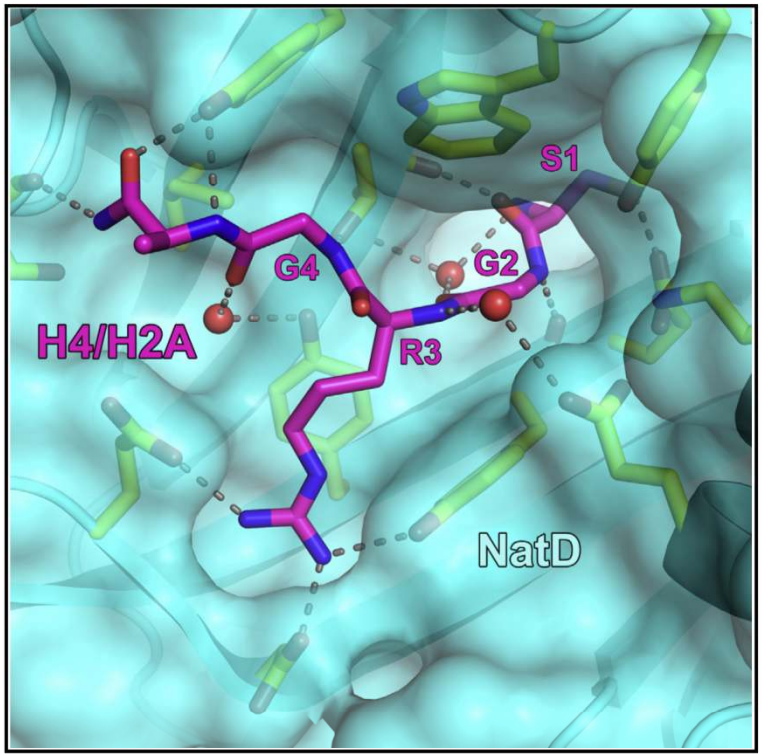
Structure

The Molecular Basis for Histone H4- and H2A-Specific Amino-Terminal Acetylation by NatD

Graphical Abstract



Authors

Robert S. Magin, Glen P. Liszczak, Ronen Marmorstein

Correspondence

marmor@mail.med.upenn.edu

In Brief

N-terminal acetyltransferase (NAT) proteins mediate N-terminal protein acetylation, and NatD is among the most selective NATs, acetylating histones H4 and H2A. Magin et al. report a crystal structure of NatD bound to an H4/H2A N-terminal substrate, highlighting the importance of a Ser-Gly-Arg-Gly recognition sequence for histone-specific acetylation.

Highlights

- NatD is among the most selective N-terminal acetyltransferases (NATs)
- NatD-specific structural features accommodate its high substrate selectivity
- NatD substrate site is tailored for a Ser-Gly-Arg-Gly histone H4/H2A N terminus
- Studies have implications for understanding substrate-specific acetylation by NATs

The Molecular Basis for Histone H4- and H2A-Specific Amino-Terminal Acetylation by NatD

Robert S. Magin,[1,2,3] Glen P. Liszczak,[1,5] and Ronen Marmorstein[1,2,4,*]
[1]Department of Biochemistry and Biophysics, Abramson Family Cancer Research Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA
[2]Program in Gene Expression and Regulation, The Wistar Institute, Philadelphia, PA 19104, USA
[3]Graduate Group in Biochemistry and Biophysics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA
[4]Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA
[5]Present address: Department of Chemistry, Princeton University, Princeton, NJ 08544, USA
*Correspondence: marmor@mail.med.upenn.edu
http://dx.doi.org/10.1016/j.str.2014.10.025

SUMMARY

N-terminal acetylation is among the most common protein modifications in eukaryotes and is mediated by evolutionarily conserved N-terminal acetyltransferases (NATs). NatD is among the most selective NATs; its only known substrates are histones H4 and H2A, containing the N-terminal sequence SGRGK in humans. Here we characterize the molecular basis for substrate-specific acetylation by NatD by reporting its crystal structure bound to cognate substrates and performing related biochemical studies. A novel N-terminal segment wraps around the catalytic core domain to make stabilizing interactions, and the α1-α2 and β6-β7 loops adopt novel conformations to properly orient the histone N termini in the binding site. Ser1 and Arg3 of the histone make extensive contacts to highly conserved NatD residues in the substrate binding pocket, and flanking glycine residues also appear to contribute to substrate-specific binding by NatD, together defining a Ser-Gly-Arg-Gly recognition sequence. These studies have implications for understanding substrate-specific acetylation by NAT enzymes.

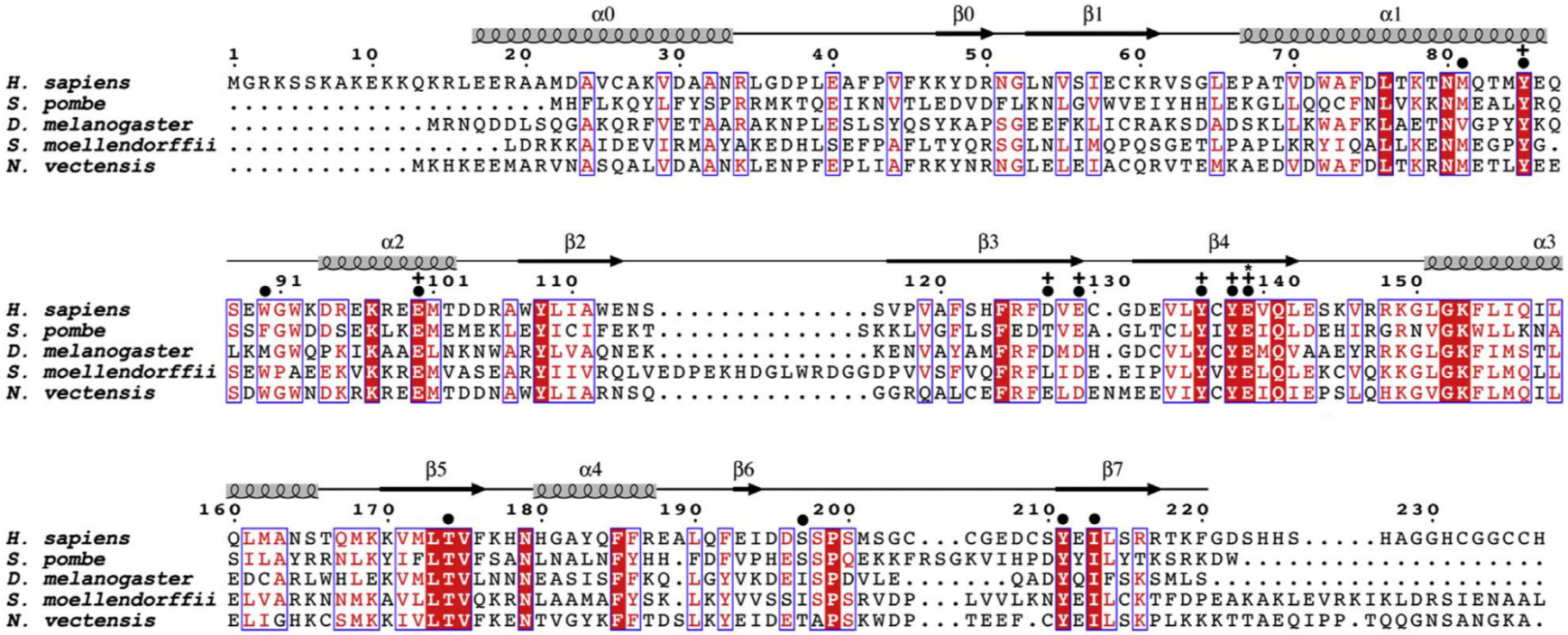


Figure 1. Sequence Alignment of NatD Orthologs

The sequence alignment contains the following NatD orthologs: human (*Homo sapiens*), fission yeast (*Schizosaccharomyces pombe*), fruit fly (*Drosophila melanogaster*), spikemoss (*Selaginella moellendorffii*), and sea anemone (*Nematostella vectensis*). The blue boxes represent conserved patches of sequence alignment. Residues in red are highly conserved, and residues in white with a red background are strictly conserved. Above the sequence alignment are indicated amino acid numbering for *H. sapiens* and secondary structure elements. Amino acid residues are indicated that make contacts to the substrate peptide (●), show mutational sensitivity (+), or are proposed to play catalytic roles (*).

Table 1. Data Statistics for NatD Crystal Structures

	Selenomethionine SpNatD	hNatD/acetyl-CoA	hNatD/CoA/H4-H2A Peptide
PDB ID	4AU3	4U9V	4U9W
Crystal Parameters			
Space group	C2	C2	$P2_1$
Cell dimensions			
a, *b*, *c*, (Å)	114.701, 42.629, 85.417	88.256, 44.064, 50.356	52.901, 93.288, 100.329
α, β, γ (°)	90.00, 98.68, 90.00	90.00, 95.44, 90.00	90.00, 96.69, 90.00
Data Collection[a]			
Resolution (Å)	50–1.85 (1.92–1.85)	50–1.78 (1.84–1.78)	50–2.50 (2.59–2.50)
Unique reflections	34807	18519	33496
R_{merge}	0.067 (0.595)	0.074 (0.471)	0.102 (0.552)
$I/\sigma I$	32.6 (6.1)	22.6 (1.7)	12.3 (2.3)
Completeness (%)	99.4 (98.7)	98.5 (87.5)	98.8 (97.6)
Redundancy	17.4 (16.7)	8.1 (4.2)	3.8 (3.7)
Refinement			
R_{work}/R_{free} (%)	17.9/23.8	16.9/20.6	17.5/22.3
No. of atoms			
Protein	3132	1582	6348
Acetyl-CoA/CoA	96	51	192
Peptide	N/A	N/A	125
Solvent	234	166	457
B factors ($Å^2$)			
Protein	64.63	23.99	26.61
Acetyl-CoA/CoA	62.97	21.59	24.92
Peptide	N/A	N/A	25.85
Solvent	77.18	32.57	27.23
Root-mean-square deviation			
Bond lengths (Å)	0.007	0.009	0.010
Bond angles (°)	1.078	1.126	1.206
Ramachandran statistics (%)			
Favored	100	97.44	98.72
Allowed	0	2.56	1.15
Outlier	0	0	0.13

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

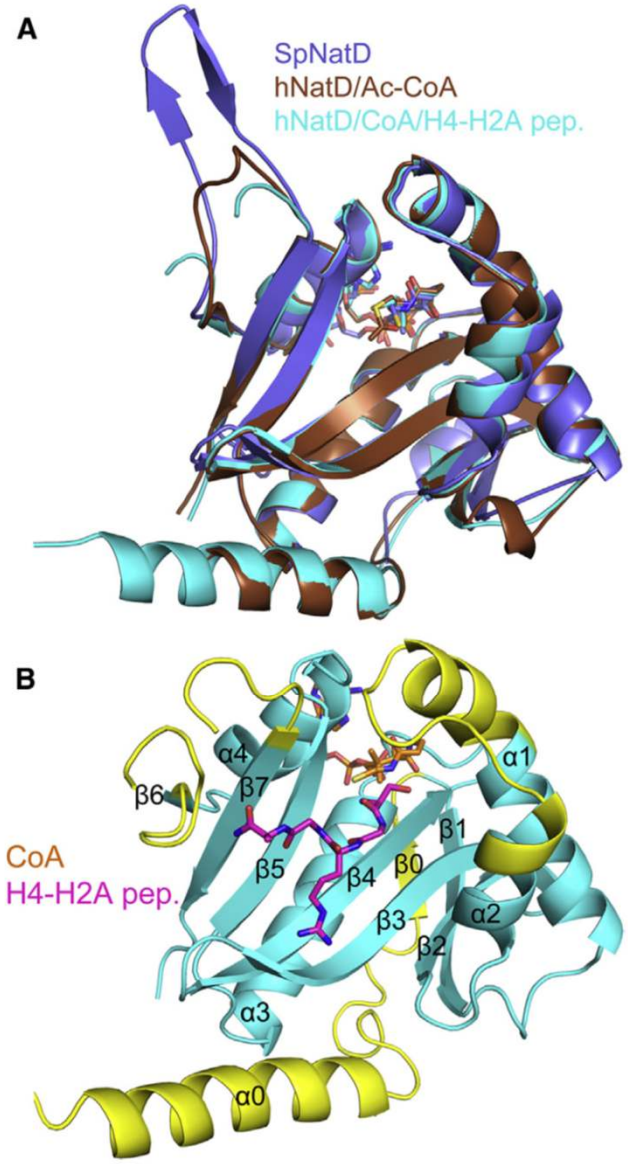


Figure 2. Overall Structure of NatD Complexes

(A) Superposition of the SpNatD/acetyl-CoA (violet), hNatD/acetyl-CoA (brown), and hNatD/CoA/H4-H2A peptide (cyan) complexes. CoA is shown as sticks, and the H4-H2A peptide is omitted for clarity.

(B) Overall structure of hNatD/CoA/H4-H2A peptide with structurally unique elements of NatD relative to other NATs highlighted in yellow. CoA is shown in orange and H4 is shown in magenta.

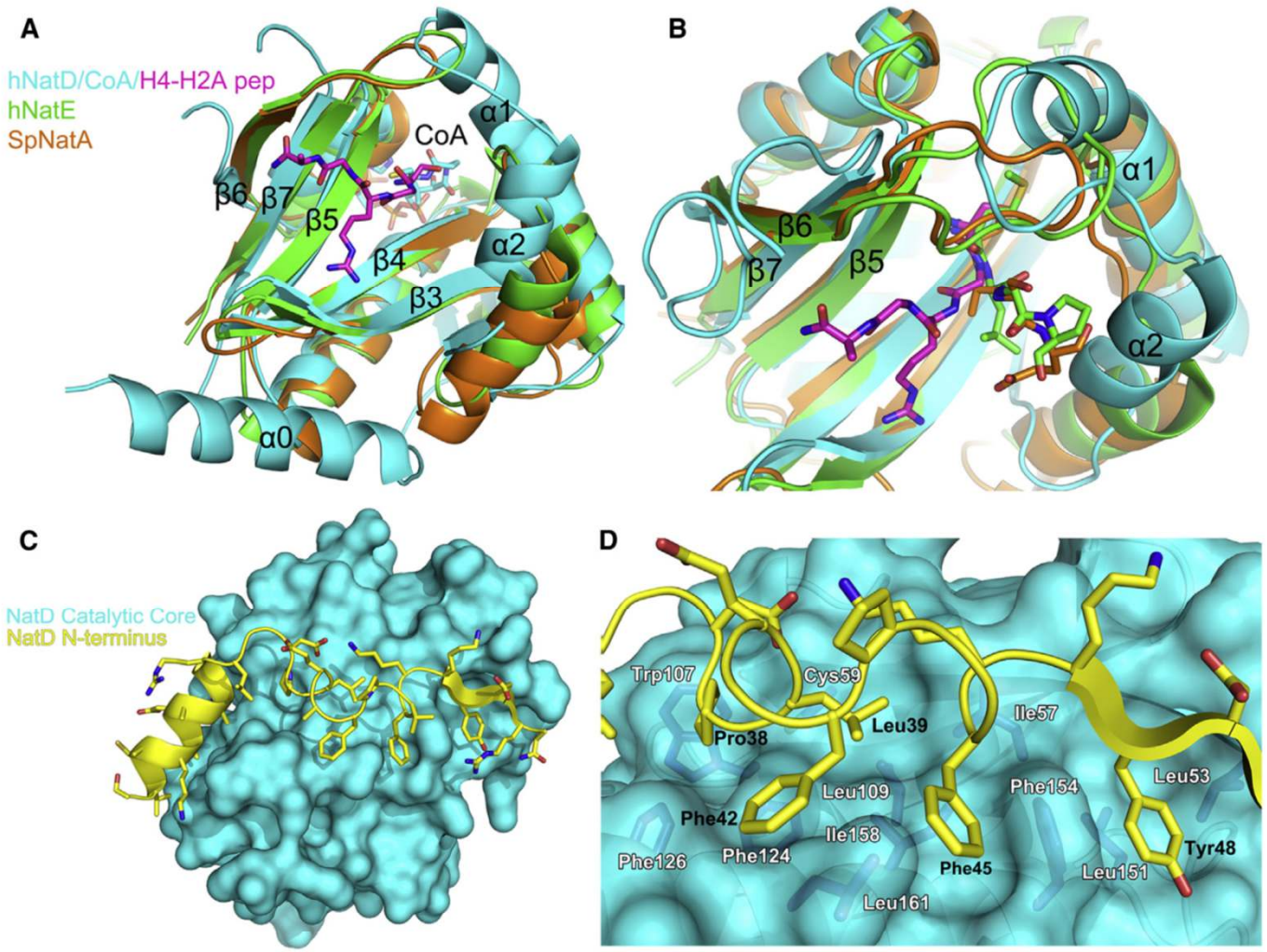


Figure 3. Unique Structural Features of NatD

(A) Superposition of hNatD (cyan), SpNatA (orange), and hNatE (green) complexes. The H4-H2A peptide is shown in magenta. Only the CoA from the hNatD/CoA/H4-H2A peptide structure is shown for clarity.

(B) Close-up view of the substrate binding groove of NatD in comparison with NatA and NatE. The color coding is as in Figure 3A. NatA substrate (SASE) and NatE substrate (MLGP) are shown as orange and green sticks, respectively.

(C) View of the interaction of the NatD N-terminal segment (yellow cartoon representation) with the catalytic core domain (cyan surface representation).

(D) Detailed interactions between the N-terminal segment and catalytic core domain of NatD. Residues from the N terminus that mediate interactions are labeled in black, and residues from the core domain are colored in dark blue and labeled in white. Met-162 is omitted for clarity.

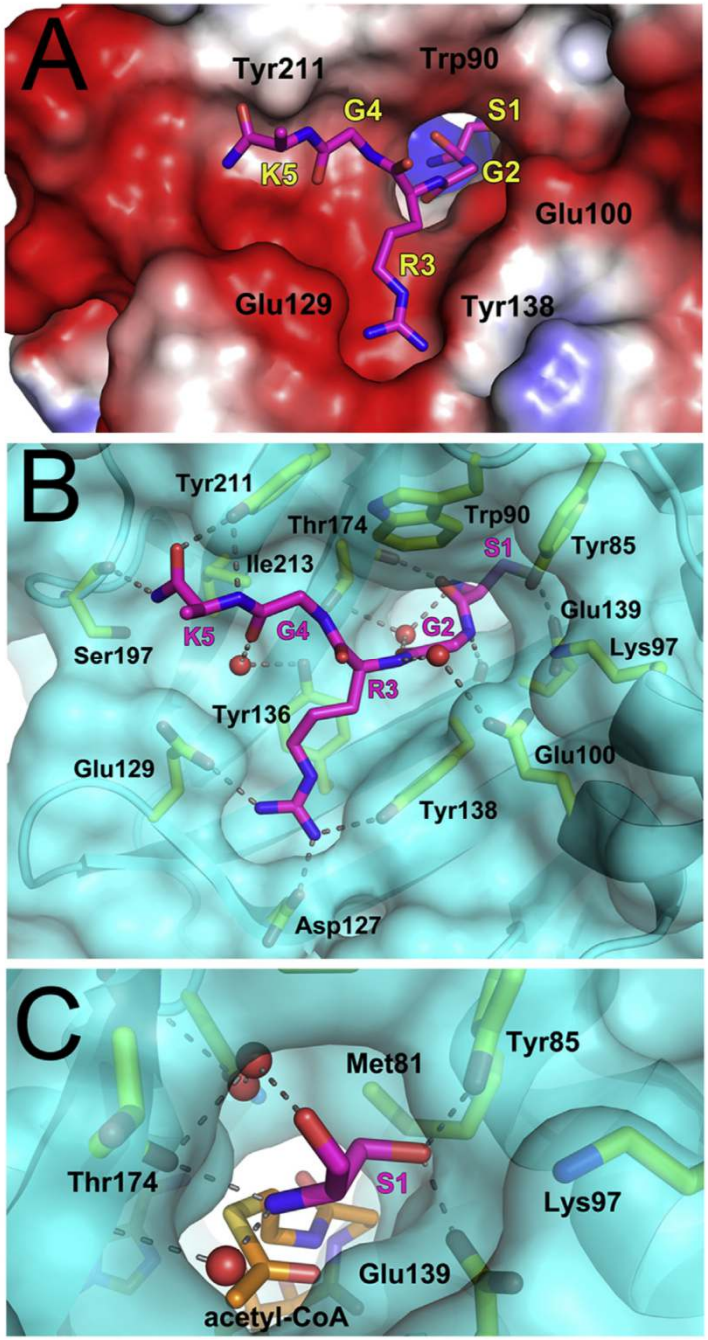


Figure 4. Peptide Binding Site of NatD

(A) Electrostatic surface of the NatD peptide binding site with the peptide shown in magenta stick figure. Residues from NatD are labeled in black and residues from the peptide are labeled in yellow with their corresponding one-letter codes and numerical positions in the peptide. The side chain of $Lys5_p$ was disordered and not modeled into the crystal structure.

(B) Detailed interactions between NatD and the H4-H2A peptide. NatD is shown as a transparent cyan surface, and residues that interact with the peptide are yellow. Hydrogen bonds are shown as dashed lines and waters are shown as red spheres.

(C) Close-up view of the NatD active site highlighting interactions made by $Ser1_p$. Acetyl-CoA was modeled into the figure by aligning the binary and ternary NatD structures.

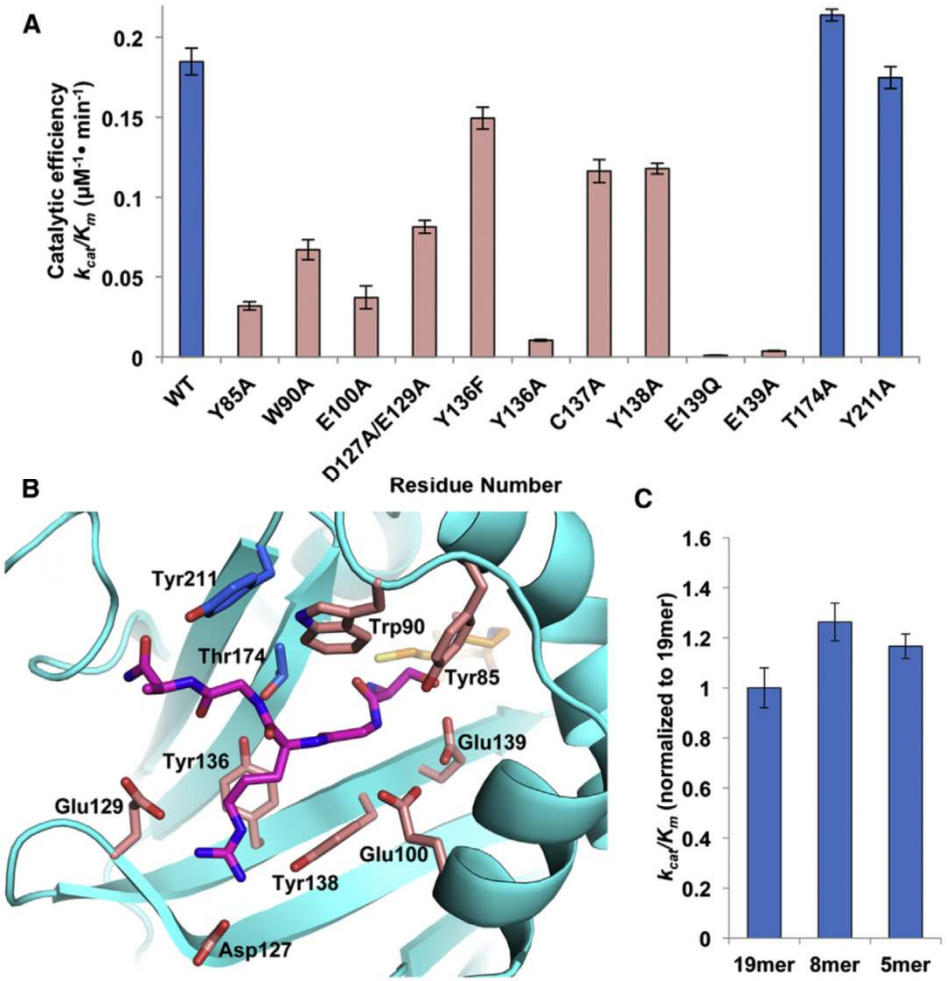


Figure 5. Mutational Analysis of NatD

(A) Catalytic efficiency of selected NatD mutants. Mutations that have negligible effect are in blue, while those that decrease the catalytic efficiency of the enzyme are in pink.

(B) Residues targeted for mutagenesis are mapped onto the NatD structure. The color scheme is the same as Figure 5A.

(C) The catalytic efficiency of wild-type NatD toward N-terminal histone H4 peptides of varying length. Data are represented as mean ± SEM. Each mutant and peptide of varying length was assayed in triplicate.

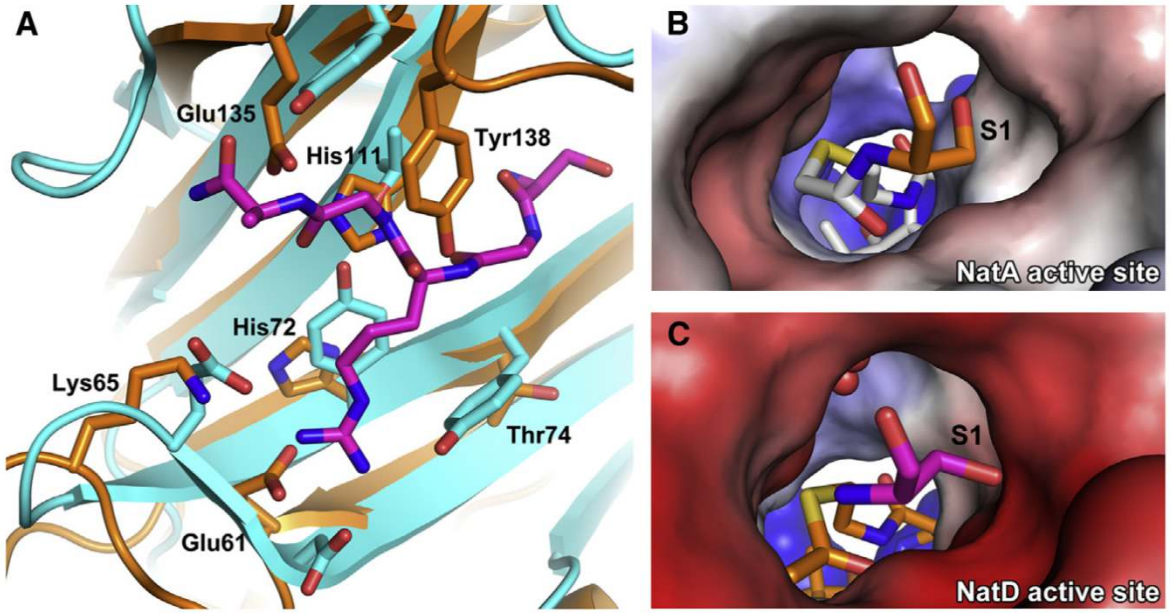


Figure 6. Comparison between Substrate Recognition of NatD and NatA

(A) Overlay of the peptide binding site of NatA (orange) and NatD (cyan with magenta histone substrate).

(B) Electrostatic potential surface of the NatA active site with a bound covalently linked bisubstrate inhibitor. The N-terminal serine of the substrate peptide is shown in orange, and the acetyl-CoA moiety of the bisubstrate inhibitor is in white.

(C) Electrostatic potential surface of the NatD active site. The N-terminal serine of the substrate peptide is shown in magenta, and acetyl-CoA is in orange. Waters are shown as red spheres. Acetyl-CoA was modeled into the figure by aligning the binary and ternary NatD structures.

EXPERIMENTAL PROCEDURES

Homo sapiens NatD (hNatD) Expression and Purification

The full-length hNatD gene (encoding residues 1–237) was a gift from Dr Thomas Arnesen from the University of Bergen. Several N-terminal and C-terminal truncation constructs were engineered into a modified PCDF vector containing an N-terminal tobacco etch virus (TEV) protease-cleavable His tag. All constructs were transformed into Rosetta (DE3)pLysS competent *Escherichia coli* cells, which were grown to an OD_{600} of 0.7–0.9 and induced with 0.5 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) at 16°C for ~16 hr. All subsequent purification steps were carried out at 4°C. Cells were isolated by centrifugation and lysed by sonication in lysis buffer containing 25 mM Tris, pH 8.0, 1 M NaCl, 10 mM β-mercaptoethanol (βME) and 10 μg/ml phenylmethanesulfonylfluoride. The lysate was clarified by centrifugation and passed over nickel resin (Thermo Scientific), which was subsequently washed with >20 column volumes of lysis buffer supplemented with 25 mM imidazole. The protein was eluted in lysis buffer supplemented with 300 mM imidazole. His tagged TEV protease was added to the eluent containing the target protein for the duration of a 14 hr dialysis into dialysis buffer containing 25 mM Tris, pH 8.5, 200 mM NaCl, 10 mM βME. This solution was passed through an additional nickel column to remove TEV protease as well as any uncut NatD. The resin was then washed with approximately seven column volumes of dialysis buffer supplemented with 25 mM imidazole, which was pooled with the initial flow-through. This solution was dialyzed into ion exchange buffer containing 25 mM Tris, pH 8.5, 50 mM NaCl and 1 mM dithiothreitol (DTT) and loaded onto a 5 ml HiTrap Q ion exchange column (GE Healthcare). The protein was eluted in the same buffer with a salt gradient (50–750 mM NaCl) over the course of 20 column volumes. Peak fractions were pooled and dialyzed into sizing buffer containing 25 mM HEPES, pH 7.0, 200 mM NaCl, and 1 mM DTT for ~16 hours. This was concentrated to a volume of 500 μl (10 kDa concentrator; Amicon Ultra, Millipore), and loaded onto and run on an s75prep gel filtration column (GE Healthcare) in sizing buffer. Peak fractions were concentrated to 10 mg/ml as measured by UV_{280} for crystallization trials.

S. pombe NatD (SpNatD) Expression and Purification

The full-length SpNatD gene (encoding residues 1–204) was cloned from the *S. pombe* genome (ATCC), and several N-terminal truncation constructs were engineered into a modified pETDUET vector containing a TEV protease-cleavable His tag. All constructs were transformed into Rosetta (DE3) pLysS competent *E. coli* cells, which were grown to an OD_{600} of 0.7–0.9 and induced with 0.5 mM IPTG at 16°C for ~16 hr. These constructs all expressed as insoluble inclusion bodies in Rosetta (DE3)pLysS *E. coli* cells. To improve solubility, a number of constructs of the SpNatD gene were subcloned into a modified pETDUET vector containing an N-terminal SUMO tag. These constructs were transformed and induced with IPTG in the same way and were soluble in *E. coli* cells. All subsequent studies were carried out at 4°C and are identical to the purification protocol used for purifying the human protein, with the following changes. After the nickel column, His tagged Ulp1 protease was used in place of TEV protease to cleave the SUMO tag. The ion exchange buffer contained 25 mM HEPES, pH 7.0, 50 mM NaCl, and 1 mM DTT, and a 5 ml HiTrap SP ion-exchange column (GE Healthcare) was used. The selenomethionine derivative was produced by expressing the SUMO tagged SpNatD (residues 13–204) in minimal media (Molecular Dimensions) with 50 μg/l selenomethionine (Sigma) purified in the same manner as above.

NatD Crystallization and Data Collection

hNatD (8 mg/ml) was incubated with acetyl-CoA (Sigma-Aldrich) at a 1:3 molar ratio. Crystals were obtained with hanging-drop vapor diffusion in a drop containing a 1:1 mixture of protein to a well solution containing 20% PEG 3350, 200 mM sodium citrate tribasic dihydrate, 100 mM citric acid, pH 4.0. Diffraction-quality crystals grew in 2–3 days. The ternary complex was crystallized using an hNatD construct containing residues 17–220 at 10 mg/ml. The protein was incubated with CoA, and the C-terminally amidated substrate peptide (SGRGK, GenScript) at a 1:3:5 molar ratio of protein:CoA:peptide. Crystals were obtained through hanging-drop vapor diffusion at 20°C at a 1:1 ratio of protein to a well solution containing 20% PEG 3350 and 200 mM sodium malonate, pH 6.0. Crystals grew after 1 day. SpNatD was crystallized at a concentration of 10 mg/ml using a construct that contained residues 13-204. The protein was mixed with acetyl-CoA at a molar ratio of 1:3 protein to acetyl-CoA. Crystals were obtained with hanging-drop vapor diffusion at 20°C at a 1:1 ratio of protein to well solution containing 20% PEG 3350 and 200 mM potassium formate. Crystals grew after 1 day. Selenomethionine crystals of SpNatD were generated in the same manner. All crystals were cryoprotected by transferring them to their respective well solutions supplemented with 20% glycerol before being flash frozen in liquid nitrogen. Data for the hNatD ternary complex were collected at a home source using a MicroMax-007 HF rotating anode (Rigaku). Data for all other crystals were collected at beamline X29A at the National Synchrotron Light Source (Brookhaven National Laboratory). All data sets were processed using HKL2000 (Otwinowski and Minor, 1997).

Structure Determination and Refinement

Initially, a number of full-length and truncated search models based on known structures of NATs (Liszczak et al., 2011; Liszczak et al., 2013; Liszczak and Marmorstein, 2013) were used in an attempt to solve either hNatD or SpNatD structure through molecular replacement (McCoy et al., 2007). However, these attempts were unsuccessful in generating a suitable starting model for either hNatD or SpNatD. A data set was collected on the selenomethionine-labeled SpNatD crystals at the selenium peak wavelength (0.9788 Å), heavy atoms were placed in the asymmetric unit, and initial phases were generated using single-wavelength anomalous diffraction phasing using SOLVE in the Phenix suite (Adams et al., 2011). The SpNatD structure was initially built into the resulting density by AutoBuild in the Phenix suite. The rest of the structures described were determined using molecular replacement. The selenomethionine-labeled SpNatD structure was used as a search model for the wild-type SpNatD crystals. However, these crystals did not yield a model of sufficient quality, so only the selenomethionine data set was refined fully. The program CHAINSAW (Stein, 2008) was used to generate a search model from the SpNatD structure for the binary hNatD structure, and the final refined hNatD structure was used as a search model for the ternary hNatD crystals. For all of the structures, the ligands acetyl-CoA/CoA and the substrate peptide for the ternary complex were built into the $F_o - F_c$ electron density map. All of the structures were refined iteratively by using Phenix.Refine and manual model building in the molecular graphics program Coot (Emsley et al., 2010). The SpNatD/acetyl-CoA complex crystallized as a dimer with the N terminus of one monomer in the active site of another. The N-terminal serine of the enzyme is acetylated, and CoA is bound to the enzyme. This is likely an artifact of crystallization, wherein the subunit packing positions the N terminus of SpNatD (SRRMK), which superficially resembles the natural substrate, in the active site. The final model and structure factors for all of the structures were submitted to the Protein Data Bank (PDB). Refinement statistics and PDB IDs can be found in Table 1.

Acetyltransferase Assays

Acetyltransferase assays were carried out in 25 mM HEPES pH 7.0, 100 mM NaCl, and 1 mM DTT. The substrate peptide used in the assay for the mutant enzymes corresponds to the first 19 residues of human H4 (NH2-SGRGKGGKGLGKGGAKRHR-COOH; GenScript). In the assay, a saturating amount (400 μM) of radiolabeled [^{14}C]acetyl-CoA (4 mCi/mmol; PerkinElmer Life Sciences) and varying concentrations of the substrate peptide (6.25–200 μM) were incubated with 75 nM hNatD (500 nM for E139A and E139Q mutants) in 75 μl of reaction volume for 20–180 min at 37°C. To quench the reaction, 20 μl of the reaction mixture was added to negatively charged P81 paper disks (Whatman), and the paper disks were immediately placed in wash buffer (10 mM HEPES, pH 7.5). The paper disks were washed three times, at 5 min per wash, to remove unreacted acetyl-CoA. The papers were then dried with acetone and added to 4 ml of scintillation fluid, and the signal was measured with a Packard Tri-Carb 1500 liquid scintillation analyzer. Background control reactions were performed in the absence of enzyme or in the absence of substrate peptide to ensure that any possible signal due to autoacetylation was negligible. Each reaction was performed in triplicate. Complications from product inhibition at high peptide concentrations kept us from generating full catalytic curves. However, by keeping the substrate concentration far below K_m, we were able to plot velocity versus [substrate] and use the slope of the resulting line to obtain k_{cat}/K_m values for the mutants based on the equation: $v = (k_{cat}/K_m)[E][S]$. The counts per minute were converted to molar units using a standard curve of known [^{14}C]acetyl-CoA concentrations in scintillation fluid. The assay was performed in the same manner for substrate peptides corresponding to the first five and eight residues of human H4 (SGRGK and SGRGKGGK, respectively) with wild-type hNatD.

AUTHOR CONTRIBUTIONS

R.S.M. and G.P.L. performed the experiments described in the manuscript. R.S.M. prepared manuscript figures and text. R.M. designed and supervised experiments by R.S.M. and G.P.L. and prepared manuscript text. All authors read and approved the submitted manuscript.

ACKNOWLEDGMENTS

This work was supported by NIH grant R01 GM060293 awarded to R.M. and NIH grant T32 GM071339 awarded to G.P.L. and R.S.M. We acknowledge support of the Proteomics core facility at the Wistar Institute (NIH P30 CA010815) and the University of Pennsylvania DNA Sequencing Facility at the Perelman School of Medicine, University of Pennsylvania (NIH P30 CA016520). We thank Petra Van Damme for critical reading of the manuscript. The coordinates of the structures have been deposited in the PDB under accession numbers 4UA3 (selenomethionine-labeled SpNatD/Ac-CoA), 4U9V (hNatD/Ac-CoA) and 4U9W (hNatD/Ac-CoA/Peptide).

Received: September 30, 2014
Revised: October 22, 2014
Accepted: October 24, 2014
Published: January 22, 2015