Structural and functional characterization of the N-terminal acetyltransferase NatC

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Stephan Grunwald

Berlin
October 01, 2019

Die vorliegende Arbeit wurde von April 2014 bis Oktober 2019 am

Max-Delbrück-Centrum für Molekulare Medizin unter der Anleitung von

PROF. DR. OLIVER DAUMKE

angefertigt.

Erster Gutachter:	PROF. DR. OLIVER DAUMKE
Zweite Gutachterin:	PROF. DR. ANNETTE SCHÜRMANN

Disputation am 26. November 2019

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen – die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; und dass diese Dissertation keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat.

Berlin, 9. November 2020

Stephan Grunwald

Acknowledgement

I would like to thank Prof. Oliver Daumke for giving me the opportunity to do the research for this project in his laboratory and for the supervision of this thesis.

I would also like to thank Prof. Dr. Annette Schürmann from the German Institute of Human Nutrition (DifE) in Potsdam-Rehbruecke for being my second supervisor.

From the Daumke laboratory I would like to especially thank Dr. Manuel Hessenberger, Dr. Stephen Marino and Dr. Tobias Bock-Bierbaum, who gave me helpful advice. I also like to thank the rest of the lab members for helpful discussions.

I deeply thank my wife Theresa Grunwald, who always had some helpful suggestions and kept me alive while writing this thesis.

Abbreviations

AcCoA / Acetyl-CoA	(Acetyl-) coenzyme A
ARFRP1	ADP-ribosylation factor-related protein 1 (a *trans*-Golgi-associated small GTPase)
Cryo-EM	Cryogenic electron microscopy
EPR	Electropositive Region
ES	Ribosomal RNA expansion segments
Gag	Major capsid protein of the *Saccharomyces cerevisiae* virus L-A
GNAT	GCN5-related N-acetyltransferase superfamily
HAT	Histone acetyl transferase
HRV 3C	Human rhinovirus 3C protease / recognition site
IMAC	Immobilized metal affinity chromatography
KAT	Lysine acetyl transferase
MALDI-TOF MS	Matrix-assisted laser desorption/ionization – time-of-flight – mass spectrometry
MetAP	Methionine aminopeptidase
NAA/Naa	Nα-acetyltransferase gene/protein, respectively.
NAT	N-terminal acetyltransferase (enzyme that the catalyzes the Nt-acetylation)
Nt-acetylation	N-terminal acetylation (also known as Nα-acetylation)
pI	Isoelectric point
PE	Ribosome nascent polypetide exit tunnel
PTM	Post-translational modification
SEC	Size-exclusion chromatography
yArl3	ADP-ribosylation factor-like protein 3 from *S. cerevisieae* (the yeast ortholog of human ARFRP1)
X-ray	X-radiation, a form of high-energy electromagnetic radiation.

The above list represents only frequently occurring abbreviations that are beneficial for text comprehension. Further abbreviations are found in the materials and methods sections or defined upon first use.

Table of contents

1 Introduction

1.1 *The 'missing' genetic information in eukaryotic genomes*

The Human Genome Project finished its goal of obtaining a highly accurate sequence of the vast majority of the euchromatic[1] human genome in 2004 (International Human Genome Sequencing 2004). The 2.85 billion nucleotides of the haploid human genome encode only 20,000–25,000 protein-coding genes, which is approximately one fifth of the number that was originally envisaged (Council 1988; Antequera and Bird 1993; Liang et al. 2000). Similarly, other eukaryotic genomes contain only a limited number of genes, e.g. ~13,600 in *Drosophila melanogaster* (Adams et al. 2000), ~25,500 in *Arabidopsis thaliana* (Arabidopsis Genome 2000) and roughly 6000 genes in *Saccharomyces cerevisiae* (Goffeau et al. 1996). A central question that has to be answered is how such limited number of genes can lead to the development and functionality of a living being as complex as a human.

Nilsen and Graveley (2010) have concluded that the 'missing' genetic information is in large part provided by alternative splicing. During the splicing reaction, eukaryotic gene transcripts are processed and assembled into mature mRNA through excision of non-coding introns. Alternative splicing patterns determine the inclusion of different exons in the mRNA, giving rise to multiple protein isoforms that differ in sequence and subsequently have different chemical and biological properties (Black 2003).

Approximately 92–95% of human multiexon gene transcripts undergo alternative splicing (Pan et al. 2008; Wang et al. 2008). Through analysis of RNA sequencing data, Hu et al. (2015) concluded that the expression of alternate transcripts is mostly tissue specific and estimated that there are more than 200,000 mRNA transcripts with protein-coding potential, which correspond to roughly ten variants per annotated human gene. However, it is still debated to which extent alternative splicing contributes to protein diversity (Chaudhary et al. 2019). Weatheritt, Sterne-Weiler, and Blencowe (2016) suggested that a majority of splice variants are translated, and that specific cellular functions like cell cycle control are subject to alternative splicing-dependent modulation of translation output. Furthermore, Yang et al. (2016) showed that alternative splicing can produce protein isoforms with vastly different protein-protein interaction profiles. However, another proteomic study suggests that there is only a single dominant protein isoform for the vast majority of protein-coding genes, irrespective of tissue or cell type (Ezkurdia et al. 2015). After reviewing the current evidence, Tress, Abascal, and Valencia (2017) conclude that alternative splicing may not be the key to proteome complexity.

[1] The euchromatic genome (~2.88 Gb) covers about 93.5% of overall human genome (~3.08 Gbs) (International Human Genome Sequencing 2004). Euchromatin represent the lightly packed chromatin, which is enriched in genes and more actively transcribed than heterochromatin.

1.2 *Post-translational modifications — a major source of protein diversity?*

The second route of proteome expansion is achieved through post-translational modifications (PTMs), which are covalent modifications that occur after the translation of mRNA into proteins or, in some cases, even during the translation process (co-translational modification). The nascent or folded proteins are subjected to a variety of enzyme-catalyzed modifications on their side chains or backbones. PTMs can be divided into two categories: enzyme-catalyzed covalent additions of a chemical group to a side chain or backbone of a protein and covalent cleavage of peptide backbones either by proteases or by autocatalytic cleavage. Fifteen out of the twenty common proteogenic amino acids are post-translationally modified. The five most common types are phosphorylation, glycosylation, acylation, alkylation and oxidation (Walsh, Garneau-Tsodikova, and Gatto 2005). In total, about 5% of the genome of higher eukaryotes encodes enzymes that carry out PTMs (Walsh, Garneau-Tsodikova, and Gatto 2005).

The most actively researched PTM type is phosphorylation, which was already discovered sixty years ago (Fischer et al. 1959). Kinases, the enzymes that catalyze the phosphorylation reaction, form a superfamily of more than 500 members, which is termed the kinome (Manning et al. 2002). Protein phosphorylation is a reversible PTM, which is tightly controlled by phosphatase enzymes. This reversibility is consistent with the importance of kinases and phosphatases in receptor-mediated signal transduction and cell cycle control (Hunter 1995). Other reversible PTMs include glycosylation, alkylation, lysine-acylation and disulfide formation.

In the literature, different terms have been used to describe the highly related, but chemically different, protein molecules that are derived from a single gene, including the terms: protein forms, protein isoforms, protein species and protein variants. However, the terms “variants” and “isoforms” were intended to describe proteins derived from distinct DNA or RNA and the term “protein species” does not distinguish between proteins originating from different genes and those originating from a single gene. Thus a new term, ‘proteoform’, was proposed to designate all of the different molecular forms in which the protein product of a single gene can be found, including changes due to genetic variations, alternatively spliced RNA transcripts and post-translational modifications (Smith, Kelleher, and Consortium for Top Down 2013).

The question of how many proteoforms are created from the roughly 20,000 human genes is hard to answer. The theoretical number of proteoforms created through post-translational modifications is 2^n, where n is the number of PTMs (Aebersold et al. 2018). This calculation works for site-specific PTMs that are binary, like phosphorylation and acetylation. However, lysine residues can be unmodified, acetylated or methylated at three different levels: mono-, di- and trimethylated (Wang and Liu 2017) and thus lysines can exist in five different states. Aebersold et al. (2018) used the specific example of the human histone H4, for which the 58 SwissProt-annotated PTM modifications at seventeen different sites would yield more than 10^{10} theoretical proteoforms. Even when using the 13 most common PTM sites from the literature, plus an allelic variant (E64Q), the authors calculate more than 98,000 possible proteoforms for histone H4. However, in an extensive mass spectrometry

project only 74 proteoforms were reported, and only 21 unambiguously (Dang et al. 2014). Thus, even though the potential complexity of the proteome is enormous, the few available studies suggest that only a fraction of this complexity is used.

1.3 *Protein acetylation*

Protein acetylation is one of the major PTMs in eukaryotes, in which the acetyl group from acetyl coenzyme A (acetyl-CoA) is transferred to either the protein Nα-terminus or the ε-amino group of a lysine side chain (Drazic et al. 2016) (see Fig. 1). The modification of the N-terminal amino group (N-terminal acetylation or Nt-acetylation) is catalyzed by N-terminal acetyltransferases (NATs) and is considered to be irreversible as no Nt-deacetylase has thus far been identified. The catalytic subunits of NATs belong to the GCN5-related N-acetyltransferase (GNAT) superfamily (Vetting et al. 2005) (see section 1.8). Protein acetylation of lysine ε-amino groups is catalyzed by lysine acetyltransferases (KATs), which were originally named histone acetyltransferases (HATs), because they were the first proteins for which a reversible acetylation of the ε-amino group was identified (Allfrey, Faulkner, and Mirsky 1964). KATs can be grouped into three major families GCN5, CBP/p300 and MYST, which are reviewed by Berndsen and Denu (2008). The removal of the acetyl groups is catalyzed by lysine deacetylases (KDACs), which belong to two different families; Zn^{2+}-dependent histone deacetylases and NAD^{+}-dependent sirtuin deacetylases (Choudhary et al. 2014).

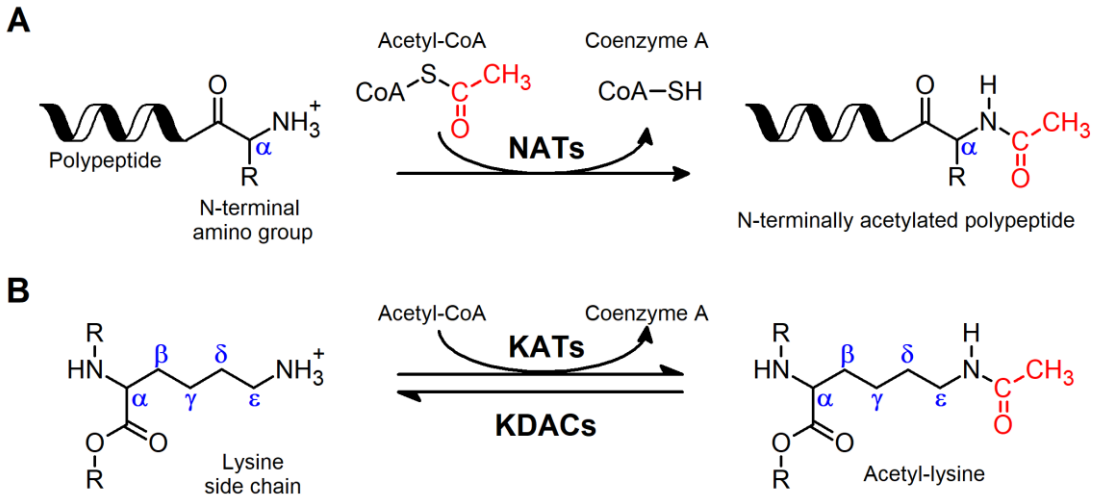


Fig. 1 Schematic overview of acetylation and deacetylation. (A) Transfer of an acetyl group from acetyl-CoA to the N-terminal amino group of a polypeptide, catalyzed by N-terminal acetyltransferases (NATs). (B) Reversible acetylation of the ε-amino group of a lysine residue by lysine acetyltransferases (KATs). The deacetylation of lysine residues is catalyzed by lysine deacetylases (KDACs). Adapted from (Drazic et al. 2016).

1.4 *Evolutionary development of N-terminal acetylation*

N-terminal acetylation occurs in all kingdoms of life, but the frequency increases dramatically with the complexity of the organism. In *E. coli*, five proteins have long been known to be N-terminally acetylated, three of them being ribosomal proteins from the small or large ribosomal subunit (Gordiyenko et al. 2008). These proteins, with N-termini Ala-His-, Ala-Arg-, and Ser-Ile- are each acetylated by a separate, specific N-terminal acetyltransferase (Yoshikawa et al. 1987; Tanka et al. 1989). In bacteria and eukaryotic organelles, the initiator methionine of nascent chains is Nα-terminally formylated. The formyl moiety is co-translationally removed by a ribosome-bound peptide deformylase (Piatkov et al. 2015). The resulting methionine can be cleaved off subsequently by methionine aminopeptidases (MetAPs), if the subsequent residue at position 2 is not larger than valine (Frottin et al. 2006). Only then can the N-termini be acetylated by their dedicated N-terminal acetyltransferase. Both peptide deformylase and MetAP compete for the same binding site close to the exit tunnel of the *E. coli* ribosome. Their extremely fast association and dissociation kinetics allow them to frequently sample the ribosome and ensure processing of nascent chains after their emergence, followed by later maturation events, including chaperone recruitment and folding (Sandikci et al. 2013). A recent proteomics study has now identified a total of 31 N-terminally acetylated proteins in *E. coli* (Schmidt et al. 2016), comprising less than 1% of the *E. coli* genome[2] (Grenier et al. 2014).

In archaea, acetylated N-termini are far more common, affecting about 15% of the proteins (Falb et al. 2006). The haloarchaeal N-terminal acetyltransferase(s) reveal a narrow substrate specificity, which is limited to N-termini starting with Ala or Ser, resembling the activity of the eukaryotic NatA complex (see below). Consequently, they also require the prior removal of the initial methionine by MetAP. Many proteins are only partially Nt-acetylated and complete acetylation is seen in only half of the acetylated proteins (Falb et al. 2006). Another archaea, the crenarchaeote *Sulfolobus solfataricus*, revealed a more relaxed substrate specificity and its single NAT, ssArd1, was able to acetylate Ala-, Ser- and Thr-, and also Met-Glu- amino termini (Mackay et al. 2007). The authors suggest that ssArd1 represents an ancestral form of the eukaryotic NATs and that the latter have evolved a complex NAT machinery, that contains multiple NAT paralogs with a much more restricted substrate specificity.

In the lower eukaryote *Saccharomyces cerevisiae*, about 50% and in humans about 80% to 90% of all proteins are completely or partially Nt-acetylated (Polevoda and Sherman 2003; Arnesen et al. 2009; Van Damme, Hole, et al. 2011). Arnesen et al. (2009) found that only 12% of yeast proteins are fully acetylated at their N-terminus, whereas 45% are partially acetylated and 43% are non-acetylated. In contrast, 76% of human proteins are fully acetylated, 8% partially acetylated and 16% non-acetylated.

[2] The *E. coli* genome contains approximately 4,096 genes (https://www.uniprot.org/proteomes/UP000029103).

1.5 *The eukaryotic NAT machinery*

Eukaryotes have at least five different highly conserved NATs, termed NatA to NatE, suggesting an increase in complexity of the NAT machinery during evolution (Rathore et al. 2016). Higher eukaryotes additionally express NatF and plants and animals even possess a seventh, kingdom-specific NAT: Naa70 (Dinh et al. 2015) or Naa80, respectively (Drazic et al. 2018).

Rathore et al. (2016) suggested that all six major eukaryotic N-terminal acetyltransferases were present in the Last Eukaryotic Common Ancestor (LECA), i.e. NAT diversification did not occur in the eukaryotes. However some NATs were secondarily lost during evolution: for example, NatF is absent in *S. cerevisiae* (Van Damme, Hole, et al. 2011).

The eukaryotic NATs differ from each other in terms of subunit composition and substrate specificity (Table 1). Each NAT consists of at least a unique catalytic subunit and up to two auxiliary subunits that contribute to substrate specificity or mediate NAT-ribosome interactions (Aksnes et al. 2016). A simplified nomenclature for N-terminal acetyltransferase genes was introduced in 2009: The symbol *NAA*/Naa (Nα-acetyltransferase) was assigned to each gene/protein of the corresponding NAT complex (Polevoda, Arnesen, and Sherman 2009). For example, the NatC complex, analyzed in this thesis, consists of the catalytic subunit Naa30 and two auxiliary subunits Naa35 and Naa38. Orthologous genes have the same name, e.g. hNaa30 (human), yNaa30 (yeast, *S. cerevisiae*), and are assigned not only by their sequence similarity but also by their substrate specificity, i.e. they acetylate the same set of proteins. Each NAT recognizes a specific subset of protein N-termini, depending mainly on the amino acid sequence of the first two amino acids.

Table 1 Composition and substrate specificity of eukaryotic NATs Adapted from (Ree, Varland, and Arnesen 2018). [a]Amino acids are listed according to *in vivo* Nt-acetylation specificity. [b]NatF is located at the Golgi-membrane and performs post-translational Nt-acetylation. Common subunit names are given in parentheses.

	NatA	**NatB**	**NatC**	**NatD**	**NatE**	**NatF[b]**	**NatG**	**NatH**
Prevalence	All eukaryotes (with some clade-specific losses)					Only higher eukaryotes	Plant chloroplasts	Animals
Catalytic subunit[a]	Naa10 (Ard1)	Naa20 (Nat3)	Naa30 (Mak3)	Naa40 (Nat4)	Naa50 (Nat5)	Naa60	Naa70	Naa80
Auxiliary subunit	Naa15 (Nat1)	Naa25 (Mdm20)	Naa35 (Mak10) Naa38 (Mak31)	—	Naa10 Naa15	—	—	—
Specificity	Ala- Ser- Thr- Cys- Val- Gly-	MD- MN- ME- MQ-	ML- MI- MF- MY- MW- MV- MM- MH- MK-	SGG- (H2 and H4 histones)	MK- MV- MA- MY- MF- ML- MS- MT-	MK- MS- MV- ML- MQ- MI- MY- MT-	M- A- S- T-	DDD-/EEE- (β- and γ-actin)

The NatA complex, which consists of the catalytic subunit Naa10 and the auxiliary subunit Naa15, co-translationally acetylates N-termini that bear a small amino acid after the initiator methionine is removed by methionine aminopeptidases[3] (Arnesen et al. 2005; Arnesen et al. 2009; Liszczak et al. 2013). NatA is the major NAT in view of the number of its potential substrates, which represent 46% of the human proteome (Fig. 2B), of which 83% is (partially or fully) Nt-acetylated and 17% is unacetylated. Thus, NatA is estimated to Nt-acetylate 38% (Fig. 2A) of the human proteome and an estimated 8% of potential NatA substrates are not acetylated (Ree, Varland, and Arnesen 2018). Notably, purified monomeric Naa10 has been shown to post-translationally acetylate acidic N-termini (Asp- or Glu-) *in vitro* with β- (DDDI-) and γ-actin (EEEI-) being the most prominent substrates (Van Damme, Evjenth, et al. 2011; Liszczak et al. 2013).

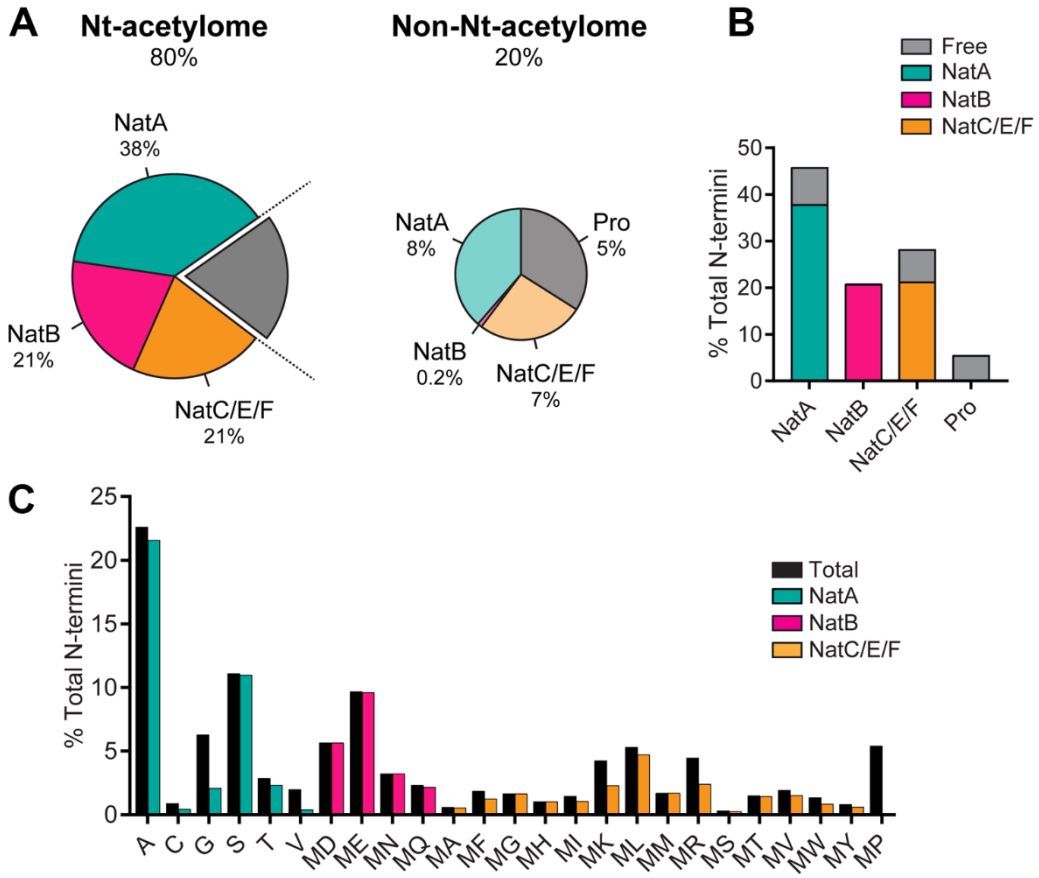


Fig. 2 The human Nt-acetylome. Adapted from Ree, Varland, and Arnesen (2018). (A) Prevalence of Nt-acetylation in humans predicted by combining experimental data on all SwissProt entries (version 57.8.), based on the occurrence of the first two amino acids. (B) NAT substrate class and frequency. (C) N-terminal amino acid frequency.

The NatB complex is formed by the catalytic subunit Naa20 and the auxiliary subunit Naa25. It co-translationally acetylates N-termini starting with methionine, followed by an acidic residue (MD-, ME-) or their amide (MN-, MQ) (Starheim et al. 2008; Van Damme et al. 2012). NatB substrates

[3] The aminopeptidase removes methionine when it precedes a small amino acid with a gyration radius of 1.29 Å or less, i.e. Ala, Cys, Gly, Pro, Ser, Thr or Val (Sherman, Stewart, and Tsunasawa 1985).

represent 21% of the human proteome (Fig. 2A) and almost every protein of this substrate class is acetylated (Fig. 2B, C).

The heterotrimeric NatC complex, whose structure and function are the focus of this thesis, is composed of a catalytic subunit Naa30, a large auxiliary subunit Naa35 and a small auxiliary subunit Naa38. Polevoda and Sherman (2001) showed that deletion of any of the three subunits in yeast strains results in a lack of Nt-acetylation of a NatC substrate *in vivo*.

NatC acetylates protein N-termini starting with methionine, followed by a hydrophobic amino acid (ML-, MI-, MF-, MW-) (Tercero, Dinman, and Wickner 1993; Polevoda et al. 1999; Kimura et al. 2000; Polevoda and Sherman 2001; Starheim et al. 2009). Behnia et al. (2004) have shown that the N-terminus of the *S. cerevisiae* GTPase yArl3 (starting with MF-) is Nt-acetylated by NatC. They also showed that the human ortholog of yArl3, Arf-related protein 1 (ARFRP1), interacts with hSys1 and that the alanine substitution of its tyrosine at position 2 disrupts this interaction, suggesting that MY-termini are additional targets of NatC. Human NatC has been shown to additionally acetylate MM-, MV-, MK- and MH-termini *in vivo* (Van Damme et al. 2016).

Starheim et al. (2009) purified an MBP-tagged human Naa30 subunit which displayed N-terminal acetyltransferase activity toward ML- termini, suggesting that in humans, NatC activity and substrate specificity are at least in part contained within the catalytic core alone. They also showed that hNaa30 knockdown cells (HeLa and CAL-62) demonstrated a stronger phenotype – a fragmentation of the Golgi stack – than hNaa35 and hNaa38 knockdown cells, suggesting that Naa30 has an additional function independent of Naa35 and Naa38. Also the *Arabidopsis thaliana* ortholog of Naa30, atNaa30 can functionally replace yeast Naa30, Naa35 and Naa38 (Pesaresi et al. 2003). While an atNaa30 mutant showed a decreased synthesis of the PSII core proteins D1 and CP47, resulting in decreased photosynthesis, no phenotype was observed for an atNaa35 mutant. These results indicate that Naa30 function does not strictly require NatC complex formation in higher eukaryotes, as it does in yeast.

The monomeric NatD (Naa40) was identified as a highly selective NAT that Nt-acetylates the SG-sequence of histones H4 and H2A in all species from yeast to mammals (Song et al. 2003; Hole et al. 2011). N-termini starting with serine are also typical NatA substrates. However, Mullen et al. (1989) found that the Nt-acetylation of histones H4 and H2A is not affected in *S. cerevisiae* Naa10 or Naa15 mutant strains.

NatE is composed of the catalytic subunit Naa50 and both NatA subunits Naa10 and Naa15 (Gautschi et al. 2003; Arnesen et al. 2006). Gautschi et al. (2003) were the first to notice that yeast NatA binds to another NAT, Naa50, forming a trimeric complex in a 1:1:1 ratio with Naa10 and Naa15. However, they observed no phenotype for a ΔNaa50 yeast strain, compared to ΔNaa10 or ΔNaa15 mutant strains, which show temperature sensitivity and reduced mating efficiency. However, yeast Naa50 lacks the optimal acetyl-CoA binding motif (Arnesen et al. 2006) and is enzymatically inactive (Deng et al. 2019). A recent cryo-EM structure of the yeast NatA/Naa50 complex showed that yeast Naa50 contributes to the NatA-ribosome interaction (Knorr et al. 2019).

Arnesen et al. (2006) demonstrated that the same heterotrimeric assembly (Naa10/Naa15/Naa50) exists in *H. sapiens* and, in contrast to yNaa50, is catalytically active. Uncomplexed human Naa50 has been shown to efficiently acetylate ML- and MI- termini, thus partly overlapping with NatC substrate specificity, implying a functional redundancy; however a phenylalanine at position 2 led to a dramatic increase in the K_m value for Naa50, in contrast to NatC (Evjenth et al. 2009). Due to the distinct NAT activity of Naa50, which is different from NatA, it is named NatE. Arnesen et al. (2006) furthermore showed that uncomplexed Naa50 also has a low lysine acetyltransferase (KAT) activity *in vitro*: it can Nϵ-acetylate histone H4 and also auto-acetylates three of its own lysines, albeit at a rate 2000 times slower than its NAT activity. They showed that Naa50 autoacetylation leads to an increased enzyme specificity toward NAT substrates, but decreased KAT activity. Van Damme et al. (2015) showed that ectopic expression of human Naa50 in *S. cerevisiae* resulted in Nt-acetylation of N-termini, starting with MK, MV, MA, MY, MF, ML, MS, and MT. The identification of acetylated iMet followed by a small residue (MA, MS, MT, MV), revealed a kinetic competition between Naa50 and methionine aminopeptidases (MetAPs). The authors have shown *in vitro* that Nt-acetylated iMet cannot be removed by human MetAP2 and thus Naa50-mediated Nt-acetylation may act to retain the iMet of proteins, which are normally processed by MetAPs. They conclude that the physical association of Naa50 with the NatA complex might confer an ideal positioning to enable the two NATs to Nt-acetylate the very same N-termini, either with or without their iMet processed. Deng et al. (2019) showed that NatA/Naa50 complex formation enables a catalytic crosstalk between the Naa10 and Naa50 subunits, significantly increasing the catalytic efficiency of hsNaa50 in comparison to its uncomplexed state by dramatically increasing its substrate binding affinity. The catalytic efficiency of NatA was only slightly increased upon binding of Naa50.

Naa60 (NatF) is only present in higher eukaryotes, where it contributes to the higher abundance of Nt-acetylation compared to lower eukaryotes. Like NatC and NatE, NatF can process Met-starting N-termini, with a preference for MK-, MA, MV- and MM- sequences (Van Damme, Hole, et al. 2011). Aksnes et al. (2015) showed that Naa60 differs from all other NATs, as it is anchored to the cytosolic side of the Golgi, where it specifically acetylates transmembrane proteins, with a preference for cytosol-facing N-termini.

Together NatC, E and F substrates cover ca. 28% of the human proteome (Fig. 2B), with approximately 75% of the potential substrates being partially or fully Nt-acetylated (Fig. 2A,C) (Ree, Varland, and Arnesen 2018).

A seventh NAT, Naa70 (NatG), was recently discovered in *A. thaliana* (Dinh et al. 2015). AtNaa70 is located in the chloroplast, where it acetylates iMet and non-Met starting substrates, with a preference for M-, A-, S-, or T-starting N-termini. Furthermore, the authors also detected an Nϵ-acetylation activity of AtNaa70 on three internal lysines, similar to the autoacetylation activity observed for hsNaa50.

Finally, Drazic et al. (2018) recently discovered Naa80, which is responsible for the post-translational Nt-acetylation of processed actin in animals. Unprocessed cytosolic β- (MDDD-) or γ-

actin (MEEE-) N-termini are co-translationally Nt-acetylated by NatB (Van Damme et al. 2012). Rubenstein and Martin (1983) have shown that the N-terminal acetyl group is removed by a still unknown acetylmethionine aminopeptidase. Finally, the new N-termini (DDD-/EEE-) are post-translationally acetylated by the newly discovered Naa80, leading to the mature β- (Ac-DDD-) or γ-actin (Ac-EEE-).

1.6 *Co-translational Nt-acetylation at the ribosome*

Gautschi et al. (2003) first revealed that NatA (Naa10/Naa15) is quantitatively anchored to ribosomes via Naa15. By using chemical peptide crosslinking, they showed that Naa15 contacts the nascent chain, when about 40 amino acids have emerged from the exit tunnel (resembling a total length of the nascent chain of ca. 80 amino acids). They furthermore identified a direct interaction of Naa50 with NatA, which is thus also placed in close proximity to the ribosomal exit tunnel. Neither Naa10 nor Naa50 alone were able to bind to ribosomes without Naa15. Thus, the function of the NatA auxiliary subunit Naa15 seems to be the guidance of the growing peptide chain to the catalytic subunit Naa10 and, possibly, also Naa50, for co-translational acetylation of their corresponding substrates.

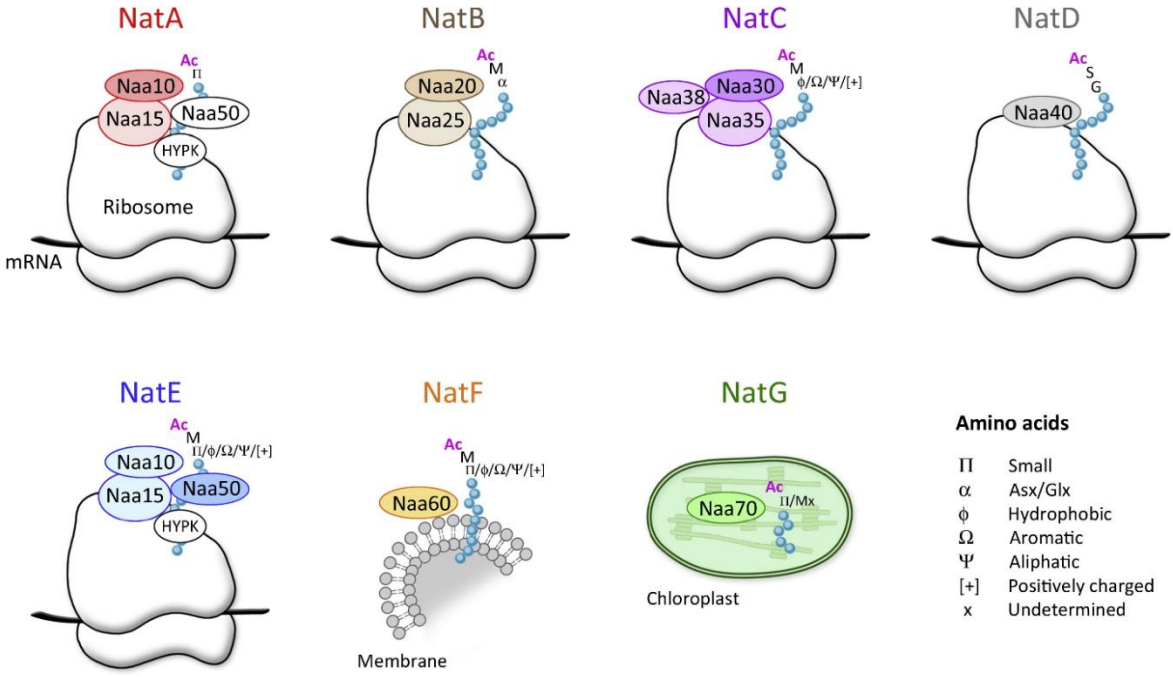


Fig. 3 Subunit composition and location of the NAT machinery Adapted from Aksnes et al. (2016). The catalytic subunits are typically associated with up to two auxiliary subunits that contribute to NAT complex activity through ribosome anchoring, complex stabilization and/or modulation of substrate specificity.

Later, ribosome interaction was demonstrated for all major eukaryotic NATs, NatA to NatE (Polevoda et al. 2008; Polevoda, Hoskins, and Sherman 2009; Hole et al. 2011), except for NatF, which specifically associates with Golgi membranes and acetylates transmembrane proteins (Aksnes et al. 2015) (see Fig. 3). Furthermore, Polevoda et al. (2008) constructed yeast strains with deletions of the NAT auxiliary subunits and could show that, in contrast to NatA, the catalytic subunit of NatC, Naa30, also binds in the absence of its auxiliary subunits Naa35 and Naa38. Moreover, they

identified the ribosomal proteins uL14 and uL23 to be associated with NatA. Both proteins are located in close proximity to the ribosomal exit tunnel. A recent cryo-EM structure of the NatA/Naa50-ribosome complex (Knorr et al. 2019) showed that both Naa15 and Naa50 interact with ribosomal RNA expansion segments, which positions NatA/Naa50 dynamically in close proximity to the ribosomal exit tunnel.

1.7 *Diverse functions of N-terminal acetylation*

Originally, it was thought that N-terminal acetylation might have a protective function towards proteolysis (Persson et al. 1985; Jornvall 1975), which was supported by initial experiments showing a stabilizing role of the acetylated N-termini (Hershko et al. 1984). In 2004, it was shown that proteins without a lysine can be degraded by the ubiquitin system via conjugation of the first ubiquitin moiety to the N-terminal residue, instead of the much more common conjugation to internal lysine residues (Ben-Saadon et al. 2004). Thus, Nt-acetylation had been proposed as a mechanism for blocking N-terminal ubiquitination and therefore preventing degradation. However, the exact opposite function for Nt-acetylation has been identified as well: The Doa10 ubiquitin ligase specifically targets acetylated protein N-termini for subsequent proteasome-dependent degradation (Hwang, Shemorry, and Varshavsky 2010).

Nt-acetylation furthermore affects membrane targeting and the subcellular location of proteins: Nt-acetylation of α-synuclein was demonstrated to stabilize its N-terminal helix, which increased its affinity to lipid vesicles of moderate charge similar to synaptic vesicles (Dikiy and Eliezer 2014).

The small GTPase ARFRP1 and its yeast counterpart yArl3 are both Nt-acetylated by the NatC complex. The acetylated N-termini bind directly to the *trans*-Golgi specific transmembrane protein hSys1/Sys1p, thus determining the localization of hARFRP1/yArl3 at the *trans* face of the Golgi apparatus (Behnia et al. 2004; Setty et al. 2004). Moreover, Starheim et al. (2009) showed that depletion of hNaa30 induces a fragmentation of the Golgi stack in HeLa and CAL-62 cells.

Using an *S. cerevisiae* Naa30 mutant strain, Tercero and Wickner (1992) found that the *MAK3* gene is necessary for N-terminal acetylation of the major capsid protein Gag of the yeast L-A virus and that this modification is necessary for viral particle assembly. Polevoda and Sherman (2001) showed that deletion of any NatC subunit in yeast results in lack of acetylation, diminished growth and the defective assembly of L-A viral particles.

In 2011, the first human genetic disorder related to a NAT gene was identified. A point mutation within the X-linked *NAA10*, coding for the catalytic subunit of NatA, causes a missense mutation (S37P), which leads to a significant decrease of Nt-acetylation activity. The afflicted boys suffering from the associated Ogden syndrome had an aged appearance, craniofacial anomalies, hypotonia, global development delays, cardiac arrhythmias and died during infancy (Rope et al. 2011).

1.8 *The structure of NAT enzymes*

During the last years, crystal structures of several Nα-acetyltransferases and NAT complexes have shed light on the structural mechanism of Nt-acetylation (Table 1). The structure of all major eukaryotic NATs, with the exception of the heterotrimeric NatC complex, have been solved in the presence of (acetyl-) coenzyme A and several different peptides, resembling their natural substrates

Table 2 Published structures of N-terminal acetyltransferases. With PDB, Protein Data Bank and (EMD), Electron Microscopy Data Bank accession codes. [a]Depositions without an associated publication.

PBD (EMD)	Structure	Res (Å)	Organism	Author
Archaeal NAT complexes				
4LX9	Archaeal NAT bound to AcCoA	2.0	*S. solfataricus*	(Liszczak and Marmorstein 2013
NatA complex (Naa10/Naa15), Naa10 monomer, NatA-Naa50, NatA-HYPK, NatA-Ribsome				
4HNY	Apo NatA complex	2.3	*S. cerevisiae*	(Neubauer et al. 2012)[a]
4HNX	NatA bound to ppGpp	2.3		
4HNW	NatA bound to Inositol Hexakisphosphate	2.8		
4KVX	Naa10 bound to AcCoA	2.0	*S. pombe*	(Liszczak et al. 2013)
4KVM	NatA bound to a bisubstrate analog	2.6		
4KVO	NatA bound to AcCoA	3.2		
5C88	Apo Naa10	2.5	*S. solfataricus*	(Chang and Hsu 2016)
5NNP, 5NNR	NatA bound to HYPK-THB	2.6 3.1	*C. thermophilum*	(Weyer et al. 2017)
6C9M	NatA complex	2.8	*H. sapiens*	(Gottlieb and Marmorstein 2018)
6C95	NatA bound to HYPK	3.2		
6HD7	Ribosome-NatA/Naa50 Cryo-EM structure	3.4	*S. cerevisiae*	(Knorr et al. 2019)
6O07	Dual enzyme NatA/Naa50 complex	2.7	*S. cerevisiae*	(Deng et al. 2019)
NatB complex (Naa20/Naa25)				
5K04	NatB bound to CoA and MES	2.4	*C. albicans*	(Hong et al. 2017)
5K18	NatB bound to bisubstrate inhibitor	2.7		
Naa40 (NatD)				
4U9V	Naa40 bound to AcCoA	1.8	*H. sapiens*	(Magin, Liszczak, and Marmorstein 2015)
4U9W	Naa40 bound to H4/H2A peptide and CoA	2.5		
4UA3	Selenomethionine labeled Naa40	1.9	*S. pombe*	
Naa50 (NatE)				
3TFY	Naa50 bound to substrate peptide and CoA	2.8	*H. sapiens*	(Liszczak, Arnesen, and Marmorstein 2011)
4X5K	Naa50 with CoA and an acetylated peptide	2.5	*H. sapiens*	(Reddi, Neubauer et al. 2014)[a]
4XNH	Naa50 (IP6) bound to a bisubstrate	2.1	*S. cerevisiae*	(Dong, Neubauer et al. 2015)[a]
Naa60 (NatF)				
5HGZ	Naa60 bound to AcCoA	1.4	*H. sapiens*	(Chen et al. 2016)
5HH0	Naa60 bound to CoA	1.6		
5ICW	Naa60 homodimer bound to CoA	2.0	*H. sapiens*	(Stove et al. 2016)
5ICV	Naa60 bound to a bisubstrate analogue	1.5		
Naa80 (NatH)				
5WJD	Naa80 bound to AcCoA	2.0	*D. melanogaster*	(Goris et al. 2018)
5WJE	Naa80 bound to a bisubstrate analogue	1.8		

The catalytic subunits of NATs belong to the GCN5-related N-acetyltransferase (GNAT) superfamily, which is an enormous superfamily of enzymes that uses acyl-CoAs (e.g. acetyl-CoA, propionyl-CoA and myristoyl-CoA) to acylate their cognate substrates. The superfamily includes, among others, Nα-terminal acetyltransferases (NATs), lysine acetyltransferases (KATs), aminoglycoside N-acetyltransferases (AACs), serotonin N-acetyltransferases (SNATs) and N-myristoyl transferases (NMTs) (Vetting et al. 2005).

All NATs share the structural GNAT domain (see Fig. 4), which is composed of an N-terminal strand (β1), followed by two helices (α1 and α2), three antiparallel strands (β2-β4), followed by a "signature" central helix (α3), a fifth strand (β5), a fourth helix (α4) and a final strand (β6) (Fig. 4B,C) (Vetting et al. 2005). Beta strands four and five splay apart, as a result of a β-bulge in β4.

The GNAT domain contains a central conserved CoA-binding motif, with the sequence Q/R-x-x-G-x-G/A (where x is any amino acid) and is located in the β4–α3 loop (Lu, Berkey, and Casero 1996; Wolf et al. 1998). Substrate specificity of NAT enzymes is mediated by the N-terminal α1–loop–α2 region and a C-terminal β-hairpin loop (β6–loop–β7) (Liszczak, Arnesen, and Marmorstein 2011; Liszczak et al. 2013).

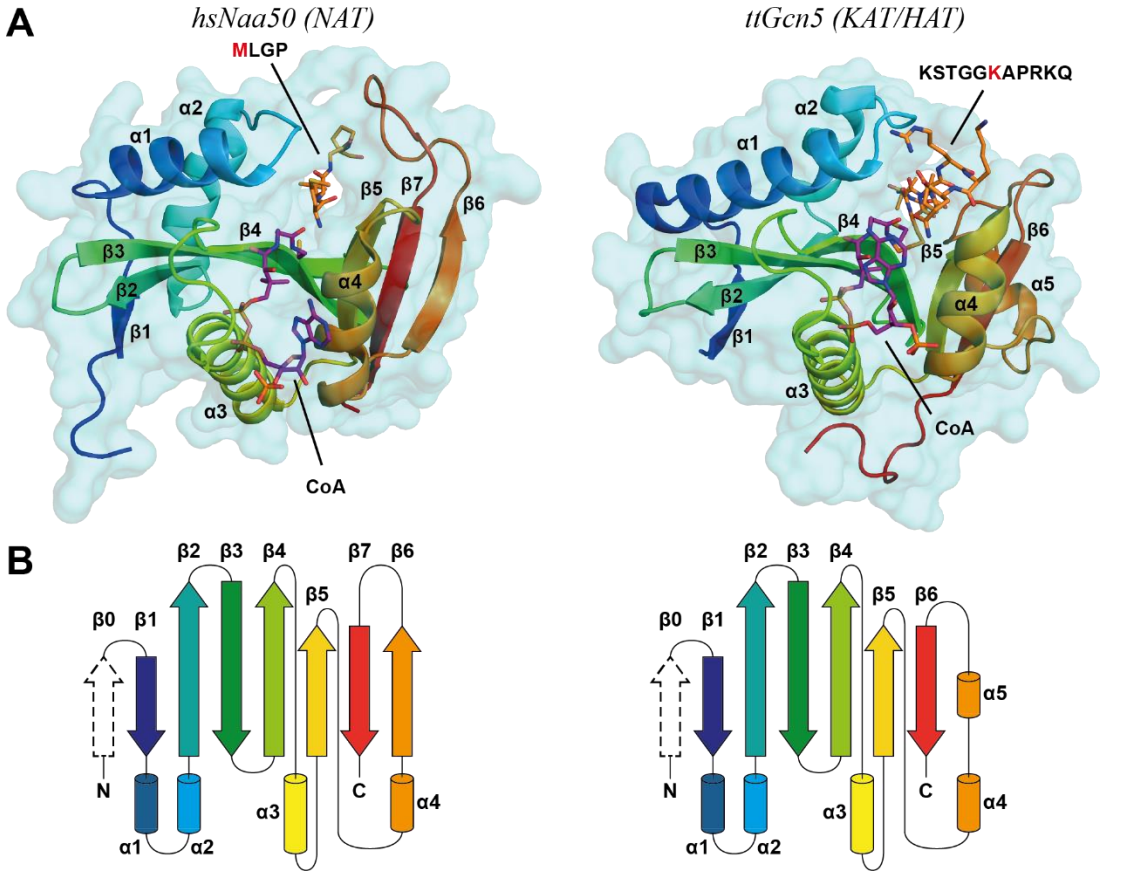


Fig. 4 NATs and KATs share a conserved GNAT fold. (A) Crystal structures of human Naa50 (left, PDB accession code 3TFY) and *Tetrahymena thermophila* Gcn5 (right, PDB code 1QSN) as representative of NAT or KAT enzymes, respectively. (B) Topology plot, showing the conserved GNAT-fold and the structural differences between NATs and KATs at the C-terminal end.

Liszczak, Arnesen, and Marmorstein (2011) solved the first crystal structure of an Nα-terminal acetyltransferase and showed that Naa50 can only accommodate an α-amino substrate and not a side chain lysine, which is acetylated by lysine acetyltransferases (KATs) like the histone acetyltransferase Gcn5. They have mapped the largest region of structural divergence between NATs and KATs to the C-terminal end of the two proteins (compare the two structures in Fig. 4). In hsNaa50, this region forms a β-hairpin (β6–loop–β7) containing a relatively long 14-residue loop, that blocks off part of the NAT substrate binding-site. In the KAT Gcn5, this region forms a loop–helix–loop–strand motif, which facilitates the formation of a relatively wide substrate-binding site (Liszczak, Arnesen, and Marmorstein 2011; Rojas et al. 1999).

As mentioned earlier, NatF/Naa60 differs from the ribosome-associated NatA to NatE as it specifically associates with Golgi membranes and acetylates transmembrane proteins. The subcellular location of Naa60 depends on a unique C-terminal membrane-integration region (Aksnes et al. 2015).

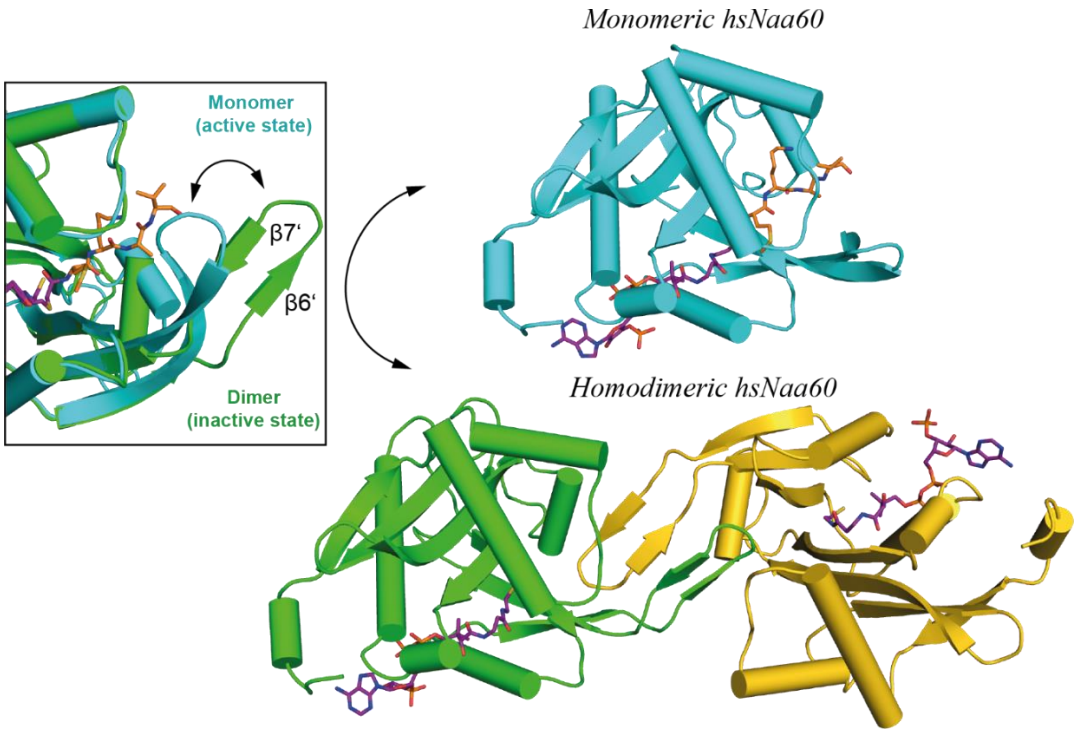


Fig. 5 Comparison of the monomeric and dimeric states of Naa60 Monomeric human Naa60 (cyan) is bound to the bisubstrate analog CoA-Ac-MKAV. Homodimeric Naa60 (in green and gold) bound to CoA. A superposition of the two structures reveals the movements in the β6–β7 loop. The PDB accession codes for monomeric and dimeric hsNaa60 are 5ICW and 5IVC, respectively.

The crystal structure of human Naa60 furthermore revealed the presence of two unique elongated loops within the conserved NAT region (Stove et al. 2016). The elongated β3–β4 loop folds back toward the enzyme core and is essential for protein stability. The β6–β7 loop of Naa60 hangs over the peptide binding site in a manner characteristic of all NATs. However, in Naa60, the β-hairpin structure of this loop contains two additional short strands, which the authors named β6' and β7'. This loop mediates dimerization of hsNaa60 in the absence of substrate peptide (Fig. 5), a feature that has not been seen for any other NAT. The β6–β7 loops of both protomers extend in the active

sites of each other, thereby competing with peptide binding. Stove et al. (2016) suggest that this could be a way to regulate substrate specificity, as the affinity for a potential substrate would need to be higher than the affinity for the other protomer of the dimer.

Liszczak et al. (2013) solved the structure of the heterodimeric *Schizosaccharomyces pombe* NatA complex, which is composed of the catalytic subunit Naa10 and the large auxiliary subunit Naa15. They showed that Naa15 adopts a ring-like topology, that wraps around Naa10 (Fig. 6, top). This interaction alters the catalytic activity of Naa10 for substrate-specific Nt-acetylation.

The heterodimeric NatB complex (Naa20/Naa25) forms a very similar tertiary and quaternary structure (Hong et al. 2017) (Fig. 6, bottom) with the auxiliary subunit Naa25 forming a 'horseshoe-like deck' that specifically hold its catalytic subunit Naa20. In contrast to NatA, the auxiliary subunit of NatB is essential for the solubility of its catalytic subunit Naa20.

The α-helices in the NatA and NatB auxiliary subunits form several helical bundle tetratricopeptide repeats (TPR). The TPR motif adopts a basic helix-turn-helix fold and consists of 34 amino acids with a degenerate consensus sequence, containing small and large hydrophobic amino acids (Zeytuni and Zarivach 2012). These repeats form a tertiary structure, which is found in numerous proteins and serves as an interaction module and mediator for heteromeric complex formation.

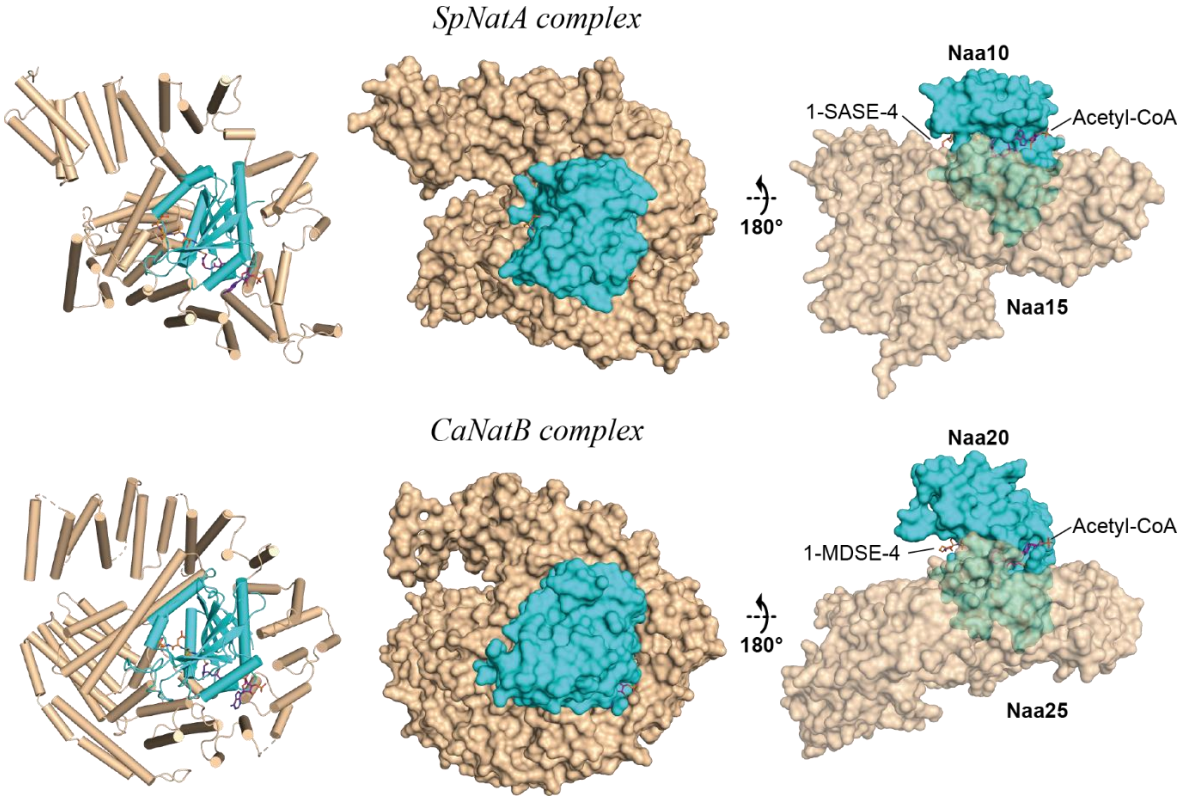


Fig. 6 Structures of the heterodimeric NatA and NatB complexes. Cartoon and surface presentations of *Schizosaccharomyces pombe* NatA (top) and *Candida albicans* NatB (bottom). NatA and NatB were both co-crystallized with a bisubstrate analog, CoA-SASE and CoA-MDEV, respectively. The catalytic subunits of NatA and NatB, Naa10 and Naa20 (cyan), are surrounded by their corresponding auxiliary subunit Naa15 or Naa25, respectively.

Deng et al. (2019) recently solved the structure of the trimeric NatA/Naa50 complex which comprises subunits Naa10, Naa15 and Naa50 (see Fig. 7, left). They could show that NatA/Naa50 complex formation enables a catalytic crosstalk between the Naa10 and Naa50 subunits, significantly increasing the catalytic efficiency of hsNaa50 in comparison to its uncomplexed state, by drastically increasing its substrate binding affinity. The catalytic efficiency of NatA was only slightly increased upon binding of Naa50. They concluded that the physical association of Naa50 with the NatA complex might infer an ideal positioning to enable the two NATs to Nt-acetylate the very same N-termini, either with or without their iMet processed. A recent cryo-EM structure of the trimeric NatA/Naa50 complex bound to the *S. cerevisiae* ribosome (Knorr et al. 2019), revealed that the trimeric NatA/Naa50 complex interacts with ribosomal RNA expansion segments via Naa15 and Naa50 (Fig. 7, right). Both catalytic subunits Naa10 and Naa50 are positioned dynamically in close proximity to the ribosomal exit tunnel.

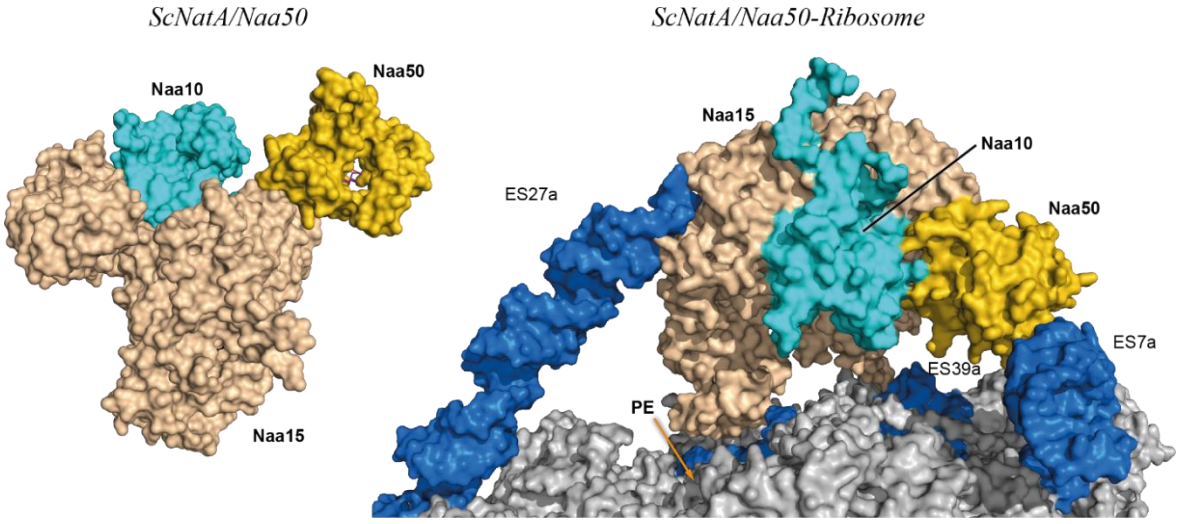


Fig. 7 Structures of heterotrimeric yeast NatA/Naa50 and the yNatA/Naa50-ribosome complex. Crystal structures of yeast NatA/Naa50 (PDB accession code 6O07, left). Molecular model of the NatA/Naa50-ribosome complex solved by cryo-EM (6HD7, right). The NatA/Naa50 complex binds to ribosomal RNA expansion segments (blue) near the nascent polypeptide exit tunnel (PE).

As described in an earlier section (see section 1.5), NatA is the major NAT in eukaryotes and is estimated to Nt-acetylate 38% of the human proteome (Ree, Varland, and Arnesen 2018). In search for possible NatA regulators, Arnesen et al. (2010) identified the Huntingtin yeast two-hybrid protein K (HYPK) as a stable interactor of NatA. HYPK was first identified in a yeast two-hybrid screen as an interaction partner of Huntingtin protein, which is associated with the neurodegenerative Huntigton's disease (Faber et al. 1998). HYPK was shown to be intrinsically disordered (Raychaudhuri et al. 2008) and Arnesen et al. (2010) demonstrated that NatA is required for the stability of HYPK and that both proteins prevent aggregation of the Huntingtin protein. Furthermore, HPYK is required for N-terminal acetylation of a known *in vivo* NatA substrate protein, PCNP (Arnesen et al. 2010). Recent studies of the *Chaetomium thermophilum* (Weyer et al. 2017) and human (Gottlieb and Marmorstein 2018) NatA/HYPK complexes, revealed that the HYPK C-terminal region is responsible for high affinity association with the C-terminal part of the NatA auxiliary subunit Naa15 (see Fig. 8, right). The N-terminus of HYPK binds to the Naa10 active site

entrance and inhibits the NatA catalytic activity *in vitro* (Weyer et al. 2017; Gottlieb and Marmorstein 2018). Gottlieb and Marmorstein (2018) conclude, that while HYPK inhibits NatA activity *in vitro*, it likely plays a more nuanced regulatory role *in vivo*, where it functions as part of an hNatA substrate proofreading mechanism.

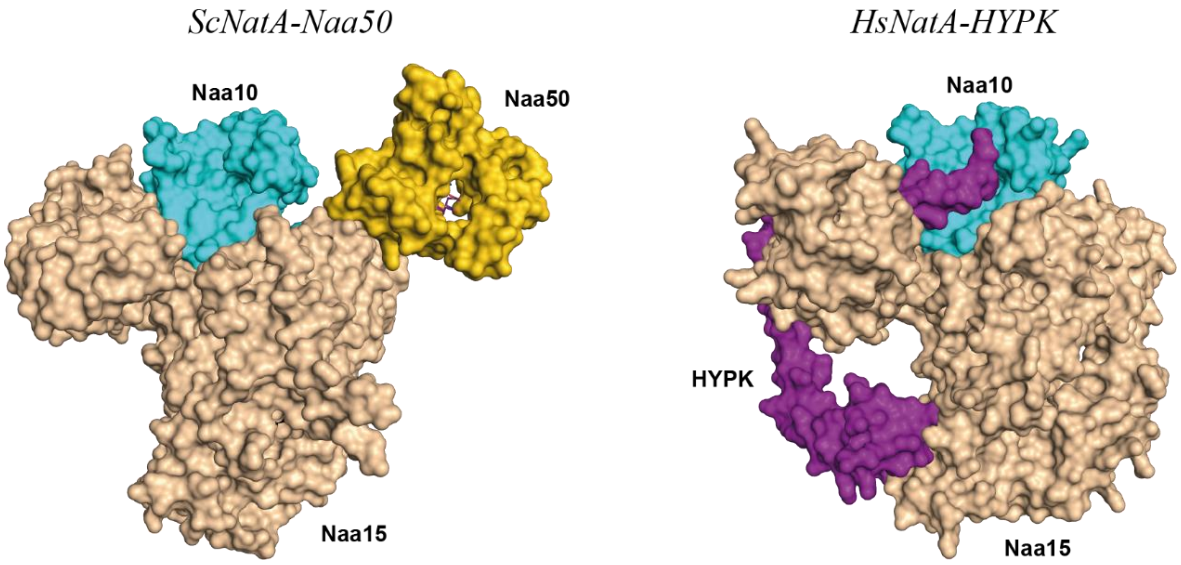


Fig. 8 Structures of the heterotrimeric NatA/Naa50 and NatA/HYPK complexes Crystal structures of yeast NatA/Naa50 (left, PDB accession code 6O07) and human NatA in complex with the Huntington yeast two-hybrid protein (HYPK) (right, PDB code 6C96).

Although HYPK and Naa50 bind to different NatA surfaces (compare both images in Fig. 8), they mediate NatA conformational changes, which prevent simultaneous binding of HYPK and Naa50 to NatA in humans and was confirmed by competitive pull-down studies. Thus, hHYPK binding blocks hNaa50 targeting to hNatA, thereby indirectly inhibiting the co-translational activity of hNatE. Gottlieb and Marmorstein (2018) suggest that an additional cellular protein might facilitate the simultaneous binding of HYPK and Naa50 to NatA in metazoans.

1.9 *General catalytic mechanism*

On the basis of many structural and functional studies of GNAT enzymes, a common general catalytic mechanism was identified, which involves the direct transfer of an acyl group from a donor substrate (usually acetyl CoA) to the amino group of an acceptor substrate (Fig. 9) (Salah Ud-Din, Tikhomirova, and Roujeinikova 2016).

The first step of the catalytic mechanism usually involves the deprotonation of the amino group ($-NH3^+$) of the acceptor substrate, which is mediated by an amino acid (usually Glu, Asp or Ser) near the active site that serves as a general base (B) and extracts a proton from the amino group of the peptide.

Second, the deprotonated amino group attacks the carbonyl carbon of the enzyme bound AcCoA, forming a transient zwitterionic tetrahedral intermediate. The breakdown of the intermediate occurs through a proton transfer from a general acid (H-A), usually Tyr or Ser and may involve

deprotonation of the tetrahedral intermediate by a general base. Finally, the acetylate product is released. Deprotonation of the acceptor substrate may also occur via ordered water molecules.

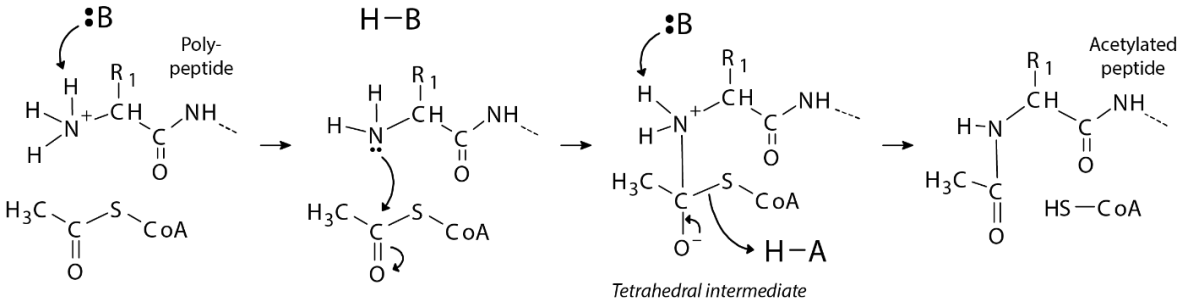


Fig. 9 General direct acetyl-transfer mechanism of GNAT superfamily members. B, general base (usually a Glu, Asp or Ser; H-A, general acid, usually Tyr or Ser). Adapted from (Salah Ud-Din, Tikhomirova, and Roujeinikova 2016)

1.10 *Objectives of this work*

The only major eukaryotic N-terminal acetyltransferase that has thus far eluded structural characterization is the heterotrimeric NatC complex, which consist of the catalytic subunit Naa30 and the two auxiliary subunits Naa35 and Naa38. It is the only known NAT that strictly requires two auxiliary subunits for Nt-acetylation *in vivo*. The known substrate specificity profile of NatC includes methionine N-termini followed by a hydrophobic amino acid or lysine and partially overlaps those of NatE and NatF (see above).

In order to obtain insights into the tertiary and quaternary assembly of NatC, its structure was solved by X-ray crystallography. To unravel the structural mechanism for substrate-specific recognition, additional structures of NatC in complex with its cognate substrates were determined. Independent biochemical studies with recombinant purified protein provided further insights into substrate specificity and catalysis, culminating into a reaction scheme for the catalytic mechanism of NatC.

Finally, this work aimed to characterize the interaction of NatC with the ribosome and confirm ribosome binding patches on the NatC surface by co-sedimentation assays.

2 Materials

2.1 *Instruments*

ÄKTA™FPLC	GE Healthcare Life Sciences, Chicago, IL, USA
ÄKTA™purifier	
ÄKTA™prime plus	
HiLoad™ 26/600 Superdex™ 200 pg	
Model M-110L Microfluidizer Processor	Microfluidics, Newton, MA, USA
NanoDrop 2000 Spectrophotometer	Thermo Fisher Scientific Inc., Waltham, MA, USA
Avanti® J-26 XP Centrifuge	Beckman Coulter Inc., Brea, CA, USA
J-Lite® JLA-8.1000 rotor	
Optima™ L-100K Ultracentrifuge	
Type 45 Ti rotor	
Optima™ Max-XP Tabletop Ultracentrifuge	
TLA-100 Fixed-Angle Rotor	
New Brunswick™ Innova® 44 Incubator	Eppendorf AG, Hamburg, Germany
and Excella E24® Incubator	
ImageQuant™ LAS 4000 imaging system	GE Healthcare Life Sciences, Chicago, IL, USA
Crystal Gryphon	Art Robbins Instruments, Sunnyvale, CA, USA
FORMULATOR® - Protein Crystallography Screen Builder	FORMULATRIX®, Bedford, MA, USA
ROCK IMAGER® - Protein Crystallization Imagers	
BioPhotometer®	Eppendorf AG, Hamburg, Germany
Infinite® M1000 Pro Microplate reader	Tecan Group Ltd., Männedorf, Switzerland
C1000 Touch™ Thermal Cycler	Bio-Rad Laboratories, Inc., Hercules, CA, USA
Gel Doc™ XR+ Gel Documentation System	Bio-Rad Laboratories, Inc., Hercules, CA, USA
PowerPac™ 300 Basic Power Supply	Bio-Rad Laboratories, Inc., Hercules, CA, USA
mircoflex™ LRF MALDI-TOF MS	Bruker Corporation, Billerica, MA, USA
MSP 96 target ground steel BC plate	

2.2 *Consumables and Kits*

innuPREP Plasmid Mini Kit 2.0	Analytik Jena AG, Jena, Germany
Monarch® PCR & DNA Cleanup Kit	New England Biolabs Inc., Ipswich, MA, USA
Monarch® DNA Gel Extraction Kit	
Whatman™ ME24 ST Sterile Cellulose Ester Membrane Filter, Pore Size: 0.2 µm, d= 4.7 cm	GE Healthcare Life Sciences, Chicago, IL, USA
CrystalQuick™ LP Plates, 96 Well (609171)	Greiner Bio-One International GmbH, Kremsmünster, Austria
Vapor diffusion crystallization (VDX™) plate with sealant	Hampton Research, Aliso Viejo, CA, USA
22 mm Thick Circular Cover Slides - siliconized	Jena Bioscience GmbH, Jena, Germany
Seed Bead Kit, PTFE	Hampton Research, Aliso Viejo, CA, USA
Mounted CryoLoop™ - 20 µm diameter nylon (0.05-0.1 mm, 0.1-0.2 mm, 0.2-0.3 mm loops)	Hampton Research, Aliso Viejo, CA, USA
NuPAGE™ 4-12% Bis-Tris Protein Gel	Thermo Fisher Scientific Inc., Waltham, MA, USA
NuPAGE™ 12% Bis-Tris Protein Gel	
NuPAGE™ MES SDS Running Buffer (20X)	
NuPAGE™ MOPS SDS Running Buffer (20X)	
PageRuler™ Plus Prestained Protein Ladder (10 to 250 kDa)	Thermo Fisher Scientific Inc., Waltham, MA, USA
Unstained Protein Molecular Weight Marker (14.4 to 116 kDa)	
GSTrap™ HP Column	GE Healthcare Life Sciences, Chicago, IL, USA
Ni Sepharose High Performance	
Amicon® Ultra-15 Centrifugal Filter Unit (50 kDa MWCO)	Merck KGaA, Darmstadt, Germany
Amersham™ Protran™ 0.45 µm nitrocellulose membrane	GE Healthcare Life Sciences, Chicago, IL, USA
Extra Thick Blot Filter Paper, Precut, 14 x 16 cm	Bio-Rad Laboratories, Inc., Hercules, CA, USA
SuperSignal™ West Dura Extended Duration Substrate	Thermo Fisher Scientific Inc., Waltham, MA, USA
ProteoMass™ Peptide and Protein MALDI-MS Calibration Kit	Sigma-Aldrich, Inc. St. Louis, MO, USA

2.3 *Chemicals*

Acetic acid, 100%	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Acetyl coenzyme A, sodium salt, *≥93.0%*	Sigma-Aldrich, Inc. St. Louis, MO, USA
4-(2-Aminoethyl)-benzolsulfonylfluorid (AEBSF) - Hydrochlorid BioChemica	AppliChem GmbH, Darmstadt, Germany
Agar-Agar, Kobe I	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Agarose for DNA Electrophoresis	SERVA Electrophoresis GmbH, Heidelberg, Germany
Ammonium chloride, ≥99.5 %	Sigma-Aldrich, Inc. St. Louis, MO, USA
Ammonium iodide, ≥99.0%, Honeywell™ Fluka™	Thermo Fisher Scientific, Waltham, MA, USA
Boric acid, ≥99.8 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Bromophenol blue sodium salt, for electrophoresis	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Calcium chloride dihydrate, ≥99.5%	Sigma-Aldrich, Inc. St. Louis, MO, USA
Citric acid anhydrous, cell culture tested	Sigma-Aldrich, Inc. St. Louis, MO, USA
Cobalt(II)-chloride hexahydrate, ≥99 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Coomassie® Brilliant Blue G 250	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Copper(II) sulfate pentahydrate, puriss.	Sigma-Aldrich, Inc. St. Louis, MO, USA
5,5′-Dithiobis(2-nitrobenzoic acid) (DTNB), ≥98 %	Sigma-Aldrich, Inc. St. Louis, MO, USA
1,4-Dithiothreitol (DTT), ≥99 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Dimethyl sulfoxide (DMSO)	New England Biolabs Inc., Ipswich, MA, USA
Ethylenediamine tetraacetic acid (EDTA) disodium salt dihydrate, ≥99 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Ethylene glycol, for analysis EMSURE®	Sigma-Aldrich, Inc. St. Louis, MO, USA
Formaldehyde, 16% (w/v), methanol-free, Pierce™	Thermo Fisher Scientific, Waltham, MA, USA
D(+)-Glucose monohydrate, for microbiology	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Glycerol, ≥99.5 % anhydrous	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Glycine, ≥99 %, for synthesis	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Guanidine hydrochloride, ≥99.5 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Imidazole, PUFFERAN® ≥99 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Iron(II) sulfate heptahydrate, ≥99 %	Sigma-Aldrich, Inc. St. Louis, MO, USA
L-Isoleucine, ≥98%	Sigma-Aldrich, Inc. St. Louis, MO, USA
Isopropyl-β-D-thiogalactopyranoside (IPTG), ≥99 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), ≥99.5%	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Hydrochloric acid, 32%, extra pure	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Kanamycin sulfate ≥750 I.U./mg, for biochemistry	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
L-Leucine, ≥99.5%	Sigma-Aldrich, Inc. St. Louis, MO, USA
L-Lysine crystallized, ≥98.0%	Sigma-Aldrich, Inc. St. Louis, MO, USA
Magnesium sulfate, anhydrous	Merck KGaA, Darmstadt, Germany
Magnesium chloride, hexahydrate, ≥99.0%	Sigma-Aldrich, Inc. St. Louis, MO, USA
Manganese(II) chloride, tetrahydrate, ≥99.0%	Sigma-Aldrich, Inc. St. Louis, MO, USA
2-Mercaptoethanol, ≥99.0%	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Methanol, ≥ 99.8 %	Th. Geyer GmbH & Co. KG, Renningen, Germany
L-Phenylalanine, ≥99.0%	Sigma-Aldrich, Inc. St. Louis, MO, USA
Poly(ethylene glycol) BioUltra, 4,000	Sigma-Aldrich, Inc. St. Louis, MO, USA
Ponceau S, for electrophoresis	Merck KGaA, Darmstadt, Germany
Potassium dihydrogen phosphate, ≥99 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
L-Selenomethionine, ≥97 %	Merck KGaA, Darmstadt, Germany
Silver nitrate solution 5% (w/v)	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Skim milk powder, for microbiology	Sigma-Aldrich, Inc. St. Louis, MO, USA
Sodium carbonate, anhydrous, ≥98 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Sodium chloride, ≥99.8 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Sodium citrate tribasic dihydrate, ≥99.5 %	Sigma-Aldrich, Inc. St. Louis, MO, USA
Sodium hydroxide, ≥98 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Sodium phosphate monobasic dihydrate, ≥99 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Sodium phosphate dibasic, ≥99 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Terrific broth, modified	Melford Laboratories Ltd., Chelsworth, UK
L-Threonine, ≥99.0%	Sigma-Aldrich, Inc. St. Louis, MO, USA
Trifluoroacetic acid (TFA), ≥99 %	Sigma-Aldrich, Inc. St. Louis, MO, USA
Tris-(hydroxymethyl)-aminomethan (TRIS), ≥99.9 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
Tween® 20, Ph. Eur.	Carl Roth GmbH + Co. KG, Karlsruhe, Germany
L-Valine, ≥99.0%	Sigma-Aldrich, Inc. St. Louis, MO, USA
Zinc sulfate, heptahydrate, ≥99.5 %	Carl Roth GmbH + Co. KG, Karlsruhe, Germany

2.4 *Antibodies*

Monoclonal ANTI-FLAG® M2 antibody, mouse	Sigma-Aldrich, Inc. St. Louis, MO, USA
Anti-RPL7 antibody (ab72550), mouse	Abcam plc, Cambridge, UK
Goat IgG anti-Mouse IgG (H+L)-HRPO, MinX (115-035-003)	DIANOVA GmbH, Hamburg Germany
Goat IgG anti-Rabbit IgG (H+L)-HRPO, MinX (111-035-003)	

2.5 *Enzymes*

Taq DNA polymerase with ThermoPol® Buffer	New England Biolabs Inc., Ipswich, MA, USA
Phusion® High-Fidelity DNA polymerase (2000 U/mL) with Phusion® HF Buffer	
Shrimp alkaline phosphatase (rSAP)	
T4 DNA Ligase with T4 DNA ligase Buffer	
Restriction Enzymes: BamHI-HF, MfeI-HF, NdeI, NotI, PvuI-HF, XhoI,	
PreScission® protease for GST-tag removal	GE Healthcare Life Sciences, Chicago, IL, USA
Benzonase® endonuclease, >90%, 25 U/µL	Merck KGaA, Darmstadt, Germany
DNase I, grade II, from bovine pancreas	Roche Diagnostics GmbH, Mannheim, Germany
Endoproteinase Glu-C (*S. aureus* V8)	Roche Diagnostics GmbH, Mannheim, Germany

2.6 *Bacterial strains*

E. coli BL21(DE3)	New England Biolabs Inc., Ipswich, MA, USA

2.7 *Plasmids and cDNA clones*

pRSFDuet-1™ *(Novagen)*	Merck KGaA, Darmstadt, Germany

Table 3 cDNA clones

Clone	Manufacturer
pRSFDuet-1™ *(Novagen)*	Merck KGaA, Darmstadt, Germany
***E. coli* clones from Dharmacon yeast ORF collection**	
S. cerevisiae NAA35 (MAK10) Clone ID: YEL053C Accession: NM_001178868.3 / NP_010861.3 Uniprot ID: Q02197	
S. cerevisiae NAA30 (MAK3) Clone ID: YPR051W Accession: NM_001184148.1 / NP_015376.1 Uniprot ID: Q03503	Dharmacon, a Horizon Discovery Group Co., Cambridge, UK
S. cerevisiae NAA38 (MAK31 / LSM9 / SMX1) Clone ID: YCR020C-A Accession: NM_001178734.1 / NP_009948.1 Uniprot ID: P23059	

2.8 *Peptides*

All peptides were synthesized by ProteoGenix (Schiltigheim, France) with a purity of ≥95% and with trifluoroacetic acid (TFA) exchanged for HCl. The first five amino acids of all peptides correspond to the N-terminal sequence of known NatA (SASEA), NatB (MDESV) or NatC (MFHLV, MYTLL and MLRFV) substrates as detailed in Table 4. The remaining C-terminal residues correspond to a short GS-linker followed by three arginine residues to increase solubility of the peptide.

Table 4 Peptides used for co-crystallization and acetyltransferase assays. [a]All peptides are decamers ending on the C-terminal five residues: GSRRR.

N-terminal five residues[a]	N-terminal sequence protein origin	Protein species	Uniprot-ID	MW (Da)
SASEA	Threonine-tRNA ligase (ThrRS)	*S. cerevisiae*	P04801	1076.1
MDSEV	Actin	*S. cerevisiae*	P60010	1192.3
MYTLL	ADP-ribosylation factor-related protein 1 (Arfrp1)	*H. sapiens*	Q13795	1252.5
MLRFV	Major capsid protein (Gag)	*S. cerevisiae virus L-A*	P32503	1277.65
MFHLV	ADP-ribosylation factor-like protein 3 (Arl3)	*S. cerevisiae*	Q02804	1258.5
Arl3 variants with aromatic amino acid substitutions at position 2				
MYHLV	Arl3 variant F2→Y	-	-	1274.61
MWHLV	Arl3 variant F2→W	-	-	1297.64
Arl3 variants with alanine substitutions at positions 2, 3 and 4				
MAHLV	Arl3 variant F2→A	-	-	1182.51
MFALV	Arl3 variant H3→A	-	-	1192.54
MFHAV	Arl3 variant L4→A	-	-	1216.41
Arl3 variants with single amino acid substitution to Gag peptide at position 2, 3 or 4				
MLHLV	Arl3 variant F2→L	-	-	1224.48
MFRLV	Arl3 variant H3→R	-	-	1277.65
MFHFV	Arl3 variant L4→F	-	-	1292.63

2.9 *DNA oligonucleotides*

DNA oligonucleotides were ordered from *Eurofins Genomics* at a synthesis scale of 0.01 - 0.05 µmol in High Purity Salt Free (HPSF) quality.

Table 5 Oligonucleotides for NatC expression vectors

Name	Sequence 5' → 3'	5' overhang
Naa30 Fw	GGC CGATCG AAGGAG ATATACC ATGGAAATAGTGTACAAGCCATTG	RBS+PvuI
Naa30 Rv	GGC CTCGAG TTA TGTGGCCAGCCGGCC	XhoI
Naa30ΔC17 Rv	GGC CTCGAG TCATTA AGTCAACGGCAATATCAGTTTAAAG	XhoI
Naa35 Fw	GGC GGATCC G CTGGAAGTTCTGTTCCAGGGGCCC ATGGAAGTAGACAGTATATTAG	BamHI-HRV 3C
Naa35ΔN17 Fw	GGC GGATCC G CTGGAAGTTCTGTTCCAGGGGCCC CAATTGGTTGACGTCACAAG	BamHI-HRV 3C
Naa35ΔN44 Fw	GGC GGATCC G CTGGAAGTTCTGTTCCAGGGGCCC GATTTGTTCGAGGGTACAC	BamHI-HRV 3C
FLAG-Naa35 Fw	GGC GGATCC G CTGGAAGTTCTGTTCCAGGGGCCC GATTATAAAGATGATGATGATAAA ATGGAAGTAGACAGTATATTAG	BamHI-HRV 3C-FLAG
Naa35 Rv	GGC GCGGCCGC TTATTTATAGCGGTCTTGC	NotI
Naa38 Fw	GGC CATATG GACATCTTGAAACTGTC	NdeI
Naa38 Rv	GGC CAATTG CTAAACAATATTAGCCATCAATTC	MfeI
Naa38ΔC11 Fw	CTGCAGGAGCTTACTTAGCAATTGGATATC	Overlap for Site-directed mutagenesis
Naa38ΔC11 Rv	GATATCCAATTGCTAAGTAAGCTCCTGCAG	

Table 6 Oligonucleotides for Naa35 ribosome binding mutants

Name	Sequence 5' → 3'
Naa35 subunit mutants	
K500A_K501A_ K503A_K504A_Fw	GAACGCAGCACTTGCAGCATTGAAAGCTGGTGAAAAG
K500A_K501A_ K503A_K504A_Rv	GCTGCAAGTGCTGCGTTCATGGAATGAATGGC
K511A_R515A_R519A_Fw	GACCAATTGGCGCTAAAGTACGCGTTTGCTATGGATAATGAAATG
K511A_R515A_R519A_Rv	CTTTAGCGCCAATTGGTCGGCCTTTTCACCAGCTTTCAAT
R636A_Fw	CTTTATCGAAACTGAAGTGGCCAATGTCGTTAGTTCTATTG
R636A_Rv	CAATAGAACTAACGACATTGGCCACTTCAGTTTCGATAAAG
K681A_K688A_Fw	GCATTATCAGTCAACTCAGCAGTTATCGTCAACACTTTG
K681A_K688A_Rv	CTGAGTTGACTGATAATGCTGCAGAAGTATGTTTTAGTTTACAATAG
K696A_K699A_K701A_Fw	GCGAACATTGCAAACGCGGACTCCCACGAATACAAG
K696A_K699A_K701A_Rv	CGCGTTTGCAATGTTCGCTTTCAAAGTGTTGACGATAAC

Table 7 Oligonucleotides for Naa30 and Naa35 mutants

Name	Sequence 5' → 3'
Naa30 subunit mutants	
L27A_Fw	GTATTAAAAAACTAATAGATGCGGATGCATCAGAGCCGTACTCGATATAC
L27A_Rv	GTATATCGAGTACGGCTCTGATGCATCCGCATCTATTAGTTTTTTAATAC
S28A_Fw	CTAATAGATGCGGATCTAGCAGAGCCGTACTCGATATAC
S28A_Rv	GTATATCGAGTACGGCTCTGCTAGATCCGCATCTATTAG
E29A_Fw	GGATCTATCAGCGCCGTACTCG
E29A_Rv	CGAGTACGGCGCTGATAGATCC
E29Q_Fw	GGATCTATCACAGCCGTACTCG
E29Q_Rv	CGAGTACGGCTGTGATAGATCC
Y31F_Fw	GGATCTATCAGAGCCGTTCTCGATATACGTATATAGGTATTTC
Y31F_Rv	GAAATACCTATATACGTATATCGAGAACGGCTCTGATAGATCC
Y80A_Fw	CGTGAGACTGAGAGGAGCTATAGGAATGCTAGCCGTAG
Y80A_Rv	CTACGGCTAGCATTCCTATAGCTCCTCTCAGTCTCACG
Y80F_Fw	GACTGAGAGGATTTATAGGAATGC
Y80F_Rv	GCATTCCTATAAATCCTCTCAGTC
E118A_Fw	GATCATGTTAGCAACAGAGGTGG
E118A_Rv	CCACCTCTGTTGCTAACATGATC
E118Q_Fw	GATCATGTTACAAACAGAGGTGG
E118Q_Rv	CCACCTCTGTTTGTAACATGATC
E120A_Fw	GTTAGAAACAGCGGTGGAAAATTC
E120A_Rv	GAATTTTCCACCGCTGTTTCTAAC
E120Q_Fw	GTTAGAAACACAGGTGGAAAATTC
E120Q_Rv	GAATTTTCCACCTGTGTTTCTAAC
Y130A_Fw	GCGGCTCTAAACTTAGCTGAAGGAATGGGTTTCATC
Y130A_Rv	GATGAAACCCATTCCTTCAGCTAAGTTTAGAGCCGC
Y130F_Fw	GCGGCTCTAAACTTATTTGAAGGAATGGGTTTCATC
Y130F_Rv	GATGAAACCCATTCCTTCAAATAAGTTTAGAGCCGC
Y145F_Fw	GAAACGGATGTTCCGCTACTTCTTGAATGAAGGGGACG
Y145F_Rv	CGTCCCCTTCATTCAAGAAGTAGCGGAACATCCGTTTC
Naa35 subunit mutants	
F47A_Fw	GATCCCAGGTTTGATTTGGCCGAGGGTACACATTCTTTAG
F47A_Rv	CTAAAGAATGTGTACCCTCGGCCAAATCAAACCTGGGATC
H51A_Fw	GATTTGTTCGAGGGTACAGCTTCTTTAGAGGTCAACAACTC
H51A_Rv	GAGTTGTTGACCTCTAAAGAAGCTGTACCCTCGAACAAATC
K59A_Fw	CTTTAGAGGTCAACAACTCCGCATTAGACTCCAGTCTGATAG
K59A_Rv	CTATCAGACTGGAGTCTAATGCGGAGTTGTTGACCTCTAAAG
E171A_E172A_Fw	GAGTGCTGGTATTGTTTTTGCAGCGGAGGATTTAAATTTTAATAAC
E171A_E172A_Rv	GTTATTAAAATTTAAATCCTCCGCTGCAAAAACAATACCAGCACTC

2.10 *Growth media*

Lysogeny broth (LB) medium was used for *E. coli* precultures and during transformations for cell recovery after heat shock treatment. LB medium consisted of 0.5% (w/v) yeast extract, 1% (w/v) tryptone, and 1 % (w/v) NaCl. TB medium (*Melford Laboratories Ltd.*) was used for *E. coli* expression cultures and consisted of 12 g/L casein digest peptone, 24 g/L yeast extract, 9.5 g/L K_2HPO_4, 2.2 g/L KH_2PO_4 and was supplemented with 0.4% (v/v) glycerol. M9 minimal medium was used for expression of Selenomethionine-labelled NatC and contained the components listed in Table 8 to Table 10.

Table 8 M9 minimal medium. Components were added according to the table order (i.e. top to bottom).

Volume	Solution	Component	c (final)
~865.7 mL	ddH_2O (autoclaved)		
		Na_2HPO_4	100 mM
100 mL	10 × M9 salts	KH_2PO_4	22 mM
		NaCl	8.56 mM
20 mL	20% (w/v) glucose (filter sterilized)	“	0.4% (w/v) ≙ 4 g/L
2 mL	1M $MgSO_4$ (filter sterilized)	“	2 mM
2 mL	5.0 g /20 mL NH_4Cl (filter sterilized)		0.5 g/L ≙ 9.35 mM
300 µL	1M $CaCl_2$		0.3 mM
		EDTA, anhydrous	171 µM
		Fe_2SO_4	33 µM
		$ZnCl_2$ / $ZnSO_4$	3.7 µM
10 mL	100× Trace elements	$CoSO_4$ / $CoCl_2$	0.64 µM
		$CuCl_2$ / $CuSO_4$	0.74 µM
		H_3BO_3	1.62 µM
		$MnCl_2$	81 µM

Table 9 M9 salt solution (10× concentrate, pH 7.0-7.4)

Salt / (substitute)	MW (g/mol)	β	c (10× conc.)
Na_2HPO_4	141.96	142 g/L	1 M
KH_2PO_4	136.09	30 g/L	220 mM
NaCl	58.44	5 g/L	85.6 mM

Table 10 Trace element solution (100× concentrate)

Salt / (substitute)	MW (g/mol)	β	c (100× conc.)
EDTA (anhydrous)	292.24	5 g/L	17.1 mM
$Fe_2SO_4 \times 7\ H_2O$	278.01	915.0 mg/L	3.3 mM
$ZnSO_4 \times 7\ H_2O$	287.56	105.5 mg/L	0.37 mM
$CoCl_2 \times 6\ H_2O$	237.93	15.3 mg/L	64 µM
$CuSO_4 \times 5\ H_2O$	249.69	18.6 mg/L	74 µM
H_3BO_3	61.83	10 mg/L	162 µM
$MnCl_2 \times 4\ H_2O$	197.91	1.6 g/L	8.1 mM

2.11 *Software*

Software	Manufacturer
flexAnalysis3.4 flexControl3.4	Bruker Corporation, Billerica, MA, USA
Rock maker	Formulatrix, Bedford, MA, USA
PyMOL, Version 2.3.1	Schrödinger, LLC.
ImageJ, Version 1.52a	(Schneider, Rasband, and Eliceiri 2012)
Adxv, Version 1.9.14	https://www.scripps.edu/tainer/arvai/adxv.html
Albula	Dectris Ltd., Baden-Daettwil, Switzerland
Phenix, Version 1.13-2998	Adams et al. (2010)
Phenix.refine	(Afonine et al. 2012)
Phaser	(McCoy et al. 2007)
Geneious® 11.0.2	(Kearse et al. 2012)
MXCuBE2	(Oscarsson et al. 2019)
Coot/WinCoot, Version 0.8.9.1	(Emsley et al. 2010)
iMOSFLM, Version 7.2.2	(Battye et al. 2011)
XDS, Version Mar 15, 2019	(Kabsch 2010)
HKL2Map (graphical user interface)	(Pape and Schneider 2004)
SHELXC/D/E	(Sheldrick 2010)
CCP4 suite	(Winn et al. 2011)
Freerflag	(Brunger 1992)
Buccaneer	(Cowtan 2006)
pdbcur	Eugene Krissinel, European Bioinformatics Institute, Cambridge, UK. Martyn Winn, Daresbury Laboratory, UK - some additional keywords.

3 Methods

3.1 *Cloning of the NatC complex*

3.1.1 PCR reactions

All routine PCR reactions were performed with the *Phusion High-Fidelity* DNA polymerase due to its 50× higher fidelity compared to *Taq* DNA polymerase.[4] The final concentrations in a 50-µL reaction were: 1 U/50 µL *Phusion* DNA polymerase, 1× *Phusion HF* buffer, 200 µM dNTPs, 0.5 µM forward and reverse primer, 3% (v/v) DMSO and 20 pg/µL template DNA. The thermocycling conditions are displayed in Table 11.

Site directed mutagenesis (SDM) was performed to make specific, targeted changes in double stranded plasmid DNA, including insertions, deletions and substitutions of DNA. For single base-pair substitutions, the same reaction mix as for the routine PCR was used in a 20-µL reaction, except with a 100× higher final template DNA concentration of 2 ng/µL. For SDM reactions that aimed to substitute multiple subsequent base pairs, a two-step protocol was used. The reaction mix was split into 2×10 µL and 0.5 µM of either forward or reverse primer were added to the separate reactions. The thermocycler protocol was run for three cycles, after which the individual reactions were combined, and the protocol was continued for another 15 cycles.

Colony PCR was used to determine the presence or absence of insert DNA in plasmid constructs. The final concentrations in a typical 20-µL reaction were: 0.5 U/20 µL *Taq* DNA polymerase[5], 1× *ThermoPol* buffer, 200 µM dNTPs, 0.2 µM forward and reverse primer and 3 % (v/v) DMSO. A single colony was slightly touched with a sterile pipette tip, which was then resuspended in a PCR tube containing the complete PCR reaction mix.

Table 11 Thermocycling conditions for different PCR applications

	Routine PCR	Site-directed mutagenesis	Colony-PCR
DNA Polymerase	*Phusion*	*Phusion*	*Taq*
Initial Denaturation	98 °C, 30 s	98 °C, 30 s	95 °C, 5 min
Denaturation	98 °C, 10 s	98 °C, 10 s	95 °C, 20 s
Annealing	45-72 °C, 20 s	45-72 °C, 20 s	45-68 °C, 30 s
Extension	72 °C, ~ 30-60 s (30 s/ 1000 bp)	72 °C, ~ 3 min (30 s/ 1000 bp)	68 °C, ~1-5 min (1 min/1000 bp)
Number of cycles	30	30 (one-step variant) 3 and 15 (two-step variant)	25
Final extension	72 °C, 5 min	72 °C, 8 min	68 °C, 5 min

[4] https://international.neb.com/products/m0530-phusion-high-fidelity-dna-polymerase (*Phusion* DNA polymerase is provided at 2000 U/µL. 1 U (unit) is defined for Phusion DNA Polymerase as the amount of enzyme that will incorporate 10 nmol of dNTP into acid insoluble material in 30 min at 74 °)

[5] *Taq* DNA polymerase is provided at 5000 U/µL. 1 U is defined as the amount of enzyme that will incorporate 15 nmol of dNTP into acid insoluble material in 30 min at 75 °C (https://international.neb.com/products/m0267-taq-dna-polymerase-with-thermopol-buffer)

3.1.2 Enzymatic assembly of DNA using restriction digestion and ligation

Classical enzymatic assembly using restriction digestion and ligation was used to insert DNA into expression vectors.

Both the plasmid DNA and the insert were digested with the same set of restriction enzymes. A 50 µL restriction digestion reaction was composed of 1 µg DNA, 1× Restriction buffer (typically NEB *CutSmart* buffer), and 1 µL of the selected restriction enzymes. The DNA was digested for 1-2 h at 37°C. Afterwards 1 µL of shrimp alkaline phosphatase was added to the plasmid digestion reaction to dephosphorylate the plasmid for another 1 h at 37°C. Both the digested plasmid DNA and the digested, dephosphorylated vector DNA were purified with the *Monarch® PCR & DNA Cleanup Kit.*

Afterwards purified insert and vector DNA were ligated with T4 DNA ligase. A 20 µL ligation reaction was composed of : 1× T4 DNA ligase buffer, 1 µL of T4 DNA ligase, 100 ng of vector DNA and a 5-fold molar excess of insert DNA. Ligations were carried out at room temperature for 1 h.

3.1.3 Cloning of NatC constructs

The first structures of the heterodimeric NatA (Liszczak et al. 2013) and NatB (Hong et al. 2017) complexes were obtained using bacterially expressed orthologous proteins from the yeast species *Schizosaccharomyces pombe* or *Candida albicans*, respectively. Thus, the yeast model organism *Saccharomyces cerevisiae* was chosen for a biochemical and structural characterization of the heterotrimeric NatC complex. In case of the NatA complex (Liszczak et al. 2013), a C-terminal truncation of a non-conserved regions of its catalytic subunit Naa10 led to a soluble, active complex, that could be purified and crystallized. A sequence alignment of Naa30 orthologs (Fig. S1) revealed a similar non-conserved region of seventeen amino acids and thus Naa30 was expressed as the C-terminally truncated construct Naa30ΔC17.

Genes encoding *S. cerevisiae* NatC subunits Naa30 (Uniprot ID: Q03503), Naa35 (Q02197) and Naa38 (P23059) were obtained from *Dharmacon* yeast ORF collection clones YPR051W, EL053C and YCR020C-A, respectively.

The NatC complex, containing Naa30ΔC17 (residues 1-159), full-length Naa35 (residues 1-733) and Naa38 (residues 1-88) was cloned into the pRSFDuet-1 expression vector (Novagen), which contains two multiple cloning sites (MCS), each of which is preceded by a T7*lac* promotor and ribosome binding site (RBS). The genes encoding full-length Naa38 and Naa30ΔC17, were both cloned into the second multiple cloning site (MCS) of the pRSFDuet-1 vector. Naa38 was inserted using the restriction sites for NdeI and MfeI. Naa30ΔC17 was cloned behind Naa38, between the restriction sites for PvuI and XhoI. An additional RBS with a 7 bp spacer (AAGGAGATATACC) was added in front of the ATG start codon for Naa30ΔC17. DNA encoding full-length Naa35, preceded by the human rhinovirus 3C (HRV 3C) protease cleavage site (LEVLFQGP) was cloned in frame with the

sequence for the His-Tag (MGSSHHHHHHSQDP) in the first MCS of pRSFDuet-1, using the restriction sites for BamHI and NotI. For NatC-ribosome co-sedimentation assays, a Naa35 was cloned in the same manner, but with an additional sequence coding for a FLAG-tag (DYKDDDDK), inserted in between the HRV 3C coding sequence and the start codon.

During purification of the NatC complex, a partial (proteolytic) degradation of the Naa38 subunit was observed. Using MALDI-MS, a mass difference of 1183 Da between the complete and partially degraded Naa38 protein species was determined. This difference corresponds to the C-terminal eleven residues (ANKVELMANIV) of Naa38. In order to ensure a homogenous NatC complex preparation, the Naa38 gene was shortened by 33 bp through site directed mutagenesis to encode for the truncated Naa38ΔC11 protein.

All further NatC complex constructs contain Naa30ΔC17 (residues 1-159), full-length Naa35 (residues 1-733) and Naa38ΔC11 (residues 1-77), which will be henceforth referred to as the "wild type" (WT) NatC complex.

Two wild type NatC complexes with N-terminally truncated Naa35 subunits were created by replacing the full-length Naa35 gene with the coding sequence for Naa35ΔN17 or Naa35ΔN44 by restriction cloning using the BamHI and NotI restriction sites.

All NatC mutant constructs used in the acetyltransferase assay and the NatC-ribosome co-sedimentation assay are based on the WT NatC complex and were created by site-directed mutagenesis.

3.2 *Expression and purification of the NatC complex*

3.2.1 Preparation of chemically competent E. coli cells

Chemically competent *E. coli* BL21(DE3) cells were prepared for use in the subsequent heat shock transformations (section 3.2.2). A single colony (plated on LB + 2% agar) was used to inoculate a 5 mL LB culture and incubated overnight at 37°C on a rotary shaker. A 500 mL Erlenmeyer flask with 100 mL of LB medium was inoculated with 1 mL of the overnight culture and grown at 37°C, 250 rpm on a rotary shaker. After cells reached an OD_{600} of 0.6 to 0.8 they were transferred into two 50 mL falcon tubes and incubated on ice for 10 min. Afterwards the cell cultures were centrifuged for 10 min at 2600×*g* at 4°C and the cells were resuspended in 10 mL ice cold 0.1 M $CaCl_2$/10% glycerol solution. After 15 min incubation on ice, the cells were centrifuged as before. The cell pellets were resuspended in 1 ml ice cold 0.1 M $CaCl_2$/10% glycerol and 50 µL aliquots were prepared and transferred to -80°C.

3.2.2 Transformation of chemically competent E. coli BL21(DE3)

For the transformation of chemically competent *E. coli* BL21(D3) cells, 10 µL of a ligation reaction or 1 µL purified vector DNA was added to 50 µL competent cells and incubated on ice for 30 min.

The cells were heat shocked at 42°C for 45 s and immediately put back on ice for 5 min. 300 µL LB medium was added and the cell suspension was incubated for 1 h at 37°C, shaking at 1000 rpm. Afterwards the cells were centrifuged at 3000×*g* for 2 min, resuspended in 50 µL LB medium and plated on LB agar plates with 50 µg/mL Km. Agar plates were incubated overnight at 37 °C.

3.2.3 Small-scale IPTG induction test

A small-scale protein expression test was performed to screen several colonies in order to verify protein expression in different bacterial clones, to exclude clones that show no protein expression or to select clones that show a higher protein expression compared to other cultivated clones. Several single colonies were used to inoculate separate 5 mL LB medium cultures in 15 mL conical tubes. Cultures were grown to an optical density (OD_{600}) of ~0.6 and then split into 2× 2.5 mL in separate 15 mL conical tubes. 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to one of the fractions (*+IPTG* culture) incubated together with the non-induced fraction (*-IPTG* culture) for another 2 h at 37°C. 30 µL sample was taken from both *-IPTG* and *+IPTG* cultures, mixed with 10 µL 4× SDS sample buffer (250 mM TRIS-HCl, pH 6.8, 10% (w/v) SDS, 30% glycerol, 10% (v/v) 2-mercaptoethanol, 0.5% (w/v) bromophenol blue) and loaded on an SDS-PAGE.

After Coomassie staining, lanes with *-IPTG* and *+IPTG* samples were analyzed visually for the presence of a protein band in the IPTG-induced sample at the expected molecular weight. The intensity of the recombinant protein band was compared between different bacterial clones. The clone that showed the strongest protein expression was selected and a glycerol stock was prepared by mixing 500 µL of glycerol with 1000 µL of the non-induced culture and stored in a cryogenic vial at -80 °C.

3.2.4 Expression of the native NatC complex

A pre-culture with 80 ml LB medium and 50 µg/mL kanamycin was inoculated by transferring ~ 50 µL frozen glycerol stock with a sterile pipette tip. The culture was incubated overnight at 37 °C in a 250 mL Erlenmeyer flask, shaking at 300 rpm.

Protein expression was performed in 6×1 L TB medium with 50 µg/mL kanamycin, which was inoculated with 1% (v/v) of the preculture. Expression cultures were grown in 5L baffled Erlenmeyer flasks at 37°C, shaking at 110 rpm. At an OD_{600}=0.6 the temperature was lowered to 18°C and cultures were induced with 100 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated overnight.

Bacterial cells were pelleted for 15 min at 4500 rpm (~3876 ×*g*) with a *JLA-8.1000* rotor in an *Avanti J-26 XP* centrifuge (*Beckman Coulter*).

3.2.5 Expression of the selenomethionine-labeled NatC complex

For the expression of selenomethionine-labeled proteins, a modified protocol from Van Duyne et al. (1993) was used, which relies on the feedback inhibition of methionine biosynthesis through addition of lysine, phenylalanine, threonine, isoleucine, leucine and valine, which are supplied together with selenomethionine to a M9 minimal medium culture before induction of protein expression (Doublie 1997).

A 5 mL LB medium preculture (with 50 µg/mL kanamycin) was inoculated from an *E. coli* BL21(DE3) glycerol stock carrying the pRSFDuet-1_NatC expression plasmid and grown at 37°C for ~6h. An overnight culture with 100 mL LB-medium (and 50 µg/mL Kan) was inoculated with 1 mL of the preculture and grown at 37° in a 500 mL Erlenmeyer Flask. After approx. 24 h, 6×1L pre-warmed, *M9 minimal medium* cultures were inoculated with 10 ml or 1% (v/v) of the overnight culture and grown at 37°C, 110 rpm. Final 1× *M9 minimal medium* was prepared by adding 1× *M9 salt solution*, 4 g/L glucose, 2 mM $MgSO_4$, 0.5 g/L NH_4Cl, 0.3 mM $CaCl_2$, 1× *trace elements solution* and antibiotics (i.e. 50 µg/mL kanamycin) to autoclaved ddH_2O in the aforementioned order and as detailed in Table 8.

The composition of *10× M9 salt solution* and *10× trace element solution* is described in Table 9 and Table 10, respectively.

At mid-log phase ($OD_{600} \approx 0.6$) the amino acid mixture detailed in Table 12 was added to all minimal medium cultures, which grew for another 15 min before induction with 100 µM isopropyl β-D-1-thiogalactopyranoside (IPTG). The temperature was lowered to 18°C and the cultures were incubated overnight before harvesting the cells on the following day. The success of selenomethionine incorporation was examined by MALDI-MS. Selenomethionine-labeled NatC was purified, exactly like the native complex, but with the addition of 2.5 mM 2-mercaptoethanol to all purification buffers due to a higher oxidation sensitivity of selenium compared to sulfur, which can result in less stable crystals (Walden 2010).

Table 12 **Amino acid supplement** (added per 1L of 1× M9 Minimal medium)

Mass	Amino acid
100 mg	L-lysine
100 mg	L-phenylalanine
100 mg	L-threonine
50 mg	L-isoleucine
50 mg	L-leucine
50 mg	L-valine
50 mg	L-selenomethionine

3.2.6 Purification of the NatC complex

All subsequent purification steps were performed with cooled buffer solutions (~ 4–10 °C) or on ice. All buffer solutions were supplemented with 2.5 mM 2-mercaptoethanol, unless the proteins were intended for use in the colorimetric DTNB assay, in which case 2-mercaptoethanol was omitted from all buffers, due to the reaction of its thiol groups with the DTNB dye. Pellets were resuspended in a total of 120 mL Lysis Buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 10 mM Imidazole) and supplemented with 2 mL DNAse I solution (1 mg/mL in 1 M $MgCl_2$, pH 6.5) and a spatula tip of 4-(2-Aminoethyl)-benzolsulfonylfluorid hydrochloride (AEBSF, BioChemica). The cell suspension was lysed by passing it twice through a cell homogenizer (*Model M-110L Microfluidizer Processor, Microfluidic*s). Afterward the cell lysate was centrifuged for 45 min at 35,000 rpm (~95834 ×*g*) in a Type 45 Ti rotor with a *Optima*™ L-100K Ultracentrifuge (*Beckman Coulter*) and the supernatant was passed through a 0.2 µm membrane filter (GE healthcare). The clarified supernatant (~ 200 mL) was supplemented with 1 mM $MgCl_2$ (final concentration) and 40 µL *Benzonase* endonuclease (*Merck*) and incubated for ~ 30 min. Afterwards 6 mL resuspended nickel affinity agarose (GE Healthcare) (corresponding to ~ 3mL sedimented beads) was added to the supernatant and incubated for another 15 min. The supernatant, containing the nickel beads was transferred in 4×50 mL falcon tubes and centrifuged for 5 min at 500 × *g* and the supernatant was discarded. Nickel beads were washed by adding 50 mL Lysis buffer to each falcon, followed by centrifugation and discarding the supernatant. The washing step was repeated once more before finally resuspending the Ni^{2+} beads in 10 mL Lysis buffer and transferring them into a glass column. Now beads were washed with 100 mL Wash Buffer (20 mM HEPES, pH 7.5, 300 mM NaCl, 25 mM Imidazole) using a peristaltic pump. The His-tagged NatC complex was eluted with Elution Buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 60 mM Imidazole). Buffer was exchanged with SEC buffer (20 mM HEPES, pH 7.5, 150 mM NaCl) using a centrifugal filter unit (Amicon Ultra-15, 50 kDa MWCO, Merck) to bring imidazole concentration below 1 mM. The 6×HIS-tag at the N-terminus of the Naa35 subunit of NatC was removed by cleavage at the human rhinovirus 3C (HRV 3C) protease cleavage site using *PreScission* protease (GE Healthcare). *PreScission* protease was added at a ratio of 1:100 (w/w) together with 10 µL *Benzonase* endonuclease and 1 mM $MgCl_2$ and incubated overnight at 4°C. The untagged NatC complex was loaded again on a Ni^{2+} column followed by a GSTrap column (GE Healthcare) to remove the GST-tagged *PreScission* protease. The column was washed with 100 mL SEC-Buffer and the untagged NatC complex was eluted with Lysis Buffer. The eluate was concentrated to ~ 2mL and loaded on a 26/600 Superdex 200 size exclusion column, equilibrated in SEC buffer. NatC complex-containing elution fractions were concentrated to ~ 10 mg/mL and snap frozen as 30-50 µL aliquots in liquid nitrogen. NatC concentrations were determined photometrically, using the sum of the molecular weights and molar extinction coefficients of all three subunits, as calculated with the *ProtParam* tool[6] (ExPASy). For example, the HRV 3C digested NatC WT complex has a computed total molar extinction coefficient (assuming all cysteines are reduced) of 103,600 M^{-1} cm^{-1} (with

[6] https://web.expasy.org/protparam/

Naa30dC17 = 21,890 M^{-1} cm^{-1}, Naa35, cut = 81,710 M^{-1} cm^{-1} and Naa38 = 0 M^{-1} cm^{-1}) and a total molecular weight of ≈111534 Da (with 18537 Da, 84457 Da and 8541 Da for Naa30dC17, Naa35 and Naa38ΔC17, respectively).

3.3 *MALDI-TOF mass spectrometry*

A Matrix assisted laser desorption/ionization (MALDI) mass spectrometer with a time-of-flight (TOF) analyzer was used to verify the mass of purified protein constructs, to analyze the products of a partial proteolytic digest and to verify the N-terminal acetylation of NatC substrate peptides.

Protein or peptide samples were diluted in 0.1% trifluoroacetic acid (TFA) to a final concentration of 10 µM. Similarly, up to five standards proteins or peptides were combined in 0.1% TFA with final concentrations ranging between 0.5-10 µM as indicated in Table 13. MALDI matrices were prepared by adding saturating amounts of *α-Cyano-4-hydroxycinnamic acid* (CHCA) or *Sinapinic acid* to a small volume (e.g. 50 µL) of 50% acetonitrile/0.05% TFA. Equal volumes (e.g. 2 µL + 2 µL) of prepared protein or peptide samples were mixed with saturated solution of the recommended matrix (see last column of Table 13). 0.5 µL of the sample/matrix mixture was spotted on a MALDI target plate (MSP 96 target ground steel BC plate, Bruker). Samples were allowed to dry completely, before loading the target plate into a mircoflex™ LRF MALDI-TOF mass spectrometer (Bruker). Peptides samples of 3495 Da (Insulin) or smaller were analyzed in Reflectron/Positive ion mode (RP_PepMix), whereas larger protein samples were analyzed in Linear/Positive ion modes LP_12kDa, LP_44kDa or LP_66kDa depending on the maximum molecular weight to be analyzed. The instrument was calibrated with the peptide or protein standard, before measuring unknown samples. Data was analyzed using the flexAnalysis3.4 software (Bruker).

Table 13 Standard peptide and proteins for MS. Monoisotopic (mono) and average (ave) molecular weights and recommended matrices. CHCA, α-Cyano-4-hydroxycinnamic acid.

Standard peptide / protein	Final conc. (µM)	$(M+H)^+$ (Da) monoisotopic (mono) or average (ave)	Recommended Matrix
Bradykinin fragment 1-7	1.5	757.3997 (mono)	CHCA
Angiotensin II (human)	1	1,046.5423 (mono)	CHCA
$P_{14}R$ (synthetic peptide)	1	1,533.8582 (mono)	CHCA
ACTH fragment 18-39 (human)	0.5	2,465.1989 (mono)	CHCA
Insulin oxidized B chain (bovine)	2 to 5	3,494.6513 (mono)	CHCA
Insulin (bovine)	1	5,735 (ave)	CHCA or sinapinic acid
Cytochrome c (equine)	1	12,362 (ave)	CHCA or sinapinic acid
Apomyoglobin (equine)	1	16,952 (ave)	CHCA or sinapinic acid
Aldolase (rabbit muscle)	2	39,212 (ave)	sinapinic acid
Albumin (bovine serum)	10	66,430 (ave)	sinapinic acid

3.4 *Purification of yeast 80S ribosomes*

Yeast 80S ribosomes, used in the NatC-ribosome sedimentation assay, were purified by Ciara Lally[7]. The ribosome purification protocol was adapted from Magin et al. (2017).

Saccharomyces cerevisiae strain YA2488 was plated-out on YPD-agar plates (1% yeast extract, 2% pepton, 2% agar, 2% glucose), and used to inoculate a 20 mL pre-culture of YPD (1% yeast extract, 2% peptone, 2% glucose), which was incubated at 30ºC for 7 h. 8 L of YPD was inoculated with preculture to an OD_{600} of 0.001 and grown at 30ºC to an OD_{600} of 2 (stationary phase).

The cells were pelleted at 8000 ×g, 5min, 4ºC, (Beckman Rotor JLA9.1000), resuspended in YP medium (1% yeast extract, 2% peptone, no glucose) and incubated for 10 min, 250 rpm, 30ºC, to ensure that all ribosomes would be in the "apo" form, without nascent chain or tRNA.

The following cell lysis steps were all carried out at 4ºC: Cells were pelleted by centrifugation at 8000 ×g, 5 min, 4ºC, (Beckman Rotor JLA9.1000) and resuspended in Buffer A (30mM HEPES, pH 7.5, 50 mM KCl, 10 mM $MgCl_2$, 8.5% sorbitol, 2 mM DTT, 0.5 mM EDTA). Cells were centrifuged again at 8000 ×g, 5 min, 4ºC, (Beckman Rotor JLA9.1000), the pellet was weighed and resuspended in lysis buffer (Buffer A + one cOmplete mini EDTA-free protease inhibitor tablet (Roche 11836170001), 1 Unit/µl RNAsin (N261B, Promega), 800 µg/ml heparin) to a final concentration of 200% (w/v). Glass beads (G8772 sigma) were added to the resuspension (such that there was an equal weight of glass beads to pellet), and lysis was performed by vortexing for 30 s, followed by 30 s on ice, repeated 4 times. To obtain the yeast lysates, the resuspension was centrifuged at 9000 ×g, 10 min, 4ºC (rotor Beckman TA-10.250), and the supernatant was collected. The absorbances at 260 nm, and 280 nm were measured.

To isolate the 80S monosomal ribosome fraction, yeast lysate (A_{260} = 200) was underlaid with 1 ml 30% sucrose in 80S Buffer (20 mM HEPES, pH 7.6, 100 mM potassium acetate, 5mM $MgCl_2$, 2 mM DTT) and centrifuged 70000 ×g, 18 h (Beckman ultracentrifuge, rotor mla80). The pellet was collected, and added at a concentration of A_{260} = 100, to the top of a 15-30% sucrose gradient, which was prepared by underlying 15% sucrose in 80S buffer with 30% sucrose in 80S buffer, and mixing (Gradient Master, BIOCOMP, short settings). The gradients were then spun at 23318 rpm, 6 h, using a SW32 Ti Beckman swing-out rotor, and then loaded to a polysome collector (LKB), which allowed isolation of the 80S peak, via monitoring the absorbance at 260 nm, at increasing levels of sedimentation.

Fractions containing 80S monosomes were pooled and centrifuged at 43000 rpm, 14 h (Beckman MLA-80 rotor). The pellet containing the 80S yeast ribosomes were resuspended in 20 mM HEPES pH 7.6, 50 mM potassium acetate, 5 mM $MgCl_2$, 2 mM DTT to give a final concentration of 2 µM.

[7]Ciary Lally, Phd, Institute of Medical Physics and Biophysics - Charité, Charitéplatz 1, 10117 Berlin

3.5 *NatC-ribosome sedimentation assay*

A sedimentation assay was performed to analyze the interaction between NatC and the yeast ribosome. The assay was adapted from the protocol described by Magin, Deng et al. (2017). NatC and ribosomes were combined to final concentrations of 1 µM and 0.8 µM, respectively in 80S Buffer (20 mM HEPES, pH 7.5, 50 mM KOAc, 5 mM $Mg(OAc)_2$, 2 mM DTT, 0.003% (v/v) Tween-20). Tween-20 was added to reduce unspecific binding of NatC to plastic surfaces. 35 µL of the NatC/ribosome solution was underlaid with 80 µL of 80S Buffer supplemented with 30% (w/v) sucrose. Samples were centrifuged for 70 min at 95,000 rpm (352,769 ×*g*) in a pre-cooled (4°C), TLA-100 rotor (Beckman). The supernatant (i.e. 115 µl) was carefully removed and mixed with 30 µL of 5× SDS sample buffer (250 mM TRIS-HCl, pH 6.8, 10% (w/v) SDS, 30% glycerol, 10% (v/v) 2-mercaptoethanol, 0.5% (w/v) bromophenol blue). The pellet was resuspended in 145 µL resuspension buffer (35 µL 80S Buffer, 80 µL 80S Buffer with 30% sucrose, 30 µL 5x SDS sample buffer) and run on an SDS-PAGE gel to analyze the amount of NatC that co-sedimented with the ribosome by silver staining and Western blot (sections 3.5.1 and 3.5.2, respectively).

3.5.1 Silver staining

For silver staining, SDS-PAGE gels were fixed for 30 min or overnight in Gel-fix (50% ethanol, 10% acetic acid). Afterwards gels were washed 3× 5 min (i.e. three times for 5 min intervals each) with Milli-Q water[8]. Gels were incubated for 1.5 min in Gel sensitizer (20 mg / 100 mL water) and afterwards washed 3× 30 s with Milli-Q water. Gels were then incubated in silver solution (0.2% (w/v) silver nitrate) for 25 min, followed by 2× 30s washes in 2-3 L Milli-Q water each. Gels were developed by incubating in developing solution (6 g Na_2CO_3, 2 ml Gel sensitizer, 50 µL of 37% (w/v) formaldehyde ad 100 mL water) until the desired level of staining is achieved (~ 1min). Development is stopped by discarding the developing solution and incubating in 6% acetic acid for 10 min. Gels are finally washed for another 4×10 min.

3.5.2 Western Blot, Ponceau S- and immunostaining

Western blots were performed through "wet" electro-transfer using a tank blotting system. Proteins were transferred from SDS-PAGE gels to a 0.4 µm nitrocellulose membrane (GE Healthcare) in cold *Transfer buffer* (25 mM Tris Base, 192 mM glycine, 0.1 mM EDTA and 20% methanol) by running at 80 V for 1h (or 100 V for 1.5h if two blots were run simultaneously in the same tank blotting chamber). Afterwards gels were washed 2× 5 min with Milli-Q water and a Ponceau S staining was performed in order to verify that all lanes of the SDS-PAGE gel were transferred evenly onto the nitrocellulose membrane. To this end, the nitrocellulose membrane was incubated in *Ponceau S solution* (0.2% (w/v) Ponceau S in 3% acetic acid) for 1 min and washed with water to clear the

[8] Milli-Q® is a registered trademark of MilliporeSigma (Merck). Milli-Q water refers to any type of "ultrapure" water of "Type1" which has a resistivity at 25°C of 18.2 MΩ · cm.

background. The membrane was washed in *1× TBST* (20 mM Tris-HCL, pH 7.5, 150 mM NaCl, 0.1% (v/v) Tween-20) for 2× 5 min. The membrane was then incubated in *Blocking buffer* (5% milk powder in TBST) for 1–2 h at room temperature or overnight at 4 °C. The membrane was then cut horizontally between the 55 and 70 kDa molecular weight marker bands of the *PageRuler™ Plus Prestained Protein* Ladder to allow incubation of the two membrane halves with different primary antibodies. The Blocking buffer was discarded and replaced with fresh Blocking buffer containing mouse, α-FLAG M2 (1:10,000) for incubation with the upper membrane half and rabbit, α-RPL7 (1:3,000) for the lower membrane half. The primary antibody was incubated for 1-2 h or overnight at 4°C. The membrane was rinsed twice and washed for 2× 5 min with TBST. HRP (horseradish peroxidase)-conjugated secondary antibodies in fresh *Blocking buffer* (1:10,000) were added and incubated with the corresponding membrane halves for 1–2 h at room temperature. The membrane was again rinsed twice and washed for 2× 5 min with TBST, followed by a 2× 5 min wash with ddH_2O. A working solution of *SuperSignal™ West Dura Extended Duration Substrate* (ThermoFisher) was prepared by mixing equal volumes of Stable Peroxide Solution and Luminol/Enhancer Solution (300 µL + 300 µL). The membrane was covered with working solution, placed in clear plastic wrap and chemiluminescence images were taken with the ImageQuant™ LAS 4000 imaging system (GE Healthcare). Chemiluminescence intensities of protein bands were analyzed with *ImageJ* (Schneider, Rasband, and Eliceiri 2012).

3.6 *Acetyltransferase assay*

The acetyltransferase activity was determined using the Ellman method, adapted from Thompson et al. (2004). In Fig. 10, the NatC-catalyzed Nα-terminal acetylation (Nt-acetylation) reaction (1) is shown. The two substrates: acetyl-CoA and a short, ten-residue peptide are converted to the two products of the reaction: coenzyme A (CoA) and the N-terminal acetylated peptide. The employed assay is *discontinuous*, which means that the reaction is *quenched* at regular time intervals through the addition of the chaotropic guanidinium hydrochloride, which immediately denatures the enzyme. The *Quenching buffer* is supplemented with Ellman's reagent (5,5-dithiobis(2-nitrobenzoic acid) or DTNB), which reacts with the thiol groups of CoA to produce the yellow 2-nitro-5-thiobenzoate (TNB^{2-}) anion (2), whose concentration was determined using the Beer–Lambert law ($A = \varepsilon \times c \times l$), with $\varepsilon_{412\,nm} = 14150\ M^{-1}\ cm^{-1}$.

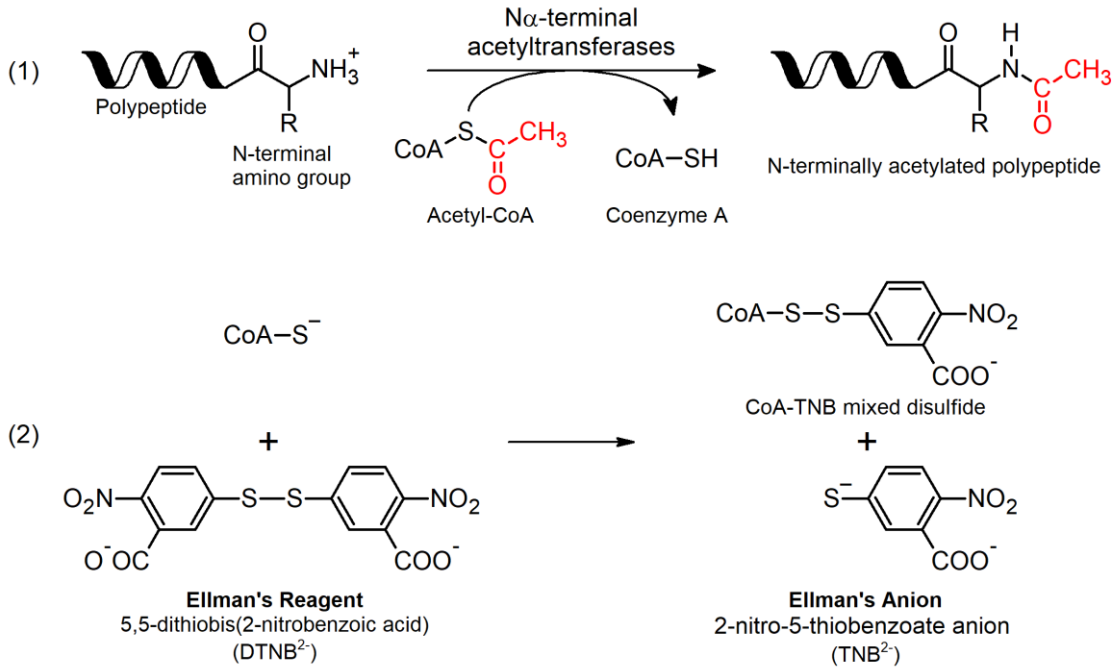


Fig. 10 Separate reactions of the colorimetric acetyltransferase assay. (1) Nα-terminal acetylation of a protein/peptide N-terminal amino group catalyzed by Nα-terminal acetyltransferases (NAT). The acetyl group is transferred from acetyl coenzyme A (acetyl-CoA). (2) Coenzyme A (CoA) thiol groups react with 5,5-dithiobis(2-nitrobenzoic acid) (DTNB or Ellman's reagent) to produce the yellow 2-nitro-5-thiobenzoate anion (TNB^{2-}) and CoA-TNB mixed disulfides.

All acetylation reactions were carried out at 25 °C in *Acetylation buffer* (50 mM HEPES, pH 7.5, 150 mM NaCl, 0.2 mM EDTA, 0.003% (v/v) Tween-20). Tween-20 was added to the buffer to reduce protein surface adsorption (Deechongkit et al. 2009).

The substrate peptide k_{cat} and K_m parameters for wild type NatC (Naa30 1–159/ Naa35 full-length/ Naa38 1–77) or NatC mutants were determined using the yArl3 substrate peptide (MFHLVGSRRR), containing the N-terminal five residues of the *Saccharomyces cerevisiae* ADP-ribosylation factor-like protein 3 (Arl3; Uniprot ID Q02804), followed by a GS-linker and a triple-arginine at the C-terminus to facilitate peptide solubility. Reactions were carried out as duplicates at six different peptide concentrations, typically between 100 to 1500 µM. Acetyl-CoA was used at a fixed concentration of 500 µM ($\geq 10\times K_m$). Thus, a total of twelve individual reactions were carried out simultaneously using a 12-channel pipette. Individual reactions (110 µL total volume) were initiated by addition of 99 µL substrate solution to 11 µL enzyme solution with final NatC concentrations of 50–2500 nM (50 nM for wild type NatC and more for NatC mutant complexes, that were catalytically deficient). The reactions were carried out for a total of 150 s (NatC wild type) up to 100 min (for catalytically impaired NatC mutant complexes). At regular time intervals, 20-µL aliquots were taken from the reaction and quenched in 40 µL *Quenching buffer* (3.2 M guanidinium-HCl, 5 mM EDTA, 100 mM sodium phosphate, pH 6.8), containing freshly added 5,5-dithiobis(2-nitrobenzoic acid) (DTNB) at a final concentration of 0.5 mM. The free thiol groups (–SH) of CoA produced in the acetylation reaction react with DTNB to produce the yellow 2-nitro-5-thiobenzoate (TNB^{2-}) anion. 50 µl of the quenched reactions were transferred into 384-well plates and absorbances were measured at 412 nm with a *M1000 Pro Microplate reader* (Tecan), assuming ε = 14150 M^{-1} cm^{-1} (Riddles,

Blakeley, and Zerner 1983). Reaction background absorbances were determined for each reaction and subtracted from the absorbances of the individual reactions. Reaction background samples had the same composition as the reaction samples, but with enzyme added to the *Quenching buffer* aliquot, to account for the cysteine thiol groups of the enzyme. Moreover, plate background absorbances of each individual empty 384-well microplate were determined and subtracted from all reactions and reaction background absorbances, to account for plate-specific imperfections (micro-scratches, impurities, etc.).

Turnover of limiting substrate did not exceed 20%. Initial velocities were fitted by nonlinear least fit squares to the Michaelis-Menten equation ($v_0 = v_{max}[S]/(K_m + [S])$) using GraphPad, version 6, to determine k_{cat} and K_m parameters. Each complete acetylation assay was performed at least in triplicate and the average k_{cat} and K_m and s.d. were calculated from the individual values of the replicates.

The K_m of acetyl-CoA for wild type NatC was determined as described before, but with a fixed Arl3 peptide concentration of 1.5 mM and 50–500 µM acetyl-CoA.

3.6.1 Assay adaptations for different peptide substrates

In addition to the Arl3 peptide (MFHLVGSRRR), k_{cat} and K_m parameters of various other substrate peptides were determined for wild type NatC.

The NatA substrate peptide (SASEAGSRRR) containing the N-terminal five residues (after removal of the initiator methionine) of the *S. cerevisiae* Threonyl-tRNA synthetase (ThrRS) (Uniprot ID: P04801); and the NatB substrate peptide (MDSEVGSRRR) containing the N-terminal five residues of the *S. cerevisiae* Actin (Uniprot ID: P60010) were tested with wild type NatC. No activity could be measured for either NatA or NatB peptides under standard conditions and even after incubating for 2 h and using a large excess (2 µM) of wild type NatC.

Arl3 peptides variants with substitutions at the second position (MWHLV-, MYHLV-, MLHLV-, MAHLV-(GSRRR)) were measured as described for the yArl3 peptide.

The Gag peptide (MLRFVGSRRR), containing the N-terminal five residues of the Major capsid protein (Gag) of the *S. cerevisiae* virus L-A (Uniprot ID: P32503), represents another known NatC substrate (Tercero, Dinman, and Wickner 1993). Arl3/Gag-hybrid peptides are based on the yArl3 peptide with single amino acid substitutions at position 3 or 4, replacing the yArl3 amino acid residues with the corresponding Gag peptide residues (MFRLVGSRRR and MFHFVGSRRR). The Gag peptide, Arl3/Gag-hybrid peptides and yArl3 variants with alanine substitutions at position 3 (MFALVGSRRR) or position 4 (MFHAVGSRRR) showed a much higher catalytic efficiency and thus the acetyltransferase assay was correspondingly adapted. To account for the higher k_{cat}, final NatC concentrations were lowered to 10-80 nM and total reaction times were as short as 60 s, with aliquots quenched every 15 s.

To determine the low peptide K_m values, peptide concentrations were lowered to between 30 µM to 500 µM. Instead of measuring duplicates for six different concentrations, the lowest two

concentrations were measured in triplicates and the three highest concentrations were measured as duplicates (e.g. three times 30 µM and 60 µM and two times 100, 200 and 300 µM). The total reaction volume was increased to up to 226 µL, so that larger aliquots of 72 µL could be taken. The aliquots were quenched in half the volume (36 µL) of 2× *Quenching buffer* (6.4 M guanidinium-HCl, 10 mM EDTA, 200 mM sodium phosphate, pH 6.8), giving a total volume of 108 µL. Two times 50 µl of each quenched reaction were transferred into separate wells of the 384-well microplates to obtain duplicate absorbance readings, which were averaged to reduce random errors.

3.6.2 Determination of peptide concentrations

Peptides were solubilized in *Acetylation buffer* to a concentration of approximately 5 mM, based on gravimetric analysis. However, since net peptide content can vary between 50-80% of the gross peptide weight and not all peptides were soluble to the same final concentration, peptide concentrations were determined photometrically from TNB^{2-} product concentration, formed in an end-point acetylation reaction. For each peptide stock, five different dilutions (25 to 125 µM, based on gravimetric estimates) were incubated together with an excess of Coenzyme A (200 µM) and 2 µM wild type NatC in *Acetylation buffer*. Reactions aliquots were quenched after 1.5 h and 2 h with a two-fold excess of *Quenching buffer*, containing 0.5 mM DTNB (final concentration). Absorbances were determined, background absorbances were subtracted and concentrations were calculated as described for the Michaelis-Menten assays. The average concentration of the five end-point reactions was taken to calculate the net peptide concentrations of the peptide stock. The 2 h measurements did not show a significant increase of absorbance, compared to the 1.5 h measurements and both readings were within 5% of one another.

3.6.3 Pathlength calculations for the acetyltransferase assay

The TNB^{2-} anion product concentration, was determined using the Beer–Lambert law ($A = \varepsilon \times c \times l$), with $\varepsilon_{412\ nm}$= 14150 M^{-1} cm^{-1}. Since the calculated product concentration is linearly dependent on the pathlength, an accurate determination of the latter is mandatory. As the absorbance was measured in 384-well microplates, the pathlength is equivalent to the height of the liquid in the well (h_l). The liquid height is dependent on the shape of the well and the volume of the measured solution. The wells of the employed microplate (Cat. No. 781 162, Greiner Bio one) have the shape of a frustum of an (upside-down) pyramid with rounded lateral edges. The well height (h_{well}) was specified by the manufacturer as 11.5 mm and the top (d_{top}) and bottom (d_{bottom}) diameters of the wells were given as 3.7 mm and 3.3 mm, respectively. A microscope image of the well top was taken to determine the top corner radius ($r_{\mathrm{c,top}}$). Calibrating the image dimensions with the diameter of the well top, a corner radius of $r_{\mathrm{c,top}}$=1.0 mm was measured. Given the pyramidal shape of the well, the bottom and top corner radii are linearly correlated with the well diameter. Thus, the bottom corner radius was calculated as $r_{\mathrm{c,bottom}} \approx 0.89$ mm. The rounded square areas of the well top and bottom were calculated with equations (1) and (2), respectively. Finally, the height of the liquid h_l (i.e. the

pathlength) was calculated using the equation for the volume of a frustum, solved for h_l (3). A detailed explanation for the derivation of these equations is given in the supplement (sections 10.9.1 to 10.9.5). Moreover, the results from the final model for the shape of the well are compared to simpler geometrical description of the well shape in supplement sections 10.9.6 and 10.9.7.

$$A_{\text{top}} = d_{\text{top}}^2 - r_{\text{c,top}}^2 \times (4 - \pi) \quad (1)$$

$$A_{\text{bottom}} = d_{\text{bottom}}^2 - \frac{d_{\text{bottom}}}{d_{\text{top}}} \times r_{\text{c,top}}^2 \times (4 - \pi) \quad (2)$$

$$h_l = \sqrt[3]{3V_l \times \frac{h_{\text{well}}^2}{\left(\sqrt{A_{\text{top}}} - \sqrt{A_{\text{bottom}}}\right)^2} + \left(\frac{h_{\text{well}} \times \sqrt{A_{\text{bottom}}}}{\left(\sqrt{A_{\text{top}}} - \sqrt{A_{\text{bottom}}}\right)}\right)^3} - \frac{h_{\text{well}} \times \sqrt{A_{\text{bottom}}}}{\left(\sqrt{A_{\text{top}}} - \sqrt{A_{\text{bottom}}}\right)} \quad (3)$$

3.7 *Crystallization of NatC*

Crystals of the NatC complex, containing subunits Naa30ΔC17 (residues 1–159), full-length Naa35 (residues 1–733) and Naa38ΔC11 (residues 1–77) were obtained through vapor diffusion at 20 °C as detailed in Table 14. NatC protein solutions were diluted to concentrations between 5–5.8 mg/mL in SEC Buffer (20 mM HEPES, pH7.5, 150 mM NaCl, 2.5 mM 2-mercaptoethanol). The selenomethionine-labeled NatC protein was crystallized in its apo form. Coenzyme A (CoA) was added at a 1:3 molar ratio to all native NatC proteins. In addition to CoA, the substrate peptides MFHLVGSRRR or MLRFVGSRRR were added to two further crystallization setups at a molar ratio of 1:5 or 1:3 respectively. SeMet and native crystals were obtained by mixing protein solution in a 1:1 mixture with well solution containing 14.5–16.5% PEG 4000, 150 mM ammonium iodide and 100 mM sodium citrate, pH 6.1–6.4.

Crystallization plates were setup either manually in a 24-well hanging-drop format by mixing 1 µL protein solution with 1 µL reservoir solution above a 1000-µL reservoir. Or plates were setup as sitting-drops in a 96-well format with a *Crystal Gryphon* dispensing robot (Art Robbins Instruments) by mixing 0.2 µL protein solution with 0.2 µL reservoir solution above an 80-µL reservoir.

Although not mandatory, NatC crystal growth could be facilitated through micro seeding. A seed stock was prepared from NatC crystals of low quality (i.e. thin needles) grown in a 2-µL drop (hanging-drop setup, 24-well plate). The entire drop was transferred into 100 µl well solution and crushed with Seed Bead (Hampton) by vortexing for 1 min. The seed stock was diluted 1:100 in well solution to be used for microseeding. To this end, 0.5 µL seeding solution was admixed with a 2-µL drop in the 24-well format or < 0.2 µL was added on top of a 0.4-µL drop in the 96-well format using a 12-channel pipette.

Diffraction quality crystals appeared after 12 – 48 h and required another five days to reach maximum dimensions. Crystals were soaked for approx. 10 s in well solution containing 20 % (v/v) glycerol or 20 % (v/v) ethylene glycol and quick frozen in liquid nitrogen.

Table 14 NatC crystallization conditions. WT, wild type (Naa30ΔC17, Naa35, Naa38ΔC17). Arl3, peptide with the sequence MFHLVGSRR. Gag, peptide with the sequence MLRFVGSRRR. EG, ethylene glycol. [a]CoA and the Gag peptide, were not observed in the electron density of the Naa30_L27A crystal.

	NatC-apo (SeMet)	NatC-CoA (native)	NatC-CoA-MFHLV (native)	NatC-CoA-MLRFV (native)	NatC_ Naa30_L27A (native)
Protein solution					
Protein concentration	52 µM	45 µM	45 µM	45 µM	45 µM
Ligands	-	135 µM CoA	135 µM CoA 225 µM Arl3	135 µM CoA 135 µM Gag	-
Buffer	20 mM HEPES, pH 7.5, 150 mM NaCl				
Reservoir solution					
PEG 4000	14.5 %	16.5 %	16 %	14.5 %	15.5 %
pH (100 mM sodium citrate)	pH 6.1	pH 6.3	pH 6.2	pH 6.3	pH 6.4
Ammonium iodide	150 mM	150 mM	150 mM	150 mM	150 mM
Crystallization setup					
Method	Sitting drop	Sitting drop	Sitting drop	Hanging Drop	Sitting drop
Format	96-well	96-well	96-well	24-well	96-well
Reservoir volume	80 µL	80 µL	80 µL	1 mL	80 µL
Drop Volume (µL)	0.2 + 0.2	0.2 + 0.2	0.2 + 0.2	1.0 + 1.0	0.2 + 0.2
Seed solution	no	no	no	0.5 µL	< 0.2 µL
Cryo-protectant	20 % EG	20 % EG	20 % glycerol	20 % EG	20 % EG

3.8 *Data collection*

All data sets were collected at the Macromolecular Crystallography (MX) beamlines BL14.1 at BESSY II in Berlin-Adlershof (Helmholtz-Zentrum Berlin für Materialien und Energie 2016). Beamline parameters are detailed in Table 15. The control software used was MxCuBE2 (Oscarsson et al. 2019).

Table 15 BESSYII – BL14.1 Parameter. From Helmholtz-Zentrum Berlin für Materialien und Energie (2016)

Source	7T-WLS-2
Monochromator	Si111-DCM with sagital bender
Energy range	5 - 15.5 keV
Energy resolution	< 2 eV
Flux	1.4 x 10^{11} (Photons/s/100 mA)
Polarization	Horizontal
Divergence horizontal	1.5 mrad
Divergence vertical	0.5 mrad
Focus size (hor. x vert.)	150 x 100 µm²
Experiment in vacuum	no
Temperature	90 - 293 K
X-ray fluorescence detector	Amptek XR123 SDD
Pixel detector	Pilatus 6M
Sample manipulators	CATS sample changer
Goniometer	Microdiffractometer MD2 with Minikappa goniometer MK3

A selenomethionine (SeMet)-labeled NatC protein crystal was mounted and an x-ray fluorescence (XRF) energy scan[9] was performed (Fig. 11).

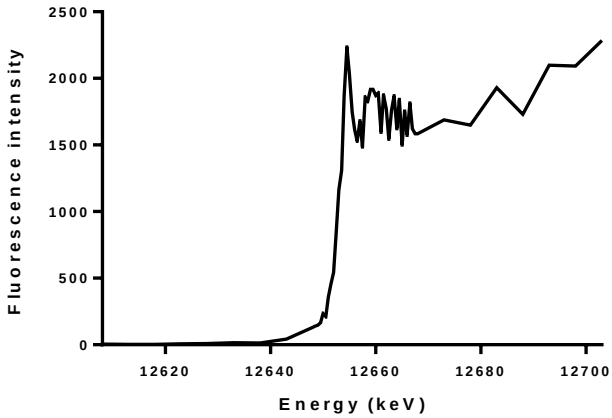


Fig. 11 X-ray fluorescence scan of a SeMet-labeled NatC crystal

The data was analyzed with the *MXCube2*-integrated software *CHOOCH* (Evans and Pettifer 2001) to obtain the anomalous scattering factors f' and f'' and to calculate the selenium (K-edge) peak (12.655 keV), inflection (12.653 keV) and remote (12.705 keV)[10] energies. Afterwards, beamline energy was shifted to the selenium peak energy to perform a single-wavelength anomalous diffraction (SAD) experiment. An initial characterization of all SeMet-labeled NatC crystals was conducted by taking two diffraction images with an exposure time of 3 s and 90° apart from one another. The diffraction images were evaluated with the X-ray diffraction image viewers *adxv* [11] or *albula*[12] (Dectris) to visually assess the quality of the crystal diffraction pattern, including maximal resolution, spot shape, mosaicity, anisotropic diffraction, multiple lattices and ice rings. The two images from the crystal with the best diffraction pattern (i.e. highest resolution, low anisotropy, low mosaicity, no ice rings, only one visible lattice) were analyzed with *iMOSFLM* (Battye et al. 2011) to index the diffraction spots and to calculate a data collection strategy.

A dataset of the SeMet-labeled NatC crystal was taken at peak energy of 12.655 keV (0.9797 Å) and consisted of 3600 images, with 0.1° oscillation range and an exposure time of 0.2 s. Oscillation was started at the rotation angle recommended by *iMOSFLM* but extended to a total oscillation range of 360°, in order to obtain a high multiplicity for SAD phasing.

For the native NatC crystals, the wavelength was shifted to 13.5 keV (0.9184 Å) and an initial characterization was performed for all crystals as described for the SeMet-labeled crystals. The two test images were indexed with *iMOSFLM* and a data collection strategy was calculated.

Datasets from all native NatC crystals were taken at 13.5 keV and consisted of 1100 to 1800 images for crystals in space group $P2_12_12_1$ and 3600 images for the crystal that grew in space group

[9] The x-ray energy was changed in a range close to the absorbance edge of the chosen element.

[10] The energy of the "remote" data set is usually taken 50 eV above peak energy ("high energy remote").

[11] Adxv - A program to display X-Ray diffraction images (https://www.scripps.edu/tainer/arvai/adxv.html)

[12] ALBULA software (https://www.dectris.com/products/albula-software)

*P*1 (see Table 16). Exposure time was between 0.3 s and 0.5 s at 0.1° oscillation range. Oscillation was started about 10° before the oscillation angle recommended by *iMOSFLM* and extended above the minimal recommendation for a complete dataset to a total of 110° to 180° for the native $P2_12_12_1$ crystals and to 360° for the *P*1 crystal, in order to increase the multiplicity for the native datasets as well.

Table 16 Data collection parameters. (for SeMet-labeled and native NatC crystals)

	NatC-apo (SeMet)	NatC-CoA (native)	NatC-CoA-MFHLV (native)	NatC-CoA-MLRFV (native)	NatC_ Naa30_L27A (native)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	*P*1
Energy (eV)	12,655	13,500	12,500	13,500	13,500
Wavelength (Å)	0.9797	0.9184	0.9184	0.9184	0.9184
Number of images	3600	1800	1100	1800	3600
Exposure time (s)	0.2	0.3	0.5	0.3	0.3
Oscillation range (°)	0.1	0.1	0.1	0.1	0.1

3.9 *Data processing, model building and refinement*

3.9.1 Anomalous dataset

The peak energy dataset from the SeMet-labeled NatC crystals was processed with XDS (Kabsch 2010). The reflection file XDS_ASCII.HKL, containing the corrected *intensities of all reflections* (= "unmerged data") was used for single-wavelength anomalous diffraction (SAD) phasing with *SHELXC/D/E* (Sheldrick 2010) using the *HKL2Map* (Pape and Schneider 2004) graphical user interface.

SHELXC estimates the marker-atom structure factors F_A and the phase shift α. The main marker atoms, that are present in the SeMet-labeled NatC crystals are the selenium atoms (Se) of the selenomethionine side chains (MSE). In addition to Selenium, iodide atoms (present in the reservoir solution in the form of 150 mM ammonium iodide) provided additional anomalous signal. At the theoretical K-edge (~12.667 keV) of selenium, iodide has an anomalous scattering factor f'', that is about 83% of the selenium scattering factor f'' as calculated with the *Anomalous Scattering server*[13] and shown in Fig. 12.

[13] The Anomalous Scattering by Ethan A. Merritt (http://skuld.bmsc.washington.edu/scatter/) uses the subroutine library by Brennan and Cowan (1992) and the theoretical approximation developed by Cromer and Liberman (1981) to calculate f' and f''.

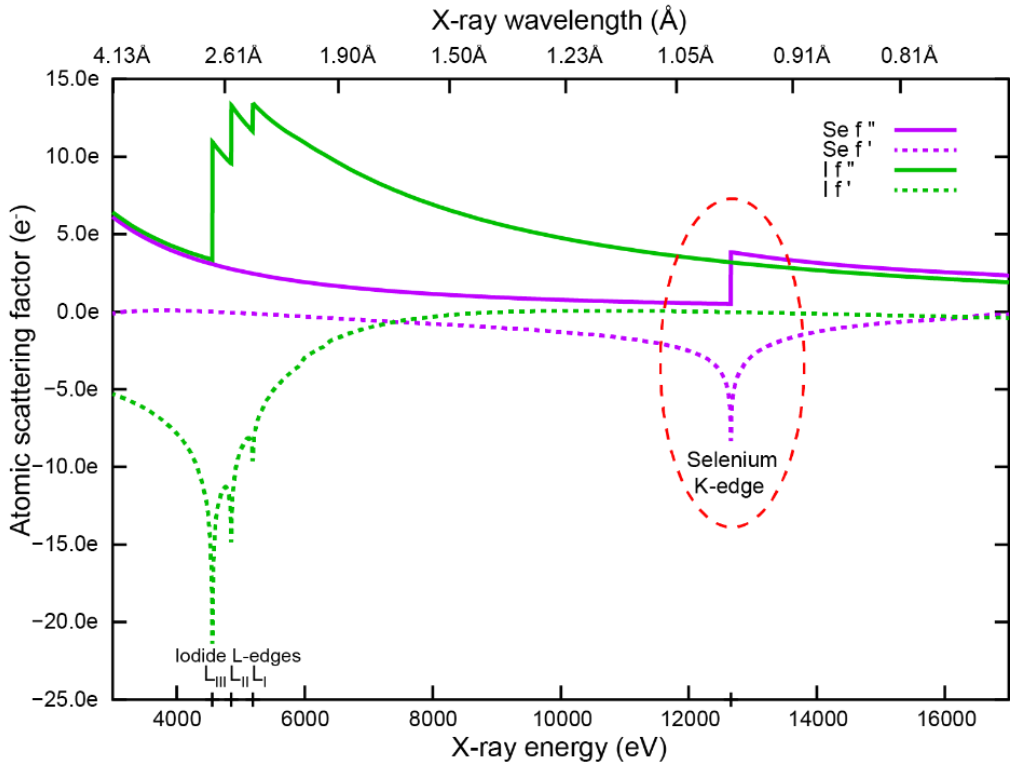


Fig. 12 Calculated atomic scattering factors for iodide and selenium. The graph was generated using the X-ray Anomalous Scattering server.

SHELXC/D calculated 1000 possible solutions for the substructure with the best solution having a CC_{all} of 47.2. For the best solution, 28 sites with occupancy > 0.2 were identified, which were three more than the 25 MSE residues present in the NatC construct, which could be attributed to the additional anomalous signal from coordinated iodide ions. From the substructure phases ϕ_A and the estimated phase shift α, the approximate phases for the macromolecule ϕ_T were calculated. Together with the observed structure factors F_T, an initial electron-density map was calculated. *SHELXE* performed an iterative phase improvement by density modification, combined with an auto-tracing of a poly-Ala chain into the electron density.

The *phs*-file, containing the electron density map produced by SHELXE, was converted to *mtz*-format using the C-shell script *phs2mtz*[14].

FreeR flags were added to 5% of the reflections with the software *FreeRflag* (Brunger 1992) from the *CCP4 suite* (Winn et al. 2011), before performing an automated model building and refinement with *Buccaneer* (Cowtan 2006).

Before performing any further refinement, the *CORRECT* step of XDS was repeated, in order to truncate the dataset to a resolution of 2.4 Å (Friedels law = false), so that the highest resolution shell shows a signal-to noise ratio (I/σ) of >1 and a Pearson correlation coefficient $CC_{1/2}$ > 50% as written in the CORRECT.LP file. The new XDS_ASCII.HKL (truncated to 2.4 Å), was converted to CCP4's mtz-format, containing the *mean structure factor amplitudes* and their standard deviations (=

[14] http://webapps.embl-hamburg.de/hkl2map/phs2mtz

"merged data") using *XDSconv*. The FreeR flags already used during the automated model-building and refinement steps in *Buccaneer* were transferred to the mtz-file from XDS with CCP4's *FreeRflag* software.

The initial model (.pdb) obtained from *buccaneer* was then completed manually in *Coot* (Emsley et al. 2010) and refined against the truncated dataset using *Phenix* (Adams et al. 2010).

The final model of the selenomethionine-substituted NatC complex (*NatC-apo)* complex contained the completely modeled Naa30ΔC17 (residues 1-159). No model could be built for two flexible loops in Naa35 (129-TTAG-132 and 376-EFESSI-381) and five residues at the Naa35 C-terminus (729-QDRYK-733). Two residues (1-ME-2) of Naa38 could not be modelled, due to the flexibility of the N-terminus.

The anomalous density map from SHELXE was used to identify selenium atoms of the selenomethionine side chains to assist model building. Additional strong anomalous density was found in between well-ordered protein chains and attributed to coordinated iodide ions (the reservoir solution contains 150 mM ammonium iodide). At the theoretical selenium K-edge (~12.667 keV), iodide has an anomalous scattering factor f'', that is about 83% of the selenium scattering factor f'' as calculated with the *Anomalous Scattering server*[15]. Iodide occupancies were included in the refinement with *Phenix*.

3.9.2 Native datasets

Datasets of all native NatC crystals were processed with *XDS* (Friedel's law = true) and truncated resolution ranges using the same criteria ($I/\sigma > 1$ and $CC_{1/2} > 50\%$ for the highest resolution shell) as described for the SeMet dataset.

The FreeR flags from the anomalous dataset was transferred to all native datasets belonging to the same space group ($P2_12_12_1$) using CCP4s *FreeRflag*. New FreeR flags were generated for the native dataset that was indexed in space group *P*1, during generation of the merged dataset with *XDSconv*.

Native NatC datasets were solved by molecular replacement (MR) using *Phaser* (McCoy et al. 2007). The refined structure of the SeMet-labeled NatC complex was used as an input model after removing all non-protein chains and replacing the selenomethionine with methionine residues using the *Phenix* software *PDB tools*.

Phaser was set to look for one copy of NatC for the native datasets that crystallized in space group $P2_12_12_1$ (identical to the SeMet crystal), which gave a unique and unambiguous solution. The unit cell of the native *P*1 crystal occupied almost exactly half the volume (50.71%) compared to the SeMet $P2_12_12_1$ crystal, which contains four unit cells per asymmetric unit (z = 4). Thus, two copies

[15] The Anomalous Scattering by Ethan A. Merritt (http://skuld.bmsc.washington.edu/scatter/)

of NatC were expected to be found in *P*1 and *Phaser* gave a unique solution using this search parameter.

Afterwards, several manual model building (*Coot*) and refinement (*Phenix*) rounds were performed to address the deviations of the protein backbone and amino acid side chains between the native and the SeMet crystals. A strong electron density was found at most of the iodide positions, observed in the SeMet structure and thus iodide ions were placed at these positions as well. A few additional iodides were placed at coordinates with strong spherical electron densities, next to well defined protein sidechains, that could not be explained by other components of the crystal buffer or reservoir solution (i.e. HEPES, citrate, PEG4000, glycerol, ethylene glycol, ammonium, sodium or chloride ions). Occupancies of iodides were refined for the native structures as well.

A clear positive difference density was visible in the active center of the NatC structures that were co-crystallized with coenzyme A (NatC-CoA structure) and the structures that were co-crystallized with both CoA and a substrate peptide (NatC-CoA-MFHLV or NatC-CoA-MLRFV). Ligand occupancies were refined. Omit maps for coenzyme A, the different substrate peptides and a water molecule in the active site of the NatC complex were generated using *phenix.polder* (Liebschner et al. 2017).

The final model of the *NatC-CoA* complex contained the completely modeled Naa30ΔC17 (residues 1-159) and Naa38ΔC11 (residues 1-77) subunits and 724 out of the 733 residues of the full-length Naa35. The nine residues that are missing in the model are located at the Naa35 C-terminus (731-RYL-733), in the α7-α8 loop (130-TA-131) and in the α16-α17 loop (376-EFES-379).

To judge the accuracy of all NatC models, the *Phenix* integration of the validation program *MolProbity* (Chen et al. 2010) was used.

3.10 *Secondary structure annotations*

Secondary structure assignment for all NatC structures was done manually by comparing the outputs of different automatic assignment methods and evaluating them through a careful visual inspection of the model and electron density. Automatic assignments from the DSSP, STRIDE and KAKSI algorithms were obtained for the Coenzyme A-bound NatC structure using the *2Struc* secondary structure server (Klose, Wallace, and Janes 2010)[16]. Prior to the analysis with the *2Struc* secondary structure server, alternative side chains were removed from the PDB file, using the *CCP4i* tool *pdbcur*[17]. Afterwards, ATOM serial numbers were reset using *Phenix's PDB file editor*. Moreover, *KAKSI* required a PDB file with SEQRES records, which were added manually.[18]

[16] http://2struc.cryst.bbk.ac.uk/twostruc

[17] http://www.ccp4.ac.uk/html/pdbcur.html

[18] SEQRES records contain the primary sequence of all polymeric molecules (i.e. proteins, peptides, and nucleic acids) that are present in the PDB entry (https://www.wwpdb.org/documentation/file-format-content/format33/v3.3.html)

Secondary structure assignments from the *PyMOLs dss* algorithm were obtained through *PyMOLs* command line[19].

Initial HELIX and SHEET records were generated with the open-source DSSP-derived program *ksdssp*[20], which is incorporated into the *phenix.secondary_structure_restraints*[21] tool (Adams et al. 2011). HELIX and SHEET lines were then manually adjusted after comparing the results of all four secondary structure assignment algorithms. Finally, HELIX and SHEET lines were added to the secondary structure section of the PDB file.[22] Furthermore REMARKS 650 and 700 were added to indicate the author-provided secondary structure assignments for both helices and sheets, respectively.

3.11 *Sequence alignments and conservation analysis*

To analyze the residue-specific conservation of NatC, multiple sequence alignments (MSA) were generated separately for Naa30 (Fig. S1), Naa35 (Fig. S2) and Naa38 (Fig. S3). Global sequence alignment with free end gaps (Blosum45) were generated with *Geneious* (https://www.geneious.com/) and visualized using *ESPript3.0* (http://espript.ibcp.fr/). Amino acid sequences of NatC orthologs were selected from six different species, covering different clades from the eukaryotic tree of life. Two species were selected from the Fungi kingdom: *Saccharomyces cerevisiae* (Sc) and the closely related *Kluyveromyces lactis* (Kl), which can assimilate lactose and convert it into lactic acid. The mouse-ear cress *Arabidopsis thaliana* (At) was selected as a model organism for the plant kingdom. *Homo sapiens* (Hs), the zebra fish *Danio rerio* (Dr) and the red flour beetle *Tribolium castaneum* (Tc) were selected for the holozoan clade, which includes animals and single-celled relatives, but excludes fungi.

The degree of conservation (conservation score) of each amino acid was calculated with the *ConSurf* (Ashkenazy et al. 2016) web server[23] using a user-provided multiple sequence alignment. To generate the MSA, NatC subunit orthologs were selected manually from eighteen different species across the eukaryotic tree of life, using the publication of Rathore et al. (2016) as a guide. The eighteen species are represented in a phylogenetic tree diagram in Fig. S4.

[19] Using the command: “iterate n. CA, print chain + ';' + resi + ';' + resn + ';' + alt + ';' + ss”

[20] *Phenix’s kdssp* can be executed from the *Phenix Command Prompt* Using the command: “phenix.secondary_structure_restraints model.pdb format=pdb”

[21] https://www.phenix-online.org/documentation/reference/secondary_structure_restraints.html

[22] A detailed description can be found in the secondary structure section of the Atomic Coordinate Entry Format document (Version 3.3, https://www.wwpdb.org/documentation/file-format-content/format33/sect5.html)

[23] http://consurf.tau.ac.il/

3.12 *Structure analysis methods*

The approximate Minimal Bounding Box (MBB) was calculated with the *PyMOL* script "Draw Protein Dimensions", which calculates an Inertia Axis Aligned Bounding Box (IABB) and was written by Pablo Guardado Calvo[24].

Interfaces between the subunits Naa30, Naa35 and Naa38 were analyzed for the *NatC-CoA* structure using the *PISA* (Krissinel and Henrick 2007) web server[25]. Buried surface areas were also calculated by the *PISA* server.

The structures of the three NatC subunits were compared against the complete PDB database using the *DALI* web server[26] (Holm and Rosenström 2010).

Structural superpositions and RMSD calculations for structures from *different* proteins (e.g. NAT homologs and *DALI* matches) were performed with the *PyMOL* command *super*[27]. To compute the C_α RMSD between two structures of the *same* protein (e.g. *NatC-CoA* vs. *NatC-apo*) the command *fit*[28] was used.

The diameters of the tunnels with a roughly spherical diameter were calculated with the *PyMOL* plugin *Caver* (Chovancova et al. 2012). Dimension of tunnels with elliptical cross section were determined by measuring the distance between pointer atoms, placed manually on the *PyMOL*-generated Conolly surface[29].

The electrostatic surface potential of the NatC-CoA complex was calculated with the *PyMOL APBS* plugin (Jurrus et al. 2018)[30].

[24] https://pymolwiki.org/index.php/Draw_Protein_Dimensions

[25] 'Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute. (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html)

[26] http://ekhidna2.biocenter.helsinki.fi/dali/

[27] *Super* (https://pymolwiki.org/index.php/Super) performs a sequence-independent alignment and is the method of choice for structural alignment of two proteins with a low sequence similarity.

[28] *Fit* (https://pymolwiki.org/index.php/Fit.)will superimpose two models by aligning *matching atoms*, which need to have the exact same atom identifier (segi, chain, resn, resi, name and alt). Any deviations between atoms identifiers of the two structures, for example, alternate locations of amino acids, were removed beforehand.

[29] The Connolly surface, shown in the *PyMOL* surface representation, is the solvent/protein contact surface, which is traced out by surfaces of waters in contact with the protein (https://pymolwiki.org/index.php/Surface)

[30] APBS writes out the electrostatic potential in dimensionless units of kT/e, where k is the Boltzmann's constant (1.381×10^{-23} J·K^{-1} or 8.617×10^{-5} eV·K^{-1}), T is the temperature of the calculation in K (default is 298.15 K, i.e. 25°C) and e is the charge of an electron ($1.60217646 \times 10^{-19}$ C).

4 Results

4.1 *Purification of the NatC complex*

To obtain sufficient amounts of the NatC complex for structural and biochemical studies, the three NatC subunits (Naa30, Naa35 and Naa38) from *S. cerevisiae* were co-expressed in *E. coli* using a single expression vector (see sections 3.13.1.1 and 3.2). Secondary structure predictions revealed a highly flexible C-terminus in Naa30. Moreover, sequence alignments of Naa30 paralogs (Fig. S1), showed that the corresponding C-terminal residues are not conserved, pointing to a non-essential role for catalysis. Since a deletion of similar non-conserved C-terminal residues of the Naa10 catalytic subunit was required for the crystallization of the NatA complex (Liszczak et al. 2013), a truncated construct was also used for Naa30 (Naa30ΔC17).

Furthermore, during initial purification experiments, a partial proteolytic degradation of the smallest subunit, Naa38, was observed, as indicated by a double band on an SDS-PAGE gel (Fig. 13A). MALDI-MS analysis (see 3.3) of the degradation fragment demonstrated a mass difference of 1183.0 Da, compared to full length Naa38, which is closest to the mass of the eleven C-terminal residues of Naa38 (Fig. 13B).

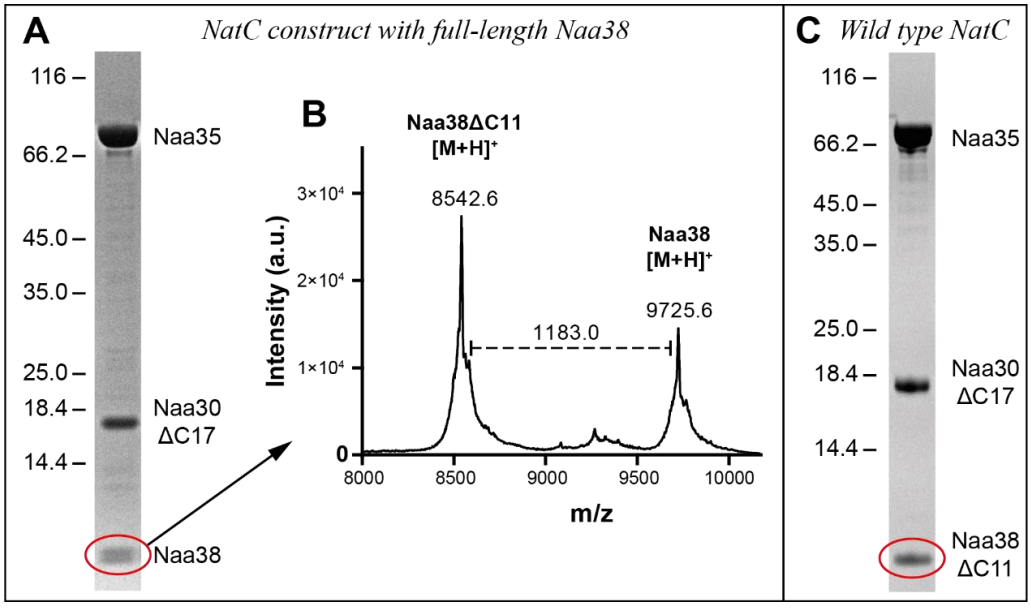


Fig. 13 MALDI-MS analysis of the Naa38 partial proteolytic degradation. (A) Purified NatC complex with full-length Naa38 on a 4–12% SDS-PAGE gel, showing a clear double band (red circle) for the small subunit. (B) MALDI-MS of the purified NatC complex with partially degraded Naa38. The right peak corresponds to full-length Naa38 and was used together with its doubly charged species (not seen on the plot) for calibration. The left peak corresponds to the partially degraded Naa38 subunit. A mass difference of 1183 Da between full-length and degraded Naa38 was measured. (C) 12% SDS-PAGE gel showing purified wild type NatC with truncated Naa38ΔC11, displaying only a single protein band (red circle).

The presence of two Naa38 protein species results in an inhomogeneous NatC complex. Moreover, a partial proteolytic digest is indicative of an unstructured region in a protein and both factors decrease the likelihood of crystallization (McPherson and Gavira 2014). As for Naa30, the C-terminus of Naa38 is not conserved (Fig. S2). Therefore, the eleven C-terminal residues were

truncated in the Naa38ΔC11 construct. After expression and purification, only a single protein band was visible for the Naa38ΔC11 subunit on the SDS-PAGE (Fig. 13C). Thus, all further NatC complex constructs contained subunits Naa30ΔC17 (residues 1–159), full-length Naa35 (residues 1–733) and Naa38ΔC11 (residues 1–77), which will be henceforth referred to as the wild type NatC complex.

In addition to suitable construct boundaries, a reproducible protein purification protocol needed to be established (see Fig. 14). To achieve this, the large auxiliary subunit Naa35 was equipped with an N-terminal 6×His-tag for affinity purification by Immobilized metal ion affinity chromatography (IMAC) (see 0). The remaining NatC subunits Naa30ΔC17 and Naa38ΔC11 were untagged, and co-purified with Naa35, indicating that the three proteins form a stable complex.

Following bacterial expression, Naa30ΔC17 and Naa38ΔC11 were in molar excess over the tagged Naa35 in the soluble fraction, because a large proportion of the untagged subunits passed the affinity material and appeared in the flow-through fraction (Fig. 14A). The excess amounts of Naa30ΔC17 and Naa38ΔC11 allowed for the purification of a homogenous NatC complex, with a subunit ratio of 1:1:1.

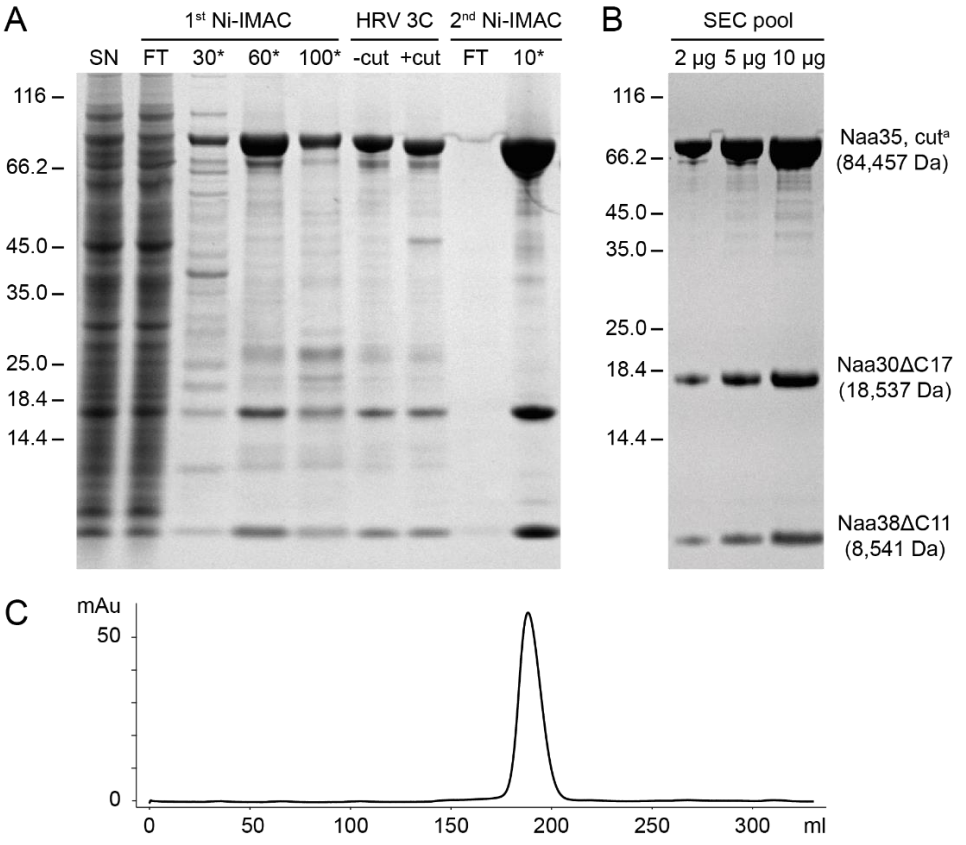


Fig. 14 Purification of the wild type NatC complex. (A) Samples loaded on a 4–12% gradient gel. Ni-IMAC, Ni^{2+}-Immobilized metal ion affinity chromatography; SN, supernatant; FT, flow-through; *, Imidazole concentrations during wash and elution steps. HRV 3C, human rhinovirus 3C protease cleavage with fractions before (-cut) and after (+cut) digest. (B) Different amounts of purified NatC complex, loaded on a 12% gel, after a final size exclusion chromatography (SEC) polishing step with an S200 column. (C) [a]Naa35, cut refers to the full-length Naa35 subunit, with residues GP at the N-terminus, which remain after HRV 3C protease cleavage.

The majority of *E. coli* proteins were removed during a first washing step of the affinity column using low imidazole concentrations. When eluted from the column with higher imidazole concentrations, the NatC complex exhibited a 260/280-ratio of ≈1.4, which suggested a significant nucleic acid contamination with an extent of roughly 20%.[31] An overnight endonuclease treatment in parallel to proteolytic cleavage of the 6×His-tag, followed by a second IMAC, allowed for the successful removal of the nucleic acid contaminants, resulting in a low 260/280 ratio of ≈ 0.6. During the final size exclusion chromatography (SEC) step, further *E. coli* protein contaminants were removed (Fig. 14B) and the 260/280 ratio was decreased to ≈ 0.5, which correspond to a protein purity of almost 100%. Furthermore, the S200 SEC profile (Fig. 14C) showed one symmetrical peak in the center of the chromatogram, indicating the presence of one uniform NatC complex species. NatC was typically concentrated to 10 mg/mL, but also higher concentration (up to 30 mg/mL) were stable and did not aggregate over time. The typical yield for a NatC complex purification was about 1–3 mg NatC per liter of *E. coli* culture. A seleno-methionine substituted variant of the NatC complex was expressed as described in section 3.2.5, and purified in the same way as the WT NatC complex.

In conclusion a stable heterotrimeric *S. cerevisiae* NatC complex could be purified to homogeneity for use in functional and structural studies.

[31] https://www.biotek.com/resources/docs/PowerWave200_Nucleic_Acid_Purity_Assessment.pdf

4.2 *Structures of the apo and CoA-bound NatC complex*

4.2.1 Crystallization

To obtain insights into the atomic structure of NatC, the purified selenomethionine-labeled NatC complex was crystallized through vapor diffusion in both hanging- and sitting-drop format. Also, native NatC complex was crystallized together with coenzyme A, a natural product of acetyltransferases.

First crystals of the NatC complex were obtained in several conditions of an initial PEG screen, containing PEG 3350 or PEG 4000 with the pH varying from 5.6 to 8.5. These crystals grew as thin needles from a single nucleation center (Fig. 15, left). By fine screening the PEG concentration, buffer conditions and through the addition of different ammonium salts, plate-shaped crystals were obtained (Fig. 15, middle). These plates were initially thin and entangled, but were further increased in size by changing the drop and reservoir volumes, until thicker, single plates were obtained (Fig. 15A, right). However, these plates diffracted X-rays to only about 8 Å. By further fine-tuning the conditions, crystals diffracting X-rays between 2.4–3.0 Å were obtained in a crystallization solution containing 14.5–16.5% PEG 4000, 150 mM ammonium iodide and 100 mM sodium citrate, pH 6.1–6.3 (see 3.7). Under these conditions, both SeMet-labeled and native NatC protein, supplemented with CoA formed rod-shaped crystals, which appeared after 12–48 h and required another five days to reach their maximum dimensions (Fig. 15B).

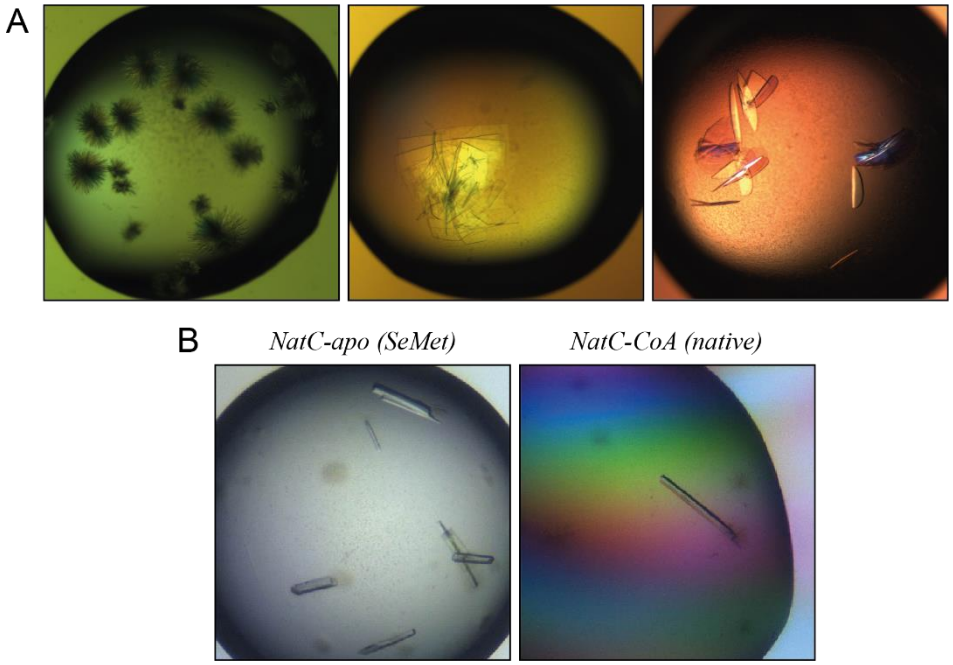


Fig. 15 Initial and diffraction quality crystals of the NatC complex (A) Early NatC crystals, forming needles and thin plates, diffracted to a resolution of about 8 Å only. (B) Rod-shaped, single, diffraction quality crystals of SeMet-labeled and native NatC-CoA protein.

4.2.2 Structure determination

X-ray data were collected at BESSY as described in 3.8. The best selenomethionine-substituted NatC crystal (*NatC-apo*) diffracted X-rays to a maximum resolution of 2.4 Å and belongs to the orthorhombic space group $P2_12_12_1$. The data collection statistics are listed in Table 17.

Table 17 Data collection statistics for a crystal of selenomethionine-substituted and CoA-bound NatC. Values in parentheses are for highest-resolution shell.

Data collection	NatC-apo (SeMet)	NatC-CoA (native)
Space group	$P2_1\ 2_1\ 2_1$	$P2_1\ 2_1\ 2_1$
Cell dimensions		
a, *b*, *c* (Å)	48.12 140.42 166.41	48.04 139.79 166.56
α, *β*, *γ* (°)	90 90 90	90 90 90
Wavelength (Å)	0.9797	0.9184
Resolution range (Å)	46.23 - 2.40 (2.48 - 2.40)	45.43 - 2.45 (2.54 - 2.45)
Completeness (%)	99.6 (97.9)	99.9 (99.4)
R_{sym} (%)	12.8 (133.2)	14.8 (159.5)
Redundancy	7.0 (6.4)	6.6 (6.8)
I/σ (I)	11.3 (1.2)	10.9 (1.1)

The crystallized selenomethionine (MSE)-labeled NatC construct contained 25 selenium (Se) atoms. In the initial substructure of the anomalous scatterers, 28 sites with occupancy > 0.2 were identified, some of them resulting from iodide atoms that also have a weak anomalous signal at the chosen peak wavelength of the K-absorption edge of selenium. Together, selenium and iodide delivered enough anomalous signal for successful experimental phasing by a single-wavelength anomalous diffraction (SAD) approach (see 3.9.1).

Exemplary selections of the calculated electron maps including the anomalous signals are shown in Fig. 16A. Alpha-helices were clearly visible and their orientation could be traced. At some positions, it was even possible to discriminate between the three large aromatic side chains. For example, the hydroxyl group of Tyr472 was clearly visible (Fig. 16A, top). The experimental electron density map was used to automatically build a model for the entire NatC complex, that was then manually completed and refined to a R_{work}/R_{free}[32] of 19.5% / 22.3% (see Table 18). Both *R* values are lower than the median *R* values of PDB structures with similar resolution (see Fig. 17B). Thus, the overall model provided a good explanation of the experimental data.

[32] R_{work} and R_{free} values measure the agreement between the model and the experimental diffraction data. To avoid an over-modeling of the data, a R_{free} value is calculated for 5% of the reflections that are not included in the refinement. The better the model describes the data, the lower will be the R values.

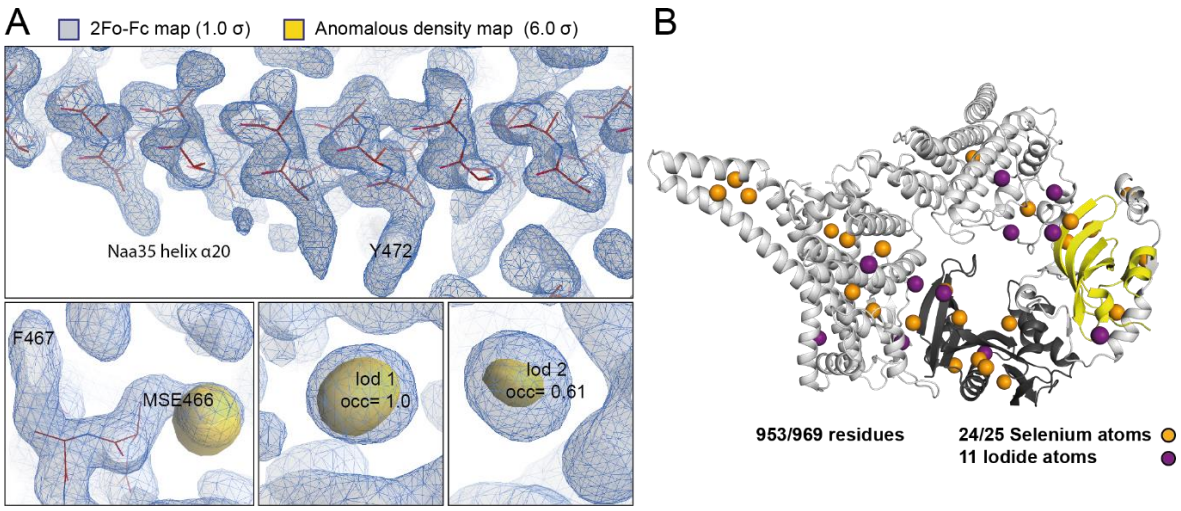


Fig. 16 Data processing and model building of the SeMet-labeled NatC complex. (A) Selected examples of the final *SHELX* $2F_o$-F_c map (1 σ) and anomalous map (6 σ). (B) Final model of *NatC-apo* (SeMet), built with *COOT* and refined with *PHENIX*.

The native CoA-bound NatC complex diffracted X-rays to a resolution of 2.45 Å and crystallized in the same orthorhombic space group as the SeMet-labeled crystal. Data collection statistics are listed in Table 17. The refined structure of the SeMet-labeled NatC complex was used as an input model for solving the structure of the native NatC complex by molecular replacement (see 3.9.2).

The final model of the *NatC-CoA* complex contained the completely modeled Naa30ΔC17 (residues 1-159) and Naa38ΔC11 (residues 1-77) subunits and 724 out of the 733 residues of the full-length Naa35. It was refined to a R_{work}/R_{free} of 20.3% / 23.6% (see Table 18).

Table 18 Refinement statistics of the apo and CoA-bound NatC structures. Values in parentheses are for highest-resolution shell. [a]R_{free} was calculated for 5% of the reflection data.

Refinement	NatC-apo (SeMet)	NatC-CoA (native)
Resolution range (Å)	46.23 – 2.40 (2.48 – 2.40)	45.43 – 2.45 (2.54 – 2.45)
No. of reflections	85037 (8268)	42170 (4132)
R_{work} / R_{free} (%)[a]	19.5 / 22.3	20.3 / 23.6
No of atoms		
Protein	7718	7806
Ligands	100	122
Water	181	190
B factors (Å^2)		
Average	65.7	64.5
Protein	66.0	64.6
Ligands	72.5	79.4
Water	51.6	50.2
RMSD		
Bond lengths (Å)	0.004	0.004
Bond angles (°)	0.90	0.92
Ramachandran plot (%)		
Most favored	98.5	98.9
Allowed	1.5	1.1
Outlier	0	0
Rotamer outliers (%)	0.3	0.5
Clashscore	4.0	3.9

For both models, >98% of all residues were in the most favored region of the Ramachandran plot (Fig. 17A) (Ramachandran and Sasisekharan 1968) and less than 0.2% of were Ramachandran outliers (see Table 18). Moreover, less than 1% of all side chains were rotamer outliers and the models displayed a low clashscore[33] compared to models of similar resolution (Fig. 17B), pointing to an excellent stereochemistry of the refined models.

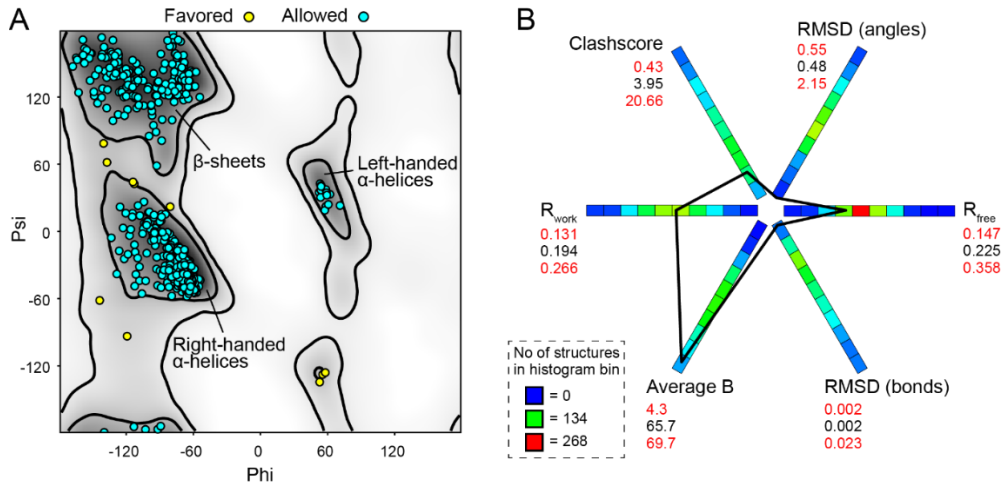


Fig. 17 Evaluation of the NatC-apo (SeMet) structure. (A) Ramachandran plot for all non-Pro/Gly residues, with favored residues in cyan and allowed residues in yellow. No residues fell in the disallowed region. (B) Data statistics for *NatC-apo* compared with 722 PDB structures of similar resolution. Values for *NatC-CoA* are printed in black and the upper and lower limit for each distribution of the PDB parameters is printed in red. The connecting polygon (black) shows were the NatC values fall in each distribution.

[33] The clashscore is defined as the number of overlaps (more than 0.4 Å) between non-hydrogen bonded atoms per 1000 atoms and should be as low as possible.

4.3 *Overall structure of the NatC complex*

The root-mean-square deviation (RMSD) of the Cα atom positions between the selenomethionine-substituted apo structure (*NatC-apo*) and *NatC-CoA* was only 0.231 Å for a total of 953 compared residues, and no difference in the overall secondary, tertiary and quaternary structure was observed. Therefore, the better-defined CoA-bound NatC structure is presented in the following. A cartoon-representation gives an overview of both tertiary and quaternary structure of the NatC complex (Fig. 18) and a topology (Fig. 19) plot shows the arrangement of secondary structure elements, especially of β-strands.

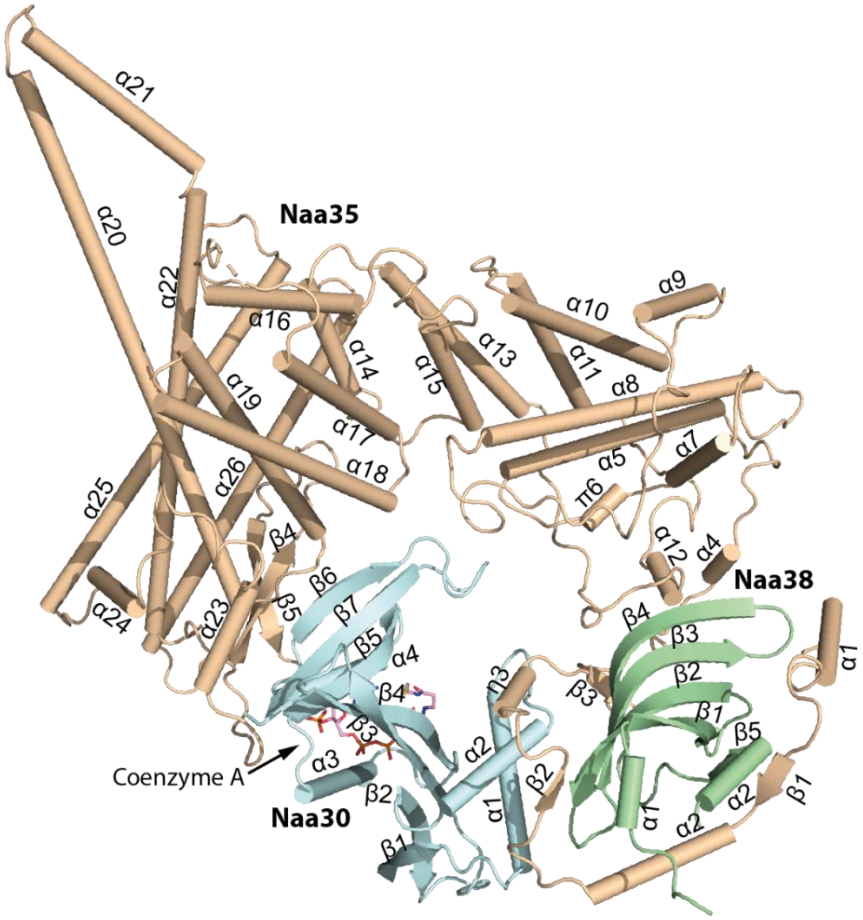


Fig. 18 Overall structure of the native NatC complex, bound to CoA Cartoon representation of *NatC-CoA* with Naa30ΔC17 (pale cyan), Naa35 (wheat) and Naa38 (pale green), bound to coenzyme A (in purple). Naa30 helices and sheets are labeled α1–α4 and β1–β7, respectively. Naa35 is labeled α1–α26 and β1–β5, with helix three (η3) being a short 3-10 helix and helix six (π6), being a short pi-helix. Naa38 is labeled with α1–α2 and β1–β5.

The catalytic subunit Naa30 adopts the typical Gcn5-related N-acetyltransferase (GNAT) fold: an N-terminal strand (β1), followed by two helices (α1 and α2), three antiparallel strands (β2-β4), followed by a "signature" central helix (α3), a fifth strand (β5), a fourth helix (α4) and a final strand (β6) (Vetting et al. 2005). The seventh β strand of Naa30 lies antiparallel in between β5 and β6 (see Fig. 18 and Fig. 19). Naa30 also contains the conserved CoA-binding motif (GNAT motif A) with the sequence Q/R-x-x-G-x-G/A (where x is any amino acid), which is located in the β4–α3 loop (see Appendix, Fig. S1) (Lu, Berkey, and Casero 1996; Wolf et al. 1998).

The auxiliary subunit Naa35 is composed of 26 helices ranging from 5 to 45 residues in length. Among these is one five-residue long 3_{10}-helix (η3) and a seven residue π-helix (π6).[34] Helix α21, together with the C-terminal end of helix 20 (Naa35 subunit) extrude by approx. 29 Å from the central body, forming an extension, which will be referred to as the "*Naa35 tip*" in this thesis.[35]

The small auxiliary subunit Naa38 adopts the conserved Sm fold (Kambach et al. 1999), containing a N-terminal α-helix, followed by a strongly bent five-stranded β-sheet. Naa38 contains an additional short α-helix at the C-terminus. The five β-strands (β1-β5) of Naa38 interact with three short β-strands (β1-β3) of subunit Naa35 to form a bifurcated, anti-parallel β-sheet. The first sheet of the bifurcated Naa35-Naa38 β-sheet consists of seven β-strands and the second sheet of five β-strands (Fig. 19).

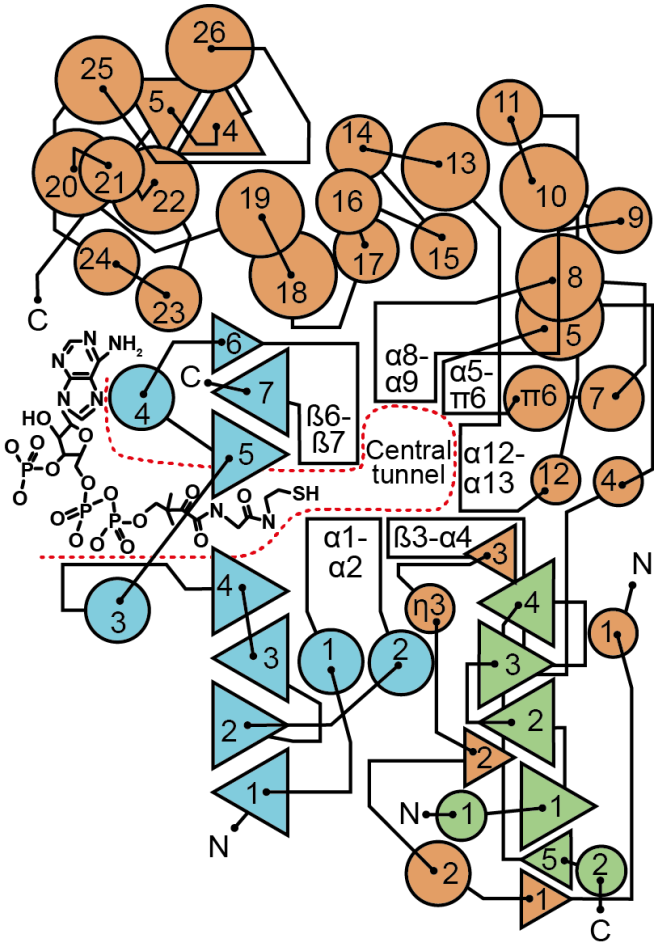


Fig. 19 Topology plot of NatC Manually drawn TOPS-like cartoon of NatC. Triangles represent β-strands, with smaller triangles indicating sheets with ≤6 residues. Adjacent triangles pointing in the same direction, indicate parallel β-strands; triangles facing in different directions indicate antiparallel β-strands. Circular symbols represent helices containing ≤5 residues (small circles), 6-19 residues (medium circles) or ≥20 residues (large circles).

[34] Helices are labeled separately for each subunit by their type (α, 3_{10} or π) followed by the total helix number.

[35] The length of the *Naa35 Tip* is approximated by the Cα atom distance (28.9 Å) between Ala507, located in the α20–α21 loop and Leu529, located in helix α22.

A minimal bounding box was calculated for the CoA-bound NatC complex which reveals its approximate dimensions of 119 Å × 86 Å × 69 Å (Fig. 20A). The surface representation shows another remarkable feature of the NatC complex: two tunnels, which are located roughly at the center of the entire complex at the Naa30-Naa35 interface. The thiol group of coenzyme A is located in the "*side tunnel*", which contains the active site of the catalytic subunit Naa30 (Fig. 20 B, C). The *side tunnel* is separated from the "*central tunnel*", which is surrounded entirely by loop-regions of the subunits Naa30 and Naa35. The abundance of loop regions near the *central tunnel* indicates a large flexibility of this region, which may play a role in substrate binding and is further explored in section 4.8.3.

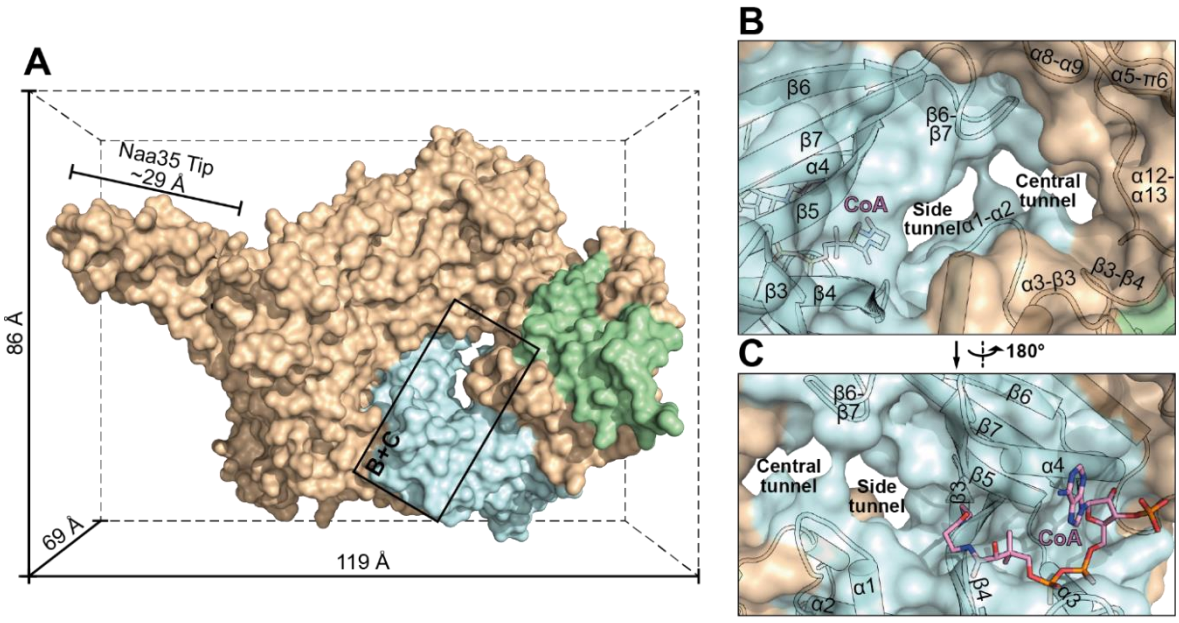
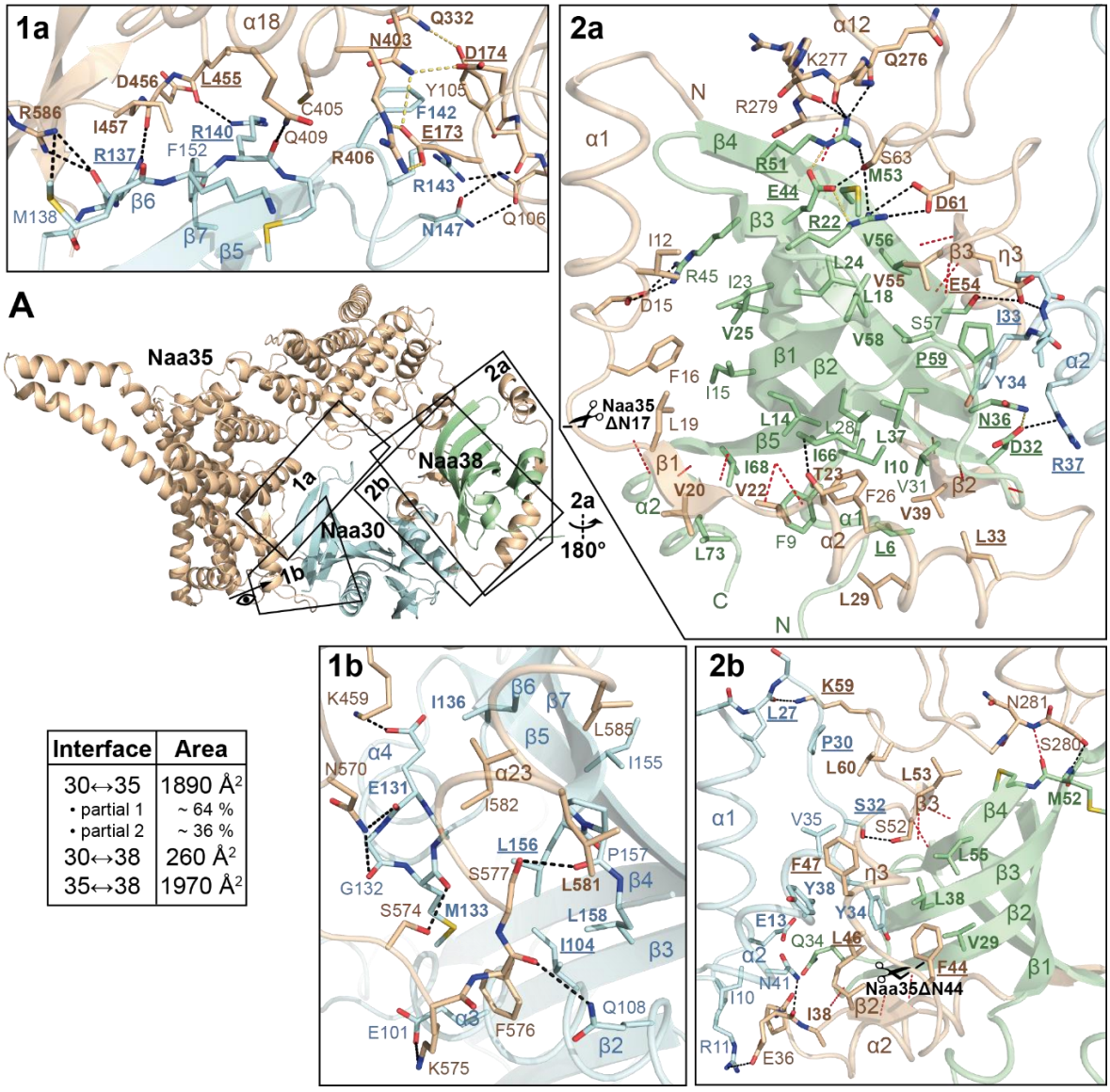


Fig. 20 Surface representation of NatC (A) *NatC-CoA* surrounded by its minimal bounding box, representing the dimensions of the complex. Helix α21 and the C-terminal end of α20 (Naa35 subunit) form the *Naa35 Tip*. (B) Zoom view, showing the *central* and *side tunnel* and surrounding loop regions. (C) Back view showing the CoA binding groove that connects directly to the end of the side tunnel, which harbors the active site.

4.4 *NatC subunit interfaces*

The Naa35 subunit wraps around the catalytic Naa30 subunit in an open ring-like tertiary structure, that covers about three fourth of the circumference of the roughly spherical Naa30 subunit (interface 1, Fig. 21). The Naa38 subunit is primarily tethered to the NatC complex by the N-terminus (residues 1–64) and a short stretch of subunit Naa35 (residues 276–282, containing helix α12) forming interface 2 (Fig. 21). Both segments of Naa35 wrap around almost the entire circumference of Naa38, forming another ring-like tertiary structure.



Interface	Area
30↔35	1890 Å²
• partial 1	~ 64 %
• partial 2	~ 36 %
30↔38	260 Å²
35↔38	1970 Å²

Fig. 21 NatC subunit interfaces. (A) Cartoon representation of *NatC-CoA*, showing the two main interface regions, enlarged in four different zoom views (1a,b and 2a,b). Key residues (stick format) are labeled in bold (highly conserved) and are underlined (completely conserved). Hydrogen bonds and salt bridges involving side chains are colored in black; backbone-backbone interactions are colored in red. The Naa30-Naa35 interface is divided into two partial interfaces, comprising residues 97–159 or residues 10–44 of subunit Naa30, respectively. Scissors symbols represent two Naa35 constructs boundaries.

The buried surface area between Naa30 and Naa38 is 1,890 Å^2, while the interface between Naa35 and Naa38 measures 1,970 Å^2. There is only a minor interface of 260 Å^2 between the catalytic subunit Naa30 and the small subunit Naa38.

The first interface region comprises the C-terminal half of subunit Naa30 (residues 97-159) and different segments of Naa35, which are far apart in sequence. The most intimate interactions between the two subunits are made between α3 and the α4–loop–β6–loop–β7 segment of subunit Naa30 and α18, the α19-loop-α20 segment, the α22–α23-loop and helix α23 of subunit Naa35. For example, two hydrogen bonds are formed between the completely conserved arginines Arg137 and Arg140 in β6 of subunit Naa30 and the backbone carbonyl oxygens of Asp456 and Leu455 of subunit Naa35, respectively. The β6–β7 loop of Naa30, important for its enzymatic activity, forms one face of the central tunnel and shows only weak interactions with several other loops of the Naa35 subunit: Two hydrogen bonds are formed between the side chains of Asn147 (Naa30) and Gln106 (Naa35) and one additional hydrogen bond is formed between Arg143 of Naa30 and the carbonyl oxygen of Tyr105 (Naa35). As will be discusses later, the binding of the Gag peptide substrate induces a conformational switch of the β6–β7 loop, demonstrating its general flexibility.

The second interface region comprises the N-terminal half of subunit Naa30, the Naa35 N-terminus and a short stretch of helix Naa35α12, plus the entire Naa38 subunit. As mentioned above, the N-terminus of Naa35 is wrapped around almost the entire circumference of Naa38. The N-terminal alpha helix (α1) and the succeeding loop are not conserved among Naa35 orthologs and form very limited interactions with the adjacent Naa38 subunit. However, the subsequent β-strand (β1) of Naa35 is highly conserved and constitutes one end of the bifurcated Naa38-Naa35 β-sheet. The Naa35 β1-α2-loop-β2 segment forms a highly conserved, extremely hydrophobic interface with several Naa38 side chains, which are distributed almost throughout the entire primary sequence of Naa38. The second β-strand of Naa35 is located on the opposite site of Naa38 (compared to the β1-strand) and forms the second end of the bifurcated Naa38-Naa35 β-sheet. An extended segment (loop-β2-loop-η3-loop) of Naa35 is sandwiched in between the two N-terminal helices α1 and α2 of Naa30 and the Naa38 subunit. The mainly hydrophobic interfaces between the three subunits involve highly or completely conserved residues from Naa30, Naa35 and Naa38. The third β-strand of Naa35 represents the first strand of the bifurcated Naa35-Naa38 β-sheet and forms three backbone hydrogen bonds with Naa38. Two salt bridges are formed between the side chains of the completely conserved Asp61 (Naa35) and Arg22 (Naa38). This interaction is supported by Ser63 (Naa35) which is within hydrogen bonding distance of Glu44, Arg22 and Arg51 of Naa38. Finally, Arg51 (Naa38) is within hydrogen bonding distance of three backbone carbonyl oxygens of Gln276, Lys277 and Arg279 from the C-terminal end of the short α12 helix of Naa35.

In summary, the auxiliary subunit Naa35 is a central assembly hub of the NatC complex, which forms extensive interactions with the catalytic subunit and the second auxiliary subunit Naa38.

4.4.1 The Naa35 N-terminus is critical for NatC integrity

To test the significance of the Naa35 N-terminus for NatC integrity, structure-guided truncation constructs were designed (see 3.1.3). Wild type NatC (Naa30ΔC17, Naa38ΔC11, full-length Naa35) runs as a complex on a size exclusion chromatography (SEC) with all three subunits present, as shown by an SDS-PAGE gel (Fig. 22, left lane). A NatC complex lacking the first seventeen non-conserved residues of Naa35 (Naa35ΔN17) also runs as a complex on the SEC, with all three subunits visible on the subsequent SDS-PAGE gel (Fig. 22, middle lane). As shown in the previous section (4.4) this non-conserved N-terminal region of Naa35 (see sequence alignment in Fig. S3 of the appendix) forms very limited interactions with the adjacent Naa38 subunit. Thus, the non-conserved N-terminal α1 helix of Naa35 is not essential for structural integrity of NatC.

The Naa35ΔN44 construct additionally lacks the β1-α2-loop-β2 segment, which forms extensive hydrophobic interactions with Naa38 and participates in the common Naa35-Naa38 β-sheet. In line with the structural data, a NatC complex affinity purified via the Naa35ΔN44 subunit only forms a heterodimer with Naa30 (Fig. 22, right lane). Thus, the integrity of the heterotrimeric NatC complex is dependent on the N-terminal region of Naa35, except for its very first α1 helix.

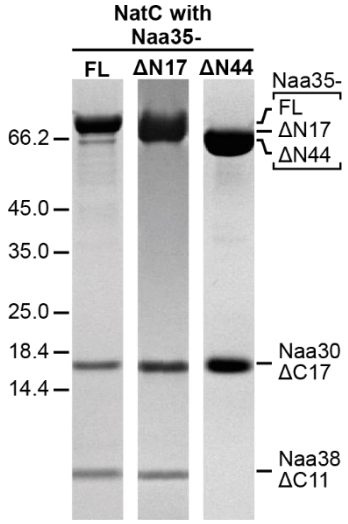


Fig. 22 The Naa35 N-terminus is critical for NatC integrity SDS-PAGE gel of purified NatC complexes containing full-length Naa35 (residues 1–733), Naa35ΔN17 (residues 18–733) or Naa35ΔN44 (residues 44–733).

4.5 *Comparison of NatC with heterodimeric NatA and NatB complexes*

A comparison of the heterotrimeric NatC complex with the heterodimeric NatA and NatB complexes shows remarkable differences in their tertiary and quaternary structures (Fig. 23).

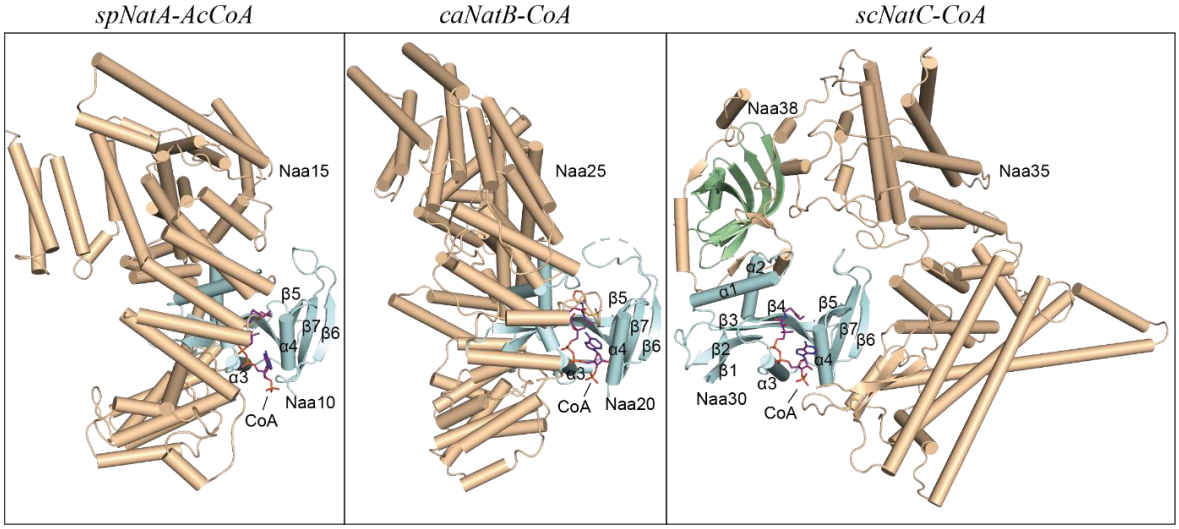


Fig. 23 Comparison of NatA, NatB and NatC complexes NatA (4KVO) and NatB (5K04) were superimposed with NatC using the corresponding catalytic subunits (cyan). Afterwards, the entire complexes were translated horizontally for a side-by-side comparison. The Cα RMSD between Naa30 and Naa10 is 1.2 Å (125 Cα atoms) and between Naa30 and Naa20 is 1.3 Å (116 Cα atoms).

At its core, all NAT complexes have the evolutionary conserved catalytic subunit, with its conserved GNAT architecture (Vetting et al. 2005). In all three NAT complexes, the additional β7-strand lies antiparallel in between β5 and β6. However, while NatA and NatB contain a single related, auxiliary subunit (called Naa15 and Naa25, respectively), NatC contains two auxiliary subunits, Naa35 and Naa38, which have different folds. The α-helices in subunits Naa15 and Naa25 form thirteen or twelve so-called helical bundle tetratricopeptide repeats (TPR), respectively. The TPR motif adopts a basic helix-turn-helix fold and consists of 34 amino acids with a degenerate consensus sequence, containing small and large hydrophobic amino acids (Zeytuni and Zarivach 2012). These repeats form a tertiary structure, which is found in numerous proteins and serves as an interaction module and mediator for heteromeric complex formation. However, in contrast to NatA and NatB, a sequence analysis with the *TPRpred* server[36] revealed no such motifs in NatC. Moreover, while TPR motifs typically consist of helices containing 12–15 residues (Zeytuni and Zarivach 2012), NatC has a very heterogenous distribution of helix lengths, with nine helices having 11 or fewer residues, three helices in the range of 12–15 residues and fourteen helices ranging from 16 to 45 residues.

Strikingly, the relative position of the NatC auxiliary subunit compared to the catalytic subunit, is exactly opposite to the positions of the NatA and NatB auxiliary subunits. Whereas Naa15 and Naa25 primarily engulf the N-terminal part of their catalytic subunits, Naa35 additionally encloses the C-terminal half of Naa30. In NatA and NatB, the catalytic β6–β7 loop makes no contact with the

[36] https://toolkit.tuebingen.mpg.de/#/tools/tprpred

corresponding auxiliary subunit. In NatC, this loop-region, which is critical for binding of the peptide substrate (see 4.8.3), is in direct contact with the auxiliary subunit Naa35 (see 4.4).

4.5.1 Comparison of the catalytic subunits α1-loop-α2 regions with their auxiliary subunit(s)

While the monomeric NAT ortholog NatE is active on its own (Liszczak, Arnesen, and Marmorstein 2011), the catalytic subunits of the heterodimeric (NatA, NatB) or heterotrimeric (NatC) NAT complexes require their corresponding auxiliary subunit(s) for proper catalytic function. The auxiliary subunits are necessary for a proper positioning of the α1–loop–α2 region for full catalytic activity for their native substrates.

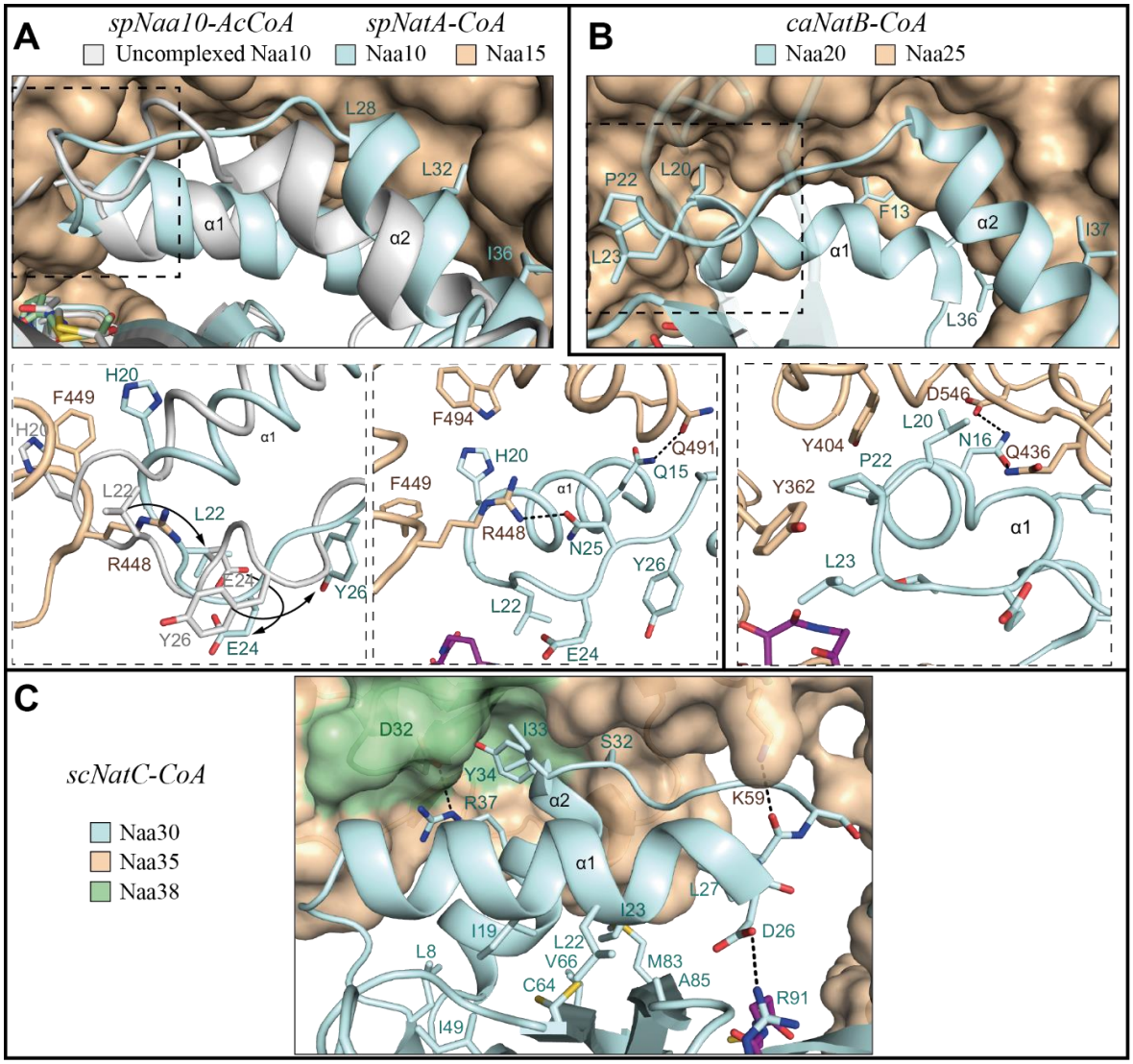


Fig. 24 Comparison of the α1–loop–α2 region of NatA, NatB and NatC complexes (A) Superposition of monomeric, uncomplexed *S. pombe* Naa10 (spNaa10-AcCoA) with the SpNaa10/Naa15 complex (*SpNatA-CoA*). (B) Naa20/Naa25 complex from *C. albicans* (*caNatB-CoA*). (C) Naa30/Naa35/Naa38 complex from *S. cerevisiae* (*scNatC-CoA*).

Liszczak et al. (2013) have solved the structure of both the uncomplexed, monomeric Naa10 and the Naa10/Naa15 complex. A superposition of the uncomplexed Naa10 with the Naa10/Naa15 complex (NatA) shows that the α1–loop–α2 region undergoes significant conformational changes in the

presence of its auxiliary subunit Naa15 (Fig. 24A), which results in large movements of Leu22, Glu24 and Tyr26 (left zoom view), all of which are essential for catalysis. The α1–loop–α2 region of complexed Naa10 makes multiple hydrophobic interactions with Naa15, in addition to a few hydrogen bonds.

In the NatB (Naa20/Naa25) complex, similar interactions of the Naa20 α1–loop–α2 region with the auxiliary subunit Naa25 are found (Hong et al. 2017). Fig. 24B shows that the interface is dominated by hydrophobic interactions from both the α1 and α2 helices.

In NatC, α2 of Naa30 is in contact with the two auxiliary subunits (Fig. 24C, described in detail in section 4.4). Most importantly, Tyr34 from Naa30 is placed in a hydrophobic pocket that is lined by residues from both Naa35 and Naa38. Additionally, Lys59 of Naa35 makes a hydrogen bond with the carbonyl oxygen backbone of Naa30, stabilizing the α1–loop–α2.

In contrast to the NatA and NatB complexes, α1 of the catalytic subunit has no contact to any of the two auxiliary subunits but makes several intramolecular hydrophobic interactions with α2 and the β-sheet. Moreover, Asp26 forms a hydrogen bound with Arg91 from the conserved CoA-binding motif (Q/R-x-x-G-x-G/A) in the β4–α3 loop. These comparisons show that although the overall architecture of the auxiliary subunits is completely different in the three NAT complexes, the regions important for catalysis in the catalytic subunit are stabilized in a similar fashion, e.g. similar mechanistic principles apply for all of the NAT complexes.

4.6 *Structural comparisons of NatC*

A comparative analysis of the three NatC subunits structures with known structures from the PDB database was used to detect related protein folds that may not have been previously derived from multiple sequence alignments. Thus, although protein sequences can be altered over the time course of evolution, in particular when proteins adopt new functions, their 3-dimensional folds often remain conserved (Holm and Sander 1996).

The structures of the three NatC subunits were compared against the complete PDB database using the *DALI* web server[37] (Holm and Rosenström 2010) (see 3.12). The top-ranked match and a few additional matches with non-redundant protein sequences are summarized in Table 19.

Table 19 Top structural similarities of NatC subunits to the PDB database. *NatC-CoA* was used as a query structure (*DALI* search). Selected non-redundant hits are shown. Z, *DALI* Z-score (significance of pairwise comparison); rmsd, root-mean-square-deviation of aligned Cα atoms in Å; lali, number of aligned positions; nres, residue number of the matched structure; %id sequence identity between aligned structures. Full species names are *Saccharolobus solfataricus* and *Schizosaccharomyces pombe*.

#	PDB-chain	Z	rmsd	lali	nres	%id	PDB structure	Res	Species	Author
Selected Naa30 hits										
1	4lx9-A	22.7	1.6	149	157	27	Archaeal Ard1 ortholog (ssNAT) bound to AcCoA	2.0	*S. solfataricus*	(Liszczak and Marmorstein 2013)
12	4kvo-E	22.2	1.7	150	153	25	NatA (**Naa10**/Naa15) complex bound to a AcCoA	3.2	*S. pombe*	(Liszczak et al. 2013)
70	4kvx-B	19.3	2.3	147	152	22	Crystal structure of Naa10 (Ard1) bound to AcCoA	2.0		
Selected Naa35 hits										
1	3zxp-C	11.5	3.8	225	384	7	Bro1 Domain-Containing Protein Brox;	2.5	*H. sapiens*	(Zhai et al. 2011)
5	5mk1-B	11.1	3.8	214	358	4	Tyrosine-Protein Phosphatase Non-Receptor Type 23	2.5	*H. sapiens*	(Gahloth et al. 2017)
8	1zb1-A	11.0	3.9	207	367	8	Bro1 protein	2.0	*S. cerevisiae*	(Kim et al. 2005)
42	2r02-A	9.0	3.6	205	697	5	Alix / Programmed Cell Death 6-Interacting Protein;	2.6	*H. sapiens*	(Zhai et al. 2007)
Selected Naa38 hits:										
1	3jcm-h	11.6	1.8	72	77	24	Spliceosomal U4/U6.U5 tri-snRNP (Cryo-EM structure)	3.8	*S. cerevisiae*	(Wan et al. 2016)
3	4m7d-G	11.5	1.8	72	77	24	Lsm2-8 complex bound to RNA fragment CGUUU	2.6	*S. cerevisiae*	(Zhou et al. 2013)
28	4m7a-G	11.0	1.7	69	74	23	Lsm2-8 complex bound to RNA fragment CGUUUU	2.8		

4.6.1 The Naa30 subunit adopts the typical Gcn5-related N-acetyltransferase (GNAT) fold

The top hit for the Naa30 subunit is the archaeal Ard1 ortholog from *Sulfolobus solfataricus* (ssNAT), which is the only NAT enzyme identified in archaea and believed to be the ancestral NAT variant from which the eukaryotic NAT machinery evolved (Liszczak and Marmorstein 2013).

[37] http://ekhidna2.biocenter.helsinki.fi/dali/

The RMSD of Cα atom positions (149 aligned atoms) between Naa30 and ssNAT is 1.6 Å. With a sequence identity of 27% and a Z-score[38] of 22.7, ssNAT is considered a close homolog (Table 19). ssNAT and Naa30 contain the same number of secondary structure elements and have a very similar fold (Fig. 25, left). The greatest difference between the two structures is the extended β3–β4 loop found in ssNAT, which is the only non-conserved core region relative to eukaryotic NATs and has a limited stabilizing effect on the 3D fold of the ssNAT active site (Liszczak and Marmorstein 2013).

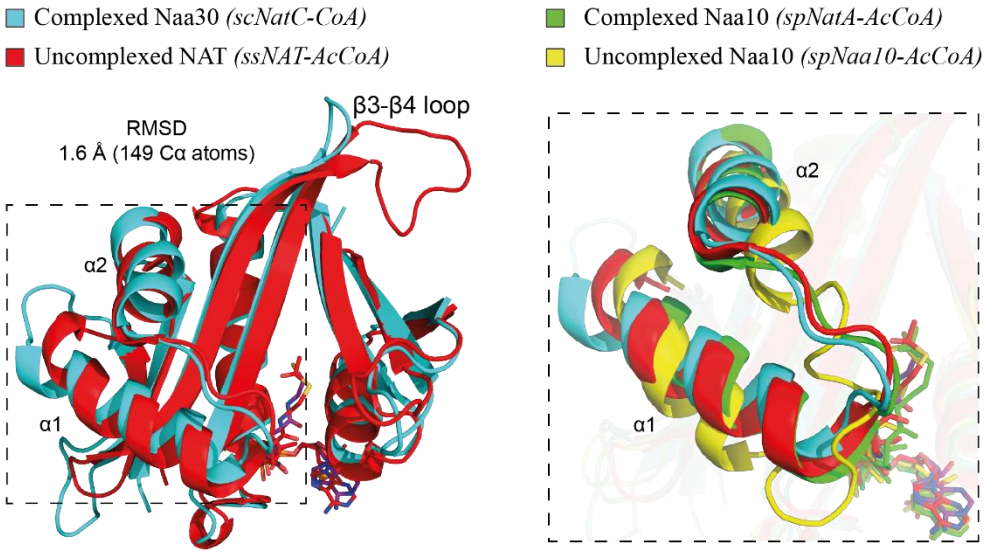


Fig. 25 Selected DALI matches for the Naa30 subunit. Naa30 from the *NatC-CoA* structure was used for a structural alignment with the complete PDB database. A superposition of the best match (ssNAT) is shown on the left. A zoom view (right) of the α1–loop–α2 region with two additionally aligned strong DALI matches.

The next best, non-redundant match for Naa30 is the Naa10 subunit of the *Schizosaccharomyces pombe* NatA (Naa10/Naa15) complex with an RMSD of 1.7 Å (over 150 Cα atoms). The same spNaa10 protein, in its uncomplexed monomeric form, has a significantly higher RMSD of 2.3 Å (147 Cα atoms), mostly due to significant changes in the α1–loop–α2 region of uncomplexed Naa10 (Fig. 25, right). This region undergoes significant changes upon complex formation with Naa15 that are essential for catalysis. The α1–loop–α2 region of Naa30 adopts the same conformation as the complexed Naa10, as does the monomeric ssNAT, which is catalytically active without the aid of an auxiliary subunit (Liszczak et al. 2013) (Mackay et al. 2007).

In summary, the catalytic subunit Naa30 takes up the conserved GNAT fold and is most similar to the monomeric archaeal ssArd1 ortholog, followed by the catalytic subunit Naa10 of the spNatA (Naa10/Naa15) complex. Through complex formation with Naa35, the catalytic α1–loop–α2 region of Naa30 is positioned in the same orientation as in the ssArd1 and spNatA structures, which most likely facilitates the proper catalytic function and/or substrate recognition of NatC.

[38] The similarity between structures is ranked by a "Z-score" (Holm et al. 2008). Significant similarities have a Z-score above 2 and usually correspond to similar folds. "Strong matches" have a sequence identity above 20% or a Z-score above a cutoff that was empirically set to n/10−4, where n is the number of residues in the query structure.

4.6.2 Naa35 contains a fold reminiscent of the Bro1 domain

DALI identified part of the Naa35 subunit to resemble the Bro1 domain of a variety of proteins. Among these Bro1 domain-containing proteins is the *S. cerevisiae* Bro1 protein (Kim et al. 2005) and three human proteins, including the mammalian ortholog of yeast Bro1, Alix (Zhai et al. 2007), which is also known as the programmed cell death 6-interacting protein (PDCD6IP) or the ALG-2-interacting protein 1 (AIP1). Further hits include the Bro1 domains from the Bro1 Domain-Containing Protein Brox (Zhai et al. 2011) and the Tyrosine-Protein Phosphatase Non-Receptor Type 23 (Gahloth et al. 2017).

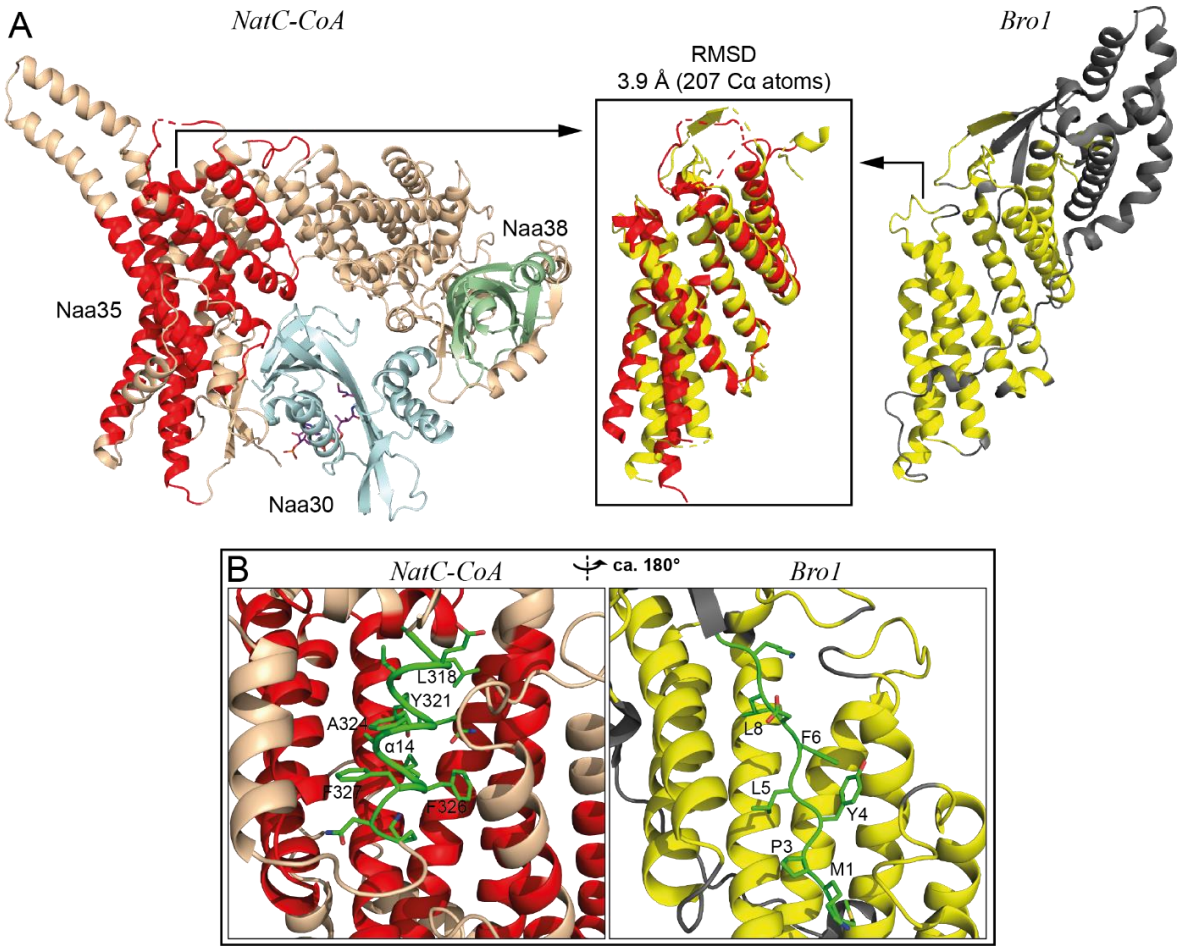


Fig. 26 A segment of Naa35 exhibits structural similarity to the Bro1 domain. (A) Naa35 from *NatC-CoA* aligned with the Bro1 domain of the *S. cerevisiae* Bro1 protein. The Bro1 protein is shown on the right. A superposition (middle) shows only the aligned regions, which are colored in red (Naa35) or yellow (Bro1). A total of 207 residues out of 725 (Naa35) or 367 (Bro1) were aligned. (B) Zoom view showing the TPR-like pocket, which in Bro1 is filled by its N-terminal eight residues (right). In NatC, this pocket is occupied by helix α14.

The aligned Naa35 segment comprises helices α15 to α19, the N-terminal half of α20, the C-terminal half of α22, as well as helices α25 and α26. The Cα atom RMSD between Naa35 and the different Bro1 domains is between 3.6 and 3.8 Å (Table 19). The sequence identity is only 4-8% and the Z-score of 9–11.5 is well below the cutoff (≥ 68.4), suggesting only a distant relationship.

The Bro1 domain of the yeast Bro1 protein was chosen for the alignment with the relevant segments of Naa35 and is shown in Fig. 26A. The Bro1 domain was first structurally characterized by (Kim et

al. 2005) for the identically named *S. cerevisiae* Bro1 protein. The domain contains a folded conserved core of nearly 370 residues which forms 14 α-helices and three β-sheets, which resembles a boomerang- or banana-shape. The bulk of the structure consists of an 11-helical solenoid. The Bro1 domain shows limited similarity to the tetratricopeptide repeat domain (TPR), which was found to be present in NatA and NatB but could not be identified for NatC. However, whereas the helices in the Bro1 domain are typically 22-30 residue long, TPR repeats have much shorter helices of only 12-15 residues (Zeytuni and Zarivach 2012). The N-terminal eight residues of Bro1 fill a TPR-like pocket on the convex side of Bro1, which is necessary for the stability of the Bro1 domain. In NatC, this pocket is filled by helix α14 (Fig. 26B).

In summary, structural comparisons of the Naa35 subunit identified a variety of proteins with distantly related folds and apparently different functions. Possible implications are discussed in section 5.2.

4.6.3 The Naa38 subunit resembles the fold of an Sm-like protein

Even before Naa38 was identified as a part of the NatC complex (Polevoda and Sherman 2001), the protein was suggested to be a Sm-like protein based on protein sequence similarity (Séraphin 1995). The current structural work confirms the structural similarity of Naa38 with Sm-like proteins. With a RMSD of 1.8 Å, Naa38 is most similar to the Sm-like protein Lsm4. Out of the 72 aligned residues, 17 are identical, which gives a sequence identity of 23.6%. With a Z-score of 11.6, Lsm4 is considered a strong match (Table 19).

Lsm4 is part of a heptameric complex that consist of seven different Lsm proteins. Together, they form a heptameric, donut-shaped assembly (Zhou et al. 2013), which binds a specific small nuclear rRNA (snRNA), U6.[39] This RNA-protein complex, referred to as “small nuclear ribonucleoprotein” (snRNP) is part of the spliceosome.

In Fig. 27, Naa38 is superimposed (aligned over 69 Cα atoms with an RMSD of 1.7 Å) with the Lsm4 protein of the heptameric, doughnut-shaped Lsm2-8 ring (Zhou et al. 2013). Each Lsm protein adopts the highly conserved Sm fold (Kambach et al. 1999), which contains an N-terminal α-helix and an antiparallel β-sheet, with the order β4–β3–β2–β1–β5. The β-sheet of each Lsm protein is continued through intermolecular interactions of the β4 and β5 sheets of neighboring Lsm proteins (Fig. 27B, left).

Naa38 exhibits the same five-stranded anti-parallel β-sheet as the Lsm proteins. However, the β–sheet of Naa38 is extended by three short β-strands from the Naa35 subunit (Fig. 27B, right).

One end of the Naa38 β-sheet is extend by the anti-parallel β1-strand of Naa35. The other end of the Naa38 β-sheet is completed by the parallel β3-strand of Naa35.

[39] The designation “U” reflects the large fraction of uracil nucleobases in these RNAs.

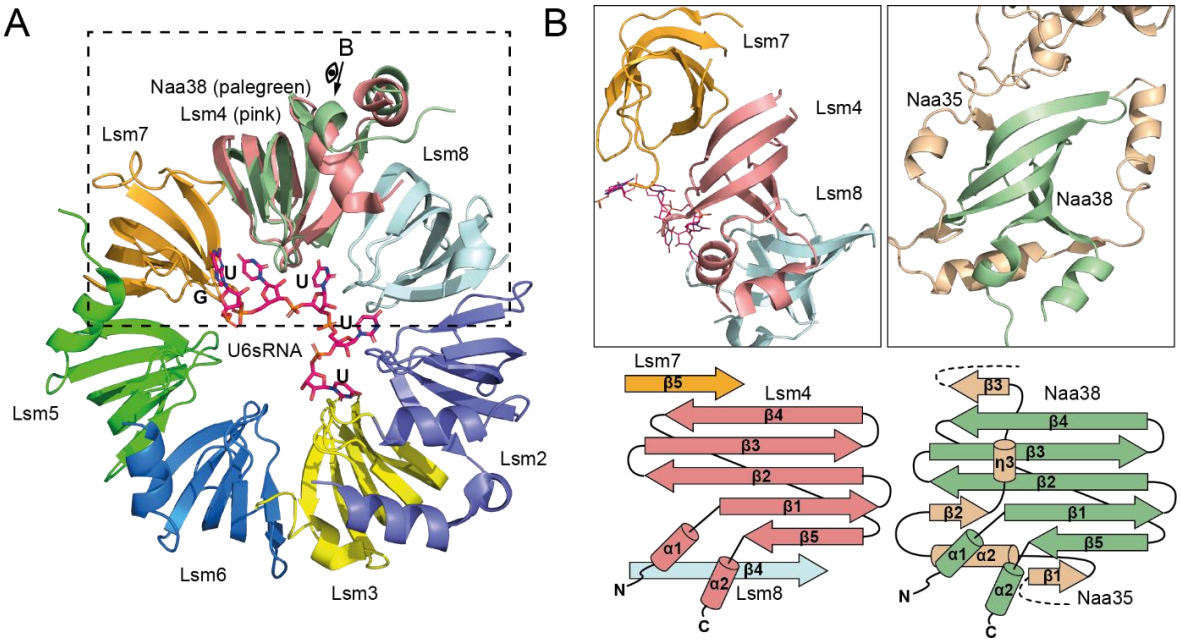


Fig. 27 Superposition of Naa38 with the yeast Lsm2-8 complex (A) Naa38 from *NatC-CoA* aligned with Lsm4 of the Lsm2-8 heptameric complex (PDB accession code 4M7A). (B) Comparison of the β-sheet topology of Lsm4 in contact with its neighboring Lsm units (Lsm7 and Lsm8) and the Naa35-Naa38 β-sheet. (C) Sequence alignment of all yeast Sm-like (Lsm) proteins and Naa38.

In summary, structural comparisons of the Naa38 subunit confirmed the previously described similarity of Naa38 with Sm-like proteins (Séraphin 1995). However, the intermolecular β-sheet hydrogens bonds between Naa35 and Naa38, are located roughly at the same position as the intermolecular hydrogen bonds formed between neighboring Lsm subunits. Thus, these Naa38 interfaces are not available for homo- or hetero-oligomerization with other Lsm proteins. Nevertheless, the structural similarity of Naa38 with the RNA-associated Lsm proteins might suggest that Naa38 is also capable of binding (small nuclear) RNA, which is further discussed in section 5.2.

4.7 *Evolutionary conservation of the NatC complex*

A surface representation of *NatC-CoA*, colored according to the degree of conservation of the underlying amino acids is shown in Fig. 28. In contrast to a multiple sequence alignment, this representation allows for the identification of conserved residues not only proximate in sequence but also in three-dimensional space. Conserved surfaces, in turn, have often a conserved function.

The conserved surfaces of NatC are located around the two prominent tunnels in the center of the NatC complex. CoA and the substrate peptide meet head-to head in the *side tunnel*, which contains the active site of the catalytic subunit Naa30 and is surrounded by the majority of conserved residues. Their role in substrate binding and catalytic activity will be explored in sections 4.8.2 to 4.8.3 and sections 4.10.3 to 4.10.4, respectively.

The *central tunnel* is lined by loop regions from both Naa30 and Naa35, including the highly conserved catalytic α1–α2 and β6–β7 loops, which separate the two tunnels. While the majority of residues in this region are likely to be involved in substrate binding and catalytic function, the high degree of conservation indicates a broader function of the tunnel structure, which will be explored in section 4.9.

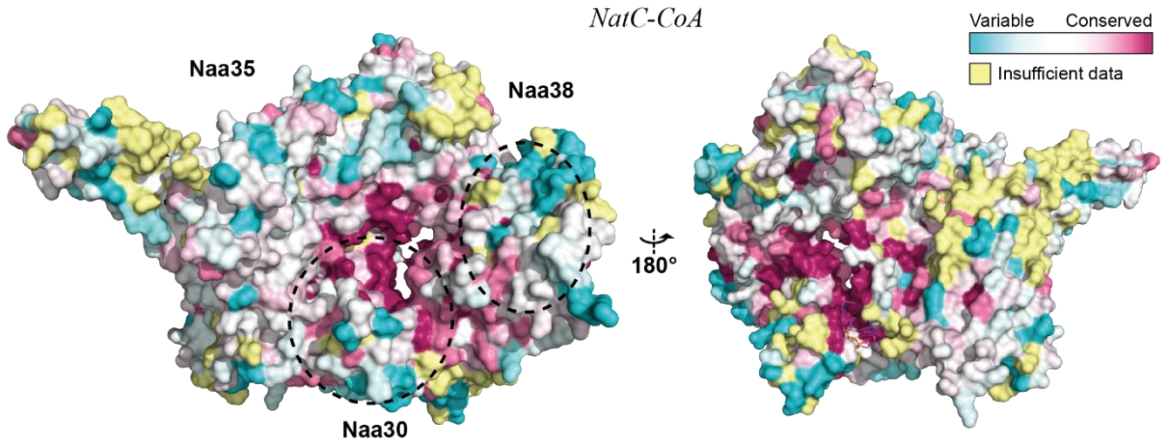


Fig. 28 Amino acid conservation projected on the surface of NatC. Analyzed for the *NatC-CoA* structure with the *ConSurf* server. Variable regions are colored in cyan; average conservation is shown in white and completely conserved regions are highlighted in purple. Conservations scores were calculated separately for each NatC subunit using multiple sequence alignments (generated with *MUSCLE)* of amino acid sequences from 18 different species. Approximate positions of the Naa30 and Naa38 subunits are indicated by circles.

4.8 *Structures of the NatC complex bound to peptide ligands*

4.8.1 Structure determination

To obtain further insights into substrate binding and the catalytic mechanism, NatC was co-crystallized with Coenzyme A and a decameric peptide, containing the N-terminal five residues of either of two proteins known to be acetylated by the NatC complex, followed by a GS-linker and a triple-arginine at the C-terminus to facilitate peptide solubility. The first substrate peptide (MFHLVGSRRR) contains the N-terminal five residues of the *Saccharomyces cerevisiae* ADP-ribosylation factor-like protein 3 (yArl3). N-terminal acetylation of yArl3 is required for its binding to the *trans*-Golgi specific transmembrane protein Sys1, thus determining the localization of yArl3 at the *trans* face of the Golgi apparatus (Behnia et al. 2004; Setty et al. 2004). The second substrate (MLRFVGSRRR) contains the N-terminal five residues of the Major capsid protein (Gag) of the *S. cerevisiae* virus L-A, another known natural NatC substrate (Tercero, Dinman, and Wickner 1993).

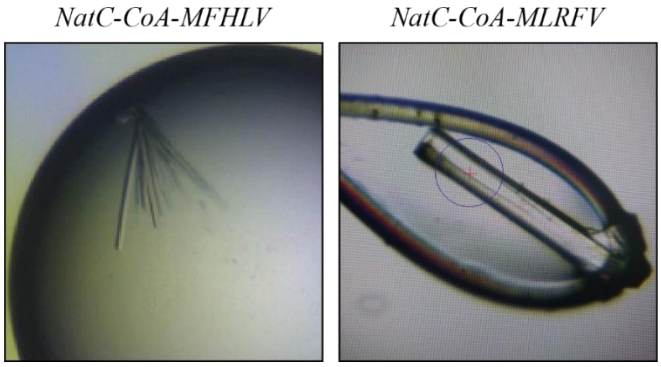


Fig. 29 Diffraction quality crystals of the NatC complex bound to CoA and peptide ligands.

Crystals of both complexes were obtained using a crystallization solution containing 14.5–16.5% PEG 4000, 150 mM ammonium iodide and 100 mM sodium citrate, pH 6.2–6.3 (see 3.7). Both crystals appeared after 12 – 48 h and required another five days to reach their maximum dimensions (Fig. 29). The crystal comprising NatC bound to coenzyme A and the yArl3 peptide (*NatC-CoA-MFHLV*) diffracted X-rays to a maximum resolution of 2.99 Å and belongs to the orthorhombic space group $P2_12_12_1$ with cell dimensions a=48.3 Å, b=139.7 Å, c=166.8 Å, α=β=γ=90°. The crystal with NatC bound to CoA and the Gag peptide (*NatC-CoA-MLRFV*) diffracted X-rays to a maximum resolution of 2.75 Å and belongs to the same space group with dimensions a=48.1 Å, b=134.9 Å, c=165.7 Å, α=β=γ=90°. Data collection statistics for both crystals are found in Table 20.

Table 20 Data collection statistics for crystals of NatC bound to CoA and the yArl3 or Gag peptide. Values in parentheses are for highest-resolution shell.

Data collection	NatC-CoA-MFHLV (native)	NatC-CoA-MLRFV (native)
Space group	$P2_1\ 2_1\ 2_1$	$P2_1\ 2_1\ 2_1$
Cell dimensions		
a, *b*, *c* (Å)	48.30 139.72 166.82	48.14 134.93 165.74
α, *β*, *γ* (°)	90 90 90	90 90 90
Wavelength (Å)	0.9184	0.9184
Resolution range (Å)	45.65 – 2.99 (3.10 – 2.99)	43.41 – 2.75 (2.85 – 2.75)
Completeness (%)	98.7 (98.6)	99.5 (96.0)
R_{sym} (%)	16.6 (94.6)	19.4 (15.1)
Redundancy	4.0 (4.2)	6.6 (6.6)
I/σ (I)	8.4 (1.5)	9.3 (1.2)

The structures of the peptide ligand-bound NatC complexes were solved by molecular replacement (see 3.9.2), using the refined structure of the SeMet-labeled NatC complex as an input model. The structures of *NatC-CoA-MFHLV* and *NatC-CoA-MLRFV* were refined to a R_{work}/R_{free} of 19.9%/24.8% and 20.4%/24.8% , respectively (see Table 21).

Table 21 Refinement statistics for NatC bound to CoA and the yArl3 or Gag peptide. Values in parentheses are for highest-resolution shell. [a]R_{free} was calculated for 5% of the reflection data.

Refinement	NatC-CoA-MFHLV (native)	NatC-CoA-MLRFV (native)
Resolution range (Å)	45.65 – 2.99 (3.10 – 2.99)	43.41 – 2.75 (2.85 – 2.75)
No. of reflections	23299 (2274)	28783 (2749)
R_{work} / R_{free} (%)[a]	19.9 / 24.8	20.4 / 24.8
No of atoms		
Protein	7753	7734
Ligands	86	84
Water	52	93
B factors (Å^2)		
Average	63.8	67.4
Protein	63.9	67.6
Ligands	67.8	78.3
Water	40.0	44.4
RMSD		
Bond lengths (Å)	0.003	0.004
Bond angles (°)	0.86	0.85
Ramachandran plot (%)		
Most favored	98.6	98.5
Allowed	1.4	1.5
Outlier	0	0
Rotamer outliers (%)	0.34	0.68
Clashscore	3.8	5.4

In Fig. 30, electron density maps (2Fo-Fc, contoured at 1 σ) for ligands in the active site are shown for all three ligand-bound structures (*NatC-CoA*, *NatC-CoA-MFHLV* and *NatC-CoA-MLRFV*) described in this work. The electron density was sufficient to model a coenzyme A molecule in all three structures. However, the electron density was clearly strongest for *NatC-CoA-MFHLV*, followed by *NatC-CoA-MLRFV* and weakest for *NatC-CoA*, which is partially reflected by the different ligand occupancies of 100%, 93% and 86% respectively. In the *NatC-CoA* structure, the adenosine diphosphate moiety, and the pantoic acid moiety of CoA are clearly identified in the 2Fo-

FC map (Fig. 30, right part of the molecule). However, no electron density can be seen for the connecting β-alanine moiety (NH-CH_2-CH_2-CO) and the final β-mercapto-ethylamine (NH-CH_2-CH_2-SH) moiety, except for the final thiol group, which is again clearly defined. The absence of electron density for parts of CoA is explained by a higher B-factor of the atoms in the corresponding region (Fig. 31). The electron density of the peptide ligands was sufficient to model the first four amino acids of the peptide 1-MFHLVGSRRR-10 and the first six residues of the peptide 1-MLRFVGSRRR-10 in the *NatC-CoA-MFHLV* or *NatC-CoA-MLRFV* structures, respectively.

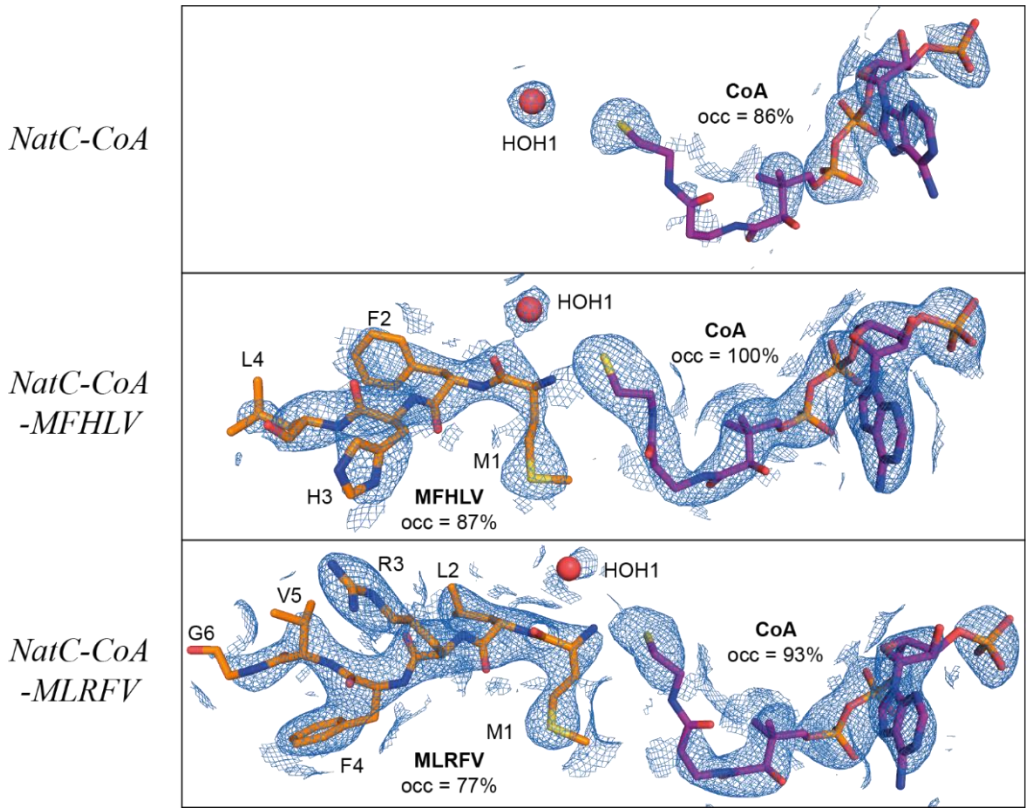


Fig. 30 2Fo-Fc electron density maps for ligands in the active site of the NatC complex. 2Fo-Fc maps (contoured at 1σ) are shown for CoA, peptide ligands and a highly coordinated water molecule in the center of the active site (HOH1). Average ligand occupancies are written next to each ligand.

Moreover, a clear electron density was visible for a tightly coordinated water molecule (red sphere) in the active site of the NatC complex. In all three structures, the B-factor of the water molecule was very similar to the average B-factor of surrounding macromolecule atoms (Fig. 31), indicating that this water is well coordinated. The tight coordination, together with the close proximity to the amino group of the peptide substrate, suggests a potential catalytic role of this water molecule.

In contrast, the B-factors of CoA and peptide ligands are higher than those of the surrounding macromolecule atoms and thus ligand occupancies might be lower than estimated. However, at the given resolutions (2.5–3.0 Å), it is not possible to differentiate between a reduced ligand occupancy and a higher B-factor. The partial ligand occupancies are most likely a result of the relatively low ligand concentrations (135–225 µM) that were used in the crystallization setup (see Table 14, section 3.7). To ensure full occupancy for all ligands, it thus advisable to use higher ligand concentrations (>10× K_m) in future crystallization setups.

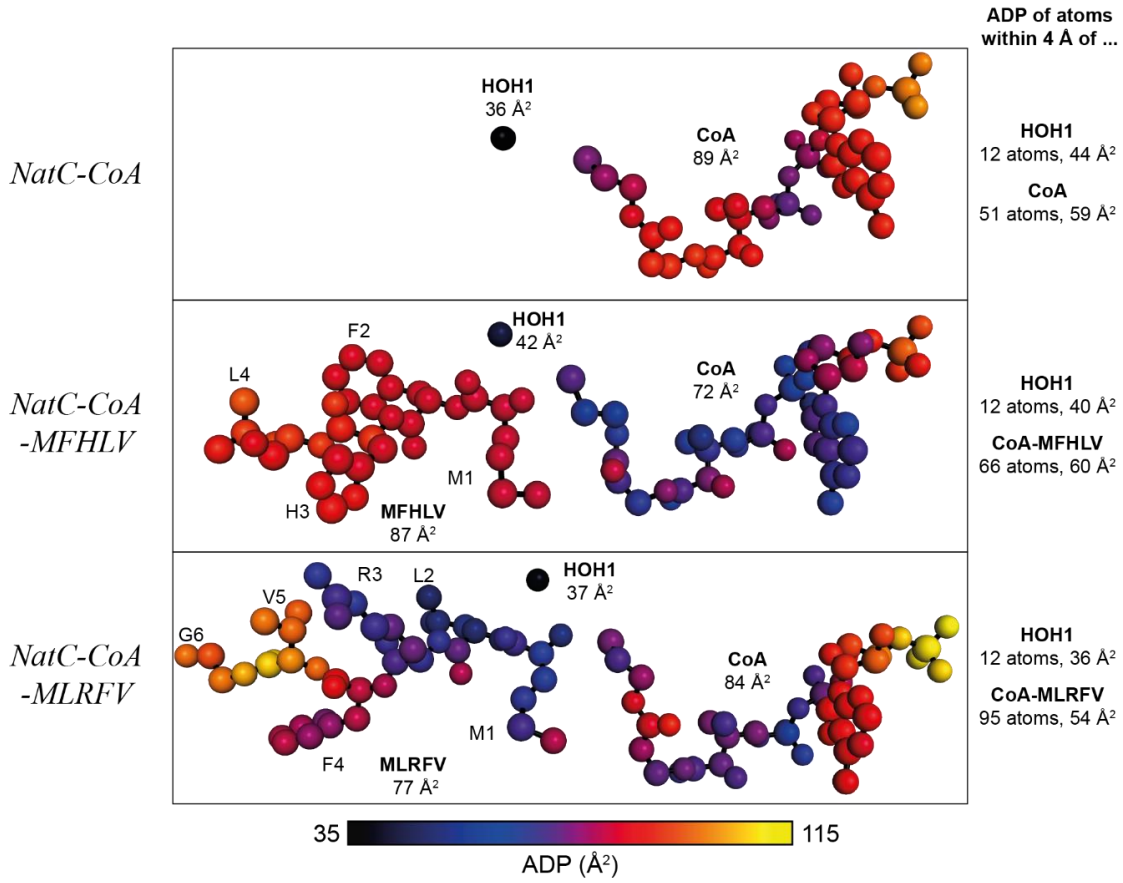


Fig. 31 Atomic displacement factors of ligands in the active site of the NatC complex. Ligand atoms are colored with a black-blue-red-yellow spectrum based on their B-factor. Average ligand B-factors are written next to the corresponding ligand.

4.8.2 Peptide binding to the NatC complex

The two peptide-bound NatC structures reveal the orientation of the peptide N-terminus in the center of the active site. Furthermore, they reveal how NatC establishes its substrate specificity toward protein N-termini that start with their initiator methionine, followed by a hydrophobic amino acid at position two.

A detailed analysis of the interactions of the yArl3 peptide (MFHLVGSRRR) with the NatC complex (Fig. 32 B, left) reveals that the yArl3 peptide forms a total of five hydrogen bonds with Naa30, which are mostly mediated via the yArl3 peptide backbone. Thus, the yArl3 amine group of Met1 forms a hydrogen bond with the backbone carbonyl oxygen of Glu118 and the Met1 carbonyl oxygen interacts with the hydroxyl group of Tyr145. The backbone amide of Phe2 forms a hydrogen bond with the carbonyl oxygen of Gly82 and the Phe2 carbonyl oxygen with Tyr31. The only hydrogen bond between a yArl3 side chain and Naa30 is formed between the imidazole ring of His3 and the carboxyl group of Glu29.

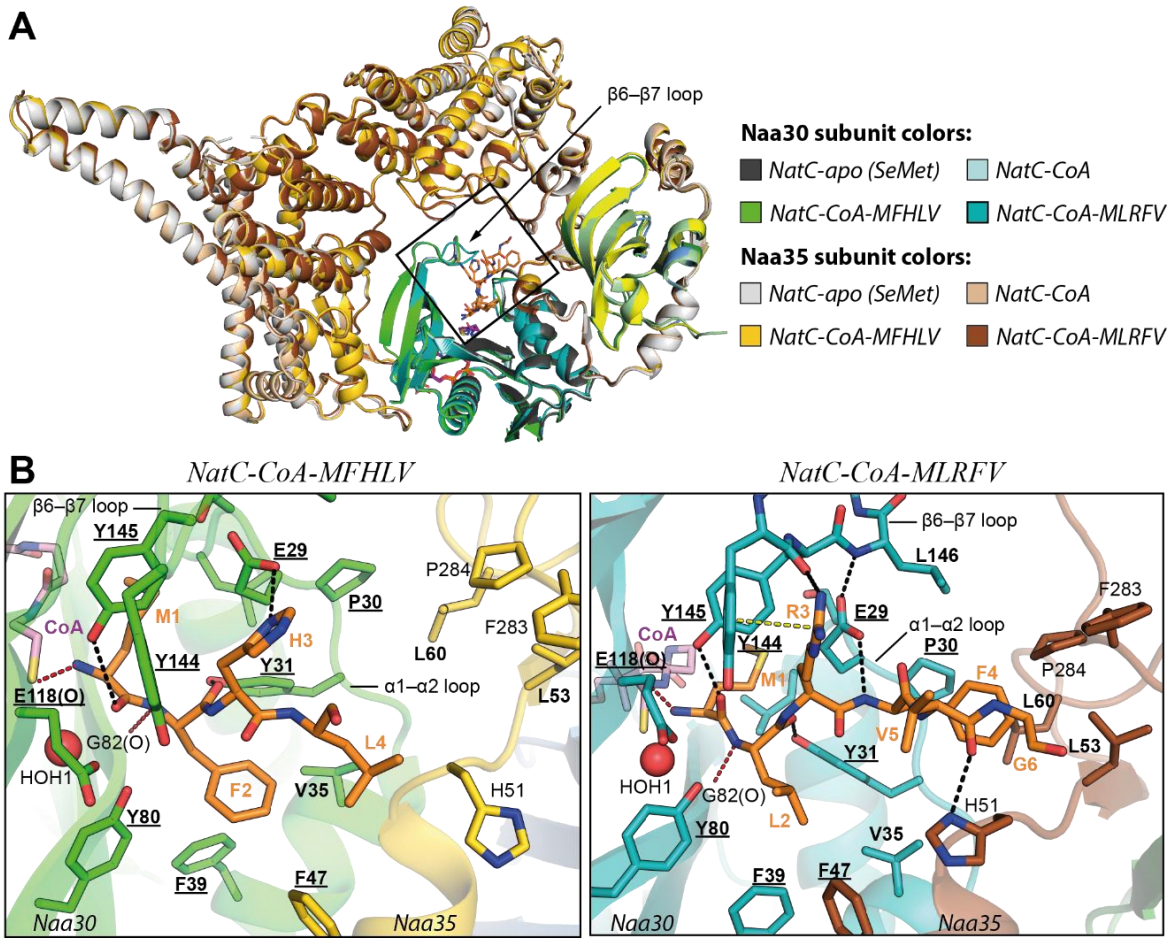


Fig. 32 The peptide binding site of NatC. (A) Superposition of *NatC-apo*, *NatC-CoA*, *NatC-CoA-MFHLV* and *NatC-CoA-MLRFV* showing the large conformational change of the β6-β7 loop in the *NatC-CoA-MLRFV* structure. (B) Interactions of the two peptide substrates (in orange): yArl3 (MFHLVGSRR) and Gag (MLRFVGSRR) with Naa30 and Naa35. The proposed catalytic water (HOH1) is shown as a sphere. Hydrogen bonds formed between two backbone chains are colored in red. Cation-π interactions are yellow and all other H-bonds are black.

The Gag peptide (MLRFVGSRRR) with the N-terminal five residues from the *S. cerevisiae* L-A virus forms a total of six hydrogen bonds with Naa30 (Fig. 32 B, right). Like the yArl3 peptide, the Gag peptide Met1 amine group forms a hydrogen bond with the backbone carbonyl oxygen of Glu118 and the oxygen forms a hydrogen bond with the Tyr145 hydroxyl. The backbone amide of Leu2 also forms the hydrogen bond with Gly82 and its carbonyl oxygen forms an H-bond with Tyr31.

The guanidyl group of Arg3 forms a hydrogen bond with the backbone carbonyl oxygen of Naa30 Tyr144. Moreover, a cation-π interaction is formed between the Arg3 guanidyl group and the aromatic ring of Tyr144, in a parallel geometry. In contrast to the yArl3-bound structure, the carboxyl group of Glu29 in the Gag-bound structure forms a hydrogen bond with the backbone amide of Phe4. This hydrogen bond cannot be formed in the yArl3 structure because the bulky imidazole group of yArl3-His3 sterically hinders Glu29 from interacting with the fourth peptide backbone amide. Last but not least, one more hydrogen bond is formed for the Gag peptide between the carbonyl oxygen of Val5 and the imidazole ring of His51 from subunit Naa35.

In addition to hydrogen bonds, both substrate peptides exhibit hydrophobic interactions of their side chains with a large hydrophobic pocket in the NatC complex. One side of the hydrophobic ligand-binding pocket is formed by the conserved Naa30 residues: Val35, Phe39, Tyr80 and, slightly further away, residues Tyr38 (Naa30) and Phe47 from Naa35. Both the Phe2 side chain of the yArl3 peptide and Leu2 side chain of the Gag peptide are positioned in this pocket. However, the hydrophobic pocket extends further and exhibits a second partially hydrophobic surface on the opposite side. This side is lined by the side chains of Naa30 residues Pro30 (its hydrophobic face) and Leu146 and the Naa35 residues Leu53, Leu60, Phe283 and the hydrophobic face of Pro284. This pocket is only explored by Phe4 of the Gag peptide, while the side chains of yArl3-Leu4 is positioned 4.7 Å away.

In summary, both peptides interact with the catalytic subunit Naa30 through formation of multiple hydrogen bonds, four of which involve the backbone amides and carbonyl oxygen atoms of the first two residues of the peptide. Both their hydrophobic side chains at position two are positioned in a conserved hydrophobic pocket of the catalytic subunit. In comparison, the Gag peptide forms more interactions with NatC, including a cation-π interaction of Arg3 and an additional hydrogen of the fourth backbone nitrogen. In addition, Phe4 explores a second hydrophobic pocket and thus the Gag peptide of the L-A virus is likely to be the stronger binder, which is in line with the data from the acetyltransferase assays (see sections 4.10.3 and 4.10.4).

4.8.3 Peptide ligand-induced conformational changes

The *NatC-CoA-MFHLV* and *NatC-CoA-MLRFV* structures undergo significant structural changes in the catalytic α1–α2 loop compared to the *NatC-CoA* reference structure (Fig. 33 A). While almost no movement can be seen between the *NatC-apo* and *NatC-CoA* structures, binding of the yArl3 or Gag peptides induces a α1–α2 loop movement of 2.8 Å or 3.5 Å, respectively, compared to the ligand-free *NatC-apo* structure. Upon binding of the yArl3 ligand, the hydroxyl group of Ser28 (α1–α2 loop) comes into hydrogen bonding distance with the carboxyl group of Glu148 (β6–β7 loop), thereby stabilizing catalytic residues of both loops and priming them for catalysis.

However, the most significant structural rearrangements of the catalytic β6-β7 loop can be seen in the *NatC-CoA-MLRFV* structure. In all NatC structures, the Naa30 β6-β7 loop is in contact with subunit Naa35. In the *NatC-CoA* structure, two hydrogen bonds are formed between the conserved Asn147 of subunit Naa30 and Gln106 of subunit Naa35. Asn147 is stabilized by additional hydrogen bonds from Arg143 (Naa30) and Arg286 (Naa35). This interaction is almost identical in the yArl3 peptide-bound *NatC-CoA-MFHLV* structure. However, in the Gag peptide-bound *NatC-CoA-MLRFV* structure the β6–β7 loop undergoes large rearrangements, so that Glu148 takes up the position of Asn147 and forms two hydrogen bonds with Gln106. As a result, the conserved Asn147 is now facing the α1–α2 loop. Compared to the *NatC-apo* structure, the side chain amide group has traversed a distance of 8.8 Å. Due to this structural rearrangement, the side chain of Asn147 can now form a bifurcated intramolecular hydrogen bonds with the hydroxyl group of Ser28 and the carboxyl group of Glu120, thereby stabilizing the entire β6–β7 loop.

The significant structural changes in the β6–β7 loop are a direct consequence of the binding of the Gag peptide (MLRFVGSRRR). The strongest driving force for this rearrangement can most likely be attributed to the Arg3 side chain of the peptide, as detailed in the previous section (4.8.2). In addition, the backbone amide of Gag-Phe4 forms a hydrogen bond with the carboxyl group of Glu29 (loop α1–α2), which itself forms another hydrogen bond with the backbone amide of Leu146, thereby providing additional stability for the β6–β7 loop.

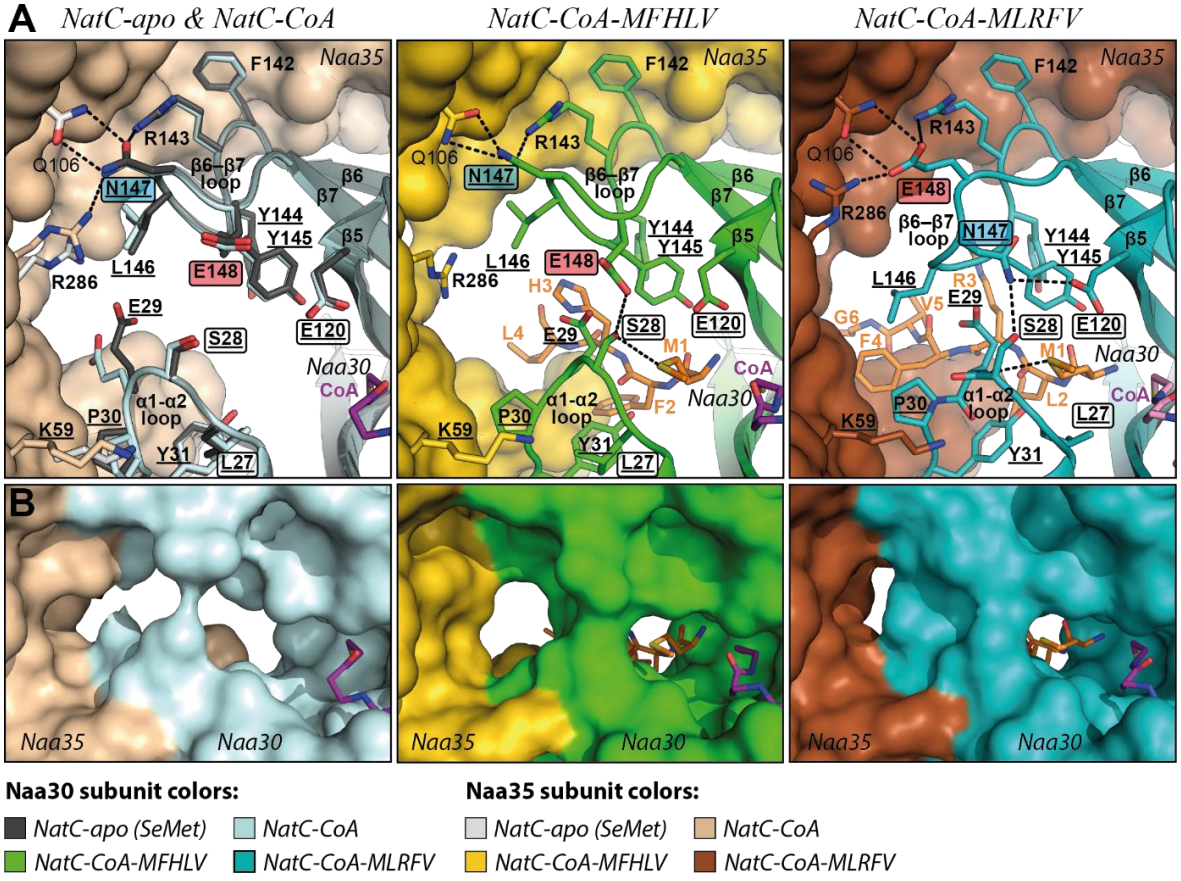


Fig. 33 Peptide ligand-induced conformational changes (A) Mixed surface (Naa35) and cartoon/stick representation (Naa30), showing the structural rearrangements of the α1–α2 loop and the β6–β7 loop of Naa30 upon ligand binding. (B) Surface projection of NatC, showing the effect of ligand binding on the shape and diameter of the central and side tunnel.

The drastic rearrangements of the α1–α2 loop and particularly the β6–β7 loop upon binding of the Gag peptide can be viewed as an induced fit-mechanism, which, as will be shown in section 4.10.4, enhances the activity of the NatC complex toward this substrate. Furthermore, the peptide-induced conformational changes also effect the shape of the *central tunnel* and the *side tunnel*, which will be explained in detail in the following section.

4.9 *Structural features close to the active site of the NatC complex*

A cross-sectional view through a surface representation of NatC reveals the location and dimensions of the two tunnels in the center of the NatC complex (Fig. 34). CoA and the substrate peptide meet head-to head in the *side tunnel*, which contains the active site of the catalytic subunit Naa30. The narrowest constriction of the *side tunnel* surrounds the peptide substrate.

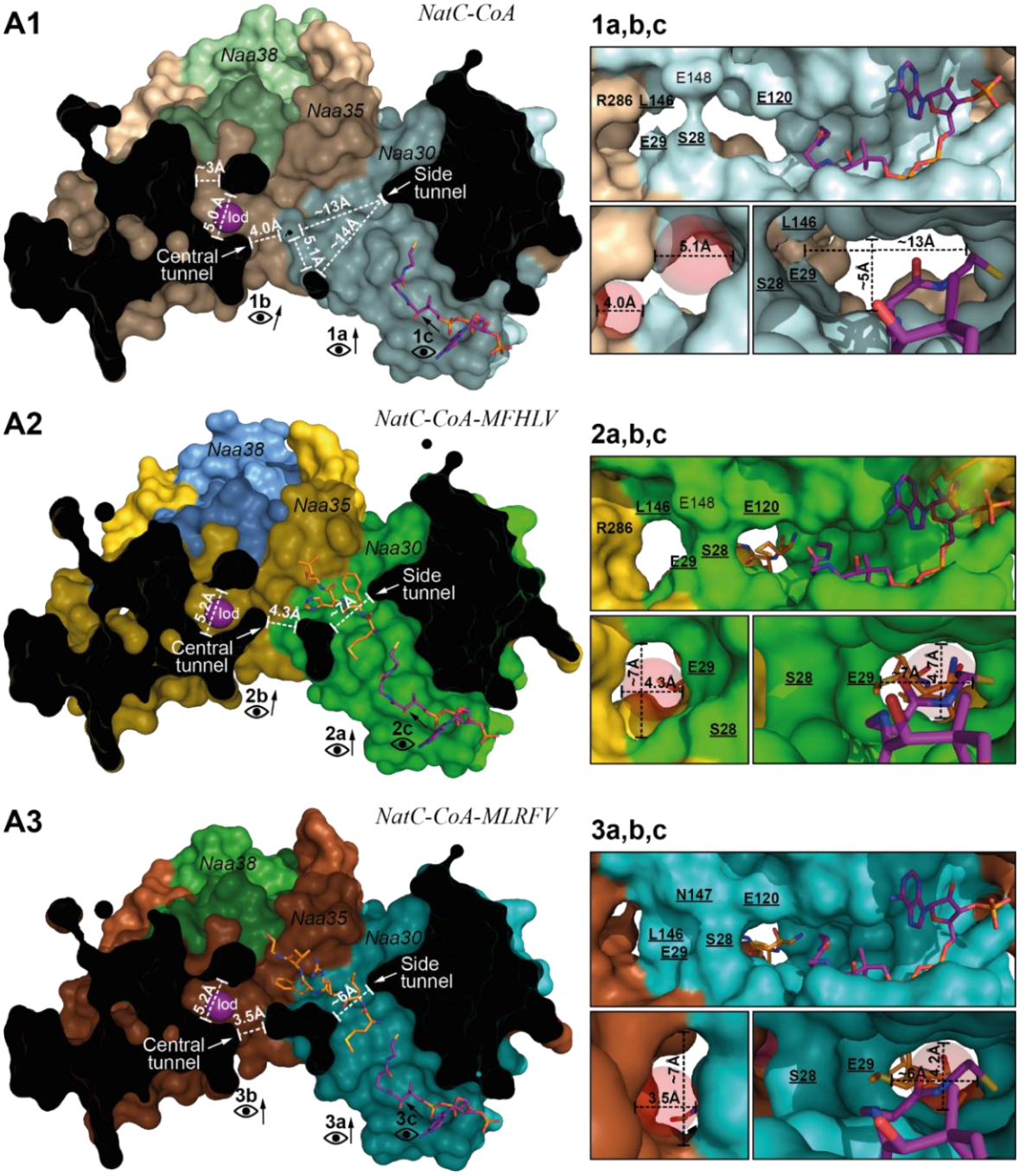


Fig. 34 Morphology of the NatC central and side tunnel (A) Cross-section through a surface projection of *NatC-CoA*, *NatC-CoA-MFHLV* and *NatC-CoA-MLRFV*, revealing the central and side tunnels. (B) Zoom views showing an overview (top) of the two tunnels, key residues separating the two tunnels and tunnel dimensions. Red spheres were generated for circular tunnels with the *PyMOLCaver* plugin (https://www.caver.cz/).

The *central tunnel* is roughly spherical and its radius at the narrowest point is approximately 4 Å in all structures. The *side tunnel* has a more elliptical cross-section with a height of ~5 Å and a width of ~ 13 Å at its narrowest constriction in both the *NatC-CoA* (Fig. 34, A1) as well as the *NatC-apo* structure (not shown). Upon peptide binding, the tunnel constricts further to ~7 Å in the *NatC-CoA-MFHLV* structure (Fig. 34, A2) and 6 Å in the *NatC-CoA-MLRFV* (Fig. 34, A3) structure, respectively.

The *central tunnel* and the *side tunnel* are separated from each other by side chains from the α1–α2 and β6–β7 loop (Fig. 34, zoom views 1-3). In case of the *NatC-CoA* and *NatC-apo* structures, side chains Ser28 and Glu29 are within 4–5 Å proximity to Glu148 and Leu146, without forming significant interactions. Upon binding of the yArl3 ligand (*NatC-CoA-MFHLV* structure), the hydroxyl group of Ser28 is in hydrogen bonding distance with the carboxyl group of Glu148 (3.7 Å), thus closing the gap between the α1–α2 loop and the β6–β7 loop and reinforcing the separation of the *central* and *side tunnel*. Upon binding of the Gag peptide, the loops move even closer together, enabling the formation of a hydrogen bond between the side chain of Asn147 and the hydroxyl group of Ser28.

Taken together, both the α1–α2 and the β6–β7 loop separate the *central* tunnel from the *side* tunnel and concurrently restrict their diameters. In the ligand-free or CoA-bound structures, no significant interactions are observed between the two loop regions. As shown in section 4.8.3, both loops can undergo significant movements upon peptide binding, showing their flexibility. It is thus conceivable that the two loops can open up even further and link the *central* and *side tunnel* completely, which may play an important role in the co-translational acetylation of nascent chain peptide substrates emerging from the ribosomal exit tunnel. This is further discussed in section 5.5.1.

4.10 *Acetyltransferase activity of the NatC complex*

4.10.1 Design of active site mutants

A structure-based mutagenesis approach was performed to study the catalytic mechanism of the NatC complex. To this end, several NatC constructs with single point mutations in the active site were designed. Conserved residues within a ~5 Å radius of the peptide N-terminus and the Coenzyme A thiol group were considered to participate in catalysis. Conserved residues were identified using a multiple sequence alignment for the catalytic subunit Naa30 (Fig. S1).

A detailed view of the active site of the *NatC-CoA-MLRFV* structure can be seen in Fig. 35. The active site is surrounded by residues from the α1–α2 loop, β4, β5 and the β6-β7 loop of the catalytic subunit Naa30. Nearly all residues in proximity of the active site are completely conserved among NatC orthologs, emphasizing their importance for proper catalytic function.

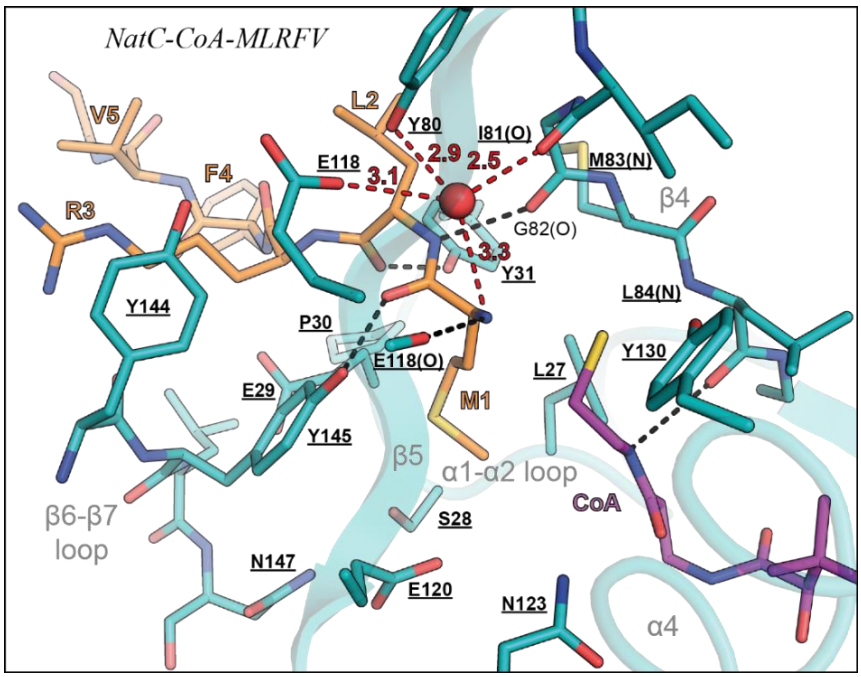


Fig. 35 The NatC active site with bound CoA and Gag peptide. (A) Detailed view of the active site of the *NatC-CoA-MLRFV* structure, bound to the Gag peptide (MFHLVGSRRR) and coenzyme A. Completely conserved residue are bold and underlined. The position of the proposed catalytic water (HOH1) is indicated by a red sphere. Hydrogen bond distances of HOH1 with surrounding residues are shown in red.

A well-coordinated water molecule (red sphere) is positioned in the vicinity of the peptide substrate N-terminus. It is within hydrogen-bonding distance of the Met1 amino-terminus of the peptide substrate. The tight coordination and the vicinity to the peptide amino group indicated a function of this water molecule in electron transfer during the enzyme catalyzed transfer of the acetyl- moiety, which is further discussed in section 5.4.

Additional evidence for the identification of catalytically important residues can be obtained by comparing Naa30 with catalytic subunits from known NAT paralogs, i.e. genes/proteins that are the results of a gene duplication events in the last common ancestor (LCA) (see Fig. 36). Residues that

are essential for the general catalytic function of all GNAT superfamily members should be highly conserved among paralogs. For example, Tyr30, Tyr130, Tyr140 and Tyr145 are almost entirely conserved among all NAT paralogs. Likewise, the coenzyme A binding motif RxxGxG/A is conserved across the entire GNAT superfamily as displayed in Fig. 36B.

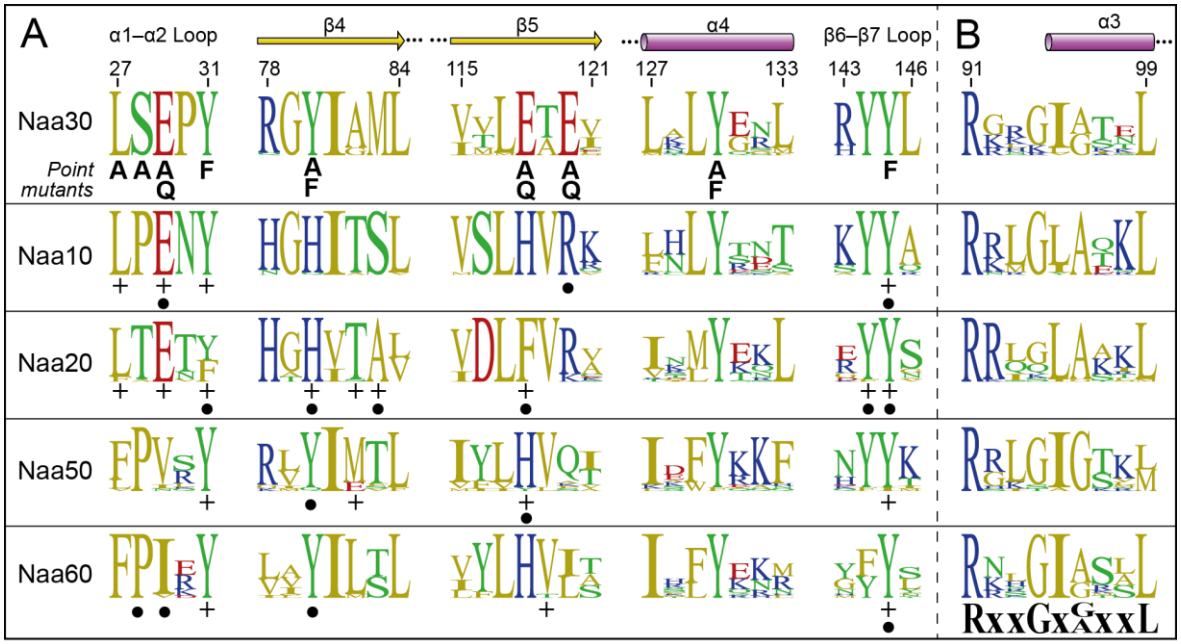


Fig. 36 Structural alignment of NAT paralogs. (A) Sequence logos[40] of the catalytic subunits of five different NAT paralogs showing identical regions (α1–α2 loop, β4, β5 and β6-β7 loop) near the active site. Sequence logos were generated from multiple sequence alignments using corresponding orthologs from 18 different species. ●, mutational sensitive residues and +, substrate binding residues (as described in the respective publications for each NAT paralog). (B) Conserved acetyl coenzyme A binding motif RxxGxG/A.

Residues that are responsible for a paralog-specific function, for example, different substrate specificities, should only be conserved among orthologs, i.e. genes/protein that are descended from the same ancestral sequence, separated by a speciation event. Two completely conserved Naa30 residues (i.e. among Naa30 orthologs) are of particular interest, because they are specific for Naa30: Glu118 and Glu120 in the β5 strand. In all other NAT paralogs, the two glutamate positions are occupied by residues with different electrostatic properties. The position of Glu118 is occupied by a histidine residue in Naa10, Naa50 and Naa60 and a phenylalanine in Naa20, which have both been shown to be relevant for the catalytic activity in these paralogs. Similarly, Naa30 residue Glu120 is replaced by Arginine in Naa10 and Naa20.

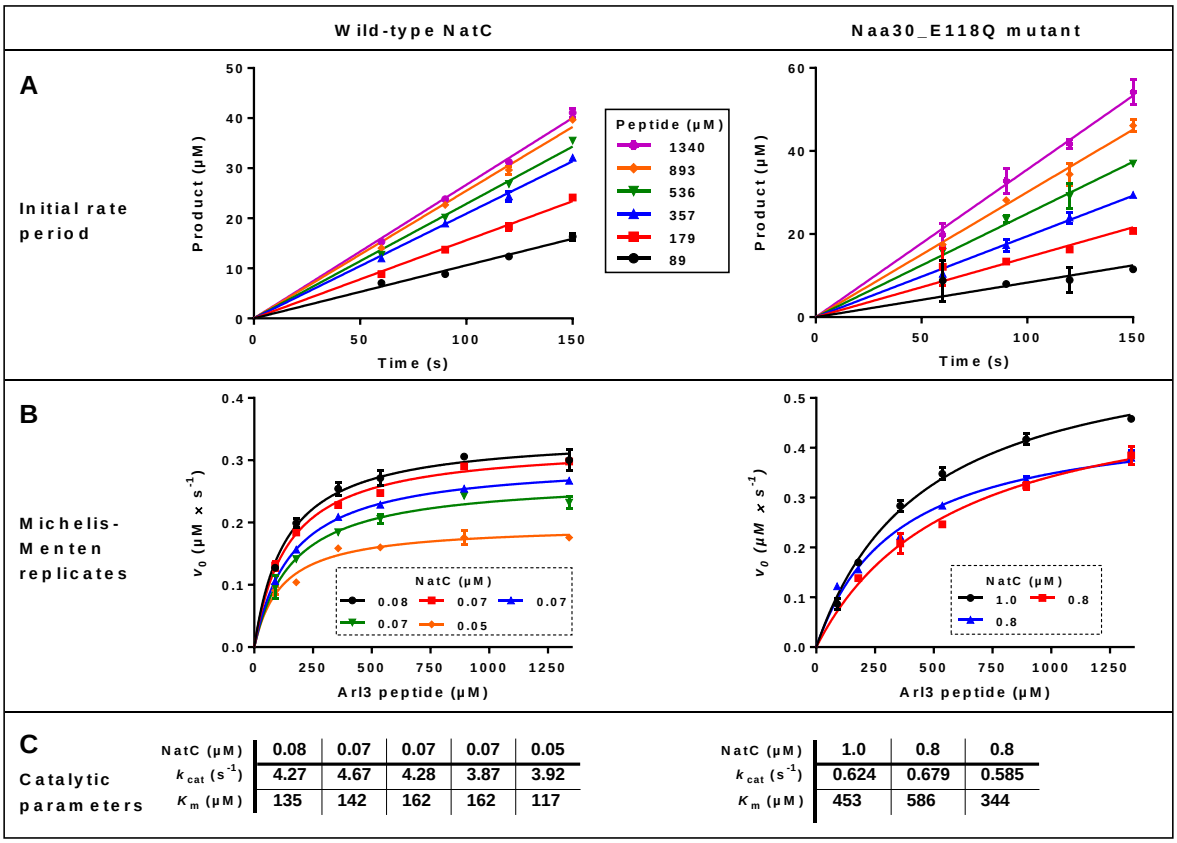
To unravel the catalytic mechanism and the paralog-specific properties of Naa30, nine different residues were selected for mutational studies: Leu27, Ser28, Glu29, Tyr80, Glu118, Glu120 and Glu145.

[40] "A sequence logo is a graphical display of a multiple sequence alignment consisting of color-coded stacks of letters representing amino acids at successive positions. Sequence logos provide a richer and more precise description of sequence similarity than consensus sequences and can rapidly reveal significant features of the alignment that could otherwise be difficult to perceive." (https://prosite.expasy.org/sequence_logo.html).

Single point mutations (see Fig. 36A) were introduced in the wild type NatC construct and the mutant NatC complexes were purified as the wild type construct (see 3.2.6). Point mutations included alanine substitutions, E→Q substitutions and Y → F substitutions. Alanine substitutions affect both electrostatic (i.e. charge and polarity) and stereochemical (i.e. size) properties of the corresponding position in the structure. E→Q and Y→F mutations were intended to only affect the functional properties of the side chain, while roughly maintaining the size of the corresponding residue.

4.10.2 Overview of the acetyltransferase assay

In vitro activity assays were performed to kinetically characterize wild type NatC and a series of NatC mutants (see 3.6). To achieve this, N-terminal acetylation of model peptides was followed by the conversion of acetyl-CoA to CoA and quantified using a subsequent colorimetric reaction. A complete exemplary data analysis is shown in Fig. 37 for wild type NatC and one NatC mutant (Naa30_E118Q) using the yArl3 peptide as a substrate (MFHLVGSRRR).



C Catalytic parameters

NatC (µM)	0.08	0.07	0.07	0.07	0.05
k_{cat} (s^{-1})	4.27	4.67	4.28	3.87	3.92
K_m (µM)	135	142	162	162	117

NatC (µM)	1.0	0.8	0.8
k_{cat} (s^{-1})	0.624	0.679	0.585
K_m (µM)	453	586	344

Fig. 37 Data analysis overview for the acetyltransferase assay. Data analysis is shown for wild type NatC and one exemplary NatC mutant (Naa30_E118Q). Enzyme reactions were carried out at six different peptide concentrations (89–1340 µM). In the chosen examples, the yArl3 peptide (MFHLVGSRRR) was used. The acetyl-CoA concentration was 500 µM. (A) Initial rate period of the reactions. Initial velocities ($\boldsymbol{v_0}$) are obtained from linear regressions ($[\boldsymbol{P}] = \boldsymbol{v_0} \times \boldsymbol{t}$). (B) Initial velocities plotted against peptide concentrations were fitted to the Michaelis-Menten equation ($\boldsymbol{v_0} = \boldsymbol{v_{max}}[\boldsymbol{S}]/(\boldsymbol{K_m} + [\boldsymbol{S}])$). Replicates (n ≥ 3) are shown in different colors. (C) Catalytic parameters ($\boldsymbol{K_m}$, $\boldsymbol{k_{cat}}$) of all replicates, which were performed at varying enzyme concentrations.

The first four data points of the reaction process are in the linear, initial rate period of the reaction (Fig. 37A). Initial reaction velocities (v_0), measured at six different peptide concentrations, are

plotted in Fig. 37B. The six data points were fitted to the Michaelis-Menten equation to obtain the catalytic parameters k_{cat} and the Michaelis-Menten constant K_m. Complete Michaelis-Menten kinetics were performed at least in triplicate (differently colored graphs in Fig. 37B) and the catalytic parameters K_m and k_{cat} of each replicate are summarized in tabular form (Fig. 37C). The complete Michaelis-Menten kinetics for all tested NatC constructs and all tested substrate peptides are summarized in Table 22 and Table 23, respectively. Additionally, all Michaelis-Menten plots are shown in the supplement (Fig. S7 - Fig. S12).

4.10.3 Kinetic analysis of wild type and mutant NatC constructs

The *in vitro* acetyltransferase assays were performed with the help of a former Master student, Mr. Linus Hopf, who has published parts of the Michaelis-Menten data in his Master thesis.[41]

Table 22 Catalytic parameters for wild type and mutant NatC. K_m values are calculated for the yArl3 substrate peptide (MFHLVGSRRR). Acetyl-CoA was used at a fixed concentration of 500 µM (≥ 10× K_m). [a]WT, wild type NatC construct (Naa30ΔC17, full-length Naa35 and Naa38ΔC11). [b]The acetyl CoA K_m was calculated for wild type NatC (K_m = 31±6 µM, n = 5). [c]The Naa35 *Tip1* mutant contains point mutations: K500A, K501A, K503A and K504A.[d]The NatC-Naa38 deletion consist of Naa30ΔC11 and full-length Naa35 only. Errors represent s.d. with number of replicated (*n*) indicated in the last column.

Enzyme	k_{cat} (s^{-1})	k_{cat}/ WT	K_m (µM)	K_m/ WT	*n*
NatC WT[a,b]	4.2 ± 0.3	1.0	140 ± 20	1.0	5
Naa30 mutations					
L27A	0.010 ± 0.005	0.0023	80 ± 20	0.53	4
S28A	1.8 ± 0.1	0.43	270 ± 30	1.9	3
E29A	0.24 ± 0.01	0.06	160 ± 30	1.1	4
E29Q	0.18 ± 0.02	0.04	250 ± 40	1.8	3
Y31F	0.45 ± 0.01	0.11	230 ± 50	1.6	3
Y80A	2.5 ± 0.3	0.59	550 ± 70	3.9	3
Y80F	1.68 ± 0.01	0.40	430 ± 60	3.0	3
E118A	0.40 ± 0.03	0.10	320 ± 30	2.3	3
E118Q	0.60 ± 0.02	0.14	430 ± 60	3.0	4
E120A	0.88 ± 0.04	0.21	260 ± 10	1.8	3
E120Q	1.01 ± 0.05	0.24	300 ± 60	2.1	4
Y130A	0.057 ± 0.002	0.01	200 ± 60	1.4	4
Y130F	0.084 ± 0.01	0.02	120 ± 20	0.80	4
Y145F	1.02 ± 0.09	0.24	230 ± 10	1.6	3
Naa35 mutations					
F47A	2.3 ± 0.1	0.55	160 ± 10	1.1	3
K59A	2.2 ± 0.3	0.53	191 ± 8	1.3	3
E171A E172A	5.8 ± 0.5	1.4	660 ± 50	4.6	3
R406A	3.0 ± 0.1	0.70	120 ± 20	0.83	3
Tip1[c]	4.2 ± 0.4	1.0	175 ± 9	1.2	3
Naa38 deletion[d]	0.27 ± 0.04	0.06	110 ± 10	0.75	3

[41] Hopf, Linus Valentin Moritz (2019), "Biochemical and mechanistic characterization of NatC", submitted to the Department of Biology, Chemistry and Pharmacy of the Freie Universität Berlin

For the wild type NatC complex, a k_{cat} of 4.2 ± 0.3 s^{-1} and a K_m of 140 ± 20 µM were determined for the yArl3 peptide (MFHLVGSRRR). The k_{cat} of the NatC complex is comparable to the published values for wild type NatA (Liszczak et al. 2013) and NatB (Hong et al. 2017), each measured with their cognate peptides substrates: for NatA, a k_{cat} of 3.0 ± 0.5 s^{-1} and for NatB, a k_{cat} of 2.7 ± 0.2 s^{-1} were determined by the authors. The corresponding K_m values are 340 ± 50 µM for NatA and 660 ± 50 µM for NatB, and thus slightly higher compared to the K_m determined for NatC and yArl3 peptide.

Single point mutations at amino acid positions close to the active site of the Naa30 subunit were introduced to identify residues that are crucial for the catalytic activity of NatC. All mutations close to the active site of the Naa30 subunit led to a reduction of k_{cat} of at least 40% (Fig. 38). The most significant reductions to ≤ 10% of wild type activity were seen for residues Leu27, Glu29, Tyr31, Glu118 and Tyr130, indicating their importance in the catalytic mechanism. The strongest increase in K_m compared to wild type Naa30 was observed for residues Tyr80 and E118A, suggesting a crucial function of these residues in substrate binding.

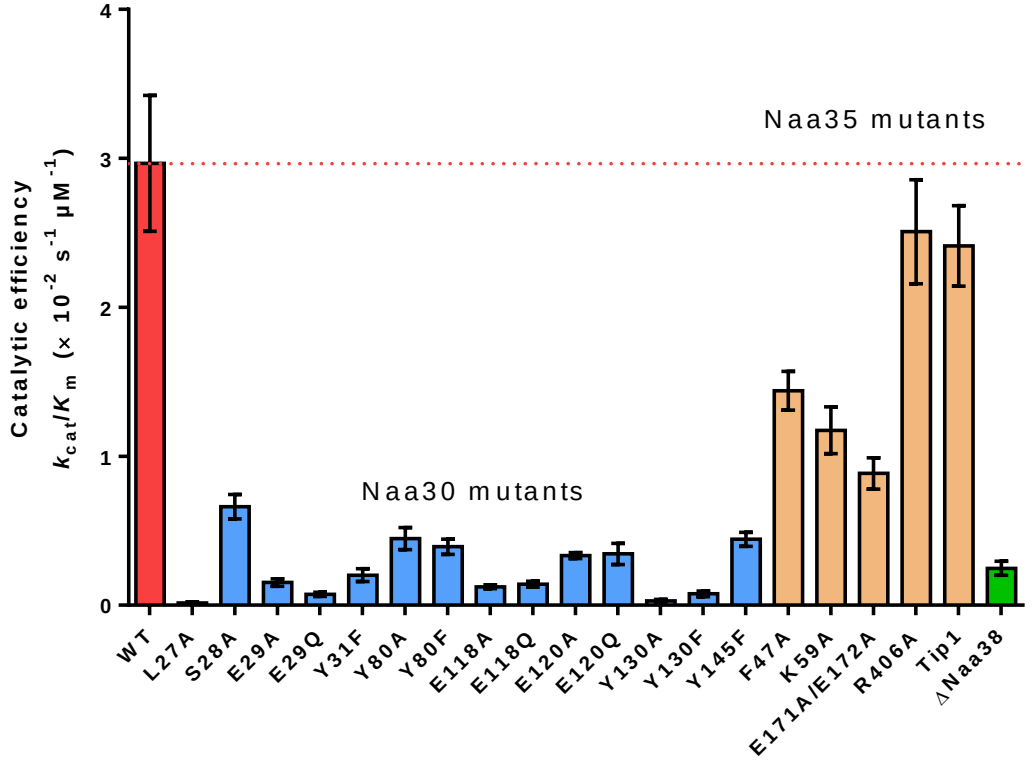


Fig. 38 Catalytic efficiency (k_{cat}/K_m) of NatC wild type and mutant constructs. The s.d. for the catalytic efficiency was calculated from the s.d. values of k_{cat} and K_m, considering the propagation of uncertainty[42].

Mutations in the large auxiliary subunit Naa35 had only a minor effect on the k_{cat}. For example, the K59A mutant, which stabilizes the catalytic α1–α2 loop, lead to a roughly 50% reduction of the k_{cat}. Phe47 from Naa35, that forms a distal end of the hydrophobic substrate-binding pocket, shows a roughly 45% reduction of the k_{cat}. Interestingly, the double mutant E171A/E172A, located in a conserved loop near the C-terminal end of the MFHLVGSRRR peptide, displayed a 40% increased

[42] Equation for the calculation of the s.d. of the catalytic efficiency: $\sigma_{kcat/Km} = \frac{k_{cat}}{K_m} \times \sqrt{\left(\frac{\sigma_{kcat}}{k_{cat}}\right)^2 + \left(\frac{\sigma_{Km}}{K_m}\right)^2}$

k_{cat}, coupled with a 4.6-fold increase in K_m. Even though the peptides' triple arginine tag could not be modeled due to missing electron density, electrostatic interactions of the arginine side chains with the negative charges of the 170-FEEED-174 loop are likely to increase the affinity of the peptide with the NatC complex. The disruption of the negative charge (E171A/E172A) thus resulted in an observed increase in K_m. Arg406 stabilized the 170-FEEED-174 loop through cation-π interactions with Phe170 and two salt bridges with Glu174. The R406A mutant was designed to decrease the stability of the loop. However, no significant effect on either K_m or k_{cat} was observed, indicating that the conserved 170-FEEED-174 loop has no direct function for NatC catalysis.

Interestingly, a NatC construct that was expressed and purified without its small auxiliary subunit Naa38 (ΔNaa38) showed a strong decrease in k_{cat} to 6% of wild type activity, while the K_m remained almost identical to WT. These data suggest that the small subunit has a direct impact on catalytic activity of the NatC complex.

4.10.4 Analysis of the NatC substrate specificity

By using a variety of substrate peptides in the acetyltransferase assay, the peptide specificity of the NatC complex was analyzed. The results are summarized in Table 23 and visualized in Fig. 39. NatC activity was tested for two peptides, which are native substrates of the NatA and NatB complex. The sequence SASEA corresponds to the N-terminus of the *S. cerevisiae* protein SytC after the removal of its initiator methionine by a methionine aminopeptidase (MetAP) and is a NatA substrate that is efficiently acetylated in *S. cerevisiae* (Arnesen et al. 2009). SASEA was also used for co-crystallization as a peptide-CoA conjugate with the NatA complex (Liszczak et al. 2013). The sequence MDSEV corresponds to the yeast actin N-terminus, a NatB substrate, which was co-crystallized as a peptide-CoA conjugate with the NatB complex (Hong et al. 2017). For both the NatA and NatB model peptides, no activity could by detected in the acetyltransferase assay for wild type NatC.

A peptide with the N-terminal sequence of the human ortholog of yArl3, hARFRP1 (MYTLL) showed a comparable activity to yArl3, with a k_{cat} of 3.2 ± 0.3 s^{-1} and a K_m of 180 ± 40 µM. This suggests that the species-specific changes in the N-terminal sequences between the yeast Arl3 and the human ARFRP1 proteins are not the results of an adaption to their respective NatC ortholog.

Three yArl3 peptide variants, substituting Phe2 with tyrosine (MYHLV), tryptophan (MWHLV) or leucine (MLHLV), were designed to analyze which hydrophobic amino acid is preferred at position two. While a F2W substitution lead to comparable k_{cat} and K_m values, the F2Y variant showed a 50% reduction in k_{cat} and a 3.7-fold increase in K_m, indicating a negative impact of the polar tyrosine hydroxyl group for binding to the hydrophobic pocket. An F2L substitution at position 2 (MLHLV) showed a 30% increase of the k_{cat} compared to the wild type yArl3 peptide. However, the K_m showed a five-fold increase to 730 ± 50 µM, suggesting that the small leucine at position 2 only explores a fraction of the hydrophobic peptide biding pocket of NatC, which seems to prefer bulky aromatic amino acids at the second position. Interestingly, an F2A variant showed a 50% decrease of the K_m

compared to the wild type, suggesting that the reduced size of the alanine side chain allows the peptide to bind in a unique way, without exploring the hydrophobic pocket. However, a structural analysis would be necessary to determine the exact binding mode of this MAHLV peptide.

Table 23 Catalytic parameters for wild type NatC with various substrate peptides. [a]N-terminal five residues of the decameric peptides ending on -GSRRR.. [b]All normalizations are relative to wild type NatC catalytic parameters, generated with the Arl3 substrate. Where k_{cat} is ND (not determined), activity could not be detected. Errors represent s.d. with number of replicated (*n*) indicated in the last column.

Substrate N-terminus[a]	Substrate origin	k_{cat} (s^{-1})	k_{cat}/ yArl3[b]	K_m (µM)	K_m/ yArl3[b]	*n*
MFHLV-	**Arl3**	4.2 ± 0.3	1.0	140 ± 20	1.0	5
MYTLL-	Arfrp1	3.2 ± 0.3	0.77	180 ± 40	1.3	4
MLRFV-	Gag	12 ± 1	2.8	4 ± 3	0.030	4
	Arl3 variants					
MYHLV-	F2Y	2.5 ± 0.1	0.60	520 ± 60	3.7	4
MWHLV-	F2W	3.4 ± 0.4	0.80	140 ± 60	1.0	4
MLHLV-	F2L	5.6 ± 0.2	1.3	730 ± 50	5.1	3
MAHLV-	F2A	5.1 ± 0.3	1.2	80 ± 10	0.5	3
MFALV-	H3A	7.0 ± 0.5	1.7	4 ± 3	0.029	3
MFHAV-	L4A	8 ± 1	1.9	6 ± 5	0.044	4
	Arl3/Gag hybrids					
MFRLV-	H3R	9 ± 2	2.1	27 ± 2	0.19	4
MFHFV-	L4F	10 ± 1	2.5	28 ± 4	0.20	5
SASEA-	ThrRS (NatA substrate)	ND				
MDSEV-	Actin (NatB substrate)	ND				

The Major capsid protein (Gag) of the *S. cerevisiae* virus L-A is another known native NatC substrate (Tercero, Dinman, and Wickner 1993). With a k_{cat} of 12 ± 1 s^{-1}, the turnover number of the Gag peptide (MLRFV) was 2.8-fold higher compared to the yArl3 peptide. In addition, the K_m was reduced to 4 ± 3µM, which is at the detection limit of the colorimetric acetyltransferase assay. The strong reduction of the K_m suggests a higher affinity of the Gag peptide toward the catalytic subunit of NatC.

To identify the Gag peptide residues that convey the large increase in catalytic efficiency (k_{cat}/K_m) compared to yArl3, two yArl3/Gag hybrid peptides with single amino acid exchanges from Gag in a yArl3 background were designed and tested. Both, the H3R and the L4F yArl3/Gag hybrids showed marked increases in k_{cat} and strongly reduced K_m values to below 30 µM. This can be explained by the Gag-Naa30 ligand interactions, observed in the *NatC-CoA-MLRFV* structure, in which Arg3 interacts with the β6–β7 loop and Phe4 with the opposite side of the extended hydrophobic pocket (Fig. 32, section 4.8.2).

Interestingly, also the H3A and L4A yArl3 variants exhibited strong increases in k_{cat} and strongly reduced K_m values (< 10 µM). From the *NatC-CoA-MFHLV* structure, it is apparent, that the bulky imidazole group of yArl3-His3 sterically hinders Naa30-Glu29 from forming a hydrogen bound with the fourth peptide backbone amide. A smaller amino acid at peptide positions 3 or 4 is likely to give

Glu29 free access to interact with the substrate peptide thereby increasing the peptide affinity. Similar to the *NatC-CoA-MLRFV* structure, the Glu29 could in turn stabilize the catalytic β6–β7 loop, resulting in the observed increase of k_{cat}.

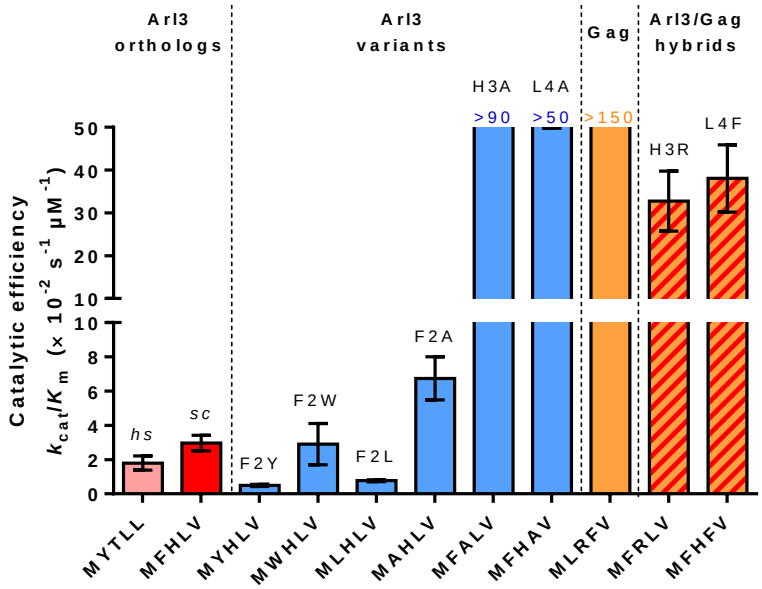


Fig. 39 Catalytic efficiency (k_{cat}/K_m) of wild type NatC with various substrates. The s.d. for the catalytic efficiency was calculated from the s.d. values of k_{cat} and K_m, considering the propagation of uncertainty. Arl3 single amino acid substitutions in the Arl3 variants and Arl3/Gag hybrid peptides are written above each bar. Three peptides exhibited K_m values at the lower detection limit of the employed colorimetric assay, resulting in very large catalytic efficiencies. Due to the large s.d. values of these datasets, the graph is cut a 50×10^{-3} µM s^{-1} and the lower limit for each of the three datasets (calculated as their mean minus 1 s.d.) is given above the respective column.

4.11 *Structure of the NatC_Naa30_L27A mutant*

4.11.1 Structure determination

The *NatC_Naa30_L27A* mutant construct exhibited the lowest catalytic activity compared to the wild type NatC construct, with a more than 100-fold reduction of the k_{cat}. The mutant was crystallized to assess if this mutation causes structural changes in the α1-α2 loop region, which could explain the reduced activity of this mutant.

A single crystal of *NatC_Naa30_L27A* was obtained using a crystallization solution containing 15.5% PEG 4000, 150 mM ammonium iodide and 100 mM sodium citrate, pH 6.4 (see 3.7). As with all other NatC constructs, the crystal grew at 20°C, appeared after 12 – 48 h and required roughly five days to reach maximum dimensions (Fig. 40).

Fig. 40 Diffraction quality crystals of the apo *NatC_Naa30_L27A* mutant construct.

The crystal of the *NatC_Naa30_L27A* mutant construct diffracted X-rays to a maximum resolution of 2.69 Å and belongs to the space group *P*1 with cell dimensions a=48.2 Å, b=89.5 Å, c=138.6 Å, α=85.1° β=80.9° γ=75.2°. The data collection statistics are found in Table 24.

Table 24 Data collection statistics for a crystal of the NatC_Naa30_L27A mutant. Values in parentheses are for highest-resolution shell.

Data collection	NatC Naa30_L27A (native)
Space group	*P*1
Cell dimensions	
a, *b*, *c* (Å)	48.22 89.54 138.55
α, *β*, *γ* (°)	85.14 80.89 75.20
Wavelength (Å)	0.9184
Resolution range (Å)	46.14 - 2.69 (2.79 - 2.69)
Completeness (%)	94.6 (86.4)
R_{sym} (%)	8.541 (95.1)
Redundancy	3.7 (3.4)
I/σ (I)	11.6 (1.2)

The structure of the *NatC_Naa30_L27A* mutant was solved by molecular replacement using the refined structure of the SeMet-labeled NatC complex as an input model. The structure of *NatC_Naa30_L27A* was refined to a R_{work}/R_{free} of 19.8%/22.8% (see Table 25). The final model contains no Ramachandran outliers, with 98.7% of its residues located in the favored region.

Table 25 Refinement statistics for the NatC_Naa30_L27A mutant structure. Values in parentheses are for highest-resolution shell. [a]R_{free} was calculated for 5% of the reflection data.

Refinement	NatC Naa30_L27A (native)
Resolution range (Å)	46.14 - 2.69 (2.79 - 2.69)
No. of reflections	58142 (5292)
R_{work} / R_{free}[a]	19.8 / 22.8
No of atoms	
Protein	15451
Ligands	91
Water	182
B factors (Å^2)	
Average	82.0
Protein	82.3
Ligands	92.5
Water	55.6
RMSD	
Bond lengths (Å)	0.004
Bond angles (°)	0.86
Ramachandran plot (%)	
Most favored	98.6
Allowed	1.4
Outlier	0
Rotamer outliers (%)	0.4
Clashscore	5.0

The asymmetric unit of the *NatC_Naa30_L27A* crystal contained two independent copies of NatC, which are referred to as biological assemblies #1 and #2. The root-mean-square deviation of the Cα positions between the two biological assemblies is low (0.89 Å, over 950 compared Cα atoms), indicating an almost similar fold.

In contrast to all other structures that crystallized in space group $P2_12_12_1$, the *NatC_Naa30_L27A* construct crystallized in space group *P*1. There are no crystal contacts around the α1–α2 region, which contains the mutated Leu27 residue, in either the *P*1 (Fig. S6) or the $P2_12_12_1$ (Fig. S5) structures. Thus, it is unlikely that the leucine-to-alanine substitution in the α1–α2 loop has a direct influence on the crystal packing. Thus, the appearance of a NatC crystal in space group *P*1 might be purely coincidental.

4.11.2 Conformational changes in the α1–loop–α2 region

The reason for crystallizing the *NatC_Naa30_L27A* mutant, was to assess if the 100-fold reduction of the k_{cat} compared to wild type NatC is caused by structural changes in the α1–loop–α2 region. A superposition of *NatC_Naa30_L27A* with *NatC-apo* and *NatC-CoA* reveals, however, that the side chain of the Ala27 substitution points in the same direction – toward the two alpha helices α1 and α2 – as does the wild type Leu27 residue (Fig. 41, 1a).

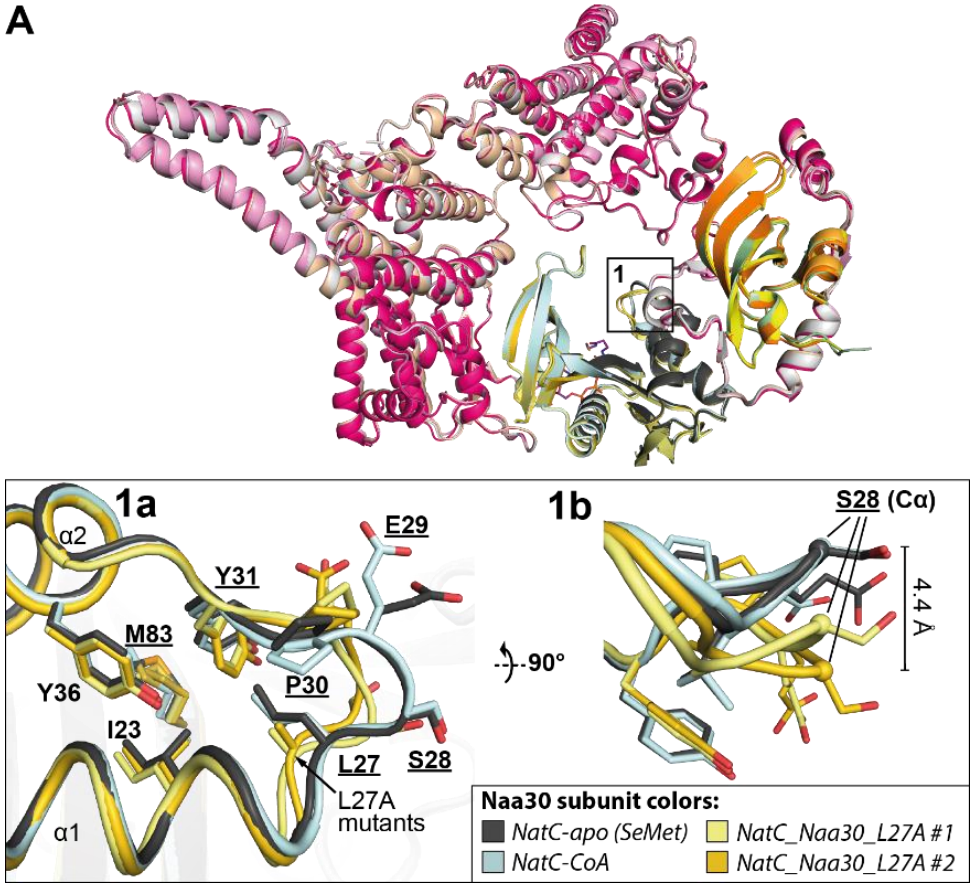


Fig. 41 Conformational changes in the Naa30 α1–α2 loop of the L27A mutant. (A) Superposition of *NatC-apo*, *NatC-CoA* and *NatC_Naa30_L27A* (assemblies #1 and #2) with zoom views (1a, b), showing the α1–α2 loop of the Naa30 catalytic subunit. Highly conserved residues are labeled in bold and completely conserved residues are underlined. The position of the L27 to alanine mutants is indicated by an arrow.

The wild type L27 side chain is positioned in a hydrophobic pocket, composed of the residues Ile23, Tyr31, Tyr36 and Met83. Since the side chain of alanine is much smaller and less hydrophobic compared to the leucine, it is less restricted in its placement. As a result, the entire α1-α2 loop moves up to 4.4 Å in *NatC_Naa30_L27A*, as judged by the Ser28 Cα position (Fig. 41, 1b).

However, it is unlikely that the slight shift of the α1–α2 loop is responsible for the drastic reduction of the catalytic efficiency of *NatC_Naa30_L27A* compared to wild type NatC. It is more likely that the amino-terminal methionine side chain of the peptide substrate needs to be stabilized by Leu27 and Tyr145, which form a hydrophobic pocket that encompasses the methionine side chain (Fig. 35).

4.12 *NatC-ribosome interactions*

Polevoda et al. (2008) showed that all major NATs, NatA, NatB, NatC and NatE are associated with mono- and polyribosomes. Using co-sedimentation assays with purified ribosomes and NatA, Magin et al. (2017) identified two conserved electropositive regions (EPR) in NatA that are responsible for its ribosome interaction. A recent cryo-EM structure of the NatA/Naa50-ribosome complex (comprising Naa10, Naa15 and Naa50) (Knorr et al. 2019) revealed that the complex interacts with ribosomal RNA expansion segments via its large auxiliary subunit Naa15 and Naa50. To identify the regions responsible for ribosome binding in NatC, the electrostatic surface potential of NatC was calculated (Fig. 42).

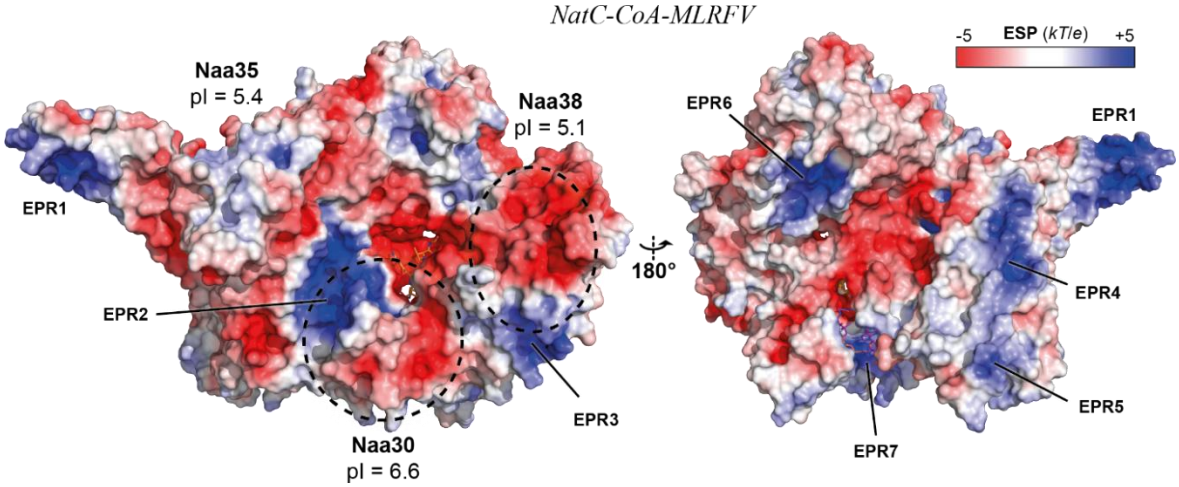


Fig. 42 Electrostatic surface potential of NatC. The electrostatic surface potential (ESP) is projected on the surface of the *NatC-CoA* structure. Approximate positions of the Naa30 and Naa38 subunits are indicated by circles.

Seven distinct electropositive regions were identified on the surface of NatC (Fig. 42). The previously described *Naa35 tip* region (α20–α21), that extends by approx. 29 Å from the central body, includes one of these electropositive regions (EPR1).

EPR1 contains a total of eleven positively charged residues, starting with H496 and ending at R519 (Fig. 43A). Due to the low overall sequence conservation of the large auxiliary subunit Naa35 (see Fig. S3, appendix), only two of the positively charged residues align perfectly (K511 and R519) among the eight different Naa35 orthologues. However, the overall positive charge of this region is conserved (Fig. 43B).

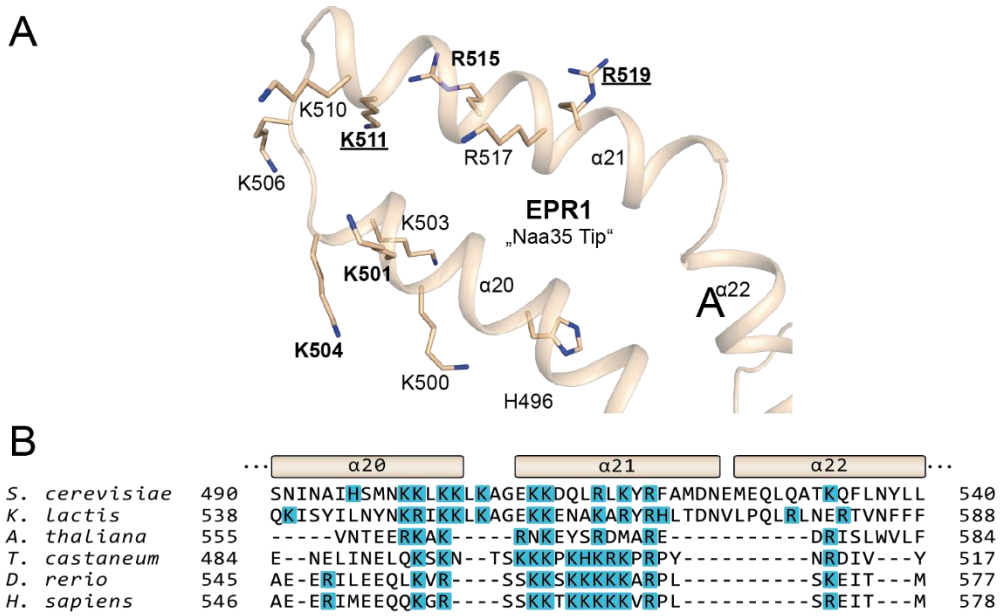


Fig. 43 Residues in the electropositive region EPR1 of NatC. (A) Electropositive region EPR1. Highly conserved residues are labeled in bold, with completely conserved residues underlined. (B) Sequence alignment of the *Naa35 Tip* region, using six different Naa35 orthologues with positively charged residues highlighted in blue.

4.12.1 Characterization of the ribosome binding surfaces in NatC

NatC-ribosome sedimentation assays were performed with the help of a former Master student, Mr. Linus Hopf, who has published part of the data in his Master thesis.[43]

Three of the identified electropositive regions: EPR1, EPR2 and EPR4 were tested in ribosome sedimentation assays (Fig. 44A). The *Naa35 tip* region contains a total of eleven positively charged residues, from which two sets from opposite sides were selected for mutational studies (Fig. 44B). The NatC *Tip1* construct contains alanine substitutions for Lys511, Arg515 and Arg519. In the *Tip2* construct, Lys500, Lys501, Lys503 and Lys504 were substituted with alanine residues. The EPR4 patch lies in the middle of the two parallel helices α25 and α26. The corresponding *Mid* mutant construct contains alanine substitutions for Arg636, Lys681 and Lys688. Finally, EPR5 was selected as an electropositive patch, that is even more distant to the Naa35 Tip. The corresponding *Distal* mutant construct contains alanine substitutions for Lys696, Lys699 and Lys701.

All NatC mutants tested in a ribosome co-sedimentation assay are based on the wild type NatC construct, with an additional FLAG-tag inserted that was used for quantification in Western blots (see 3.1). Purified NatC mutants were mixed with yeast ribosomes, underlaid with a sucrose cushion and subjected to ultracentrifugation. Supernatant and pellet fractions were separated on SDS-PAGE gels and analyzed by silver staining and western blot (see section 3.5).

[43] Hopf, Linus Valentin Moritz (2019), “Biochemical and mechanistic characterization of NatC”, submitted to the Department of Biology, Chemistry and Pharmacy of the Freie Universität Berlin

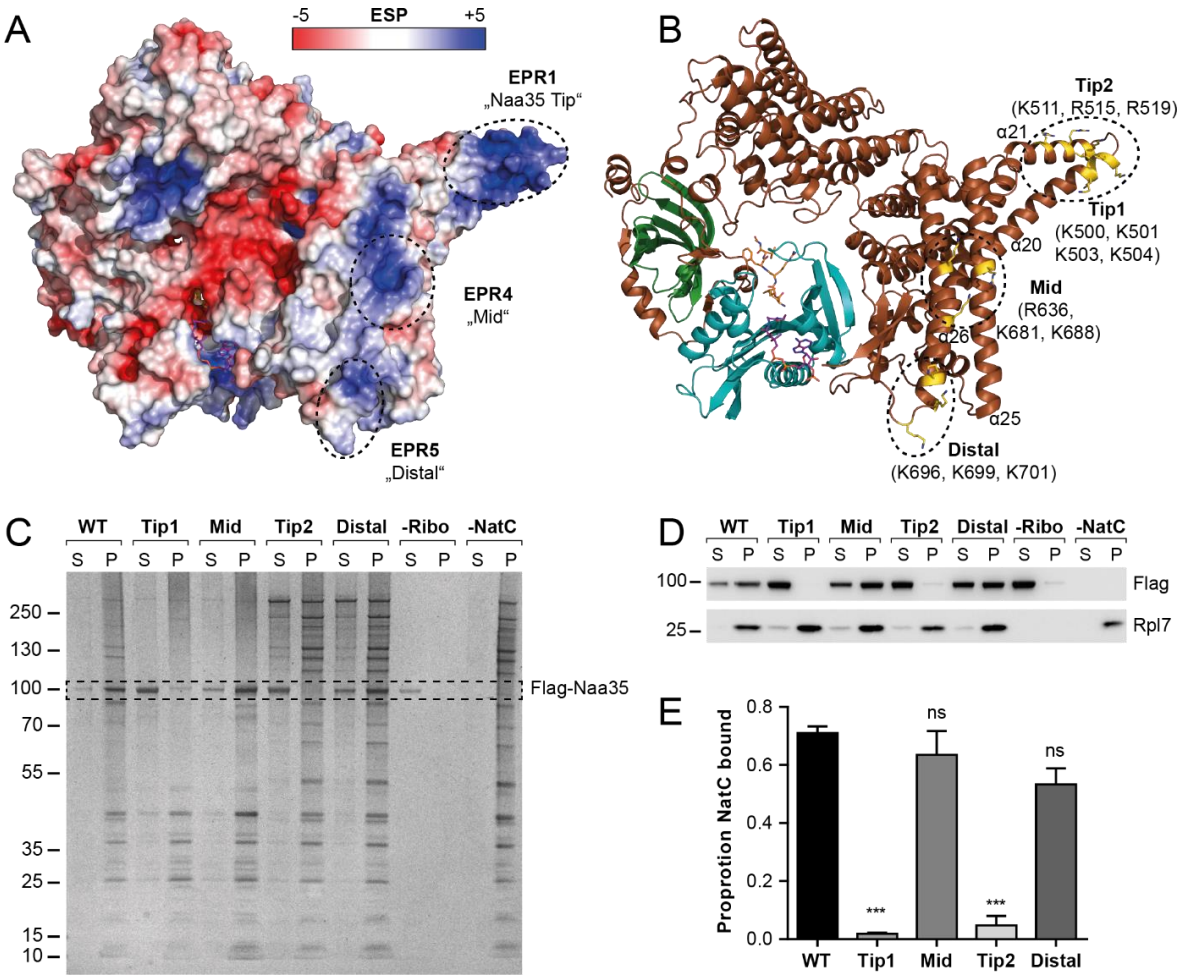


Fig. 44 NatC-ribosome co-sedimentation assay. (A) Electrostatic surface potential (ESP) projected on the surface of the *NatC-CoA* structure, showing three electropositive regions (EPR) on the back side of the NatC complex. (B) Cartoon presentation of NatC, showing the residues (colored in gold) of the four NatC mutant constructs, used in the sedimentation assays. All selected residues in the mutant constructs were substituted with alanine. (C) Silver stain gel of the supernatant (S) and pellet (P) fractions of the ribosome sedimentation assay. (D) Western blot of the sedimentation assay targeting the FLAG-tagged Naa35 subunit of NatC (Naa30 and Naa38 are untagged) and the Rpl7 protein of the large subunit of the yeast ribosome. (E) Quantitative analysis of the Western blot. The proportion of NatC bound to the ribosome was calculated as the percentage of NatC in the pellet (P/S+P) divided by the percentage of ribosome in the pellet (P/S+P). Plots show mean values and error bars represent s.d. (n=2). The result of a one-way ANOVA analysis is represented, with ns = not significant and ***, $P \leq 0.001$.

The negative control without ribosomes (-Ribo) shows that NatC stays in the supernatant, whereas the opposite negative control (-NatC), shows the ribosomal proteins in the pellet fraction (Fig. 44C, D). In co-sedimentation experiments, a large fraction of wild type NatC co-sediments together with the ribosomes. The *Mid* and *Distal* mutants show no significant difference (Fig. 44E) compared to wild type NatC and are also found in the pellet fractions. In contrast, the Naa35 Tip mutants *Tip1* and *Tip2* behave significantly different ($P \leq 0.01$) compared to wild type NatC and are primarily found in the supernatant, indicating their inability to bind to the ribosome.

In order to test if mutations in the Naa35 Tip region negatively affect the overall integrity of the NatC complex, the catalytic activity of the *Tip1* mutant was analyzed. No significant difference in k_{cat} and K_m could be seen between the *Tip1* mutant and wild type NatC (see Table 22, section 4.10.3). Thus, mutants in the *Naa35 Tip* are unlikely to affect the structural integrity of the NatC complex and the

observed loss of NatC-ribosome association is the result of the altered electrostatic properties of this region.

The NatC-ribosome co-sedimentation assay clearly shows that the electropositive *Naa35 tip* region is necessary for association of NatC to the ribosome *in vitro*, most likely via the same RNA expansion segments that interact with the auxiliary subunit Naa15 of the NatA complex (Knorr et al. 2019).

4.12.2 Movements of the Naa35 Tip region

It was established in the previous section, that the *Naa35 Tip* region is required for the interaction of NatC with the ribosome. A comparison of the different NatC structures (Fig. 45A) reveals, that the tip region can undergo substantial movements, which could facilitate the interaction of NatC with the ribosome. Helix α21 rotates in two directions around a hinge point, located approximately at Glu525 (Fig. 45B). The angle formed between helices α20 and α21 varies up to 9° between different NatC structures. The second rotation is perpendicular to the first and measured between α21 in different NatC structures, which can rotate up to 12°. Atomic displacement factors (ADPs) of the *Naa35 tip* are significantly larger in the second biological assembly of *NatC_Naa30_L27A* compared to the other NatC structures (Fig. 45C). The larger flexibility of the *Naa35 tip* in biological assembly #2 is explained by a reduced number of crystal contacts.

The large flexibility of the *Naa35 tip* region might be necessary to facilitate the NatC-ribosome interaction.

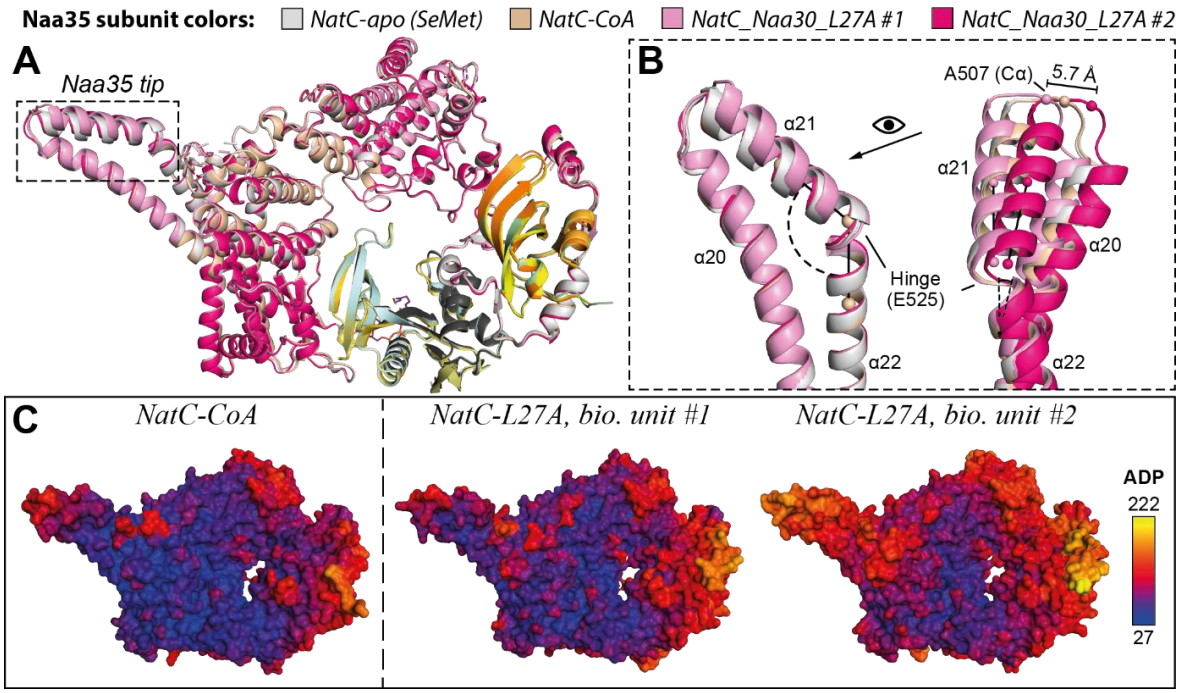


Fig. 45 Movements of the Naa35 Tip region. (A) Superposition of *NatC-apo*, *NatC-CoA* and *NatC_Naa30_L27A* (assemblies #1 and #2). (B) Detailed view of the *Naa35 tip* movements. Angles were measure between helices α21 and α22 for all NatC structures. Additionally, the angle formed by α21 in assembly#1 and the same helix in assembly #2 was determined. (C) Heat maps, showing the atomic displacement parameters (ADP) on the surface of *NatC-CoA* and *NatC_Naa30_L27A* (#1 and #2).

5 Discussion

The family of Nα-terminal acetyltransferases mediates the acetylation of N-termini in a wide range of target proteins. This irreversible post-translational modification is co-translationally attached and is of tremendous biological importance since it modifies the stability of these proteins, controls their interaction with binding partners and influences their functionality.

In yeast and mammals, NATs comprise a family of five proteins or protein complexes, whose members modify N-termini in target proteins with different consensus recognition sequences. To reveal the catalytic mechanism and substrate specificity, several structures from single NATs and heteromeric NAT complexes have been determined in the presence and absence of their cognate substrates (Liszczak, Arnesen, and Marmorstein 2011; Liszczak et al. 2013; Magin, Liszczak, and Marmorstein 2015; Chen et al. 2016; Stove et al. 2016; Hong et al. 2017; Goris et al. 2018). Most recently, a cryo-EM structure of NatA/Naa50 in complex with the ribosome has been presented (Knorr et al. 2019).

The only major eukaryotic NAT that has thus far eluded structural characterization, is the heterotrimeric NatC complex, consisting of the catalytic subunit Naa30 and the two auxiliary subunits Naa35 and Naa38. In this work, five structures of the *Saccharomyces cerevisiae* NatC complex in the absence and presence of CoA and two peptide substrates have been elucidated. These structures provide insights into the highly intertwined architecture of the NatC subunits and into their evolutionary origin. Furthermore, the structures show how N-terminal substrates with the NatC-specific consensus sequence are recognized at the interface of two NatC subunits and provide insights into the catalytic machinery of NatC and, consequently, the catalytic mechanism of NatC and other Nat complexes. Finally, by comparing the NatC structure to the ribosome-bound NatA structure, a model for the ribosome-bound NatC complex is presented, which provides structural insights into N-terminal acetylation during translation.

5.1 *Architecture of the NatC complex*

The structural analyses in this thesis show that the large auxiliary subunit Naa35 is the central assembly hub of NatC and forms extensive interactions with the catalytic subunit Naa30 and the small auxiliary subunit Naa38. Naa35 wraps around Naa30 in an open ring-like tertiary structure and covers about three fourth of its circumference. Naa38 is primarily tethered to NatC by the N-terminus of Naa35, which wraps around almost the entire circumference of Naa38, forming another ring-like tertiary structure (Fig. 18).

The interactions between the NatC subunits are evolutionary conserved among NatC orthologs and comprise a mixture of hydrophobic and hydrophilic interactions (Fig. 21). Here it was shown, that the N-terminal β1–α2–loop–β2 segment of Naa35 forms extensive hydrophobic interactions with Naa38 and participates in the common Naa35/Naa38 β-sheet. In accordance with the involvement of the Naa35 N-terminus in the NatC structure, deletion of its first 44 residues disrupted the interaction

with Naa38. Most strikingly, the architecture of the NatC complex completely differs from that of the heterodimeric NAT complexes NatA and NatB (Fig. 23) (Liszczak et al. 2013; Hong et al. 2017). Most importantly, the relative position of the NatC auxiliary subunit compared to the catalytic subunit is exactly opposite to the positions of the NatA and NatB auxiliary subunits. Whereas Naa15 and Naa25 primarily engulf the N-terminal part of their catalytic subunits, Naa35 additionally encloses the C-terminal half of Naa30. In NatA and NatB, the catalytic β6–β7 loop makes no contact with the corresponding auxiliary subunit. In NatC, this loop-region, which is critical for binding of the peptide substrate is in direct contact with the auxiliary subunit Naa35.

5.2 *Evolutionary relation of NatC subunits*

A detailed structural analysis of the individual subunits revealed expected and surprising evolutionary relationships of the individual NatC subunits that are discussed in the following.

As predicted from sequence comparisons, the catalytic subunit Naa30 is closely related to the catalytic subunits of NAT paralogs. Indeed, Naa30 adopts the typical Gcn5-related N-acetyltransferase (GNAT) fold, containing a central acetyl-CoA binding region (Vetting et al. 2005; Lu, Berkey, and Casero 1996; Wolf et al. 1998).

As shown in section 4.6.2, part of the Naa35 subunit structurally resembles the Bro1 domain of several proteins. This includes the Bro1 domain of the *S. cerevisiae* Bro1 protein (Kim et al. 2005) and its mammalian ortholog Alix (Zhai et al. 2007).

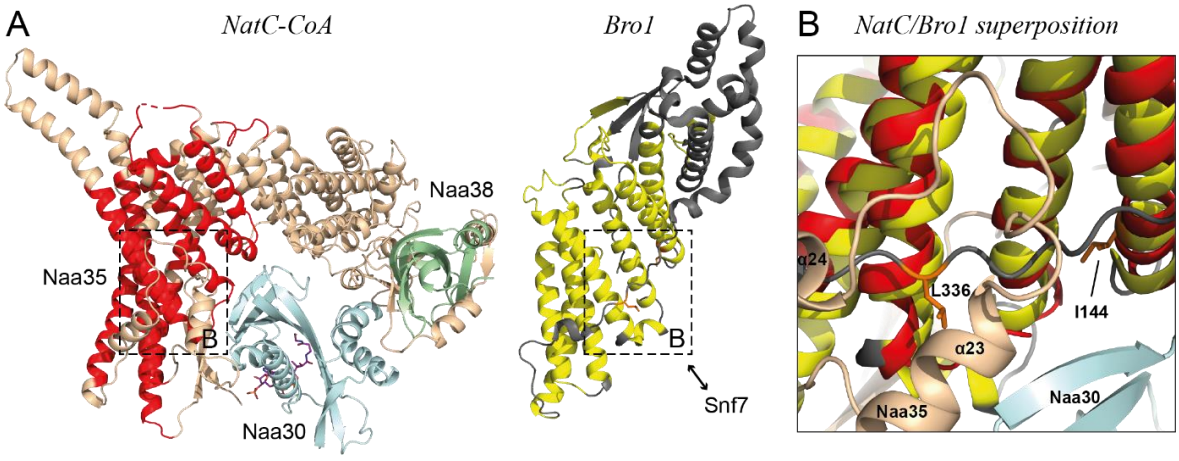


Fig. 46 Absence of a Snf7 interface in the Bro1-like domain of Naa35. (A) Part of the Naa35 subunit of NatC (red region in the *NatC-CoA* structure), resembles a large fraction of the Bro1 domain of the *S. cerevisiae* Bro1 protein (yellow region of the *Bro1* structure). A hydrophobic patch in the Bro1 protein (containing residues L336 and I144) is responsible for its association with the Snf7 component of the ESCRT-III complex. (B) A superposition of *NatC-CoA* and *Bro1* reveals that the corresponding region in NatC is not accessible for intermolecular interactions.

Yeast Bro1 is a cytoplasmic protein that associates with endosomes by binding to the Snf7 component of the ESCRT-III complex (Odorizzi et al. 2003). The ESCRT (endosomal sorting complex required for transport) pathway is a key mediator of multivesicular body (MVB) biogenesis at the endosome. MVBs deliver cargo destined for degradation to the yeast vacuole or mammalian

lysosome (Henne, Buchkovich, and Emr 2011). Kim et al. (2005) have identified a hydrophobic patch in the Bro1 domain which is necessary for binding to Snf7. The corresponding region in NatC partly overlaps with the electropositive EPR2 patch (Fig. 42, section 4.12) and is additionally covered by the α23–α24 loop region of Naa35 (Fig. 46). It is therefore unlikely that NatC can associate in a similar manner to Bro1 with the ESCRT complex at the endosomes.

The structure of the NatC complex revealed, that Naa38 adopts the conserved Sm fold (Kambach et al. 1999) which contains a strongly bent five-stranded β-sheet. In the NatC complex, the five Naa38 β-strands are extended on both sides by three short β-strands from the Naa35 subunit, forming a bifurcated, anti-parallel β-sheet (see Fig. 19, section 4.3). Within the Sm-like proteins, Lsm4 exhibits the highest structural similarity to Naa38. Lsm4 is part of a heptameric donut-shaped assembly that consist of seven different Lsm proteins and binds the small nuclear RNA (snRNA) U6 (Zhou et al. 2013) which is part of the spliceosome.

The structural similarity of Naa38 with the RNA-associated Lsm proteins may suggest that Naa38 can also assemble into multimeric structures and/or bind RNA. However, as shown in section 4.6.3, Naa38 forms tight interactions with Naa35. The intermolecular β-sheet hydrogens bonds between Naa35 and Naa38 are roughly located at the same position as the intermolecular hydrogen bonds formed between neighboring Lsm subunits and thus, these Naa38 interfaces are not available for homo- or hetero-oligomerization with other Lsm proteins (see Fig. 27).

The second question relates to the possible interaction of a Naa38 with RNA. This may be relevant since NatA was shown to contact the ribosomal RNA extension segments (Knorr et al. 2019) (see section 5.5 for further discussion), and thus, the Naa38 subunit may have evolved to mediate this RNA interaction in NatC. Séraphin and coworkers found that Naa38 is more divergent compared to other yeast Lsm proteins and lacks a conserved R-G/C motif in the C-terminal part (Séraphin 1995). In a co-immunoprecipitation experiment, tagged Lsm-constructs, including Naa38, were introduced into yeast cells, to analyze their ability to bind snRNAs. However, for Naa38, none of the snRNAs (U1, U2, U4, U5 or U6) was identified.

By comparing the snRNA-bound LSm2-8 ring with the NatC structure solved in this work, a potential Naa38-RNA interaction may still be envisaged: In the Lsm2-8 ring, the U6 snRNA is bound in a positively charged groove on the inside of the ring (Fig. 47A). The uracil bases of the U6 snRNA are each recognized by conserved residues from two sequence motifs DxXxN and IRGX (Fig. 47B). The DxXxN but not the IRGX motif is also conserved in Naa38.

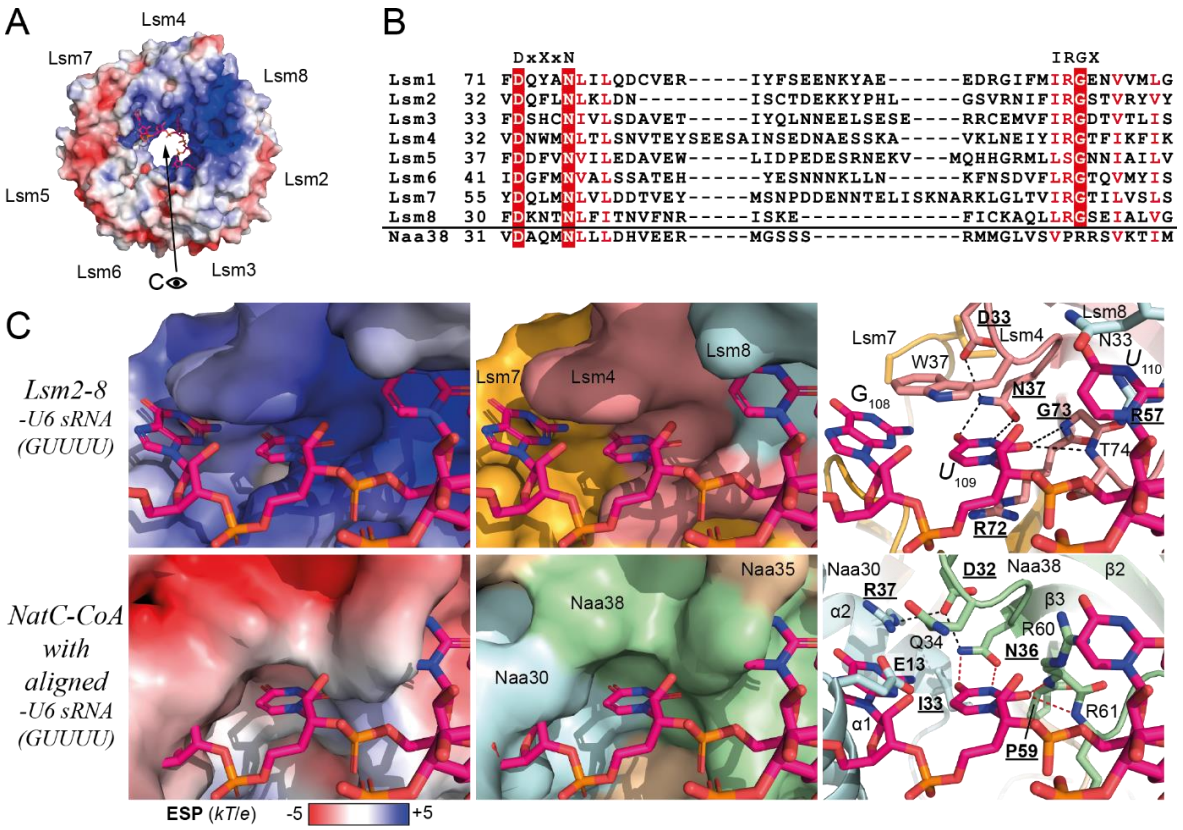


Fig. 47 Absence of U6S sRNA recognition motifs in NatC (A) Electrostatic surface potential projected on the Lsm2-8 heptameric complex (PDB accession code 4M7A). (B) Protein sequence alignment of yeast Lsm paralogs in comparison with Naa38. (C) U6 sRNA recognition by Lsm2-8 (top row). *NatC-CoA* complex showing the theoretical position of U6 sRNA from the Lsm2-8 structure, generated by superposition of Naa38 with Lsm4 (bottom row). Both complexes are projected with their surface charge (left), with differently colored subunits (middle) and in cartoon/stick presentation (right).

In analogy to the snRNA-bound Lsm4 subunit, a superposition with Naa38 reveals a potential RNA binding site in NatC (Fig. 47C). Interestingly, an RNA base would neatly fit into a pocket formed at the interface between subunits Naa35 and Naa38. However, this pocket is surrounded by hydrophobic and positively charged residues and has thus different electrostatic properties compared to the electropositive binding groove in the Lsm2-8 complex.

The tight incorporation of Naa38 in the NatC complex, its unfavorable electrostatic charge properties and the absence of coordinating residues from the IRGX motif strongly indicate that Naa38 is unable to bind RNA.

Starheim et al. (2009) found that, while all the human NatC subunits localize to the cytoplasm, Naa38 is also significantly enriched in the nucleus. They concluded that Naa38 may have a nuclear role linked to RNA processing independent of the NatC complex. However, even if monomeric Naa38 should exists in the nucleus, its overall low isoelectric point of 5.1 (see Fig. 42, section 4.12) suggest that it has a negative net charge in the cell, which makes its interaction with RNA less likely.

5.3 *Insights into NatC substrate recognition and specificity*

While NatA and NatD, require removal of the substrate initiator methionine prior to Nt-acetylation, NatB, NatC, NatE and NatF acetylate the N-terminal methionine with further specificity determined by the second amino acid (Ree, Varland, and Arnesen 2018). The NatB complex has the unique ability to Nt-acetylate methionines followed by an acidic residue (MD-, ME-) or their amide (MN-, MQ) (Starheim et al. 2008; Van Damme et al. 2012). In contrast, the substrate specificity profile of NatC, NatE and NatF is partially overlapping and includes methionine N-termini followed by a hydrophobic amino acid (ML-, MI-, MV-, MY-, MF-, MW-) or lysine (MK-) (compare Table 1, section 1.5).

To obtain insights into the substrate recognition of NatC, two structures have been solved in complex with coenzyme A and two different peptides representing the N-terminal five residues of cognate, well-characterized NatC substrates. The first peptide represents the N-terminus (MFHLV-) of the small GTPase *S. cerevisiae* Arl3, whose N-terminal acetylation is required for its binding to the *trans*-Golgi specific transmembrane protein Sys1, thus determining the localization of yArl3 at the *trans* face of the Golgi apparatus (Behnia et al. 2004; Setty et al. 2004). The second substrate contains the N-terminal five residues (MLRFV) of the Major capsid protein (Gag) of the *S. cerevisiae* virus L-A. Nt-acetylation of Gag was shown to be required for viral particle assembly (Tercero, Dinman, and Wickner 1993). Both peptides interact with the catalytic subunit Naa30 through formation of multiple hydrogen bonds with the peptide backbone (see 4.8.2). The N-terminal methionine, which determines one part of the substrate specificity of NatC, NatE and NatF, is sandwiched in between L27 and Y145. The L27A mutant resulted in the strongest reduction of k_{cat} to an almost undetectable activity (see Table 22, section 4.10.3). In the peptide-free NatC structures, Leu27 is positioned in a hydrophobic pocket in between helices α1 and α2. While binding of the yArl3 peptide induces only minor movements of the α1–α2 loop, binding of the Gag peptide induces a flipping movement of Leu27, which facilitates its interaction with the peptide's Met1. The hydrophobic side chains at position 2 of yArl3 (Phe) and Gag (Leu) are both positioned in a conserved hydrophobic pocket of the catalytic subunit (Fig. 32, section 4.8.2). Tyr80 of the Naa30 subunit contributes to the formation of the hydrophobic pocket and a Y80A mutant exhibited an approximately four times higher K_m compared to wild type NatC, emphasizing its importance for substrate recognition. Tyr80 is conserved among NAT paralogs NatC, NatE and NatF, all of which acetylate methionine N-termini followed by a hydrophobic residue. In contrast, NatA and NatB – which prefer polar or charged residues at position two, respectively, have both a histidine at the corresponding position.

While the substrate's first two amino acids are the major determinants for the specificity of most NATs, amino acid residues at positions 3 and 4 of yArl3 and Gag also contribute to the interaction with NatC. When comparing the two peptides, Gag forms more interactions with NatC, including a side chain interaction of Gag-Arg3 and a main chain interaction of residue 4. Additionally, Gag-Phe4 explores the distal end of the hydrophobic substrate binding pocket. As a result, Gag peptide binding induced drastic rearrangements of the α1–α2 and particularly the β6–β7 loop which can be viewed

as an induced fit-mechanism. Consequently, the Gag peptide displayed a greatly increased catalytic efficiency, compared to the yArl3 peptide. While NatC wild type catalytic parameters for yArl3 are very similar to the values seen for other NATs (Liszczak et al. 2013; Hong et al. 2017), the Gag peptide displays an almost three times higher activity and a low micromolar Michaelis-Menten constant (K_m) (see Table 23, section 4.10.4). As a result, the catalytic efficiency is more than one order of magnitude above the activity observed for the yArl3 peptide.

The high affinity of the Gag-peptide is likely an evolutionary adaption related to the viral origin of the Gag protein. The major coat protein Gag is encoded by the L-A double-stranded (ds) RNA virus of *S. cerevisiae*. Since the L-A virus lacks the outer virion layer which is typical for retroviruses, they do not have an extracellular phase in their life cycle and are only transferred by yeast cell-cell mating (Wickner, Fujimura, and Esteban 2013). Terecero and coworkers have shown that NatC recognizes the N-terminal four residues of the L-A virus Gag protein, which is necessary for viral assembly (Tercero, Riles, and Wickner 1992; Tercero and Wickner 1992; Tercero, Dinman, and Wickner 1993). Since N-terminal acetylation is a requirement for viral particle assembly, the N-terminal sequence of Gag is under strong evolutionary pressure, which explains the excellent catalytic efficiency observed for this substrate. Moreover, it is shown here that the N-terminal four residues make indeed the strongest contribution to Gag binding. A surface representation of the NatC substrate binding pocket reveals that residues 2-3 fill the entire extended substrate pocket that is formed in the Naa30-Naa35 interface (see Fig. 50, top left).

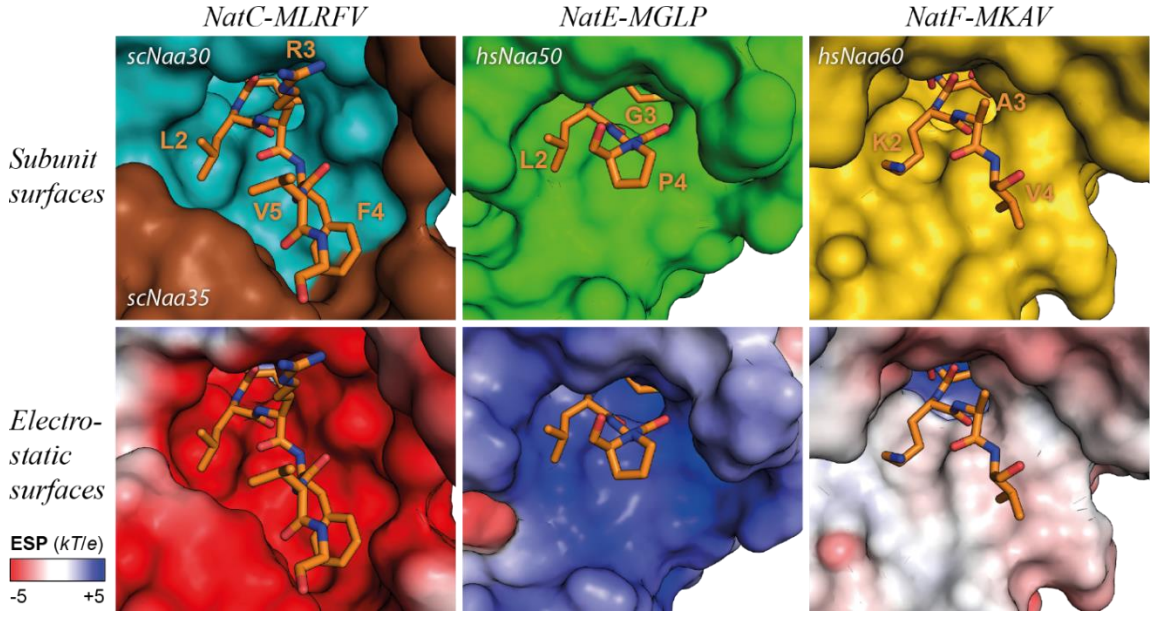


Fig. 48 Comparison between substrate recognition of NatC, NatE and NatF.(A) Surface projection of *NatC-CoA-MLRFV*, *Naa50-Co-Ac-MGLP* (3TFY) and *Naa60*-Co-Ac-*MKAV* (5ICV), showing the active site with bound peptide substrate (NatC) or bisubstrate analogs (Naa50 and Naa50), respectively. (B) Electrostatic surface potential of the active sites.

Even though the substrate pocket is lined by several conserved hydrophobic amino acids (Fig. 32, section 4.8.2), the electrostatic surface potential (see Fig. 50, bottom left) reveals a strongly negative surface charge, which is likely to contribute to the NatC substrate selectivity. Indeed, both yArl3 and Gag have a positively charged residue at position 3, which interact with Naa30 side chains.

Even though NatC, NatE and NatF share substantially overlapping substrate specificity (Table 1, section 1.5), a comparison of their substrate pockets reveals strong differences in the corresponding electrostatic surface potentials (see Fig. 50, bottom row). Since NatF is localized to the Golgi-membrane, where it specifically acetylates transmembrane proteins (Aksnes et al. 2015), the functional redundancy with NatC and NatE is likely to be minor. The opposite charge properties of the NatC and NatE pockets are likely to result in different substrate selectivity, which is not only inferred by the second position, but also subsequent positions of the peptide N-terminus.

5.4 *Catalytic mechanism of NatC*

On the basis of the structure-function studies presented in this work, a catalytic mechanism for the NatC complex was developed, which is presented in the following and compared against related NAT orthologs in the subsequent section (see 5.4.1). Different views of the active site of ligand-bound Naa30 show the residues that are involved in the catalytic mechanism (Fig. 49 A-C). Since no structure of NatC in complex with acetyl-CoA was produced in this work, the *Naa10-AcCoA* structure was aligned with Naa30, to show the approximate position of the acetyl-group in the proposed catalytic mechanism (Fig. 49 A). Based on this information, a reaction scheme was proposed (Fig. 49).

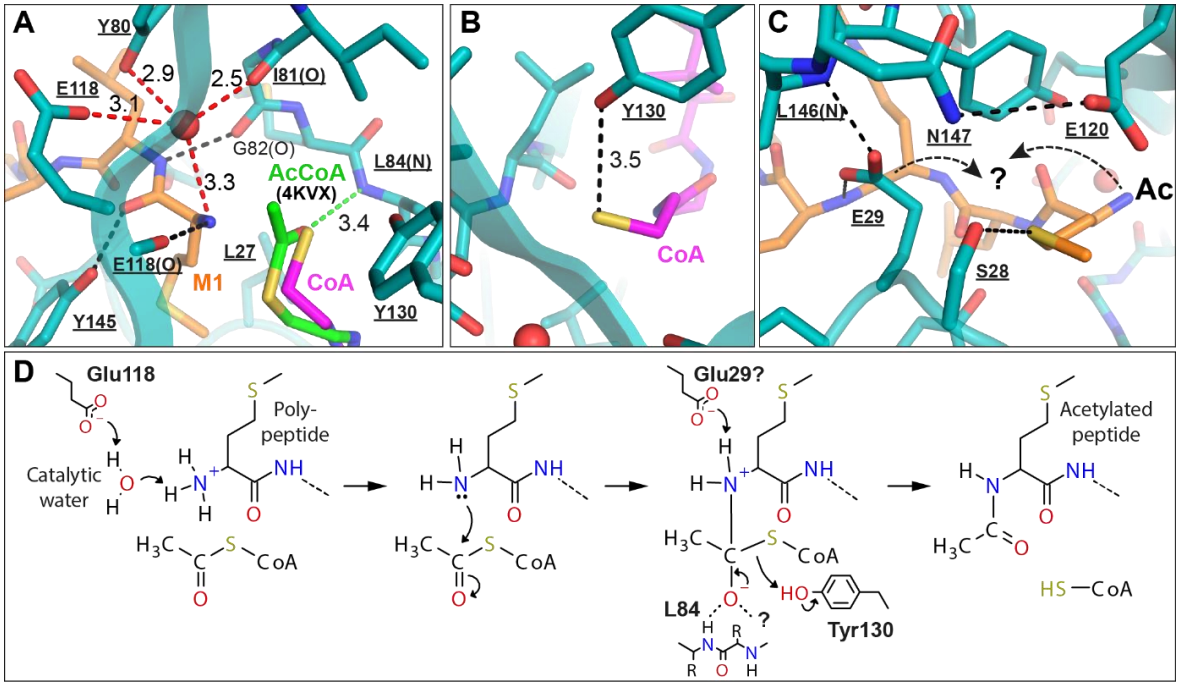


Fig. 49 Proposed catalytic mechanism for the NatC complex. (A) N-terminal acetylation in NatC starts with the deprotonation of the peptide amino terminus by the general base Glu118 and is mediated through an ordered water molecule. The activated amino group attacks the carbonyl oxygen of the enzyme-bound acetyl-CoA, resulting in a transient tetrahedral intermediate, which is stabilized by the backbone amide of Leu84 (oxyanion hole). The AcCoA molecule in the active site is obtained from a superposition of the *Naa10-AcCoA* structure (4KVX) with Naa30 (RMSD = 0.98Å over Cα 100 atoms) and shows the approximate location of the oxyanion in the tetrahedral intermediate. (B) Tyr130 serves as a general acid and donates a proton to break the thioester bond of the tetrahedral intermediate. (C) Glu29 is likely to act as a second general base, which deprotonates the acetylated peptide at the former amino terminus. (D) Complete mechanism for NatC-catalyzed transfer of an acetyl group from AcCoA to a peptide amino nitrogen.

Deprotonation of the peptide amino group

The first step of the proposed catalytic mechanism involves the deprotonation of the amino group of the acceptor substrate peptide by a general base[44], which in case of NatC, is likely mediated via a coordinated water molecule. In all NatC structures, a clear electron density for a water molecule was visible in the active site (see Fig. 30, section 4.8.1). The B-factor of the water molecule was similar to the average B-factor of surrounding macromolecule atoms (Fig. 31, section 4.8.1), indicating that this water is tightly coordinated. This coordination, together with the close vicinity to the amino group of the peptide substrate, suggests a potential role of this water molecule in catalysis.

The proposed catalytic water (red sphere, Fig. 49A), is positioned 3.3 Å away from the peptide amino group and coordinated by the hydroxyl group of the Tyr80 side chain, the backbone carbonyl oxygen of Ile81 and the carboxyl group of the Glu118 side chain. Glu118 is likely to serve as the general base, as E118A or E118Q substitutions result in a ~90% reduction of the k_{cat} compared to wild type Naa30. Y80A or Y80F led to a ~50% reduced k_{cat}, which supports the prosed role of Tyr80 in coordinating the catalytic water.

Nucleophilic attack on the carbonyl carbon of acetyl-CoA

In a second step, the activated (i.e. deprotonated) peptide amino group performs a nucleophilic attack on the carbonyl carbon of the enzyme bound acetyl coenzyme A (AcCoA), resulting in a transient zwitterionic tetrahedral intermediate (third step of Fig. 49 D). Simultaneously, the resultant negative charge on the carbonyl oxygen (oxyanion) is most probably stabilized by the backbone amide of Leu84.

Proton transfer and breakdown of the tetrahedral intermediate

The breakdown of the intermediate is likely to occur through a proton transfer from Tyr130, which serves as the general acid (Fig. 49B). At the same time, the electron pair of the oxyanion flips over to form the double bond of the carbonyl oxygen, thereby breaking the thioester bond of the CoA-intermediate. In acetyltransferase assays, Y130A and Y130F mutants exhibited an extremely reduced k_{cat}, corresponding to ~1-2% of wild type NatC turn over numbers, while showing no significant changes in K_m, emphasizing the crucial role of Tyr130.

Second deprotonation of the peptide amide

After the nucleophilic attack, the nitrogen atom of the former peptide carries a positive charge and needs to remove its excess hydrogen atom.

The most likely candidate for this second deprotonation step is the highly conserved Glu29, which would represent the second general base. E29A and E29Q mutants exhibited a strongly reduced k_{cat}, corresponding to ~4-6% of wild type NatC turn over numbers and no other function has been

[44] General base catalysis refers to the partial proton abstraction by a proton acceptor and can be performed by the unprotonated forms of Asp, Glu, His, Tyr, Cys, Lys. The same amino acids can act as general acids in their protonated form (http://faculty.smu.edu/svik/5310/5310lectures/lect14.html).

assigned for this residue in the catalytic mechanism. In the *NatC-CoA-MLRFV* structure, the amino group of the substrate peptide and the carboxyl group of Glu29 are about 9 Å apart. However, Glu29 is part of the flexible α1–α2 loop, which allows for large movements of the loop residues. Furthermore, it is possible that the acetylated peptide encounters Glu29 during the product dissociation process.

In summary, N-terminal acetylation in NatC starts with the deprotonation of the peptide amino terminus by the general base Glu118 and is mediated through an ordered water molecule. The activated amino group attacks the carbonyl oxygen of the enzyme-bound acetyl-CoA, resulting in a transient tetrahedral intermediate, which is most probably stabilized by the backbone amide of Leu84. Tyr130 serves as a general acid and donates a proton to break the thioester bond of the tetrahedral intermediate. Finally, Glu29 is likely to act as a second general base, which deprotonates the acetylated peptide at the former amino terminus.

5.4.1 Comparison of the catalytic mechanism

The active site of Naa30 is now compared to catalytic subunits from the following NAT paralogs: *S. pombe* NatA (Liszczak et al. 2013), *C. albicans* NatB (Hong et al. 2017) human Naa50 (Liszczak, Arnesen, and Marmorstein 2011) and human Naa60 (Stove et al. 2016). A superposition of scNaa30 with these NAT structures shows the highest structural similarities with the complexed catalytic subunits of NatA and NatB. Only two active side residues are universally conserved among the major eukaryotic NAT paralogs Naa10, Naa20, Naa30, Naa50 and Naa60 (Rathore et al. 2016) (see Fig. 36, section 4.10.1), which correspond to Tyr130 and Tyr145 in scNaa30 (Fig. 50). In addition, Tyr31 is completely conserved in all paralogs except Naa20 in which it is either a tyrosine or a phenylalanine.

Tyr31 and Tyr145 form hydrogen bonds with the peptide backbone and phenylalanine substitutions (Y31F, Y145F) showed significant reductions in k_{cat} (Table 22, section 4.10.3). Similar strong or total losses of catalytic activity for corresponding residues in NAT paralogs were also observed by others for *S. pombe* NatA (Liszczak et al. 2013), hsNaa50 (Liszczak, Arnesen, and Marmorstein 2011) and hsNaa60 (Stove et al. 2016; Chen et al. 2016). In *C. albicans* NatB, Tyr31 is replaced by a phenylalanine (F27), which makes a hydrophobic interaction instead of a hydrogen bond with the substrate backbone (Hong et al. 2017).

In the present work, Glu118 is proposed to deprotonate the peptide amino group via a well-ordered water molecule in the center of the active site. Liszczak, Arnesen, and Marmorstein (2011) have identified a similar well-ordered water molecule for hsNaa50 (Fig. 50C) and proposed that the nearby Y73 and H112 cooperate in the deprotonation of the amino group via the observed water molecule. In Naa60, an identical catalytic water and an equal pair of amino acids (Y97 and H138) are present and thus the same mechanism was suggested for proton abstraction from the peptide amino terminus (Stove et al. 2016; Chen et al. 2016). A glutamate was also suggested to function as the general base in the GNAT histone acetyltransferase Gcn5 from *S. cerevisiae* (Trievel et al. 1999).

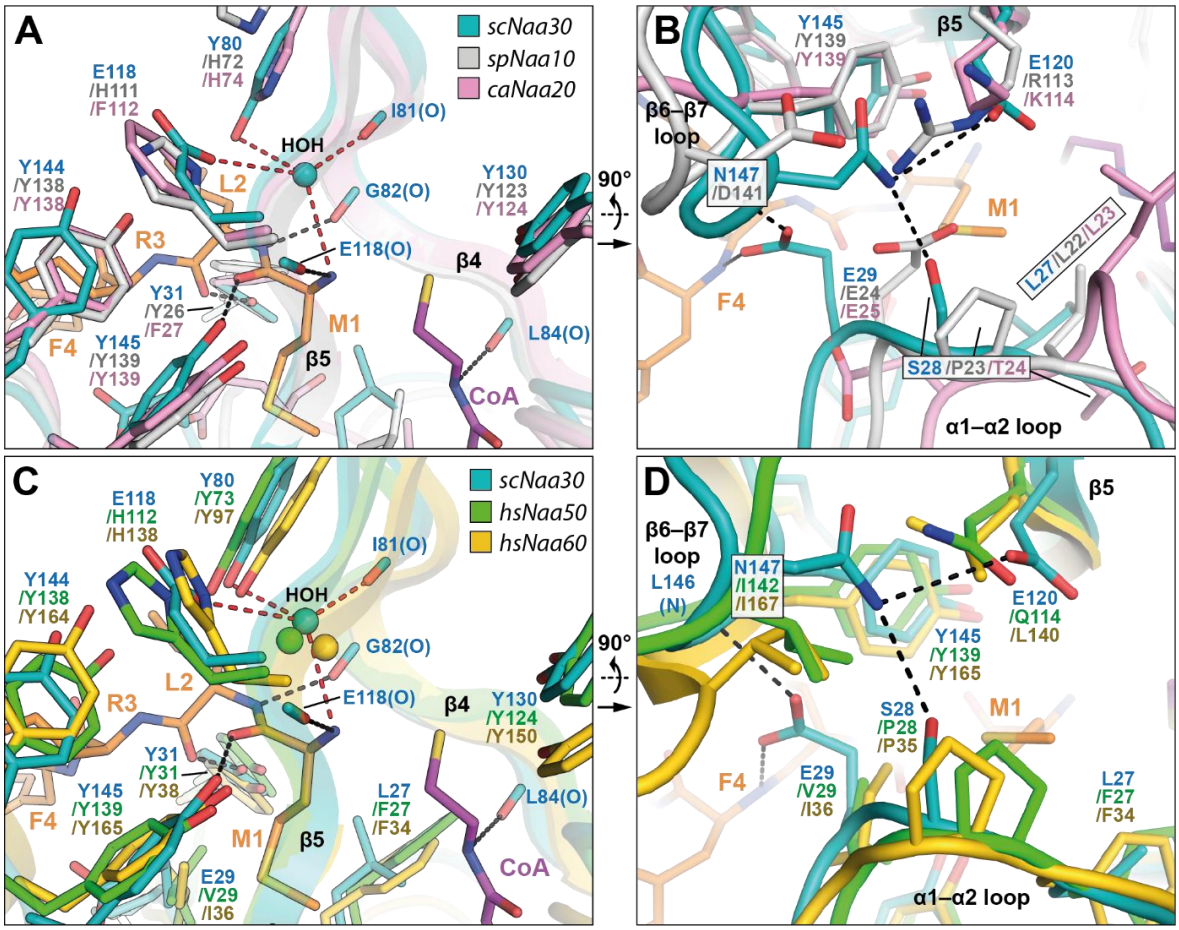


Fig. 50 Comparison of the active sites of NatC, NatE and NatF. Superposition of *NatC-CoA-MLRFV*, *spNatA-CoA-Ac-SASE* (4KVM), *caNatB-CoA-Ac-MDSE* (5K18), *Naa50-CoA-Ac-MGLP* (3TFY) and *Naa60-CoA-Ac-MKAV* (5ICV). For simplicity, only the NatC CoA (purple) and Gag peptide (orange) are shown. (A) View of the active site. The catalytic water in Naa30, Naa50 and Naa60 is shown as a sphere. Hydrogen bonds show the coordination of the water in Naa30 (red lines) and the NatC Gag substrate peptide (black lines), respectively. (B) 90° rotated view, showing the coordination of the Met1 residue. Additional hydrogen-bonding involving the α1-α2 loop, β6–β7 loop, β5 and one backbone amide of the substrate peptide (F4) are shown for NatC.

In the catalytic mechanism proposed here, Tyr130 in NatC was suggested to serve as a general acid for the breakdown of the tetrahedral intermediate. Even though Tyr130 is completely conserved in all major eukaryotic NATs, very few mutational studies are available, except for the NatB complex, where a mutation to phenylalanine reduced the k_{cat} to 4% of wild type activity, with only a minor increase of the K_m (Hong et al. 2017). Thus, Tyr130 is crucial for catalysis in NatC and potentially in all major NATs. The catalytic mechanism proposed in this work now assigns a potential role to this highly conserved tyrosine.

Glu29 in NatC was suggested to act as a second general base. In NatA, the corresponding glutamate (E24) showed similar strong reductions in k_{cat} with negligible effects on K_m (Liszczak et al. 2013). However, a mutation of the corresponding glutamate in NatB (E25A) increased its enzymatic activity by ~60% (Hong et al. 2017). Given the observed differences for Glu29 in different NAT paralogs and its large distance from the peptide amino group, the function of Glu29 in the final deprotonation step remain elusive.

5.5 *NatC-ribosome interaction*

Gautschi et al. (2003) first revealed that NatA is quantitatively anchored to ribosomes via the auxiliary subunit Naa15 to enable co-translational Nt-acetylation of NatA substrates. Polevoda et al. (2008) showed that all major NATs, e.g. NatA, NatB, NatC and NatE, are associated with mono- and polyribosomes. The authors furthermore showed ribosomal proteins uL23 and uL14 to associate with NatA in pulldown experiments. Both proteins are located in close proximity to the ribosomal exit tunnel. Using co-sedimentation assays with purified ribosomes and NatA, Magin et al. (2017) identified two conserved electropositive regions (EPR) in NatA, that are responsible for its ribosome interaction. A recent cryo-EM structure of the trimeric NatA/Naa50 complex (comprising Naa10, Naa15 and Naa50) bound to the *S. cerevisiae* ribosome (Knorr et al. 2019) revealed that the trimeric NatA/Naa50 complex interacts with ribosomal RNA expansion segments via its large auxiliary subunit Naa15 and Naa50 (Fig. 51A). The authors have identified four interaction sites on the ribosome, all of which contain ribosomal RNA. One interaction is formed between Naa15 and rRNA helices H24 and H47 (Fig. 51, colored in green) close to the peptide exit tunnel (PE). All other interactions are formed with RNA expansion segments (ES) (Fig. 51, colored in blue) which are specific for eukaryotic ribosomes. Thus, Naa15 interacts with ES27a and ES39a, whereas Naa50 binds ES7a. The catalytic Naa10 does not form contacts with the ribosome and is positioned dynamically in close proximity to the ribosomal exit tunnel (Knorr et al. 2019).

In this work, NatC-sedimentation assays showed that the electropositive *Naa35 tip* region is necessary for association of NatC with the ribosome *in vitro*. Thus, it is conceivable that NatC also interacts with negatively charged ribosomal RNAs. In order to obtain first insight into the possible interactions of NatC with the ribosome, two different superpositions of NatC with the NatA/Naa50-ribosome cryo-EM structure were generated: The catalytic subunit of NatC, Naa30, was aligned with either Naa10 (Fig. 51B) or Naa50 (Fig. 51C) of the trimeric NatA/Naa50 complex, both of which feature the conserved GNAT fold.

The superposition of the two catalytic subunits, Naa10 and Naa30, reveals that large parts of the auxiliary subunit Naa35 would be in close contact with the surface of the ribosome, while exhibiting only minor clashes (Fig. 51B). Most importantly, the *Naa35 Tip* region would be in close proximity with ribosomal RNA, which is in line with its essential role in NatC-ribosome binding as established in this work. Interestingly, this putative interaction would involve rRNA helices H18-H20, which are not involved in binding of the NatA/Naa50 complex.

The second possible superposition aligns Naa30 of NatC with Naa50 of the NatA/Naa50-ribosome complex (Fig. 51C). In this configuration, hardly any interaction between NatC and the ribosome was observed. Moreover, the *Naa35 Tip* is positioned far away from the ribosome surface. Therefore, this orientation is very unlikely to exist *in vivo*.

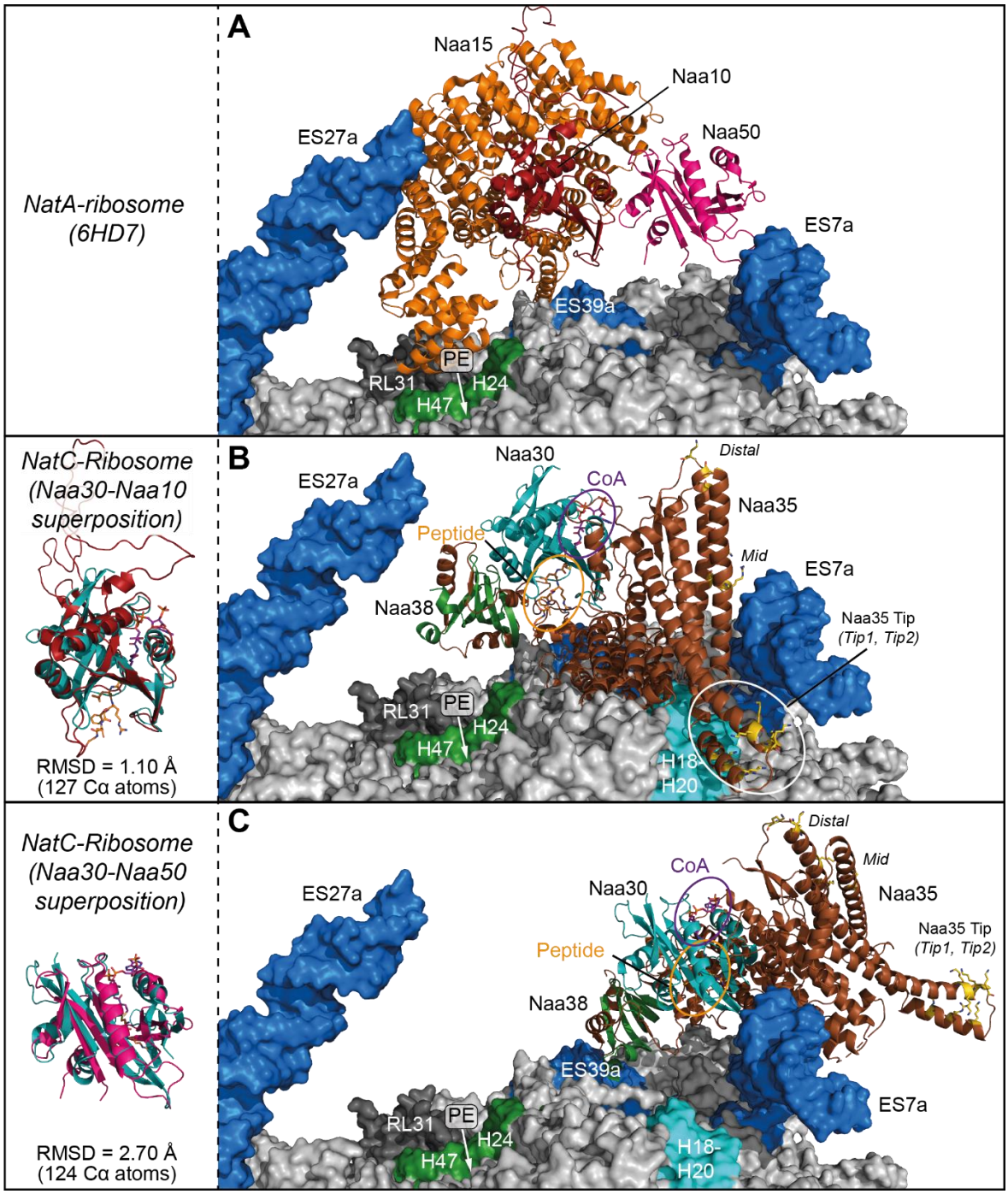


Fig. 51 Potential NatC-ribosome interfaces. (A) Molecular model of the NatA/Naa50-ribosome complex (6HD7) near the peptide exit tunnel (PE). The catalytic subunit Naa30 of the NatC-CoA-MLRFV structure was used for a superposition with either Naa10 (B) or Naa50 (C) in the NatA/Naa50-ribosome complex. The superpositions of Naa30 with either Naa10 or Naa15 are shown in the smaller boxes on the left. The NatA/Naa50-complex is omitted in the NatC-ribosome models for better clarity. The Naa35 Tip mutants are highlighted in gold.

The secondary structure of the 28S rRNA of *S. cerevisiae* is shown in Fig. 52. Expansion segments involved in the NatA/Naa50-ribosome interaction are detailed in Fig. 52B and rRNA segements involved in putative NatC-ribosome interactions are detailed in Fig. 52C.

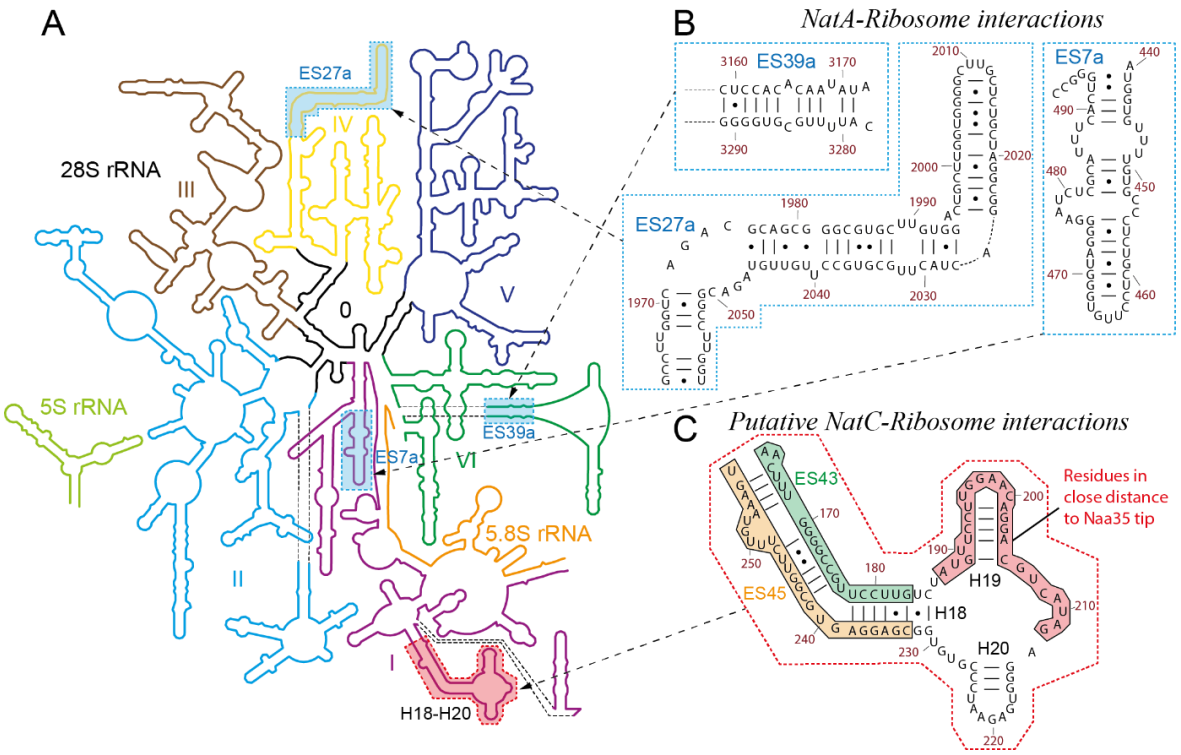


Fig. 52 Secondary structure of the 28S rRNA of *S. cerevisiae*. (A) Overview of the 28S rRNA (adapted from http://apollo.chemistry.gatech.edu/RibosomeGallery/). Expansion segments involved in the NatA/Naa50-ribosome interaction are labeled in blue. Additional rRNA segments involved in the putative NatC-ribosome interaction are colored in red. Zoom views are shown in (B) and (C) respectively.

Returning to the superposition of Naa30 with Naa10, more potential NatC-ribosome interactions can be predicted from the electrostatic surface potential of NatC. In addition to the electropositive region in the *Naa35 Tip* (EPR1), two more electropositive patches on the NatC surface are potential candidates for ribosome interactions (Fig. 53A). Both EPR2 and EPR3 could interact with the highly flexible ES27a RNA expansion segment, with EPR2 being the more likely candidate due to its larger dimensions and stronger conservation: EPR2 contains ionizable side chain groups from six basic residues of Naa30 and one from Naa35 (Fig. 53B). Three Naa30 residues are universally conserved (H72, R78 and K139) and two more Naa30 residues (R73 and K153) are highly conserved. In between these residues, electron density indicated the presence of a ligand which was attributed to two PEG molecules from the crystallization solution. It is conceivable that the PEG binding site overlaps with an RNA binding site in Naa30.

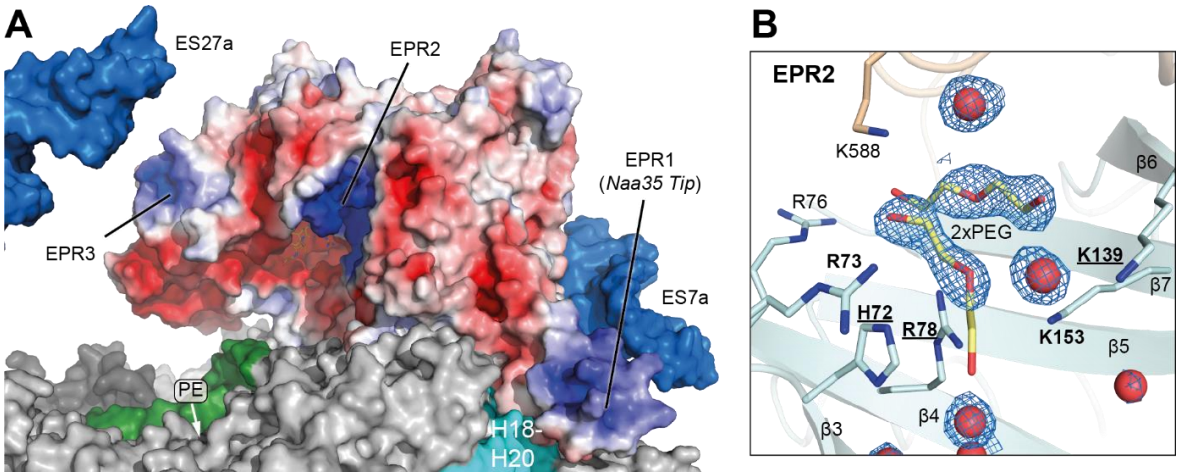


Fig. 53 Electrostatic surface potential of NatC reveals further putative ribosome interaction sites. (A) Superposition of Naa30 from NatC-CoA-*MLRFV* with Naa10 from the NatA/Naa50-ribosome complex. The electrostatic surface potential (ESP) is projected on the surface of NatC and reveals three electropositive regions (EPR) as potential ribosome binding sites. (B) Detailed view of EPR2 in the *NatC-CoA* structure. Highly conserved residues are labeled in bold, with completely conserved residues underlined. A 2Fo-Fc map, contoured at 1σ, is shown for the PEG and water molecules modeled near EPR2.

In summary, the electropositive *Naa35 tip* region was shown to be crucial for NatC-ribosome interactions *in vitro*. A superposition of NatC with the NatA/Naa50-ribosome structure revealed ribosomal RNA helices H18-H20 as the potential binding site for the *Naa35 tip* region of NatC. The conserved EPR2 region in NatC was identified as another promising candidate for the putative interaction of NatC with the ribosomal expansion segment ES27a. Further NatC-ribosome sedimentation assays would be necessary to confirm whether or not EPR2 is involved in ribosome binding.

5.5.1 Possible roles of the central tunnel for NatC interactions with the nascent chain

In this work, another remarkable feature of the NatC complex was discovered: two tunnels, which are located roughly at the center of the entire complex at the Naa30-Naa35 interface (Fig. 34, section 4.9). CoA and the substrate peptide meet head-to head in the *side tunnel*, which contains the active site of the catalytic subunit Naa30. The circular *central tunnel* is separated from the *side tunnel* by the α1-α2 loop and the β6-β7 loop and is constricted to 4 Å at its narrowest point. The *side tunnel* has a more elliptical cross-section with a height of ~5 Å and a width of ~ 13 Å in the peptide-free state. Upon peptide binding, the *side tunnel* becomes significantly constricted due to movements of the flexible α1-α2 and β6-β7 loops.

This high degree of flexibility suggests that, conversely, the two loops can also open up further, fully connecting the *central* and *side tunnel*. The opposite site of the *central tunnel* is lined by four additional Naa35 loop regions, which further shows the expandability of this region. It is thus possible, that the *central tunnel* can accommodate the nascent polypeptide chain as it emerges from the ribosomal exit tunnel, directing its N-terminus to the center of the active site. The large flexibility of the *central tunnel* might be necessary to accommodate NatC substrates that contain a bulky hydrophobic amino acid at position two (e.g. MY-, MF- or MW-).

As outlined in a current review (Waudby, Dobson, and Christodoulou 2019), nascent polypeptide chains emerging from the exit tunnel will begin to fold and assemble their tertiary and quaternary structure during biosynthesis in a vectorial manner, starting at their N-terminus[45]. While simple secondary structures (e.g. small α helices) may form within the ribosomal exit tunnel (Bhushan et al. 2010), tertiary structure formation can only occur upon reaching the exit tunnel vestibule (Kudva et al. 2018). Upon emerging from the exit tunnel, the nascent chain is exposed to the surface of the ribosome itself, which (i) constraints the nascent chain to regions adjacent to the exit tunnel (Cabrita et al. 2016) and (ii) slows down folding of the nascent chain (Kaiser et al. 2011). However, soon after the nascent chain emerges, it can collapse into a co-translational folding intermediate[46] (Kelkar et al. 2012). The supposed close proximity of NatC with the exit tunnel, would allow it to N-terminally acetylate its substrate before the nascent chain starts to collapse into a folding intermediate (e.g. the special case of the molten globule). The formation of the kinetic molten globule is initiated by a collapse of the unfolded protein chain and generally thought to be driven by hydrophobic interactions (i.e. hydrophobic collapse) (Udgaonkar 2013). Since some NatC substrates contain a bulky hydrophobic amino acid at position two (e.g. MY-, MF- or MW-), they are likely to collapse quickly (and consequently will be inaccessible for Nt-acetylation). It is possible that the structure of the NatC complex, especially of its *central tunnel* region, has evolved to accommodate and delay the collapse of the hydrophobic substrates until they have been Nt-acetylated by the catalytic subunit Naa30.

[45] Today protein folding is seen as a "parallel microscopic multi-pathway diffusion like process" (Dill and Chan 1997) along an energy landscape that resembles the movement of a ball down the surface of a "rugged" funnel, containing local minima that resemble non-native conformations. The bottom of the funnel represents the global minimum free energy state of the native folded protein.

[46] A specific type of folding intermediate is the (kinetic) 'molten globule', which is as compact as the native protein and may contain secondary structure elements, but has a fluctuating tertiary structure (Ohgushi and Wada 1983).

6 Conclusion

Based on a structural and biochemical characterization of NatC, its structure, substrate preference, catalytic mechanism and ribosome interaction surfaces have been characterized in this work. Future studies need to address the exact binding modes of NatC on the ribosome to understand the detailed mechanism of co-translational Nt-acetylation of target substrates, for example, by cryo-EM analyses. Furthermore, based on the structural work, it may now be possible to design structure-based NatC mutants which specifically abrogate N-terminal acetylation of specific subsets of NatC substrates. This may provide exciting opportunities to further study the biological function of N-terminal acetylation in specific target proteins.

7 Summary

N-terminal acetylation is one of the most abundant protein modifications in eukaryotes, affecting an estimated 80% of all human proteins and has diverse implications for the stability, folding, half-life, interaction ability and subcellular targeting of the acetylated protein. This irreversible modification is catalysed by a highly conserved family of N-terminal acetyltransferases (NATs). Eight different NATs have been described in eukaryotes so far, which differ in their substrate specificity, their subunit composition and their subcellular localization. The only major eukaryotic NAT that has thus far eluded structural determination, is the heterotrimeric NatC complex, consisting of the catalytic subunit Naa30 and the two auxiliary subunits Naa35 and Naa38.

In this work, five structures of the *Saccharomyces cerevisiae* NatC complex in the absence and presence of CoA and two peptide substrates are presented. The structural analyses show that the large auxiliary subunit Naa35 is the central assembly hub of NatC that forms extensive interactions with the catalytic subunit Naa30 and the small auxiliary subunit Naa38. Thus, Naa35 wraps around Naa30 in an open ring-like tertiary structure and its N-terminus embraces Naa38 around almost its entire circumference.

Most strikingly, the architecture of the NatC complex completely differs from that of the heterodimeric NAT complexes NatA and NatB. Most importantly, the position of the NatC auxiliary subunit relative to the catalytic subunit shows the exact opposite arrangement seen in the NatA and NatB auxiliary subunits.

NatC structures in complex with two cognate peptides corresponding to the N-terminus of yeast Arl3 and the viral Gag protein provided structural insight into NatC substrate specificity and show that the peptides are bound in an elongated hydrophobic substrate binding pocket that is formed in between subunits Naa30 and Naa35. Binding of the high-affinity Gag peptide, but not of the Arl3 peptide, resulted in large ligand-induced structural rearrangements of the substrate binding loop. Accordingly, in acetyltransferase assays, the Gag peptide displayed a 10-fold increased catalytic efficiency compared to Arl3.

Based on a comprehensive structure-function study, a catalytic mechanism for the NatC complex was proposed. Finally, by comparing the NatC structure to the ribosome-bound NatA structure and supported by structure-based mutagenesis, a model for the ribosome-bound NatC complex is presented.

8 Zusammenfassung

Die N-terminale Acetylierung ist eine der am häufigsten vorkommenden Proteinmodifizierungen in Eukaryoten, welche geschätzte 80% aller humanen Proteine betrifft und weitreichende Auswirkungen auf die Stabilität, Faltung, Lebensdauer, das Interaktionsvermögen und die subzelluläre Verteilung des acetylierten Proteins hat. Diese irreversible Modifikation wird von einer stark konservierten Familie N-terminaler Acetyltransferasen (NATs) katalysiert. Acht verschiedene NATs wurden bisher in Eukaryonten beschrieben, welche sich durch ihre Substratspezifität, Untereinheiten-Zusammensetzung und ihre subzelluläre Lokalisation unterscheiden. Die einzige eukaryotische N-terminale Acetyltransferase die bisher nicht strukturell charakterisiert werden konnte, ist der heterotrimere NatC-Komplex, welcher aus der katalytischen Untereinheit Naa30 und den beiden unterstützenden Untereinheiten Naa35 und Naa38 besteht.

In dieser Arbeit werden fünf verschiedene Strukturen des *Saccharomyces cerevisiae* NatC-Komplexes in Gegenwart und Abwesenheit von Coenzyme A und zwei Substratpeptiden vorgestellt.

Die strukturelle Analyse zeigt, dass die große Neben-Untereinheit Naa35 die zentrale strukturgebende Funktion einnimmt und umfangreiche Interaktionen mit der katalytischen Untereinheit Naa30 und der kleinen Untereinheit Naa38 eingeht. Naa35 umhüllt auf diese Weise Naa30 in einer offen-ringförmigen Tertiärstruktur. In ähnlicher Weise wickelt sich der N-terminus von Naa35 fast vollständig um den gesamten Kreisumfang von Naa38.

Zwei NatC Strukturen im Komplex mit Peptiden der natürlichen Substrate Arl3 (aus Hefe) und dem viralen Gag-Protein ermöglichen einen strukturellen Einblick in die Substratspezifität von NatC und zeigen weiterhin die Bindung der beiden Peptide in einer langgestreckten, hydrophoben Substrat-Bindetasche, welche zwischen den beiden Untereinheiten Naa35 und Naa38 gebildet wird.

Die Bindung des Gag-Peptides, aber nicht die von Arl3, resultiert in einer großen Ligand-induzierten strukturellen Neuordnung der Substrat-Bindeschlaufe. Demgemäß zeigt das Gag-Peptid eine zehnfach erhöhte katalytische Effizienz im Vergleich zu Arl3.

Auf Basis der umfangreichen Struktur-Funktions-Studien wurde ein katalytischer Mechanismus für den NatC Komplex vorgeschlagen. Aufgrund des Vergleichs der NatC-Struktur mit der Ribosom-gebundenen NatA-Struktur, unterstützt durch strukturbasierte Mutagenese, wird ein Modell für den Ribosomen-gebundenen NatC Komplex präsentiert.

9 References

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, ..., and J. C. Venter. 2000. 'The genome sequence of Drosophila melanogaster', *Science*, 287: 2185-95.

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.

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, N. Echols, J. J. Headd, L. W. Hung, S. Jain, G. J. Kapral, R. W. Grosse Kunstleve, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. 2011. 'The Phenix software for automated determination of macromolecular structures', *Methods*, 55: 94-106.

Aebersold, R., J. N. Agar, I. J. Amster, M. S. Baker, C. R. Bertozzi, E. S. Boja, C. E. Costello, B. F. Cravatt, C. Fenselau, B. A. Garcia, Y. Ge, J. Gunawardena, R. C. Hendrickson, P. J. Hergenrother, C. G. Huber, A. R. Ivanov, O. N. Jensen, M. C. Jewett, N. L. Kelleher, L. L. Kiessling, N. J. Krogan, M. R. Larsen, J. A. Loo, R. R. Ogorzalek Loo, E. Lundberg, M. J. MacCoss, P. Mallick, V. K. Mootha, M. Mrksich, T. W. Muir, S. M. Patrie, J. J. Pesavento, S. J. Pitteri, H. Rodriguez, A. Saghatelian, W. Sandoval, H. Schluter, S. Sechi, S. A. Slavoff, L. M. Smith, M. P. Snyder, P. M. Thomas, M. Uhlen, J. E. Van Eyk, M. Vidal, D. R. Walt, F. M. White, E. R. Williams, T. Wohlschlager, V. H. Wysocki, N. A. Yates, N. L. Young, and B. Zhang. 2018. 'How many human proteoforms are there?', *Nat Chem Biol*, 14: 206-14.

Afonine, P. V., R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, and P. D. Adams. 2012. 'Towards automated crystallographic structure refinement with phenix.refine', *Acta Crystallogr D Biol Crystallogr*, 68: 352-67.

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.

Aksnes, Henriette, Adrian Drazic, Michaël Marie, and Thomas Arnesen. 2016. 'First Things First: Vital Protein Marks by N-Terminal Acetyltransferases', *Trends in Biochemical Sciences*, 41: 746-60.

Allfrey, V. G., R. Faulkner, and A. E. Mirsky. 1964. 'Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis', *Proc Natl Acad Sci U S A*, 51: 786-94.

Antequera, F., and A. Bird. 1993. 'Number of CpG islands and genes in human and mouse', *Proc Natl Acad Sci U S A*, 90: 11995-9.

Arabidopsis Genome, Initiative. 2000. 'Analysis of the genome sequence of the flowering plant Arabidopsis thaliana', *Nature*, 408: 796-815.

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., 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.

Ashkenazy, H., S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, and N. Ben-Tal. 2016. 'ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules', *Nucleic Acids Res*, 44: W344-50.

Battye, T. G., L. Kontogiannis, O. Johnson, H. R. Powell, and A. G. Leslie. 2011. 'iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM', *Acta Crystallogr D Biol Crystallogr*, 67: 271-81.

Behnia, R., B. Panic, J. R. Whyte, and S. Munro. 2004. 'Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p', *Nat Cell Biol*, 6: 405-13.

Ben-Saadon, R., I. Fajerman, T. Ziv, U. Hellman, A. L. Schwartz, and A. Ciechanover. 2004. 'The tumor suppressor protein p16(INK4a) and the human papillomavirus oncoprotein-58 E7 are naturally occurring lysine-less proteins that are degraded by the ubiquitin system. Direct evidence for ubiquitination at the N-terminal residue', *J Biol Chem*, 279: 41414-21.

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

Bhushan, S., M. Gartmann, M. Halic, J. P. Armache, A. Jarasch, T. Mielke, O. Berninghausen, D. N. Wilson, and R. Beckmann. 2010. 'alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel', *Nat Struct Mol Biol*, 17: 313-7.

Black, D. L. 2003. 'Mechanisms of alternative pre-messenger RNA splicing', *Annu Rev Biochem*, 72: 291-336.

Brennan, S., and P. L. Cowan. 1992. 'A suite of programs for calculating x-ray absorption, reflection, and diffraction performance for a variety of materials at arbitrary wavelengths', *Review of Scientific Instruments*, 63: 850-53.

Brunger, A. T. 1992. 'Free R value: a novel statistical quantity for assessing the accuracy of crystal structures', *Nature*, 355: 472-5.

Cabrita, L. D., A. M. E. Cassaignau, H. M. M. Launay, C. A. Waudby, T. Wlodarski, C. Camilloni, M. E. Karyadi, A. L. Robertson, X. Wang, A. S. Wentink, L. Goodsell, C. A. Woolhead, M. Vendruscolo, C. M. Dobson, and J. Christodoulou. 2016. 'A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding', *Nat Struct Mol Biol*, 23: 278-85.

Chang, Yu-Yung, and Chun-Hua Hsu. 2016. 'Multiple Conformations of the Loop Region Confers Heat-Resistance on SsArd1, a Thermophilic NatA', *Chembiochem*, 17: 214-17.

Chaudhary, S., W. Khokhar, I. Jabre, A. S. N. Reddy, L. J. Byrne, C. M. Wilson, and N. H. Syed. 2019. 'Alternative Splicing and Protein Diversity: Plants Versus Animals', *Front Plant Sci*, 10: 708.

Chen, Ji-Yun, Liang Liu, Chun-Ling Cao, Mei-Jun Li, Kemin Tan, Xiaohan Yang, and Cai-Hong Yun. 2016. 'Structure and function of human Naa60 (NatF), a Golgi-localized bi-functional acetyltransferase', *Scientific Reports*, 6.

Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. 2010. 'MolProbity: all-atom structure validation for macromolecular crystallography', *Acta Crystallogr D Biol Crystallogr*, 66: 12-21.

Choudhary, C., B. T. Weinert, Y. Nishida, E. Verdin, and M. Mann. 2014. 'The growing landscape of lysine acetylation links metabolism and cell signalling', *Nat Rev Mol Cell Biol*, 15: 536-50.

Chovancova, E., A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor, and J. Damborsky. 2012. 'CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures', *PLoS Comput Biol*, 8: e1002708.

Council, National Research. 1988. *Mapping and Sequencing the Human Genome* (The National Academies Press: Washington, DC).

Cowtan, K. 2006. 'The Buccaneer software for automated model building. 1. Tracing protein chains', *Acta Crystallogr D Biol Crystallogr*, 62: 1002-11.

Cromer, D. T., and D. A. Liberman. 1981. 'Anomalous dispersion calculations near to and on the long-wavelength side of an absorption edge', *Acta Crystallographica Section A*, 37: 267-68.

Dang, X., J. Scotcher, S. Wu, R. K. Chu, N. Tolic, I. Ntai, P. M. Thomas, R. T. Fellers, B. P. Early, Y. Zheng, K. R. Durbin, R. D. Leduc, J. J. Wolff, C. J. Thompson, J. Pan, J. Han, J. B. Shaw, J. P. Salisbury, M. Easterling, C. H. Borchers, J. S. Brodbelt, J. N. Agar, L. Pasa-Tolic, N. L. Kelleher, and N. L. Young. 2014. 'The first pilot project of the consortium for top-down proteomics: a status report', *Proteomics*, 14: 1130-40.

Deechongkit, S., J. Wen, L. O. Narhi, Y. Jiang, S. S. Park, J. Kim, and B. A. Kerwin. 2009. 'Physical and biophysical effects of polysorbate 20 and 80 on darbepoetin alfa', *J Pharm Sci*, 98: 3200-17.

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.

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.

Dill, Ken A., and Hue Sun Chan. 1997. 'From Levinthal to pathways to funnels', *Nature Structural Biology*, 4: 10-19.

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.

Doublie, S. 1997. 'Preparation of selenomethionyl proteins for phase determination', *Methods Enzymol*, 276: 523-30.

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., L. M. Myklebust, R. Ree, and T. Arnesen. 2016. 'The world of protein acetylation', *Biochim Biophys Acta*, 1864: 1372-401.

Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. 2010. 'Features and development of Coot', *Acta Crystallogr D Biol Crystallogr*, 66: 486-501.

Evans, Gwyndaf, and Robert F. Pettifer. 2001. 'CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra', *Journal of Applied Crystallography*, 34: 82-86.

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.

Ezkurdia, Iakes, Jose Manuel Rodriguez, Enrique Carrillo-de Santa Pau, Jesús Vázquez, Alfonso Valencia, and Michael L. Tress. 2015. 'Most highly expressed protein-coding genes have a single dominant isoform', *Journal of proteome research*, 14: 1880-87.

Faber, P. W., G. T. Barnes, J. Srinidhi, J. Chen, J. F. Gusella, and M. E. MacDonald. 1998. 'Huntingtin interacts with a family of WW domain proteins', *Hum Mol Genet*, 7: 1463-74.

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.

Fischer, E. H., D. J. Graves, E. R. Crittenden, and E. G. Krebs. 1959. 'Structure of the site phosphorylated in the phosphorylase b to a reaction', *J Biol Chem*, 234: 1698-704.

Frottin, F., A. Martinez, P. Peynot, S. Mitra, R. C. Holz, C. Giglione, and T. Meinnel. 2006. 'The proteomics of N-terminal methionine cleavage', *Mol Cell Proteomics*, 5: 2336-49.

Gahloth, D., C. Levy, L. Walker, L. Wunderley, A. P. Mould, S. Taylor, P. Woodman, and L. Tabernero. 2017. 'Structural Basis for Specific Interaction of TGFbeta Signaling Regulators SARA/Endofin with HD-PTP', *Structure*, 25: 1011-24 e4.

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.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. 1996. 'Life with 6000 genes', *Science*, 274: 546, 63-7.

Gordiyenko, Y., S. Deroo, M. Zhou, H. Videler, and C. V. Robinson. 2008. 'Acetylation of L12 increases interactions in the Escherichia coli ribosomal stalk complex', *J Mol Biol*, 380: 404-14.

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.

Grenier, F., D. Matteau, V. Baby, and S. Rodrigue. 2014. 'Complete Genome Sequence of Escherichia coli BW25113', *Genome Announc*, 2.

Hahn, Th. 2002. 'The 230 space groups.' in Th Hahn (ed.), *International Tables for Crystallography Volume A: Space-group symmetry* (Springer Netherlands: Dordrecht).

Helmholtz-Zentrum Berlin für Materialien und Energie. 2016. 'The MX Beamlines BL14.1-3 at BESSY II', *Journal of large-scale research facilities JLSRF*, 2.

Henne, William M, Nicholas J Buchkovich, and Scott D Emr. 2011. 'The ESCRT Pathway', *Developmental Cell*, 21: 77-91.

Hershko, A., H. Heller, E. Eytan, G. Kaklij, and I. A. Rose. 1984. 'Role of the alpha-amino group of protein in ubiquitin-mediated protein breakdown', *Proc Natl Acad Sci U S A*, 81: 7021-5.

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.

Holm, L., S. Kaariainen, P. Rosenstrom, and A. Schenkel. 2008. 'Searching protein structure databases with DaliLite v.3', *Bioinformatics*, 24: 2780-1.

Holm, L., and C. Sander. 1996. 'Mapping the protein universe', *Science*, 273: 595-603.

Holm, Liisa, and Päivi Rosenström. 2010. 'Dali server: conservation mapping in 3D', *Nucleic Acids Research*, 38: W545-W49.

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.

Hu, Z., H. S. Scott, G. Qin, G. Zheng, X. Chu, L. Xie, D. L. Adelson, B. E. Oftedal, P. Venugopal, M. Babic, C. N. Hahn, B. Zhang, X. Wang, N. Li, and C. Wei. 2015. 'Revealing Missing Human Protein Isoforms Based on Ab Initio Prediction, RNA-seq and Proteomics', *Sci Rep*, 5: 10940.

Hunter, T. 1995. 'Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling', *Cell*, 80: 225-36.

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

International Human Genome Sequencing, Consortium. 2004. 'Finishing the euchromatic sequence of the human genome', *Nature*, 431: 931-45.

Jornvall, H. 1975. 'Acetylation of Protein N-terminal amino groups structural observations on alpha-amino acetylated proteins', *J Theor Biol*, 55: 1-12.

Jurrus, E., D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G. W. Wei, M. J. Holst, J. A. McCammon, and N. A. Baker. 2018. 'Improvements to the APBS biomolecular solvation software suite', *Protein Sci*, 27: 112-28.

Kabsch, W. 2010. 'Xds', *Acta Crystallogr D Biol Crystallogr*, 66: 125-32.

Kaiser, Christian M., Daniel H. Goldman, John D. Chodera, Ignacio Tinoco, and Carlos Bustamante. 2011. 'The Ribosome Modulates Nascent Protein Folding', *Science*, 334: 1723-27.

Kambach, Christian, Stefan Walke, Robert Young, Johanna M. Avis, Eric de la Fortelle, Veronica A. Raker, Reinhard Lührmann, Jade Li, and Kiyoshi Nagai. 1999. 'Crystal Structures of Two Sm Protein Complexes and Their Implications for the Assembly of the Spliceosomal snRNPs', *Cell*, 96: 375-87.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. 'Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data', *Bioinformatics*, 28: 1647-9.

Kelkar, D. A., A. Khushoo, Z. Yang, and W. R. Skach. 2012. 'Kinetic analysis of ribosome-bound fluorescent proteins reveals an early, stable, cotranslational folding intermediate', *J Biol Chem*, 287: 2568-78.

Kim, J., S. Sitaraman, A. Hierro, B. M. Beach, G. Odorizzi, and J. H. Hurley. 2005. 'Structural basis for endosomal targeting by the Bro1 domain', *Dev Cell*, 8: 937-47.

Kimura, Y., M. Takaoka, S. Tanaka, H. Sassa, K. Tanaka, B. Polevoda, F. Sherman, and H. Hirano. 2000. 'N(alpha)-acetylation and proteolytic activity of the yeast 20 S proteasome', *J Biol Chem*, 275: 4635-9.

Klose, D. P., B. A. Wallace, and R. W. Janes. 2010. '2Struc: the secondary structure server', *Bioinformatics*, 26: 2624-5.

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.

Kudva, R., P. Tian, F. Pardo-Avila, M. Carroni, R. B. Best, H. D. Bernstein, and G. von Heijne. 2018. 'The shape of the bacterial ribosome exit tunnel affects cotranslational protein folding', *Elife*, 7.

Liang, F., I. Holt, G. Pertea, S. Karamycheva, S. L. Salzberg, and J. Quackenbush. 2000. 'Gene index analysis of the human genome estimates approximately 120,000 genes', *Nat Genet*, 25: 239-40.

Liebschner, D., P. V. Afonine, N. W. Moriarty, B. K. Poon, O. V. Sobolev, T. C. Terwilliger, and P. D. Adams. 2017. 'Polder maps: improving OMIT maps by excluding bulk solvent', *Acta Crystallogr D Struct Biol*, 73: 148-57.

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, Glen, and Ronen Marmorstein. 2013. 'Implications for the evolution of eukaryotic amino-terminal acetyltransferase (NAT) enzymes from the structure of an archaeal ortholog', *Proceedings of the National Academy of Sciences*, 110: 14652-57.

Lu, L., K. A. Berkey, and R. A. Casero, Jr. 1996. 'RGFGIGS is an amino acid sequence required for acetyl coenzyme A binding and activity of human spermidine/spermine N1acetyltransferase', *J Biol Chem*, 271: 18920-4.

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.

Manning, G., D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. 2002. 'The protein kinase complement of the human genome', *Science*, 298: 1912-34.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. 2007. 'Phaser crystallographic software', *J Appl Crystallogr*, 40: 658-74.

McPherson, A., and J. A. Gavira. 2014. 'Introduction to protein crystallization', *Acta Crystallogr F Struct Biol Commun*, 70: 2-20.

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.

Nilsen, T. W., and B. R. Graveley. 2010. 'Expansion of the eukaryotic proteome by alternative splicing', *Nature*, 463: 457-63.

Odorizzi, Greg, David J. Katzmann, Markus Babst, Anjon Audhya, and Scott D. Emr. 2003. 'Bro1 is an endosome-associated protein that functions in the MVB pathway in Saccharomyces cerevisiae', *Journal of Cell Science*, 116: 1893-903.

Ohgushi, M., and A. Wada. 1983. "Molten-globule state': a compact form of globular proteins with mobile side-chains', *FEBS Lett*, 164: 21-4.

Oscarsson, Marcus, Antonia Beteva, David Flot, Elspeth Gordon, Matias Guijarro, Gordon Leonard, ..., and Daniele de Sanctis. 2019. 'MXCuBE2: the dawn of MXCuBE Collaboration', *Journal of Synchrotron Radiation*, 26: 393-405.

Pan, Q., O. Shai, L. J. Lee, B. J. Frey, and B. J. Blencowe. 2008. 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing', *Nat Genet*, 40: 1413-5.

Pape, Thomas, and Thomas R. Schneider. 2004. 'HKL2MAP: a graphical user interface for macromolecular phasing withSHELXprograms', *Journal of Applied Crystallography*, 37: 843-44.

Persson, B., C. Flinta, G. von Heijne, and H. Jornvall. 1985. 'Structures of N-terminally acetylated proteins', *Eur J Biochem*, 152: 523-7.

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.

Piatkov, K. I., T. T. Vu, C. S. Hwang, and A. Varshavsky. 2015. 'Formyl-methionine as a degradation signal at the N-termini of bacterial proteins', *Microb Cell*, 2: 376-93.

Polevoda, B., T. Arnesen, and F. Sherman. 2009. 'A synopsis of eukaryotic Nalpha-terminal acetyltransferases: nomenclature, subunits and substrates', *BMC Proc*, 3 Suppl 6: S2.

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., J. Hoskins, and F. Sherman. 2009. 'Properties of Nat4, an N(alpha)-acetyltransferase of Saccharomyces cerevisiae that modifies N termini of histones H2A and H4', *Mol Cell Biol*, 29: 2913-24.

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.

———. 2003. 'N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins', *J Mol Biol*, 325: 595-622.

Polevoda, Bogdan, Joakim Norbeck, Hikaru Takakura, Anders Blomberg, and Fred Sherman. 1999. 'Identification and specificities of N-terminal acetyltransferases from Saccharomyces cerevisiae', *EMBO J*, 18: 6155-68.

Ramachandran, G. N., and V. Sasisekharan. 1968. 'Conformation of polypeptides and proteins', *Adv Protein Chem*, 23: 283-438.

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.

Raychaudhuri, S., P. Majumder, S. Sarkar, K. Giri, D. Mukhopadhyay, and N. P. Bhattacharyya. 2008. 'Huntingtin interacting protein HYPK is intrinsically unstructured', *Proteins*, 71: 1686-98.

Ree, Rasmus, Sylvia Varland, and Thomas Arnesen. 2018. 'Spotlight on protein N-terminal acetylation', *Experimental & Molecular Medicine*, 50: 90.

Riddles, P. W., R. L. Blakeley, and B. Zerner. 1983. 'Reassessment of Ellman's reagent', *Methods Enzymol*, 91: 49-60.

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

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.

Rubenstein, P. A., and D. J. Martin. 1983. 'NH2-terminal processing of actin in mouse L-cells in vivo', *J Biol Chem*, 258: 3961-6.

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.

Sandikci, A., F. Gloge, M. Martinez, M. P. Mayer, R. Wade, B. Bukau, and G. Kramer. 2013. 'Dynamic enzyme docking to the ribosome coordinates N-terminal processing with polypeptide folding', *Nat Struct Mol Biol*, 20: 843-50.

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.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. 'NIH Image to ImageJ: 25 years of image analysis', *Nat Methods*, 9: 671-5.

Séraphin, B. 1995. 'Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs', *EMBO J*, 14: 2089-98.

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

Sheldrick, G. M. 2010. 'Experimental phasing with SHELXC/D/E: combining chain tracing with density modification', *Acta Crystallogr D Biol Crystallogr*, 66: 479-85.

Sherman, Fred, John W. Stewart, and Susumu Tsunasawa. 1985. 'Methionine or not methionine at the beginning of a protein', *Bioessays*, 3: 27-31.

Smith, L. M., N. L. Kelleher, and Proteomics Consortium for Top Down. 2013. 'Proteoform: a single term describing protein complexity', *Nat Methods*, 10: 186-7.

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.

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., 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.

Tanka, Seiji, Yasuhiko Matsushita, Akikazu Yoshikawa, and Katsumi Isono. 1989. 'Cloning and molecular characterization of the generimL wich encodes an enzyme acetylating ribosomal protein L12 ofEscherichia coki K12', *Molecular and General Genetics MGG*, 217: 289-93.

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', *J Bacteriol*, 175: 3192-4.

Tercero, J. C., L. E. Riles, and R. B. Wickner. 1992. 'Localized mutagenesis and evidence for post-transcriptional regulation of MAK3. A putative N-acetyltransferase required for double-stranded RNA virus propagation in Saccharomyces cerevisiae', *J Biol Chem*, 267: 20270-6.

Tercero, J. C., and R. B. Wickner. 1992. 'MAK3 encodes an N-acetyltransferase whose modification of the L-A gag NH2 terminus is necessary for virus particle assembly', *J Biol Chem*, 267: 20277-81.

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

Tress, M. L., F. Abascal, and A. Valencia. 2017. 'Alternative Splicing May Not Be the Key to Proteome Complexity', *Trends Biochem Sci*, 42: 98-110.

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

Udgaonkar, J. B. 2013. 'Polypeptide chain collapse and protein folding', *Arch Biochem Biophys*, 531: 24-33.

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. 2016. 'A Role for Human N-alpha Acetyltransferase 30 (Naa30) in Maintaining Mitochondrial Integrity', *Mol Cell Proteomics*, 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. 2011. '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-M110.80.

Van Duyne, Gregory D, Robert F Standaert, P Andrew Karplus, Stuart L Schreiber, and Jon Clardy. 1993. 'Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin', *J Mol Biol*, 229: 105-24.

Vetting, M. W., S. de Carvalho LP, M. Yu, S. S. Hegde, S. Magnet, S. L. Roderick, and J. S. Blanchard. 2005. 'Structure and functions of the GNAT superfamily of acetyltransferases', *Arch Biochem Biophys*, 433: 212-26.

Walden, Helen. 2010. 'Selenium incorporation using recombinant techniques', *Acta Crystallographica Section D: Biological Crystallography*, 66: 352-57.

Walsh, C. T., S. Garneau-Tsodikova, and G. J. Gatto, Jr. 2005. 'Protein posttranslational modifications: the chemistry of proteome diversifications', *Angew Chem Int Ed Engl*, 44: 7342-72.

Wan, Ruixue, Chuangye Yan, Rui Bai, Lin Wang, Min Huang, Catherine C. L. Wong, and Yigong Shi. 2016. 'The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis', *Science*, 351: 466-75.

Wang, E. T., R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S. F. Kingsmore, G. P. Schroth, and C. B. Burge. 2008. 'Alternative isoform regulation in human tissue transcriptomes', *Nature*, 456: 470-6.

Wang, Z. A., and W. R. Liu. 2017. 'Proteins with Site-Specific Lysine Methylation', *Chemistry*, 23: 11732-37.

Waudby, C. A., C. M. Dobson, and J. Christodoulou. 2019. 'Nature and Regulation of Protein Folding on the Ribosome', *Trends Biochem Sci*.

Weatheritt, R. J., T. Sterne-Weiler, and B. J. Blencowe. 2016. 'The ribosome-engaged landscape of alternative splicing', *Nat Struct Mol Biol*, 23: 1117-23.

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.

Wickner, R. B., T. Fujimura, and R. Esteban. 2013. 'Viruses and prions of Saccharomyces cerevisiae', *Adv Virus Res*, 86: 1-36.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, and K. S. Wilson. 2011. 'Overview of the CCP4 suite and current developments', *Acta Crystallogr D Biol Crystallogr*, 67: 235-42.

Wolf, E., A. Vassilev, Y. Makino, A. Sali, Y. Nakatani, and S. K. Burley. 1998. 'Crystal structure of a GCN5-related N-acetyltransferase: Serratia marcescens aminoglycoside 3-N-acetyltransferase', *Cell*, 94: 439-49.

Yang, X., J. Coulombe-Huntington, S. Kang, G. M. Sheynkman, T. Hao, A. Richardson, S. Sun, F. Yang, Y. A. Shen, R. R. Murray, K. Spirohn, B. E. Begg, M. Duran-Frigola, A. MacWilliams, S. J. Pevzner, Q. Zhong, S. A. Trigg, S. Tam, L. Ghamsari, N. Sahni, S. Yi, M. D. Rodriguez, D. Balcha, G. Tan, M. Costanzo, B. Andrews, C. Boone, X. J. Zhou, K. Salehi-Ashtiani, B. Charloteaux, A. A. Chen, M. A. Calderwood, P. Aloy, F. P. Roth, D. E. Hill, L. M. Iakoucheva, Y. Xia, and M. Vidal. 2016. 'Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing', *Cell*, 164: 805-17.

Yoshikawa, Akikazu, Setsuko Isono, Abraham Sheback, and Katsumi Isono. 1987. 'Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12', *Molecular and General Genetics MGG*, 209: 481-88.

Zeytuni, N., and R. Zarivach. 2012. 'Structural and functional discussion of the tetra-trico-peptide repeat, a protein interaction module', *Structure*, 20: 397-405.

Zhai, Q., M. B. Landesman, H. Robinson, W. I. Sundquist, and C. P. Hill. 2011. 'Structure of the Bro1 domain protein BROX and functional analyses of the ALIX Bro1 domain in HIV-1 budding', *PLoS One*, 6: e27466.

Zhai, Qianting, Robert D. Fisher, Hyo-Young Chung, David G. Myszka, Wesley I. Sundquist, and Christopher P. Hill. 2007. 'Structural and functional studies of ALIX interactions with YPXnL late domains of HIV-1 and EIAV', *Nat Struct Mol Biol*, 15: 43-49.

Zhou, Lijun, Jing Hang, Yulin Zhou, Ruixue Wan, Guifeng Lu, Ping Yin, Chuangye Yan, and Yigong Shi. 2013. 'Crystal structures of the Lsm complex bound to the 3′ end sequence of U6 small nuclear RNA', *Nature*, 506: 116.

10 Supplement

10.1 *List of figures*

10.2 *List of tables*

10.3 *Sequence alignments of Naa30, Naa35 and Naa38 orthologs*

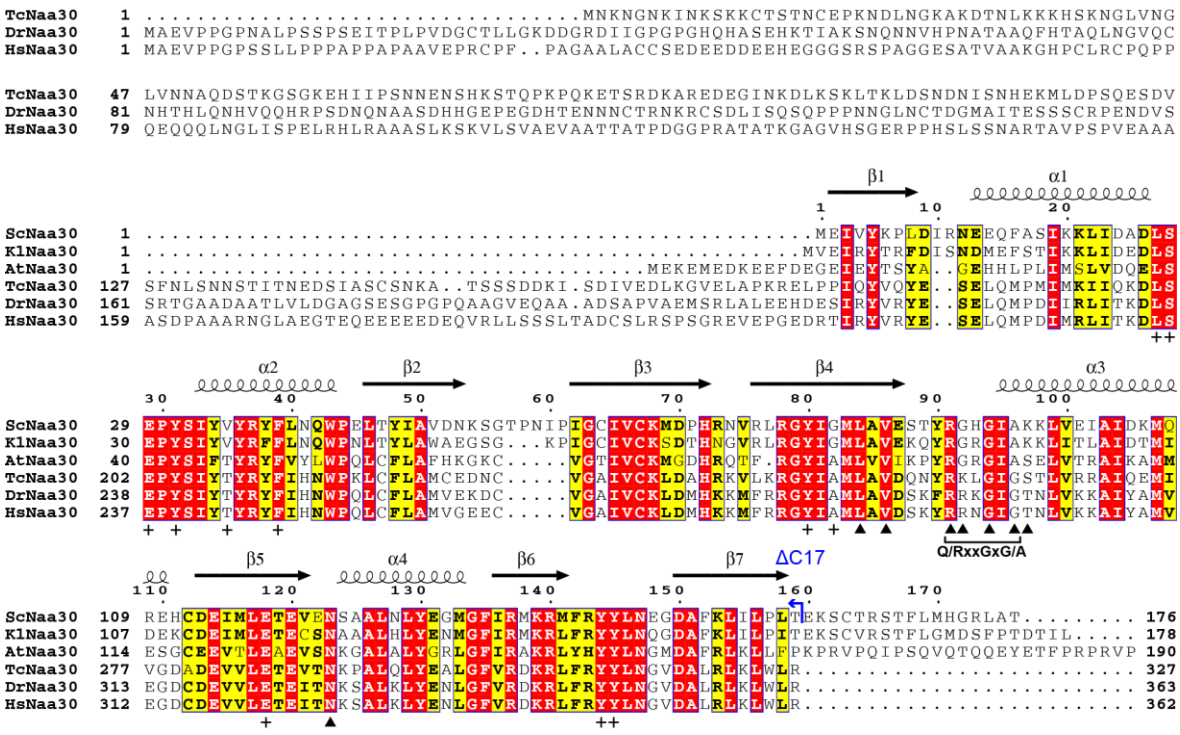


Fig. S1 Sequence alignment of Naa30 orthologs. Sc, *Saccharomyces cerevisiae*; Kl, *Kluyveromyces lactis*; At, *Arabidopsis thaliana*; Tc, *Tribolium castaneum*; Dr, *Danio rerio* and Hs, *Homo sapiens*. Completely conserved residues are highlighted with red boxes, highly conserved residues with yellow boxes. Secondary structures assignments of ScNaa30 are shown on top. The conserved coenzyme A binding motif RxxGxG/A is annotated below.▲, CoA binding residues and +, Peptide binding residues, identified with the *PISA* server (https://www.ebi.ac.uk/pdbe/pisa/). The construct boundary for Naa30ΔC17 is shown in blue (ΔC17).

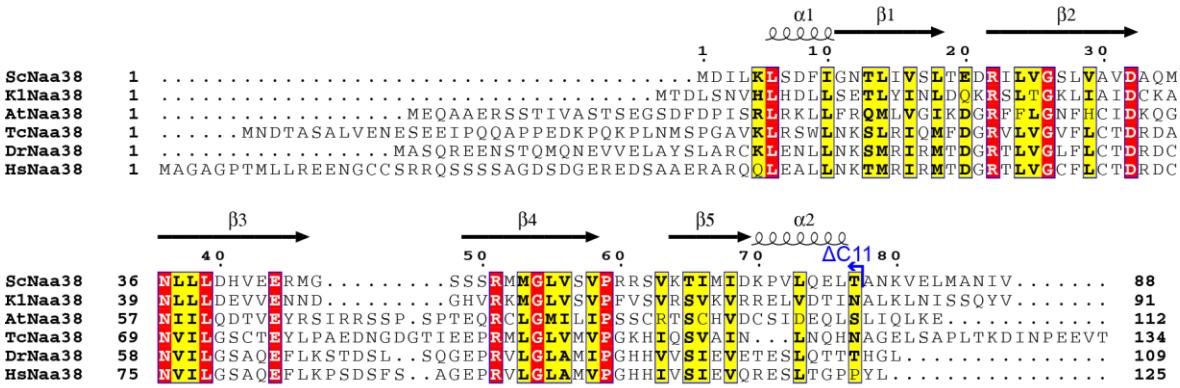


Fig. S2 Sequence alignment of Naa38 orthologs. Sc, *Saccharomyces cerevisiae*; Kl, *Kluyveromyces lactis*; At, *Arabidopsis thaliana*; Tc, *Tribolium castaneum*; Dr, *Danio rerio* and Hs, *Homo sapiens*. Global sequence alignment with free end gaps (Blosum45) made with *Geneious* (https://www.geneious.com/) and visualized using *ESPript3.0* (http://espript.ibcp.fr/). Completely conserved residues are highlighted with red boxes, highly conserved residues with yellow boxes. Secondary structures assignments of ScNaa38 are shown on top. The construct boundary for the Naa38 subunit in the crystallized NatC Complex is shown in blue (ΔC11).

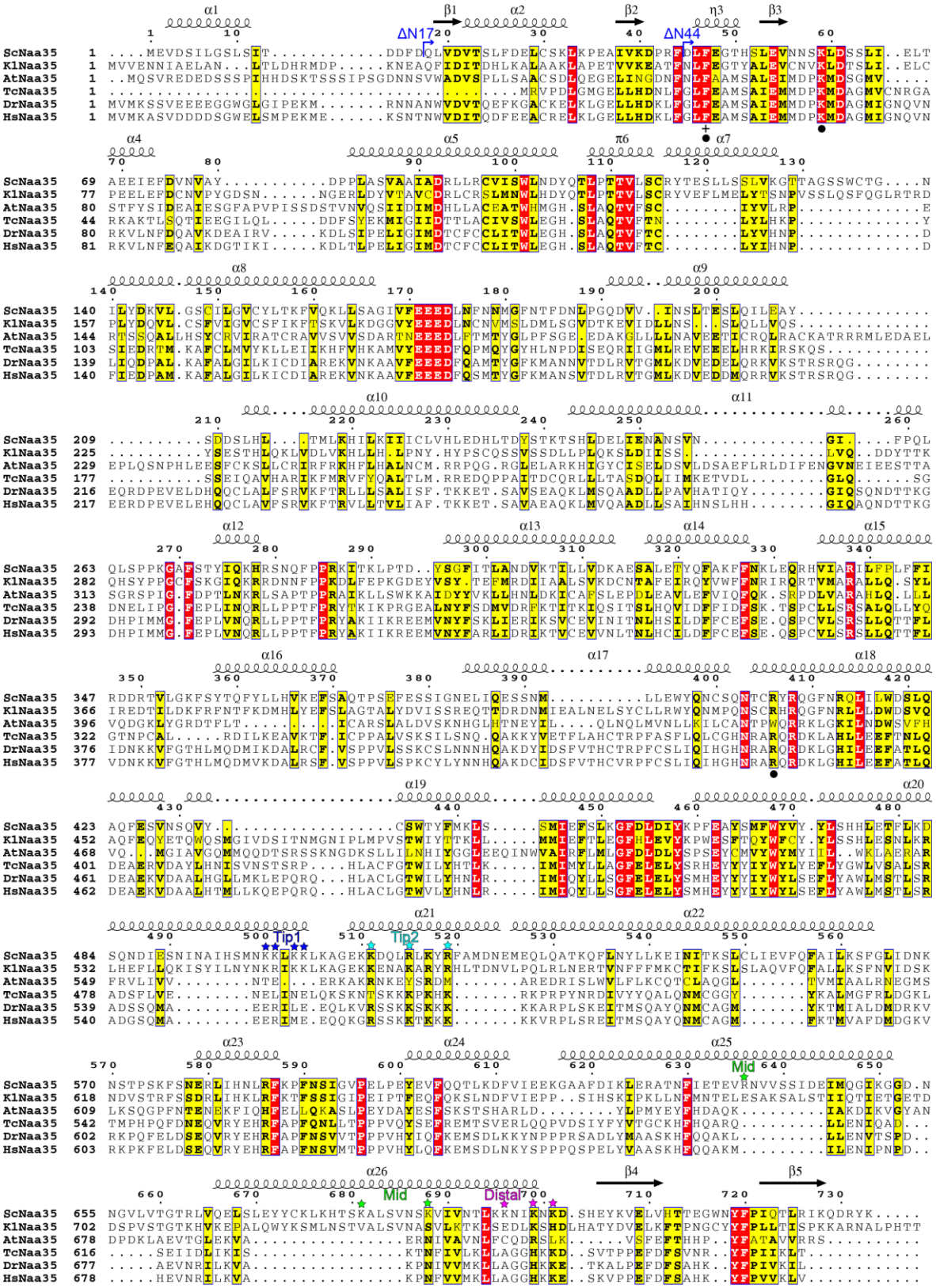


Fig. S3 Sequence alignment of Naa35 orthologs. Selected species and alignment method are identical to the Naa38 alignment (Fig. S2). Two construct boundaries, that are described in section 4.4, are shown in blue (ΔN17 and ΔN44). +, substrate binding residues and ●, mutationally sensitive residues are annotated below the sequence alignment. Selected residues (★) in positively charged patches (Tip1, Tip2, Mid and Distal) were mutated to alanine in the ribosome sedimentation assay (section 4.12).

10.4 *Phylogenetic tree of NatC orthologs*

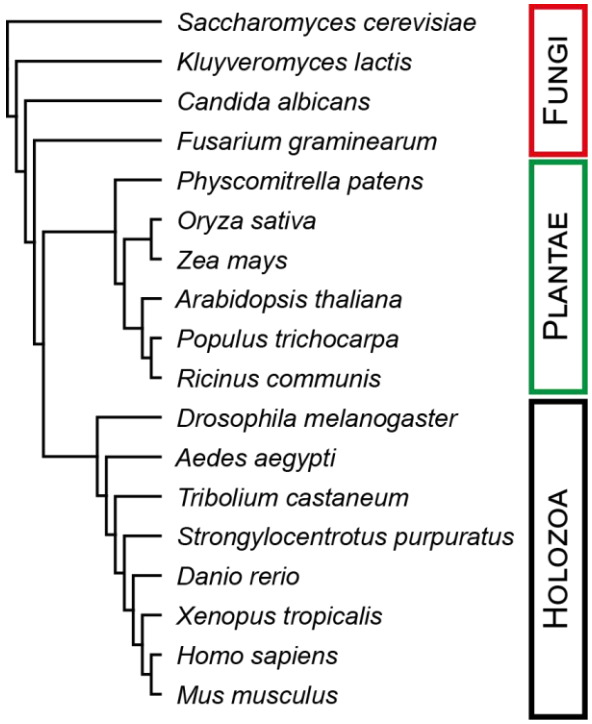


Fig. S4 Phylogenetic tree of NatC orthologs. Shown are the 18 different species for which orthologs of all three NatC subunits (Naa30, Naa35 and Naa38) could be identified. The tree was generated for Naa35 amino acid sequences with the *Geneious* software, using a BLOSUM45 cost matrix. Orthologs were used to generate Multiple sequence alignments in order to calculate *ConSurf* conservation scores.

10.5 *Unit cell content*

The *NatC-apo* and *NatC-CoA*, *NatC-CoA-MFHLV* and *NatC-CoA-MLRFV* structures crystallized in $P2_12_12_1$ with very similar unit cell dimensions, containing one NatC complex per asymmetric unit. Four asymmetric units are present per unit cell ($Z = 4$), which are generated through general position operators.[47] Two neighboring unit cells, with four NatC symmetry mates each, are shown in Fig. S5.

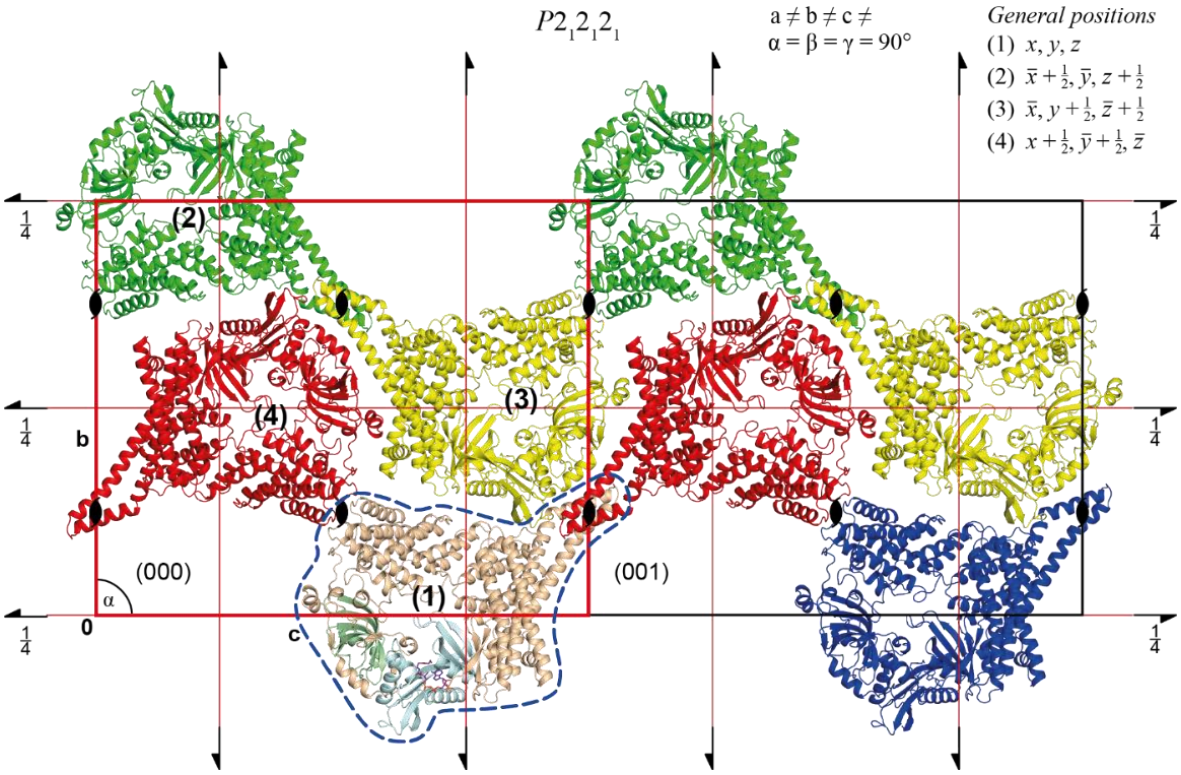


Fig. S5 Two neighboring unit cells of the NatC complex in $P2_12_12_1$. Orthoscopic projection along the a-axis, showing unit cells (000) and (001) of the *NatC-CoA* structure. Each of the four asymmetric units contains one copy of the *NatC-CoA* complex (blue outline). The four NatC copies (1)-(4) in each unit-cell are symmetry-mates, which can be generated by applying the translation and rotation operations, listed under *General positions*. Twofold screw axes are shown either parallel to the plain of projection (—◗) or pointing in the plane of projection (◖).

The structure of *NatC_Naa30_L27A* crystallized in *P*1 and its unit cell occupied approximately half the volume compared to the $P2_12_12_1$ crystals. The asymmetric unit of *NatC_Naa30_L27A* contains two independent NatC copies (biological assemblies #1 and #2). The root-mean-square deviation of the Cα positions between the two biological assemblies is 0.893 Å (over 950 compared Cα atoms).

To visualize the orientation of the two biological assemblies and crystal contacts of the two *NatC_Naa30_L27A* molecules, four neighboring unit cells are shown in Fig. S6.

[47] These operators can be found in the *International Tables for Crystallography*, *Volume A* (Hahn 2002)

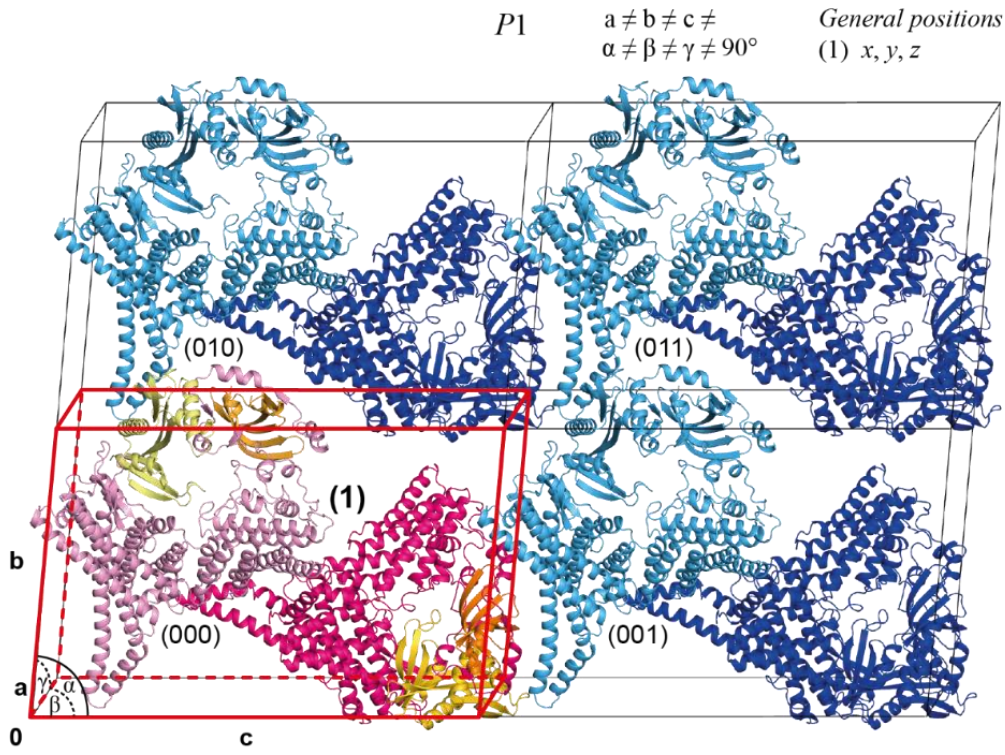


Fig. S6 Four neighboring unit cells of the NatC complex in *P*1. Orthoscopic projection along the a-axis, showing four unit cells of the *NatC_Naa30_L27A* structure. The asymmetric unit contains two different copies of the NatC complex. Biological assembly #1 (on the left side of the unit cell) is colored in pale colors. Assembly #2 (on the right) is colored in saturated colors, with subunit Naa30 in yellow, Naa35 in pink and Naa38 in orange.

10.6 *Pairwise comparison of the different NatC structures*

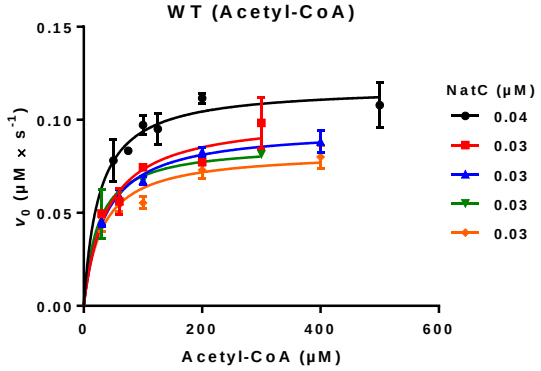
A total of six different NatC models were built — one model for each $P2_12_12_1$ dataset and two models for the *P*1 datasets, which contains two NatC complexes (biological assembly #1 and #2) in the asymmetric unit. All six NatC complexes are compared in a pairwise structural superposition of matching Cα positions (Table S 1).

Table S 1 Pairwise RMSDs of Cα positions between different NatC structures. The entire NatC complex or individual subunits of different NatC structures were superimposed (using *PyMOL's fit* command) to compare the root-mean-square-deviations (RMSDs) of their Cα atom positions (in Å). The number of matching (i.e. compared) Cα atoms is given in brackets. For identical structures (grey boxes), this number corresponds to the structure's Cα atom number. Cells are colored using a heat map gradient (with green = lowest RMSD and red = highest RMSD among all structures). The two biological assemblies (#1 and #2) of *NatC_Naa30_L27A* were analyzed separately.

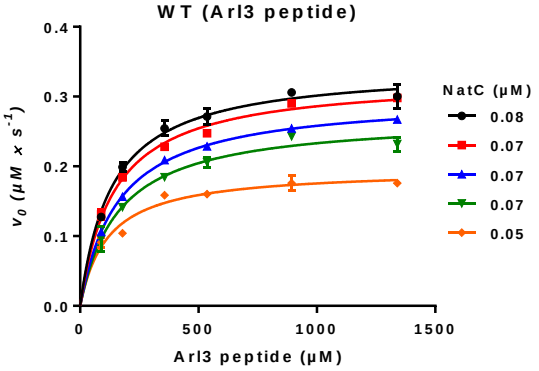
Structure	NatC-apo (SeMet)[a]	NatC -CoA	NatC-CoA -MFHLV	NatC-CoA -MLRFV	NatC_Naa30 _L27A #1	NatC_Naa30 _L27A #2
Entire NatC complex						
NatC (SeMet)	(953)					
NatC-CoA	0.231 (953)	(961)				
NatC-CoA-MFHLV	0.435 (950)	0.381 (952)	(953)			
NatC-CoA-MLRFV	0.781 (948)	0.743 (950)	0.678 (949)	(950)		
NatC_Naa30_L27A #1	0.546 (951)	0.584 (954)	0.614 (952)	0.951 (950)	(955)	
NatC_Naa30_L27A #2	0.764 (950)	0.840 (953)	0.761 (950)	1.067 (948)	0.893 (950)	(953)
Naa30 subunit only						
NatC (SeMet)	(159)					
NatC-CoA	0.218	(159)				
NatC-CoA-MFHLV	0.472	0.440	(159)			
NatC-CoA-MLRFV	0.855	0.825	0.679	(159)		
NatC_Naa30_L27A #1	0.513	0.595	0.479	0.812	(159)	
NatC_Naa30_L27A #2	0.478	0.527	0.414	0.777	0.352	(159)
Naa35 subunit only						
NatC (SeMet)	(719)					
NatC-CoA	0.205 (719)	(725)				
NatC-CoA-MFHLV	0.391 (717)	0.341 (719)	(720)			
NatC-CoA-MLRFV	0.732 (718)	0.685 (720)	0.628 (719)	(720)		
NatC_Naa30_L27A #1	0.547 (719)	0.572 (722)	0.618 (720)	0.943 (720)	(723)	
NatC_Naa30_L27A #2	0.781 (717)	0.88 (720)	0.774 (717)	1.050 (718)	0.933 (718)	(720)
Naa38 subunit only						
NatC (SeMet)	(75)					
NatC-CoA	0.307 (75)	(77)				
NatC-CoA-MFHLV	0.38 (74)	0.365 (74)	(74)			
NatC-CoA-MLRFV	0.403 (71)	0.44 (71)	0.530 (71)	(71)		
NatC_Naa30_L27A #1	0.394 (73)	0.438 (73)	0.442 (73)	0.461 (71)	(73)	
NatC_Naa30_L27A #2	0.439 (74)	0.472 (74)	0.503 (74)	0.536 (71)	0.418 (73)	(74)

10.7 *Complete Michaelis-Menten kinetics*

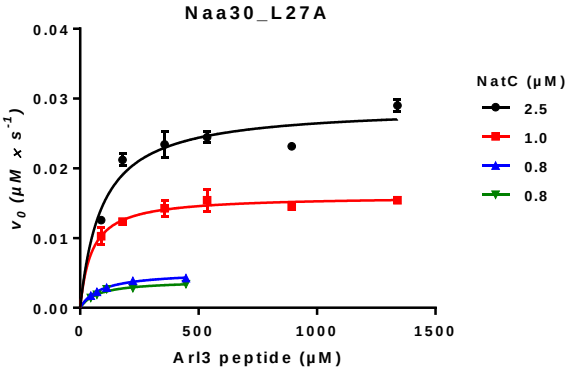
NatC (µM)	0.04	0.03	0.03	0.03	0.03
k_{cat} (s^{-1})	2.95	3.39	3.19	2.89	2.77
K_m (µM)	25.9	38.9	35.6	24.3	31.2



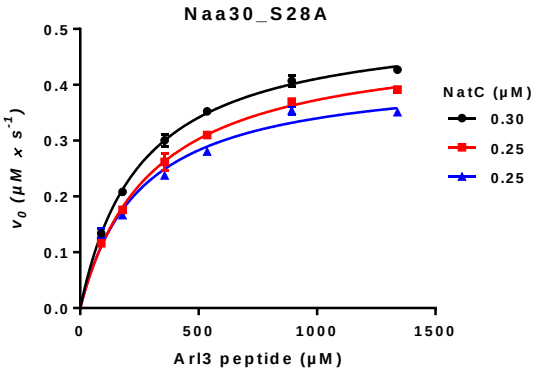
NatC (µM)	0.08	0.07	0.07	0.07	0.05
k_{cat} (s^{-1})	4.27	4.67	4.28	3.87	3.92
K_m (µM)	135	142	162	162	117



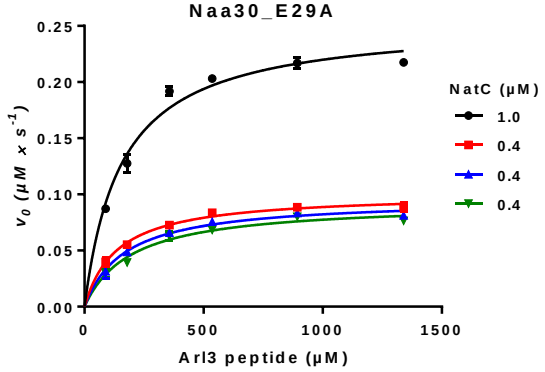
NatC (µM)	2.5	1.0	0.8	0.8
k_{cat} (s^{-1})	0.0115	0.0160	0.00646	0.00489
K_m (µM)	92.3	49.4	86.8	73.9



NatC (µM)	0.30	0.25	0.25
k_{cat} (s^{-1})	1.72	1.94	1.70
K_m (µM)	256	301	254



NatC (µM)	1.0	0.4	0.4	0.4
k_{cat} (s^{-1})	0.254	0.252	0.240	0.232
K_m (µM)	155	137	168	197



NatC (µM)	0.25	0.25	0.25
k_{cat} (s^{-1})	0.196	0.177	0.166
K_m (µM)	263	204	287

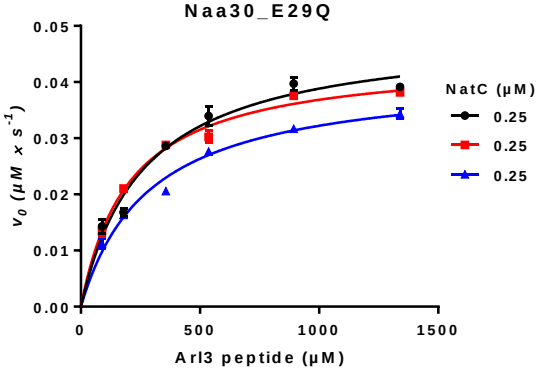
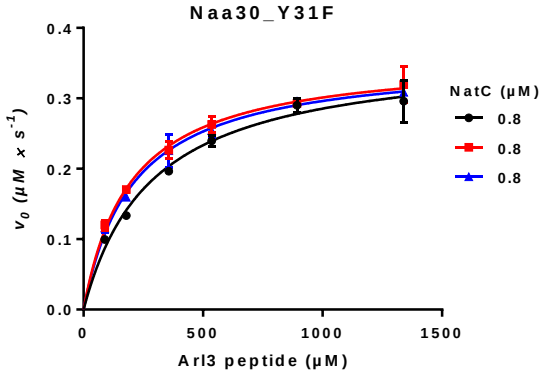
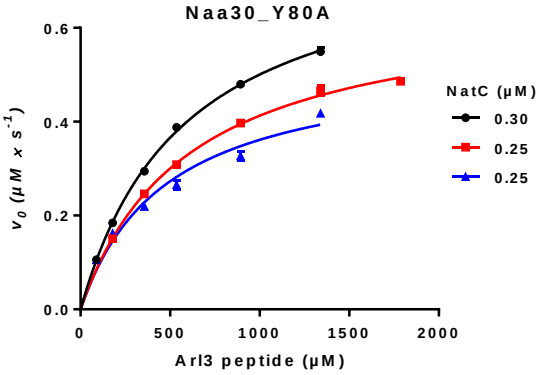


Fig. S7 Michaelis-Menten kinetics for wild type NatC and Naa30 mutants (1/3). Kinetics for either acetyl-CoA (top left – with a fixed yArl3 peptide concentration of 1340 µM) or the yArl3 peptide (top right – with 500 µM acetyl-CoA). Middle and bottom row, Michaelis-Menten kinetics for NatC mutant constructs, with single point mutations in the Naa30 subunit, using the yArl3 (MFHLVGSRRR) substrate peptide (at 500 µM acetyl-CoA). Enzyme concentrations and catalytic parameters (K_m, k_{cat}), of all replicates (n ≥ 3) are shown above each plot.

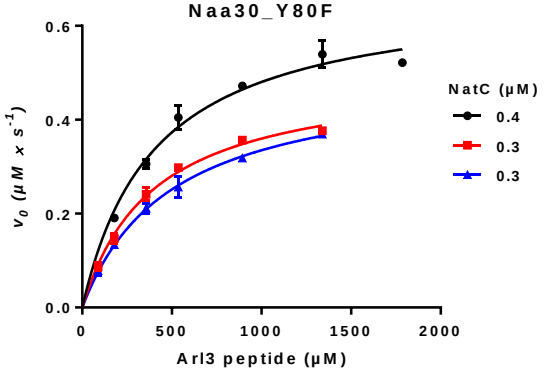
NatC (μM)	0.8	0.8	0.8
k_{cat} (s^{-1})	0.459	0.450	0.447
K_m (μM)	286	196	207



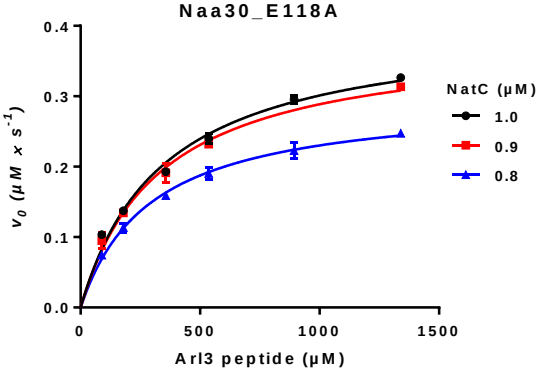
NatC (μM)	0.30	0.25	0.25
k_{cat} (s^{-1})	2.64	2.64	2.14
K_m (μM)	582	600	479



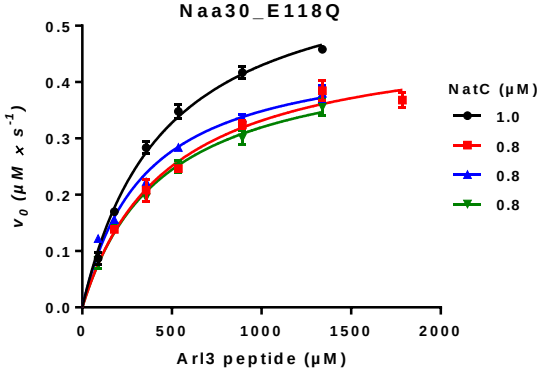
NatC (μM)	0.4	0.3	0.3
k_{cat} (s^{-1})	1.69	1.67	1.67
K_m (μM)	404	395	497



NatC (μM)	1.0	0.9	0.8
k_{cat} (s^{-1})	0.405	0.427	0.372
K_m (μM)	346	333	295



NatC (μM)	1.0	0.8	0.8	0.8
k_{cat} (s^{-1})	0.624	0.611	0.585	0.577
K_m (μM)	453	476	344	445



NatC (μM)	0.4	0.4	0.4
k_{cat} (s^{-1})	0.914	0.880	0.840
K_m (μM)	267	272	252

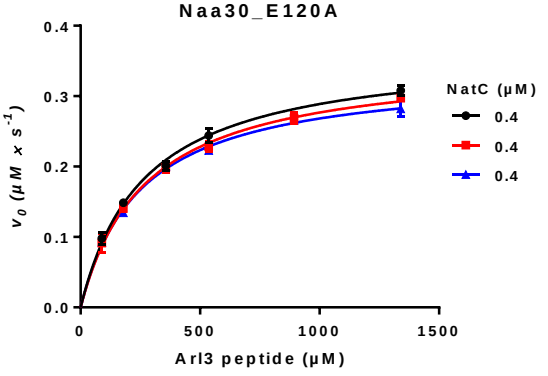


Fig. S8 Michaelis-Menten kinetics for NatC–Naa30 mutants (2/3) Single point mutations were introduced in the Naa30 subunit. Kinetics were performed with the yArl3 (MFHLVGSRRR) substrate peptide and 500 μM acetyl-CoA.

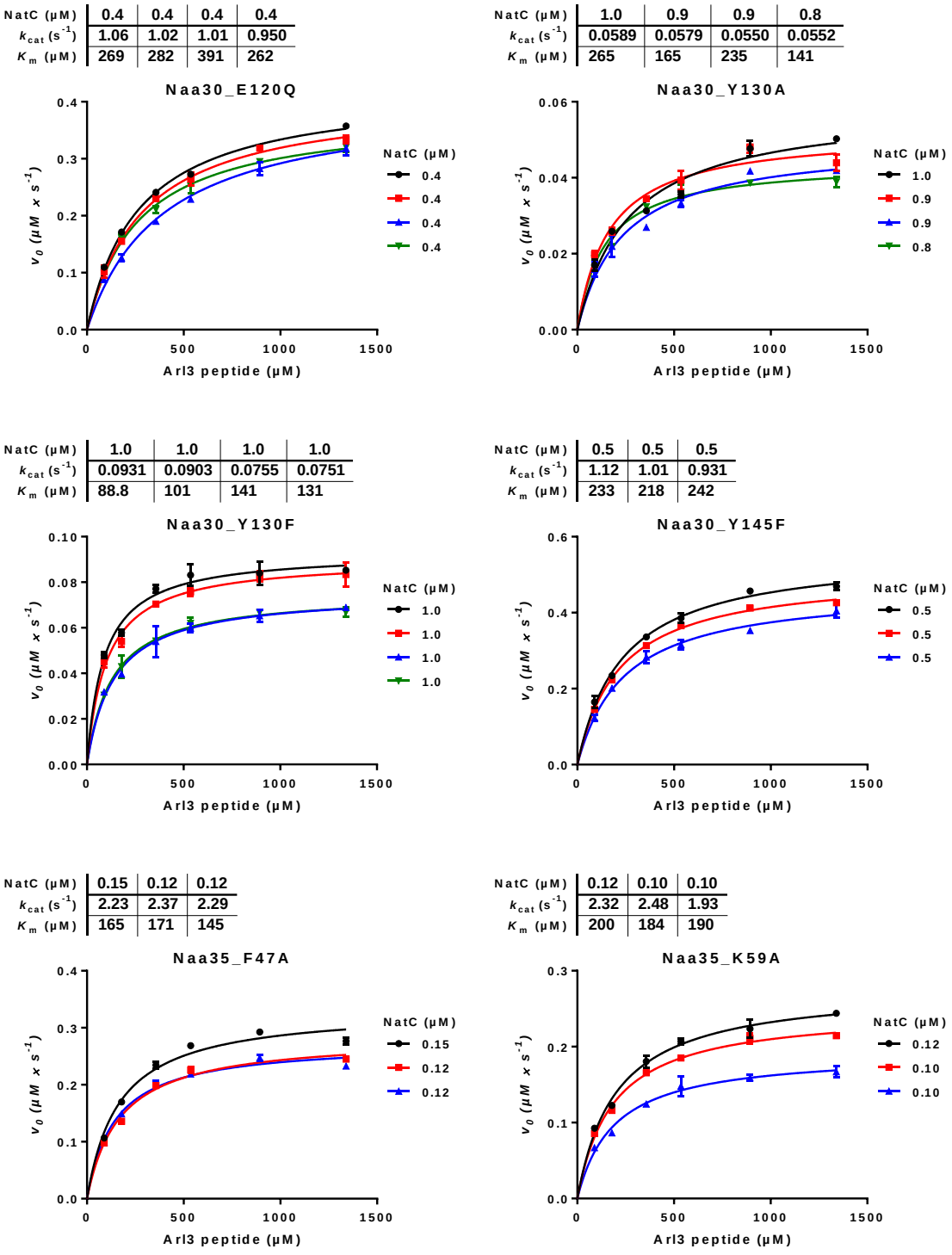


Fig. S9 Michaelis-Menten kinetics for NatC–Naa30 mutants (3/3) and NatC–Naa35 mutants (1/2) Single point mutations were introduced in the Naa30 or Naa35 subunit. Kinetics were performed with the yArl3 (MFHLVGSRRR) substrate peptide and 500 µM acetyl-CoA.

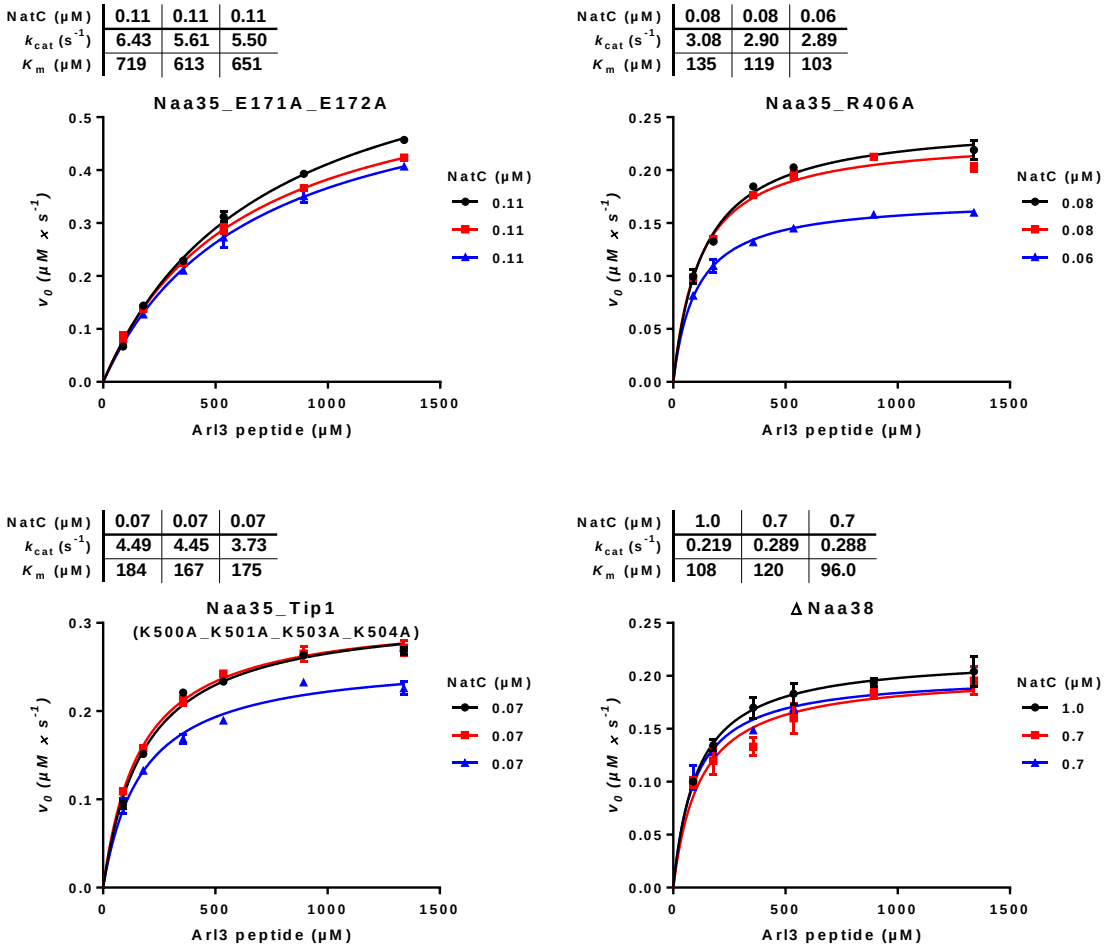
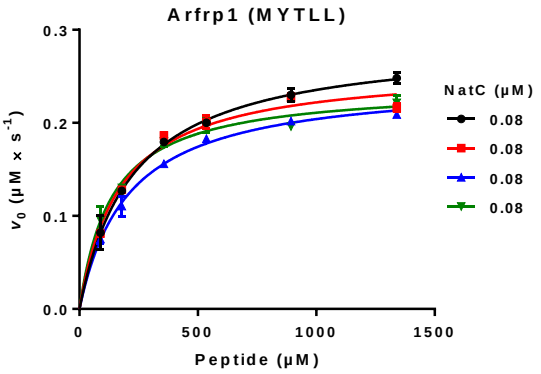
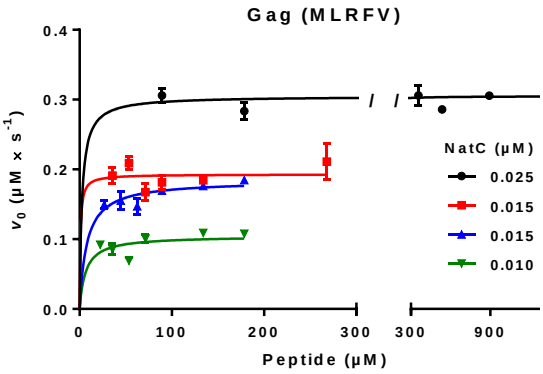


Fig. S10 Michaelis-Menten kinetics for NatC–Naa35 mutants (2/2) and NatCΔNaa38. Single point mutations were introduced in the Naa35 subunit. The *NatCΔNaa38* construct is expressed and purified without the small auxiliary subunit Naa38. Kinetics were performed with the yArl3 (MFHLVGSRRR) substrate peptide and 500 μM acetyl-CoA.

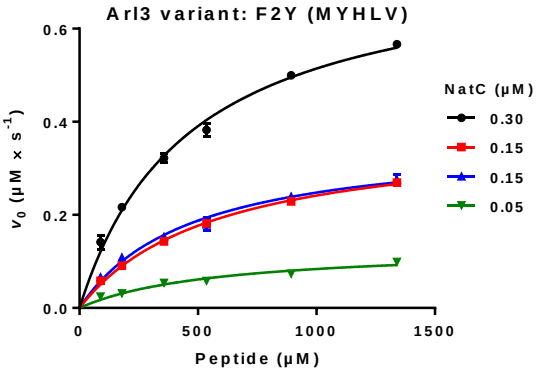
NatC (µM)	0.08	0.08	0.08	0.08
k_{cat} (s^{-1})	3.61	3.25	3.08	2.99
K_m (µM)	225	172	206	136



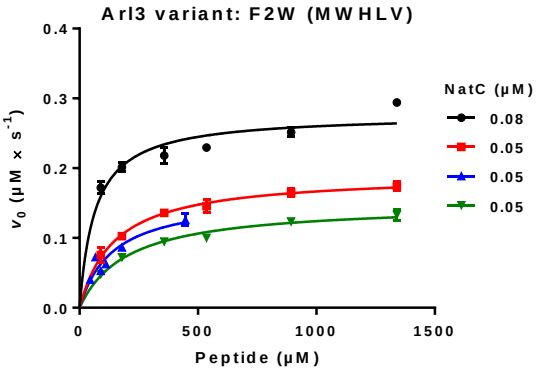
NatC (µM)	0.025	0.015	0.015	0.010
k_{cat} (s^{-1})	12.2	12.9	12.3	10.5
K_m (µM)	2.53	0.807	7.70	6.45



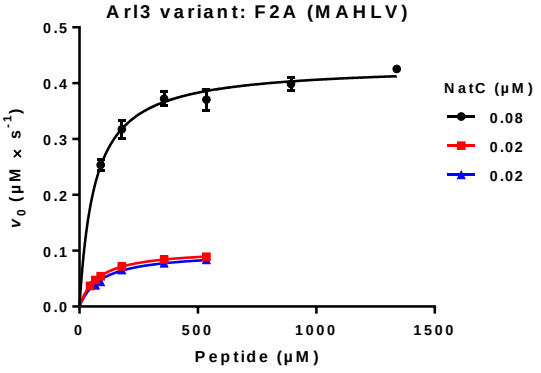
NatC (µM)	0.30	0.15	0.15	0.05
k_{cat} (s^{-1})	2.51	2.54	2.44	2.65
K_m (µM)	464	579	474	580



NatC (µM)	0.08	0.05	0.05	0.05
k_{cat} (s^{-1})	3.46	3.83	3.21	2.98
K_m (µM)	65.8	148	140	202



NatC (µM)	0.08	0.02	0.02
k_{cat} (s^{-1})	5.40	5.13	4.88
K_m (µM)	64.3	78.3	92.5



NatC (µM)	0.040	0.030	0.015
k_{cat} (s^{-1})	6.89	7.52	6.49
K_m (µM)	6.56	4.63	1.14

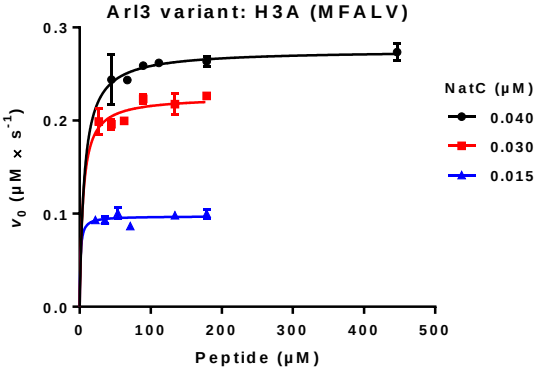


Fig. S11 Michaelis-Menten kinetics for wild type NatC, with various substrate peptides (1/2). Kinetics were performed with 500 µM acetyl-CoA. The N-terminal sequence of the corresponding substrate peptide is given in brackets above each plot.

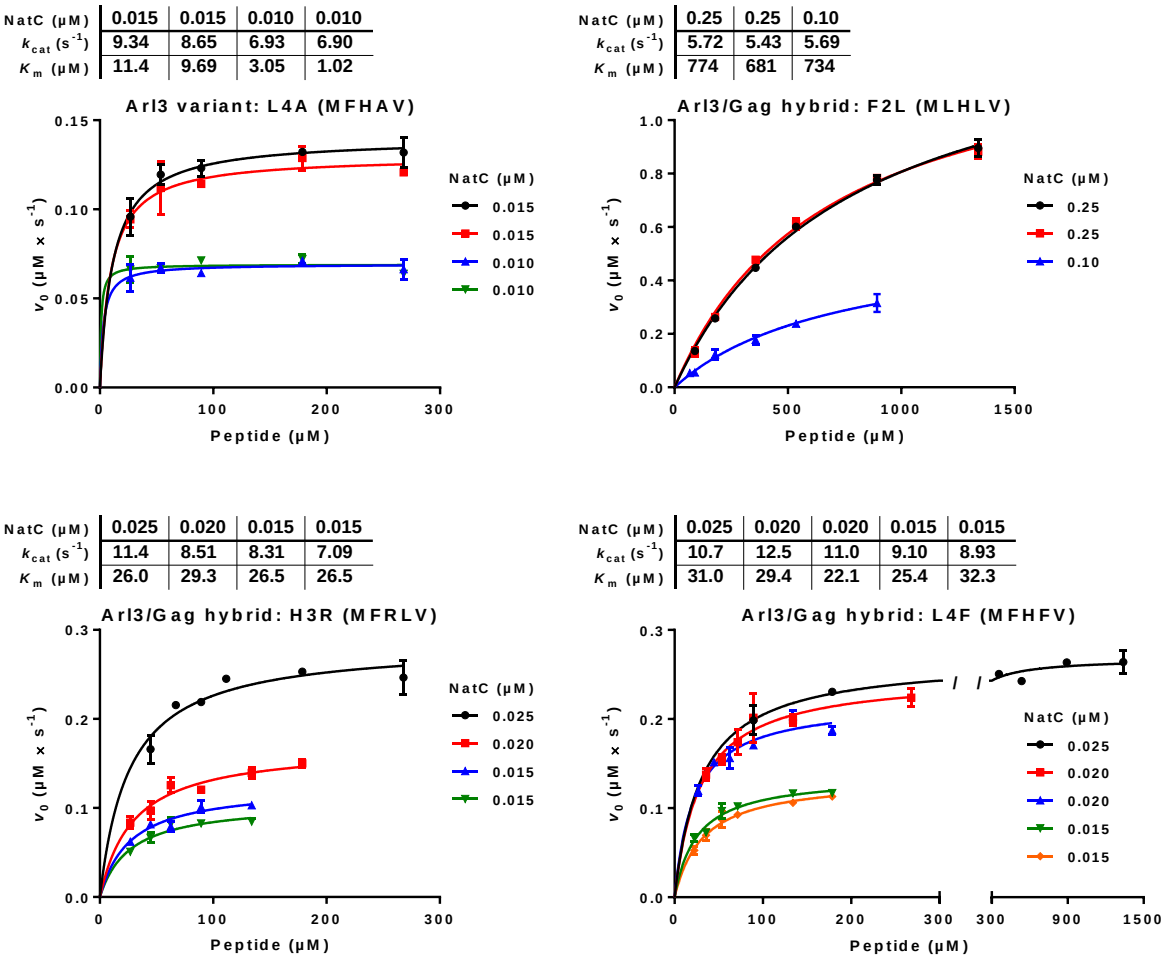


Fig. S12 Michaelis-Menten kinetics for wild type NatC, with various substrate peptides (2/2). Kinetics were performed with 500 µM acetyl-CoA. The N-terminal sequence of the corresponding substrate peptide is given in brackets above each plot

10.8 *NatC-ribosome sedimentation assay replicate*

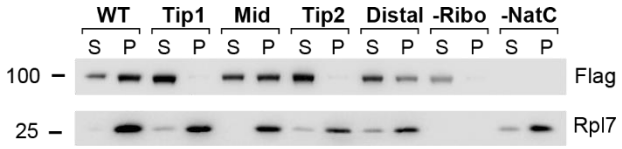


Fig. S13 NatC-ribosome sedimentation assay – Second Western blot. Western blot for the replication experiment of the sedimentation assay.

10.9 *Pathlength determination for the acetyltransferase assay*

10.9.1 Dependency of the pathlength on the well shape

The acetyltransferase assay (see methods section 3.6 and results section 4.9) employs a colorimetric read-out for the quantification of product formation. Coenzyme A thiol groups react with DTNB to produce the yellow 2-nitro-5-thiobenzoate (TNB^{2-}) anion, which absorbs at 412 nm. The TNB^{2-} concentration is determined using the Beer–Lambert law:

$$A = \varepsilon \times c \times l \qquad \text{(S1)}$$

The concentration of DTNB in solution can be directly determined from the absorbance of the dye, its molar extinction coefficient (ε= 14150 M^{-1} cm^{-1}) and the pathlength, i.e. the length of the solution that the light passes through. In a cuvette, light passes horizontally through the solution and the pathlength is defined by the cuvette's dimensions (typically 0.1–1 cm). However, in a microplate the light passes vertically and the pathlength depends on the volume of the solution to be measured. To precisely determine the concentration of an analyte with Beer–Lambert's Law, the pathlength needs to be calculated as exact as possible.

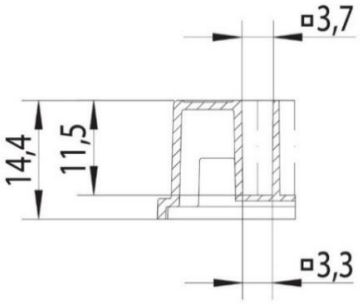


Fig. S14 Well dimensions (in millimeter) for the microplate (Cat. No. 781 162, Greiner Bio one) used in the acetyltransferase assays. The technical drawing is obtained from the Microplate Dimensions Guide.[49]

The 384-well microplate used in the acetyltransferase assay (Cat. No. 781 162, Greiner Bio one) has a rounded square well profile, i.e. the top and bottom areas of the well represent a square with rounded corners. As illustrated in Fig. S14, the well diameter is 3.7 mm at the top and 3.3 mm at the bottom and the well height is 11.5 mm. No information is given for the radii of the rounded corners. The manufacturer specifies a working volume of 10 to 130 µL. Two different specifications are given for the well volume. The information "Total volume: 131 µL" is found in the plate's datasheet[48] and "Mathematical Volume: 138 µL" is found in the Microplate Dimensions Guide[49].

In a cylindrical or cubical well, the depth or height of a liquid (i.e. its pathlength) is linearly dependent on the volume of the liquid. However, the wells of the microplate used in the acetylation reaction

[48] https://shop.gbo.com/en/germany/files/10362809/781162.pdf

[49] Microplate_Dimensions_Guide, page 7 and p. 14-15 (www.gbo.com/fileadmin/user_upload/Downloads/Brochures/Brochures_BioScience/F073027_Microplate_Dimensions_Guide.pdf)

have a conical shape and thus this relation is not valid. The shape of the well represents a frustum: the portion of a cone or pyramid, which remains after its tip has been cut off by a plane parallel to its base (shown as a transparent green volume in Fig. S15, A-C). To be precise, the well corresponds to a frustum of an (upside-down) pyramid with rounded lateral edges. It is thus neither the frustum of a pyramid or a circular cone, but the frustum of a general cone, since a cone is defined as a three-dimensional geometric shape that tapers smoothly from a flat base to a point called the apex or vertex.[50] The liquid in a frustum-shaped well will also have the form of a frustum and the liquid height corresponds to the pathlength needed to calculate the concentration of the analyte. It is possible to calculate the liquid height, with the available parameters: Top and bottom well dimension, total well height and volume of the solution.

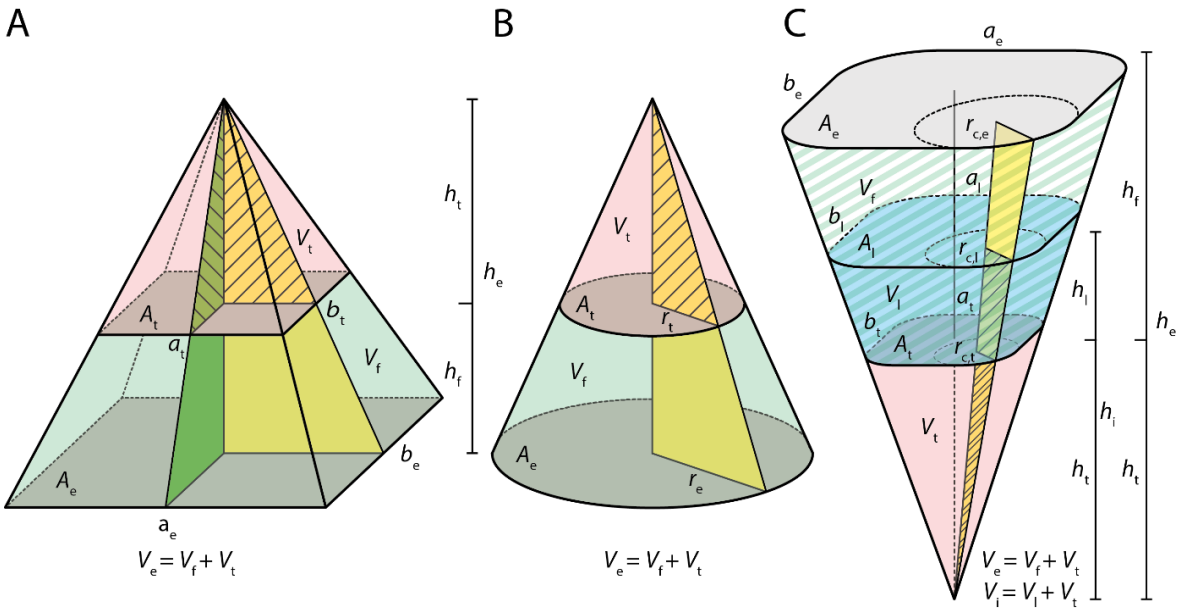


Fig. S15 Drawing of a pyramid (A), a circular cone (B) and an inverted pyramid with rounded lateral edges (C), representing an approximation of the well shape for the microplate used in the acetyltransferase assay. Base areas of the entire cone ($\boldsymbol{A_e}$) and tip ($\boldsymbol{A_t}$) are colored in grey. The edges of the square-shaped bases areas are labeled ($\boldsymbol{a_e}$, $\boldsymbol{b_e}$) and ($\boldsymbol{a_t}$, $\boldsymbol{b_t}$), respectively. The radii of the circular base areas are labeled $\boldsymbol{r_e}$ and $\boldsymbol{r_t}$. The corner radii of the rounded square bases are labeled $\boldsymbol{r_{c(e)}}$ and $\boldsymbol{r_{c(t)}}$. The heights are labeled: $\boldsymbol{h_f}$, frustum height; $\boldsymbol{h_t}$, tip height; $\boldsymbol{h_e}$, entire cone/pyramid height; $\boldsymbol{h_l}$, liquid height; $\boldsymbol{h_i}$, imaginary height. The tip volume ($\boldsymbol{V_t}$) is filled with transparent red, the volume of the frustum is filled with transparent green ($\boldsymbol{V_f}$) and together they form the entire volume ($\boldsymbol{V_e}$).The liquid volume ($\boldsymbol{V_l}$) is colored in transparent blue and together with the tip volume it forms the imaginary volume ($\boldsymbol{V_i}$). The striped triangles represent similar triangles of the yellow or green colored triangles lying below.

The names of the variables used in the calculations, are related to the geometrical objects illustrated in Fig. S15. Correspondingly, the top of a well from the 384-well plate represents the base of the entire (A_e) pyramid (which is also the base of the frustum) and the bottom of the well represents the base of the tip (A_t), which has been produced by cutting through the entire pyramid. The area of the well bottom and well top corresponds to a square with rounded corners. Since the manufacturer has not provided any information for the corner radii, a microscope photograph was taken to measure these (Fig. S16).

[50] https://en.wikipedia.org/wiki/Cone

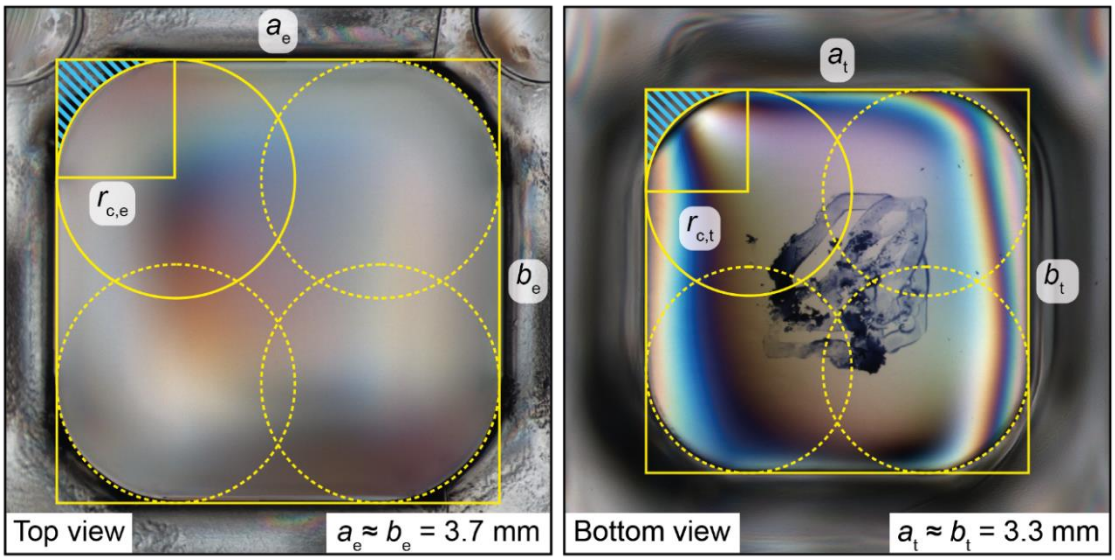


Fig. S16 A microscope photograph of one well of the microplate used in the acetyltransferase assay. The shape of the well represents a frustum of a pyramid with rounded corners. The top of the plate represents the base of the entire pyramid ($\boldsymbol{A_e}$), and the bottom of the plate represents the base of the tip ($\boldsymbol{A_t}$). The diameters of the well top ($\boldsymbol{a_e \approx b_e}$ = 3.7 mm) and well bottom ($\boldsymbol{a_t \approx b_t}$ = 3.3 mm) were used to calibrate the images in order to obtain the corner radii $\boldsymbol{r_{c,e}}$ and $\boldsymbol{r_{c,t}}$. For the well top, the radius is $\boldsymbol{r_{c,e}}$≈ 1.0 mm. The bottom radius is a bit smaller, with $\boldsymbol{r_{c(t),}}$≈ 0.9 mm. Under the assumption, that the well shape represents an upside-down pyramid with rounded corners, the radii are linearly correlated with the base length and, taking $\boldsymbol{r_{c,e}}$≈ 1.0 mm as the reference, $\boldsymbol{r_{c,t}}$ must have a length of ≈ 0.89 mm, which is in very good agreement with the measured length.

The image dimensions were calibrated using the well diameters given by the manufacturer: 3.7 mm for the well top and 3.3 mm for the well bottom. The measurements of the images have shown that the horizontal and vertical dimensions of the well top (a_e, b_e) and well bottom (a_t, b_t) differ by only 0.10 mm and 0.05 mm, respectively and thus the top and bottom well areas are approximated as square areas with side lengths a_e or a_t, respectively in further calculations. The corner radii were determined to be ≈ 1.0 mm for the top of the well (i.e. the base of the entire cone or frustum, $r_{c,e}$) and ≈ 0.9 mm for the bottom of the well (i.e. the base of the tip $r_{c,t}$). Under the assumption that the well shape represents an upside-down pyramid with rounded corners, the radii are proportional to the base lengths (equation (S2)) and, taking $r_{c,e}$≈ 1.0 mm as the reference, $r_{c,t}$ must have a length of $\frac{3.3\ \text{mm}}{3.7\ \text{mm}} \times 1.0\ \text{mm} \approx 0.89\ \text{mm}$, which is in very good agreement with the measured length.

$$\frac{a_e}{a_t} = \frac{r_{c,e}}{r_{c,t}} \qquad \text{(S2)}$$

10.9.2 The area of a rounded rectangle

As shown in Fig. S16, a rounded rectangle (in this case a rounded square) is a rectangle, that is truncated at each corner (shown as a blue-striped area) and this truncation can be described by a corner radius r_c. Each truncated corner area (A_c) can be obtained by subtracting the area of a quarter circle with radius r_c, from a square with the same side length:

$$A_c = r_c^2 - \frac{1}{4}\,\pi \times\, r_c^2 \tag{S3}$$

To obtain the area of the rounded rectangle $A_{round.rect.}$, A_c needs to be subtracted four times (once for each corner) from the area of the untruncated rectangle $A_{rect.}$ (with side length $a_{rect.}$ and $b_{rect.}$):

$$A_{round.rect.} = A_{rect.} - 4 \times A_c \tag{S4}$$

$$A_{round.rect.} = a_{rect.} \times b_{rect.} - 4 \times \left(r_c^2 - \frac{1}{4}\pi \times r_c^2\right) \tag{S5}$$

$$A_{round.rect.} = a_{rect.} \times b_{rect.} - (4\, r_c^2 - \pi \times r_c^2) \tag{S6}$$

$$A_{round.rect.} = a_{rect.} \times b_{rect.} - r_c^2(4 - \pi) \tag{S7}$$

In case of a rounded square this simplifies to:

$$\boldsymbol{A_{round.sq.} = a_{sq.}^2 - r_c^2(4 - \pi)} \tag{S8}$$

10.9.3 The volume of a frustum

Before deriving the equation for the volume (section 10.9.4) or height (section 0) of a liquid in a frustum it is a good idea to calculate the theoretical volume of the well-frustum (i.e. a completely filled well) and compare it with the manufacturer's specification of the theoretical well volume.

The volume of a cone (including the special case of a pyramid) is defined as:

$$V = \frac{1}{3} \times A \times h \tag{S9}$$

As mentioned earlier, a frustum is the portion of a cone, which remains after its tip has been cut off by a plane parallel to its base. Thus, the volume of a frustum of a cone (V_f) is simply the volume of the tip of the cone (V_t) subtracted from the volume of the entire cone (V_e):

$$V_f = V_e - V_t \tag{S10}$$

$$V_f = \frac{1}{3} \times (A_e \times h_e - A_t \times h_t) \tag{S11}$$

The height of the entire cone (h_e) corresponds to the sum of tip and frustum heights:

$$h_e = h_t + h_f \tag{S12}$$

The height of the frustum (h_f) is given and the height of the tip h_t can be readily calculated from *similar* triangles[51]. In case of the pyramid (Fig. S15, A) two similar triangles pairs can be drawn. The first pair of similar triangles lies parallel to the *a*-edges (a_e and a_t) of the two bases. The triangle that expands between the apex and the base of the entire pyramid is similar to the triangle that expands between the apex and the base of the tip volume. Perpendicular to the first pair, a second pair of similar triangles lies parallel to the *b*-edges (b_e and b_t) of the two bases. The halved lengths of the bases correspond to one site of the similar triangles and they are proportional to the heights of the pyramid tip or entire pyramid:

$$\frac{a_\mathrm{e}}{a_\mathrm{t}} = \frac{h_\mathrm{e}}{h_\mathrm{t}} \text{and} \qquad \frac{b_\mathrm{e}}{b_\mathrm{t}} = \frac{h_\mathrm{e}}{h_\mathrm{t}} \tag{S13), (S14}$$

The two equations can be combined, and the products of the base lengths can be substituted by the base areas ($A = a \times b$), of the entire pyramid and the tip:

$$\frac{a_\mathrm{e} \times b_\mathrm{e}}{a_\mathrm{t} \times b_\mathrm{t}} = \left(\frac{h_\mathrm{e}}{h_\mathrm{t}}\right)^2 \tag{S15}$$

$$\frac{A_\mathrm{e}}{A_\mathrm{t}} = \frac{h_\mathrm{e}^2}{h_\mathrm{t}^2} \tag{S16}$$

In case of a circular cone (Fig. S15, B), one pair of similar triangles can be found, with the base radii representing one side of the similar triangles and the heights representing the second side:

$$\frac{r_\mathrm{e}}{r_\mathrm{t}} = \frac{h_\mathrm{e}}{h_\mathrm{t}} \tag{S17}$$

The entire equation can be squared and the fraction on the left side of the equation can be expanded by π and the products in the numerator and denominator can be replaced by the areas of a circle ($A = \pi \times r^2$), leading to the same equation as shown for the pyramid:

$$\frac{\pi \times r_\mathrm{e}^2}{\pi \times r_\mathrm{t}^2} = \left(\frac{h_\mathrm{e}}{h_\mathrm{t}}\right)^2 \tag{S18}$$

$$\frac{A_\mathrm{e}}{A_\mathrm{t}} = \frac{h_\mathrm{e}^2}{h_\mathrm{t}^2} \tag{S19}$$

In case of a square pyramid with rounded lateral edges (Fig. S15, C) – in addition to the two similar triangles found in the pyramid – we can draw another pair of similar triangles. One side of the similar triangles corresponds to the corner radii, and the other side corresponds to the heights of the entire pyramid or pyramid tip, respectively (S20). Since the base lengths are proportional to the heights (S13), they are also proportional to the corner radii (S21).

$$\frac{h_\mathrm{e}}{h_\mathrm{t}} = \frac{r_\mathrm{c,e}}{r_\mathrm{c,t}} \text{and} \qquad \frac{a_\mathrm{e}}{a_\mathrm{t}} = \frac{r_\mathrm{c,e}}{r_\mathrm{c,t}} \tag{S20), (S21}$$

[51] In geometry, two triangles $\triangle$ ABC and $\triangle$ A′B′C′ are similar ($\triangle$ ABC $\sim$ $\triangle$ A′B′C′) if they have congruent (i.e. identical) angles. That is $\angle$ABC = $\angle$A′B′C′ and $\angle$BAC = $\angle$B′A′C′ and ACB = $\angle$A′C′B′. The lengths of corresponding sides of similar triangles are proportional.

It was shown for both the pyramid and the circular cone, that the base areas are proportional to the squared heights. Thus, one can postulate that the same relation is true for a pyramid with rounded lateral edges:

$$\frac{A_\mathrm{e}}{A_\mathrm{t}} = \frac{h_\mathrm{e}^2}{h_\mathrm{t}^2} \tag{S22}$$

In order to prove the above, the rounded square area (S8) of the entire pyramid (A_e) can be divided by the rounded square area of the tip (A_t),

$$\frac{A_\mathrm{e}}{A_\mathrm{t}} = \frac{a_\mathrm{e}^2 - r_\mathrm{c,e}^2(4-\pi)}{a_\mathrm{t}^2 - r_\mathrm{c,t}^2(4-\pi)} \tag{S23}$$

and expanded by h_e^2 and h_t^2:

$$\frac{A_e}{A_t} = \frac{h_e^2}{h_t^2} \times \frac{\frac{a_e^2}{h_e^2} - \frac{r_{c,e}^2}{h_e^2}(4-\pi)}{\frac{a_t^2}{h_t^2} - \frac{r_{c,t}^2}{h_t^2}(4-\pi)} \tag{S24}$$

Eq. (S22) is true, if the second term on the right-hand side of eq. (S24) is one:

$$\frac{\frac{a_e^2}{h_e^2} - \frac{r_{c,e}^2}{h_e^2}(4-\pi)}{\frac{a_t^2}{h_t^2} - \frac{r_{c,t}^2}{h_t^2}(4-\pi)} = 1 \tag{S25}$$

Thus, the numerator and denominator must be equal:

$$\frac{a_e^2}{h_e^2} - \frac{r_{c,e}^2}{h_e^2}(4-\pi) = \frac{a_t^2}{h_t^2} - \frac{r_{c,t}^2}{h_t^2}(4-\pi) \tag{S26}$$

a_t on the right-hand side of the equation can be substituted using eq. (S21),

$$\frac{a_e^2 - r_{c,e}^2}{h_e^2}(4-\pi) = \frac{a_e^2 \times r_{c,t}^2}{h_t^2 \times r_{c,e}^2} - \frac{r_{c,t}^2}{h_t^2}(4-\pi) \tag{S27}$$

and the last term on the right side can be expanded by $\mathrm{r_{c,e}^2}$:

$$\frac{a_e^2 - r_{c,e}^2}{h_e^2}(4-\pi) = \frac{a_e^2 \times r_{c,t}^2}{h_t^2 \times r_{c,e}^2} - \frac{r_{c,t}^2 \times r_{c,e}^2}{h_t^2 \times r_{c,e}^2}(4-\pi) \tag{S28}$$

$$\frac{a_e^2 - r_{c,e}^2}{h_e^2}(4-\pi) = \frac{a_e^2 \times r_{c,t}^2 - r_{c,t}^2 \times r_{c,e}^2}{h_t^2 \times r_{c,e}^2}(4-\pi) \tag{S29}$$

Finally h_e^2 on the left side of the equation can be replaced using eq. (S20):

$$\frac{r_{c,t}^2 \times \left(a_e^2 - r_{c,e}^2\right)}{h_t^2 \times r_{c,e}^2}(4-\pi) = \frac{r_{c,t}^2 \times \left(a_e^2 - r_{c,e}^2\right)}{h_t^2 \times r_{c,e}^2}(4-\pi) \tag{S30}$$

The left- and right-hand side of the above equation are equal and thus eq. (S22) is true.

It could be shown that the base areas are proportional to the squared heights for the geometrical shape of a pyramid (S16), as well as a cone (S19) and also for a pyramid with rounded lateral edges (S22). Solving this equation for h_e gives:

$$h_\mathrm{e} = h_\mathrm{t} \times \frac{\sqrt{A_\mathrm{e}}}{\sqrt{A_\mathrm{t}}} \tag{S31}$$

, which can be used to substitute h_e in eq. (S11):

$$V_f = \frac{1}{3} \times \left(A_e \times h_t \times \frac{\sqrt{A_e}}{\sqrt{A_t}} - A_t \times h_t \right) \tag{S32}$$

$$V_f = \frac{1}{3} \times h_t \times \left(A_e \times \frac{\sqrt{A_e}}{\sqrt{A_t}} - A_t \right) \tag{S33}$$

$$V_f = \frac{1}{3} \times h_t \times \left(\frac{A_e \times \sqrt{A_e} - A_t \times \sqrt{A_t}}{\sqrt{A_t}} \right) \tag{S34}$$

h_e in equation (S31) can be substituted by eq. (S12) and solved for h_t:

$$h_t + h_f = h_t \times \frac{\sqrt{A_e}}{\sqrt{A_t}} \tag{S35}$$

$$h_f = h_t \times \frac{\sqrt{A_e}}{\sqrt{A_t}} - h_t \tag{S36}$$

$$h_f = h_t \times \left(\frac{\sqrt{A_e}}{\sqrt{A_t}} - 1 \right) \tag{S37}$$

$$h_f = h_t \times \left(\frac{\sqrt{A_e} - \sqrt{A_t}}{\sqrt{A_t}} \right) \tag{S38}$$

$$h_t = h_f \frac{\sqrt{A_t}}{\sqrt{A_e} - \sqrt{A_t}} \tag{S39}$$

Substituting h_t in eq. (S34) with (S39) gives the **equation for the volume of a general frustum (S43)**:

$$V_f = \frac{1}{3} \times h_f \left(\frac{\sqrt{A_t}}{\sqrt{A_e} - \sqrt{A_t}} \right) \times \left(\frac{A_e \times \sqrt{A_e} - A_t \times \sqrt{A_t}}{\sqrt{A_t}} \right) \tag{S40}$$

$$V_f = \frac{1}{3} \times h_f \left(\frac{A_e \times \sqrt{A_e} - A_t \times \sqrt{A_t}}{\sqrt{A_e} - \sqrt{A_t}} \right) \tag{S41}$$

$$V_f = \frac{1}{3} \times h_f \left(\frac{\left(\sqrt{A_e} - \sqrt{A_t}\right) \times \left(A_e + \sqrt{A_e} \times \sqrt{A_t} + A_t\right)}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} \right) \tag{S42}$$

$$\boldsymbol{V_f = \frac{1}{3} \times h_f \times \left(A_e + A_t + \sqrt{A_e \times A_t}\right)} \tag{S43}$$

Substituting eq. (S43) with the area of a square gives the **equation for the volume of a frustum of a pyramid with a square base (S45)**:

$$V_{f,pyramid} = \frac{1}{3} \times h_f \times \left(a_e^2 + a_t^2 + \sqrt{a_e^2 \times a_t^2} \right) \tag{S44}$$

$$V_{f,pyramid} = \frac{1}{3} \times h_f \times (a_e^2 + a_t^2 + a_e \times a_t) \tag{S45}$$

Substituting eq. (S43) with the area of a circle, gives the **equation for the volume of a frustum of a circular cone (S48) or (S49)**:

$$V_{f,cone} = \frac{1}{3} \times h_f \times \left(\pi r_e^2 + \pi r_t^2 + \sqrt{\pi r_e^2 \times \pi r_t^2} \right) \quad \text{(S46)}$$

$$V_{f,cone} = \frac{1}{3} \times h_f \times (\pi r_e^2 + \pi r_t^2 + \pi \times r_e \times r_t) \quad \text{(S47)}$$

$$\boldsymbol{V_{f,cone} = \frac{\pi}{3} \times h_f \times (r_e^2 + r_t^2 + r_e \times r_t)} \quad \text{(S48)}$$

or the alternate form of the equation, using diameters instead of radii:

$$\boldsymbol{V_{f,cone} = \frac{\pi}{12} \times h_f \times (d_e^2 + d_t^2 + d_e \times d_t)} \quad \text{(S49)}$$

10.9.4 Liquid volume in a frustum

The volume of a liquid in a frustum can be calculated from the liquid height (h_l**)** and the well dimensions (h_f, A_e and A_t). A liquid filled in a frustum represents another frustum itself. The top of the liquid corresponds to the base of that frustum. In order to obtain the volume of that liquid, the volume of the tip cone (shown in red) needs to be subtracted from the cone that is formed from the sum of the tip volume and the liquid volume (blue), which is here referred to as V_i, for imaginary (cone) volume:

$$V_l = V_i - V_t \quad \text{(S50)}$$

V_i and V_t can be substituted by the equation for the volume of a cone or pyramid (S9):

$$V_l = \frac{1}{3} \times (A_l \times h_i - A_t \times h_t) \quad \text{(S51)}$$

The height of the imaginary cone corresponds to the sum of the tip and liquid heights:

$$\mathrm{h_i} = \mathrm{h_l} + \mathrm{h_t} \quad \text{(S52)}$$

$\mathrm{h_t}$ can be substituted using eq. (S39):

$$h_i = h_l + h_f \frac{\sqrt{A_t}}{\sqrt{A_e} - \sqrt{A_t}} \quad \text{(S53)}$$

The last variable in equation (S51), that still needs to be addressed, is the area of the imaginary cone which corresponds to the liquid surface area and is thus termed A_l. As established before, the base area of a cone is proportional to the squared height of the cone (S22). From another pair of similar triangles, with the first extending from the base of the entire pyramid and the second extending from the base of the liquid (i.e. the upper surface), the following relation can be obtained:

$$\frac{A_l}{A_e} = \frac{h_i^2}{h_e^2} \quad \text{(S54)}$$

Taking the square root and solving for $\sqrt{A_l}$ gives:

$$\sqrt{A_l} = \sqrt{A_e} \times \frac{h_i}{h_e} \quad \text{(S55)}$$

Substituting h_i with eq. (S52) gives:

$$\sqrt{A_l} = \sqrt{A_e} \times \frac{h_l + h_t}{h_e} \tag{S56}$$

Placing eq. (S39) into (S31) gives the **equation for the height of the entire cone**:

$$h_e = h_f \frac{\sqrt{A_t}}{\sqrt{A_e} - \sqrt{A_t}} \times \frac{\sqrt{A_e}}{\sqrt{A_t}} \tag{S57}$$

$$h_e = h_f \frac{\sqrt{A_e}}{\sqrt{A_e} - \sqrt{A_t}} \tag{S58}$$

Substituting h_e in the denominator of (S56) by eq. (S58) allows $\sqrt{A_e}$ to be reduced:

$$\sqrt{A_l} = \sqrt{A_e} \times \frac{(h_l + h_t) \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f \times \sqrt{A_e}} \tag{S59}$$

$$\sqrt{A_l} = \frac{(h_l + h_t) \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f} \tag{S60}$$

Using eq. (S39) to substitute the tip height h_t and rewriting gives:

$$\sqrt{A_l} = \frac{\left(h_l + \frac{h_f \times \sqrt{A_t}}{\sqrt{A_e} - \sqrt{A_t}}\right) \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f} \tag{S61}$$

$$\sqrt{A_l} = \frac{h_l \times \left(\sqrt{A_e} - \sqrt{A_t}\right) + h_f \times \sqrt{A_t}}{h_f} \tag{S62}$$

$$\sqrt{A_l} = \sqrt{A_t} + \frac{h_l \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f} \tag{S63}$$

$$A_l = \left(\sqrt{A_t} + \frac{h_l \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f}\right)^2 \tag{S64}$$

Substituting eq. (S51) with equations (S39), (S53) and (S64) gives **the equation for the volume of a liquid in a frustum**:

$$\boldsymbol{V_l = \frac{1}{3} \times \left(\left(\sqrt{A_t} + h_l \times \frac{\left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f} \right)^2 \times \left(h_l + h_f \frac{\sqrt{A_t}}{\sqrt{A_e} - \sqrt{A_t}} \right) - h_f \times \frac{A_t \times \sqrt{A_t}}{\sqrt{A_e} - \sqrt{A_t}} \right)} \tag{S65}$$

Or an even shorter form of the equation[52]:

$$\boldsymbol{V_l = \frac{1}{3} \times \left(\left(\sqrt{A_t} + \frac{h_l \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f} \right)^2 \times \left(h_l + \frac{h_f \times \sqrt{A_t}}{\sqrt{A_e} - \sqrt{A_t}} \right) - \frac{h_f \times \sqrt{A_t^3}}{\sqrt{A_e} - \sqrt{A_t}} \right)} \tag{S66}$$

[52] With $a^m \times a^n = a^{m+n}$ and $a^{\frac{m}{n}} = \sqrt[n]{a^m}$, it can be derived that $A_t \times \sqrt{A_t} = {A_t}^1 \times {A_t}^{\frac{1}{2}} = {A_t}^{\frac{3}{2}} = \sqrt[2]{{A_t}^3}$

10.9.5 Liquid height in a frustum

In order to obtain the pathlength, which corresponds to the height of the liquid in a frustum, eq. (S66) needs to be solved for h_l. This requires expanding the equation first:

$$\boldsymbol{V_l} = \frac{1}{3} \times \left(\left(A_t + 2\sqrt{A_t} \times \frac{h_l \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f} + \left(\frac{h_l \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f} \right)^2 \right) \times \left(h_l + \frac{h_f \times \sqrt{A_t}}{\sqrt{A_e} - \sqrt{A_t}} \right) - \frac{h_f \times \sqrt{{A_t}^3}}{\sqrt{A_e} - \sqrt{A_t}} \right) \quad \text{(S67)}$$

$$\boldsymbol{V_l} = \frac{1}{3} \times \left(\left(A_t + 2h_l \times \frac{\sqrt{A_t} \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f} + h_l^2 \times \frac{\left(\sqrt{A_e} - \sqrt{A_t}\right)^2}{h_f^2} \right) \times \left(h_l + \frac{h_f \times \sqrt{A_t}}{\sqrt{A_e} - \sqrt{A_t}} \right) - \frac{h_f \times \sqrt{{A_t}^3}}{\sqrt{A_e} - \sqrt{A_t}} \right) \quad \text{(S68)}$$

$$\begin{aligned} \boldsymbol{V_l} = \frac{1}{3} \times \Bigg(& h_l A_t + 2h_l^2 \times \frac{\sqrt{A_t} \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f} + h_l^3 \times \frac{\left(\sqrt{A_e} - \sqrt{A_t}\right)^2}{h_f^2} + \frac{h_f \times \sqrt{{A_t}^3}}{\sqrt{A_e} - \sqrt{A_t}} + 2h_l A_t \\ & + h_l^2 \times \frac{\sqrt{A_t} \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f} - \frac{h_f \times \sqrt{{A_t}^3}}{\sqrt{A_e} - \sqrt{A_t}} \Bigg) \end{aligned} \quad \text{(S69)}$$

$$\boldsymbol{V_l} = \frac{1}{3} \times \left(h_l^3 \frac{\left(\sqrt{A_e} - \sqrt{A_t}\right)^2}{h_f^2} + 3h_l^2 \times \frac{\sqrt{A_t} \times \left(\sqrt{A_e} - \sqrt{A_t}\right)}{h_f} + 3h_l\, A_t \right) \quad \text{(S70)}$$

$$\boldsymbol{V_l} = \frac{1}{3} \times \frac{\left(\sqrt{A_e} - \sqrt{A_t}\right)^2}{h_f^2} \times \left(h_l^3 + 3h_l^2 \times \frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} + 3h_l \frac{h_f^2 \times A_t}{\left(\sqrt{A_e} - \sqrt{A_t}\right)^2} \right) \quad \text{(S71)}$$

$$\boldsymbol{V_l} = \frac{1}{3} \times \frac{\left(\sqrt{A_e} - \sqrt{A_t}\right)^2}{h_f^2} \times \left(h_l^3 + 3h_l^2 \times \frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} + 3h_l \left(\frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} \right)^2 \right) \quad \text{(S72)}$$

After expanding the equation, the resulting polynomial, consisting of the first four terms in the bracket, can be factored to a binomial raised to the third power[53]:

$$V_l = \frac{1}{3} \times \frac{\left(\sqrt{A_e} - \sqrt{A_t}\right)^2}{h_f^2} \times \left(h_l^3 + 3h_l^2 \times \frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} + 3h_l \left(\frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} \right)^2 + \left(\frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} \right)^3 - \left(\frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} \right)^3 \right) \quad \text{(S73)}$$

$$V_l = \frac{1}{3} \times \frac{\left(\sqrt{A_e} - \sqrt{A_t}\right)^2}{h_f^2} \times \left(\left(h_l + \frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} \right)^3 - \left(\frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} \right)^3 \right) \quad \text{(S74)}$$

Which can be solved for h_l, to obtain the **equation for the height of a liquid in a frustum (S77)**:

$$3V_l \times \frac{h_f^2}{\left(\sqrt{A_e} - \sqrt{A_t}\right)^2} = \left(h_l + \frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} \right)^3 - \left(\frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} \right)^3 \quad \text{(S75)}$$

$$3V_l \times \frac{h_f^2}{\left(\sqrt{A_e} - \sqrt{A_t}\right)^2} + \left(\frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} \right)^3 = \left(h_l + \frac{h_f \times \sqrt{A_t}}{\left(\sqrt{A_e} - \sqrt{A_t}\right)} \right)^3 \quad \text{(S76)}$$

$$\boldsymbol{h_l} = \sqrt[3]{\boldsymbol{3V_l} \times \frac{\boldsymbol{h_f^2}}{\left(\sqrt{\boldsymbol{A_e}} - \sqrt{\boldsymbol{A_t}}\right)^2} + \left(\frac{\boldsymbol{h_f} \times \sqrt{\boldsymbol{A_t}}}{\left(\sqrt{\boldsymbol{A_e}} - \sqrt{\boldsymbol{A_t}}\right)} \right)^3} - \frac{\boldsymbol{h_f} \times \sqrt{\boldsymbol{A_t}}}{\left(\sqrt{\boldsymbol{A_e}} - \sqrt{\boldsymbol{A_t}}\right)} \quad \text{(S77)}$$

[53] $(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$

10.9.6 Calculation of the well volume

Two contradictory specifications for the total well volume: 131 µL[54] and 138 µL[55], are given by the manufacturer. In the following, three different models describing the conical, frustum-shaped well are used to calculate the total well volume and compared with the manufacturer specifications, to identify the model that describes the well shape most accurately. In section 10.9.3, the equation (S43) for the volume of a general frustum (i.e. the complete well) is derived. The first model represents a frustum of a circular cone (S49). The variables are renamed in eq. (S78) to represent the shape of a well, rather than the corresponding geometrical object. With the given well dimensions ($d_{\mathrm{top}} = 3.7$ mm, $d_{\mathrm{bottom}} = 3.3$ mm and $h_{\mathrm{well}} = 11.5$ mm),

$$V_{\mathrm{f,cone}} = \frac{\pi}{12} \times \mathrm{h_{well}} \times \left(d_{\mathrm{top}}^2 + d_{\mathrm{bottom}}^2 + d_{\mathrm{top}} \times d_{\mathrm{bottom}}\right) \tag{S78}$$

$$V_{\mathrm{f,cone}} = \frac{\pi}{12} \times 11.5\ \mathrm{mm} \times ((3.7\ \mathrm{mm})^2 + (3.3\ \mathrm{mm})^2 + 3.7\ \mathrm{mm}\ \times 3.3\ \mathrm{mm}) \tag{S79}$$

$$V_{\mathrm{f,cone}} \approx 110.76\ \mathrm{mm}^3 \tag{S80}$$

the well volume is approx. 110.7 µL. This volume is much smaller than the theoretical 131 µL or 138 µL, provided by the manufacturer, because the well has a square profile and the volume of a circular cone is only $\frac{\pi}{4}$ of the volume of pyramid with base lengths equal to the cones diameter.

$$V_{\mathrm{cone}} = \frac{\pi}{4} \times V_{\mathrm{pyramid}} \tag{S81}$$

Thus, the second model represents a frustum of a square pyramid and is described by eq. (S45). Again, the variables are renamed in eq. (S82) to represent the shape of a well:

$$V_{\mathrm{f,pyramid}} = \frac{1}{3} \times h_{\mathrm{well}} \times \left(d_{\mathrm{top}}^2 + d_{\mathrm{bottom}}^2 + d_{\mathrm{top}} \times d_{\mathrm{bottom}}\right) \tag{S82}$$

$$V_{\mathrm{f,pyramid}} = \frac{1}{3} \times 11.5\ \mathrm{mm} \times ((3.7\ \mathrm{mm})^2 + (3.3\ \mathrm{mm})^2 + 3.7\ \mathrm{mm}\ \times 3.3\ \mathrm{mm}) \tag{S83}$$

$$V_{\mathrm{f,pyramid}} \approx 141.03\ \mathrm{mm}^3 \tag{S84}$$

Now, the well volume is approx. 141.0 µL. This volume is much closer to the two mentioned theoretical volumes of 131 µL or 138 µL, but still too large. This deviation can be explained by the fact, that the well has a "rounded square design", which means that the top and bottom areas of the well are truncated at all four corners. $r_{\mathrm{c,top}}$ (≈ 1.0 mm) was determined from a microscope photograph (Fig. S16) and $r_{\mathrm{c,bottom}}$ (≈ 0.89 mm) is calculated with eq. (S2). Considering the *rounded* square profile of the well, the second model can be refined to yield the third and final model for the shape of the well: a frustum of a square pyramid *with rounded lateral edges*. The total well volume can be calculated with the equation for the volume of a general frustum:

$$V_{\mathrm{f,general}} = \frac{1}{3} \times h_{\mathrm{well}} \times \left(A_{\mathrm{top}} + A_{\mathrm{bottom}} + \sqrt{A_{\mathrm{top}} \times A_{\mathrm{bottom}}}\right) \tag{S85}$$

[54] Microplate datasheet (https://shop.gbo.com/en/germany/files/10362809/781162.pdf)

[55] Microplate_Dimensions_Guide, page 7 and p. 14-15
https://www.gbo.com/fileadmin/user_upload/Downloads/Brochures/Brochures_BioScience/F073027_Microplate_Dimensions_Guide.pdf

With the rounded rectangular area of the well top (A_{top}),

$$A_{\mathrm{top}} = d_{\mathrm{top}}^2 - r_{\mathrm{c,top}}^2(4 - \pi) \tag{S86}$$

$$A_{\mathrm{top}} = (3.7\ \mathrm{mm})^2 - (1.0\ \mathrm{mm})^2(4 - \pi) \tag{S87}$$

$$A_{\mathrm{top}} = 12.83\ \mathrm{mm}^2 \tag{S88}$$

and the area of the well bottom (A_{bottom}),

$$A_{\mathrm{bottom}} = d_{\mathrm{bottom}}^2 - \left(\frac{d_{\mathrm{bottom}}}{d_{\mathrm{top}}} \times r_{\mathrm{c,top}}\right)^2 (4 - \pi) \tag{S89}$$

$$A_{\mathrm{bottom}} = (3.3\ \mathrm{mm})^2 - (\frac{3.3\ \mathrm{mm}}{3.7\ \mathrm{mm}} \times 1.0\ \mathrm{mm})^2(4 - \pi) \tag{S90}$$

$$A_{\mathrm{bottom}} = 10.21\ \mathrm{mm}^2 \tag{S91}$$

inserted into eq. (S85), we obtain:

$$V_{\mathrm{well}} = \frac{1}{3} \times 11.5\ \mathrm{mm} \times \left(12.83\ \mathrm{mm}^2 + 10.21\ \mathrm{mm}^2 + \sqrt{12.83\ \mathrm{mm}^2 \times 10.21\ \mathrm{mm}^2}\right) \tag{S92}$$

$$V_{\mathrm{well}} = 132.2\ \mu\mathrm{L} \tag{S93}$$

Under the assumption, that the well represent a frustum of a square pyramid with rounded lateral edges, the well volume is approx. 132.2 µL – a value that lies in between the two manufacturer-specified well volumes of 131 µL or 138 µL. Thus, this model provides the closest approximation for the well shape and will be used to calculate the liquid height (i.e. the pathlength) for the microplates employed in the acetyltransferase assays.

10.9.7 Calculation of the liquid height

The derivation of the equation for the height of a liquid in a frustum-shaped well (S77) is described in sections 10.9.4 and 10.9.5. The variables in the eq. (S77) are renamed in eq. (S94) to represent the shape of a well, rather than the corresponding geometrical object. Combined with the equations for the rounded square areas of the top (S86) and bottom (S89) of the well, the liquid height (i.e. the pathlength) can be calculated for a well, shaped like a frustum of a square pyramid with rounded lateral edges. Using the interim results for the rounded square areas: $A_{\mathrm{top}} = 12.83\ \mathrm{mm}^2$ and $A_{\mathrm{bottom}} = 10.21\ \mathrm{mm}^2$, the well height of $h_{\mathrm{well}} = 11.5\ \mathrm{mm}$ and an exemplary liquid volume of $V_{\mathrm{l}} = 50\ \mu\mathrm{L}$ gives:

$$h_{\mathrm{l}} = \sqrt[3]{3V_{\mathrm{l}} \times \frac{h_{\mathrm{well}}^2}{\left(\sqrt{A_{\mathrm{top}}} - \sqrt{A_{\mathrm{bottom}}}\right)^2} + \left(\frac{h_{\mathrm{well}} \times \sqrt{A_{\mathrm{bottom}}}}{\sqrt{A_{\mathrm{top}}} - \sqrt{A_{\mathrm{bottom}}}}\right)^3} - \frac{h_{\mathrm{well}} \times \sqrt{A_{\mathrm{bottom}}}}{\sqrt{A_{\mathrm{top}}} - \sqrt{A_{\mathrm{bottom}}}} \tag{S94}$$

$$h_{\mathrm{l}} = \sqrt[3]{\frac{3 \times 50\ \mathrm{mm}^3 \times (11.5\ \mathrm{mm})^2}{\left(\sqrt{12.83\ \mathrm{mm}^2} - \sqrt{10.21\ \mathrm{mm}^2}\right)^2} + \left(\frac{11.5\ \mathrm{mm} \times \sqrt{10.21\ \mathrm{mm}^2}}{\sqrt{12.83\ \mathrm{mm}^2} - \sqrt{10.21\ \mathrm{mm}^2}}\right)^3} - \frac{11.5\ \mathrm{mm} \times \sqrt{10.21\ \mathrm{mm}^2}}{\sqrt{12.83\ \mathrm{mm}^2} - \sqrt{10.21\ \mathrm{mm}^2}} \tag{S95}$$

$$h_{\mathrm{l}} = 4.67\ \mathrm{mm} \tag{S96}$$

Using the model of a pyramidal frustum with rounded lateral edges, the pathlength for 50 µL analyte solution in one well of the microplate used in the acetyltransferase assay is 4.67 mm. In comparison, the model for a simple pyramidal frustum gives a 6% shorter pathlength of only 4.39 mm (Table S 2). Instead of using the well dimension (d_{top}, d_{bottom}, h_{well}), one could also calculate the liquid height as the fraction of liquid volume divided by the total well volume (V_{total}), times the well height ($h_{\mathrm{l}} = \frac{V_{\mathrm{l}}}{V_{\mathrm{total}}} \times h_{\mathrm{well}}$). However, this model assumes that the well is a rectangular cuboid and completely neglects the conical shape of the well. Depending on which value is used for V_{total}, the calculated pathlengths for 50 µL analyte solution are 4.39 mm ($V_{\mathrm{total}} = 131$ µL) or 4.17 mm ($V_{\mathrm{total}} = 138$ µL), which is 5.9% or even 10.7% shorter, respectively, than the value obtained with the reference model of a pyramidal frustum with rounded lateral edges.

Table S 2 Calculated liquid heights (h_l) for four different well shapes and the percent height difference to the pyramidal frustum with rounded lateral edges. Both frustum liquid heights were calculated using eq. (S94), with square areas $A_{\mathrm{top}} = d_{\mathrm{top}}^2$ and $A_{\mathrm{bottom}} = d_{\mathrm{bottom}}^2$ for the pyramidal frustum and rounded square areas (eq. (S86) and (S89)) for the pyramidal frustum with rounded lateral edges, with $h_{\mathrm{well}} = 11.5$ mm $d_{\mathrm{top}} = 3.7$ mm, $d_{\mathrm{bottom}} = 3.3$ mm and $r_{\mathrm{c,top}} = 1.0$ mm. Liquid heights for the cuboids were calculated as $h_{\mathrm{l}} = \frac{V_{\mathrm{l}}}{V_{\mathrm{total}}} \times h_{\mathrm{well}}$, with a V_{total} of either 131 µl or 138 µL, reflecting the two different specifications by the manufacturer.

Liquid volume, V_{l} (µL)	h_{l}, frustum, pyramid, round lateral edges (mm) (V_{total} = 132.2 µL)	h_{l}, frustum, pyramid (mm) (V_{total} = 141.0 µL)	h_{l}, right cuboid (mm) (V_{total} = 131 µL)	h_{l}, right cuboid (mm) (V_{total} = 138 µL)
10	0.97	0.91 (-6.2%)	0.88 (-9.5%)	0.83 (-14.1%)
20	1.92	1.80 (-6.2%)	1.76 (-8.6%)	1.67 (-13.2%)
30	2.85	2.68 (-6.1%)	2.63 (-7.7%)	2.50 (-12.4%)
40	3.77	3.54 (-6.0%)	3.51 (-6.8%)	3.33 (-11.5%)
50	4.67	4.39 (-6.0%)	4.39 (-5.9%)	4.17 (-10.7%)
100	8.93	8.41 (-5.8%)	8.78 (-1.7%)	8.33 (-6.7%)
130	11.33	10.69 (-5.7%)	11.41 (0.7%)	10.83 (-4.4%)

In summary, using the reference model of a pyramidal frustum with rounded lateral edges and the well dimensions given by the manufacturer, the total well volume was calculated as 132.2 µL. A value that lies in between the manufacturers specifications for the total well volume of 131 µL or 138 µL. In comparison to the reference model, the calculated pathlengths, are about 6% shorter when neglecting the rounded edges (pyramidal frustum model) and differ widely in a range of -14.1% to +0.7% for the rectangular cuboid models, depending on the liquid volume and assumed total well height (131 µL or 138 µL). Thus, the model of a pyramidal frustum with rounded lateral edges was used to calculate the pathlength for all acetyltransferase assays.

10.10 *Sequence alignments of secondary structure assignments*

NatC_CoA, subunit Naa30ΔC17

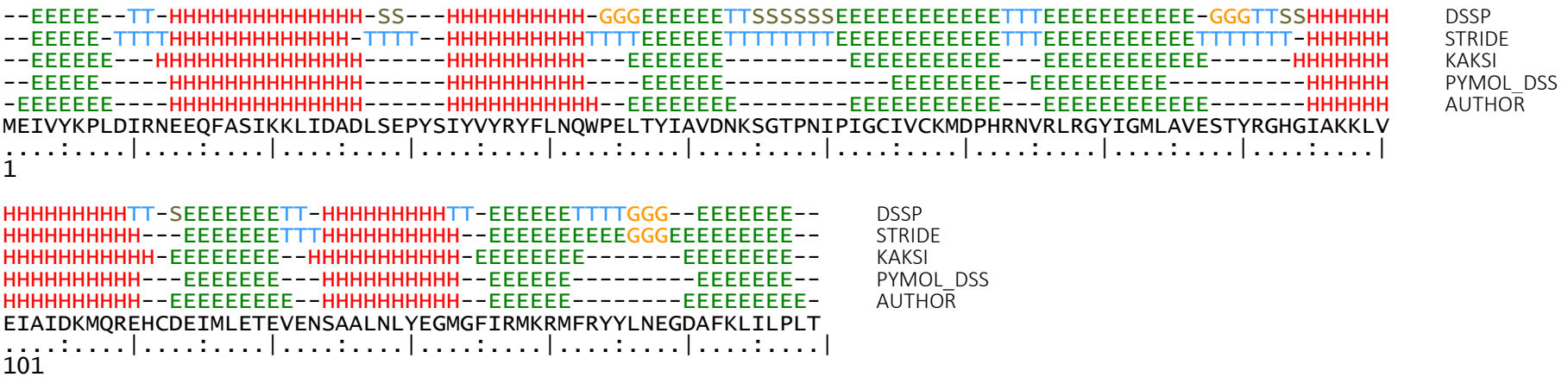
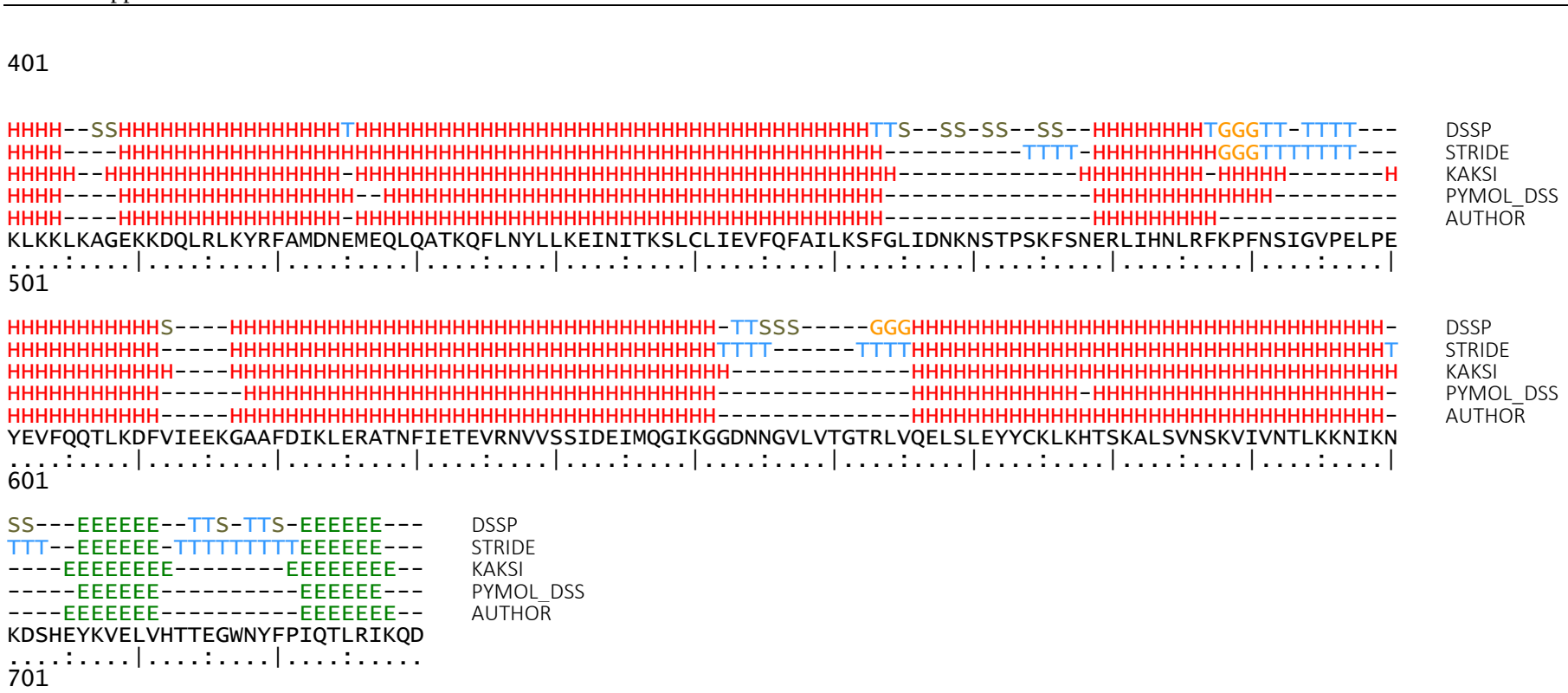


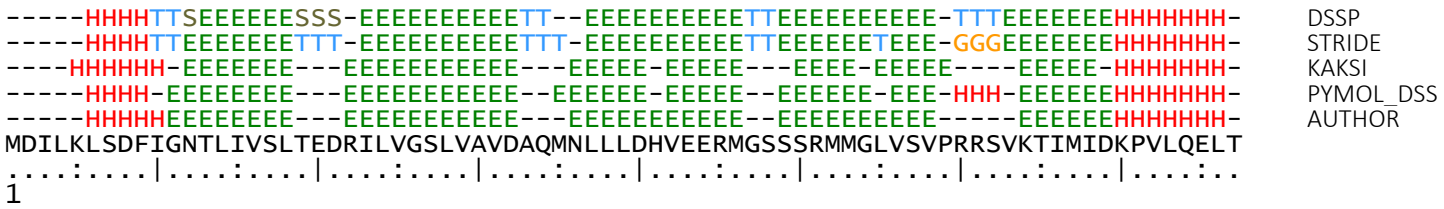
Fig. S17 Sequence alignment of the secondary structure estimations for the three subunits Naa30ΔC17, Naa38ΔC11 and Naa35 of the Coenzyme A bound NatC model (NatC-CoA). Assignments were performed using DSSP, STRIDE, PyMOLs default assignment algorithm *dss* (PYMOL_DSS) and an author provided assignment (AUTHOR). The secondary structure server (http://2struc.cryst.bbk.ac.uk/) was used for DSSP and STRIDE assignment. PYMOL_DSS and AUTHOR assignments were added manually. Original outputs were used for DSSP and STRIDE, with H = alpha helix, I = 5-helix (pi helix), G = 3-helix (3/10 helix), B = residue in isolated beta-bridge, E = extended conformation, T = hydrogen bonded turn, S = bend, - = other or 'coil'). The three-state PYMOL_DSS output (H = Helix, S = beta sheet, L = loop) was converted to H, E, -, respectively, to match the DSSP/STRIDE assignment. The AUTHOR assignment uses the same three-state representation. The alignment for subunits Naa35 and Naa38 are shown on the following two pages. X = amino acids for which no model was built due to insufficient electron density and for which the coil/other (-) secondary structure class is assigned.

NatC_CoA, subunit Naa35

```
--HHHHHHHHTTTTT-S-EEE-HHHHHHHHHHS-TT-EEE-SS--GGGGGG-EETT-TTT-GGG----HHHHH--TTS--SSHHHHHHHHHHHHHHHHH  DSSP
--HHHHHHHHHTTTTTTTEEEHHHHHHHHHHH-TTTTEEETTTT-GGGGTTTEETTTTTTTTTTT---HHHHH-TTTT-TTTHHHHHHHHHHHHHHHHH  STRIDE
-HHHHHHHHH-----------HHHHHHHHHHHH------------HHHHHH-----------------HHHHHHH-------HHHHHHHHHHHHHHHHH  KAKSI
--HHHHHHHHHHH-----EEE-HHHHHHHHHH-------------HHHHH-----------------HHHHH---------HHHHHHHHHHHHHHHHH  PYMOL_DSS
--HHHHHHHHH-------EEEHHHHHHHHHHH-----EEE-----GGGGG--EEE-----------HHHHH---------HHHHHHHHHHHHHHHHH  AUTHOR
MEVDSILGSLSITDDFDQLVDVTSLFDELCSKLKPEAIVKDPRFDLFEGTHSLEVNNSKLDSSLIELTAEEIEFDVNVAYDPPLASVAAIADRLLRCVIS
....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|
1

HHTS---STTTGGGBHHHHHHHHHHHH------S---S-HHHHTHHHHHHHHHHHHHHHHHHHHTTSS-BTTTB----SS---TTTS--HHHHHHHHHH  DSSP
HHHH---GGG-GGGBHHHHHHHHHHHH------------HHHHHHHHHHHHHHHHHHHHHHHHHH---TTTTTTT--TTTT-TTTT---HHHHHHHHHH  STRIDE
HHHH----------HHHHHHHHHHHHHH----------HHHHHHHHHHHHHHHHHHHHHHHHHHHH----------------------HHHHHHHHHH  KAKSI
HHHH---HHHHHHH-HHHHHHHHHHHH------------HHHHHHHHHHHHHHHHHHHHHHHHHHHH----------------------HHHHHHHHHH  PYMOL_DSS
HHHH---IIIIIII-HHHHHHHHHHHH------------IIIIIIIHHHHHHHHHHHHHHHHHHHHH----------------------HHHHHHHHHH  AUTHOR
WLNDYQTLPTTVLSCRYTESLLSSLVKGTXXGSSWCTGNILYDKVLGSCILGVCYLTKFVQKLLSAGIVFEEEDLNFNNMGFNTFDNLPGQDVVINSLTE
....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|
101

HHHHHHH-SS--HHHHHHHHHHHHHHHHHGGGHHHHS-----HHHHHHHHHHHHHHHHGGGS-----TTSBSSHHHHHS---S---------SSSHHHHH  DSSP
HHHHHHH-----HHHHHHHHHHHHHHHHHH-HHHHHH----HHHHHHHHHHHHHHHHHGGG-----TTTTBTHHHHHH----------------HHHHHH  STRIDE
HHHHHHHH---HHHHHHHHHHHHHHHHHHHHHHHHHH----HHHHHHHHHHHHHHHHHHH-----------HHHHHH----------------HHHHHH  KAKSI
HHHHHHH-----HHHHHHHHHHHHHHHHHHHHHHHHH------HHHHHHHHHHHHHHHHH------------HHHHH----------------HHHHH  PYMOL_DSS
HHHHHHH-----HHHHHHHHHHHHHHHHHGGGHHHHH-----HHHHHHHHHHHHHHHHGGG------------HHHHH----------------HHHHH  AUTHOR
SLQILEAYSDDSLHLTMLKHILKIIICLVHLEDHLTDYSTKTSHLDELIENANSVNGIFPQLQLSPPKGAFSTYIQKHRSNQFPPRKITKLPTDYSGFIT
....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|
201

HHHHHHHHHGGGG--SHHHHHHHHHHTTTSS---HHHHHHHHHHHB-TTSEETTTEEHHHHHHHHHHHHH-----------S-HHHHHHHHHHHHHHHT  DSSP
HHHHHHHHHGGGG---HHHHHHHHHHTTTTT--HHHHHHHHHHEETTTEETTTB-HHHHHHHHHHHHH--TTT----TTTTHHHHHHHHHHHHHHHH  STRIDE
HHHHHHHHHHHHH--HHHHHHHHHHHH------HHHHHHHHHHHH----------HHHHHHHHHHHHH------------HHHHHHHHHHHHHHHH  KAKSI
HHHHHHHHHHHHH---HHHHHHHHHH--------HHHHHHHHHHH-----------HHHHHHHHHHHHH------------HHHHHHHHHHHHHHHH  PYMOL_DSS
HHHHHHHHHGGGG---HHHHHHHHHH--------HHHHHHHHHHH-----------HHHHHHHHHHHHH------------HHHHHHHHHHHHHHHH  AUTHOR
LANDVKTILLVDKAESALETYQFAKFFNKLEQRHVIARILFPLFFIRDDRTVLGKFSYTQFYLLHVKEFSAQTPSXXXXSIGNELIQESSNMLLEWYQNC
....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|
301

TS-HHHHHHHHHHHHHHHHHHHHHHHTT-HHHHHHHHHHHHHHHHHHHHHHHTT-S-GGGHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH  DSSP
---HHHHHHHHHHHHHHHHHHHHHHHH--HHHHHHHHHHHHHHHHHHHHHHHH----GGGHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH  STRIDE
----HHHHHHHHHHHHHHHHHHHHHHH----HHHHHHHHHHHHHHHHHHHHHHH---HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-HHHHHH  KAKSI
H---HHHHHHHHHHHHHHHHHHHHHHH--HHHHHHHHHHHHHHHHHHHHHHHH----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH  PYMOL_DSS
H--HHHHHHHHHHHHHHHHHHHHHHHH--HHHHHHHHHHHHHHHHHHHHHHHH----GGGHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH  AUTHOR
SQNTCRYRQGFNRQLILWDSLQAQFESVNSQVYCSWTYFMKLSSMIEFSLKGFDLDIYKPFEAYSMFWYVYYLSHHLETFLKDSQNDIESNINAIHSMNK
....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|....:....|
```



NatC_CoA, subunit Naa38ΔC11



10.11 *HELIX and SHEET records*

```
REMARK 650
REMARK 650 HELIX
REMARK 650 DETERMINATION METHOD: AUTHOR PROVIDED
         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
HELIX    1  A1 GLU A   13  ASP A   26  1                                  14
HELIX    2  A2 ILE A   33  TRP A   43  1                                  11
HELIX    3  A3 ILE A   95  GLU A  110  1                                  16
HELIX    4  A4 SER A  124  MET A  133  1                                  10
HELIX    5  B1 VAL B    3  SER B   11  1                                   9
HELIX    6  B2 VAL B   22  LYS B   32  1                                  11
HELIX    7  B3 LEU B   46  THR B   50  5                                   5
HELIX    8  B4 ALA B   69  GLU B   73  1                                   5
HELIX    9  B5 PRO B   83  ASP B  104  1                                  22
HELIX   10  B6 LEU B  108  SER B  114  3                                   7
HELIX   11  B7 ARG B  116  LYS B  127  1                                  12
HELIX   12  B8 ILE B  140  LEU B  146  3                                   7
HELIX   13  B8 GLY B  147  ALA B  166  1                                  20
HELIX   14  B9 GLN B  191  ALA B  207  1                                  17
HELIX   15 B10 LEU B  213  VAL B  229  1                                  17
HELIX   16 B10 HIS B  230  GLU B  232  5                                   3
HELIX   17 B10 ASP B  233  ASP B  237  1                                   5
HELIX   18 B11 SER B  243  ILE B  258  1                                  16
HELIX   19 B11 PHE B  259  GLN B  261  5                                   3
HELIX   20 B12 TYR B  274  HIS B  278  1                                   5
HELIX   21 B13 SER B  296  LEU B  309  1                                  18
HELIX   22 B13 LEU B  310  LYS B  313  5                                   4
HELIX   23 B14 ALA B  317  PHE B  327  1                                  11
HELIX   24 B15 VAL B  335  ILE B  346  1                                  12
HELIX   25 B16 TYR B  358  SER B  370  1                                  13
HELIX   26 B17 GLU B  384  SER B  401  1                                  18
HELIX   27 B18 THR B  404  VAL B  428  1                                  25
HELIX   28 B19 GLN B  431  LEU B  455  1                                  25
HELIX   29 B20 PRO B  460  GLU B  462  5                                   3
HELIX   30 B20 ALA B  463  LYS B  504  1                                  42
HELIX   31 B21 GLU B  509  ASN B  524  1                                  16
HELIX   32 B22 MET B  526  PHE B  563  1                                  38
HELIX   33 B23 GLU B  579  PHE B  587  1                                   9
HELIX   34 B24 TYR B  601  PHE B  611  1                                  11
HELIX   35 B25 GLY B  617  GLY B  651  1                                  35
HELIX   36 B26 GLN B  666  LYS B  699  1                                  34
HELIX   37  C1 LEU C    6  ILE C   10  1                                   5
HELIX   38  C2 LYS C   70  LEU C   76  1                                   7
```

Fig. S 18 Manually added (author) HELIX records for the NatC-CoA structure. The numberings above the HELIX lines are not included in the PDB file and are only included to aid with column assignment: 1-6, Record name “HELIX”; 8-10, Serial number of the helix, starts at 1 and increases incrementally through all protein chains; 12-14, helix identifier (1= right-handed alpha, 3= right-handed pi, 5= right-handed 3-10). May contain any alphanumeric character; 16-18, Name of the initial residue (3-letter code); 20 and 32, chain identifier for the chain containing this helix; 22-25, Sequence number of the initial residue; 28-30, name of the terminal residue of the helix; 34-36, Sequence number of the terminal residue; 39-40, helix class (see below); 41-70, Comments about this helix; 72-76, Length of this helix.

```
REMARK 700 SHEET
REMARK 700 DETERMINATION METHOD: AUTHOR PROVIDED
REMARK 700 THE SHEET STRUCTURE OF THE MOLECULE NAA38 (CHAIN C) IS BIFURCATED.
REMARK 700 IN ORDER TO REPRESENT THIS FEATURE IN THE SHEET RECORDS BELOW,
REMARK 700 TWO SHEETS ARE DEFINED. STRANDS 1, 2, 3 AND 4
REMARK 700 OF SHEET *B* AND *C* ARE IDENTICAL.

         1         2         3         4         5         6         7         8
12345678901234567890123456789012345678901234567890123456789012345678901234567890
SHEET    1   A 7 GLU A   2  LEU A   8  0
SHEET    2   A 7 LEU A  46  ASN A  53 -1  N  ASN A  53   O  GLU A   2
SHEET    3   A 7 ILE A  62  HIS A  72 -1  N  CYS A  67   O  LEU A  46
SHEET    4   A 7 ARG A  76  GLU A  87 -1  N  GLU A  87   O  ILE A  62
SHEET    5   A 7 ASP A 113  VAL A 121  1  O  MET A 116   N  ILE A  81
SHEET    6   A 7 ASP A 150  LEU A 158 -1  N  LEU A 158   O  ASP A 113
SHEET    7   A 7 ILE A 136  MET A 141 -1  N  MET A 141   O  ALA A 151
SHEET    1   B 7 LEU B  53  VAL B  55  0
SHEET    2   B 7 SER C  49  VAL C  58  1  O  SER C  57   N  VAL B  55
SHEET    3   B 7 ASN C  36  MET C  46 -1  O  LEU C  37   N  VAL C  58
SHEET    4   B 7 ARG C  22  ASP C  32 -1  N  VAL C  25   O  GLU C  43
SHEET    5   B 7 GLY C  11  LEU C  18 -1  N  LEU C  18   O  ARG C  22
SHEET    6   B 7 LYS C  64  ASP C  69 -1  O  MET C  67   N  ILE C  15
SHEET    7   B 7 LEU B  19  ASP B  21 -1  N  VAL B  22   O  ILE C  68
SHEET    1   C 5 LEU B  53  VAL B  55  0
SHEET    2   C 5 SER C  49  VAL C  58  1  O  SER C  57   N  VAL B  55
SHEET    3   C 5 ASN C  36  MET C  46 -1  O  LEU C  37   N  VAL C  58
SHEET    4   C 5 ARG C  22  ASP C  32 -1  N  VAL C  25   O  GLU C  43
SHEET    5   C 5 ILE B  38  LYS B  40 -1  N  VAL B  39   O  VAL C  31
SHEET    1   D 2 GLU B 705  VAL B 711  0
SHEET    2   D 2 ILE B 722  LYS B 728 -1  N  ARG B 726   O  LYS B 707
```

Fig. S 19 Manually added (author) SHEET records for the NatC-CoA structure. The numberings above the SHEET lines are not included in the PDB file and are only included to aid with column assignment: 1-6, Record name “SHEET”; 8-10 Strand number within a sheet. Starts at 1 for every sheet and increases by one; 12-14, Sheet identifier, which can be any alphanumeric character; 15-16, number of strands in sheet; Description of the initial residue in strand with 18-20, residue name; 22, chain ID and 23-26, sequence number; Description of the terminal residue in strand with 29-31, residue name; 33, chain ID and 34-37, sequence number; 39-40 Sense of the strand, with respect to previous strand in the sheet, with 0 = first strand, 1 = parallel and -1 = antiparallel; Registration of current strand with 42-45, Atom name; 46-48, residue name; 50, chain ID and 51-54, sequence number; Registration of previous strand with 57-60, Atom name; 61-63, residue name; 65, chain ID and 66-69, sequence number.

Publications

(Addendum, added on 11 November 2020)

Parts of this thesis have been published after the submission (01 October 2019) and disputation (26 November 2019) of this dissertation:

1) The core findings of this thesis have been published online in *Nature Communications*:

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.

Received: 01 April 2020 Accepted: 06 October 2020 Published: 02 November 2020

2) Coordinates and diffraction data have been deposited (27 March 2020) and released (28 October 2020) in the Protein Data Bank (PDB) with the following accession codes:

6YGA – Crystal structure of the apo NatC complex
DOI: 10.2210/pdb6YGA/pdb

6YGB – Crystal structure of the NatC complex bound to CoA
DOI: 10.2210/pdb6YGB/pdb

6YGC – Crystal structure of the NatC complex bound to Arl3 peptide and CoA
DOI: 10.2210/pdb6YGC/pdb

6YGD – Crystal structure of the NatC complex bound to Gag peptide and CoA
DOI: 10.2210/pdb6YGD/pdb

Deposition Authors:
Grunwald, S., Hopf, L., Bock-Bierbaum, T., Lally, C.C., Spahn, C.M.T., Daumke, O.