Amino-terminal acetylation of proteins: role in human disease and biology

Gholson J. Lyon, M.D. Ph.D.
Figure 4. NAT activity of recombinant hNaa10p WT or p.Ser37Pro towards synthetic N-terminal peptides. A) and B) Purified MBP-hNaa10p WT or p.Ser37Pro were mixed with the indicated oligopeptide substrates (200 µM for SESSS and 250 µM for DDDIA) and saturated levels of acetyl-CoA (400 µM). Aliquots were collected at indicated time points and the acetylation reactions were quantified using reverse phase HPLC peptide separation. Error bars indicate the standard deviation based on three independent experiments. The five first amino acids in the peptides are indicated, for further details see materials and methods. Time dependent acetylation reactions were performed to determine initial velocity conditions when comparing the WT and Ser37Pro NAT-activities towards different oligopeptides. C) Purified MBP-hNaa10p WT or p.Ser37Pro were mixed with the indicated oligopeptide substrates (200 µM for SESSS and AVFAD, and 250 µM for DDDIA and EEEIA) and saturated levels of acetyl-CoA (400 µM) and incubated for 15 minutes (DDDIA and EEEIA) or 20 minutes (SESSS and AVFAD), at 37°C in acetylation buffer. The acetylation activity was determined as above. Error bars indicate the standard deviation based on three independent experiments. Black bars indicate the acetylation capacity of the MBP-hNaa10p wild type (WT), while white bars indicate the acetylation capacity of the MBP-hNaa10p mutant p.Ser37Pro. The five first amino acids in the peptides are indicated.
I moved to Utah in July 2009 to find at least one new human disease, thus revealing new biology.

- **July 2009-December 2009:** Attended weekly genetics case conference in which 10-30 genetic cases are presented weekly, led by Dr. Alan Rope and attended by Drs. John Carey and John Opitz.

- **There are indeed MANY idiopathic disorders not described in the literature, many of which have neuropsychiatric manifestations.** I thought about hundreds of such cases, looking for the ideal first family to sequence.
Discovering a new syndrome and its genetic basis.
I met the entire family on March 29, 2010

Photo of mother with son in late 1970’s
This is the first boy in the late 1970’s.

First boy. Called "a little old man" by the family. Died around ~1 year of age, from cardiac arrhythmias.
prominence of eyes, down-sloping palpebral fissures, thickened eyelids, large ears, beaking of nose, flared nares, hypoplastic nasal alae, short columella, protruding upper lip, micro-retrognathia
Family now in October 2011, with five mutation-positive boys dying from the disease.
An unrelated second family was also identified, due to sharing the same genotype, i.e. the same mutation.
We found the SAME mutation in two unrelated families, with a very similar phenotype in both families, helping prove that this genotype contributes to the phenotype observed.
These are the Major Features of the Syndrome.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth</td>
<td>post-natal growth failure</td>
</tr>
<tr>
<td>Development</td>
<td>global, severe delays</td>
</tr>
<tr>
<td>Facial</td>
<td>prominence of eyes, down-sloping palpebral fissures, thickened lids, large ears, beaking of nose, flared nares, hypoplastic alae, short columella, protruding upper lip, micro-retrognathia</td>
</tr>
<tr>
<td>Skeletal</td>
<td>delayed closure of fontanels, broad great toes</td>
</tr>
<tr>
<td>Integument</td>
<td>redundancy / laxity of skin, minimal subcutaneous fat, cutaneous capillary malformations</td>
</tr>
<tr>
<td>Cardiac</td>
<td>structural anomalies (ventricular septal defect, atrial level defect, pulmonary artery stenoses), arrhythmias (Torsade de points, PVCs, PACs, SVtach, Vtach), death usually associated with cardiogenic shock preceded by arrhythmia.</td>
</tr>
<tr>
<td>Genital</td>
<td>inguinal hernia, hypo- or cryptorchidism</td>
</tr>
<tr>
<td>Neurologic</td>
<td>hypotonia progressing to hypertonia, cerebral atrophy, neurogenic scoliosis</td>
</tr>
</tbody>
</table>

Shaded regions include features of the syndrome demonstrating variability. Though variable findings of the cardiac, genital and neurologic systems were observed, all affected individuals manifested some pathologic finding of each.
A

B

II-1  II-6  III-4  III-6  III-7

C

D

II-1  III-2
◆ We performed X-chromosome exon capture with Agilent, followed by Next Gen Sequencing with Illumina.

◆ We analyzed the data with ANNOVAR and VAAST (Variant Annotation, Analysis and Search Tool). New computational tools for identifying disease-causing mutations by individual genome sequencing.


VAAST integrates AAS & Variant frequencies in a single probabilistic framework

- non-coding variants scored using allele frequency differences

- \( n_i \): frequency of variant type among all variants observed in Background and Target genomes

- \( a_i \): frequency of variant type among disease causing mutations in OMIM

- This approach means that every variant can be scored, non-synonymous, synonymous, coding, and non-coding. Phylogenetic conservation not required.
Analysis with VAAST readily identified a few likely candidates.

<table>
<thead>
<tr>
<th>SNV calling pipeline</th>
<th>GATK</th>
<th>Samtools</th>
<th>GNUMAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>III-4 (total SNVs)</td>
<td>1546</td>
<td>1499</td>
<td>2168</td>
</tr>
<tr>
<td>III-4 (nsSNVs)</td>
<td>146</td>
<td>114</td>
<td>155</td>
</tr>
<tr>
<td>VAAST candidate genes (NAA10 ranking)</td>
<td>4 (3)</td>
<td>3 (2)</td>
<td>5 (2)</td>
</tr>
<tr>
<td>Present in III-4 and mother II-2 (nsSNVs)</td>
<td>122</td>
<td>107</td>
<td>116</td>
</tr>
<tr>
<td>VAAST candidate genes (NAA10 ranking)</td>
<td>3 (2)</td>
<td>2 (1)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Present in III-4, mother II-2, and grandmother I-2 (nsSNVs)</td>
<td>115</td>
<td>95</td>
<td>104</td>
</tr>
<tr>
<td>VAAST candidate genes (NAA10 ranking)</td>
<td>2 (1)</td>
<td>2 (1)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Present in III-4, II-2, and I-2, absent in brother III-2 and uncle II-8 (nsSNVs)</td>
<td>8</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>VAAST candidate genes (NAA10 ranking)</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td>2 (1)</td>
</tr>
</tbody>
</table>
This is the mutation we found… one nucleotide change out of 6 billion nucleotides in a diploid genome…

\[
\begin{align*}
\text{Pro37} & \quad \text{C C C C} \\
\text{Ser37} & \quad \text{C T T T T C C T G G} \\
\end{align*}
\]
Proving Relevance of the mutation

- Present in two unrelated families with very similar phenotype of affected boys.

- Blinded Sanger sequencing showed perfect segregation of the mutation with the disease. Mutation present in Proband, Carrier Mother, Carrier Grandmother and other carrier mothers. Absent in unaffected brother and unaffected uncle.

- Also present in DNA from formalin-fixed paraffin-embedded tissue from two other deceased affected boys, found in pathology department, saved in one case for 30 years.

- Mutation NOT present in ~6000 exomes or genomes sequenced at BGI, CHOP and Utah for other projects.
Ogden Syndrome, in honor of where the first family lives, in Ogden, Utah
The mutation is a missense resulting in Serine to Proline change in Naa10p

- Ser 37 is conserved from yeast to human
- Ser37Pro is predicted to affect functionality (SIFT and other prediction programs)
- Structural modelling of hNaa10p wt (cyan) and S37P (pink)
The mutation disrupts the N-terminal acetylation machinery (NatA) in human cells.

Slide courtesy of Thomas Arnesen
**Nε-acetylation: Lysine Acetylation (KATs, HATs)**

![Chemical structure of Nε-acetylation](image)

**Figure 1.3: Nε-acetylation.** Lysine acetyltransferases (KATs) transfer the acetyl group of acetyl-CoA to the side chain of lysine, and thus remove the positive charge of the amino acid. Deacetylases (HDACs) catalyze the reverse reaction.

---

Annette Katharina Brenner  
Dissertation for the degree philosophiae doctor (PhD)  
2012  
at the University of Bergen
N$^\alpha$-acetylation: Amino-terminal Acetylation

In N$^\alpha$-acetylation, the acetyl group of acetyl-CoA is transferred to the backbone of the N-terminal amino acid of the target protein (figure 1.4).

**Figure 1.4:** N$^\alpha$-acetylation. The acetyl group of acetyl-CoA is irreversibly transferred to the N-terminal amino acid of the target protein by an N-terminal acetyltransferase (NAT).
N-terminal processing in human cells

Van Damme P et al., PLoS Genet, 2011
Protein acetylation in higher eukaryotes

1. Cotranslational N-alpha acetylation
   - Irreversible
   - N-terminus
   - Cytoplasm

2. Posttranslational N-epsilon acetylation
   - Reversible
   - Lysine
   - Nucleus-cytoplasm

3. Posttranslational N-alpha acetylation
   - Irreversible?
   - N-terminus
   - Vesicles etc.
NAT activity of recombinant hNaa10p WT or p.Ser37Pro towards synthetic N-terminal peptides

Assay performed in Thomas Arnesen lab
Protein Expression and purification

Expression plasmid was transformed in chemically competent E. coli BL21 DE3 (NEB). A starting culture (100 ml LB media + antibiotic) was incubated for 12 h at 37°C. 30 ml of this culture were added to 400 ml expression culture (LB media + 1 % (m/v) glucose + antibiotic) and were incubated at 37°C until an optical density of 0.8 was reached. Expression was induced with 0.5 mM IPTG (Milipore) at 30°C for 1h. The bacteria were pelleted at 6000 x g for 15 min and resuspended in 8 ml lysis buffer (40 mM Tris/Cl, pH 8.0; 100 mM NaCl) or HIS-lysis buffer (50 mM NaH2PO4, 300 mM NaCl, pH 8,0). The cells were ultrasonificated on ice and cellular debris was pelleted at 20,800 x g for 30 min at 4°C. Protein was bound to 0.7 ml GSH beads/Glutathione sepharose (Sigma-Aldrich) or Ni-NTA (Qiagen), the beads were washed twice with lysis buffer or HIS-lysis buffer + 50 mM imidazole and purified protein was eluted with 20 mM L-Glutathione (Sigma-Aldrich) or 250 mM imidazole. Fractions of 0.5 ml were collected and aliquots separated on SDS-PAGE.
Peptide substrate: SYSMEHFRWGKPVGKRRPVKYP and corresponds to amino acids 1-24 of the human adrenocorticotropic hormone (ACTH)
Big Questions though:

Simulated structure of S37P mutant

What is the molecular basis of Ogden syndrome?
- Naa10/Naa15 complex
- Naa10 localisation
- Naa10 function

what can we learn from ogden syndrome?
- characterizing different model systems (fibroblasts, yeast, *C. elegans*)
These are the Major Features of the Syndrome.

<table>
<thead>
<tr>
<th>Table 1. Features of the syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Growth</strong></td>
</tr>
<tr>
<td><strong>Development</strong></td>
</tr>
<tr>
<td><strong>Facial</strong></td>
</tr>
<tr>
<td><strong>Skeletal</strong></td>
</tr>
<tr>
<td><strong>Integument</strong></td>
</tr>
<tr>
<td><strong>Cardiac</strong></td>
</tr>
<tr>
<td><strong>Genital</strong></td>
</tr>
<tr>
<td><strong>Neurologic</strong></td>
</tr>
</tbody>
</table>

Shaded regions include features of the syndrome demonstrating variability. Though variable findings of the cardiac, genital and neurologic systems were observed, all affected individuals manifested some pathologic finding of each.
Family 1: II-1
Family 1: II-1
Family 1: II-6
Family 1: III-4
Family 1: III-6
Family 1: III-6
These are the Major Features of the Syndrome.

<table>
<thead>
<tr>
<th>Table 1. Features of the syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Growth</strong></td>
</tr>
<tr>
<td><strong>Development</strong></td>
</tr>
<tr>
<td><strong>Facial</strong></td>
</tr>
<tr>
<td><strong>Skeletal</strong></td>
</tr>
<tr>
<td><strong>Integument</strong></td>
</tr>
<tr>
<td><strong>Cardiac</strong></td>
</tr>
<tr>
<td><strong>Genital</strong></td>
</tr>
<tr>
<td><strong>Neurologic</strong></td>
</tr>
</tbody>
</table>

Shaded regions include features of the syndrome demonstrating variability. Though variable findings of the cardiac, genital and neurologic systems were observed, all affected individuals manifested some pathologic finding of each.
Family 2: II-1
Family 2: II-1
Family 2: III-2
Family 2: III-2
Family 2: III-4
NAA10/NatA is essential for life

*C. elegans*

*T. Brucei*

*D. melanogaster*

The human N-Acetylome

- A majority of soluble human proteins are N-terminally acetylated
- NatA is a major protein modifying enzyme of the human proteome

Figure 1. Overview of human NATs and their substrate specificities. The human NATs are composed of catalytic subunits (yellow) and auxiliary subunits (green), and all are associated with ribosomes (blue). NatA potentially acetylates Ser-, Ala-, Thr-, Gly- and Val- N-termini after the iMet has been removed by methionine aminopeptidases (MetAPs). NatB potentially acetylates Met-Asp-, Met-Glu- and Met-Asn-, whereas NatC may target Met-Leu-, Met-Ile-, Met-Phe- and Met-Trp-. NatD apparently only acetylates the Ser- N-termini of histones H2A and H4. NatE and NatF demonstrate some specificity towards NatC-type substrates as well as Met-Lys-, Met-Ala-, Met-Met- and Met-Val-. *NatA or hNaa10p may also mediate post-translational acetylation of mature actins harboring Asp- and Glu- N-termini.

REVIEW
Protein N-terminal acetyltransferases in cancer

TV Kalvik¹ and T Arnesen¹,²
Naa10 function

cytosolic  
(co-translational)

• Rho (Rac/Cdc42 cell migration)  
  • Caspase2 (apoptosis)  
• TSC2 (mTOR cell proliferation)  
  • HIF-1α (angiogenesis)  
    • α-Tubulin

protein stability, activity and sorting
Naa10 function

**Cytosolic (co-translational)**
- Rho (Rac/Cdc42 cell migration)
  - Caspase2 (apoptosis)
- TSC2 (mTOR cell proliferation)
  - HIF-1α (angiogenesis)
  - α-Tubulin

**Nuclear transcription factors**
- DNMT1 (E-Cad transformation)
- RelA/p65 (MCL-1 survival)
- Rip1/NEMO (DNA damage)
- β-Catenin/Lef-1/TEF (wnt-pathway)
  - AR (proliferation)

Protein stability, activity, and sorting
localization of Naa10

Immunofluorescence

- HEK293 cells +/- V5-hNaa10 wt or mutants
- IF with anti-V5

V5-Naa10 wt
interaction of Naa10 and Naa15

Co-IP of Naa10 and Naa15 in HEK293 cells
• precipitating antibody: α-Myc

→ reduced interaction of Naa15 and Naa10 S37P?
interaction of Naa10 and Naa15

Co-IP of Naa10 and Naa15 in HEK293 cells
  • precipitating antibody: α-V5

reduced interaction of Naa15 and Naa10 S37P?
Molecular basis for N-terminal acetylation by the heterodimeric NatA complex

Glen Liszczak¹,², Jacob M Goldberg², Håvard Foyn³, E James Petersson², Thomas Arnesen³,⁴ & Ronen Marmorstein¹,²

Figure 1 Overall structure of the NatA complex bound to acetyl CoA. (a) Naa10p (teal) and Naa15p (brown) subunits, shown in cartoon, bound to acetyl CoA (CPK coloring and stick format). Only Naa15p helices that contact Naa10p are labeled. The dashed brown line represents a disordered loop region in Naa15p. The dimensions of the complex are 107 Å × 85 Å × 70 Å. (b) A 90° rotation of the view in a. Helices that are depicted in c and d are labeled. (c) Zoom view highlighting key residues that compose the predominantly hydrophobic interface between Naa10p α1–α2 and Naa15p α29–α30. (d) Zoom view of the intersubunit interface at the C-terminal region of Naa10p α1 and the Naa15p α25–α27–α28 helices.
Crystal Structure of NAA50

FIGURE 1. Overall structure of the ternary Naa50p·CoA·peptide complex. A, structure of the ternary complex showing Naa50p in teal; CoA as magenta sticks with carbon, nitrogen, and sulfur atoms in Corey-Pauling-Koltun coloring; and the substrate peptide as yellow sticks with carbon, nitrogen, and sulfur atoms in Corey-Pauling-Koltun coloring. Substrate peptide electron density obtained from a composite omit map (blue) is shown contoured to 1.5σ. B, superposition of the ternary Naa50p complex with the ternary Gcn5

Glen Liszczak‡§, Thomas Arnesen¶‖, and Ronen Marmorstein‡§

© 2011 by The American Society for Biochemistry and Molecular Biology, Inc. Published, JBC Papers in Press, September 6, 2011, DOI 10.1074/jbc.M111.282863
De novo mutations in histone-modifying genes in congenital heart disease

Samir Zaidi1,2*, Murim Choi1,2*, Hiroko Wakimoto3, Lijiang Ma4, Jianming Jiang3,5, John D. Overton1,6,7, Angela Romano-Adesman8, Robert D. Bjornson7,9, Roger E. Breitbart10, Kerry K. Brown3, Nicholas J. Carriero7,9, Yee Him Cheung11, John Deanfield12, Steve DePalma3, Khalid A. Fakhro1,2, Joseph Glessner13, Hakon Hakonarson13,14, Michael J. Italia15, Jonathan R. Kaltman16, Juan Kaski12, Richard Kim17, Jennie K. Kline18, Teresa Lee4, Jeremy Leipzig15, Alexander Lopez1,6,7, Shrikant M. Mane1,6,7, Laura E. Mitchell9, Jane W. Newburger10, Michael Parfenov3, Itsik Pe’er20, George Porter21, Amy E. Roberts10, Ravi Sachidanandam22, Stephan J. Sanders1,23, Howard S. Seiden24, Mathew W. State1,23, Sailakshmi Subramanian22, Irina R. Tikhonova1,6,7, Wei Wang15,25, Dorothy Warburton4,26, Peter S. White14,15, Ismee A. Williams4, Hongyu Zhao1,27, Jonathan G. Seidman3, Martina Brueckner1,28, Wendy K. Chung4,29, Bruce D. Gelb22,24,30, Elizabeth Goldmuntz14,31, Christine E. Seidman3,5,32 & Richard P. Lifton1,2,6,7,33
lacked cardinal disease manifestations or had atypical cardiac features. Genes known to cause Mendelian CHD; however, affected subjects

Del, deletion; Dx, diagnosis; fs, frameshift mutation; fs

* syndrome;

NAA15 inhibits NEDD8, a cofactor for cullin-based ubiquitin ligases. Last, canals) for CHARGE syndrome

controls (expected by chance (in these genes (Table 2 and Supplementary Table 13), more than

system studies. There were 13 CHD probands with

ate CHD genes (Supplementary Table 12) from human and model

once (for example, mutations had dextrocardia with unbalanced complete atrioventricu-

role in left–right axis determination

Supplementary Table 10). Other structural, neurodevelopmental and

H3K4me pathway revealed diverse cardiac phenotypes (Table 2 and

Table 2 | Genes of interest with de novo mutations in probands

<table>
<thead>
<tr>
<th>ID</th>
<th>Gene</th>
<th>Mutation</th>
<th>Dx</th>
<th>Other structural/neuro/ht-wt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-00596</td>
<td>MLL2†</td>
<td>p.Ser1722Arg fs*9</td>
<td>LVO</td>
<td>Y/Y/N</td>
</tr>
<tr>
<td>1-00853</td>
<td>WDR5†</td>
<td>p.Lys7Gln</td>
<td>CTD</td>
<td>N/Y/N</td>
</tr>
<tr>
<td>1-00534</td>
<td>CHD7†</td>
<td>p.Gln1599*</td>
<td>CTD</td>
<td>Y/Y/Y</td>
</tr>
<tr>
<td>1-00230</td>
<td>KDM5A†</td>
<td>p.Arg1508Trp</td>
<td>LVO</td>
<td>N/N/Y</td>
</tr>
<tr>
<td>1-01965</td>
<td>KDM5B†</td>
<td>p.IVS12 + 1 G&gt;A</td>
<td>LVO</td>
<td>N/N/Y</td>
</tr>
<tr>
<td>1-0197</td>
<td>UBE2B†</td>
<td>p.Arg8Thr</td>
<td>CTD</td>
<td>N/N/N</td>
</tr>
<tr>
<td>1-00075</td>
<td>RNF20†</td>
<td>p.Gln83*</td>
<td>HTX</td>
<td>Y/Y/Y</td>
</tr>
<tr>
<td>1-01260</td>
<td>USP44†</td>
<td>p.Glu71Asp</td>
<td>LVO</td>
<td>N/N/N</td>
</tr>
<tr>
<td>1-02020</td>
<td>SMAD2‡‡</td>
<td>p.IVS6 + 1 G&gt;A</td>
<td>HTX</td>
<td>Y/N/N</td>
</tr>
<tr>
<td>1-02621</td>
<td>SMAD2‡‡</td>
<td>p.Trp24Cys</td>
<td>HTX</td>
<td>Y/NA/N</td>
</tr>
<tr>
<td>1-01451</td>
<td>MED20</td>
<td>p.IVS2 + 2 T&gt;C</td>
<td>HTX</td>
<td>N/Y/Y</td>
</tr>
<tr>
<td>1-01151</td>
<td>SUV420H1</td>
<td>p.Arg143Cys</td>
<td>CTD</td>
<td>N/Y/N</td>
</tr>
<tr>
<td>1-00750</td>
<td>HUWE1</td>
<td>p.Arg3219Cys</td>
<td>LVO</td>
<td>N/Y/N</td>
</tr>
<tr>
<td>1-00577</td>
<td>CUL3</td>
<td>p.Iso145Phe fs*23</td>
<td>LVO</td>
<td>Y/Y/N</td>
</tr>
<tr>
<td>1-00116</td>
<td>NUB1</td>
<td>p.As310His</td>
<td>CTD</td>
<td>Y/Y/Y</td>
</tr>
<tr>
<td>1-01828</td>
<td>DAPK3</td>
<td>p.Pro193Leu</td>
<td>CTD</td>
<td>N/N/NA</td>
</tr>
<tr>
<td>1-03151</td>
<td>SUPT5H</td>
<td>p.Glu451Asp</td>
<td>LVO</td>
<td>N/NA/N</td>
</tr>
<tr>
<td>1-00455</td>
<td>NAA15†</td>
<td>p.Lys336Lys fs*6</td>
<td>HTX</td>
<td>Y/Y/N</td>
</tr>
<tr>
<td>1-00141</td>
<td>NAA15†</td>
<td>p.Ser761*</td>
<td>CTD</td>
<td>N/NA/Y</td>
</tr>
<tr>
<td>1-01138</td>
<td>USP34</td>
<td>p.Leu472pro</td>
<td>LVO</td>
<td>N/NA/N</td>
</tr>
<tr>
<td>1-00448</td>
<td>NF1</td>
<td>p.IVS6 + 4 del A</td>
<td>CTD</td>
<td>N/NA/N</td>
</tr>
<tr>
<td>1-00802</td>
<td>PTCH1</td>
<td>p.Arg1351Gln</td>
<td>LVO</td>
<td>N/NA/N</td>
</tr>
<tr>
<td>1-02458</td>
<td>SOS1</td>
<td>p.Thr266Lys</td>
<td>Other</td>
<td>Y/Y/Y</td>
</tr>
<tr>
<td>1-02952</td>
<td>PITX2</td>
<td>p.Ala47Val</td>
<td>LVO</td>
<td>N/NA/N</td>
</tr>
<tr>
<td>1-01913</td>
<td>RAB10</td>
<td>p.As112Ser</td>
<td>Other</td>
<td>N/NA/N</td>
</tr>
<tr>
<td>1-00638</td>
<td>FBN2</td>
<td>p.As191Asn</td>
<td>CTD</td>
<td>N/NA/N</td>
</tr>
<tr>
<td>1-00197</td>
<td>BCL9</td>
<td>p.Met1395Lys</td>
<td>LVO</td>
<td>N/NA/N</td>
</tr>
<tr>
<td>1-02598</td>
<td>LRP2</td>
<td>p.Glu4372Lys</td>
<td>HTX</td>
<td>N/NA/N</td>
</tr>
</tbody>
</table>

Gene symbols are as in NCBI RefSeq database. Other structural/neuro/ht-wt denotes presence (Y) or absence (N) of other structural abnormalities, impaired cognitive speech or motor development, and height (ht) and/or weight (wt) less than 5th percentile for age, respectively. Further clinical details in Supplementary Tables 10 and 11. Associated syndromes: MLL2, Kabuki syndrome; CHD7, CHARGE syndrome; CUL3, pseudohypoaldosteronism, type 2E.

* Premature termination mutation.
† Gene involved in production, removal or reading of H3K4 methylation mark.
‡‡ Gene involved in removal of H3K27 methylation mark.
Del, deletion; Dx, diagnosis; fs, frameshift mutation; fs*n, frameshift mutation followed by premature termination n codons later; NA, data not available.
Table S10. Chromatin modifying and other genes of interest with de novo mutations in CHD probands

<table>
<thead>
<tr>
<th>ID</th>
<th>Gene</th>
<th>Heart Exp</th>
<th>Mutation</th>
<th>Primary Classification: Specific Cardiovascular Diagnoses</th>
<th>Extracardiac Structural Anomalies</th>
<th>Neuro-Developmental</th>
<th>Somatic Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-00455</td>
<td>NAA15</td>
<td>214</td>
<td>p.Lys335Lys fs*6</td>
<td>HTX: Dextrocardia, TAPVR, LSVC, hypoplastic TV, DORV, hypoplastic RV, D-TGA, PS</td>
<td>Hydronephrosis, asplenia, malrotation</td>
<td>normal</td>
<td>50</td>
</tr>
<tr>
<td>1-00141</td>
<td>NAA15</td>
<td>214</td>
<td>p.Ser761*</td>
<td>CTD: TOF, single LCA</td>
<td>No</td>
<td>n/a</td>
<td>&lt;5</td>
</tr>
</tbody>
</table>

¹Heart expression refers to # reads per million at murine e14.5. Mutation denotes the impact on encoded protein in three letter code; * denotes termination mutation. Frameshift mutation in MLL2, CUL3, and NAA15. ‘IVS’ stands for intervening sequence. ‘fs’ stands for frameshift. Splice site mutation in KDM5B, MED20, and SMAD2 occur at 1st base of canonical splice donor of intron 12, at 2nd base of canonical splice donor of intron 2 and 1st base of canonical splice donor of intron 6 respectively.

⁸HLHS-hypoplastic left heart syndrome; Dbl AA-double aortic arch; TOF-tetralogy of Fallot; PAPVR-partial anomalous pulmonary venous return; LSVC-left superior vena cava; LA-left atrium; CAVC-complete atroventricular canal defect; TAPVR-total anomalous pulmonary venous return; MV-mitral valve; BAV-bicuspid aortic valve; ASD-atrial septal defect; VSD-ventricular septal defect; PA-pulmonary atresia; RA-right atrial isomerization.

Table S4. All de novo mutations with Bayesian QS≥50

<table>
<thead>
<tr>
<th>ID</th>
<th>Primary Cardiac Classification</th>
<th>Gene</th>
<th>Mutation</th>
<th>Amino Acid Change</th>
<th>dbSNP</th>
<th>RefSeq NM Accession IDs</th>
<th>RefSeq NP Accession IDs</th>
<th>Chr</th>
<th>Position (hg19)</th>
<th>Base change</th>
<th>Mean Heart Exp.</th>
<th>Variant Quality Score</th>
<th>Ref Cov</th>
<th>Nonref Cov</th>
<th>Ref Cov</th>
<th>Nonref Cov</th>
<th>Ref Cov</th>
<th>Nonref Cov</th>
<th>Ref Cov</th>
<th>Nonref Cov</th>
<th>Bayesian Quality Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-00141</td>
<td>CTD</td>
<td>NAA15</td>
<td>Nonsense</td>
<td>S761X</td>
<td>Novel</td>
<td>NM_057175</td>
<td>NP_476516</td>
<td>4</td>
<td>140306112</td>
<td>C&gt;A</td>
<td>213.74</td>
<td>196</td>
<td>16</td>
<td>12</td>
<td>30</td>
<td>0</td>
<td>41</td>
<td>0</td>
<td>92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-00455</td>
<td>HTX</td>
<td>NAA15</td>
<td>Frameshift</td>
<td>D335</td>
<td>Novel</td>
<td>NM_057175</td>
<td>NP_476516</td>
<td>4</td>
<td>140272757</td>
<td>-AAAG</td>
<td>213.74</td>
<td>Indel-Pass</td>
<td>68</td>
<td>17</td>
<td>57</td>
<td>0</td>
<td>72</td>
<td>0</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-01119</td>
<td>CTD</td>
<td>NAA16</td>
<td>Missense</td>
<td>R70C</td>
<td>Novel</td>
<td>NM_024561</td>
<td>NP_078837</td>
<td>13</td>
<td>41893010</td>
<td>C&gt;T</td>
<td>12.47</td>
<td>228</td>
<td>142</td>
<td>121</td>
<td>234</td>
<td>0</td>
<td>219</td>
<td>0</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: For the table, additional columns are not shown in the image. The table includes information on the type of mutation, its impact on the encoded protein, and the associated phenotypic features.
“Model Organisms” to study this in?

- S. cerevisiae (yeast)
- S. pombe (yeast)
- C. elegans (worms)
- D. melanogaster (flies)
- M. musculus (mouse)
- Microcebus (mouse lemur- primate)
- C. Jacchus (marmoset- primate)
- Cell lines from the Ogden Syndrome boys
- Induced Pluripotent Stem (iPS) cells - human
S. cerevesiae (yeast)

Figure 1. Life cycles of heterothallic and homothallic strains of S. cerevisiae. Heterothallic strains can be stably maintained as diploids and haploids, whereas homothallic strains are stable only as diploids, because the transient haploid cells switch their mating type, and mate.
hNaa10p-S37P is functionally impaired *in vivo* using a yeast model. Unpublished data, do not further distribute.
Open question:
Function of N-terminal acetylation?

Protein stability? Protein secretion?

Figure courtesy of Kris Gevaert
Naa10p and Naa15p interact at the C-terminal region of Naa10p (residues 1-156), which make intramolecular interactions with residues in a substantially different conformation in the presence of Naa15p. The Naa10p-Naa15p interface appears to adopt a specific conformation that is essential for catalysis.

Overall structure of the NatA complex bound to acetyl CoA. (Fig. 1a) Naa10p (teal) and Naa15p (brown) subunits, shown in cartoon, bound to a NatA complex and had only modest effects on substrate binding and catalysis. (Fig. 2a) Naa10p (teal) and Naa15p (brown) subunits, shown in cartoon, bound to a NatA complex and had only modest effects on substrate binding and catalysis.

To determine the molecular basis for substrate-specific peptide binding by NatA, and in particular how, unlike most other NATs, it is able to disrupt NatA complex formation. Several additional scattered point mutations at either Naa15p Arg448 or Naa15p Phe449 were able to disrupt NatA complex formation. This is evident from the observation that alanine mutations in this region of Naa15p directly stabilize the position of the Naa10p V50 bond between Naa10p Gln25 and Naa15p Arg448. This region and Naa15p Phe449 and Trp494 and is supplemented with a hydrophobic interaction with Naa10p Ile36, which make intramolecular interactions with residues in a substantially different conformation in the presence of Naa15p.

Table for NAA10 and NAA15 peptide binding.

Overall structure of the NatA complex bound to acetyl CoA. (Fig. 1) Naa10p (teal) and Naa15p (brown) subunits, shown in cartoon, bound to a NatA complex and had only modest effects on substrate binding and catalysis. (Fig. 2) Naa10p (teal) and Naa15p (brown) subunits, shown in cartoon, bound to a NatA complex and had only modest effects on substrate binding and catalysis.

Table for NAA10 and NAA15 peptide binding.

Overall structure of the NatA complex bound to acetyl CoA. (Fig. 1) Naa10p (teal) and Naa15p (brown) subunits, shown in cartoon, bound to a NatA complex and had only modest effects on substrate binding and catalysis. (Fig. 2) Naa10p (teal) and Naa15p (brown) subunits, shown in cartoon, bound to a NatA complex and had only modest effects on substrate binding and catalysis.

Table for NAA10 and NAA15 peptide binding.
Cell Fractionation

**Strain 18**
MATa ade2-1 ura3-1 his3-11,15 leu2-3,112 trp1-1 can1-100 ard1::LEU2 nat1::KanMX p[BEVY hNAT1 hARD1 URA3]

- G-6-PDH
- hNaa10
- hNaa15
- RPL3

**Strain 94**
MATa ade2-1 ura3-1 his3-11,15 leu2-3,112 trp1-1 can1-100 ard1::LEU2 nat1::KanMX p[BEVY hNAT1 hARD1-S37P URA3]

- G-6-PDH
- hNaa10
- hNaa15
- RPL3
Cell Fractionation with HEK293 cells

Exp 1
- GAPDH
- Naa10
- Naa15
- Naa50
- RPL3
- RPS6

Exp 2
- GAPDH
- Naa10
- Naa15
- Naa50

Exp 3
- GAPDH
- Naa10
- Naa15
- Naa50
- RPL3
- RPS6
Cell Fractionation with HEK293 cells

7-47% Linear Sucrose Gradient
HEK293 Cells
All antibodies against endogenous proteins
Experiment 1 of 3 - 10-12-13
Lysis conditions as per Arnesen 2009, BMC Biochem 10:15.
Abstract

Amino-terminal acetylation is probably the most common protein modification in eukaryotes with as many as 50%-80% of proteins reportedly altered in this way. Here we report a systematic analysis of the predicted N-terminal processing of cytosolic proteins versus those destined to be sorted to the secretory pathway. While cytosolic proteins were profoundly biased in favour of processing, we found an equal and opposite bias against such modification for secretory proteins. Mutations in secretory signal sequences that led to their acetylation resulted in mis-sorting to the cytosol in a manner that was dependent upon the N-terminal processing machinery. Hence N-terminal acetylation represents an early determining step in the cellular sorting of nascent polypeptides that appears to be conserved across a wide range of species.
Proteomics Analysis of EBV-transformed cell lines from family members

II.A.

= male ♂

= female ♀

= FFPE DNA (for patient III.7.) or DNA from blood available (and for some of them: EBV transformed cell lines available + skin fibroblast of patient III.6.)

= stillborn

= proband

= patient samples analyzed by N-terminal COFRADIC analyses (#1 to #5)

= patient samples prepared for N-terminal COFRADIC analyses (but still to be analyzed) (#8 and #9)

II.1. mother of proband, carrier (89324) (#2)
II.2. married-in father of proband, WT(89325)
III.1. sister of proband, WT (90527) (#4)
III.2. brother of proband, WT (90526) (#3)
I.1. married-in grandfather of proband, WT(90529) (#5)
II.3. aunt of proband, WT (90530) (#6)
II.7. uncle of proband, WT(90688) (#9)
II.5. aunt of proband, carrier with deceased boy (90797) (#7)
Selecting protein N-terminal peptides by combined fractional diagonal chromatography

An Staes¹,², Francis Impens¹,², Petra Van Damme¹,², Bart Ruttens¹,², Marc Goethals¹,², Hans Demol¹,², Evy Timmerman¹,², Joël Vandekerckhove¹,² & Kris Gevaert¹,²

¹Department of Medical Protein Research, Vlaams Instituut voor Biotechnologie (VIB), Ghent, Belgium. ²Department of Biochemistry, Ghent University, Ghent, Belgium. Correspondence should be addressed to K.G. (kris.gevaert@vib-ugent.be).

Published online 14 July 2011; doi:10.1038/nprot.2011.355

In recent years, procedures for selecting the N-terminal peptides of proteins with analysis by mass spectrometry have been established to characterize protease-mediated cleavage and protein α-N-acetylation on a proteomic level. As a pioneering technology, N-terminal combined fractional diagonal chromatography (COFRADIC) has been used in numerous studies in which these protein modifications were investigated. Derivatization of primary amines—which can include stable isotope labeling—occurs before trypsin digestion so that cleavage occurs after arginine residues. Strong cation exchange (SCX) chromatography results in the removal of most of the internal peptides. Diagonal, reversed-phase peptide chromatography, in which the two runs are separated by reaction with 2,4,6-trinitrobenzenesulfonic acid, results in the removal of the C-terminal peptides and remaining internal peptides and the fractionation of the sample. We describe here the fully matured N-terminal COFRADIC protocol as it is currently routinely used, including the most substantial improvements (including treatment with glutamine cyclotransferase and pyroglutamyl aminopeptidase to remove pyroglutamate before SCX, and a sample pooling scheme to reduce the overall number of liquid chromatography—tandem mass spectrometry analyses) that were made since its original publication. Completion of the N-terminal COFRADIC procedure takes ~5 d.
Proteomics Strategy
With Thomas Arnesen, Petra van Damme and Kris Gevaert
Some downstream substrates for NatA

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Q5VT52</th>
<th>Q92616</th>
<th>Q14241</th>
<th>P78346</th>
<th>Q9UBE0</th>
<th>Q72456</th>
<th>Q94UG6</th>
<th>Q92989</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAAEEDGGPEGPNR</td>
<td>RNFS_HUMAN</td>
<td>GCN1L_HUMAN</td>
<td>ELOA1_HUMAN</td>
<td>RPP30_HUMAN</td>
<td>SAE1_HUMAN</td>
<td>KI21A_HUMAN</td>
<td>PDRG1_HUMAN</td>
<td>CLP1_HUMAN</td>
</tr>
<tr>
<td>AADTQVSETLKR</td>
<td>66</td>
<td>86</td>
<td>92</td>
<td>91</td>
<td>87</td>
<td>84</td>
<td>91</td>
<td>Membrane; Multi-pass membrane protein. Mitochondrion membrane. Endoplasmic reticulum membrane.</td>
</tr>
<tr>
<td>AAESALQVVEKLQAR</td>
<td>52</td>
<td>80</td>
<td>84</td>
<td>84</td>
<td>80</td>
<td>81</td>
<td>85</td>
<td>Translational activator GCN1</td>
</tr>
<tr>
<td>AVFADDLDR</td>
<td>58</td>
<td>89</td>
<td>92</td>
<td>92</td>
<td>87</td>
<td>87</td>
<td>92</td>
<td>Nucleus.</td>
</tr>
<tr>
<td>MVEKEEAGGGISEEEAAQYDR</td>
<td>66</td>
<td>95</td>
<td>96</td>
<td>96</td>
<td>96</td>
<td>95</td>
<td>100</td>
<td>Ribonuclease P protein subunit p30</td>
</tr>
<tr>
<td>MLGAPEESSVR</td>
<td>69</td>
<td>90</td>
<td>91</td>
<td>94</td>
<td>91</td>
<td>95</td>
<td>96</td>
<td>Nucleus; nucleolus.</td>
</tr>
<tr>
<td>MLSPEAER</td>
<td>51</td>
<td>79</td>
<td>74</td>
<td>79</td>
<td>70</td>
<td>72</td>
<td>80</td>
<td>Cytoplasm; cytoskeleton.</td>
</tr>
<tr>
<td>AAGGGGGSSKASSSSASSAGALESSLDR</td>
<td>74</td>
<td>97</td>
<td>97</td>
<td>97</td>
<td>96</td>
<td>97</td>
<td>97</td>
<td>p53 and DNA damage-regulated protein 1</td>
</tr>
<tr>
<td>GEEANDKKPTTKFELER</td>
<td>72</td>
<td>85</td>
<td>84</td>
<td>85</td>
<td>82</td>
<td>84</td>
<td>87</td>
<td>Regulation of nuclear pre-mRNA domain-containing protein 2</td>
</tr>
<tr>
<td></td>
<td>79</td>
<td>91</td>
<td>93</td>
<td>93</td>
<td>94</td>
<td>93</td>
<td>92</td>
<td>Polyrribonucleotide 5'-hydroxyl-kinase Clp1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Nucleus.</td>
</tr>
</tbody>
</table>
Design, Synthesis, and Kinetic Characterization of Protein N-Terminal Acetyltransferase Inhibitors

Håvard Foyn,†‡§ Justin E. Jones,§ Dan Lewallen,§ Rashmi Narawane,† Jan Erik Varhaug,‡,‖ Paul R. Thompson,§ and Thomas Arnesen*,†‡

†Department of Molecular Biology, University of Bergen, N-5020 Bergen, Norway
‡Department of Surgery, Haukeland University Hospital, N-5021 Bergen, Norway
§Department of Chemistry, The Scripps Research Institute, 120 Scripps Way, Jupiter, Florida 33458, United States
‖Department of Surgical Sciences, University of Bergen, N-5020 Bergen, Norway
Summary

• Found first human genetic disease involving Nt-acetylation of proteins

• Characterizing the Nt-acetylation pathway both \textit{in vitro} and \textit{in vivo}

• Working toward identifying interacting components and more downstream substrates of NatA complex.