2021

The Molecular And Regulatory Mechanism Of Multi-Subunit N-Terminal Acetyltransferases

Sunbin Deng
University of Pennsylvania

The Molecular And Regulatory Mechanism Of Multi-Subunit N-Terminal Acetyltransferases

Abstract

N-terminal acetylation (NTA) is one of the most widespread protein modifications, which occurs on most eukaryotic proteins, but is significantly less common on bacterial and archaea proteins. This modification is carried out by a family of enzymes called N-terminal acetyltransferases (NATs). To date, 12 NATs have been identified, harboring different composition, substrate specificity, and in some cases, modes of regulation. In the first chapter, we review the molecular features of NATs.NatA/E, NatB and NatC, are multi-subunit enzymes, responsible for the majority of eukaryotic protein NTA. Their mechanisms of action and regulation remain poorly understood before this dissertation. In the second chapter, we determined the X-ray crystal structure of yeast NatA/Naa50 as a scaffold to understand coregulation of NatA/Naa50 activity in both yeast and human. We found that Naa50 makes evolutionarily conserved contacts to both the Naa10 and Naa15 subunits of NatA. These interactions promote catalytic crosstalk within the human complex, but do so to a lesser extent in the yeast complex, where Naa50 activity is compromised. Thirdly, we reported the Cryo-EM structures of human NatE and NatE/HYPK complexes and associated biochemistry. We revealed that NAA50 and HYPK exhibit negative impacts on their binding to NAA15 in vitro and in human cells by inducing NAA15 shifts in opposing directions. NAA50 and HYPK each contribute to NAA10 activity inhibition through structural alteration of the NAA10 substrate binding site. Fourthly, we reported the Cryo-EM structure of hNatB bound to a CoA-αSyn conjugate, together with structure-guided analysis of mutational effects on catalysis. This analysis revealed functionally important differences with human NatA and Candida albicans NatB, resolved key hNatB protein determinants for αSyn N-terminal acetylation, and identified important residues for substrate-specific recognition and acetylation by NatB enzymes. Lastly, we report the Cryo-EM structure of S. pombe NatC with a NatE/C-type bi-substrate analogue and inositol hexaphosphate (IP6), and associated biochemistry. We find that all three subunits are prerequisite for normal NatC acetylation activity, IP6 binds tightly to NatC to stabilize the complex, and we determine the molecular basis for IP6-mediated stability of the complex and the overlapping yet distinct substrate profiles of NatC and NatE.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Chemistry

First Advisor

Ronen Marmorstein

Keywords

Catalysis mechanism, Cryo-EM, Crystallography, N-terminal acetylation, N-terminal acetyltransferase

Subject Categories

Biochemistry

THE MOLECULAR AND REGULATORY MECHANISM OF MULTI-SUBUNIT N-TERMINAL ACETYLTRANSFERASES

Sunbin Deng

A DISSERTATION

in

Chemistry

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2021

Supervisor of Dissertation

_____

Dr. Ronen Marmorstein

George W. Raiziss Professor of Biochemistry and Biophysics

Graduate Group Chairperson

_____

Dr. Daniel J. Mindiola

Brush Family Professor of Chemistry

Dissertation Committee

Dr. Emmanuel Skordalakes, Associate Professor of Chemistry, Committee Chair

Dr. Barry S. Cooperman, Professor of Chemistry

Dr. Vera Moiseenkova-Bell, Associate Professor of Pharmacology

Dedicated to my wife, Yiming Xu

ACKNOWLEDGMENTS

I want to begin this thesis with deep gratitude to my incredible mentor Dr. Ronen Marmorstein. As I always feel it, the best decision I have ever made in graduate school is that I joined his lab. Nothing could be accomplished by a novice researcher like me, without his kindness, patience, belief, effort and support in me. For the past years, I can't remember how many times I have walked into his office, and bothered him with my experiments, presentations, papers, even my personal life troubles. Ronen helped me immensely and words are clearly not enough to describe here what Ronen has taught me and how he influenced me as a mentor.

I began to work with Dr. Robert S. Magin right after Ronen took me in and Rob taught me a lot about biochemistry and protein crystallography. Since then, I also received very warm welcomes from Dr. Ryan P. Emptage, Dr. Austin Vogt, Dr. Dan Ricketts, Dr. Michael Grasso, Kevin Damico, Dr. Gleb Bazilevsky, Dr. Adam Olia, Angel Payan, Dr. Joe Han and Dr. Julie Barber-Rotenberg. I am grateful that they helped me numerous times with all kinds of problems I was having when I started. We had so much laughter together that I always keep those fun times in my heart and miss you a lot. I also want to thank Rob and Dr. Leah Gottlieb for their previous work on NATs, which made several of my projects possible. Other people in the lab supported me later. I worked with Dr. Xuepeng Wei to learn Cryo-EM techniques. Xuepeng was very helpful and almost taught me everything I know about Cryo-EM. Sarah Gardner and Julianna Supplee are two terrific rotation students, and Elaine Zhou, Meng Li, Mary Szurgot, Kollin Schultz, and Dr. Andrea Acevedo were great lab mates. I have learned so much from them and I thank them for putting up with me.

To my dissertation committee, Dr. Emmanuel Skordalakes, Dr. Barry S. Cooperman, Dr. Vera Moiseenkova-Bell, thank you so much for the great suggestions and advice. To Buyan Pan, Nina Mc Tiernan, Dr. Youchao Deng, Janice Reynaga, Dr. E. James Petersson, Dr. Thomas Arnesen, Dr. Rong Huang, Dr. Benjamin A. Garcia, you are amazing collaborators and I'm so glad I have worked with you.

I also want to thank many others for their help before the start of my graduate school career. My unusual path to this science journey would not have been possible without the generous help from Mrs. Yunzhen Jiang. She had the belief in me and encouraged me a lot when I first came up with the thought of pursuing studies in the biological sciences. I also feel extremely thankful and lucky to have the support of Dr. Alice Haddy and Dr. Jerry Walsh. And, to Jian-gang, thank you for being a long-time great friend and always a very good listener. To my parents and family, thank you for raising me.

Lastly, my wife has been my lucky star ever since we met in 2010. I was once a lost soul wandering without a life purpose after the failure in my first college experience, and you helped me rediscover my passion. Thank you for staying with me wherever I go, for supporting me whatever I decide, for pushing me to do my best, for making me laugh when I'm sad, for developing our mutual hobbies, and for growing old with me. I have become a better man because you are in your life. You make me feel life is too short to squander, and I will continue to keep you happy.

ABSTRACT

THE MOLECULAR AND REGULATORY MECHANISM OF MULTI-SUBUNIT N-TERMINAL ACETYLTRANSFERASES

Sunbin Deng

Ronen Marmorstein

N-terminal acetylation (NTA) is one of the most widespread protein modifications, which occurs on most eukaryotic proteins, but is significantly less common on bacterial and archaea proteins. This modification is carried out by a family of enzymes called N-terminal acetyltransferases (NATs). To date, 12 NATs have been identified, harboring different composition, substrate specificity, and in some cases, modes of regulation. In the first chapter, we review the molecular features of NATs.

NatA/E, NatB and NatC, are multi-subunit enzymes, responsible for the majority of eukaryotic protein NTA. Their mechanisms of action and regulation remain poorly understood before this dissertation. In the second chapter, we determined the X-ray crystal structure of yeast NatA/Naa50 as a scaffold to understand coregulation of NatA/Naa50 activity in both yeast and human. We found that Naa50 makes evolutionarily conserved contacts to both the Naa10 and Naa15 subunits of NatA. These interactions promote catalytic crosstalk within the human complex, but do so to a lesser extent in the yeast complex, where Naa50 activity is compromised.

Thirdly, we reported the Cryo-EM structures of human NatE and NatE/HYPK complexes and associated biochemistry. We revealed that NAA50 and HYPK exhibit negative impacts on their binding to NAA15 in vitro and in human cells by inducing

NAA15 shifts in opposing directions. NAA50 and HYPK each contribute to NAA10 activity inhibition through structural alteration of the NAA10 substrate binding site.

Fourthly, we reported the Cryo-EM structure of hNatB bound to a CoA-αSyn conjugate, together with structure-guided analysis of mutational effects on catalysis. This analysis revealed functionally important differences with human NatA and *Candida albicans* NatB, resolved key hNatB protein determinants for αSyn N-terminal acetylation, and identified important residues for substrate-specific recognition and acetylation by NatB enzymes.

Lastly, we report the Cryo-EM structure of *S. pombe* NatC with a NatE/C-type bi-substrate analogue and inositol hexaphosphate (IP_6), and associated biochemistry. We find that all three subunits are prerequisite for normal NatC acetylation activity, IP_6 binds tightly to NatC to stabilize the complex, and we determine the molecular basis for IP_6-mediated stability of the complex and the overlapping yet distinct substrate profiles of NatC and NatE.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Protein N-Terminal Acetylation: Structural Basis, Mechanism, Versatility, and Regulation

1.1 Evolutionarily conserved NTA has diverse biological functions

The majority of eukaryotic proteins are subject to several N-terminal modifications at the early stage of their biogenesis by the ribosome, including Nt-methionine excision (NME), Nt-acetylation (NTA), Nt-myristoylation (MYR), Nt-methylation and Nt-arginylation (Eldeeb and Ragheb 2018; Varland, Osberg, and Arnesen 2015; Giglione, Fieulaine, and Meinnel 2015). The evolutionarily conserved and irreversible NME and NTA modifications affect the functions of many proteins and increase the diversity of the proteome (Varland, Osberg, and Arnesen 2015; Giglione, Fieulaine, and Meinnel 2015). NTA occurs when the iMet is removed or retained, co-translationally or post-translationally (Aksnes, Ree, and Arnesen 2019; Aksnes et al. 2016; Linster and Wirtz 2018) and on 50-90% of eukaryotic proteins but only 10-29% of bacterial and archaea proteins (Arnesen et al. 2009; Bienvenut et al. 2012; Kelkar et al. 2011; Bienvenut, Giglione, and Meinnel 2015; Schmidt et al. 2016; Kirkland et al. 2008; Aivaliotis et al. 2007), correlating the abundance of NTA with organism complexity.

NTA changes the chemical properties of the protein N-termini: neutralizing the charge, creating a new h-bond acceptor, changing the α-amino nitrogen nucleophilicity and basicity, and increasing its hydrophobicity and size. These changed chemical properties have diverse biological consequences on protein function including protein half-life, folding, complex formation, and localization (Aksnes et al. 2016; Oh, Hyun, and Varshavsky 2017). Correlating with these diverse functions, knockout of individual eukaryotic N-terminal acetyltransferases (NATs) display different phenotypes, presumably associated with the differential misregulation of their respective client proteins (Aksnes et al. 2016). Alteration in NTA activity is also linked to disease such as various developmental and neurodegenerative disorders and cancers (Myklebust, Stove,

and Arnesen 2015; Kalvik and Arnesen 2013; Dorfel and Lyon 2015) in humans and stress response in plants (Linster and Wirtz 2018; Huber et al. 2020). While several other reviews cover the biology of NATs (Aksnes, Ree, and Arnesen 2019; Ree, Varland, and Arnesen 2018; Aksnes et al. 2016; Dorfel and Lyon 2015) , this chapter will focus on their structure, mechanism and regulation over evolution.

1.2 General molecular features of NATs

The **GCN5 related N-acetyltransferases (GNAT)** superfamily of proteins is one of the largest protein families, with over 10,000 protein members present in all domains of life. These proteins catalyze a bi-reactant process: transferring an acetyl group from the donor acetyl-CoA cofactor to the primary amine group of a variety of biomolecules including small molecules and protein or peptides lysine sidechains and N-termini. GNATs share a common structurally conserved α/β fold, despite their low degree of sequence identity; and mode of acetyl-CoA binding (Salah Ud-Din, Tikhomirova, and Roujeinikova 2016). GNATs typically contain distinct flanking N and C terminal regions that directly contributing to their distinct functions.

NTA is carried out by the NAT subfamily of GNATs, which typically contain four α-helices and seven β-strands, although addition secondary structural elements are often present at their N or C termini (**Figure 1.1A**). β1-β4 and β5-β6 are arranged antiparallelly, while the parallel β4 and β5 strands split to create a splay in between for binding peptide and acetyl-CoA (**Figure 1.1A).** The peptide binding site of NATs is typically flanked by the β6-β7 and α1-α2 loops, which directly participate in N-terminal substrate binding and take on a relatively closed configuration relative to GNATs that

acetylate internal lysine substrates, thereby disfavoring lysine sidechain binding (Magin, March, and Marmorstein 2016; Evjenth et al. 2009; Abboud et al. 2020). NATs typically make extensive h-bond interactions with the backbone atoms of the first 2-3 N-terminal residues further contributing to specificity for protein N-termini. These two distinct features distinguish NATs from **KATs (lysine acetyltransferases)** and other members of GNATs. NTA by NATs follows an ordered Bi-Bi reaction mechanism, where acetyl-CoA cofactor binding promotes subsequent N-termini binding prior to direct acetyl-group transfer from the cofactor to protein (Evjenth et al. 2012). One or two residues function as a general base (s) (often through an intervening water molecule) to deprotonate the terminal amino group to facilitate its nucleophilic attack of the acetyl group of acetyl-CoA (**Figure 1.1B**). This reaction proceeds through a tetrahedral intermediate, thus making CoA-peptide conjugate bi-substrate analogues in which N-terminal peptides are linked to CoA through an acetaldehyde group potent NAT inhibitors (**Figure 1.1C**) (Foyn et al. 2013).

A wealth of mechanistic information about the substrate specific activities of NATs has been derived from their structures and associated biochemical studies. The NATs that have been structurally characterized to date are listed in **Table 1.1.** As described below, each NAT employs a unique strategy to achieve substrate-specific acetylation and regulation.

Table 1.1 N-terminal acetyltransferases are characterized in all domains of life

	NATs	Subunits	Substrates specificity	PDB	
Bacteria	**RimI**	monomer	Ribosomal protein S18, relaxed	*S. typhimurium*	2CNM
	RimL	homodimer	Ribosomal protein L7/L12	*S. typhimurium*	1S7N
	RimJ	monomer	Ribosomal protein S5, relaxed	*Mycobacterium*	6C32
Archaea	**ssNAT**	monomer	relaxed	*S. solfataricus*	4LX9
Eukary-otes	**NAA10**	monomer	Acidic substrates	*S. pombe*	4KVX
	NatA	NAA10, NAA15	A-, S-, T-, V-, G-	*S. pombe* *Human*	4KVM 6C9M
	NatA/HYPK	NAA10, NAA15, HYPK		*C. thermophilum* *Human*	5NNR 6C95
	NatB	NAA20, NAA25	M-D/E/N/Q-	*C. albicans* *Human*	5K18 6VP9
	NatC	NAA30, NAA35, NAA38	M-L/I/F/Y/K-	*S. cerevisiae* *S. pombe*	6YGD 7L1K
	NatD	NAA40, monomer	Histones H2A/H4 (S-G-R-G-)	*S. pombe* *Human*	4UA3 4U9W
	NAA50	monomer	M-S/T/A/V/L/I/F/Y/K-	*Human*	3TFY[c,e]
	NatE	NAA50, NAA10, NAA15	NatA- and NAA50-type	*S. cerevisiae* *Human*	6O07[c] 6PPL
	NatE/HYPK	NAA50, NAA10, NAA15, HYPK		*Human*	6PW9
	NatF	Homodimer/monomer	M-L/I/F/Y/K-	*Human* *A. thaliana*	5ICW,5ICV 6TGX
	NatG	NAA70	M-, A-, S-, T-	-	

	NatH	NAA80, monomer	Actins (D/E-D/E-D/E-)	*Drosophila*	5WJE
	NatH/profilin /actin	NAA80, profilin, actin		*Human*	6NBE

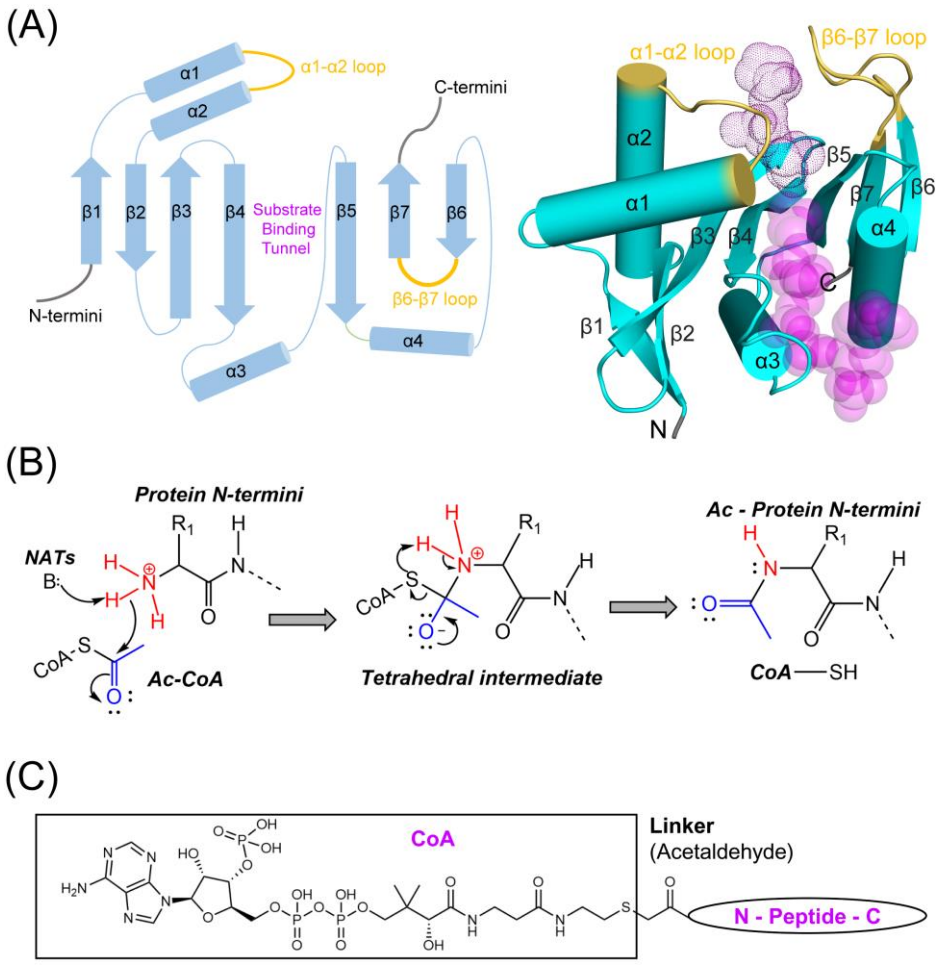


Figure 1.1 NAT catalytic domains share common topology and related catalytic mechanism

(A) The general topology of NAT catalytic subunits is depicted as in 2D cartoon on the left and 3D on the right panel with secondary structures shown. NATs usually contain seven β strands and four helices, but additional secondary structural elements within this topology and at their N-, C- termini are sometimes present. The α1- α2 and β6-β7 substrate binding loops are highlighted in yellow. The length of α helices and β strands

does not accurately reflect actual scale in the 2D representation. The transparent magenta dots and spheres in the 3D representation represent the peptide and acetyl-CoA substrates, respectively. The 3D representation was generated using human NAA50 PDB: 3TFY. (B) General catalytic mechanism of NATs, which transfer an acetyl group from acetyl-CoA to protein Nt-amino group is depicted. The acetyl and Nt-amino groups are colored as blue and red, respectively. A general base (or two, sometimes through a coordinated water molecule) is utilized to deprotonate the protein Nt-amino group, which subsequently attacks the acetyl group to form a tetrahedral intermediate. The deprotonated CoA is then deprotonated by a general acid (not shown) and released as the tetrahedral intermediate collapses. (C) The chemical structure of a CoA-peptide conjugate Bi-substrate analogue is shown. A linker (acetaldehyde group) is used to covalently link the CoA and peptide substrate. Abbreviations: NAT, N terminal acetyltransferase; CoA, Coenzyme A; PDB, Protein Data Bank.

1.3 Bacterial NATs

The three *E. coli* homologs, RimI, RimJ and RimL were among the first characterized NATs, showing post-translational NTA activity towards the ribosomal proteins S18, S5, and L7/L12, respectively (Favrot, Blanchard, and Vergnolle 2016). Some recombinant proteins overexpressed *in E coli* were later found to undergo NTA (Bernal-Perez et al. 2012), and more recent proteome-wide analysis found that ~30-100 *E. coli* N-termini are at least partially acetylated, predominantly at N-terminal serine, alanine, methionine and threonine residues (Bienvenut, Giglione, and Meinnel 2015; Schmidt et al. 2016). The abundance of NTA in *Pseudomonas aeruginosa PA14* and *Mycobacterium tuberculosis* appears to be in the range of 10% ~ 29% (Kelkar et al.

2011; Ouidir et al. 2015), although their biological importance have not been rigorously evaluated.

Bacterial orthologs of each of the Rim proteins have been structurally characterized. The *Salmonella typhimurium* RimI (*St*RimI) structure was determined in complex with CoA, acetyl-CoA or a CoA-peptide conjugate showing a typical GNAT fold (**Figure 1.2A)** (Vetting et al. 2008). Several *St*RimI residues change conformation when in complex with the peptide portion of the conjugate with several residues making backbone h-bonds to carbonyl and nitrogen atoms of the first three residues of the peptide portion, while several RimI residues make van der Waals and h-bonds to side chains 2-4 (**Figure 1.3**). RimI Glu103 is proposed to function as a general base through an intervening water molecule, while Tyr115 is proposed to act as the general acid (**Figure 1.3)**. A more recent biochemical characterization of *Mycobacterium tuberculosis* RimI demonstrated a conventional eukaryotic NatA/NatC/NatE substrate preference, consistent with a relaxed substrate specificity for RimI (Pathak et al. 2016).

Crystal structures of *St*RimL in apo form and in complex with CoA reveals a similar overall fold to bacterial *St*RimI, however *St*RimL forms a homodimer through antiparallel b-strands (b6) from opposing subunits to form a contiguous b-sheet that spans the entire dimer (**Figure 1.2B)** (Vetting et al. 2005). This 2-fold symmetry appears to be precisely configured to acetylate its cognate dimeric L7 N-termini (Wahl et al. 2000; Bocharov et al. 2004). Compared to the apo *St*RimL structure, the bound of CoA appears to stabilize the α1-α2 loop, which could also facilitate the cooperative binding of the dimeric L7/L12. Like *St*RimI, *St*RimL is proposed to use a ternary complex mechanism involving a glutamate general base residue (Glu160), while a tyrosine

(Tyr98) is proposed to stabilize the tetrahedral intermediate, with a serine residue (Ser141) functioning as a general acid.

Mycobacterium smegmatis RimJ structures alone and in complex with cofactor or cofactor analogs has recently been deposited to the PDB (PDB: 6C32, 6C30, 6C37. Taken together, the bacterial NAT proteins reveal a general GNAT fold in either monomeric or homodimeric form with substrate protein recognition occurring through backbone and sidechain interaction to the first few N-terminal residues, with acid-base catalysis proceeding through dedicated residues.

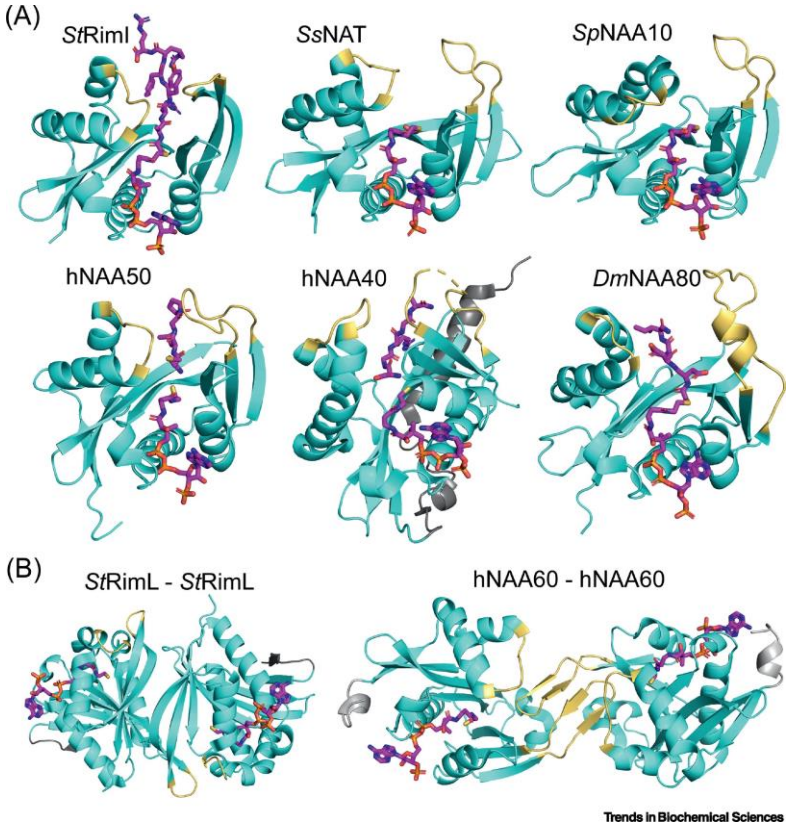


Figure 2.2 Some NATs function independently, either in monomeric or homodimer form

(A) Structures of monomeric StRimI (PDB: 2CNM), SsNAT (PDB: 4LX9), uncomplexed SpNAA10 (PDB: 4KVX), hNAA50 (PDB: 3TFY), hNAA40 (PDB: 4U9W), and DmNAA80 (PDB: 5WJE) are shown in cartoon and color in cyan. The α1- α2 and β6-β7 substrate binding loops are highlighted in yellow. Substrates in the structures are shown in stick and colored in magenta. The hNAA40-specific N terminal domain is highlighted in grey. (B) Dimeric RimL (PDB:1S7N) and hNAA60 (PDB: 5ICW) are shown. The α1-α2 and β6-β7 substrate binding loops are highlighted in yellow. Substrates in the structures are shown in stick and colored in magenta. To form a dimer, StRimL utilizes the two β6 strands from each subunit, while hNAA60 uses the extended β6-β7 loops. The NAA60-specific N terminal domain is highlighted in grey. NAT catalytic subunits that have not been shown to function independently are not shown.

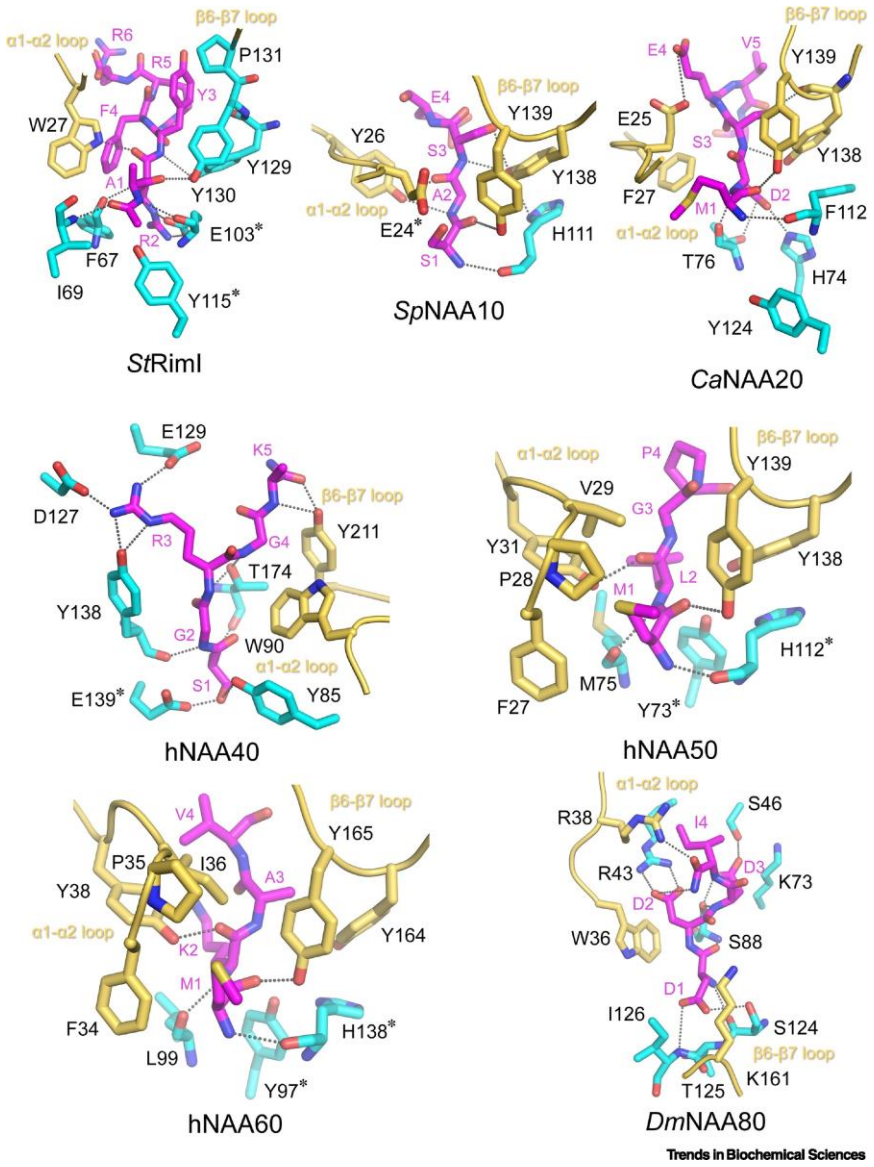


Figure 3.3 NATs use related but distinct mechanism to recognize their peptide substrate N-termini

Peptide binding sites of StRimI (PDB: 2CNM), complexed SpNAA10 (PDB: 4KVM, with SpNAA15 hidden), CaNAA20 (PDB: 5K04, with CaNAA25 hidden), hNAA40 (PDB: 4U9W), hNAA60 (PDB: 5ICV), and DmNAA80 (PDB: 5WJE) are shown in cartoon. Peptide substrates are shown in magenta sticks. The residues labeled with a * symbol are proposed catalytic residues. Broken lines indicate h-bonds formed between atoms.

The α1- α2 and β6-β7 substrate binding loops are highlighted in yellow. Water-mediate interactions in the PDB structures are not shown.

1.4 Archaea NATs

Like in bacteria, while earlier studies indicated that NTA occurs on only ribosomal proteins in archaea, more recent N-terminomics studies reveal that NTA occurs on 14–29% of archaeal proteins (Falb et al. 2006; Aivaliotis et al. 2007; Kirkland et al. 2008). Archaea contains only one conserved NAT, which exhibits a very relaxed substrate spectrum including substrates by eukaryotic NatA/B/C/E (Mackay et al. 2007; Liszczak and Marmorstein 2013). The *Sulfolobus solfataricus* NAT (*Ss*NAT) shows the greatest sequence conservation with the NAA10 catalytic subunit of the eukaryotic NatA complex, with 33% sequence identity (Liszczak and Marmorstein 2013), as confirmed by its structure bound to acetyl-CoA (PDB: 4LX9) (**Figure 1.2A**) (Liszczak and Marmorstein 2013). Interestingly, the monomeric *Ss*NAT is more structurally similar to the active form of Naa10 within the binary NatA complex, rather than NAA10 alone, which is not active towards NatA substrates. A notable place where *Ss*NAT and NAA10 diverge is in the C-terminal end of the a1 helix that is part of the a1-a2 loop substrate binding region of NATs. This region of *Ss*NAT is more similar to the eukaryotic NAA50, which is also active towards NatE substrates as a monomer. This suggests that *Ss*NAT is a hybrid of the NAA10 and NAA50 eukaryotic NATs. This is consistent with biochemical studies on *Ss*NAT demonstrating that different catalytic residues are used to acetylate NatA (Glu84) and NAA50/NatE (Tyr73/His112) -like substrates, analogous to corresponding residues in these eukaryotic NATs (Liszczak and Marmorstein 2013). It is possible that the archaea NAT serves as one of the ancestors of eukaryotic NATs, which subsequently

evolved to be substrate-specific. An unusually long β3-β4 loop in *Ss*NAT, was also shown to play a more selective role in acetylating NAA50-type substrates, although the mechanism for this is still unclear. A more recent *Ss*NAT structure (called Ard1) (Chang and Hsu 2015) and *Thermoplasma volcanium* Ard1 (Ma et al. 2014), were consistent with the earlier study, reinforcing the substrate flexibility of the archaeal NATs. Interestingly, it was demonstrated that mutations of the catalytic Glu residue of ssNAT could shift its substrate preference (Chang and Hsu 2015).

1.5 Eukaryotic NATs

NatA: NatA acetylates ~ 40% of human proteins, the most of any NAT (Arnesen et al. 2009). NatA is a heterodimer of catalytic NAA10 (also named ARD1) and auxiliary NAA15 subunits (Mullen et al. 1989; Park and Szostak 1992). Notably, NAA10 can also exist independently (Arnesen et al. 2005; Van Damme, Evjenth, et al. 2011a). When the second residue of a protein is relatively small and uncharged (**Table 1.1),** it will be acetylated by NatA after the initiator methionine is cleaved (Polevoda et al. 1999; Arnesen et al. 2009). In contrast, the NAA10 monomer is not able to acetylate canonical NatA substrate, but instead exhibits specificity for acidic N-terminal sequences in vitro (Van Damme, Evjenth, et al. 2011a; Drazic et al. 2018; Goris et al. 2018).

The structure of *Schizosaccharomyces pombe* NatA (*Sp*NatA) bound to a CoA-peptide conjugate reveals that the *Sp*NAA15 auxiliary subunit adopt a ring-like structure of tetratricopeptide repeat (TPR) motifs, which completely wraps around the *Sp*NAA10 catalytic subunit, with the two proteins making extensive hydrogen bonding and van der Waals interactions (**Figure 1.4A)** (Liszczak et al. 2013). The *Chaetomium thermophilum*

(Weyer et al. 2017) and *Saccharomyces cerevisiae* (Deng et al. 2019) NatA display similar overall folds. A structural comparison of the *Sp*NatA complex with *Sp*NAA10 alone, reveals that *Sp*NAA15 binding to *Sp*NAA10 induces a conformational change of the a1-a2 loop region of *Sp*Nat10 to position key catalytic residues for substrate-specific recognition and catalysis (Liszczak et al. 2013). The *Sp*NAA10 active site reveals that h-bonds to the backbone carbonyl groups of the first two residues and van der Waals contacts to Ser1 anchor the peptide portion of the CoA-peptide conjugate (**Figure 1.3**). Glu24 was also proposed to play role as a general base for catalysis. Each of the residues revealed to play important roles in substrate binding and catalysis are highly conserved within NatA orthologs and shown to result in decreased NatA activity when mutated (Liszczak et al. 2013). Importantly, modeling of amino acid side chains in the Ser1 position, reveals that only small uncharged amino acids could be accommodated, consistent with the known substrate specificity of NatA (Liszczak et al. 2013). Interestingly, this same study showed that a non-cognate EEE peptide could be acetylated by the *Sp*NAA10-E24A mutant but not by WT *Sp*NatA, suggesting that free NAA10 uses a different catalytic strategy to acetylate NatH-type substrates.

The structure of the human NatA (hNatA) shows a high degree structural conservation with *Sp*NatA (Gottlieb and Marmorstein 2018), except for the presence of an extended metazoan-specific Sel1-like repeat region at the C-terminal end of the hNAA15 auxiliary subunit, which was shown to play a role in overall hNatA stability (Gottlieb and Marmorstein 2018). In addition, the hNatA structure reveals the presence of an endogenously bound inositol hexaphosphate (IP_6) molecule at an interface region between hNAA10 and hNAA15 via electrostatic interactions (Gottlieb and Marmorstein 2018; Deng, McTiernan, et al. 2020). A recently identified hNatA mutant K450E displays

defects in IP_6 binding and leads to NatA activity loss which can be rescued by the addition of IP_6 (Cheng et al. 2019), consistent with a role of IP_6 in hNatA stability and activity.

NatB: NatB, conserved from yeast to man, acetylates ~21% of human proteins, such as actin, tropomyosin, CDK2, and α-synuclein. NatB is a binary complex containing catalytic NAA20 and auxiliary NAA25 subunits and has specificity for N termini containing MD-, ME-, MN- and MQ- sequences (Van Damme et al. 2012). The structure of *Candida albicans* NatB (*Ca*NatB) was determined alone and in complex with a CoA-peptide conjugate (Hong et al. 2017). The catalytic *Ca*Naa20 subunit shows a typical GNAT fold, and the *Ca*NAA25 auxiliary subunit also displays high structural similarity to *Sp*NAA15, particularly in the way it wraps around the catalytic subunit (**Figure 1.4A)**. Unlike NAA10, NAA20 is unstable in the absence of NAA25. Extensive interactions are observed between *Ca*NAA25 and *Ca*NAA20, with the major contact interface mediated by the α1- α2 loop of *Ca*NAA20 (Hong et al. 2017), similar to the NatA complex. The interaction between NAA20 and the peptide substrate mainly involves h-bond interactions to the backbone amides of the first three residues of the peptide portion of the CoA-peptide conjugate, with side chain contacts to residues Met1 and Asp2 through van der Walls and h-bond interactions (**Figure 1.3**). Mutation of contact residues generally reduce NatB acetylation activity, consistent with the structural observations (Hong et al. 2017). A more recent Cryo-EM structure of human NatB bound to a CoA-a-synulein peptide conjugate shows a high degree of structural similarity to *Ca*NatB, although provides structural and biochemical data consistent with a role for Tyr124 (hNAA20-Tyr123) acting as a general base for catalysis (Deng, Pan, et al. 2020).

NatC: NatC is a heterotrimer of a catalytic subunit NAA30 and two auxiliary subunits NAA35 and NAA38 (Polevoda and Sherman 2001). NatC acts co-translationally on the peptide substrate staring with ML-, MF, MI, MW-, playing roles in viral particle assembly, maintaining mitochondrial integrity and apoptosis (Van Damme et al. 2016a; Starheim et al. 2009). Recent x-ray crystal structures of *S. cerevisiae* NatC with cofactor and ligand bound revealed that it has distinct divergent architecture compared to other NATs (Grunwald et al. 2020). Overall, the larger auxiliary subunit NAA35 serves as the central assembly hub to form extensive interaction with the smaller auxiliary subunit NAA38 and NAA30. NAA38 displays an Sm fold, consisting of an α-helix followed by a five-strand β-sheet, and interacts with the N-terminal region of NAA35. The absence of NAA38 compromises NatC activity in vitro, indicating its indispensable role in the whole complex. NAA30 adopts a typical GNAT fold, however, unlike NatA or NatB, its α1–α2 and β6–β7 loops are both buried inside the interface between NAA30 and NAA35. Biochemical analysis suggested that NatC preferably recognizes its substrates not just by the identity of the second residue, but also residues beyond position 2 (Grunwald et al. 2020). It was proposed that *Sc*Naa30-Glu118 functions as the general base residue to deprotonate the substrate and *Sc*Naa30-Tyr130 serves as the general acid to re-protonate the initially formed thiolate anion of CoA (Grunwald et al. 2020). Our recent structural and biochemical studies on *Sp*NatC showed some distinct features but agreed mostly with the results on *Sc*NatC describe above (See **chapter 5**).

NatD: NatD (NAA40), which functions as a monomer, is conserved from yeast to human, and is the most selective NAT, carrying out NTA of only histones H4 and H2A (Song et al. 2003; Hole et al. 2011), which most commonly contain the sequence SGRGK. The structure of hNatD bound to acetyl-CoA and CoA and an N-terminal

histone H4/H2A peptide, reveals that while NatD adopts a GNAT fold, it also contains a unique N terminal helix-loop-strand segment, which wraps around the GNAT domain and plays an important role in hNatD stability, partially mimicking similar roles of the auxiliary subunits of NatA and NatB (Magin, Liszczak, and Marmorstein 2015) (**Figure 1.2A)**. Another unique feature of NatD is found in its substrate binding loops, where the α1- α2 loop is extended and flipped towards the peptide substrate, such that the opposing β6-β7 loop is flipped away (Magin, Liszczak, and Marmorstein 2015). This alters the peptide binding site to make it uniquely suited for its cognate N-termini. Indeed, the structure of hNatD bound to CoA and peptide reveals that nearly every hydrogen bond donor and acceptor atom within the first four residues are engaged in hydrogen bounding interactions, with Arg3 playing a particularly important role, and the small sizes of Gly2 and Gly4 also being critical (**Figure 1.3**). Activity assays also revealed that Glu139 is essential for catalyst, possibly acting as a general base (**Figure 1.3**).

NatE: NatE refers to the complex between the NAA50 catalytic subunit and the NatA complex (also containing the NAA10 catalytic subunit). Monomeric NAA50 is also stable. However, yeast NAA50 is inactive, while in higher eukaryotes, such as human and *Drosophila*, NAA50 is active in the absence or presence of NatA (Evjenth et al. 2009; Liszczak, Arnesen, and Marmorstein 2011; Van Damme et al. 2015; Rong et al. 2016; Deng et al. 2019). NAA50 has specificity for N-terminal Met residues followed by a less restricted set of amino acids at the second position (**Table 1.1**) (Evjenth et al. 2009) . The structure of hNAA50 bound to CoA and an N-terminal peptide reveals that the peptide is recognized almost exclusively through interactions to residues one and two (Liszczak, Arnesen, and Marmorstein 2011) (**Figures 1.2A and 1.3)**. While the carbonyl

and amino groups of residues Met1 and Leu2 engage in h-bond interactions, the side chains make extensive van der Walls interactions. Consistent with the known specificity of NAA50, the active site appears exquisitely suited for Met1, while the Val2 binding site appears to have greater flexibility to accommodate several amino acid substitutions (**Figure 1.3**). A combination of mutagenesis, kinetics and structural observations suggest that hNAA50 Tyr73 and His112 are essential for catalysis via an ordered water molecule, likely serving as general acid or base residues (Liszczak, Arnesen, and Marmorstein 2011) (**Figure 1.3**). Interestingly, these two catalytic residues of hNAA50 are absent in the corresponding positions of yeast homologous, which likely explains the inactivity of the yeast NAA50 proteins. A recent structure of *Sc*NatE and related biochemical studies also shows that *Sc*NAA50 has relatively poor acetyl-CoA binding capacity, consistent with its catalytic inactivity (Deng et al. 2019).

The binding affinity between NAA50 and NatA is in the nanomolar range (Deng et al. 2019), consistent with the fact that a significant portion of NAA50 in the cell is bound to NatA (Hou et al. 2007). The molecular basis for association between NAA50 and NatA from yeast and human were recently revealed (Deng et al. 2019; Deng, McTiernan, et al. 2020) , demonstrating that NAA50 docks onto a unique surface of the NAA15 auxiliary subunit primarily through hydrophobic interactions, while making significantly more modest electrostatic interactions with NAA10 (**Figure 1.4)**. Nonetheless biochemical experiments reveal that NAA10 and NAA50 can influence each other's activity within the NatE complex (Deng et al. 2019; Deng, McTiernan, et al. 2020) .

NatF: NatF (NAA60), found only in higher eukaryotes (Rathore et al. 2016; Van Damme, Hole, et al. 2011), is located on the cytosolic side of Golgi membranes to

acetylate transmembrane protein (Aksnes et al. 2015; Aksnes et al. 2017). It has an unique membrane binding domain at its C-termini (Aksnes et al. 2015; Chen et al. 2016; Stove et al. 2016; Aksnes et al. 2017), which mediates Golgi membrane binding (Aksnes et al. 2017). Recently, *Arabidopsis thaliana* NatF was demonstrated to associate with the plasma membrane also via its C-termini (Linster et al.). NatF has specificity for N-terminal Met residues followed by Leu, Ile, Phe, Tyr and Lys residues (**Table 1.1**). Unlike other NATs, NatF is homodimeric in solution but shifts to a monomer state when peptide substrate binds (**Figure 1.2B)** (Stove et al. 2016). The structure of hNatF in complex with CoA or a CoA-peptide conjugate reveals that it has a unique extended β -β7 loop, which mediates dimerization and sterically occludes peptide binding, but participates in peptide recognition of the NatF monomer when cofactor and peptide are both bound (Stove et al. 2016). hNatF peptide recognition in hNatF is mainly to the first two residues, through backbone h-bonds and a hydrophobic pocket (similar to NAA50) that recognizes Met1 with less specificity for residue Lys2 (**Figure 1.3**) (Stove et al. 2016). Tyr97 and His138 are proposed to function as general base residues in hNatF, similar to roles played by residues of Tyr73 and His112 from hNAA50 (Stove et al. 2016), which is consistent with mutagenesis studies (**Figure 1.3**) (Chen et al. 2016).

NatG: NatG is localized within the chloroplast of plant, processing M-, A-, S-, T-staring-termini as substrates (Dinh et al. 2015). As of the writing of this review, a detailed structure/function analysis of NatG had not yet been reported.

NatH: NatH (NAA80) is widespread only in animals (Drazic et al. 2018), and post-translationally active toward processed cytoplasmic β- and γ-actin *in vivo* with acidic N termini DDDI and EEEI, respectively (Drazic et al. 2018; Arnesen, Marmorstein, and Dominguez 2018; Wiame et al. 2018). The structure of *Drosophila melanogaster* Naa80

(*Dm*Naa80) bound to a CoA-peptide conjugate reveals a more open and highly basic substrate binding site than other NATs, which is specifically configured to bind its acidic substrates (Goris et al. 2018) (**Figure 1.2A**). H-bonds extend through the first three backbone and side-chain residues (**Figure 1.3**), explaining the specificity for highly acidic substrates. An unusual intramolecular hydrogen bond is also observed between the backbone amide nitrogen of I_4 and the sidechain of D_2 (**Figure 1.3**). Mutational and activity assays is consistent with substrate preference deriving mostly from the acidic residues at positions two and three (Goris et al. 2018), which is unusual for NAT proteins that typically specify residues one and two.

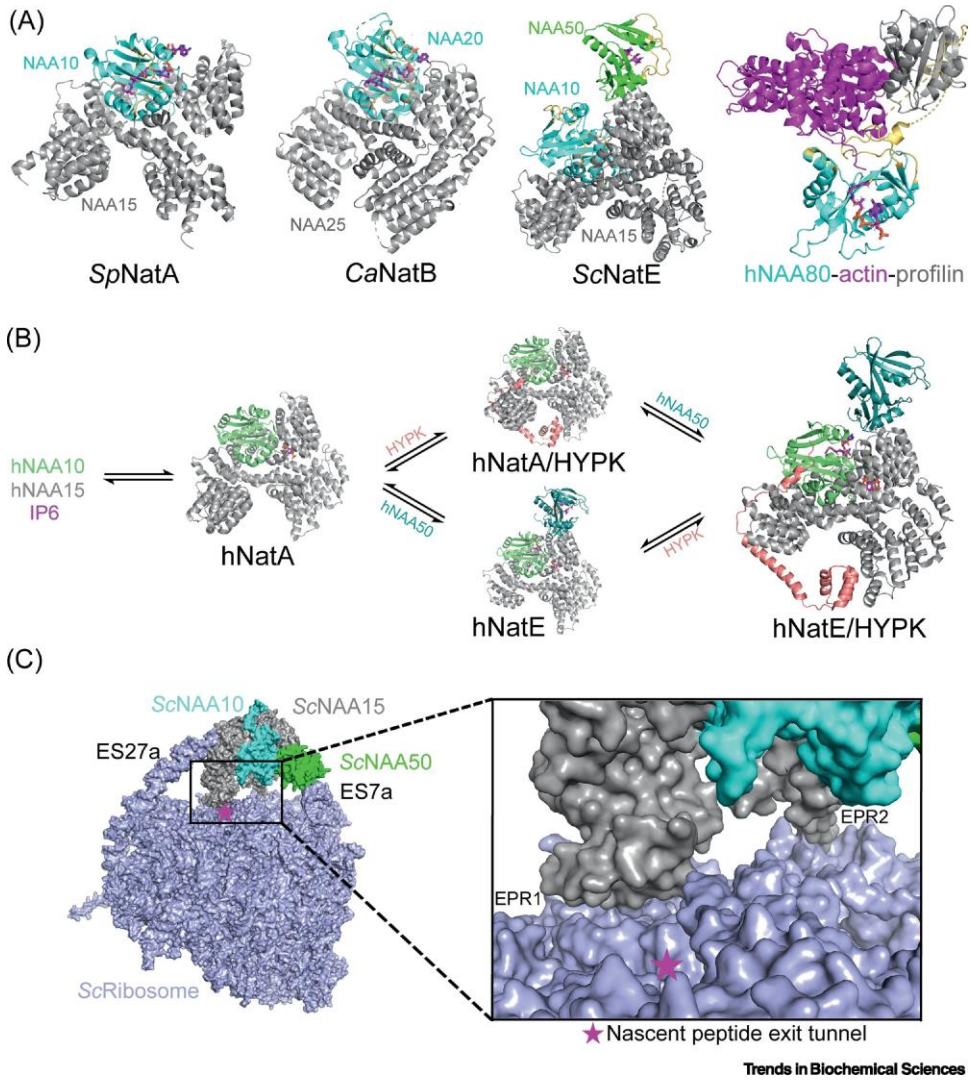


Figure 4.4 Some NATs function by forming complexes with auxiliary subunits or regulatory proteins

(A) Structures of SpNatA (PDB:4KVM), CaNatB (PDB:5K04), ScNatE (PDB: 6O07), and hNAA80-actin-profilin complexes (PDB:6NBE) are shown in cartoon. (B) In humans, the dynamics and interplay between hNatA, hNAA50, and HYPK are shown. hNAA10 can exist independently. Two subunits and Inositol hexaphosphate (IP_6) form hNatA (PDB: 6C9M) complex. HYPK and hNAA50 each can associate with hNatA to form competing complexes (hNatA/HYPK PDB: 6C95 and hNatE PDB:6PPL). Tetrameric complex

hNatE/HYPK (PDB: 6PW9) can be formed when both HYPK and hNAA50 are bound to hNatA. hNAA10, HYPK, hNAA50, and hNAA15 are shown in chartreuse, salmon, teal, and grey, respectively. (C) Structure of ScNatE bound to ribosome (PDB: 6HD7) is shown. Two rRNA expansion segments ES27a and ES7a are contacting NAA15 and NAA50, respectively. Two electropositive regions ERP1 (N terminus of NAA15), and EPR2 (internal basic helix) on NAA15, which directly contact ribosome, are shown in the zoom-in view.

1.6 Regulation of NATs

1.6.1 Huntingtin-Interacting Protein K regulates both NatA and NatE activity

The Huntingtin-Interacting Protein K (HYPK) **,** was found to associate with the human NatA complex (Arnesen et al. 2010). It is absent in most yeast , but found in *Chaetomium thermophilum* (Weyer et al. 2017) and human (Deng, McTiernan, et al. 2020; Gottlieb and Marmorstein 2018) where it was structurally characterized bound to NatA. In both cases, HYPK was shown to have intrinsic NatA inhibitory activity, and in the study of the human complex proposed to contribute to cognate substrate specificity as HYPK was shown to be partially uncompetitive and noncompetitive with respect to acetyl-CoA and peptide substrate, respectively (Gottlieb and Marmorstein 2018). When bound to NatA, HYPK forms a bipartite structure: a C-terminal ubiquitin-associated domain (a3-a5) binds extensively to the NAA15 auxiliary subunit, a N-terminal loop-a1-helix region binds across the catalytic NAA10 subunit to distort its active site, and a long a2 helix connects these two ends of HYPK. Biochemical data is consistent with the structural findings that the ubiquitin-associated domain forms the high affinity interaction,

while the loop-a1-helix region harbors the catalytic inhibitory activity. More recently, structural and biochemical studies reveal that HYPK can also form a stable complex with the NatE complex (**Figure 1.4B**), although HYPK and NAA50 where shown to allosterically reduce the binding affinity of each other (Deng, McTiernan, et al. 2020).

1.6.2 NAA80 activity is regulated by the actin chaperone profilin

Naa80 is preferentially active towards monomeric actin and the presence of profilin has been shown to increase its catalytic efficiency (Rebowski et al. 2020). The crystal structure of hNaa80 in complex with monomeric actin and profilin with cofactor and cofactor analogs reveals the first structure of a NAT bound to its intact protein substrate (Rebowski et al. 2020). The structure shows that a Naa80-specific extended β6-β7 proline-rich loop and α2-helix of hNAA80 is utilized to meditate interaction with both actin and profilin (Rebowski et al. 2020). Thus formation of hNAA80-actin-profilin complex orients the N termini of actin perfectly into the substrate binding groove of hNAA80 **(Figure 1.4A)** (Rebowski et al. 2020). It is proposed that, with the aid of profilin, hNAA80 can act more efficiently to compensate for the low abundance of hNAA80 relative to actin in the cell.

1.6.3 Ribosome association by NATs

In eukaryotes, NatA/B/C/E associate with the ribosome for co-translational activity, utilizing their auxiliary subunits as the predominant anchoring point (Polevoda et al. 2008), while NatD could potentially use its extended N terminal region (Magin,

Liszczak, and Marmorstein 2015; Hole et al. 2011). This proximity of the NATs to the ribosome likely facilitates the relatively high stoichiometry of NTA that is observed *in vivo*.

The current molecular understanding of the interactions between NATs and the ribosome is limited to NatA or NatE. Early studies suggested that NatA contacts ribosomal protein uL23 and uL29 around the peptide exit tunnel (Polevoda et al. 2008) to interact with the nascent chain (Gautschi et al. 2003). The ratio of ribosome to NatA is around 40:1 in yeast, leading to questions about the dynamics of NatA/ribosome interactions (Raue, Oellerer, and Rospert 2007). Purified non-translating ribosome can also interact with NatA *in vitro* via two conserved electropositive regions (EPR) on NAA15 (Magin et al. 2017), which is confirmed through mutational studies (Magin et al. 2017) and *in vivo* (Varland and Arnesen 2018). A recent structure of a *Sc*NatE/ribosome complex reveals that *Sc*NatE docks at the nascent peptide exit tunnel near uL31 and uL22, with rRNA expansion segments making key contacts to *Sc*NAA15 and *Sc*Naa50 (**Figure 1.4C**) (Knorr et al. 2019).

1.7 Remaining Questions

NTA occurs on countless proteins in three domains of life both co-translationally and post-translationally, impacting diverse cellular functions of their client protein. Although sharing an evolutionarily conserved GNAT fold and related catalytic mechanism, each NAT employs unique elements to mediate substrate-specific activity to carry out their distinct cellular functions. NAT auxiliary and regulatory subunits also play NAT-type specific roles. While some NATs have dedicated co-translational and post-

translational roles, the mechanisms that may toggle some NATs such as NAA10 and NAA50 between the two roles requires further investigation. Still missing from the PDB are the chloroplast localized NATs, which may have unique functions. Do other NAT agonists or antagonists exist and how might they work? Finally, a greater molecular understanding of how NATs carry out and navigate between each other co-translational NTA on the ribosome warrants further study (See **Chapter 6** for more details).

CHAPTER 2: Structure and Mechanism of Acetylation by the N-Terminal Dual Enzyme NatA/Naa50 Complex

2.1 Introduction

The majority of nascent peptides biosynthesized by the eukaryote cytosolic ribosome undergo several N-terminal co-translational modifications before they become fully functional and mature proteins. Along with initial methionine cleavage by methionine aminopeptidase (iMAP) and N-terminal myristoylation, N-terminal acetylation (Nt-acetylation), which transfers an acetyl moiety from acetyl-CoA to the N-terminal amino group of proteins, is an irreversible, highly conserved and abundant modification (Giglione, Fieulaine, and Meinnel 2015). About 50-70% of *S. cerevisiae* proteins, 70%-75% of *A. thaliana* proteins, and 70-90% of *H. sapiens* proteins are subject to Nt-acetylation by a family of N-terminal acetyl transferases (NATs) (Starheim, Gevaert, and Arnesen 2012). Nt-acetylation is generally considered to be a co-translational process, with most NATs making direct interactions with both nascent peptide and the ribosome (Gautschi et al. 2003; Polevoda et al. 2008; Magin et al. 2017; Knorr et al. 2019).

Numerous studies have demonstrated that Nt-acetylation can affect diverse protein and cellular activities including gene regulation, apoptosis, protein folding, protein degradation, protein complex formation and subcellular localization (Aksnes et al. 2016; Holmes et al. 2014; Shemorry, Hwang, and Varshavsky ; Behnia et al. 2004; Yang et al. 2013; Dikiy and Eliezer 2014; Hwang, Shemorry, and Varshavsky 2010; Schiza et al. 2013; Starheim et al. 2009; Arnesen, Gromyko, et al. 2006; Setty et al. 2004b; Pavlou and Kirmizis 2016). Knockout of NATs frequently results in severe cellular defects including slow growth and reduced mating efficiency in yeast. In higher eukaryotes, misregulation of N-terminal acetylation can lead to numerous developmental disorders and cancers (Kalvik and Arnesen 2013; Myklebust, Stove, and Arnesen 2015; Lee et al.

2010; Yu et al. 2009; Arnesen, Gromyko, et al. 2006; Ametzazurra et al. 2008; Starheim et al. 2008; Starheim et al. 2009; Fluge et al. 2002; Hua et al. 2011). For example, the NatA catalytic subunit, Naa10, plays important roles in mammalian cells including cellular hypoxia, bone formation and DNA damage (Dorfel and Lyon 2015; Jeong et al. 2002; Yoon et al. 2014; Yi et al. 2007). Mutation in human Naa10 causes the genetic disorder Ogden syndrome (Rope et al. 2011).

Several NATs, including NatA, NatB, NatC, NatD, NatE, and NatF are conserved across eukaryotes (Aksnes et al. 2016) (Aksnes et al. 2015; Aksnes et al. 2017; Stove et al. 2016). NatG is resident in chloroplasts of plant cells (Dinh et al. 2015), and the recently identified NatH/Naa80 exists only in animals to acetylate a subset of processed actins (Drazic et al. 2018). The presence of NatF, NatG, and NatH indicates that this modification can also occur post-translationally. NATs can acetylate the N-termini either when the initiator methionine (iMet) is cleaved or retained, mostly depending on the identity of the first two amino acids of the nascent peptide (Aksnes et al. 2016), which further enriches the complexity of the eukaryote proteome. Both NatA and NatD complete the process after iMet is cleaved. NatA shows specificity toward the N-termini of exposed A-, S-, T-, V-, C-, and sometimes G-, while NatD can only acetylate the N-terminal serine of histones H2A and H4 (Magin, Liszczak, and Marmorstein 2015; Gottlieb and Marmorstein 2018; Liszczak et al. 2013; Arnesen et al. 2009; Song et al. 2003). Other NATs like NatB, NatC, NatE, and NatF modify peptides with the iMet retained. NatB prefers the Met-Asx/Glx type N-termini (MD-, ME-, MN-, and MQ-starting) , whereas NatC, NatE, and NatF have some substrate overlap, and they acetylate peptides with “hydrophobic/amphipathic” residues like L, I, F, Y, K directly following iMet (Liszczak, Arnesen, and Marmorstein 2011; Stove et al. 2016; Van Damme et al. 2016b;

Tercero, Dinman, and Wickner 1993; Evjenth et al. 2009; Van Damme, Hole, et al. 2011). NatA, B and C are the three major contributors to the Nt-acetylome, as they cover about 80% of all NAT substrates (Aksnes et al. 2016). Notably, NatA, B and C contain at least one auxiliary subunit to couple with the catalytic subunit for enzymatic activity towards cognate substrates and anchoring of the catalytic subunit to the ribosome, while NatD, F, G, H are each composed of only one single catalytic subunit. NatE is composed to two catalytic and one auxiliary subunit.

NatA is comprised of the catalytic subunit Naa10 and the auxiliary subunit Naa15 (Liszczak et al. 2013; Gottlieb and Marmorstein 2018). Naa10 can also exist independently of Naa15, exhibiting altered substrate specificity and might also be involved in transcriptional activity inside the nucleus (Arnesen et al. 2005; Van Damme, Evjenth, et al. 2011b). Alternatively, NatA can also form a complex with another catalytic subunit Naa50 to form a ternary NatA/Naa50 complex (also called NatE) (Gautschi et al. 2003), and NatA from some eukaryotic species including *Chaetomium thermophilum* and human exhibit tight binding affinity with a regulatory protein called HYPK (Gottlieb and Marmorstein 2018; Weyer et al. 2017; Arnesen et al. 2010). Deletion of Naa50 in yeast shows no phenotype (Gautschi et al. 2003), while Naa50 knockout can affect sister chromatid cohesion in higher organisms like *Drosophila* and human (Hou et al. 2007; Chu et al. 2011; Ribeiro et al. 2016; Williams et al. 2003). Endogenous pulldown of NatA/Naa50 complexes from human cells indicates that over 80% of Naa50 is not associated with NatA *in vivo* (Hou et al. 2007), whereas the remaining ~20% forms a stable and stoichiometric complex. Furthermore, another recent study demonstrates that loss of Naa50 in *Drosophila* leads to a decrease of NatA *in vivo* acetylation activity (Rathore et al. 2016). A recent Cryo-EM structure of a yeast NatA/Naa50-ribosome

complex with an 8 Å NatA local resolution reveals that both Naa50 and the Naa15 regulatory subunit of NatA contribute to ribosome interaction, with the Naa10 active site about 50 Å from the ribosome exit tunnel, and the Naa50 active site about 2-times that distance (Knorr et al. 2019). The underlying molecular basis for the potential functional coupling and interplay between the catalytic subunits of the NatA/Naa50 complex is still poorly understood.

To understand the molecular basis for NatA/Naa50 co-regulation of N-terminal protein acetylation across species, we determined the X-ray crystal structure of a yeast NatA/Naa50 complex and used it as a scaffold to understand coregulation of NatA/Naa50 activity in yeast and human. We find that NatA/Naa50 forms a stable complex in yeast and human through conserved interactions: Naa50 makes contacts to both the Naa10 and Naa15 subunits of NatA. We demonstrate that NatA-Naa50 interactions can significantly promote the catalytic activities of the two catalytic subunits within the human complex relative to NatA or Naa50 alone. In contrast, in the yeast complex, Naa50 modestly increases the activity of NatA, and Naa50 is defective in catalytic activity, likely through compromised peptide binding. These studies have implications for understanding the role of the NatA/Naa50 complex in modulating the majority of the N-terminal acetylome in diverse species.

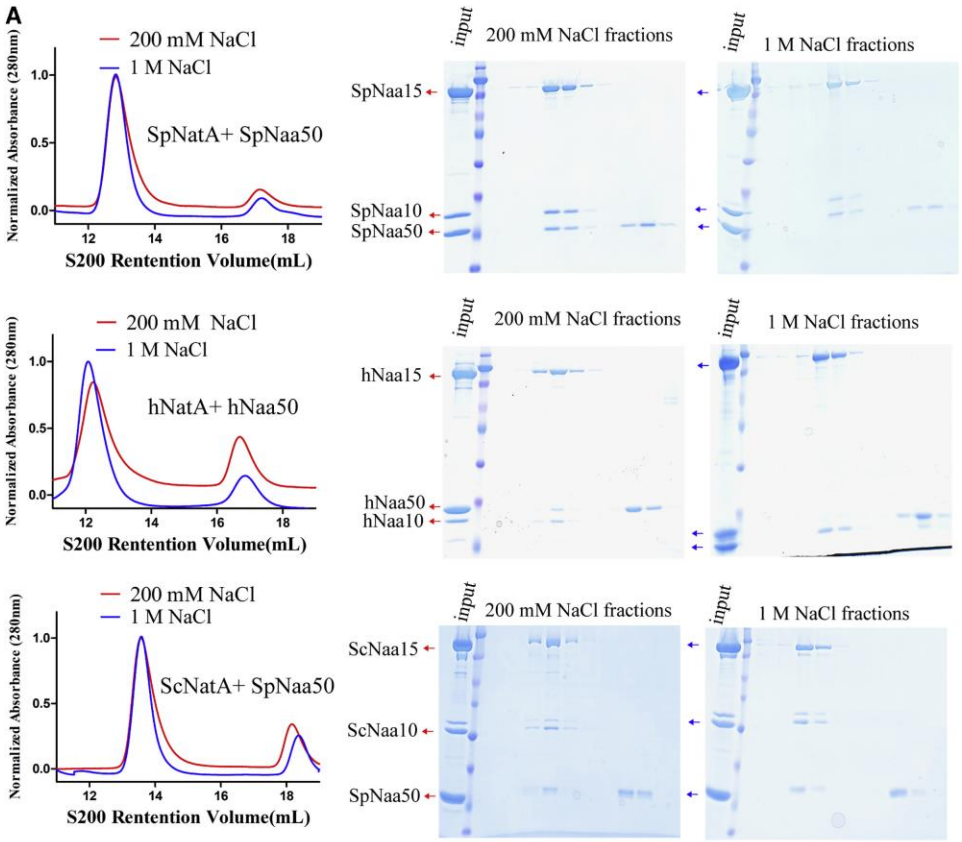
2.2 Results

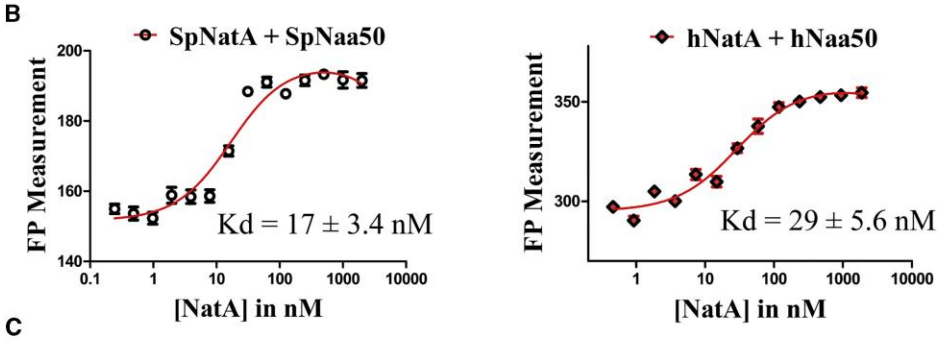
2.2.1 NatA and Naa50 form stable complexes in yeast and human

We used recombinant proteins to characterize the interaction between Naa50 and NatA from human (h) and *S. pombe (Sp*) *in vitro*. Recombinant SpNatA, SpNaa50, and human Naa50 were overexpressed and purified from *E. coli*, while hNatA was prepared from baculovirus-infected Sf9 insect cells. Consistent with previous data reporting a direct interaction between NatA and Naa50 (Gautschi et al. 2003; Hou et al. 2007), we found that N-terminal GST-tagged Naa50 can pull-down NatA. We also found that human and *S. pombe* NatA and Naa50 co-migrate by size-exclusion chromatography to form stochiometric complexes under near native salt concentration (200 mM NaCl) **(Figure 2.1A)**. Interestingly, *Sp*Naa50 maintained the ability to co-migrate with *Sp*NatA in sizing buffer with NaCl concentration as high as 1 M (**Figure 2.1A)**, suggesting that this interaction in *S. pombe* is robust and not dominated by electrostatic interactions, while hNatA and hNaa50 failed to form a stochiometric complex in the high salt buffer (**Figure 2.1A)** suggesting some divergence of the NatA-Naa50 interaction between the yeast and human complexes. In addition, we observed that SpNaa50 displays similar comigration with ScNatA under both native and high NaCl concentrations (**Figure 2.1A**), strongly indicating that the NatA-Naa50 interaction within yeast is highly conserved.

To more quantitatively assess the interaction between NatA and Naa50, we carried out a Fluorescence Polarization (FP) assay, using Naa50 labeled with Fluorescein-5-Maleimide via exposed surface cysteines. We observed that Naa50 tightly binds to NatA in both the *S. pombe* and human systems with K_d values of 17 ± 3.4 nM

and 29 ± 5.6 nM, respectively (**Figure 2.1B**). These data, together with the high degree of sequence conservation of the interaction interfaces between Naa50 and NatA in the two species, suggest that Naa50 and NatA from yeast and human are very likely to have similar association mechanisms. However, we found that GST-hNaa50 failed to pull down *Sp*NatA, and that hNaa50 and *Sp*NatA could not form a stoichiometric complex in size-exclusion chromatography (data not shown), indicating that the details of the hNatA/Naa50 interaction have diverged over evolutionary time. This was further validated using differential scanning fluorimetry experiments, where we found that the addition of SpNaa50 to *Sp*NatA and hNaa50 to hNatA had thermal melting temperature increases of 37.8 °C to 40.4 °C and 44 °C to 46.6 °C, respectively; while the addition of hNaa50 to SpNatA had a thermal melting temperature of 37.5 °C, very close to *Sp*NatA alone (**Figure 2.1C**). Notably, SpNaa50 significantly extended the melting temperature of ScNatA by 7.4°C, further confirming the existence of a conserved robust interaction across these two yeast species (**Figure 2.1C)**.


A
200 mM NaCl
1 M NaCl
SpNatA+ SpNaa50
Normalized Absorbance (280nm)
S200 Rentention Volume(mL)
input
200 mM NaCl fractions
1 M NaCl fractions
SpNaa15
SpNaa10
SpNaa50
hNatA+ hNaa50
hNaa15
hNaa50
hNaa10
ScNatA+ SpNaa50
ScNaa15
ScNaa10
SpNaa50


B
SpNatA + SpNaa50
Kd = 17 ± 3.4 nM
hNatA + hNaa50
Kd = 29 ± 5.6 nM
FP Measurement
[NatA] in nM

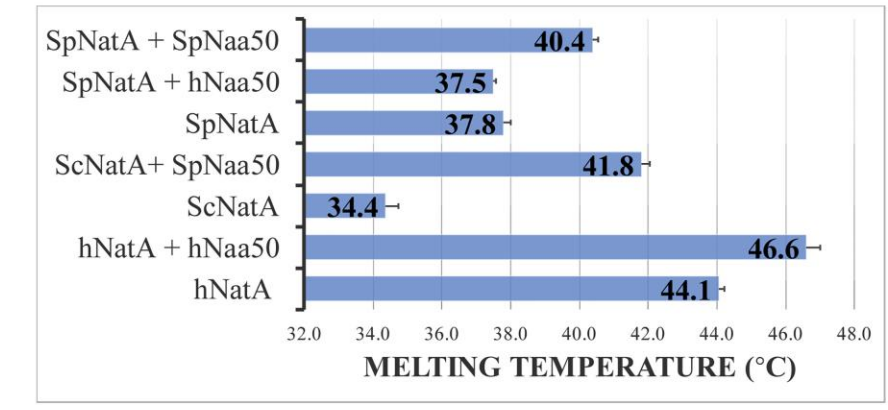

C
SpNatA + SpNaa50
SpNatA + hNaa50
SpNatA
ScNatA+ SpNaa50
ScNatA
hNatA + hNaa50
hNatA
40.4
37.5
37.8
41.8
34.4
46.6
44.1
MELTING TEMPERATURE (°C)

Figure 5.1 NatA and Naa50 form stable complexes in yeast and human

(A) Gel filtration elution profiles of Naa50 with NatA in S. pombe (top), human (middle), and cross-species between yeast S. pombe and S. cerevisiae (bottom), using a SuperdexS200column, with either 200 mM or 1MNaCl in sizing buffer. Coomassie-stained SDS-PAGE of peak fractions are shown to the right of the chromatograms. (B) Fluorescence polarization assays with NatA titrated into fluorescein-5-maleimide-labeled Naa50 in both S. pombe (left) and human (right) systems. The data are fit to calculate a dissociation constant between NatA and Naa50. Errors bars are reported in SD with n = 3. (C) Differential scanning fluorimetry assays of NatA alone or with Naa50 in both yeast and human systems. Recorded melting temperature transitions are indicated. Error bars in the figure indicate the SD of each sample, n = 3.

2.2.2 Crystal structure of the ternary NatA/Naa50 complex shows that Naa50 contacts both subunits of NatA

To reveal the molecular basis for the association and regulation mechanism of the NatA/Naa50 complex, we determined its crystal structure. We were unable to obtain crystals of the human complex, and crystals of the *S. pombe* complex diffracted X-rays poorly. However, we found that full-length ScNaa15 (residues 1–854) from *S. cerevisiae* (*Sc*) with a C-terminal truncation of ScNaa10 (1–226 out of 238 total residues) and full-length ScNaa50 (residues 1–176) produced crystals with good diffraction quality, in the presence of inositol hexaphosphate (IP_6) and bi-substrate analogs for both Naa50 and Naa10. The crystals formed in the $P2_12_12_1$ space group with one ternary complex in the asymmetric unit. The diffraction data set was collected to 2.7 Å resolution, and the

structure was determined by molecular replacement with an unpublished *Sc*NatE complex (Protein Data Bank code: 4XNH) as an initial search model. The final structure was refined to R_{work} and R_{free} values of 22.21% and 25.03%, respectively. Refinement statistics can be found in **Table 2.1**.

Not surprisingly, the *Sc*Naa15 auxiliary subunit of NatA displays a high degree of structure conservation with *Sp*Naa15 and hNaa15, with root-mean square deviation of 1.409 Å (over 577 common C_α atoms) and 1.537 Å (over 640 common C_α atoms), respectively. Shaped like a horseshoe, ScNaa15 is comprised of 15 TPR motifs, which often mediate protein-protein interactions. The electron density for the first two alpha helices (N terminal 1-53 residues) is poorly resolved and not built in our model. This auxiliary subunit, consisting of a total 42 α-helices, serves as the binding scaffold for both catalytic subunits. *Sc*Naa10 is completely wrapped by the Naa15 helices (from α11 to α30, encompassing residues Lys198- Gly595) with extensive interactions, with an overall contact surface area of 3555.5 $Å^2$ (**Figure 2.2A**). The longest alpha helix is α29, expanding to be as long as 70 Å across almost the entire length of the protein. This extremely long helix bridges the overall ring-like tertiary structure of Naa15 and locks Naa10 into the well-formed cradle made by the surrounding helices. Naa50 sits adjacent to Naa10 and interacts mostly with several Naa15 helices (α21, α22, α23, α24, encompassing residues Pro380-Asp448), and more modestly with the □2-□3 loop of Naa10, with an overall contact surface area of only 632.3 $Å^2$ (**Figure 2.2B**).

The significantly less contact surface between Naa50 and Naa15, relative to between Naa10 and Naa15, is probably one reason that Naa50 in this ternary complex displays surprising flexibility, as indicated by its high average B factor (87.6 $Å^2$) compared to Naa10 (48.9 $Å^2$) and Naa15 (56.2 $Å^2$). Several Naa50 residues away from

the Naa50-Naa15 interface have poor side chain density and are therefore modeled as alanine residues and some parts of α helices and β strands are built as loops. Interestingly, the NatA bisubstrate analog used in the cocrystallization could not be resolved due to its poor electron density and was therefore omitted from the model, while only the CoA portion of the Naa50 bisubstrate inhibitor was well resolved and are therefore included in the model.

Previous studies have revealed that human NatA contains a bound IP_6 molecule that bridges interactions between helices α19, α24, α25 of hNaa15 and the β2-loop-β3 of hNaa10 (Gottlieb and Marmorstein 2018). To identify whether a similar IP_6 binding pocket exists in *Sc*NatA, IP_6 was added to the crystallization of the ternary *Sc*NatA/Naa50 complex. Clear electron density for IP_6 was visible between Naa10 and Naa15, in almost the same position identified in the corresponding human NatA complex (data not shown). Extensive electropositive and hydrogen-bond interactions contribute to this interaction including K80, Y85, and K91 from Naa10, and K429, K457, R464, R426, H430 and Y461 from Naa15 (**Figure 2.2C**). All these residues in Naa15 are conserved from yeast to human, which suggests that this binding mode is structurally conserved. We validated the conserved intimate binding of IP_6 by also showing its robust binding to *Sp*NatA using isothermal titration calorimetry (ITC), measuring a dissociation constant of ~130 nM, with a 1:1 stoichiometry (**Figure 2.2D)**.

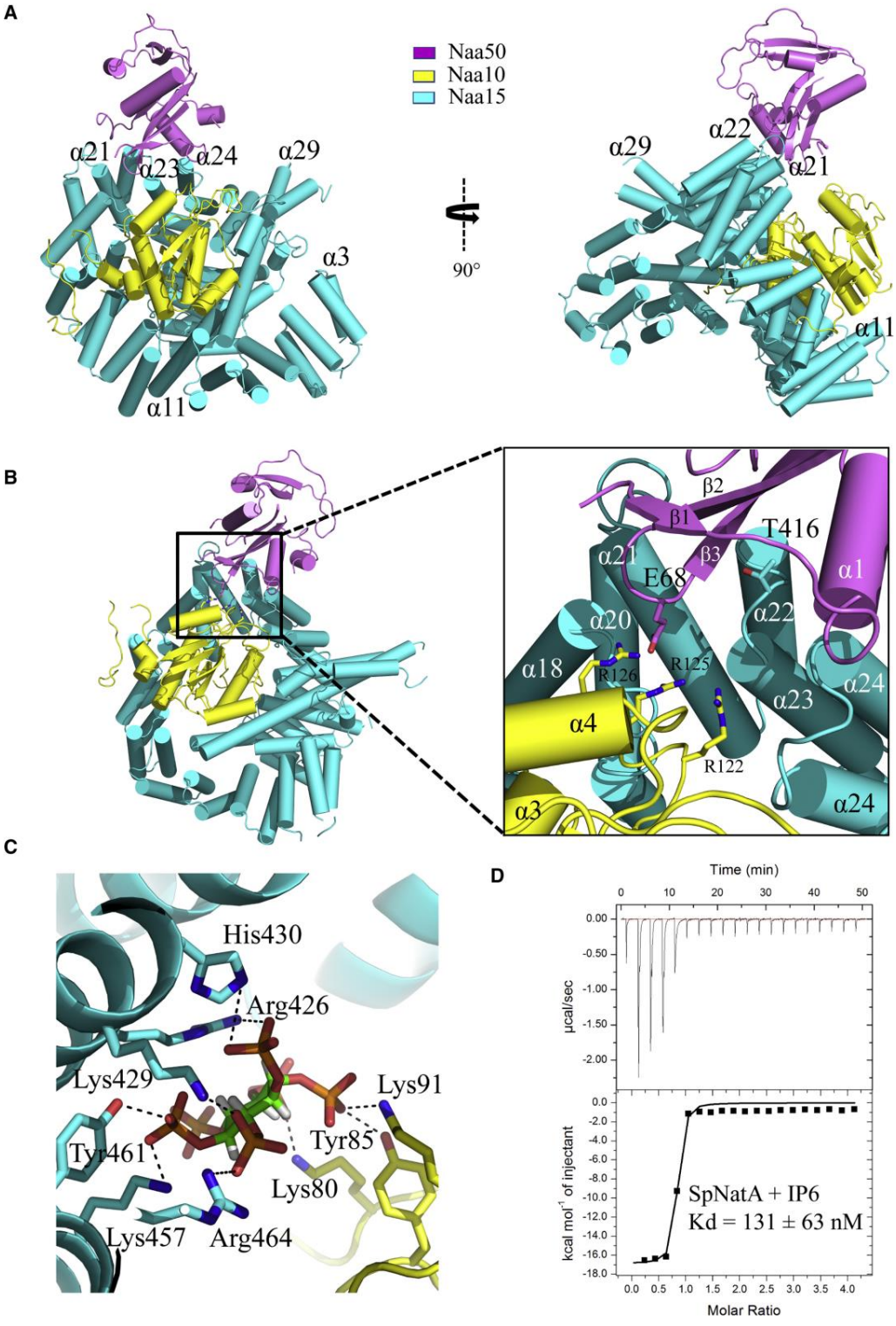

A
Naa50
Naa10
Naa15
α21
α23
α24
α29
α3
α11
90°
α29
α22
α21
α11
B
β2
β1
T416
α21
β3
E68
α1
α20
α22
α18
R126
R125
α24
α23
α4
R122
α3
α24
C
His430
Arg426
Lys429
Lys91
Tyr461
Tyr85
Lys80
Lys457
Arg464
D
Time (min)
0
10
20
30
40
50
0.00
-0.50
-1.00
-1.50
-2.00
µcal/sec
0.0
-2.0
-4.0
-6.0
-8.0
-10.0
-12.0
-14.0
-16.0
-18.0
kcal mol-1 of injectant
SpNatA + IP6
Kd = 131 ± 63 nM
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Molar Ratio

Figure 6.2 Crystal structure of the ternary NatA/Naa50 complex shows that Naa50 contacts both subunits of NatA

(A) *Sc*Naa10 (yellow), *Sc*Naa15 (cyan), and *Sc*Naa50 (violet) are shown in cartoon. The N-terminal two alpha helices (residues 1-53) of Naa15 are not resolved and not shown in the structure. Several alpha helices of Naa15 that contribute to Naa10 and Naa50 binding are labeled. (B) Zoom-in view of the interface between Naa50, Naa10 and Naa15. Residues that contribute to interactions between Naa10 and Naa50 are shown. (C) Zoom-in view showing key residues involved in interactions with IP_6. (D) Representative isothermal titration calorimetry (ITC) of curve of IP_6 titrated into *Sp*NatA with the calculated dissociation constant indicated.

Table 2.1 Data statistics for ScNatA/Naa50 crystal structure (PDB: 6O07)

Crystal Parameters	
Space group	$P2_12_12_1$
Unit cell dimension	
a, b, c (Å)	84.44, 125.73, 145.92
α, β, γ (°)	90, 90, 90
Data Collection	
Resolution (Å)	50–2.7 (3.0–2.7[a])
Unique reflections	41,731
R_{merge}	0.11 (0.18[a])
I/σ	13.11(1.923[a])
Completeness (%)	99.3 (99.1[a])
Redundancy	6.1 (5.7[a])
Refinement	
R_{work}/R_{free}	0.2221/0.2503
Root-mean-square deviation	
Bonds (Å)	0.002
Angels (°)	0.424
Average B factors (Å^2)	
Protein	58.26
Solvent	60.44
Ligand	83.58
Ramachandran statistics (%)	
Favored	94.3%
Allowed	5.7%
Outliers	0

[a]Values in parentheses are for highest-resolution shell.

2.2.3 Hydrophobic interactions dominate the ScNatAScNaa50 binding interface

The NatA/Naa50 structure reveals that Naa15 is the primary docking site for Naa50. The contact interface is mediated primarily by the Naa50 β2-β3 and β4-α2 loops with the Naa15 α22-α23 loop. These regions form a network of hydrophobic interactions mediated by several non-polar residues, including Met18, Leu22, Tyr65, Pro70, and Val71 from Naa50 and Leu447, Val418, Thr416, Leu417 from Naa15 (**Figure 2.3A**). A few H-bonds are observed: the Naa15 Thr414 sidechain bridges and interaction with the side chain of Naa50 Tyr102 and Naa15 His413, and the Naa15 Thr416 sidechain interacts with the backbone carbonyl group of Naa50 Pro70 (**Figure 2.3A**). The largely hydrophobic nature of the interface is consistent with our biochemical observation that the *Sp*NatA and Naa50 interaction can be maintained under high salt. Importantly, we found that the TPTLXE (where X represents either V or I) motif in the Naa15 α20-α21 loop is highly conserved in both yeast and human, thus suggesting that yeast and human share a similar NatA-Naa50 binding mode.

Unlike Naa10, which is surrounded on all sides by interactions with Naa15 helical segments, thereby locking Naa10 in place, contacts between Naa50 and Naa10 are more modest and mediated by the α3-β4 loop of Naa10. Specifically, contact between Naa10 and Naa50 involves several basic residues within the α3-β4 loop of Naa10 (Arg122, Arg125 and Arg126) and Glu68 within the β2-β3 of Naa50, which make a network of electrostatic interactions (**Figure 2.3B)**. The distance of Glu68 to Arg122, Arg125 and Arg126 is 2.3 Å,3.8 Å, and 6.2 Å, respectively. The relative paucity of Naa50-NatA interactions likely allows Naa50 to be conformationally flexible to allow it to impose regulatory effects on Naa10 activity. Supporting this notion is the fact that Naa10 Arg125 is the first conserved residue of the acetyl-CoA binding motif Q/RxxGxG/A.

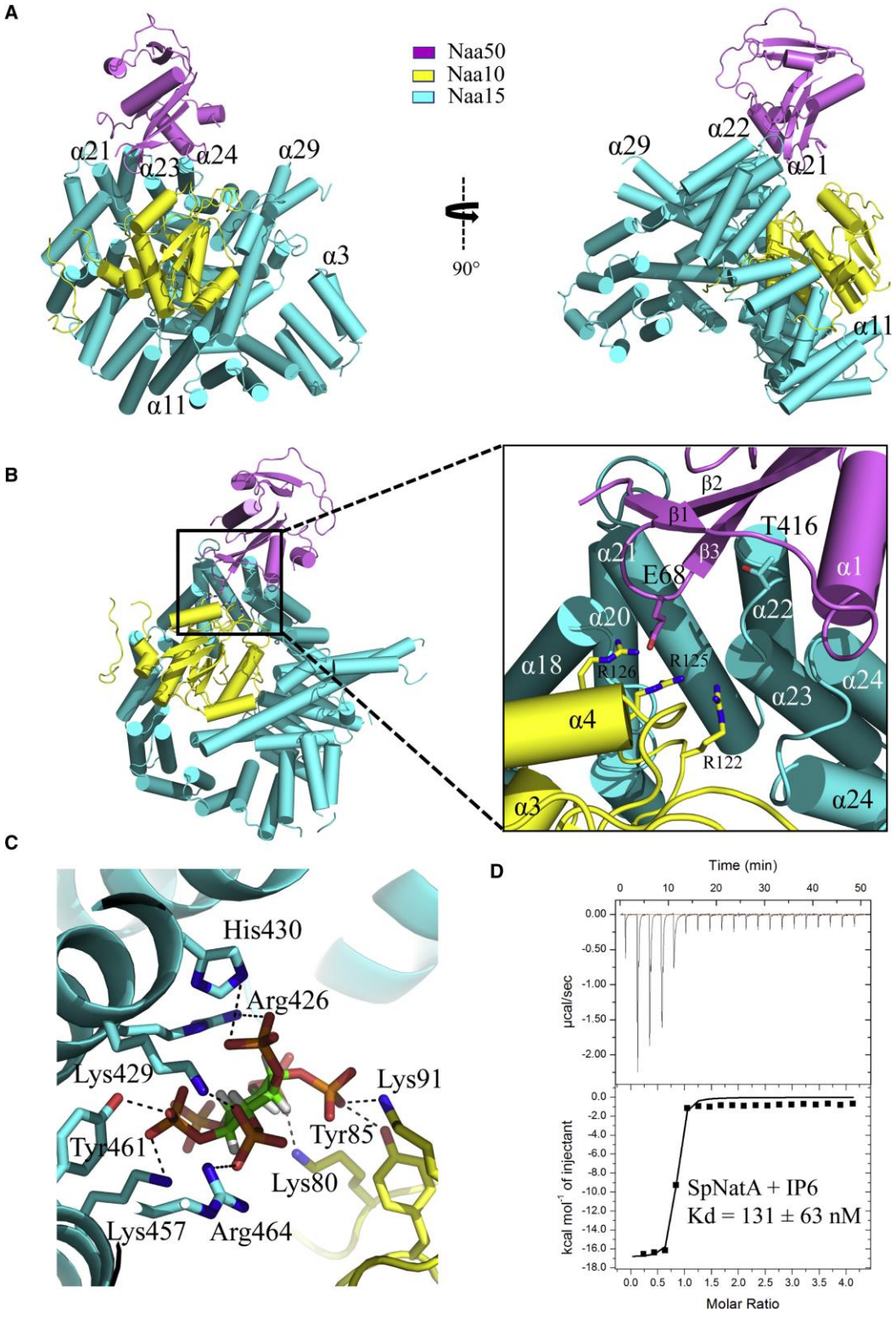

A
Naa50
Naa10
Naa15
α21
α23
α24
α29
α3
α11
90°
α29
α22
α21
α11
B
β2
β1
T416
α21
β3
α1
E68
α20
α22
α18
R126
R125
α23
α24
α4
R122
α3
α24
C
His430
Arg426
Lys429
Lys91
Tyr461
Tyr85
Lys80
Lys457
Arg464
D
Time (min)
0
10
20
30
40
50
0.00
-0.50
-1.00
-1.50
-2.00
µcal/sec
0.0
-2.0
-4.0
-6.0
-8.0
-10.0
-12.0
-14.0
-16.0
-18.0
kcal mol-1 of injectant
SpNatA + IP6
Kd = 131 ± 63 nM
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Molar Ratio

Figure 7.3 Hydrophobic interactions dominate the ScNatA-ScNaa50 binding interface

(A) Zoom-in view of the major hydrophobic binding interface between Naa15 and Naa50 with residues that participate in interaction shown. (B) Zoom-in view of the contacts between Naa10 and Naa50 with residues that participate in interaction shown.

2.2.4 The NatA/Naa50 complex promotes catalytic crosstalk

Since deletion of ScNaa50 shows no phenotype (Gautschi et al. 2003), while Naa50 knockout in higher organisms has been shown to perturb sister chromatid cohesion (Hou et al. 2007; Chu et al. 2011; Ribeiro et al. 2016) and NatA acetylation activity *in vivo* (Rathore et al. 2016), we hypothesized that there might be fundamental functional differences in the catalytic properties of Naa50 within the NatA/Naa50 complexes in yeast and human. To test this hypothesis, we took advantage of the fact that NatA and Naa50 have distinct substrate preferences. Canonical NatA substrates include peptides with the first four amino acids of SASE (Arnesen et al. 2009; Liszczak et al. 2013) (hereinafter called “SASE peptide”) and Naa50 substrates include peptides with the first four amino acids MLGP (Liszczak, Arnesen, and Marmorstein 2011; Evjenth et al. 2009) (hereinafter called “MLGP peptide”). These peptides were used because of their high specificity by their cognate NATs. For our kinetic analysis, we ignored any background activity contribution of NatA toward MLGP peptide and Naa50 toward SASE peptide, based on our data that NatA has negligible activity towards an MLGP peptide (**Figure 2.7**) and previously published data showing that Naa50 has negligible activity towards non-methionine containing N-termini (Evjenth et al. 2009).

We found that hNaa50, either alone or in complex with hNatA, harbored robust *in vitro* acetylation activity toward the Naa50 cognate peptide, MLGP (**Figure 2.4A**). Unlike hNaa50, we found that both SpNaa50 alone or with SpNatA, and ScNaa50 alone or with ScNatA, were enzymatically inactive toward the MLGP peptide (**Figure** 2.**4A**), consistent with the reported lack of *in vivo* phenotype of ScNaa50 deletion (Gautschi et al. 2003). Notably, we found that that hNaa50 activity was significantly promoted within the ternary hNatA/Naa50 complex (**Figure 2.4B)**. Indeed, kinetic analysis as a function of peptide substrate demonstrated that hNaa50 harbored a 14-fold decrease in K_m value and 11-fold increase of catalytic efficiency when bound to hNatA (**Figure 2.4B, Table 2.2)**. The dramatic change in K_m value suggested that either hNaa15 or hNaa10 can affect hsNaa50 activity by increasing its substrate binding affinity.

We also investigated the effect of Naa50 binding on hNatA activity towards its cognate SASE peptide in both the yeast and human systems. We observed that hNat*A* showed about a 2.29-fold decrease of K_m value and 1.75-fold decrease of V_{max} when it bound to hNaa50 (**Figure 2.4C, Table 2.2)**. Similarly, we found that SpNatA showed about a 2.45-fold decrease of K_m value and 1.60-fold decrease of V_{max} when bound to *Sp*Naa50 (**Figure 2.4D, Table 2.2)**. Overall, *Sp*NatA and hNatA showed comparably significant but modest increases in catalytic efficiencies of 1.53-fold and 1.31-fold, respectively. It is noteworthy that both the Naa10 catalytic subunit of the NatA complex and hNaa50 display a lower K_m for peptide when in the ternary complex, while *Sp*Naa10 also displays a lower K_m in the yeast complex. Taken together, we conclude that the acetylation activities of both catalytic subunits- Naa10 and Naa50 are promoted in this

ternary human NatA/Naa50 complex and Naa10 activity is more modestly promoted in the *S. pombe* complex.

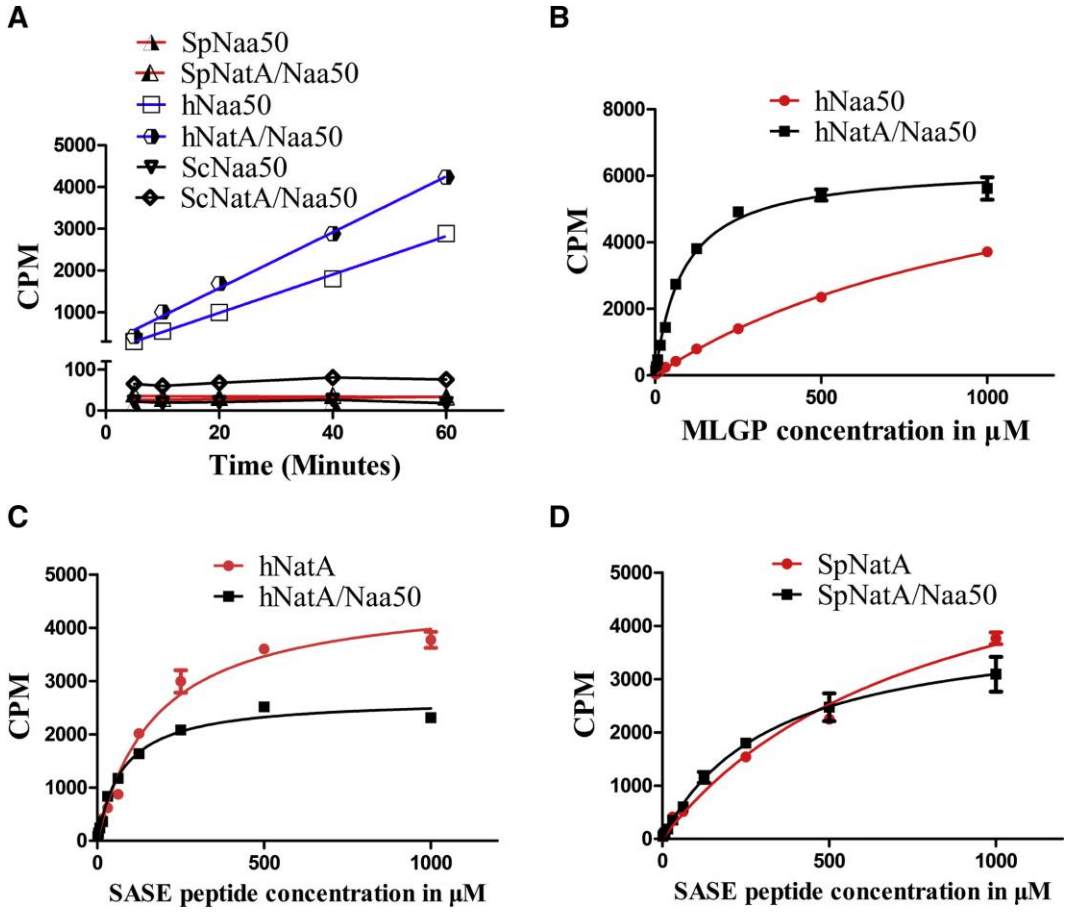


Figure 8.4 The NatA/Naa50 complex promotes catalytic crosstalk

(A) Time course acetylation activity of SpNaa50, SpNatA/Naa50, hNaa50, hNatA/Naa50, ScNaa50 and ScNatA/Naa50 against the MLGP peptide substrate. (B) Michaelis–Menten kinetic curve of hNaa50 and hNatA/Naa50 against the MLGP peptide substrate. (C) Michaelis–Menten kinetic curve of hNatA and hNatA/Naa50 against the SASE peptide substrate. (D) Michaelis–Menten kinetic curve of SpNatA and SpNatA/Naa50 against the SASE peptide substrate.

Table 3.2 Catalytic parameters for NatA, Naa50, and NatA/Naa50 against peptide substrate

	Normalized K_m	Normalized V_{max}	Normalized Catalytic Efficiency
SpNatA (SASE peptide)	1	1	1
SpNatA/Naa50 (SASE peptide)	0.41	0.63	1.54
hNatA (SASE peptide)	1	1	1
hNatA/Naa50 (SASE peptide)	0.43	0.57	1.33
hNaa50 (MLGP peptide)	1	1	1
hNatA/Naa50 (MLGP peptide)	0.071	0.77	10.8

2.2.5 Narrow substrate binding sites likely contribute to the catalytic inactivity of yeast Naa50s

Since we were unable to observe catalytic activity for *Sp*Naa50 and ScNaa50, we were surprised to observe electron density in the *Sc*Naa50 active site that could be used to model the acetyl-CoA portion of the acetyl-CoA-MLGP bisubstrate inhibitor that was used in the cocrystallization. Moreover, sequence alignment of Naa50 in human and yeast reveals that *Sp*Naa50 and *Sc*Naa50 do not contain an optimal Q/RxxGxG/A consensus acetyl-CoA binding motif, where in both *Sp*Naa50 and *Sc*Naa50 the first glycine residue is replaced with a serine residue. Using ITC, we were able to detect acetyl-CoA binding to hNaa50 (**Figure 2.5A**) but not to *Sp*Naa50 (**Figure 2.5B**). We also observed that the *Sp*NatA/Naa50 complex bound acetyl-CoA about 4.5-fold more strongly than *Sp*NatA alone (**Figures 2.5C and 2.5D**). Given that SpNaa50 alone binding is not detected in ITC, we propose that the 4.5-fold increase in acetyl-CoA binding affinity for *Sp*NatA/Naa50 relative to *Sp*NatA alone suggests that *Sp*Naa50 binding to *Sp*NatA facilitates acetyl-CoA binding activity to *Sp*NatA. This is consistent with our structural observation that *Sc*Naa50 sits closely to the Q/RxxGxG/A acetyl-CoA binding motif of *Sc*Naa10, and thereby able to modulate the acetyl-CoA binding activity of *Sc*NatA. Surprisingly, binding signal was detected by ITC between GST-ScNaa50 and acetyl-CoA (**Figure 2.5E)**, but not the free GST or GST-*Sp*Naa50 controls (**Figure 2.5F and 2.5G)**. This result suggests that the lack of acetyltransferase activity of *Sc*Naa50 is unlikely due to a defect in acetyl CoA binding.

Consistent with the catalytic inactivity of *Sp*Naa50 and ScNaa50, we found that the α1-α2 and β6-β7 loops of ScNaa50 that typically flank the peptide substrate, form a much narrower groove compared to hNaa50 for peptide binding, with the most intimate

distance between the loops of less than 2.8 Å (**Figure 2.5H)**. In addition, both *Sp*Naa50 and *Sc*Naa50 lack the characteristic YY motif found in the β6- β7 loop which is important for substrate binding in all NATs except for Naa40 and Naa80 (Liszczak et al. 2013; Liszczak, Arnesen, and Marmorstein 2011; Magin, Liszczak, and Marmorstein 2015; Goris et al. 2018; Hong et al. 2017; Aksnes et al. 2015). Taken together, it appears that the acetyl-CoA and N-terminal peptide substrate binding sites of *Sp*Naa50 are not properly configured for catalytic activity, while *Sc*Naa50 binds acetyl-CoA but likely does not bind peptide substrate.

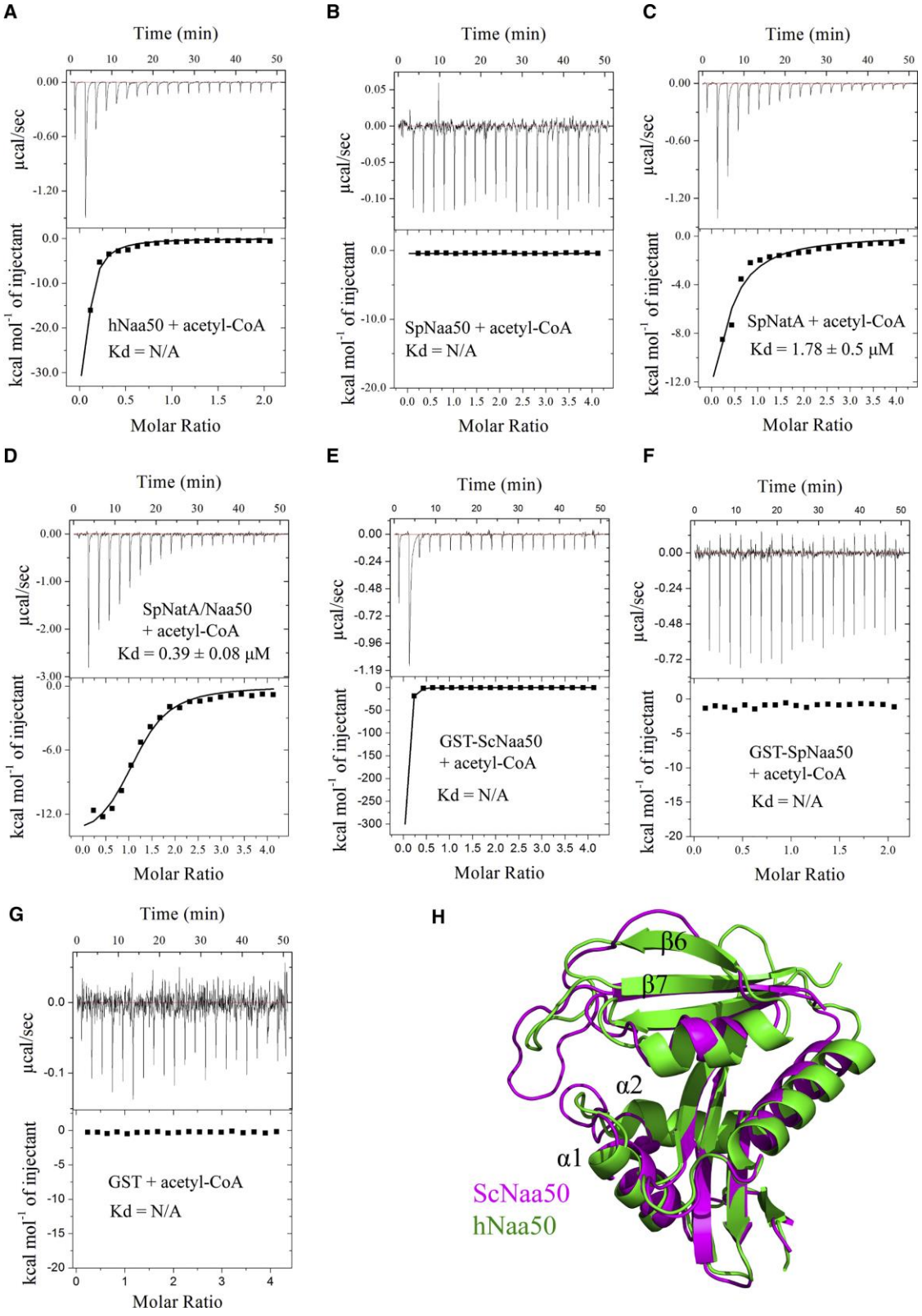

A
Time (min)
µcal/sec
kcal mol-1 of injectant
hNaa50 + acetyl-CoA
Kd = N/A
Molar Ratio
B
Time (min)
µcal/sec
kcal mol-1 of injectant
SpNaa50 + acetyl-CoA
Kd = N/A
Molar Ratio
C
Time (min)
µcal/sec
kcal mol-1 of injectant
SpNatA + acetyl-CoA
Kd = 1.78 ± 0.5 µM
Molar Ratio
D
Time (min)
µcal/sec
kcal mol-1 of injectant
SpNatA/Naa50
+ acetyl-CoA
Kd = 0.39 ± 0.08 µM
Molar Ratio
E
Time (min)
µcal/sec
kcal mol-1 of injectant
GST-ScNaa50
+ acetyl-CoA
Kd = N/A
Molar Ratio
F
Time (min)
µcal/sec
kcal mol-1 of injectant
GST-SpNaa50
+ acetyl-CoA
Kd = N/A
Molar Ratio
G
Time (min)
µcal/sec
kcal mol-1 of injectant
GST + acetyl-CoA
Kd = N/A
Molar Ratio
H
β6
β7
α2
α1
ScNaa50
hNaa50

Figure 9.5 Weak acetyl-CoA binding activity and a narrow substrate binding grove contributes to the catalytic inactivity of yeast Naa50

(A) Representative ITC curve of acetyl-CoA titrated into hNaa50. (B) Representative ITC curve of acetyl-CoA titrated into *Sp*Naa50. (C) Representative ITC curve of acetyl-CoA titrated into SpNatA. The calculated dissociation constant is indicated. (D) Representative ITC curve of acetyl-CoA titrated into SpNatA/Naa50. The calculated dissociation constant is indicated. (E) Representative ITC curve of acetyl-CoA titrated into GST-*Sc*Naa50. (F) Representative ITC curve of acetyl-CoA titrated into GST-*Sp*Naa50. (G) Representative ITC curve of acetyl-CoA titrated into free GST. (H) Superimposition of *Sc*Naa50 and hNaa50 (PDB: 3TFY) structures.

2.2.6 NatA-Naa50 from yeast and human make conserved interactions

To evaluate the degree of evolutionary conservation of the NatA-Naa50 interface, we first used the *Sc*NatA/Naa50 complex as a scaffold to carry out mutagenesis and binding studies of the yeast NatA/Naa50 complex, using *S. pombe* for mutagenesis for consistency with the earlier biochemical studies in *S. pombe*. Superimposition of the *Sc*NatA/Naa50 and *Sp*NatA structures highlighted a few potential key *S. pombe* candidate residues for interaction, including *Sp*Naa10 R117, *Sp*Naa15 T412, and *Sp*Naa50 Y49 and V53 (**Figures 2.3**). Using pull-down assays, we demonstrated that *Sp*Naa10-R117E had an insignificant effect on *Sp*NatA binding, while *Sp*Naa50-V53K and -Y49A displayed some decrease in *Sp*NatA binding (**Figure 2.6A, left**). As expected, the double mutation, *Sp*Naa50- Y49A/V53K was significantly defective in *Sp*NatA and *Sc*NatA binding (**Figure 2.6A**). As these two residues are conserved from

yeast to human, it was not surprising that the double mutation of hNaa50-Y50A/I54K displayed similar defects in hNatA binding. The *Sc*NatA/Naa50 crystal structure reveals that *Sc*NatA-T412 sits roughly in the middle position of the conserved Naa15 TPTLXE motif and is located in the center of the binding interface within Naa15 and Naa50 **(Figure 2.3A**). Interestingly, we found that a single *Sp*Naa15-T412 mutation to either tyrosine or lysine completely disrupted complex formation with *Sp*Naa50 (**Figure 2.6A**). As expected, sequence alignment demonstrates that *Sp*Naa15-T412 is well conserved in both human and yeast. We further observed that both *Sc*Naa15-T416Y and hNaa15-T406Y (hereinafter referred to as TY mutants) lost their ability to association with Naa50 in pull down assays (**Figure 2.6A)**. Consistent with the pull-down results (**Figure 2.6A**), the Naa15 TY mutant from *S. pombe*, *S. cerevisiae* and human did not co-migrate with Naa50 by size-exclusion chromatography (**Figure 2.6B**). For further studies, we focused on *Sp*NatA T412Y and the conserved hNatA T406Y and characterized their roles in NatA-Naa50 association. While the addition of Naa50 to NatA showed robust binding using a FP assay (**Figure 2.1A**) and increased thermal melting temperature of NatA by about 2.5 °C using DSF assay (**Figure 2.1B**), addition of Naa50 to the NatA TY mutants did not show detectable binding (**Figure 2.7A**) or increase in NatA TY mutant thermal melting temperature (**Figure 2.7B**). We further tested the effect of the NatA TY mutants on NatA acetylation activity and found that they maintained acetylation activity towards cognate substrates although the hNatA TY mutant showed an apparent increase in acetylation activity relative to WT (**Figure 2.7C**). The reason for this is unclear. Taken together, the Naa15 TY mutants can disrupt the association between NatA and Naa50 in yeast and human. These studies highlight the important roles of this conserve threonine residue in the association between Naa50 and NatA.

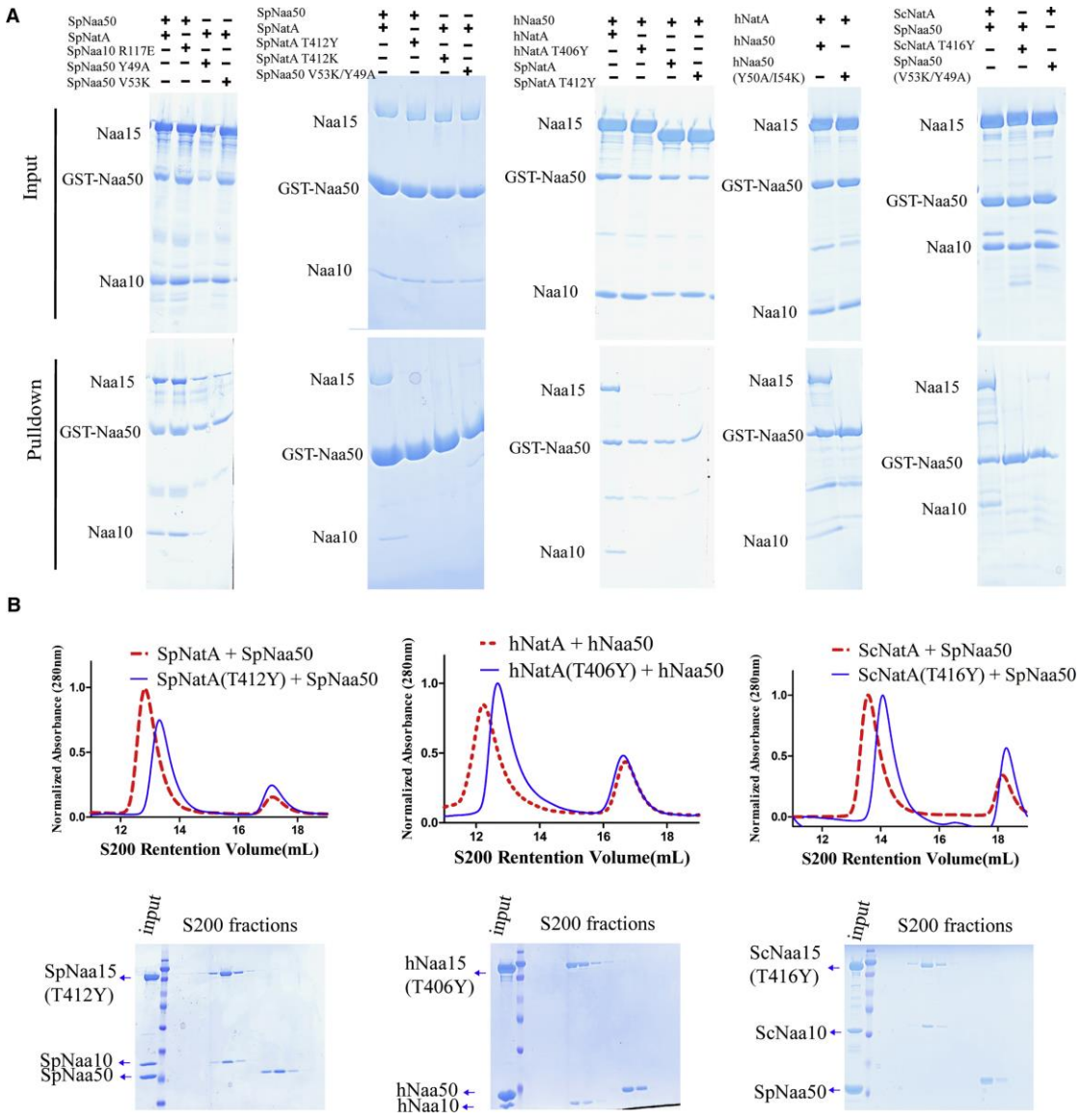


Figure 10.6 NatA TY mutants are unable to pull-down Naa50 and co-migrate with Naa50

(A) GST pull-down assay with NatA/Naa50 mutants to interrogate the contribution of residues in Naa50-NatA association. (B) Gel filtration elution profiles of Naa50 with either wild-type (dotted red line taken from Figure 1A) or TY mutants (blue line) of NatA using a Superdex S200 column. Coomassie-stained SDS–PAGE of peak fractions of Naa50 with NatA TY mutants shown below the chromatograms.

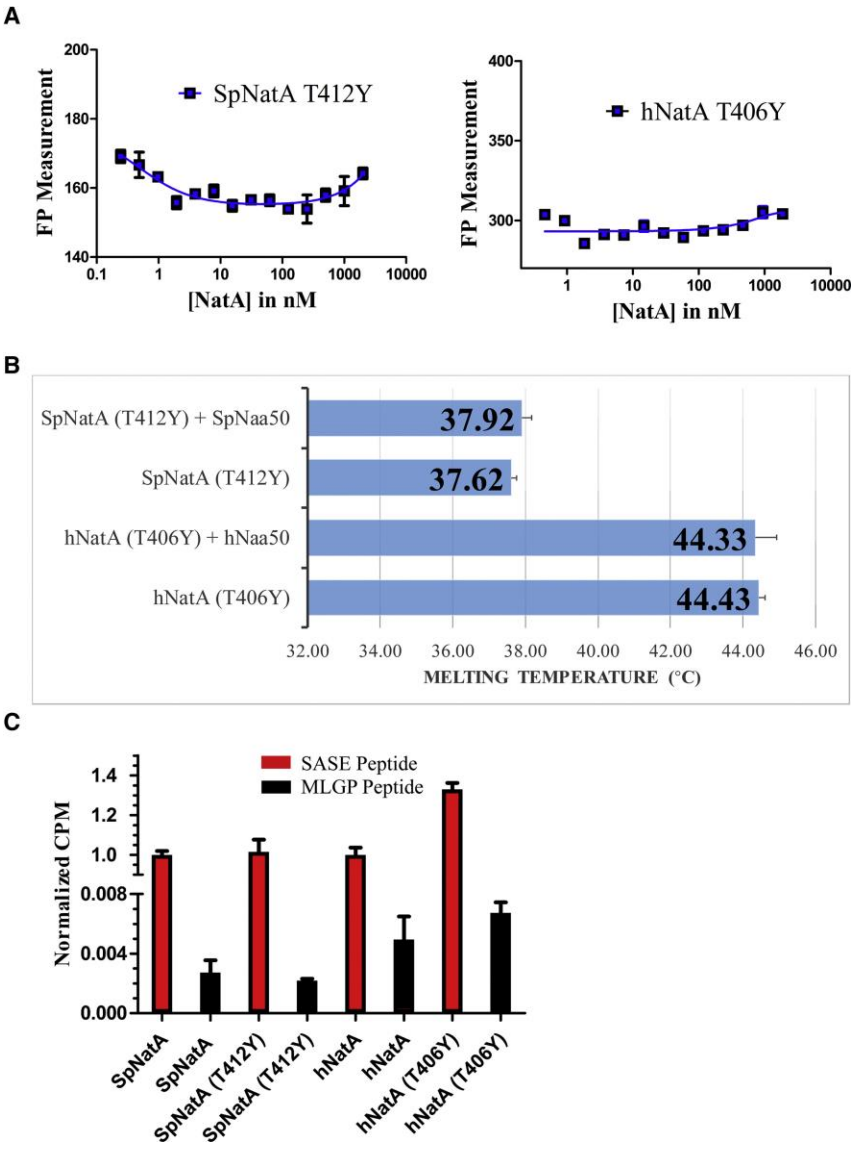


Figure 11.7 NatA TY mutants disrupt NatA/Naa50 complex interactions

(A) Fluorescence polarization assays of NatA TY mutants and Naa50 in *S. pombe* and human. (B) Differential scanning fluorimetry assays of NatA TY mutants with and without Naa50. Recorded melting temperature transitions are indicated. (C) Comparison of acetylation activity of NatA wild-type and TY mutants. Activities are normalized to WT protein activity level.

2.3 Discussion

While the molecular basis for acetylation activity by either NatA or Naa50 independently has been previously described, the underlying molecular mechanism of how these enzymes work in complex was poorly understood before this study. Although similar to an unpublished structure of *Sc*NatE that has been deposited in the PDB (4XNH), our reported crystal structure of the *Sc*NatA/Naa50 complex in combination with related biochemical data provides a peer reviewed and comprehensive analysis of the structure and mechanism of acetylation by the N-terminal dual enzyme NatA/Naa50 complex across evolution. Overall, the Naa50 electron density is poorly resolved relative to NatA, which suggested that its association with NatA is likely dynamic. Indeed, the major contact region between Naa50 and Naa15 is mediated by loop-loop interactions, unlike the extensive interactions made between Naa10 and Naa15 within the NatA complex. We also found that *Sc*Naa50 makes direct electrostatic interactions with Naa10, providing a path of contact for Naa10 and Naa50 to influence each other's catalytic functions.

Because of their conserved association motif, the hNaa50 docking site on hNaa15 could be identified, and was further supported by mutational analysis. We propose that the proper distance between Naa10 and Naa50 that is required for cognate acetylation of N-terminal protein substrates by the respective NATs, coupled to the less rigid positioning of Naa50, relative to Naa10, may allow Naa50 to adopt multiple conformations within the NatE complex to facilitate both Naa50 and Naa10 cognate substrate binding. Indeed, we found that hNaa50 and hNatA displayed significant enhancement in substrate affinity and catalytic efficiency in the context of the complex, although such enhancement was not observed in the yeast complex, where we found

that Naa50 is catalytically defective. This is consistent with our structural data showing that unlike Naa10, Naa50 harbors a narrow substrate binding site that is incompatible with N-terminal protein substrate binding. However, we cannot exclude the possibility that yeast Naa50 is enzymatically active against another substrate. Regardless, our findings suggest some evolutionary divergence between yeast and human NatE complexes. Nonetheless, we also found that SpNaa50 does not bind human NatA and vice versa, yet from conservation analysis there is no obvious reason why this should be the case. There may be subtle differences in the yeast and human systems that have co evolved, perhaps to reflect the regulatory role of the HYPK protein in the human but not *S. pombe* and *S. cerevisiae* yeast systems (Arnesen et al. 2009).

Based on our *in vitro* activity data, it is noteworthy that both human Naa10 and Naa50 catalytic subunits display lower K_m values for their respective peptide substrates and increased catalytic efficiencies, when they are in the NatA/Naa50 relative to when they are not in complex. Meanwhile, we observed that the other substrate, acetyl-CoA, while not showing detectable binding to *Sp*Naa50, binds to *Sp*NatA/Naa50 about 4.5-fold tighter than *Sp*NatA alone. It is likely that *Sp*Naa50 facilitates acetyl-CoA binding to *Sp*NatA and the structural observation that *Sc*Naa50 sits closely to the Q/RxxGxG/A acetyl-CoA binding motif of *Sc*Naa10 provides a plausible explanation for how *Sp*Naa50 may modulate the acetyl-CoA binding activity of *Sp*NatA. Taken together, we conclude that in human, NatA and Naa50 have cooperative effects on their acetylation activities when associating together, while in yeast, the inactive Naa50 still contributes to NatA activity, which is consistent with previous result that reduction of some NatA type N-termini was detected in yeast Naa50 deletion mutants (Van Damme et al. 2015) We note that the cooperative effects between NatA and Naa50 that we observed *in vitro* are

based only on single cognate NatA and Naa50 peptide substrates. It is currently unknown if the degree of cooperation varies with different peptide substrates.

One surprising aspect of the *Sc*NatA/Naa50 structure is the identification of Inositol hexaphosphate (IP_6), which bridges interactions between the Naa10 and Naa15 subunits of NatA with a dissociation constant in the sub-mircomolar range. A similar role for IP_6 was identified in the recently reported hNatA/HYPK complex (Gottlieb and Marmorstein 2018), suggesting an evolutionarily conserved role of IP_6 in stabilizing the NatA complex. Since IP_6 has long been appreciated to have many signaling properties in diverse organisms including yeast and mammals (Kalam Shamsuddin and Bose 2012), it is also possible that IP_6 may have a yet unidentified signaling role in NatA function.

HYPK is an important regulatory protein that binds to hNatA but is not conserved in *S. pombe* or *S. cerevisiae*. Recent structural studies suggest that HYPK was a nanomolar dissociation constant NatA binding partner and that HYPK binding likely inhibits Naa50 binding (Weyer et al. 2017; Gottlieb and Marmorstein 2018). Here we have demonstrated that hNaa50 can promote hNatA acetylation activity. We propose that the slightly divergent association mechanism between NatA and Naa50 in yeast and human may be related to the presence of HYPK in human and higher eukaryotes but not in most yeast. It will be of interest to further investigate the interplay among these binding partners within the tetrameric human NatA/Naa50/HYPK complex.

We find in our study that *Sp*Naa50 is catalytically inactive, and this is likely due to its reduced affinity for acetyl-CoA and the relative narrowness of the peptide substrate binding site of *Sc*Naa50 observed in the *Sc*NatA/Naa50 structure. Surprisingly, we observed binding between *Sc*Naa50 and acetyl-CoA, although the optimal acetyl-CoA

binding motif is missing in both *Sc*Naa50 and *Sp*Naa50. Nonetheless, both *Sc*Naa50 and *Sp*Naa50 are catalytically inactive in out acetyltransferase assays. We propose that the yeast lineage lost catalytically active Naa50 at least due to the lack of peptide substrate binding. Indeed, unlike metazoans, *Sc*Naa50 knock out displays no phenotype (Gautschi et al. 2003). In another study, no impact on Met-starting yeast N-termini was observed where yeast Naa50 was deleted (Van Damme et al. 2015). Meanwhile, Naa50 has functional roles in sister chromatic cohesion in higher organism, independent of the NatA complex, which is not observed in yeast (Rong et al. 2016; Ribeiro et al. 2016; Williams et al. 2003). Thus, yeast Naa50 is very likely enzymatically inactive, only providing extra stability and regulatory roles for NatA, including participation in ribosome binding (see below). We propose that Naa50 substrates are processed by NatC, which recognizes a similar cognate sequence.

A recent Cryo-EM structure of a native *Sc*NatA/Naa50-ribosome structure, where the ribosome is resolved to 3.4 Å and the NatA/Naa50 portion is resolved to about 8 Å resolution, indicates that NatA/Naa50 associates with the ribosome through interaction with the ribosomal RNA expansion segments involving predominantly Naa15, but also Naa50 (Knorr et al. 2019). While Naa10 is positioned proximal to the ribosome peptide exit tunnel, Naa50 is remote from the peptide exit tunnel. A docking of our 2.7 Å resolution NatE structure onto the Cryo-EM structure (PDB: 6HD7) provides a pseudo-3.4 Å resolution structure of the NatE-ribosome complex (**Figure 2.8**). This superposition reveals that Naa50 is positioned ~80 Å away from the peptide exit tunnel of the ribosome, consistent with the finding of Knorr et al. that approximates this distance to be ~ 85 Å (Knorr et al. 2019). As this observation suggests that peptides emerging from the ribosome exit tunnel encounter the Naa10 active site prior to the Naa50 active site, we

speculate that it is likely that Naa50 requires a longer nascent chain for acetylation. However, we found that *Sp*Naa50 binding slightly promotes *Sp*NatA acetylation catalytic efficiency. This effect could be more significant in the context of the ribosome, given that Knorr et al. observe that Naa50 makes contact to the ribosome RNA expansion segment, thus also playing a role in NatA recruitment to the ribosome. Such a contribution to NatA-ribosome recruitment by Naa50 is consistent with a recent observation that *D. melanogaster* Naa50 knockout impairs NatA acetylation activity *in vivo* (Rathore et al. 2016). We placed the Naa10 acetyl CoA binding site by overlapping our structure to the *Sp*NatA structure in complex with bi-substrate analogue (PDB:4KVM) and found that the acetyl-CoA binding sites for both Naa10 and Naa50 are oriented in the same direction and facing the ribosome nascent peptide exit tunnel and separated by a distance of only 26.6 Å (**Figure 2.8**). It is likely that the relative position of these two catalytic subunits, Naa10 and Naa50, might change on the ribosome to accommodate the cognate peptide substrate that emerges out of the ribosome exit tunnel. We propose that the flexibility of the hNatA-Naa50 interface (as suggested by the relatively high B-factor of Naa50 relative to NatA) may also facilitate movement of Naa50 closer to the ribosome exit tunnel once its cognate substrate is in proximity for N-terminal acetylation. This might be facilitated further if human Naa50 did not participate in NatA anchoring to the ribosome as it does in the yeast system. Additional interactions might be compensated through the binding of HYPK.

Taken together, our biochemical and structural characterization of ternary NatA/Naa50 complexes in yeast and human in this study has provided new and significant insights into the association and regulatory mechanism of this unusual dual enzyme system. These studies set the stage for future studies to decipher the molecular

mechanism of the human NatA/Naa50/HYPK complex. In addition, our identification of mutants that prevent NatA/Naa50 association in yeast and humans will facilitate analysis of the uncoupling of co- and post-translational protein acetylation by Naa50.

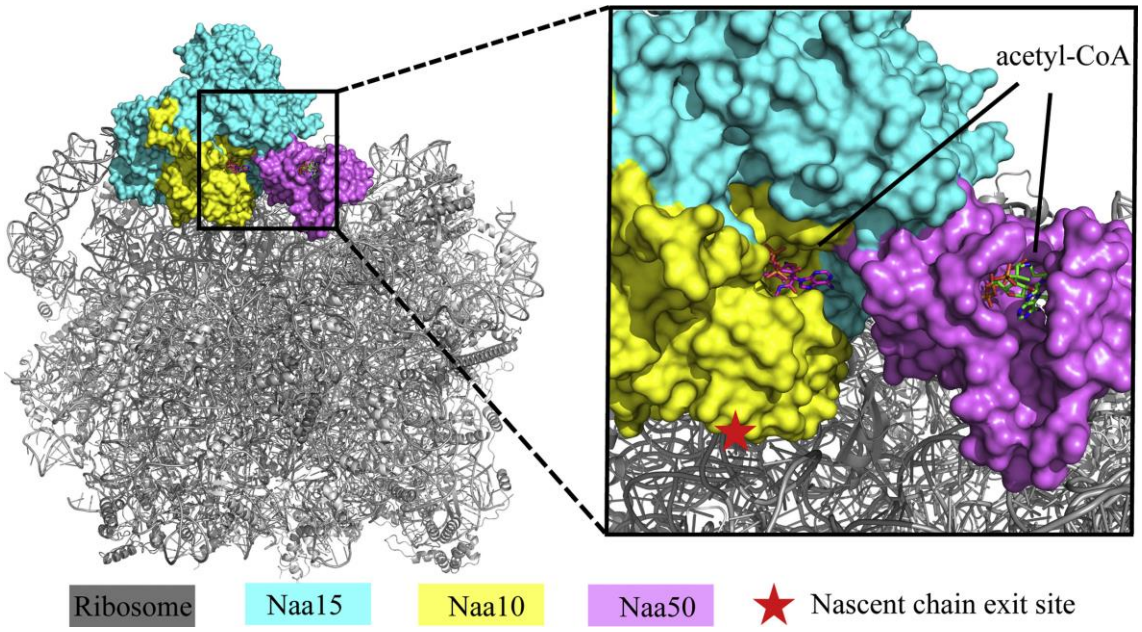


Figure 12.8 Docking of our ScNatA/Naa50 crystal structure onto the ribosome-NatA/Naa50 Cryo-EM structure

The *Sc*NatA/Naa50 crystal structure is aligned to the NatA/Naa50-ribosome structure (PDB: 6HD7). The magnified view shows that Naa10 is most proximal to the ribosome peptide exit tunnel, while the acetyl-CoA binding sites of Naa10 and Naa50 are facing the same side.

2.4 Materials and methods

2.4.1 NatA and Naa50 expression and purification

Human NatA (hNatA), human Naa50 (hNaa50), and *Schizosaccharomyces pombe* NatA (SpNatA) were expressed and purified as described previously (Liszczak et al. 2013; 2018; Liszczak, Arnesen, and Marmorstein 2011). Full-length *Schizosaccharomyces pombe* Naa50 (spNaa50) was cloned into pRSF with a N-terminal GST fusion, and expressed and purified with the same method as described for spNatA (Liszczak et al. 2013). Full-length *Saccharomyces cerevisiae* Naa15 (ScNaa15) was cloned into pRSF_Duet vector with a N terminal PolyHis tag and a Tobacco Etch Virus (TEV) cleavage site. C-terminal truncated ScNaa10 (1-226) with no tag was cloned into pET_Duet, and un-tagged full-length *Sc*Naa50 was cloned into pCDF-Duet. To express the ternary *Sc*Naa10-*Sc*Naa15-*Sc*Naa50 complex (*Sc*NatA/Naa50 complex), these three plasmids were co-transformed into Rosetta (DE3)pLysS competent *E. coli* cells, which were grown to an OD_{600} of 1 and induced with 0.5 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 16 °C for ~16 h. ScNatA/Naa50 complex purification procedure was modified from SpNatA as described (Liszczak et al. 2013). Cells were isolated by centrifugation for 20 mins at 4500 rpm and lysed by sonication in a buffer containing 25 mM Tris, pH 8.0, 500 mM NaCl, 0.1 mg/mL PMSF. After centrifugation, the supernatant was isolated and passed over Ni-resin (Thermo Scientific), which was subsequently washed with 10 column volumes of lysis buffer supplemented with 25 mM imidazole and 10 mM 2-mercaptoethanol. Protein was eluted with 300 mM imidazole and dialyzed into buffer containing 25 mM sodium citrate monobasic, pH 5.5, 10 mM NaCl and 10 mM 2-mercaptoethanol. Protein was purified with a 5-mL HiTrap SP ion-exchange column and eluted in a salt gradient (10–700 mM NaCl). Peak fractions were

concentrated to ~ 0.5 mL with a 100-kDa concentrator (Amicon Ultra, Millipore), and loaded onto an S200 gel-filtration column (GE Healthcare) in a buffer containing 25 mM HEPES, pH 7.0, 200 mM NaCl, 5% glycerol and 1 mM TCEP. Proteins were aliquoted, snap-frozen in liquid nitrogen, and stored at −80 °C for further use. Full-length *Sc*Naa50 was cloned into pRSF with a N-terminal GST tag and a TEV cleavage site, to express and purify *Sc*Naa50 alone. We observed that *Sc*Naa50 was not stable in the absence of the GST tag and absorbs minimally at 280 nm. We therefore only worked with GST-*Sc*Naa50 in biochemical experiments. GST-*Sc*Naa50 was purified using the same method as described above. Concentrations of hNatA, hNaa50, *Sp*NatA, *Sp*Naa50, *Sc*NatA/Naa50 and GST-*Sc*Naa50 were determined by absorbance at 280 nm and using their extinction coefficient of 132,480 $cm^{-1}M^{-1}$,13,410 $cm^{-1}M^{-1}$, 120,000 $cm^{-1}M^{-1}$, 14,440 $cm^{-1}M^{-1}$, 151,800 $cm^{-1}M^{-1}$, and 58,790 $cm^{-1}M^{-1}$, respectively. All protein mutants were generated using the QuikChange protocol from Stratagene and obtained following the expression and purification protocols described above.

2.4.2 Fluorescence polarization (FP) assays

There are five cysteine residues present in *Sp*Naa50 (Cys22, Cys54, Cys60, and Cys96) and four cysteine residues in hNaa50 (Cys60, Cys61, Cys79, and Cys100). We found that we did not have to make mutations to obtain the fluorescein labeled Naa50, which suggested that at least one cysteine in each is exposed to solvent. 25-fold excess of Fluorescein-5-Maleimide (ThermoFisher) was added into about 100 µL of 1 mg/ml purified hNaa50 or *Sp*Naa50 in sizing buffer. The reaction was quenched with 2-mercaptoethanol after overnight incubation at 4°C. Excess Fluorescein-5-Maleimide was

removed by S75 gel-filtration chromatography. To run the assays,10 nM of labeled Naa50 was used in all reactions, and NatA concentrations were varied to determine the dissociation constant (K_d). 5 mg/mL BSA and 0.2% v/v Tween were added into the reaction buffer (25 mM HEPES, pH 7.0, 200 mM NaCl and 10 mM DTT) to prevent non-specific binding. FP readings were recorded with a Perkin Elmer EnVision and each curve was repeated in triplicate. GraphPad Prism, version 5.01, was used for all data fitting. Errors were reported in SD with $n = 3$.

2.4.3 Acetyltransferase activity assays

All the acetyltransferase assays were performed essentially as previously described (Liszczak et al. 2013; Liszczak, Arnesen, and Marmorstein 2011) and carried out at room temperature in the reaction buffer containing 75 mM HEPES, pH 7.0, 120 mM NaCl, 1 mM DTT. The SASE substrate peptide (NH_2-SASEAGVRWGRPVGRRRRP-COOH; GenScript) and the MLGP substrate peptide (NH_2-MLGPEGGRWGRPVGRRRRP-COOH; GenScript) were used to determine the enzymatic activity of NatA and Naa50, respectively. For Naa50 activity test in time course manner, 500 nM of SpNaa50, SpNatA/SpNaa50, hNaa50, hNatA/hNaa50, ScNaa50, and ScNatA/Naa50 were mixed with 300 μM of C^{14} labeled acetyl-CoA (4 mCi $mmol^{-1}$; PerkinElmer Life Sciences), 500 μM MLGP peptide in a 100 μL reaction. To quench the reaction at specific time, 15 μL of the reaction mixture was added to P81 paper discs (Whatman), and the paper discs were immediately placed in wash buffer with 10 mM HEPES, pH 7.5. Unreacted acetyl CoA should be removed by washing at least three times. Paper discs was then dried with acetone and mixed with 4 scintillation

fluid for signal measurement by a Packard Tri-Carb 1500 liquid scintillation analyzer. Data was plotted with GraphPad Prism, version 5.01. To determine steady-state catalytic parameters, enzymes were mixed with 500 μM C^{14} labeled acetyl-CoA and corresponding peptide substrate concentration was varied. 20 μL reaction were performed but only 15 μL reaction mixture was quenched onto the paper discs after specific time. For hNaa50 and hNatA/hNaa50 against MLGP peptide, 300 nM enzyme was used in 40-minute reaction. For hNatA and hNatA/hNaa50 against SASE peptide, 100 nM enzyme was used in 12-minute reaction. For SpNatA and SpNatA/SpNaa50 against SASE peptide, 50 nM enzyme was used in 10-minute reaction. Paper discs washes and signal reading steps were performed in the same manner as mentioned above. GraphPad Prism, version 5.01, was used for all kinetic data fitting to the Michaelis–Menten equation. Each curve was repeated at least 3 times. In NatA TY mutant activity tests, 100 nM of SpNatA, SpNatA-T412Y, hNatA, hNatA-T406Y were mixed with 300 μM of C^{14} labeled acetyl-CoA and 500 μM peptide substrate for a 30-minutes reaction. Signals were normalized to the WT activity, and errors were reported in SEM with $n = 3$.

2.4.4 Size exclusion chromatography assays

A total of 500 μl of purified NatA (4.2 μM) and purified Naa50 (12.6 μM) was injected onto a S200 gel-filtration column (GE Healthcare). UV absorbance values were normalized to the maximum value at first elution peak, then plotted against the elution volume in GraphPad Prism, version 5.01. Only elution volume around two peaks were shown in the figure. Corresponding peak fractions were run on SDS-PAGE and stained with colloidal Coomassie blue for imaging.

2.4.5 GST pull-down assays

1 µM of GST tagged Naa50 and 5 µM of NatA was incubate with 50 µL of glutathione agarose resin at 4 °C for 1 hour in reaction buffer containing 25 mM HEPES, pH 7.0, 200 mM NaCl and 1 mM TCEP. Resin was then washed extensively with the same buffer and eluted with glutathione. Samples were collected for analysis on SDS-PAGE.

2.4.6 Differential scanning fluorimetry assays

Sypro Orange (5000X stock, ThermoFisher Scientific) was diluted 1:200 and 4 µL was mixed with 16 µL solution with 5 µM NatA and 5 µM Naa50 in sizing buffer. Fluorescent readings were recorded using a qPCR (ABI 7900 RealTime PCR) with a 2% ramp rate, while heated from 20 °C to 95 °C. Melting curves were generated from these readings and melting temperatures were determined by taking the first derivative of the curves. DSF scans of all samples were performed in triplicates. Error bars in the figure indicated the SD of each sample.

2.4.7 ITC measurements

Measurements were recorded on a MicroCal iTC200 at 20 °C. Samples were dialyzed into buffer containing 25 mM HEPES pH 7.0, 200 mM NaCl, 1 mM DTT. Protein samples (*Sp*Naa50, *Sp*NatA, *Sp*NatA/Naa50, GST-SpNaa50, GST-ScNaa50, GST) with

concentrations of 50 μM in the cell and 1mM of acetyl-CoA in the syringe were used in the experiments. For hNaa50, 100 μM protein and 0.5 mM of acetyl-CoA were used. The raw data was analyzed with MicroCal ITC analysis software.

2.4.8 ScNatA/Naa50 crystallization and data collection

Ternary *Sc*NatA/Naa50 complex was mixed with 3 molar excess of NatA bisubstrate inhibitor (CoA-SASEA), Naa50 bisubstrate inhibitor (CoA-MLGPE) and IP_6. Bisubstrate inhibitors were prepared as previously described (Liszczak et al. 2013). Crystals were grown within ~3 days by hanging drop vapor diffusion at 20 °C at 7.5 mg/ml, using one to one drop ratios against reservoir solution containing 10% PEG 3350, 0.1 M Sodium Malonate (pH 5.0). Crystals were transferred into mother liquor supplemented with 25% glycerol and flash-frozen in liquid nitrogen. Data were collected at the Advanced Photon Source (beamline 24-ID-C) and processed using HKL2000 (Otwinowski and Minor 1997).

2.4.9 ScNatA/Naa50 structure determination and refinement

The collected diffraction data showed strong anisotropy and ellipsoidal truncation and anisotropic scaling was performed (Strong et al. 2006). The resolution limit along a*, b*, c* are 2.7 Å, 3.4 Å, 2.7 Å, respectively. The crystal structure of ScNatA/Naa50 was determined by molecular replacement using a structure of unpublished ScNatE (PDB ID Code: 4XNH) without ligand or solvent molecules as a search model. Molecular replacement was done using Phaser in Phenix (Adams et al. 2010).

Initial Manual model building was done in Coot (Emsley and Cowtan 2004) and all subsequent rounds of refinement were performed using Phenix refine and Coot interchangeably. Refinement statistics can be found in **Table 2.1**. The final model and structure factors were submitted to the Protein Data Bank with code 6O07 (Research Collaboratory for Structural Bioinformatics).

Distance calculations, as well as three-dimensional alignment r.m.s. deviations and graphics were generated in PyMOL (http:// www.pymol.org/)

Sequence alignments in the manuscript were performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and visualized using ESPript 3.0 (Robert and Gouet 2014a) (http://espript.ibcp.fr/ESPript/ESPript/)

All surface area calculations were performed using PDBePISA (Proteins, Interfaces, Structures and Assemblies) (http://www.ebi.ac.uk/pdbe/pisa/)

CHAPTER 3: Molecular Basis for N-terminal Acetylation by Human NatE and Its Modulation by HYPK

This work is adapted from *Nat Commun.* (Deng, McTiernan, et al. 2020) © Nature Research

3.1 Introduction

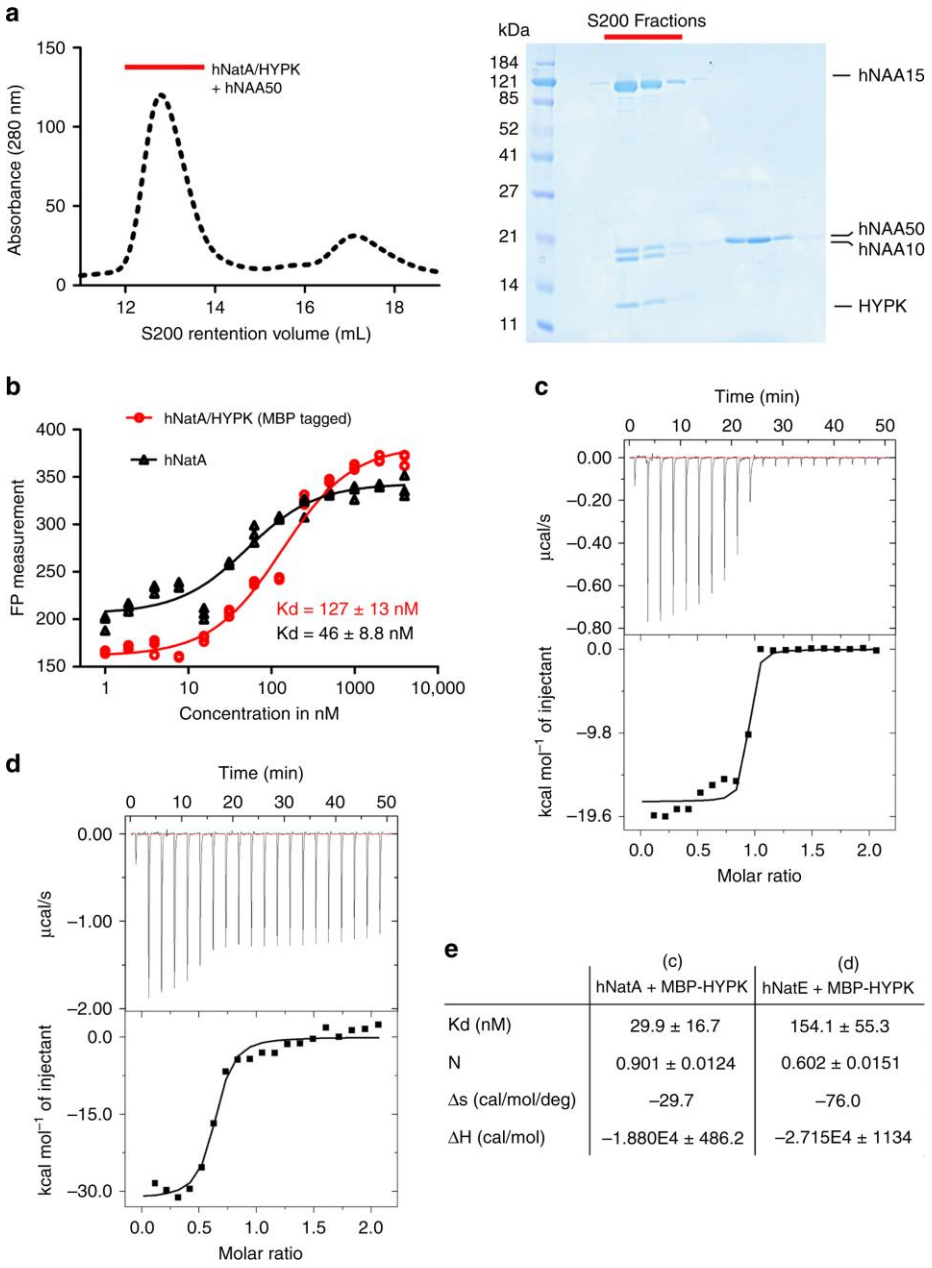
About one-half of the human N-terminal acetylome is mediated by hNatA. Mutations in either the NAA10 catalytic or NAA15 auxiliary subunits have been correlated with a broad spectrum of pathologies including intellectual disabilities, developmental delay, autism spectrum disorders, craniofacial dysmorphology, congenital cardiac anomalies and Ogden syndrome (Dorfel and Lyon 2015; Lee et al. 2010; Yoon et al. 2014; Saunier et al. 2016; Popp et al. 2015; Esmailpour et al. 2014; Casey et al. 2015; Cheng et al. 2019; Ree et al. 2019; Rope et al. 2011). Aberrant NatA activity has also been correlated with neurodegenerative disorders and cancer although a causative role of NatA is less clear (Kalvik and Arnesen 2013). NatA forms complexes with at least two other proteins, NAA50 and the chaperone protein Huntingtin interacting protein K (HYPK) (Arnesen et al. 2010; Gautschi et al. 2003; Arnesen, Anderson, et al. 2006; Liszczak et al. 2013; Weyer et al. 2017). The trimeric hNAA10/hNAA15/hNAA50 complex is referred to as hNatE, and we refer to the tetrameric hNAA10/hNAA15/hNAA50/HYPK complex as hNatE/HYPK. Previous studies have demonstrated that HYPK may be important for cellular NatA activity and that NAA50 and NatA can affect each other's function (Weyer et al. 2017; Arnesen et al. 2010; Gottlieb and Marmorstein 2018; Rathore et al. 2016; Deng et al. 2019). Previous studies also demonstrated that hNatA can physically associate with hNAA50 (Gautschi et al. 2003) and HYPK (Arnesen et al. 2010), separately. Moreover, the crystal structures of hNatA with HYPK and related pull-down experiments indicated that HYPK and hNAA50 binding to hNatA might be mutually exclusive (Gottlieb and Marmorstein 2018). However, another study suggested that HYPK and NatE form a tetrameric complex in *Drosophila* cells (Rathore et al. 2016).

To delineate the mechanistic basis for how the NAA10 and NAA50 catalytic subunits of the NatE complex coordinate function and how this is regulated by HYPK, we characterized the human NatE and NatE/HYPK complexes biochemically, structurally and in cells. We show that NAA50 and HYPK exhibit negative cooperative binding to NatA *in vitro* and in human cells, by inducing NAA15 shifts in opposing directions. We show that NAA50 and HYPK both mediate NAA10 inhibition through structural alteration of the NAA10 substrate binding site. We show that NatE is about 10-fold more active than NAA50, likely due to a reduced entropic cost for substrate binding through NatA tethering, but is inhibited by HYPK through structural alteration of the NatE substrate binding site. Taken together, these studies reveal the molecular basis for coordinated N-terminal acetylation by the NAA10 and NAA50 catalytic subunits of NatE and its modulation by HYPK.

3.2 Results

3.2.1 hNatE and HYPK form a tetrameric complex

To determine if human NAA50 and HYPK can simultaneously bind to NatA to form a NatE/HYPK complex, we prepared the recombinant hNatA/HYPK complex from insect cells and mixed it with recombinant hNAA50 for analysis on size exclusion chromatography (**Fig. 3.1a – left**). Analysis of the peak fractions on SDS-PAGE revealed that the four protein components (hNAA10, hNAA15, hNAA50 and HYPK) co-eluted in a single major peak with excess hNAA50 eluting in later fractions (**Fig. 3.1a**). This result demonstrates that HYPK and hNAA50 can bind to hNatA simultaneously to form a tetrameric hNatE/HYPK complex.



e

	(c) hNatA + MBP-HYPK	(d) hNatE + MBP-HYPK
Kd (nM)	29.9 ± 16.7	154.1 ± 55.3
N	0.901 ± 0.0124	0.602 ± 0.0151
Δs (cal/mol/deg)	–29.7	–76.0
ΔH (cal/mol)	–1.880E4 ± 486.2	–2.715E4 ± 1134

Figure 13.1 HYPK and hNatE form a tetrameric complex

(a) Left - gel filtration elution profile of the hNatA/HYPK complex with excess hNAA50, using a Superdex S200 column. Right - Coomassie-stained SDS-PAGE of peak fractions. Red bars indicated the peak of complex. (**b)** Fluorescence polarization assays with either hNatA or hNatA/MBP-HYPK titrated into fluorescein-5-maleimide labeled

hNAA50. The data is fit to calculate a dissociation constant (Kd). Replicates were shown in the curve with n = 3 independent experiments. Source data are provided as a Source Data file. **(c)** Representative ITC curve of MBP-HYPK titrated into hNatA. The calculated dissociation constant is indicated**. (d)** Representative ITC curve of MBP-HYPK titrated into hNatE. **(e)** The ITC fitting information and calculated dissociation constant (Kd) is provided for (c) and (d) curves.

3.2.2 HYPK and hNAA50 display negative cooperative binding to hNatA

Previous studies demonstrated that both human HYPK and hNAA50 bind tightly to hNatA, with dissociation constants in the nanomolar range (Gottlieb and Marmorstein 2018; Deng et al. 2019). We set out to determine if the HYPK and hNAA50 binding properties to hNatA are altered in the context of the tetrameric complex. Using fluorescence polarization (FP) assays, we observed that hNAA50 bound to hNatA with a K_d of 46 ± 8.8 nM (**Fig. 3.1b)**, consistent with previous results. However, in the presence of MBP-HYPK, this binding K_d decreased about ~3-fold to 127 ± 13 nM (**Fig. 3.1b)**. This data demonstrates that HYPK can negatively affect the stability of the hNatE complex by about 3-fold. Based on this data, we hypothesized that in the presence of hNAA50, the binding affinity between hNatA and HYPK would also decrease. Using isothermal titration calorimetry (ITC), we determined that the K_d between hNatA and MBP-HYPK was 29.9± 16.7 nM (**Fig. 3.1c)**, which is consistent with the previously reported K_d in the thermophilic fungus *Chaetomium thermophilum (Weyer et al. 2017)*.

The reason for the apparent two-binding transition when HYPK is titrated into hNatA is unknown. We further determined the K_d between HYPK and hNatE and observed that the affinity was 154 ± 55.3 nM, ~ 5-fold weaker than HYPK binding to hNatA (**Fig. 3.1d)**. A control ITC run of MBP titrated into hNatA failed to show binding**.** Taken together, these data indicate that HYPK and NAA50 display negative cooperative binding with respect to hNatA.

3.2.3 hNAA50 and HYPK inhibit hNatA activity, and HYPK is dominant

While our previous studies demonstrated that yeast and human NatA have slightly elevated catalytic efficiencies in the presence of NAA50, we set out to make a more direct comparison of NatA activity as a function of added NAA50 vs. HYPK. To do this, we used an *in vitro* acetylation assay using a cognate SASE NatA substrate to compare the hNatA catalytic activity as a function of added hNAA50 or MBP-HYPK. This comparison revealed that while MBP-HYPK showed about 90% inhibition of hNatA activity at an MBP-HYPK concentration beyond ~100 nM, hNAA50 only decreased hNatA activity by about 50% at concentrations between ~ 100 nM - 1µM (**Fig. 3.2a),** consistent with previous data (Gottlieb and Marmorstein 2018; Deng et al. 2019). Titrating both MBP-HYPK and hNAA50 into hNatA showed a similar degree to hNatA inhibition as did titrating in MBP-HYPK alone, illustrating that HYPK inhibition of hNatA overrides the incomplete inhibitory effect of hNAA50 (**Fig. 3.2a**).

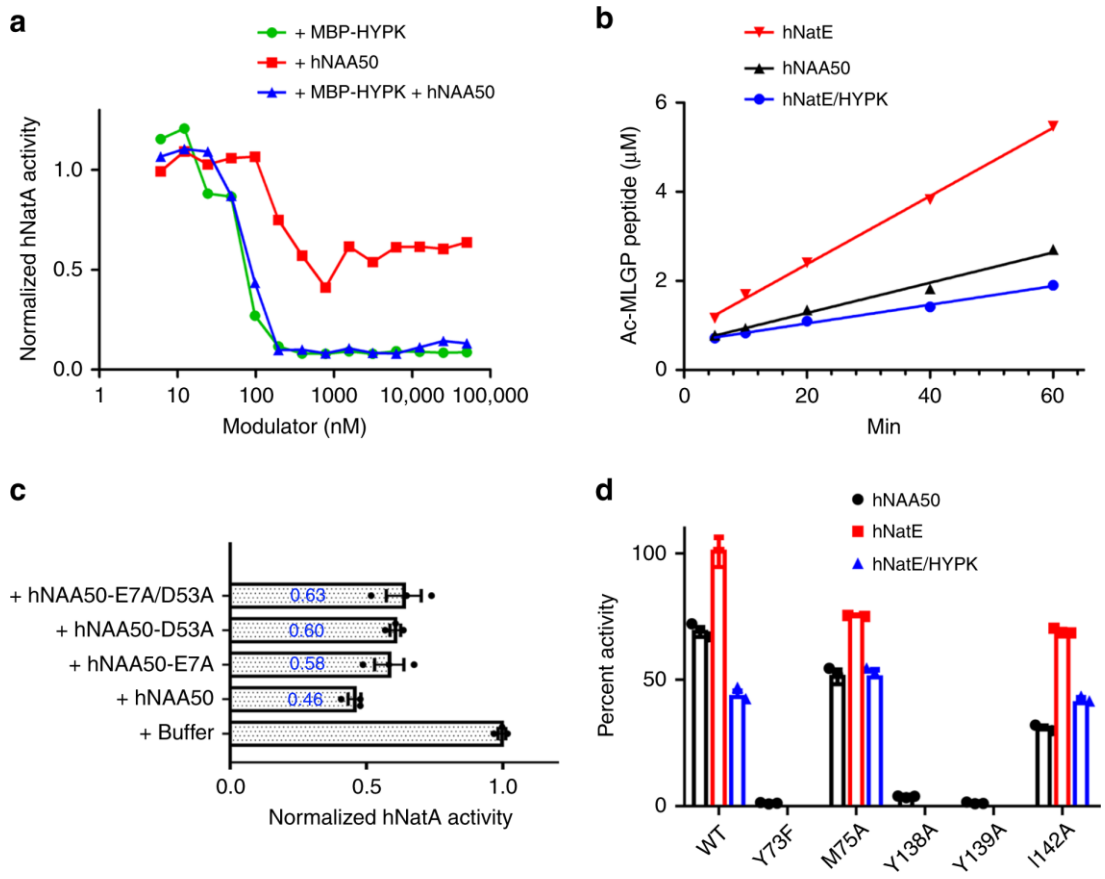


Figure 14.2 HYPK binding negatively affects hNatE acetylation activity

(a) Either MBP-HYPK (green), hNAA50 (red), or both (blue) are titrated into hNatA (100 nM) to evaluate their modulatory effect on hNatA activity against an SASE peptide substrate. A best line is drawn through the data points for ease of visualization. Source data are provided as a Source Data file. (**b)** Comparison of time course acetylation activity of hNAA50 (black), preformed hNatE (red), and preformed hNatE/HYPK (blue; all 500 nM) against an MLGP peptide substrate. Source data are provided as a Source Data file. (**c)** hNatA acetylation activity against an SASE peptide with addition of buffer, wild type (WT), or hNAA50 mutants. Errors were reported in SEM with $n = 3$ independent experiments. Source data are provided as a Source Data file. **(d)** Activity of hNAA50 WT and mutants against MLGP peptide was tested, either alone (black), in the context of

hNatE (+hNatA; red), or in the context of preformed hNatE/HYPK (+hNatA/HYPK; blue). Data were normalized to WT hNatE activity as 100%, represented as mean ± SD, $n = 3$ independent experiments.

3.2.4 HYPK inhibits hNatE activity

Given that HYPK was shown to be a negative regulator for hNatA acetylation activity (Weyer et al. 2017; Gottlieb and Marmorstein 2018), we asked if it also inhibited the activity hNatE. To do this, we used the same *in-vitro* acetyltransferase activity assay but employed a cognate MLGP hNatE substrate in the absence or presence of HYPK, as well as a hNAA50 only control. This analysis revealed that while hNatE was more active than hNAA50, this improvement was essentially nullified with the addition of HYPK (**Figs. 3.2b, 3.2c and 3.2d**). A full kinetic analysis of hNAA50, hNatE, and hNatE/HYPK revealed the kinetic parameters responsible for this modulation of hNAA50 function. Specifically, we found that hNatE displayed ~8.6-fold decrease of K_m, and ~1.1-fold decrease of k_{cat}, with an overall ~7.7-fold increase of catalytic efficiency, compared to hNAA50 (**Table 3.1)**. These findings are in general agreement with previous studies with the orthologous yeast proteins (Deng et al. 2019). A comparison of the activities of hNatE and hNatE/HYPK revealed that in the presence HYPK, hNatE displayed an ~1.3-fold decrease in K_m, and an ~3.8-fold decrease of k_{cat}, with an overall ~2.9-fold decrease in catalytic efficiency (**Table 3.1)**. This data suggests that HYPK can compromise hNatE activity. Taken together, while hNatA significantly enhances the catalytic efficiency of hNatE, HYPK binding to hNatE largely nullifies this effect, demonstrating that HYPK can indirectly inhibit hNatE activity.

Table 4.1 Catalytic parameters for hNAA50, hNatE, and hNatE/HYPK

	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (μM^{-1} min^{-1})
hNAA50*	3.45	831.5	4.15×10^{-3}
hNatE*	3.09	96.2	3.21×10^{-2}
hNatE/HYPK	0.81	73.1	1.11×10^{-2}

*Calculation of kinetic parameters for hNAA50 and hNatE is based on an independent replicate of experiments previously reported[46]. Source data are provided in the Source Data file.

3.2.5 hNatA forms competing complexes with hNAA50 and HYPK in cells

To further confirm the existence of the tetrameric hNatE/HYPK complex observed *in vitro* in human cells, we immunoprecipitated hNAA15 (c-terminal V5 tagged) from HeLa cells to evaluate the associated proteins. Mass spectrometry analysis of this sample revealed that endogenous hNAA10, hNAA50 and HYPK were all co-immunoprecipitated with NAA15-WT-V5, in agreement with the tetrameric hNatE/HYPK complex formation (**Fig. 3.3a)**.

In prior studies, we reported that a hNAA15-T406Y mutant can disassociate hNAA50 from hNatA *in vitro* (Deng et al. 2019), while a hNAA15-L814P mutant is defective for HYPK inhibition and reduces hNatA thermostability (Cheng et al. 2019). Based on these observations, we also carried out a mass spectrometry analysis of the V5-immunoprecipitates of the hNAA15-T406Y-V5 and hNAA15-L814P-V5 mutants to evaluate the effect of the mutation's on formation of the hNatE/HYPK tetrameric

complex. For relative quantification of the hNatA proteins, the IBAQ intensities of the hNatA components in each sample were normalized to the IBAQ intensity of hNAA15 in the respective sample and to the corresponding protein in the WT sample **(Fig. 3.3a)**. This analysis demonstrated that hNAA15-T406Y-V5 and hNAA15-L814P-V5 had reduced binding of hNAA50 and HYPK respectively, indicating that the mutations negatively affected the ability of hNAA15 to bind hNAA50 and HYPK, respectively. Furthermore, it appeared that the reduced binding of HYPK or hNAA50 resulted in greater binding of the other component: hNAA15-T406Y-V5 bound less hNAA50, but more HYPK than hNAA15-WT, while hNAA15-L814P-V5 bound less HYPK and more hNAA50 than hNAA15-WT (**Fig. 3.3a)**. These observations are in agreement with the *in vitro* FP and ITC assays, demonstrating that HYPK and hNAA50 display negative cooperative binding with respect to hNatA. Taken together, these cellular studies agree with the *in vitro* studies by confirming the existence of the tetrameric hNatE/HYPK complex and indicating that HYPK and hNAA50 display negative cooperative binding with respect to hNatA.

The intrinsic hNatA catalytic activities of the hNAA15-T406Y-V5 and hNAA15-L814P-V5 variants were also tested in a Nt-acetylation assay and westernbloted (**Fig. 3.3b)**. Importantly, equal levels of hNAA10 were pulled out with the lysates supporting that none of the mutants affect hNAA10 binding. The hNAA15-T406Y-V5 hNatA complex displayed a decreased catalytic activity towards the hNatA substrate $SESS_{24}$ compared to WT hNatA. This could be explained by an increased binding and inhibitory effect of HYPK and/or due to the reduced binding of hNAA50. The hNAA15-L814P-V5 hNatA complex showed an increased catalytic activity compared to WT hNatA, which is

consistent with the *in vitro* findings of reduced binding of HYPK and consequently less inhibition of the catalytic activity and/or an increased binding of hNAA50.

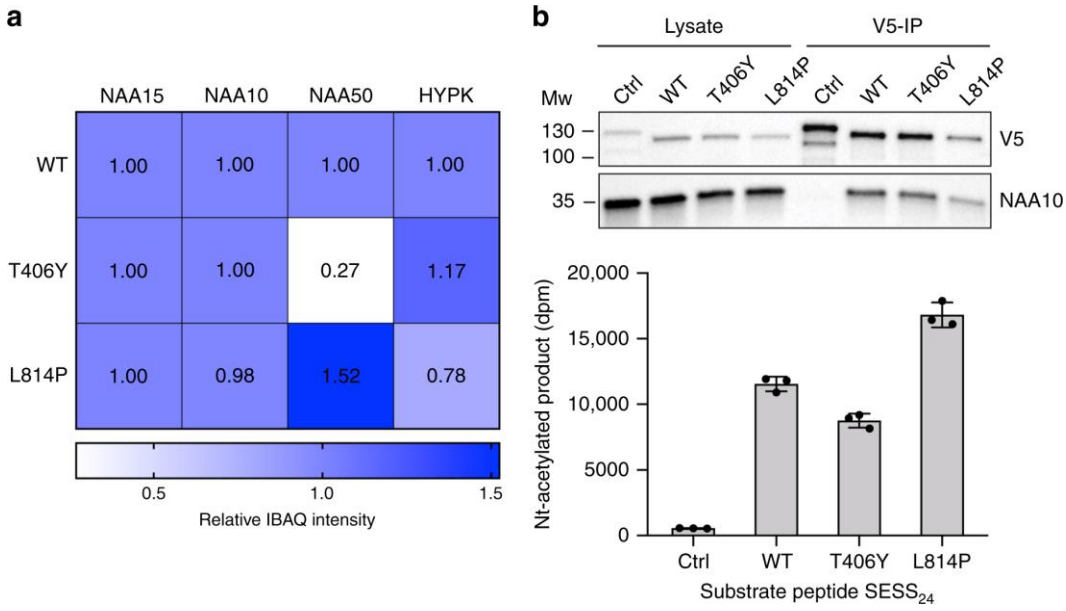


Figure 15.3 hNatA forms competing complexes with hNAA50 and HYPK

(a) Heat map showing the mean relative IBAQ intensity of NatA components from three independent MS analyses of immunoprecipitated NAA15 variants, NAA15-WT-V5, NAA15-T406Y-V5, and NAA15-L814P-V5. The IBAQ intensities of each component were normalized to the IBAQ intensity of NAA15 in the respective sample and to the corresponding NatA WT protein. Source data are provided as a Source Data file. (**b)** Western blot analysis and NatA Nt-acetylation assay of V5-immunoprecipiated NAA15 variants. The measured DPM signal for each reaction was normalized to the corresponding V5-band in the IP. The immunoprecipitation and activity assay were performed in $n = 3$ independent experiments, each with three technical replicates (indicated by dot plots) per assay.

3.2.6 hNatE structure reveals molecular basis for enzyme crosstalk

To determine the molecular basis for interaction and enzymatic crosstalk between human NAA10 and NAA50 enzyme subunits within the NatE complex, we determined the single particle Cryo-EM structure of the hNatE complex in the presence of inositol hexaphosphate IP_6 and bi-substrate analogues of both hNatA and hNAA50. The model of the human complex was rigid-body fitted with the coordinates of the crystal structure of hNatA (PDB: 6C9M) and hNAA50 (PDB: 3TFY), with all ligands removed. After refinement of the model, we observed clear density of IP_6 and acetyl-CoA cofactors in each catalytic subunit, but the peptide portion of the bisubstrate inhibitors were not resolved. The structure was refined to an overall resolution of 3.0 Å, with most residues of hNAA10 and hNAA50 subunits resolved to 2.5 Å. The refinement statistics can be found in **Table. 3.2**. The N-terminal 112 residues of hNAA15 subunit were not modeled due to a lack of traceable Cryo-EM density for this region.

Consistent with previous studies with the orthologous yeast proteins (Deng et al. 2019), hNAA50 predominantly interacts with hNAA15 (α21, α22, α23 and α24), and makes relatively few interactions with hNAA10 (**Fig. 3.4a),** burying a solvent excluded surface of 34,316 $Å^2$ with hNAA15, and 9,100 $Å^2$ with hNAA10. Based on the structure of the *Sc*NatE complex, a hNAA15-T406Y mutant was shown to disrupt the hNatA-NAA50 interaction (Deng et al. 2019). Indeed, we find that hNAA15-Thr406 is located at the center of the hNAA50-hNAA15 interface, making hydrogen bond interactions with the backbone carbonyl group of hNAA50-Ala55 (**Fig. 3.4b)**. Additional hydrogen bonds are formed between the backbone carbonyl of hNAA50-His14 with the backbone amide of hNAA15-Thr439, between hNAA50-Arg21 and the backbone carbonyl of hNAA15-Pro405, and between hNAA50-Asn52 and hNAA15-Thr371 (**Fig. 3.4b)**. hNAA50-Arg21

also makes direct ionic interaction with hNAA15-Glu433, which is absent in yeast (**Fig. 3.4b)**. It is possible that this ionic interaction underlies the increased salt sensitivity of the human vs. yeast NatE complex (Deng et al. 2019). Interestingly, we do not observe significant hydrophobic interactions at the hNAA15-hNAA50 contact interface.

hNAA50-hNAA10 interactions are less extensive than hNAA50-hNAA15, yet still significant. Interactions are observed between hNAA50-Glu7 and hNAA10-Arg116, and between hNAA50-Asp53 and hNAA10-Arg83 (**Fig. 3.4c)**. This structural observation highlights the recent identification of a missense mutation of hNAA10 R83H which exhibits decrease acetylation activity in boys with intellectual disability and developmental delay (Ree et al. 2019). It is noteworthy that hNAA10-Arg83 also makes interaction with the 3' phosphate of the acetyl-CoA molecule bound to hNAA10 (**Fig. 3.4c)**.

To assess the importance of the NatA-NAA50 interactions on the catalytic crosstalk of the two enzymes, we prepared hNAA50-E7A, -D53A and -E7A/D53A mutants and evaluated their effect on hNatA activity (**Fig. 3.2c**). We found that the single and double mutants were able to restore between 12-17% of hNatA activity relative to the activity of hNatA bound to WT-hNAA50 (**Fig. 3.2c)**. These results demonstrate that the observed hydrogen bond interactions between NatA and NAA50 make a small, albeit significant, contribution to the catalytic crosstalk between the two enzymes.

An overlay of the *Saccharomyces cerevisiae* (PDB: 6O07) and human NatE structures reveals a high degree of structural conservation with a root-mean square deviation (RMSD) of 1.517 Å (over 622 common Cα atoms). However, it is noteworthy

that when bound to NatA, we observe that NAA50 shifts significantly closer to NAA10 in the human over the yeast complex, resulting in more significant and intimate NAA10-NAA50 interactions in the human complex (**Fig. 3.5a**). α3 and β7 of hNAA50 shift toward NAA10 about 10 Å and 11 Å, respectively (**Fig. 3.5a**). Such significant shift in NAA50 position relative to NatA is surprising and may reflect a functional importance when these proteins are associated with the ribosome (see discussion).

To understand other structural contributions that hNAA50 might have on hNatA activity, we superimposed the hNatE structure with the hNatA crystal structure (PDB:6C9M). We found that hNAA15 and hNAA10 in the ternary hNatE complex displayed an RMSD of 0.747 Å (over 624 common Cα atoms), and 0.897 Å (over 141 common Cα atoms), respectively (**Fig. 3.5b**). The NAA10 β6-β7 loop displayed about a 4 Å shift toward hNAA50 in the hNatE complex relative to the hNatA complex (**Fig. 3.5b)**. Given that the β6-β7 loop of NAA10 plays an important role in peptide substrate recognition, we propose that this shift in position also contributes to the inhibitory effect that hNAA50 binding has on hNatA activity.

To understand the molecular basis for why hNatA binding increases the catalytic efficiency of hNatE by nearly 8-fold (**Fig. 3.2b and Table 3.1)**, we superimposed the hNAA50 crystal structure (PDB: 3TFY) with the hNAA50 subunits of the NatE structure. Remarkably, we did not observe any significant structural changes in hNAA50 with an overall RMSD for all atoms of 0.437 Å (**Fig. 3.5b**). Notably, the β6-β7 and α1-α2 peptide substrate binding loops of hNAA50 within the hNatE complex superimposed well with hNAA50 bound to CoA and substrate peptide **(Fig. 3.5b)**. Based on this observation, we hypothesize that the increased activity of hNatE is due to a reduced entropic cost for

substrate binding to the hNAA50 subunit due to hNatA tethering. This is consistent with the observations of Wand and colleagues for other protein systems (Caro et al. 2017).

Table 5.2 Cryo-EM data collection, refinement, and validation statistics

	20501 PDB: 6PW9	20442 PDB: 6PPL
Data collection and processing		
Magnification	36,000	165,000
Voltage (keV)	200	300
Electron exposure (e/Å^2)	40	40
Defocus range (μm)	−1.5 to −3.0	−1.5 to −3.0
Pixel size (Å)	1.169	0.832
Symmetry imposed	C1	C1
Initial particles(no.)	477,608	1,229,331
Final particles (no.)	168,536	353,541
Map resolution (Å)	4.03	3.02
FSC threshold	0.143	0.143
Map resolution range (Å)	3.5–5.0	2.5–4.5
Refinement		
Initial model used (PDB code)	6C95 and 3TFY	6C9M and 3TFY
Model resolution (Å)	4.3	3.1
FSC threshold	0.5	0.5
Model resolution range (Å)	—	—
Map sharpening *B* factor (Å^2)	−154.684	−119.78
Model composition		
Non-hydrogen atoms	8836	8155
Protein residues	1075	980
Ligands	2	3
B factors (Å^2)		
Protein	91.05	38.81
Ligand	68.44	46.42
RMSD		
Bonds lengths (Å)	0.004	0.006
Bond angles (°)	0.990	0.811
Validation		
MolProbity score	1.52	1.96
Clash score	4.71	3.88
Poor rotamers (%)	1.49	0.42
Ramachandran plot		
Favored (%)	96.53	95.65
Allowed (%)	100	100
Disallowed (%)	0	0

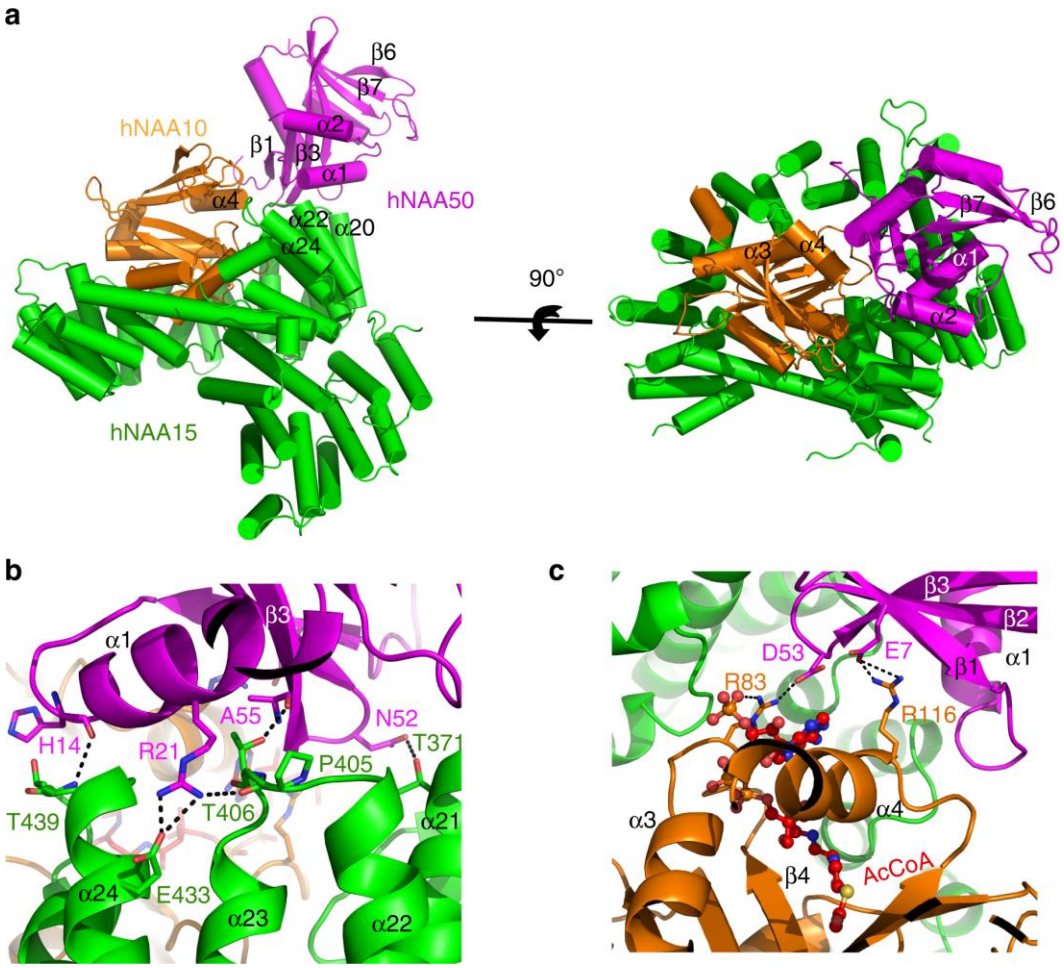


Figure 16.4 Cryo-EM structure of the hNatE complex

(a) hNAA50 (magenta), hNAA15 (green), and hNAA10 (orange) within the hNatE complex shown in cartoon. (**b)** Zoom-in view of the contacts between hNAA15 and hNAA50 with residues that participate in interaction labeled. (**c)** Zoom-in view of the contacts between hNAA10 and hNAA50 with residues that participate in interaction shown. Acetyl-CoA bound to hNAA10 is shown in ball and stick.

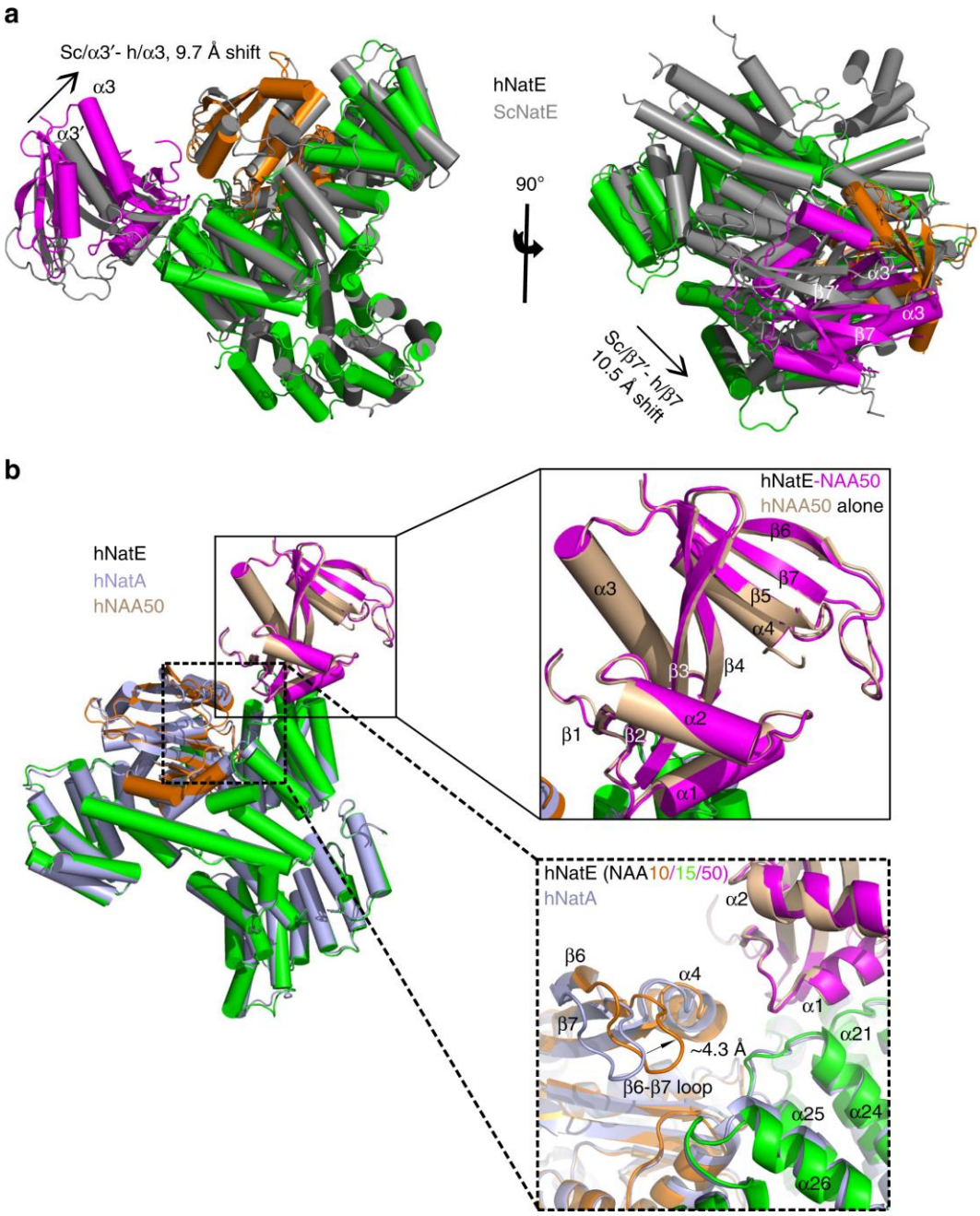


Figure 17.5 Subunit crosstalk within the hNatE complex

(a) hNAA50 (magenta), hNAA15 (green), and hNAA10 (orange) within the hNatE complex overlay with ScNatE (grey, PDB: 6O07). The α3 helix and β7 strand of Naa50 is shown to shift towards Naa10 in the human structure **(b)** hNatE aligned with hNatA (light

blue, PDB:6C9M) and hNAA50 (wheat, PDB: 3TFY). The top zoom-in area shows the alignment of free hNAA50 and hNatE. The below zoom-in area shows the hNAA10 conformational change induced by hNAA50 binding.

3.2.7 hNatE/HYPK structure reveals negative cooperative mechanism

To obtain a molecular understanding of how NAA50 and HYPK can both bind NatA we determined the structure of the hNatE/HYPK complex by Cryo-EM, which we were able to resolve to an overall resolution of 4.0 Å. The starting structure was modeled using the Cryo-EM NatE structure and the HYPK portion from the hNatA/HYPK crystal structure (PDB: 6C95), which was placed into the Cryo-EM map through rigid-body fitting into the EM density map, followed by adjustment and refinement. The refinement statistics can be found in **Table. 3.2**.

In the tetrameric structure, we observe that both hNAA50 and HYPK simultaneously dock onto the same binding regions on hNatA as previously identified in the absence of the other protein (Gottlieb and Marmorstein 2018). While both hNAA50 and HYPK contact both subunits of hNatA, no direct interactions are observed between hNAA50 and HYPK **(Fig. 3.6a)**. HYPK binds NatA mainly through interactions with the hNAA15 subunit of hNatA: only the α1 helix of HYPK interacts with hNAA10, while the α2 and C-terminal UBA domain (α3, α4, and α5 helices) of HYPK interact with hNAA15 (**Fig. 3.6a)**. Previous studies demonstrated that the UBA domain and α2 helix play key roles in hNatA binding, while α1 is essential for hNatA activity inhibition (Gottlieb and Marmorstein 2018; Weyer et al. 2017).

Over the HYPK and hNatA interaction interface within the tetrameric complex, we observe polar interactions between HYPK-Glu74 and hNAA15-Tyr158, between the backbone carbonyl of HYPK-Thr100 and hNAA15-Lys687, between the backbone carbonyl of HYPK-Asn129 and hNAA15-Arg697, between HYPK-Asn129 and hNAA15-Lys696, which were all overserved similarly in the hNatA/HYPK crystal structure (Gottlieb and Marmorstein 2018) (**Fig. 3.6b)**. However, HYPK-Glu103 , hNAA15-Lys685 and hNAA15-Glu655 which form a salt bridge in the hNatA/HYPK structure do not show a similar contact in the tetrameric hNatE/HYPK complex (Gottlieb and Marmorstein 2018) (**Fig. 3.6b)**. We speculate that the absence of this interaction might contribute to the slightly weaker affinity of HYPK for hNatE over hNatA.

hNAA15 and hNAA50 also make similar interactions in the absence or presence of HYPK. The backbone carbonyl of hNAA50-His14 hydrogen bonds with the backbone amide of hNAA15-Thr439, and hNAA50-Asn52 hydrogen bonds with hNAA15-Thr371 (**Fig. 3.6c)**. Interestingly however, a key residue that stabilizes the hNatA-hNAA50 interaction, hNAA15-Thr406, shifts from in interaction with the carbonyl of hNAA50-A55 to a hydrogen bond to the hNAA50-Gln18 side chain in the tetrameric complex (**Fig. 3.6c)**. An additional H-bond is formed between hNAA50-Gln18 and the backbone carbonyl of hNAA15-Thr406 (**Fig. 3.6c)**. hNAA50-Arg21, a residue that makes extensive contacts to NAA15 in the hNatA/NAA50 complex (**Fig. 3.4b**) is only contacting hNAA15-Glu433 in the tetrameric complex (**Fig. 3.6c)**. In addition, while hNAA50-Asp53 and hNAA10-Arg83 maintain interaction as observed in the hNatA/hNAA50 complex, hNAA50-Glu7 and hNAA10-Arg116 loose interaction in the tetrameric complex (**Fig. 3.6d)**. These observations correlate with the observed weaker binding interaction between hNAA50 and hNatA in the presence of HYPK.

To understand the molecular basis for the biochemical findings that the binding of hNAA50 and HYPK to hNatA exhibit negative cooperativity, we superimposed the ternary hNatA/HYPK and hNatE complexes onto the hNatE/HYPK complex **(Fig. 3.7).** To align the hNatE and hNatE/HYPK complexes we superimposed the hNAA50 position to delineate how the HYPK binding regions are affected by hNAA50 binding (**Fig. 3.7a)**. The overall overlay suggests that both the hNAA10 and hNAA15 subunits of hNatA shift in the direction that is away from the hNAA50 binding region to make closer and more optimal contacts with HYPK (**Fig. 3.7a)**. The hNAA15 α36-α40 helices, which bind to the C-terminal UBA domain (α3-α5) of HYPK in the tetrameric hNatE/HYPK complex, clearly shift away from the binding interface, when only hNAA50 is bound to hNatA (**Fig. 3.7c)**. Notably, α40, α38, α37, α36 display shift of ~4.5 Å, ~2.3 Å, ~1.6 Å, and ~1.8 Å, respectively. The important residues of hNAA15 (Lys696, Arg697, Lys687, Lys685), which mediate HYPK interactions in this region suggested that most of them in the hNatE structure exhibit an ~ 2 Å shift away from HYPK (**Fig. 3.7c)**. Meanwhile, hNAA15 N-terminal helices of α7, α8, α9, α10 which are in close contact with HYPK-α2 also show ~3.0 Å, ~4.1 Å, ~1.9 Å, ~3.5 Å shifts, respectively (**Fig. 3.7d)**, while Tyr 158 is about ~2 Å further away. Lastly, we observe that α2 and β3- β4 of hNAA50-bound hNAA10 move ~2.5 Å closer to contact the HYPK N-terminal α1 helix (**Fig. 3.7e)**. Thus, it appears that hNAA50 binding to hNatA, destabilizes hNatA-HYPK interactions through moving hNatA away from its optimal binding position to HYPK.

Similarly, a superposition of hNatE/HYPK with the crystal structure of hNatA/HYPK by aligning to the HYPK position to delineate how the hNAA50 binding regions are affected by HYPK binding reveals that hNAA15 is shifted away from HYPK and closer to hNAA50 in the hNatE/HYPK complex (**Fig. 3.7b)**. A further zoom-in view of

the hNAA50 binding regions in hNAA15 demonstrates that residues Glu433, Thr406, Thr371 and Thr439 of hNAA15 are shifted by ~ 2-3 Å away from hNAA50 when HYPK is bound relative to their positions in the absence of hNAA50 (**Fig. 3.7f)**. Regarding the hNAA50-hNAA10 binding interface, a 4.3 Å shift of hNAA10-Arg116 was observed, with a 1.7 Å for hNAA10-Arg83 (**Fig. 3.7g)**. It is also noteworthy that significant conformational shift of the C-terminal hNAA15 helices α41- α45 (~2.2 - 4.2 Å) was induced in order to have hNAA50 bound in the tetrameric hNatE/HYPK complex (**Fig. 3.7h**). This suggests that HYPK bound hNatA is also not optimal for hNAA50 binding.

Taken together, these comparisons reveal that hNAA50 and HYPK destabilize the binding of the other protein to hNatA, despite their independent binding surfaces to hNatA, and explain the observed negative cooperative binding of hNAA50 and HYPK to hNatA.

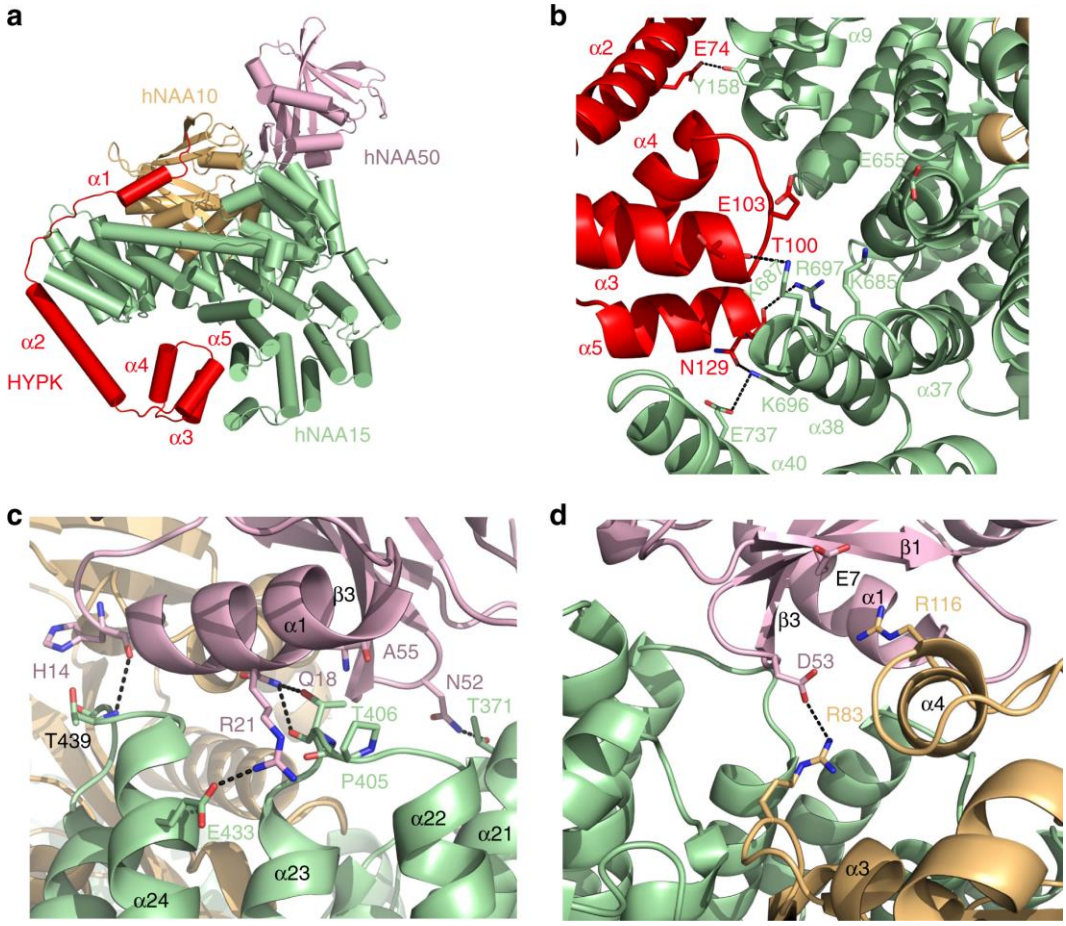


Figure 18.6 Overall structure of the hNatE/HYPK complex

(a) hNAA50 (pink), hNAA15 (green), hNAA10 (orange), and HYPK (red) within the hNatE complex is shown in cartoon. (**b)** Zoom-in view of the contacts between HYPK and hNAA15 with residues that participate in interaction shown. (**c)** Zoom-in view of the contacts between hNAA15 and hNAA50 with residues that participate in interaction shown. (**d)** Zoom-in view of the contacts between hNAA10 and hNAA50 with residues that participate in interaction shown.

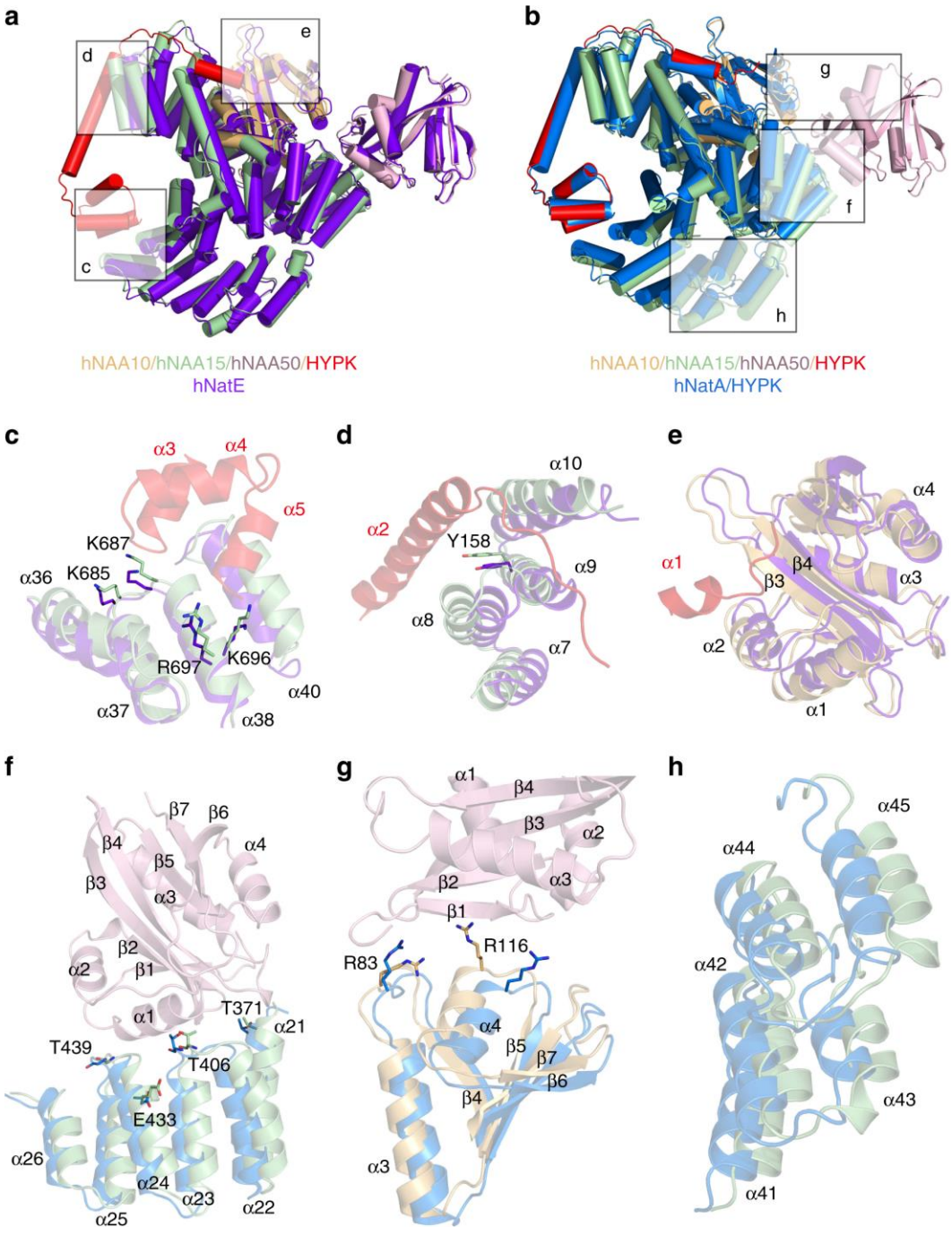


Figure 19.7 Molecular basis for NAA50 and HYPK binding to NatA

(a) hNatE/HYPK overlayed onto hNatE (purple) with NAA50 aligned. **(b)** hNatE/HYPK overlayed onto hNatA/HYPK (PDB: 6C95, blue) with HYPK aligned. **(c)** zoom-in view of HYPK C-terminal binding region as indicated in (a). (**d)** zoom-in view of HYPK α2 binding region as indicated in (a). (**e)** zoom-in view of HYPK N terminal α1 domain binding region as indicated in (a). **(f)** zoom-in view shows the NAA50 binding region on NAA15 as indicated in (b). (**g)** zoom-in view shows the NAA50 binding region on NAA10 as indicated in (b). **(h)** zoom-in view shows the NAA15 C-terminal helices conformal changes when HYPK bound as indicated in (b).

3.2.8 Molecular basis for decrease of hNatE activity by HYPK

Our kinetic data demonstrated that the catalytic efficiency of hNatE is greater than hNAA50 but that this increase is partially nullified in the presence of HYPK (**Table 3.1 and Fig. 3.2b**). To understand the molecular basis for this, we overlay structures of hNatE and hNatE/HYPK with the common hNatA subunits aligned. This superposition reveals that the largest structural differences within the hNAA50 subunit map to the β6-β7 and α1-α2 loops, which shift such that the hNAA50 peptide substrate binding groove is ~2.4 Å wider in the hNatE/HYPK complex relative to the hNatE complex (**Fig. 3.8)**. In addition, the hNAA50 β3-β4 loop is shifted in the direction away from hNatA by about 3.5 Å in the presence of HYPK (**Fig. 3.8)**. Previous studies demonstrated that Tyr73 and Met75 from hNAA50-β4 and Tyr138, Tyr139 and Leu142 from the hNAA50 β6-β7 loop are important for hNAA50 catalytic function and substrate binding (Liszczak, Arnesen, and Marmorstein 2011). To directly test this, we prepared the following mutants: hNAA50-Y73F, -M75A, -Y138A, -Y139A and -I142A, and tested their effects on hNatE

activity when HYPK was bound (**Fig. 3.2d)**. Consistent with previous studies, Y73F, Y138A and Y139A decreased hNAA50 activity to undetectable levels (Liszczak, Arnesen, and Marmorstein 2011). For hNAA50-M75A and -I142A, we observed increased activity when bound to hNatA to form the hNatE complex, as we observed for hNAA50-WT. As expected, the hNatE/HYPK complex displayed reduced activity compared to both hNAA50 alone and hNatE. However, we observed that hNatE-M75A/HYPK and hNatE-I142A/HYPK exhibited similar activity, compared to their corresponding hNAA50 mutants alone. This data is consistent with the conclusion that perturbation of the hNAA50 β3-β4 and β6-β7 loops within the hNatE/HYPK complex underlies the observed reduced catalytic efficiency of hNatE in the presence of HYPK.

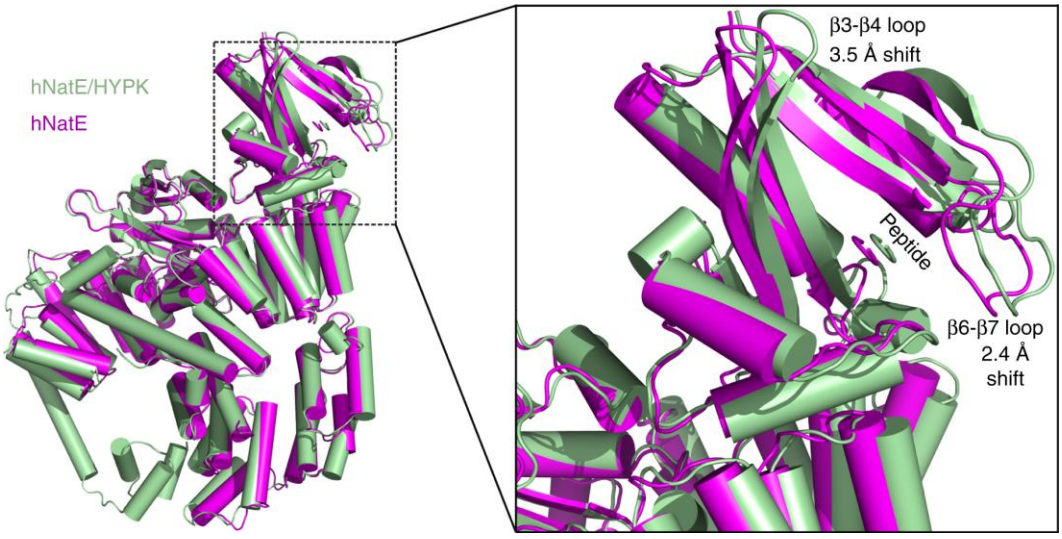


Figure 20.8 Molecular basis for the decrease of NAA50 activity by HYPK

hNatE/HYPK (green) overlayed onto hNatE (magenta) with hNatA aligned. Zoom-in view shows the local conformational change of hNAA50.

3.3 Discussion

While the interaction and catalytic crosstalk within NatE in the yeast system had previously been reported, the corresponding human system and the influence of HYPK binding on hNatE activity had not previously been characterized. Here, we biochemically and structurally characterize the mechanistic interplay of the hNAA10 and hNAA50 catalytic subunits within the hNatE complex and the role of HYPK in regulating hNatE activity. Similar to the crystal structure of *Sc*NatE (PDB:6O07), we find that hNAA50 binds hNatA mainly through hNAA15, using highly conserved residues, including a key threonine residue located at the center of the binding interface. Notably, the distance between hNAA50 and hNAA10 in the human complex is significantly shorter than in the yeast complex. We find that these two acidic hNAA50 residues, Glu7 and Asp53 contact two hNAA10 arginine residues, inducing a conformational change in hNAA10, which contributes to a decrease in hNatA enzymatic activity.

While both hNAA50 and hNatA are conserved from yeast to human, it had long been considered that HYPK is not present in yeast until the recently identified HYPK in thermophilic fungus *Chaetomium thermophilum (Weyer et al. 2017)*. Nevertheless, it appears that evolutionarily NAA50 appeared as a NatA binding partner earlier than HYPK did. In both budding and fission yeast, NAA50 is inactive but regulates NatA acetylation activity (Deng et al. 2019). While in higher eukaryotes, NAA50 is enzymatically active, either with or without NatA (Liszczak, Arnesen, and Marmorstein

2011; Evjenth et al. 2009). Thus, within the tetrameric NatE/HYPK complex, it is likely that the regulation of NatA activity by NAA50 was replaced by HYPK.

Here, we have reconstituted the tetrameric hNatE/HYPK complex to demonstrate that HYPK and hNAA50 can bind to hNatA simultaneously. We also confirmed the physiological relevance of the tetrameric hNatE/HYPK complex through immunoprecipitation of the complex from HeLa cells. We find that while HYPK inhibits hNatA activity directly, it also indirectly, but less potently, inhibits hNatE. We also find that hNAA50 and HYPK exhibit negative cooperativity with respect to hNatA binding *in vitro* and *in vivo*. The structure of the hNatE/HYPK complex suggests that this is due to the ability of the hNAA50 and HYPK protein to bind to the hNAA15 subunit of hNatA in a way that indirectly destabilizes binding of the other protein. This observation, together with our findings that HYPK has a dominant regulatory effect on hNatA activity relative to hNAA50 may explain why excess free hNAA50 is observed over hNatA in higher organism and HeLa cells (Hou et al. 2007).

The nano-molar dissociation constants of HYPK and hNAA50 for hNatA is consistent with our immunoprecipitation of the tetrameric hNatE/HYPK complex from cells. Previous studies in *Drosophila* also reported that a dNAA50 knockout decreases *in-vivo* dNatA catalytic activity in the presence of HYPK (Rathore et al. 2016). In terms of the biological function of this tetrameric complex, we propose two non-mutually exclusive regulatory mechanism for how hNAA50 protein levels could affect hNatA activity. First, since hNAA50 has a negative effect on binding between hNatA and HYPK, the absence of hNAA50, would promote HYPK binding to hNatA to inhibit hNatA activity. Second, since a recent yeast NatE-ribosome structure demonstrated that yeast NAA50 participates in ribosome association (Knorr et al. 2019), it is likely that hNAA50 in higher

eukaryotes also contributes to the association between tetrameric hNatE/HYPK and ribosome. The lower level of hNAA50 could therefore decrease the fraction of ribosome associated hNatA to regulate its co-translational activity.

The overall structures of hNAA50 and ScNAA50 are well conserved with a mix of □-helices and □-strands. However, the overlay of hNatE and ScNatE suggested that NAA50 shifts closer to NAA10 in the human over the yeast complex. It is possible that in the human system NAA10 and NAA50 have more intimate contact and crosstalk in enzymatic activity. However, previous data indicated that the degree of NAA50 and NAA10 activity crosstalk is similar in yeast and human (Deng et al. 2019). In a recently reported yeast ScNatE-ribosome structure, the ScNAA10 and ScNAA50 catalytic sites are 50 Å and 85 Å away from the peptide exit site on the ribosome (Knorr et al. 2019). While unlike ScNAA50, hNAA50 is co-translationally enzymatically active (Evjenth et al. 2009). Thus, we propose that this shift of NAA50 closer to NAA10 in the human over the yeast complex may have an impact on the relative distance between the nascent chain and active sites of hNAA10 and hNAA50.

With respect to hNatA binding to the ribosome, it is possible that HYPK may also contribute to ribosome binding by hNatA. This could then indirectly affect hNAA50 recruitment to the ribosome, which would be promoted by hNatA but inhibited by HYPK. HYPK contains a UBA domain, which is also present at the C-terminus of the alpha subunit of the Nascent Polypeptide-Associated Complex (NAC). Further studies are required to test these hypotheses for hNatA recruitment to the ribosome.

The hNatE/HYPK complex co-translationally acetylates nascent peptides with N termini of either methionine maintained or cleaved, which accounts for the largest

number of substrates among all NATs. Given that this complex acetylates about 40-60% of the human proteome and has altered function in many human diseases, the studies presented here provides an important molecular scaffold for potential therapeutic development.

3.4 Materials and methods

3.4.1 Plasmid construction.

A mammalian expression vector *pcDNA3.1/NAA15-V5* was modified by site-directed mutagenesis (Q5® Site Directed Mutagenesis Kit, New England Biolabs) to encode hNAA15 variants c.1216-1218ACA>TAC (p.T406Y) and c.2441T>C (p.L814P). Primer sequences for p.T406Y are *AAGTACACCTTACTTAATAGAACTCTTTCTCGTG* and *TCAATAGCAGTATTTATGTAC*. Primer sequences for p.L814P are *TGATGGTAGCCCAGGAGACTGTA* and *TACAAGGCTTCCAATACC*. The mutated plasmid sequences were verified by sequencing.

3.4.2 Proteomics sample preparation of NAA15-V5 immunoprecipitates.

For immunoprecipitation of NAA15-V5 variants, two 10 cm dishes of HeLa cells (ATCC CCL-2) were transfected with 10 µg of either *pcDNA3.1/NAA15-WT-V5*, *pcDNA3.1/NAA15-T406Y-V5* or *pcDNA3.1/NAA15-L814P-V5* using X-tremeGENE™ 9 DNA Transfection Reagent (Roche). The growth medium was replaced after 24 h. 48 h post transfection, cells were harvested and lysed in 200 µl IPH lysis buffer (50 mM Tris-

HCl pH 8.0, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1× complete EDTA-free protease inhibitor cocktail (Roche)) per cell dish for 15 min at 4°C on a rotating wheel. Cell debris was pelleted by centrifugation (17000 × g, 4°C, 5 min) and the supernatant was mixed with 2 µg of V5-tag mouse monoclonal antibody (Invitrogen, R960-25) and incubated for 2 h at 4°C on a rotating wheel. Thereafter, 20 µl of Dynabeads™ Protein G (Invitrogen) washed in IPH lysis buffer was added to the cell lysates and incubated overnight. The next day, beads were washed three times in IPH lysis buffer before the retrieved immunocomplexes were eluted in 60 µl FASP buffer (2% SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT) and heated for 5 min at 95 °C. Filter-aided sample preparation (FASP) method was performed to process the eluates for LC-MS/MS analysis (Wisniewski et al. 2009). The protein samples were mixed with 200 µl UA buffer (8 M urea, 100 Mm Tris-HCl pH 8.0), transferred to Microcon 30kDa MWCO filters and centrifuged. All centrifugation steps were carried out at 23 °C and 14 000 x g for 15 min. The filters were washed three times with 200 µl UA by spinning. Proteins were then Cys-alkylated by incubation with 100 µl 50 mM iodoacetamide (IAA) in UA for 20 min, before IAA was removed by centrifugation. Filters were then washed three times with 100 µl UA. Finally, the UA buffer was exchanged by three washes with 100 µl 50 mM ammonium bicarbonate (ABC), before 500 ng trypsin (Sequencing Grade Modified Trypsin, Promega) was added to the filters and incubated overnight at 37 °C. Digested proteins were collected by centrifugation and the filter was washed once with 75 µl ABC. Peptides were acidified with 5% formic acid (FA) and desalted using Pierce™ C18-Tips (Thermo Scientific) according to manufacturer´s protocol. The final eluates were dried by speed vacuum and diluted to desired concentration with 5% FA.

3.4.3 MS analysis of NAA15-V5 immunoprecipitates.

Mass spectrometric analysis was performed with an Ultimate 3000 RSLC system (Thermo Scientific) coupled to a Q-Exactive HF mass spectrometer (Thermo Scientific) equipped with EASY-spray nano-electrospray ion source (Thermo Scientific). About 1 µg tryptic peptides were loaded and desalted on a pre-column (Acclaim PepMap 100, 2cm x 75µm ID nanoViper column, packed with 3 µm C18 beads) with 0.1% TFA (flow rate 5 µl/min, 5 min). Peptides were separated during a biphasic ACN gradient from two nanoflow UPLC pumps (flow rate of 200 nl/min) on an analytical column (PepMap RSLC, 50cm x 75 µm i.d. EASY-spray column, packed with 2 µm C18 beads). Solvent A and B were 0.1% FA (vol/vol) in water and 100% ACN respectively. The gradient composition was 5% B for 5 min, 5-8% B for 0.5 min, 8–24% B for 109.5min, 24–35% B for 25min, and 35–80% B for 15 min, 80% B over 15 min for isocratic elution and 5% B over 20 min for conditioning. The eluting peptides were ionized in the electrospray and analyzed by the Q-Exactive HF. The mass spectrometer was operated in data-dependent mode to automatically switch between full scan MS and MS/MS acquisition. MS spectra (m/z 375-1500) were obtained with a resolution of 120 000 at m/z 200, automatic gain control (AGC) target of 3 x 106 and maximum injection time (IT) of 100 ms. The twelve most intense peptides above a threshold of 50 000 counts and charge states 2 to 5 were isolated (window of 1.6 m/z, AGC target of 1 x 105 and maximum IT of 110 ms) for fragmentation at a normalized collision energy of 28%. Fragments were detected in the orbitrap at a resolution of 15 000 at m/z 200, with first mass fixed at m/z 100. Precursor masses selected for MS/MS analysis were excluded by dynamic exclusion for 25 s with “exclude isotopes” enabled. Lock-mass internal calibration (m/z 445.12003) was used. The raw data acquired was processed with MaxQuant v. 1.6.2.6 and Andromeda search

engine. The spectra were searched against a database of Swiss-Prot annotated human protein sequences (20431 sequences, retrieved 25.06.2019) and a reverse decoy database. Cystein carbamidomethylation was selected as a fixed modification and variable modifications included methionine oxidation and protein N-terminal acetylation. The false discovery rate (FDR) was set to 1% for peptide and protein identification, minimum peptide length allowed was 7 and match between runs (0.7 min match time window, 20 min alignment time window) was enabled. Label-free quantification (LFQ) (Cox et al. 2014) and intensity-based absolute quantification (IBAQ) were selected. All other parameters were set to default values. The resulting proteingroups.txt file was analysed using Perseus software v. 1.6.5.0. Proteins only identified by site, common contaminants and reverse hits were filtered away. The IBAQ intensities of hNAA10, hNAA50 and HYPK were normalized to the IBAQ intensity of NAA15 in each IP sample. Thereafter, the IBAQ intensity of hNAA15, hNAA10, hNAA50 and HYPK in each sample was normalized to the IBAQ intensity of the corresponding protein in the hNAA15-WT-V5 IP sample.

3.4.4 [^{14}C]-Ac-CoA-based acetylation and western blot analysis

In order to test the NatA Nt-acetylation activity of hNAA15-T406Y and hNAA15-L815P variants, immunoprecipitation of hNAA15-V5 was performed as described above with minor adjustments. Five 10 cm dishes of HeLa cells (ATCC CCL-2) were transfected with 8 µg *pcDNA3.1/NAA15-WT-V5*, 8 *µg pcDNA3.1/NAA15-T406Y-V5* or 10 µg *pcDNA3.1/NAA15-L814P-V5*. *NAA15-WT-V5* and *NAA15-T406Y-V5* were co-transfected with 2 µg of empty *pcDNA3.1/V5*

vector to ensure equal conditions for the cells. As a negative control, five dishes of HeLa cells were transfected with 10 µg of *pcDNA3.1/LacZ-V5*. Immunoprecipitation from harvested and lysed cells was performed using 3 µg of V5-tag mouse monoclonal antibody (Invitrogen) and 30 µl of Dynabeads™ Protein G (Invitrogen, R960-25). The beads were washed three times in IPH lysis buffer and resuspended in 95 µl acetylation buffer (50 mM Tris-HCl pH 8.5, 1 mM EDTA, 10% Glycerol). [14C]-Ac-CoA–based Nt-acetylation assays were performed as described (Drazic and Arnesen 2017). In brief, three reaction mixture replicates were prepared containing 10 µl of immunoprecipitated enzyme, 200 µM synthetic 24-mer oligopeptide $SESS_{24}$: NH_2-SESSSKSRWGRPVGRRRRPVRVYP-COOH (BioGenes), 50 µM [^{14}C]-Ac-CoA (Perkin-Elmer) and acetylation buffer to a final volume of 25 µl. Reaction mixtures without oligopeptide were used as negative controls. The reaction mixtures were incubated at 37 °C at 1400 rpm shaking for 30 min. The reaction was stopped by isolating beads on a magnet and transferring 23 µl of the supernatant onto P81 phosphocellulose filter discs (Millipore). The filter discs were washed three times in 10 mM HEPES buffer (pH 7.4) and air dried, before they were added to 5 ml Ultima Gold F scintillation mixture (Perkin-Elmer) and the incorporated [^{14}C]-Ac was measured by a Perkin-Elmer TriCarb 2900TR Liquid Scintillation Analyzer. The immunoprecipitate input was determined by Western blot analysis. Proteins were separated by SDS-PAGE, transferred onto a nitrocellulose membrane (Amersham Protran 0.2 µM NC) and probed with V5-tag mouse monoclonal antibody (1:5000, Invitrogen, R960-25) and NAA10 rabbit monoclonal antibody

(1:1000, Cell Signaling #13357). Protein bands were imaged by ChemiDocTM XRS+ system (Bio-Rad) and quantified using ImagelabTM Software (Bio-Rad). The measured DPM signal for each reaction was normalized to the amount of NAA15-V5 in the respective IP sample.

3.4.5 Protein expression and purification

N-his tagged hNatA, N-his tagged hNatA/HYPK, were expressed in sf9 cells (ThermoFisher, cat# 12659017) and purified as described previously (Gottlieb and Marmorstein 2018). hNAA50 (Liszczak, Arnesen, and Marmorstein 2011) and MBP-HYPK (Gottlieb and Marmorstein 2018) were expressed in BL21 (DE3)*E. Coli* cells (Thermo Scientific) and Rosetta (DE3)pLysS *E. coli* cells (Millipore Sigma), respectively, and purified as described previously. For hNatA and hNatA/HYPK sf9 expression, High density (2 x106 cells/ml) suspension cultures of Sf9 cells were infected at a multiplicity of infection (MOI) of 1 for 48hrs in Fernbach Shake flasks at 27 °C. Cells pellets were resuspended in lysis buffer of 25 mM Tris, pH 8.0, 500 mM NaCl, 10 mM Imidazole, 10 mM β-ME, 10 mg/ml PMSF (phenylmethanesulfonylfluoride), DNase, and complete, EDTA-free protease inhibitor tablet (Roche). After sonication, clarified lysate was passed on nickel resin (Thermo Scientific), washed with 10 CV of lysis buffer, and eluted with 25 mM Tris, pH 8.0, 200 mM NaCl, 200 mM Imidazole, 10 mM β-ME. Elute was further purified with HiTrap SP ion-exchange column with the salt gradient (200mM to 1M). Peak fractions were pooled and run on a Superdex 200 Increase 10/300 GL gel filtration column in sizing buffer containing 25 mM HEPES, pH 7.0, 200 mM NaCl, and 1 mM TCEP. Peak fractions were pooled and flash-frozen for storage in -80 °C until use.

For *E. coli* protein expression, transformed cells were cultured at 37 °C until the absorbance A_{600} reached ~ 0.7, induced with 0.5 mM IPTG (isopropyl 1-thio-β-D-galactopyranoside), and grown overnight at 16 °C. *E. coli* cells overexpressing MBP-HYPK were lysed in 25 mM Tris, pH 8.0, 150 mM NaCl, 10 mM β-ME, 10 mg/ml PMSF. Clarified lysate was passed on to amylose agarose resin (New England Biolabs), washed with lysis buffer, and elute in lysis buffer supplementary with 20 mM maltose. Eluted protein was purified with a 5 mL HiTrap Q ion-exchange column (GE Healthcare) in the same buffer with a gradient (150 mM to 1M NaCl). Peak fractions were pooled and loaded to a Superdex 75 Increase 10/300 GL gel filtration column (GE Healthcare) in buffer of 25 mM HEPES, pH 7.0, 200 mM NaCl, and 1 mM TCEP. Peak fractions were pooled and flash-frozen for storage in -80 °C until use. *E. coli* cells over-expressing GST tagged NAA50 were lysed in 25mM HEPES pH 7.5, 100 mM NaCl, and 10 mM β-ME by sonication. The supernatant was isolated and loaded to GST-binding resin (Clontech), washed with 10 CV lysis buffer. TEV (tobacco etch virus protease) was added to the resin for on-column cleavage overnight at room temperature. Untagged Naa50 was washed off the column with lysis buffer and collected for overnight dialysis in buffer with 25mM HEPES (pH 7.5), 50 mM NaCl, and 10 mM β-ME. Ion exchange was done with a 5-ml HiTrap SP ion exchange column (GE Healthcare) in gradient of 50 – 750 mM NaCl. Peak fractions were further purified with Superdex 75 gel filtration column (GE Healthcare) to homogeneity in buffer containing 25 mM HEPES, pH 7.5, 100 mM NaCl, and 10 mM DTT (dithiothreitol). Protein was flash-frozen for storage in -80 °C until use. Protein harboring mutations was generated with QuickChange protocol (Stratagene) (Braman, Papworth, and Greener 1996) and obtained following the same expression and purification protocol as wild type. Primers for making hNAA50-E7A: *AAGGTAGCCGGATCGCGCTGGGAGATGTG* and

CACATCTCCCAGCGCGATCCGGCTACCTT. For hNAA50-D53A: *CTTGCCTATTTCAATGCTATTGCTGTAGGTGC* and *GCACCTACAGCAATAGCATTGAAATAGGCAAG.* For hNAA50-M75A: *GACTTTACATCGCGACACTAGGATG* and *CATCCTAGTGTCGCGATGTAAAGTC*. For hNAA50-Y138A: *ACAAAGAAGAACGCCTATAAGAGGATAG* and *CTATCCTCTTATAGGCGTTCTTCTTTGT*. hNAA50-Y73F, hNAA50-I142A, hNAA50-Y139A were generated as previously described (Liszczak, Arnesen, and Marmorstein 2011).

3.4.6 Acetyltransferase activity assays

Acetyltransferase assays were modified from previous studies (Liszczak et al. 2013; Liszczak, Arnesen, and Marmorstein 2011) and carried out at room temperature in a reaction buffer containing 75 mM HEPES, pH 7.0, 120 mM NaCl, 1 mM DTT. The SASE substrate peptide (NH_2-SASEAGVRWGRPVGRRRRP-COOH; GenScript) and the MLGP substrate peptide (NH_2-MLGPEGGRWGRPVGRRRRP-COOH; GenScript) were used to determine the enzymatic activity of hNatA and hNAA50, respectively. Each curve was repeated at least 3 times. To test the effect of hNAA50 and HYPK on hNatA activity, 100 nM of hNatA was mixed with 500 µM SASE peptide, 300 µM C^{14} labeled acetyl-CoA (4 mCi $mmol^{-1}$; PerkinElmer Life Sciences), and varied concentrations of hNAA50 or HYPK modulator were added for 12-minute reactions. Signals were normalized against enzyme without modulator. For time course activity assays, 500 nM of hNAA50, hNatE, or hNatE/HYPK were mixed with 500 µM MLGP peptide, and 300 µM C^{14} labeled acetyl-CoA (4 mCi $mmol^{-1}$; PerkinElmer Life Sciences) in a 20 µL

reaction volume. Only 15 µL of each reaction mixture was quenched at specific times. For kinetic assays of hNAA50, hNatE and hNatE/HYPK against MLGP peptide, 300 nM enzyme was mixed with 300 µM C^{14} labeled acetyl-CoA (4 mCi $mmol^{-1}$; PerkinElmer Life Sciences) and the concentration of the peptide substrate was varied for a 40-minute reaction. Data was fit to a Michaelis–Menten equation in GraphPad Prism for determination of kinetic parameters.

3.4.7 FP binding assays

Fluorescence polarization (FP) binding assays were performed essentially as described previously (Deng et al. 2019). 10 nM of fluorescein labeled hNAA50 (Deng et al. 2019) was used in all reactions, and hNatA or hNatA/MBP-HYPK concentrations were varied to determine the dissociation constant (K_d). 5 mg/mL BSA and 0.2% v/v Tween were added into the reaction buffer (25 mM HEPES, pH 7.0, 200 mM NaCl and 10 mM DTT) to prevent non-specific binding. FP readings were taken with a Perkin Elmer EnVision and each curve was repeated in triplicate. GraphPad Prism (version 5.01) was used for all data fitting to determine K_d. A single-site specific binding model was used, with equation 1,

$$F_i = F_0 + F_1 \times \frac{K_d + [M] + [L] - \sqrt{(K_d + [M] + [L])^2 - 4 \times [M] \times [L]}}{2 \times [M]} \quad (1)$$

where F_i is the fluorescence reading at ligand (either hNatA or hNatA/HYPK) concentration $[L]_i$; K_d is the equilibrium dissociation binding constant; $[M]$ is the concentration of hNaa50; F_0 is the fluorescence reading extrapolated to no ligand; F_1 is

the maximum fluorescence increase at saturating ligand concentration. Fit parameters for hNatA/HYPK (MBP tagged) binding were F_0 = 161.3 ± 3.3, F_1 = 222.0 ± 5.2, K_d =127 ± 13 nM, and R^2 = 0.98. For hNatA, F_0 = 205.8 ± 4.4, F_1 = 137.2 ± 5.4, K_d = 46 ± 8.8 nM, and R^2 = 0.95.

3.4.8 ITC measurements

ITC measurements were carried out using a MicroCal iTC200 at 20 °C. Samples were dialyzed into buffer containing 25 mM HEPES pH 7.0, 200 mM NaCl, 1 mM DTT. Protein samples (hNatA, hNatE) with concentrations of 15 μM in the cell and 150 μMof MBP-HYPK in the syringe were used in the experiments. The raw data was analyzed with the MicroCal ITC analysis software.

3.4.9 Cryo-EM sample preparation and data collection

To prepare hNatE complex, purified hNatA and three molar excess of hNAA50 were mixed and loaded onto a Superdex 200 10/30 GL column (GE Healthcare). Peak fractions with all subunits present, as confirmed with SDS-PAGE analysis, were concentrated to 1 mg/ml. Fresh sample incubated with 3 molar excess of both hNatA and hNAA50 bi-substrate analogues for ~30 min on ice was applied to Quantinfoil R1.2/1.3 holey carbon support grids. To prepare hNatE/HYPK complex, purified hNatA/HYPK and three molar excess of hNAA50 were mixed and loaded onto an S200 gel-filtration column (GE Healthcare). Peak fractions with all subunits present, as confirmed with SDS-PAGE were concentrated to 1 mg/ml. Fresh sample with 3 molar

excess of acetyl-CoA and hNAA50 bi-substrate analogues for ~30 min in ice was applied to Quantinfoil R1.2/1.3 holey carbon support grids.

Both sample grids were blotted for 10-12 seconds (blot force = 2) under 100% humidity at 16°C before the sample was plunged into liquid ethane, using a FEI Vitrobot Mark IV. An FEI TF20 was used for screening the grids and data collection was performed either with a Talos Arctica microscope or Titan Krios equipped with a K2 Summit direct detector (Gatan).

3.4.10 Cryo-EM data processing

Original image stacks were summed and corrected for drift and beam-induced motion at the micrograph level using MotionCor2 (Zheng et al. 2017). Defocus estimation and the resolution range of each micrograph were performed with Gctf (Zhang 2016). For hNatA/hNAA50, ~3,000 particles were manually picked to generate several rough 2D class averages. Representative 2D classes were used to autopick ~1,229,331 particles from 4,025 micrographs using Relion (Kimanius et al. 2016). All particles were extracted and binned to accelerate the 2D and 3D classifications. After bad particles were further removed by 2D and 3D classification, 353,541 particles were used for auto refinement, particle polishing and per particle CTF refinement. The final map of the hNatE complex was refined to an overall resolution of 3.02 Å, with local resolution estimated by Resmap (Kucukelbir, Sigworth, and Tagare 2014). For the hNatE/HYPK complex, image processing workflow was similar as described above. 477,608 particles were picked from 1,004 micrographs and 168,536 particles were used for refinement. The final resolution of the hNatE/HYPK complex was 4.03 Å.

3.4.11 Cryo-EM model building and refinement

For hNatE model building, the crystal structures of hNatA (PDB: 6C9M) and hNAA50 (PDB: 3TFY) were fit into the 3.02 Å EM map as rigid bodies, followed by manual adjustment in coot (Emsley and Cowtan 2004) and real-space refinement in PHENIX (Adams et al. 2010). For the hNatE/HYPK model building, the crystal structure of hNatA/HYPK (PDB: 6C95) and hNAA50 (PDB: 3TFY) were used. All representations of Cryo-EM density and structural models were performed with Chimera (Pettersen et al. 2004) and PyMol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.).

CHAPTER 4: Molecular Basis for N-Terminal Alpha-Synuclein Acetylation by Human NatB

4.1 Introduction

NatB is conserved from yeast to human in both complex composition and in its substrate specificity profile (Starheim, Gevaert, and Arnesen 2012). In *Saccharomyces cerevisiae*, the deletion of NatB subunits produces more severe phenotypes, compared to the knockout of NatA or NatC subunits. Deletion of either NAA20 or NAA25 leads to similar phenotypes including slower growth rate, diminished mating, defects in actin cable formation, and aberrant mitochondrial and vacuolar inheritance (Polevoda et al. 2003). These observations suggest that the proper function of actin and tropomyosin requires NTA by the intact NatB complex (Polevoda et al. 2003). In humans, disruption of NatB (hNatB) by knockout leads to defects in proper actin cytoskeleton structure, cell cycle progression and cell proliferation (Ametzazurra et al. 2008; Starheim et al. 2008; Ametzazurra et al. 2009; Neri et al. 2017). In addition, hNatB is upregulated in human hepatocellular carcinoma (Ametzazurra et al. 2008), where it has been suggested as a potential therapeutic target as silencing of this complex can block cell proliferation and tumor formation (Neri et al. 2017). hNatB-mediated NTA of α-synuclein (αSyn) has been shown to increase αSyn stability and lipid binding, and to reduce aggregation capacity (Watson and Lee 2019; Mason et al. 2016; Maltsev, Ying, and Bax 2012; Dikiy and Eliezer 2014; Trexler and Rhoades 2012; Fernández and Lucas 2018; Fauvet et al. 2012; Kang et al. 2012; Iyer et al. 2016). Since αSyn is a key protein in Parkinson's disease (PD) (Halliday et al. 2011; Spillantini et al. 1998), hNatB might play an indirect role in PD pathogenesis *in vivo* as supported by a recent study (Vinueza-Gavilanes et al. 2020). It was also recently demonstrated that NTA of αSyn increases its propensity for lipid membrane binding without altering its structural properties of the bound state (Runfola et al. 2020).

Compared to the comprehensive structural and biochemical characterization of NatA (Liszczak et al. 2013; Weyer et al. 2017; Gottlieb and Marmorstein 2018; Deng et al. 2019; Knorr et al. 2019; Magin et al. 2017), the study of NatB has been limited, particularly in humans. Recently, the crystal structure of *Candida albicans* (*Ca*) NatB bound to a bisubstrate CoA-peptide conjugate was determined, providing important insights into substrate specificity and NTA by caNatB (Hong et al. 2017). However, hNAA20 and hNAA25 share only ~40% and ~20% sequence identity with the *Candida albicans* homologue (Hong et al. 2017), respectively. Moreover, nearly all biological studies of NatB have been conducted in *Saccharomyces cerevisiae* (Lee et al. 2014; Caesar, Warringer, and Blomberg 2006; Singer and Shaw 2003), Arabidopsis (Ferrandez-Ayela et al. 2013; Huber et al. 2020), mouse (Ohyama et al. 2012) and human (Neri et al. 2017; Starheim et al. 2008; Ametzazurra et al. 2008; Ametzazurra et al. 2009) as model organisms. As a result, the mode of human NatB-mediated catalysis and αSyn-specific NatB recognition remains unresolved. In this study, we report the 3.5 Å resolution cryo-electron microscopy (Cryo-EM) structure of the ~130 KDa hNatB bound to a bisubstrate CoA-αSyn conjugate, together with a structure-guided analysis of mutational effects on catalytic activity. This analysis reveals functionally important structural differences between hNaB and related NAT enzymes, as well as insights into the molecular mechanisms that define αSyn and related substrates that are recognized for hNatB-mediated N-terminal acetylation.

4.2 Results

4.2.1 hNatB is potently inhibited by a CoA-aSyn conjugate

While attempts to express recombinant hNatB in *E. coli* were unsuccessful, we found that overexpression of hNatB complex with full-length hNAA25 (residues 1-972) and C-terminally truncated hNAA20 (residue 1-163 out of 178 total residues) in baculovirus-infected Sf9 insect cells produced soluble protein that could be purified to homogeneity (**Fig. 4.1A)**. To evaluate the activity of the recombinant hNatB, we tested it against different peptide substrates. αSyn with an N-terminal sequence of “MDVF” has been widely considered as an *in vivo* hNatB substrate (Anderson et al. 2006; Ohrfelt et al. 2011; Theillet et al. 2016; Van Damme et al. 2012). We therefore incorporated this sequence into a peptide substrate named “MDVF” for an *in vitro* acetyltransferase assay (“MDVF” peptide sequence: NH_2-MDVFMKGRWGRPVGRRRRP-COOH). In agreement with *in vivo* studies (Theillet et al. 2016; Anderson et al. 2006; Ohrfelt et al. 2011; Van Damme et al. 2012), we observed that the purified recombinant hNatB was active against this “MDVF” peptide, while no activity could be observed in the absence of either the enzyme or peptide (**Fig. 4.1B)**. hNatB also showed no observable activity if either the first residue “M” or the first two residues “MD” in this αSyn peptide substrate was removed (“DVFM” peptide sequence: NH_2-DVFMKGLRWGRPVGRRRRP-COOH; “VFMK” peptide sequence: NH_2- VFMKGLSRWGRPVGRRRRP-COOH) (**Fig. 4.1B)**, suggesting that peptide substrate recognition by NatB is highly dependent on the first two N-terminal residues. To further confirm the substrate specificity of hNatB, we tested it against several previously identified peptide substrates for other NATs (“SASE” peptide sequence (NatA-type): NH_2-SASEAGVRWGRPVGRRRRP-COOH; “MLRF” peptide sequence (NatC-type): NH_2-MLRFVTKRWGRPVGRRRRP-COOH; “SGRG”/H4

peptide sequence (NatD-type): NH_2-SGRGKGGKGLGKGGAKRHR-COOH; “MLGP” peptide sequence (NatE-type): NH_2-MLGPEGGRWGRPVGRRRRP-COOH) (**Fig. 4.1B)**. Consistent with previous results (Van Damme et al. 2012), hNatB is only active toward its unique canonical substrate type, displaying no overlapping activity towards other NAT substrates.

In order to understand the mechanism of hNatB substrate recognition, we synthesized a bi-substrate inhibitor in which the first 10 residues of αSyn are covalently linked to CoA (Liszczak et al. 2013) for enzymatic and structural studies. Half-maximum inhibitory concentration (IC_{50}) determinations revealed that this CoA-αSyn conjugate had an IC_{50} of about 1.63 ± 0.13 μM (**Fig. 4.1C)**, significantly lower than the K_m values we had determined for hNatB toward a “MDVF” peptide (45.08 ± 3.15 μM) and acetyl-CoA (47.28 ± 5.70 μM) (**Table 4.1).**

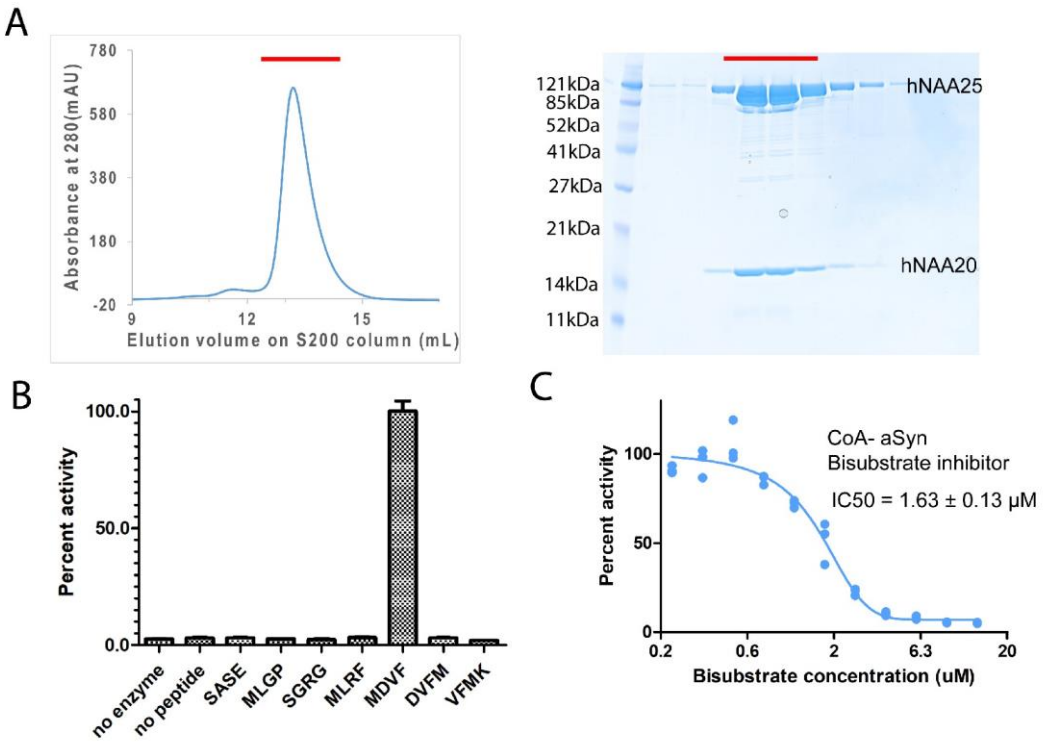


Figure 21.1 hNatB is active toward an α-Synuclein peptide and can be inhibited by a CoA-Syn conjugate

A. Gel filtration elution profile of hNatB, using a Superdex S200 column. Coomassie-stained SDS-PAGE of peak fractions is reproduced to the right of the chromatograms. **B.** Comparison of hNatB activity toward different peptide substrates. All the activities are normalized to the activity of hNatB toward αSyn peptide (MDVF). **C.** Dose-response curve corresponding to the titration of CoA-αSyn conjugate (CoA-MDVFMKGLSK) into hNatB acetyltransferase reactions. The calculated IC_{50} value is indicated. Reactions were performed in triplicate; replicates are shown in the graph as vertical dots.

Table 6.1 Catalytic parameter of wild-type hNatB and mutants

Substrate	Protein	K_{cat} (min^{-1})	K_{cat} (normalized to WT)	K_m (μM)	K_m (normalized to WT)	K_{cat}/K_m (normalized to WT)
Acetyl-CoA	WT	9.25 ± 0.29	1	47.28 ± 5.70	1	1
MDVF peptide	WT	7.63 ± 0.14	1.0	45.08 ± 3.15	1.0	1.0
	E25A	8.31 ± 0.35	1.1	39.30 ± 6.62	0.87	1.3
	Y27A	16.73 ± 2.11	2.2	75.03 ± 33.26	1.7	1.3
	H73A	0.89 ± 0.14	0.12	54.76 ± 32.81	1.2	0.10
	R84A	13.86 ± 1.75	1.8	320.8 ± 96.7	7.1	0.25
	R85A	18.65 ± 1.02	2.4	109.6 ± 19.33	2.4	1.0
	G87A	14.86 ± 0.56	1.9	78.16 ± 10.25	1.7	1.1
	N116A	0.90 ± 0.06	0.12	39.86 ± 11.84	0.88	0.14
	Y123A	0.34 ± 0.03	0.045	9.04 ± 3.81	0.20	0.23
	Y123F	0.94 ± 0.06	0.12	42.43 ± 11.52	0.94	0.13
	Y137A	6.43 ± 0.23	0.84	53.67 ± 9.63	1.2	0.70
	Y138A	3.09 ± 0.25	0.40	44.23 ± 13.87	0.98	0.41

4.2.2 hNatB reveals potentially biologically significant structural differences with hNatA and caNatB

We performed single particle Cryo-EM of hNatB in the presence of the CoA-αSyn conjugate. A 3.46 Å-resolution Cryo-EM three-dimensional (3D) map was determined from 982,420 particles, selected from 5,281 raw electron micrographs (**Table 4.2, Fig. 4.2)**. The central core region of the EM map contains excellent side chain density with a local resolution of ~ 2.5 Å, particularly around the catalytic subunit, hNAA20.

Consistent with previous NAT structural studies, the atomic model of hNatB features a catalytic subunit, hNAA20, that adopts a canonical Gcn5-related N-Acetyltransferase (GNAT) fold (Neuwald and Landsman 1997). Additionally, the model

reveals that the auxiliary subunit hNAA25 is composed of a total of 39 α-helices among where the predicted first and second α-helices were built as poly-alanine due to a lack of resolvable side-chain density (**Fig. 4.2)**. The 39 α-helices can be roughly divided into three groups: an N-terminal region: α1- α8; a core region: α9- α29; and a C-terminal region: α30-α39 **(Fig. 4.2A)**. The N terminal region (residue 1-164) displays relatively weak EM density compared to other regions, suggesting that it is relatively flexible. The eight helices of the N terminal region form four helical bundle tetratricopeptide repeat (TPR) motifs, which often participate in protein-protein interactions. While there are no visible contacts between the N-terminal TPR motifs and hNAA20, it is possible that this region participates in ribosome association, similar to the N-terminal region of the NAA15 auxiliary subunits of *Schizosaccharomyces pombe* (Magin et al. 2017) and *Saccharomyces cerevisiae* (Knorr et al. 2019) NatA. The 21 helices of the core region also form a number of TPR motifs, which come together to form a ring that completely wraps around and extensively contacts hNAA20 within its hollow center **(Fig. 4.2A)**. Indeed. the interaction between hNAA20 and the TPR motifs of this core region buries a total interface area is about 2300 $Å^2$. In the core region, it is noteworthy that there is a long α-helix (α28, ranging 30 residues) that traverses almost from one side of the complex to the other. The α28 helix closes the core ring structure, locking hNAA20 in position, and bridging the N- and C-terminal and regions. This is similar to the role played by α29-α30 of the hNAA15 auxiliary subunit of hNatA (**Fig. 4.2A and 4.2B).** The C terminal region features helices that bundle together to protrude out of the plane of the core ring structure at an angle of ~ 45° (**Fig. 4.2B**).

The sequence identity of the catalytic and auxiliary subunits of hNatA and hNatB are 20%, and 15 %, respectively. To understand how this translates to key

structural differences, we superimposed the crystal structure of hNatA (PDB: 6C9M) with our model. Between the catalytic subunits, there is a high degree of superposition (1.151 Å root-mean-square deviation [RMSD] over 105 common C_α atoms), except for an additional helix, α5, on the C-terminus of hNatA-NAA10, which is absent in hNatB-NAA20 **(Fig. 4.2B)**. Between the auxiliary subunits, the core and N-terminal regions of both hNatA-NAA15 and hNatB-NAA25 display similar topology, although a higher degree of deviation than the catalytic subunits. The core regions both wrap around their respective catalytic subunits (8.369 Å RMSD over 262 common C_α atoms), while the N-terminal regions jut off to the side (7.360 Å RMSD over 55 common C_α atoms). In contrast, the C-terminal regions of hNatA-NAA15 and hNatB-NAA25 diverge significantly from one another. For hNA25, the C-terminal region of hNAA25 is oriented towards its N-terminal region, while the C-terminal region of hNAA15 is positioned ~ 50Å away from the relative position of the superimposed C-terminal domain of hNAA25 (**Fig. 4.2B)**. The positioning of hNAA25 may serve to promote hNAA25 intra-termini communication, which is similar to the interaction of hNatA and its regulatory protein HYPK (Gottlieb and Marmorstein 2018). HYPK, which does not interact with hNatB, interacts with both the N- and C-terminal domains of hNatA-NAA15, potentially serving as bridge to enable closer communication between these two domains (**Fig. 4.2)**. Recent reports have described the role of the small molecule IP_6 (inositol hexaphosphate) in hNatA activity, where it is found to act as “glue” between the C-terminal and core domains in hNAA15 and hNAA10 via a series of hydrogen bonds and electrostatic interactions (Gottlieb and Marmorstein 2018; Cheng et al. 2019). While no corresponding small molecules have been identified to play a similar role in hNatB, our model shows that this interaction is replaced by an extended loop that connects the □31 helix with the □32 helix of hNatB-NAA25. This loop, which is not present in hNatA-NAA15, appears to mediate

hydrophobic interactions between hNatB-NAA25 and -NAA20, likely to serve a similar role as IP_6**.**

We also compared the structures from human and the previously described *C. albicans* NatB (*Ca*NatB, PDB: 5K18). Although the two superimposed structures revealed a high degree of structural conservation (NAA20: 0.698 Å RMSD over 125 common C_α atoms; NAA25 Core region: 3.267 Å RMSD over 266 common C_α atoms), the N-terminal region of hNatB-NAA25 appears to overlay more closely to hNatA-NAA15 than to *Ca*NatB-NAA25 (**Fig. 4.2B and C)**. Compared to *Ca*NAA25, the N-terminal regions of hNatB-NAA25 and hNatA-NAA15 are positioned more closely to the peptide substrate binding sites of the respective catalytic subunits. Based on the role that the N-terminal yeast NatA-Naa15p regions play in ribosome docking (Magin et al. 2017; Knorr et al. 2019), we propose that the relative shift in position of the N-terminal regions of the human NAT auxiliary subunits, hNAA15 and hNAA25, may reflect a difference in mechanism for ribosome association and co-translational NTA in *C. albicans* compared with humans. In addition, the overlay of C terminal regions of hNAA25 and CaNAA25 displays a RMSD of 15.960 Å over 133 common C_α atoms. We observe that the main difference that contributes to this deviation in this region is the length of helices.

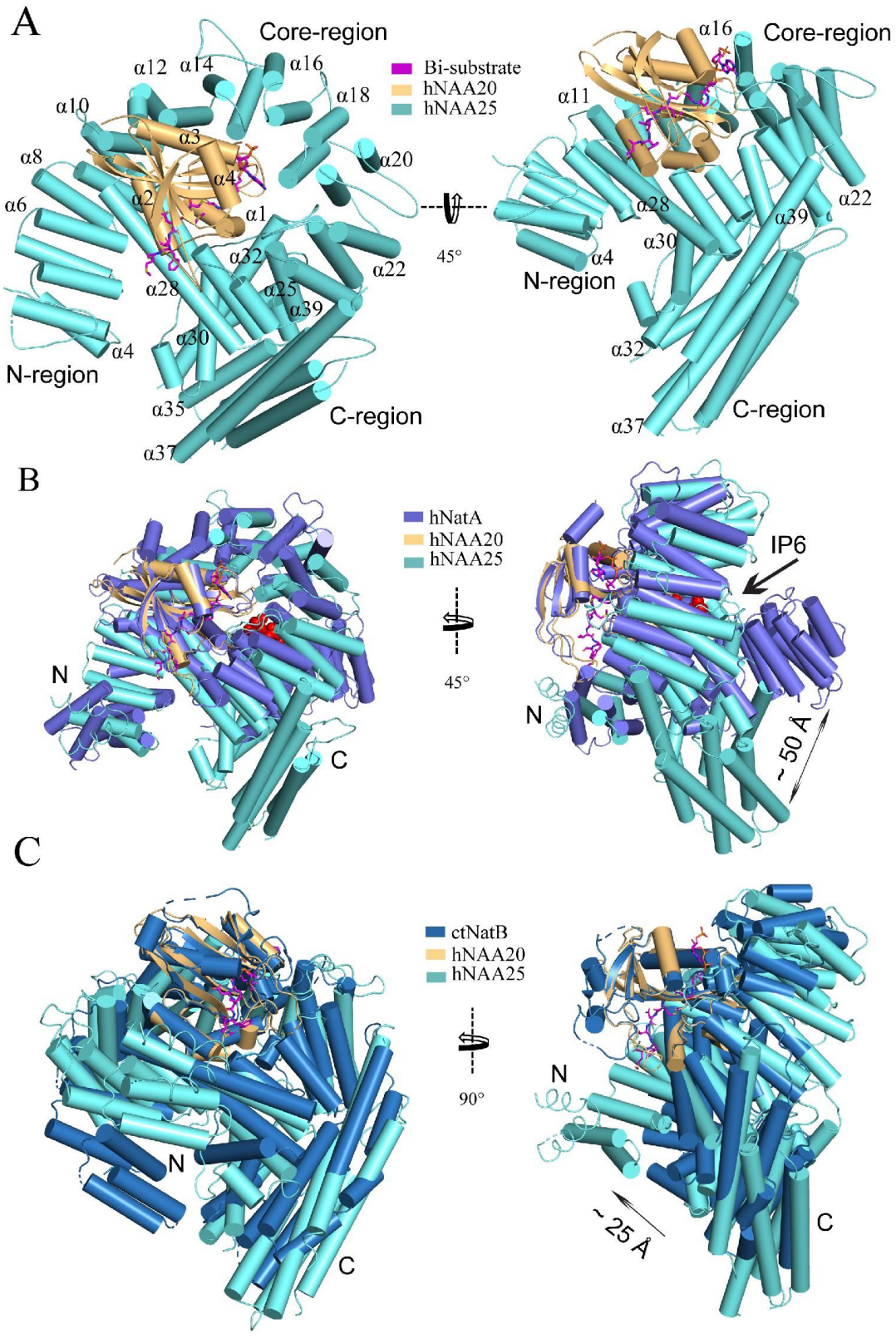

A
Core-region
α12
α14
α16
α18
α10
α3
α8
α20
α4
α2
α6
α1
α32
α22
α28
α25
α39
α30
α4
N-region
α35
C-region
α37
Bi-substrate
hNAA20
hNAA25
45°
α16
Core-region
α11
α28
α22
α39
α30
α4
N-region
α32
C-region
α37
B
hNatA
hNAA20
hNAA25
45°
N
C
IP6
N
~ 50 Å
C
ctNatB
hNAA20
hNAA25
90°
N
C
N
~ 25 Å
C

Figure 22.2 hNatB shows structural differences with hNatA and C. albicans NatB

hNaa20 (light orange) and hNaa25 (cyan) are shown in cartoon. The CoA-□Syn conjugate inhibitor is shown in sticks and colored as magenta. The N- and C- terminal regions are indicated as “N” and “C”, respectively. Some helices are as labeled. **B.** hNaa20 (light orange) and hNaa25 (cyan) are shown overlapped with hNatA (marine blue, PDB: 6C9M). Small molecule IP_6 bound to hNatA is shown as surface representation (red). **C.** hNaa20 (light orange) and hNaa25 (cyan) are shown superimposed on *Ca*NatB (slate blue, PDB:5K04).

Table 7.2 Cryo-EM data collection, refinement, and validation statistics

Data collection and processing	
Magnification	105,000
Voltage (keV)	300
Electron exposure (e/Å^2)	40
Defocus range (μm)	−1.5 to −2.5
Pixel size (Å)	0.83
Symmetry imposed	C1
Initial particles (no.)	1,927,675
Final particles (no.)	982,420
Map resolution (Å)	3.46
FSC threshold	0.143
Map resolution range (Å)	2.5–4.5
Refinement	
Initial model used (PDB code)	-
Model resolution (Å)	3.5
FSC threshold	0.5
Model resolution range (Å)	-
Map sharpening *B* factor (Å^2)	−191.177
Model composition	
Non-hydrogen atoms	8611
Protein residues	1077
Ligands	2
B factors (Å^2)	
Protein	2.51/84.68/33.39
Ligand	17.07/19.91/19.74
R.M.S. deviations	
Bonds lengths (Å)	0.011
Bond angles (°)	1.105
Validation	
MolProbity score	1.73
Clash score	3.98
Poor rotamers (%)	0.55
Ramachandran plot	
Favored (%)	90.10
Allowed (%)	9.71
Disallowed (%)	0.19

4.2.3 hNAA25 and hNAA20 make intimate interactions within hNatB

The hNatB/CoA-αSyn structure reveals an extensive interaction interface between the core region of the auxiliary hNAA25 and catalytic hNAA20 subunits (**Fig. 4.3A**). The most intimate contact between the two proteins are made by the α28 - α29 segment of hNAA25 and almost the entire length of hNAA20 α2, creating a large

hydrophobic interface (**Fig. 4.3B**). Residues that contribute to interaction include Thr26, Gly28, Ile29, Pro30, Leu33, Gln34, Leu36, Ala37, His38, Glu41 of hNAA20 and Lys535, His536, Ile537, Phe569, Asp576, Thr577, Tyr580, Ala584, Tyr 587, Lys592, Phe596, Phe599, Leu603 of hNAA25 (**Fig. 4.3B**). Another region of interaction, involving predominantly van der Waals interaction occurs between the hNAA20 β4- α3 loop and the hNAA25 α18- α19 loop. Here, residues Lys362, Pro363, Lys362-Thr367, Cys364, and Pro363 from hNAA25 and residues Arg85, Phe83, Glu82 and Arg84 from hNAA20 contribute to this small batch of hydrophobic interface (**Fig. 4.3C**).

Additional intimate contacts between hNAA20 and hNAA25 are mediated between hNAA20 α1 helix and α26-α27 loop, and the hNAA25 α24-α25 loop, consisting of a mix of polar and non-polar interactions. A few hydrogen bonds are formed, centered around hNAA20-Asn16, between the sidechain nitrogen atoms of hNAA20-Asn16 and sidechain oxygen of hNAA25-Thr541, between the sidechain oxygen atom of hNAA20-Asn16 and sidechain nitrogen atom of hNAA25-Gln538, and between the sidechain nitrogen of hNAA20-Asn16 and the sidechain of hNAA25-Asp657 (**Fig. 4.3D**). A salt bridge is found between the sidechain of hNAA20-Asp10 and the sidechain of hNAA25-Arg602. In addition, there are hydrophobic interactions between residues Phe15, Leu12, Phe13, Leu19, Pro22 from hNAA20 and residues Ile537, Asp540, His536, Ala506, Phe662, Ser504, Ile542, Leu603 from hNAA25 (**Fig. 4.3D**).

Different sides of hNAA20 feature several potentially hNAA25-stabilizing polar interactions. Hydrogen bonds are formed between the Asp261 side chain of the hNAA25 α14-α15 loop and the hNAA20-Thr2 backbone nitrogen atom (**Fig. 4.3E**), between the Glu227 backbone carbonyl group of hNAA25 α12-α13 loop and the hNAA20-Arg102

sidechain (**Fig. 4.3F**), and between the Lys535 sidechain from NAA25 α26-α27 loop and the hNAA20-Ser141 backbone carbonyl group (**Fig. 4.3G**).

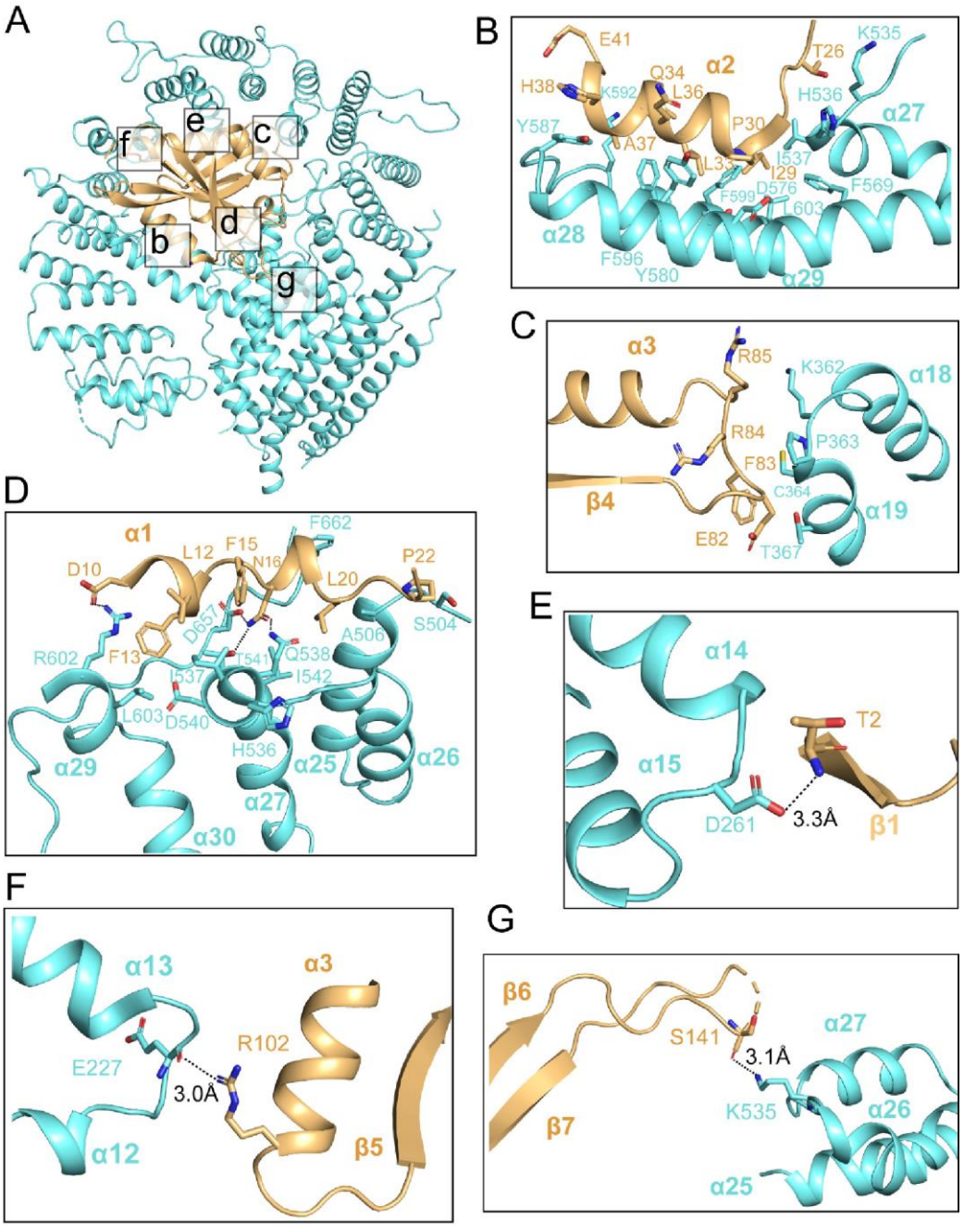


Figure 23.3 hNAA20 and hNAA25 make intimate interactions within hNatB

A. hNAA20 (light orange) and hNAA25 (cyan) are shown in cartoon with major associated interface denoted. **B-G.** Zoom-in views of the hydrophobic interface regions as indicated in **A.**

4.2.4 hNatB makes specific interactions with the first 4 N-terminal residues of aSyn

In the Cryo-EM map, density for the CoA-αSyn conjugate bisubstrate inhibitor is well resolved, allowing us to confidently model the CoA portion and the first five N-terminal residues (of ten residues present) of the αSyn portion (**Fig. 4.4A and 4.4B**). Similar to other NATs, CoA enters the catalytic active site through a groove formed by α3 and α4 of the catalytic subunit, while the peptide substrate enters the active site on the opposite side of the catalytic subunit flanked by the α1-2 and β6- β7 loops (**Fig. 4.4A)**.

hNAA20 contains a conserved acetyl-CoA binding motif among NATs: $R_{84}R_{85}XG_{87}XA_{89}$. Here, we observe that the positively charged hNAA20-Arg 85 interacts with the negatively charged 3'-phosphorylated ADP portion of CoA to form a salt bridge while Arg84 makes Van der Waals interactions (**Fig. 4.4C and 4.4D)**. A hydrogen bonding network is formed mainly between the 5'-diphosphate group and backbone atoms of a few residues including Val79, Gly87, Ala89, and Ala90 (**Fig. 4.4C and 4.4D)**, and mediated by the side chains of Arg85 and Gln125. The CoA molecule anchors to the binding pocket through a series of van der Waals contacts formed by residues Ser67, Val79, Leu77, Leu88, Val118, Met 122, and Tyr123 (Fig. **4.4D)**.

Four N-terminal residues of αSyn participate in hNAA20 interactions. Anchoring of the αSyn peptide is mediated by protein hydrogen bonds with the backbone atoms of Met1 and Asp2 of αSyn. Hydrogen bonds are formed between the backbone N-H group of αSyn-Met1 and the backbone carbonyl group of hNAA20-Phe111, as well as the backbone carbonyl group of αSyn-Met1 with the sidechain of hNAA20-Tyr138. The backbone N-H and carbonyl of αSyn-Asp2 also form hydrogen bonds to the sidechain of Tyr27, and between the backbone carbonyl group of Asp2 and sidechain of hNAA20-Tyr27 (**Fig. 4.4D)**. Remarkably, hNAA20 contacts each of the first four N-terminal residue side chains of αSyn via van der Waals interactions. The only side chain that forms a hydrogen bond with hNAA20 is αSyn-Asp2, which hydrogen bonds with a hNAA20-His73 ring nitrogen and the hNAA20-Thr75 side chain (**Fig. 4.4D)**. The more extensive van der Waals interactions include the following: αSyn-Met1 interacts with hNAA20 residues Glu25, Phe27, Tyr56 and Ala76; αSyn-Asp2 interacts with hNAA20 residues Try27, Thr75, His73, Phe111, and Tyr138; αSyn-Val3 interacts with hNAA20 residues Tyr137 and Tyr138; and αSyn-Phe4 interacts with hNAA20 residues Glu25 and Als140. αSyn-Met5 does not appear to make specific interactions (**Fig. 4.4D)**. Consistent with the importance of the residues that mediate αSyn binding, most of the residues are highly conserved from yeast to humans.

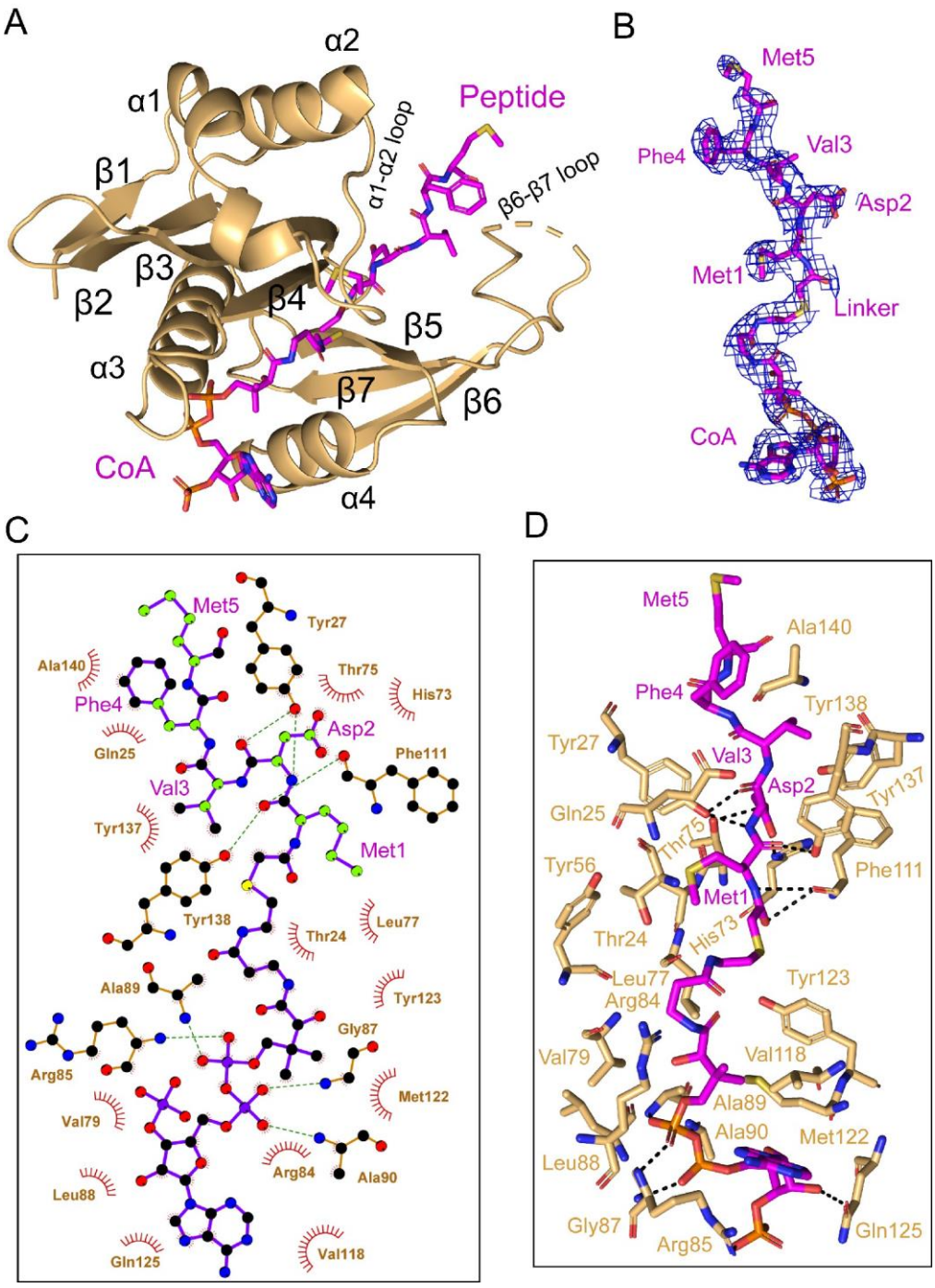


Figure 24.4 hNAA20 makes key CoA- and substrate peptide-interactions

A. The structure of hNAA20 bound to the CoA-□Syn conjugate bound is shown in cartoon, with corresponding secondary structures labeled. **B.** The fit of the CoA-□Syn

conjugate ligand in the EM density map. The contour level is 4.0 sigma. **C.** Interaction between CoA-□Syn conjugate and hNAA20 residues is generated with LIGPLOT (Laskowski and Swindells 2011). Hydrogen bonds are indicated by dashed green lines, and van der Waals interactions are indicated with red semicirlcles. **D.** Highlighted polar and hydrophobic interactions between CoA-□Syn conjugate and the hNAA20 are depicted in 3D view.

4.2.5 Mutational analysis identifies key residues for hNatB catalysis and cognate substrate binding

To determine the functional importance of hNAA20 residues that appear to make important peptide or CoA substrate contacts in our model, we used an *in vitro* acetyltransferase assay to kinetically characterize WT and mutant hNatB proteins. Each mutant was purified to homogeneity and displayed identical gel filtration chromatography elution profiles (data not shown), indicating that they were all properly folded. We prepared alanine mutants of several residues involved in the CoA binding including Arg84, Arg85, Gly87 and Tyr123. Among them, R84A, R85A and G87A did not show significant defects in overall protein catalytic function (**Table 4.1**). However, a Y123A mutant nearly abolished protein activity, with a 95% loss of protein activity, affecting both k_{cat} and K_m (**Table 4.1**). To further interrogate the properties of this residue, we prepared a Y123F mutant which features a similar aromatic bulky side chain but not the polar p-hydroxyl group. We observed that Y123F displayed a similar ~88% loss of k_{cat}, but had a negligible effect on the peptide K_m (**Table 4.1**). These data suggested that the Tyr123 hydroxyl group is critical for catalysis but not required for substrate binding, while the aromatic ring of Tyr123 plays a role in peptide substrate binding. Given that the hydroxyl

group of Tyr123 is about 3.5 Å from the sulfur atom of the CoA-Syn conjugate and 6.3 Å away from the αSyn N-terminus, it is in a position to play a role as a general base or acid for catalysis, potentially through an intervening water molecule (**Fig. 4.5A).** This is analogous to the proposed general base role of Tyr31 as a general base for hNAA50 catalysis (Liszczak, Arnesen, and Marmorstein 2011) (See discussion).

We also prepared alanine substitutions for residues that appeared to play important roles in αSyn binding: Glu25, Tyr27, His73, Tyr137, Tyr138. We were surprised to find that mutations of hNatB residues that mediated backbone hydrogen bond interactions, Y27A and Y138A, had relatively modest effects on αSyn peptide NTA with Y27A showing ~2-fold higher K_m and Y138A showing ~2-fold reduced K_{cat}, together suggesting that side chain contacts might dominate the binding energy (**Table 4.1**). Consistent with this, and our structural observations, we found that H73A produced a ~90% reduction in activity (**Table 4.1**). This correlates with the importance of the His73 hydrogen bond and van der Waals contacts with αSyn-Asp2. Of note, other cognate hNatB side chain residues at position 2, Glu, Gln, and Asn would also be well positioned to form hydrogen bonds with His73. Together, hNAA20-His73 appears to play a critical role in cognate substrate recognition by hNatB.

hNatB-Asn116 is a highly conserved NatB residue that caps the hNAA20 4 helix, which also harbors the putative catalytic residue, Tyr123, and is also in position to make a water-mediated hydrogen bond with to CoA pantetheine nitrogen (**Fig. 4.5A).** This observation suggests that Asn116 could play an important functional role. To test this, we prepared and evaluated an N116A mutant and, consistent with our hypothesis, Ans116, we found that this mutant leads to ~ 90% loss in activity (**Table 4.1**).

Taken together, our structural and mutational analysis of hNatB highlight the functional importance of hNatB-NAA20 residues His73, Asn116, and Tyr123 in hNatB-mediated N-terminal acetylation. While Tyr123 appears to play a critical catalytic role, potentially as a general base and/or acid for catalysis; His73 appears to play an important role in the recognition of substrate residue 2 and Asn116 likely plays a structural role (**Fig. 4.5A).** Each of these residues could employ a bridging water molecule to mediate their functional roles, although these putative water molecules are not visible at the current resolution of our structure.

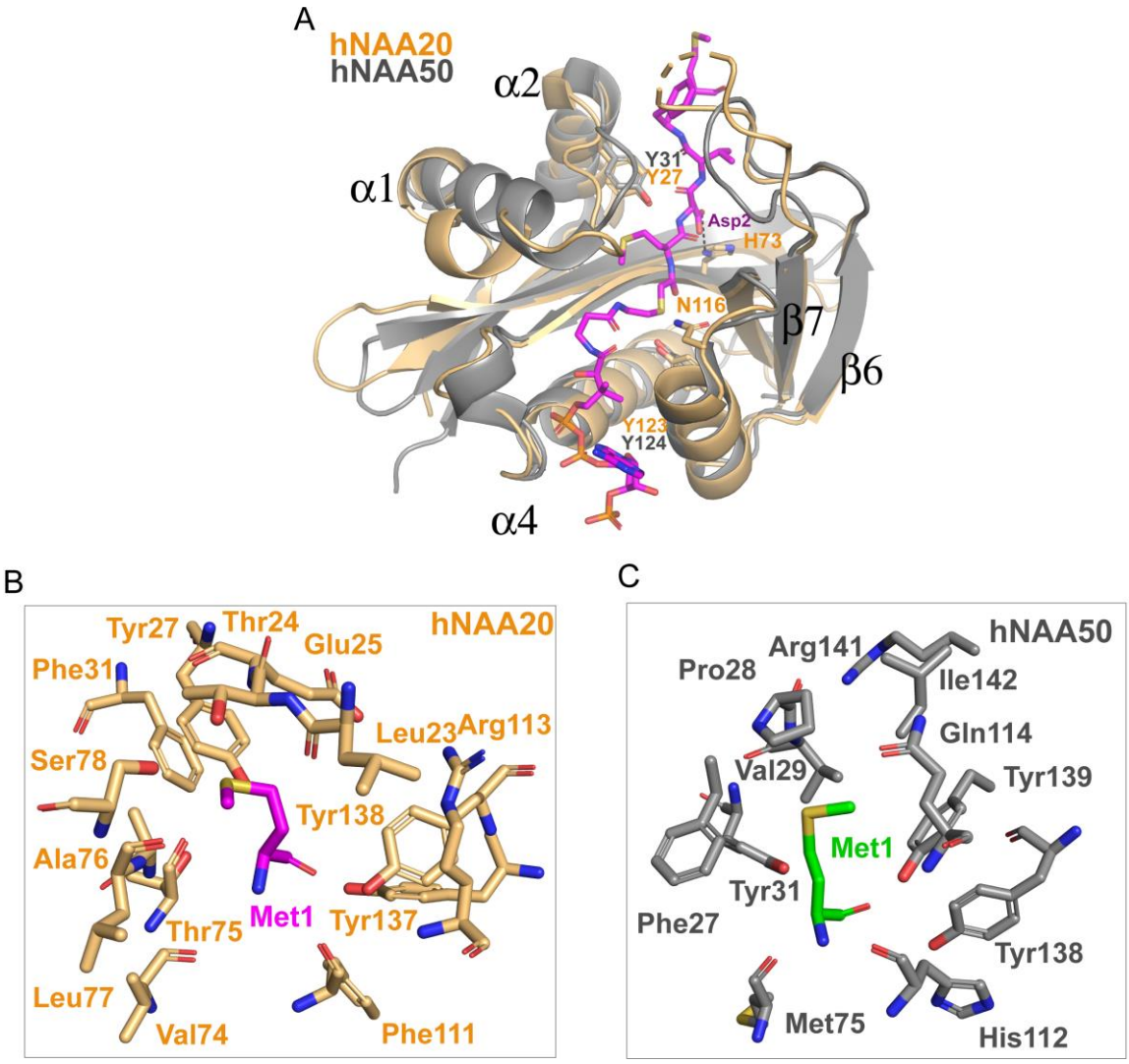


Figure 25.5 Structural comparison between hNAA20 and hNAA50

A. Bi-substrate inhibitor-bound hNAA20 (light orange) is shown superimposed with hNAA50 (grey, PDB: 3TFY). H73, N116, and Y123 (sticks), mediate important functional roles in hNatB catalysis. **B.** Residues forming the Met1 binding pocket of hNAA20 are depicted. **C.** Residues forming the Met1 binding pocket of hNAA50 are depicted.

4.3 Discussion

Since the identification of hNatB more than a decade ago, many studies have shown that it N-terminally acetylates important proteins such as actin, tropomyosin, CDK2, and α-Synuclein, and its function has connections to diseases such as hepatocellular carcinoma and Parkinson disease (Spillantini et al. 1998; Polevoda et al. 2003; Ametzazurra et al. 2008; Starheim et al. 2008; Halliday et al. 2011; Neri et al. 2017). Despite its clear biological importance, hNatB-mediated NTA by hNatB remained poorly understood. Here, we developed a CoA-αSyn conjugate hNatB inhibitor, determined the Cryo-EM structure of CoA-αSyn inhibitor-bound hNatB, and carried out associated structure-guided mutagenesis and activity assays. This has led to the identification of functionally important differences with human NatA and *C. albicans* NatB. These studies have also provided evidence for important hNatB-specific elements responsible for αSyn recognition and N-terminal acetylation, providing direct implications for NatB recognition of other canonical substrate proteins.

Consistent with previous studies, we have demonstrated that hNatB acetylates a cognate “MD” N-terminus, and is unable to N-terminally acetylate non-cognate N-termini that are substrates for other NATs such as NatA, NatC, NatE. We demonstrated for the first time that hNatB can acetylate an α-Syn peptide *in vitro*, directly linking hNatB to NTA of α-Syn. We have also demonstrated that αSyn peptides lacking the Met1 or both the Met1 and Asp2, do not serve as hNatB substrates, confirming the strict substrate specificity of hNatB. This is consistent with our structural model showing significant interactions between hNatB-NAA20 and both the first and second N-terminal residues of an αSyn peptide, with important but less extensive interactions with the third

and fourth residues. This hierarchy of interactions likely explains how NatB enzymes can accommodate cognate substrates that diverge at positions three and four.

Here, we have presented hNatB-αSyn interactions that can be used to rationalize the substrate specificity of hNatB: N-terminal sequences containing “MD-”, “ME-”, “MN-”, and “MQ-”. αSyn-Met1 sits in a hydrophobic pocket that comfortably accommodates a methionine residue, whereas shorter side chains or longer polar or charged side chains would fit poorly (**Fig. 4.5B-C**). The nature of this binding pocket is similar to the described hNAA50 recognition of Met1 (Liszczak, Arnesen, and Marmorstein 2011) (**Fig. 4.5B-C**). Although both of hNAA50 and hNAA20 can N-terminally acetylate peptides with Met at the first position, no overlapping activity has been observed. This can be rationalized based on the chemical properties of the second residue in the cognate peptide. We find that the αSyn-Asp2 sidechain forms hydrogen bonds with the hNAA20 side chains His73 and Thr75. These polar residues in the peptide binding site of hNAA20 would likely serve as poor acceptors of the largely hydrophobic residues targeted by hNAA50. The hNatB substrate client profile featuring D-, E-, N- or Q-residues in position two is consistent with the mechanisms of substrate recognition observed in the binding pocket for αSyn-Asp2. The aliphatic regions of each of these side chains (D, E, N, Q) would all benefit from the extensive hNatB van der Waals interactions surrounding the aliphatic region for αSyn-Asp2 (Tyr27, His73, Thr75, Phe111, and Tyr138). This second residue would also form a hydrogen bond interaction with His73 and Thr75, which may accommodate the carboxyl side chains of both the shorter D- and N- and longer E- and Q- side chains. Notably, His73 and Thr75 are strictly conserved from yeast to man (**Fig. 4.5)**. In contrast, shorter polar or nonpolar side chains would less efficiently fill the pocket for residue two, while larger polar or charged

side chains would likely result in steric clashes. In agreement with this, we have demonstrated that H73A mutation has severe impact on hNatB catalysis (**Table 4.1**).

The hNatB/CoA-αSyn structure has implications for the mode of hNatB catalysis. While previous studies have suggested that hNAA50-Tyr31 plays an important role in catalysis, mutation of the corresponding hNaa20 residue, Tyr27, had minimal effects on hNatB kinetic parameters. Strikingly, our mutational analysis has identified the functional importance of hNatB-NAA20 residues His73, Asn116 and Tyr123, although Tyr123 is the only residue that is in position to play a catalytic role (**Fig. 4.5A**). Specifically, Tyr123 is in position to play a catalytic role, potentially as a general base and/or acid, through a bridging water molecule (although a water molecule is not visible at the current resolution). Interestingly, hNAA50 contains a tyrosine residue at the same position (hNAA50-Tyr124), although the mechanistic significance of this tyrosine residue has not yet been described (Liszczak, Arnesen, and Marmorstein 2011). It would be of interest to determine if hNAA50-Tyr124 also plays an important catalytic role (possibly in combination with hNAA50-Tyr31), similar to the corresponding hNAA20-Tyr123 of hNatB.

The biological importance of hNatB and its connection to various disease processes highlights it as an important target for probe and inhibitor development. Indeed, a recent study highlights hNatB as a therapeutic target for aSyn toxicity (Vinueza-Gavilanes et al. 2020). Our development of a CoA-αSyn conjugate bisubstrate with an IC_{50} of ~ 1.6 μM represents a step in this direction, although the structural information provided here could further aid to the rational development of more drug-like hNatB inhibitors with possible therapeutic applications.

4.4 Materials and methods

4.4.1 Protein expression and purification

hNAA20 with a C-terminal truncation (1-163 out of 178 residues) and full-length hNAA25 were cloned into two separate insect cell expression vectors pFASTBac HTA. hNAA20 was untagged, while hNAA25 contained a Tobacco-etch virus (TEV)-cleavable N-terminal 6xHis-tag. Human NatB complex ($hNAA20^{1-163}$/$hNAA25^{FL}$) was obtained by co-expressing these two plasmids in Sf9 (*S. frugiperda*) cells (ThermoFisher, cat# 12659017), and purified as described previously (Gottlieb and Marmorstein 2018). Sf9 cells were grown to a density of $1x10^6$ cells/ml and infected using the amplified $hNAA20^{1-163}$/$hNAA25^{FL}$ baculovirus to an MOI (multiplicity of infection) of 1-2. The cells were grown at 27 °C and harvested for 48 hours post-infection by centrifugation. Cell pellets were resuspended in lysis buffer (25 mM Tris, pH 8.0, 300 mM NaCl, 10 mM Imidazole, 10 mM β-ME, 0.1 mg/mL PMSF, DNase, and complete, EDTA-free protease inhibitor tablet) and lysed by sonication. After centrifugation, the supernatant was isolated and passed over Ni-NTA resin (Thermo Scientific), which was subsequently washed with 10 column volumes of lysis buffer. Protein was eluted with a buffer with 25 mM Tris, pH 8.0, 300 mM imidazole, 200 mM NaCl, 10 mM β-ME, which was dialyzed into buffer with 25 mM HEPES pH 7.5 50 mM NaCl 10 mM β-ME. Ion-exchange was carried out with an SP ion-exchange column (GE Healthcare) in dialysis buffer with a salt gradient (50-750 mM NaCl). Peak fractions were concentrated to ~ 0.5 mL with a 50-kDa concentrator (Amicon Ultra, Millipore), and loaded onto an S200 gel-filtration column (GE Healthcare) in a buffer with 25 mM HEPES, pH 7.5, 200 mM NaCl, and 1 mM TCEP. Proteins were aliquoted, snap-frozen in liquid nitrogen, and stored at −80 °C for further use. Protein harboring mutations were generated with the QuickChange protocol

(Stratagene) and obtained following the same expression and purification protocol as described for the wild-type protein.

4.4.2 Acetyltransferase activity assays

All acetyltransferase assays were carried out at room temperature in a reaction buffer containing 75 mM HEPES, pH 7.5, 120 mM NaCl, 1 mM DTT as described (Deng, McTiernan, et al. 2020; Deng et al. 2019). The “MDVF” peptide substrate was based on the first 7 amino acid of α-Synuclein (“MDVF” peptide: NH_2-MDVFMKGRWGRPVGRRRRP-COOH; “SASE” peptide: NH_2-SASEAGVRWGRPVGRRRRP-COOH; “MLGP” peptide: NH_2-MLGPEGGRWGRPVGRRRRP-COOH; “SGRG”/H4 peptide: NH_2-SGRGKGGKG LGKGGAKRHR-COOH; “MLRF” peptide: NH_2-ML RFVTKRWGRPVGRRRRP-COOH; “DVFM” peptide: NH_2- DVFMKGLRWGRPVGRRRRP-COOH; “VFMK” peptide: NH_2-VFMKGLSRWGRPVGRRRRP -COOH; GenScript). Reactions were performed in triplicate. To determine steady-state catalytic parameters of hNatB with respect to acetyl-CoA, 100 nM hNatB was mixed with a saturating concentration of “MDVF” peptide substrate (500 μM) and varying concentrations (1.95 μM to 1 mM) of acetyl-CoA (^{14}C-labeled, 4 mCi mmol^{-1}; PerkinElmer Life Sciences) for 10-minute reactions. To determine steady-state catalytic parameters of hNatB with respect to peptide substrate, 100 nM hNatB was mixed with saturating concentrations of acetyl- CoA (300 μM, ^{14}C-labeled) and varying concentrations of “MVDF” peptide (1.95 μM to 1 mM) for 10 minutes. Reactions were quenched by adding the solution to P81 paper discs (Whatman). Unreacted acetyl-CoA was removed by washing the paper discs in buffer

with 10 mM HEPES, pH 7.5, at least three times, each 5 minutes. The paper discs were then dried with acetone and transferred to 4 mL scintillation fluid for signal measurement (Packard Tri-Carb 1500 liquid scintillation analyzer). Data was fitted to a Michaelis–Menten equation in GraphPad Prism to calculate kinetic parameters. Kinetic parameters on mutants with respect to peptide were carried out in the same condition as for wild type, with 300 μM ^{14}C labeled acetyl-CoA and varied peptide concentration (1.95 μM to 1 mM). All radioactive count values were converted to molar units with a standard curve created with known concentrations of radioactive acetyl-CoA added to scintillation fluid. GraphPad Prism (version 5.01) was used for all data fitting to the Michaelis–Menten equation. For IC_{50} determination of the CoA-αSyn conjugate, 100 nM hNatB was mixed with 500 μM "MVDF" peptide and 300 μM ^{14}C labeled acetyl-CoA, and inhibitor concentrations were varied (0.23 μM to 13.44 μM). Data were fit to a sigmoidal dose-response curve with GraphPad Prism (version 5.01). Errors represent s.d. (n = 3).

4.4.3 Cryo-EM data collection

For initial sample screening, 0.6 mg/ml fresh hNatB sample with three-molar excess bisubstrate was used. hNatB particles on these grids exhibited a severe preferred orientation, which generated an incorrect 3D initial model (data not shown). To solve this issue, 1 μL of 0.05% NP-40 was mixed with 20 μL of hNatB (4 mg/mL). 3 μL of this sample was applied to glow-discharged Quantinfoil R1.2/1.3 holey carbon support grids, blotted and plunged into liquid ethane, using an FEI Vitrobot Mark IV. An FEI TF20 was used for screening the grids and data collection was performed with a Titan Krios equipped with a K3 Summit direct detector (Gatan), at a magnification of 105,000 x, with defocus values from -0.1 to -2.0 μm. Each stack was exposed in super-resolution mode

with a total dose of 45 $e^-/Å^2$, resulting in 35 frames per stack. Image stacks were automatically collected with Latitude software (Gatan, Inc).

4.4.4 Cryo-EM data processing

Original image stacks were summed and corrected for drift and beam-induced motion at the micrograph level using MotionCor2 (Zheng et al. 2017), and binned twofold, resulting in a pixel size of 0.83 Å/pixel. Defocus estimation and the resolution range of each micrograph were performed with Gctf (Zhang 2016). About 3000 particles were manually picked to generate several rough 2D classaverages. Representative 2D classes were used to automatically pick ~1,927,673 particles from 5281 micrographs in Relion 3.0 (Kimanius et al. 2016; Zivanov et al. 2018). All particles were extracted and binned to accelerate the 2D and 3D classification. After bad particles were removed by 2D and 3D classification, 982, 420 particles were used for auto-refinement and per particle CTF refinement. After refinement, a mask was created in Relion with an initial binarization threshold of 0.005, covering the protein complex and extending the binary map and soft-edge by 12 pixels. The map was sharpened with the created mask by estimating B-factor automatically in Relion. The final map was refined to an overall resolution of 3.46 Å, with local resolution estimated by Resmap (Kucukelbir, Sigworth, and Tagare 2014). We attempted particle polishing on this data set, but this surprisingly resulted in artifactual density in the resulting map. We believe that this was due to some small defects in the K3 camera during data collection, which corrupted the particle polishing process. We therefore did not perform particle polishing on this data set. Raw micrographs were deposited in EMPIAR with access ID of EMPIAR-10477**.**

4.4.5 Cryo-EM model building and refinement

The hNatB atomic model was manually built de novo using the program COOT (Emsley and Cowtan 2004) according to the Cryo-EM map, with the guidance of predicted secondary structure and bulky residues such as Phe, Tyr, Trp and Arg. The first two alpha helices of hNAA25 were built as poly-alanine, due to the lack of tracible density in the 3D map. The complete model was then refined by real-space refinement in PHENIX (Adams et al. 2010). All representations of Cryo-EM density and structural models were performed with Chimera (Pettersen et al. 2004) and PyMol (https://pymol.org/2/). The sequence alignments with secondary structure display were created by ESPript 3.0 (Robert and Gouet 2014b). hNAA25 TPR predictions were performed using the TPRpred server (Karpenahalli, Lupas, and Söding 2007; Zimmermann et al. 2018) (https://toolkit.tuebingen.mpg.de/#/tools/tprpred). The surface area calculation was performed using PDBePISA (Krissinel and Henrick 2007) (Proteins, Interfaces, Structures and Assemblies) (http://www.ebi.ac.uk/pdbe/pisa/).

CHAPTER 5: Molecular Mechanism of N-terminal Acetylation by the Ternary NatC Complex

This work is adapted from *bioRxiv* preprint (Deng et al. 2021)

5.1 Introduction

Substrate recognition by eukaryotic NATs is usually dictated by the first few residues of the substrate via the active site structure (Liszczak et al. 2013; Hong et al. 2017; Magin, Liszczak, and Marmorstein 2015; Stove et al. 2016; Goris et al. 2018; Deng et al. 2019; Deng and Marmorstein 2021; Deng, Pan, et al. 2020). The active site of NatA contains a relatively small binding pocket to accommodate a small, uncharged residue at position 1 such as alanine, valine or threonine residues that remains following initiator methionine (iMet) - cleavage (Liszczak et al. 2013). NatB has a hydrophobic pocket for a retained iMet and uses a histidine residue to recognize sequences with D/E/N/Q at position 2 via hydrogen bonding (Arnesen et al. 2009; Deng and Marmorstein 2021). Interestingly, NatC and NatE share similar and relatively versatile substrate profiles toward substrates containing a retained iMet followed by residues other than D/E/N/Q (Van Damme et al. 2016a; Arnesen et al. 2009). NatD is a highly selective enzyme with histones H4 and H2A as its only known substrates (Hole et al. 2011; Song et al. 2003), arising from extensive and specific interactions with each of the first four residues of its substrate (Magin, Liszczak, and Marmorstein 2015). Although NATs vary in terms of their modes of substrate recognition, their mechanism of catalysis usually involves the use of dedicated residues to serve as general base and acid residues (Deng and Marmorstein 2021).

The activity of some NATs is further regulated by other protein modulators. For example, within the human NatE complex, there is catalytic crosstalk between the two NAA10 and NAA50 catalytic subunits (Deng et al. 2019). Moreover, the activities of both NatA and NatE are inhibited to a different extent by a small protein binding partner called Huntingtin-interacting protein K (HYPK) (Deng, McTiernan, et al. 2020; Gottlieb and

Marmorstein 2018; Weyer et al. 2017; Arnesen et al. 2010). Inositol hexaphosphate (IP_6, also named inositol hexaphosphate or phytic acid), a small molecule, usually reserved for cellular signaling, appears to act as a stabilizing ligand for both NatA and NatE, by binding in a pocket between NAA15 and NAA10, and, thus, plays an indirect role in promoting NatA and NatE acetylation activity (Gottlieb and Marmorstein 2018; Deng, McTiernan, et al. 2020; Deng et al. 2019; Cheng et al. 2019).

NatC is distinct from other NATs since it contains two eukaryotic conserved auxiliary subunits – Naa38 and Naa35 – that act together with the Naa30 catalytic subunit (Polevoda and Sherman 2001; Ochaya et al. 2019; Starheim et al. 2009). Initial studies reported that NatC activity requires the interaction of all three subunits, since knock out of any single subunit produced a similar phenotype in yeast that was accompanied by diminished N-terminal acetylation of a cognate N-terminal substrate, MIRLK- (Polevoda and Sherman 2001). However, it was subsequently reported that the Naa38 subunit is not required for *in vivo* acetylation of an Arl3p substrate with an N-terminal protein sequence of MPHLV- (Setty et al. 2004a). In plants, knockout of *At*NAA35 did not lead to similar phenotypic effects as the deletion of *At*NAA30 (Pesaresi et al. 2003), and the single subunit of the *A. thaliana* analogue *At*NAA30 can rescue the knockout of yeast NatC complex subunits (Pesaresi et al. 2003). Similarly, recombinant human NAA30 (hMak3) was shown to have substrate-specific acetylation activity, even in the absence of its auxiliary subunits (Starheim et al. 2009). Thus, the functional roles of the NatC auxiliary subunits are not clear, particularly with respect to Naa38, which is not well-conserved across species. Although the recently reported crystal structures and associated biochemical studies of *S. cerevisiae* NatC (*Sc*NatC) by Grunwald *et al.* have provided some important insights into NatC substrate specificity and its catalytic

mechanism (Grunwald et al. 2020), several questions remain unanswered. In particular, those regarding the mechanisms dictating the overlapping yet distinct substrate profiles of NatC and NatE and the influence of IP_6 on NatC.

In this study, we report that yeast NatC complex formation of all three subunits is a prerequisite for normal NatC acetylation activity and that, like NatA and NatE, IP_6 binds tightly to and stabilizes NatC. We also report the Cryo-EM structure of ~100 kDa *S. pombe* NatC (*Sp*NatC) bound to a NatE/C-type bisubstrate analogue and IP_6, with related biochemistry, to reveal the molecular basis for IP_6-mediated stabilization of the complex and the similar substrate profiles of NatE and NatC. Comparison of these studies with the recently published structural and functional studies of the *Sc*NatC complex (Grunwald et al. 2020) reveal evolutionarily conserved and divergent features of NatC.

5.2 Results

5.2.1 NAA38 is required for normal NatC acetylation activity in yeast

We found that overexpression of the *Sp*NatC complex containing full-length *Sp*Naa30 (residues 1–150), N-terminally truncated *Sp*Naa35 (residues 31–708, with deletion of divergent residues, 1-30) and *Sp*Naa38 (residues 48-116, with deletion of divergent residues, 1-47) in *E. coli* produced soluble protein that could be purified to homogeneity (**Figure 5.1a**). To evaluate the activity of the recombinant ternary *Sp*NatC, we used an *in-vitro* acetyltransferase activity assay with different peptide substrates. Consistent with previous studies, the recombinant *Sp*NatC is active toward both its own canonical substrate - “MLRF peptide” and the canonical NatE substrate – “MLGP peptide”, but with preference for the “MLRF peptide” (Grunwald et al. 2020) (**Figure**

5.1b, see methods section for full peptide sequences). To investigate the catalytic roles of Naa38, we could readily purify both binary (*Sc*Naa30$^{1-161}$ and *Sc*Naa35$^{19-733}$) and ternary *Sc*NatC complexes (*Sc*Naa30$^{1-161}$ and *Sc*Naa35$^{19-733}$, and *Sc*Naa38$^{1-70}$) from *Saccharomyces cerevisiae* to compare their activities. Consistent with the studies by Grunwald *et al (Grunwald et al. 2020)*, evaluation of the activities of the binary and ternary *Sc*NatC complexes toward the canonical NatC substrate (MLRF peptide) revealed that the ternary complex showed robust activity, while the binary complex showed compromised activity (**Figure 5.1b**). Thus, it appears that Naa38 is required for optimal NatC activity in yeast.

5.2.2 Inositol hexaphosphate (IP_6) binding contributes to yeast NatC complex stability

Earlier reports demonstrated that IP_6 interacts with and stabilizes both the yeast and human NatA and NatE complexes (Deng, McTiernan, et al. 2020; Deng et al. 2019; Gottlieb and Marmorstein 2018; Cheng et al. 2019), but does not appear to interact with human NatB (Deng, Pan, et al. 2020). To determine if IP_6 could bind to the recombinant NatC complex, we employed isothermal titration calorimetry (ITC). We found that IP_6 binds to *Sp*NatC with a K_d of ~225 nM (with a stoichiometry of ~1) (**Figure 5.1c**), similar to its K_d value for yeast NatA (Deng et al. 2019). Differential scanning fluorimetry (DSF) was used to evaluate a potential role of IP_6 binding in NatC complex stability. We found that addition of IP_6 increased the thermal stability of ternary *Sp*NatC, binary *Sc*NatC, and ternary *Sc*NatC, by ~4.9°C, 2.7°C, and 2.1°C, respectively. Thus, IP_6 can provide additional thermal stability to the yeast NatC complexes (**Figure 5.1d**).

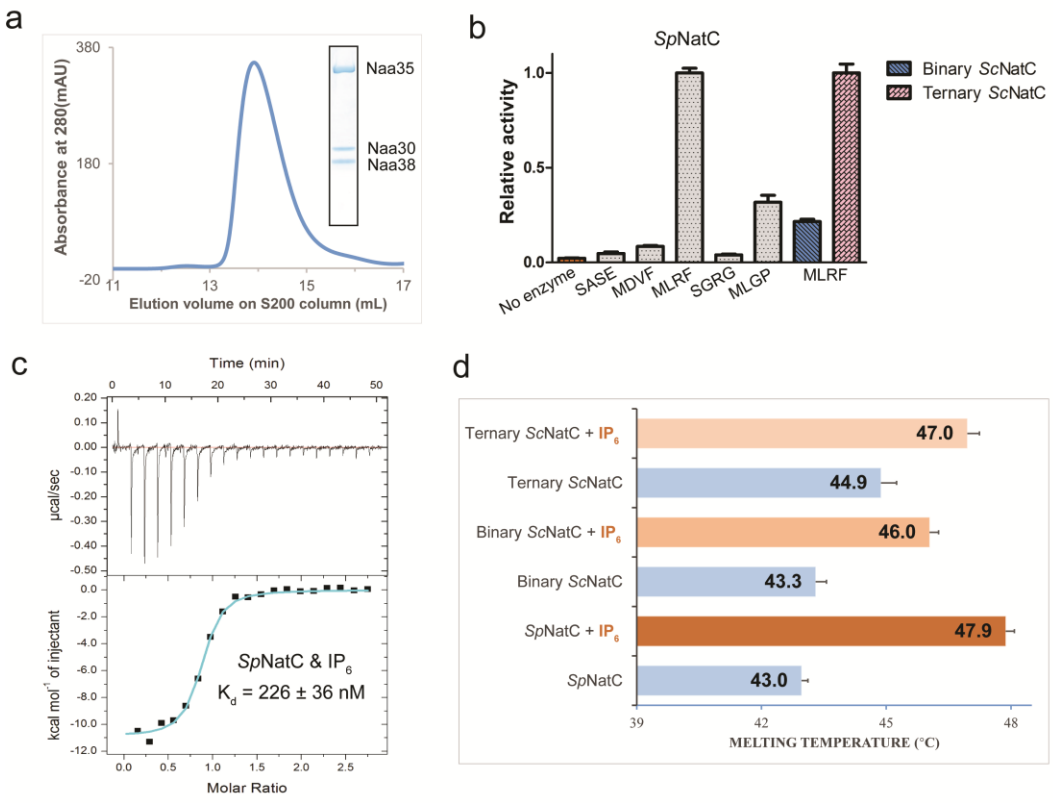


Figure 26.1 Binary NatC displays deficiency in acetylation activity and inositol hexaphosphate (IP_6) binding stabilizes NatC complex formation

(a) Gel filtration elution profile of the ternary *Sp*NatC complex using a Superdex S200 column. Coomassie-stained SDS–PAGE of elution peak is shown in the box, with bands for corresponding *Sp*NatC subunits labeled. **(b)** Comparison of *Sp*NatC activity toward different peptide substrates. The activities shown in gray are normalized to the activity of *Sp*NatC toward MLRF peptide. The pink and blue columns show the activity of binary and ternary *Sc*NatC toward the MLRF peptide, respectively, and are normalized to the pink column. Errors are shown in SEM with n = 3. **(c)** Representative isothermal titration calorimetry (ITC) curve of IP_6 titrated into *Sp*NatC. The ITC fitting information is N = 0.830 ± 0.00943 sites, $\Delta H = -1.091 \times 10^4 \pm 177.0$ cal mol^{-1}, $\Delta S = -6.79$ cal mol^{-1} deg^{-1}. **(d)**

Differential scanning fluorimetry assays of NatC alone or with IP_6. Average calculated melting temperature transitions are indicated. Error bars in the figure indicate the SD of each sample, n = 3.

5.2.3 NatC adopts a distinct NAT architecture

In order to understand the molecular details underlying the ternary NatC complex, its overlapping yet distinct substrate specificity with NatE, and the mode of IP_6 stabilization, we performed single particle Cryo-EM analysis using ternary *Sp*NatC prepared in the presence of the CoA-Ac-MLGP bisubstrate conjugate and IP_6. The MLGP sequence was selected for synthesis of the conjugate inhibitor as it is a predicted substrate for both NatC and NatE. A 3.16 Å-resolution Cryo-EM three-dimensional (3D) map was determined from 607,131 particles, which were selected from 5,397 raw electron micrographs (**Figure 5.2).** Most areas of the EM map contained excellent sidechain density, with a local resolution of ~3 Å, suggesting the relative rigidity of the recombinant complex (**Figure 5.2 and 5.3**). The atomic model was built *de novo* based on the EM map and details for the model refinement statistics can be found in **Table 5.1 (Figure 5.3)**.

The ternary *Sp*NatC complex contains three proteins, the catalytic Naa30 subunit, the large auxiliary Naa35 subunit, and the small auxiliary Naa38 subunit. Upon superficial inspection of the complex, it appears as though it is only formed by two proteins, as Naa38 is embedded within the Naa35 fold to form a “composite” auxiliary subunit (**Figure 5.4a and 5.4b)**. While Naa30 displays a canonical NAT fold with mixed α/β secondary structure, the Naa35/38 complex forms a clamp that pinches Naa30 on

opposite sides. This clamp-like structure is distinct from the architecture of the large auxiliary subunits of NatA and NatB, which surround the base of their respective catalytic subunits (Gottlieb and Marmorstein 2018; Liszczak et al. 2013; Hong et al. 2017; Deng, Pan, et al. 2020). The Naa35 architecture is mostly helical, with 25 helices, three short β-strands in its N-terminal region and a single β-hairpin at its C-terminal region that are unique to the NatC complex (**Figure 5.4b)**. Naa38 adopts an Sm-like fold, with an N-terminal α-helix, followed by a sharply bent β-sheet consisting of 5 five antiparallel β-strands (**Figure 5.4b)**. The Naa35/Naa38 clamp contains α4-α13 from Naa35 at its base, and α14-α25 and the C-terminal β-hairpin of Naa35 flanking one side of Naa30. The opposing side of Naa30 is flanked by a composite of the N-terminal end of Naa35 (a2, a3, b1-b3) and the entirety of Naa38 (**Figure 5.4a and 5.4b).** Consistent with the composite nature of the Naa35-Naa38 interaction, the buried surface area between Naa38 and Naa35 is ~1925 $Å^2$, which is notably higher than the area buried between Naa30 and Naa35 (~1735 $Å^2$), and between Naa30 and Naa38 (~197 $Å^2$).

Table 8.1 Cryo-EM data collection, refinement and validation statistics

	*Sp*NatC with Bisubstrate and IP_6 EMD-23110 PDB: 7L1K
Data collection and processing	
Magnification	105,000
Voltage (keV)	300
Electron exposure (e/Å^2)	42
Defocus range (μm)	-1.0 to -3.0
Pixel size (Å)	0.84
Symmetry imposed	C1
Initial particles (no.)	1,860,276
Final particles (no.)	607,131
Map resolution (Å)	3.16
FSC threshold	0.143
Map resolution range (Å)	2.5-3.5
Refinement	
Initial model used (PDB code)	-
Model resolution (Å)	3.4
FSC threshold	0.5
Model resolution range (Å)	-
Map sharpening *B* factor (Å^2)	159.1
Model composition	
Non-hydrogen atoms	7216
Protein residues	883
Ligands	2
***B* factors (Å^2)**	
Protein	26.97/110.48/71.91
Ligand	86.46/102.51/93.10
R.M.S. deviations	
Bonds lengths (Å)	0.006
Bond angles (°)	1.036
Validation	
MolProbity score	1.76
Clash score	4.10
Poor rotamers (%)	0.38
Ramachandran plot	
Favored (%)	89.55
Allowed (%)	10.22
Disallowed (%)	0.23

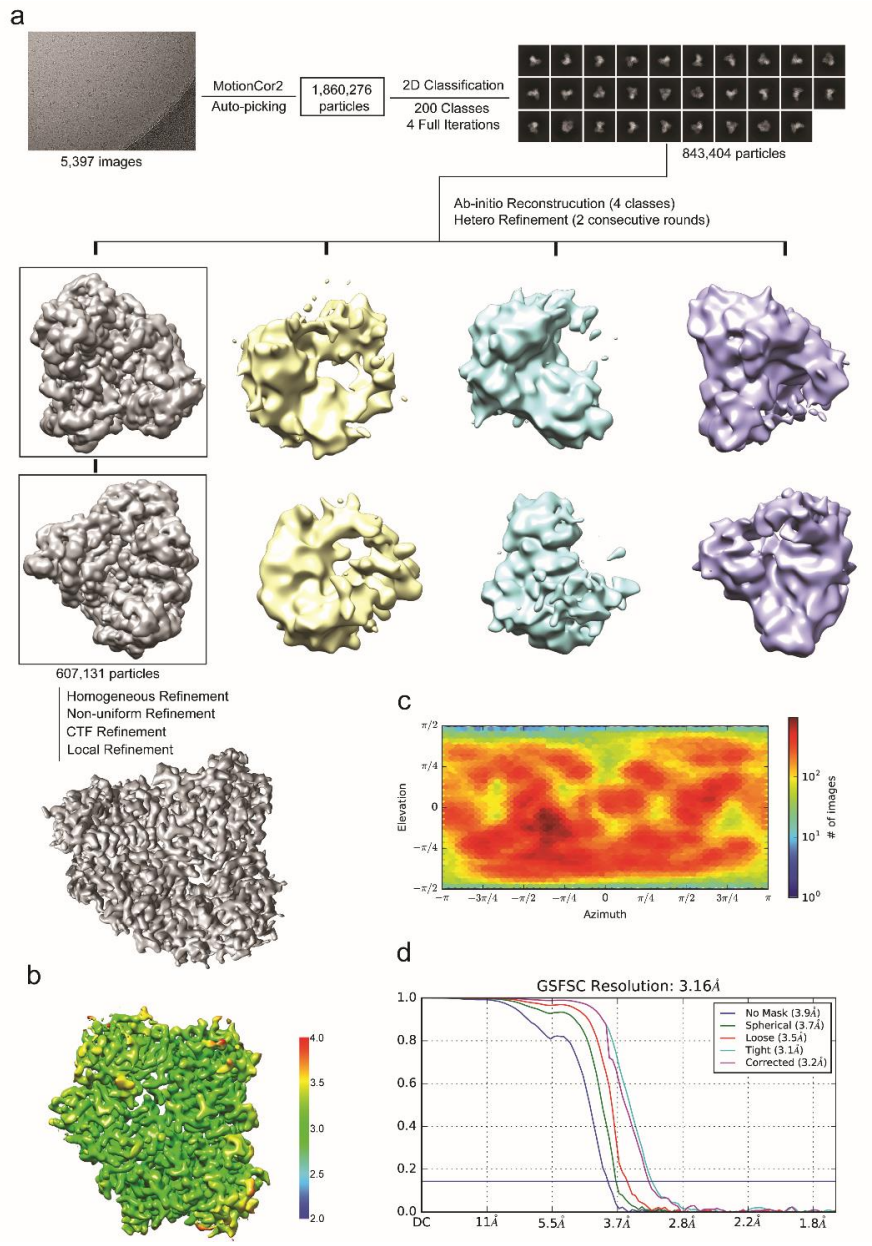


Figure 27.2 Cryo-EM workflow and resolution of SpNatC EM map

(**a**) 2D and 3D classification scheme and workflow for *Sp*NatC EM map determination. (**b**) Local resolution of the final *Sp*NatC EM map. (**c**) Viewing direction distribution of *Sp*NatC final EM map 3D reconstruction generated by cryoSPARC v2. (**d**) Gold standard

Fourier Shell Correlation (FSC) curves of *Sp*NatC EM map 3D reconstruction, generated by cryoSPARC v2.

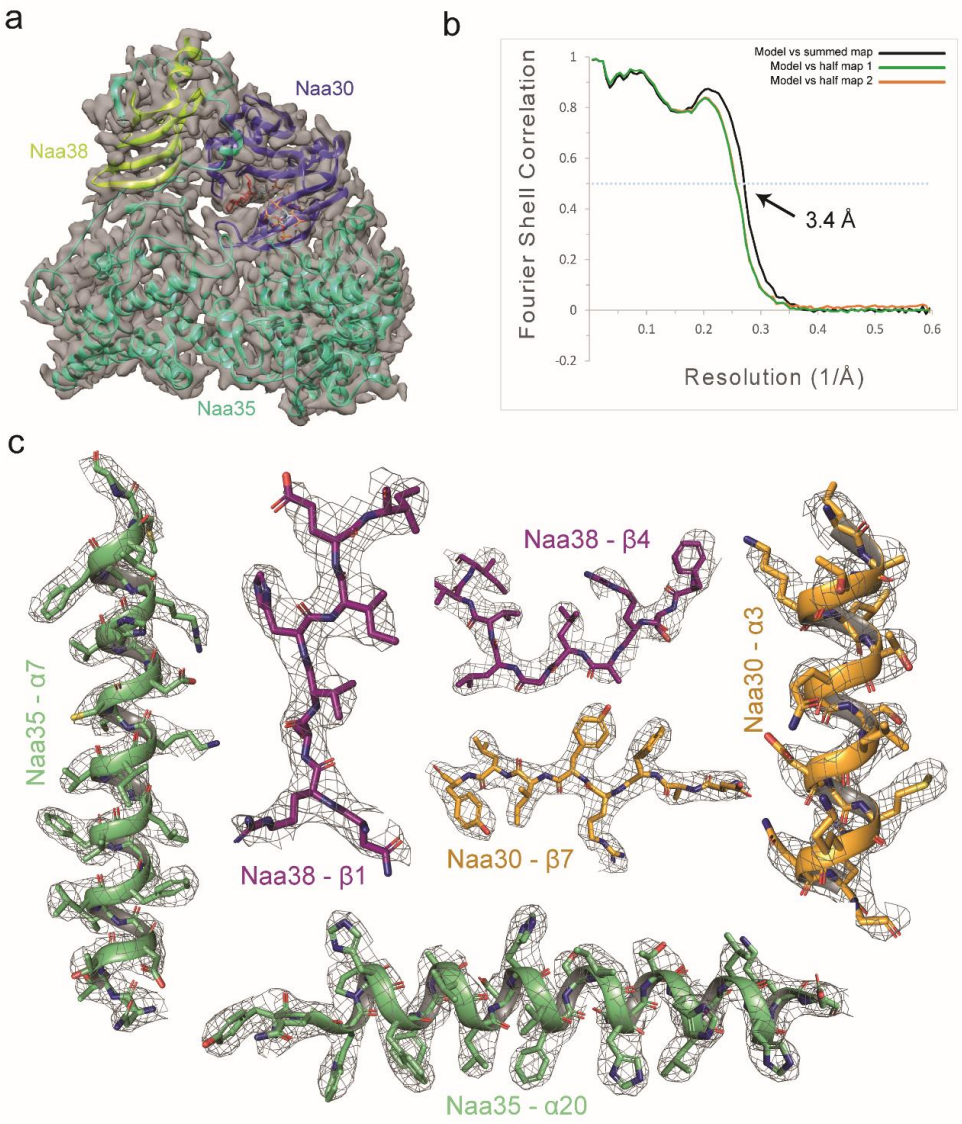


Figure 28.3 Correlation between Cryo-EM map of SpNatC EM and its atomic model

(a) Atomic model of *Sp*NatC fitted into the Cryo-EM map. **(b)** FSC curves of the refined model versus the overall map that it was refined against (black); of the model refined in the first of the two independent maps used for the gold-standard FSC versus the same

map (green); and of the model refined in the first of the two independent maps versus the second independent map (orange). (**c)** The fit of several helical segments or β-strands from all three subunits of NatC in the EM density. The contour level is 5 sigma.

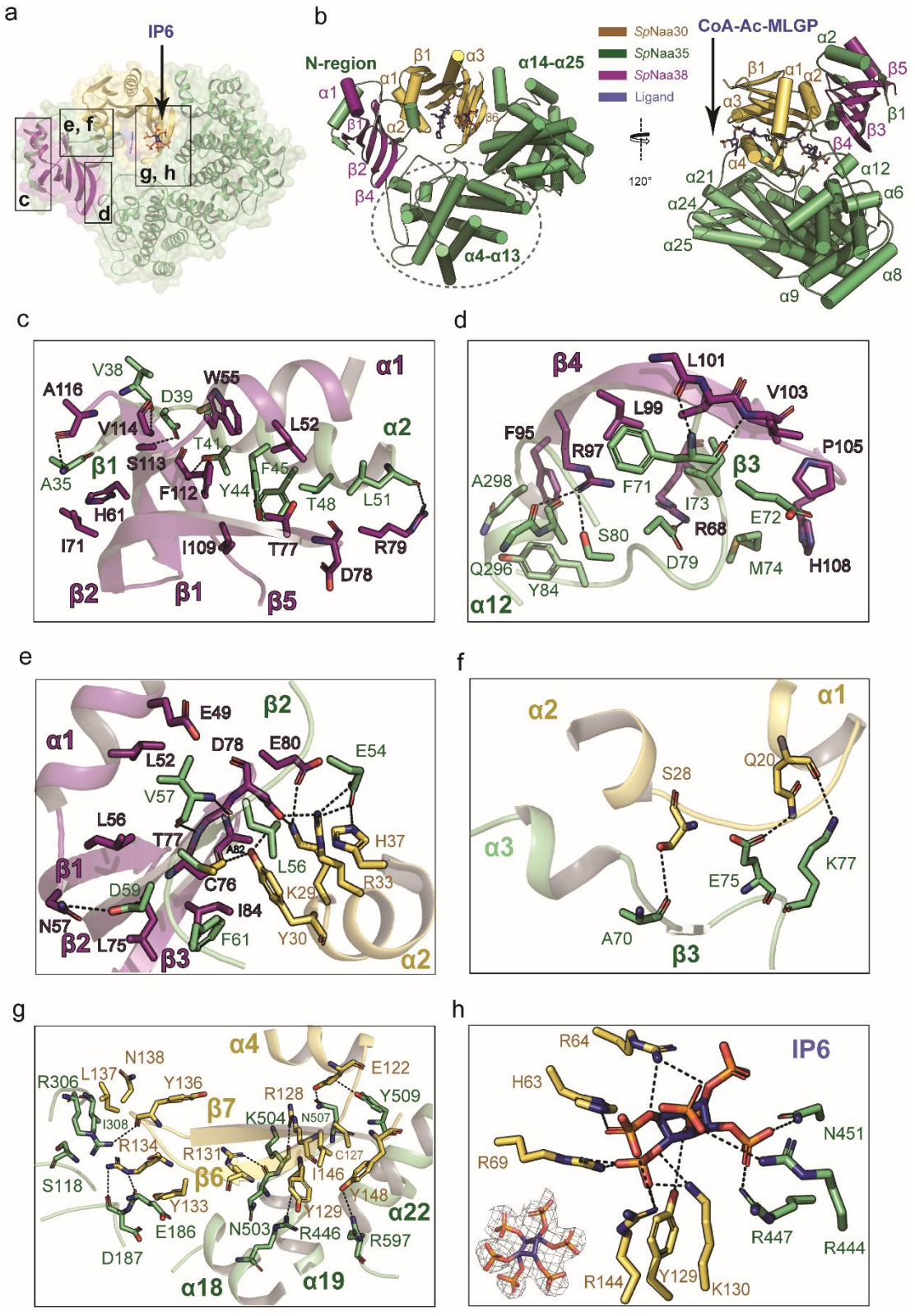

a
IP6
e, f
g, h
c
d
b
N-region
α14-α25
α4-α13
SpNaa30
SpNaa35
SpNaa38
Ligand
120°
CoA-Ac-MLGP
c
d
e
f
g
h
R64
IP6
H63
N451
R69
R447
R444
R144
Y129
K130

Figure 29.4 Overall structure of SpNatC reveals intimate interactions among all subunits

(a) NAA30 (bright orange), NAA38 (deep purple) and NAA35 (lime green) are shown in transparent surface and cartoon, with the boxed area labeled. (**b)** The CoA-MLGP bisubstrate conjugate and inositol hexaphosphate (IP_6) is shown in sticks and colored in blue. (**c)** Zoom-in view of the first sub-interface between NAA38 and NAA35. Hydrogen bonds are indicated by dashed black lines. For simplicity, only some of the hydrophobic interactions are shown. (**d)** Zoom-in view of the second sub-interface between NAA38 and NAA35. (**e)** Zoom-in view of the third sub-interface between NAA38 and NAA35, which also involves some residues from NAA30 (light orange). (**f)** Zoom-in view of the interface between NAA35 and NAA30 α1-α2 segment. (**g)** Zoom-in view of the interface between NAA35 and NAA30 β6-β7 segment. (**h)** Zoom-in view depicting key residues involved in interactions with IP_6. The lower left shows the fit of the IP_6 in the EM density map. The contour level is 5.0 sigma.

5.2.4 NAA38 plays a key role in the NatC complex

The structure of *Sp*NatC was consistent with our biochemical data suggesting that Naa38 plays important structural and functional roles within the NatC complex. A more detailed view of the Naa35-Naa38 interface reveals that it is nucleated by the formation of three anti-parallel β-sheets between Naa35 and Naa38, resulting in extensive hydrogen bonding and van der Waals interactions, which also involve the a2 and a3 helices of Naa35. These interactions can roughly be divided into three regions according to the relative position of the β-sheets (**Figure 5.4**).

The first sub-interface involves the Naa35 β1-α2 segment (**Figure 5.4c**). Hydrogen bonds are formed between the backbone nitrogen of Naa35-Ala35 and the backbone carbonyl of Naa38-Ala116, the backbone carbonyl of Naa35-Val38 and the backbone nitrogen of Naa38-Val114, the sidechains of Naa35-Asp39 and of Naa38-Ser113, the sidechain of Naa35-Thr41 and the backbone nitrogen of Naa38-Phe112, and the backbone carbonyl of Naa35-Leu51 and the sidechain of Naa38-Arg79. This region also features a large area of van der Waals interactions involving residues Ala35, Gly36, Tyr37, Tyr44, Phe45, Ala47, and Thr48 from Naa35 and residues Gly48, Leu52, Trp55, His 61, Ile71, Thr77, Asp78, Ile109, Arg115, and Ala116 from Naa38.

The second sub-interface is mediated by Naa35-β3 and α12 (**Figure 5.4d**). Hydrogen bonds are formed between the backbone nitrogen of Naa35-Phe71 and the backbone carbonyl of Naa38-Leu101, the backbone carbonyl of Naa35-Phe71 and the backbone nitrogen of Naa38-Leu103, the sidechains of Naa35-Asp79 and Naa38-Arg68, the sidechains of Naa35-Ser80 and Naa38-Arg97, and the backbone carbonyl of Naa35-Gln296 and the sidechain of Naa38-Arg97. Residues involved in van der Waals

interactions include Ala70, Glu72, Ile73, Met74, Tyr84, Ala 297, Gln 298, Val30, and Ala301, from Naa35, and residues Phe95, Ala98, Leu99, Val102, Ile104, Pro105, and His108 from Naa38.

The third sub-interface is mediated by Naa35-β2 but also involves extensive interactions with the catalytic Naa30 subunit (**Figure 5.4e**). The hydrogen bonds between Naa35 and Naa38 are formed between the sidechains of Naa35-Asp59 and Naa38-Asn57, the backbone nitrogen of Naa35-Val57 and the backbone carbonyl of Naa38-Thr77, and the backbone carbonyl of Naa35-Val57 and the backbone nitrogen of Naa38-Thr77. Naa38 engages with the α2 helix of the Naa30 subunit to form an extensive hydrogen bonding network via the Naa30-Tyr30 sidechain to the backbone carbonyl of Naa38-Ala82, and the sidechains of Naa38-Cys76 and Asp78; the sidechain of Naa30-Lys29 hydrogen bonds to the sidechains of Naa38-Asp78 and Glu80; and between the sidechains of Naa30-Arg33 and Naa38-Asp78. As the a1-a2 loop of Naa30 plays a key role in protein N-terminal substrate recognition, these Naa38-Naa30 interactions likely play a key role in substrate recognition by NatC (see discussion). Finally, the Naa30-a2 helix also interacts with Naa35 where Naa35-Glu54 forms hydrogen bonds with Naa30-Arg33 and His37. Within this sub-interface, the hydrophobic interactions are primarily mediated by residues Glu49, Leu52, Leu56, Leu75, Cys76, Asp78, and Ile84 from Naa38 and residues Leu56, Cys58, Asp59 and Phe62 from Naa35. Taken together, it appears that Naa38 plays key structural and, likely, functional roles in NatC.

5.2.5 The substrate binding α1-α2 loop and β6-β7 segments of NAA30 are buried within NAA35

Naa30 uses its α1-α2 loop (**Figure 5.4f)** and β6-β7 segments to interact with Naa35 **(Figure 5.4g)**. Analogous segments in the NatA and NatB complexes are used for N-terminal substrate recognition. In these complexes, however, the β6-β7 segment is exposed to solvent (Gottlieb and Marmorstein 2018; Liszczak et al. 2013; Deng, Pan, et al. 2020; Hong et al. 2017). By contrast, this interaction between Naa30 and Naa35 is facilitated by hydrogen bonds between the Naa30 α1-α2 loop and Naa35 α3-β3 segment: the sidechain of Naa30-Ser28 and the backbone carbonyl of Naa35-Ala70; between the sidechains of Naa30-Gln20 and Naa35-Glu75; and the backbone carbonyl of Naa30-Gln20 and the sidechain of Naa35-Lys77.

Notably, the interactions between Naa30 β6-β7 and Naa35 are significantly more extensive than those found in the NatA and NatB complexes **(Figure 5.4g)**. Here, hydrogen bonds are formed between the sidechain of Naa30-Glu122 and the sidechains of Naa35-Gln507 and Tyr509, the sidechain of Naa30-Arg128 and the backbone carbonyl of Naa35-Lys504, the backbone carbonyl of Naa30-Tyr129 and the sidechain of Naa35-Arg446, the sidechain of Naa30-Arg131 and the backbone carbonyl of Naa35-Gln503, the sidechain of Naa30-Arg134 and the backbone carbonyls of both Naa35-Glu186 and Asp187, the backbone carbonyl of Naa30-Tyr136 and the sidechain of Naa35-Arg306, and the sidechains of Naa30-Tyr148 and Naa35-Arg597. Van der Waals interactions are observed between residues Cys127, Leu132, Tyr133, Leu137, Asn138, Phe143, Ile146, and Tyr148 from Naa30, and residues Ile308, Asn441, Cys443, Leu505, Phe581, Ser590, Tyr591, and Ala594 from Naa35. Together, it appears that the Naa35

auxiliary subunit makes extensive contacts with the protein N-terminal substrate binding loops of Naa30 to influence substrate recognition.

5.2.6 IP_6 is bound to the interface between NAA30 and NAA35 in close proximity to the peptide substrate binding site

EM density corresponding to a bound IP_6 molecule is well-resolved, revealing a IP_6 binding pocket at the interface between the Naa30 and Naa35 subunits (**Figure 5.4h).** Notably, the IP_6 binding region is distinct from its binding sites in the NatA and NatE complexes (Deng, McTiernan, et al. 2020; Deng et al. 2019; Gottlieb and Marmorstein 2018). In *Sp*NatC, IP_6 engages in polar interactions with Naa30 residues His63, Arg64, Arg69, Tyr129, Lys130, and Arg144, and Naa35 residues Arg444, Arg447, and Asn451. Interestingly, although the IP_6 binding pocket of NatC features a similar chemical environment to the one found in the NatA and NatE complexes, the location in the architecture of these NAT complexes differs. In NatC, IP_6 is located close (~8 Å) to the peptide substrate entrance site, while the pocket found in Nat*A* and NatE is close to the Ac-CoA entrance site (**Figure 5.5)**. Therefore, the NatC-bound IP_6 molecule could be in position to impact NatC activity. To evaluate this possibility, we assayed the effects of IP_6 on NatC activity and found that IP_6 had a relatively modest inhibitory effect on NatC activity, comparable to its similar modest inhibitory activity on *Sp*NatA and hNatB (**Figure 5.6**). While we cannot exclude the possibility that this modest inhibitor activity is an artifact of the effect of IP_6 on the radioactive activity, it appears that IP_6 plays a predominantly structural role in NatC.

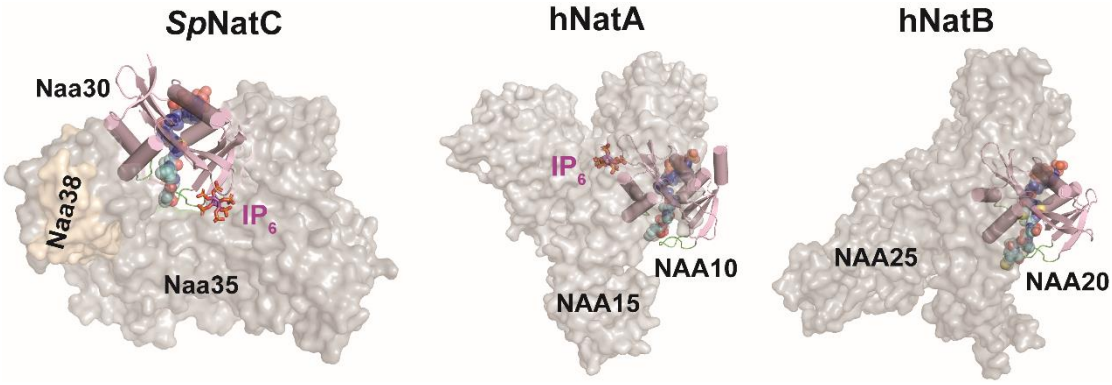


Figure 30.5 Comparison of NatC with Nat and NatB reveals divergent NatC architecture

The auxiliary subunits are shown as transparent surface in grey or brown, while the catalytic subunits are shown as cartoon in light pink (catalytic subunits are aligned in the same orientation). The blue spheres, cyan spheres, and magenta sticks represent Ac-CoA, peptide substrate, and IP_6, respectively. The highlighted green loops shown are α1-α2 and β6-β7 substrate binding loops of the catalytic subunits. The PDB models for generating this figure: *Hs*NatA with IP_6, PDB: 6C9M (the bisubstrate shown is aligned from *Sp*NatA, PDB:4KVM); *Hs*NatB with the bisubstrate analogue, PDB: 6VP9; model of *Sp*NatC with the bisubstrate analogue and IP_6, PDB: 7L1K. As NatE is a complex with dual catalytic subunits and shares the same auxiliary subunit with NatA, for simplicity, it is not compared in this figure.

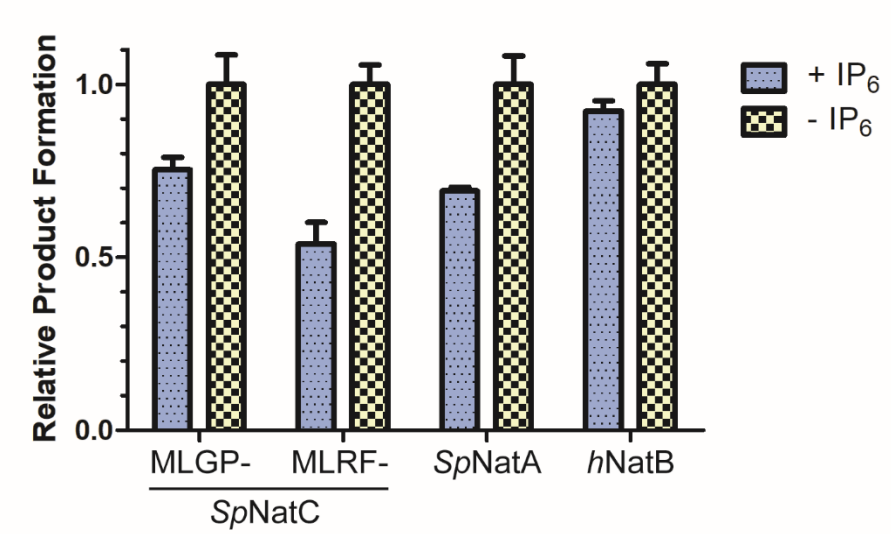


Figure 31.6 The activity impacts of IP_6 on selective NATs

50 nM of either SpNatC, SpNatA, and hNatB was mixed with 2 µM or no IP_6 for measuring production formation. Data was normalized to NATs activity without IP_6. Error bars indicates Mean with SEM, n=3.

5.2.7 NatC displays significant differences with NatA, NatE, and NatB

Overall, there are three notable differences between NatC and the other multi-subunit NATs – NatA, NatE and NatB. First, the auxiliary subunits of NatA, NatE and NatB contain only helical secondary structure, forming 10-13 conserved tetratricopeptide repeat (TPR) motifs (Gottlieb and Marmorstein 2018; Liszczak et al. 2013; Hong et al. 2017; Deng, Pan, et al. 2020; Deng, McTiernan, et al. 2020; Deng et al. 2019). In contrast, the NatC auxiliary subunits (Naa35 and Naa38) do not contain TPR repeats and but do contain several beta strands that make key interactions in the complex (**Figure 5.4b).** Secondly, the auxiliary subunits of NatA, NatE, and NatB form a closed ring-like cradle to completely wrap around their corresponding catalytic subunit. In contrast, NatC auxiliary subunits arrange themselves into a clamp-like structure to only wrap roughly half-way around the Naa30 catalytic subunit (**Figure 5.5).** Thirdly, the Naa30 β6-β7 peptide substrate recognition loop is buried within the auxiliary subunit, while the corresponding loops in NatA, NatE, and NatB are largely exposed to solvent (**Figure 5.5).** Thus, the relative orientation of the catalytic subunit to the auxiliary subunit (s) is different in NatC relative to NatA and NatB. In turn, this has implications in altering the NatC substrate binding mode: the NatC substrate binding tunnel is roughly perpendicular to the Naa30-Naa35 interface, while the substrate binding tunnel of NatA, NatE, and NatB is parallel to the catalytic-auxiliary subunit-interface (**Figure 5.5).** Taken together, the NatC complex forms an architecture that is distinct from other multi-subunit NATs.

5.2.8 Substrate recognition by NAA30 shows similarity to NAA50

Overall, Naa30 displays a typical NAT fold containing four α-helices and seven β-strands with similar substrate binding modes: Ac-CoA enters the catalytic active site through a groove formed by α3 and α4 of the catalytic subunit, while the peptide substrate enters the active site on the opposite side of the catalytic subunit flanked by the α1–α2 and β6-β7 peptide substrate binding loops (Deng and Marmorstein 2021) (**Figure 5.7a and 5.7b).** In the Cryo-EM map, density for the CoA-Ac-MLGP conjugate bisubstrate is well-resolved, allowing us to confidently model both the CoA portion and all four residues of the peptide portion (**Figure 5.7c).**

Naa30 harbors a conserved Ac-CoA binding motif $R_{82}XXG_{85}XA_{87}$ where Ac-CoA binding is mediated by a series of hydrogen bonds, mainly to the pyrophosphate group. Specifically, the sidechains of Arg82, Ser119, Arg123 and backbone atoms from Ile77, Gly85, Ile86, and Ala87 contribute to Ac-CoA hydrogen bonding (**Figure 5.7d).**

In our NatC model, Naa30 binds the peptide substrate mainly through peptide backbone hydrogen bonding and hydrophobic pockets for the sidechains of the first two peptide residues. Direct hydrogen bonds are formed between the backbone amide groups of the peptide residues 1 and 2 to Naa30 residues Tyr27, Glu109, Thr110, Tyr121, and Tyr136 (**Figure 5.7d).** The binding pocket for the peptide Met1 is surrounded by residues Leu23, Ser24, Glu25, Glu111, and Tyr136 (**Figure 5.7e)**, while the binding pocket for peptide Leu2 is half-open and surrounded by Tyr27, Val31, Tyr34, Phe35, Lys59, Tyr71, and Met74 (**Figure 5.7f).** Beyond the first two residues, there are no direct contacts between the peptide substrate and Naa30, with the exception of distant contacts between the backbones of peptide Gly3 and peptide Pro4 with Naa30-

Tyr135 (4-4.5 Å) and Naa30-Glu25 (~4 Å), respectively. These two Naa30 residues likely play a role in recognizing positions 3 and 4 of high-affinity NatC peptide substrates, as supported by the recent report by Grunwald *et al.* (Grunwald et al. 2020)

Previous studies have described the overlapping substrate profile of NatC and NAA50. Since our NatC model contains the canonical NatE (NAA50)-type peptide substrate (Met2-Leu2-Gly3-Pro4) in the bisubstrate inhibitor, we were able to evaluate this relationship by visual inspection of the inhibitor-bound *Sp*Naa30 subunit from the model. The *Sp*Naa30 model aligns well with the *Hs*NAA50 structure (PDB: 3TFY) bound to the same MLGP peptide fragment, with a root-mean-square deviation (RMSD) of 1.23 Å (over 105 common C_α atoms) (**Figure 5.7b**). A more detailed view of the peptide substrate binding site reveals that Met1 sits in similar binding pockets in both structures (**Figure 5.7e)**, consistent with their high specificity for Met1. Specificity for Leu at position 2 is also similar. Notably, residues NatC-Tyr27 and Tyr71, which are responsible for Leu2 backbone recognition align well with the corresponding residues in NAA50 (**Figure 5.7f)**. For Leu2 sidechain recognition, the residues, Naa30-Val31 and Lys59 are replaced by bulkier residues, NAA50-Phe35 and Arg62 (**Figure 5.7f)**, respectively. However, Naa30-Phe35, Ala73 and Met74 are replaced by NAA50-Phe35, Met75 and Thr76 (**Figure 5.7f),** perhaps explaining the greater tolerance of NAA50 for other residues at position 2. In both models, the peptide substrates begin to diverge from the architecture of their respective models at residue 3, where the Ca atoms of the corresponding peptide Gly3 residues of both models are ~3 Å apart. This also seems to be correlated with a shift in the *p*-hydroxyl group of a nearby conserved Tyr residue (Naa30-Tyr135 and NAA50-Tyr138) by about 2.5 Å (**Figure 5.7g**). The corresponding peptide Pro4 residues also point in different directions (**Figure 5.7g**). We propose that

these differences in peptide substrate positioning beyond residue 2 is mediated by the different active site environments of Naa30 and NAA50 in this region, thus allowing these two enzymes to harbor varying activities toward substrates with the same residues at positions 1 and 2 but differing residues at position 3 and beyond (See discussion). Taken together, this comparison explains the overlapping yet distinct substrate profiles of NatC and NatE.

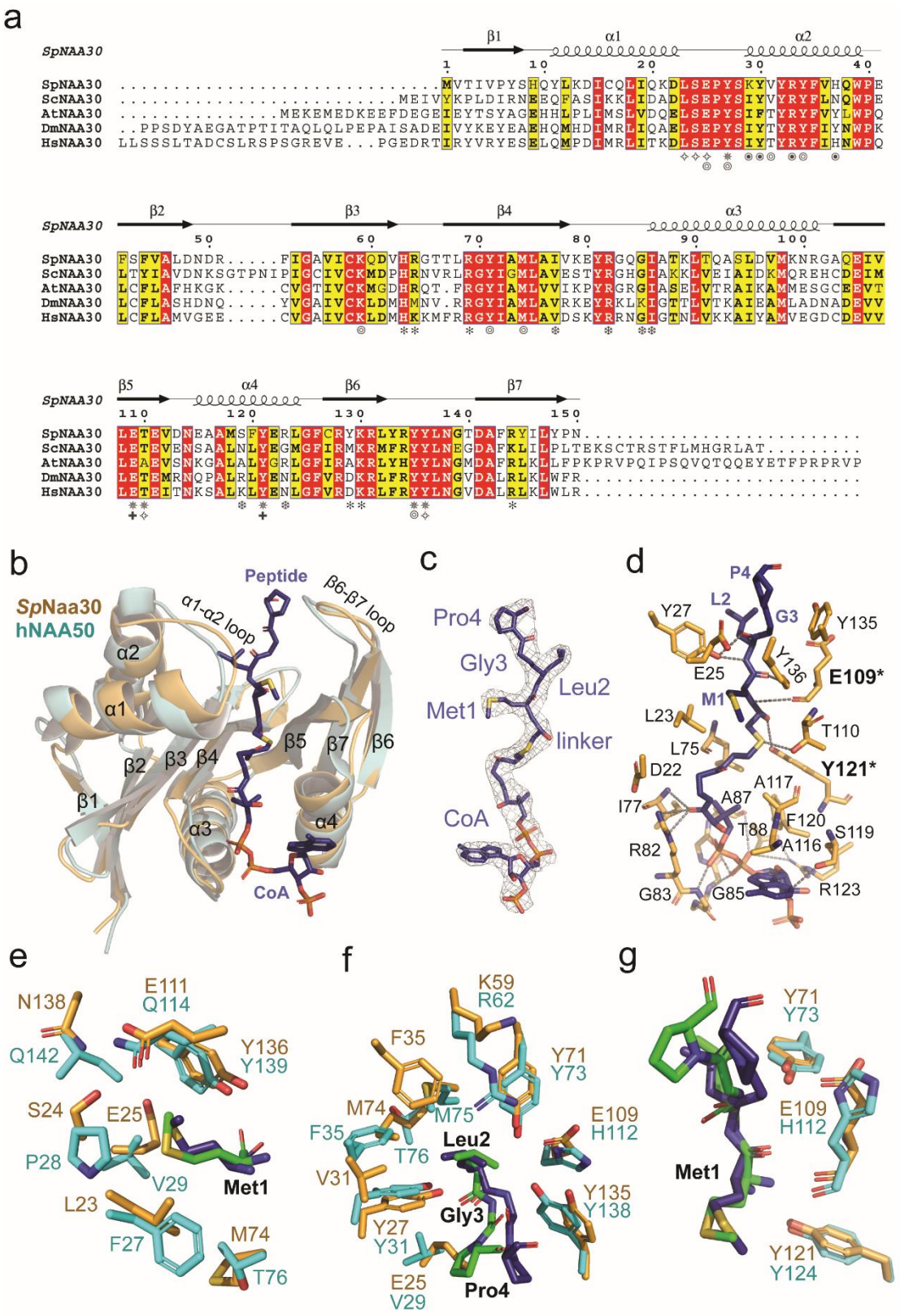

a
SpNAA30
ScNAA30
AtNAA30
DmNAA30
HsNAA30
b
SpNaa30
hNAA50
Peptide
α1-α2 loop
β6-β7 loop
CoA
c
Pro4
Gly3
Met1
Leu2
linker
CoA
d
P4
L2
G3
E109*
Y121*
M1
e
f
Leu2
Gly3
Pro4
g
Met1

Figure 32.7 Substrate recognition by NAA30 is similar to NAA50

(a) The sequence alignment of NAA30 orthologs from *Schizosaccharomyces pombe* (Sp), *Saccharomyces cerevisiae (*Sc), *Arabidopsis thaliana* (At), *Drosophila melanogaster (*Dm), and *Homo sapiens* (Hs). Numbering and secondary structure elements for NAA30 are indicated above the sequence alignment. Residues of NAA30 that contact the peptide backbone (✳), CoA (❄), Met1 (✧), Leu2 (◎), NAA38 (◉), IP_6 (✻), and catalytic residues (✚) are indicated. (**b)** Structural alignment of *Sp*NAA30 (bright orange) with *Hs*NAA50 (cyan), with secondary structure elements indicated. (**c)** The fit of the bi-substrate inhibitor in the EM density map. The contour level is 5.0 sigma. (**d)** Highlighted polar and hydrophobic interactions between CoA-Ac-MLGP and NAA30 are depicted in the 3D view. (**e)** Residues forming a hydrophobic pocket surrounding the substrate peptide Met1 side chain are shown in sticks (Orange, *Sp*NAA30; Cyan, *Hs*NAA50). (**f)** Residues forming a hydrophobic pocket surrounding the substrate peptide Leu2 side chain are shown in sticks. (Orange, *Sp*NAA30; Cyan, *Hs*NAA50). (**g)** Residues proposed as catalytic residues are shown in sticks. (Orange, *Sp*NAA30, E109 and Y121; Cyan, *Hs*NAA50, Y73 and H112).

5.2.9 Mutational analysis reveals that NAA30 Tyr-121 and Glu-109 play key catalytic roles

In order to evaluate key Naa30 residues identified in our model and their functional roles in substrate binding and catalysis, we used a radioactive *in vitro* acetyltransferase activity assay in conjunction with the MLRF peptide as the peptide substrate to kinetically characterize WT and mutant *Sp*NatC proteins. Each mutant was

purified to homogeneity and displayed identical gel filtration chromatography elution profiles (data not shown), consistent with their native folding and complex formation. Consistent with our structural observations, mutation of a majority of residues involved in peptide substrate recognition resulted in an increase in K_m (**Table 5.2)**. As the "YY motif" is shown to be conserved in most NATs (Stove et al. 2016) (except NatD/NAA40), single mutation of either Naa30-Tyr135 or Tyr136 to an alanine resulted in a significant loss in activity, with the Y136A mutant displaying almost no detectable activity. In previous studies, NAA50 residues Tyr73 and His112 were proposed to contribute to catalysis (Liszczak, Arnesen, and Marmorstein 2011). The corresponding residues in Naa30, Tyr71 and Glu109, however, had an unexpected effect on catalysis. Instead, Naa30-Y71A or Y71F primarily altered the K_m, which ultimately resulted in a minor decrease (12-67%) in catalytic efficiency in comparison with WT. On the other hand, mutation of Naa30-Glu109 to either alanine or glutamine led to >90% loss of catalytic efficiency, impacting both K_m and k_{cat}. NAA50-Tyr124 and hNatB-Tyr123 were recently proposed to play analogous catalytic roles (Deng, Pan, et al. 2020), and the corresponding mutation of Naa30-Tyr121 to alanine demonstrated that NatC activity was almost completely abolished (**Table 5.2).** While Glu109 is positioned to function as a general base to deprotonate the a-amino group of the peptide substrate, Tyr121 is positioned to function as a general acid to re-protonate the CoA leaving group. This is consistent with the proposed roles of the analogous Glu118 and Tyr130 residues in *Sc*NatC *(Grunwald et al. 2020),* as well as the conserved nature of these residues (**Figure 5.7a**). In addition, Naa30-Asn114 is highly conserved among all NATs (data not shown) and it was recently suggested to play a structural role in orienting the catalytic tyrosine residues (*Sp*Naa30-Tyr135 and Tyr136) by ensuring the proper position of the active site helix (a4) for interaction with Ac-CoA (Deng, Pan, et al. 2020). Consistent with the importance of

Asn114, we observed that Naa30-N114A exhibited an ~50% loss in activity. Taken together, the conserved Naa30-Glu109 and Tyr121 residues likely play key catalytic roles in NatC complex activity (**Figure 5.7g).**

Table 9.2 Catalytic parameter of wild-type and mutant NAA30 in *Sp*NatC

Substrate	Protein (*Sp*NatC)	k_{cat}* (min^{-1})	Relative k_{cat} (normalized to WT)	K_m* (μM)	Relative K_m (normalized to WT)	k_{cat}/K_m (normalized to WT)
gag Peptide (MLRF-)	WT	23.9 ± 1.4	1.0	7.6 ± 2.1	1.0	1.0
	S24A	22.7 ± 1.6	0.95	10.3 ± 3.2	1.4	0.70
	E25A	49.3 ± 2.3	2.1	19.7 ± 3.6	2.6	0.79
	Y27A	7.33 ± 0.5	0.31	3.0 ± 1.0	0.39	0.79
	Y71A	44.7 ± 2.8	1.9	16.1 ± 4.2	2.1	0.88
	Y71F	30.2 ± 2.3	1.3	29.6 ± 8.5	3.9	0.32
	M74A	13.7 ± 1.1	0.57	18.7 ± 6.1	2.5	0.23
	E109A	10.5 ± 1.6	0.44	28.7 ± 16	3.8	0.11
	E109Q	3.2 ± 0.9	0.13	201 ± 130	27	0.0050
	N114A	9.5 ± 0.6	0.40	5.6 ± 1.7	0.74	0.53
	Y121A	1.2 ± 0.1	0.050	16.0 ± 5.4	2.1	0.023
	Y121F	1.4 ± 0.2	0.060	7.9 ± 4.3	1.0	0.057
	Y135A	6.3 ± 0.8	0.26	9.5 ± 5.7	1.3	0.21
	Y136A	N.A	N.A	N.A	N.A	N.A
*values indicated represent mean ± S.D. of best-fit curves						

5.3 Discussion

Here, were report that the notably small *Sp*Naa38 auxiliary subunit cooperates with the large Naa35 auxiliary subunit to stabilize the active N-terminal acetyltransferase NatC complex for its robust N-terminal acetylation activity. Through structure/function studies, we observed that Naa38 interacts with Naa30-α2, a region that directly contributes to recognition of peptide substrate backbone and sidechain residues. Based on these observations and similar *Sc*NatC complex biochemical and structural findings described by Grunwald *et al. (Grunwald et al. 2020),* it appears that Naa38 is essential for normal NatC acetylation activity. Therefore, we propose that Naa38 plays a conserved role in yeast. Future studies are needed to further interrogate this in the plant or human systems, as some reports suggest that Naa38 may not be as important in higher eukaryotes (Pesaresi et al. 2003; Starheim et al. 2009), but there has also been a report of a human developmental disorder linked to *NAA38* deletion (Zhao et al. 2016).

In *S. pombe*, we have found that the ternary *Sp*NatC complex displays acetylation activity toward both canonical NatC and NatE-type peptide substates, but with a preference for NatC-type substrates. While recognition of residues 1 and 2 of the cognate substrate appears highly homologous between the Naa30 and NAA50 catalytic subunits of NatC and NatE, respectively, the substrate binding pockets responsible for recognizing substrate residues beyond the penultimate sidechain diverge significantly. This likely explains the divergence in substrate profiles observed between these two NATs.

IP_6 was previously found to play a stabilizing role at the interface between the NAA15 and NAA10 subunits of the NatA and NatE complexes (Deng, McTiernan, et al. 2020; Deng et al. 2019; Gottlieb and Marmorstein 2018), (Cheng et al. 2019). Here, we

find that IP_6 also appears to play a stabilizing role in both *Sc*NatC and *Sp*NatC, and that the *Sp*NatC structure reveals that IP_6 binds at the Naa35-Naa30 interface located near the peptide substrate entrance site, although it does not appear to play a major direct role in modulating NatC activity.

Grunwald *et al.* recently reported X-ray crystal structures of *Sc*NatC in several liganded states along with associated biochemical studies *(Grunwald et al. 2020)*. Overall, their findings are in agreement with our findings on *Sp*NatC reported here. There are, however, some notable differences when comparing NatC from *S. pombe* and *S. cerevisiae*, which have implications for the conserved and unique features of NatC.

The *Sc*NatC structures bound to cognate NatC-type peptide substrates shows specificity for the first four residues, which is consistent with their reported peptide substrate mutation and binding data. In contrast, our structure, featuring *Sp*NatC bound to a peptide sequence that is optimal for NatE (NAA50) acetylation, shows *Sp*Naa30's specificity for only the first two residues. This mode of recognition is consistent with recognition for these same residues by NAA50, while there is significant divergence in the regions of the peptide binding grooves of Naa30 and NAA50 that are responsible for interaction with substrates residues beyond residue 2. This comparison suggests that discrimination between Naa30 and NAA50 substrates is largely dictated by the identities of the third and fourth residues in an N-terminal substrate. As NatC-type and NatE-type substrates can include many different hydrophobic/amphipathic N-termini (ML- MI-, MF, MW-, MV-, MM-, MH-, and MK-) (Ree, Varland, and Arnesen 2018), it may be difficult for a single enzyme to cover acetylation of such a broad repertoire of protein N-termini. Given that Naa50 is inactive in yeast (Deng et al. 2019), we propose that the overlapping

substrate profile for residues at positions 1 and 2 by NatC and NatE may have evolved in higher eukaryotic cells to fully cover the acetylation on these types of N-terminal substrates.

Grunwald *et al.* identified several electropositive regions (EPR) on *Sc*NatC that are implicated in ribosome association. Of these regions, labeled EPR1-EPR4 (Grunwald et al. 2020) (**Figure 5.8**), mutation of only EPR2 was found to influence ribosome association. Comparison with the electrostatic surface potential of *Sp*NatC shows notable differences with *Sc*NatC (**Figure 5.8)**. Specifically, EPR2 is not present in *Sp*NatC, since the *Sc*NatC "*Sc*Naa35 tip" is missing in *Sp*NatC. While EPR1 is conserved, this electropositive patch harbors the IP_6-binding pocket in *Sp*NatC and, therefore, is likely the IP_6 binding site in *Sc*NatC. In addition, the surface region of *Sp*NatC corresponding to EPR3 is broader compared to the corresponding surface of *Sc*NatC and is located close to where *Sc*NatC-EPR2 would be. It is therefore possible that *Sp*NatC utilizes EPR3 to compensate for the absence of EPR2. Taken together, this comparison suggests that the mode of ribosome association utilized by *Sp*NatC and *Sc*NatC and, by extension, NatC complexes from other species, may differ. Future studies would investigate the molecular details of these differing interactions.

NatC has been shown to be important for proper chloroplast (Pesaresi et al. 2003) and mitochondrial (Van Damme et al. 2016a) function, essential for embryonic development (Wenzlau et al. 2006), cell viability and p53-dependent apoptosis (Starheim et al. 2009), and as a potential therapeutic target in cancer (Mughal et al. 2015). In light of these important connections to NatC function, the studies reported here have important implications for health and disease.

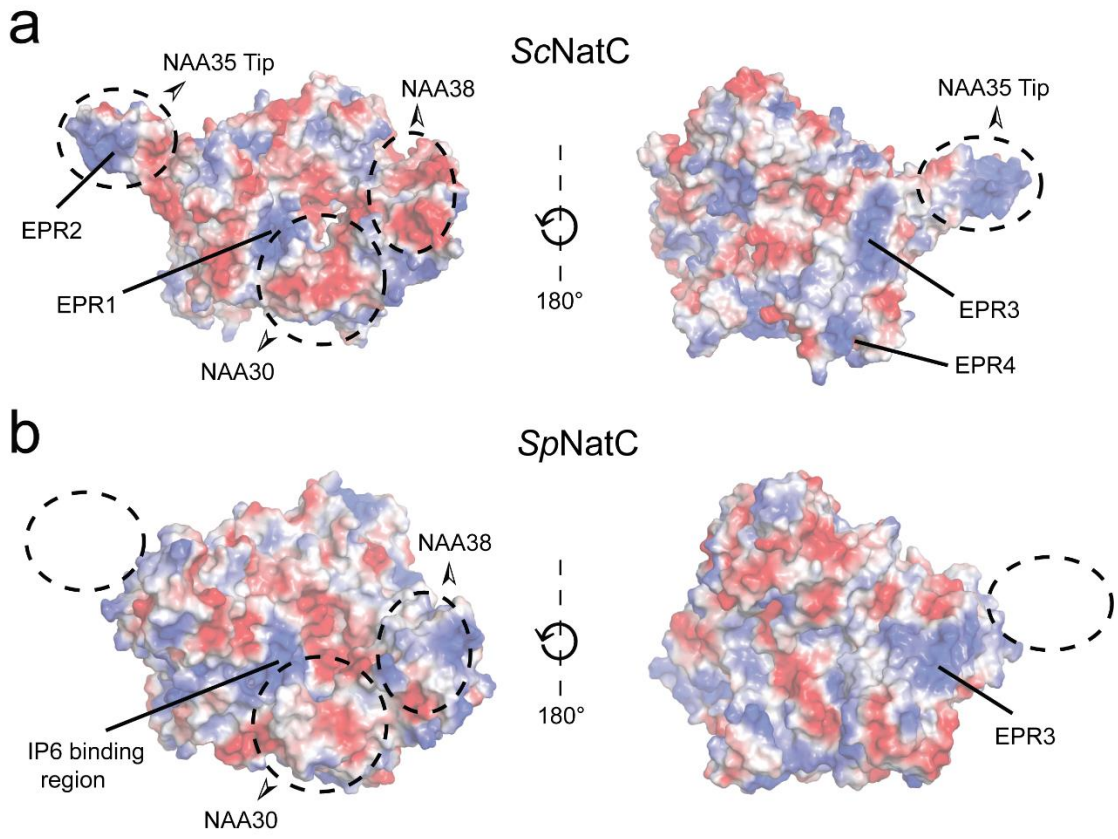


Figure 33.8 Divergent structural features exist between NatC in S. pombe and S. cerevisiae

(a) Electrostatic surface potential of the *Sc*NatC structure (PDB: 6YGB) with previous EPRs indicated. (**b)** Electrostatic surface potential of *Sp*NatC (PDB: 7L1K) aligned with the orientations shown in (a).

5.4 Materials and Methods

5.4.1 SpNatC expression and purification

*Sp*Naa35$^{31\text{-}708}$ (with truncation of the N-terminal 30 residues) and full-length *Sp*Naa30^{FL} were cloned into a modified pET DUET vector containing an N-terminal His_6-SUMO tag. *Sp*Naa38$^{48\text{-}116}$ (with truncation of the N-terminal 47 residues) was cloned into a pCDF vector with an N-terminal His_6 tag. To obtain the heterotrimeric NatC complex (WT ternary *Sp*NatC), these two plasmids were used to co-transform Rosetta (DE3)pLysS competent *E. coli* cells, which was then cultured to grow at 37 °C in the presence of both ampicillin (100 µg/mL) and streptomycin (50 µg/mL). When the absorbance OD_{600} reached ~ 0.7, protein expression was induced at 16°C with 0.5 mM isopropyl 1-thio-β-galactopyranoside (IPTG) overnight. Cells were harvested by centrifugation, resuspended, and lysed by sonication in lysis buffer containing 25 mM Tris, pH 8.0, 300 mM NaCl, 10 mg/ml PMSF (phenylmethanesulfonylfluoride). The lysate was clarified by centrifugation and passed over Nickel-NTA resin (Thermo Scientific), which was subsequently washed with ~10 column volumes of wash buffer containing 25 mM Tris, pH 8.0, 300 mM NaCl, 20 mM imidazole, 10 mM 2-mercaptoethanol (βME). The protein was eluted in batches with buffer containing 25 mM Tris, pH 8.0, 300 mM NaCl, 200 mM imidazole, 10 mM βME. After elution, His_6-tagged Ulp1 protease was added to the eluate to cleave the His_6-SUMO tag and dialyzed further into buffer containing 25 mM sodium citrate monobasic, pH 5.5, 10 mM NaCl and 10 mM βME. Protein was purified with a 5-mL HiTrap SP cation-exchange column (GE Healthcare) and eluted with a salt gradient (10–1000 mM NaCl). Peak fractions were concentrated to ~0.5 mL with a 50 kDa concentrator (Amicon Ultra, Millipore), and loaded onto an S200 gel-filtration column (GE Healthcare) in a buffer containing 25 mM HEPES, pH 7.0, 200

mM NaCl, 1 mM dithiothreitol (DTT). Peak fractions were pooled and concentrated to ~15 mg/ml as measured by UV_{280} using a Nanodrop and flash-frozen for storage at −80 °C until use. All protein mutants were generated using the QuikChange protocol from Stratagene and obtained following the expression and purification protocols described above.

5.4.2 Binary and ternary ScNatC expression and purification

*Sc*Naa30$^{1-161}$ and *Sc*Naa35$^{19-733}$ were cloned into a pET DUET vector containing an TEV cleavable N-terminal His_6-tag. Binary complex (*Sc*Naa30$^{1-161}$/*Sc*Naa35$^{19-733}$) was obtained by transforming this plasmid in ScarabXpress T7lac (Scarab Genomics) competent *E. coli* cells, which were grown in terrific broth media (DOT Scientific) supplemented with ampicillin (100 µg mL^{-1}) at 37°C to an OD_{600} of ~0.9 and induced by addition of 0.5 mM IPTG at 17°C for 16 hr. All subsequent purification steps were carried out at 4°C. Cells were isolated by centrifugation, lysed by sonication in lysis buffer containing 25 mM Tris, pH 8.0, 300 M NaCl, 10 mM β-ME and 10 mg/mL PMSF. The lysate was clarified by centrifugation and passed over nickel resin, which was subsequently washed with >20 CV of lysis buffer supplemented with 25 mM imidazole. The protein was eluted in lysis buffer supplemented with 200 mM imidazole by batch elution. TEV protease (~1 mg/ml) was added to the eluted protein and dialyzed into buffer containing 25 mM Tris, pH 8.0, 300 mM NaCl, 5 mM Imidazole, 10 mM β-ME. This solution was passed through an additional nickel column to remove TEV protease as well as any uncut binary *Sc*NatC. The resin was then washed with approximately 3 CV of dialysis buffer supplemented with 25 mM imidazole, which was then pooled with the initial flowthrough. This solution was dialyzed into ion exchange buffer containing 25 mM HEPES, pH 7.5, 50 mM NaCl and 10 mM β-ME and loaded onto a 5 ml HiTrap SP anion

exchange column (GE HealthcarE). The binary complex *Sc*NatC was then eluted using a salt gradient (50–750 mM NaCl). Peak fractions were concentrated using a 50-kDa MWCO concentrator (Amicon) and further purified Superdex 200 Increase 10/300 GL gel filtration column (GE Healthcare) in a buffer containing 25 mM HEPES, pH 7.0, 200 mM NaCl, 1 mM TCEP. Peak fractions were concentrated to a UV_{280} of ~4-5 mg mL^{-1} as measured by nanodrop. The protein was then flash-frozen and stored at -80°C until use.

To obtain the ternary *Sc*NatC complex, *Sc*Naa38$^{1-70}$ was cloned into a pRSF vector with a TEV-cleavable N-terminal STREP-tag. This plasmid was co-transformed with the pET DUET plasmid containing *Sc*Naa30$^{1-161}$ and *Sc*Naa35$^{19-733}$ and cultured as described above except with the addition of kanamycin (50 µg mL^{-1}) to select for pRSF *Sc*Naa38$^{1-70}$ plasmid. The purification of the ternary *Sc*NatC complex was the same as described for the binary *Sc*NatC complex.

5.4.3 Acetyltransferase activity assays

All acetyltransferase assays were carried out at room temperature in a reaction buffer containing 75 mM HEPES, pH 7.0, 120 mM NaCl, 1 mM DTT as described previously (Deng et al. 2019). The full sequence of peptide substrates are listed below: “SASE” peptide (NatA-type): NH_2-SASEAGVRWGRPVGRRRRP-COOH; “MDVF” peptide (NatB-type): NH_2-MDVFMKGRWGRPVGRRRRP-COOH; “MLRF” peptide (NatC-type): NH_2-MLRFVTKRWGRPVGRRRRP-COOH; “MLGP” peptide (NatE-type): NH_2-MLGPEGGRWGRPVGRRRRP-COOH; “SGRG”/H4 peptide (NatD-type): NH_2-

SGRGKGGKG LGKGGAKRHR-COOH). All peptides were purchased from GenScript. To evaluate *Sp*NatC activity against these peptides, 50 nM *Sp*NatC was mixed with 50 μM Ac-CoA and 500 μM peptide, and allowed to react for 10 min. To determine steady-state catalytic parameters of *Sp*NatC (WT or mutants) with respect to the peptide substrate, 50 nM *Sp*NatC (WT or mutants) was mixed with 50 μM Ac-CoA (^{14}C-labeled, 4 mCi mmol^{-1}; PerkinElmer Life Sciences) and varying peptide concentrations (ranging from 1.95 μM to 500 μM, 9-data points) for 5-minute reactions. P81 paper was purchased from SVI (St. Vincent's Institute Medical Research). All radioactive count values were converted to molar units with a standard curve created with known concentrations of radioactive Ac-CoA added to scintillation fluid. GraphPad Prism (version 5.01) was used for all data fitting to the Michaelis–Menten equation. The errors in **Table 2** represent the standard deviation of the best fit values of the curves.

For the comparison of ternary and binary *Sc*NatC complex activity, 10 nM of either binary *Sc*NatC or ternary *Sc*NatC was mixed with 100 μM of ^{14}C-labeled Ac-CoA and "MLRF" peptide, for a 12-minute reaction in buffer containing 50 mM HEPES, pH 7.5, 150 mM NaCl and 1 mM DTT.

For the activity comparison of ternary *Sp*NatC and other NATs in the presence or absence of IP_6, 50 nM of *Sp*NatC, *Sp*NatA or hNatB was mixed with 50 μM of ^{14}C-labeled Ac-CoA, 50 μM of their peptide substrate, in reaction buffer containing 75 mM HEPES, pH 7.0, 120 mM NaCl, 1 mM DTT, supplemented with either 0 or 2 μM IP_6 (Sigma-Aldrich), for a 5-minute reaction. The peptide substates for *Sp*NatA and hNatB are SASE peptide and MDVF peptide, respectively. *Sp*NatA (Liszczak et al. 2013) and hNatB (Deng, Pan, et al. 2020) are purified as previously described.

5.4.4 ITC measurements

ITC measurements were carried out using a MicroCal iTC200 at 20 °C. Samples were dialyzed into buffer containing 25 mM HEPES pH 7.0, 200 mM NaCl, and 1 mM DTT. 15 μM of SpNatC in the cell and 300 μM of IP_6 in the syringe were used. The raw data were analyzed with the MicroCal ITC analysis software.

5.4.5 Differential scanning fluorimetry assays

Sypro Orange (50,000X stock, ThermoFisher Scientific) was diluted 1:200, and 4 μL was mixed with 16 μL solution with 0.1 mg/ml of various NatC samples in pH 7.0, 200 mM NaCl, and 1 mM DTT, with or without 10 μM IP_6. Fluorescent readings were recorded using a qPCR (ABI 7900 RealTime PCR) with a 1% ramp rate, while heated from 20 °C to 95 °C. Melting curves were generated from these readings and melting temperatures were determined by taking the first derivative of the curves. DSF scans of all samples were performed in triplicate as technical replicates. Error bars in the figure indicates the Standard Deviation (SD) of each sample.

5.4.6 Cryo-EM data collection

For initial sample screening, 0.05 mg/ml fresh SpNatC sample was prepared with 3-molar excess of both bisubstrate and IP_6. SpNatC particles on cryo grids exhibited a severe preferred orientation, which generated an incorrect initial 3D model (data not

shown). Further screening by addition of detergent NP-40 in sample did not improve this situation. To solve this issue, 1 µL of 80 mM CHAPSO was mixed with 20 µL of 12 mg/mL SpNatC. 3 µL of this mixed sample was applied to Quantifoil R1.2/1.3 holey carbon support grids, blotted and plunged into liquid ethane, using an FEI Vitrobot Mark IV. An FEI TF20 was used for screening the grids and data collection was performed with a Titan Krios equipped with a K3 Summit direct detector (Gatan), at a magnification of 105,000 ×, with defocus values from −0.1 to −3.0 µm. Each stack was exposed and counted in super-resolution mode with a total dose of 42 e-/Å^2, resulting in 35 frames per stack. Image stacks were automatically collected with EPU. A full description of the Cryo-EM data collection parameters can be found in **Table 5.1**.

5.4.7 Cryo-EM data processing

Original image stacks were summed and corrected for drift and beam-induced motion at the micrograph level using MotionCor2 (Zheng et al. 2017), and binned twofold, resulting in a pixel size of 0.84 A˚ /pixel. After motion correction in Relion 3.0 (Kimanius et al. 2016), corrected micrographs were imported into CryoSPARC (Punjani et al. 2017) to perform defocus estimation, the resolution range of each micrograph with Gctf (Zhang 2016) and all the subsequent data analysis. 2D classifications were performed on the particles auto-picked by “blob picker” with particle diameter of 100 - 200 Å. The representative 2D classes were used as templates to further auto-pick 1,860,276 particles from 5514 micrographs. After bad particles were removed by two runs of 2D classification, 849,844 particles were used to generate four Ab-Initio models and two rounds of heterogeneous refinement were performed using the four models.

508, 298 particles were used for auto refinement and per particle CTF refinement. The final map was refined to an overall resolution of 3.20 Å, with local resolution estimated in Cryo-SPARC (Punjani et al. 2017).

5.4.8 Cryo-EM model building and refinement

The *Sp*NatC atomic model was manually built de novo using the program COOT (Emsley and Cowtan 2004) according to the Cryo-EM map, with the guidance of predicted secondary structure and bulky residues such as Phe, Tyr, Trp and Arg. The complete model was then refined by real-space refinement in PHENIX (Adams et al. 2010). All representations of Cryo-EM density and structural models were performed with Chimera (Pettersen et al. 2004) and PyMol (Schrodinger 2015) (https://pymol.org/2/). The sequence alignments with secondary structure display were created by ESPript 3.0 (Robert and Gouet 2014a). The surface area calculation was performed using PDBePISA (Krissinel and Henrick 2007) (Proteins, Interfaces, Structures and Assemblies) (http://www.ebi.ac.uk/pdbe/pisa/).

CHAPTER 6: Conclusions and Future Directions

While Nt-acetylation is a protein modification present in all kingdoms of life, it's much more abundant and complicated in eukaryotes than in bacterial and archaea, in term of target substrate profile. Meanwhile, the biological relevance of Nt-acetylation is also comparably more significant in eukaryotes, impacting diverse cellular processes and diseases. In correspondence to these, eukaryotic NATs have evolved some distinct features. Among these features, the major one is that some eukaryotic NATs, like NatA, B, C, and E collaborate with at least one auxiliary subunit, a small molecule (IP_6, for NatA, NatC, and NatE), and in the case of NatA and NatE, the HYPK regulatory protein to form multi-subunit enzyme complexes and perform the acetylation co-translationally in a controlled manner. Overall, these enzymes are collectively responsible for Nt-acetylation of over 80% human proteins and abnormal activities of them have been reported to be corelated with developmental diseases such as Ogden syndrome, neurodegenerative diseases, and cancers (Dorfel and Lyon 2015; Myklebust, Stove, and Arnesen 2015; Kalvik and Arnesen 2013). Motivated by the biological significance and disease connections of these protein complexes, this thesis set out to provide comprehensive studies towards understanding the molecular and regulatory mechanism of these multi-subunit NATs, as described in previous chapters.

While important insights are provided here, some questions remain unanswered. How do NatA and NatE toggle between co-translational activity and post-translational activities? NAA10 has been reported to act alone towards acidic substrates (distinct from NatA-type), while human NAA50 *per se* is also active towards NatE-type substrates. Moreover, controversial reports indicate that both NAA10 and NAA50 can perform lysine acetylation and auto-acetylation (Magin, March, and Marmorstein 2016; Seo et al. 2010;

Evjenth et al. 2009). Thus, the biological relevance of these potential versatile activities of NAA10 and NAA50 requires further investigation.

Can NATs carry out cross regulation amongst themselves? Nt-acetylation is known to be important for protein complex formation and protein function in some cases. NATs themselves are also subjected to Nt-acetylation depending on their N-terminal sequence identity. A recent report highlighted that NAA20 maturation mediated by NatA activity is essential for NatB complex formation (Lasa et al. 2020).

What is the molecular mechanism of co-translational acetylation on the ribosome of NatA/E, and other NATs that function co-translationally such as NatC/B/D? How do NATs with co-translational activities navigate the same or different ribosomes? How do ribosomes distinguish and selects the correct NAT for co-translational nascent chain Nt-acetylation? How do NATs cooperate or compete with other N-terminal modification enzymes like NME? These are challenging but important questions that remain to be answered. For the known yeast NatE-ribosome architecture, *Sc*NAA50 is inactive and remote from the ribosome exit tunnel. However, in humans, hNAA50 is active, and how hNAA50 shifts its distance relative to the exit tunnel when the nascent chain contains a NatE-type substrate is also an interesting question that remains to be answered. Recently, Electropositive regions (EPRs) on the surface of NatB and NatC were proposed to participate in ribosome association (Magin et al. 2017; Grunwald et al. 2020), structural studies are required to confirm these results

As HYPK reduces ~ 90% of hNatA activity and influences NatE activity as well in vitro (Gottlieb and Marmorstein 2018; Deng, McTiernan, et al. 2020), why are such strong inhibition effects on hNatA not observed in vivo? What is the biological function of

the presence of HYPK in humans, as HYPK is absent in most yeast? Do other NATs agonists or antagonists exist? Learning from HYPK, can we develop novel and specific inhibitors for NATs? In humans, NATs are frequently correlated with disease and cancers, especially hNatA. The development of bi-substrate analogues has been important for understanding NAT catalytic mechanism but are unsuitable as drugs. It will be of great interest to develop drug-like NAT-specific inhibitors.

Lastly, to date, the structure and mechanism of the chloroplast localized NATs are unknown. What are the unique features of these NATs? These, among many other remaining questions, regarding NAT mechanism will keep the field vibrant for decades to come.

REFERENCES

Abboud, A., P. Bedoucha, J. Byska, T. Arnesen, and N. Reuter. 2020. 'Dynamics-function relationship in the catalytic domains of N-terminal acetyltransferases', *Comput Struct Biotechnol J*, 18: 532-47.

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. 2010. 'PHENIX: a comprehensive Python-based system for macromolecular structure solution', *Acta Crystallogr D Biol Crystallogr*, 66: 213-21.

Aivaliotis, M., K. Gevaert, M. Falb, A. Tebbe, K. Konstantinidis, B. Bisle, C. Klein, L. Martens, A. Staes, E. Timmerman, J. Van Damme, F. Siedler, F. Pfeiffer, J. Vandekerckhove, and D. Oesterhelt. 2007. 'Large-scale identification of N-terminal peptides in the halophilic archaea Halobacterium salinarum and Natronomonas pharaonis', *J Proteome Res*, 6: 2195-204.

Aksnes, H., A. Drazic, M. Marie, and T. Arnesen. 2016. 'First Things First: Vital Protein Marks by N-Terminal Acetyltransferases', *Trends Biochem Sci*, 41: 746-60.

Aksnes, H., M. Goris, O. Stromland, A. Drazic, Q. Waheed, N. Reuter, and T. Arnesen. 2017. 'Molecular determinants of the N-terminal acetyltransferase Naa60 anchoring to the Golgi membrane', *J Biol Chem*, 292: 6821-37.

Aksnes, H., R. Ree, and T. Arnesen. 2019. 'Co-translational, Post-translational, and Non-catalytic Roles of N-Terminal Acetyltransferases', *Mol Cell*, 73: 1097-114.

Aksnes, H., P. Van Damme, M. Goris, K. K. Starheim, M. Marie, S. I. Stove, C. Hoel, T. V. Kalvik, K. Hole, N. Glomnes, C. Furnes, S. Ljostveit, M. Ziegler, M. Niere, K. Gevaert, and T. Arnesen. 2015. 'An organellar nalpha-acetyltransferase, naa60, acetylates cytosolic N termini of transmembrane proteins and maintains Golgi integrity', *Cell Rep*, 10: 1362-74.

Ametzazurra, A., C. Gazquez, M. Lasa, E. Larrea, J. Prieto, and R. Aldabe. 2009. 'Characterization of the human Nalpha-terminal acetyltransferase B enzymatic complex', *BMC Proc*, 3 Suppl 6: S4.

Ametzazurra, A., E. Larrea, M. P. Civeira, J. Prieto, and R. Aldabe. 2008. 'Implication of human N-alpha-acetyltransferase 5 in cellular proliferation and carcinogenesis', *Oncogene*, 27: 7296-306.

Anderson, J. P., D. E. Walker, J. M. Goldstein, R. de Laat, K. Banducci, R. J. Caccavello, R. Barbour, J. Huang, K. Kling, M. Lee, L. Diep, P. S. Keim, X. Shen, T. Chataway, M. G. Schlossmacher, P. Seubert, D. Schenk, S. Sinha, W. P. Gai, and T. J. Chilcote. 2006. 'Phosphorylation of Ser-129 is the dominant pathological modification of alpha-synuclein in familial and sporadic Lewy body disease', *J Biol Chem*, 281: 29739-52.

Arnesen, T., D. Anderson, C. Baldersheim, M. Lanotte, J. E. Varhaug, and J. R. Lillehaug. 2005. 'Identification and characterization of the human ARD1-NATH protein acetyltransferase complex', *Biochem J*, 386: 433-43.

Arnesen, T., D. Anderson, J. Torsvik, H. B. Halseth, J. E. Varhaug, and J. R. Lillehaug. 2006. 'Cloning and characterization of hNAT5/hSAN: an evolutionarily conserved component of the NatA protein N-alpha-acetyltransferase complex', *Gene*, 371: 291-5.

Arnesen, T., D. Gromyko, F. Pendino, A. Ryningen, J. E. Varhaug, and J. R. Lillehaug. 2006. 'Induction of apoptosis in human cells by RNAi-mediated knockdown of hARD1 and NATH, components of the protein N-alpha-acetyltransferase complex', *Oncogene*, 25: 4350-60.

Arnesen, T., R. Marmorstein, and R. Dominguez. 2018. 'Actin's N-terminal acetyltransferase uncovered', *Cytoskeleton (Hoboken)*, 75: 318-22.

Arnesen, T., K. K. Starheim, P. Van Damme, R. Evjenth, H. Dinh, M. J. Betts, A. Ryningen, J. Vandekerckhove, K. Gevaert, and D. Anderson. 2010. 'The chaperone-like protein HYPK acts together with NatA in cotranslational N-terminal acetylation and prevention of Huntingtin aggregation', *Mol Cell Biol*, 30: 1898-909.

Arnesen, T., P. Van Damme, B. Polevoda, K. Helsens, R. Evjenth, N. Colaert, J. E. Varhaug, J. Vandekerckhove, J. R. Lillehaug, F. Sherman, and K. Gevaert. 2009. 'Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans', *Proc Natl Acad Sci U S A*, 106: 8157-62.

Behnia, Rudy, Bojana Panic, James R. C. Whyte, and Sean Munro. 2004. 'Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p', *Nature Cell Biology*, 6: 405.

Bernal-Perez, L. F., F. Sahyouni, L. Prokai, and Y. Ryu. 2012. 'RimJ-mediated context-dependent N-terminal acetylation of the recombinant Z-domain protein in Escherichia coli', *Mol Biosyst*, 8: 1128-30.

Bienvenut, W. V., C. Giglione, and T. Meinnel. 2015. 'Proteome-wide analysis of the amino terminal status of Escherichia coli proteins at the steady-state and upon deformylation inhibition', *Proteomics*, 15: 2503-18.

Bienvenut, W. V., D. Sumpton, A. Martinez, S. Lilla, C. Espagne, T. Meinnel, and C. Giglione. 2012. 'Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-alpha-acetylation features', *Molecular & cellular proteomics : MCP*, 11: M111 015131.

Bocharov, E. V., A. G. Sobol, K. V. Pavlov, D. M. Korzhnev, V. A. Jaravine, A. T. Gudkov, and A. S. Arseniev. 2004. 'From structure and dynamics of protein L7/L12 to molecular switching in ribosome', *J Biol Chem*, 279: 17697-706.

Braman, J., C. Papworth, and A. Greener. 1996. 'Site-directed mutagenesis using double-stranded plasmid DNA templates', *Methods Mol Biol*, 57: 31-44.

Caesar, R., J. Warringer, and A. Blomberg. 2006. 'Physiological importance and identification of novel targets for the N-terminal acetyltransferase NatB', *Eukaryot Cell*, 5: 368-78.

Caro, J. A., K. W. Harpole, V. Kasinath, J. Lim, J. Granja, K. G. Valentine, K. A. Sharp, and A. J. Wand. 2017. 'Entropy in molecular recognition by proteins', *Proc Natl Acad Sci U S A*, 114: 6563-68.

Casey, J. P., S. I. Stove, C. McGorrian, J. Galvin, M. Blenski, A. Dunne, S. Ennis, F. Brett, M. D. King, T. Arnesen, and S. A. Lynch. 2015. 'NAA10 mutation causing a novel intellectual disability syndrome with Long QT due to N-terminal acetyltransferase impairment', *Sci Rep*, 5: 16022.

Chang, Y. Y., and C. H. Hsu. 2015. 'Structural basis for substrate-specific acetylation of Nalpha-acetyltransferase Ard1 from Sulfolobus solfataricus', *Sci Rep*, 5: 8673.

Chen, J. Y., L. Liu, C. L. Cao, M. J. Li, K. Tan, X. Yang, and C. H. Yun. 2016. 'Structure and function of human Naa60 (NatF), a Golgi-localized bi-functional acetyltransferase', *Sci Rep*, 6: 31425.

Cheng, H., L. Gottlieb, E. Marchi, R. Kleyner, P. Bhardwaj, A. F. Rope, S. Rosenheck, S. Moutton, C. Philippe, W. Eyaid, F. S. Alkuraya, J. Toribio, R. Mena, C. E. Prada, H. Stessman, R. Bernier, M. Wermuth, B. Kauffmann, B. Blaumeiser, R. F. Kooy, D. Baralle, G. M. S. Mancini, S. J. Conway, F. Xia, Z. Chen, L. Meng, L. Mihajlovic, R. Marmorstein, and G. J. Lyon. 2019. 'Phenotypic and biochemical analysis of an international cohort of individuals with variants in NAA10 and NAA15', *Hum Mol Genet*, 28: 2900-19.

Chu, C. W., F. Hou, J. Zhang, L. Phu, A. V. Loktev, D. S. Kirkpatrick, P. K. Jackson, Y. Zhao, and H. Zou. 2011. 'A novel acetylation of beta-tubulin by San modulates microtubule polymerization via down-regulating tubulin incorporation', *Mol Biol Cell*, 22: 448-56.

Cox, J., M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj, and M. Mann. 2014. 'Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ', *Molecular & cellular proteomics : MCP*, 13: 2513-26.

Deng, S., R. S. Magin, X. Wei, B. Pan, E. J. Petersson, and R. Marmorstein. 2019. 'Structure and Mechanism of Acetylation by the N-Terminal Dual Enzyme NatA/Naa50 Complex', *Structure*, 27: 1057-70 e4.

Deng, S., and R. Marmorstein. 2021. 'Protein N-Terminal Acetylation: Structural Basis, Mechanism, Versatility, and Regulation', *Trends Biochem Sci*, 46: 15-27.

Deng, S., N. McTiernan, X. Wei, T. Arnesen, and R. Marmorstein. 2020. 'Molecular basis for N-terminal acetylation by human NatE and its modulation by HYPK', *Nat Commun*, 11: 818.

Deng, S., B. Pan, L. Gottlieb, E. J. Petersson, and R. Marmorstein. 2020. 'Molecular basis for N-terminal alpha-synuclein acetylation by human NatB', *Elife*, 9: e57491.

Deng, Sunbin, Leah Gottlieb, Buyan Pan, Julianna Supplee, Xuepeng Wei, James Petersson, and Ronen Marmorstein. 2021. 'Molecular mechanism of N-terminal acetylation by the ternary NatC complex', *bioRxiv*: 2021.02.01.429250.

Dikiy, I., and D. Eliezer. 2014. 'N-terminal acetylation stabilizes N-terminal helicity in lipid- and micelle-bound alpha-synuclein and increases its affinity for physiological membranes', *J Biol Chem*, 289: 3652-65.

Dinh, T. V., W. V. Bienvenut, E. Linster, A. Feldman-Salit, V. A. Jung, T. Meinnel, R. Hell, C. Giglione, and M. Wirtz. 2015. 'Molecular identification and functional characterization of the first Nalpha-acetyltransferase in plastids by global acetylome profiling', *Proteomics*, 15: 2426-35.

Dorfel, M. J., and G. J. Lyon. 2015. 'The biological functions of Naa10 - From amino-terminal acetylation to human disease', *Gene*, 567: 103-31.

Drazic, A., H. Aksnes, M. Marie, M. Boczkowska, S. Varland, E. Timmerman, H. Foyn, N. Glomnes, G. Rebowski, F. Impens, K. Gevaert, R. Dominguez, and T. Arnesen. 2018. 'NAA80 is actin's N-terminal acetyltransferase and regulates cytoskeleton assembly and cell motility', *Proc Natl Acad Sci U S A*, 115: 4399-404.

Drazic, A., and T. Arnesen. 2017. '[(14)C]-Acetyl-Coenzyme A-Based In Vitro N-Terminal Acetylation Assay', *Methods Mol Biol*, 1574: 1-8.

Eldeeb, M. A., and M. A. Ragheb. 2018. 'Post-translational N-terminal Arginylation of Protein Fragments: A Pivotal Portal to Proteolysis', *Curr Protein Pept Sci*, 19: 1214-23.

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

Esmailpour, T., H. Riazifar, L. Liu, S. Donkervoort, V. H. Huang, S. Madaan, B. M. Shoucri, A. Busch, J. Wu, A. Towbin, R. B. Chadwick, A. Sequeira, M. P. Vawter, G. Sun, J. J. Johnston, L. G. Biesecker, R. Kawaguchi, H. Sun, V. Kimonis, and T. Huang. 2014. 'A splice donor mutation in NAA10 results in the dysregulation of the retinoic acid signalling pathway and causes Lenz microphthalmia syndrome', *J Med Genet*, 51: 185-96.

Evjenth, R. H., A. K. Brenner, P. R. Thompson, T. Arnesen, N. A. Froystein, and J. R. Lillehaug. 2012. 'Human protein N-terminal acetyltransferase hNaa50p (hNAT5/hSAN) follows ordered sequential catalytic mechanism: combined kinetic and NMR study', *J Biol Chem*, 287: 10081-8.

Evjenth, R., K. Hole, O. A. Karlsen, M. Ziegler, T. Arnesen, and J. R. Lillehaug. 2009. 'Human Naa50p (Nat5/San) displays both protein N alpha- and N epsilon-acetyltransferase activity', *J Biol Chem*, 284: 31122-9.

Falb, M., M. Aivaliotis, C. Garcia-Rizo, B. Bisle, A. Tebbe, C. Klein, K. Konstantinidis, F. Siedler, F. Pfeiffer, and D. Oesterhelt. 2006. 'Archaeal N-terminal

protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey', *J Mol Biol*, 362: 915-24.

Fauvet, B., M. B. Fares, F. Samuel, I. Dikiy, A. Tandon, D. Eliezer, and H. A. Lashuel. 2012. 'Characterization of semisynthetic and naturally Nalpha-acetylated alpha-synuclein in vitro and in intact cells: implications for aggregation and cellular properties of alpha-synuclein', *J Biol Chem*, 287: 28243-62.

Favrot, L., J. S. Blanchard, and O. Vergnolle. 2016. 'Bacterial GCN5-Related N-Acetyltransferases: From Resistance to Regulation', *Biochemistry*, 55: 989-1002.

Fernández, Ricardo D., and Heather R. Lucas. 2018. 'Mass spectrometry data confirming tetrameric α-synuclein N-terminal acetylation', *Data in brief*, 20: 1686-91.

Ferrandez-Ayela, A., R. Micol-Ponce, A. B. Sanchez-Garcia, M. M. Alonso-Peral, J. L. Micol, and M. R. Ponce. 2013. 'Mutation of an Arabidopsis NatB N-alpha-terminal acetylation complex component causes pleiotropic developmental defects', *PLOS ONE*, 8: e80697.

Fluge, O., O. Bruland, L. A. Akslen, J. E. Varhaug, and J. R. Lillehaug. 2002. 'NATH, a novel gene overexpressed in papillary thyroid carcinomas', *Oncogene*, 21: 5056-68.

Foyn, H., J. E. Jones, D. Lewallen, R. Narawane, J. E. Varhaug, P. R. Thompson, and T. Arnesen. 2013. 'Design, synthesis, and kinetic characterization of protein N-terminal acetyltransferase inhibitors', *ACS Chem Biol*, 8: 1121-7.

Gautschi, M., S. Just, A. Mun, S. Ross, P. Rucknagel, Y. Dubaquie, A. Ehrenhofer-Murray, and S. Rospert. 2003. 'The yeast N (alpha)-acetyltransferase NatA is quantitatively anchored to the ribosome and interacts with nascent polypeptides', *Mol Cell Biol*, 23: 7403-14.

Giglione, C., S. Fieulaine, and T. Meinnel. 2015. 'N-terminal protein modifications: Bringing back into play the ribosome', *Biochimie*, 114: 134-46.

Goris, M., R. S. Magin, H. Foyn, L. M. Myklebust, S. Varland, R. Ree, A. Drazic, P. Bhambra, S. I. Stove, M. Baumann, B. E. Haug, R. Marmorstein, and T. Arnesen. 2018. 'Structural determinants and cellular environment define processed actin as the sole substrate of the N-terminal acetyltransferase NAA80', *Proc Natl Acad Sci U S A*, 115: 4405-10.

Gottlieb, L., and R. Marmorstein. 2018. 'Structure of Human NatA and Its Regulation by the Huntingtin Interacting Protein HYPK', *Structure*, 26: 925-35 e8.

Grunwald, S., L. V. M. Hopf, T. Bock-Bierbaum, C. C. M. Lally, C. M. T. Spahn, and O. Daumke. 2020. 'Divergent architecture of the heterotrimeric NatC complex explains N-terminal acetylation of cognate substrates', *Nat Commun*, 11: 5506.

Halliday, Glenda M., Janice L. Holton, Tamas Revesz, and Dennis W. Dickson. 2011. 'Neuropathology underlying clinical variability in patients with synucleinopathies', *Acta Neuropathologica*, 122: 187-204.

Hole, K., P. Van Damme, M. Dalva, H. Aksnes, N. Glomnes, J. E. Varhaug, J. R. Lillehaug, K. Gevaert, and T. Arnesen. 2011. 'The human N-alpha-acetyltransferase 40 (hNaa40p/hNatD) is conserved from yeast and N-terminally acetylates histones H2A and H4', *PLOS ONE*, 6: e24713.

Holmes, W. M., B. K. Mannakee, R. N. Gutenkunst, and T. R. Serio. 2014. 'Loss of amino-terminal acetylation suppresses a prion phenotype by modulating global protein folding', *Nat Commun*, 5: 4383.

Hong, H., Y. Cai, S. Zhang, H. Ding, H. Wang, and A. Han. 2017. 'Molecular Basis of Substrate Specific Acetylation by N-Terminal Acetyltransferase NatB', *Structure*, 25: 641-49 e3.

Hou, F., C. W. Chu, X. Kong, K. Yokomori, and H. Zou. 2007. 'The acetyltransferase activity of San stabilizes the mitotic cohesin at the centromeres in a shugoshin-independent manner', *J Cell Biol*, 177: 587-97.

Hua, K. T., C. T. Tan, G. Johansson, J. M. Lee, P. W. Yang, H. Y. Lu, C. K. Chen, J. L. Su, P. B. Chen, Y. L. Wu, C. C. Chi, H. J. Kao, H. J. Shih, M. W. Chen, M. H. Chien, P. S. Chen, W. J. Lee, T. Y. Cheng, G. Rosenberger, C. Y. Chai, C. J. Yang, M. S. Huang, T. C. Lai, T. Y. Chou, M. Hsiao, and M. L. Kuo. 2011. 'N-alpha-acetyltransferase 10 protein suppresses cancer cell metastasis by binding PIX proteins and inhibiting Cdc42/Rac1 activity', *Cancer Cell*, 19: 218-31.

Huber, M., W. V. Bienvenut, E. Linster, I. Stephan, L. Armbruster, C. Sticht, D. Layer, K. Lapouge, T. Meinnel, I. Sinning, C. Giglione, R. Hell, and M. Wirtz. 2020. 'NatB-Mediated N-Terminal Acetylation Affects Growth and Biotic Stress Responses', *Plant Physiol*, 182: 792-806.

Hwang, Cheol-Sang, Anna Shemorry, and Alexander Varshavsky. 2010. 'N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals', *Science*, 327: 973-77.

Iyer, A., S. J. Roeters, N. Schilderink, B. Hommersom, R. M. Heeren, S. Woutersen, M. M. Claessens, and V. Subramaniam. 2016. 'The Impact of N-terminal Acetylation of alpha-Synuclein on Phospholipid Membrane Binding and Fibril Structure', *J Biol Chem*, 291: 21110-22.

Jeong, J. W., M. K. Bae, M. Y. Ahn, S. H. Kim, T. K. Sohn, M. H. Bae, M. A. Yoo, E. J. Song, K. J. Lee, and K. W. Kim. 2002. 'Regulation and destabilization of HIF-1alpha by ARD1-mediated acetylation', *Cell*, 111: 709-20.

Kalam Shamsuddin, Abul, and Sanchita Bose. 2012. 'IP6 (Inositol Hexaphosphate) as a Signaling Molecule', *Current Signal Transduction Therapy*, 7: 289-304.

Kalvik, T. V., and T. Arnesen. 2013. 'Protein N-terminal acetyltransferases in cancer', *Oncogene*, 32: 269-76.

Kang, L., G. M. Moriarty, L. A. Woods, A. E. Ashcroft, S. E. Radford, and J. Baum. 2012. 'N-terminal acetylation of alpha-synuclein induces increased transient helical propensity and decreased aggregation rates in the intrinsically disordered monomer', *Protein Sci*, 21: 911-7.

Karpenahalli, Manjunatha R., Andrei N. Lupas, and Johannes Söding. 2007. 'TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences', *BMC Bioinformatics*, 8: 2.

Kelkar, D. S., D. Kumar, P. Kumar, L. Balakrishnan, B. Muthusamy, A. K. Yadav, P. Shrivastava, A. Marimuthu, S. Anand, H. Sundaram, R. Kingsbury, H. C. Harsha, B. Nair, T. S. Prasad, D. S. Chauhan, K. Katoch, V. M. Katoch, P. Kumar, R. Chaerkady, S. Ramachandran, D. Dash, and A. Pandey. 2011. 'Proteogenomic analysis of Mycobacterium tuberculosis by high resolution mass spectrometry', *Molecular & cellular proteomics : MCP*, 10: M111 011627.

Kimanius, D., B. O. Forsberg, S. H. Scheres, and E. Lindahl. 2016. 'Accelerated Cryo-EM structure determination with parallelisation using GPUs in RELION-2', *Elife*, 5.

Kirkland, P. A., M. A. Humbard, C. J. Daniels, and J. A. Maupin-Furlow. 2008. 'Shotgun proteomics of the haloarchaeon Haloferax volcanii', *J Proteome Res*, 7: 5033-9.

Knorr, A. G., C. Schmidt, P. Tesina, O. Berninghausen, T. Becker, B. Beatrix, and R. Beckmann. 2019. 'Ribosome-NatA architecture reveals that rRNA expansion segments coordinate N-terminal acetylation', *Nat Struct Mol Biol*, 26: 35-39.

Krissinel, E., and K. Henrick. 2007. 'Inference of macromolecular assemblies from crystalline state', *J Mol Biol*, 372: 774-97.

Kucukelbir, A., F. J. Sigworth, and H. D. Tagare. 2014. 'Quantifying the local resolution of Cryo-EM density maps', *Nat Methods*, 11: 63-5.

Lasa, M., L. Neri, B. Carte, C. Gázquez, T. Aragón, and R. Aldabe. 2020. 'Maturation of NAA20 Aminoterminal End Is Essential to Assemble NatB N-Terminal Acetyltransferase Complex', *J Mol Biol*, 432: 5889-901.

Laskowski, Roman A., and Mark B. Swindells. 2011. 'LigPlot+: multiple ligand-protein interaction diagrams for drug discovery', *Journal of chemical information and modeling*, 51: 2778-86.

Lee, C. F., D. S. Ou, S. B. Lee, L. H. Chang, R. K. Lin, Y. S. Li, A. K. Upadhyay, X. Cheng, Y. C. Wang, H. S. Hsu, M. Hsiao, C. W. Wu, and L. J. Juan. 2010. 'hNaa10p contributes to tumorigenesis by facilitating DNMT1-mediated tumor suppressor gene silencing', *J Clin Invest*, 120: 2920-30.

Lee, K. E., J. Y. Ahn, J. M. Kim, and C. S. Hwang. 2014. 'Synthetic lethal screen of NAA20, a catalytic subunit gene of NatB N-terminal acetylase in Saccharomyces cerevisiae', *J Microbiol*, 52: 842-8.

Linster, E., D. Layer, W. V. Bienvenut, T. V. Dinh, F. A. Weyer, W. Leemhuis, A. Brunje, M. Hoffrichter, P. Miklankova, J. Kopp, K. Lapouge, J. Sindlinger, D. Schwarzer, T. Meinnel, I. Finkemeier, C. Giglione, R. Hell, I. Sinning, and M. Wirtz. 2020. 'The Arabidopsis N (alpha) -acetyltransferase NAA60 locates to the plasma membrane and is vital for the high salt stress response', *New Phytol*, 228: 554-69.

Linster, E., and M. Wirtz. 2018. 'N-terminal acetylation: an essential protein modification emerges as an important regulator of stress responses', *J Exp Bot*, 69: 4555-68.

Liszczak, G., T. Arnesen, and R. Marmorstein. 2011. 'Structure of a ternary Naa50p (NAT5/SAN) N-terminal acetyltransferase complex reveals the molecular basis for substrate-specific acetylation', *J Biol Chem*, 286: 37002-10.

Liszczak, G., J. M. Goldberg, H. Foyn, E. J. Petersson, T. Arnesen, and R. Marmorstein. 2013. 'Molecular basis for N-terminal acetylation by the heterodimeric NatA complex', *Nat Struct Mol Biol*, 20: 1098-105.

Liszczak, G., and R. Marmorstein. 2013. 'Implications for the evolution of eukaryotic amino-terminal acetyltransferase (NAT) enzymes from the structure of an archaeal ortholog', *Proc Natl Acad Sci U S A*, 110: 14652-7.

Ma, C., C. Pathak, S. Jang, S. J. Lee, M. Nam, S. J. Kim, H. Im, and B. J. Lee. 2014. 'Structure of Thermoplasma volcanium Ard1 belongs to N-acetyltransferase family member suggesting multiple ligand binding modes with acetyl coenzyme A and coenzyme A', *Biochim Biophys Acta*, 1844: 1790-7.

Mackay, D. T., C. H. Botting, G. L. Taylor, and M. F. White. 2007. 'An acetylase with relaxed specificity catalyses protein N-terminal acetylation in Sulfolobus solfataricus', *Mol Microbiol*, 64: 1540-8.

Magin, R. S., S. Deng, H. Zhang, B. Cooperman, and R. Marmorstein. 2017. 'Probing the interaction between NatA and the ribosome for co-translational protein acetylation', *PLOS ONE*, 12: e0186278.

Magin, R. S., G. P. Liszczak, and R. Marmorstein. 2015. 'The molecular basis for histone H4- and H2A-specific amino-terminal acetylation by NatD', *Structure*, 23: 332-41.

Magin, R. S., Z. M. March, and R. Marmorstein. 2016. 'The N-terminal Acetyltransferase Naa10/ARD1 Does Not Acetylate Lysine Residues', *J Biol Chem*, 291: 5270-7.

Maltsev, A. S., J. Ying, and A. Bax. 2012. 'Impact of N-terminal acetylation of alpha-synuclein on its random coil and lipid binding properties', *Biochemistry*, 51: 5004-13.

Mason, R. J., A. R. Paskins, C. F. Dalton, and D. P. Smith. 2016. 'Copper Binding and Subsequent Aggregation of alpha-Synuclein Are Modulated by N-Terminal Acetylation and Ablated by the H50Q Missense Mutation', *Biochemistry*, 55: 4737-41.

Mughal, A. A., Z. Grieg, H. Skjellegrind, A. Fayzullin, M. Lamkhannat, M. Joel, M. S. Ahmed, W. Murrell, E. O. Vik-Mo, I. A. Langmoen, and B. Stangeland. 2015. 'Knockdown of NAT12/NAA30 reduces tumorigenic features of glioblastoma-initiating cells', *Mol Cancer*, 14: 160.

Mullen, J. R., P. S. Kayne, R. P. Moerschell, S. Tsunasawa, M. Gribskov, M. Colavito-Shepanski, M. Grunstein, F. Sherman, and R. Sternglanz. 1989. 'Identification and characterization of genes and mutants for an N-terminal acetyltransferase from yeast', *EMBO J*, 8: 2067-75.

Myklebust, L. M., S. I. Stove, and T. Arnesen. 2015. 'Naa10 in development and disease', *Oncotarget*, 6: 34041-2.

Neri, L., M. Lasa, A. Elosegui-Artola, D. D'Avola, B. Carte, C. Gazquez, S. Alve, P. Roca-Cusachs, M. Inarrairaegui, J. Herrero, J. Prieto, B. Sangro, and R. Aldabe. 2017. 'NatB-mediated protein N-alpha-terminal acetylation is a potential therapeutic target in hepatocellular carcinoma', *Oncotarget*, 8: 40967-81.

Neuwald, A. F., and D. Landsman. 1997. 'GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein', *Trends Biochem Sci*, 22: 154-5.

Ochaya, S., O. Franzen, D. A. Buhwa, H. Foyn, C. E. Butler, S. I. Stove, K. M. Tyler, T. Arnesen, E. Matovu, L. Aslund, and B. Andersson. 2019. 'Characterization of Evolutionarily Conserved Trypanosoma cruzi NatC and NatA-N-Terminal Acetyltransferase Complexes', *J Parasitol Res*, 2019: 6594212.

Oh, J. H., J. Y. Hyun, and A. Varshavsky. 2017. 'Control of Hsp90 chaperone and its clients by N-terminal acetylation and the N-end rule pathway', *Proc Natl Acad Sci U S A*, 114: E4370-E79.

Ohrfelt, A., H. Zetterberg, K. Andersson, R. Persson, D. Secic, G. Brinkmalm, A. Wallin, E. Mulugeta, P. T. Francis, E. Vanmechelen, D. Aarsland, C. Ballard, K. Blennow, and A. Westman-Brinkmalm. 2011. 'Identification of novel alpha-synuclein isoforms in human brain tissue by using an online nanoLC-ESI-FTICR-MS method', *Neurochem Res*, 36: 2029-42.

Ohyama, K., K. Yasuda, K. Onga, A. Kakizuka, and N. Mori. 2012. 'Spatio-temporal expression pattern of the NatB complex, Nat5/Mdm20 in the developing mouse brain: implications for co-operative versus non-co-operative actions of Mdm20 and Nat5', *Gene Expr Patterns*, 12: 36-45.

Otwinowski, Zbyszek, and Wladek Minor. 1997. '[20] Processing of X-ray diffraction data collected in oscillation mode.' in, *Macromolecular Crystallography Part A* (Academic Press).

Ouidir, T., F. Jarnier, P. Cosette, T. Jouenne, and J. Hardouin. 2015. 'Characterization of N-terminal protein modifications in Pseudomonas aeruginosa PA14', *J Proteomics*, 114: 214-25.

Park, E. C., and J. W. Szostak. 1992. 'ARD1 and NAT1 proteins form a complex that has N-terminal acetyltransferase activity', *EMBO J*, 11: 2087-93.

Pathak, D., A. H. Bhat, V. Sapehia, J. Rai, and A. Rao. 2016. 'Biochemical evidence for relaxed substrate specificity of Nalpha-acetyltransferase (Rv3420c/rimI) of Mycobacterium tuberculosis', *Sci Rep*, 6: 28892.

Pavlou, Demetria, and Antonis Kirmizis. 2016. 'Depletion of histone N-terminal-acetyltransferase Naa40 induces p53-independent apoptosis in colorectal cancer cells via the mitochondrial pathway', *Apoptosis*, 21: 298-311.

Pesaresi, P., N. A. Gardner, S. Masiero, A. Dietzmann, L. Eichacker, R. Wickner, F. Salamini, and D. Leister. 2003. 'Cytoplasmic N-terminal protein acetylation is required for efficient photosynthesis in Arabidopsis', *Plant Cell*, 15: 1817-32.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. 'UCSF Chimera--a visualization system for exploratory research and analysis', *J Comput Chem*, 25: 1605-12.

Polevoda, B., S. Brown, T. S. Cardillo, S. Rigby, and F. Sherman. 2008. 'Yeast N (alpha)-terminal acetyltransferases are associated with ribosomes', *J Cell Biochem*, 103: 492-508.

Polevoda, B., T. S. Cardillo, T. C. Doyle, G. S. Bedi, and F. Sherman. 2003. 'Nat3p and Mdm20p are required for function of yeast NatB Nalpha-terminal acetyltransferase and of actin and tropomyosin', *J Biol Chem*, 278: 30686-97.

Polevoda, B., J. Norbeck, H. Takakura, A. Blomberg, and F. Sherman. 1999. 'Identification and specificities of N-terminal acetyltransferases from Saccharomyces cerevisiae', *EMBO J*, 18: 6155-68.

Polevoda, B., and F. Sherman. 2001. 'NatC Nalpha-terminal acetyltransferase of yeast contains three subunits, Mak3p, Mak10p, and Mak31p', *J Biol Chem*, 276: 20154-9.

Popp, B., S. I. Stove, S. Endele, L. M. Myklebust, J. Hoyer, H. Sticht, S. Azzarello-Burri, A. Rauch, T. Arnesen, and A. Reis. 2015. 'De novo missense mutations in the NAA10 gene cause severe non-syndromic developmental delay in males and females', *Eur J Hum Genet*, 23: 602-9.

Punjani, Ali, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. 2017. 'cryoSPARC: algorithms for rapid unsupervised Cryo-EM structure determination', *Nat Methods*, 14: 290-96.

Rathore, O. S., A. Faustino, P. Prudencio, P. Van Damme, C. J. Cox, and R. G. Martinho. 2016. 'Absence of N-terminal acetyltransferase diversification during evolution of eukaryotic organisms', *Sci Rep*, 6: 21304.

Raue, U., S. Oellerer, and S. Rospert. 2007. 'Association of protein biogenesis factors at the yeast ribosomal tunnel exit is affected by the translational status and nascent polypeptide sequence', *J Biol Chem*, 282: 7809-16.

Rebowski, G., M. Boczkowska, A. Drazic, R. Ree, M. Goris, T. Arnesen, and R. Dominguez. 2020. 'Mechanism of actin N-terminal acetylation', *Sci Adv*, 6: eaay8793.

Ree, R., A. S. Geithus, P. M. Torring, K. P. Sorensen, M. Damkjaer, D. D. D. study, S. A. Lynch, and T. Arnesen. 2019. 'A novel NAA10 p. (R83H) variant with impaired acetyltransferase activity identified in two boys with ID and microcephaly', *BMC Med Genet*, 20: 101.

Ree, R., S. Varland, and T. Arnesen. 2018. 'Spotlight on protein N-terminal acetylation', *Exp Mol Med*, 50: 1-13.

Ribeiro, A. L., R. D. Silva, H. Foyn, M. N. Tiago, O. S. Rathore, T. Arnesen, and R. G. Martinho. 2016. 'Naa50/San-dependent N-terminal acetylation of Scc1 is potentially important for sister chromatid cohesion', *Sci Rep*, 6: 39118.

Robert, X., and P. Gouet. 2014a. 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Res*, 42: W320-4.

Robert, Xavier, and Patrice Gouet. 2014b. 'Deciphering key features in protein structures with the new ENDscript server', *Nucleic Acids Research*, 42: W320-W24.

Rong, Z., Z. Ouyang, R. S. Magin, R. Marmorstein, and H. Yu. 2016. 'Opposing Functions of the N-terminal Acetyltransferases Naa50 and NatA in Sister-chromatid Cohesion', *J Biol Chem*, 291: 19079-91.

Rope, A. F., K. Wang, R. Evjenth, J. Xing, J. J. Johnston, J. J. Swensen, W. E. Johnson, B. Moore, C. D. Huff, L. M. Bird, J. C. Carey, J. M. Opitz, C. A. Stevens, T. Jiang, C. Schank, H. D. Fain, R. Robison, B. Dalley, S. Chin, S. T. South, T. J. Pysher, L. B. Jorde, H. Hakonarson, J. R. Lillehaug, L. G. Biesecker, M. Yandell, T. Arnesen, and G. J. Lyon. 2011. 'Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency', *Am J Hum Genet*, 89: 28-43.

Runfola, M., A. De Simone, M. Vendruscolo, C. M. Dobson, and G. Fusco. 2020. 'The N-terminal Acetylation of alpha-Synuclein Changes the Affinity for Lipid Membranes but not the Structural Properties of the Bound State', *Sci Rep*, 10: 204.

Salah Ud-Din, A. I., A. Tikhomirova, and A. Roujeinikova. 2016. 'Structure and Functional Diversity of GCN5-Related N-Acetyltransferases (GNAT)', *Int J Mol Sci*, 17.

Saunier, C., S. I. Stove, B. Popp, B. Gerard, M. Blenski, N. AhMew, C. de Bie, P. Goldenberg, B. Isidor, B. Keren, B. Leheup, L. Lampert, C. Mignot, K. Tezcan, G. M. Mancini, C. Nava, M. Wasserstein, A. L. Bruel, J. Thevenon, A. Masurel, Y. Duffourd, P. Kuentz, F. Huet, J. B. Riviere, M. van Slegtenhorst, L. Faivre, A. Piton, A. Reis, T. Arnesen, C. Thauvin-Robinet, and C. Zweier. 2016. 'Expanding the Phenotype Associated with NAA10-Related N-Terminal Acetylation Deficiency', *Hum Mutat*, 37: 755-64.

Schiza, Vassia, Diego Molina-Serrano, Dimitris Kyriakou, Antonia Hadjiantoniou, and Antonis Kirmizis. 2013. 'N-alpha-terminal Acetylation of Histone H4 Regulates Arginine Methylation and Ribosomal DNA Silencing', *Plos Genetics*, 9: e1003805.

Schmidt, A., K. Kochanowski, S. Vedelaar, E. Ahrne, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold, and M. Heinemann. 2016. 'The quantitative and condition-dependent Escherichia coli proteome', *Nat Biotechnol*, 34: 104-10.

Schrodinger, LLC. 2015. "The PyMOL Molecular Graphics System, Version 1.8." In.

Seo, J. H., J. H. Cha, J. H. Park, C. H. Jeong, Z. Y. Park, H. S. Lee, S. H. Oh, J. H. Kang, S. W. Suh, K. H. Kim, J. Y. Ha, S. H. Han, S. H. Kim, J. W. Lee, J. A. Park, J. W. Jeong, K. J. Lee, G. T. Oh, M. N. Lee, S. W. Kwon, S. K. Lee, K. H. Chun, S. J. Lee, and K. W. Kim. 2010. 'Arrest defective 1 autoacetylation is a critical step in its ability to stimulate cancer cell proliferation', *Cancer Res*, 70: 4422-32.

Setty, S. R., T. I. Strochlic, A. H. Tong, C. Boone, and C. G. Burd. 2004a. 'Golgi targeting of ARF-like GTPase Arl3p requires its Nalpha-acetylation and the integral membrane protein Sys1p', *Nat Cell Biol*, 6: 414-9.

Setty, Subba Rao Gangi, Todd I. Strochlic, Amy Hin Yan Tong, Charles Boone, and Christopher G. Burd. 2004b. 'Golgi targeting of ARF-like GTPase Arl3p requires its Nα-acetylation and the integral membrane protein Sys1p', *Nature Cell Biology*, 6: 414.

Shemorry, A., C. S. Hwang, and A. Varshavsky. 2013. 'Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway', *Mol Cell*, 50: 540-51.

Singer, J. M., and J. M. Shaw. 2003. 'Mdm20 protein functions with Nat3 protein to acetylate Tpm1 protein and regulate tropomyosin-actin interactions in budding yeast', *Proc Natl Acad Sci U S A*, 100: 7644-9.

Song, O. K., X. Wang, J. H. Waterborg, and R. Sternglanz. 2003. 'An Nalpha-acetyltransferase responsible for acetylation of the N-terminal residues of histones H4 and H2A', *J Biol Chem*, 278: 38109-12.

Spillantini, M. G., R. A. Crowther, R. Jakes, M. Hasegawa, and M. Goedert. 1998. 'alpha-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with lewy bodies', *Proc Natl Acad Sci U S A*, 95: 6469-73.

Starheim, K. K., T. Arnesen, D. Gromyko, A. Ryningen, J. E. Varhaug, and J. R. Lillehaug. 2008. 'Identification of the human N (alpha)-acetyltransferase complex B (hNatB): a complex important for cell-cycle progression', *Biochem J*, 415: 325-31.

Starheim, K. K., K. Gevaert, and T. Arnesen. 2012. 'Protein N-terminal acetyltransferases: when the start matters', *Trends Biochem Sci*, 37: 152-61.

Starheim, K. K., D. Gromyko, R. Evjenth, A. Ryningen, J. E. Varhaug, J. R. Lillehaug, and T. Arnesen. 2009. 'Knockdown of human N alpha-terminal acetyltransferase complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization', *Mol Cell Biol*, 29: 3569-81.

Stove, S. I., R. S. Magin, H. Foyn, B. E. Haug, R. Marmorstein, and T. Arnesen. 2016. 'Crystal Structure of the Golgi-Associated Human Nalpha-Acetyltransferase 60 Reveals the Molecular Determinants for Substrate-Specific Acetylation', *Structure*, 24: 1044-56.

Strong, M., M. R. Sawaya, S. Wang, M. Phillips, D. Cascio, and D. Eisenberg. 2006. 'Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis', *Proc Natl Acad Sci U S A*, 103: 8060-5.

Tercero, J. C., J. D. Dinman, and R. B. Wickner. 1993. 'Yeast MAK3 N-acetyltransferase recognizes the N-terminal four amino acids of the major coat protein (gag) of the L-A double-stranded RNA virus', *Journal of bacteriology*, 175: 3192-4.

Theillet, F. X., A. Binolfi, B. Bekei, A. Martorana, H. M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb, and P. Selenko. 2016. 'Structural disorder of monomeric alpha-synuclein persists in mammalian cells', *Nature*, 530: 45-50.

Trexler, A. J., and E. Rhoades. 2012. 'N-Terminal acetylation is critical for forming alpha-helical oligomer of alpha-synuclein', *Protein Sci*, 21: 601-5.

Van Damme, P., R. Evjenth, H. Foyn, K. Demeyer, P. J. De Bock, J. R. Lillehaug, J. Vandekerckhove, T. Arnesen, and K. Gevaert. 2011a. 'Proteome-derived peptide libraries allow detailed analysis of the substrate specificities of N (alpha)-acetyltransferases and point to hNaa10p as the post-translational actin N (alpha)-acetyltransferase', *Molecular & cellular proteomics : MCP*, 10: M110 004580.

Van Damme, P., K. Hole, K. Gevaert, and T. Arnesen. 2015. 'N-terminal acetylome analysis reveals the specificity of Naa50 (Nat5) and suggests a kinetic competition between N-terminal acetyltransferases and methionine aminopeptidases', *Proteomics*, 15: 2436-46.

Van Damme, P., K. Hole, A. Pimenta-Marques, K. Helsens, J. Vandekerckhove, R. G. Martinho, K. Gevaert, and T. Arnesen. 2011. 'NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation', *PLoS Genet*, 7: e1002169.

Van Damme, P., T. V. Kalvik, K. K. Starheim, V. Jonckheere, L. M. Myklebust, G. Menschaert, J. E. Varhaug, K. Gevaert, and T. Arnesen. 2016a. 'A Role for Human N-alpha Acetyltransferase 30 (Naa30) in Maintaining Mitochondrial Integrity', *Molecular & cellular proteomics : MCP*, 15: 3361-72.

Van Damme, P., M. Lasa, B. Polevoda, C. Gazquez, A. Elosegui-Artola, D. S. Kim, E. De Juan-Pardo, K. Demeyer, K. Hole, E. Larrea, E. Timmerman, J. Prieto, T. Arnesen, F. Sherman, K. Gevaert, and R. Aldabe. 2012. 'N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB', *Proc Natl Acad Sci U S A*, 109: 12449-54.

Van Damme, Petra, Rune Evjenth, Håvard Foyn, Kimberly Demeyer, Pieter-Jan De Bock, Johan R. Lillehaug, Joël Vandekerckhove, Thomas Arnesen, and Kris Gevaert. 2011b. 'Proteome-derived Peptide Libraries Allow Detailed Analysis of the Substrate Specificities of Nα-acetyltransferases and Point to hNaa10p as the Post-translational Actin Nα-acetyltransferase', *Molecular & Cellular Proteomics*, 10.

Van Damme, Petra, Thomas V. Kalvik, Kristian K. Starheim, Veronique Jonckheere, Line M. Myklebust, Gerben Menschaert, Jan Erik Varhaug, Kris Gevaert, and Thomas Arnesen. 2016b. 'A Role for Human N-alpha Acetyltransferase 30 (Naa30) in Maintaining Mitochondrial Integrity', *Molecular & cellular proteomics : MCP*, 15: 3361-72.

Varland, S., and T. Arnesen. 2018. 'Investigating the functionality of a ribosome-binding mutant of NAA15 using Saccharomyces cerevisiae', *BMC Res Notes*, 11: 404.

Varland, S., C. Osberg, and T. Arnesen. 2015. 'N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects', *Proteomics*, 15: 2385-401.

Vetting, M. W., D. C. Bareich, M. Yu, and J. S. Blanchard. 2008. 'Crystal structure of RimI from Salmonella typhimurium LT2, the GNAT responsible for N (alpha)-acetylation of ribosomal protein S18', *Protein Sci*, 17: 1781-90.

Vetting, M. W., L. P. de Carvalho, S. L. Roderick, and J. S. Blanchard. 2005. 'A novel dimeric structure of the RimL Nalpha-acetyltransferase from Salmonella typhimurium', *J Biol Chem*, 280: 22108-14.

Vinueza-Gavilanes, R., I. Inigo-Marco, L. Larrea, M. Lasa, B. Carte, E. Santamaria, J. Fernandez-Irigoyen, R. Bugallo, T. Aragon, R. Aldabe, and M. Arrasate. 2020. 'N-terminal acetylation mutants affect alpha-synuclein stability, protein levels and neuronal toxicity', *Neurobiol Dis*: 104781.

Wahl, M. C., G. P. Bourenkov, H. D. Bartunik, and R. Huber. 2000. 'Flexibility, conformational diversity and two dimerization modes in complexes of ribosomal protein L12', *EMBO J*, 19: 174-86.

Watson, M. D., and J. C. Lee. 2019. 'N-Terminal Acetylation Affects alpha-Synuclein Fibril Polymorphism', *Biochemistry*, 58: 3630-33.

Wenzlau, J. M., P. J. Garl, P. Simpson, K. R. Stenmark, J. West, K. B. Artinger, R. A. Nemenoff, and M. C. Weiser-Evans. 2006. 'Embryonic growth-associated protein is one subunit of a novel N-terminal acetyltransferase complex essential for embryonic vascular development', *Circ Res*, 98: 846-55.

Weyer, F. A., A. Gumiero, K. Lapouge, G. Bange, J. Kopp, and I. Sinning. 2017. 'Structural basis of HypK regulating N-terminal acetylation by the NatA complex', *Nat Commun*, 8: 15726.

Wiame, E., G. Tahay, D. Tyteca, D. Vertommen, V. Stroobant, G. T. Bommer, and E. Van Schaftingen. 2018. 'NAT6 acetylates the N-terminus of different forms of actin', *Febs j*, 285: 3299-316.

Williams, B. C., C. M. Garrett-Engele, Z. Li, E. V. Williams, E. D. Rosenman, and M. L. Goldberg. 2003. 'Two putative acetyltransferases, san and deco, are required for establishing sister chromatid cohesion in Drosophila', *Curr Biol*, 13: 2025-36.

Wisniewski, J. R., A. Zougman, N. Nagaraj, and M. Mann. 2009. 'Universal sample preparation method for proteome analysis', *Nat Methods*, 6: 359-62.

Yang, Dongxue, Qianglin Fang, Mingzhu Wang, Ren Ren, Hong Wang, Meng He, Youwei Sun, Na Yang, and Rui-Ming Xu. 2013. 'Nα-acetylated Sir3 stabilizes the conformation of a nucleosome-binding loop in the BAH domain', *Nature Structural & Molecular Biology*, 20: 1116–18.

Yi, C. H., D. K. Sogah, M. Boyce, A. Degterev, D. E. Christofferson, and J. Yuan. 2007. 'A genome-wide RNAi screen reveals multiple regulators of caspase activation', *J Cell Biol*, 179: 619-26.

Yoon, H., H. L. Kim, Y. S. Chun, D. H. Shin, K. H. Lee, C. S. Shin, D. Y. Lee, H. H. Kim, Z. H. Lee, H. M. Ryoo, M. N. Lee, G. T. Oh, and J. W. Park. 2014. 'NAA10 controls osteoblast differentiation and bone formation as a feedback regulator of Runx2', *Nat Commun*, 5: 5176.

Yu, M., M. Ma, C. Huang, H. Yang, J. Lai, S. Yan, L. Li, M. Xiang, and D. Tan. 2009. 'Correlation of expression of human arrest-defective-1 (hARD1) protein with breast cancer', *Cancer Invest*, 27: 978-83.

Zhang, K. 2016. 'Gctf: Real-time CTF determination and correction', *J Struct Biol*, 193: 1-12.

Zhao, Jianhua, Sarah E. Noon, Ian D. Krantz, and Yaning Wu. 2016. 'A de novo interstitial deletion of 7q31.2q31.31 identified in a girl with developmental delay and hearing loss', *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 172: 102-08.

Zheng, S. Q., E. Palovcak, J. P. Armache, K. A. Verba, Y. Cheng, and D. A. Agard. 2017. 'MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy', *Nat Methods*, 14: 331-32.

Zimmermann, Lukas, Andrew Stephens, Seung-Zin Nam, David Rau, Jonas Kübler, Marko Lozajic, Felix Gabler, Johannes Söding, Andrei N. Lupas, and Vikram Alva. 2018. 'A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core', *Journal of Molecular Biology*, 430: 2237-43.

Zivanov, J., T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl, and S. H. Scheres. 2018. 'New tools for automated high-resolution Cryo-EM structure determination in RELION-3', *Elife*, 7: e42166.