtuin deacetylases (SIRTs). The mammalian genome encodes 11 HDACs (HDAC1-11), which have a common deacetylase domain, the structure of which was first observed for the HDAC-like protein from the thermophile *A. aeolicus.*[45] The HDAC catalytic domain has an α/β fold, shared by arginases,[46] that features a central eight-stranded β-sheet. Conserved residues coordinate a catalytic Zn^{2+} ion that activates a bound water molecule for nucleophilic attack on the acetyl group carbonyl.[31]

HDACs are further subdivided into four classes: I, IIa, IIb, and IV (Sirtuins were previously referred to as Class III HDACs).[29] Class I HDACs (1, 2, 3, and 8) are predominantly nuclear, and have little or no extensions to their catalytic domains. Class IIA HDACs (4, 5, 7, and 9) shuttle between the nucleus and the cytoplasm, and contain large N-terminal extensions. They notably lack a conserved, active-site tyrosine, resulting in a loss of catalytic activity.[47,48] The Class III HDACs (6,10) are cytosolic. Relatively little is known about HDAC11, the sole Class IV HDAC.

Sirtuins employ a distinct chemical mechanism from HDACs, coupling deacetylation to cleavage of NAD^{+}. Utilization of NAD^{+} is energetically unnecessary to carry out such a reaction, and has been proposed to have evolved as a regulatory strategy rather than out of chemical necessity.[49] The mechanistic details of this reaction are still a matter of debate. The first crystal structure of the sirtuin catalytic domain, of a sirtuin homolog from *A. fulgidus*, revealed a larger NAD^{+}-binding subdomain with a Rossmann fold and a smaller subdomain composed of a helical module and a Zn^{2+}-binding module.[50]

Mammals have seven sirtuins (SIRT1-7), which, similar to the HDACs, feature varying N- and C-terminal extensions to their catalytic domain, subcellular localizations, and chemical activities.[51] SIRTs 1, 6, and 7, are nuclear, and SIRT2 is cytosolic. SIRTs 3, 4, and 5 are mitochondrial, and highlight the significance of acetylation in the mitochondria. With low levels of phosphorylation, acetylation seems to be the most abundant modification in this organelle.[1] Recent studies have shown that a number of sirtuins can remove other acyl modifications of lysine in addition to acetyl groups. SIRTs 1, 2, and 3 have decrotonylase activity,[52] SIRT5 has desuccinylase and demalonylase activities[53], and SIRT6 has demyristoylase activity.[54] Depending upon the enzyme, this additional activity can either occur in addition to, or in place of, deacetylase activity. The scope of acyl-lysine modifications outside of acetylation and their functional implications are an area of active research.

ACETYL-LYSINE RECOGNITION

PTMs allow for the regulation of protein-protein interactions via the recruitment of domains that specifically bind to modified residues, called readers. The canonical acetyl-lysine reader is the bromodomain. The discovery of the binding preference for bromodomains was simultaneously reported with an NMR solution structure of the bromodomain from P/CAF.[55] The structure revealed a left-handed four-helix bundle with a large hydrophobic pocket at one end, which mutational studies suggested to be the acetyl-lysine binding site. A later crystal structure of the Gcn5 bromodomain in complex with an acetylated ligand confirmed this model, and showed that hydrophobic residues lining the pocket stabilized the lysine side chain, while hydrogen bonding of the acetyl group oxygen with an Asn residue at the base of the pocket provided specificity for the modified lysine.[56] More recent large-scale structural studies have led to crystal structures for 29 of the 61 bromodomains encoded in the human genome and shed light on their various binding preferences.[57] Bromodomains are also present in tandem with each other and other PTM readers.[58]

Other readers of acetyl-lysine include plectrin homology (PH) and YEATS domains. Thus far, a single example of each has been characterized, so it remains to be seen whether these domains represent general strategies for acetyl-lysine binding. The YEATS domain, named for its founding domain-containing proteins Yaf9, ENL, AF9, Taf12, and Sas5, was shown to recognize acetyl-lysine through a serine lined aromatic cage,[59] while a structural mechanism of acetyl-lysine recognition by the PH domain remains to be established.[60]

Lastly, a number of proteins containing tandem plant homeodomain (PHD) domains have been shown to bind acetyl-lysine.[61-63] The PHD family is notable for its diversity in binding partners, including unmodified, methylated, and acetylated lysines as well as unmodified arginines[64] Crystal structures have shown that, as for other readers, tandem PHD domains bind acetyl-lysine through a primarily hydrophobic pocket that includes hydrogen bonding partners for the acetyl group.[63,65] In all observed cases, this binding pocket is on the surface of one of the two tandem PHD fingers, while the other makes additional contacts to further define binding specificity.

REFERENCES

1. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E. & Mann, M. (2014). The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol* **15**, 536-50.
2. Phillips, D. M. (1963). The presence of acetyl groups of histones. *Biochem J* **87**, 258-63.
3. Allfrey, V. G., Faulkner, R. & Mirsky, A. E. (1964). Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis. *Proc Natl Acad Sci U S A* **51**, 786-94.
4. L'Hernault, S. W. & Rosenbaum, J. L. (1983). Chlamydomonas alpha-tubulin is posttranslationally modified in the flagella during flagellar assembly. *J Cell Biol* **97**, 258-63.
5. Piperno, G. & Fuller, M. T. (1985). Monoclonal antibodies specific for an acetylated form of alpha-tubulin recognize the antigen in cilia and flagella from a variety of organisms. *J Cell Biol* **101**, 2085-94.
6. Gu, W. & Roeder, R. G. (1997). Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90**, 595-606.
7. Kiernan, R. E., Vanhulle, C., Schiltz, L., Adam, E., Xiao, H., Maudoux, F., Calomme, C., Burny, A., Nakatani, Y., Jeang, K. T., Benkirane, M. & Van Lint, C. (1999). HIV-1 tat transcriptional activity is regulated by acetylation. *EMBO J* **18**, 6106-18.

8. Ott, M., Schnolzer, M., Garnica, J., Fischle, W., Emiliani, S., Rackwitz, H. R. & Verdin, E. (1999). Acetylation of the HIV-1 Tat protein by p300 is important for its transcriptional activity. *Curr Biol* **9**, 1489-92.

9. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V. & Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-40.

10. Turner, B. M., Birley, A. J. & Lavender, J. (1992). Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in Drosophila polytene nuclei. *Cell* **69**, 375-84.

11. Turner, B. M. & Fellows, G. (1989). Specific antibodies reveal ordered and cell-cycle-related use of histone-H4 acetylation sites in mammalian cells. *Eur J Biochem* **179**, 131-9.

12. Hebbes, T. R., Thorne, A. W. & Crane-Robinson, C. (1988). A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J* **7**, 1395-402.

13. Brownell, J. E. & Allis, C. D. (1995). An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. *Proc Natl Acad Sci U S A* **92**, 6364-8.

14. Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y. & Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**, 843-51.

15. Taunton, J., Hassig, C. A. & Schreiber, S. L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **272**, 408-11.

16. Imai, S., Armstrong, C. M., Kaeberlein, M. & Guarente, L. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795-800.

17. Kurdistani, S. K., Tavazoie, S. & Grunstein, M. (2004). Mapping global histone acetylation patterns to gene expression. *Cell* **117**, 721-33.

18. Liu, C. L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S. L., Friedman, N. & Rando, O. J. (2005). Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS Biol* **3**, e328.

19. Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D. K. & Young, R. A. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517-27.

20. Lee, D. Y., Hayes, J. J., Pruss, D. & Wolffe, A. P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* **72**, 73-84.

21. Marushige, K. (1976). Activation of chromatin by acetylation of histone side chains. *Proc Natl Acad Sci U S A* **73**, 3937-41.

22. Nightingale, K. P., Wellinger, R. E., Sogo, J. M. & Becker, P. B. (1998). Histone acetylation facilitates RNA polymerase II transcription of the Drosophila hsp26 gene in chromatin. *EMBO J* **17**, 2865-76.

23. Vettese-Dadey, M., Grant, P. A., Hebbes, T. R., Crane- Robinson, C., Allis, C. D. & Workman, J. L. (1996). Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA in vitro. *EMBO J* **15**, 2508-18.

24. Tang, Y., Zhao, W., Chen, Y., Zhao, Y. & Gu, W. (2008). Acetylation is indispensable for p53 activation. *Cell* **133**, 612-26.

25. Ito, A., Kawaguchi, Y., Lai, C. H., Kovacs, J. J., Higashimoto, Y., Appella, E. & Yao, T. P. (2002). MDM2-HDAC1-mediated deacetylation of p53 is required for its degradation. *EMBO J* **21**, 6236-45.

26. Wagner, S. A., Beli, P., Weinert, B. T., Nielsen, M. L., Cox, J., Mann, M. & Choudhary, C. (2011). A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol Cell Proteomics* **10**, M111 013284.

27. Wang, Q., Zhang, Y., Yang, C., Xiong, H., Lin, Y., Yao, J., Li, H., Xie, L., Zhao, W., Yao, Y., Ning, Z. B., Zeng, R., Xiong, Y., Guan, K. L., Zhao, S. & Zhao, G. P. (2010). Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science* **327**, 1004-7.

28. Zhao, S., Xu, W., Jiang, W., Yu, W., Lin, Y., Zhang, T., Yao, J., Zhou, L., Zeng, Y., Li, H., Li, Y., Shi, J., An, W., Hancock, S. M., He, F., Qin, L., Chin, J., Yang, P., Chen, X., Lei, Q., Xiong, Y. & Guan, K. L. (2010). Regulation of cellular metabolism by protein lysine acetylation. *Science* **327**, 1000-4.

29. Haberland, M., Montgomery, R. L. & Olson, E. N. (2009). The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet* **10**, 32-42.

30. Keppler, B. R. & Archer, T. K. (2008). Chromatin-modifying enzymes as therapeutic targets--Part 1. *Expert Opin Ther Targets* **12**, 1301-12.

31. Lombardi, P. M., Cole, K. E., Dowling, D. P. & Christianson, D. W. (2011). Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes. *Curr Opin Struct Biol* **21**, 735-43.

32. Fenichel, M. P. (2015). FDA approves new agent for multiple myeloma. *J Natl Cancer Inst* **107**, djv165.

33. Lakshmaiah, K. C., Jacob, L. A., Aparna, S., Lokanatha, D. & Saldanha, S. C. (2014). Epigenetic therapy of cancer with histone deacetylase inhibitors. *J Cancer Res Ther* **10**, 469-78.

34. Yuan, H. & Marmorstein, R. (2013). Histone acetyltransferases: Rising ancient counterparts to protein kinases. *Biopolymers* **99**, 98-111.

35. Clements, A., Rojas, J. R., Trievel, R. C., Wang, L., Berger, S. L. & Marmorstein, R. (1999). Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A. *EMBO J* **18**, 3521-32.

36. Trievel, R. C., Rojas, J. R., Sterner, D. E., Venkataramani, R. N., Wang, L., Zhou, J., Allis, C. D., Berger, S. L. & Marmorstein, R. (1999). Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. *Proc Natl Acad Sci U S A* **96**, 8931-6.

37. Liu, X., Wang, L., Zhao, K., Thompson, P. R., Hwang, Y., Marmorstein, R. & Cole, P. A. (2008). The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. *Nature* **451**, 846-50.

38. Yan, Y., Barlev, N. A., Haley, R. H., Berger, S. L. & Marmorstein, R. (2000). Crystal structure of yeast Esa1 suggests a unified mechanism for catalysis and substrate binding by histone acetyltransferases. *Mol Cell* **6**, 1195-205.

39. Yan, Y., Harper, S., Speicher, D. W. & Marmorstein, R. (2002). The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. *Nat Struct Biol* **9**, 862-9.

40. Berndsen, C. E., Albaugh, B. N., Tan, S. & Denu, J. M. (2007). Catalytic mechanism of a MYST family histone acetyltransferase. *Biochemistry* **46**, 623-9.

41. Lee, K. K. & Workman, J. L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol* **8**, 284-95.

42. Berndsen, C. E. & Denu, J. M. (2008). Catalysis and substrate selection by histone/protein lysine acetyltransferases. *Curr Opin Struct Biol* **18**, 682-9.

43. Thompson, P. R., Wang, D., Wang, L., Fulco, M., Pediconi, N., Zhang, D., An, W., Ge, Q., Roeder, R. G., Wong, J., Levrero, M., Sartorelli, V., Cotter, R. J. & Cole, P. A. (2004). Regulation of the p300 HAT domain via a novel activation loop. *Nat Struct Mol Biol* **11**, 308-15.

44. Wagner, G. R. & Hirschey, M. D. (2014). Nonenzymatic protein acylation as a carbon stress regulated by sirtuin deacylases. *Mol Cell* **54**, 5-16.

45. Finnin, M. S., Donigian, J. R., Cohen, A., Richon, V. M., Rifkind, R. A., Marks, P. A., Breslow, R. & Pavletich, N. P. (1999). Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors. *Nature* **401**, 188-93.

46. Kanyo, Z. F., Scolnick, L. R., Ash, D. E. & Christianson, D. W. (1996). Structure of a unique binuclear manganese cluster in arginase. *Nature* **383**, 554-7.

47. Jones, P., Altamura, S., De Francesco, R., Gallinari, P., Lahm, A., Neddermann, P., Rowley, M., Serafini, S. & Steinkuhler, C. (2008). Probing the elusive catalytic activity of vertebrate class IIa histone deacetylases. *Bioorg Med Chem Lett* **18**, 1814-9.

48. Schuetz, A., Min, J., Antoshenko, T., Wang, C. L., Allali-Hassani, A., Dong, A., Loppnau, P., Vedadi, M., Bochkarev, A., Sternglanz, R. & Plotnikov, A. N. (2007). Structural basis of inhibition of the human NAD+-dependent deacetylase SIRT5 by suramin. *Structure* **15**, 377-89.

49. Sauve, A. A., Wolberger, C., Schramm, V. L. & Boeke, J. D. (2006). The biochemistry of sirtuins. *Annu. Rev. Biochem.* **75**, 435-65.

50. Min, J., Landry, J., Sternglanz, R. & Xu, R. M. (2001). Crystal structure of a SIR2 homolog-NAD complex. *Cell* **105**, 269-79.

51. Haigis, M. C. & Sinclair, D. A. (2010). Mammalian sirtuins: biological insights and disease relevance. *Annu. Rev. Pathol.* **5**, 253-95.

52. Bao, X. C., Wang, Y., Li, X., Li, X. M., Liu, Z., Yang, T. P., Wong, C. F., Zhang, J. W., Hao, Q. & Li, X. D. (2014). Identification of 'erasers' for lysine crotonylated histone marks using a chemical proteomics approach. *Elife* **3**.

53. Du, J., Zhou, Y., Su, X., Yu, J. J., Khan, S., Jiang, H., Kim, J., Woo, J., Kim, J. H., Choi, B. H., He, B., Chen, W., Zhang, S., Cerione, R. A., Auwerx, J., Hao, Q. & Lin, H. (2011). Sirt5 is a NAD-dependent protein lysine demalonylase and desuccinylase. *Science* **334**, 806-9.

54. Jiang, H., Khan, S., Wang, Y., Charron, G., He, B., Sebastian, C., Du, J., Kim, R., Ge, E., Mostoslavsky, R., Hang, H. C., Hao, Q. & Lin, H. (2013). SIRT6 regulates TNF-alpha secretion through hydrolysis of long-chain fatty acyl lysine. *Nature* **496**, 110-3.

55. Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K. & Zhou, M. M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491-6.

56. Owen, D. J., Ornaghi, P., Yang, J. C., Lowe, N., Evans, P. R., Ballario, P., Neuhaus, D., Filetici, P. & Travers, A. A. (2000). The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *EMBO J* **19**, 6141-9.

57. Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J. P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Muller, S., Pawson, T., Gingras, A. C., Arrowsmith, C. H. & Knapp, S. (2012). Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214-31.

58. Patel, D. J. & Wang, Z. (2013). Readout of epigenetic modifications. *Annu Rev Biochem* **82**, 81-118.

59. Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S. Y., Li, W., Li, H. & Shi, X. (2014). AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell* **159**, 558-71.

60. Su, D., Hu, Q., Li, Q., Thompson, J. R., Cui, G., Fazly, A., Davies, B. A., Botuyan, M. V., Zhang, Z. & Mer, G. (2012). Structural basis for recognition of H3K56-acetylated histone H3-H4 by the chaperone Rtt106. *Nature* **483**, 104-7.

61. Ali, M., Yan, K., Lalonde, M. E., Degerny, C., Rothbart, S. B., Strahl, B. D., Cote, J., Yang, X. J. & Kutateladze, T. G. (2012). Tandem PHD fingers of MORF/MOZ acetyltransferases display selectivity for acetylated histone H3 and are required for the association with chromatin. *J Mol Biol* **424**, 328-38.

62. Qiu, Y., Liu, L., Zhao, C., Han, C., Li, F., Zhang, J., Wang, Y., Li, G., Mei, Y., Wu, M., Wu, J. & Shi, Y. (2012). Combinatorial readout of unmodified H3R2 and acetylated H3K14 by the tandem PHD finger of MOZ reveals a regulatory mechanism for HOXA9 transcription. *Genes Dev* **26**, 1376-91.

63. Zeng, L., Zhang, Q., Li, S., Plotnikov, A. N., Walsh, M. J. & Zhou, M. M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466**, 258-62.

64. Sanchez, R. & Zhou, M. M. (2011). The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36**, 364-72.

65. Dreveny, I., Deeves, S. E., Fulton, J., Yue, B., Messmer, M., Bhattacharya, A., Collins, H. M. & Heery, D. M. (2014). The double PHD finger domain of MOZ/MYST3 induces alpha-helical structure of the histone H3 tail to facilitate

acetylation and methylation sampling and modification. *Nucleic Acids Res* **42**, 822-35.

Chapter II

Structural and Functional Characterization of the α-tubulin Acetyltransferase MEC-17

This chapter, first published in *Journal of Molecular Biology* in 2014, was written by Andrew M. Davenport[1], Hui Chiu[2], Leslie N. Collins[1], Paul J. Minor[2], Paul W. Sternberg[2], and André Hoelz[1].

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

[2]Howard Hughes Medical Institute, Division of Biology, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

ABSTRACT

Tubulin protomers undergo an extensive array of post-translational modifications to tailor microtubules to specific tasks. One such modification, the acetylation of lysine 40 of α-tubulin, located in the lumen of microtubules, is associated with stable, long-living microtubule structures. MEC-17 was recently identified as the acetyltransferase that mediates this event. We have determined the crystal structure of the catalytic core of human MEC-17 in complex with its cofactor acetyl-CoA at 1.7 Å resolution. The structure reveals that the MEC-17 core adopts a canonical Gcn5-related *N*-acetyltransferase (GNAT) fold that is decorated with extensive surface loops. An enzymatic analysis of 33 MEC-17 surface mutants identifies hot-spot residues for catalysis and substrate recognition. A large, evolutionarily conserved hydrophobic surface patch that is critical for enzymatic activity is identified, suggesting that specificity is achieved by interactions with the α-tubulin substrate that extend outside of the modified surface loop. An analysis of MEC-17 mutants in *Caenorhabditis elegans* shows that enzymatic activity is dispensable for touch sensitivity.

INTRODUCTION

Microtubules are cytoskeletal components with a diverse range of essential cellular functions. Cytoplasmic microtubules mediate cell motility, transport, and shape, while specialized microtubule-based structures include the mitotic spindle, required for chromosome segregation, and the axoneme, the core component of cilia. All microtubule filaments share a common architecture. Heterodimers of α- and β-tubulin assemble in a head-to-tail fashion into linear polymers termed protofilaments, which associate laterally to form a cylindrical filament. Polymerization of tubulin heterodimers into microtubules will occur spontaneously *in vitro* in the presence of GTP.[1]

Cells have adopted several means to tailor the common microtubule structure to distinct cellular functions. These include the activity of microtubule-associated proteins, and the expression of various α- and β-tubulin isoforms.[2] Microtubules are also subject to a wide range of evolutionarily conserved post-translational modifications. A majority of these post-translational modifications occur on the flexible C-terminal tails of α- and β-tubulin that protrude from the microtubule. These include the removal and addition of the C-terminal tyrosine of α-tubulin, the removal of the penultimate C-terminal glutamate of α-tubulin, and the addition of poly-glycine and poly-glutamate chains to any one of a number of glutamate residues near the C-termini of both α- and β-tubulin.[3-5] An additional microtubule post-translational modification, the acetylation of lysine 40 (Lys40) of α-tubulin, is believed to uniquely occur in the lumen of the microtubule filament.[3,4]

Acetylation of α-tubulin Lys40 has long been associated with stable microtubule structures, such as those in the cytoplasmic array of neuronal axons, and in the ciliary

axoneme.[4] However, a coherent picture of the effects of this acetylation event on microtubule function has yet to emerge. Genetic manipulations abrogating α-tubulin Lys40 acetylation cause no detectable phenotype in *Tetrahymena thermophila,* a loss of touch sensation in *Caenorhabditis elegans*, anatomical abnormalities in *Danio rerio,* and delayed cilia assembly in *Homo sapiens*.[6-8] Additional reports have shown that Kinesin-1 preferentially travels on acetylated microtubule tracks, the endoplasmic reticulum preferentially slides along acetylated microtubules, acetylated microtubules bind to and inhibit the Na^+/K^+ pump, and microtubule acetylation regulates protofilament number in native microtubules.[9-15]

Recent biochemical and cell biological studies have identified MEC-17, or αTAT-1, as the α-tubulin Lys40 acetyltransferase.[6,8] MEC-17 is highly conserved among ciliated organisms, and displays strong substrate specificity, with no discernable activity towards histone tails. To further aid in the understanding of MEC-17 and the effects of α-tubulin Lys40 acetylation, we have determined the crystal structure of the catalytic domain of human MEC-17 in complex with acetyl-CoA. Structure-guided mutagenesis was used to determine hot-spot residues that abolish enzymatic activity. Strikingly, a large evolutionarily conserved hydrophobic patch is required for enzymatic activity, suggesting that substrate specificity is established by an extensive interaction with α-tubulin, extending beyond the typical peptide substrate interaction. A mutational analysis in *C. elegans* establishes an additional, non-catalytic function of MEC-17 responsible for touch sensitivity.

RESULTS

Structure determination

Using secondary structure prediction and sequence conservation analyses, we designed a series of C-terminal deletion constructs of *H. Sapiens* MEC-17 for expression in *E. coli* (Table 1). Of these, a stable and well-expressing fragment encompassing the core catalytic domain was identified (residues 1 to 193, hereafter referred to as MEC-17^{CORE}). Crystals of MEC-17^{CORE} in the presence of acetyl-CoA were obtained in the orthorhombic space group $P2_12_12$. The structure was solved by single-wavelength anomalous dispersion (SAD) using X-ray diffraction data obtained from selenomethionine-labeled crystals. The final model contains residues 1 to 193 and was refined to R_{work} and R_{free} values of 17.8 % and 21.3 %, respectively. For the details of the data collection and refinement statistics, see Table 2.

Structural Overview

The MEC-17^{CORE}•acetyl-CoA crystal structure is shown in Figure 1. MEC-17^{CORE} adopts a globular α/β fold that features a central five-stranded anti-parallel β-sheet, composed of β1-3, β6, and β7. An additional β-hairpin insertion between β3 and β6 (β4 and β5) packs against a C-terminal unstructured tail marked by three conserved phenylalanines (Phe183, Phe186, and Phe190). The central β-sheet is decorated on either side by six α-helices of varying lengths (αA-F). The central β-sheet along with αD forms a structural element that, despite weak sequence identity, is strongly reminiscent of Gcn5 and other GNAT family members.[16,17]

Acetyl-CoA is encapsulated in a highly conserved, predominantly basic pocket on the surface of MEC-17^{CORE}, between αD, αF and β6. A total of 18 residues contact the cofactor, constituting a buried surface area of ~ 1440 Å^2. Two invariant residues (Arg132 and Lys162) flank the adenine ring of CoA. Arg132 and Lys169 form hydrogen bonds with the 3'-ADP phosphate moiety, while three main chain amides (Gly134, Gly136, and Arg137) form hydrogen bonds with the pyrophosphate. The acetyl-CoA pantetheine arm is bound by a combination of hydrophobic van der Waals contacts and hydrogen bonds primarily with Phe124, Gln131, and Ser160. As with the MEC-17^{CORE} protein, acetyl-CoA adopts an overall conformation similar to that seen in complex with other GNAT family members, particularly in the acetyl and pantetheine moieties.

The donor acetyl group is oriented by a hydrogen bond with the backbone amide of Phe124, and sits at the base of the acetyl-CoA binding pocket, proximal to a groove which must bind the incoming Lys40 substrate residue of α-tubulin and its surrounding residues. This groove is also predominantly basic (Fig. 2), and could mediate binding to several acidic residues which surround Lys40 on α-tubulin. Our crystal structure is in agreement with previously published structural data from several other groups,[18-20] as indicated by RMSDs of ~0.6 Å^2 over 193 Cα atoms for available human MEC-17 structures.

Mutational Analysis

To identify key residues for catalytic activity and substrate recognition by MEC-17, we used the crystal structure and a sequence conservation analysis (Fig. 3) to guide the design of two C-terminal truncations and 33 surface mutations. These mutants were

tested in an *in vitro* enzymatic assay, using purified MEC-17 proteins, α-tubulin, and an acetylated Lys40 α-tubulin specific antibody, which allowed the determination of the initial velocity, Michaelis-Menten constant, and turnover number. All mutants examined were purified to homogeneity in milligram quantities and are indistinguishable from the wild-type protein in their behavior on a gel-filtration column, demonstrating their proper folding. The results of these experiments are summarized in Figure 4 and Table 3.

We first tested the activity of the crystallized MEC-17^{CORE} (residues 1-193), as well as a slighter longer fragment (residues 1-236, hereafter referred to as MEC-17$^{\Delta TAIL}$) that encompasses the MEC-17^{CORE} and an additional evolutionarily conserved C-terminal extension (CTE). In agreement with earlier work, MEC-17$^{\Delta TAIL}$ exhibits increased activity.[8] Specifically, while both MEC-17 fragments possess similar turnover numbers, the MEC-17^{CORE} exhibits a three times higher K_M value than MEC-17$^{\Delta TAIL}$ (~ 45 *versus* ~ 15 μM). While the CTE is absent in our structure, these findings suggest that it forms additional contacts with α-tubulin. Full-length MEC-17 severely degraded during purification, consistent with the predicted lack of secondary structure elements in the remainder of the protein (Fig. 1a) and therefore was not tested. Based on these results, we used MEC-17$^{\Delta TAIL}$ for our mutational analyses. Our experimentally determined K_M value for MEC-17$^{\Delta TAIL}$ (~ 15 μM) was in the range of those previously reported (~ 2 μM and ~ 30 μM), while our turnover number (~ 4×10^{-2} s^{-1}) was ~ 50-fold higher (~ 6×10^{-4} s^{-1} and ~ 20×10^{-4} s^{-1}).[8,20]

Alanine mutations in the acetyl-CoA binding pocket, including mutations of residues that interact with the adenine ring (Arg132 and Lys162), ADP-phosphate (Lys169), and panthenine arm (Gln131 and Ser160), all have a mild effect on enzymatic

activity. Together, these results suggest very robust protein-cofactor interaction, relatively immune to individual mutations.

The MEC-17 active site sits at the base of the cofactor binding pocket. GNAT family members use a general base, a glutamate in the case of Gcn5, to activate a water molecule to deprotonate the substrate lysine, priming it for nucleophilic attack on the cofactor acetyl group.[21,22] Potential general bases in MEC-17 include Gln58, Cys120, and Asp157, which are located ~5 Å from the donor acetyl group. Mutation of each of these residues to alanine severely compromises enzymatic activity. Asp157 forms a salt bridge with Arg158, but could undergo structural rearrangements upon substrate binding. To test whether the carboxylate moiety of Asp157 is required for enzymatic activity, we analyzed an asparagine mutant (D157N) mutant and found it to have no enzymatic activity. Additional conserved residues in the active site (Leu60, Ile64, and Arg158) are presumed to position the substrate lysine and their mutation to alanine severely reduced the catalytic activity, as expected. Below the acetyl-CoA binding pocket and active site αC, β3, β6, and connecting loops contribute to a conserved surface groove. Basic residues (Arg69, Lys98, and Lys102) lining this groove prove indispensible for catalysis, and could potentially interact with a number of acidic residues adjacent to the substrate lysine, as previously suggested.[18,19,23] The active-site architecture and identified catalytic residues are in agreement with the findings in recent publications.[18-20]

MEC-17 displays two additional highly conserved surface patches that are adjacent to the substrate binding grove (Fig. 2b). Whereas alanine mutations in one of them (Ser66 and Asn73) had no discernable effect on catalytic activity, mutations in the second patch show a pronounced effect on MEC-17 catalytic activity (Leu104, Phe105,

Leu164, Leu173, Pro178, Gln179, Phe183, Val184, Phe186, and Phe190). These 10 residues are located in the β4-β5 hairpin, helix αF, and the C-terminal region following helix αF, are primarily hydrophobic, and constitute approximately 1400 $Å^2$ of solvent-accessible surface area on MEC-17^{CORE}. The size of this patch and its distance from the active site support the conclusion that MEC-17 not only engages the loop that harbors Lys40 but also makes extensive contacts with α-tubulin. We explored this possibility by testing the ability of MEC-17 to acetylate or bind to a 20-residue α-tubulin peptide centered on Lys40, consistent with earlier work.[23] Indeed, the peptide proved insufficient for either enzymatic activity or interaction *in vitro,* consistent with earlier work (Fig. 5).[23]

In summary, our data suggest that the substrate specificity of MEC-17 is the result of numerous contacts between regions of α-tubulin that extend outside the acetylated Lys40 loop and a large evolutionarily conserved surface patch on MEC-17^{CORE}, as well as a C-terminal extension following the MEC-17^{CORE}. Further atomic details of the substrate recognition by MEC-17 await additional structural characterization.

In vivo Analysis

MEC-17 and its paralog ATAT-2 acetylate Lys40 on α-tubulin and are required for touch sensitivity in *C. elegans* that is mediated by the touch receptor neurons (TRNs).[6,8,24] However, it is currently unclear whether their catalytic activity is required for mechanosensation, as different groups have found a catalytically inactive MEC-17 mutant to be either sufficient or insufficient for rescuing touch sensitive of a *mec-17;atat-2* double null mutant.[8,14] To address these contradicting results, we tested the ability of MEC-17 variants to restore touch sensitivity of a *mec-17(ok2109); atat-2(ok2415)* double

null mutant *C. elegans* strain. The expression of all MEC-17 variants was verified buy the fluorescence signal from a C-terminal GFP tag, and all worms tested contained marked and comparable levels of GFP fluorescence (Fig. 6A). The results are summarized in Figure 6.

We first tested full-length wild-type MEC-17 and MEC-17$^{\Delta TAIL}$, which both restore touch sensitivity to the same levels as the *atat-2(ok2415)* single null mutant straing (~50% and ~40% of worms, respectively; Fig. 6B). The results demonstrate that the C-terminal tail of human MEC-17, which is not present in the *C. elegans* protein, is dispensable for mechanosensation. We next introduced three mutants into the truncated *H. sapiens* MEC-17 fragment: a catalytically dead active-site mutant (D157N) and two catalytically active surface mutants (S66A and N73A). We found that the catalytically dead D157N mutant (42%) and the catalytically active S66A (42%) and N73A (36%) mutants all rescued touch sensitivity in approximately the same percentage of worms. These data are consistent with previous work despite the use of different genetic background and expression systems.[14] Together, these results support the conclusion that MEC-17 promotes mechanosensation in *C. elegans* through a non-catalytic function.

DISCUSSION

Microtubules are extensively post-translationally modified on flexible C-terminal tails that protrude from the filament. One prominent exception is the highly conserved acetylation of the lumenal Lys40 of α-tubulin. To shed further light on this post-translational modification, we have determined the crystal structure of the *H. sapiens* tubulin acetyltransferase, MEC-17, in complex with acetyl-CoA. Our structural data were used to guide a series of biochemical and *in vivo* experiments to determine residues required for tubulin acetylation and the functional significance of MEC-17 activity in *C. elegans.*

The detailed catalytic mechanism of MEC-17 remains open for debate. GNAT family members employ a catalytic base to activate a water molecule to deprotonate the substrate lysine.[21] Different groups have proposed that Gln58, Asp157, or Asp157 and Cys120 participate as general bases in MEC-17.[19,20,23] Our data cannot currently distinguish among these potential mechanisms. Each of these residues is an appropriate distance from the acetyl-CoA in our crystal structure. Alanine mutations of Gln58, Cys120 and Asp157 severely compromise enzymatic activity. Notably, Cys120 does not exhibit the same level of conservation as the other two potential general bases. Additional studies are required for a comprehensive understanding of catalysis by MEC-17.

Our mutational analysis, along with those of others, point to a basic groove below the active site that interacts with acidic residues surround lysine 40 of α-tubulin, providing significant substrate specificity relative to histone tails.[19,23] However, our further results suggest a more extensive interaction with the substrate that goes beyond the typical recognition of an extended peptide segment, as is common for other

acetyltransferases.[17,25] In particular, our biochemical data establish an extensive surface patch, particularly in the β4-β5 hairpin and C-terminal loop, that while distal from the active site is necessary for efficient catalysis. While previous results suggested the importance of the β4-β5 hairpin[20], our work greatly expands upon the size of this patch and its sensitivity to mutation. Additional evidence for this notion is provided by the inability of MEC-17 to either bind or acetylate a 20-residue α-tubulin peptide centered on Lys40 *in vitro.* A more detailed understanding of this recognition awaits further structural exploration of MEC-17 in complex with α-tubulin.

Pioneering genetic studies identified a role for MEC-17 in *C. elegans* mechanosensation more than thirty years ago.[24] Two different groups have examined the role of MEC-17 catalytic function *in vivo*.[8,14] Both studies provide strong evidence that tubulin acetylation (of specialized tubulin MEC-12) requires MEC-17 catalytic activity, but both differ in their observations of the extent to which the acetyltransferase activity of MEC-17 is required for whole animal touch sensitivity. Our results with human protein in *C. elegans* are more consistent with a dispensable function of the catalytic activity for the whole animal behavior. Close examination of the published data and our data leads us to question the utility of this whole animal behavioral assay for a quantitative assessment of touch receptor neuronal function. The touch assay has been stunningly successful in identifying gens with strong effect on touch receptor neuron function. The double acetyltransferase mutant in everyone's hands still displays some response to light mechanical stimulation, raising the possibility of another acetylation pathway, if you hold with acetylation being crucial for the behavior, or consistent with the view that acetylation helps but is not essential for whole animal mechanosensation. Future work

will be required to elucidate the mechanistic basis for the identified non-catalytic activities of MEC-17 in mechanosensation. In particular, cell-based assays, for example, using genetically encoded calcium sensors, might provide a more accurate indication of the role of acetylation. Our structural and functional analyses presented here provide the framework for these studies and will assist in further research into both MEC-17 and tubulin acetylation as a whole.

MATERIALS AND METHODS

Protein expression and purification

The DNA fragments of *H. Sapiens* MEC-17 were amplified by PCR (Addgene plasmid 27099)[8] and cloned into a modified pET28a vector (Novagen) that contained a PreScission protease site after the N-terminal hexahistidine tag.[26] MEC-17 mutants were generated using QuikChange mutagenesis (Stratagene), and confirmed by DNA sequencing. Details of the bacterial expression constructs can be found in Table 1.

All proteins were expressed in *E. coli* BL21-CodonPLus(De3)-RIL cells (Stratagene) grown in LB media supplemented with appropriate antibiotics. Seleno-L-methionine-labeled protein was produced with a methionine pathway inhibition protocol.[27] Protein expression was induced at OD_{600} of ~0.6 with 500 μM IPTG for 16-18 hours at 18 °C. Cells were harvested by centrifugation, resuspended in a buffer containing 20 mM Tris (pH 7.0), 500 mM NaCl, 5 mM β-mercaptoethanol, 2 μM bovine lung aprotinin (Sigma), and complete EDTA-free protease inhibitor cocktail (Roche) and flash frozen in liquid nitrogen.

Thawed cells were lysed with a cell disrupter (Avestin) and the lysate was centrifuged for 60 minutes at 40,000 x g. The cleared lysate was applied to a Ni-NTA column (Qiagen) and eluted via an imidazole gradient. Pooled Ni-NTA fractions were cleaved with PreScission protease for 12 hours and further purified over a HiLoad16/60 Superdex 200 column (GE Healthcare) equilibrated in 20 mM Tris (pH 7.0), 150 mM NaCl, and 5 mM DTT. Pooled fractions were concentrated to 12 mg/ml and used in subsequent crystallization trials or for biochemical activity assays.

Structure determination and refinement

Crystals of MEC-17^{CORE} were grown at 21 °C in hanging drops containing 1 μl of the protein, which contained a two-fold molar excess of acetyl-CoA (Sigma), and 1 μl of a reservoir solution consisting of 26 % PEG 3350, 100 mM Bis-Tris (pH 6.5), and 200 mM NaCl. Crystals grew in the space group $P2_12_12$ and reached a maximum size of approximately 100 μm x 100 μm x 100 μm within a week. For cryo-protection, crystals were stabilized in 30 % PEG 3350, 100 mM Bis-Tris, pH 6.5, 200 mM NaCl, and 20 % glycerol (added in 5 % increments). X-ray diffraction data were collected at 100K at beamline 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL). X-ray intensities were processed using the HKL2000 denzo/scalepack package.[28] Phases from a SeMet SAD data set were obtained with SHARP,[29] which following density modification by DM[30] yielded an electron density map of high quality. ARP/wARP was used for initial automated model building, followed by manual model building in Coot.[30,31] Refinement was carried out in Phenix and the final model contains residues 1-193 and was refined to R_{work} and R_{free} values of 17.8 % and 21.3 %, respectively.[32] The stereochemical quality of the model was assessed with PROCHECK and MolProbity.[33,34] For details of the data collection and refinement statistics, see Table 2.

MEC-17 Activity Assays

In vitro MEC-17 acetyltransferase reactions were initiated by the addition of 1 μM recombinant MEC-17 to varying concentrations of porcine brain tubulin (Cytoskeleton), in 10 μL of a buffer containing 80 mM PIPES (pH 6.9), 0.5 mM $MgSO_4$, 0.8 mM EGTA, 1 mM GDP, 5 mM DTT, 10 % glycerol, and 30 μM acetyl-CoA.

Reactions were carried out at room temperature for 30 minutes, in the previously demonstrated linear range of MEC-17 activity,[8] and stopped by the addition of SDS loading buffer. Reaction volumes were separated on a 12.5 % SDS-PAGE gel, transferred to PVDF membranes, and immunoblotted with a mouse anti-acetylated Lys40 α-tubulin antibody (Sigma, 1:3,000 dilution) followed by an IRDye-800 conjugated goat anti-mouse antibody (Li-Cor, 1:15,000 dilution). Bound IRDye-800 conjugated antibody was detected and quantified using the Odyssey Imaging System (Li-Cor). Previously demonstrated levels of porcine brain tubulin Lys40 acetylation were used to estimate concentrations from measured absorbance units.[35] K_m and k_{cat} values were determined by fitting the concentration dependence of initial velocities using GraphPad Prism software 6.0.

***In vivo* analysis**

C. elegans was handled as described previously.[36] All strains used are derivatives of *C. elegans* N2 Bristol strain. The following mutations were used: LGIV: *mec-17(ok2109)*; LGX: *atat-2(ok2415)*. DNA fragments of *H. Sapiens* and *C. elegans* MEC-17 were amplified by PCR and cloned into Fire vector pPD95.75. To drive gene expression in the TRNs, C-terminally GFP-tagged MEC-17 was then fused to 850 bp of the promoter region of *mec-7* using PCR fusion.[37] The construct was injected with a *Pmyo-2::dsRed* co-injection marker.[38,39] All worms analyzed were selected for comparable levels of GFP fluorescence. All touch assays were performed as previously described.[24] Briefly, nematode touch sensitivity was determined by stroking an eyebrow hair attached to a toothpick across the anterior or posterior half of an adult animal.

Movement away from the eyebrow stroke was considered a touch-sensitive response. Results indicate the percentage of responses to a single touch, and all trails included at least 100 worms.

For fluorescence microscopy, animals were mounted on 2% agarose pads in M9 buffer, containing 10 mM sodium azide (Sigma-Aldrich), and were examined with a 40x, 0.75 numerical aperture oil immersion objective (UPLFLN, Olympus) on a standard epifluorescence microscope (IX71, Olympus). Images were captured with an electron-multiplying CCD camera (C9100-13, Hamamatsu).

Illustration and figures

The sequence alignment of MEC-17 was generated using ClustalX and colored with Alscript.[40,41] Figures were generated using PyMOL (www.pymol.org). Electrostatic potential was calculated with APBS.[42]

Accession Numbers

The atomic coordinates and structure factors of the *H. sapiens* MEC-17 have been deposited in the Protein Data Bank under accession code 4IF5.

ACKNOWLEDGEMENTS

We thank Daniel H Lin, Alina Patke, Tobias Stuwe, Karsten Thierbach, and Yunji Wu for critical reading of the manuscript, David King for mass spectrometry analysis, Maxene Nachury and Irini Topalidou for providing material, and Jens Kaiser and the scientific staff of SSRL beamline 12-2 for their support with X-ray diffraction measurements. We acknowledge the Gordon and Betty Moore Foundation for their support of the Molecular Observatory at the California Institute of Technology. The operations at the SSRL are supported by the Department of Energy and by the National Institutes of Health. AMD and PJM are supported by a National Institutes of Health Research Service Award (5 T32 GM07616). AH was supported by the Albert Wyrick V Scholar Award of the V Foundation for Cancer Research, the 54th Mallinckrodt Scholar Award of the Edward Mallinckrodt, Jr. Foundation, and a Kimmel Scholar Award of the Sidney Kimmel Foundation for Cancer Research. Research supported in part by the Howard Hughes Medical Institute, with which P.W.S. is an investigator.

FIGURE LEGENDS

Fig. 1. Structural Overview of MEC-17. (a) Domain structure. Blue, acetyltransferase domain; Green, C-terminal extension (CTE); Gray, unstructured C-terminal tail. (b) Cartoon representation of the MEC-17^{Core}•acetyl-CoA structure, shown rotated 90º to the right and bottom. (c) Schematic representation of the architecture of MEC-17^{CORE} with secondary structure elements indicated.

Fig. 2. Multispecies Sequence Alignment of MEC-17 Homologs. The secondary structure is indicated above the sequence as blue cylinders (α-helices), green arrows (β-sheets), and gray lines (coil regions). The numbering below the alignment is relative to human MEC-17. Overall sequence conservation at each position is shaded in a color gradient from white (below 60 % identity) to yellow (60 % identity) to dark red (100 % identity). Magenta dots indicate mutated residues in our biochemical and *in vivo* analyses.

Fig. 3. Surface Properties of MEC-17. (a) Surface representation colored according to identified binding regions, with the acetyl-CoA binding pocket in yellow. (b) Surface representation of MEC-17^{CORE} colored according to a multispecies sequence alignment. The conservation at each position is mapped onto the surface and is shaded in a color gradient from white (below 60 % identity) to light yellow (60 % identity) to dark red (100 % identity). (c) Surface representation of MEC-17^{CORE} colored according to electrostatic potential. The electrostatic potential is plotted onto the surface and colored in a gradient from red (-10 k_BT/e) to blue (10 k_BT/e).

Fig. 4. Mutational Analysis. (a) WT and mutant forms of MEC-17 used in *in vitro* biochemical assays visualized on an SDS-PAGE gel. Multiple gels are shown due to size limitations. (b) Representative biochemical data for wild-type MEC-17. Reaction volumes separated by SDS-PAGE stained with Coomassie brilliant blue, and immunoblotted with an acetylated Lys40 α-tubulin specific antibody (top). Quantification of immunoblots using an Odyssey Imaging System (bottom). (c) The same as in (b), for a catalytically dead mutant (D157A). (d) Normalized initial velocities of all tested MEC-17 variants. Error bars represent standard error. All experiments were repeated at least 3 times. (e) Surface representation colored according to relative acetyltrasferase activity of mutants in a gradient from light purple (>80% activity) to dark purple (<20% activity). A 90° rotated view is shown on the right.

Fig. 5. Substrate Peptide Analysis. (a) Purified hsMEC17^{CORE} was mixed in excess with residues 25-49 of α-tubulin fused to SUMO and analyzed by size-exclusion chromatography. (b) Representative data from MEC-17 activitiy assays carried with the substrate peptide SUMO fusion in place of tubulin. Reaction volumes separated by SDS-PAGE stained with Coomassie brilliant blue, and immunoblotted with an acetylated Lys40 α-tubulin specific antibody. The anti-acetylated tubulin antibody recognizes an epitope within four residues of Lys40 when acetylated.

Fig. 6. *In vivo* analysis. (a) Merged differential interference contrast (DIC) and fluorescent image of transgenic worms expressing GFP-tagged human MEC-17 variants. (b) Touch sensitivity of *C. elegans* strains, represented as percentage of positive responses. Asterisks

indicate statistical significance based on Fisher's Exact Test. N2 denotes the wild-type strain; mec-17 the single null-mutant *mec-17(ok2109)* strain, atat-2 the single null mutant *atat-2(ok2415)* strain, and mec-17;atat-2 the double null mutant *mec-17(ok2109); atat-2(ok2415) strain.* Remaining strains are the mec-17;atat-2 double null mutant strain with the indicated hsMEC-17 variant present as a transgene. All hsMEC-17 variants, irrespective of catalytic activity, rescue the touch sensitivity of the double mutant.

Table 1. Expression Constructs

Table 2. Crystallographic analysis

[a] SSRL, Stanford Synchrotron Radiation Lightsource

[b] Highest-resolution shell is shown in parentheses

[c] As determined by Molprobity[33]

Table 3. Biochemical analysis of MEC-17 Mutants. The Michaelis-Menten K_m and V_{max} values were determined by direct fitting of linear initial velocities determined at substrate concentrations of 1-20 μM. Error values represents standard error.

REFERENCES

1. Nogales, E. (2000). Structural insights into microtubule function. *Annu. Rev. Biochem.* **69**, 277-302.
2. Luduena, R. F. (1998). Multiple forms of tubulin: different gene products and covalent modifications. *Int. Rev. Cytol.* **178**, 207-75.
3. Garnham, C. P. & Roll-Mecak, A. (2012). The chemical complexity of cellular microtubules: Tubulin post-translational modification enzymes and their roles in tuning microtubule functions. *Cytoskeleton (Hoboken)*.
4. Janke, C. & Bulinski, J. C. (2011). Post-translational regulation of the microtubule cytoskeleton: mechanisms and functions. *Nat. Rev. Mol. Cell Biol.* **12**, 773-86.
5. Szyk, A., Deaconescu, A. M., Piszczek, G. & Roll-Mecak, A. (2011). Tubulin tyrosine ligase structure reveals adaptation of an ancient fold to bind and modify tubulin. *Nat. Struct. Mol. Biol.* **18**, 1250-8.
6. Akella, J. S., Wloga, D., Kim, J., Starostina, N. G., Lyons-Abbott, S., Morrissette, N. S., Dougan, S. T., Kipreos, E. T. & Gaertig, J. (2010). MEC-17 is an alpha-tubulin acetyltransferase. *Nature* **467**, 218-22.
7. Gaertig, J., Cruz, M. A., Bowen, J., Gu, L., Pennock, D. G. & Gorovsky, M. A. (1995). Acetylation of lysine 40 in alpha-tubulin is not essential in Tetrahymena thermophila. *J. Cell Biol.* **129**, 1301-10.
8. Shida, T., Cueva, J. G., Xu, Z., Goodman, M. B. & Nachury, M. V. (2010). The major alpha-tubulin K40 acetyltransferase alphaTAT1 promotes rapid ciliogenesis and efficient mechanosensation. *Proc. Natl. Acad. Sci. USA* **107**, 21517-22.

9. Casale, C. H., Alonso, A. D. & Barra, H. S. (2001). Brain plasma membrane Na+,K+-ATPase is inhibited by acetylated tubulin. *Mol. Cell. Biochem.* **216**, 85-92.

10. Cueva, J. G., Hsin, J., Huang, K. C. & Goodman, M. B. (2012). Posttranslational acetylation of alpha-tubulin constrains protofilament number in native microtubules. *Curr. Biol.* **22**, 1066-74.

11. Friedman, J. R., Webster, B. M., Mastronarde, D. N., Verhey, K. J. & Voeltz, G. K. (2010). ER sliding dynamics and ER-mitochondrial contacts occur on acetylated microtubules. *J.Cell Biol.* **190**, 363-75.

12. Reed, N. A., Cai, D., Blasius, T. L., Jih, G. T., Meyhofer, E., Gaertig, J. & Verhey, K. J. (2006). Microtubule acetylation promotes kinesin-1 binding and transport. *Curr. Biol.* **16**, 2166-72.

13. Santander, V. S., Bisig, C. G., Purro, S. A., Casale, C. H., Arce, C. A. & Barra, H. S. (2006). Tubulin must be acetylated in order to form a complex with membrane Na(+),K (+)-ATPase and to inhibit its enzyme activity. *Mol. Cell. Biochem.* **291**, 167-74.

14. Topalidou, I., Keller, C., Kalebic, N., Nguyen, K. C., Somhegyi, H., Politi, K. A., Heppenstall, P., Hall, D. H. & Chalfie, M. (2012). Genetically separable functions of the MEC-17 tubulin acetyltransferase affect microtubule organization. *Curr. Biol.* **22**, 1057-65.

15. Zampar, G. G., Chesta, M. E., Carbajal, A., Chanaday, N. L., Diaz, N. M., Casale, C. H. & Arce, C. A. (2009). Acetylated tubulin associates with the fifth

cytoplasmic domain of Na(+)/K(+)-ATPase: possible anchorage site of microtubules to the plasma membrane. *Biochem. J.* **422**, 129-37.

16. Dyda, F., Klein, D. C. & Hickman, A. B. (2000). GCN5-related N-acetyltransferases: a structural overview. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 81-103.

17. Vetting, M. W., LP, S. d. C., Yu, M., Hegde, S. S., Magnet, S., Roderick, S. L. & Blanchard, J. S. (2005). Structure and functions of the GNAT superfamily of acetyltransferases. *Arch. Biochem. Biophys.* **433**, 212-26.

18. Kormendi, V., Szyk, A., Piszczek, G. & Roll-Mecak, A. (2012). Crystal structures of tubulin acetyltransferase reveal a conserved catalytic core and the plasticity of the essential N-terminus. *J. Biol. Chem.*

19. Taschner, M., Vetter, M. & Lorentzen, E. (2012). Atomic resolution structure of human alpha-tubulin acetyltransferase bound to acetyl-CoA. *Proc. Natl. Acad. Sci. USA* **109**, 19649-54.

20. Friedmann, D. R., Aguilar, A., Fan, J., Nachury, M. V. & Marmorstein, R. (2012). Structure of the alpha-tubulin acetyltransferase, alphaTAT1, and implications for tubulin-specific acetylation. *Proc. Natl. Acad. Sci. USA* **109**, 19655-60.

21. Rojas, J. R., Trievel, R. C., Zhou, J., Mo, Y., Li, X., Berger, S. L., Allis, C. D. & Marmorstein, R. (1999). Structure of Tetrahymena GCN5 bound to coenzyme A and a histone H3 peptide. *Nature* **401**, 93-8.

22. Tanner, K. G., Trievel, R. C., Kuo, M. H., Howard, R. M., Berger, S. L., Allis, C. D., Marmorstein, R. & Denu, J. M. (1999). Catalytic mechanism and function of

invariant glutamic acid 173 from the histone acetyltransferase GCN5 transcriptional coactivator. *J. Biol. Chem.* **274**, 18157-60.

23. Li, W., Zhong, C., Li, L., Sun, B., Wang, W., Xu, S., Zhang, T., Wang, C., Bao, L. & Ding, J. (2012). Molecular basis of the acetyltransferase activity of MEC-17 towards alpha-tubulin. *Cell Res* **22**, 1707-11.

24. Chalfie, M. & Sulston, J. (1981). Developmental genetics of the mechanosensory neurons of Caenorhabditis elegans. *Dev. Biol.* **82**, 358-70.

25. Lee, K. K. & Workman, J. L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. *Nat Rev Mol Cell Biol* **8**, 284-95.

26. Hoelz, A., Nairn, A. C. & Kuriyan, J. (2003). Crystal structure of a tetradecameric assembly of the association domain of Ca2+/calmodulin-dependent kinase II. *Mol. Cell.* **11**, 1241-51.

27. Doublie, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523-30.

28. Otwinowski, Z., Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-324.

29. La Fortelle, E. D. B., G. (1997). Maximum-likelihood heavy-atom parameteter refinement in the MIR and MAD methods. *Methods Enzymol.* **276**, 476-494.

30. CCP4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3.

31. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-32.

32. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-21.

33. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-83.

34. Laskowski, R. A. M., M.W.; Moss, D.S.; Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.

35. Edde, B., Rossier, J., Le Caer, J. P., Desbruyeres, E., Gros, F. & Denoulet, P. (1990). Posttranslational glutamylation of alpha-tubulin. *Science* **247**, 83-5.

36. Brenner, S. (1974). The genetics of Caenorhabditis elegans. *Genetics* **77**, 71-94.

37. Hobert, O. (2002). PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. *Biotechniques* **32**, 728-30.

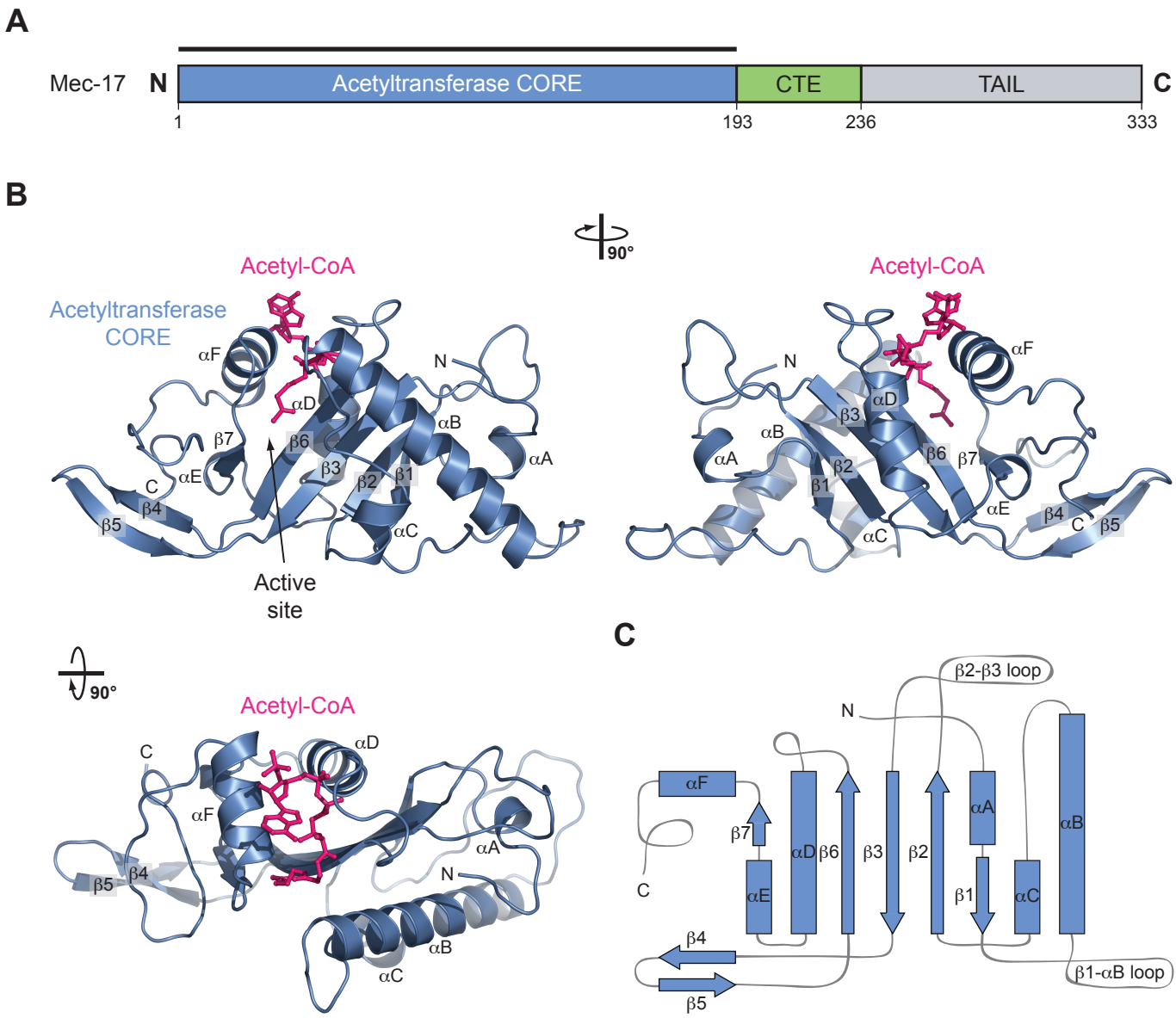
38. Mello, C. & Fire, A. (1995). DNA transformation. *Methods Cell Biol.* **48**, 451-82.

39. Shaham, S. & Horvitz, H. R. (1996). Developing Caenorhabditis elegans neurons may contain both cell-death protective and killer activities. *Genes Dev.* **10**, 578-91.

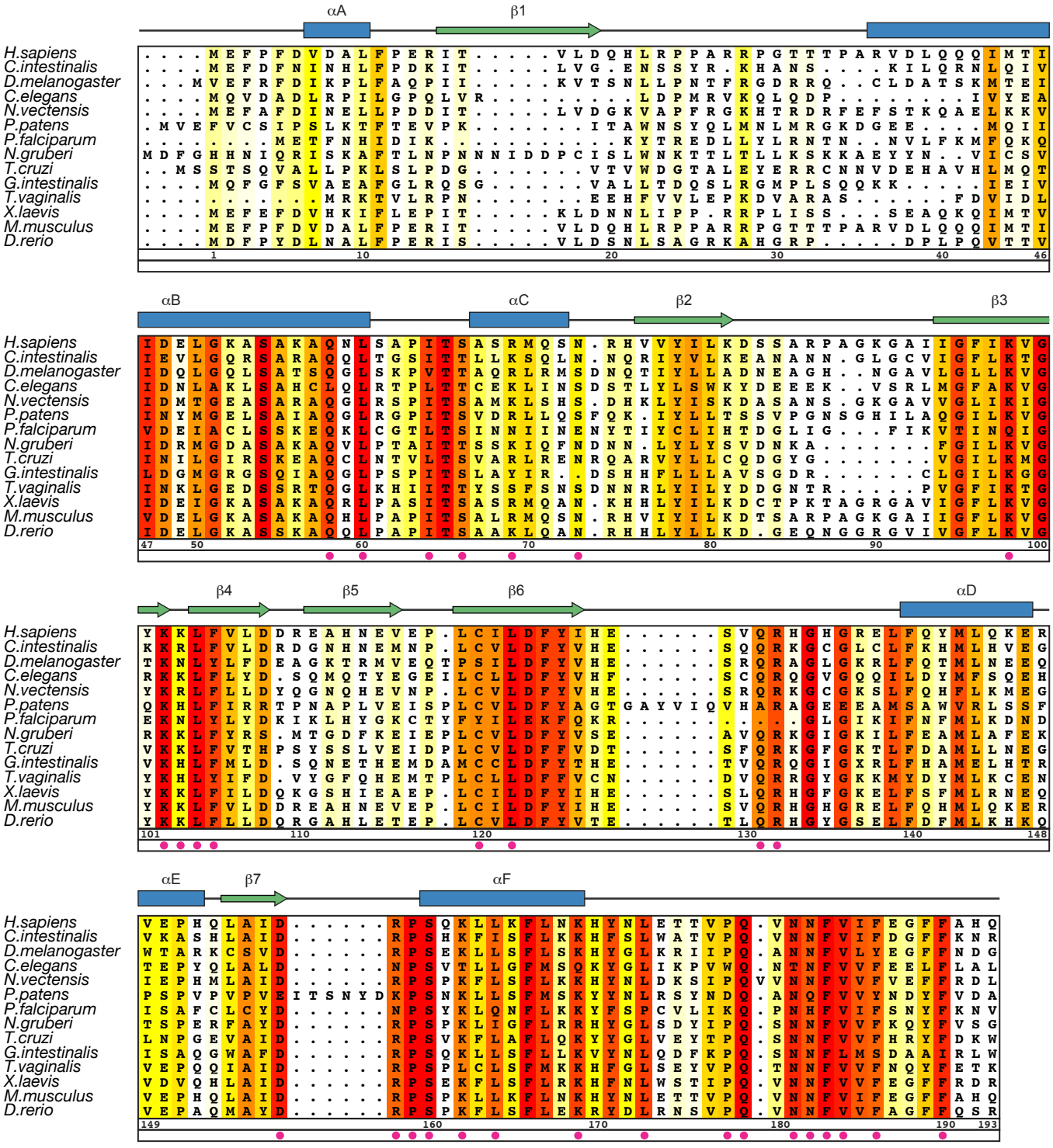
40. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**, 403-5.

41. Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37-40.

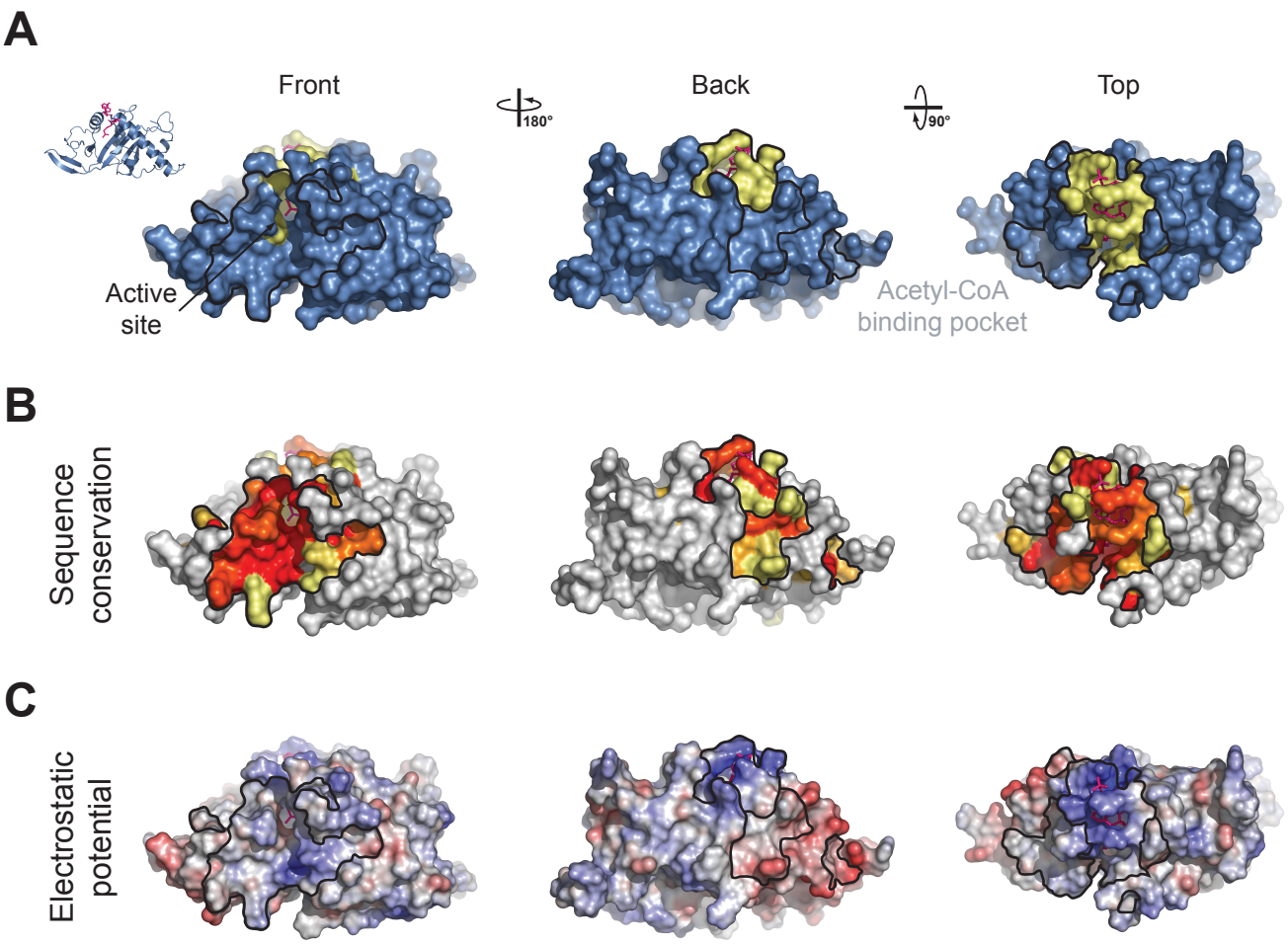
42. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037-10041.



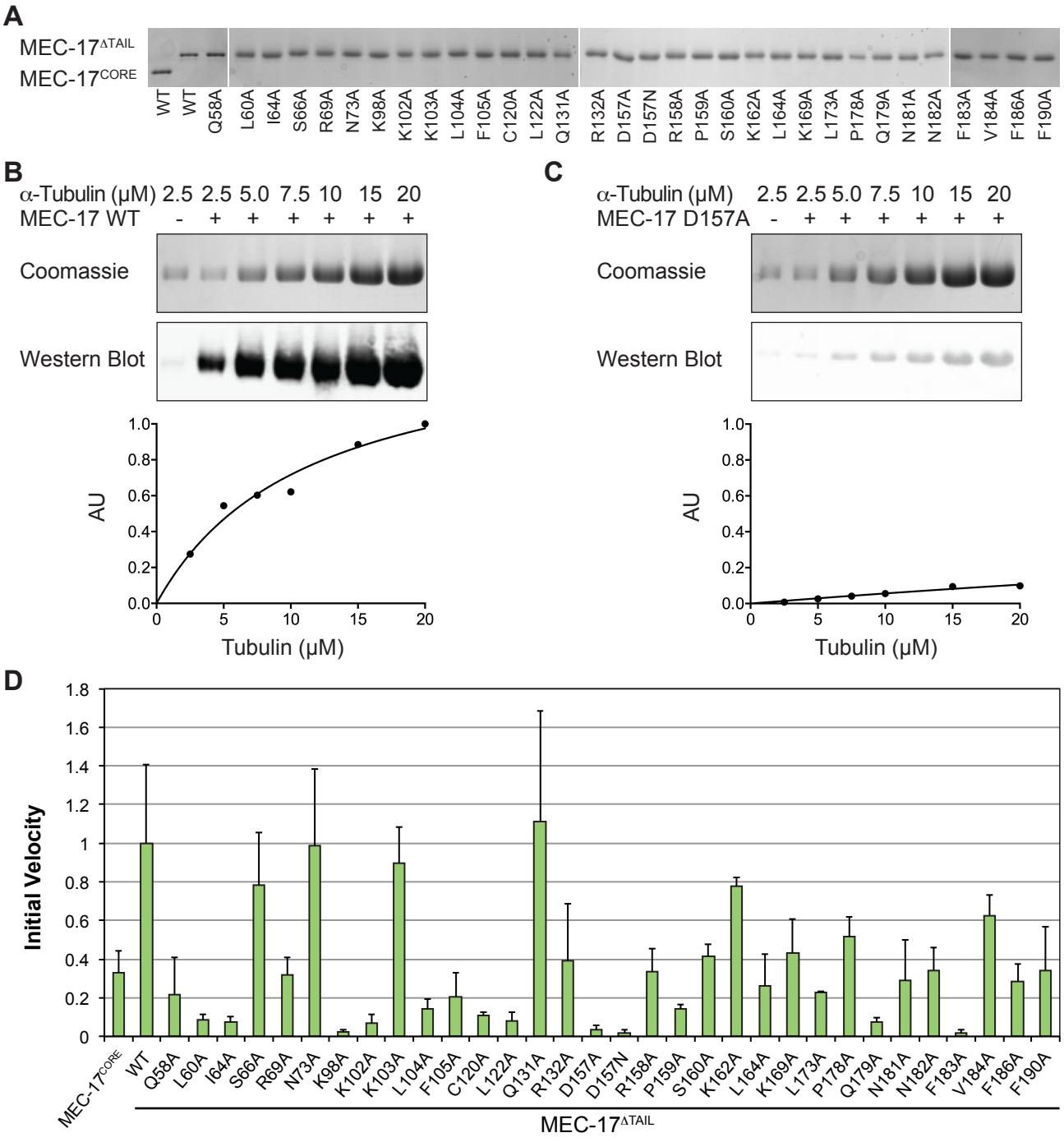
Chapter II, Figure 1



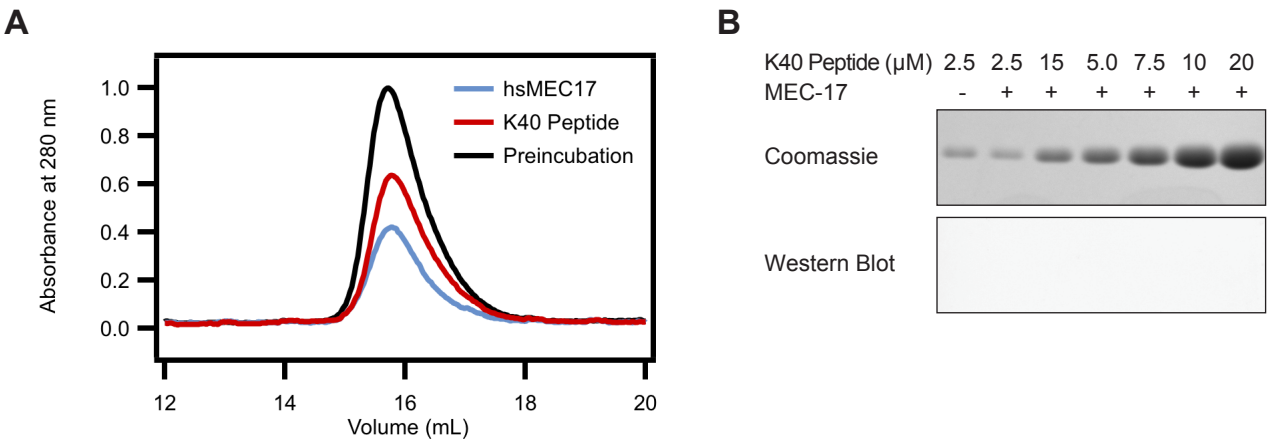
Chapter II, Figure 2



Chapter II, Figure 3

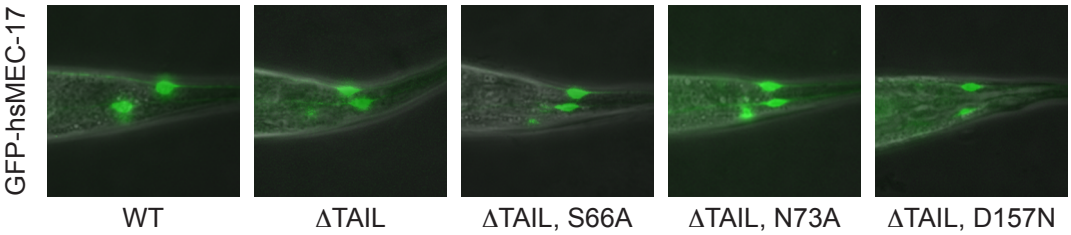


Chapter II, Figure 4

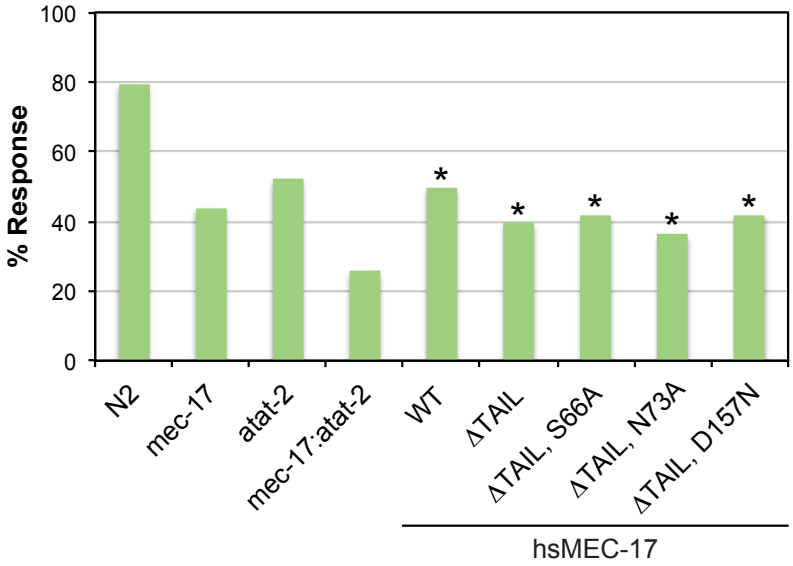


Chapter II, Figure 5

A



B



Chapter II, Figure 6

Table 1. Expression constructs

Bacterial expression constructs				
Protein	Residues (Mutations)	Expression vector	Restriction sites 5', 3'	N-terminal overhang
hsMEC-17	1-193	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (D48A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (Q58A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (L60A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (I64AA)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (S66A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (R69A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (N73A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (K98A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (K102A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (K103A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (L104A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (F105A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (C120A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (L122A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (Q131A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (R132A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (D157A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (D157N)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (R158A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (P159A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (S160A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (K162A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (L164A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (K169A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (L173A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (P178A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (Q179A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (N181A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (N182A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (F183A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (V184A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (F186A)	pET28a-PreS	NdeI, NotI	GPH
hsMEC-17	1-236 (F190A)	pET28a-PreS	NdeI, NotI	GPH
hs α-tubulin	25-49	pET28b-SUMO	BamHI, NotI	His_6-SUMO

***C. elegans* Expression Constructs**				
Protein	Residues (Mutations)	Vector	Restriction Sites 5', 3'	C-terminal overhang
hsMEC-17	1-333	pPD95.75	HindIII, BamHI	-GFP
hsMEC-17	1-263	pPD95.75	HindIII, BamHI	-GFP
hsMEC-17	1-236 (S66A)	pPD95.75	HindIII, BamHI	-GFP
hsMEC-17	1-236 (N73A)	pPD95.75	HindIII, BamHI	-GFP
hsMEC-17	1-236 (D157N)	pPD95.75	HindIII, BamHI	-GFP

Table 2. Crystallographic analysis

Data collection	
Protein	MEC-17^{CORE}
PDB code	4IF5
Synchrotron	SSRL
Beamline	BL12-2
Space group	$P2_12_12_1$
Cell parameters	
a, *b*, *c* (Å)	41.7, 122.2, 37.5
α, β, γ (°)	90.0, 90.0, 120.0
	Se peak
Wavelength (Å)	0.97940
Resolution (Å)	50.0 – 1.7
R_{sym} (%)[b]	7.5 (70.4)
$\langle I \rangle / \langle \sigma I \rangle$[b]	27.8 (2.0)
Completeness (%)[b]	97.2 (81.6)
No. observations	345,965
No. unique reflections	39,430 (3,277)
Redundancy	8.8 (6.4)
Refinement	
Resolution (Å)	40.0 – 1.7
No. reflections total	37,288
No. reflections test set	3,500 (9.4%)
R_{work} / R_{free} (%)	17.8 / 21.3
No. atoms	1,826
Protein	1,567
Ligand	52
Water	209
B-factors	36.2
Protein	35.9
Ligand	22.5
Water	42.4
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.1
Ramachandran plot[c]	
Favored (%)	94.3
Additionally allowed (%)	5.7
Outliers (%)	0.0

[a] SSRL, Stanford Synchrotron Radiation Lightsource
[b] Highest resolution shell is shown in parentheses
[c] As determined by MolProbity[33]

Table 3. Biochemical Analysis of MEC-17 Mutants

Biochemical Analysis of MEC-17 Mutants			
Mutation	*Km* (μM)	$k_{cat} \times 10^{-2}$ (sec^{-1})	$k_{cat} / Km \times 10^{3}$ ($sec^{-1}\mu M^{-1}$)
WT (1-236)	13.64 ± 2.70	3.79 ± 0.39	2.78 ± 0.84
WT (1-193)	45.37 ± 55.76	3.24 ± 3.03	0.71 ± 1.53
Q58A	N/A	N/A	N/A
L60A	N/A	N/A	N/A
I64A	N/A	N/A	N/A
S66A	84.26 ± 66.24	9.23 ± 6.16	1.10 ± 1.59
R69A	98.42 ± 48.78	2.76 ± 1.18	0.28 ± 0.26
N73A	23.39 ± 3.24	4.91 ± 0.43	2.10 ± 0.47
K98A	N/A	N/A	N/A
K102A	N/A	N/A	N/A
K103A	31.24 ± 18.55	2.79 ± 1.15	0.89 ± 0.61
L104A	N/A	N/A	N/A
F105A	N/A	N/A	N/A
C120A	N/A	N/A	N/A
L122A	N/A	N/A	N/A
Q131A	14.64 ± 4.06	2.75 ± 1.20	1.88 ± 1.34
R132A	52.03 ± 29.31	3.32 ± 1.46	0.63 ± 0.92
D157A	NA	N/A	N/A
D157N	NA	N/A	N/A
R158A	79.48 ± 67.3	4.81 ± 3.42	0.61 ± 0.94
P159A	36.88 ± 18.73	0.97 ± 0.35	0.26 ± 0.23
S160A	50.94 ± 22.51	3.67 ± 1.26	0.72 ± 0.56
K162A	48.40 ± 13.66	6.49 ± 1.46	1.34 ± 0.68
L164A	3.45 ± 5.25	0.56 ± 0.12	1.62 ± 2.81
K169A	43.28 ± 20.78	3.45 ± 1.25	0.80 ± 0.67
L173A	36.75 ± 54.19	1.99 ± 2.12	0.54 ± 1.38
P178A	18.80 ± 8.39	2.39 ± 0.63	1.27 ± 0.90
Q179A	N/A	N/A	N/A
N181A	N/A	N/A	N/A
N182A	58.17 ± 37.87	4.14 ± 2.15	0.71 ± 0.83
F183A	N/A	N/A	N/A
V184A	47.12 ± 55.95	5.77 ± 5.24	1.22 ± 2.57
F186A	15.23 ± 15.17	0.77 ± 0.43	0.51 ± 0.79
F190A	107.30 ± 99.03	5.97 ±4.82	0.56 ± 0.96

Chapter III

Structural and Functional Analysis of the Putative Proto-Oncoprotein DPF2

This chapter was written by Andrew M. Davenport[1], Ferdinand M. Huber[1], and André Hoelz[1].

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

ABSTRACT

DPF2 is a highly conserved member of the d4 family of proteins that is present in all human tissues, and has recently been shown to repress myeloid differentiation in acute myelogenous leukemia (AML). We have solved the crystal structure of the C-terminal tandem PHD finger domain of human DPF2 to a 1.6 Å resolution. Each PHD finger displays the canonical central two-stranded b-sheet and two zinc atoms anchored by a His-Cys3 motif followed by a Cys4 motif. We biochemically show that DPF2 preferentially binds H3 tail peptides acetylated at Lys14, and binds H4 tail peptides with no preference for acetylation state. Through a structural and mutational analysis we identify the molecular basis of histone recognition. A structural comparison of DPF2 and a recently determined solution structure of DPF3b suggest the molecular determinants for ligand specificity are distributed throughout the entire domain, rather than at the binding sites. We propose a model for the role of DPF2 in AML and identify the DPF2 tandem PHD finger domain as a promising novel target for anti-leukemia therapeutics.

INTRODUCTION

Acetylation is a widespread posttranslational modification (PTM) that can serve to affect a range of cellular processes, including transcriptional regulation, protein localization, enzymatic regulation, and proteasomal degradation.[1,2] One mechanism through which acetylation functions is the recruitment of specific binding partners. Cells have evolved a number of mechanisms for the molecular recognition of acetyl-lysine modifications, including bromodomains, plextrin homology (PH) domains, and plant homeodomain (PHD) fingers.[3] PHD fingers, which contain two zinc finger motifs, are notable for their ability to read a diverse number of PTMs, including unmodified, methylated, or acetylated lysines as well as unmodified arginines.[4]

d4, zinc and double PHD finger 2 (DPF2, also known as requiem/REQ or ubi-d4), the founding member of the d4 family of proteins, was initially discovered as a factor required for apoptosis in myeloid cells.[5] d4 proteins, which in humans include DPF1 and DPF3 in addition to DPF2, are characterizd by an N-terminal requiem domain, a central C2H2-type zinc finger domain, and a C-terminal tandem PHD finger domain.[6] DPF2 is expressed ubiquitously in human tissues.[7,8] Relatively little is known about the function or regulation of DPF2 or its individual domains. DPF2 has been shown to bridge SWI/SNF components and RelB/p52 to effect noncanonical NF-kB signaling, repress the transcriptional activity of Estrogen-related receptor a,[9] act as a globin switching factor,[10] and be a target for Staufen-1 mediated mRNA decay.[10]

DPF2, along with its fellow d4 family members DPF1 and DPF3, has been implicated in a range of different cancer types.[11-16] RUNX1 is one of the most frequently altered genes in acute myelogenous leukemia (AML), with the t(8;21) translocation that

produces the RUNX1-ETO fusion protein alone present in 4-10 % of adult and 12-30 % of pediatric patients.[17] Recent work has shown that RUNX1 interacts with DPF2 in AML patient samples.[11] This interaction was shown to be dependent upon methylation of RUNX1 on Arg223 by the type I arginine methyltransferase PRMT4, as mutations of Arg223 or chemical inhibition of PRMT4 abolished or reduced the RUNX1-DPF2 interaction, respectively. The interaction was further mapped in pull-down assays to a 12-residue RUNX1 peptide that includes Arg223. Knockdown of either DPF2 or PRMT4 led to increased expression of RUNX1 target genes and enhanced myeloid differentiation, and as such a model was proposed whereby DPF2 and RUNX1 form a methylation dependent repressive complex in AML. As the experiments above were carried out with cell lysate and not recombinant protein, it is currently unclear if DPF2 directly interacts with methylated RUNX1 or they act concertedly as part of a larger complex. A crystal structure of the DPF2 central C2H2-type zinc finger suggested a putative DNA binding surface,[18] but further structural or mechanistic understanding of DPF2 or its role in AML is lacking. Here, we present the crystal structure of the human DPF2 tandem PHD finger domain. We demonstrate that it binds H3 and H4 histone tails, and through a mutational analysis identify the primary determinants of histone recognition. We develop a model for the role of DPF2 in AML and propose the disruption of the DPF2-histone interaction as a novel strategy for the treatment of AML.

RESULTS

Structure determination

Based on secondary structure prediction and sequence conservation analyses, we designed a series of N-terminal truncation constructs of *H. sapiens* DPF2 for expression in *E. coli* (Table 1). We identified a fragment corresponding to the tandem PHD finger domain, encompassing residues 270 to 391, hereafter referred to as $DPF2^{PHD}$. $DPF2^{PHD}$ was prone to oxidation as observed by mass spectrometry (data not shown) and required immediate use in any experiments following purification. Crystals of $DPF2^{PHD}$ were obtained in the tetragonal space group *I*422 and diffracted to 1.6 Å resolution. The structure was solved by single-wavelength anomalous dispersion using anomalous X-ray diffraction data obtained by taking advantage of the four endogenously bound Zn^{2+} ions per $DPF2^{PHD}$ molecule. The final model contains residues 270 to 386, has excellent stereochemical properties, and was refined to R_{work} and R_{free} values of **15.5** % and **18.3** %, respectively. For the details of the data collection and refinement statistics, see Table 2.

Structural overview

$DPF2^{PHD}$ contains two PHD fingers, the first is comprised of residues 270-329 and the second of residues 330-386 (hereafter referred to as PHD1 and PHD2, respectively) (Fig. 1a). Each PHD finger contains the canonical architecture first described for the Williams Syndrome Transcription Factor (Fig. 1 and 2).[19] This includes a two-stranded anti-parallel β-sheet (β1-2 in PHD1 and β3-4 in PHD2) and two zinc atoms coordinated by a His–Cys3 motif followed by a Cys4 motif in a cross-brace topology (Cys273, Cys276, His303,

Cys306 and Cys295, Cys298, Cys324, Cys327 for PHD1, and Cys330, Cys333, His353, Cys356 and Cys345, Cys348, Cys371, Cys374 for PHD2) (Fig. 2a). PHD1 and PHD2 adopt a similar conformation, with an RMSD of ~1.9 $Å^2$ over 45 Cα atoms (Fig. 2b). The two PHD fingers mainly differ by the presence of an additional short α-helix in PHD1 (αB, residues 311-319) resulting in three flanking helices in PHD1 (αA, αB, and αC) vs. two in PHD2 (αD, and αE).

A multispecies sequence alignment shows $DPF2^{PHD}$ to be highly conserved from *H. sapiens* to *C. elegans*, with 61% identity (68/111 residues) among all species tested (Fig. 3a). While the surface of $DPF2^{PHD}$ maintains the high level of conservation observed for the entire domain, the front surface is the most invariant region of the protein (Fig. 3b). This surface contains a large hydrophobic patch on its N-terminal half and a large negatively charged patch on its C-terminal half, corresponding to PHD1 and PHD2, respectively (Fig. 3c), and is believed to mediate recognition of histone binding partners (see below). $DPF2^{PHD}$ shows a high level of sequence conservation among human d4 family members as well as across species. $DPF2^{PHD}$ shares 80% identity (93/116 residues) with the tandem PHD finger domain of the DPF3 isoform DPF3b, for which a solution NMR structure was recently determined (Fig. 4a).[20] A comparison of the $DPF3b^{PHD}$ and $DPF2^{PHD}$ structures revealed a movement in the relative position of the individual PHD domains, corresponding to a rotation by ~25° (Fig. 5a). This is best illustrated by a root-mean square deviation (RMSD) of ~1.4 $Å^2$ over 59 Cα atoms and ~1.3 $Å^2$ over 49 Cα atoms when each PHD domain is aligned individually, but an RMSD of ~2.3 $Å^2$ over 108 Cα atoms when the tandem PHD domains are aligned as a whole. The solution structure of DPF3b was determined in complex with various histone peptides, while our DPF2 crystal

structure is in an *apo* form. Given this fact, the high sequence identity between the two proteins, and no obvious differences in the PHD1-PHD2 interface, the combined structures may reveal a conformational change that occurs upon ligand binding. This notion is supported by the presence of hinge region at Cys327 between the two PHD domains in $DPF2^{PHD}$ as predicted by HingeProt.[21] Alternatively, this difference may represent a persistent structural feature of each individual protein that could account for differences in ligand preference. Indeed, the tandem PHD domain of the acetyltransferase MOZ/MYST3, for which there exist both *apo*- and ligand-bound states, shows no conformational change upon ligand binding (Fig. 6 and 7a).[22] Further structural studies will be required to distinguish between these two possibilities.

Our sequence conservation analysis shows $DPF2^{PHD}$ to be highly conserved; however, the N-terminal 270 residues show much greater sequence divergence both across species and across family members (Fig. 8).

Interaction Analysis

Recent reports of an interaction between DPF2 and a methylated form of RUNX1,[11] combined with remarkable diversity of PHD domains as recognition modules,[4] led us to explore whether $DPF2^{PHD}$ might directly recognize methylated RUNX1. Specifically, we carried out isothermal titration calorimetry (ITC) experiments to test whether purified $DPF2^{PHD}$ could interact with monomethylated or asymmetric dimethylated RUNX1 peptides (N-TPNPR^{Me}ASLNHSC-C or N-TPNPR^{ADMA}ASLNHSC-C). Our results demonstrate that there is no direct interaction between $DPF2^{PHD}$ and the RUNX1 peptides tested (Fig. 9 and 10). As most PHD fingers have been shown to read the N-terminal tail of

the histone protein H3,[4] we next explored the possibility that DPF2^{PHD} recognizes histone tails. Indeed ITC experiments with DPF2^{PHD} and H3 (N-ARTKQTARKSTGGKA-C) or H4 (N-GLGKGGAKRHRKVLR-C) peptides showed binding with a dissociation constant (K_d) of ~5.4 μM and ~0.4 μM, respectively. We further tested if DPF2^{PHD} binding to H3 or H4 was modulated by acetylation. DPF2^{PHD} bound to H3 acetylated at Lys14 (N-ARTKQTARKSTGGK^{Ac}A-C, H3K14^{Ac}) with a K_d of ~2.0 μM and to H4 acetylated at Lys16 (N-GLGKGGAK^{Ac}RHRKVLR-C, H4K16^{Ac}) with a K_d of ~0.4 μM. These data differ from the previously determined binding affinities for DPF3b^{PHD}.[20] Specifically, while both DPF2^{PHD} and DPF3b^{PHD} bind to H3 in the low micromolar range, and preferentially bind to acetylated H3 peptides, DPF2^{PHD} shows an ~100 fold increase in affinity for H4 relative to DPF3b^{PHD}, and this interaction is not influenced by acetylation. The results establish that, despite high sequence conservation, different members of the d4 family differ in their specificity for histone binding partners. Moreover, to our knowledge, DPF2^{PHD} represents the first PHD domain to bind H4 with higher affinity than H3, adding to the impressive functional diversity of this structural unit.

Despite major effort, we were unsuccessful in obtaining a crystal structure of DPF2^{PHD} in complex with a histone peptide. However, by using available structures of tandem PHD fingers in complex with histone peptides as references, we were able to generate models for the DPF2^{PHD}-histone interactions. We first generated a model for the DPF2-H3K14^{Ac} interaction based on the solution NMR structure of the DPF3b•H3K14^{Ac} complex.[20,23] Our model shows the peptide primarily engages two regions on the conserved front face of the protein (Figs. 5 and 7). The acetyl-lysine binds in a highly conserved hydrophobic pocket on the surface of PHD1, primarily composed of Phe275, Leu307, and

Trp322. The hydrophobic pocket is further lined by Asp275, which is in hydrogen bond distance to the acetyl amide group, and by Arg300, which interacts with the acyl-moeity of the acetyl-lysine (Fig. 5b). The N-terminal part of the H3 peptide binds along the surface of PHD2. The highly conserved residues Glu326 and Asp346 make electrostatic interactions with Lys4 and Arg2 residues of the histone tail, respectively, while Leu342, Leu343, and Phe344 form an antiparallel β-sheet with histone residues Arg2, Thr3, and Lys4. We next generated an alternative model for the DPF2-H3K14^{Ac} interaction based on the MOZ/MYST3•H3K14^{Ac} complex crystal structure.[22,23] In this second model, DPF2 utilizes the same hydrophobic pocket composed of Phe275, Leu307, and Trp322 for acetyl-lysine binding, and similarly uses a network of electrostatic interaction between Glu326, Asp346, Asp349 of DPF2 and Arg2 and Lys4 of H3 for binding. The main difference between the two models is the presence of an α-helical segment in H3 (Lys4-Thr11). Further structural studies will be required to demonstrate whether DPF2 interacts with histones in an extended peptide conformation manner reminiscent of DPF3b or induces an α-helical structure similar to MOZ/MYST3. However, as both DPF3b and MOZ/MYST3 use similar structural features to engage their binding partners, we used a mutational analysis to validate our models for the DPF2^{PHD}•H3 and DPF2^{PHD}•H4 complexes. Indeed, an alanine mutation of Asp346 on the negatively charged DPF2^{PHD} surface patch reduced the affinity as determined by ITC of the DPF2^{PHD}-H3 and DPF2^{PHD}-H4 interactions by ~6 fold from ~2.0 μM to 11.8 μM, and by ~10-fold from 0.4 μM to 4.51 μM, respectively (Fig. 9). Individual alanine mutations of Phe275 or Arg300 at the acetyl-lysine binding pocket had a negligible effect on DPF2^{PHD}-H3K14^{Ac} or DPF2^{PHD}-H4K16^{Ac} binding. However, a combination of the Phe275, Arg300, and Asp346 alanine mutations completely abolished

both interactions. These results demonstrate that both the hydrophobic pocket on PHD1 and the negative surface patch on PHD2 contribute to the interactions with H3 or H4, with the negatively charged surface patch being the stronger determinant.

Together, these results demonstrate that while DPF2, DPF3b, and MOZ/MYST3 differ in their preference for binding partners, they use similar structural features to engage histone peptides. Further structural studies will be required to determine the molecular basis for the increased affinity of DPF2 for H4.

DISCUSSION

We have determined the crystal structure of the tandem PHD domain of human DPF2, and have shown that it binds to the tails of histones H3 and H4. A combination of a mutational analysis and a structural comparison to DPF3b and MOZ/MYST3 suggests that the DPF2 tandem PHD domain simultaneously utilizes two binding sites: a hydrophobic pocket on PHD1 that binds the acetylated lysine residue, and a negatively charged surface patch on PHD2 that binds additional basic residues in histone tails.

DPF2 shows different preferences for histone binding compared to DPF3b or MOZ/MYST3. Both DPF3b and MOZ/MYST3 bind H3 stronger than H4, and this interaction is strengthened by acetylation of Lys14. While DPF2 also binds H3 in an acetylation-sensitive manner, it binds H4 stronger than H3, unaffected by acetylation. DPF2 is notably the first PHD domain to bind preferentially to H4. These biochemical properties are particularly striking given the high degree of both sequence conservation and structural similarities between DPF2 and DPF3b. Indeed, there are no apparent structural features at the binding sites to account for these differences. Given this fact, we prefer a model whereby the increased affinity of DPF2 for H4 is not a result of differences at the binding pocket, but rather an interaction network present throughout the entire domain, as previously observed for the small molecule kinase inhibitor Gleevec.[24-26] In this case, distributed structural differences affected a conformational change and subsequent induced fit responsible for high affinity. More detailed analyses would be required to determine if DPF2 utilizes a similar mechanism.

In addition to varying specificities in histone binding, variation among the requiem and C2H2 domains is likely to further account for the diversity of function within the d4

family by forming specific protein-protein or protein-nucleic acid interactions. Further studies are required to address the functional role of the DPF2 requiem and C2H2 domains.

Recent work has highlighted a role for DPF2 in myeloid differentiation and acute myelogenous leukemia (AML) mediated by a methylation dependent interaction between DPF2 and RUNX1. We have shown that the DPF2 tandem PHD domain binds to histone tails, and is not capable of directly interacting with methylated RUNX1 peptides that are sufficient to pull-down DPF2 from cell lysate. It remains possible that the additional N-terminal regions of DPF2 mediate this interaction, but given that the protein contains no homology to domains known to recognize methylated arginine, we prefer a model whereby DPF2 and RUNX1 (or altered forms of RUNX1 such as the RUNX1-ETO fusion) interact indirectly as part of a larger complex (Fig. 11). DPF2 could serve to mistarget this complex to specific chromatin sites. There are several candidates for the proteins that bridge the interaction between DPF2 and RUNX1, including the SWI/SNF chromatin remodeling complex, which has been shown to interact with both proteins.[27-29] A focused effort to identify the remaining components of the DPF2•RUNX1 complex should shed light on the mechanisms of AML and provide future potential therapeutic targets.

Our model suggests that DPF2 mistargets RUNX1 to specific chromatin sites to repress transcriptional activity, ultimately contributing to AML. A detailed molecular understanding of this process will require further experimentation, but given this framework, disrupting the RUNX1-DPF2 or DPF2-histone interactions represents a promising strategy for treatment of AML.

MATERIALS AND METHODS

Protein expression and purification

DNA fragments of *H. sapiens* DPF2 were amplified by PCR and cloned into the vector pGEX-6P-1 (GE Healthcare). DPF2 mutants were generated by QuikChange mutagenesis (Stratagene) and confirmed by DNA sequencing. Details of all constructs are listed in Table S1. All proteins were expressed in *E. coli* BL21-CodonPlus(DE3)-RIL cells (Stratagene) grown in LB media supplemented with 0.1 mM zinc acetate and appropriate antibiotics. Protein expression was induced at an OD_{600} of approximately 0.6 with 500 mM IPTG for 16-18 hours at 18 °C. Cells were harvested by centrifugation and resuspended in a buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, 5 mM DTT, 2 mM bovine lung aprotinin (Sigma), and complete EDTA-free protease inhibitor cocktail (Roche), and subsequently flash frozen in liquid nitrogen. Thawed cells were lysed with a cell disrupter (Avestin) and the lysate was centrifuged for 1 hour at 40,000 x g. The cleared lysate was applied to a glutathione sepharose column equilibrated in buffer containing 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM DTT (GE Healthcare) and eluted via a glutathione gradient. Pooled fractions were cleaved with PreScission protease (GE Healthcare) for 12 hours. Cleaved protein was bound to a HiTrapQ HP (GE Healthcare) column and eluted via a linear NaCl gradient, concentrated, and injected onto a HiLoad 16/60 Superdex 75 column equilibrated in 20 mM TRIS (pH 8.0), 100 mM NaCl, and 5 mM DTT. Pooled fractions were concentrated to 15 mg/ml and used in subsequent crystallization trials or for *in vitro* activity assays.

Structure determination and refinement

Crystals of DPF2^{PHD} were grown at 21 °C in hanging drops containing 1 µl of the protein and 1 µl of a reservoir solution consisting of 2.2 M ammonium sulfate and 0.15 M ammonium citrate. Crystals grew in the tetragonal space group *I*422 and reached a maximum size of approximately 500 µm x 500 µm x 500 µm within a week. For cryo protection, crystals were stabilized in the crystallization solution supplemented with 20 % (v/v) ethylene glycol (added in 5 % increments). X-ray diffraction data were collected at 100 K at beamline 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL). X-ray intensities were processed using XDS.[30]

The structure was solved by single wavelength anomalous dispersion (SAD) using X-ray diffraction data collected utilizing the anomalous scattering of the endogenous Zn^{2+} ions. Identification of heavy metal sites, calculation of phases, and density modification were performed in PHENIX.[31] Iterative rounds of model building and refinement in Coot[32] and PHENIX[31] yielded final models containing all residues of DPF2^{PHD} fragment except for residues **387 to 391**, for which no electron density was observed. The final R_{work} and R_{free} values were **15.5 % and 18.4 %**, respectively. The model possesses excellent stereochemical parameters with no residues in the generously allowed or disallowed regions of the Ramachandran plot as assessed with MolProbity.[33] For details of the data collection and refinement statistics, see Table 1.

ITC measurements

Isothermal titration calorimetry (ITC) measurements were performed at 25 °C using a MicroCal ITC200 Calorimeter (Malvern). Protein samples were stored in a buffer

containing 25 mM HEPES (pH 8.0), 2 mM TCEP and 10 % (v/v) glycerol. All histone peptide concentrations were determined by amino acid analysis. The titration was performed by injecting 0.5-3.5 mM histone peptide into 10-90 μM DPF2. Titrations consisted of 17 injections of 2.45 μl with a spacing of 180 seconds. Reference power was cal/sec 11.4 μcal/sec and the feedback mode/gain was low. The heat generated from dilution was subtracted for baseline correction. Baseline corrected data were analyzed with Origin 7.0 software.

Illustration and figures

The sequence alignment of DPF2 was generated using ClustalX2[34] and colored with Alscript.[35] Figures were generated using PyMOL (www.pymol.org). Electrostatic potential was calculated with Adaptive Poisson-Boltzmann Solver (APBS) software.[36]

ACKNOWLEDGEMENTS

We thank David King for mass spectrometry analysis, Shaun Bevers (Biophysics Core Facility of the University of Colorado, Denver) for isothermal titration calorimetry experiments, Jens Kaiser and the scientific staff of SSRL Beamline 12-2, and the scientific staff of APS Beamline GM/CA-CAT for their support with X-ray diffraction measurements.

FIGURE LEGENDS

Fig. 1. Structure of *H. sapiens* DPF2^{PHD}. (a) Domain structure. Light Grey, predicted unstructured regions (U); Blue, requiem domain (REQUIEM); Dark Grey, C2H2 zinc finger domain (C2H2); Yellow, tandem PHD finger 1 (PHD1); Orange, tandem PHD finger 2 (PHD2). The bar above the domain structure marks the crystallized fragment. (b) Structure of the DPF2 PHD domain. 90°-rotated views are shown below and to the right. (c) Representative final $2|F_O|-|F_C|$ electron density map countered at 1.0 σ around a stick representation of the indicated residues.

Fig. 2. (a) Final $2|F_O|-|F_C|$ electron density map countered at 1.0 σ around a stick representation of the Zn^{2+} coordinating residues of DPF2^{PHD1} (left) and DPF2^{PHD2} (right). The Zn^{2+} atoms are shown as orange spheres. (b) Superposition of DPF2^{PHD1} and DPF2^{PHD2}, colored as in Fig. 1. (c) Superposition of the DPF2^{PHD1} and DPF2^{PHD2}, colored as in Fig. 1., along with representative PHD domains (PDB codes 2KWJ, 2PUY, 3SOU, 4LK9, 4L7X, 4QF3)[20,22,37-39], all colored in grey.

Fig. 3. Multispecies sequence alignment and surface properties of DPF2^{PHD}. (a) The secondary structure is indicated above the sequence as blue cylinders (α-helices), green arrows (β-sheets), and gray lines (coil regions). Residue numbering below the alignment corresponds to human DPF2. Overall sequence conservation at each position is shaded in a color gradient from white (below 40 % similarity) to yellow (60 % similarity) to dark red (100 % identity). Blue dots indicate Zn^{2+} coordinating residues, purple dots indicate residues mutated in our biochemical analysis, grey dots indicate residues for which there

was no observed electron density. (b) Surface representation of DPF2^{PHD} colored according to the above multispecies sequence alignment. (c) Surface representation of DPF2^{PHD} colored according to electrostatic potential. The electrostatic potential is plotted onto the surface and colored in a gradient from -10 k_BT/e (red) to 10 k_BT/e (blue).

Fig. 4. Sequence alignment of human d4 family members. (a) The secondary structure is indicated above the sequence as blue cylinders (α-helices), green arrows (β-sheets), and gray lines (coil regions). The numbering below the alignment is relative to human DPF2. Overall sequence conservation at each position is shaded in a color gradient from white (below 40 % similarity) to yellow (60 % similarity) to dark red (100 % identity). Purple dots indicate residues mutated in our biochemical analysis, and grey dots indicate residues for which there was no observed electron density. (b) Surface representation of a model of DPF2^{PHD} in complex with the indicated histones peptides based on the DPF3b solution structures (PDB IDs 2KWJ and 2KWN).[20] The model is colored according to the above multispecies sequence alignment.

Fig. 5. Structural comparison of *H. sapiens* DPF2 to *H. sapiens* DPF3b. (a) A ribbon representation of DPF2^{PHD} (left, colored as in Fig. 1a), DPF3b•H3K14^{Ac} (middle, colored in gray, PDB ID 2KWJ)[20], and a superposition based on PHD2 (right). (b) A surface representation of the DPF3b•H3K14^{Ac} NMR structure (top, PDB ID 2KWJ)[20] and the DPF3b•H4K16^{Ac} solution NMR structure (bottom, PDB ID 2KWN)[20]. Boxes show a cartoon representation of the indicated binding sites superposed with our DPF2^{PHD} crystal structure. Residues mutated in our biochemical analysis are shown in stick representation.

Fig. 6. Sequence alignment of *H. sapiens* DPF2 and *H. sapiens* MOZ/MYST3. (a) The secondary structure is indicated above the sequence as blue cylinders (α-helices), green arrows (β-sheets), and gray lines (coil regions). The numbering below the alignment is relative to human DPF2. Overall sequence conservation at each position is shaded in a color gradient from white (below 40 % similarity) to yellow (60 % similarity) to dark red (100 % identity). Purple dots indicate residues mutated in our biochemical analysis, and grey dots indicate residues for which there was no observed electron density. (b) Surface representation of a model of $DPF2^{PHD}$ in complex with the indicated histones peptides based on the $MOZ/MYST3^{PHD}$ crystal structure (PDB ID 4LLB).[22] The model is colored according to the above multispecies sequence alignment.

Fig. 7. Structural comparison of *H. sapiens* DPF2 to *H. sapiens* MOZ/MYST3. (a) A ribbon representation of $DPF2^{PHD}$ (left, colored as in Fig. 1a), $MOZ/MYST3^{PHD}$•$H3K14^{Ac}$ (middle, $MOZ/MYST3^{PHD}$ colored in gray and $H3K14^{Ac}$ colored in purple, PDB ID 4LLB)[22], and a superposition of the two (right). (b) A surface representation of the $MOZ/MYST3^{PHD}$•$H3K14^{Ac}$ crystal structure (top, PDB ID 4LLB).[22] Boxes show a cartoon representation of the indicated binding sites superposed with our $DPF2^{PHD}$ crystal structure. Residues mutated in our biochemical analysis are shown in stick representation.

Fig. 8. Multi-species sequence alignment of DPF2 homologues. An expanded sequence alignment that shows the full-length protein rather than the tandem PHD finger domain. The secondary structure is indicated above the sequence as cylinders (α-helices), arrows (β-sheets), and lines (coil regions). The numbering below the alignment is relative to human

DPF2. Overall sequence conservation at each position is shaded in a color gradient from white (below 40 % identity) to yellow (60 % identity) to dark red (100 % identity). The secondary structure elements above the tandem PHD finger domain (residues 270 to 391) are as observed in the structure presented here and colored in blue, those above the C2H2 zinc finger domain (residues 206 to 232) are from the previously solved structure (PDB ID 3IUF)[18] and colored in green.

Fig. 9. Biochemical interaction analyses. (a) Sequences and modifications of histone peptides used in interaction studies (b) Representative ITC data for interactions between an unacetylated H3 peptide (left) and an acetylated H3K14^{Ac} peptide (middle) and wildtype DPF2, and H3K14^{Ac} and mutant DPF2 (F275A, R300A, and D346A) (right). Upper parts of each box show raw data and lower parts show integrated heat changes corrected for heat from dilution. (c) Representative ITC data for interactions between an unacetylated H4 peptide (left) an acetylated H4K16^{Ac} (middle) and wildtype DPF2, and H4K16^{Ac} and mutant DPF2 (F275A, R300A, and D346A) (right). Upper parts of each box show raw data and lower parts show integrated heat changes corrected for heat from dilution. (d) Table summarizing ITC interaction analysis. The dissociation constant (K_d), binding enthalpy (ΔH), and entropy (ΔS) were derived by curve fitting using a single-site model.

Fig. 10. Biochemical interaction analyses. (a) Sequences and modifications of peptides used in interaction experiments (b) ITC data for interactions between an acetylated H3K14^{Ac} peptide and the indicated DPF2 variants. (c) ITC data for interactions between an acetylated H4K16^{Ac} peptide and the indicated DPF2 variants. (d) ITC data for interactions

between RUNX1 peptides and DPF2. Upper parts of each box show raw data and lower parts show integrated heat changes corrected for heat from dilution.

Fig. 11. Model of the role for DPF2 in myeloid differentiation. In HSPCs and AML, DPF2 forms a repressive multiprotein complex with RUNX1 dependent upon the RUNX1 methylation state. The interaction between DPF2 and methylated RUNX1 is bridged by a yet to be identified protein. DPF2 recruits the complex to specific chromatin sites by preferentially binding acetylated histone tails. In differentiated myeloid cells, a reduction in PRMT4 expression results in disassembly of the repressive transcriptional complex.

Table 1. Bacterial expression constructs

[a]Construct that was used for crystallization of *hs*DPF2

Table 2. Crystallographic analysis

[a]SSRL, Stanford Synchrotron Radiation Lightsource

[b]Highest resolution shell is shown in parentheses

[c]As determined by Molprobity[33]

REFERENCES

1. Verdin, E. & Ott, M. (2014). 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nat Rev Mol Cell Biol*.
2. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E. & Mann, M. (2014). The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol* **15**, 536-50.
3. Patel, D. J. & Wang, Z. (2013). Readout of epigenetic modifications. *Annu Rev Biochem* **82**, 81-118.
4. Sanchez, R. & Zhou, M. M. (2011). The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36**, 364-72.
5. Gabig, T. G., Mantel, P. L., Rosli, R. & Crean, C. D. (1994). Requiem: a novel zinc finger gene essential for apoptosis in myeloid cells. *J Biol Chem* **269**, 29515-9.
6. Chestkov, A. V., Baka, I. D., Kost, M. V., Georgiev, G. P. & Buchman, V. L. (1996). The d4 gene family in the human genome. *Genomics* **36**, 174-7.
7. Kulikova, D. A., Mertsalov, I. B. & Simonova, O. B. (2013). [D4 family genes in vertebrates: genomic organization and expression]. *Ontogenez* **44**, 3-9.
8. Gabig, T. G., Crean, C. D., Klenk, A., Long, H., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Quincey, D., Parente, F., Lespinasse, F., Carle, G. F., Gaudray, P., Zhang, C. X., Calender, A., Hoeppener, J., Kas, K., Thakker, R. V., Farnebo, F., Teh, B. T., Larsson, C., Piehl, F., Lagercrantz, J., Khodaei, S., Carson, E. & Weber, G. (1998). Expression and chromosomal localization of the Requiem gene. *Mamm Genome* **9**, 660-5.

9. Matsuyama, R., Takada, I., Yokoyama, A., Fujiyma-Nakamura, S., Tsuji, N., Kitagawa, H., Fujiki, R., Kim, M., Kouzu-Fujita, M., Yano, T. & Kato, S. (2010). Double PHD fingers protein DPF2 recognizes acetylated histones and suppresses the function of estrogen-related receptor alpha through histone deacetylase 1. *J Biol Chem* **285**, 18166-76.

10. Kim, M. Y., Park, J., Lee, J. J., Ha, D. H., Kim, J., Kim, C. G., Hwang, J. & Kim, C. G. (2014). Staufen1-mediated mRNA decay induces Requiem mRNA decay through binding of Staufen1 to the Requiem 3'UTR. *Nucleic Acids Res* **42**, 6999-7011.

11. Vu, L. P., Perna, F., Wang, L., Voza, F., Figueroa, M. E., Tempst, P., Erdjument-Bromage, H., Gao, R., Chen, S., Paietta, E., Deblasio, T., Melnick, A., Liu, Y., Zhao, X. & Nimer, S. D. (2013). PRMT4 blocks myeloid differentiation by assembling a methyl-RUNX1-dependent repressor complex. *Cell Rep* **5**, 1625-38.

12. Theodorou, M., Speletas, M., Mamara, A., Papachristopoulou, G., Lazou, V., Scorilas, A. & Katsantoni, E. (2013). Identification of a STAT5 target gene, Dpf3, provides novel insights in chronic lymphocytic leukemia. *PLoS One* **8**, e76155.

13. Loi, S., Haibe-Kains, B., Desmedt, C., Lallemand, F., Tutt, A. M., Gillet, C., Ellis, P., Harris, A., Bergh, J., Foekens, J. A., Klijn, J. G., Larsimont, D., Buyse, M., Bontempi, G., Delorenzi, M., Piccart, M. J. & Sotiriou, C. (2007). Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* **25**, 1239-46.

14. Lo, K. C., Rossi, M. R., LaDuca, J., Hicks, D. G., Turpaz, Y., Hawthorn, L. & Cowell, J. K. (2007). Candidate glioblastoma development gene identification

using concordance between copy number abnormalities and gene expression level changes. *Genes Chromosomes Cancer* **46**, 875-94.

15. Hoyal, C. R., Kammerer, S., Roth, R. B., Reneland, R., Marnellos, G., Kiechle, M., Schwarz-Boeger, U., Griffiths, L. R., Ebner, F., Rehbock, J., Nelson, M. R. & Braun, A. (2005). Genetic polymorphisms in DPF3 associated with risk of breast cancer and lymph node metastases. *J Carcinog* **4**, 13.

16. Choi, Y. P., Kang, S., Hong, S., Xie, X. & Cho, N. H. (2005). Proteomic analysis of progressive factors in uterine cervical cancer. *Proteomics* **5**, 1481-93.

17. Muller, A. M., Duque, J., Shizuru, J. A. & Lubbert, M. (2008). Complementing mutations in core binding factor leukemias: from mouse models to clinical applications. *Oncogene* **27**, 5759-73.

18. Zhang, W., Xu, C., Bian, C., Tempel, W., Crombet, L., MacKenzie, F., Min, J., Liu, Z. & Qi, C. (2011). Crystal structure of the Cys2His2-type zinc finger domain of human DPF2. *Biochem Biophys Res Commun* **413**, 58-61.

19. Pascual, J., Martinez-Yamout, M., Dyson, H. J. & Wright, P. E. (2000). Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *J Mol Biol* **304**, 723-9.

20. Zeng, L., Zhang, Q., Li, S., Plotnikov, A. N., Walsh, M. J. & Zhou, M. M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466**, 258-62.

21. Emekli, U., Schneidman-Duhovny, D., Wolfson, H. J., Nussinov, R. & Haliloglu, T. (2008). HingeProt: automated prediction of hinges in protein structures. *Proteins* **70**, 1219-27.

22. Dreveny, I., Deeves, S. E., Fulton, J., Yue, B., Messmer, M., Bhattacharya, A., Collins, H. M. & Heery, D. M. (2014). The double PHD finger domain of MOZ/MYST3 induces alpha-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification. *Nucleic Acids Res* **42**, 822-35.

23. London, N., Raveh, B., Cohen, E., Fathi, G. & Schueler-Furman, O. (2011). Rosetta FlexPepDock web server--high resolution modeling of peptide-protein interactions. *Nucleic Acids Res* **39**, W249-53.

24. Agafonov, R. V., Wilson, C., Otten, R., Buosi, V. & Kern, D. (2014). Energetic dissection of Gleevec's selectivity toward human tyrosine kinases. *Nat Struct Mol Biol* **21**, 848-53.

25. Seeliger, M. A., Nagar, B., Frank, F., Cao, X., Henderson, M. N. & Kuriyan, J. (2007). c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty. *Structure* **15**, 299-311.

26. Wilson, C., Agafonov, R. V., Hoemberger, M., Kutter, S., Zorba, A., Halpin, J., Buosi, V., Otten, R., Waterman, D., Theobald, D. L. & Kern, D. (2015). Using ancient protein kinases to unravel a modern cancer drug's mechanism. *Science* **347**, 882-886.

27. Tando, T., Ishizaka, A., Watanabe, H., Ito, T., Iida, S., Haraguchi, T., Mizutani, T., Izumi, T., Isobe, T., Akiyama, T., Inoue, J. & Iba, H. (2010). Requiem protein links RelB/p52 and the Brm-type SWI/SNF complex in a noncanonical NF-kappaB pathway. *J Biol Chem* **285**, 21951-60.

28. Middeljans, E., Wan, X., Jansen, P. W., Sharma, V., Stunnenberg, H. G. & Logie, C. (2012). SS18 together with animal-specific factors defines human BAF-type SWI/SNF complexes. *PLoS One* **7**, e33834.

29. Bakshi, R., Hassan, M. Q., Pratap, J., Lian, J. B., Montecino, M. A., van Wijnen, A. J., Stein, J. L., Imbalzano, A. N. & Stein, G. S. (2010). The human SWI/SNF complex associates with RUNX1 to control transcription of hematopoietic target genes. *J Cell Physiol* **225**, 569-76.

30. Kabsch, W. (2010). Xds. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 125-32.

31. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* **66**, 213-21.

32. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* **60**, 2126-32.

33. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-83.

34. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem Sci* **23**, 403-5.

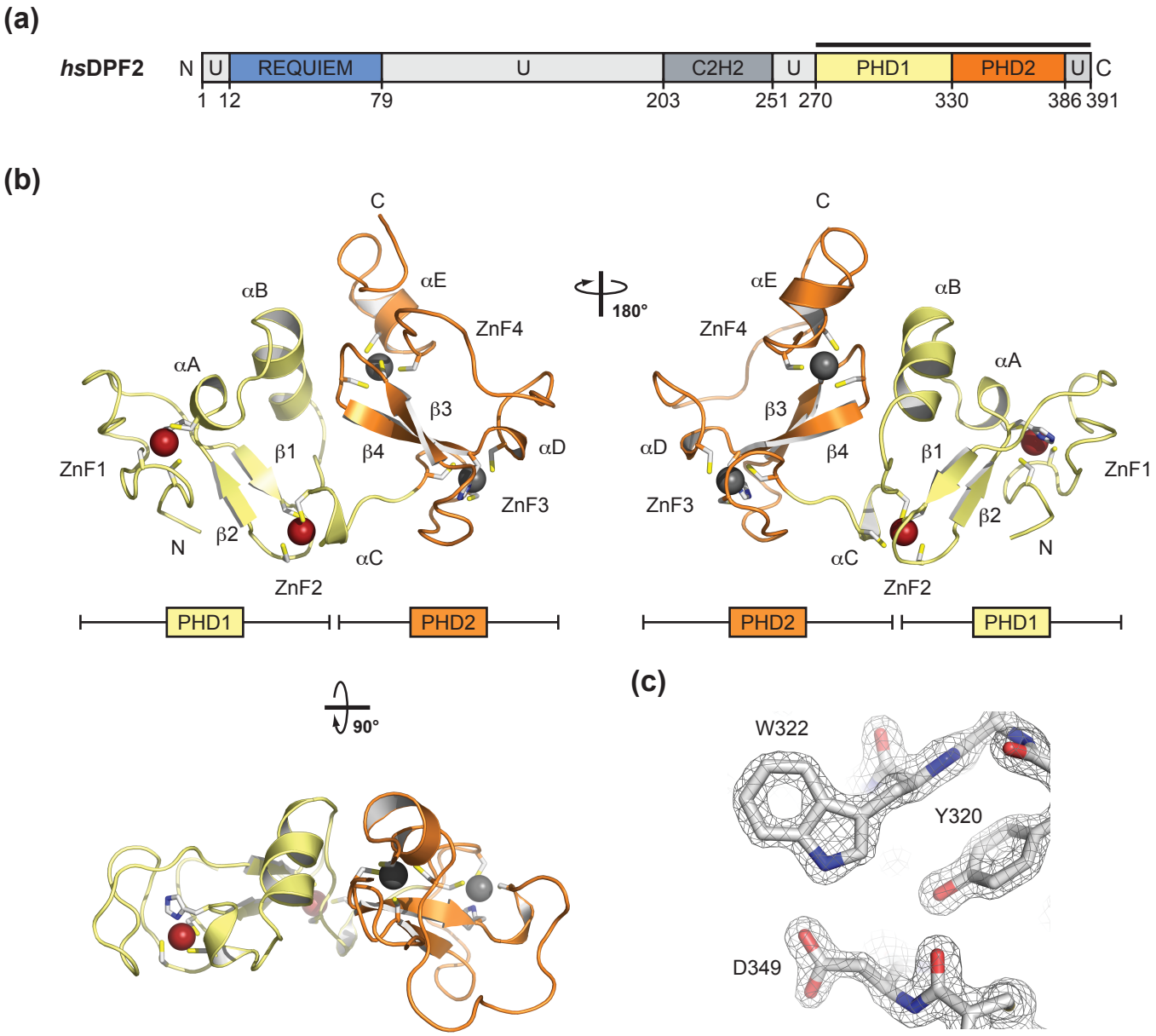
35. Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng* **6**, 37-40.

36. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-41.

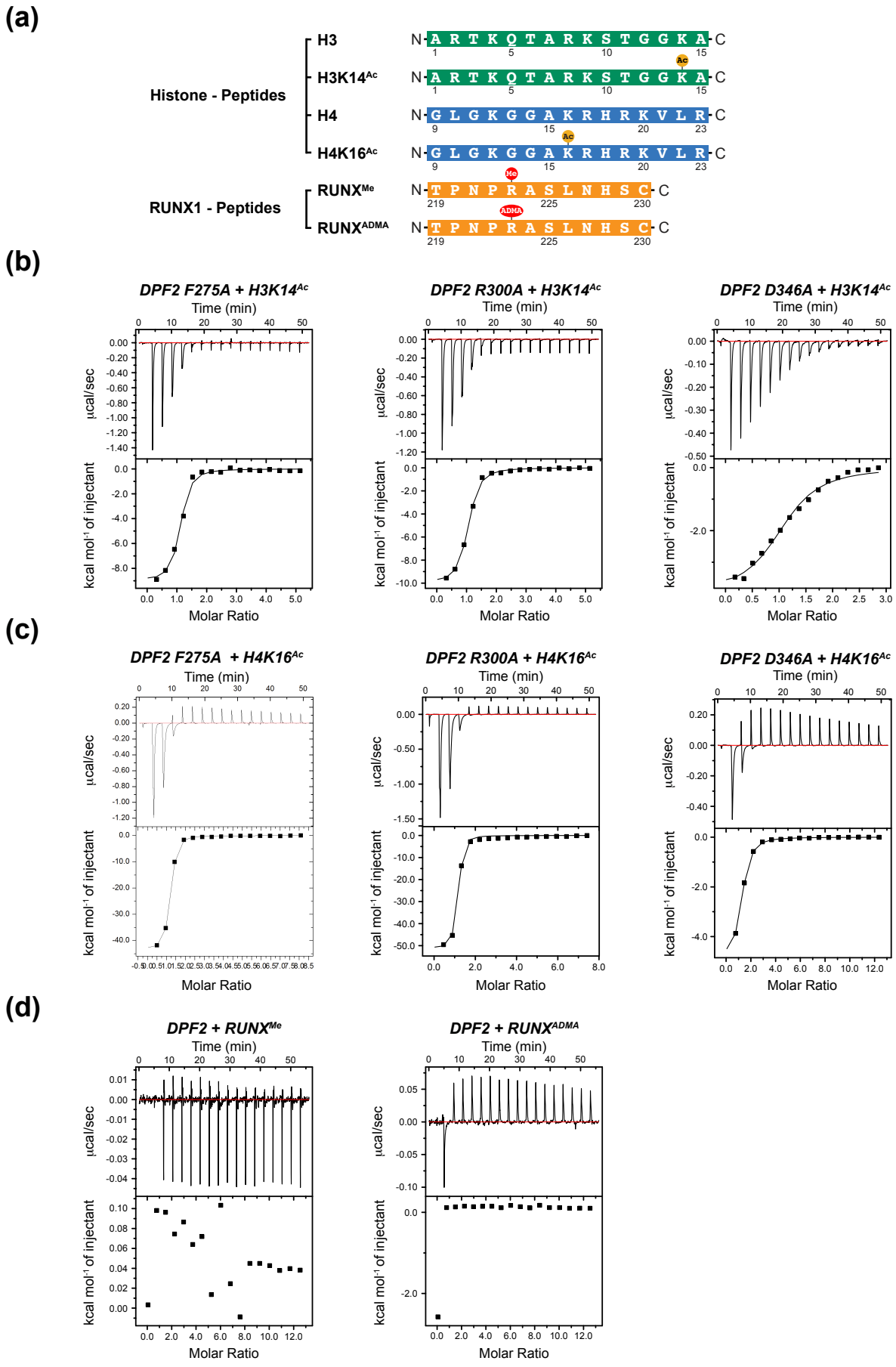
37. Gatchalian, J., Futterer, A., Rothbart, S. B., Tong, Q., Rincon-Arano, H., Sanchez de Diego, A., Groudine, M., Strahl, B. D., Martinez, A. C., van Wely, K. H. & Kutateladze, T. G. (2013). Dido3 PHD modulates cell differentiation and division. *Cell Rep* **4**, 148-58.

38. Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D. & Patel, D. J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**, 91-5.

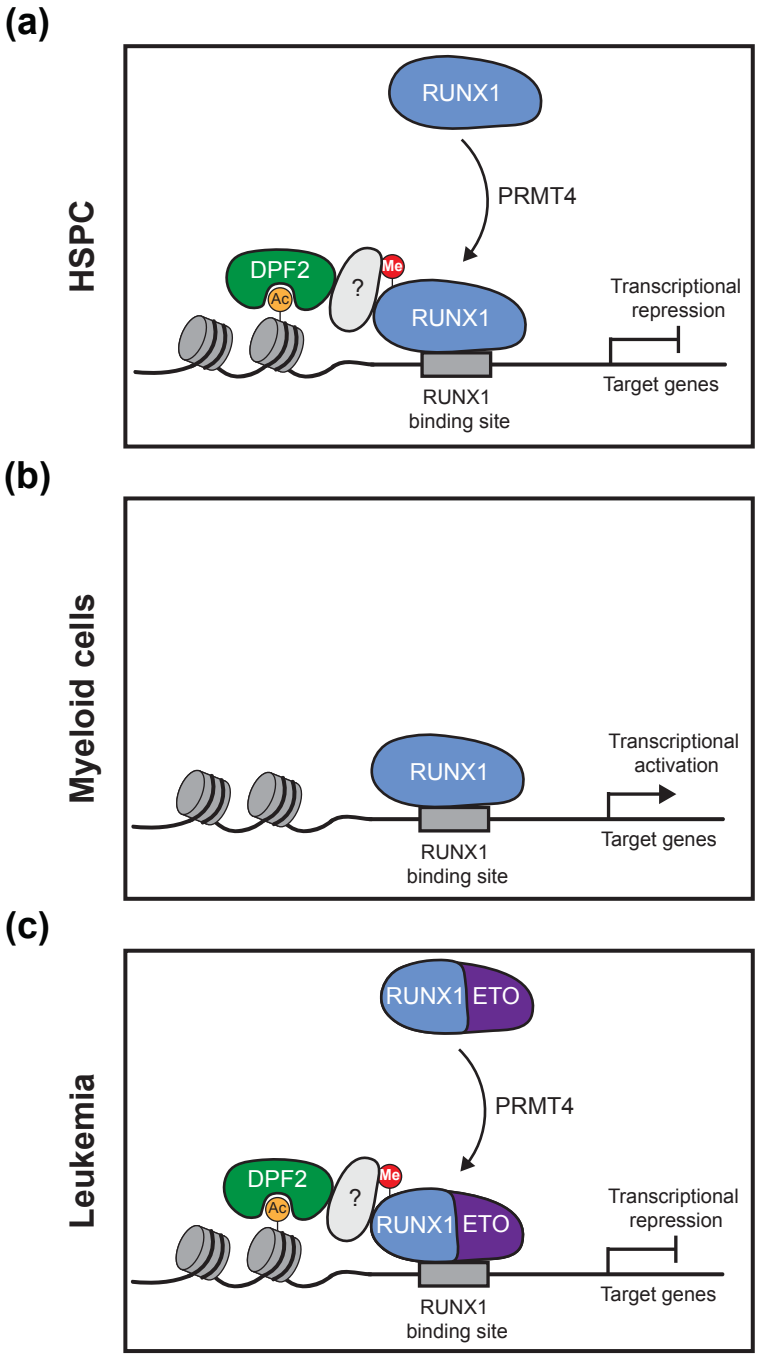
39. Rajakumara, E., Wang, Z., Ma, H., Hu, L., Chen, H., Lin, Y., Guo, R., Wu, F., Li, H., Lan, F., Shi, Y. G., Xu, Y., Patel, D. J. & Shi, Y. (2011). PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Mol Cell* **43**, 275-84.



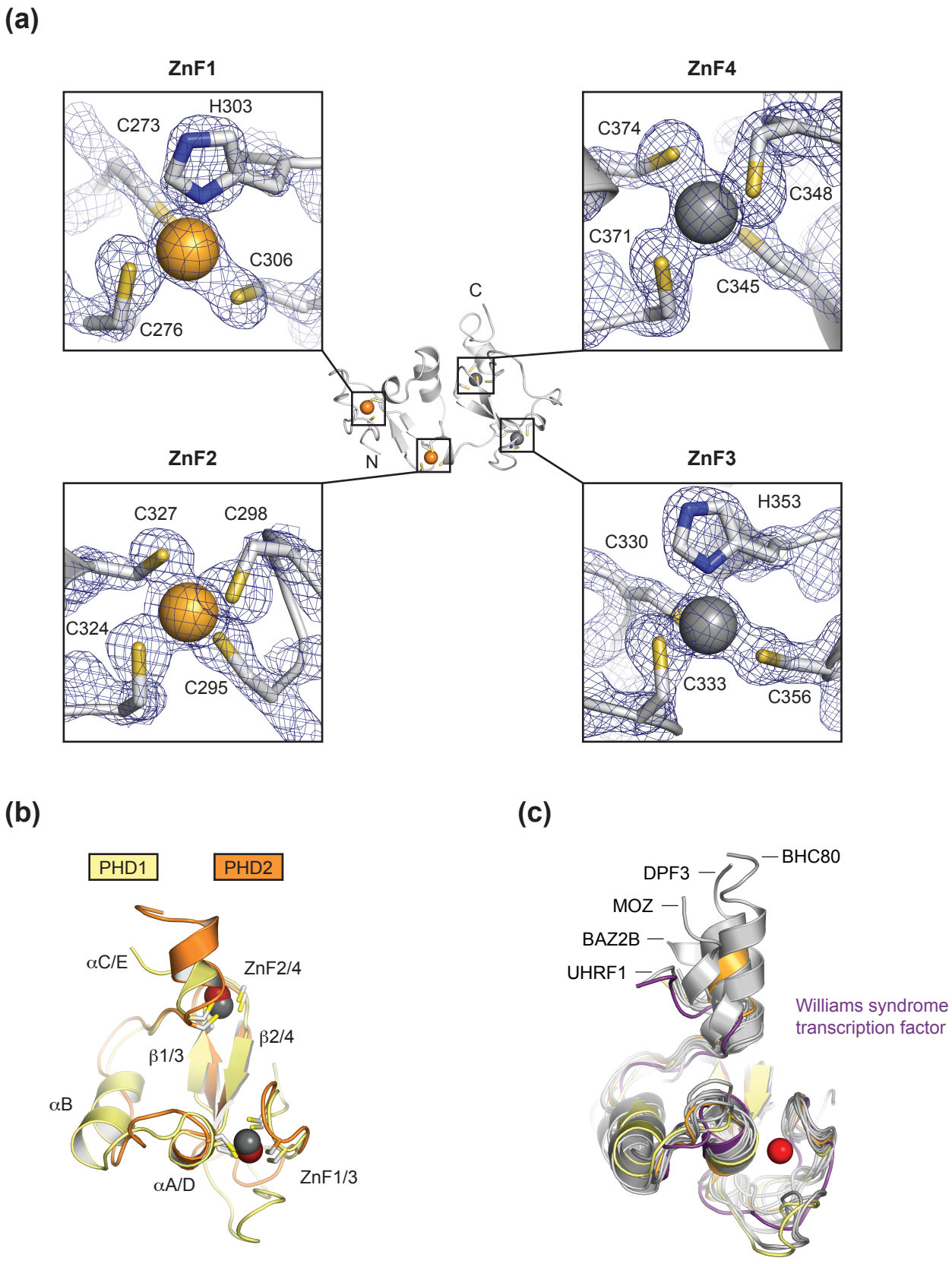
Chapter III, Figure 1



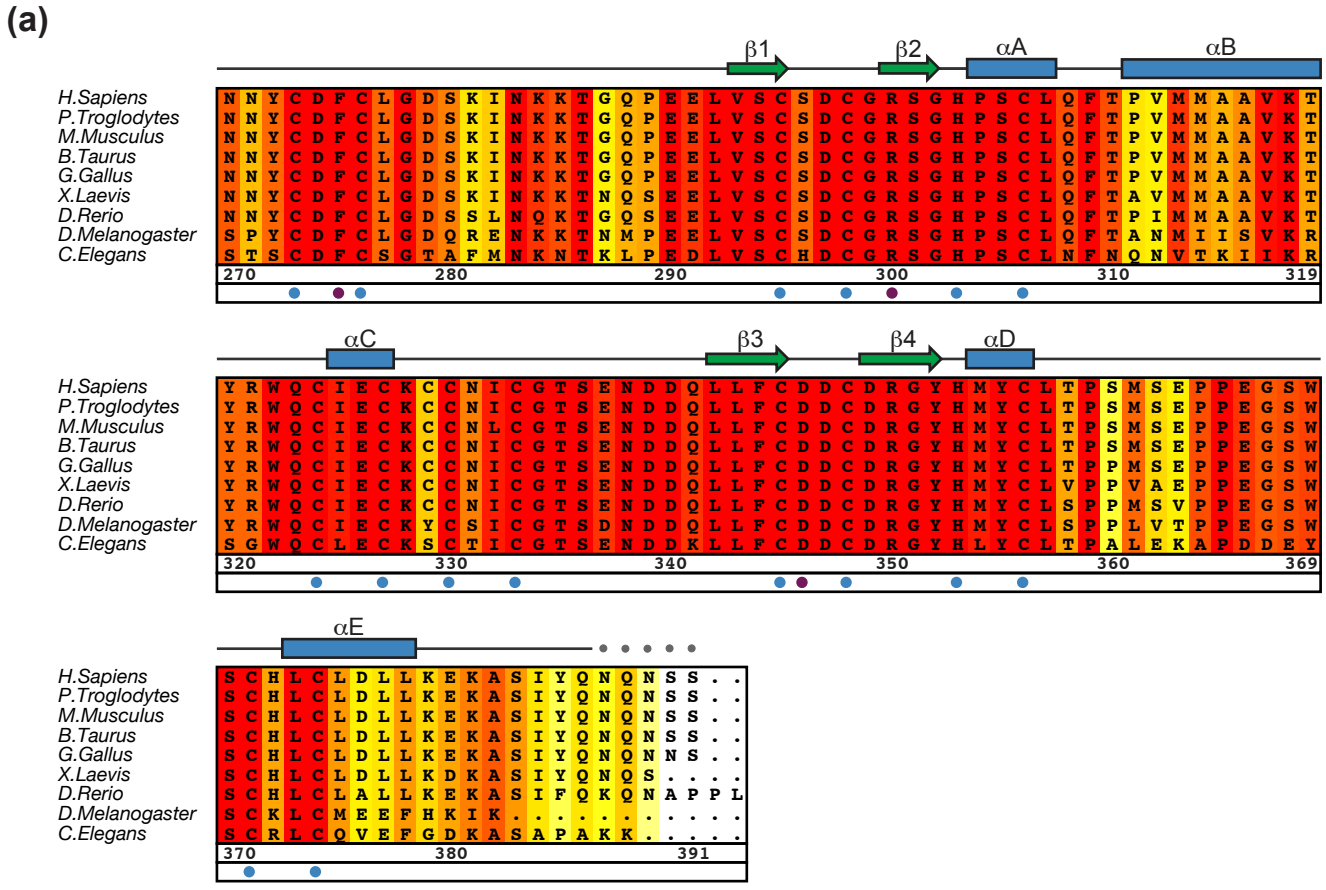
Chapter III, Figure 10

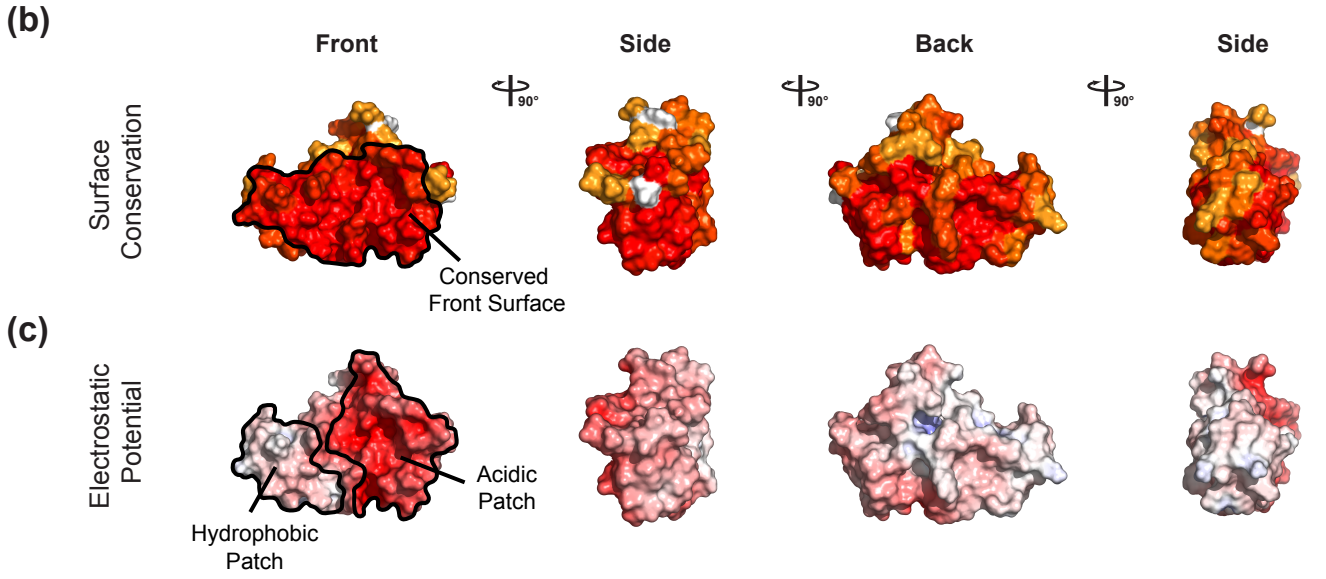


Chapter III, Figure 11



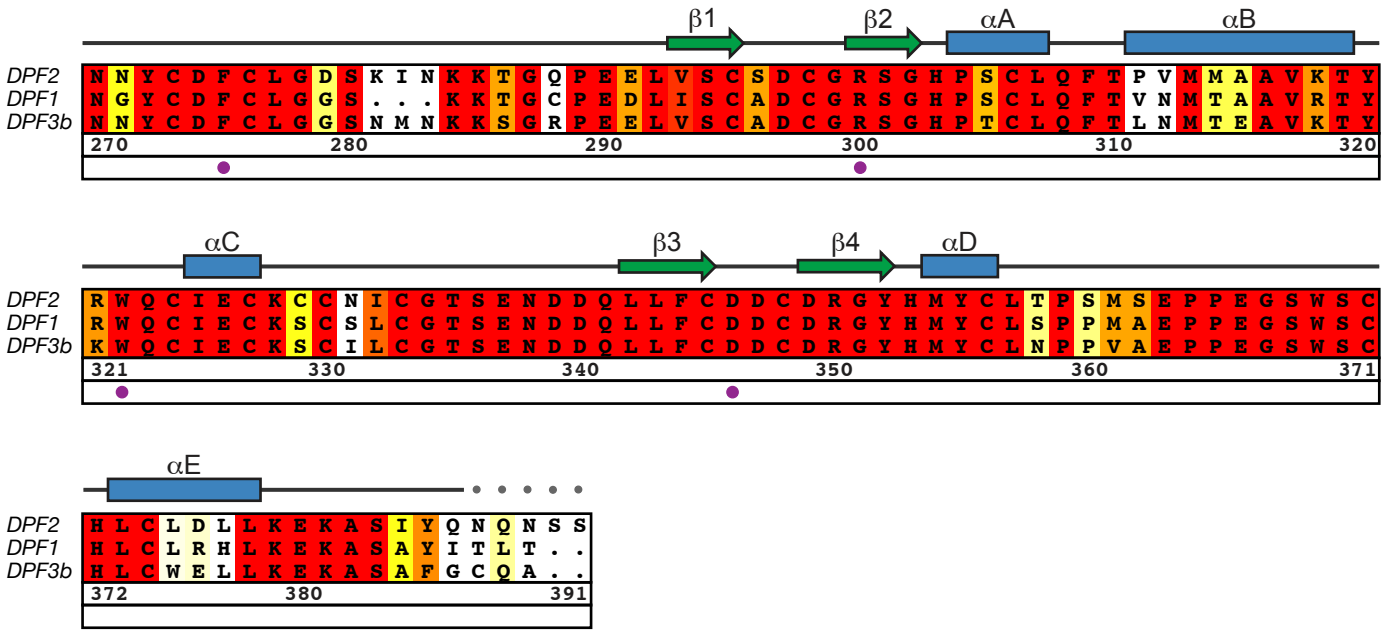
Chapter III, Figure 2



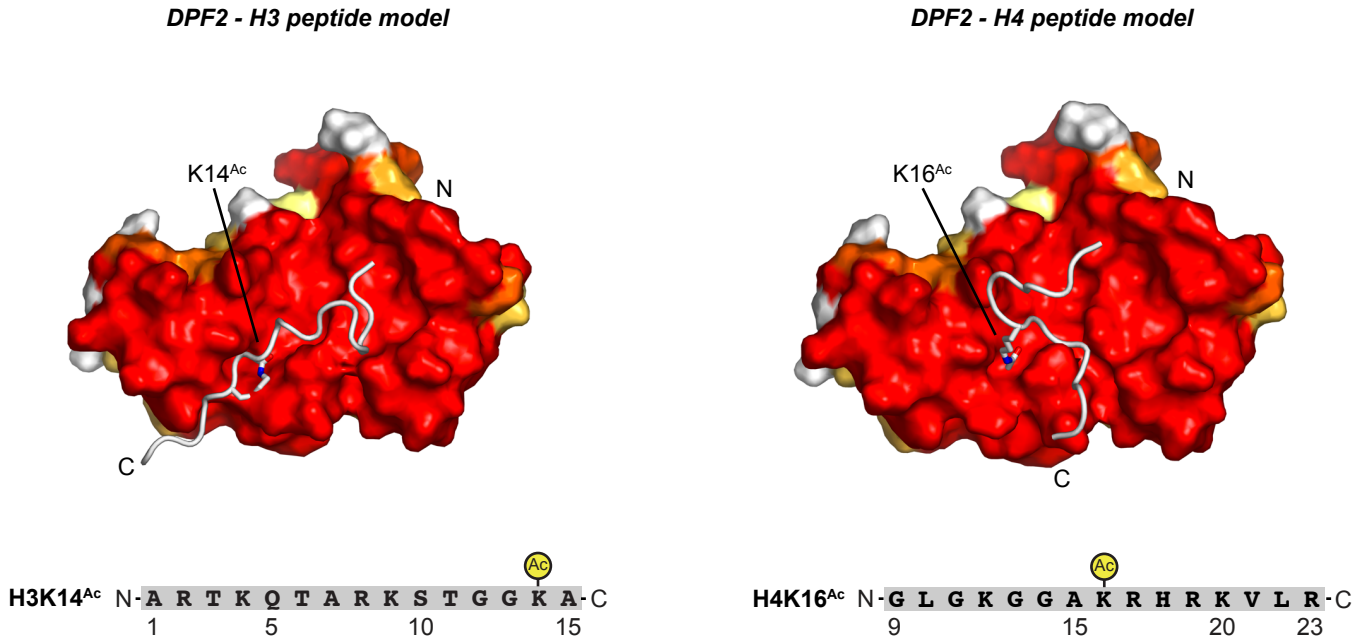


Chapter III, Figure 3

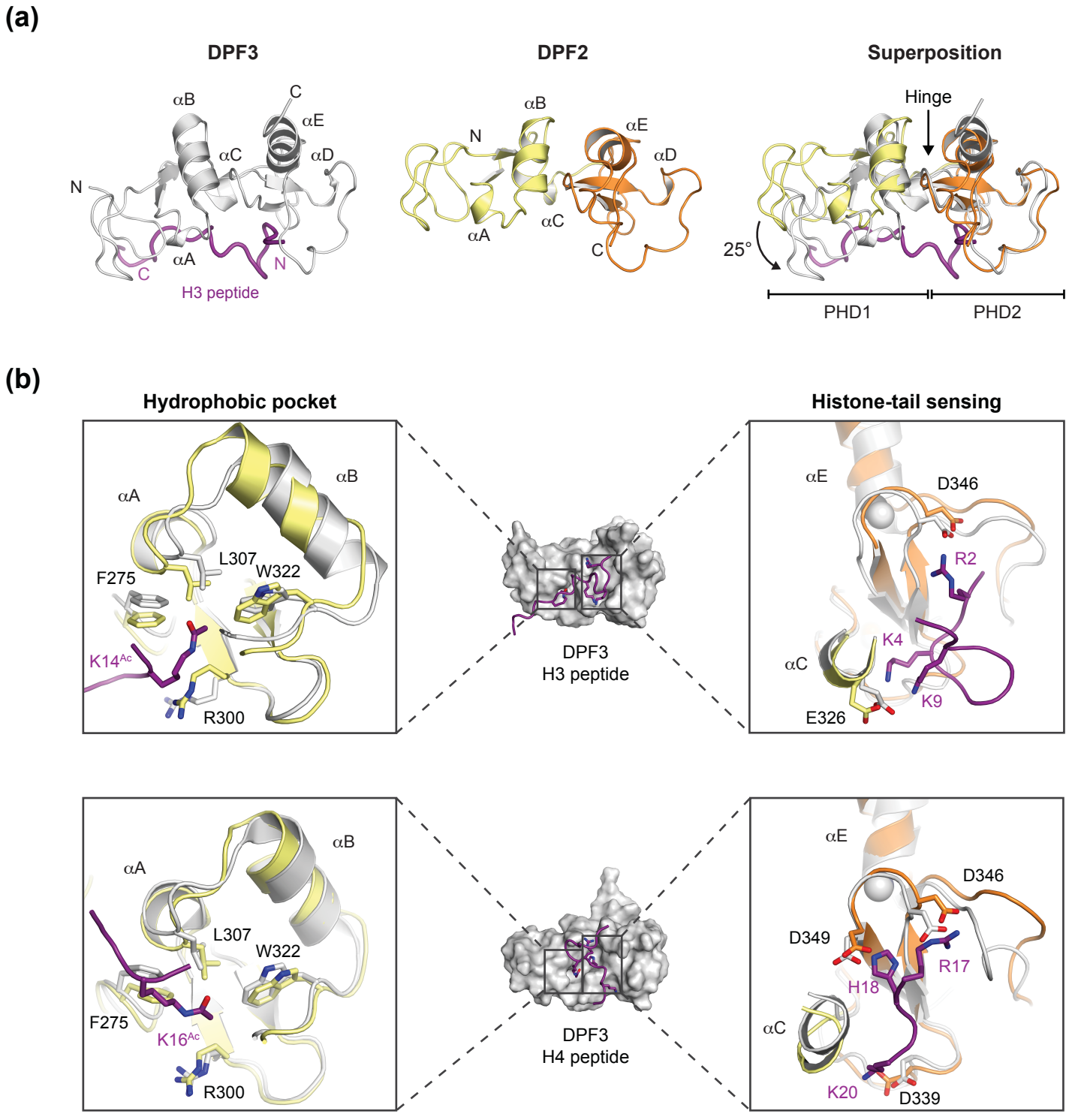
(a)



(b)

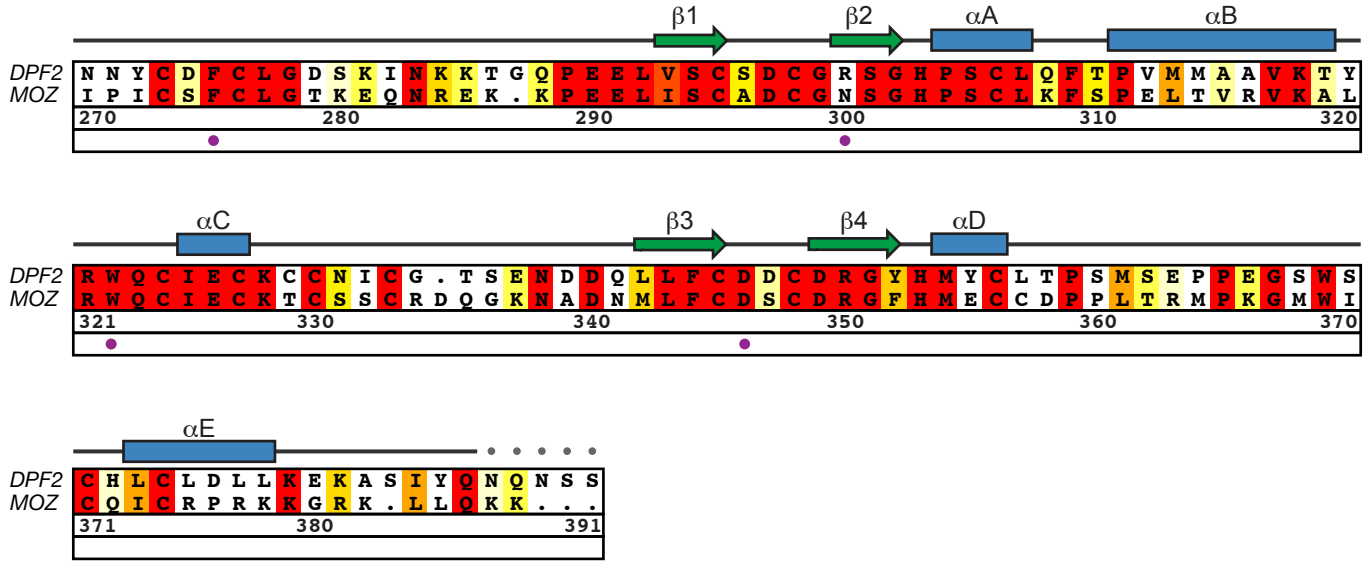


Chapter III, Figure 4

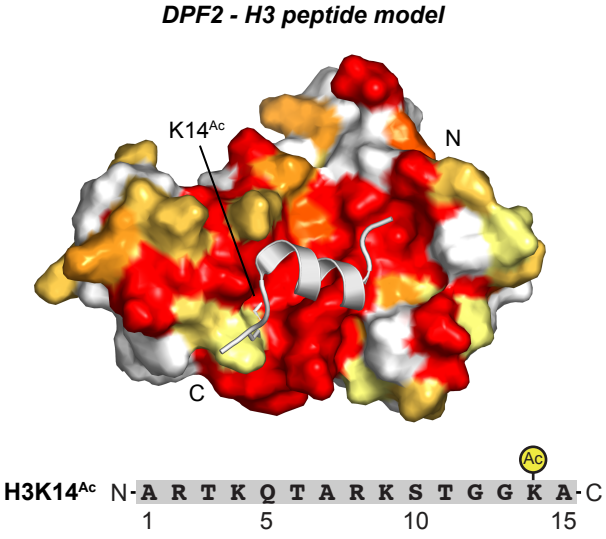


Chapter III, Figure 5

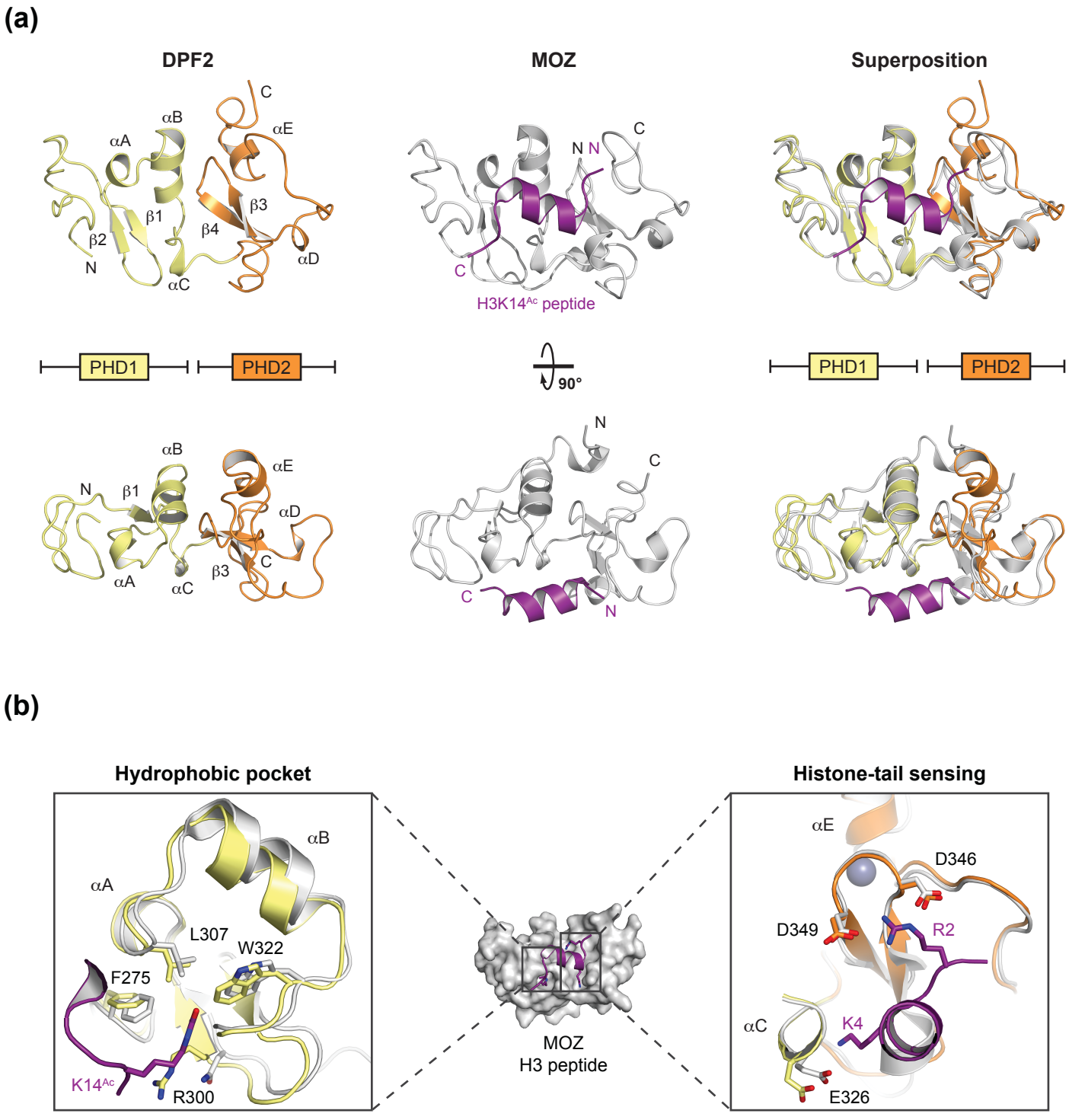
(a)



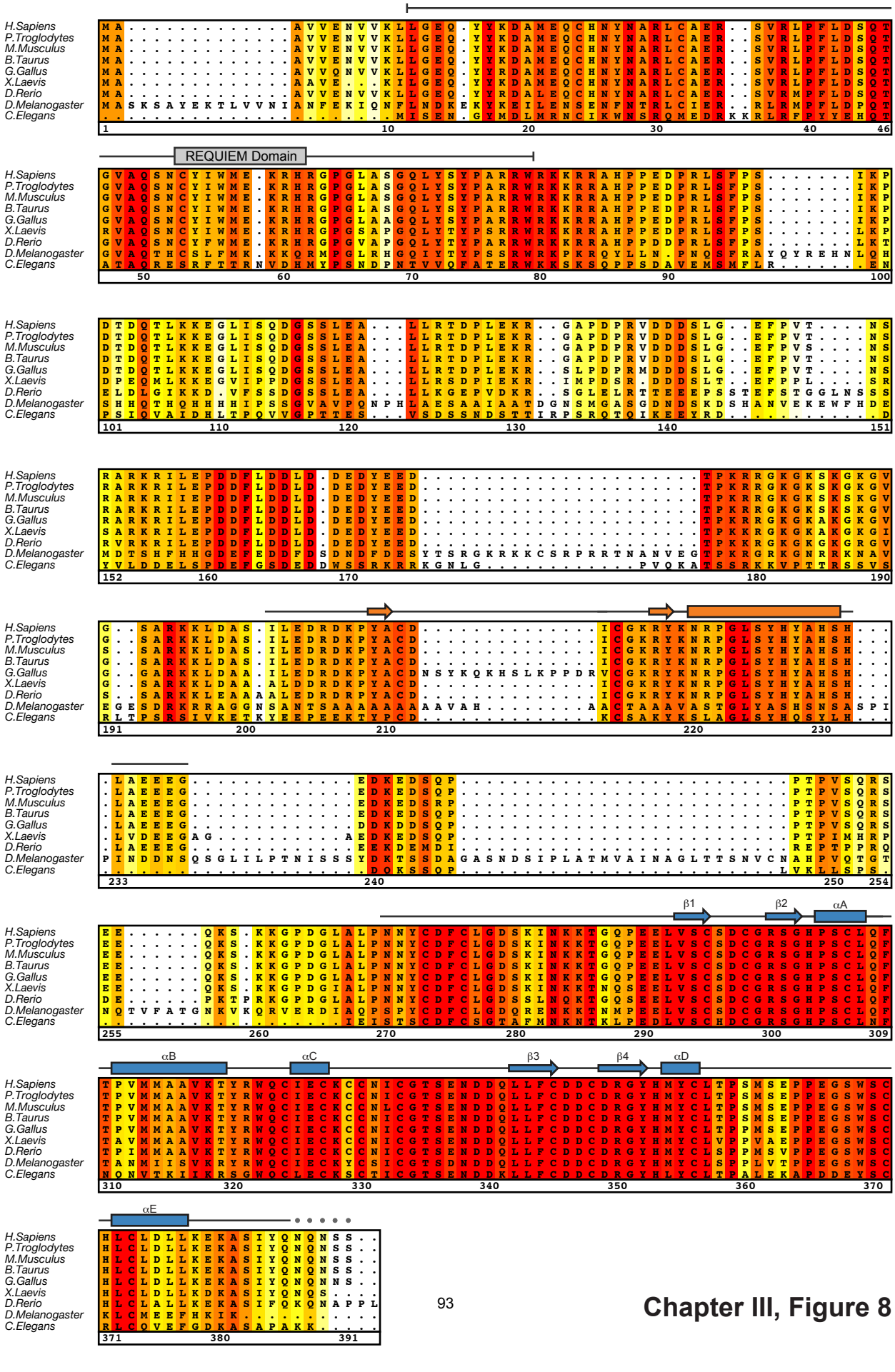
(b)



Chapter III, Figure 6

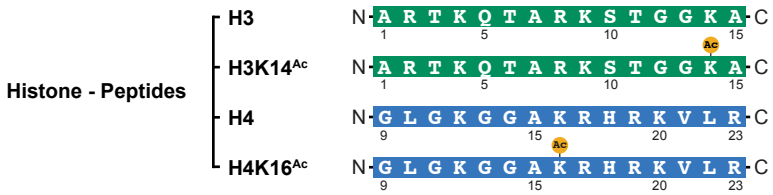


Chapter III, Figure 7

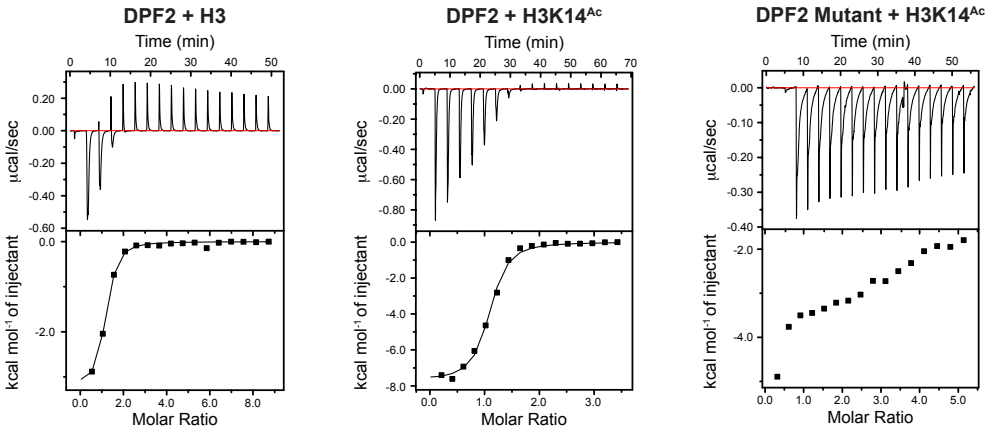


Chapter III, Figure 8

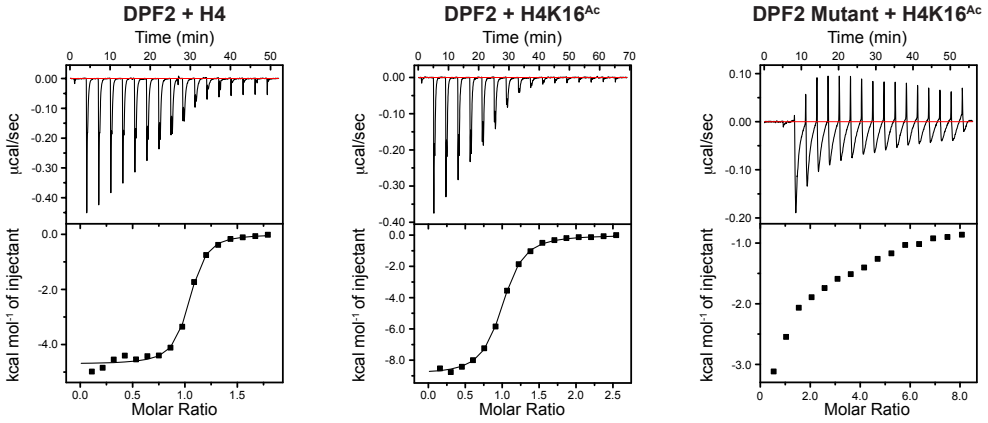
(a)



(b)



(c)



(d)

Protein	Peptide	K_d (µM)	ΔH (kcal/mol)	ΔS (cal/mol/deg)
DPF2	H3	5.38 ± 0.74	- 3.25 ± 0.08	13.20
DPF2	H3K14^{Ac}	1.98 ± 0.29	- 7.71 ± 0.13	0.25
DPF2 F275A	H3K14^{Ac}	1.36 ± 0.30	- 9.02 ± 0.24	-3.40
DPF2 R300A	H3K14^{Ac}	1.75 ± .018	- 10.00 ± 0.14	-7.32
DPF2 D346A	H3K14^{Ac}	11.81 ± 2.16	- 3.99 ± 0.18	9.18
DPF2 F275A, R300A, D346A	H3K14^{Ac}	N/A	N/A	N/A
DPF2	H4	0.42 ± 0.07	- 4.71 ± 0.06	13.40
DPF2	H4K16^{Ac}	0.44 ± 0.04	- 9.35 ± 0.08	-2.28
DPF2 F275A	H4K16^{Ac}	0.19 ± 0.01	- 43.31 ± 0.28	-114
DPF2 R300A	H4K16^{Ac}	0.14 ± 0.03	- 51.28 ± 1.00	-141
DPF2 D346A	H4K16^{Ac}	4.51 ± 0.20	- 5.08 ± 0.06	7.41
DPF2 F275A, R300A, D346A	H4K16^{Ac}	N/A	N/A	N/A
DPF2	RUNX1R223^{Me}	N/A	N/A	N/A
DPF2	RUNX1R223^{ADMA}	N/A	N/A	N/A
DPF3b[a]	H3	2.3 ± 0.7	- 4.4 ± 0.2	11.1
DPF3b[a]	H3K14^{Ac}	0.5 ± 0.1	- 6.6 ± 0.1	6.7
DPF3b[a]	H4	28.8 ± 7.8	-3.3 ± 0.4	9.7
DPF3b[a]	H4K16^{Ac}	46.9 ± 10.0	-4.6 ± 0.6	4.4

Chapter III, Figure 9

Table 1. Bacterial expression constructs

Protein	Residues (Mutations)	Expression Vector	Restriction Sites 5', 3'	N-terminal Overhang
*hs*DPF2[a]	270-391	pGEX-6P-1	BamHI, NotI	GMGS
*hs*DPF2	270-391 F275A	pGEX-6P-1	BamHI, NotI	GMGS
*hs*DPF2	270-391 R300A	pGEX-6P-1	BamHI, NotI	GMGS
*hs*DPF2	270-391 D346A	pGEX-6P-1	BamHI, NotI	GMGS
*hs*DPF2	270-391 F275A R300A D346A	pGEX-6P-1	BamHI, NotI	GMGS

[a]Construct that was used for crystallization of *hs*DPF2

Table 2. Crystallographic analysis

Data collection	
Protein	$DPF2^{PHD}$
Synchrotron	SSRL
Beamline	BL12-2
Space group	I422
Cell parameters	
a, *b*, *c* (Å)	107.3, 107.3, 53.3
α, β, γ (°)	90.0, 90.0, 120.0
	Zn peak
Wavelength (Å)	1.28210
Resolution (Å)	37.9 – 1.45
R_{sym} (%)[b]	5.7 (162.5)
$\langle I \rangle / \langle \sigma I \rangle$[b]	30.8 (3.3)
Completeness (%)[b]	99.9 (100.0)
No. observations	675,632
No. unique reflections	27,349 (3,932)
Redundancy	24.7 (25.5)
Refinement	
Resolution (Å)	37.9 – 1.6
No. reflections total	20,624
No. reflections test set	1,814 (8.80%)
R_{work} / R_{free} (%)	15.2 / 18.9
No. atoms	1,106
Protein	1,003
Ligand	16
Water	87
B-factors	
Protein	35.1
Ligand	56.3
Water	49.3
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.2
Ramachandran plot[c]	
Favored (%)	98.4
Additionally allowed (%)	1.6
Outliers (%)	0.0
MolProbity score	2.08

[a] SSRL, Stanford Synchrotron Radiation Lightsource
[b] Highest resolution shell is shown in parentheses
[c] As determined by MolProbity[62]

Chapter IV

Structural and Functional Analysis of Human SIRT1

This chapter, first published in *Journal of Molecular Biology* in 2014, was written by Andrew M. Davenport[1], Ferdinand M. Huber[1], and André Hoelz[1].

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

ABSTRACT

SIRT1 is a NAD^+-dependent deacetylase that plays important roles in many cellular processes. SIRT1 activity is uniquely controlled by a C-terminal regulatory segment (CTR). Here we present crystal structures of the catalytic domain of human SIRT1 in complex with the CTR in an open *apo* form and a closed conformation in complex with a cofactor and a pseudo-substrate peptide. The catalytic domain adopts the canonical sirtuin fold. The CTR forms a β-hairpin structure that complements the β-sheet of the NAD^+-binding domain, covering an essentially invariant, hydrophobic surface. The *apo* form adopts a distinct open conformation, in which the smaller subdomain of SIRT1 undergoes a rotation with respect to the larger NAD^+-binding subdomain. A biochemical analysis identifies key residues in the active site, an inhibitory role for the CTR, and distinct structural features of the CTR that mediate binding and inhibition of the SIRT1 catalytic domain.

INTRODUCTION

Sirtuins are a large family of protein-modifying enzymes highly conserved throughout bacteria, archaea, and eukaryotes. The founding member of the sirtuin family, silent information regulator 2 (Sir2), was first identified through a genetic screen in *Saccharomyces cerevisiae* as necessary for silencing of the mating-type information locus, *HM*.[1] Three additional sirtuin genes were subsequently discovered in *S. cerevisiae*,[2,3] the products of which collectively act along with Sir2 to silence transcription at telomeres and rDNA clusters as well as *HM* loci.[4-7] Later work showed Sir2 and its homologs to function primarily as nicotinamide adenine dinucleotide (NAD^+)-dependent deacetylases,[8-10] with specific family members reported to possess mono-ADP ribosyl transferase,[11-16] demalonylase, or desuccinylase activity.[17]

In the sirtuin deacetylation reaction, the substrate acetyl group is transferred onto the ribose moiety of NAD^+, generating nicotinamide (NAM) and 2'-*O*-acetyl-adenosine disphosphoribose (2-*O*'-AADPR).[18] While the details of this reaction are still a matter of debate, a proposed mechanism that proceeds through an adenosine diphosphoribose (ADPR)-peptidyl-imidate intermediate is consistent with available biochemical data.[19] Crystal structures of a number of sirtuin family members have provided insight into substrate and cofactor binding as well as catalysis.[11,20-40]

The mammalian genome encodes seven different sirtuins (SIRT1-7), with varying subcellular localizations and chemical activities.[17,41] The direct human homolog of *S. cerevisiae* Sir2 is SIRT1. SIRT1 deacetylates a wide range of substrates, including p53, NF-κB, FOXO transcription factors, and PGC-1α, with roles in cellular processes ranging

from energy metabolism to cell survival.[42] As such, SIRT1 is implicated in a wide range of human diseases and is a prominent therapeutic target.

Despite progress over the last decade, relatively little is known about the regulatory mechanism of SIRT1. Like all sirtuins, SIRT1 is strongly inhibited by NAM through a base-exchange mechanism that reforms cleaved NAD^+.[43] Active Regulator of SIRT1 (AROS) and Deleted in Breast Cancer 1 (DBC1) have been identified as endogenous proteins that promote or inhibit SIRT1 activity, respectively.[44-46] Additionally, various regions in the long and mostly unstructured N- and C-termini that flank the SIRT1 catalytic domain have been shown to affect SIRT1 deacetylation activity.[47,48]

To shed light on the regulation of human SIRT1 activity, we have determined the crystal structure of SIRT1 in complex with its C-terminal regulatory segment (CTR) in its *apo* form and in a quaternary complex with the NAD^+ hydrolysis product ADPR and a substrate-mimicking peptide at 2.65 Å and 1.85 Å resolution, respectively. The structures reveal that the CTR binds at the lower edge of the larger NAD^+-binding domain, complementing the central parallel β-sheet of its Rossmann fold. The substrate-bound closed state completely encapsulates the cofactor and forms a binding site with a hydrophobic tunnel for the substrate residue that leads to a shielded active site in the interior of the enzyme. The overall conformation and mode of substrate binding confirms previous predictions of how human SIRT1 interacts with peptide substrates. In the absence of bound cofactor and substrate, the smaller domain of the SIRT1 catalytic domain undergoes a striking ~25° rotation that is accompanied by an ~15 Å shift of the residues of the domain, generating a wide open interdomain groove, while the larger domain and CTR interface remain mostly unchanged. A mutational analysis identifies key residues for

enzymatic activity of SIRT1 and supports the previously proposed imidate reaction mechanism. Further biochemical experiments establish an inhibitory role for the CTR and define corresponding binding and inhibitory regions. Our results provide a promising avenue for the development of novel SIRT1 activators that take advantage of the distinct features of the catalytic domain-CTR interface.

RESULTS

Reconstitution of active SIRT1 and structure determination

Our attempts to express various fragments of the catalytic domain of *Homo sapiens* SIRT1 in bacteria yielded protein prone to aggregation. Based on previous findings that a C-terminal region is required for SIRT1 activity,[47,48] we generated a series of expression constructs for various C-terminal fragments that were tested for their ability to interact with the catalytic domain. We identified residues 234 to 510 and 641 to 665 of the catalytic domain (CAT) and the C-terminal regulatory segment (CTR), respectively, which formed a heterodimeric complex as determined by size exclusion chromatography (Fig. 1a, b). Coexpression of the two SIRT1 fragments greatly improved the solubility, stability, and behavior of the catalytic domain in solution (Tables 1 and 2). An analysis by size exclusion chromatography coupled to multiangle light scattering (SEC-MALS) revealed that the heterodimer is monomeric in solution with a measured molecular mass of 34.8 kDa (theoretical 35.1 kDa) (Fig. 1c). We refer to this complex as the $SIRT1^{CAT}$•CTR heterodimer in the following text.

Crystals of the *apo* form of $SIRT1^{CAT}$•CTR appeared in the tetragonal space group $P4_32_12$, contained four copies of the heterodimer in the asymmetric unit, and diffracted to 2.65 Å resolution. We also obtained crystals of the $SIRT1^{CAT}$•CTR heterodimer in the presence of NAD^+ in the trigonal space group $P3_221$ that diffracted to 1.85 Å and contained one copy of the complex in the asymmetric unit. Both structures were solved by single-wavelength anomalous dispersion (SAD) using anomalous X-ray diffraction data obtained by taking advantage of the endogenous bound Zn^{2+} ions. The final models essentially contain all residues of both protein fragments in the substrate- and cofactor-bound states

and possess excellent stereochemical parameters. The models were refined to R_{work} and R_{free} values of 23.0 % and 26.5 % (*apo* form) and 16.8 % and 18.9 % (quaternary complex), respectively. For details of the data collection and refinement statistics, see Table 3.

Structural overview

The SIRT1 catalytic domain possesses the canonical sirtuin fold as first described in the archaeal Sir2•Af1•NAD^+ complex structure,[25] composed of a larger NAD^+-binding subdomain with a Rossmann fold and a smaller subdomain composed of a helical module and a Zn^{2+}-binding module, which are both insertions in the NAD^+-binding domain.

The 25-residue CTR can be divided into a larger N-terminal region, followed by a shorter C-terminal extension, which each make distinct interactions with the catalytic domain. The N-terminal region forms a β-hairpin that complements the central parallel β-sheet of the Rossmann fold domain with two additional strands. This interface between the catalytic domain and the CTR is large, hydrophobic, and highly conserved in evolution. The C-terminal CTR extension forms additional contacts with the helical module that are only observed in one of the two crystallized states.

Despite the fact that we obtained crystals of $SIRT1^{CAT}$•CTR in the presence of NAD^+, the heterodimer is bound to adenosine diphosphate ribose (ADPR), likely a product of non-enzymatic NAD^+ hydrolysis or disorder of the nicotinamide moiety. Surprisingly, the heterodimer engages the extended tail region of the catalytic domain of an adjacent molecule in the crystal lattice, yielding a quaternary substrate-mimicking complex. When compared to the substrate- and cofactor-bound state of the heterodimer, the *apo* form

adopts a wide-open conformation, in which the smaller subdomain undergoes a striking ~25° rotation, leading to the disruption of all interactions between the large and small subdomains and the formation of an ~15 Å wide interdomain cleft. As such, we refer to these two states as "open" and "closed" throughout the text. This conformational change leads to the exposure of the hydrophobic interior and the cofactor-binding site of the enzyme in the open state, and to the disruption of the substrate-binding groove.

The structure of SIRT1 catalytic domain

The SIRT1 catalytic core is composed of 277 residues and is divided into two subdomains, a larger NAD^+-binding domain, which adopts a Rossmann fold, and a smaller domain, which is generated by two insertions in the NAD^+-binding domain, a helical module (residues 269 to 324) and a Zn^{2+}-binding module (residues 362 to 419) (Figs. 2 and 3). The NAD^+-binding domain is composed of a central six-stranded parallel β-sheet, comprising strands β1-3 and β7-9, and eight α-helices, αA, αB, αG, αH, and αJ-M that pack against the β-sheet core of the domain (Figs. 2 and 3). The helical module is composed of four α-helices, αC-F, and the Zn^{2+}-binding module is composed of three β-strands, β4-6, and a single α-helix, αI. The Zn^{2+} ion is tetrahedrally coordinated by four invariant cysteine residues, Cys371, Cys374, Cys395, and Cys398. An extensive hydrophobic interface holds the helical and Zn^{2+}-binding modules together, generating the tightly associated small subdomain of the catalytic core. In total, 40 residues, primarily of helices αD, αE, αF, and the αI- β6 region, are involved in the interaction between the two modules, burying ~1,800 $Å^2$ of surface area. The relative orientation and the interaction between the large and small

domains are dramatically different in the open and closed states and are described in detail below.

Interaction between SIRT1^{CAT} and the CTR

The larger N-terminal region of the CTR (residues 641 to 653) forms a β-hairpin that binds to the lower edge of the NAD^+-binding domain of SIRT1^{CAT}, complementing the central six-stranded parallel β-sheet of the Rossmann fold with two additional β-strands, β10 and β11 (Fig. 4). The β-hairpin covers an essentially invariant hydrophobic patch, which is generated by residues of helices αA and αL, β-strand β9, and the β8-β9 connector, forming an entirely hydrophobic interface that is composed of 17 and 11 residues of SIRT1^{CAT} and CTR, respectively. In total, the two proteins bury ~1,150 $Å^2$ of surface area. The CTR-binding site on the Rossmann fold is located close to the binding site for the adenine ring of NAD^+ (termed the A pocket) and the CTR directly interacts with Arg466 of the β8-β9 loop, which forms extensive van der Waals contacts with one face of the adenine ring (Fig. 4). CTR binding does not alter the conformation of the catalytic domain, as evidenced by a root-mean-square deviation (RMSD) of ~0.6 Å (276 Ca atoms) between the closed state and a recently determined structure of the catalytic domain in the absence of the CTR in complex with a small molecule inhibitor (Fig. 5).[39] In the closed state the C-terminal extension of the CTR (residues 654 to 660) undergoes a conformational change and forms additional interactions with the helical module (see below).

Substrate interaction

To shed light on the conformational changes that SIRT1 undergoes upon cofactor binding, we also crystallized SIRT1^{CAT}•CTR in the presence of NAD^+. While we were unable to obtain crystals of SIRT1^{CAT}•CTR in the presence of an acetylated substrate peptide, the structure surprisingly revealed that SIRT1^{CAT}•CTR engages the extended seven-residue tail segment (residues 504 to 510) of an adjacent heterodimer in the crystal (Figs. 2 and 6). The sequence of the tail segment, PVK**L**SEI, resembles the sequence and chemical nature of the SIRT1 substrate region in p53, HK**K**$^{\mathbf{Ac}}$LMF, but possesses a hydrophobic leucine residue in place of the substrate acetyl-lysine residue that is inserted into the hydrophobic tunnel which leads to the internal active site of the enzyme (Fig. 7). The tail region binds SIRT1^{CAT}•CTR by forming anti-parallel β-sheet-like interactions on one side with the β7-αK connector (residues 444 to 446) of the Rossmann fold domain and on the other side with the Phe-Gly-Glu (FGE) motif (residues 414 to 416) of the Zn^{2+}-binding module (Fig. 7a, b). While substrate binding is primarily achieved by backbone hydrogen bonding, the residues of both regions are evolutionarily highly conserved, suggesting that different substrates interact with SIRT1 in a similar manner. Overall, the *trans* binding of the pseudo substrate tail region to SIRT1^{CAT}•CTR faithfully mimics the "β-staple" interaction that was previously observed in the structure of the non-physiological ternary complex between the thermophilic bacterium *Thermotoga maritima Tm*Sir2, ADPR, and a six-residue p53 peptide (Fig. 7).[29] Confirmation of these results will require additional structural analysis of a SIRT1 substrate interaction.

While we crystallized SIRT1^{CAT}•CTR in the presence of NAD^+, the electron density unambiguously establishes that only the NAD^+ hydrolysis product ADPR, which

lacks nicotinamide, is bound (Fig. 7c). The presence of ADPR may be the result of slow non-enzymatic hydrolysis of NAD^+, or disorder of the nicotinamide moiety.[49,50] The ADPR molecule is well ordered and makes extensive hydrogen bond and van der Waals interactions with the Rossmann fold domain, essentially identical to those seen in other previously determined sirtuin structures.[20,21,27,29,30,39,51]

In order to determine whether the *trans* interaction between the tail region of one $SIRT1^{CAT}$•CTR with the substrate-binding site of another $SIRT1^{CAT}$•CTR may be physiologically relevant, we tested whether $SIRT1^{CAT}$•CTR is capable of forming dimers or other higher order oligomers in solution in the presence of NAD^+. A SEC-MALS analysis revealed that $SIRT1^{CAT}$•CTR remains monomeric in the presence of 5 mM NAD^+ with no higher order oligomeric states detectable at concentrations as high as 20 mg/ml (Fig. 8). Further studies are required to address whether the tail region is important for oligomerization and regulation of SIRT1 *in vivo*, especially in the context of the full-length protein and oligomeric substrates.

Conformational changes of SIRT1 upon cofactor and substrate binding

We have determined the structure of human SIRT1 in two different states, the *apo* form of $SIRT1^{CAT}$•CTR (open state), and a quaternary complex of $SIRT1^{CAT}$•CTR with ADPR and a pseudo-substrate peptide (closed state) (Fig. 2). The four copies of the *apo* form in the asymmetric unit make different crystal packing interactions and align with a maximum RMSD of ~0.5 $Å^2$ over 286 Cα atoms, suggesting that SIRT1 possesses a distinct open conformation that is independent of crystal packing interactions (Fig. 6b). Moreover, the tail region that engages an adjacent $SIRT1^{CAT}$•CTR in the closed state is disordered in all

four copies of the open state, further demonstrating that it is not a structural element of the catalytic core.

A comparison between the open and closed states reveals the conformational changes that SIRT1 undergoes upon substrate and cofactor binding. The larger NAD^+-binding domain, together with the N-terminal region of the CTR, does not undergo any major structural changes, as illustrated by an RMSD of ~0.5 $Å^2$ over 170 Cα atoms (Fig. 9). In the closed state, the smaller domain forms an extensive interface with the larger domain, resulting in the complete closure of the interdomain groove and the formation of a hydrophobic tunnel that engages the substrate residue (Fig. 9d). In the absence of cofactor and substrate, SIRT1 adopts an open state with a large interdomain groove, exposing the invariant hydrophobic interior and the NAD^+-binding site (Fig. 9e and 9f). Overall, the smaller domain rotates by ~25° with a maximum displacement of the domain atoms of ~15 Å. Altogether, the two modules that generate the small subdomain rotate essentially as a rigid body with only minor changes to the backbone and side-chain conformations as illustrated by an RMSD of ~1.3 $Å^2$ over 114 Cα atoms. Notable exceptions are the αB-αC interdomain connector and helix αC, which undergo an additional conformational change to form a lid over the cofactor, leading to its encapsulation, and the FGE motif of the β6-αJ connector which forms the upper β-strand of the substrate-binding β-staple. The lower β-strand of the β-staple that is formed by the β7-αK connector of the Rossmann fold only undergoes minor conformational changes. Whereas the N-terminal region of the CTR is static, the C-terminal extension adopts a distinct conformation in the closed state that allows for the formation of a salt bridge between Glu656 and Arg276 of the helical module.

Taken together, SIRT1 undergoes dramatic conformational changes upon substrate and cofactor binding. The open state observed here for SIRT1 is substantially more open than previously determined *apo* states of other sirtuins, suggesting that each member of the sirtuin family adopts a unique open state, while the closed states are essentially identical.

Mutational analysis

To determine the importance of various residues for the deacetylase activity of SIRT1, we generated mutants based on the structures presented here, a multi-species sequence alignment (Fig. 3), and previous biochemical and structural work.[18] We mutated seven residues in the SIRT1 active site that are evolutionarily invariant in bacteria, archaea, and eukaryotes, and which have been shown in other sirtuins to either capture the nicotinamide released from the NAD^+ cofactor (Ser265, Asn346, Ile347, and Asp348), bind and orient the NAD^+ cofactor (Gln345), mediate the interaction with the substrate peptide (Phe414), or are catalytically required to abstract a proton from the activated NAD^+ (His363) (Fig. 10a). To clarify the role of the CTR in SIRT1 regulation, we generated an additional five mutants in the $SIRT1^{CAT}$-CTR interface. These included two residues of the catalytic domain which contact both the CTR and the adenine ring of ADPR (Arg466 and Asp481), two residues that form a salt bridge between the catalytic domain and the CTR in the closed conformation (Arg276 and Glu656), as well as a truncation of the twelve C-terminal residues of the CTR ($CTR^{\Delta C}$) (Fig. 10b). Of the twelve mutants, ten were indistinguishable from the wild-type $SIRT1^{CAT}$•CTR heterodimer in their behavior on a gel filtration column. Two mutants, N346A and D348N, were insoluble following recombinant expression and could therefore not be tested. The ten $SIRT1^{CAT}$•CTR mutants (S265A, R276A, Q345A,

I347A, H363A, F414A, R466A, D481A, E656A, and $CTR^{\Delta C}$) along with the crystallized wild-type $SIRT1^{CAT}$•CTR heterodimer and $SIRT1^{CAT}$ in the absence of the CTR, were tested in two different *in vitro* deacetylase activity assays that employ modified, acetylated p53-based peptides (Fig. 11). The results of both assays were consistent and are summarized in Fig. 10c.

All five active site mutants display either weak (S265A) or no deacetylase activity (Q345A, I347A, H363A, and F414A), demonstrating their involvement in the catalytic reaction. These results are in line with previous studies on other sirtuins and support the previously proposed ADPR-peptidyl-imidate mechanism for the deacetylation reaction.[19] The two mutants R466A and D481A, in which the wild-type residues are sandwiched between the ADPR and the C-terminal extension of the CTR, yield proteins with substantially reduced catalytic activity, highlighting their role in co-factor binding.

$SIRT1^{CAT}$ alone, being prone to aggregation, showed variability in activity between experiments, and, hence, any results obtained in the absence of CTR need to be interpreted with caution. However, despite its fragility, $SIRT1^{CAT}$ in isolation consistently displayed higher activity compared to the $SIRT1^{CAT}$•CTR heterodimer, indicating an inhibiting effect of the CTR. This inhibitory effect is mediated by the C-terminal 10 residues of the CTR ($CTR^{\Delta C}$), as evidenced by the loss of $SIRT1^{CAT}$ inhibition in the $SIRT1^{CAT}$•$CTR^{\Delta C}$ heterodimer. Similar results were obtained for the R276A and E656A mutants, indicating that the salt bridge between these residues is the primary determinant of inhibition by the CTR. The R276A mutant did not significantly increase the activity of $SIRT1^{CAT}$ in the absence of the CTR, suggesting that the primary means through which this residue regulates activity is through its interaction with the CTR. Moreover, both Arg276 and

Glu656 are invariant in mammalian SIRT1 enzymes, indicating that this regulatory feature is evolutionarily conserved.

Biochemical analysis of the SIRT1^{CAT}-CTR interaction

As a result of the observed inhibitory effect of the CTR on SIRT1 activity, we sought to determine whether the CTR is a persistent structural feature of the catalytic domain, as previously observed for the tail region of the catalytic domain of *S. cerevisiae* Sir2 (Fig. 5b).[40] Due to the fact that SIRT1^{CAT} is prone to aggregation, we were unable to determine the dissociation constant between SIRT1^{CAT} and the CTR by isothermal titration calorimetry. However, we tested whether the CTR can be displaced from the SIRT1^{CAT}•CTR complex. To test this hypothesis we incubated the SIRT1^{CAT}•CTR heterodimer with SUMO-tagged CTR and monitored exchange by size-exclusion chromatography. Indeed, a SUMO-CTR can displace the CTR from a preformed SIRT1^{CAT}•CTR complex (Fig. 12a). We additionally incubated SUMO-CTR with a GST-CTR•SIRT1^{CAT} complex and performed pull-down experiments. Similarly, increasing amounts of SUMO-CTR resulted in decreasing amounts of pulled-down SIRT1^{CAT} (Fig. 12b). To further probe the SIRT1^{CAT}•CTR interaction, we designed mutations at the binding interface and assayed their ability to form a complex in a GST pull-down assay (Fig. 12c). Mutation of three hydrophobic CTR residues (Phe644, Tyr650, Ile651) to alanine or aspartate abolished CTR binding to SIRT1^{CAT}. Mutation of Tyr642 to alanine resulted in reduced binding that was completely lost upon mutation to aspartate. Based on these results we conclude that the CTR binding may represent a regulated interaction to inhibit SIRT1 activity.

Thermostability

During our *in vitro* deacetylation assays, we noticed a rapid decay of enzymatic activity and a variability between different purifications for $SIRT1^{CAT}$, which was not observed for $SIRT1^{CAT}$•CTR. In addition, we obtained substantially lower yields for $SIRT1^{CAT}$ in isolation relative to the coexpressed $SIRT1^{CAT}$•CTR complex (Table 2). These findings prompted us to explore whether the CTR stabilizes the catalytic domain. We found that the $SIRT1^{CAT}$•CTR heterodimer was substantially more resistant to temperature denaturation than $SIRT1^{CAT}$ in isolation or in the presence of a non-interacting CTR mutant. Specifically, whereas $SIRT1^{CAT}$ already begins to precipitate at 30 °C, the $SIRT1^{CAT}$•CTR heterodimer remains completely soluble up to 39 °C (Fig. 13a). In line with these results, the isolated SIRT1 catalytic domain displays an almost complete loss of activity when incubated for two hours at 37 °C, whereas activity is maintained for the $SIRT1^{CAT}$•CTR heterodimer (Fig. 13b). These data support a stabilizing role for the CTR on the SIRT1 catalytic domain and provide an explanation for previous difficulties in characterizing the biochemical and structural properties of SIRT1.

DISCUSSION

We have determined the crystal structure of the catalytic domain of human SIRT1 in complex with a C-terminal regulatory segment (CTR) in two different conformations, the *apo* form of the heterodimer and a quaternary complex with ADPR and a pseudo-substrate, the tail of another $SIRT1^{CAT}$ that mimics a p53 substrate sequence. The comparison of the two structures reveals that the interdomain angle and association between the two domains of the catalytic core are dramatically different upon substrate and cofactor binding. Further conformational changes occur in the C-terminal extension of the CTR that inhibit the activity of the catalytic domain.

The observation that sirtuins engage unique regions flanking the canonical catalytic domain has been made previously in the structure of *S. cerevisiae* Hst2.[26] In that case, a homotrimer has been observed in the crystal and in solution, which is mediated by an N-terminal region that interacts with an adjacent catalytic domain in a substrate-mimicking fashion. In human SIRT1, a C-terminal seven-residue tail segment that is disordered in the *apo* state engages an adjacent molecule in the NAD^+-bound closed state. The overall mode of interaction faithfully mimics the interaction between a p53 substrate peptide, the main difference being that a hydrophobic leucine residue, instead of an acetylated lysine residue, is inserted into the hydrophobic tunnel that leads to the internal active site of the enzyme. The substrate-mimicking interaction observed here might be useful for the development of competitive SIRT1 inhibitors.

Our biochemical analysis expands upon recent findings on the regulatory nature of the CTR.[47,48] Namely, we establish an inhibitory role for the CTR primarily mediated by the formation of a salt bridge between the C-terminal extension and the helical module that

is only observed in our closed SIRT1^{CAT}•CTR•ADPR•Substrate structure. We propose that this interaction tunes the activity of the catalytic domain by influencing the conformational changes that are required for substrate and cofactor turnover (Fig. 14). This is in contrast to recent work stating that the CTR is required for SIRT1 activity.[47] However, we believe this conclusion may be a result of the propensity of the SIRT1^{CAT} to aggregate in solution in the absence of the CTR. Despite the overall hydrophobic nature and large surface of the SIRT1^{CAT}-CTR interface, we demonstrate that the SIRT1^{CAT}-CTR interaction is not merely a structural augmentation of the catalytic domain, but rather possesses the necessary characteristics of a regulated interaction. Such a regulatory mechanism could involve the phosphorylation of the CTR, which contains four predicted phosphorylation sites (Tyr650, Tyr658, Ser659, and Ser661).[52]

Our structural and functional analysis presented here allow us to propose a model for the regulation of SIRT1 and provide the framework for the development of potent and selective SIRT1 activity modulators that promise to be a new therapeutic avenue for the treatment of a variety of human diseases.

MATERIALS AND METHODS

Protein expression and purification

DNA fragments of *H. sapiens* SIRT1 were amplified by PCR and cloned into the vector pGEX-6P-1 (GE Healthcare), a modified pET28a vector (Novagen) that contained a PreScission protease site after the N-terminal hexahistidine tag (pET28a-PreS),[53] and a modified pET28b vector that contained an N-terminal hexahistidine-SUMO tag (pET28b-SUMO).[54] SIRT1 mutants were generated by QuikChange mutagenesis (Stratagene) and confirmed by DNA sequencing. Details of all constructs are listed in Table S1.

All proteins were expressed in *E. coli* BL21-CodonPlus(DE3)-RIL cells (Stratagene) grown in LB media supplemented with appropriate antibiotics. Protein expression was induced at an OD_{600} of approximately 0.6 with 500 μM IPTG for 16-18 hours at 18 °C. Cells were harvested by centrifugation.

The $SIRT1^{CAT}$•CTR complex and all $SIRT1^{CAT}$•CTR mutants were made by coexpressing $SIRT1^{CAT}$ in pET28a-PreS and CTR in pGEX-6P-1 (Table S2). Cells were resuspended in a buffer containing 20 mM Tris (pH 8.0), 100 mM NaCl, 5 mM DTT, 2 μM bovine lung aprotinin (Sigma), and complete EDTA-free protease inhibitor cocktail (Roche), and flash frozen in liquid nitrogen. Thawed cells were lysed with a cell disrupter (Avestin) and the lysate was centrifuged for 1 hour at 40,000 x g. The cleared lysate was applied to a glutathione sepharose column equilibrated in 20 mM Tris (pH 8.0), 100 mM NaCl, and 5 mM DTT (GE Healthcare) and eluted via a glutathione gradient. Pooled fractions were cleaved with PreScission protease (GE Healthcare) for 12 hours and further purified over an additional glutathione sepharose column. The protein was concentrated and injected onto a HiLoad 16/60 Superdex 75 column (GE Healthcare)

equilibrated in 20 mM Tris (pH 8.0), 100 mM NaCl, and 5 mM DTT. Pooled fractions were concentrated to 12.5 mg/ml and used in subsequent crystallization trials or for *in vitro* activity assays. The GST-tagged $SIRT1^{CAT}$•CTR used in pull-down experiments was purified in the same manner, omitting the protease cleavage step.

$SIRT1^{CAT}$ in isolation and all $SIRT1^{CAT}$ mutants were made by expressing $SIRT1^{CAT}$ in pET28a-PreS. Cells were resuspended in a buffer containing 20 mM Tris (pH 8.0), 500 mM NaCl, 5 mM b-mercaptoethanol, 2 mM bovine lung aprotinin (Sigma), and complete EDTA-free protease inhibitor cocktail (Roche), and flash frozen in liquid nitrogen. Thawed cells were lysed with a cell disrupter and the lysate was centrifuged for 1 hour at 40,000 x g. The cleared lysate was loaded onto a Ni-NTA column (Qiagen) and eluted via an imidazole gradient. Protein-containing fractions were pooled, dialyzed against a buffer containing 20 mM Tris (pH 8.0), 100 mM NaCl, and 5 mM DTT, and cleaved with PreScission protease for 12 hours. Cleaved protein was bound to a HiTrap Q HP (GE Healthcare) column and eluted via a NaCl gradient, concentrated, and injected onto a HiLoad 16/60 Superdex 75 column equilibrated in 20 mM Tris (pH 8.0), 100 mM NaCl, and 5 mM DTT. SUMO-CTR, and all SUMO-CTR mutants, followed the same purification as $SIRT1^{CAT}$ in isolation.

Structure determination and refinement

Crystals of $SIRT1^{CAT}$•CTR were grown at 21 °C in hanging drops containing 1 μl of the protein and 1 μl of a reservoir solution consisting of 100 mM Tris (pH 7.4), and 11 % PEG 20,000. Crystals grew in the tetragonal space group $P4_32_12$ and reached a maximum size of approximately 700 μm x 100 mm x 100 mm within a week. Crystals of

SIRT1^{CAT}•CTR•ADPR were grown in similar conditions in the presence of a two-fold molar excess of NAD^+ (Sigma), grew in the trigonal space group $P3_221$, and reached a maximum size of approximately 300 mm x 100 mm x 100 mm within two weeks. For cryo protection, crystals were stabilized in 100 mM Tris (pH 7.4), 13 % PEG 20,000, and 20 % ethylene glycol (added in 1 % increments). X-ray diffraction data were collected at 100 K at beamline 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL), and at beamline 8.2.2 at the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory (LBNL). X-ray intensities were processed using the HKL2000 denzo/scalepack package.[55]

Both structures were solved by single wavelength anomalous dispersion (SAD) using X-ray diffraction data collected utilizing the anomalous scattering of the endogenous Zn^{2+} ions. Heavy metal sites were identified with SHELXD.[56] Phases were calculated in SHARP,[57] followed by density modification in DM[58] with solvent flattening, histogram matching, and NCS averaging. This procedure yielded excellent electron density maps that allowed for the placement of the coordinates of the *S. cerevisiae* Sir2 (PDB code 2HJH). Iterative rounds of model building and refinement in Coot[59] and PHENIX[60] yielded final models in which all residues of both protein fragments were resolved in the substrate- and cofactor-bound states with the exception of the C-terminal tails of the catalytic domain (residues 503 to 510; *apo* state) and CTR (residues 661 to 665; both states). The final R_{work} and R_{free} values of the *apo* and the substrate- and cofactor-bound states of SIRT1 were 23.0 % and 26.5 %, and 15.6 % and 18.7 %, respectively. Both models possess excellent stereochemical parameters with no residues in the generously allowed or disallowed

regions of the Ramachandran plot as assessed with PROCHECK and MolProbity.[61,62] For details of the data collection and refinement statistics, see Table 3.

Analytical size-exclusion chromatography

Protein interaction experiments were carried out on a Superdex 200 10/300 GL gel filtration column (GE Healthcare) equilibrated in a buffer containing 20 mM Tris (pH 8.0), 100 mM NaCl, and 5 mM DTT. $SIRT1^{CAT}$ and the $SIRT1^{CAT}$•CTR heterodimer were mixed with an approximately two-fold molar excess of SUMO-tagged CTR, and incubated for 30 minutes on ice. Complex formation was monitored by injection of the preincubated proteins and the individual components in isolation, and confirmed by SDS-PAGE of the protein-containing fractions, followed by Coomassie brilliant blue staining.

GST pull-down assays

12.5 μg of a GST-CTR•$SIRT1^{CAT}$ complex was mixed with varying amounts of SUMO-CTR, added to approximately 20 ml of Glutathione Sepharose beads (GE Healthcare) equilibrated in a buffer containing 20 mM Tris (pH 8.0), 100 mM NaCl, and 5 mM DTT, and incubated on ice for 1 hour. The beads were isolated via centrifugation at 750 x g, and washed five times by resuspension in 200 mL of buffer, followed by subsequent centrifugation. All samples were analyzed by SDS-PAGE, followed by Coomassie brilliant blue staining. For CTR mutational analysis, lysate from *E. coli* transformed with the indicated constructs was added to 250 ml of beads.

Deacetylase activity assays

Deacetylase activity of SIRT1 proteins was determined using the SIRT-Glo (Promega) and *Fluor de Lys* SIRT1 (Enzo) assays, according to the manufacturers' protocols. SIRT-Glo reactions were carried out with 1 mM of pure protein and were incubated at room temperature for 45 minutes. *Fluor de Lys* reactions contained 0.2 mM pure protein, 75 mM substrate, and 3 mM NAD^+, were incubated at 37 °C for 45 minutes, and developed at room temperature for 1 hour. Luminescence (SIRT-Glo) and fluorescence (*Fluor de Lys)* were measured using a Flexstation-3 microplate reader (Molecular Devices) and autofluorescence was subtracted.

Thermostability Assay

15 μg of $SIRT1^{CAT}$ in isolation or mixed with an approximately two fold molar excess of SUMO-CTR or SUMO-$CTR^{Y650D, I651D}$ was incubated at the indicated temperatures for 1 hour. Samples were pelleted by centrifugation at 30,000 x g and analyzed by SDS-PAGE followed by Coomassie brilliant blue staining.

Multiangle light scattering

Purified proteins were characterized by multiangle light scattering following size-exclusion chromatography.[63] Proteins were injected onto a Superdex 200 10/300 GL gel filtration chromatography column equilibrated in a buffer containing 20 mM Tris (pH 8.0), 100 mM NaCl, and 5 mM DTT. The chromatography system was connected in series with an 18-angle light-scattering detector (DAWN HELEOS II; Wyatt Technology), a dynamic light-scattering detector (DynaPro Nanostar; Wyatt Technology), and a refractive index detector

(Optilab t-rEX; Wyatt Technology). Data were collected every 1 second at a flow rate of 0.5 mL/min at 25 °C. Data analysis was carried out using the program ASTRA 6, yielding the molar mass and mass distribution (polydispersity) of the sample.

Illustration and figures

The sequence alignment of SIRT1 was generated using ClustalX and colored with Alscript.[64,65] Figures were generated using PyMOL (www.pymol.org). Electrostatic potential was calculated with Adaptive Poisson-Boltzmann Solver (APBS) software.[66]

Accession numbers

The atomic coordinates and structure factors of the *H. sapiens* $SIRT1^{CAT}$•CTR heterodimer and the quaternary $SIRT1^{CAT}$•CTR•ADPR•Substrate complex have been deposited in the Protein Data Bank under accession codes 4IF6 and 4KXQ.

ACKNOWLEDGEMENTS

We thank Erik W. Debler, Daniel H. Lin, Alina Patke, Pete Stavropoulos, Tobias Stuwe, and Yunji Wu for critical reading of the manuscript, Leslie N. Collins for technical support, David King for mass spectrometry analysis, Anders Näär for providing material, and Jens Kaiser and the scientific staff of SSRL beamline 12-2, and Corie Ralston and scientific staff of ALS beamline 8.2.2 for their support with X-ray diffraction measurements. We acknowledge the Gordon and Betty Moore Foundation for their support of the Molecular Observatory at the California Institute of Technology. The operations at the SSRL are supported by the Department of Energy and by the National Institutes of Health. AMD is supported by a National Institutes of Health Research Service Award (5 T32 GM07616). AH was supported by the Albert Wyrick V Scholar Award of the V Foundation for Cancer Research, the 54th Mallinckrodt Scholar Award of the Edward Mallinckrodt, Jr. Foundation, and a Kimmel Scholar Award of the Sidney Kimmel Foundation for Cancer Research.

FIGURE LEGENDS

Fig. 1. Biochemical analysis of the SIRT1^{CAT}•CTR complex. (a) Domain structure. Blue, NAD^+-binding domain; yellow, helical module; green, Zn^{2+}-binding module; red, pseudo-substrate peptide (T, tail); purple, C-terminal regulatory segment (CTR); grey, predicted unstructured regions (U). The bars above the domain structure mark the crystallized fragments. (b) Size-exclusion chromatography interaction analysis of SIRT1^{CAT} with SUMO-CTR. The analyzed proteins and complexes are indicated in each gel filtration profile. For analysis of complex formation, the SUMO-CTR was mixed at approximately 2-fold molar excess of SIRT1^{CAT} and injected onto a Superdex 200 10/300 GL gel filtration column. Gray bars and colored lines designate the analyzed fractions. Molecular mass standards and the positions of the proteins are indicated. The asterisk indicates a degraded SUMO-CTR fragment. (c) Multiangle light scattering (MALS) analysis of the SIRT1^{CAT}•CTR heterodimer. The differential refractive index is plotted against the elution volume from a Superdex 200 10/300 GL gel filtration column and overlaid with the determined molecular mass for the peak.

Fig. 2. Structure of *H. sapiens* SIRT1. (a) Structure of the quaternary SIRT1^{CAT}•CTR•ADPR•Substrate complex in ribbon representation, colored as in Fig. 1a. (b) Structure of the SIRT1^{CAT}•CTR heterodimer, displayed in the same orientation as in panel (b). 90°-rotated views are shown on the right.

Fig. 3. Multispecies sequence alignment of SIRT1 homologs. (a) A multispecies sequence alignment of the SIRT1 catalytic domain and (b) the C-terminal regulatory segment (CTR).

The numbering below the alignment is relative to *H. Sapiens* SIRT1. The overall sequence conservation at each position is shaded in a color gradient from yellow (60 % identity) to dark red (100 % identity). The secondary structure is indicated above the sequence as arrows (b strands) or rectangles (a helices), and colored as in Fig. 1a. Residues that comprise the interface between $SIRT1^{CAT}$ and the N-terminal region of the CTR (magenta and light blue dots), or between $SIRT1^{CAT}$•CTR and the pseudo-substrate peptide (green and dark blue dots) are indicated below the aligned sequences. Gray dots indicate the invariant cysteine residues that mediate Zn^{2+}-binding.

Fig. 4. Interface between $SIRT1^{CAT}$ and the C-terminal regulatory segment. Ribbon representation of the interface between the $SIRT1^{CAT}$ and the CTR colored according to Fig. 1a. The inset illustrates the position of the CTR-binding site on the larger NAD^+-binding domain and its interacting residues and is expanded on the right. Residues participating in the interface are shown in ball-and-stick representation.

Fig. 5. Structural comparison of SIRT1 structures. (a) Ribbon representations of the closed $SIRT1^{CAT}$•CTR•ADPR•Substrate structure (left), the recently determined $SIRT1^{CAT}$•NAD^+•EX527 analog structure (middle) and a superposition of the two (right). (b) Ribbon representations of the closed $SIRT1^{CAT}$•CTR•ADPR•Substrate structure (left), the recently determined *sc*Sir2•ADPR•*sc*Sir4 structure (middle) and a superposition of the two (right). For clarity, *sc*Sir4 and the N-terminal domain of *sc*Sir2 have been removed. The PDB codes for the $SIRT1^{CAT}$•NAD^+•EX527 analog and *sc*Sir2•ADPR•*sc*Sir4 structures are 4I5I and 4IAO, respectively.[39,40]

Fig. 6. Crystallographic analysis. (a) Ribbon representation of the SIRT1^{CAT}•CTR•ADPR•Substrate complex, colored according to Fig. 1. A symmetry-related molecule, colored in gray, provides the p53-mimicking pseudo-substrate peptide (red). (b) Superposition of the four copies of the *apo* SIRT1^{CAT}•CTR heterodimer in the asymmetric unit.

Fig. 7. Structural comparison of *H. sapiens* SIRT1 to a bacterial Sir2 substrate complex. (a) A ribbon representation of SIRT1^{CAT}•CTR•ADPR•Substrate (left, colored as in Fig. 1a), *T. maritima* Sir2 in complex with a p53 peptide (middle), and a superposition of the two (right). (b) A stick representation of the pseudo-substrate and p53 substrate peptide from the corresponding structures in panel (a) and their superposition. The PDB code of the *Tm*Sir2 structure is 2H59.[29] (c) Final 2|fo|-|fc| electron density map around the ADPR and pseudo-substrate peptide rendered at 1.0 σ.

Fig. 8. Analysis of oligomeric state. MALS analysis of the SIRT1^{CAT}•CTR complex in the presence of 5 mM NAD^+. The differential refractive indices of SIRT1^{CAT}•CTR are plotted against the elution volumes from a Superdex 200 10/300 GL gel-filtration column (GE Healthcare) and are overlaid with the determined molecular mass for the selected peak.

Fig. 9. Structural comparison of open and closed SIRT1 states. (a) Ribbon representations of the closed SIRT1^{CAT}•CTR•ADPR•Substrate state, (b) the open SIRT1^{CAT}•CTR conformation, and (c) a superposition of the two. (d, e) Surface representations of the corresponding structures in panels (a) and (b). A 90°-rotated view is shown on the right. (f)

A model of pseudo-substrate peptide and ADPR binding to the open $SIRT1^{CAT}$•CTR state. A 90°-rotated view is shown on the right.

Fig. 10. Mutational analysis. (a) Ribbon representation of the $SIRT1^{CAT}$•CTR•ADPR•Substrate complex, with a stick representation of all mutated active side residues. The Zn^{2+}-binding and helical modules were removed for clarity. (b) Ribbon representation of the $SIRT1^{CAT}$•CTR•ADPR•Substrate complex, with a stick representation of all mutated $SIRT1^{CAT}$-CTR interface residues. (c) Deacetylase activity of SIRT1 proteins and mutants towards modified p53-based peptides. Activity was determined in a luminescence (red) or fluorescence (blue) based assay and normalized against wild-type $SIRT1^{CAT}$ activity. Each data point represents the mean of at least three independent measurements. Error bars represent standard deviations.

Fig. 11. SIRT1 mutant proteins. The purified proteins used in all experiments visualized by SDS-PAGE, followed by Coomassie brilliant blue staining. The molecular mass standards and the positions of the proteins are indicated. The CTR is too small to be visualized on the gel.

Fig. 12. Exchange of the C-terminal regulatory segment. (a) Size-exclusion chromatography interaction analysis of the $SIRT1^{CAT}$•CTR heterodimer with SUMO-CTR. The analyzed proteins and complexes are indicated in each gel filtration profile. For analysis of complex formation, SUMO-CTR was mixed at approximately 2-fold molar excess and injected onto a Superdex 200 10/300 GL gel filtration column. Gray bars and

colored lines designate the analyzed fractions. Molecular mass standards and the positions of the proteins are indicated. Untagged CTR is too small to be visualized on the SDS-PAGE. (b) GST pull-downs of a preformed GST-CTR•SIRT1^{CAT} complex with increasing amounts of SUMO-CTR. The loaded (top) and pulled-down (bottom) samples are visualized by SDS-PAGE, followed by Coomassie brilliant blue staining. As reference, the first two lanes of each gel contain the loaded samples of GST-CTR•SIRT1^{CAT} and SUMO-CTR. The relative molar amounts of GST-CTR•SIRT1^{CAT} and SUMO-CTR, molecular mass standards and the positions of the proteins are indicated. (c) CTR mutational analysis. His-tagged SIRT1^{CAT} and GST-tagged CTR variants were coexpressed in *E. coli* and the resulting soluble lysate fraction was analyzed in GST pull-down assays. Samples were visualized by SDS-PAGE, followed by Coomassie brilliant blue staining. Molecular mass standards and the positions of the proteins are indicated. The asterisk indicates a degraded GST-CTR fragment

Fig. 13. Thermostability assay. (a) SIRT1^{CAT} was incubated alone and in the presence of SUMO-CTR or a non-interacting SUMO-CTR$^{Y650D,\ I651D}$ mutant at the indicated temperatures for one hour and pelleted by centrifugation. Pellet and supernatant fractions were visualized by SDS-PAGE, followed by Coomassie brilliant blue staining. Molecular mass standards and the positions of the proteins are indicated. (b) Deacetylase activity of SIRT1^{CAT} and the SIRT1^{CAT}•CTR heterodimer under normal conditions (red) and following incubation at 37 °C for approximately 2 hours prior to the assay (blue). Activity was determined in a fluorescence-based assay and normalized against wild-type SIRT1^{CAT} activity. Each data point represents the mean of at least three independent measurements.

Error bars represent standard deviations. The loss of deacetylase activity is attributed to the aggregation of the catalytic domain.

Fig. 14. Model for the conformational changes and regulation of SIRT1. A cartoon representation of the *apo* SIRT1^{CAT}•CTR heterodimer (left) and the SIRT1^{CAT}•CTR•ADPR•Substrate complex (right), colored as in Fig. 1a. Substrate and co-factor binding leads to the closure of the SIRT1 catalytic domain. The substrate peptide is primarily bound by backbone interactions and the formation of a three-stranded anti-parallel β-staple. The hydrophobic acetylated lysine substrate residue reaches into the secluded internal active site through the hydrophobic tunnel where it is oriented in close proximity to the activated NAD^+. In the closed state, the CTR forms a salt-bridge with the helical module, thereby reducing the efficiency of catalysis.

Table 1. Bacterial expression constructs

[a] Constructs that were used for crystallization of the SIRT1^{CAT}•CTR heterodimer and the SIRT1^{CAT}•CTR•ADPR•Substrate complex

Table 2. Purification yields

Table 3. Crystallographic analysis

[a] SSRL, Stanford Synchrotron Radiation Lightsource

[b] Highest resolution shell is shown in parentheses

[c] As determined by Molprobity[62]

REFERENCES

1. Klar, A. J., Fogel, S. & Macleod, K. (1979). MAR1-a Regulator of the HMa and HMalpha Loci in SACCHAROMYCES CEREVISIAE. *Genetics* **93**, 37-50.
2. Haber, J. E. & George, J. P. (1979). A mutation that permits the expression of normally silent copies of mating-type information in Saccharomyces cerevisiae. *Genetics* **93**, 13-35.
3. Rine, J., Strathern, J. N., Hicks, J. B. & Herskowitz, I. (1979). A suppressor of mating-type locus mutations in Saccharomyces cerevisiae: evidence for and identification of cryptic mating-type loci. *Genetics* **93**, 877-901.
4. Aparicio, O. M., Billington, B. L. & Gottschling, D. E. (1991). Modifiers of position effect are shared between telomeric and silent mating-type loci in S. cerevisiae. *Cell* **66**, 1279-87.
5. Bryk, M., Banerjee, M., Murphy, M., Knudsen, K. E., Garfinkel, D. J. & Curcio, M. J. (1997). Transcriptional silencing of Ty1 elements in the RDN1 locus of yeast. *Genes Dev.* **11**, 255-69.
6. Fritze, C. E., Verschueren, K., Strich, R. & Easton Esposito, R. (1997). Direct evidence for SIR2 modulation of chromatin structure in yeast rDNA. *EMBO J.* **16**, 6495-509.
7. Smith, J. S. & Boeke, J. D. (1997). An unusual form of transcriptional silencing in yeast ribosomal DNA. *Genes Dev.* **11**, 241-54.
8. Imai, S., Armstrong, C. M., Kaeberlein, M. & Guarente, L. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795-800.

9. Landry, J., Sutton, A., Tafrov, S. T., Heller, R. C., Stebbins, J., Pillus, L. & Sternglanz, R. (2000). The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases. *Proc. Natl. Acad. Sci. USA* **97**, 5807-11.

10. Smith, J. S., Brachmann, C. B., Celic, I., Kenna, M. A., Muhammad, S., Starai, V. J., Avalos, J. L., Escalante-Semerena, J. C., Grubmeyer, C., Wolberger, C. & Boeke, J. D. (2000). A phylogenetically conserved NAD+-dependent protein deacetylase activity in the Sir2 protein family. *Proc. Natl. Acad. Sci. USA* **97**, 6658-63.

11. Garcia-Salcedo, J. A., Gijon, P., Nolan, D. P., Tebabi, P. & Pays, E. (2003). A chromosomal SIR2 homologue with both histone NAD-dependent ADP-ribosyltransferase and deacetylase activities is involved in DNA repair in Trypanosoma brucei. *EMBO J.* **22**, 5851-62.

12. Liszt, G., Ford, E., Kurtev, M. & Guarente, L. (2005). Mouse Sir2 homolog SIRT6 is a nuclear ADP-ribosyltransferase. *J. Biol. Chem.* **280**, 21313-20.

13. Tsang, A. W. & Escalante-Semerena, J. C. (1998). CobB, a new member of the SIR2 family of eucaryotic regulatory proteins, is required to compensate for the lack of nicotinate mononucleotide:5,6-dimethylbenzimidazole phosphoribosyltransferase activity in cobT mutants during cobalamin biosynthesis in Salmonella typhimurium LT2. *J. Biol. Chem.* **273**, 31788-94.

14. Kowieski, T. M., Lee, S. & Denu, J. M. (2008). Acetylation-dependent ADP-ribosylation by Trypanosoma brucei Sir2. *J. Biol. Chem.* **283**, 5317-26.

15. Merrick, C. J. & Duraisingh, M. T. (2007). Plasmodium falciparum Sir2: an unusual sirtuin with dual histone deacetylase and ADP-ribosyltransferase activity. *Eukaryot. Cell* **6**, 2081-91.

16. Haigis, M. C., Mostoslavsky, R., Haigis, K. M., Fahie, K., Christodoulou, D. C., Murphy, A. J., Valenzuela, D. M., Yancopoulos, G. D., Karow, M., Blander, G., Wolberger, C., Prolla, T. A., Weindruch, R., Alt, F. W. & Guarente, L. (2006). SIRT4 inhibits glutamate dehydrogenase and opposes the effects of calorie restriction in pancreatic beta cells. *Cell* **126**, 941-54.

17. Du, J., Zhou, Y., Su, X., Yu, J. J., Khan, S., Jiang, H., Kim, J., Woo, J., Kim, J. H., Choi, B. H., He, B., Chen, W., Zhang, S., Cerione, R. A., Auwerx, J., Hao, Q. & Lin, H. (2011). Sirt5 is a NAD-dependent protein lysine demalonylase and desuccinylase. *Science* **334**, 806-9.

18. Sauve, A. A., Wolberger, C., Schramm, V. L. & Boeke, J. D. (2006). The biochemistry of sirtuins. *Annu. Rev. Biochem.* **75**, 435-65.

19. Sauve, A. A., Celic, I., Avalos, J., Deng, H., Boeke, J. D. & Schramm, V. L. (2001). Chemistry of gene silencing: the mechanism of NAD+-dependent deacetylation reactions. *Biochemistry* **40**, 15456-63.

20. Avalos, J. L., Bever, K. M. & Wolberger, C. (2005). Mechanism of sirtuin inhibition by nicotinamide: altering the NAD(+) cosubstrate specificity of a Sir2 enzyme. *Mol. Cell* **17**, 855-68.

21. Avalos, J. L., Boeke, J. D. & Wolberger, C. (2004). Structural basis for the mechanism and regulation of Sir2 enzymes. *Mol. Cell* **13**, 639-48.

22. Avalos, J. L., Celic, I., Muhammad, S., Cosgrove, M. S., Boeke, J. D. & Wolberger, C. (2002). Structure of a Sir2 enzyme bound to an acetylated p53 peptide. *Mol. Cell* **10**, 523-35.

23. Finnin, M. S., Donigian, J. R. & Pavletich, N. P. (2001). Structure of the histone deacetylase SIRT2. *Nat. Struct. Biol.* **8**, 621-5.

24. Jin, L., Wei, W., Jiang, Y., Peng, H., Cai, J., Mao, C., Dai, H., Choy, W., Bemis, J. E., Jirousek, M. R., Milne, J. C., Westphal, C. H. & Perni, R. B. (2009). Crystal structures of human SIRT3 displaying substrate-induced conformational changes. *J. Biol. Chem.* **284**, 24394-405.

25. Min, J., Landry, J., Sternglanz, R. & Xu, R. M. (2001). Crystal structure of a SIR2 homolog-NAD complex. *Cell* **105**, 269-79.

26. Zhao, K., Chai, X., Clements, A. & Marmorstein, R. (2003). Structure and autoregulation of the yeast Hst2 homolog of Sir2. *Nat. Struct. Biol.* **10**, 864-71.

27. Zhao, K., Chai, X. & Marmorstein, R. (2003). Structure of the yeast Hst2 protein deacetylase in ternary complex with 2'-O-acetyl ADP ribose and histone peptide. *Structure* **11**, 1403-11.

28. Zhao, K., Harshaw, R., Chai, X. & Marmorstein, R. (2004). Structural basis for nicotinamide cleavage and ADP-ribose transfer by NAD(+)-dependent Sir2 histone/protein deacetylases. *Proc. Natl. Acad. Sci. USA* **101**, 8563-8.

29. Hoff, K. G., Avalos, J. L., Sens, K. & Wolberger, C. (2006). Insights into the sirtuin mechanism from ternary complexes containing NAD+ and acetylated peptide. *Structure* **14**, 1231-40.

30. Chang, J. H., Kim, H. C., Hwang, K. Y., Lee, J. W., Jackson, S. P., Bell, S. D. & Cho, Y. (2002). Structural basis for the NAD-dependent deacetylase mechanism of Sir2. *J. Biol. Chem.* **277**, 34489-98.

31. Szczepankiewicz, B. G., Dai, H., Koppetsch, K. J., Qian, D., Jiang, F., Mao, C. & Perni, R. B. (2012). Synthesis of carba-NAD and the structures of its ternary complexes with SIRT3 and SIRT5. *J. Org. Chem.* **77**, 7319-29.

32. Zhou, Y., Zhang, H., He, B., Du, J., Lin, H., Cerione, R. A. & Hao, Q. (2012). The bicyclic intermediate structure provides insights into the desuccinylation mechanism of human sirtuin 5 (SIRT5). *J. Biol. Chem.* **287**, 28307-14.

33. Schuetz, A., Min, J., Antoshenko, T., Wang, C. L., Allali-Hassani, A., Dong, A., Loppnau, P., Vedadi, M., Bochkarev, A., Sternglanz, R. & Plotnikov, A. N. (2007). Structural basis of inhibition of the human NAD+-dependent deacetylase SIRT5 by suramin. *Structure* **15**, 377-89.

34. Cosgrove, M. S., Bever, K., Avalos, J. L., Muhammad, S., Zhang, X. & Wolberger, C. (2006). The structural basis of sirtuin substrate affinity. *Biochemistry* **45**, 7511-21.

35. Zhu, A. Y., Zhou, Y., Khan, S., Deitsch, K. W., Hao, Q. & Lin, H. (2012). Plasmodium falciparum Sir2A preferentially hydrolyzes medium and long chain fatty acyl lysine. *ACS Chem. Biol.* **7**, 155-9.

36. Bheda, P., Wang, J. T., Escalante-Semerena, J. C. & Wolberger, C. (2011). Structure of Sir2Tm bound to a propionylated peptide. *Protein Sci.* **20**, 131-9.

37. Hawse, W. F., Hoff, K. G., Fatkins, D. G., Daines, A., Zubkova, O. V., Schramm, V. L., Zheng, W. & Wolberger, C. (2008). Structural insights into intermediate steps in the Sir2 deacetylation reaction. *Structure* **16**, 1368-77.

38. Sanders, B. D., Zhao, K., Slama, J. T. & Marmorstein, R. (2007). Structural basis for nicotinamide inhibition and base exchange in Sir2 enzymes. *Mol. Cell* **25**, 463-72.

39. Zhao, X., Allison, D., Condon, B., Zhang, F., Gheyi, T., Zhang, A., Ashok, S., Russell, M., MacEwan, I., Qian, Y., Jamison, J. A. & Luz, J. G. (2013). The 2.5 A crystal structure of the SIRT1 catalytic domain bound to nicotinamide adenine dinucleotide (NAD+) and an indole (EX527 analogue) reveals a novel mechanism of histone deacetylase inhibition. *J. Med. Chem.* **56**, 963-9.

40. Hsu, H. C., Wang, C. L., Wang, M., Yang, N., Chen, Z., Sternglanz, R. & Xu, R. M. (2013). Structural basis for allosteric stimulation of Sir2 activity by Sir4 binding. *Genes Dev.* **27**, 64-73.

41. Haigis, M. C. & Sinclair, D. A. (2010). Mammalian sirtuins: biological insights and disease relevance. *Annu. Rev. Pathol.* **5**, 253-95.

42. Guarente, L. (2007). Sirtuins in aging and disease. *Cold Spring Harb. Symp. Quant. Biol.* **72**, 483-8.

43. Sauve, A. A. & Schramm, V. L. (2003). Sir2 regulation by nicotinamide results from switching between base exchange and deacetylation chemistry. *Biochemistry* **42**, 9249-56.

44. Kim, E. J., Kho, J. H., Kang, M. R. & Um, S. J. (2007). Active regulator of SIRT1 cooperates with SIRT1 and facilitates suppression of p53 activity. *Mol. Cell* **28**, 277-90.

45. Kim, J. E., Chen, J. & Lou, Z. (2008). DBC1 is a negative regulator of SIRT1. *Nature* **451**, 583-6.

46. Zhao, W., Kruse, J. P., Tang, Y., Jung, S. Y., Qin, J. & Gu, W. (2008). Negative regulation of the deacetylase SIRT1 by DBC1. *Nature* **451**, 587-90.

47. Kang, H., Suh, J. Y., Jung, Y. S., Jung, J. W., Kim, M. K. & Chung, J. H. (2011). Peptide switch is essential for Sirt1 deacetylase activity. *Mol. Cell* **44**, 203-13.

48. Pan, M., Yuan, H., Brent, M., Ding, E. C. & Marmorstein, R. (2012). SIRT1 contains N- and C-terminal regions that potentiate deacetylase activity. *J. Biol. Chem.* **287**, 2468-76.

49. French, J. B., Cen, Y. & Sauve, A. A. (2008). Plasmodium falciparum Sir2 is an NAD+-dependent deacetylase and an acetyllysine-dependent and acetyllysine-independent NAD+ glycohydrolase. *Biochemistry* **47**, 10227-39.

50. Sanders, B. D., Jackson, B. & Marmorstein, R. (2010). Structural basis for sirtuin function: what we know and what we don't. *Biochim Biophys Acta* **1804**, 1604-16.

51. Gertz, M., Fischer, F., Nguyen, G. T., Lakshminarasimhan, M., Schutkowski, M., Weyand, M. & Steegborn, C. (2013). Ex-527 inhibits Sirtuins by exploiting their unique NAD+-dependent deacetylation mechanism. *Proc Natl Acad Sci U S A* **110**, E2772-81.

52. Blom, N., Gammeltoft, S. & Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351-62.

53. Hoelz, A., Nairn, A. C. & Kuriyan, J. (2003). Crystal structure of a tetradecameric assembly of the association domain of Ca2+/calmodulin-dependent kinase II. *Mol. Cell* **11**, 1241-51.

54. Mossessova, E. & Lima, C. D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* **5**, 865-76.

55. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.

56. Sheldrick, G. M. (2008). A short history of SHELX. *Acta Crystallogr. A* **64**, 112-22.

57. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023-30.

58. CCP4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-3.

59. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126-2132.

60. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). PHENIX: a comprehensive Python-

based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213-221.

61. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* **26**, 283-291.

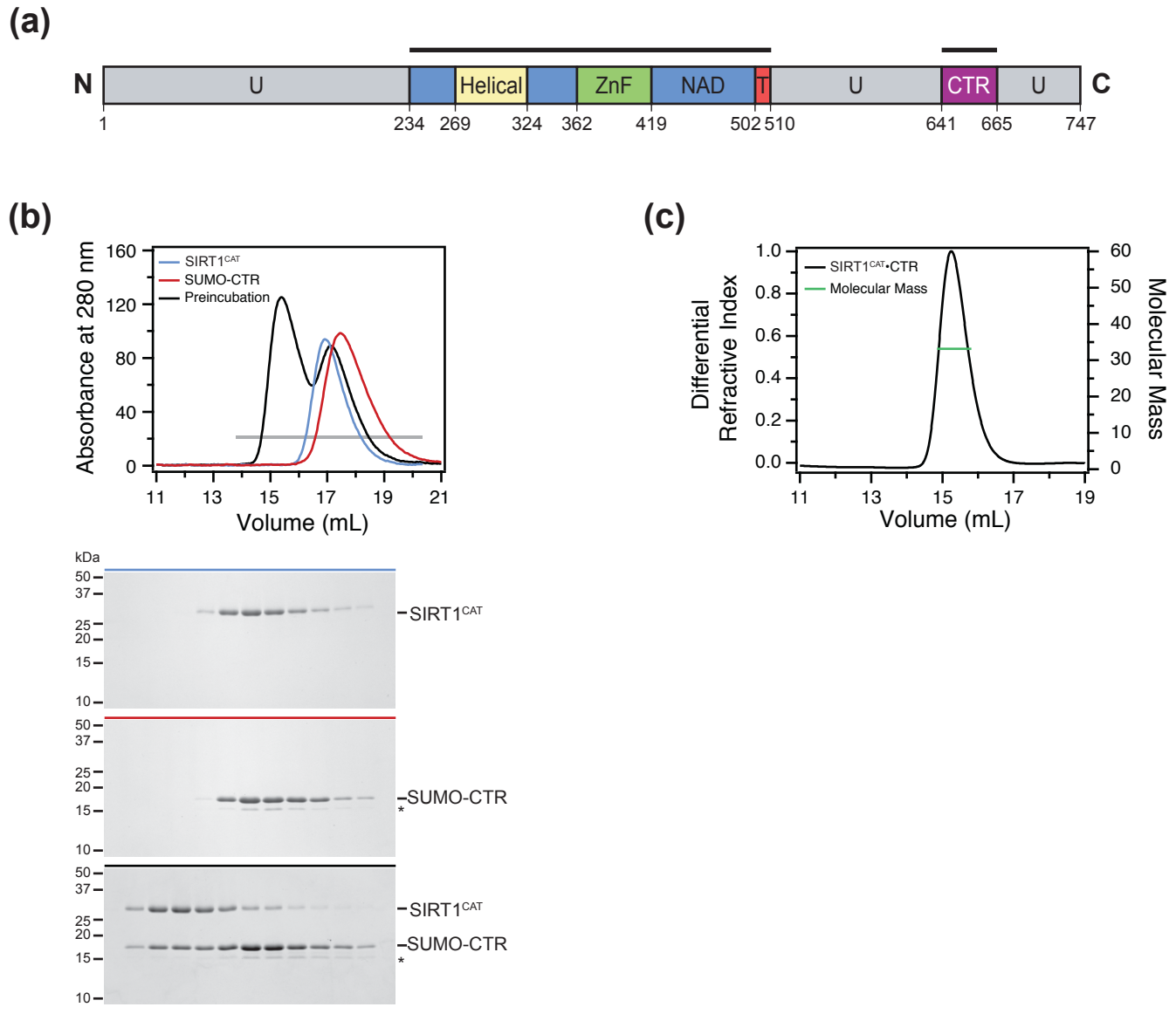
62. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375-W383.

63. Wyatt, P. J. (1997). Multiangle light scattering: The basic tool for macromolecular characterization. *Instrum. Sci. Technol.* **25**, 1-18.

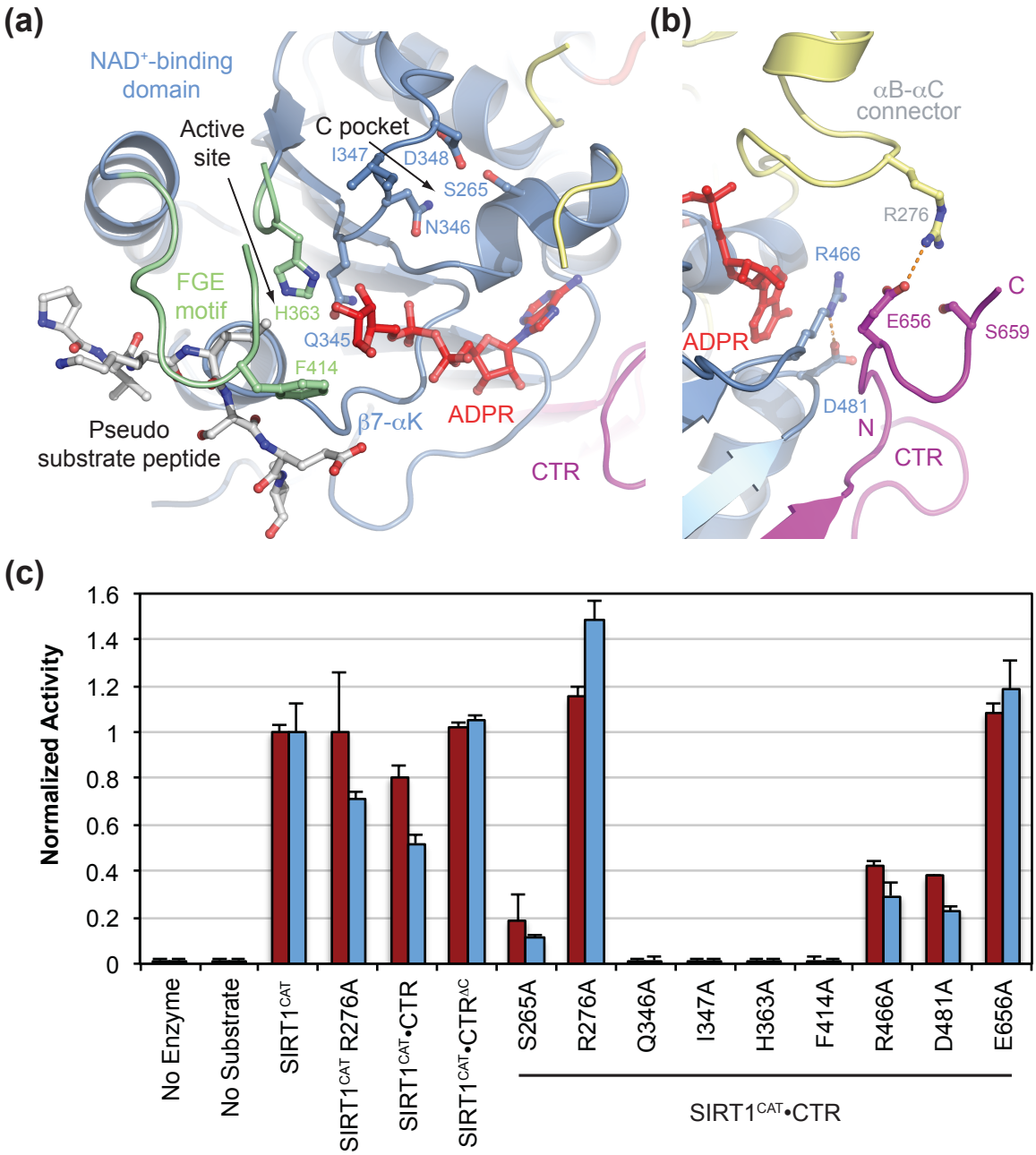
64. Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**, 403-5.

65. Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **6**, 37-40.

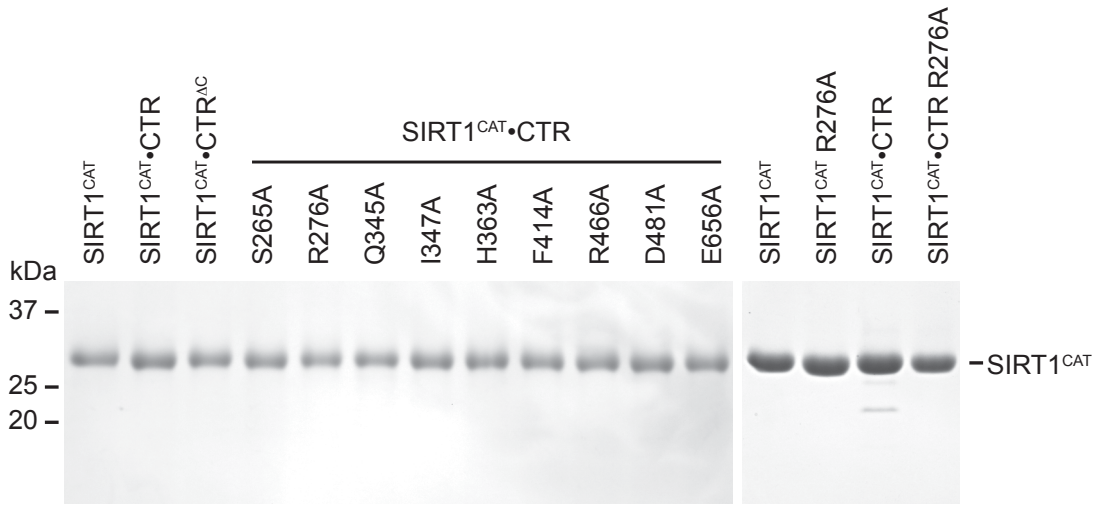
66. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037-41.



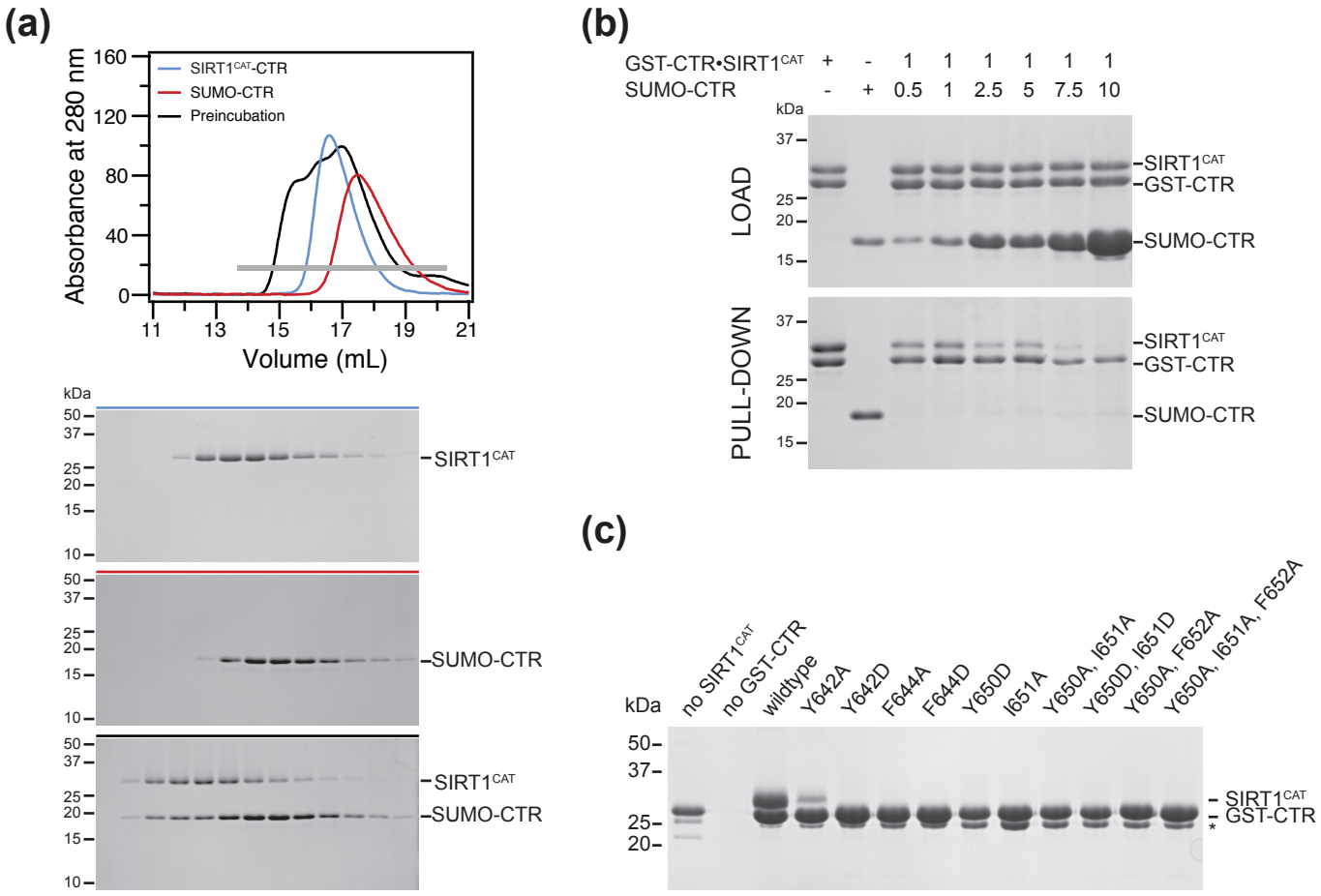
Chapter IV, Figure 1



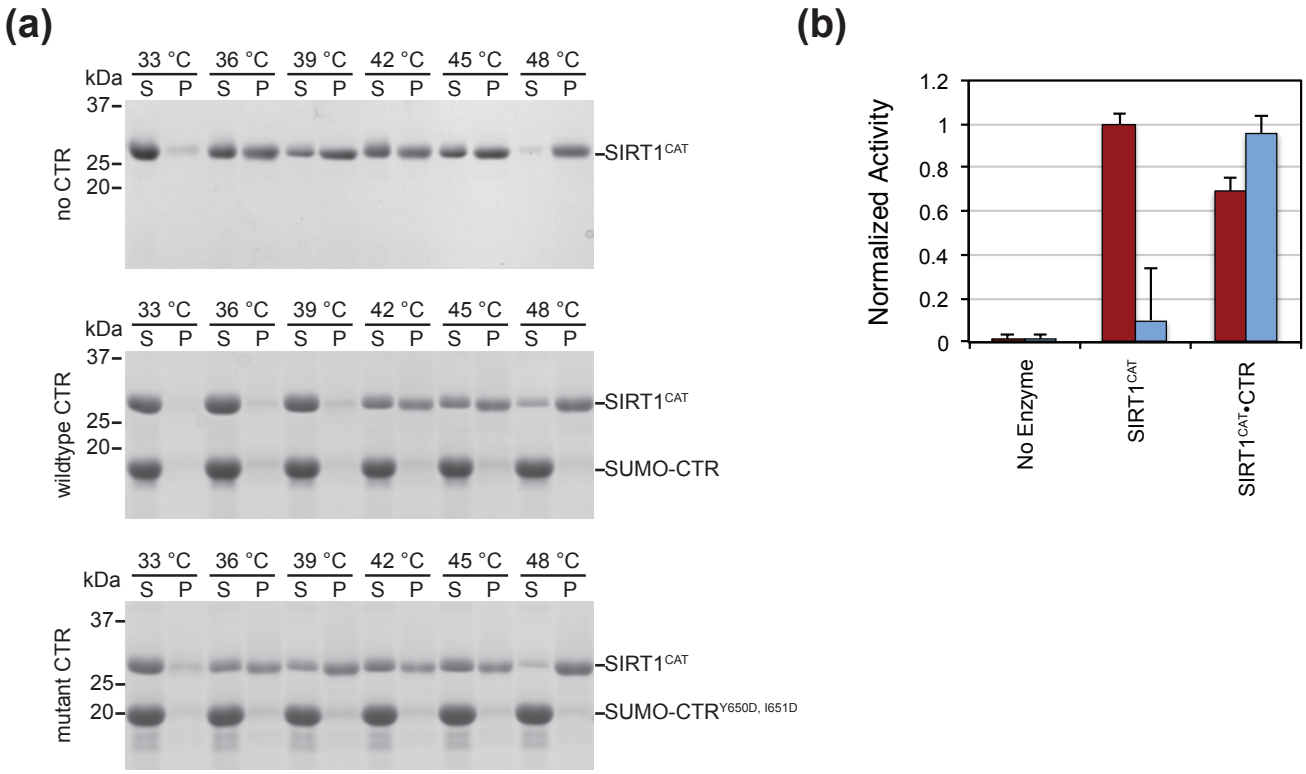
Chapter IV, Figure 10



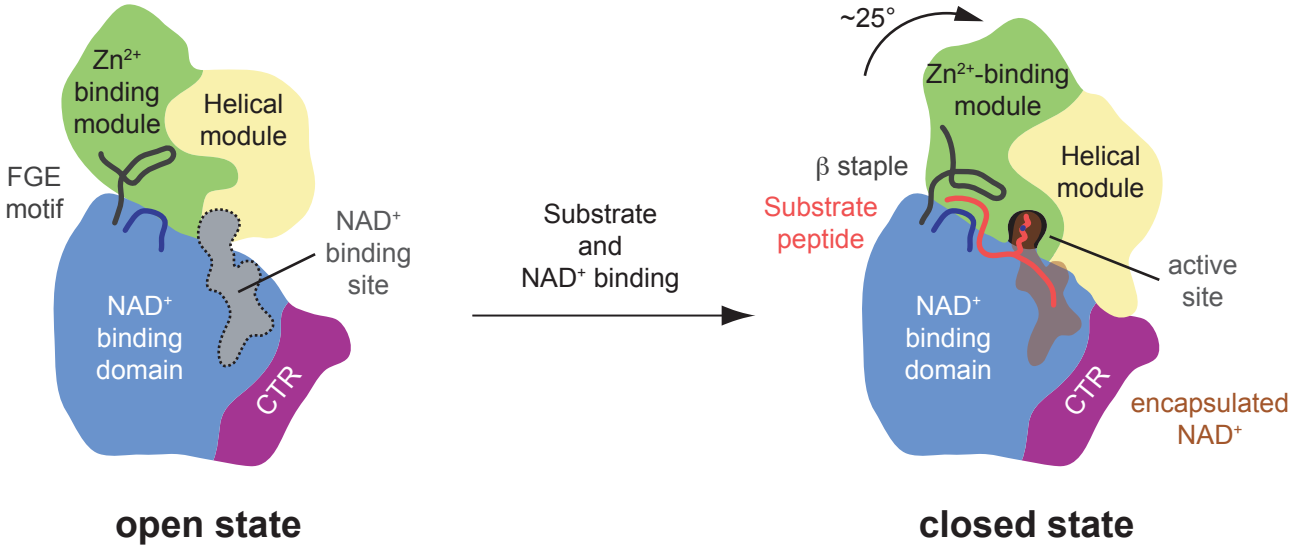
Chapter IV, Figure 11



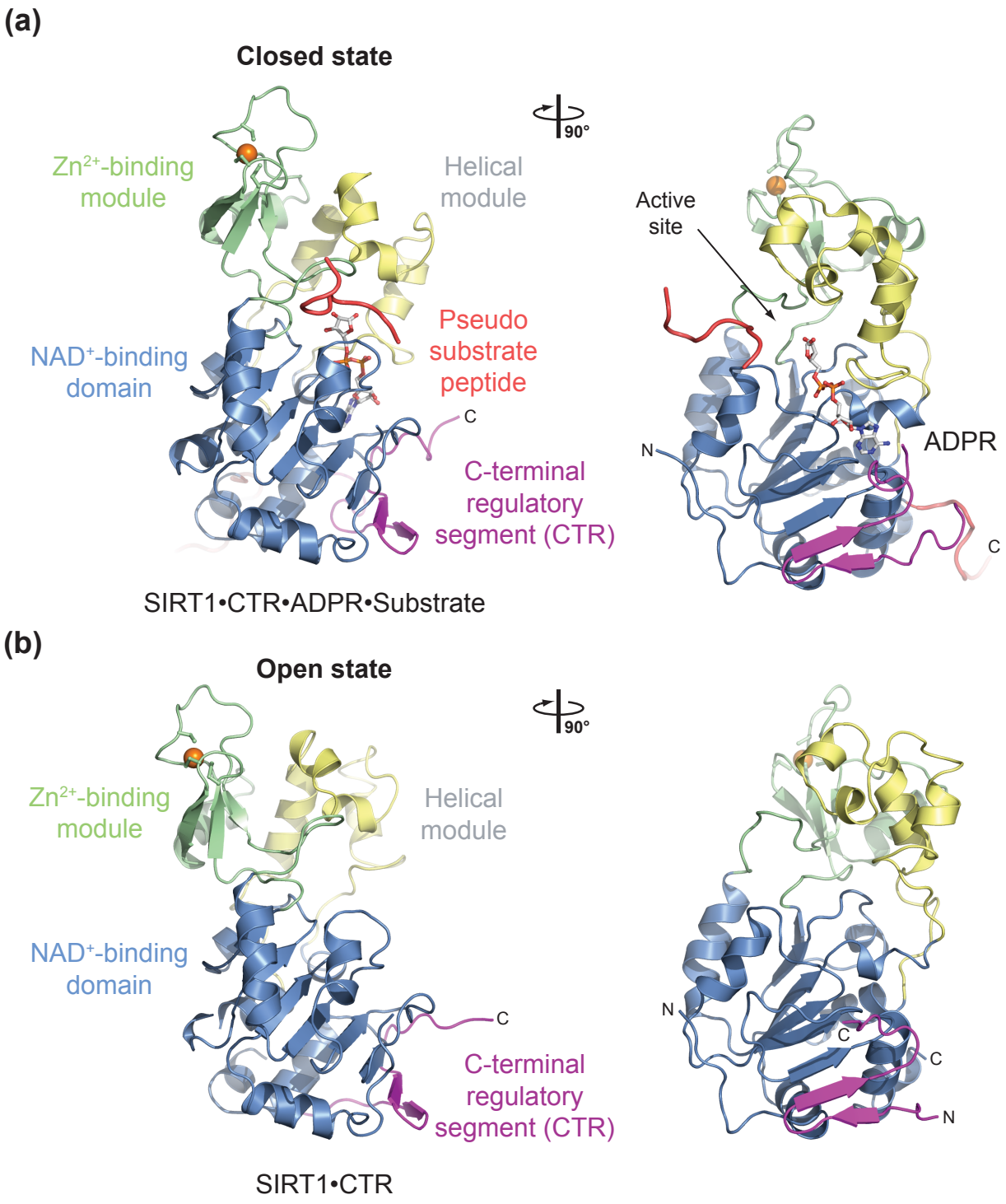
Chapter IV, Figure 12



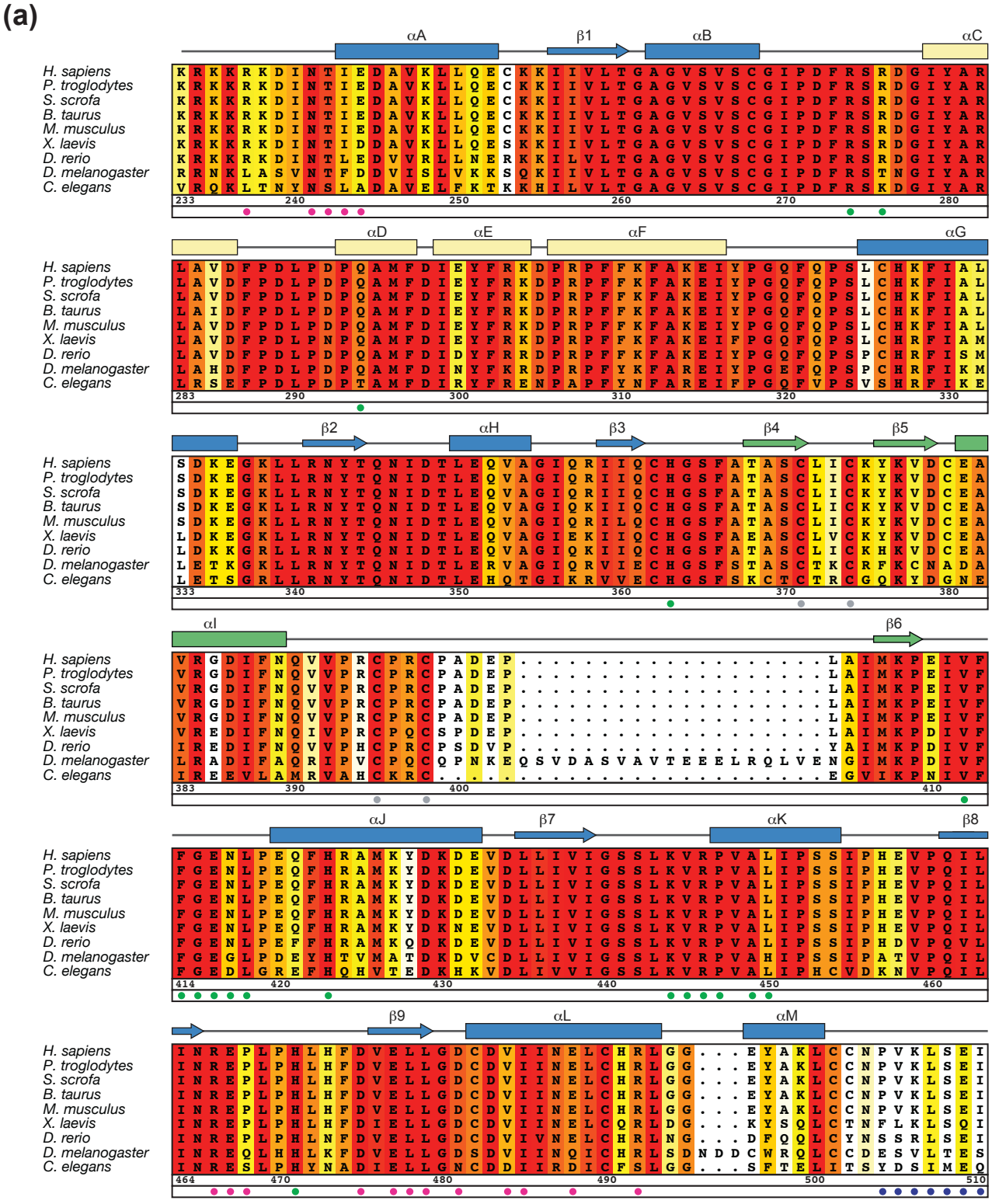
Chapter IV, Figure 13

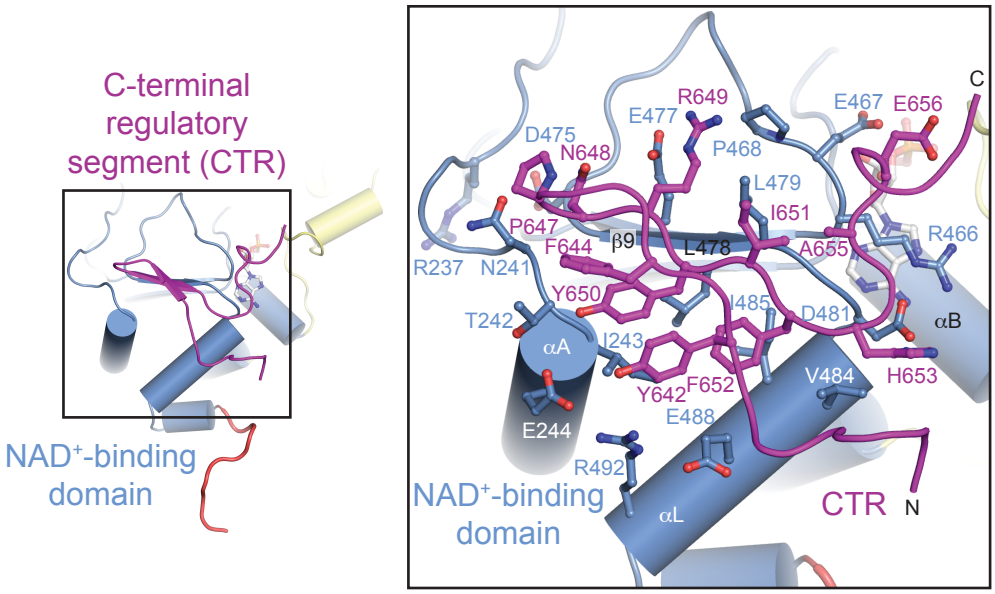


Chapter IV, Figure 14



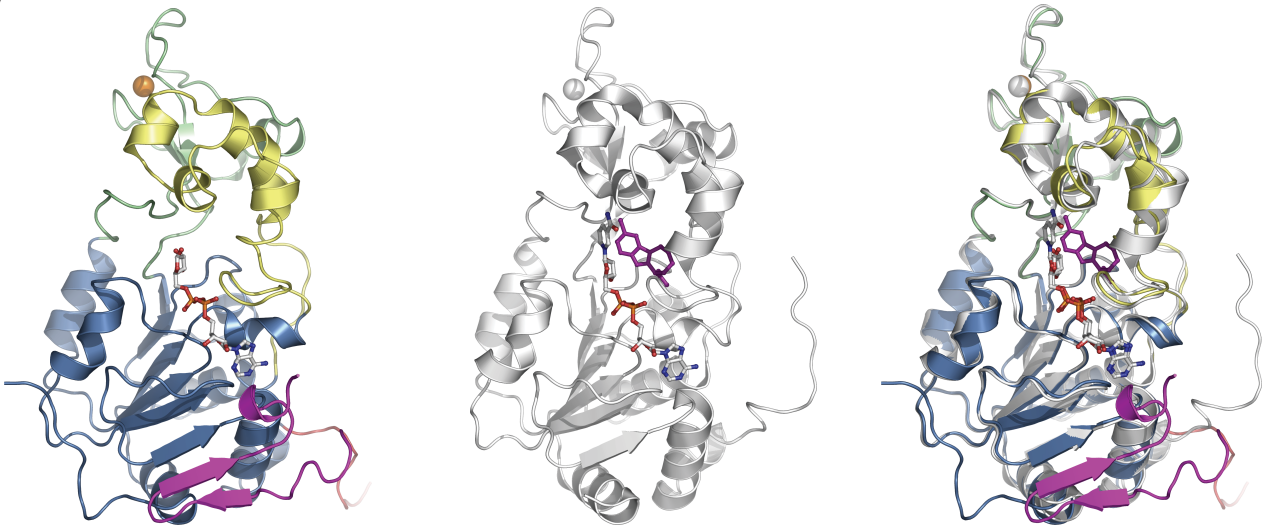
Chapter IV, Figure 2


(a)
αA
β1
αB
αC
H. sapiens
P. troglodytes
S. scrofa
B. taurus
M. musculus
X. laevis
D. rerio
D. melanogaster
C. elegans
αD
αE
αF
αG
β2
αH
β3
β4
β5
αI
β6
αJ
β7
αK
β8
β9
αL
αM
(b)
β10
β11



Chapter IV, Figure 4

(a)

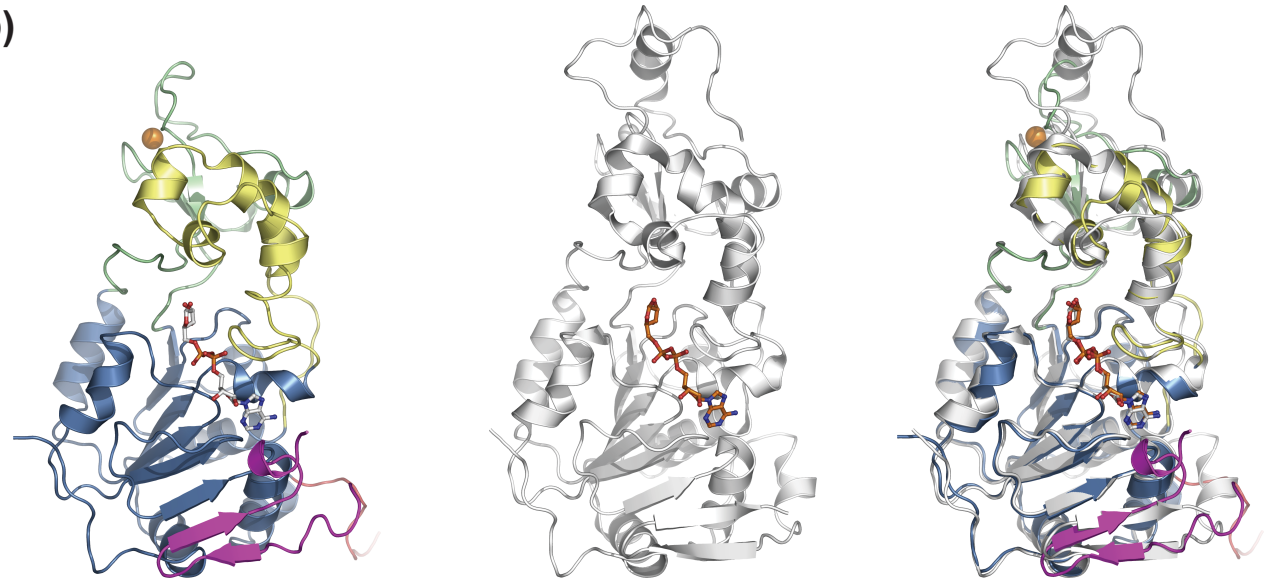


SIRT1•CTR•ADPR•Substrate

SIRT1•NAD$^+$•Ex527 analog

SUPERPOSITION

(b)



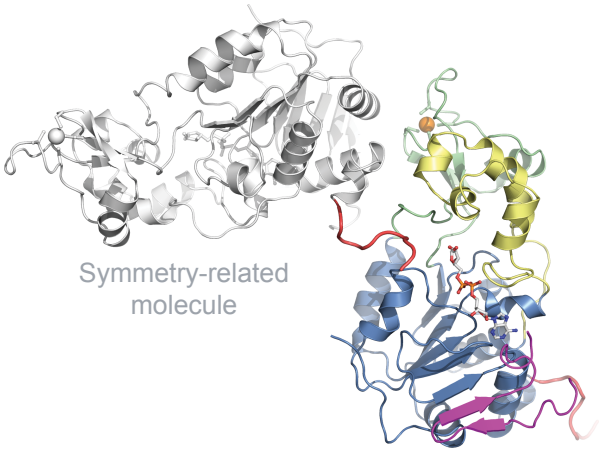
SIRT1•CTR•ADPR•Substrate

*sc*Sir2•ADPR•*sc*Sir4

SUPERPOSITION

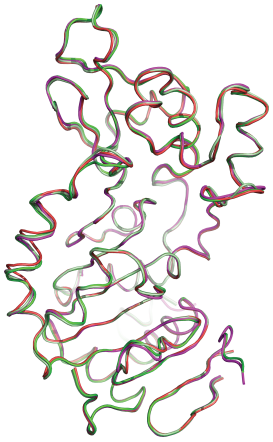
Chapter IV, Figure 5

(a)



SIRT1•CTR•ADPR•Substrate

(b)



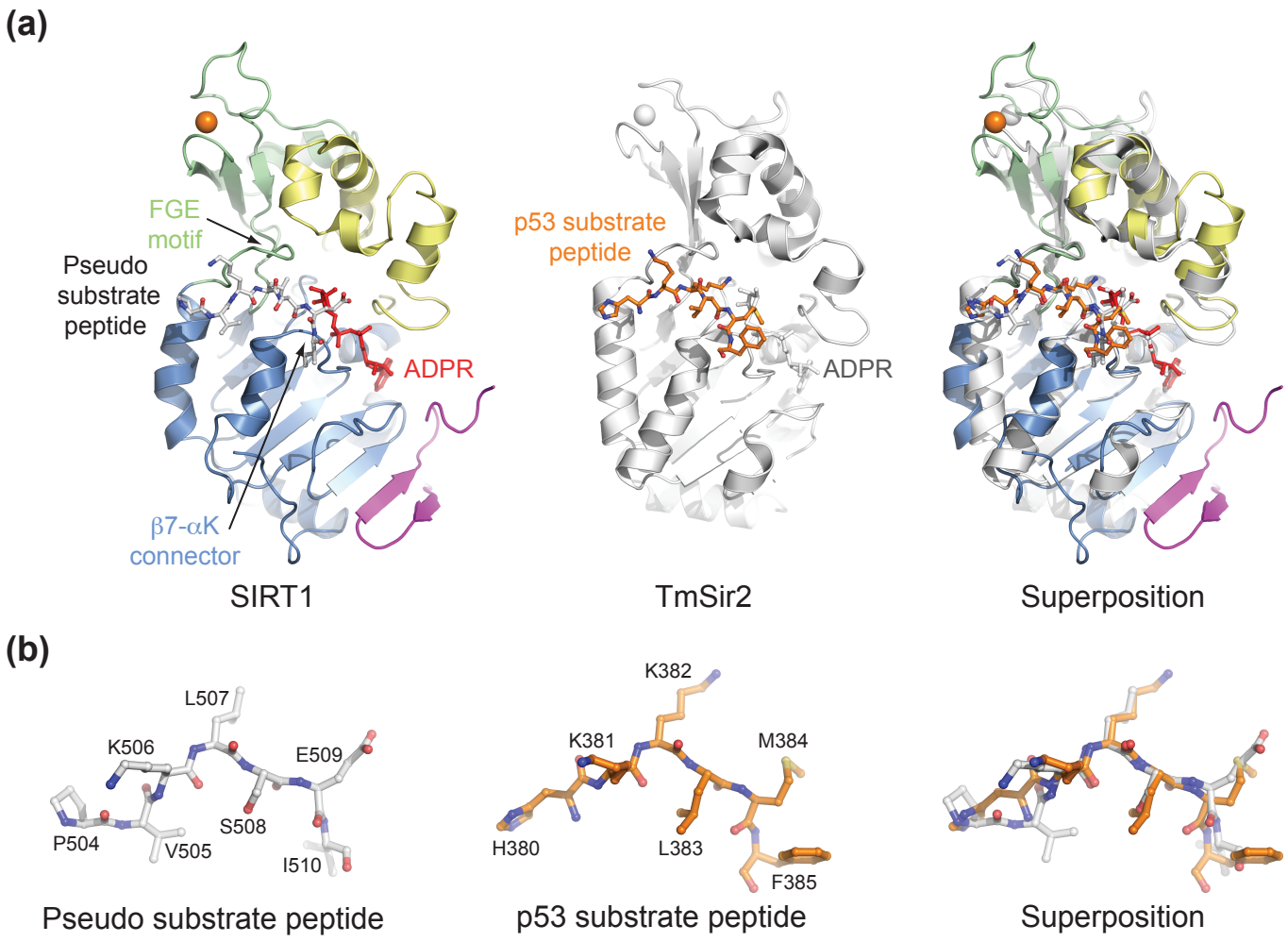
molecule 1

molecule 2

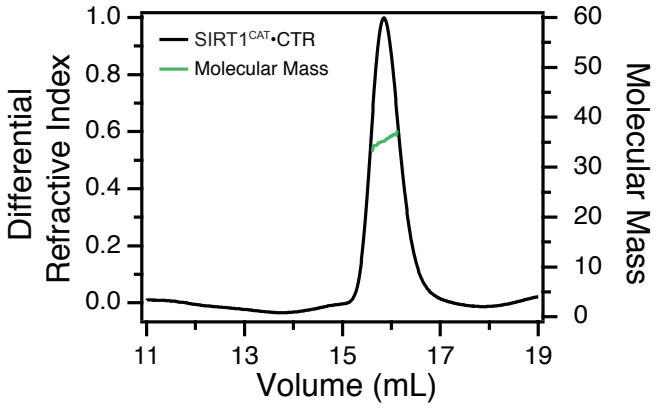
molecule 3

molecule 4

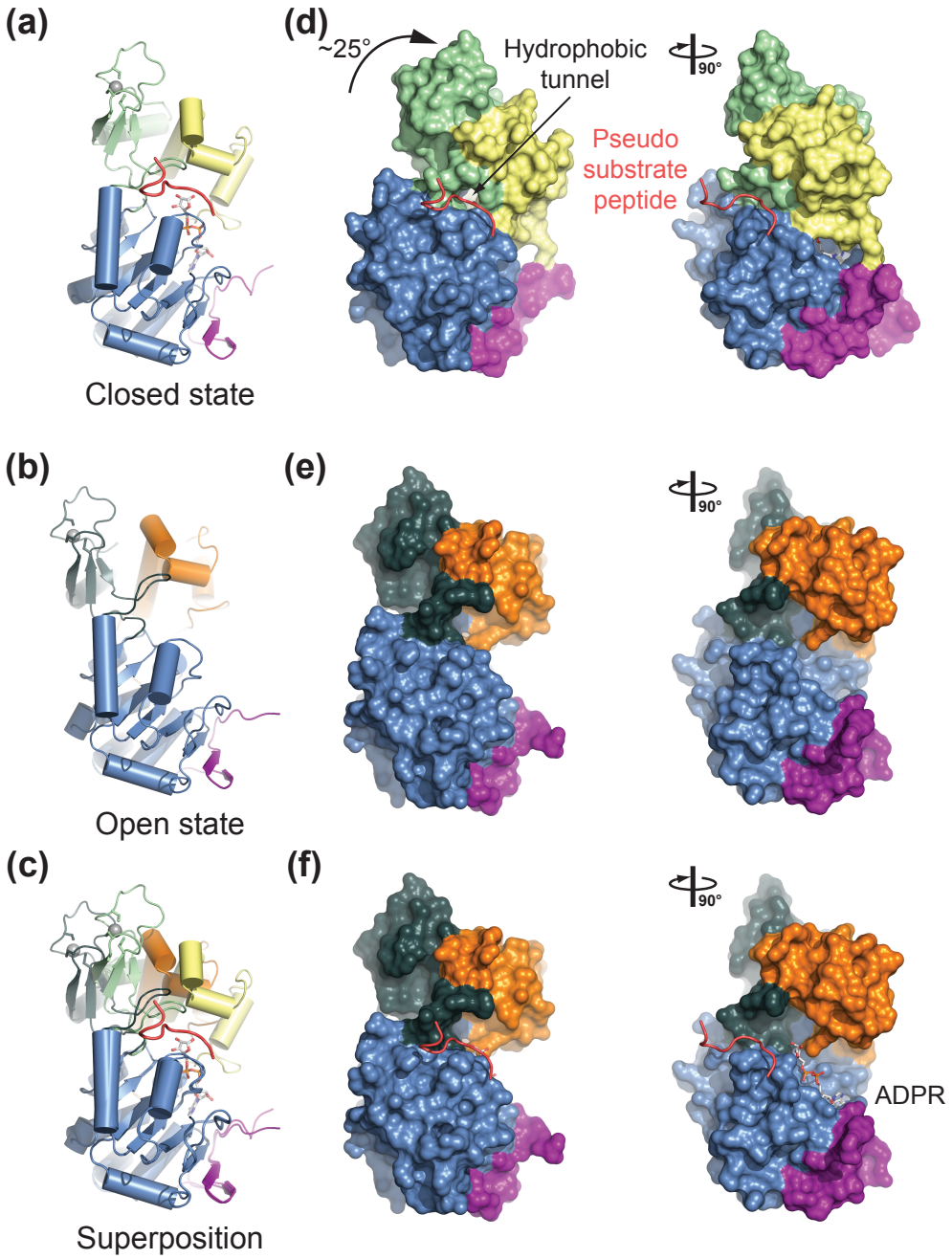
Chapter IV, Figure 6



Chapter IV, Figure 7



Chapter IV, Figure 8



Chapter IV, Figure 9

Table 1. Bacterial expression constructs

Protein	Residues (Mutations)	Expression Vector	Restriction Sites 5', 3'	N-terminal overhang
*hs*SIRT1[a]	234-510	pET28a-PreS	NdeI, NotI	GPHM
*hs*SIRT1	234-510	pGEX6P-1	BamHI, NotI	GPLGS
*hs*SIRT1	234-510 (S265A)	pET28a-PreS	NdeI, NotI	GPHM
*hs*SIRT1	234-510 (R276A)	pET28a-PreS	NdeI, NotI	GPHM
*hs*SIRT1	234-510 (Q345A)	pET28a-PreS	NdeI, NotI	GPHM
*hs*SIRT1	234-510 (N346A)	pET28a-PreS	NdeI, NotI	GPHM
*hs*SIRT1	234-510 (I347A)	pET28a-PreS	NdeI, NotI	GPHM
*hs*SIRT1	234-510 (D348N)	pET28a-PreS	NdeI, NotI	GPHM
*hs*SIRT1	234-510 (H363A)	pET28a-PreS	NdeI, NotI	GPHM
*hs*SIRT1	234-510 (F414A)	pET28a-PreS	NdeI, NotI	GPHM
*hs*SIRT1	234-510 (R466A)	pET28a-PreS	NdeI, NotI	GPHM
*hs*SIRT1	234-510 (D481A)	pET28a-PreS	NdeI, NotI	GPHM
*hs*SIRT1[a]	641-665	pGEX6P-1	BamHI, NotI	GPLGS
*hs*SIRT1	641-665	pET28b-SUMO	BamHI, NotI	His_6-SUMO
*hs*SIRT1	641-653	pGEX6P-1	BamHI, NotI	GPLGS
*hs*SIRT1	641-665 (E656A)	pGEX6P-1	BamHI, NotI	GPLGS
*hs*SIRT1	641-665 (Y642A)	pET28b-SUMO	BamHI, NotI	His_6-SUMO
*hs*SIRT1	641-665 (Y642D)	pET28b-SUMO	BamHI, NotI	His_6-SUMO
*hs*SIRT1	641-665 (F644A)	pET28b-SUMO	BamHI, NotI	His_6-SUMO
*hs*SIRT1	641-665 (F644D)	pET28b-SUMO	BamHI, NotI	His_6-SUMO
*hs*SIRT1	641-665 (Y650D)	pET28b-SUMO	BamHI, NotI	His_6-SUMO
*hs*SIRT1	641-665 (I651D)	pET28b-SUMO	BamHI, NotI	His_6-SUMO
*hs*SIRT1	641-665 (Y650A, I651A)	pET28b-SUMO	BamHI, NotI	His_6-SUMO
*hs*SIRT1	641-665 (Y650D, I651D)	pET28b-SUMO	BamHI, NotI	His_6-SUMO
*hs*SIRT1	641-665 (Y650A, F652A)	pET28b-SUMO	BamHI, NotI	His_6-SUMO
*hs*SIRT1	641-665 (Y650A, I651A, F652A)	pET28b-SUMO	BamHI, NotI	His_6-SUMO

[a]Constructs that were used for crystallization of the $SIRT1^{CAT}$•CTR heterodimer and the $SIRT1^{CAT}$•CTR•APDR•Substrate complex

Table 2. Purification yields

Residues (Mutations)	Expression Vector	Residues (Mutations)	Expression Vector	~ Yield (mg/18L)
234-510	pET28a-PreS			2
234-510 (R276A)	pET28a-PreS			1
		641-665	pET28b-SUMO	30
		641-665 (Y650D, I651D)	pET28b-SUMO	10
234-510	pET28a-PreS	641-665	pGEX6P-1	10
234-510 (S265A)	pET28a-PreS	641-665	pGEX6P-1	3
234-510 (R276A)	pET28a-PreS	641-665	pGEX6P-1	10
234-510 (Q345A)	pET28a-PreS	641-665	pGEX6P-1	10
234-510 (N346A)	pET28a-PreS	641-665	pGEX6P-1	0
234-510 (I347A)	pET28a-PreS	641-665	pGEX6P-1	1
234-510 (D348N)	pET28a-PreS	641-665	pGEX6P-1	0
234-510 (H363A)	pET28a-PreS	641-665	pGEX6P-1	10
234-510 (F414A)	pET28a-PreS	641-665	pGEX6P-1	2
234-510 (R466A)	pET28a-PreS	641-665	pGEX6P-1	3
234-510 (D481A)	pET28a-PreS	641-665	pGEX6P-1	3
234-510	pET28a-PreS	641-653	pGEX6P-1	3
234-510	pET28a-PreS	641-665 (E656A)	pGEX6P-1	3

Table 3. Crystallographic analysis

Data collection		
Protein	SIRT1•CTR•ADPR•Substrate	SIRT1•CTR *apo*
PDB code	4KXQ	4IF6
Synchrotron	SSRL	SSRL
Beamline	BL12-2	BL12-2
Space group	$P3_221$	$P4_32_12$
Cell parameters		
a, *b*, *c* (Å)	92.7, 92.7, 97.7	115.8, 115.8, 350.5
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 90.0
	Zn peak	*Zn peak*
Wavelength (Å)	1.28210	1.2676
Resolution (Å)	50.0 – 1.85	50.0 – 2.65
R_{sym} (%)[b]	4.3 (62.6)	8.8 (94.3)
$\langle I \rangle / \langle \sigma I \rangle$[b]	34.3 (2.4)	18.3 (1.8)
Completeness (%)[b]	100.0 (99.9)	97.1 (92.8)
No. observations	406,347	541,806
No. unique reflections	80,328 (8,010)	68,583 (6,409)
Redundancy	5.1 (4.7)	7.9 (7.8)
Refinement		
Resolution (Å)	50.0 – 1.85	50.0 – 2.65
No. reflections total	80,282	67,954
No. reflections test set	3,824 (4.76%)	1982 (2.92%)
R_{work} / R_{free} (%)	16.8 / 18.9	23.0 / 26.5
No. atoms	2,842	9,424
Protein	2,484	9,374
Ligand	47	4
Water	311	46
B-factors		
Protein	39.4	30.3
Ligand	33.8	60.6
Water	45.1	8.2
R.m.s. deviations		
Bond lengths (Å)	0.009	0.002
Bond angles (°)	1.2	0.630
Ramachandran plot[c]		
Favored (%)	97.7	96.1
Additionally allowed (%)	2.3	3.9
Outliers (%)	0.0	0.0
MolProbity score	1.11	2.04

[a] SSRL, Stanford Synchrotron Radiation Lightsource
[b] Highest resolution shell is shown in parentheses
[c] As determined by MolProbity[62]

Bibliography of Published Material Included in Thesis

Davenport, A.M., Huber, F.M., Hoelz, A. (2014). Structural and functional analysis of human SIRT1. *Journal of Molecular Biology* **426(3)**, 526-541. http://www.sciencedirect.com/science/article/pii/S0022283613006359

Davenport, A.M., Collins, L.N., Chiu, H., Minor, P.J., Sternberg, P.W., Hoelz, A. (2014). Structural and functional characterization of the α-tubulin acetyltransferase MEC-17. *Journal of Molecular Biology* **426(14)**, 2605-2616. http://www.sciencedirect.com/science/article/pii/S0022283614002484