# Site-specific reactivity of non-enzymatic lysine acetylation

Josue Baeza[1, 2,3], Michael J. Smallegan[2,3], and John M. Denu[1, 2, 3]

From the [1]Department of Biomolecular Chemistry, the [2]Wisconsin Institute for Discovery, University of Wisconsin, Madison, Wisconsin 53715

[3]These authors contributed equally to this work

[4]To whom correspondence should be addressed. University of Wisconsin, 330 N. Orchard St., Madison, WI 53715; Phone: 608-316-4341; Fax: 608-316-4602; E-mail: jmdenu@wisc.edu

## METHODS

### Protein samples

Purified bovine GDH and porcine αKGDH were purchased from Sigma Aldrich. Recombinant Mouse ACAT1 (amino acids 31-424) with C-terminal $(His)_6$ tag was purified according to previous methods.[1] Recombinant Histone H3.2 and H4 (Xenopus laevis) were expressed and purified according to previous methods.[2]

Recombinant mouse TFAM pQE80-L (amino acids 43-246) was expressed by transforming plasmid into BL21 DE3 PlysS *E. coli* and grown in 2XYT media. Cells were grown until mid log phase ($OD_{600}$ 0.6) at 37°C with shaking and induced with 420 μM IPTG and grown overnight at 18°C. Cells were harvested and resuspended in 30 mM sodium phosphate, 100 mM NaCl, 1mM beta-mercaptoethanol, 10 mM imidazole, 1 mM PMSF, 10 μg/mL aprotinin and 10 μg/mL leupeptin. Cells were lysed by sonication and clarified by centrifugation at 16,000 RPM for 30 minutes. Protein was purified nickel chelating chromatography (Wash buffer: 30 mM sodium phosphate, 100 mM NaCl, 1 mM BME, 10 mM imidazole; Elution buffer: 30 mM sodium phosphate, 100 mM NaCl, 1 mM BME, 250 mM imidazole) by FPLC (AKTA, GE Healthcare Life Sciences ©) using a linear gradient. A 10% step using elution buffer was performed. Protein containing fractions were pooled, concentrated and dialyzed overnight. Protein concentration was determined using Bradford reagent (Bio-Rad).

### Time-dependent acetylation

Purified protein samples were incubated at a 1 $mg{\cdot}mL^{-1}$ concentration in 100 mM ammonium bicarbonate buffer (pH 8.4) with acetyl-coa or acetyl-phosphate and incubated at 37°C on Thermomixer C at 1000 RPM (Eppendorf) in a 120 uL total volume. Time points were taken at 0, 5, 10, 30, 60 minutes and prepared for mass spectrometry analysis as described below. BSA was treated with 4.5 mM AcCoA and 8.5 mM AcP.

### Concentration-dependent acetylation

Purified protein samples were incubated at 1 $mg{\cdot}mL^{-1}$ concentration in 100 mM ammonium bicarbonate buffer (pH = 8.4) with varying concentration of AcCoA or AcP 0.5, 1, 2, 4, and 8

mM in 40 uL total volume. Reaction was incubated at 37°C on Thermomixer C at 1000 RPM for 60 minutes and prepared for mass spectrometry analysis as described below.

**Sample preparation, chemical acetylation and digestion**

Due to the nature of non-enzymatic reactions, quenching of the reaction requires the removal of the reactive chemical species. To do this, a 20uL aliquot of the reaction volume was transferred to 10 kDa MWCO filter unit (EMD Millipore) that was prefilled with 450 µL urea buffer (8M urea, 100 mM ammonium bicarbonate pH 7.4, 5 mM DTT), washed twice and transferred to a new 1.5mL microcentrifuge tube according to manufacturer's guidelines. This, in essence, lowered the concentration of both protein and acetylating reagent.

Each sample was heat denatured and reduced at 60°C on the Thermomixer C at 1000 RPM for 30 minutes followed by cysteine alkylation using 40 mM iodoacetamide for 30 minutes in the dark. The pH of the reaction volume was increased to ~8 by adding $NH_4OH$ and protein acetylation was performed by adding ~50 µmol acetic anhydride-d6 (99.8% isotopic purity, Cambridge Isotopes) and incubating at 60°C for 30 minutes. Buffer exchange occurred with 50 mM ammonium bicarbonate (pH 8.4) using 10K MWCO filter units. For the reactions using acetyl-CoA, protein was digested with a 1:80 GluC-to-protein ratio for 4 hours at 37°C while shaking at 500 RPM on Thermomixer C. Sample was dried down using speedvac and resuspended using 50 mM ammonium bicarbonate (pH 8.4) and digested with trypsin at a 1:80 ratio. For the reactions using acetyl-phosphate, protein was digested with GluC as above followed by chemical acetylation with acetic anhydride-d6 labeling N-terminus of the GluC peptides. GluC digested peptides were dried down and resuspended in 50 µL of 50 mM ammonium bicarbonate (pH 8.4) and digested with trypsin as above. Peptides were then acidified with 1% TFA, dried down and resuspended in 2% acetonitrile, 0.1% formic acid and analyzed by mass spectrometry as described below.

**Determination of 2$^{nd}$ order rate constant for purified peptide**

The purified peptide, KQTARKSTGGKAPKWW, with $N^{\alpha}$-acetylation was used to determine the rate constant of a primary amine. 1 µmol of purified peptide was incubated in 100 mM ammonium bicarbonate buffer (pH 8.5) with varying concentration of AcCoA (0.5, 1, 2, 4, 8 mM) in 50 uL total volume for 60 minutes at 37°C. To quench the reaction, 50 µmol of acetic anhydride-d6 was added to the sample and incubated at 37°C for 20 min. Sample was acidified with 1% TFA, desalted with OMIX C18 Tips (Agilent Technologies), resuspended in 2% acetonitrile, 0.1% formic acid and analyzed by mass spectrometry as described below.

**LC-MS/MS**

Peptides were separated using a Dionex Ultimate 3000 RSLCnano HPLC using a Waters Atlantis reverse phase column (100 µm x 150 mm). Mobile phase consisted of (A) 0.1% formic acid in HPLC grade $H_2O$ and (B) 0.1% formic acid in HPLC grade acetonitrile. Peptides were eluted in a linear gradient of 2 – 40% B at 700 nL/min over 60 minutes with a column temperature at 60 °C and introduced into a Thermo Q-Exactive hybrid quadrupole Orbitrap mass spectrometer by nanoelectrospray ionization. Survey scan was performed in positive ion mode with a 140,000 resolution, AGC of 1E6, max fill time 250 ms, and a scan range of 350 to 2000

m/z. Data dependent MS/MS was performed with a resolution of 17,500, AGC of 1E5, max fill time 100 ms, isolation window of 1.5 m/z, and a loop count of 10. The source voltage was set at 2000 V and capillary temperature at 250 °C.

**Database search and data analysis**

MS data was searched using high-resolution database search algorithm, Morpheus. Precursor mass tolerance was set to 10 ppm and product mass tolerance to 15 ppm with a false discovery rate of 1%. Up to 2 missed cleavages were allowed for a tandem GluC and Trypsin digest. Carbamidomethylation of cysteines was set as a fixed modification while methionine oxidation and lysine acetylation-d3 (unimod accession #: 56) was set as a variable modification. For Histone H3 and H4 only, we used the database search engine, Mascot, for peptide identification using the same settings as above, except Trypsin was used as the protease.

Extracted ion chromatograms (XICs) for the mono, di, tri acetylated peptides were obtained using Thermo Xcalibur Qual Browser. By using the appropriate mass shift, the light acetyl-peptide was also identified, matched and quantified. ($z$ = 2, 1.509 m/$z$; $z$ = 3, 1.006 m/$z$) For peptides with more than one lysine, mixed isotopic heavy and light acetylated peptides are generated. These mixed isotopic peptides are identified and quantified using mass shifts corresponding to the number of lysines present in the peptide and used for quantitation of stoichiometry, which is determined by the following equations:

1 lysine containing peptide

$$\frac{XIC_L}{XIC_L + XIC_H}$$

2 lysine containing peptides

$$\frac{XIC_L + \frac{1}{2}XIC_{LH}}{XIC_L + XIC_{LH} + XIC_H}$$

3 lysine containing peptides

$$\frac{XIC_{LLL} + \frac{2}{3}XIC_{LLH} + \frac{1}{3}XIC_{LHH}}{XIC_{LLL} + XIC_{LLH} + XIC_{LHH} + XIC_{HHH}}$$

where $XIC_L$ is the extracted ion chromatogram of the light acetyl-peptide, $XIC_H$ is the extracted ion chromatogram of the heavy acetyl-peptide and $XIC_{LH}$, when present, is the extracted ion chromatogram for the double lysine containing peptide with a light and heavy acetyl-lysine

**Acetylation proteomics data aggregation**

Proteomics datasets were retrieved from supplemental data tables accompanying.[1,3–5] Defunct International Protein Index identifiers from Sol et al. 2012 were mapped onto Uniprot Accession

numbers using DAVID's gene ID converter utility.[6,7] Peptides containing multiple lysines were split into one entry per lysine site and each was assigned the fold-change value for the whole peptide. Lysine site data from all studies was then concatenated into one data table. Conditions comparing Sirt3$^{-/-}$ mice to wild type mice were collapsed onto one data column by averaging values from the Sirt3$^{-/-}$/WT experiment in Mouse embryonic fibroblast cells (Sol 2012), Sirt3$^{-/-}$/WT experiment in mouse liver mitochondria (Rardin 2013), and the Sirt3$^{-/-}$ Control Diet/WT in mouse liver mitochondria from Hebert 2013. We subsequently merged this data with our observed lysine sites by Uniprot accession and lysine site number. If multiple data points for each study matched with one of our observed lysine sites, the fold-change value associated with the smallest peptide was taken as a representative fold-change.

***In silico* $pK_a$ prediction**

Prediction of the $pK_a$ of lysine sites of ACAT1 (Model retrieved from SWISS-MODEL Repository for UniProt: Q8QZT1 based on template pdb:2IB8)[8] and GDH (pdb:3MW9) using the H++ Version 3.1 server.[9–11] Computations were performed using the following initial parameters: external dielectric of 80, internal dielectric of 10, and salinity of 0.15 M. pKs from chain A were taken as the representative value for each lysine.

**Solvent accessible surface area calculation**

Solvent accessible surface area per residue was calculated using POPS* Version 1.6.2a.[12,13] A probe radius of 1.4 Å was used and the percent of lysine residue solvent accessible from lysines on chain A were plotted against rates in Supplementary Figure 2.

**Average B-factor calculation**

Temperature factors were extracted from the GDH structure (pdb:3MW9) and for each lysine, the B-factor values corresponding to the constituent atoms were averaged to obtain a per residue average B-factor using PyMol.[14] Values from chain A were used as representative for each lysine site.

**3D lysine motif**

Side chain interaction centers for ACAT1 (Model retrieved from SWISS-MODEL Repository for UniProt: Q8QZT1 based on template pdb:2IB8)[8] and GDH (pdb:3MW9) were calculated using PyMol in accordance with Bahar and Jernigan (1996).[14,15] For each lysine in these two proteins, the distance between the lysine of interest and all side chain interaction centers within 7 Å was calculated.

**SUPPLEMENTARY FIGURES**

**Figure 1**

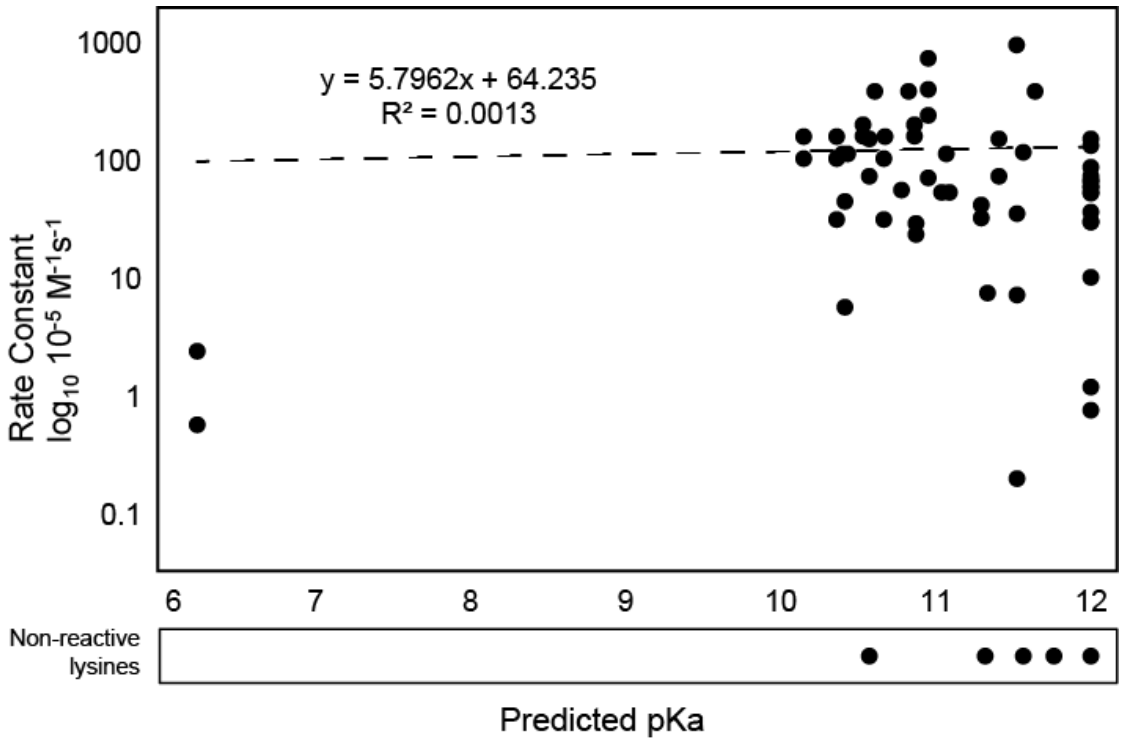


**Figure 1.** Scatter plot of predicted lysine $pK_a$ and second order acetylation rate constants with linear fit. Predicted $pK_a$ value of quantified non-reactive sites represented below and not included in regression. Sites plotted at a $pK_a$ of 12 correspond to H++ server output designated as $pK_a$ of $\geq$12 (the maximum pK value returned by the computation). Includes all observed sites from ACAT1 and GDH.

**Figure 2**

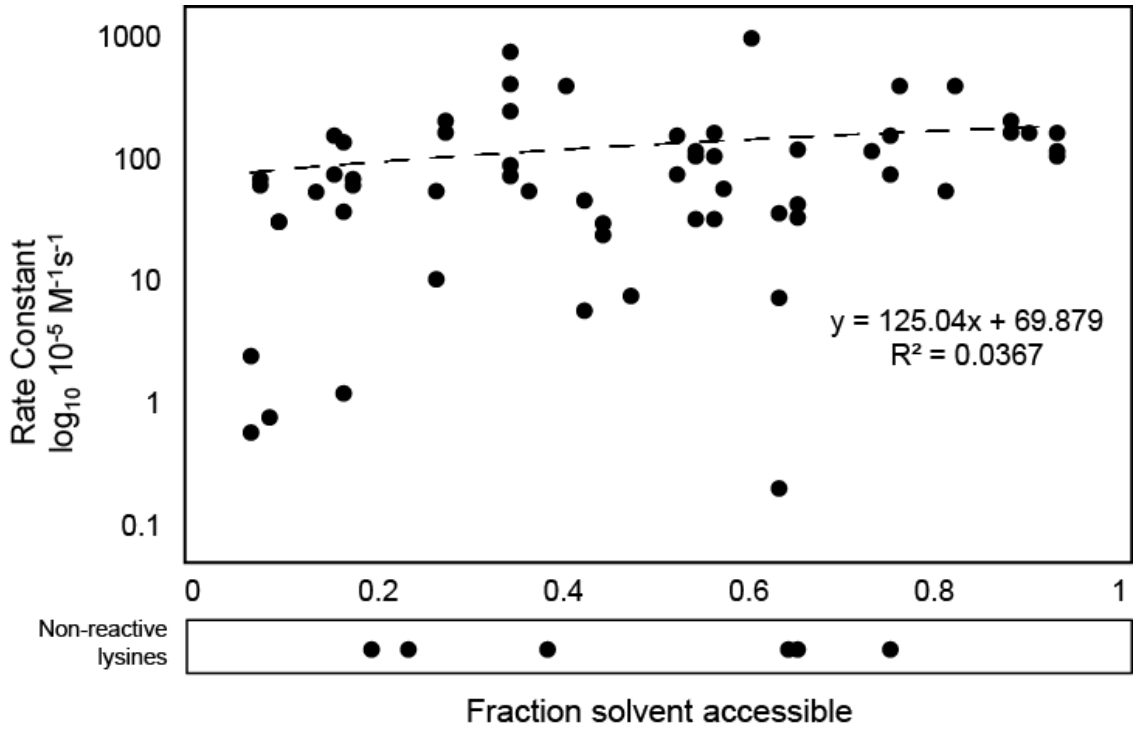


**Figure 2.** Scatter plot of solvent accessible surface area (SASA) and second order acetylation rate constants with linear fit. SASA represented as the fraction of each lysine residue accessible. SASA of non-reactive lysines shown below, and not included in regression. Analysis includes all observed sites from ACAT1 and GDH.

**Figure 3**

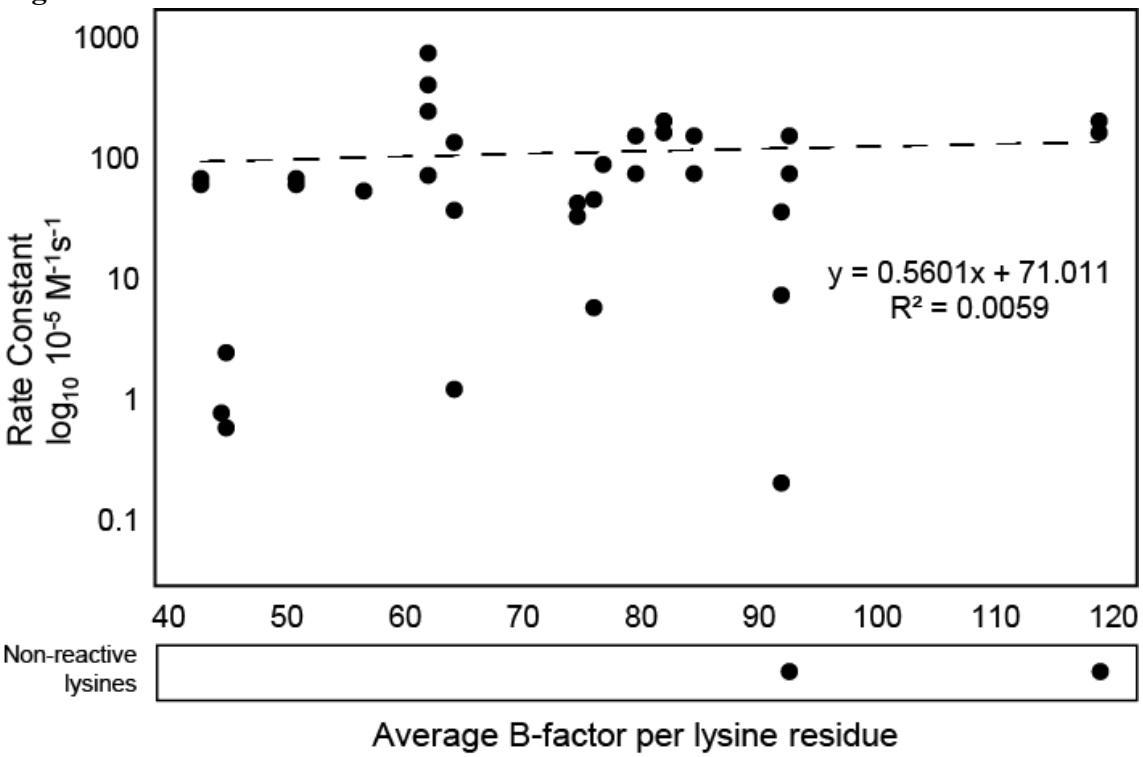


**Figure 2.** Scatter plot of average B-factor and second order acetylation rate constants with linear fit. B-factor of non-reactive lysines shown below, and not included in regression. Includes observed sites from GDH.

**Figure 4**

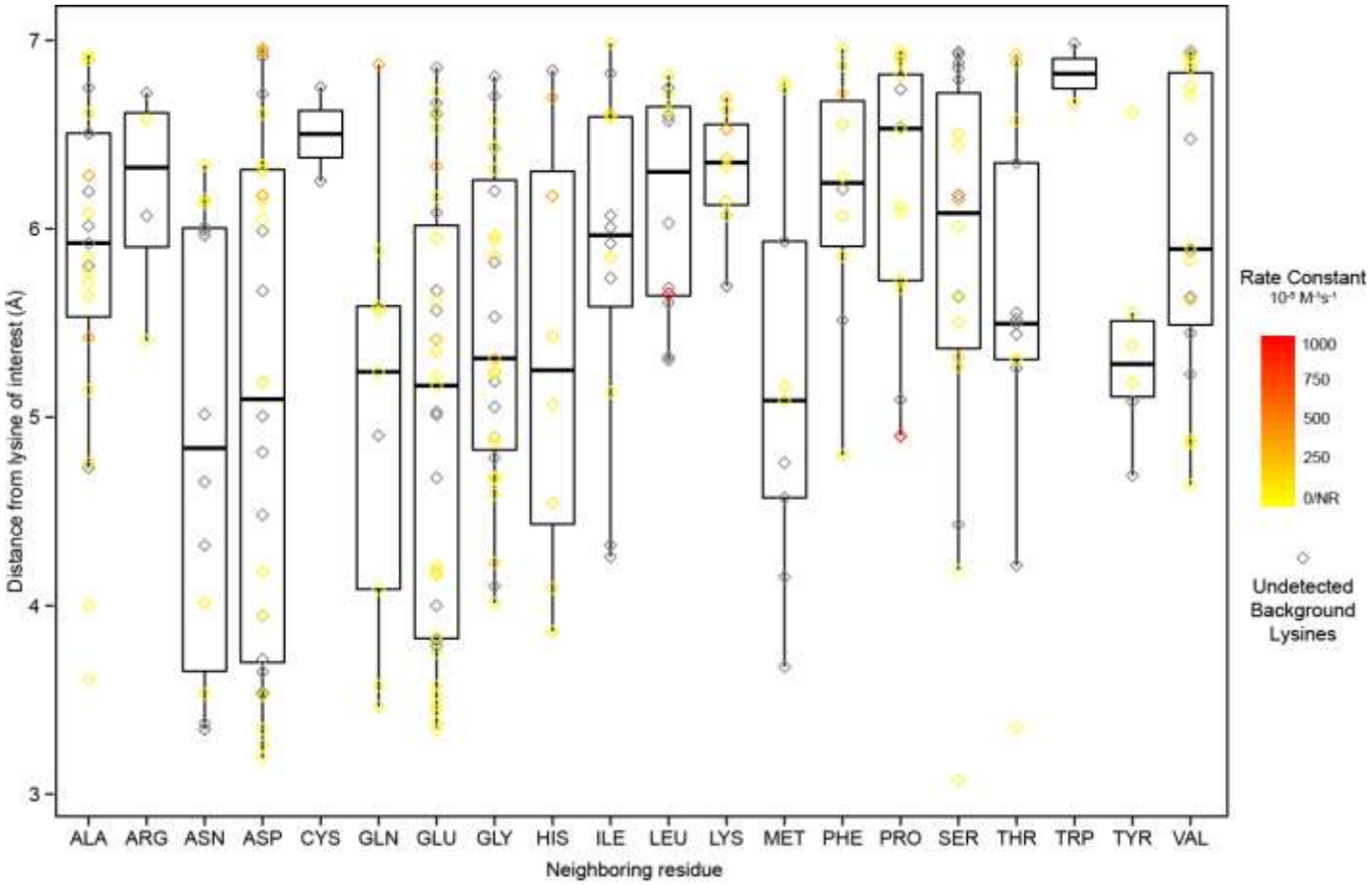


**Figure 4.** 3D structural environment of lysines in ACAT1 and GDH. Pairwise distances between the side chain interaction center of each lysine and all neighboring residues in a 7 Å sphere. Distance markers are colored in accordance with the rate constant associated with the neighboring lysine. Lysines not detected in this study are included (colored in grey) as a background to aid in interpretability. NR on rate constant color scale refers to quantified non-reactive lysines.

## References

(1) Still, A. J., Floyd, B. J., Hebert, A. S., Bingman, C. A., Carson, J. J., Gunderson, D. R., Dolan, B. K., Grimsrud, P. A., Dittenhafer-Reed, K. E., Stapleton, D. S., Keller, M. P., Westphall, M. S., Denu, J. M., Attie, A. D., Coon, J. J., and Pagliarini, D. J. (2013) Quantification of Mitochondrial Acetylation Dynamics Highlights Prominent Sites of Metabolic Regulation. *J. Biol. Chem. 288*, 26209–26219.
(2) Su, Z., Boersma, M., Lee, J.-H., Oliver, S., Liu, S., Garcia, B., and Denu, J. (2014) ChIP-less analysis of chromatin states. *Epigenetics Chromatin 7*, 7.
(3) Hebert, A. S., Dittenhafer-Reed, K. E., Yu, W., Bailey, D. J., Selen, E. S., Boersma, M. D., Carson, J. J., Tonelli, M., Balloon, A. J., Higbee, A. J., Westphall, M. S., Pagliarini, D. J., Prolla, T. A., Assadi-Porter, F., Roy, S., Denu, J. M., and Coon, J. J. (2013) Calorie restriction and SIRT3 trigger global reprogramming of the mitochondrial protein acetylome. *Mol. Cell 49*, 186–199.
(4) Sol, E. M., Wagner, S. A., Weinert, B. T., Kumar, A., Kim, H. S., Deng, C. X., and Choudhary, C. (2012) Proteomic investigations of lysine acetylation identify diverse substrates of mitochondrial deacetylase sirt3. *PLoS One 7*, e50545.
(5) Rardin, M. J., Newman, J. C., Held, J. M., Cusack, M. P., Sorensen, D. J., Li, B., Schilling, B., Mooney, S. D., Kahn, C. R., Verdin, E., and Gibson, B. W. (2013) Label-free quantitative proteomics of the lysine acetylome in mitochondria identifies substrates of SIRT3 in metabolic pathways. *Proc. Natl. Acad. Sci. U. S. A. 110*, 6601–6606.
(6) Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res. 37*, 1–13.
(7) Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc. 4*, 44–57.
(8) Kopp, J., and Schwede, T. (2004) The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res. 32*, D230–4.
(9) Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res. 40*, W537–41.
(10) Myers, J., Grothaus, G., Narayanan, S., and Onufriev, A. (2006) A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins 63*, 928–938.
(11) Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005) H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res. 33*, W368–71.
(12) Cavallo, L., Kleinjung, J., and Fraternali, F. (2003) POPS: A fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Res. 31*, 3364–3366.
(13) Fraternali, F., and Cavallo, L. (2002) Parameter optimized surfaces (POPS): analysis of key interactions and conformational changes in the ribosome. *Nucleic Acids Res. 30*, 2950–2960.
(14) Delano, W. L. (2002) The PyMOL Molecular Graphics System.
(15) Bahar, I., and Jernigan, R. L. (1996) Coordination geometry of nonbonded residues in globular proteins. *Fold. Des. 1*, 357–370.