Characterization of *NAA10* mutations in patients exhibiting phenotypes associated with N-terminal acetylation deficiency

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Science

Department of Molecular Biology
Faculty of Mathematics and Natural Sciences
University of Bergen

Takk

Denne masteroppgåva er skrive og gjennomført i perioden januar til november 2018 ved NAT-gruppa på Institutt for Biovitskap, Universitetet i Bergen. Lab har vorte utført på Institutt for Biomedisin ved Haukeland Universitetsjukehus.

Det er mange som skal ha takk for at gjennomføringa av denne masteroppgåva har vore mogleg, men ein særskild takk ynskjer eg å rette min tolmodige og venlege medrettleiar, Rasmus Moen Ree. Det er ikkje alltid lett å ha med meg å gjere når eksperimentar og skriving ikkje går slik ein ynskjer, men Rasmus har jamvel ikkje vist meg noko anna enn optimisme og velvilje i heile år. Tusen takk for det! Hovudrettleiar Thomas Arnesen skal også ha stor takk for all hjelp, særleg i skriveprosessen; helg eller solskin, det spelar inga rolle, oppgåva mi har vorte retta med strålande humør. Takk også for at eg har fått anledning til å skrive denne oppgåva og vere eit år på di gruppe, Thomas. Eg tek utan tvil med meg verdifull erfaring og kompetanse inn i eventuell jobb eller vidare studier.

Vidare må eg takke den fantastiske NAT-gruppa! Det året eg har vore på lab, har eg aldri vegra meg for å spørje kven det måtte vere om hjelp eller råd. Også i sosiale aktivitetar utafor lab har eg ikkje hatt noko anna enn gode opplevingar. Det å få vere med på seminar i nasjonalromantiske omgjevnader og trivlege middagar, har vore ei sann glede! Det er alikevel ingen ting som har bidrege meir til samhaldskjensla enn den vekentlege innebandy-treninga, organisert av Rasmus. Det er ikkje ofte eg gler meg til noko som skje 7.45 om morgonen, så tusen takk til alle som har møtt opp til trening i 2018!

Monica Hellesvik, min briljante veninne og labpartner, skal også ha stor takk for intet mindre enn å vere seg sjølv: Ein støttande og varm person med ein ekstra te-kopp på lur. Pausene med deg har vore veldig verdifulle! Ynskjer deg masse lykke til med din masteroppgåve!

Til slutt vil eg takke mine supre foreldre som ikkje eitt skund har tvilt på kapasiteten min, og til min kjæraste Leo som har holdt ut med den berg-og-dal-bana det har vore å stå på lab og skrive ein master. Kva skulle eg gjort utan deg?

Bergen, 25. november 2018

Anni Sofie Geithus

Selected abbreviations

AB	Antibody
Ac-CoA	Acetyl Coenzyme A
BSA	Bovine serum albumin
BCA	Bicinchoninic acid
CoA	Coenzyme A
CV	Column volume
ddNTP	Dideoxynucleotide
DTNB	5,5'-dithiobis(2-nitrobenzoic acid)
DTT	Dithiothreitol
FBS	Fetal bovine serum
GNAT	GCN5-related N-acetyltransferase
GCN5	General Control Nonderepressed 5
HAT	Histone acetyltransferase
His	Histidine
HRP	Horseradish peroxidase
HYPK	Huntingtin-interacting protein K
IMAC	Immobilized metal affinity chromatography
iMet	Initiator methionine
IPTG	Isopropyl-β-D-thiogalactopyranoside
KAT	Lysine acetyltransferase
kDa	Kilo dalton
KDAC	Lysine deacetylase
LMS	Lenz microphthalmia syndrome
MBP	Maltose binding protein
MetAP	Methionine aminopeptidase
MSA	Multiple sequence alignment
NAA	N-alpha acetyltransferases
NAT	N-terminal acetyltransferase
Nt	N-terminal
OD	Optical density
PenStrep	Penicillin streptomycin
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TEMED	N,N,N',N'-tetramethylethane-1,2-diamine
v/v	volume/volume
w/v	weight/volume
WB	Western blotting
WR	Working reagent
XCI	X-chromosome inactivation

Table of content

1 Summary

Approximately 80% of the proteins found in human cells are acetylated on their N-terminus, either partially or completely, by a group of enzymes called N-terminal acetyltransferases (NATs)[1]. To date, seven NATs, NatA-NatF[2–7] and NatH[8], have been identified in humans; NatA-NatE associate with ribosomes and acetylate their substrates co-translationally, NatF associate with the Golgi apparatus membrane and acetylate membrane proteins post-translationally, and NatH is found in the cytosol where it acetylates actins, also post-translationally[8,9].

The NatA complex, consisting of catalytic subunit NAA10 and auxiliary subunit NAA15 facilitating the association of the complex with the ribosome, has the most protein N-termini targets among the NATs, and is also the most studied complex[2,9]. The function of NatA and uncomplexed NAA10 in the cell is gradually being characterized by an increasing number of studies on patient identified NAA10 mutations[10–14]. The patients share a wide spectrum of phenotypes, most commonly developmental delay, growth deficiency, post-natal growth failure, and cardiac and skeletal anomalies[13].

This thesis' main focus is to functionally characterize two novel NAA10 missense mutations, C21G and R83H, identified in three male patients presenting with typical NAA10 deficiency. Characterization involves *in vitro* acetylation assay of uncomplexed recombinant MBP-NAA10, cellular stability assay of NAA10 and NatA, and bioinformatic assessments of structural functions as well as conservation and *in silico* prediction of disease. The work presented here, demonstrates that both mutations are found in highly conserved regions in NAA10 involved with Ac-CoA and substrate binding as well as NatA complex formation. The acetylation assay show that while the R83H mutant has profoundly reduced catalytic activity, the C21G mutant show an increased activity. From the cellular stability assay, none of the mutations show any apparent destabilizing effects. However, to draw conclusions about the functional impact of these mutations on the NAA10 and NatA and their link to human health, more research is needed than is presented here in this thesis.

2 Introduction

2.1 Protein modifications

The human proteome owes its complexity not only to the approximately 20 000 protein-coding genes residing in each human cell, but also to the many cellular mechanisms and physiological factors that impact which genes are expressed and how the final protein product is built [15]. A single gene may give rise to as many as 100 protein products through the processes of alternative mRNA splicing, single amino acid polymorphisms or protein modifications that the protein attains during or after synthesis[16]. The addition of certain chemical groups, such as ubiquitin, acetyl, methyl and phosphoryl, to the amino acid side chains or the N- and C-termini represents some examples of such modifications[17–19].

2.1.1 Acetylation

Acetylation as a modification in histones was first discovered in 1962[20], but gained little attention the first 30 years following its discovery [18]. It was not until the 1990's[21], when the acetylation of certain lysine residues on histone tails was connected to the regulation of gene transcription, that protein acetylation received its deserved attention. For many years, the enzymes facilitating lysine acetylation were termed histone acetyltransferases (HATs), since only acetylated histones had been found. Since then, it has been uncovered that many other proteins carry an acetyl on their lysine residues[22], and the name lysine acetyltransferase (KAT) is now more commonly used. By the action of lysine deacetylases (KDACs), facilitating the removal of lysine-attached acetyl moieties, lysine acetylation is a reversible modification[22].

Both KATs, and another common family of acetyltransferases that acetylate their substrate proteins on the N-terminus, the N-terminal acetyltransferases (NATs), use Acetyl-Coenzyme A (Ac-CoA) as an acetyl donor in the catalytic reaction (Figure 2.1). The addition of acetyl to the α-amino group on the N-terminus, or the ε-amino group of a lysine, leads to the neutralization of the positive charge and alters the electrostatic properties of the protein. Unlike KATs, there has been no discovery of N-terminal deacetylating enzymes to date, so the NAT catalysed reaction is considered irreversible.

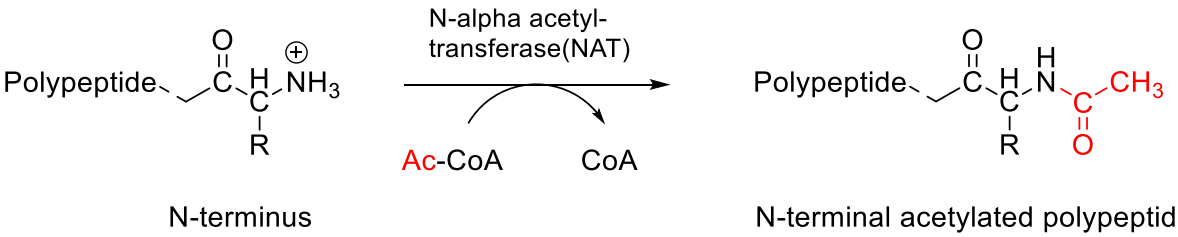

Figure 2.1 The transfer of acetyl from Ac-CoA to the N-terminus of a peptide. *Specialized enzymes called N-alpha acetyltransferases (NAAs) or N-terminal acetyltransferases (NATs) carry out acetylation on specific protein N-termini in the cell either co- or post-translationally. This catalytic reaction neutralizes the positive charge on the peptide's N-terminus, thus altering the peptide's electrostatic properties.*

Acetyl moieties on the N-terminus are present in 80-90% of human proteins and 50-70% of yeast proteins[23], but the biological function of an N-terminal acetyl group for many of these proteins remain elusive. For certain proteins, however, the acetyl moiety has been shown to play a part in protein-protein interactions and complex formation[24,25], in subcellular localization[26,27], and in protein folding[28,29]. Additionally, N-terminal acetyl groups may work as degradation signals through a branch of the N-end rule pathway, called the Ac/N-end rule[30,31], rendering the protein detectable to N-recognins and subsequent polyubiquitylation. Proteins avoid recognin detection by shielding the N-terminus in folding, forming complexes or association with chaperones. This makes N-terminal acetylation a way for the cell to quality control synthesized proteins and ensure a balanced protein stoichiometry[32]. However, a recent study in yeast reveals that acetyl on the N-terminus rarely act as a degron, and might actually in some cases have the opposite effect; rather than working as a degradation signal, N-terminal acetylation has been shown block degradation in certain pathways[33].

2.2 N-terminal acetyltransferases

Though the abundance of proteins acetylated on the N-terminus increases with the organism's complexity, the presence of this modification has been documented throughout evolution, from bacteria to yeast, plants and humans[23,34]. Facilitating this event, is the N-terminal acetyltransferase (NAT) family; of which three prokaryotic members, RimI, RimJ and RimL[35,36], and eight eukaryotic members, NatA-NatH[8,37] (Figure 2.2), have been identified. NatA-NatE can be found interacting with ribosomes in the cytosol of yeast, mammalian and plant cells, where they acetylate substrate proteins co-translationally[38]; NatF localizes to the Golgi in mammalian cells, where it acetylates the membrane proteins post-translationally[39],

and NatG is only found in the chloroplasts of plant cells[40]. NatH, the most recent discovery[8], localizes to the cytosol where it acetylates the N-termini of actins.

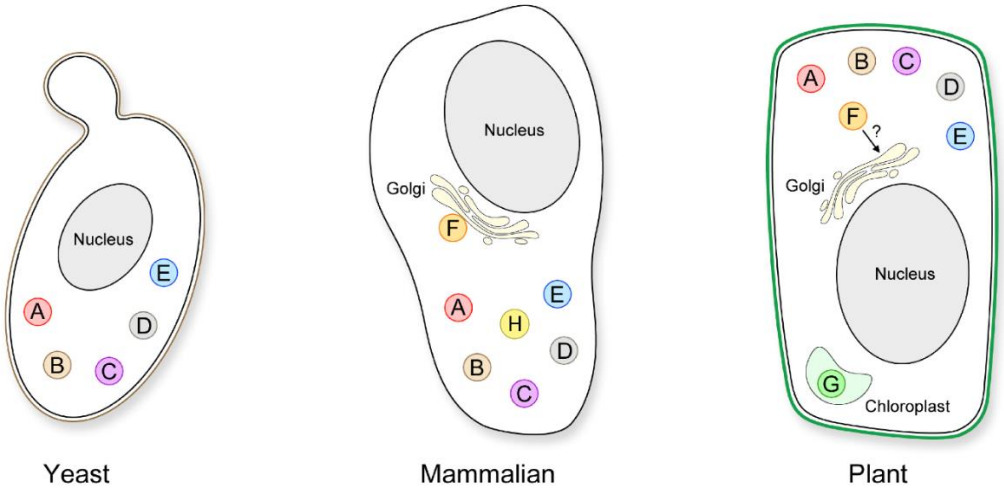

Figure 2.2 Homologues of NAT family members are found in a wide range of eukaryotic species. *The acetylation of protein N-termini is a conserved modification from bacteria to unicellular and multicellular eukaryotes. The protein family in eukaryotic species that carries out this modification, are called N-terminal acetyltransferases (NATs). NatA-E can be found in the cytosol where they associate with ribosomes, NatF associates with the Golgi apparatus of mammalian cells – it is uncertain if this is the case in plants, NatG is exclusive to plant cells, and NatH localizes to the cytosol in mammalian cells where it acetylates actin. Figure modified from Aksnes et al. 2016.*

2.2.1 Human NATs

Subunits, substrate specificity and subcellular localization vary among the different NATs (Table 2.1). NatA, for instance specializes in acetylating the N-termini of peptides starting with small, polar amino acids after the initiator methionine (iMet) has been cleaved off by a methionine aminopeptidase (MetAP)[23]. NatD, which acetylates H2A and H4, also require the removal of the iMet group prior to substrate binding[5], whereas NatB, NatC, NatE and NatF, on the other hand, are examples of NATs that acetylate N-termini with the initial methionine still attached[41,42].

Table 2.1 Composition and substrate specificity of the human NATs

NAT	Catalytic Subunit	Auxiliary subunits	Substrate N-termini
A	NAA10	NAA15, HYPK?	A, S, T, C, V, G[2,23]
B	NAA20	NAA25	MD, MN, ME, MQ[3,42]
C	NAA30	NAA35, NAA38	ML MI, MF, MW, MV, MM, MH, MK[4,43]
D	NAA40	-	SG (histones H2A and H4)[5]
E	NAA50	NAA15, NAA10?	MK, MV, MA, MY, MF, ML, MS, MT, MG, MI, MW, MM, MH, MK, MR[6,44,45]
F	NAA60	-	ML, MF, MI, MW, MK[7,39]
H	NAA80	-	β- and γ-actin [8,46]

2.2.2 NAT structure

The structures of the catalytic subunits in the NAT family are highly conserved and classified among the General Control Nonderepressed 5 (GCN5)-related N-acetyltransferase (GNAT) domains. Comprised of six to seven β-strands and four α-helices, the GNAT domain facilitates the transfer of acetyl from Ac-CoA to a primary amine of a recipient molecule [47,48]. As is the case for NATs, the acetyl recipient can be another protein, but a wide range of smaller molecules exhibiting a primary amine, such as amino glycoside and serotonin, could become acetylated provided the GNAT domain retain specificity for it [48].

A particularly conserved motif in the GNAT domains, is the Ac-CoA-binding region Q/RxxGxG/A[49,50]. A study investigating the substrate binding of NAA50, utilizing NMR and enzyme kinetics, showed that association of Ac-CoA with the Q/RxxGxG/A motif results in a conformational change and a subsequent increase in affinity toward the peptide substrate[51]. This suggests that NAA50, and therefore likely other NATs as well, follow a Theorell-Chance type of Bi Bi sequential mechanism[51]. The amino acids determining the specificity of the peptide substrate, however, are less conserved between the various catalytic domains of the NATs [52].

2.3 Human NatA

Of all the NATs found in human cells, the NatA complex is by far the most promiscuous, acetylating around 40% of human proteins[23]. The acetylation is carried out by the 235 amino acids long catalytic subunit, NAA10, and ribosomal anchorage is facilitated by 866 amino acid long NAA15[2,53]. Additional interacting partners to NatA are Huntingtin-interacting protein K (HYPK) and NatE.

2.3.1 Structure

The 45 α-helices that make up the NAA15 subunit are arranged into a ring-like tertiary structure, taking the shape of 13 tetracopeptide repeat (TPR) motifs and one C-terminal Sel1-like repeat (SLR)[54]. These structures are known to participate in the protein-protein interaction between NatA, the ribosome and auxiliary subunits NAA50 and HYPK[52,53] (Table 2.1); some are also involved in the binding of NAA10[52], docking it to the centre of the NAA15 ring-

structure and stabilizing it. The complex formation of NAA15 and NAA10 facilitates a conformational change in the NAA10 catalytic site, thereby altering the substrate specificity, as well as ensuring a necessary proximity of NAA10 to the nascent polypeptide chain exit on the ribosome[52,54].

2.3.2 Auxiliary interaction partners

The complex also associates with the chaperone-like protein HYPK[28], and the catalytic subunit NAA50, also known as NatE[55] (Table 2.1). Research on NatA and HYPK association by Arnesen and collegues[28] showed that the interaction was essential for HYPK stability and the acetylation activity of NatA. Two other studies, in which solved structures of HYPK-associating NatA were published, found that HYPK displayed an inhibitory function toward NatA by blocking the active site[54,56], indicating that HYPK might acts as a regulator on NatA activity. NAA50 is recognized as its own NAT enzyme, NatE[6], because it retains its own distinct substrate specificity (Table 2.1), but its role as a NatA auxiliary unit, however, is not yet quite understood. The identification, in a 2015 study, of the N-terminal iMet residue followed by a small amino acid as NAA50 substrate, revealed that NAA50 competes against MetAPs in modifying/processing the iMet[45]. As MetAPs appear unable to process the acetylated iMet, this indicates a function of NAA15 toward the retainment of certain proteins' iMet residue[45].

2.3.3 NAA10

The catalytic subunit of NatA does not only exist in a ribosome-associated complex with NAA15 and its auxiliary interacting subunits; it can also be found in monomeric form, localized to the nucleus[2,44]. Early *in vitro* studies of NAA10, revealed that the monomeric form retained specificity for acidic N-termini, thereby making NAA10 a likely candidate for the N-terminal acetylation of actin[44]. Since then, actin acetylation has been attributed to the catalytic actions of NatH. Other potential NAA10 targets have been suggested, however, indicating involvement of NAA10 in, among others, bone and blood vessel formation[57] and organism development[58–60]. Several studies have also claimed that NAA10 exerts KAT activity[57,61,62], but contradictory analyses of the NAA10 structure, suggesting that the substrate binding site is too narrow to acetylate internal lysine residues[63], have caused a long-standing debate concerning this function. Research published earlier this year show that NAA10 interacts with the dioxygenase catalytic domain of FIH, leading to the hydroxylation of NAA10 at W38[64]. This hydroxylation

is then suggested to cause a widening of the NAA10 binding site, adapting it for lysine acetylation[64].

2.3.4 Disease

Given the substantial number of NatA targets in the cell, in addition to the individual roles of NAA10, it is no surprise that a malfunctioning NAA10 may have serious consequences at the organismal level. Studies on several model organisms show that the expression of NAA10 is essential for viability[58–60,65]. In zebrafish, the loss of NAA10 function was associated with serious phenotypes during embryonic development[58]. Underdeveloped eyes and reduced pigmentation were among the early stage adversities. Later stages revealed a decrease in growth and activity relative to the wild type, underdeveloped mandibles and pericardial oedema[58]. A mutation in the *NAA10 g*ene studied in homozygous fruit flies, reported fatality at the second-instar larval stage, and several defects in oogenesis, egg chamber capsulation and nurse cell chromatin dispersion[59]. The two life stage forms of the parasite *T. brucei* also depend on the expression of *NAA10* for viability[60]. Some recently developed *NAA10* knockout mouse models, however, seem to not depend on NAA10 for survival[57,66], despite suffering from growth retardation and some neuronal defects[66]. A possible explanation for the survival of the knockout models is the compensatory expression of the paralogous NAA10 variant, NAA11, ensuring the continued forming of the NatA complex and thus survival of the models[67].

The malfunction of NatA and NAA10 has in many cases been connected to various human disorders and abnormalities, ranging from cancer to development-related syndromes and neurodegenerative diseases[13,68,69]. Studies have identified NAA10 and NAA15 as oncoproteins, where upregulated expression is linked to tumour aggressiveness and poor prognosis in various cancer types[22,70]. Elevated NAA15 levels has, for instance, been reported in papillary thyroid carcinoma[71], neuroblastomas[72] and gastric cancer[68], whereas elevated NAA10 expression has been documented in studies on breast, lung and prostate cancer[73,74], among others. Neurodegenerative disorders, such as Huntington's disease, arise from the aggregation of proteins to fibrilic macrostructures that puncture and degenerate neural cells[75]. The chaperone activity of HYPK when associated with NatA, is shown to prevent the aggregation of huntingtin[28], and NatA is reported to stabilize β-amyloid precursor protein (APP), thus hindering it in becoming the plaque-forming amyloid β-protein potentially leading to Alzheimer's disease[76].

In 2011, the *NAA10* missense mutation, S37P, was identified as the cause of a recessive X-linked lethal disorder, Ogden syndrome[10] (Figure 2.3, left). As S37 is a conserved residue in the NAA10 protein, its substitution to proline was suggested to affect structure and enzymatic activity, thus causing disease[10]. Later investigation also showed that the S37P mutation impaired complex formation of NatA and that NatA mediated Nt-acetylation was impaired in cells derived from Ogden syndrome boys[77]. Female carriers of the S37P mutation presented with an X-chromosome inactivation (XCI) of the mutated allele, either completely, or nearly completely, skewing towards the wild type[77]. A 2014 study of Lenz microphthalmia syndrome (LMS) revealed a *NAA10* splice mutation, 471+2T→A, leading to the truncation of the NAA10 C-terminus (Figure 2.3, right). Unlike Ogden syndrome, LMS is not lethal in infancy, but severe phenotypes, such as blindness due to underdeveloped or absence of eyes, digit, teeth and ear abnormalities, scoliosis and renal anomalies are present in affected individuals[11].

Figure 2.3 NAA10 mutations cause physiological and syndromic phenotypes. *The missense mutation S37P is causative for the Ogden syndrome (left); a lethal condition in which the patient presents with global developmental delay, cryptorchidism, hypotonia, craniofacial anomalies and cardiac arrythmia. Females with the mutation are unaffected*[10]*. The splice mutation c.471+2T→A (right), causes Lenz microphthalmia syndrome, and is characterized by anophthalmia/microphthalmia; digit, ear and teeth anomalies and skeletal abnormalities. 60% of affected males exhibit intellectual disability and seizure disorders*[11]*. Images from Rope et al. 2011 (left) and Esmailpour et al. (2014).*

As a result of increasing identification of NAA10 mutations in patients over the past seven years, some shared phenotypic features have been identified in the affected individuals (Table 2.2), among which are developmental delay, intellectual disability (ID), postnatal growth deficiency, hypotonia, cardiac arrythmia and various skeletal, organ and morphological

anomalies. Some attempts have been made at identifying consistency between the specific mutations and phenotypic features of affected individuals; a study investigating the phenotypes and NAA10 mutations of 11 affected females, however, found no evident phenotypic pattern between the seven girls presenting with one mutation (R83C) as opposed to the rest (V107F, R116W, F128L and F128I) [13].

Early functional analyses NAA10 mutants, showed a correlation between reduced NAA10 activity and phenotype severity. The 2011 study identifying the NAA10 mutation S37P as the cause of Ogden syndrome, which resulted in severe phenotypes and infant lethality, presented a reduction in NAA10 catalytic activity of 20-80%[10] (Table 2.2). Researchers therefore hypothesized that the reduction in activity could explain the severity of the mutation; A 2015 study investigating the Y43S mutation in two brothers, however, found an 85% reduction in the NAA10 catalytic activity, but where the affected individuals presented relatively mild phenotypes[78] (Table 2.2). Additionally, the R116W mutant, reported in studies by Rauch et al. (2012) and Popp et al. (2015), had a reduction in acetylation activity of only 15%, yet the phenotypes of affected individuals were characterized as severe[12] (Table 2.2). These comparisons indicate that the severity of NAA10 mutations rely on other factors in addition to catalytic activity; NatA complex formation, protein stability and the ability of the mutant NAA10 to associate with Ac-CoA or substrate, for instance, are functions that have been shown to be consequential for severity[13,77]. The forming of the NatA complex not only depends on the integrity of NAA10; a study published earlier this year reported that many of the typical phenotypes seen in patients with NAA10 mutations were also found in patients with truncated NAA15 variants [79]. The researchers hypothesized that this was due to impaired NatA complex formation. A selection of the identified variants and corresponding phenotypes are given in Table 2.2, but functional consequences of these variants are not yet characterized.

Table 2.2 Effects on protein function and phenotypic features of NAA10 and NAA15 variants

Variants	Mutations	Function	Phenotype
NAA10	S37P[10] *	Reduced catalytic activity (20-80%), impaired complex formation, impaired substrate binding	Severe NAA10 deficiency, truncal hypotonia, scoliosis, prominent eyes, aged appearance, lethal cardiac arrythmia and cardiomyopathy
	c.471+2T>A[11]	C-terminal truncation, impaired PPI, dysregulation of pathways	NAA10 deficiency, anophthalmia or microphthalmia, scoliosis, hypotonia, renal abnormalities
	Y43S[78]	Reduced catalytic activity (85%), reduced protein stability, form aggregates	Mild NAA10 deficiency, hypotonia, scoliosis, prolonged QT
	R83C[13]	Reduced catalytic activity (60%), impaired Ac-CoA binding	Severe/moderate NAA10 deficiency, hypotonia, limited speech and mobility, facial dysmorphism, microcephaly
	R116W[12,80]	Reduced catalytic activity (15%), Impaired Ac-CoA binding	NAA10 deficiency, hypotonia, minor facial features, behavioral anomalies
	V107F[12,13]	Reduced catalytic activity (95%), reduced protein stability	NAA10 deficiency, hypotonia, minor facial features, behavioral anomalies
	F128I[13,81]	Reduced catalytic activity, reduced protein stability	NAA10 deficiency, microcephaly, hypotonia, limited speech and mobility, facial dysmorphism
	F128L[13]	Reduced catalytic activity (>90%), reduced protein stability, form aggregates	NAA10 deficiency, hypotonia, limited speech and mobility, facial dysmorphism
	V111G[14]	Reduced protein stability (85%), reduced NAA10 activity (no change in NatA activity)	Mild NAA10 deficiency, delayed motor- and language development
NAA15	D76Efs*20[79]		Developmental delay, ID, autism, constipation
	T55Hfs*2[79]		Abnormal heart rhythm, developmental delay, autism, minor cerebral anomalies
	H80Rfs*17[79]		Behavioral abnormalities, facial dysmorphism, delayed speech development, multiple allergies,
	W83*[79]		Autism, epiloia, skin abnormalities, behavioural abnormalities, delayed speech development,
	G290*[79]		Global developmental delay, delayed speech devolvement, autistic traits, behavioral abnormalities

*) lethal, ID = intellectual disability, PPI = protein-protein interactions, FS = Frameshift, NAA10 deficiency = developmental delay, growth deficiency, post-natal growth failure, cardiac and skeletal anomalies[13]

2.4 Aim of thesis

The acetylation of protein N-termini is an evolutionary conserved protein modification, carried out by NATs, which prepare target proteins for subsequent cellular tasks and fates. In humans, NatA, either completely or partially, acetylates 40% of the proteome, more specifically MetAP processed N-termini starting with a small amino acid. The malfunction of NatA as a consequence of *NAA10* mutations has become an increasingly researched topic, since the lethal S37P mutation of Ogden syndrome was identified in 2011[10]. Some studies have attempted to uncover links between the observed phenotypes of the affected individuals and their specific mutations, but so far, few consistent patterns have been identified[13]. There is however a range of phenotypes shared among the patients retaining the various mutations, such as intellectual disability, growth deficiency, hypotonia, and skeletal and cardiac anomalies, hinting at which cellular pathways and developmental stages are most heavily affected by NatA dysfunction[22].

Though research on NAA10 mutations have provided us with a broader understanding of NAA10 structure and function, many aspects of the disorders and phenotypes accompanying NAA10 mutations remain elusive. In this thesis, molecular biological methods and bioinformatics was used to investigate whether two novel NAA10 variants, C21G and R83H, had an impact on NAA10 function in terms of activity and stability. To this end, wild-type and mutant NAA10 was purified before being submitted in an *in vitro* acetylation assay for activity assessment, as well as transfected into human cells to undergo subsequent stability analysis. Additionally, disease prediction, conservation assessments and structural examinations were conducted to investigate the mutations *in silico*.

3 Materials

3.1 Antibodies

Table 3.2 Primary and secondary antibodies used for Western blotting

Supplier	Antibody	Order	Species	Cat. no.
Invitrogen	Anti-V5	1º	Mouse	46-0705
Abcam	Anti-pan-actin	1º	Mouse	AB14128
ECL™	Anti-mouse	2º	Sheep	NA931V

3.2 Bacteria strains

Table 3.1 Strains of bacteria used for construct propagation and expression of MBP-NAA10 variants

Name	Cat. no.	Use	Suppplier
NEB® 5-alpha Competent E. coli	C2987H	Transformation and cloning	New England Biolabs
One Shot® TOP10 Chemically Competent *E. coli*	C404003	Transformation and cloning	Invitrogen
One Shot® BL21 Star™ (DE3) Chemically Competent *E. coli*	C601003	Protein expression	Invitrogen

3.3 Buffers, solutions and media

3.3.1 Growth media for bacteria

1x SOC outgrowth medium (New England Biolabs)

- 2% vegetable peptone
- 0.5% yeast extract
- 10 mM NaCl
- 2.5 mM KCl
- 10 mM MgCl2
- 10 mM MgSO4
- 20 mM glucose

LB-media

- 10 g/L tryptone
- 5 g/L yeast extract
- 5 g/L NaCl
- Autoclaved

LB-Agar

- 15 g/L agar in LB-medium
- Autoclaved

3.3.2 Buffers used in protein purification

***E. coli* lysis buffer, pH 7.4**

- 50 mM Tris-HCl, pH 8
- 300 mM NaCl
- 2 mM DTT
- 20 mM imidazole
- 1 tablet of C0mplete EDTA free protease inhibitor per 50 mL of buffer (added right before use)

IMAC wash buffer, pH 7.4

- 50 mM Tris-HCl, pH 8
- 300 mM NaCl
- 2 mM DTT
- 20 mM imidazole
- Filtered and degassed before use

IMAC elution buffer, pH 7.4

- 50 mM Tris-HCl, pH 8
- 300 mM NaCl
- 2 mM DTT
- 350 mM imidazole
- Filtered and degassed before use

NAA10 gel filtration buffer, pH 7.4

- 50 mM Tris-HCl, pH 8
- 300 mM NaCl
- 2 mM DTT
- Filtered and degassed before use

3.3.3 Buffers used in DTNB-based acetylation assay

2x Acetylation buffer, pH 8.5

- 100 mM Tris-HCl, pH 8
- 2 mM EDTA
- 20% glycerol

Quenching buffer, pH 6.8

- 3.2 M guanidine-HCl
- 100 mM Na_2HPO_4

DTNB buffer, pH 6.8

- 100 mM Na_2HPO_4
- 10 mM EDTA

3.3.4 Buffers and solutions used in SDS-PAGE and Western blotting

1x PBS

- 8 g/L NaCl
- 0.2 g/L KCl
- 2 g/L Na_2HPO_4
- 0.4 g/L KH_2PO_4

1x PBS-TWEEN

- 1x PBS
- 0.1% (v/v) Tween-20

SDS-PAGE running buffer

- 10% Bio-Rad 10x TGS

Towbin buffer, pH 8.3

- 10% Bio-Rad 10x TG

5x protein sample buffer

- 0.17 M Tris-HCl (pH 6.8)
- 5.2% (v/v) SDS
- 25.7% (v/v) glycerol
- 62.5 mM (v/v) DTT
- 0.025% (w/v) bromophenol blue
- 4.17 mM KCl

3.3.5 Buffers and media used in HeLa cell work

IPH buffer, pH 8

- 50 mM Tris-HCl
- 150 mM NaCl
- 5 mM EDTA
- 0.5% NP-40

Dulbecco's Modified Eagle's Medium (DMEM) (Sigma-Aldrich)

- 4500 mg/L glucose, sodium pyruvate, and sodium bicarbonate
- 10% FBS
- 2% L-glutamine
- 2% Penicillin Streptomycin (Pen-Strep)

3.4 Chemicals and reagents

Table 3.3 Various chemicals and reagents

Supplier	Name	Cat. no.
AppliChem	Isopropylthio-β-galactoside (IPTG)	A1008,005
Bio-Rad	Bio-Safe™ Coomassie Stain	161-0786
	10x Tris/glycine/SDS	161-0772
	10x Tris/glycine	161-0734
Bristol-Myers Squibb	Ampicillin	N/A
Lonza	L-glutamine	BE17-605E
	Pen-Strep	DE17-602E
Merck	2-propanol	33539-1L-M
	Ethylenediaminetetraacetic acid (EDTA)	1.80417.0250
	Imidazole	1.04716.1000
	Kanamycin	1.05177.0025
	Potassium dihydrogen phosphate (KH_2PO_4)	529568
	Tryptone	107213
New England Biolabs	1x SOC Outgrowth medium	B9020S
Riedel-de Hëin	Bromophenol blue	32712
Roche	50x C0mplete EDTA free protease inhibitor (tablet)	11 873 580 001
	X-tremeGENE™ 9 DNA Transfection Reagent	06 365 809 001
Sigma-Aldrich	2-amino-hhydroxymethyl-propane-1,3-diol (TRIS base)	T1503-1KG
	5,5-dithiobis(2-nitrobenzoic acid) (DTNB)	D8130-25G
	Acetyl-CoA	10101893001
	Acrylamide/bisacrylamide	A3699
	Agar	A1296
	Cycloheximide	C4859
	Dithiothreitol (DTT)	43816-250ML
	Dulbecco's Modified Eagle's medium (DMEM)	D6546-500ML
	Fetal Bovine Serum (FBS)	F752
	Glycerol	G5516-500ML
	ICEPAL® CA-630 (NP-40)	18896-100ML
	N,N,N',N'-tetramethylethane-1,2-diamine (TEMED)	T9281
	Potassium chloride (KCl)	P9541-1KG
	Sodium chloride (NaCl)	31434-1KG
	Sodium dodecyl sulphate (SDS)	74255-250G
	Yeast extract	89526
Thermo Scientific	PageRuler™ Plus Prestained Protein Ladder	26619
	Gibco™ Trypsin	15400-054
	Gibco™ Opti-MEM®	31985-070
VWR Chemicals	Ethanol 100%	20821.330

3.5 Commercial kits

Table 3.4 Commercially available kits employed in various methods

Supplier/ Name	Cat. no.	Use
Applied Biosystems/ Big Dye v3.1 Cycle Sequencing kit	4337456	Verification of correct mutation after mutagenesis
Macherey-Nagel/ Nucleobond® Xtra Midi	740410.100	Purification of plasmid DNA
New England Biolabs/ Q5 ® Site-Directed Mutagenesis Kit	E0554S	Introduction of specific point-mutations into plasmid vector
Thermo Scientific/ Micro BSA™ Protein Assay Kit	23235	Determining concentrations of purified protein
Thermo Scientific/ SuperSignal™ West Pico PLUS Chemiluminescent Substrate	34080	Chemiluminescence for Western blots
Thermo Scientific/ GeneJet Plasmid Miniprep kit	K0503	Purification of plasmid DNA

3.6 General materials and equipment

Table 3.5 Miscellaneous consumable material

Supplier	Name	Cat. no.	Use
Bio-Rad	Mini-PROTEAN® TGX Stain-free™ Precast Gels, 8-16% 10/15 wells	4568104/4568106	SDS-PAGE
GE Healthcare	Amersham™ Protran™ 0.2 µm NC	10600001	Western blot
	Chromatography paper 3MM CHR (10 cm x 100 m)	3030-672	
Merck	Amicon® Ultra 15 mL filter unit (30 kDa)	PRO4192	Concentrating purified protein
Thermo Scientific	Nunclon™ Delta Surface	167008	96-microwell plate

3.7 Instruments

Table 3.6 Instruments and machines used

Supplier	Name	Use
BioChrom	Ultraspec 10	Spectrophotometry
Bio-Rad	GelDoc™ EZ imager ChemiDoc XRS+ Power-Pac HC	Protein and DNA gel analyses Western blot membrane analyses SDS-PAGE and Western blot
Eppendorf	Mastercycler gradient	PCR
GE Healthcare	ÄKTA Pure	Protein purification
Saveen Werner	NanoDrop ND-1000	Absorbance measurements of DNA and protein
SLM Aminco.	FA-078FI French Press	Hydraulic laboratory press
Thermo Scientific	FRENCH® Press pressure cell 40K	Bacteria lysis
TECAN™	Infinite M Nano	Spectrophotometry
VWR	Microflex	Protein purification, IMAC

Table 3.7 Centrifuges

Supplier	Name
Eppendorf	Centrifuge 5415 R AG 5811
Sorvall®	RC 5B Plus
Thermo Scientific	Heraeus Fresco 17

3.8 Peptides

Table 3.8 Peptides used in *in vitro* DTNB-based acetylation assay

Sequence	Protein of origin
[H] EEEIAAL RWGRPVGRRRRPVRVYP [OH]	γ-actin
[H] DDDIAAL RWGRPVGRRRRPVRVYP [OH]	β-actin
[H] SESSSKS RWGRPVGRRRRPVRVYP [OH]	HMGA1
[H] MLGPEGG RWGRPVGRRRRPVRVYP [OH]	hnRNP F

3.9 Protein purification columns

Table 3.9 IMAC and SEC purification columns

Name	Cat. no.	Supplier
HisTrap™ HP 5 ml	17-5248-01	GE Healthcare
HiLoad™ 16/600 Superdex™ 200 pg	45-002-490PM	GE Healthcare

3.10 Plasmids

Table 3.10 MBP-NAA10 and NAA10-V5 plasmid constructs

Name	#	Protein	Properties	Expression	Supplier
pcDNA 3.1V5His TOPO-*NAA10*	01	hNAA10-V5-His	Ampicillin resistance, 6x His-tag, C-terminal V5	Mammalian	Invitrogen/ T. Arnesen[2]
pETM41-*NAA10*	67	MBP-His-hNAA10	Kanamycin resistance, 6x His-tag, N-terminal MBP-tag	Prokaryotic	G. Stier, EMBL/ R. Evjenth[44]
pcDNA 3.1V5His TOPO-*NAA10*-C21G	474	hNAA10-Cys21Gly-V5-His	Ampicillin resistance, 6x His-tag, C-terminal V5	Mammalian	Made as part of this thesis
pcDNA 3.1V5His TOPO-*NAA10*-R83H	745	hNAA10-Arg83His-V5-His	Ampicillin resistance, 6x His-tag, C-terminal V5	Mammalian	Made as part of this thesis
pETM41-*NAA10*-C21G	746	MBP-His-hNAA10-Cys21Gly	Kanamycin resistance, 6x His-tag, N-terminal MBP-tag	Prokaryotic	Made as part of this thesis
pETM41-*NAA10*-R83H	747	MBP-His-hNAA10-Arg83His	Kanamycin resistance, 6x His-tag, N-terminal MBP-tag	Prokaryotic	Made as part of this thesis

Number in plasmid library

3.11 Primers

Table 3.11 NEBaseChanger generated plasmids used in site-directed mutagenesis

Name	Sequence (5'-3')	T_m (˚C)	Supplier
NAA10 T61G p.C21G fwd.	AATGCCTTCCCCACAGCCC	69	Sigma-Aldrich
NAA10 T61G p.C21G rev.	GGTGGCAGGCAGCCGTCT	69	
NAA10 G248A p.R83H fwd.	CAACCTCCTCGGCCTGCCCGA	69	
NAA10 G248A p.R83H rev.	CAGTGCTGCATGTTCATTAGGTC	69	

3.12 Software

Table 3.12 Desktop programs

Supplier	Name	Use
Bio-Rad	Image Lab™ 6.0.1	SDS-PAGE analyses
Nanodrop Technologies Inc.	ND-1000 v3.5.2	Protein and DNA concentration determination
GE Healthcare	Unicorn v.7.1	Collection of spectrophotometric data from SEC purification
GSL Biotech LLC	SnapGene v.4.0	Analyses of sequenced DNA
Jalview	Jalview Desktop 2.10.5	Analysis and visualization of MSA
Perkin-Elmer	ChemDraw Professional 16.0	Illustrations of chemical reactions
Schrödinger	PyMOL v.2.2.0	Structural analyses of proteins
TECAN	I-control 2.0	Spectrophotometry of microwell plate

Table 3.13 Online resources

Name	URL	Use
Clustal Omega	https://www.ebi.ac.uk/Tools/msa/clustalo/	MSA
MutationTaster	http://www.mutationtaster.org/	Mutant prediction
PolyPhen-2	http://genetics.bwh.harvard.edu/pph2/	Mutant prediction
SIFT	http://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html	Mutant prediction
RSCB PDB	https://www.rcsb.org/	Structural information
UniProt	https://www.uniprot.org/	FASTA sequences

4 Methods

An overview of the general workflow of the project is outlined in Figure 4.1.

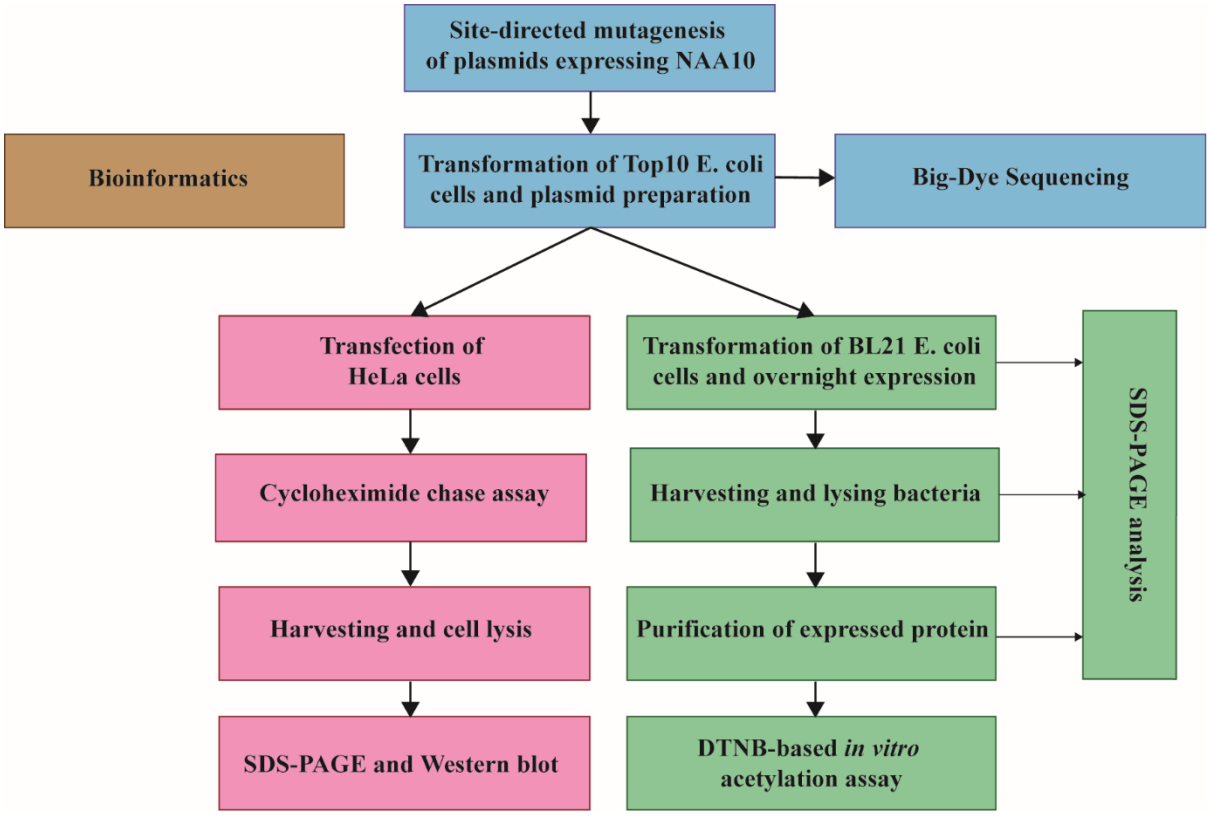

Figure 4.1 The general workflow, detailing the main steps and procedures of this project. *Initial DNA work (blue) was performed to generate mutated constructs. These constructs were then introduced to bacteria, where the in vitro activity was assessed (green); then they were transfected into HeLa cells, where the stability of the protein was investigated (pink). Bioinformatic analyses was conducted throughout the study.*

4.1 Bioinformatics

The amino acid conservation, probability for disease-causing properties and an overall structural assessment of the studied NAA10 mutations, C21G and R83H, was investigated by employing bioinformatic methods. Multiple sequence alignment (MSA) and the use of *in silico* prediction tools was used to provide insight in the degree of C21 and R83 conservation and probability of adverse consequences in affected individuals. A structural visualization of the microenvironment surrounding the mutations, was conducted to investigate possible effects on weak interactions. Similarly, APBS analyses, of the R83-site in particular, was performed to examine possible changes in electrostatic potential.

4.1.1 Multiple sequence alignment (MSA)

Multiple sequence alignment is a highly useful tool that enables the identification of conserved regions by comparing homologous sequences of proteins or genes. In this thesis, the MSA tool, Clustal Omega[82], was used to align the NAA10 protein sequences of *Homo sapiens, Mus musculus, Rattus norvegicus, Xenopus laevis, Caenorhabditis elegans, Arabidopsis thaliana, Saccharomyces pombe and Saccharomyces cerevisiae*. The sequences were retrieved in FASTA format from UniProt[83] (https://www.uniprot.org/) and NCBI[84] (https://www.ncbi.nlm.nih.gov/), and input to the Clustal Omega web program[82] (https://www.ebi.ac.uk/Tools/msa/clustalo/). The results were opened and analysed with Jalview[85].

4.1.2 *Online* in silico *mutation predictions*

This thesis employed three online prediction tools to assess the potential damage caused by the mutation: MutationTaster, PhenoPoly-2 and SIFT. All methods required input of a FASTA sequence or the accession code (PMID, HGNC or Ensembl gene ID, NCBI gene ID, or similar) and an indication of the mutation to be investigated. The entry initiated a calculation process carried out by the programs' algorithms, eventually leading to a summarized document revealing the predictions.

MutationTaster employs various analyses tools to the input, gathers biomedical information from databases and feed the outcomes to a specifically trained naïve Bayes classifier[86]. The classifier then compiles a summary of scores implicating the likelihood that the mutation leads to disease based on amino acid conservation, splice site alterations, lost or retained protein features, and potential truncation of the protein[86]. Three scores from the summaries of NAA10 mutations C21G and R83H are included in this study: The probability that the mutation is disease-causing, The PhastCons score and the PhyloP score. The probability that the input mutation has disease causing properties, is indicated in a score from 0 to 1, where a score close to 0 more confidently predict that the mutation is a "harmless polymorphism", i.e. benign; a score close to 1, more confidently predict that the mutation is "disease-causing". The PhastCons and PhyloP scores are indicative of the amino acid conservation, and are based on the MSA of 36 different species[87]. PhastCons assigns scores from 0 to 1, where 0 is attributed non-conserved, and 1 to highly conserved residues; it also factors in the conservation of flanking amino acids. The PhyloP score, on the other hand, is not only based on the amino

acid's degree of conservation, but the evolutionary rate (faster or slower than expected under neutral genetic drift) of the amino acid on that site[87]. The residue is assigned a score between -14 and +6, where a negative score indicates rapid evolution, whereas a positive score indicates a slow evolution[87].

The second *in silico* prediction tool applied to the mutations, was the PolyPhen-2. This tool uses an iterative-greedy algorithm to choose sequence- and structure based predictive features, involving a comparison of a wild type allele property with the corresponding property of the mutant allele[88]. Further, an MSA of homologous proteins is generated, and the functional implications of the mutation is calculated by a naïve Bayes classifier[88]. From the calculation, the PolyPhen-2 tool compiles a prediction based on a score ranging from 0 to 1. Mutations attaining scores close to 1 are predicted to be "probably damaging", whereas scores close to 0 are predicted to be benign.

The Sorting Intolerable From Tolerable (SIFT) tool was the final *in silico* prediction method applied in this thesis. This method employs sequence homology to assess the adversity of given amino acid substitutions on the overall function of the protein. The general assumption in the SIFT algorithm is that the more evolutionary conserved a region in a protein is, the more likely it is that a mutation in this region will affect the function[89]. The score given input mutations ranges from 0 to 1; the closer the score comes to 1, the higher becomes the degree of certainty with which a mutation can be predicted as "tolerated". Mutations are tolerated as long as they score above the 0.05 threshold; scores below the threshold are predicted to be deleterious[89].

4.1.3 In silico *mutagenesis and electrostatic analyses*

The structures 6C9M[54] and 4KVM[52] were retrieved from the RCSB protein data bank (www.rcsb.org)[90] and uploaded to the molecular visualization program, PyMOL. The 6C9M Chain B structure, i.e. hNAA10, was saved to a separate PDF-file and subsequently aligned with the Chain A structure of 4KVM to investigate the structural relationships between the amino acids around the mutation sites. The alignment allowed an approximate placement of the bisubstrate inhibitor within the hNAA10 binding site, imitating the Ac-CoA-peptide substrate of NAA10. An *in silico* mutagenesis of C21 to glycine and R83 to histidine was also conducted to get a visual impression of how the mutations might affect overall structure and

substrate bonding. Mutated hNAA10 structures were also saved as separate PDB files for subsequent electrostatic analyses.

To calculate the electrostatic potential of the hNAA10 wild type and variants C21G and R83H, the PDB files saved from the 6C9M Chain B were uploaded to the Adaptive Poisson-Boltzmann Solver (APBS) PDB2PQR webserver[91–93] (http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0/). PDB2PQR make preparatory changes to the PDB file by optimizing hydrogen bond, repairing heavy atoms, assessing pKa values and assigning charge and radius parameters[91]. The resulting PQR file can then be solved for continuum electrostatic potential by the APBS software. Output files from the APBS can be visualized in offline molecular visualization programs such as PyMOL.

All three PDB-files were prepared and calculated with the same parameters, optimized for cellular conditions. The molecular mechanical forcefield, Amber[94], and the heuristic pKa calculation software PROPKA[95], was used to generate the PQR in pH = 7.4. Following the PQR preparations, APBS was set to utilize a manually-configured multigrid calculation, where the automatically suggested dimensions were applied. "Ionic strength of the solvent" was set to that of 0.15 M NaCl with ion radius of 1.5 Å. No electrostatic energies were calculated, and the output was set to "write out the electrostatic potential in units of k_bT/e_c (multigrid and finite element)". The APBS was then launched with the remaining parameters unchanged. The output PQR file from the APBS was then opened in PyMOL and visualized in surface view and aligned to S. pombe structure 4KVX[52].

4.2 Mutagenesis, plasmid preparation and DNA sequencing

Studying the impact of the point mutations C21G and R83H on NAA10 protein activity, stability and function, were the focus of this thesis. In order to do so, these mutations were introduced to the wild type *NAA10* gene through Site-Directed Mutagenesis on plasmid vectors pETM-41/His-*MBP-NAA10* (prokaryotic) and pcDNA3.1/*NAA10*-V5-His (mammalian), resulting in four mutated constructs. The constructs were transformed into chemically competent *E. coli* cells optimized for DNA propagation and cloned in overnight incubation.

Cloned DNA was subsequently isolated, concentration was measured, and the mutated sequence was confirmed through DNA sequencing.

4.2.1 Generating NAA10 mutant plasmids using site-directed mutagenesis, transformation and cloning

Site-directed mutagenesis is a PCR-based method in which mutations are introduced to a plasmid by incorporating the mutation, be it substitution, insertion or deletion, into the primer sequences. The Q5® Site-Directed Mutagenesis kit (New England Biolabs) employed in this thesis utilizes back-to-back primes, which has the advantage of generating non-nicked PCR product that is amplified exponentially (Q5® Site-Directed Mutagenesis, user manual). Following the PCR, is a Kinase-Ligase-DpnI reaction step that circularizes the plasmids and digests the non-mutated DNA templates, preparing the plasmids for transformation into bacterial cells optimized for cloning DNA.

DNA can be efficiently transformed into chemically competent bacteria using heat shock transformation. This is a procedure in which DNA and bacteria is mixed and exposed to an abrupt temperature change from 0-42ºC. It is suggested that the change in temperature decreases the membrane potential of the bacteria, which could facilitate the entry of the foreign DNA into the bacteria[96].

Mutagenesis, transformation and cloning procedures

Site-directed Mutagenesis was performed using the Q5® Site-Directed Mutagenesis kit (New England Biolabs) scaled down by half relative to the manufacturer's protocol. Back-to-back primers (designed using the NEBaseChanger tool) containing either of the mutations as well as 12.5 ng plasmid, were used in a PCR amplification. Apart from an extended elongation time at 3 min (~32 sec/kb) the cycling conditions complied with protocol instructions. The amplified plasmid was subsequently added to a Kinase-Ligase-DpnI (KLD) enzyme mix. 25 µL NEB® 5-alpha Competent *E. coli* (New England Biolabs) or 50 µL One Shot® TOP10 (Invitrogen) cells were then subjected to heat shock transformation with the mutated plasmids. Cells recovered for 1 hour with 250 µL SOC medium (New England Biolabs) (Materials, section 3.3.1) at 37ºC whilst gently shaking (200 rpm). 50 µL recovered cells were spread on LB-agar (Materials, section 3.3.1) dishes infused with appropriate selection drug (100 µg/mL ampicillin for pcDNA3.1-plasmids, 30 µg/mL kanamycin for pETM41-plasmids) and placed in 37ºC

overnight incubation. Bacterial colonies formed from successfully transformed *E. coli* cells were picked and inoculated in 1 mL LB-media (Materials, section 3.3.1) containing either kanamycin (30 µg/mL.) or ampicillin (100 µg/mL), depending on the resistance gene present in the vector. Inoculation of bacteria transformed with wild type plasmids was also performed using a small portion of culture stored in glycerol at -80ºC. Cultures incubated overnight at 37 ºC whilst shaking at 250 rpm.

4.2.2 Plasmid DNA isolation

DNA was isolated from the bacteria using a commercial kit based on the 1979 extraction method developed by Birnboim and Doly[97], in which bacteria are treated with lysozyme, SDS and NaOH to disrupt the cell wall and plasma membrane, and denature DNA. Upon the addition of neutralizing reagent sodium acetate, DNA renatures and form insoluble aggregates. Neutralization also leads to the precipitation of protein-SDS complexes and high-weight RNA, which effectively can expelled from the mixture by centrifugation[97]. The kit used in this thesis is supplied with RNase, thus eliminating RNA early in the extraction procedure.

Plasmid preparation procedure

Plasmids were isolated using the GeneJet Plasmid Miniprep Kit (Thermo Scientific). Resuspension of bacteria pellets was done with 125 µL Resuspension Solution; 125 µL Lysis Solution was added to resuspended bacteria; lysed cells were then neutralized with 175 µL Neutralization Solution, and the plasmid DNA was released from the spin column membrane with 30 µL Elution Buffer. Otherwise, the purification was carried out according to the protocol provided by the manufacturer. Additionally, plasmids containing the *NAA10* wild type were isolated according to the NucleoBond Xtra midi, plasmid purification protocol. DNA concentrations were measured using NanoDrop.

4.2.3 Verifying mutation by DNA sequencing

Verification that the correct mutations had been inserted was accomplished through DNA sequencing using the Big Dye Terminator v3.1 Cycle Sequencing kit. This kit is based on Sanger sequencing[98,99], where a PCR of the desired DNA sequence is run with chain-terminating nucleotides, dideoxynucleotides (ddNTPs), resulting in DNA strands of random lengths ending in a ddNTP. Each of four ddNTPs (ddATP, ddGTP, ddCTP and ddTTP) retain

a characteristic fluorescent dye, so that when the DNA strands are separated in capillary electrophoresis, the terminating base and its position in the DNA sequence can be identified[100].

A sequencing of the newly generated plasmids using the Big Dye terminator v3.1 sequencing kit was performed to verify that the correct mutations had been made. A mix of the components listed in Table 4.1 was run with the PCR program detailed in Table 4.2. Subsequently, 10 µl MilliQ water was added to each reaction prior to analysis by the sequencing facility at the University of Bergen.

Table 4.1 Big Dye terminator v3.1 reaction mix

Components	Amounts
Big Dye	1.0 µL
Sequencing buffer	1.0 µL
Primer	3.2 pmol
Template	70-400 ng
MilliQ H_2O	To total volume of 10 µL

Table 4.2 Big Dye terminator v3.1 PCR program

Steps	Cycles	Temperatures	Time
1	1	96˚C	5 min
2	25	96˚C	10 sec
		50˚C	5 sec
		60˚C	4 min
3	1	4˚C	∞

4.3 Expression of recombinant MBP-NAA10 variants

The pETM41-His-*MBP-NAA10* constructs were generated in order to produce high amounts of protein so that the activity of mutated monomeric NAA10 fused to maltose binding protein (MBP) could be studied *in vitro*. To this end, the constructs were transformed into BL21 Star™ DE3 competent *E. coli* cells optimized for protein expression; upon induction, the cells incubated overnight. Culture samples pre- and post-induction were taken for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis.

4.3.1 Inducing His-MBP-NAA10 expression using the T7 promoter expression system

His-*MBP-NAA10* was expressed in BL21 Star™ DE3 competent *E. coli* cells, adapted for high-level expression, using the T7 promoter expression system. This expression system is activated by inducing the expression of the lac-operon controlled T7 RNA polymerase with an

allolactose analogue, such as IPTG[101]. Once the T7 RNA polymerase is active, genes downstream of T7 promoter sites will become expressed. Since the T7 promoter site is included in the pETM42 vector plasmid sequence, the induction of the T7 RNA polymerase will ultimately lead to the expression of His-*MBP-NAA10*. The advantage of using this expression system rather than the E. coli RNA polymerase, is that the T7 RNA polymerase synthesizes RNA at a much higher rate and terminates the transcription less frequently[101].

NAA10 is in this study expressed as a recombinant protein consisting of a 6xHis-tag (His) and a maltose binding protein (MBP). The N-terminal His-tag provides exposed imidazole moieties that enables the adhesion of MBP-NAA10 to the IMAC column. To avoid aggregation and improper folding of NAA10 in *E. coli*, it is expressed in recombination with MBP, which has been found to enhance protein solubility and promote proper protein folding of its fusion partners[102].

4.3.2 Sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE is a long-standing electrophoretic analysis technique[103] that permits the separation of proteins in a sample based on their molecular weight, in a polyacrylamide gel. This is made possible by denaturing proteins by heat exposure and treatment with the detergent Sodium-dodecyl sulphate (SDS); the heat will increase molecular movement, allowing SDS to associate with the peptide bond, which leads to protein linearization and the saturation of the protein in negative charge. To reduce disulphide bonds and disrupt quaternary structure, reducing agents are often added to the samples as well. Denatured proteins are then loaded onto the polyacrylamide gel and placed in an electric field, where proteins will migrate toward the anode; larger proteins will spend longer on migration, whereas smaller proteins migrate faster. Separation time, and therefore quality of the separation for the protein of interest, is chiefly determined by the acrylamide concentration in the gel. Higher concentrations give a more compact gel, which is ideal for separation of small proteins, whereas lower concentrations are more ideal for large proteins

Expression and SDS-PAGE analysis procedure

Applying the heat shock transformation method, ~100 ng isolated pETM-NAA10 mutant constructs were added to 50 µL BL21 Star™ DE3 competent *E. coli* cells (Invitrogen) and

placed in water at 42˚C for 30 sec. Subsequently, the cells incubated on ice for 5 min. 250 µL SOC Outgrowth medium (New England Biolabs) (Materials, section 3.3.1) was added to the cells, which then recovered at 37˚C whilst gently shaking (220 rpm) for 1 hour. 10 µL recovered cells were then added to 5 mL LB-medium (Materials, section 3.3.1) containing 30 µg/mL kanamycin, which incubated overnight at 37˚C, shaking at 220 rpm.

The overnight culture was transferred to 100 mL LB-medium containing (30 µg/mL) kanamycin, which incubated at 37˚C and shaking at 250 rpm. Every 30 min, the optical density (OD), λ = 600 nm, of the culture was measured until it reached an OD of 0.6. Then, IPTG was added to the culture for a final concentration of 1 mM, and the culture incubated at 37˚C and 250 rpm overnight. Next day, bacteria were harvested by centrifugation at 3202 *x g* and 4˚C for 20 min, and subsequently stored at -20˚C until further use. Samples of 200 µL were taken out pre- and post-induction to be analysed on SDS-PAGE.

The culture samples from the pre- and post-induced cultures were mixed with protein sample buffer (Materials, section 3.3.5) to a final concentration of 1x. Samples then incubated at 95ºC for 5 min, before being loaded onto a Mini-PROTEAN® TGX Stain-free™ precast gel (Bio-Rad). The electrophoresis ran for 5 min at 100 V and then for 35 min at 200 V. Subsequent imaging and analysis was done using the Bio-Rad GelDoc™ EZ imager and the Image Lab™ 6.0.1

4.4 The harvest and purification of MBP-NAA10 from *E. coli*, using French® Press and two-step chromatography

For the assessment of MBP-NAA10 activity, it was important to attain a purified protein product as free of contaminants as possible. First, the proteins were harvested from the *E. coli* cells; the French® Press method was used to lyse the cells and generate a lysate. Clarified lysate was then loaded onto an immobilized metal ion chromatography (IMAC) column, separating proteins based on their ability to bind the matrix, and finally, eluted proteins from the IMAC was loaded onto a size exclusion chromatography (SEC) column where the proteins were sorted based on size. The harvest- and purification process were analysed on SDS-PAGE, as with samples collected in the expression.

4.4.1 French® Press applies pressure to disrupt cells

Following His-MBP-NAA10 expression, the proteins were harvested from the in BL21 Star™ DE3 competent *E. coli* cells using the FRENCH® Press method[104]. The method requires two components, the Laboratory Press and the French Pressure Cell[105] (Figure 4.2). The Pressure Cell is composed of three main parts: A piston (Figure 4.2, I), a cell body (Figure 4.2, II), and the closure plug (Figure 4.2, III), which seals desired cellular suspension within the cell body. The Laboratory Press applies hydraulic pressure to the piston, which forces the suspension through a narrow outlet in the closure plug. By increasing the pressure in the cell body chamber, the intracellular pressure in the suspension is also increased; the difference between internal and atmospheric pressure when cells escape through the outlet, causes the cell membrane to burst.

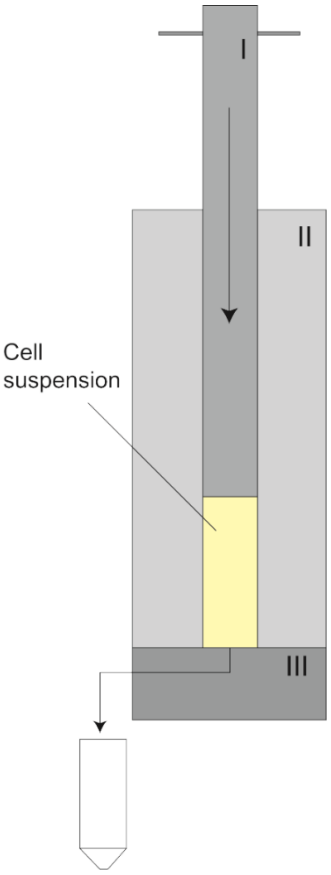

Figure 4.2 Schematic overview of the 40K FRENCH Pressure Cell. *The FRENCH Press method relies on the combined effort of the Laboratory Press and the Pressure Cell for cell lysis. First, the piston (I) is inserted into the cell body (II) and propped in an upside-down position. Then, a bacteria culture (yellow) of maximum 40 mL is poured into the cell body. When the closure plug (III) seals the chamber containing the culture, the Pressure Cell can be turned the correct way and placed in the Laboratory Press. Hydraulic pressure is then applied on the piston, increasing the internal pressure in the cell body and the intracellular pressure of the bacteria. When the cells escape through the narrow outlet in the closure plug, the difference between the intracellular and the external pressure rapture the cells, releasing cytoplasmic content into the suspension.*

French® Press procedure

The bacteria pellet was resuspended in 33.3 mL *E. coli* lysis buffer and transferred to a 40K French Pressure Cell to which hydraulic pressure was applied, yielding a bacterial lysate from the cell suspension. The lysate was then clarified by centrifugation at 20 000 *x g* and 4˚C for 25 min.

4.3.2 Two-step chromatography: Protein separation based on specific affinity and size

The chelation between metal ions and amino acids is an important feature, often required for certain proteins to function[106]. This affinity, that certain amino acid residues such as histidine, cysteine and tryptophan, retains toward some divalent transition metal ions, is utilized in the IMAC method to purify a protein of interest[107] (Figure 4.3). Prior to IMAC purification, the protein of interest is therefore often expressed in recombination with a histidine-tag, which improves the association of the protein with the metal-bound resin. To eventually release the protein of interest from the column, the addition of either a chelating agent or a competitive substance can be supplied, but a reduction in pH will also do the trick. Co-purification of non-specifically binding proteins can be reduced by washing the column with buffer of low imidazole concentrations prior to elution[107].

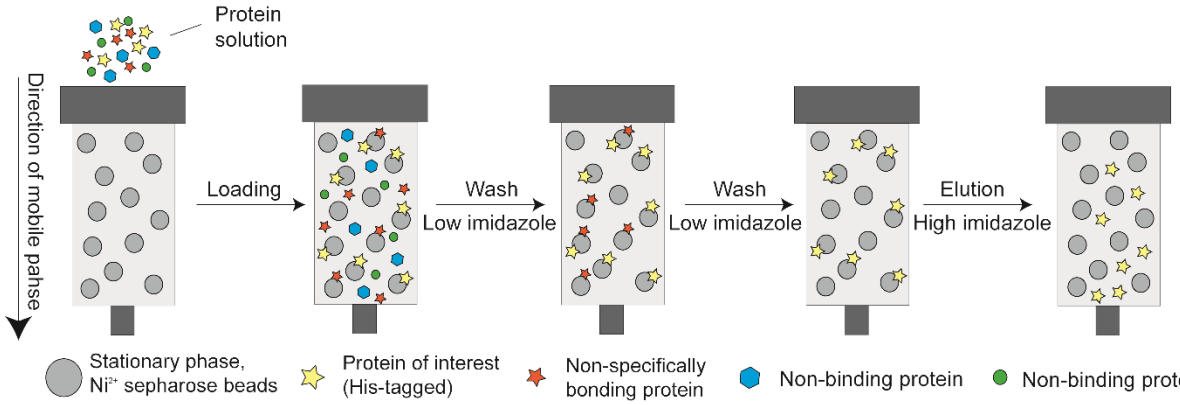

Figure 4.3 Exposed imidazole groups on proteins interact with stationary phase of IMAC. *A bacteria culture overexpressing the protein of interest, MBP-NAA10, was lysed, and then clarified lysate was added to a HisTrap HP column packed with Ni^{2+}-Sepharose beads. Proteins with exposed histidine residues, or His-tagged proteins, will interact with the beads, as the Ni^{2+} ions retain a high affinity for imidazole. After the protein solution is added to the column, a washing buffer containing low levels of imidazole will gradually remove proteins with non-specific interaction toward the matrix, so that only the recombinant his-tagged protein of interest remains attached. The his-tagged proteins are eventually eluted using a buffer high in imidazole.*

SEC, which separates proteins based on size (Figure 4.4), is often used as the secondary purification method, after IMAC, in two-step purifications. This ensures the elimination of co-purified proteins from the IMAC. The stationary phase of the SEC column consists of porous beads made from cross-linked agarose and dextran; the exterior and interior structures of these beads interact with small proteins in the solution, making them migrate slower. Large proteins, on the other hand, will be too massive to interact with the column and thus migrate faster.

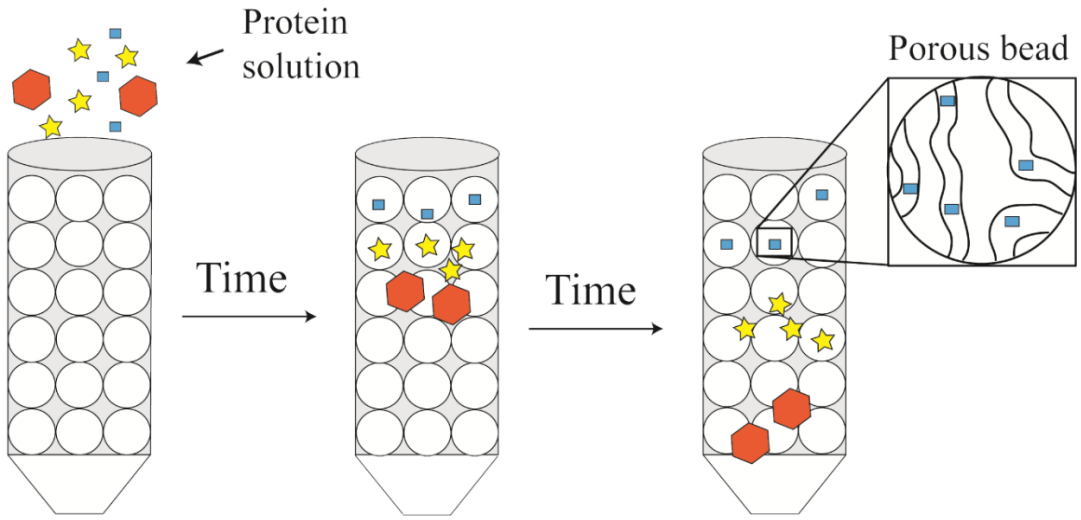

Figure 4.4 Size exclusion chromatography separates proteins based on size and shape. *The proteins eluted from the IMAC purification was subsequently loaded onto a Superdex200 column. The stationary phase of the Superdex column consists of beads made from cross-linked agarose and dextran, providing a porous matrix that smaller proteins can enter, thus delaying their migration. Proteins too large to enter, or otherwise interact with, the porous beads are eluted with the mobile phase. Consequently, larger proteins are eluted early, and small proteins late in the SEC run.*

Two-step purification procedure

Clear lysate was added to a HisTrap column equilibrated with three column volumes of IMAC wash buffer (Materials section, 3.3.2) using a VWR Microflex peristaltic pump. Subsequently, the IMAC column was washed with five column volumes (CV) (25 mL) of IMAC wash buffer, before resin-bound protein was eluted by passing three CV of IMAC elution buffer (Materials section, 3.3.2) through the column. The eluted fractions containing the most protein were then pooled and concentrated by centrifugation at 3202 *x g* and 4˚C for 15 min in an Amicon® Ultra 15 mL 30 kDa MWCO-filter.

The concentrated fractions from the IMAC purification was added to a 120 mL Superdex 200 chromatography column equilibrated with 1.1 CV of NAA10 gel filtration buffer (Materials

section, 3.3.2). The software monitoring the purification process, indicated the fractions in which the most protein was retained. Samples of these fractions were analysed on SDS-PAGE to investigate the quality of the purification and assess molecular size of the proteins. Based on purifications carried out in previous studies of NAA10, the expected elution volume of MBP-NAA10 was at 70 mL. Fractions collected eluted around this volume were pooled and concentrated (as fractions in the IMAC purification). 500 µL concentrated protein was added glycerol to a 50% concentration and stored at -20ºC; the rest was stored at 4ºC.

4.5 The determination of MBP-NAA10 concentration and *in vitro* acetylation activity, using colorimetric assays

Colorimetric techniques are used to measure the concentration of a coloured compound in solution, and they are frequently applied in determining protein and nucleic acid concentrations, and in the assessment of enzymatic activity by, for instance, using a colour generating substrate. To ensure accurate measurements of the *in vitro* acetylation activity of MBP-NAA10, the concentration of the protein was determined using a bicinchoninic acid (BCA) assay, in which peptide bonds turn violet upon contact with BCA. The activity was determined using Ellman's reagent, which reacts with CoA and generates a yellow biproduct.

4.5.1 BCA protein assay and BSA standard curve to determine MBP-NAA10 concentration

After obtaining purified MBP-NAA10, the concentration of the protein was determined using the Micro BCA™ Protein Assay kit (Thermo Scientific). This method is based on the colorimetric detection of the purple coloured chelation between two bicinchoninic acid (BCA) molecules and one Cu^{1+} ion (Figure 4.5). The Cu^{1+} ions are generated from a biuret reaction that takes place between peptide bonds and Cu^{2+} present in the solution[108]. The advantage of employing a method that rely on the association with the peptide bond for concentration determination, as opposed to the amino acid residues, is the limited protein-to-protein variation. In the Bradford protein assay, for instance, the colorimetric detection is based on the ready association of the reagent with arginine and lysine residues[109]. This way, two different protein solutions with the same concentration could potentially be given entirely different concentrations upon measurement.

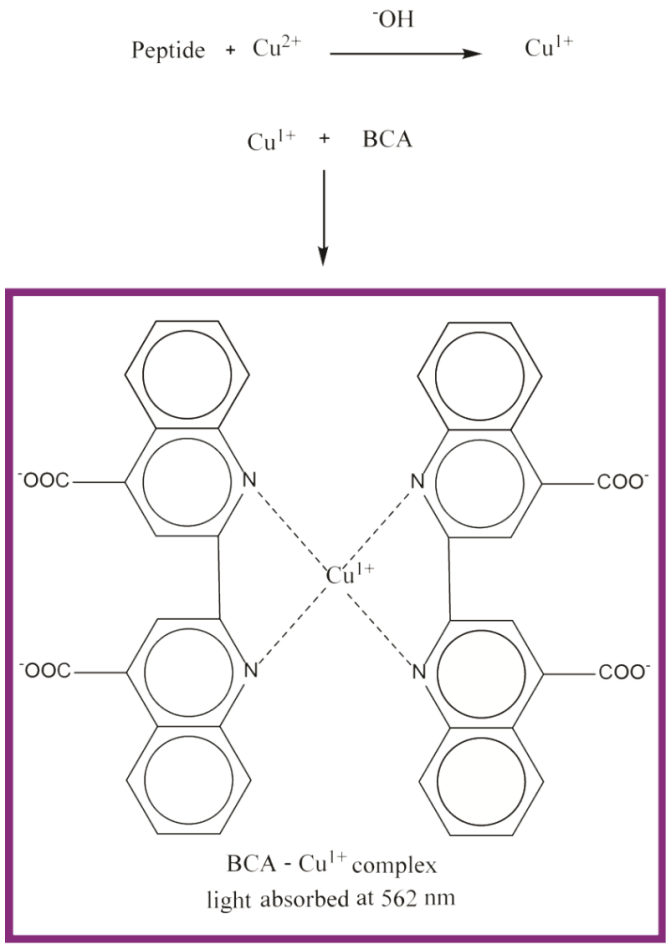

Figure 4.5 Coloured chelation between a cuprous ion and bicinchoninic acid permit protein concentration determination. *The reaction between proteins in alkaline solution and divalent cuprous ions (Cu^{2+}) generate univalent cuprous ions (Cu^{1+}), which then form chelation complexes with bicinchoninic acid. The chelation is purple in colour and absorb light at 562 nm, which enables the determination of protein concentration using spectrophotometry.*

BCA protein assay procedure

A BSA standard curve with concentrations 1000, 500, 250, 100, 50 and 25 µg/mL, was prepared, alongside three MBP-NAA10 samples where MBP-NAA10-C21G samples were diluted 1:10, 1:100, and 1:1000, and MBP-NAA10-R83H samples were diluted 1:2, 1:4 in addition to one undiluted sample. As per the instructions of the Micro BCA™ Protein Assay Kit (Thermo Scientific), a working reagent (WR) was composed of reagents MA:MB:MC in a ratio of 50:48:2. Samples were added WR in a 20:1 ratio, as per the instructions of Thermo Scientific "Protocol for BCA protein assay using a NanoDrop 2000/2000c spectrophotometer" (high-range assay), and incubated at 37˚C for 1 hour. All reactions were measured at λ = 562 nm with NanoDrop. The linear range of the BSA standard curve was used to determine the concentration of protein in the samples.

4.5.2 In vitro *acetylation of MBP-NAA10 variants determined using a DTNB-based acetylation assay*

The protein concentration was subsequently used to calculate the amount of protein required for the reactions in the *in vitro* DTNB-based acetylation assay[110,111]. This assay allows assessment of *in vitro* acetylation activity of an acetyltransferase, by adding Ellman's reagent, 5,5-dithio-bis-(2-nitrobenzoic acid) (DTNB), to a quenched enzyme reaction. DTNB reacts with the by-product of the enzymatic reaction, CoA-SH, forming the yellow substance, 2-nitro-5-thiobenzoate (NTB^-) (Figure 4.6). NTB^- has a maximum absorption at $\lambda = 412$ nm[112], and its concentration can be determined through spectrophotometry. The amount of NTB^- generated will correspond to the number of peptides acetylated by the NAA10 variants, and thus the enzymatic activity.

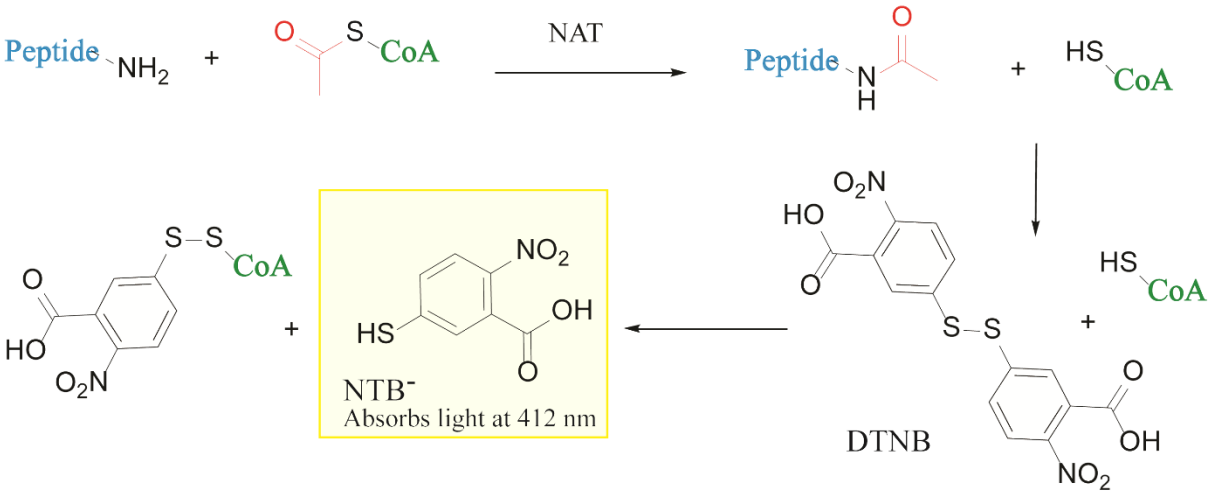

Figure 4.6 The reaction between DTNB and CoA-SH generates the yellow compound NTB^-. *The acetylation of a peptide by an N-terminal acetyltransferase generates CoA-SH as a by-product. Once added, 5,5-dithio-bis-(2-nitrobenzoic acid) (DTNB) reacts with CoA-SH to produce 2-nitro-5-thiobenzoate (NTB^-): A yellow coloured compound that absorbs light at $\lambda = 412$ nm.*

In vitro acetylation assay procedure

Two acetylation assays were performed: A substrate-dependent assay, where the acetylation of the substrate peptides EEEIA, DDDIA, MLGPE and SESSS by mutated MBP-NAA10 was compared with the wild type; and a time-dependent assay, where wild type and mutant reactions with only EEEIA were stopped after 10 minutes, 20 minutes and 30 minutes. All reactions were made from the contents detailed in table 4.3 and was incubated at 37˚C.

Table 4.3 DTNB-based acetylation reaction mix

Components	Initial concentrations	Final concentrations
Acetylation buffer	2x	1x
Peptide	5 mM	300 µM
Ac-CoA	10 mM	300 µM
MBP-NAA10	1 µM	0.1 µM
MilliQ H_2O	-	Until total volume of 50

In the substrate-dependent assay, the reactions were stopped after 20 minutes by the addition of 100 µL quenching buffer. Enzyme was added to the blank reactions after all reactions had been quenched. 25 µL of a saturated DTNB solution in DTNB buffer (Materials, section 3.2.3), was added right before measuring the absorbance at λ = 412 nm in a 96 microwell plate. The absorption (A) and extinction coefficient[112] (ε) of NTB^- (13600 $M^{-1}*cm^{-1}$) was used to solve the Beer-Lamberts law (Eq. 1) for NTB^- concentration (C). The spectrophotometer employed for the measurements used a corrected pathlength (L), in which the measurements for water absorption (λ = 977 nm) and background absorption of plastic (λ = 900 nm) were used to establish the actual distance from the light source to the sample.

$$A = \varepsilon * C * L \qquad \text{Eq.1}$$

$$C = \frac{A}{(\varepsilon * L)}$$

4.6 Maintenance, transfection and stability analyses of NAA10 variants in HeLa cells

To investigate how the mutations affected cellular stability of the NAA10, wild type, C21G and R83H versions of pcDNA3.1/*NAA10*-V5-His constructs were transfected into HeLa cells. The stability was tested by cycloheximide chase assay, where cells were treated with cycloheximide and then harvested after specific time intervals. Harvested cells were lysed and protein content analysed by Western blot.

4.6.1 Maintenance of HeLa cells

HeLa cells were cultured in 10 cm dishes with 10 mL DMEM containing FBS (10%), L-glutamine (2 mM) and PenStrep (50 U/mL penicillin, 50 µg/mL streptomycin) at 37ºC and 5%

CO_2. Cell treatment was conducted in sterile environments using a laminar flow (LAF) hood. When the cells reached approximately 90-100% confluency, they were washed with 1x PBS (Materials, section 3.3.4) and treated with 500 µL trypsin for 6-12 min. Detached cells were resuspended in medium appropriate for subsequent use. The most frequently used concentrations for subculturing in 10 cm dishes were 1:10, 1:8, and 1:4; when split to 6-well plates, 1 day prior to transfection, the concentration was 1:3 or 1:4.

4.6.2 Transfection

HeLa cells were split in 6-well plates 24 hours prior to transfection with pcDNA3.1/*NAA10*-V5-His constructs. The transfection reagent was prepared from 3.6 µl XtremeGENE™ and 1.2 µg DNA diluted in 100 µL room temperature Opti-MEM® and incubated for 30 min at room temperature. Meanwhile, Dulbecco's modified eagle medium (DMEM) (Materials, section 3.3.4) was removed from the wells and cells were washed with 1 mL fresh DMEM medium; 3 mL fresh DMEM medium was added to the cells. The transfection reagent/Opti-MEM mix was then added dropwise to the cells. The cells were incubated with the reagent for 48 hours at 37ºC and 5% CO_2.

4.6.3 Cycloheximide (CHX) chase assay

Cycloheximide (Figure 4.7) is a toxic substance produced by *Streptomyces griseus*, which arrests eukaryotic protein synthesis by binding to the E site of the 60S ribosomal unit, thus blocking translation[113,114]. This compound is often used as a protein synthesis inhibitor for studies of eukaryotic cells *in vitro*, where the inhibitory effects are reversed by replacing culture medium[115].

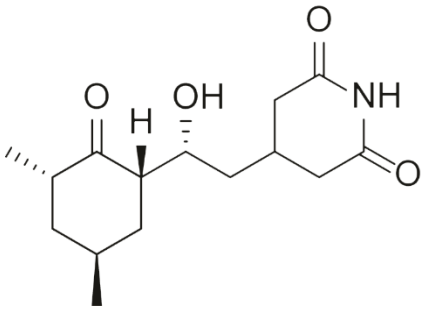

Figure 4.7 Cycloheximide (CHX) inhibits protein synthesis. *By associating with the 60S ribosomal unit E site in eukaryotic cells, thus preventing tRNA units from binding, CHX inhibits the translational elongation of peptides, effectively arresting protein synthesis*[114].

Following transfection of HeLa cells with the wild type and mutated pcDNA3.1/*NAA10*-V5-His constructs, the stability of the NAA10 variants was assessed using a CHX chase assay. In

this assay, CHX is added to halt protein synthesis for various time periods before cells are harvested, lysed and analysed on Western blotting (WB)[116]. The analysis reveals the protein abundance of the various samples; a decrease in protein content over time indicates protein degradation[117].

CHX chase assay and cell harvest procedure

DMEM media in wells were removed and cells were washed with 1 mL fresh medium. Room tempered DMEM medium supplemented with cycloheximide (50 µg/mL) was added to HeLa cells in 6-well plates at different time intervals: 0, 2 and 4 hours. After 6 hours all cells were harvested simultaneously by removing the CHX infused medium, washing cells with 1 mL 1x PBS, detaching cells with 200 µL trypsin for approximately 10 minutes, and suspending cells in 1 mL ice cold FBS-, L-glutamine-, and PenStrep deficient DMEM medium. Cell suspensions were from this point kept on ice. Subsequently, the cells were centrifuged at 1000 *xg* and 4ºC for 5 min; the supernatant was discarded, and cell pellet resuspended in 1mL ice cold 1x PBS. The centrifugation step was repeated, and cell pellets were stored at -80ºC until further use.

4.6.4 Western blotting

The cells harvested after the cycloheximide chase assay were analysed by Western blotting. This technique allows detection of proteins in samples; first, SDS-PAGE (Materials, section 3.2.5) is used to separate the proteins and mark their size, then the polyacrylamide gel is placed against a nitrocellulose membrane in an electric field, with the membrane facing the anode. The proteins, which retain a negative charge from their association with SDS, migrate towards the anode, and will thus adhere to the membrane. To detect the proteins, the membrane must be treated with specialized antibodies presenting features that allow detection, such as fluorescence or chemiluminescence. Horseradish peroxidase (HRP)-conjugated antibodies are frequently used as HRP substrates becomes chemiluminescent upon HPR catalysed oxidation. To prevent non-specific binding of the antibodies, membranes are often incubated with a dry-milk solution prior to the antibody treatment. The proteins in the dry-milk occupies most of the sites where the antibodies potentially could bind. They also have a low degree of immunogenicity, which mean that they will not be target for the antibodies.

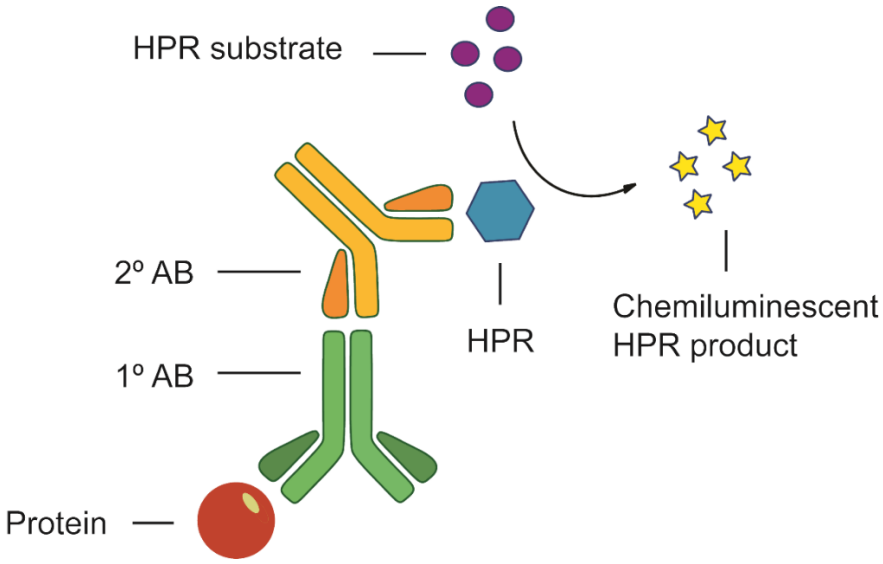

Figure 4.7 Protein detection in nitrocellulose membrane permitted by chemiluminescence. *Proteins on a nitrocellulose membrane can be detected by treatment of specific antibodies. First, a primary antibody (1º AB) recognizes an epitope on the target protein, then a secondary antibody (2º AB) recognizes the primary antibody. Attached to the secondary antibody is a Horseradish peroxidase (HRP), which catalyses the oxidation of a substrate that becomes chemiluminescent. This enables the indirect visualization of the target protein.*

Western blotting procedure

HeLa cell pellets stored at -80ºC were retrieved and resuspended in 100 µL IPH-buffer (Materials, section 3.3.4) supplied with 1x C0mplete EDTA free protease inhibitor. The suspension incubated 30 min on ice before centrifugation at 17 000 *x g*, 4ºC for 5 min. The supernatant was transferred to fresh microcentrifuge tubes, and protein content was separated on SDS-PAGE (Methods, section 4.3.2). Following electrophoresis, the proteins in the gel were transferred to a nitrocellulose membrane in an electroblotting chamber submerged in 1x cold Towbin buffer. The transfer ran at 100 V for 35 min. Subsequently, the membrane was blocked with 5% dry-milk in 1x PBS-Tween (Materials, section 3.3.4) for 1 hour at room temperature whilst gently shaking, before being treated with a primary antibody (Materials, section 3.1) in 1% dry-milk and 1x PBS-Tween at 4ºC overnight.

The following morning, primary antibodies were removed, and the membrane washed with 1x PBS-Tween; first a quick rinse, then 3 x 5 min. The membrane was then added secondary antibody (Materials, section 3.1) in 3% dry-milk and 1x PBS-Tween and placed on shaking for 2 hours at room temperature. After removing the secondary antibody, the membrane was washed with 1x PBS-Tween as done after removal of primary antibody. An additional washing step with 1x PBS for 5 min was carried out before treating the membrane with SuperSignal®

West Pico Chemiluminescent Substrate (Thermo Scientific) working reagent (WR) for 4 minutes. After removing the WR, the membranes were imaged using the ChemiDoc™ XRS+ and ImageLab™ v3.0 (Bio-Rad). The treatment of the nitrocellulose membrane with antibodies and subsequent imaging was conducted two times, first for the V5-tagged NAA10 proteins, using an α-V5 primary antibody (mouse); the second time for the loading control, Pan-actin, using an α-pan-actin primary antibody (mouse) (Materials, section 3.1).

5 Results

5.1 Clinical data

The NAA10 mutations studied in this thesis were derived from the clinical data of three male patients; two with the R83H mutation and one with the C21G mutation. Not much is known of the phenotypic profile of the C21G patient, as clinical data obtained from this case was very limited. All that has been shared so far is that the patient exhibit phenotypic features similar to those identified in previous studies of NAA10 mutations. For the R83H mutation, a more detailed description was provided; the two boys affected by the R83H mutation, had mainly intellectual disability, behavioural abnormalities and poor speech development as common denominators. Patient 1, a teenaged boy with maternally inherited mutation, otherwise presents with developmental delay, mild learning disability, coeliac disease, chronic constipation, hypermobility, a mild concentric left ventricular hypertrophy, microcephaly and astigmatism. His father is also diagnosed with microcephaly and coeliac disease, and has mild learning difficulties. Patient 2 is 12 years old and attained the R83H mutation *de novo* with a 75% mosaic degree. Apart from the features shared with patient 1, he has autism, dystonia and muscle stiffness, and large low-set ears.

5.2 Bioinformatics

Bioinformatic analyses were carried out as part of an initial assessment of the two variants studied in this thesis. Among the methods utilized were multiple sequence alignment (MSA), mutation predictions of protein function and structure, visualization and *in silico* mutagenesis of NAA10, and finally an electrostatic assessment of the alterations that occur in the R83H mutant.

5.2.1 Multiple sequence alignment (MSA)

NAA10 protein FASTA sequences from species *Homo sapiens, Mus musculus, Rattus norvegicus, Xenopus laevis, Caenorhabditis elegans, Arabidopsis thaliana, Saccharomyces pombe and Saccharomyces cerevisiae* were entered into the Clustal Omega multiple sequence alignment tool[82]. The resulting alignment file was then visualized and analysed in Jalview[85] (Figure 5.1) This visualization gave an overview of the conserved regions of the NAA10

primary structure, and an opportunity to evaluate the conservation of the mutated amino acids studied in this thesis. NAA10 C21 (Figure 5.1, red box) lies close to a conserved linker region involved with binding the N-terminus of the substrate protein (Figure 5.1, *). Amino acid substitutions in this region were studied in the 2013 paper by Glen Liszczak and colleagues, and were found to greatly impact the activity of NAA10[52]. C21 itself is not among the very most conserved amino acids in this region and is given a conservation score of 8 by the Clustal alignment. R83 (Figure 5.1, dark blue box) is located in the Ac-CoA binding motif Q/RxxGxG/A (Figure 5.1, purple box), next to the highly conserved R82. R83 itself is given a conservation score of 8, as the unicellular organism *S. pombe* retains a histidine residue in this position of NAA10. The substitution of arginine to histidine is interestingly enough the mutation studied in this thesis.

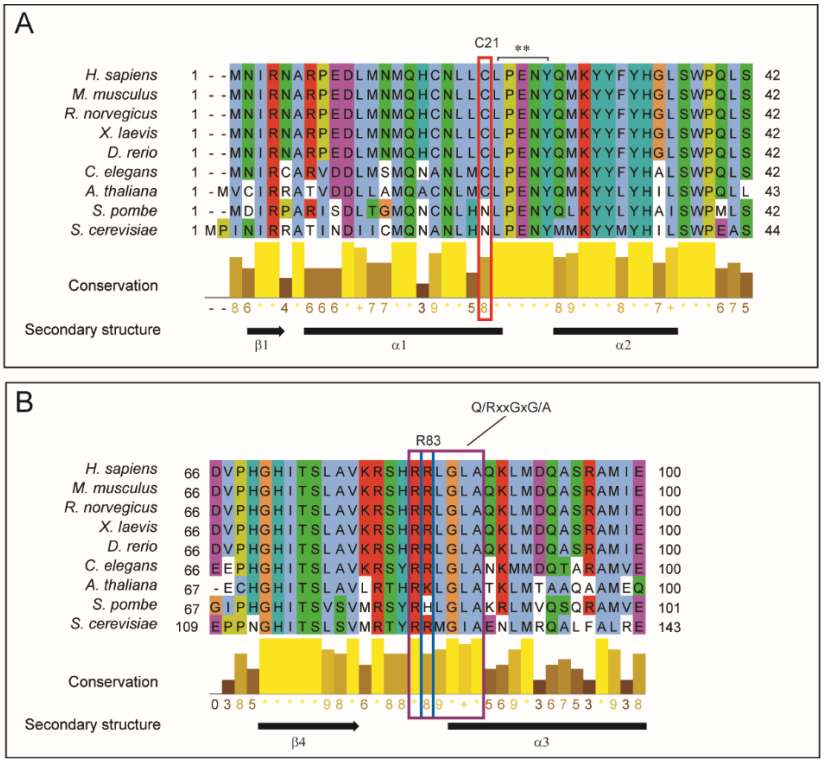

Figure 5.1 MSA reveal conserved regions of NAA10 around mutation sites C21 and R83. *FASTA sequences from the NAA10 protein of species* Homo sapiens, Mus musculus, Rattus norvegicus, Xenopus laevis, Caenorhabditis elegans, Arabidopsis thaliana, Saccharomyces pombe and Saccharomyces cerevisiae, *were entered into the Clustal Omega online MSA software to generate a multiple sequence alignment. C21 (A, red box) is conserved for all multicellular organisms included in the alignment, and resides adjacent to a highly conserved helix-loop-helix region (**). R83 (B, dark blue box) is conserved for all but the* S. pombe *sequence and is located in the conserved Ac-CoA binding motif, Q/RxxGxG/A (B, purple box). The colouring scheme chosen for the alignment is "Clustal" which attributes blue to hydrophobic (Ala, Ile, Leu, Met, Phe, Trp, Val), red to positively charged (Lys, Arg), magenta to negatively charged (Glu, Asp), green to polar (Asn, Gln), orange to glycines, yellow to prolines and cyan to aromatic (His, Tyr) residues. White indicate non-conserved residues. Conservation scoring ranges from 0 to 11, where '+', assigned scores of 10, indicate conserved amino acid properties, and '*', assigned scores of 11, indicate conserved amino acid residue.*

5.2.2 *in silico* disease predictions

Three *in silico* mutation prediction tools, MutationTaster[86], PolyPhen-2[88] and SIFT[89], were utilized to assess whether the genetic mutations studied in this thesis had deleterious effects on the protein. Employing a naïve Bayes classifier, the MutationTaster method is able to predict and score certain properties of an input mutation. Among these properties are the mutated amino acid's conservation, given as PhastCons and PhyloP values (Methods, section 4.1.2), and the general prediction of whether the mutation is disease-causing or benign. MutationTaster predicted both the C21G and the R83H mutations to be disease causing, scoring with a probability close to 1 (Table 5.1). Both amino acid positions C21 and R83 were given PhastCons scores of 1, which indicates a high degree of conservation. C21 gained a PhyloP score of 2.764, a lower score than that assigned R83, which was 5.359.

Table 5.1 MutationTaster prediction of mutations' physiological impact and conservation

	C21G	R83H
Prediction	Disease causing	Disease causing
Prediction probability *	0.9999999984	0.9999999999
PhastCons **	1	1
PhyloP ***	2.764	5.359

*) Prediction probability ranges from 0.0 to 1.0, where scores close to 1.0 predicted the mutation to be disease causing, and scores close to 0.0 predict the mutation to be a harmless polymorphism (not disease-causing).
**) PhastCons scores ranges from 0 to 1, where scores close to 1 indicate a high degree of conservation, and scores close to 0 indicate low degree conservation.
***) PhyloP scores range from -14 to 6, and indicate the rate with which amino acids are mutated at the given site. Negative scores indicate a higher rate of mutations, whereas positive scores indicate a low mutation rate.

PolyPhen-2 predicts, like MutationTaster, whether the mutations entered into the program are physiologically damaging. This is assessed by generating structure- and sequence-based alignments (Methods, section 4.1.2), where properties of the wild type and the mutants are compared. A HumDiv-trained PolyPhen-2 dataset was chosen for the prediction of C21G and R83H. Both mutations were evaluated by the algorithm to be "probably damaging" with probability scores of 0.998 and 1.00, respectively (Table 5.2).

Table 5.2 PolyPhen-2 prediction of mutations' physiological impact

	C21G	R83H
Prediction	Probably damaging	Probably damaging
Prediction probability *	0.998	1.000

*) Prediction probability ranges from 0.000 to 1.000. Scores close to 1.000 indicate a high probability that the mutation is damaging. Scores close to 0.000 will classify the mutation as benign.

The final *in silico* prediction tool employed in this thesis, SIFT (Sorting Intolerant from Tolerant), is based on an algorithm that predicts the functional implications that an amino acid substitution can have on a protein. Probability scores between 0 and 0.05 are characterized as "affecting protein function". Substitutions studied in this thesis, C21G and R83H were both predicted by the SIFT algorithm to be "tolerated" with scores of 0.4 and 0.06, respectively (Table 5.3).

Table 5.3 SIFT predictions of mutations' impact on protein function

	C21G	R83H
Prediction	Tolerated	Tolerated
Prediction probability *	0.4	0.06

*) Prediction probability ranges from 0.00 to 1.00. Scores between 0.00 and 0.05 are predicted to affect protein function. Scores from 0.05 to 1.00 are characterized as tolerated.

5.2.3 *in silico* mutagenesis and electrostatics analysis

The recently solved structure of human NatA[54] was used to obtain a general overview of the amino acid neighbourhoods surrounding the mutated residues discussed in this study. The following structural visualization (Figure 5.2) employs the NAA10 structure from PDB entry 6C9M[54] (human NatA) positioned in the NAA15 structure from PDB entry 4KVM[52] (*S. pombe* NatA). The Ac-CoA structure, along with a short peptide sequence (Ser-Ala-Ser-Glu) obtained from 4KVM, was also included in the structural analysis.

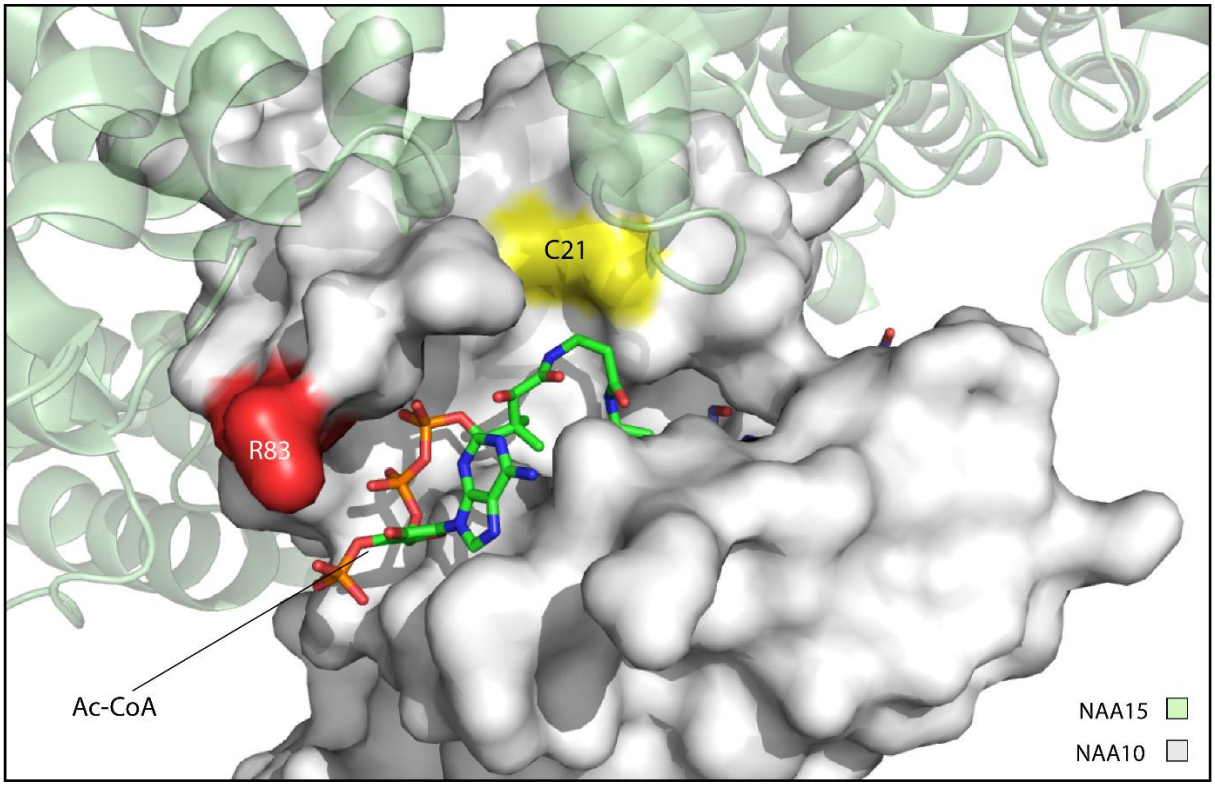

Figure 5.2 Structural visualization of the Ac-CoA binding site of human NAA10. *The structure of human NatA (6C9M) was viewed in the molecular visualization software PyMol and aligned with the structure of 4KVM (*S. pombe *NatA) to situate Ac-CoA in the binding site of human NAA10 (light gray, surface view); the NAA15 structure (pale green, cartoon, reduced visibility) is also taken from the 4KVM entry. Bound to Ac-CoA in 4KVM, is a short peptide sequence, Ser-Ala-Ser-Glu (not visible in this image). Mutation sites, R83 (red) and C21 (yellow) are highlighted.*

An *in silico* mutagenesis was conducted on the human NatA structure (6C9M) in alignment with Ac-CoA and the peptide Ser-Ala-Ser-Glu from 4KVM structure. Here, an image of the amino acid neighbourhoods surrounding the mutation sites C21 and R83 are presented (Figure 5.3). Figure 5.3 A and B show the surroundings of the C21 site; A representing the wild type condition and B an outlook on how C21 mutated to glycine might turn out. Figure 5.3 C and D present the R83 site, where C show the wild type structure, and D gives an impression on how the mutation of R83 to histidine could present. It should be noted that the *in silico* mutagenesis can in no way account for the potential three-dimensional alterations that might accompany such mutations. Only the amino acid has been modified; hydrogen bonds, hydrophobic interactions, and other weak bonds remain as they were before mutagenesis.

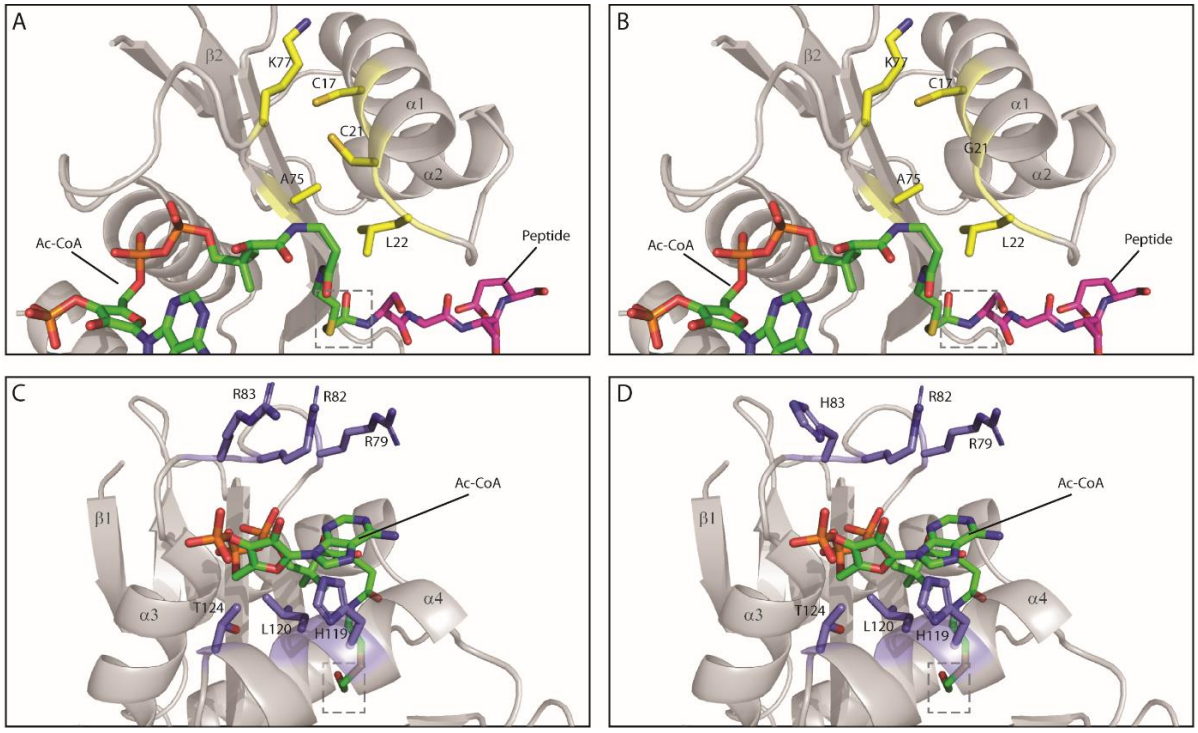

Figure 5.3 Structural visualization of mutation sites and surrounding residues. *The NAA10 structure belonging to the human NatA (6C9M) was viewed in PyMol. An alignment of the NAA10 structure with 4KVM (*S. pombe *NatA) allowed the accompanying structure of Ac-CoA to be situated in the binding site. A short peptide sequence is seen in image A and B, associated to Ac-Coa (the bond between the peptide and Ac-CoA highlighted by stippled box), representing the N-terminus of a protein substrate. A) The wild type condition of the site surrounding C21, B) The mutation of C21 to glycine. C) The wild type condition of the site surrounding R83, D) The mutation of R83 to histidine. Stippled boxes mark the bond between the Ac-CoA and the N-terminus of the peptide substrate.*

Following up on the structural analyses, an assessment of the possible alterations in electrostatic potential for the mutants relative to the wild type was made. In PyMOL, the NAA10 chain from the 6C9M structure was extracted and new PDB files for the wild type, the C21G and the R83H mutants were made. These PDB files were uploaded to the APBS server and converted to PQR files, containing approximate electrostatic data solved using the Poisson-Boltzmann equation[92,118]. The generated PQR files were then opened in PyMOL where the electrostatic potential was analysed. An analysis of the difference in electrostatic potential between the wild type and the C21G mutant is not included here, since this difference was not particularly great and since a change in the electrostatic potential was not believed to be the cause of damage to the protein function. The electrostatics analysis of the R83H mutant, however, indicate notable change from the wild type (Figure 5.4 A), mainly that there is a reduction of positive charge (blue colour) at the mutation site (Figure 5.4 B).

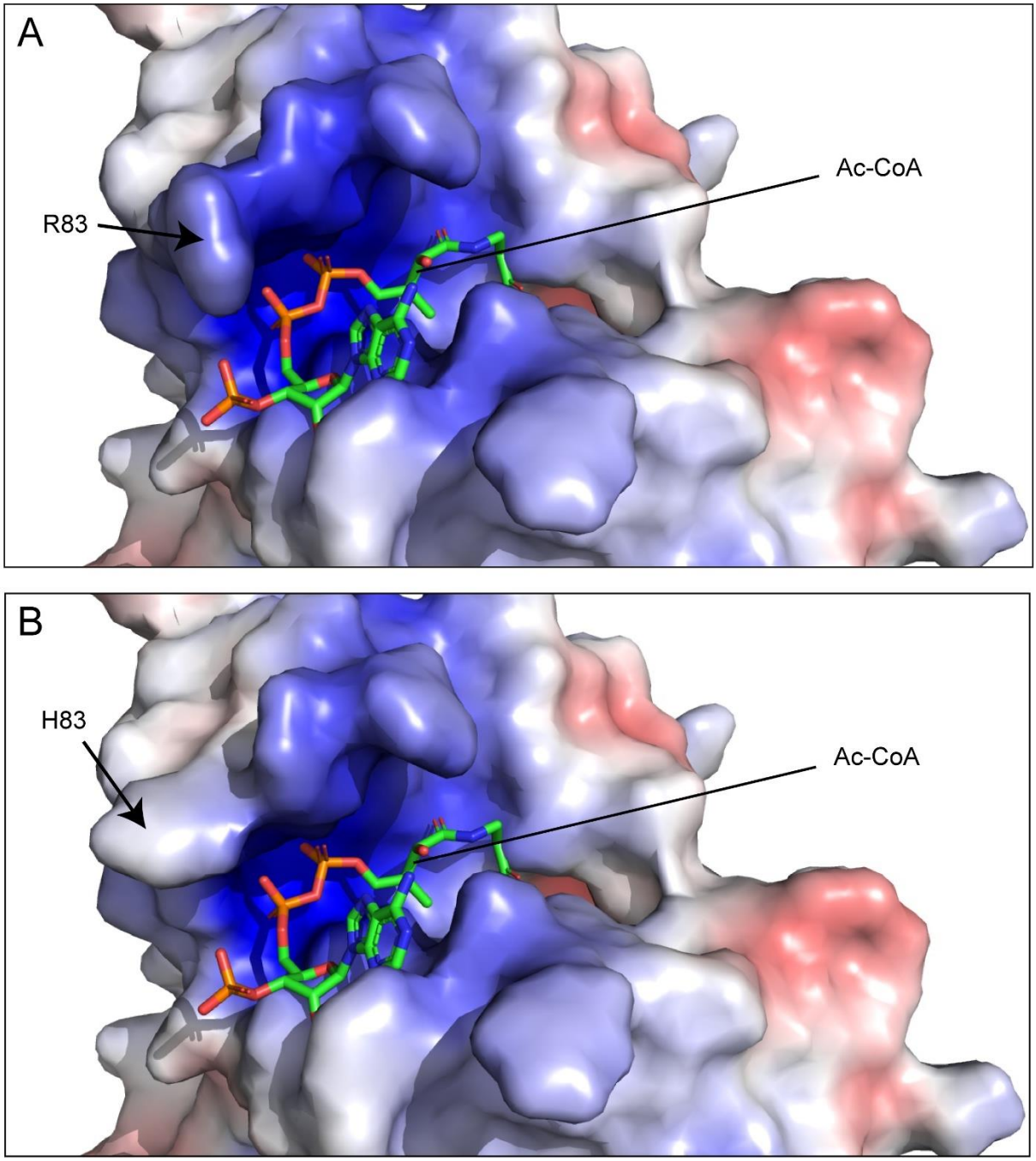

Figure 5.4 The electrostatic potential of the NAA10 Ac-CoA binding site is altered when R83 is substituted with histidine. *The wild type and mutated NAA10 structures of the 6C9M pdb file was uploaded to the APBS server, and converted to a PQR file, detailing the electrostatic potential of each residue, representing it as blue (positive charges) and red (negative charges) in the PyMOL surface view. The 6C9M structure was aligned with the* S. pombe *4KVX structure to visually place Ac-CoA in the binding site. A) Visualized electrostatic potential of the Ac-CoA binding site in wild type NAA10, B) the electrostatic potential of the Ac-CoA binding site in NAA10-R83H.*

5.3 Sequencing of mutated NAA10 expressing plasmid vectors

Following the DNA mutagenesis and preparation (Methods, sections 4.2.1 and 4.2.2), plasmids were sequenced using the Big-Dye terminator v3.1 kit (Methods, sections 4.2.3) and subsequent analysis at the sequencing facilities at the Department of Biosciences (BIO) at UiB. The results, visualized in SnapGene, confirmed that the desired mutations were introduced in all four plasmids (Figure 5.5).

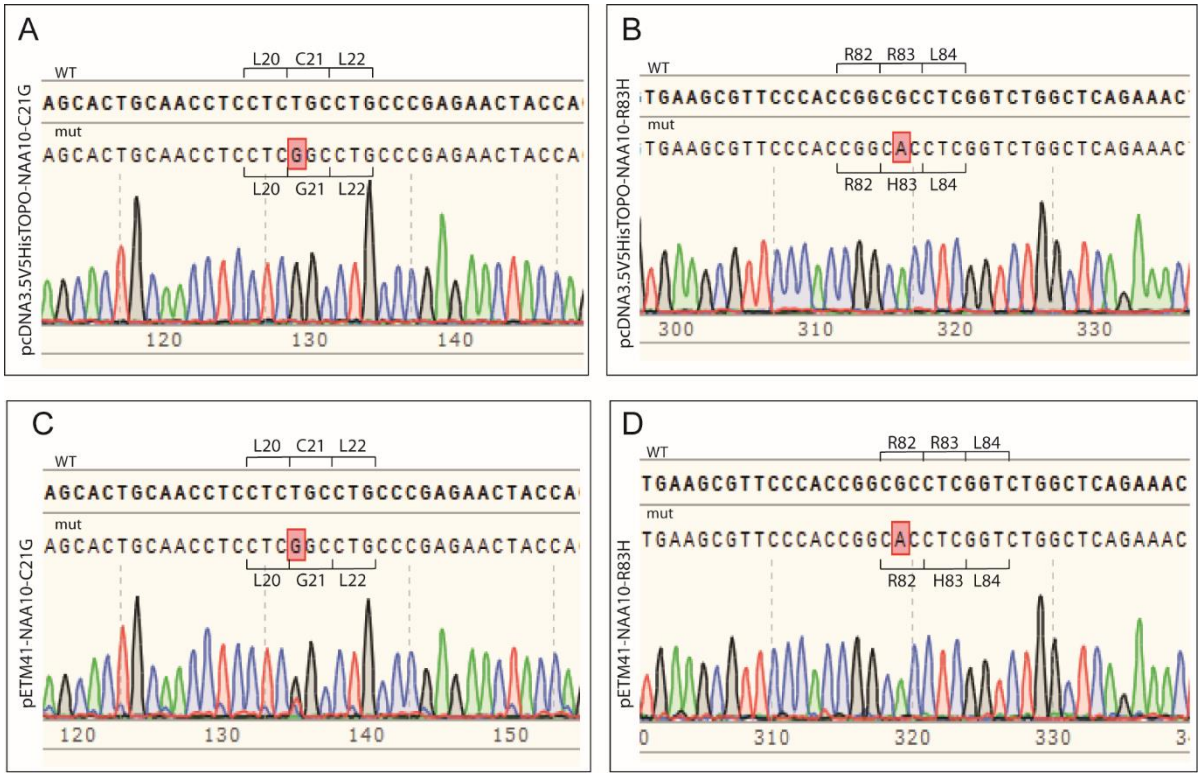

Figure 5.5 The correct mutations in plasmids expressing NAA10 verified through sequencing. *The plasmids that had been harvested from surviving colonies on selection drug-containing LB-agar plates, were sequenced. Results were visualized in SnapGene.* ***A, C)*** *The base T of codon TGC (Cys) has been substituted by base G, generating the codon GGC (Gly) in the NAA10 C21G plasmid constructs.* ***B, D)*** *The base G of codon CGC (Arg) has been substituted by base A, generating the codon CAC (His) in the NAA10 R83H plasmid constructs. A and B represent mutated sequences of plasmid pcDNA3.5V5HisTOPO-NAA10*[2]*, C and D represent mutated sequences of plasmid pETM-NAA10-His.*

5.4 Expression, purification and chromatograms

Once the mutations had been confirmed, the pETM41-NAA10 constructs were transformed into BL21 *E. coli* cells by heat shock (Methods, section 4.2.1). The transformed bacteria were then cultivated in LB-medium. Upon addition of IPTG, expression of MBP-NAA10 was initiated. The cultures were incubated overnight, and aliquots were analysed by SDS-PAGE to

confirm the protein expression. On the resulting gel image (Figure 5.6), bands (marked: *) close to the weight of the marker band of 70 kDa can be observed. This is close to the theoretical molecular weight of MBP-NAA10 (71 kDa) and can be seen in the lanes corresponding to culture samples collected after expression had been induced. Such bands were not present in the lanes with culture samples without induced expression. Therefore, assuming that the desired protein had been expressed, the bacteria from the analysed cultures were harvested.

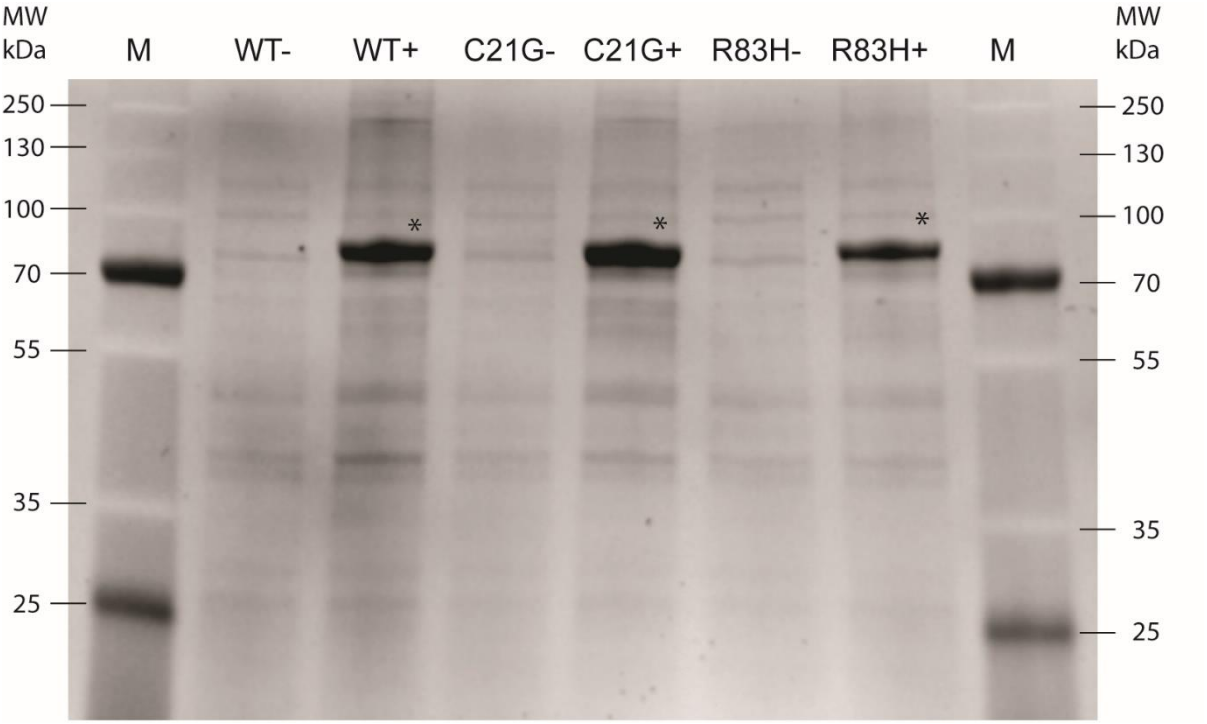

Figure 5.6 MBP-NAA10 variants became expressed upon addition of IPTG. *BL21* E. coli *cells were transformed with pETM41 constructs containing MBP-fused NAA10, either non-mutated (WT), mutated from cysteine in position 21 to glycine (C21G), or arginine in position 83 to histidine (R83H). To induce MBP-NAA10 expression, the culture was added IPTG and set to incubate overnight at 18˚C. Aliquots were collected from the cultures before IPTG had been added (-) and after the cultures had incubated overnight (+). Induced cultures gained prominent bands close to 70 kDa (*).*

Harvested bacteria was passed through a French® pressure cell (Methods, section 4.4.1), generating a bacterial lysate. Cleared lysate was then purified by immobilized metal ion chromatography (IMAC) and size exclusion chromatography (SEC) (Methods, section 4.3.2). The purification process for the wild type and the two variants were run on SDS-PAGE and subsequently imaged and analysed; here represented by the gel image generated for the WT (Figure 5.7). The image shows recurring bands, throughout the purification process, close to the marker band of 70 kDa, assumed to belong to the overexpressed MBP-NAA10 of 71 kDa. The band is absent in the flowthrough (FT) and waste (W) samples, implying that no protein

was lost during the loading and washing step of the IMAC purification. No prominent band is seen in the sample from peak D of the SEC chromatogram either.

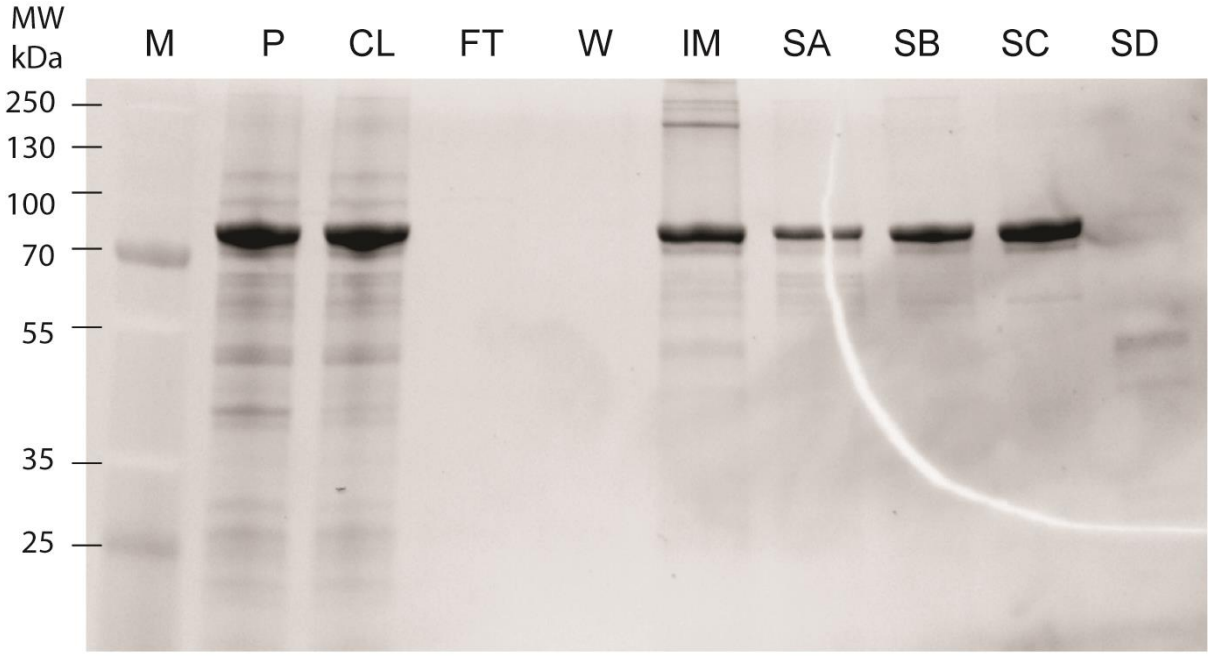

Figure 5.7 MBP-NAA10 wild type purification from cells to purified protein product. *BL21 E. coli cells expressing MBP-NAA10 were harvested, lysed and centrifuged. The pellet (P) was discarded, and the clear lysate (CL) was further purified on IMAC. Aliquots from the flowthrough (FT) and waste (W) were collected from IMAC purification, as well as from the protein containing fractions (IM). Concentrated protein fractions from the IMAC purification were then loaded onto a SEC column, which gave four peaks: A (SA), B (SB), C (SC) and D (SD). The peaks represented fractions in which the majority of the eluted protein was found. The molecular weight (MW in kDa) of the proteins were indicated using a protein ladder (M).*

The SEC samples chosen for gel analysis corresponded to peaks A, B, C and D from a plot of eluate absorbance (mAu) at 280 nm against the volume (mL) eluted from the SEC column. The peaks in this plot, or chromatogram, indicate which fractions contain the most protein, providing suggestions for which fractions should be considered for analysis. Based on standard protein profiles established in earlier studies, expected elution volume for monomeric MBP-NAA10 protein was around 70 mL (peak C). A comparison between the SDS-PAGE image of the different peaks and the chromatogram can be viewed in Figure 5.8.

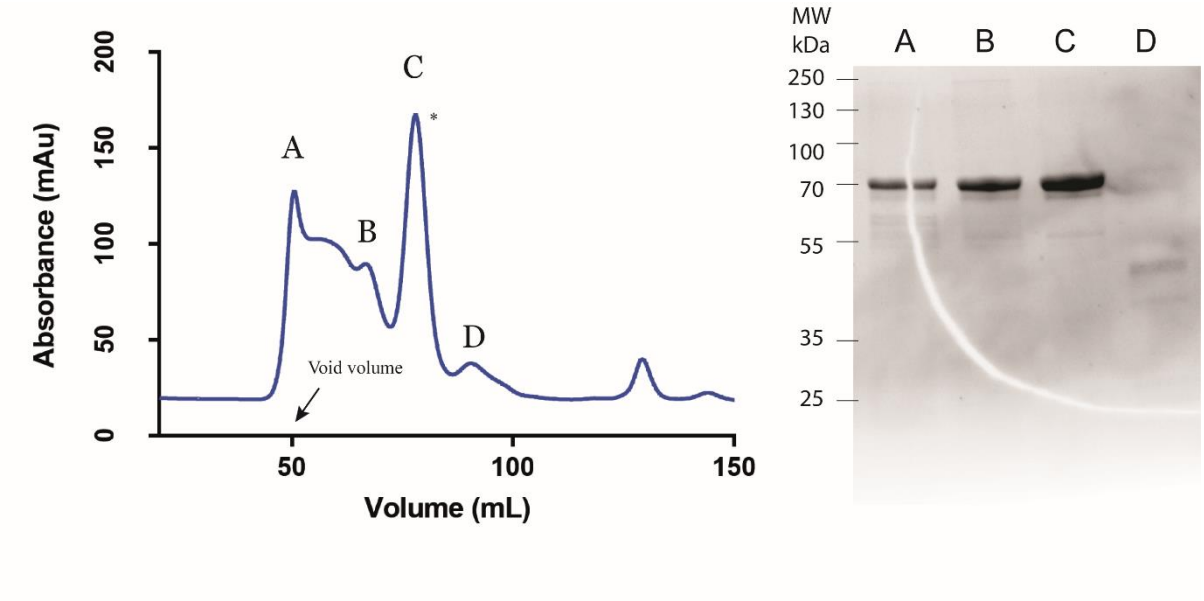

Figure 5.8 Analysis of SEC peaks on SDS-PAGE reveal fraction content. *The chromatogram (left) show primarily four peaks: A. B, C and D, of which the contents were analysed on SDS-PAGE (right). At void volume (50 mL), peak A emerges from the column, followed by peak B at approximately 60 mL; the protein content of these peaks analysed on SDS-PAGE (right) show proteins around the size of monomeric MBP-NAA10 (72.4 kDa), meaning the peaks likely contained aggregated NAA10 proteins. Peak C (*) eluted at around 70 mL, which resembles the predicted elution volume of a protein with the mass of the monomeric MBP-NAA10. Peak D show no apparent band around 70 kDa*

The presence of aggregates in the SEC eluates of the mutants' purifications were also evident, with related chromatograms (Figure 5.9) showing high peaks at lower elution volumes relative to the expected elution volume of 70 mL for MBP-NAA10. The more protein eluted as aggregated complexes early in the SEC, the less protein is left to elute as monomeric subunits. This is confirmed when comparing the three chromatograms in Figure 5.9 the monomeric wild type retains the highest concentration with an absorption maxima of 167.293 at 77.9 mL, whereas C21G, which has the highest peak of the three variants when aggregates are eluted (A = 211.31, V = 49.8 mL), has an absorption maxima of 73.2 at 77.5 mL for the monomeric MBP-NAA10. High peaks at lower elution volumes correlate to smaller peaks at higher elution volumes.

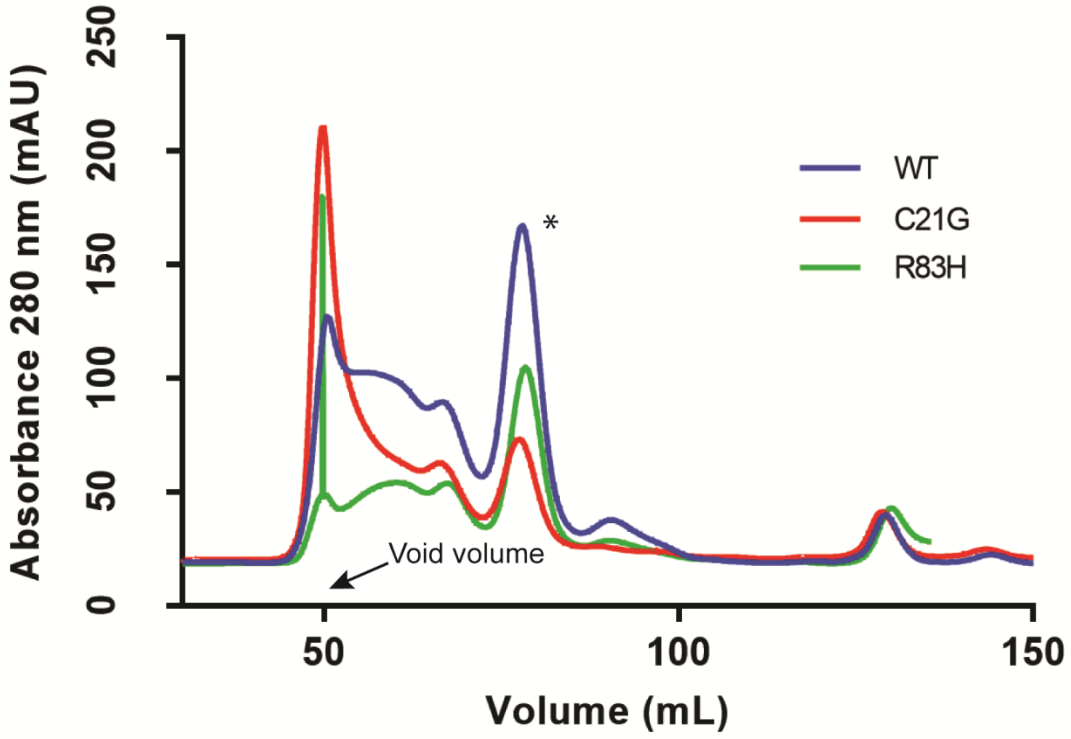

Figure 5.9 Purified NAA10 variants by SEC eluted at ~70 mL. *The three NAA10 variants, wild type (wt, blue), C21G (red), R83H (green), that were expressed and harvested from the BL21* E. coli *cells were purified in a two-step chromatography purification, first by IMAC, then SEC (Methods, section 4.3.2). Eluted protein fractions were represented as absorption peaks at λ=280 nm. The largest proteins were eluted at void volume (50 mL, arrow); monomeric MBP-NAA10 eluted at ~70 mL (*).*

5.5 Colorimetric determination of MBP-NAA10 concentration and activity

The purified protein fractions from the SEC were initially concentrated, before having the protein concentration determined using a BCA protein assay (Methods, section 4.5.1). Here, bovine serum albumin (BSA) samples of increasing concentrations were prepared to generate a standard curve from which the concentrations of MBP-NAA10 samples could be assessed. The MBP-NAA10 wild type had a concentration of 1401.2 µg/mL, the C21G variant had a concentration of 530.4 µg/mL, and the R83H variant had a concentration of 857.2 µg/mL. The molarity of these concentrations corresponded to 19.7, 7.5 and 12.1 µM, respectively.

Proteins that had been stored at -20ºC with glycerol, were retrieved and applied to an *in vitro* DTNB based acetylation assay in 100 nM concentrations (Methods, section 4.5.2). The DTNB-

based acetylation assay was carried out in two different ways: One was used to test the activity of the NAA10 variants at 37˚C together with EEEIA after 10, 20 and 30 minutes (Figure 5.10); in the other acetylation assay, the three NAA10 variants were tested in 20 min reactions at 37˚C with the peptides EEEIA, DDDIA, SESSS and MLGPE (Figure 5.11). The activity was assessed by spectrophotometrically measuring the concentration of NTB^-, generated after DTNB had been added. The time-based assay showed a steady increase in activity over time, with a slight plateauing tendency toward the 30 min mark, likely indicating a decrease in substrate. The peptide-based assay was therefore conducted with 20 min reactions, ensuring that enzyme activity was assessed close to maximum velocity. At 20 minutes, the NTB^- concentration were 22.80 µM for the MBP-NAA10 wild type, 35.76 µM for the C21G variant, and 0.85 µM for the R83H variant.

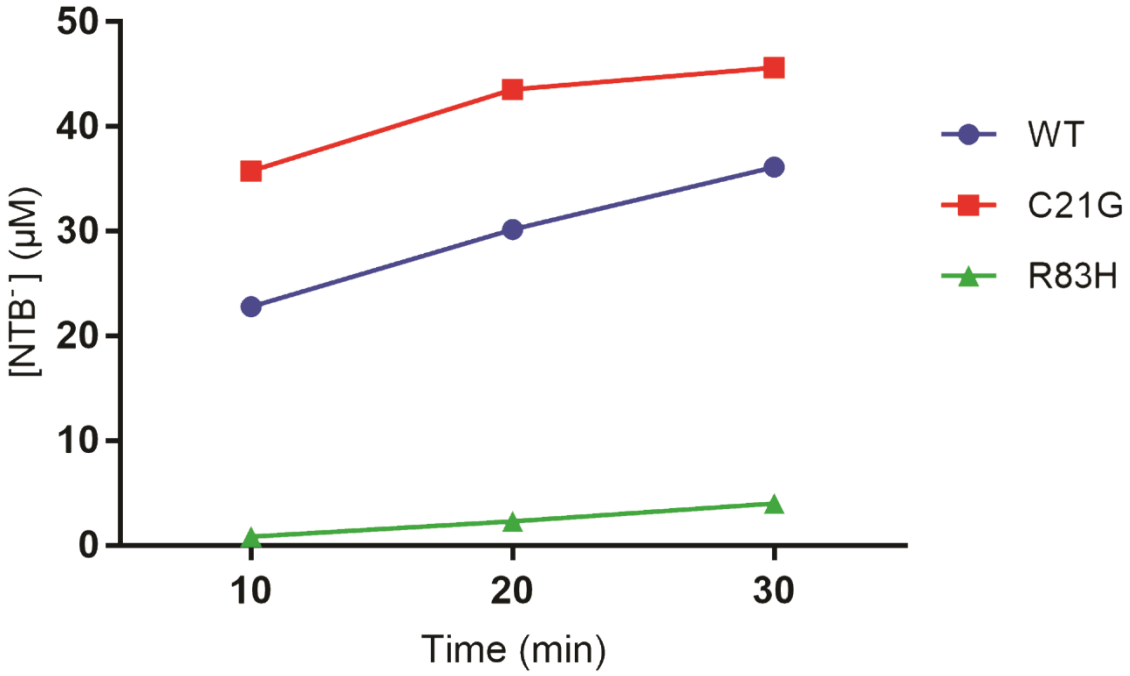

Figure 5.10 ***in vitro*** **DTNB-based acetylation assay determine NAA10 activity over time.** in vitro *NAA10 reactions of wild type (blue), C21G (red) and R83H (green) with peptide substrate EEEIA ran for 10, 20 and 30 minutes. Subsequent addition of saturated DTNB solution, generated the yellow compound* NTB^- *as a bi product of the reaction between DTNB and CoA, enabling the spectrophotometric measurement of enzymatic activity.*

In the peptide-based assay (Figure 5.11), NTB^- concentration after 20 minute reactions with EEEIA were 29.48 µM, 41.48 µM and 2.24 µM for the wild type, the C21G and the R83H variants, respectively. For the DDDIA, NTB^- was at 12.4 µM for the wild type, 20.21 µM for

the C21G mutant and 2.79 µM for the R83H mutant. The measurements for MLGPE and SESSS were too low to assess confidently. These measurements demonstrate an increase in catalytic activity for C21G at 40.7% for EEEIA and 62.9% for DDDIA, and a decrease in catalytic activity for R83H at 97.9% for EEEIA and 77.4% for DDDIA.

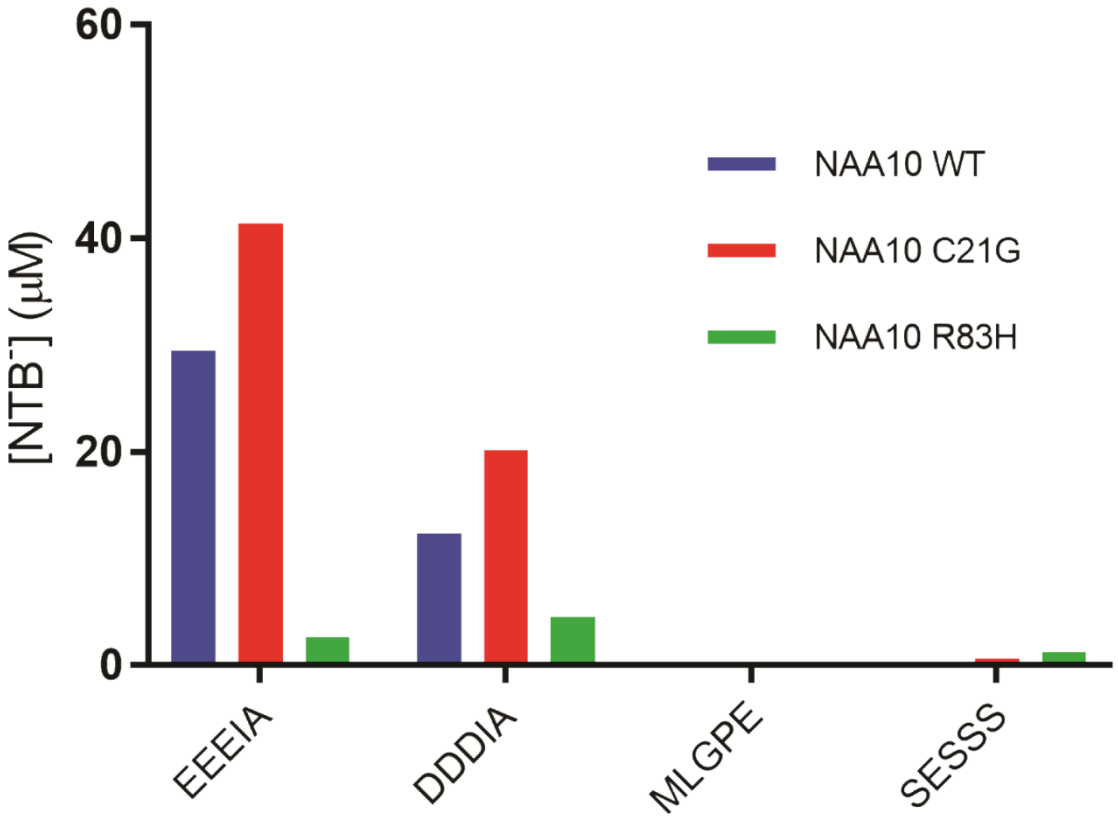

Figure 5.11 in vitro DTNB-based acetylation assay determine NAA10 variants' activity toward different peptides. in vitro *NAA10 reactions with peptide substrates EEEIA, DDDIA, MLGPE and SESS ran for 20 minutes. Subsequent addition of saturated DTNB solution, generated the yellow compound NTB^- as a bi product of the reaction between DTNB and CoA, enabling the spectrophotometric measurement of enzymatic activity.*

5.6 NAA10 protein stability assessment

HeLa cells transfected with the V5-plasmids (Methods, section 4.6.2) encoding the different NAA10 variants were treated with the translation inhibitor, cycloheximide (CHX), and harvested after 2, 4 and 6 hours. Control cells were harvested without CHX treatment (0 hours). Harvested cells were then lysed and analysed by Western blotting (Methods, section 4.6.4). The experiment was carried out in triplicates: Here, only the 3rd replicate is presented due to difficulties with the analysis of the two other replicates. After treatment with primary and

secondary antibodies towards the V5-signal and the pan-actin loading control, blots were imaged (Figure 5.12 A) and the signals from the various bands were quantified and normalized to the control signals (0h) of their respective protein (Figure 5.12 B). The normalized results showed that C21G appears to follow the same rate of degradation as the wild type for the 6 hours after CHX addition (Figure 5.12 B), though generally yielding a stronger V5-signal throughout the time frame. R83H show a more gradual degradation than the wild type protein, yielding a V5-signal of 81.5% after 6 hours, whereas the wild type yields a V5 signal of 22.3% after 6 hours. The NAA10 C21G V5-signal after 6 hours was at 43.9%.

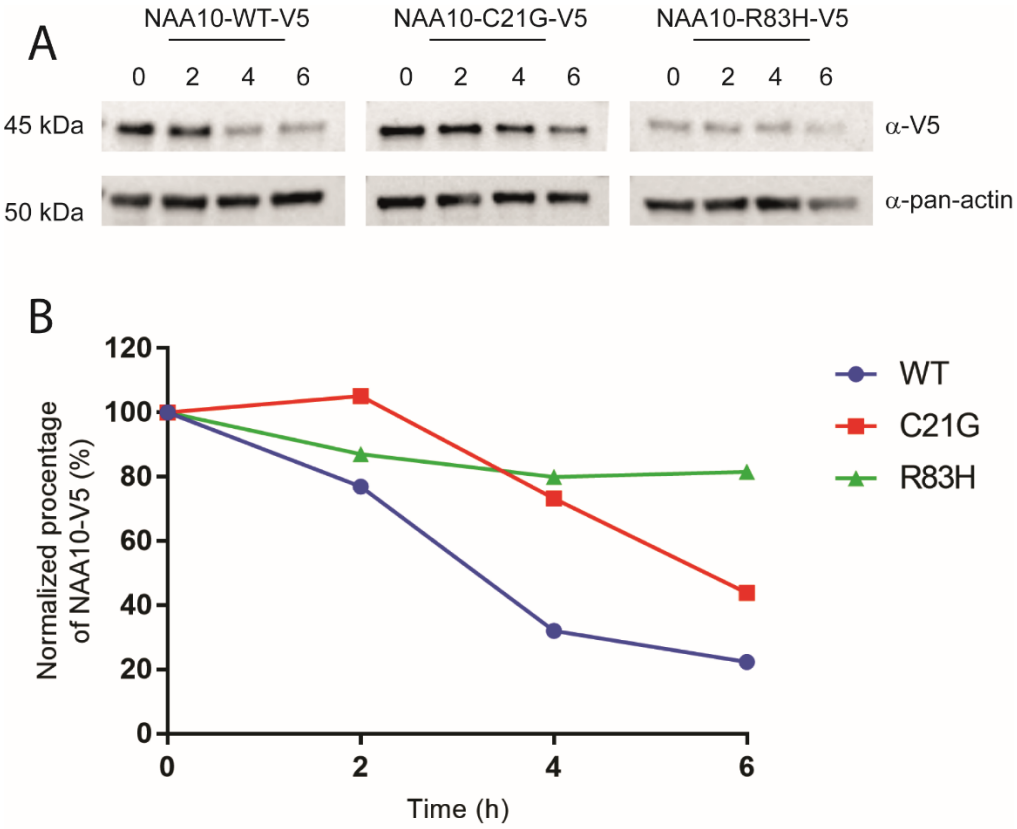

Figure 5.12: Western blots of lysed HeLa cells following CHX chase assay show protein stability over time. *Three replicates from each of the protein variants were treated with CHX for 0 (no treatment), 2, 4 and 6 hours to investigate the stability of the mutant NAA10 proteins over time. Pan-actin was used as a loading control. All membranes were imaged; shown here are the third replicates (A), and the signals were quantified using the 0h sample of V5-NAA10-WT as reference. The quantified signals were then normalized to the 0h samples of their respective protein variant and plotted against the time interval of the CHX treatment.*

6 Discussion

Nt-acetylation is a common and essential protein modification carried out by the NAT protein family in eukaryotic cells. Of the NATs, NatA, composed of the catalytic subunit NAA10 and a ribosome binding unit NAA15, acetylates the most protein targets in the human cell. Recent studies characterizing NAA10 and NAA15 mutations expressed in human patients have communicated the importance of Nt-acetylation in human health. Phenotypes shared among the affected individuals are typically hypotonia, mild to severe intellectual disabilities (ID), postnatal growth deficiency, and various skeletal and organ anomalies. Functional *in vitro* studies of NAA10 and NatA have shown that reduction in catalytic activity, reduced protein stability, impaired complex formation, and impaired binding to substrate or Ac-CoA are tied to the severity of the patients' phenotypic profiles. In this study, I aimed to investigate two novel NAA10 mutations, C21G and R83H, identified in three male patients. By applying various methods in molecular biology and bioinformatics, the *in vitro* activity and cellular stability of these protein variants were characterized, coupled with investigations of mutated residues' conservation and in silico analysis of electrostatic potential.

6.1 Previously identified and characterized NAA10 variants

Since the identification of NAA10 mutation S37P in Ogden syndrome by Rope and colleagues in 2011[10], and the linking of LMS to a C-terminally truncated NAA10 protein in 2014 by Esmailpour and colleagues[11], several studies identifying and aiming to functionally characterize novel NAA10 mutations have emerged (see Introduction, Table 2.2 for details on phenotypes). Two such studies, Popp et al. (2015) and Saunier et al. (2016), characterized missense mutations V107F, F128I and F128L, situated in the hydrophobic core of NAA10 and were generally associated with a destabilized protein structure and a reduced catalytic activity[12,13]. Another 2015 study characterized the functional impact of NAA10 mutation Y43S; this mutation was found to reduce catalytic activity and destabilize the protein, because it holds a structurally important position between the β-sheet and the 3rd α-helix[78]. Less structurally impactful were the R116W and R83C mutations documented in studies of Popp et al. (2015) and Saunier et al. (2016), respectively. R116 and R83 were both found to be positively charged and associated with the Ac-CoA binding site of NAA10; substitutions to tryptophan and cysteine resulted in reduced catalytic activity as a consequence of impaired association with Ac-CoA[12,13].

6.2 Disease predictions and residue conservation

The MSA performed for the human NAA10 protein, and the eight other species (*M. musculus, R. norvegicus, X. laevis, C. elegans, A. thaliana, S. pombe and S. cerevisiae*) by Clustal Omega assigned conservation scores of 8 (out of 11) to the two mutation sites studied in this thesis. Conservation scores retrieved from *in silico* mutant predictions, however, typically showed a near complete conservation; for instance, the PhastCons score of MutationTaster[86], indicating conservation of the residue considering adjacent amino acids and an MSA of 36 species, gave the scores of 1, indicating the highest degree of conservation[87], to both C21 and R83. The difference in conservation scored by the various alignments, could be explained by the inclusion of distantly related NAA10 proteins from unicellular organisms (*S. pombe and S. cerevisiae*) in the MSA.

Additionally, the MSA revealed that the NAA10 sequence of *S. pombe* had a histidine at position 83. When the human NAA10 gains a histidine in position 83, *in silico* predictions, MutationTaster and PolyPhen-2 (Results, section 5.2.2), predicted that the studied mutation would be "disease causing" and "probably damaging", respectively. This posed the question: Why is histidine in position 83 in the *S. pombe* NAA10 accordant with a functional protein, whereas an R83H mutation in human NAA10 causes disease? An alignment of the Ac-CoA associating *S. pombe* NAA10 (4KVX)[52] and Ac-CoA-free human Nat A (6C9M)[54] (Figure 6.1) showed that H83 of the *S. pombe* protein orients toward the Ac-CoA and appear to interact weakly with indole ring and phosphate group of Ac-CoA (top image); bond lengths are, however, slightly too long (4.0 and 4.2 Å) for proper π-stacking and hydrogen bonds, possibly indicating van der Waals interactions. As a structure for human NatA associated with Ac-CoA is not yet solved, the orientation and bond length of R83 toward Ac-CoA in this alignment (Figure 6.1, bottom image) is not representative for the Ac-CoA-bound state of NatA. However, the slight shift of R83 in the C-terminal direction relative to H83 in the *S. pombe* protein (Figure 6.1, bottom image), directs the residue towards other parts of Ac-CoA, indicating that H83 and R83, though in the same position, interacts with different chemical groups of Ac-CoA.

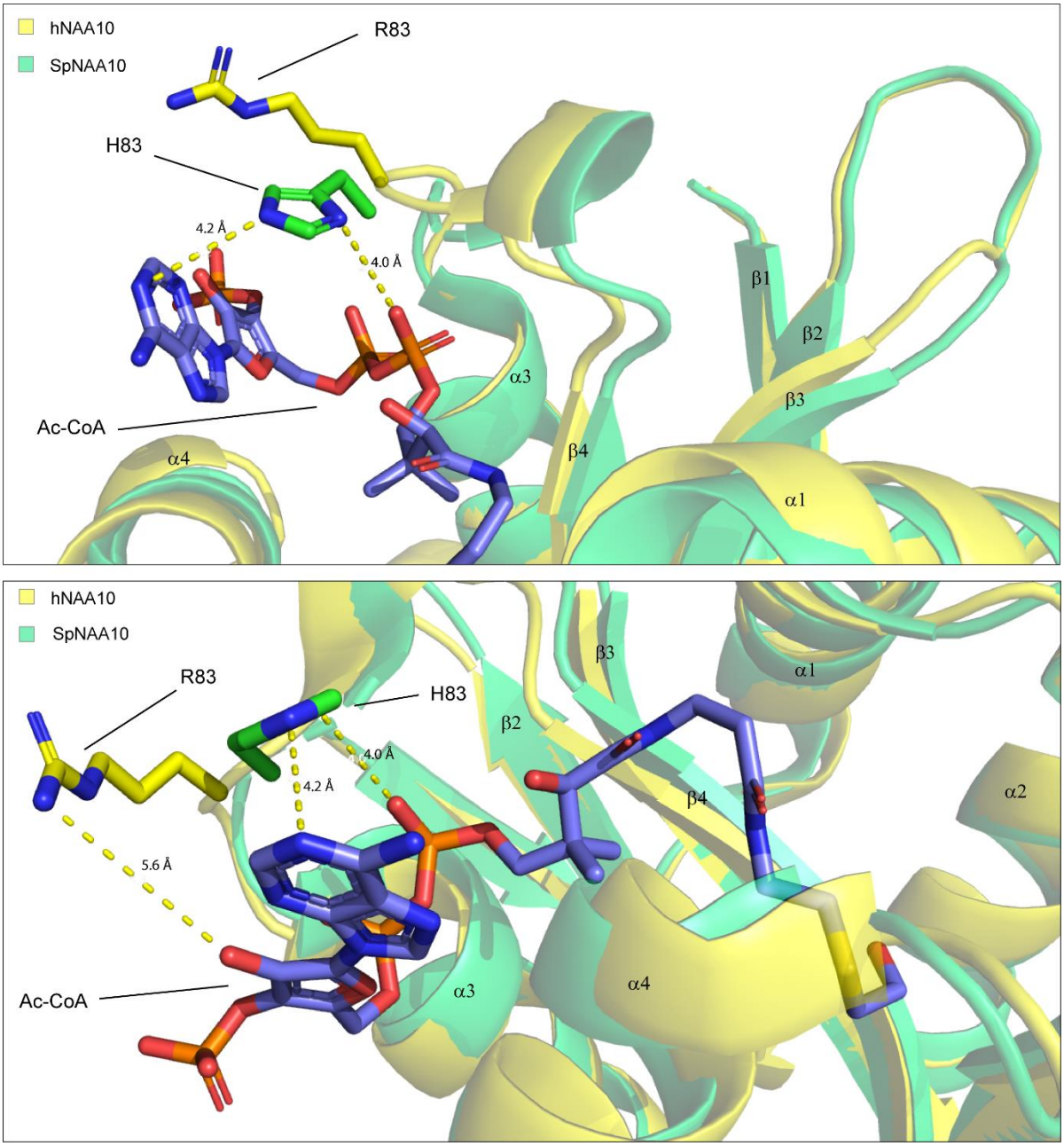

Figure 6.1 H83 in *S. pombe* NAA10 interacts with different parts of Ac-CoA compared with R83 in human NAA10. *An alignment between* S. pombe *NAA10 (green) in complex with Ac-CoA (4KVM) (blue, sticks) and human NAA10 (6C9M) (yellow), show that the spatial orientation of the amino acids in position 83 for the two structures are slightly different. H83 of the* S. pombe *NAA10 positions itself between the phosphate groups and the indole ring of Ac-CoA, possibly participating weak interactions (top image). The R83 residue of hNAA10 orients more toward the outer functional groups of Ac-CoA (bottom image), possibly indicating that these two amino acids, despite being in the same position, interacts with different parts of the Ac-CoA.*

6.3 Clinical data and functional characterization

6.3.1 R83C and R83H are localized in the Ac-CoA binding site

One of the studied NAA10 mutations in this thesis is R83H, located in the same position as R83C, previously characterized by Saunier and colleagues in 2016[13]. The seven female patients retaining the NAA10 R83C mutant had no common denominator phenotype other than a general NAA10-related Nt-acetylation deficency[13] and microcephaly which was shared by all patients in the study. One of the girls inherited the R83C mutation maternally, and also had a brother with severe phenotypes who died at week 1; all others had *de novo* mutations. From available clinical data on the two boys affected by the R83H mutation, intellectual disability, behavioural abnormalities and poor speech development appear to be the main shared phenotypes. Patient 1, teenaged, inherited the mutation maternally, whereas patient 2, aged 12, attained the mutation *de novo* with a 75% mosaic degree (Results, section 5.1). Explanations for the observed inconsistencies in phenotypic profiles of R83-mutated patients are many and complex. Individual genetic make-up and variations in development can potentially steer the mutation toward a more, or less, severe phenotype. Additionally, monomeric NAA10 has been reported to exhibit a role in genetic imprinting concerning expression of genes involved with growth, development and neuron survival[66]. A NAA10 mutated on R83 can thus lead to altered expression of genes, though not necessarily identical for individuals with the same mutation.

Functional characterization of the R83C mutant by Saunier et al. showed a 60% reduction in catalytic activity and no decrease in structural stability relative to the wild type. Structural assessments of a NAA10 homology model, placed R83 at the Ac-CoA binding site, thus suggesting that reduced catalytic activity is a consequence of impaired Ac-CoA binding[13]. In comparison, acetylation assays conducted in this thesis show that NAA10 R83H has a catalytic reduction of almost 100% (Results, Figure 5.11), and potentially an increase in cellular stability relative to the wild type (Results, Figure 5.12). Visualization of the human NAA10 structure, 6C9M[54], as well as electrostatic analyses of the R83H mutant (Results, Figures 5.2, 5.3 and 5.4) show that when R83 is substituted with histidine, the positive charge at the site is lost; additionally, histidine, as a more bulky and shorter residue, might not be able to retain the weak interactions with Ac-CoA facilitated by the original arginine residue. Though findings functionally characterizing the R83H mutant are more extreme than the R83C findings, both indicate that the decline in catalytic activity likely arise from impaired Ac-CoA binding. Unfortunately, no concrete conclusions can be drawn from the findings presented in the

stability assay of this thesis since biological replicates have not been produced and reproducibility is not assured. The same goes for the in silico structural assessments, as the human NAA10 structure employed in this thesis is solved for a different conformational state (associated with HYPK) than the NAA10 structure of *S. pombe* (associated with Ac-CoA, or bisubstrate analogue) with which it is aligned.

6.3.2 NAA10 C21G is a novel mutation site showing an increased catalytic activity

Mutations of NAA10 C21 have not previously been characterized, so no earlier studies can be used to make general comparisons with the functional assessments of this variant. Not much is known of the affected individual's clinical assessment either, other than him harbouring phenotypes similar to the previously described non-syndromous NAA10 deficient patients. In the in vitro acetylation assay of MBP-NAA10, this mutant, quite surprisingly, showed a prominent increase in catalytic activity relative to the wild type. For the EEEIA peptide, NAA10 C21G had a 41% higher activity, and a 63% higher activity for the reactions with DDDIA (Results, section 5.5, Figure 5.11). The stability assays gave no indication that C21G reduced the structural integrity of NAA10 as a monomer.

Structural assessments (Results, Figure 5.3) and multiple sequence alignment (Results, Figure 5.1) place C21 at the C-terminal end of α1 when complexed with NAA15. This is in the middle of the highly conserved α1-loop-α2 region suggested by Liszczak et al. (2013) to form important hydrophobic interfaces between NAA10 and NAA15, in addition to hydrogen bonds and van der waals interactions with the substrate N-termini of NatA[52]. Further, the study suggests that upon NatA complex formation, the C-terminal end of α1 gains an additional turn, and thereby shifts the positions of the substrate interacting amino acids (L22A, E24Q and Y26A) from a surface-exposed conformation and into the active site[52]. This implies that C21, the C-terminal residue in α1 of complexed NAA10, goes from being in the α1-α2 loop (uncomplexed structure), to being incorporated into α1 (NatA complex). An alignment between the *S. pombe* monomeric NAA10 (4KVX)[52], *S. pombe* NatA complex (4KVM)[52] and human NatA (complexed with HYPK, 6C9M)[54], illustrate the structural difference between NAA10 in monomeric and NatA complex conformation (Figure 6.2). In monomeric form, NAA10 has a shortened α1-α2 loop, and an apparently elongated α2 structure. N21, of the *S. pombe* protein displays a slight shift toward the α1 when complexed, and the substrate interacting amino acids, L22, E24 and Y26, also clearly retain different orientations when complexed. When C21

residue is mutated to glycine, this could potentially reverse the extension of α1 in NatA conformation and introduce higher flexibility in the α1-α2 loop, thereby possibly affecting substrate binding or even global structural integrity.

Figure 6.2 also display the important shift of S. pombe residue H20, which participate in a small hydrophobic interface crucial for NatA complex formation. It is not known if L20 retains the same function in the human structure, but if so, a C21G mutation might impact its orientation, and thereby impact functional NatA complexing.

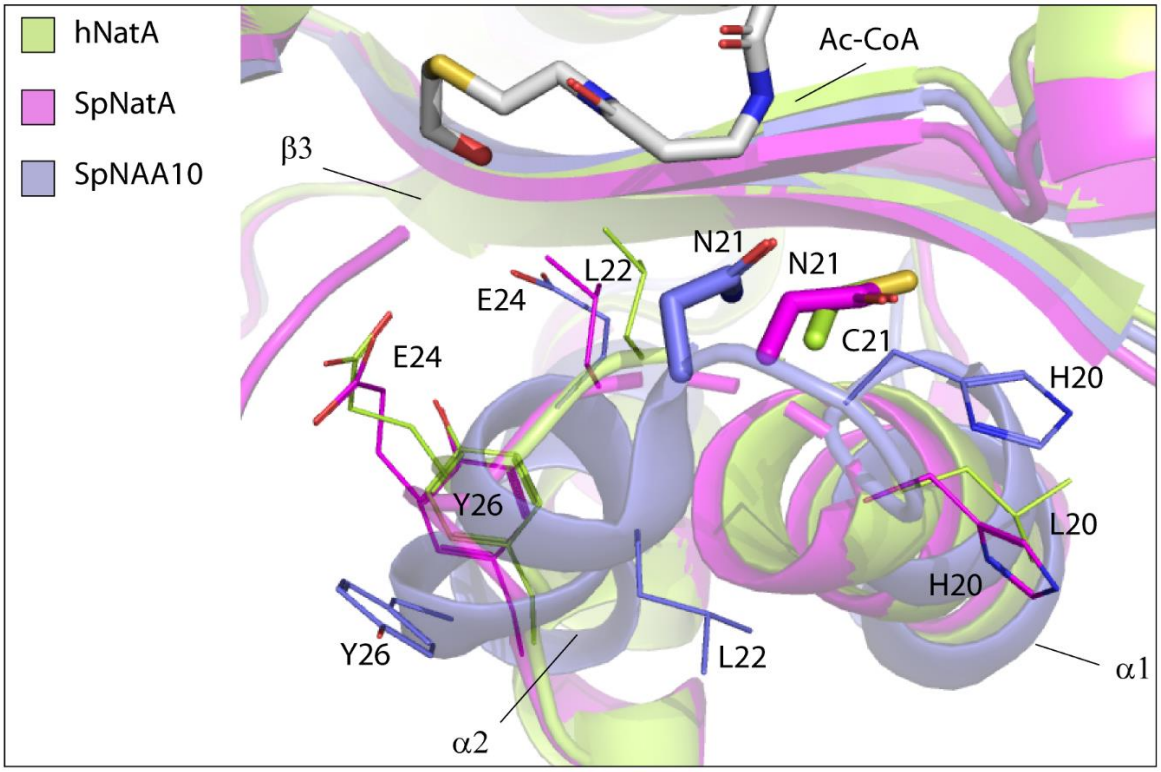

Figure 6.2 Complexed NAA10 gains additional turn in C-terminal end of α1 and prepare α1-α2 loop for substrate binding. *When NAA10 forms the NatA complex with NAA15, structural analysis by Liszczak et al. show that the C-terminal end of α1 is extended by one turn*[52]*. This leads the C21 and N21 residues of human (green) and* S. pombe *(NatA = magenta, uncomplexed NAA10 = blue) proteins, respectively, from a loop region to an α-helix (α1). Complexing also shifts residues L22, E24 and Y26 from a surface-oriented conformation, into position for substrate interaction. The mutation of C21 to glycine can potentially reverse the α1 extension and increase flexibility in the loop-region between α1 and α2. Increased flexibility could further impact substrate binding and overall structure of both NAA10 and NatA.* S. pombe *residue H20 was shown by Liszczak and co-workers to participate in a hydrophobic interface crucial for NatA complex formation in* S. pombe[52]*; if L20 has a similar role in humans, a C21G mutation could also impact the NatA complex formation in this respect.*

6.4 Experimental procedures and limitations

6.4.1 Predictions and structural assessment

The *in silico* mutation prediction tools, MutationTaster and PolyPhen-2, characterized both C21G and R83H as "disease causing" and "probably damaging", respectively, with scores exactly or close to 1 (high prediction confidence). SIFT, on the other hand, evaluated the mutations to be "tolerated" (non-deleterious) with scores of 0.4 for the C21G and 0.06 for the R83H mutant. According to SIFT documentation, "tolerated" predictions are attributed with more confidence as the score approaches 1. The prediction that R83H is "tolerated" is, therefore, not very reliable as the score is barely above the threshold value for "deleterious" ($\leq$0.05). The C21G mutation being "tolerated" is more likely, but 0.4 is not a decidedly confident score either. The PyloP scores (from -14 to +6) generated by the MutationTaster tool, showed a similar trend, where R83 were scored 5.359 and C21 2.764. The scores are positive, thus indicating a slow rate of evolution compared to neutral genetic drift (Methods, section 4.1.2), but R83 is closer to 6, suggesting that this site is historically less frequently mutated compared to, for instance, C21. These results give an indication that mutations in R83 are generally more severe or "intolerable" to the organism, than mutations at C21.

Solved structures used in this thesis are NAA10 (chain B) of the HYPK-associated human NatA structure 6C9M[54], NAA10 (chain E) of the Ac-CoA-associated *S. pombe* NAA10 structure 4KVX[52], and NAA10 (chain E) of the bisubstrate inhibitor-associated *S. pombe* NatA structure 4KVM[52]. As these structures exhibit different conformational states of NAA10, the alignments of 6C9M with 4KVX or 4KVM for visual placement of Ac-CoA or the bisubstrate inhibitor in the active site (Results, Figures 5.2, 5.3, 5.4), will therefore not give a true representation of human NAA10 bound to these compounds. These alignments were conducted to get a rough impression of where the Ac-CoA and peptide would associate relative to mutation sites.

In silico mutagenesis was performed in the molecular visualization program, PyMOL, to assess possible outlooks of the studied substitutions relative to adjacent or near-by amino acids in the microenvironment (Results, Figure 5.3). Consequential changes the in secondary and tertiary structure are not accounted for in this method; proper assessment of these parameters could be

achieved by employing homology modelling of the mutant proteins, or comparing wild type secondary structure with the mutant using circular dichroism.

6.4.2 Protein expression and purification

Bacterial culture pre- and post-induction with IPTG were sampled and analysed on SDS-PAGE (Results, Figure 5.6). Bands at approximately 70 kDa formed in IPTG-induced culture samples, and as monomeric, His-tagged, MBP-NAA10 retains a molecular weight of 71 kDa, it was concluded that these bands were likely to derive from successfully expressed proteins, and could be harvested from the cultures.

Throughout the purification process, samples were collected in order to monitor the various stages on subsequent SDS-PAGE (Results, Figure 5.7). The resulting gel image of the wild type MBP-NAA10 purification show bands close to the 70 kDa marker in the preparatory stages (Figure 5.7, P, CL) prior to the purification. No bands are seen in the lanes representing samples extracted from the loading (FT) and washing (W) step in the IMAC purification, but a quite clear band close to 70 kDa turn up in the sample taken from the eluted fraction of the IMAC. The same sized bands appear in the A, B and C fractions of the SEC purification as well. As concluded from the gel image of the expression, these recurring ~70 kDa bands likely belong to MBP-NAA10 (71.0 kDa). This progression suggests the presence of MBP-NAA10 in the cellular remains after the cell lysis (P), and in the clear lysate (CL). The absence of MBP-NAA10 in the IMAC flowthrough and waste fractions, indicate that the His-tags of the loaded MBP-NAA10 successfully adhered to the column matrix; this is also evident from the prominent band appearing in the lane containing the IMAC fraction (IM). The same prominent band appeared in SEC fractions A, B and C; based on standard protein profiles established in earlier studies and previous purification experiments with MBP-NAA10, for instance presented in the 2017 master's thesis by Nina McTiernan, the protein was expected to elute from the SEC at approximately 70 mL, making fraction C the likely candidate from which to extract MBP-NAA10. As SEC separates protein based on size, peak A and B fractions were expected to contain proteins or protein complexes larger than MBP-NAA10. From the SDS-PAGE analysis, which denatures and disrupts quaternary structure of the sampled proteins, peak A, B and C all appeared to contain MBP-NAA10. From these observations, it was likely that MBP-NAA10 at some point during the purification process formed aggregates of various sizes, which were disbanded in SDS-PAGE, but not in the SEC purification. In Nina McTiernan's thesis,

however, no large peaks around 50 and 60 mL (peak A and B respectively), of the wild type MBP-NAA10 purification were seen. Though no clear explanation for the differences in MBP-NAA10 aggregation between our two theses spring to mind, a few methods and experimental parameters differed in our purification procedures. One, or a combination, of these differences could maybe have had an impact on the aggregation. Flawed experimental conduction on my part might also have led to a higher rate of aggregation than expected.

6.4.3 Catalytic activity and protein stability assays

Acetylation reactions of purified MBP-NAA10 were conducted some time after concentration measurements by BCA protein assay (Methods, section 5.5). It was brought to my attention that aggregation taking place in the protein sample after purification might have affected the *in vitro* activity of the protein. Roughly, the suggestion entailed centrifugation of the protein sample, remeasuring the concentration of the supernatant solution, and then redoing the acetylation assay. There was unfortunately not enough time left in this thesis to carry out this experiment, but in future purification experiments of NAA10, this procedure should be followed prior to acetylation assays.

The cycloheximide chase experiments conducted to assess the protein stability of NAA10 variants in the cellular environment of HeLa cells, gave no indication that either of the mutants impacted the integrity of NAA10 protein structure. Rather, the results demonstrated a highly unlikely increase in V5-signal for both variants; the C21G variant gained 5% signal strength from 0 to 2 hours after CHX addition, the R83H variant gained 0.4% signal strength from 4 to 6 hours after CHX addition (Results, Figure 5.12). A possible explanation for this is that the CHX reagent was ineffective as a consequence of prolonged storage in solution at -20ºC, leading to continued protein synthesis after the reagent had been added.

6.5 Concluding remarks and future prospects

The NAA10 C21G and R83H mutations have been partially characterized in this thesis in terms of monomeric acetylation activity, stability and *in silico* structural and functional analysis. Nevertheless, more work remains to fully pinpoint these mutations' impacts on NAA10 and NatA function. For now, the variants have been predicted with a high degree of certainty by two out of three online *in silico* mutation prediction tools to have a deleterious effect on NAA10 function. Further strengthening these predictions, are the structural assessments in human

HYPK-NatA structure, which places the positively charged R83 at the rim of the Ac-CoA binding site, and C21 in the middle of the highly conserved α1-loop-α2 region that make up parts of the NatA substrate binding site and forms hydrophobic interfaces upon association with NAA15[52]. It is my suggestion that the C21G mutant will increase the local flexibility, possibly disrupting NatA complex formation, and impairing NatA substrate binding; in monomeric form, the mutation might lead to a more active acetylation as the flexibility potentially "opens up" the substrate binding site and making it more accessible. These effects could also have negative implications for the overall protein stability of NAA10 C21G, potentially leading to a more rapid degradation relative to the wild type. It's difficult to predict whether the C21G will impair the potential KAT and non-acetylating (protein interaction) roles of NAA10, but if C21G indeed leads to loss in structural integrity, it is my belief that this might also lead to reduced KAT activity and impaired protein-protein interaction. As for the NAA10 R83H variant, I propose that the reason for reduced acetylation activity is chiefly due to impaired binding to Ac-CoA; as the mutation unlikely affects the overall structure of NAA10, a reduced activity should be observed in both monomeric and complexed NAA10. For the same reason, non-acetylating roles of NAA10 will likely not become affected by this mutation. The NAA10 KAT activity, on the other hand, will likely also become reduced as this reaction also depend on the association of NAA10 with Ac-CoA.

For continued functional characterization of these variants, other methods can be applied, such as enzyme kinetics, immunoprecipitation (IP) and ^{14}C-assay to determine NatA activity for mutants relative to the wild type. To investigate the variants at endogenous levels, and attain a broader analysis of the various roles of NAA10 in the cell, CRISPR-Cas9 mutated NAA10 cell lines for each variant could be made. Then, cells could be checked for NatA complex formation, proteome-wide N-terminal- and potential lysine acetylation patterns, and possibly non-acetylating functions of NAA10. Further, to investigate the possibly increased mobility in the active site of the NAA10 C21G variant, homology modelling of the mutant and wild type could be employed. Potential changes in the secondary structure of the C21G variant could also be detected using circular dichroism (CD).

7 References

1. Aksnes H, Drazic A, Marie M, Arnesen T. First Things First: Vital Protein Marks by N-Terminal Acetyltransferases. *Trends Biochem Sci*. 2016;41(9):746-760. doi:10.1016/j.tibs.2016.07.005.

2. Arnesen T, Anderson D, Baldersheim C, Lanotte M, Varhaug JE, Lillehaug JR. Identification and characterization of the human ARD1–NATH protein acetyltransferase complex. *Biochem J*. 2005;386(3):433-443. doi:10.1042/BJ20041071.

3. Starheim KK, Arnesen T, Gromyko D, Ryningen A, Varhaug JE, Lillehaug JR. Identification of the human *N* $^{\alpha}$ -acetyltransferase complex B (hNatB): a complex important for cell-cycle progression. *Biochem J*. 2008;415(2):325-331. doi:10.1042/BJ20080658.

4. Starheim KK, Gromyko D, Evjenth R, et al. Knockdown of Human N -Terminal Acetyltransferase Complex C Leads to p53-Dependent Apoptosis and Aberrant Human Arl8b Localization. *Mol Cell Biol*. 2009;29(13):3569-3581. doi:10.1128/MCB.01909-08.

5. Hole K, van Damme P, Dalva M, et al. The human N-Alpha-acetyltransferase 40 (hNaa40p/hNatD) is conserved from yeast and N-terminally acetylates histones H2A and H4. *PLoS One*. 2011;6(9):1-11. doi:10.1371/journal.pone.0024713.

6. Evjenth R, Hole K, Karlsen OA, Ziegler M, Amesen T, Lillehaug JR. Human Naa50p (Nat5/San) displays both protein Nα- and Nε-acetyltransferase activity. *J Biol Chem*. 2009;284(45):31122-31129. doi:10.1074/jbc.M109.001347.

7. van Damme P, Hole K, Pimenta-Marques A, et al. NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation. *PLoS Genet*. 2011;7(7). doi:10.1371/journal.pgen.1002169.

8. Drazic A, Aksnes H, Marie M, et al. NAA80 is actin's N-terminal acetyltransferase and regulates cytoskeleton assembly and cell motility. *PNAS*. 2018;115(17):201718336. doi:10.1073/pnas.1718336115.

9. Ree R, Varland S, Arnesen T. Spotlight on protein N-terminal acetylation. *Exp Mol Med*. 2018;50(7):90. doi:10.1038/s12276-018-0116-z.

10. Rope AF, Wang K, Evjenth R, et al. Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *Am J Hum Genet*. 2011;89(1):28-43. doi:10.1016/j.ajhg.2011.05.017.

11. Esmailpour T, Riazifar H, Liu L, et al. A splice donor mutation in NAA10 results in the dysregulation of the retinoic acid signaling pathway and causes Lenz microphthalmia syndrome. *J Med Genet*. 2014;51(3):185-196. doi:10.1021/nl061786n.Core-Shell.

12. Popp B, Støve SI, Endele S, et al. De novo missense mutations in the NAA10 gene cause severe non-syndromic developmental delay in males and females. *Eur J Hum Genet*. 2015;23(5):602-609. doi:10.1038/ejhg.2014.150.

13. Saunier C, Støve SI, Popp B, et al. Expanding the Phenotype Associated with NAA10-Related N-Terminal Acetylation Deficiency. *Hum Mutat*. 2016;37(8):755-764. doi:10.1002/humu.23001.

14. McTiernan N, Støve SI, Aukrust I, et al. NAA10 dysfunction with normal NatA-complex activity in a girl with non-syndromic ID and a de novo NAA10 p.(V111G) variant - a case report. *BMC Med Genet*. 2018;19(1):1-9. doi:10.1186/s12881-018-0559-z.

15. Ezkurdia I, Juan D, Rodriguez JM, et al. Multiple evidence strands suggest that theremay be as few as 19 000 human protein-coding genes. *Hum Mol Genet*. 2014;23(22):5866-5878. doi:10.1093/hmg/ddu309.

16. Ponomarenko EA, Poverennaya E V., Ilgisonis E V., et al. The Size of the Human Proteome: The Width and Depth. *Int J Anal Chem*. 2016;2016. doi:10.1155/2016/7436849.

17. Cohen P. The origins of protein phosphorylation. *Nat Cell Biol*. 2002;4:E127. http://dx.doi.org/10.1038/ncb0502-e127.

18. Verdin E, Ott M. 50 years of protein acetylation: From gene regulation to epigenetics, metabolism and

beyond. *Nat Rev Mol Cell Biol*. 2015;16(4):258-264. doi:10.1038/nrm3931.

19. Baslé E, Joubert N, Pucheault M. Protein Chemical Modification on Endogenous Amino Acids. *Chem Biol*. 2010. doi:10.1016/j.chembiol.2010.02.008.

20. Philips DM. The presence of acetyl groups in histones. *Biochem J*. 1963;87(2):258-263. doi:10.1042/bj0870258.

21. Brownell JE, Allis CD. An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. *PNAS*. 1995;92(14):6364-6368. doi:10.1073/pnas.92.14.6364.

22. Drazic A, Myklebust LM, Ree R, Arnesen T. The world of protein acetylation. *Biochim Biophys Acta - Proteins Proteomics*. 2016;1864(10):1372-1401. doi:10.1016/j.bbapap.2016.06.007.

23. Arnesen T, Van Damme P, Polevoda B, et al. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *PNAS*. 2009;106(20):8157-8162. doi:10.1073/pnas.0901931106.

24. Arnaudo N, Fernández IS, McLaughlin SH, Peak-Chew SY, Rhodes D, Martino F. The N-terminal acetylation of Sir3 stabilizes its binding to the nucleosome core particle. *Nat Struct Mol Biol*. 2013;20(9):1119-1121. doi:10.1038/nsmb.2641.

25. Scott DC, Monda JK, Bennett EJ, Harper JW, Schulman BA. N-Terminal Acetylation Acts as an Avidity Enhancer Within an Interconnected Multiprotein Complex. *Science*. 2011;334:674-678. doi:10.1126/science.1209307.

26. Behnia R, Panic B, Whyte JRC, Munro S. Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p. *Nat Cell Biol*. 2004;6(5):405-413. doi:10.1038/ncb1120.

27. Dikiy I, Eliezer D. N-terminal Acetylation stabilizes N-terminal Helicity in Lipid- and Micelle-bound α-Synuclein and increases its affinity for Physiological Membranes. *J Biol Chem*. 2014;289(6):3652-3665. doi:10.1074/jbc.M113.512459.

28. Arnesen T, Starheim KK, Van Damme P, et al. The Chaperone-Like Protein HYPK Acts Together with NatA in Cotranslational N-Terminal Acetylation and Prevention of Huntingtin Aggregation. *Mol Cell Biol*. 2010;30(8):1898-1909. doi:10.1128/MCB.01199-09.

29. Holmes WM, Mannakee BK, Gutenkunst RN, Serio TR. Loss of amino-terminal acetylation suppresses a prion phenotype by modulating global protein folding. *Nat Commun*. 2014;5:1-11. doi:10.1038/ncomms5383.

30. Cheol-Sang H, Shemorry A, Varshavsky A. N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals. *Science*. 2010;327(5968):973-977. doi:10.1126/science.1183147.N-Terminal.

31. Nguyen KT, Mun S-H, Lee C-S, Hwang C-S. Control of protein degradation by N-terminal acetylation and the N-end rule pathway. *Exp Mol Med*. 2018;50(7):91. doi:10.1038/s12276-018-0097-y.

32. Shemorry A, Hwang C-S, Varshavsky A. Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway. *Mol Cell*. 2013;50(4):540-551. doi:10.1016/j.molcel.2013.03.018.

33. Kats I, Khmelinskii A, Kschonsak M, et al. Mapping Degradation Signals and Pathways in a Eukaryotic N-terminome. *Mol Cell*. 2018;70(3):488-501.e5. doi:10.1016/j.molcel.2018.03.033.

34. Soppa J. Protein acetylation in archaea, bacteria, and eukaryotes. *Archaea*. 2010;2010. doi:10.1155/2010/820681.

35. Tanka S, Matsushita Y, Yoshikawa A, Isono K. Cloning and molecular characterization of the generimL wich encodes an enzyme acetylating ribosomal protein L12 ofEscherichia coki K12. *MGG Mol Gen Genet*. 1989;217(2-3):289-293. doi:10.1007/BF02464895.

36. Yoshikawa A, Isono S, Sheback A, Isono K. Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12. *MGG Mol Gen Genet*. 1987;209(3):481-488. doi:10.1007/BF00331153.

37. Aksnes H, Drazic A, Marie M, Arnesen T. First Things First: Vital Protein Marks by N-Terminal

Acetyltransferases. *Trends Biochem Sci*. 2016;41(9):746-760. doi:10.1016/j.tibs.2016.07.005.

38. Polevoda B, Brown S, Cardillo TS, Rigby S, Sherman F. Yeast Nα-terminal acetyltransferases are associated with ribosomes. *J Cell Biochem*. 2008;103(2):492-508. doi:10.1002/jcb.21418.

39. Aksnes H, Van Damme P, Goris M, et al. An organellar nα-acetyltransferase, naa60, acetylates cytosolic n termini of transmembrane proteins and maintains golgi integrity. *Cell Rep*. 2015. doi:10.1016/j.celrep.2015.01.053.

40. Dinh T V., Bienvenut W V., Linster E, et al. Molecular identification and functional characterization of the first Nα-acetyltransferase in plastids by global acetylome profiling. *Proteomics*. 2015;15(14):2426-2435. doi:10.1002/pmic.201500025.

41. Polevoda B, Norbeck J, Takakura H, Blomberg A, Sherman F. Identification and specificities of N-terminal acetyltransferases from Saccharomyces cerevisiae. *EMBO J*. 1999;18(21):6155-6168. doi:10.1093/emboj/18.21.6155.

42. Van Damme P, Lasa M, Polevoda B, et al. N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. *PNAS*. 2012;109(31):12449-12454. doi:10.1073/pnas.1210303109.

43. Van Damme P, Kalvik T V., Starheim KK, et al. A Role for Human N-alpha Acetyltransferase 30 (Naa30) in Maintaining Mitochondrial Integrity. *Mol Cell Proteomics*. 2016;15(11):3361-3372. doi:10.1074/mcp.M116.061010.

44. Van Damme P, Evjenth R, Foyn H, et al. 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. *Mol Cell Proteomics*. 2011;10(5):M110.004580. doi:10.1074/mcp.M110.004580.

45. Van Damme P, Hole K, Gevaert K, Arnesen T. N-terminal acetylome analysis reveals the specificity of Naa50 (Nat5) and suggests a kinetic competition between N-terminal acetyltransferases and methionine aminopeptidases. *Proteomics*. 2015;15(14):2436-2446. doi:10.1002/pmic.201400575.

46. Wiame E, Tahay G, Tyteca D, et al. NAT6 acetylates the N-terminus of different forms of actin. *FEBS J*. 2018;5:3299-3316. doi:10.1111/febs.14605.

47. Vetting MW, Luiz LP, Yu M, et al. Structure and functions of the GNAT superfamily of acetyltransferases. *Arch Biochem Biophys*. 2005;433(1):212-226. doi:10.1016/j.abb.2004.09.003.

48. Dyda F, Klein DC, Hickman AB. GCN5-Related N-Acetyltransferases: A Structural Overview. *Annu Rev Biophys Biomol Struct*. 2000. doi:10.1146/annurev.biophys.29.1.81.

49. Neuwald AF, Landsman D. GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. *Trends Biochem Sci*. 1997;22(5):154-155. doi:10.1016/S0968-0004(97)01034-7.

50. Trievel RC, Rojas JR, Sterner DE, et al. Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. *PNAS*. 1999;96(16):8931-8936. doi:10.1016/j.cbi.2017.04.011.

51. Evjenth RH, Brenner AK, Thompson PR, Arnesen T, Frøystein NÅ, Lillehaug JR. Human protein N-terminal acetyltransferase hNaa50p (hNAT5/hSAN) follows ordered sequential catalytic mechanism: Combined kinetic and NMR study. *J Biol Chem*. 2012;287(13):10081-10088. doi:10.1074/jbc.M111.326587.

52. Liszczak G, Goldberg JM, Foyn H, Petersson EJ, Arnesen T, Marmorstein R. Molecular basis for N-terminal acetylation by the heterodimeric NatA complex. *Nat Struct Mol Biol*. 2013;20(9):1098-1105. doi:10.1038/nsmb.2636.

53. Gautschi M, Just S, Mun A, et al. The yeast N(alpha)-acetyltransferase NatA is quantitatively anchored to the ribosome and interacts with nascent polypeptides. *Mol Cell Biol*. 2003;23(20):7403-7414. doi:0.1128/MCB.23.20.

54. Gottlieb L, Marmorstein R. Structure of Human NatA and Its Regulation by the Huntingtin Interacting Protein HYPK. *Structure*. 2018;26(7):925-935.e8. doi:10.1016/j.str.2018.04.003.

55. Arnesen T, Anderson D, Torsvik J, Halseth HB, Varhaug JE, Lillehaug JR. Cloning and characterization of hNAT5/hSAN: An evolutionarily conserved component of the NatA protein N-α-acetyltransferase complex. *Gene*. 2006;371(2):291-295. doi:10.1016/j.gene.2005.12.008.

56. Weyer FA, Gumiero A, Lapouge K, Bange G, Kopp J, Sinning I. Structural basis of HypK regulating N-Terminal acetylation by the NatA complex. *Nat Commun*. 2017;8. doi:10.1038/ncomms15726.

57. Yoon H, Kim HL, Chun YS, et al. NAA10 controls osteoblast differentiation and bone formation as a feedback regulator of Runx2. *Nat Commun*. 2014;5(May):1-14. doi:10.1038/ncomms6176.

58. Ree R, Myklebust LM, Thiel P, Foyn H, Fladmark KE, Arnesen T. The N-terminal acetyltransferase Naa10 is essential for zebrafish development. *Biosci Rep*. 2015;35(5):e00249-e00249. doi:10.1042/BSR20150168.

59. Wang Y, Mijares M, Gall MD, et al. Drosophila variable nurse cells encodes arrest defective 1 (ARD1), the catalytic subunit of the major N-terminal acetyltransferase complex. *Dev Dyn*. 2010;239(11):2813-2827. doi:10.1002/dvdy.22418.

60. Ingram AK, Cross GAM, Horn D. Genetic manipulation indicates that ARD1 is an essential Nα-acetyltransferase in Trypanosoma brucei. *Mol Biochem Parasitol*. 2000;111(2):309-317. doi:10.1016/S0166-6851(00)00322-4.

61. Shin SH, Yoon H, Chun YS, et al. Arrest defective 1 regulates the oxidative stress response in human cells and mice by acetylating methionine sulfoxide reductase A. *Cell Death Dis*. 2014;5(10):1-12. doi:10.1038/cddis.2014.456.

62. Shin DH, Chun YS, Lee KH, Shin HW, Park JW. Arrest defective-1 controls tumor cell behavior by acetylating myosin light chain kinase. *PLoS One*. 2009;4(10). doi:10.1371/journal.pone.0007451.

63. Magin RS, March ZM, Marmorstein R. The N-terminal acetyltransferase Naa10/ARD1 does not acetylate lysine residues. *J Biol Chem*. 2016;291(10):5270-5277. doi:10.1074/jbc.M115.709428.

64. Kang J, Chun YS, Huh J, Park JW. FIH permits NAA10 to catalyze the oxygen-dependent lysyl-acetylation of HIF-1α. *Redox Biol*. 2018;19(August):364-374. doi:10.1016/j.redox.2018.09.002.

65. Sönnichsen B, Koski LB, Walsh A, et al. Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature*. 2005;434(7032):462-469. doi:10.1038/nature03353.

66. Lee CC, Peng SH, Shen L, et al. The Role of N-α-acetyltransferase 10 Protein in DNA Methylation and Genomic Imprinting. *Mol Cell*. 2017;68(1):89-103.e7. doi:10.1016/j.molcel.2017.08.025.

67. Arnesen T, Betts MJ, Pendino F, et al. Characterization of hARD2, a processed hARD1 gene duplicate, encoding a human protein N-α-acetyltransferase. *BMC Biochem*. 2006;7:1-12. doi:10.1186/1471-2091-7-13.

68. Linē A, Stengrēvics A, Slucka Z, Li G, Rees RC. Serological identification and expression analysis of gastric cancer-associated genes. *Br J Cancer*. 2002;86(11):1824-1830. doi:10.1038/sj.bjc.6600321.

69. Midorikawa Y, Tsutsumi S, Taniguchi H, et al. Identification of genes associated with dedifferentiation of hepatocellular carcinoma with expression profiling analysis. *Japanese J Cancer Res*. 2002;93(6):636-643. doi:10.1111/j.1349-7006.2002.tb01301.x.

70. Kalvik T V., Arnesen T. Protein N-terminal acetyltransferases in cancer. *Oncogene*. 2013;32(3):269-276. doi:10.1038/onc.2012.82.

71. Fluge Ø, Bruland O, Akslen LA, Varhaug JE, Lillehaug JR. NATH, a novel gene overexpressed in papillary thyroid carcinomas. *Oncogene*. 2002;21(33):5056-5068. doi:10.1038/sj.onc.1205687.

72. Martin DT, Gendron RL, Jarzembowski JA, et al. Tubedown expression correlates with the differentiation status and aggressiveness of neuroblastic tumors. *Clin Cancer Res*. 2007;13(5):1480-1487. doi:10.1158/1078-0432.CCR-06-1716.

73. Yu M, Ma M, Huang C, et al. Correlation of expression of human arrest-defective-1 (hard1) protein with breast cancer. *Cancer Invest*. 2009;27(10):978-983. doi:10.3109/07357900902769723.

74. Wang Z, Wang Z, Guo J, et al. Inactivation of androgen-induced regulator ARD1 inhibits androgen receptor acetylation and prostate tumorigenesis. *PNAS*. 2012;109(8):3053-3058. doi:10.1073/pnas.1113356109.

75. Chiti F, Dobson CM. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu Rev Biochem*. 2006;75(1):333-366. doi:10.1146/annurev.biochem.75.101304.123901.

76. Asaumi M, Iijima K, Sumioka A, et al. Interaction of N-terminal acetyltransferase with the cytoplasmic domain of β-amyloid precursor protein and its effect on Aβ secretion. *J Biochem*. 2005;137(2):147-155.

doi:10.1093/jb/mvi014.

77. Myklebust LM, Van Damme P, Støve SI, et al. Biochemical and cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects. *Hum Mol Genet*. 2014;24(7):1956-1976. doi:10.1093/hmg/ddu611.

78. Casey JP, Støve SI, McGorrian C, et al. NAA10 mutation causing a novel intellectual disability syndrome with Long QT due to N-terminal acetyltransferase impairment. *Sci Rep*. 2015;5(October):1-14. doi:10.1038/srep16022.

79. Cheng H, Dharmadhikari A V., Varland S, et al. Truncating Variants in NAA15 Are Associated with Variable Levels of Intellectual Disability, Autism Spectrum Disorder, and Congenital Anomalies. *Am J Hum Genet*. 2018;102(5):985-994. doi:10.1016/j.ajhg.2018.03.004.

80. Rauch A, Wieczorek D, Graf E, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: An exome sequencing study. *Lancet*. 2012;380(9854):1674-1682. doi:10.1016/S0140-6736(12)61480-9.

81. Thevenon J, Duffourd Y, Masurel-Paulet A, et al. Diagnostic odyssey in severe neurodevelopmental disorders: Toward clinical whole-exome sequencing as a first-line diagnostic test. *Clin Genet*. 2016;89(6):700-707. doi:10.1111/cge.12732.

82. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7(539). doi:10.1038/msb.2011.75.

83. Bateman A, Martin MJ, O'Donovan C, et al. UniProt: The universal protein knowledgebase. *Nucleic Acids Res*. 2017;45(D1):D158-D169. doi:10.1093/nar/gkw1099.

84. Coordinators NR. Database resources of the National Center for Biotechnology Information [ftp://ftp.ncbi.nih.gov/genomes/Bacteria]. 2016;44(November 2015):7-19. doi:10.1093/nar/gkv1290.

85. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2: A multiple sequence alignment and analysis workbench. *Bioinformatics*. 2009;25(9):1189-1191. doi:10.1093/bioinformatics/btp033.

86. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7(8):575-576. doi:10.1038/nmeth0810-575.

87. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 2010;(20):110-121. doi:10.1101/gr.097857.109.

88. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249. doi:10.1038/nmeth0410-248.

89. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 2012;40(W1):452-457. doi:10.1093/nar/gks539.

90. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. 2000;28(1):235-242. doi:10.1093/nar/28.1.235.

91. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*. 2004;32(WEB SERVER ISS.):665-667. doi:10.1093/nar/gkh381.

92. Dolinsky TJ, Czodrowski P, Li H, et al. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res*. 2007;35(Web Server issue):W522-W525. doi:10.1093/nar/gkm276.

93. Jurrus E, Engel D, Star K, et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci*. 2018;27(1):112-128. doi:10.1002/pro.3280.

94. Weiner SJ, Kollman PA, Singh UC, et al. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J Am Chem Soc*. 1984;106(3):765-784. doi:10.1021/ja00315a051.

95. Søndergaard CR, Olsson MHM, Rostkowski M, Jensen JH. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p Kavalues. *J Chem Theory Comput*. 2011;7(7):2284-2295. doi:10.1021/ct200133y.

96. Panja S, Saha S, Jana B, Basu T. Role of membrane potential on artificial transformation of E. coli with plasmid DNA. *J Biotechnol*. 2006;127:14-20. doi:10.1016/j.jbiotec.2006.06.008.

97. Birnboim HC, Doly J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res*. 1979;7(6):1513-1523. doi:10.1093/nar/7.6.1513.

98. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*. 1975;94(3):441-448. doi:10.1016/0022-2836(75)90213-2.

99. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *PNAS*. 1977;74(12):5463-5467. doi:10.1073/pnas.74.12.5463.

100. Smith LM, Sanders ZJ, Kaiser RJ, et al. Fluorescense detection in automated DNA sequence analysis. *Nature*. 1986;321(12):674-679. doi:10.1038/321674a0.

101. Ausubel FM, Brent R, Kingston RE, et al. Expression Using the T7 RNA Polymerase/Promoter System. *Curr Protoc Mol Biol*. 1990;11(1):4767-16.2.11. doi:10.1002/0471142727.mb1602s11.

102. Kapust RB, Waugh DS. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*. 1999;8(8):1668-1674. doi:10.1110/ps.8.8.1668.

103. LAEMMLI UK. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*. 1970;227(5259):680-685. doi:10.1038/227680a0.

104. French CS, Milner HW. Disintegration of bacteria and small particles by high-pressure extrusion. *Methods Enzymol*. 1955;1:64-67. doi:10.1016/0076-6879(55)01013-6.

105. Vanderheiden GJ, Fairchild AC, Jago GR. Construction of a Laboratory Press for Use with the French Pressure Cell. *Appl Microbiol*. 1970;19(5):875-877.

106. Lodish HF. Post-translational modification of proteins. *Enzym Microb Technol*. 1981;3:178-188. doi:10.1016/0141-0229(81)90084-3.

107. Cheung RCF, Wong JH, Ng TB. Immobilized metal ion affinity chromatography: A review on its applications. *Appl Microbiol Biotechnol*. 2012;96(6):1411-1420. doi:10.1007/s00253-012-4507-0.

108. Smith PK, Krohn RI, Hermanson GT, et al. Measurement of protein using bicinchoninic acid. *Anal Biochem*. 1985;150(1):76-85. doi:10.1016/0003-2697(85)90442-7.

109. Bradford MM. Rapid and sensitive method for the quantification of microgram quantities of protein untilizing the principle of protein dye-binding. *Anal Biochem*. 1976;(72):248-254. doi:10.1016/0003-2697(76)90527-3.

110. Foyn H, Thompson PR, Arnesen T. DTNB-Based Quantification of In Vitro Enzymatic N-Terminal Acetyltransferase Activity. In: O. S, ed. *Protein Terminal Profiling. Methods in Molecular Biology, Vol 1574*. New York: Press, Humana; 2017:9-15. doi:https://doi.org/10.1007/978-1-4939-6850-3_2.

111. Thompson PR, Wang D, Wang L, et al. Regulation of the p300 HAT domain via a novel activation loop. *Nat Struct Mol Biol*. 2004;11(4):308-315. doi:10.1038/nsmb740 308.

112. Ellman GL. Tissue sulfhydryl groups. *Arch Biochem Biophys*. 1959;82(1):70-77. doi:10.1016/0003-9861(59)90090-6.

113. Obrig TG, Culp WJ, McKeehan WL, Hardesty B. The mechanism by which cycloheximide and related glutarimide antibiotics inhibit peptide synthesis on reticulocyte ribosomes. *J Biol Chem*. 1971;246(1):174-181. http://www.ncbi.nlm.nih.gov/pubmed/2823261.

114. Schneider-Poetsch T, Ju J, Eyler DE, et al. Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat Chem Biol*. 2010;6(3):209-217. doi:10.1038/nchembio.304.

115. Müller F, Ackermann P, Margot P. Fungicides, agricultural, 2. individual fungicides. In: *Ullmann's Encyclopedia of Industrial Chamistry*. 6th ed. Chicago/Turabian: Wiley VCH.; 2012:407. doi:10.1002/14356007.o12_o06.

116. Tran JR, Brodsky JL. Assays to measure ER-associated degradation in yeast. *Methods Mol Biol*. 2012;832(10):505-518.

117. Buchanan BW, Lloyd ME, Engle SM, Rubenstein EM. Cycloheximide Chase Analysis of Protein Degradation in Saccharomyces cerevisiae *J Vis Exp*. 2016;(110):1-9. doi:10.3791/53975.

118. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to

microtubules and the ribosome. *PNAS*. 2001;(98):10037-10041. doi:https://doi.org/10.1073/pnas.181342398.