Novel bisubstrate inhibitors for protein N-terminal acetyltransferase D

Youchao Deng[†], Sunbin Deng[§], Yi-Hsun Ho[†], Sarah M. Gardner[#], Zhi Huang[†], Ronen Marmorstein[§,#], and Rong Huang[*,†]

[†]Department of Medicinal Chemistry and Molecular Pharmacology, Purdue Institute for Drug Discovery, Purdue University Center for Cancer Research, Purdue University, West Lafayette, Indiana 47907, United States

[§]Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States

[#]Department of Biochemistry and Biophysics, Abramson Family Cancer Research Institute, Graduate Group in Biochemistry and Molecular Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, United States

[*]Corresponding Author: huang-r@purdue.edu

ABSTRACT: Protein N-terminal acetyltransferase D (NatD, NAA40, Nat4) that specifically acetylates the N-terminus of histone H4 and H2A has been implicated in various diseases, but no inhibitor has been reported for this important enzyme. Based on the acetyl transfer mechanism of NatD, we designed and prepared a series of highly potent NatD bisubstrate inhibitors by covalently linking coenzyme A to different peptide substrates via an acetyl or propionyl spacer. The most potent bisubstrate inhibitor displayed a K_i of 170 ± 16 pM. We also demonstrated that these inhibitors are highly specific towards NatD, displaying 10,000-fold selectivity over other closely-related acetyltransferases. High resolution crystal structures of NatD bound to two of these inhibitors revealed the molecular basis for their selectivity and inhibition mechanisms, providing a rational path for future inhibitor development.

INTRODUCTION

α-N-terminal acetylation (Nα-acetylation) is a ubiquitous protein modification that occurs on 80 – 90% of human proteins.[1] It is essential for various biological functions including protein-protein interactions, protein complex formation, cellular apoptosis, rDNA transcriptional regulation, protein subcellular localization, and degradation.[2-4] This modification is catalyzed by protein N-terminal acetyltransferases (NATs) that transfer an acetyl group from the donor acetyl coenzyme A (AcCoA) onto the Nα-amino group of protein substrates. To date, eight members of eukaryotic NATs (NatA-NatH) have been reported.[5] Among them, NatA NatD acetylate the nascent chain after the initiator methionine is cleaved.[6] NatA is a multisubunit enzyme, acetylating ~40% of human proteins that contain small and uncharged first residues.[1,6] In contrast, NatD is a monomeric protein showing extremely high substrate specificity for histone proteins H2A and H4 that have the same N-terminal sequence SGRGK.[7,8]

NatD mediated Nα-acetylation on H4 has diverse biological functions and implications in tumorigenesis.[2,9,7] Depletion of NatD induces apoptosis through the mitochondrial pathway in colorectal cancer cells.[10] In addition, NatD is downregulated in hepatocellular carcinoma tissues and upregulated in primary human lung cancer tissues.[11] As the function of histone H4 Nα-acetylation has recently come to light, it has been shown to regulate crosstalk with arginine methylation, lysine acetylation, and serine phosphorylation on H4.[12,13] For example, Nα-acetylation of H4 stimulates ribosomal DNA expression by inhibiting asymmetric dimethylation of Arg3 on H4 (H4R3me2a), consistent with its critical role in cell growth.[12,13] Nα-acetylation of H4 suppresses Ser1 phosphorylation and induces the expression of Slug transcription to promote the epithelial-to-mesenchymal transition in lung cancer, suggesting that the acetyltransferase activity of NatD is critical for Slug regulation.[13] In addition, Nα-acetylation of H4 promotes the expression of oncogenes through upregulation of protein arginine methyltransferase 5 in colorectal cancer cells.[10] Bases on these disease connections, NatD has surfaced as a new therapeutic target. Hence, potent and selective NatD inhibitors would be valuable probes to interrogate its functions and therapeutical potential. However, there is no NatD inhibitor available to date.

The co-crystal structure of the ternary NatD\CoA\SGRGK complex (PDB: 4U9W) revealed that its

substrate peptide (SGRGK) is inserted into a highly acidic binding pocket of NatD, and biochemical studies supported the importance of the first 4 residues (SGRG) of the cognate substrate for specific recognition.[8] Although the kinetic mechanism of NatD has not yet been elucidated, both NatA and NatE follow an ordered Bi-Bi mechanism.[14–16] Structural alignment of NatD with NatA infers a similar mechanism as NatA.[8] Furthermore, upon mutation of an active-site C137, a previously suggested catalytic residue that forms an acetyl-NatD intermediate, hNatD remained largely active, supporting a direct transfer of the acetyl group through a Bi-Bi mechanism.[8] Based on these observations, we hypothesized that bisubstrate analogues could provide potent inhibitors for NatD. Bisubstrate analogues have previously been used to help elucidate the catalytic mechanism of other NATs and to develop valuable tool compounds, usually displaying IC_{50} values within a low micromolar to high nanomolar range against *Sp*NatA,[14] hNatB,[17] hNatF,[18] NatH[19]. In the study presented here, highly potent and selective NatD bisubstrate inhibitors (K_i values of 170 pM to 1.6 nM) have been developed. Moreover, the X-ray crystal structures of NatD bound to two of these inhibitors have been determined to confirm the key design features, thus paving the way for the structure-based development of more drug-like NatD inhibitors.

RESULTS AND DISCUSSION

Inhibitor Design. Bisubstrate analogues that covalently connect cofactor and short peptide substrate with a linker have been reported for targeting a variety of transferases, such as methyltransferases and acetyltransferases.[20,21] Based on the co-crystal structure of the NatD/CoA/SGRGK ternary complex (Figure 1), the linear distance between the Nα-amino nitrogen atom of Ser and the sulfur atom of CoA was measured to be about 3.1 Å.[8] Therefore, we proposed to connect the Nα-amines of different peptide substrates and the thiol group of CoA with either an acetyl or propionyl linker to prepare bisubstrate analogues for NatD.

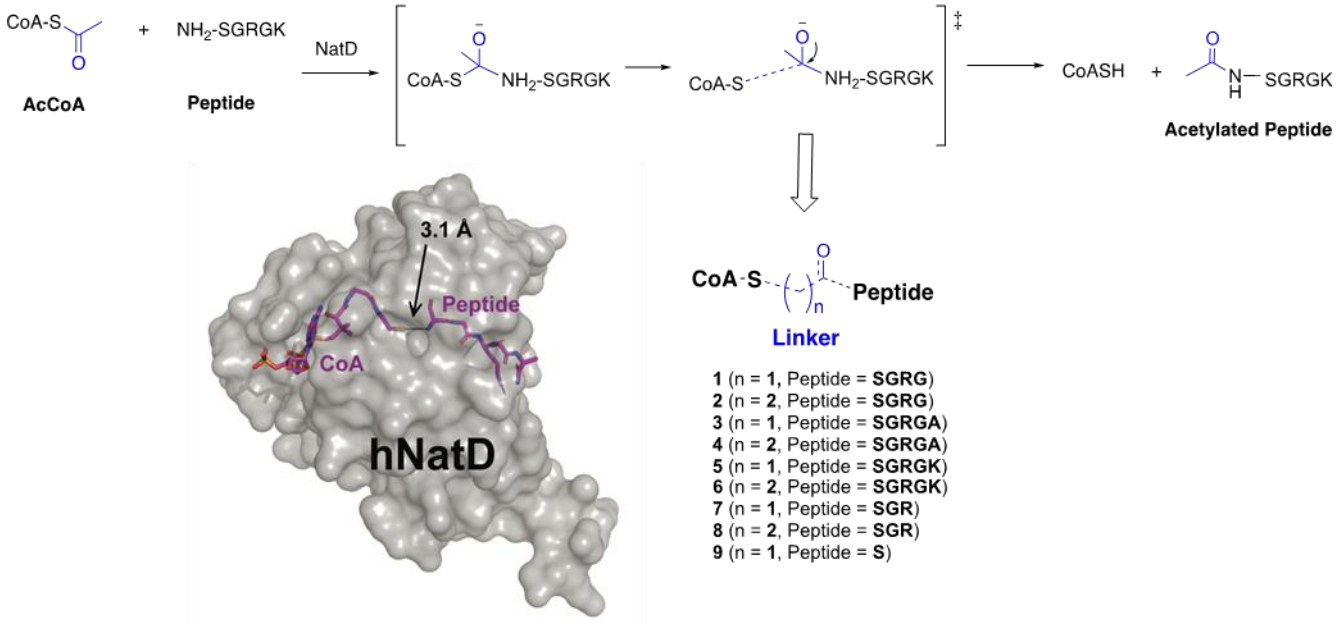
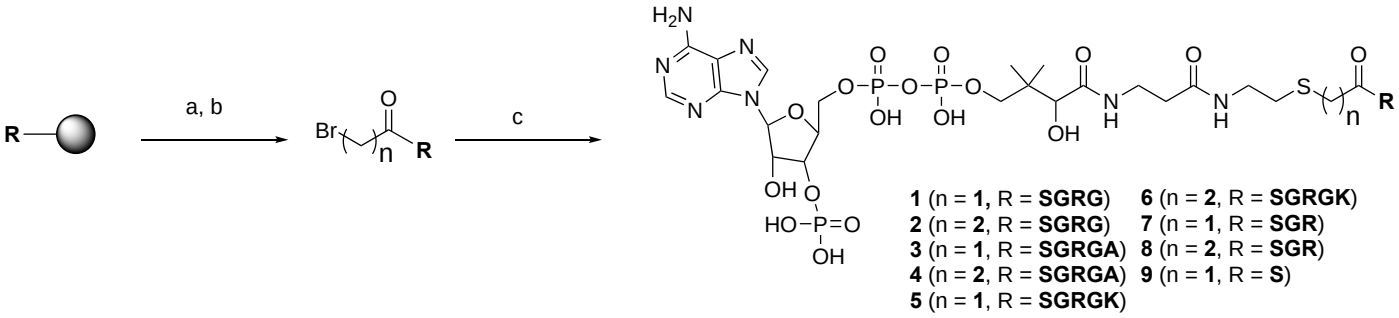


Figure 1. Inhibitor design strategy. The reported crystal structure of the NatD/CoA/SGRGK ternary complex (PDB: 4U9W) is shown in grey transparent surface, with the CoA and peptide fragment shown as magenta sticks. The distance between the sulfur atom of CoA and peptide N terminal nitrogen atom is 3.1 Å. The designed NatD bisubstrate analogues are illustrated.

Inhibitor Synthesis. Peptides were prepared on rink resin following a standard Fmoc solid-phase peptide synthesis protocol. 2-bromoacetic acid or 3-bromopropionic acid was coupled with the free N-terminal amine of the peptide on the resin. Subsequent cleavage with the cocktail consisting of trifluoroacetic acid (TFA)/ water (H_2O)/triisopropylsilane (95:2.5:2.5) and purification through high-performance liquid chromatography (HPLC) provided the purified bromopeptides, which were then reacted with coenzyme A trilithium salt dihydrate in the triethylammonium bicarbonate buffer. The resulting mixture was purified by HPLC to afford the desired bisubstrate analogues **1**-**9**.

Scheme 1. Synthetic route for the bisubstrate analogues.



Scheme 1. (a) 2-Bromoacetic acid or 3-Bromopropionic acid, DIC, HOBt, DMF, r.t., overnight; (b) TFA / TIPS / water (95 / 2.5 / 2.5), r.t., 4h, 34-56% in two steps; (c) CoASH, triethylammonium bicarbonate buffer, pH 8.4±0.1, r.t., 48 h, 33-55%.

Structure-Activity Relationship Studies. All synthesized bisubstrate analogues were evaluated with an established fluorescence assay.[22] Initial testing was performed in the presence of both AcCoA and peptide substrate SGRGK at their respective Michaelis constant (K_m) values, but resulting IC_{50} values were close to the enzyme concentration. Thus, we optimized the condition to characterize IC_{50} values in the presence of both concentrations of AcCoA and SGRGK peptide at 4xK_m values. All bisubstrate analogues **1**-**9** displayed IC_{50} values ranging from 79 nM to 218 nM for NatD, except when only Ser was incorporated in the substrate moiety, which resulted in no detectable inhibition (Table 1). When three to five amino acids were incorporated in the substrate moiety, there was less than a three-fold difference among all bisubstrate analogues between acetyl and propionyl linker. This suggested that both linkers can be accommodated by NatD. In addition, it indicated that the first three amino acids contribute significantly to the interaction with NatD. A bisubstrate analogue (**9)** that only contained a Ser showed an IC_{50} of over 500 uM, supporting the conclusion that the first three amino acids are important for inhibitor potency. Removal of the CoA moiety also abolished the inhibitory activity, as propionyl-SGRGK **10** exhibited an IC_{50} of over 500 uM, also supporting the importance of the CoA moiety for inhibitor potency.

To validate the inhibitory activities of all bisubstrate analogues, we applied an orthogonal radioactive assay to directly monitor the production of acetylated peptide under a similar condition with both AcCoA

and peptide substrate at their 4x*Km* values. All bisubstrate inhibitors showed IC_{50} values less than 50 nM in the radioactive assay. In addition, the difference among this series of compounds was more salient, likely due to the higher sensitivity of the radioactive assay than the fluorescence assay. However, the IC_{50} values collected from these two different methods shared the same trends for the synthesized inhibitors. Compound **6** that links CoA and SGRGK peptide through a propionyl linker showed the most potent inhibition activity (K_i of 170 pM) against NatD, which is 2-fold and 10-fold higher than **2** and **8**, respectively. For bisubstrate analogues containing either a tetrapeptide SGRG or pentapeptide SGRGK, a propionyl linker was more favorable than an acetyl linker as inhibitory activity was about 5-fold better. However, acetyl and propionyl linkers showed more comparable inhibitor activity for the bisubstrate ananlogs that only contained a tripeptide SGR, as **7** was only about 3-fold more potent than **8**. For the bisubstrate analogs containing a pentapeptide SGRGA, both linkers were acceptable and showed even more similar potency.

Table 1. Inhibition activities of the bisubstrate inhibitors

ID	Structure	**Fluorescence assay**		**Radioactive assay**	
		IC_{50} (nM)*	K_i (nM)	IC_{50} (nM)*	K_i (nM)
1	CoA-C2-SGRG	99.6 ± 11.9	3.9 ± 0.47	26.7 ± 3.90	1.1 ± 0.15
2	CoA-C3-SGRG	119 ± 18.2	4.7 ± 0.73	7.11 ± 1.82	0.28 ± 0.073
3	CoA-C2-SGRGA	128 ± 10.9	5.1 ± 0.43	25.2 ± 3.35	1.0 ± 0.13
4	CoA-C3-SGRGA	112 ± 15.4	4.4 ± 0.61	35.4 ± 5.39	1.4 ± 0.21
5	CoA-C2-SGRGK	218 ± 39.3	8.7 ± 1.6	19.5 ± 3.23	0.78 ± 0.12
6	CoA-C3-SGRGK	164 ± 15.0	6.5 ± 0.60	4.13 ± 0.39	0.17 ± 0.016
7	CoA-C2-SGR	79.0 ± 4.82	3.1 ± 0.19	15.8 ± 4.90	0.63 ± 0.20
8	CoA-C3-SGR	118 ± 13.0	4.7 ± 0.52	41.2 ± 6.72	1.6 ± 0.26
9	CoA-C2-S	> 500 uM		> 370 uM	
10	Propionyl-SGRGK	> 500 uM		> 500 uM	

Note: * both AcCoA and peptide substrate were at 4 Km.

* IC_{50} values were determined in triplicates (n = 3) and presented as mean ± standard deviation (SD).

Selectivity Study. Two of the most potent bisubstrate analogues **6** and **7** were chosen to examine their selectivity against a panel of closely-related N-terminal acetyltransferases including NatA, NatB, NatC, NatE; and protein lysine acetyltransferases PCAF and hMOF, which all harbor an Ac-CoA binding motif. Both **6** and **7** did not show any inhibition towards NatB, NatC, NatE, and PCAF at 10 uM, indicating over

10,000-fold selectivity for NatD. Compound **8** demonstrated 10-40% of inhibition against both NatA and hMOF at 1.0 μM, indicating over 1,000 fold selectivity for NatD over NatA and hMOF. **7** displayed IC_{50} values over 0.1 uM against both NatA and hMOF, suggesting that it was less selective than **6**.

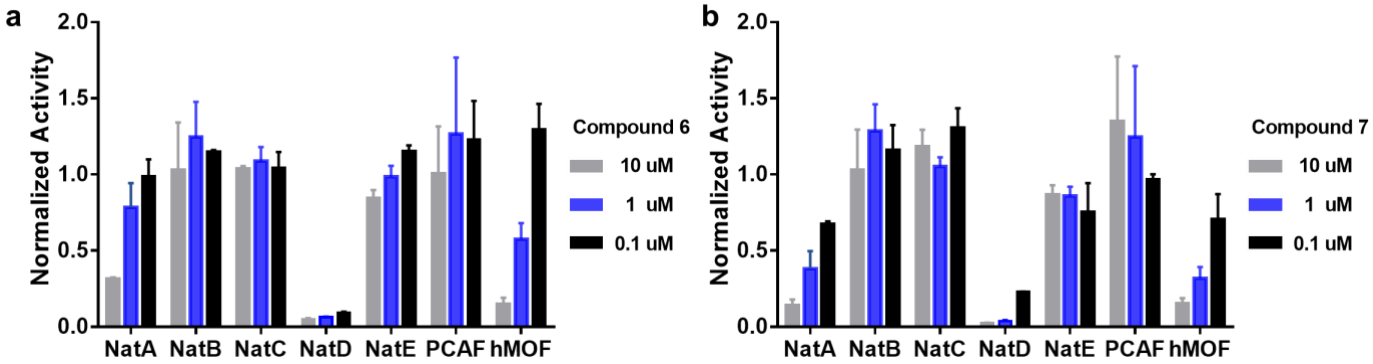


Figure 2. Selectivity study of CoA-C3-SGRGK (**6)** and CoA-C2-SGR (**7)** against a panel of protein acetyltransferases in triplicate (n=3).

Co-crystal Structures of Compounds 5 and 6 with NatD. To understand the molecular interactions between the bisubstrate inhibitors and NatD, the X-ray co-crystal structures of NatD complexed with **5** (PDB ID: 7KD7) and **6** (PDB ID: 7KPU) were obtained at 1.44 Å and 1.43 Å resolution, respectively **(Figure 3a-b)**. The structure refinement statistics for these structures can be found in **Table 2**. Overall, both structures showed high similarity to the previously reported NatD/CoA/SGRGK ternary complex (PDB ID: 4U9W).[8] Structural alignment of the **5-NatD** and **6-NatD** binary complexes with the NatD/CoA/SGRGK ternary complex gave RMSD values of 0.255 Å and 0.217 Å, respectively. Interestingly, despite the different length of the linker region within the two bisubstrate inhibitors, the sulfur atom of the CoA and the nitrogen atom of the peptide Nα-amino group remain in the same position (**Figure 3c)**. Thus, such similar positioning of these two atoms explains the comparable potency of the inhibitors with two different linkers as described above. Also noteworthy is that CoA-C2-SGRGK and CoA-C3-SGRGK adopt slightly different CoA binding modes, especially with respect to the β-mercaptoethylamine group **(Figure 3c)**. Thus, we speculate that the flexibility of this CoA binding region endows NatD the ability to accept a bisubstrate inhibitor CoA-C3-SGRGK with a relatively long linker. Nevertheless, tight binding of bisubstrate analogues

supports the Bi-Bi mechanism of NatD.

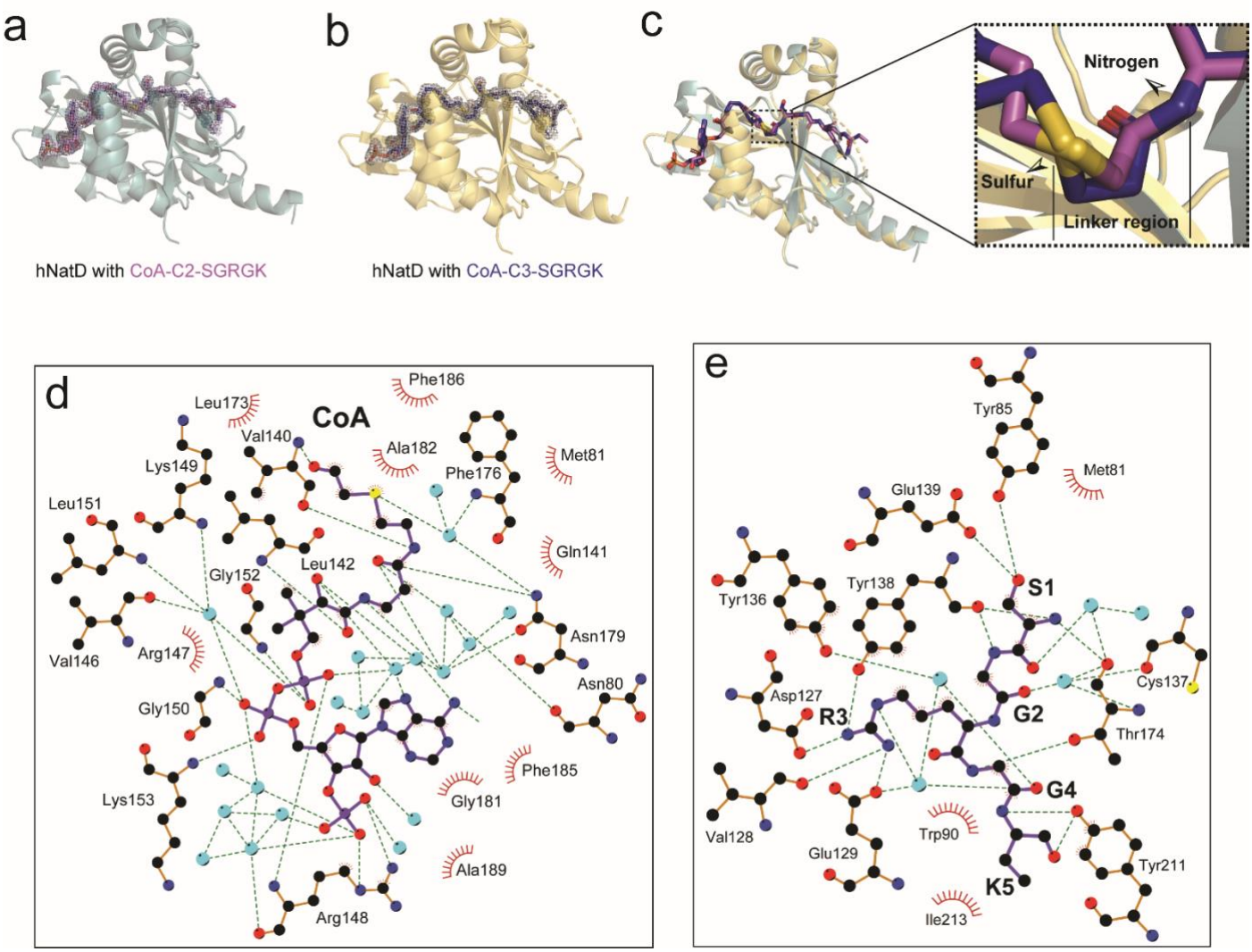


Figure 3. Structures of hNatD with Inhibitors.

(a) Structure of hNatD with **5** (CoA-C2-SGRGK) bound. hNatD is shown in cyan as a cartoon and CoA-C2-SGRGK as magenta sticks. The $2mF_{obs}$-DF_{cal} electron density around the bisubstrate is shown at a contour level of 1σ (gray mesh).

(b) Structure of hNatD with **6** (CoA-C3-SGRGK) bound. hNatD is shown in yellow as a cartoon and CoA-C3-SGRGK as blue sticks. The $2mF_{obs}$-DF_{cal} electron density around the bisubstrate is shown at a contour level of 1σ (gray mesh).

(c) Overlay of the structure in (a) and (b). The zoom-in view shows the positions of the sulfur atom of CoA and the nitrogen of the peptide amino group.

(d) Interaction between CoA and hNatD residues is generated with LIGPLOT Hydrogen bonds are indicated

by dashed green lines, and van der Waals interactions are indicated with red semicircles. Waters molecules are shown as cyan spheres.

(e) Interaction between SGRGK and hNatD residues is generated with LIGPLOT. Hydrogen bonds are indicated by dashed green lines, and van der Waals interactions are indicated with red semicircles. Waters molecules are shown as cyan spheres.

Given the similarlity between both NatD/inhibitor structures , we focused on the structure of NatD with CoA-C2-Ser_1-Gly_2-Arg_3-Gly_4-Lys_5 to further examine the molecular basis of NatD inhibition. An extensive H-bonding network is observed for CoA recognition involving several water molecules and NatD residues Asn80, Val140, Leu142, Val146, Arg148, Lys149, Gly150, Leu151, Gly152, Lys153, Phe176, and Asn179 **(Figure 3d)**. Hydrophobic interactions are also observed from NatD residues Met81, Gln141, Arg 147, Leu173, Gly181, Ala182, Phe185, Phe186, and Ala189. These extensive interactions explain why inhibitory activity drops significantly when the CoA portion is removed (**Table. 1**, compare **6** and **10**). NatD interacts with the peptide portion of the bisubstrate inhibitor through hydrogen bonding to the backbone of residues 1-5, and the sidechains of Ser_1 and Arg_3. These interactions involve several water molecules and NatD residues Tyr85, Asp127, Val128, Glu129, Tyr136, Cys137, Tyr138, Gln139, Thr174, and Tyr121 **(Figure 3e)**. In contrast, van der Waals interactions are significantly more limited, involving only several residues including Met80, Trp90, Ile213. The extensive NatD interaction with the first three residues is consistent with the biochemical inhibitory results: inhibition potency depends on the inclusion of Ser_1 through Arg_3 (**Table. 1**, compare **9** to **1-8**). The NatD/inhibitor structures also reveal that the sidechain of Lys_5 does not mediate any interactions with the enzyme, which is consistent with the biochemical findings that a minimal effect on inhibitor potency was observed when Lys_5 was replaced with Ala (**Table. 1**, compare **3-4** to **5-6**).

Table 2. Data Statistics for NatD Crystal structures with inhibitors.

	NatD bound with CoA-C2-SGRGK	NatD bound with CoA-C3-SGRGK
PDB	**7KD7**	**7KPU**
Crystal Parameters		
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimension a,b,c (Å)	46.433 (90)	46.317 (90)
	74.349 (90)	74.164 (90)
α, β, γ (°)	126.821 (90)	127.022 (90)
Data collection		
Wavelength	0.97918	0.97918
Resolution (Å)	48.25 - 1.439 (1.49 - 1.439)	43.51 - 1.429 (1.48 - 1.429)
Unique reflections	79843 (7732)	81274 (7793)
R_{merge}	0.038 (0.661)	0.034 (0.244)
I/σ	23.1 (2.5)	35.9 (6.2)
Completeness	99.23 (97.07)	99.34 (96.72)
Redundancy	6.4 (6.0)	6.5 (5.7)
Refinement		
R_{work}/R_{free}	0.1614/0.1829	0.1626/0.1827
R.m.s. deviations		
Bonds (Å)	0.018	0.012
Angels (°)	1.587	1.24
Average B factors (Å²)	25.24	20.26
Protein	24.05	18.26
Solvent	36.27	31.54
Ligand	36.13	30.25
Ramachandran statistics (%)		
Favored	97.79	98.72
Allowed	2.21	1.28
Outliers	0	0

Values in parentheses are for the highest-resolution shell.

CONCLUSION

In this work, we designed and synthesized the first series of potent and selective NatD bisubstrate inhibitors **1-8**, exhibiting K_i values ranging from 0.17-1.6 nM. The most potent inhibitor CoA-C3-SGRGK

showed a K_i of 170 ± 16 pM in the radioactive assay, exhibiting over 10,000-fold selectivity for NatD over other acetyltransferases including NatA-C, NatE, hMOF, and PCAF. Compared to reported bisubstrate inhibitors for other NATs including NatA and NatE using a similar bisubstrate strategy,[20] high potency of NatD bisubstrate inhibitors further establish the uniqueness of NatD and strengthen the possibility to develop potent and specific inhibitors for NatD. Furthermore, the co-crystal structures of NatD in complex with CoA-C3-SGRGK and CoA-C2-SGR clearly demonstrate that the bisubstrate inhibitors engaged both substrate and cofactor AcCoA binding sites. The structural observations reveal that the NatD active site is specifically tailored for its histone substrate, explaining the selectivity of the NatD bisubstrate inhibitors. Comparable inhibitory activities and minor difference in the AcCoA binding site for bisubstrate analogues with either an acetyl or propionyl linker suggest active site plasticity, which may result from a subtle conformational change within the cofactor binding site of NatD. Moreover, these co-crystal structures of NatD-inhibitor binary complexes provide a structural foundation to guide the future development of drug-like NatD inhibitors.

EXPERIMENTAL SECTION

Chemistry General Procedures. All chemicals and solvents were purchased from commercial suppliers and used without further purification unless stated otherwise. Preparative high pressure liquid chromatography (RP-HPLC) was performed on an Agilent 1260 Series system. Systems were run with 0-95% methanol/water gradient with a 0.1% TFA modifier. High-resolution Matrix-assisted laser desorption/ionization (MALDI) spectra were performed on a 4800 MALDI TOF/TOF mass spectrometry (Sciex) at the Mass Spectrometry and Purdue Proteomics Facility (PPF), Purdue University. Peptides were synthesized on a CEM Liberty Blue peptide synthesizer. Compounds were also characterized and confirmed by TLC-MS or MALDI-MS. The purity of final compounds was confirmed on a Waters LC-MS system and/or Agilent 1260 Series system. Systems were run with 0-40% methanol/water gradient with 0.1% TFA. The purity of all target compounds showed >95%.

General procedure A for solid-phase peptide synthesis. The Peptides were synthesized using a Liberty Blue automated microwave peptide synthesizer (CEM Corp., Matthews, NC, USA) following a standard

Fmoc protocol.[23] A Rink Amide MBHA resin (0.05 mmol) was used as solid support and placed in the microwave tube. Standard couplings of amino acids were carried out at 0.2 M in DMF and external amino acids at 0.1 M in DMF using 0.5 M DIC and 1.0 M Oxyma in DMF for activation and 20% piperidine in DMF for deprotection. The resin was transferred to a filter-equipped syringe, washed with CH_2Cl_2 (3 mL) and MeOH (3 mL) three times, and dried under air.

General procedure B for the synthesis of bromoacetylated peptide.[20] To a suspension of peptide on resin (0.05 mmol, 1.0 equiv) in DMF (2 mL) were added corresponding acid (0.1 mmol, 2 equiv), DIC (15.5 μL, 0.1 mmol, 2 equiv), and HOBt (13.5 mg, 0.1 mmol, 2 equiv). The suspension was shaken at room temperature for 20 h. The solvent was filtered, and the resulting resin was washed with CH_2Cl_2 (3 mL) and MeOH (3 mL) three times and dried under air. The dried peptide was cleaved from the resin using a cleavage cocktail (TFA/TIPS/ddH2O = 95/2.5/2.5 v/v) (5 mL) for 0.05 mmol scale of resin for 4 h. The suspension was filtered, washed with TFA (2 mL) and the volatiles of the filtrate was removed under N_2. The peptide solution was precipitated with cold anhydrous ether (10 mL), centrifuged at 4200 rpm for 10 min. The supernatant was discarded and the pellet was washed with cold ether, centrifuged, removed supernatant, and air-dried. The dried peptide was dissolved in ddH2O (5 mL) and filtered through a 0.2 μM filter membrane. The filtered sample solution was purified by preparative reversed-phase high performance liquid chromatography (RP-HPLC) using an Agilent 1260 Series system with a C18 column (5 μm, 10 mm × 250 mm) at a flow rate of 4.0 mL/min. Two mobile phases (mobile phase A consisting of 0.1% trifluoroacetic acid in ddH2O; mobile phase B consisting of methanol) were used and monitored at 214 and 256 nm. An injection volume of 400 μL of the solution was injected into the column. The desired fractions were evaporated and the resulting solution was frozen (–85 °C) and lyophilized.

General procedure C for synthesis of the bisubstrate analogues.[20] The bromoacetylated peptide (1 equiv) was dissolved in 1 mL triethylammonium bicarbonate buffer, PH 8.4±0.1. The mixture is added with coenzyme A trilithium salt dihydrate (2 equiv) and the solution was allowed to react for 4 h at room temperature, and then overnight at 4 ℃. 100 uL of the reaction solution was purified by HPLC to obtain the desired product, which was detected by MALDI.

Compound **1** (5.1 mg, 43% yield) was synthesized by following the general procedure C as a foamy white

solid. MALDI-TOF (positive) *m/z*: calcd for $C_{36}H_{63}N_{15}O_{22}P_3S^+$ $[M + H]^+$ *m/z* 1182.3200, found *m/z* 1182.3802.

Compound **2** (4.3 mg, 36 % yield) was synthesized by following the general procedure C as a foamy white solid. MALDI-TOF (positive) *m/z*: calcd for $C_{37}H_{65}N_{15}O_{22}P_3S^+$ $[M + H]^+$ *m/z* 1196.3357, found *m/z* 1196.3724.

Compound **3** (5.6 mg, 45% yield) was synthesized by following the general procedure C as a foamy white solid. MALDI-TOF (positive) *m/z*: calcd for $C_{39}H_{68}N_{16}O_{23}P_3S^+$ $[M + H]^+$ *m/z* 1253.3571, found *m/z* 1253.3676.

Compound **4** (4.8 mg, 38% yield) was synthesized by following the general procedure C as a foamy white solid. MALDI-TOF (positive) *m/z*: calcd for $C_{40}H_{70}N_{16}O_{23}P_3S^+$ $[M + H]^+$ *m/z* 1267.3728, found *m/z* 1267.4188.

Compound **5** (6.2 mg, 47% yield) was synthesized by following the general procedure C as a foamy white solid. MALDI-TOF (positive) *m/z*: calcd for $C_{42}H_{75}N_{17}O_{23}P_3S^+$ $[M + H]^+$ *m/z* 1310.4150, found *m/z* 1310.4417.

Compound **6** (4.8 mg, 36% yield) was synthesized by following the general procedure C as a foamy white solid. MALDI-TOF (positive) *m/z*: calcd for $C_{43}H_{77}N_{17}O_{23}P_3S^+$ $[M + H]^+$ *m/z* 1324.4306, found *m/z* 1324.4221.

Compound **7** (4.7 mg, 42% yield) was synthesized by following the general procedure C as a foamy white solid. MALDI-TOF (positive) *m/z*: calcd for $C_{34}H_{60}N_{14}O_{23}P_3S^+$ $[M + H]^+$ *m/z* 1125.2986, found *m/z* 1125.2793.

Compound **8** (3.8 mg, 33% yield) was synthesized by following the general procedure C as a foamy white solid. MALDI-TOF (positive) *m/z*: calcd for $C_{35}H_{62}N_{14}O_{23}P_3S^+$ $[M + H]^+$ *m/z* 1139.3142, found *m/z* 1139.2950.

Compound **9** (4.7 mg, 51% yield) was synthesized by following the general procedure C as a foamy white solid. MALDI-TOF (positive) *m/z*: calcd for $C_{26}H_{45}N_9O_{19}P_3S^+$ $[M + H]^+$ *m/z* 912.1760, found *m/z* 912.1606.

Compound **10** (3.1 mg, 55% yield) was synthesized by following the general procedure B as a foamy white solid. MALDI-TOF (positive) *m/z*: calcd for $C_{22}H_{43}N_{10}O_7^+$ $[M + H]^+$ *m/z* 559.3311, found *m/z* 559.3316.

Protein Expression and Purification.

Expression and purification of human PCAF, hMOF, NatA-E except NatC, were performed as previously described.[8,17,24-27]

***NatC*:** Ternary *S. pombe* NatC encoding NAA30^{FL}, NAA35$^{31-708,}$ and NAA38$^{48-116}$, was expressed in *E. coli* cells, and purified as follows. Cells were harvested by centrifugation, resuspended, and lysed by sonication in lysis buffer containing 25 mM Tris, pH 8.0, 300 mM NaCl, 10 mg/ml PMSF (phenylmethanesulfonylfluoride). The lysate was clarified by centrifugation and passed over a nickel resin (Thermo Scientific), which was subsequently washed with ~10 column volumes of wash buffer containing 25 mM Tris, pH 8.0, 300 mM NaCl, 20 mM imidazole, 10 mM 2-mercaptoethanol. The protein was eluted with buffer containing 25 mM Tris, pH 8.0, 300 mM NaCl, 200 mM imidazole, 10 mM βME. After elution, His-tagged Ulp1 protease was added to the eluent to cleave the SUMO tag. The eluent was further dialyzed into a buffer containing 25 mM sodium citrate monobasic, pH 5.5, 10 mM NaCl and 10 mM 2-mercaptoethanol. Protein was purified with a 5-mL HiTrap SP ion-exchange column and eluted with a salt gradient (10–1000 mM NaCl). Peak fractions were concentrated to ~ 0.5 mL with a 50 kDa concentrator (Amicon Ultra, Millipore), and loaded onto an S200 gel-filtration column (GE Healthcare) in a buffer containing 25 mM HEPES, pH 7.0, 200 mM NaCl, 1 mM dithiothreitol (DTT). Peak fractions were concentrated to ~ 15 mg/ml as measured by UV280 and flash-frozen for storage in −80 °C until use.

Fluorescence Assays. A fluorescence-based assay was applied to study the IC_{50} values for all the compounds. The assay was performed under the following conditions in a final well volume of 40 µL: 25 mM HEPES (pH = 7.5), 150 mM NaCl, 0.01% Triton X-100, 0.05 µM NatD, 0.5 µM AcCoA, and 15 µM ThioGlo4. The inhibitors were added at concentrations ranging from 0.15 nM to 10 µM. After 10 min incubation, reactions were initiated by the addition of 5.0 µM H4-8 peptide. Fluorescence was monitored on a BMG ClariOtar microplate reader with excitation 400 nm and emission 465 nm. Data were processed by using GraphPad Prism software 7.0.

Radioisotopic Acetyltransferase Assay. FL NatD was used for activity assays. NatD Acetyltransferase assays were carried out in 25 mM HEPES, pH 7.5, 150 mM NaCl, and 1 mM DTT. The H4 substrate peptide used in the assay corresponds to the first 19 residues of human H4 (NH2-SGRGKGGKGLGKGGAKRHR-COOH; GenScript). In the assay, 50 nM of hNatD was mixed with 2 μM of radiolabeled [^{14}C]acetyl-CoA (4 mCi/mmol; PerkinElmer Life Sciences), 20 μM of the peptide, and inhibitors of concentrations ranging from 0.15 nM to 10 μM, for a reaction of 30 minutes at room temperature. To quench the reaction, the reaction solution was applied onto negatively charged P81 paper disks (SVI, St vincent's institute medical research) to trap the peptides, 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 minutes 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. Each reaction was performed in triplicate. IC50 statistic values were determined with GraphPad Prism software 7.0.[8]

Selectivity Assays. The selectivity studies of NatA, NatB, NatC, NatE, PCAF, and hMOF were performed as follows. 100 nM of hNatA was mixed with 30 μM of [^{14}C]acetyl-CoA and 30 μM of either H4 peptide or SASE peptide (NH_2-SASEAGVRWGRPVGRRRRP-COOH; GenScript), with none, 0.1 μM, 1 μM or 10 μM of inhibitors, for a 12-minute reaction at room temperature, in the buffer containing 75 mM HEPES, pH 7.5, 120 mM NaCl, 1mM DTT.

100 nM of hNatB was mixed with 50 μM [^{14}C]acetyl-CoA and 50 μM of MDVF peptide (NH_2-MDVFMKGRWGRPVGRRRRP-COOH, GenScript), with none, 0.1 μM, 1 μM or 10 μM of inhibitors, for a 10-minute reaction at room temperature, in the buffer containing 75 mM HEPES, pH 7.5, 120 mM NaCl, 1 mM DTT.

50 nM of *Sp*NatC was mixed with 30 μM [^{14}C]acetyl-CoA and 10 μM of MLRF peptide (NH_2-ML RFVTKRWGRPVGRRRRPCOOH, GenScript), with none, 0.1 μM, 1 μM or 10 μM of inhibitors, for a 5-

minute reaction at room temperature, in the buffer containing 75 mM HEPES, pH 7.0, 120 mM NaCl, 1 mM DTT.

300 nM of hNatE was mixed with 50 μM [^{14}C]acetyl-CoA and 100 μM of MLGP peptide (NH_2-MLGPEGGRWGRPVGRRRRP-COOH, GenScript), with none, 0.1 μM, 1 μM or 10 μM of inhibitors, for a 40-minute reaction at room temperature, in the buffer containing 75 mM HEPES, pH 7.0, 120 mM NaCl, 1 mM DTT.

100 nM of PCAF was mixed with 50 μM [^{14}C]acetyl-CoA and 400 μM of H4 peptide, with none, 0.1 μM, 1 μM or 10 μM of inhibitors, for a 20-min reaction at room temperature, in the buffer containing 40 mM Tris, pH 8.0, 100 mM NaCl, 1 mM DTT and 2 mg/mL BSA.

50 nM of hMOF was mixed with 50 μM [^{14}C]acetyl-CoA and 400 μM of H4 peptide, with none, 0.1 μM, 1 μM or 10 μM of inhibitors, for a 20-min reaction at room temperature, in the buffer containing 100 mM Tris, pH 8.0, 50 mM NaCl, 1 mM DTT, 800 μM cysteine and 0.25 mg/ml BSA.

.

Co-crystallization and Structure Determination. For co-crystallization and structure determination of **5** and **6** with NatD, a truncation construct of hNatD$^{17–220}$ was used and purified similarly as described above. 10 mg/ml of purified hNatD$^{17-220}$ was incubated with 1 mM of either **5** or **6** for 30 minutes in ice before the crystal setup**.** The best crystals of hNatD$^{17-220}$ with **5** bound was obtained with hanging-drop vapor diffusion at 20 °C in a well containing 0.1 M Bis-Tris, pH 5.5, 2 M Ammonium sulfate, in a drop containing a 1:1.5 mixture of protein to a well solution. The best crystals of hNatD$^{17-220}$ with **6** bound was obtained with hanging-drop vapor diffusion at 20 °C in a well solution containing 0.1 M Bis-Tris, pH 5.5, 2 M Ammonium sulfate, in a drop containing a 1.25:1 mixture of protein to a well solution. All crystals were cryoprotected by transferring them to their respective well solutions supplemented with 20% glycerol before being flash frozen in liquid nitrogen. Data were collected at the Advanced Photon Source (beamline 24-ID-E) and processed using HKL2000.[28]

Structure Determination and Refinement. Both structures were determined by molecular replacement using the structure of NatD/CoA/SGRGK (PDB: 4U9W) with ligands and solvent molecules removed from the search model. Molecular replacement was done using Phaser in Phenix.[29] Initial Manual model building was done in Coot[30] and all subsequent rounds of refinement were performed using Phenix refine and Coot interchangeably. Refinement statistics can be found in **Table 2.** The final model and structure factors were submitted to the Protein Data Bank. Distance calculations, as well as three-dimensional alignment r.m.s. deviations and graphics were generated in PyMOL (http:// www.pymol.org/).

ASSOCIATED CONTENT

Supporting Information

HRMS and HPLC spectra of compounds **1**-**10**.

Accession Codes

The coordinates for the structure of human NatD in complex with compound **5 (**PDB ID: 7KD7**) and 6 (**PDB ID: 7KPU**)** have been deposited in the Protein Data Bank. Authors will release the atomic coordinates and experimental data upon article publication.

Author Information

Corresponding Author

*Phone: (765) 494 3426. E-mail: huang-r@purdue.edu

ORCID

Rong Huang: 0000-0002-1477-3165

Ronen Marmorstein: 0000-0003-4373-4752

Author Contributions

Y.D. synthesized and characterized all compounds described in the manuscript. Y.H. and Y.D. performed fluorescence assay. Z.H. contributed to the synthesis. Y.D. and S.D. prepared manuscript figures and text. S.D. performed the radioactivity assay and obtained the co-crystal structures. S.D. and S.M.G. performed the selectivity study. R.M. designed and supervised experiments by S.D. and S.M.G. and prepared manuscript text. R.H. developed the concept, designed and supervised experiments by Y.D., Y.H., and Z.H., and prepared manuscript text and figures. All authors have read and approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors acknowledge the support from NIH grant R35GM118090 (RM) and proteomic core facility at Purdue University Center for Cancer Research (P30 CA023168). We also thank the support from the Department of Medicinal Chemistry and Molecular Pharmacology (RH) and Leah Gottlieb for her assistance in obtaining proteins of PCAF and hMOF for the inhibitor selectivity studies.

ABBREVIATIONS

NatA, protein N-terminal acetyltransferase A; CoA, coenzyme A; AcCoA, acetyl coenzyme A; rt, room temperature; TFA, trifluoroacetic acid.

REFERENCES

(1) Arnesen, T.; Van Damme, P.; Polevoda, B.; Helsens, K.; Evjenth, R.; Colaert, N.; Varhaug, J. E.; Vandekerckhove, J.; Lillehaug, J. R.; Sherman, F.; Gevaert, K. Proteomics Analyses Reveal the Evolutionary Conservation and Divergence of N-Terminal Acetyltransferases from Yeast and Humans. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 8157-8162. https://doi.org/10.1073/pnas.0901931106.

(2) Scott, D. C.; Monda, J. K.; Bennett, E. J.; Harper, J. W.; Schulman, B. A. N-Terminal Acetylation Acts as an Avidity Enhancer within an Interconnected Multiprotein Complex. *Science* **2011**, *334*,

674-678. https://doi.org/10.1126/science.1209307.

(3) Behnia, R.; Panic, B.; Whyte, J. R. C.; 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*, 405–413. https://doi.org/10.1038/ncb1120.

(4) Aksnes, H.; Ree, R.; Arnesen, T. Co-Translational, Post-Translational, and Non-Catalytic Roles of N-Terminal Acetyltransferases. *Molecular Cell*. **2019**, *73*, 1097-1114. https://doi.org/10.1016/j.molcel.2019.02.007.

(5) Deng, S.; Marmorstein, R. Protein N-Terminal Acetylation: Structural Basis, Mechanism, Versatility, and Regulation. *Trends in Biochemical Sciences*. **2021**, *46*, 15-27 . https://doi.org/10.1016/j.tibs.2020.08.005.

(6) Ree, R.; Varland, S.; Arnesen, T. Spotlight on Protein N-Terminal Acetylation. *Experimental and Molecular Medicine*. **2018**, *50*, 1-13. https://doi.org/10.1038/s12276-018-0116-z.

(7) Hole, K.; van Damme, P.; Dalva, M.; Aksnes, H.; Glomnes, N.; Varhaug, J. E.; Lillehaug, J. R.; Gevaert, K.; Arnesen, T. The Human N-Alpha-Acetyltransferase 40 (HNaa40p/HNatD) Is Conserved from Yeast and N-Terminally Acetylates Histones H2A and H4. *PLoS One* **2011**, *6*, e24713. https://doi.org/10.1371/journal.pone.0024713.

(8) Magin, R. S.; Liszczak, G. P.; Marmorstein, R. The Molecular Basis for Histone H4- and H2A-Specific Amino-Terminal Acetylation by NatD. *Structure* **2015**, *23*, 332–341. https://doi.org/10.1016/j.str.2014.10.025.

(9) Polevoda, B.; Norbeck, J.; Takakura, H.; Blomberg, A.; Sherman, F. Identification and Specificities of N-Terminal Acetyltransferases from Saccharomyces Cerevisiae. *EMBO J.* **1999**, *18*, 6155-6168. https://doi.org/10.1093/emboj/18.21.6155.

(10) Pavlou, D.; Kirmizis, A. Depletion of Histone N-Terminal-Acetyltransferase Naa40 Induces P53-Independent Apoptosis in Colorectal Cancer Cells via the Mitochondrial Pathway. *Apoptosis* **2016**, *21*, 298-311. https://doi.org/10.1007/s10495-015-1207-0.

(11) Liu, Z.; Liu, Y.; Wang, H.; Ge, X.; Jin, Q.; Ding, G.; Hu, Y.; Zhou, B.; Chen, Z.; Ge, X.; Zhang, B.; Man, X.; Zhai, Q. Patt1, a Novel Protein Acetyltransferase That Is Highly Expressed in Liver and Downregulated in Hepatocellular Carcinoma, Enhances Apoptosis of Hepatoma Cells. *Int. J. Biochem. Cell Biol.* **2009**, *41*, 2528–2537. https://doi.org/10.1016/j.biocel.2009.08.009.

(12) Schiza, V.; Molina-Serrano, D.; Kyriakou, D.; Hadjiantoniou, A.; Kirmizis, A. N-Alpha-Terminal Acetylation of Histone H4 Regulates Arginine Methylation and Ribosomal DNA Silencing. *PLoS Genet.* **2013**, *9*, e1003805. https://doi.org/10.1371/journal.pgen.1003805.

(13) Ju, J.; Chen, A.; Deng, Y.; Liu, M.; Wang, Y.; Wang, Y.; Nie, M.; Wang, C.; Ding, H.; Yao, B.; Gui, T.; Li, X.; Xu, Z.; Ma, C.; Song, Y.; Kvansakul, M.; Zen, K.; Zhang, C. Y.; Luo, C.; Fang, M.; Huang, D. C. S.; Allis, C. D.; Tan, R.; Zeng, C. K.; Wei, J.; Zhao, Q. NatD Promotes Lung Cancer Progression by Preventing Histone H4 Serine Phosphorylation to Activate Slug Expression. *Nat. Commun.* **2017**, *8*, 1-14. https://doi.org/10.1038/s41467-017-00988-5.

(14) Liszczak, G.; Goldberg, J. M.; Foyn, H.; Petersson, E. J.; Arnesen, T.; Marmorstein, R. Molecular Basis for N-Terminal Acetylation by the Heterodimeric NatA Complex. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1098. https://doi.org/10.1038/nsmb.2636.

(15) Evjenth, R. H.; Brenner, A. K.; Thompson, P. R.; Arnesen, T.; Frøystein, N. Å.; Lillehaug, J. R. Human Protein N-Terminal Acetyltransferase HNaa50p (HNAT5/HSAN) Follows Ordered Sequential Catalytic Mechanism: Combined Kinetic and NMR Study. *J. Biol. Chem.* **2012**, *287*, 10081–10088. https://doi.org/10.1074/jbc.M111.326587.

(16) Weyer, F. A.; 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*, 1-10.

https://doi.org/10.1038/ncomms15726.

(17) Deng, S.; Pan, B.; Gottlieb, L.; Petersson, E. J.; Marmorstein, R. Molecular Basis for N-Terminal Alpha-Synuclein Acetylation by Human Natb. *Elife* **2020**, *9*, e57491. https://doi.org/10.7554/ELIFE.57491.

(18) Støve, S. I. I.; Magin, R. S. S.; Foyn, H.; Haug, B. E. E.; Marmorstein, R.; Arnesen, T. Crystal Structure of the Golgi-Associated Human Nα-Acetyltransferase 60 Reveals the Molecular Determinants for Substrate-Specific Acetylation. *Structure* **2016**, *24*, 1044-1056. https://doi.org/10.1016/j.str.2016.04.020.

(19) Goris, M.; Magin, R. S.; Foyn, H.; Myklebust, L. M.; Varland, S.; Ree, R.; Drazic, A.; Bhambra, P.; Støve, S. I.; Baumann, M.; Haug, B. E.; Marmorstein, R.; Arnesen, T. Structural Determinants and Cellular Environment Define Processed Actin as the Sole Substrate of the N-Terminal Acetyltransferase NAA80. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 4405-4410. https://doi.org/10.1073/pnas.1719251115.

(20) Foyn, H.; Jones, J. E.; Lewallen, D.; Narawane, R.; Varhaug, J. E.; Thompson, P. R.; Arnesen, T. Design, Synthesis, and Kinetic Characterization of Protein N-Terminal Acetyltransferase Inhibitors. *ACS Chem. Biol.* **2013**, *8* (6), 1121–1127. https://doi.org/10.1021/cb400136s.

(21) Chen, D.; Dong, G.; Noinaj, N.; Huang, R. Discovery of Bisubstrate Inhibitors for Protein N-Terminal Methyltransferase 1. *J. Med. Chem.* **2019**, *62*, 3773–3779. https://doi.org/10.1021/acs.jmedchem.9b00206.

(22) Ho, Y. H.; Chen, L.; Huang, R. Development of a Continuous Fluorescence-Based Assay for n-Terminal Acetyltransferase D. *Int. J. Mol. Sci.* **2021**, *22*, 1–12. https://doi.org/10.3390/ijms22020594.

(23) Collins, J. M.; Porter, K. A.; Singh, S. K.; Vanier, G. S. High-Efficiency Solid Phase Peptide Synthesis (He -Spps). *Org. Lett.* **2014**, *16*, 940-943. https://doi.org/10.1021/ol4036825.

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

(25) Deng, S.; Magin, R. S.; Wei, X.; Pan, B.; Petersson, E. J.; Marmorstein, R. Structure and Mechanism of Acetylation by the N-Terminal Dual Enzyme NatA/Naa50 Complex. *Structure* **2019**, *27*, 1057-1070. e4. https://doi.org/10.1016/j.str.2019.04.014.

(26) Deng, S.; McTiernan, N.; Wei, X.; Arnesen, T.; Marmorstein, R. Molecular Basis for N-Terminal Acetylation by Human NatE and Its Modulation by HYPK. *Nat. Commun.* **2020**, *11*, 1-14. https://doi.org/10.1038/s41467-020-14584-7.

(27) McCullough, C. E.; Song, S.; Shin, M. H.; Johnson, F. B.; Marmorstein, R. Structural and Functional Role of Acetyltransferase HMOF K274 Autoacetylation. *J. Biol. Chem.* **2016**, *291*, 18190-18198. https://doi.org/10.1074/jbc.M116.736264.

(28) Otwinowski, Z.; Minor, W. Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* **1997**, *276*, 307-326. https://doi.org/10.1016/S0076-6879(97)76066-X.

(29) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 213-221. https://doi.org/10.1107/S0907444909052925.

(30) Emsley, P.; Cowtan, K. Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60*, 2126-2132. https://doi.org/10.1107/S0907444904019158.

Table of Contents graphic