An Unbiased Approach to Identify Endogenous Substrates of "Histone" Deacetylase 8

Supporting Information

David E. Olson[1], Namrata D. Udeshi[2], Noah A Wolfson[5], Carol Ann Pitcairn[6], Eric D. Sullivan[6], Jacob D. Jaffe[2], Tanya Svinkina[2], Ted Natoli[3], Xiaodong Liu[3], Joshiawa Paulk[4], Patrick McCarren[4], Florence F. Wagner[1], Doug Barker[1], Eleanor Howe[4], Fanny Lazzaro,[1] Jennifer P. Gale,[1] Yan-Ling Zhang,[1] Aravind Subramanian[3], Carol A. Fierke[5,6,7], Steven A. Carr[2], & Edward B. Holson[1]*

*[1]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [2]Proteomics Platform, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [3]Cancer Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [4]Center for the Science of Therapeutics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [5]Department of Biological Chemistry, University of Michigan, Ann Arbor, MI, USA. [6]Interdepartmental Program in Chemical Biology, University of Michigan, Ann Arbor, MI, USA. [7]Department of Chemistry, University of Michigan, Ann Arbor, MI, USA. *email: edholson@broadinstitute.org*

Online Methods

Compound Synthesis.

PCI-34051 is commercially available. BRD3811 was synthesized according to the following procedure: A solution of methyl 5-methyl-1H-indole-6-carboxylate (100 mg, 0.529 mmol, 1.0 equiv), potassium iodide (8.8 mg, 0.053 mmol, 0.1 equiv), and sodium hydride (60% dispersion, 23.3 mg, 0.581 mmol, 1.1 equiv) in DMF (1.6 mL) was stirred at 0 ºC for 1 h. Next, a solution of 1-(chloromethyl)-4-methoxybenzene (124 mg, 0.793 mmol, 1.5 equiv) in DMF (1 mL) was added, and the reaction mixture was heated to 65 ºC for 2h. The reaction was diluted in sat. $Na_2CO_{3(aq)}$ (50 mL) and extracted with EtOAc (3 x 25 mL). The combined organic extracts were washed with brine, dried over $MgSO_4$, filtered, and concentrated under reduced pressure. The resulting residue was purified by flash chromatography (hexanes/EtOAc) to yield methyl 1-(4-methoxybenzyl)-5-methyl-1H-indole-6-carboxylate (99.0 mg, 0.320 mmol, 61%). To a solution of methyl 1-(4-methoxybenzyl)-5-methyl-1H-indole-6-carboxylate (99.0 mg, 0.320 mmol, 1.0 equiv) and sodium hydroxide (64.0 mg, 1.60 mmol, 5.0 equiv) in 1:1 MeOH:THF (1.3 mL) was added 50% aqueous hydroxylamine (0.628 mL, 10.2 mmol, 32 equiv), and the resulting solution was stirred for 7 h. Upon reaction completion, 5 mL of water was added followed by removal of organic solvents under reduced pressure. Neutralization of the remaining aqueous solution with 1 M $HCl_{(aq)}$ resulted in the precipitation of product, which was filtered, washed with cold water, and dried under reduced pressure to yield BRD3811 (45.0 mg, 0.145 mmol, 45%) as a white powder in >95% purity (as determined by LCMS, ESI^+ MS**:** m/z: 311.7 $[M+H]^+$). ^1H NMR (300 MHz, DMSO-d_6): δ 10.72 (br s, 1H), 8.97 (br s, 1H), 7.51 (d, *J* = 2.8 Hz, 1H), 7.41 (s, 1H), 7.37 (s, 1H), 7.16 (d, *J* = 8.4 Hz, 2H), 6.86 (d, *J* = 8.4 Hz, 2H), 6.40 (d, *J* = 2.8 Hz, 1H), 5.31 (s, 2H), 3.70 (s, 3H), 2.36 (s, 3H) ppm. For a ^1H NMR spectrum, see **Supplementary Fig. 5a**. For an HPLC trace demonstrating compound purity and detection of the appropriate mass, see **Supplementary Fig. 5b**.

Caliper Assay.

The biochemical determination of HDAC IC_{50}s for compounds was performed as described previously.[30] All HDACs were purchased from BPS Bioscience. The substrates Broad Substrate A and Broad Substrate B were synthesized in house, but can be purchased from PerkinElmer (Product number CLS960006 and CLS960007, respectively). All other reagents were purchased from Sigma. Caliper EZ reader II system was

used to collect all data. Compounds were tested in duplicate in a 12-point dose curve with 3-fold serial dilution starting from 33.33 μM. Purified HDACs were incubated with 2 μM carboxyfluorescein (FAM)-labeled acetylated or trifluoroacetylated peptide substrate (Broad Substrate A and B, respectively) and test compound for 60 min at room temperature, in an HDAC assay buffer that contained 50 mM HEPES (pH 7.4), 100 mM KCl, 0.01% BSA and 0.001% Tween-20. Reactions were terminated by the addition of the known pan HDAC inhibitor LBH-589 (panobinostat) with a final concentration of 1.5 μM. Substrate and product were separated electrophoretically and fluorescence intensity in the substrate and product peaks was determined and analyzed by Labchip EZ Reader. The percent inhibition was plotted against the compound concentration, and the IC_{50} values were automatically fitted by Genedata Screener software using 4-parameter logistic dose response model. Inhibition of HDAC10 and 11 was not measured due to either low purity of the available recombinant HDAC enzyme preparations and/or lack of activity of the enzymes and low substrate conversion.

Molecular Docking.

The best pose of compound PCI-34051 in the HDAC8 binding site was determined using the zinc HDAC8 structure deposited in the PDB by Somozo et al. (accession code 1T64, Structure 2004, 1235-1334).[31] Unresolved residues (33, 60, 69, 81, 85, 87-89, 221, 238, 377) were added using Prime, however, these residues would not be expected to be important interacting residues. Only one, Lys33, was near the ligand solvent-exposed end of the ligand after addition. Protonation of the structure and formation of sidechain-metal interactions was performed using the Protein Preparation Wizard followed by manual correction in the Schrodinger Drug Discovery Suite 2014-2.[32] Histidines 142 and 143 were protonated at the delta nitrogen and histidine 180, which interacts with zinc at the delta position was protonated at the epsilon position. The binding ligand, Trichostatin A, was used as the center of the docking grid and the final cubic docking grid had the dimensions 10x10x10Å^3 constructed using Schrodinger Glide 6.3 with default parameters. Glide XP was used to dock PCI-34051 and BRD3811 with options to enhance planarity of aromatic groups and to perform post-docking minization.[33] The best docking pose for each was depicted in Figure 1 using PyMol and coloring the surface within 5 Å by hydrophobicity using the color_h.py script detailed on the PyMol Wiki (http://www.pymolwiki.org/index.php/Color_h). PCI-34051 had a Glide XP score of -9.1 and the methyl analog, BRD3811, had a score of only -3.7 with significant decreases in the hydrogen bonding and Coloumbic score

contributions. The docked geometry of BRD3811 does not permit binding with the metal center and the score should be considered exponentially worse than PCI-34051.

Cell Culture and Compound Treatment for SILAC[34] Proteomics Experiments.

MCF7 cells were grown and expanded from the same frozen vial stock, followed by differentially labeling with non-radioactive stable isotopic amino acids by growing in light, medium, and heavy SILAC media (See below for recipe), respectively. MCF7 cells grown in SILAC medium were plated (2 millions cells/plate, 10 mL per plate) into 10-cm tissue culture treated plates and incubated 24 h prior to treatment. For treatment, 10 μl of compounds (10 mM stocks in DMSO) or DMSO vehicle control were added to the plates, and the cells were incubated for 24h. Next, the growth medium was aspirated, and the monolayers of cells were rinsed twice with cold PBS. Cells were detached using a cell scraper and collected with 1 ml of cold PBS. Cell pellets were harvested by centrifugation at 1,500 rpm for 1 min and flash frozen in liquid nitrogen. All pellets were stored in -80 ºC freezer prior to lysis.

SILAC media:
450 mL DMEM
50 mL FBS (Sigma, F-0392)
5 mL 100x Pen/Strep/Glutamine (Gibco 10378)
3.9 mL 45% Glucose solution (Sigma, G8769)
500 μL Methionine 0 (stock 30g/L, final 30mg/L)
500 μL Proline 0 (stock 20g/L, final 20mg/L)
500 μL Lysine 0 or Lysine 4 or Lysine 8 (stock 146g/L, final 146mg/L)
500 μL Arginine 0 or Arginine 6 or Arginine 10 (stock 84g/L, final 84mg/L)

Proteomics.

SILAC-labeled MCF-7 cells were lysed in ice-cold 8 M urea, 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 2 μg/ml aprotinin (Sigma-Aldrich), 10 μg/ml leupeptin (Roche Applied Science), 1 mM phenylmethylsulfonyl fluoride (PMSF), and 5 mM sodium butyrate (Sigma-Aldrich). Lysates were centrifuged at 20,000 x g for 10 min at 4 ºC to remove insoluble material. Protein concentrations were measured using a bicinchoninic acid (BCA) protein assay. For each replicate, 10 mg of protein per SILAC state was used for acetylation profiling. Proteins were reduced with 5 mM dithiothreitol for 45 min at RT. After reduction, proteins were alkylated using 10 mM iodoacetamide for 30 min at RT in the dark. Samples were diluted to 2 M urea and digested overnight with sequencing grade trypsin (Promega) using an enzyme to substrate ratio of 1:50 (w/w). TFA was used to quench digests. Peptide samples were desalted on tC18 SepPak SPE cartridges

(Waters) exactly as previously described[35]. Peptides were fractionated by basic pH reversed-phase (bRP) chromatography exactly as previously described[35]. Briefly, a Zorbax 300 Extend-C18 column (9.4 × 250 mm, 300 Å, 5 µm; Agilent) was used for the separation. Peptides were reconstituted in 5 mM ammonium formate (pH 10.0)/2% (vol/vol) acetonitrile (bRP Buffer A). Using the exact method parameters previously described,[35] a total of 96 2 ml fractions were collected across the bRP separation. For acetylated (Kac) peptide analysis, each fraction was combined in a non-contiguous manner such that every eighth fraction was combined (final fraction 1 = 1,9,17,25,33,41,49,57,65; final fraction 2 = 2,10,18,26,34,42,50,58,66; …) to create 8 final fractions. Pooled fractions were dried using vacuum centrifugation. An acetyl lysine antibody (Immunchem) was used for enrichment of Kac peptides from fractionated samples. Dried samples were reconstituted in 1.5 ml of 50 mM MOPS (pH 7.2), 10 mM sodium phosphate and 50 mM NaCl (IP buffer). Peptides were incubated with 120 µg of anti-Kac antibody beads for 1 hr at 4 ºC with end-over-end rotation. Antibody beads were washed twice with 1.5 ml of ice-cold IP buffer followed by three washes with ice-cold PBS. Kac peptides were eluted from the antibody with 2 x 50 µl of 0.15% TFA. Enriched peptides were desalted using StageTips exactly as previously described[35]. Samples were analyzed by nanoflow-UPLC-HCD-MS/MS using an Easy-nLC 1000 system (Proxeon) coupled online to a Q Exactive mass spectrometer (Thermo Fisher Scientific). Samples were reconstituted in 9 µl of 3% MeCN/0.1% formic acid and 4 µl was injected for analysis. Samples were injected at a flow rate of 500 nl/min onto a PicoFrit column (360 µm (OD) × 75 µm (ID)), 10 µm ID tip, 50 cm length (New Objective) self-packed with 24 cm of ReproSil-Pur 120 Å, 1.9 µm C18-AQ beads and heated to 50 ºC using a column heater (Pheonix S&T). The gradient and flow rate settings used were as previously described[35]. The Q Exactive was operated by acquiring an MS1 scan (R=70,000) followed by MS/MS scans on the 12 most abundant ions. For MS acquisition, ion targets of 3×10^6 and 5×10^4 ions were used for MS1 and MS2 scans, respectively. A maximum ion time of 20 ms and 120 ms was used for MS1 and MS2 scans, respectively. The HCD collision energy was set to 25. The dynamic exclusion time was set to 20 s and the peptide match and isotope exclusion functions were enabled. The MaxQuant software package (version 1.3.0.5) was used for identification and quantification of MS data. For searching, the enzyme specificity was set to trypsin, the maximum number of missed cleavages was set to 2, the precursor mass tolerance was set to 20 ppm for the first search, and the tolerance was set to 6 ppm for the main search. Carbamidomethylation of cysteines was searched as a fixed modification and oxidation of methionines, N-terminal acetylation of

proteins, and acetylation of lysines were searched as variable modifications. The minimum peptide length was set to 6, and false discovery rate for peptide, protein, and site identification was set to 1%.

***In vitro* Deacetylation Assay.**

Peptides were purchased from Peptide 2.0 with a purity of >75% with an acetylated N-terminus and an amidated C-terminus and were resuspended in water. The concentrations of peptides containing an unmodified lysine were measured using the fluorescamine assay as previously described[36]. All peptide concentrations were within two-fold of the calculated concentration based on weight. The CSRP2BP peptide contains no amine or aromatic amino acids; and therefore, the concentration was calculated based on the weight provided by Peptide 2.0. Recombinant human HDAC8 was either purchased or purified from *E. coli* as previously described[37] and all other HDAC homologues were purchased from BPS Biosciences. HDAC assays were performed using an enzyme-coupled system to measure acetate production as previously described[36]. The reactions were measured under standard HDAC reaction conditions (137 mM NaCl, 2.7 mM KCl, 25 mM HEPES, pH = 7.8, 30 ºC). Reactions measuring deacetylation of acetylated peptides (0 – 1600 μM) were initiated by addition of recombinant Zn(II)-HDAC8 (0.5 – 2.0 μM). The reactions were quenched by the addition of acid, and the acetate product, as reflected by an increase in the NADH fluorescence, was measured at 4 time points (up to 50 min). Recombinant HDAC isozymes 1-9 (0.4 μM), prepared by BPS Biosciences from baculovirus expression, were mixed with acetylated peptides (100 μM) and the formation of acetate product was measured as a function of time. The initial velocities (v_0) were calculated from a linear fit of the time-dependent increase in NADH fluorescence. The kinetic parameters were determined from fitting either a line or the Michaelis-Menten equation (v_0/[HDAC8] = (k_{cat}[S]) / (K_M + [S])) to the dependence of the initial velocity on the peptide concentration. Inhibition of HDAC8-catalyzed deacetylation of ARID1A (100 μM, below K_M) by PCI-34051 was measured using the acetate assay in 2.7 mM KCl, 137 mM NaCl, 50 mM HEPES, pH = 7.8, 0.001% BSA, 0.001% Tween 20. The value of K_i was determined from a fit of $v_{obs} = v_o/(1 + [I]/K_i)$ to the data.

Cell Culture and Compound Treatment for Gene Expression Experiments.

MCF7 cells (ATCC, # HTB-22) were cultured in DMEM medium (Gibco, #11995) containing 10% fetal bovine serum (Sigma, F4135) and 1x Pen Strep Glutamine (Gibco, #10378). Cells were plated into 384-well tissue culture treated plates (Corning, #3707) using a Multidrop Combi (Thermo, # 5840300) at 2,000 cells per well. Cells were incubated for 24 hours at 37°C in a humidified incubator containing 5% CO_2 before treatment.

Prior to treatment, 10 mM DMSO stock solutions of compounds were diluted to multiple doses in DMSO and arrayed into a 384-well plate (Abgene,# AB-1056). These 1,000x stock solutions were first diluted (100-fold) in culture medium, and then the diluted compounds were transferred to the cell culture plates using CyBi-well vario 384-well tips (another 10-fold dilution). Ultimately, all compounds were diluted 1,000-fold to their desired serial concentrations with a final DMSO concentration of 0.1%. Treated cells were incubated for 24 hours prior to lysis. Cells were lysed by partial removal of the culture media (15 μl remaining) followed by the addition of TCL lysis buffer (Qiagen, #1031576) using a liquid handling system. Cell lysate plates were sealed using a plate sealer, kept at room temperature for 30 minutes, and then frozen at -80°C until L1000 gene expression profiling was performed. Detailed cell culture and treatment protocols for L1000 can be found at http://lincscloud.org.

Gene Expression Profiling.

In this study, we utilize L1000, a high-throughput, bead-based gene expression assay in which mRNA is extracted from cultured human cells treated with various chemical or genomic perturbagens (small molecules, gene knockdowns, or gene over-expression constructs) as previously described.[38] This mRNA is reverse-transcribed into first-strand cDNA. Gene specific probes containing barcodes and universal primer sites are annealed to the first strand cDNA. The probes are ligated to form a template for PCR. The template is PCR amplified with biotinylated-universal primers. The end products are biotinylated, fixed length, barcoded amplicons. The amplicons can then be mixed with Luminex beads that contain complementary barcodes to those encoded in each of the 1000 amplified landmark genes. These 1000 landmark genes were chosen as a reduced representation of the transcriptome and account for the majority of expression variation across many cellular contexts (Subramanian, *et al.*, manuscript in preparation). These beads are then stained with fluorescent streptavidin-phycoerythrin (SAPE) and detected in 384-well plate format on a Luminex FlexMap

flow cytometry-based scanner. The resulting readout is a measure of mean fluorescent intensity (MFI) for each landmark gene.

Connectivity Map Query Analysis.

The raw expression data are log2-scaled, quantile normalized, and z-scored, such that a differential expression value is achieved for each gene in each well. In the standard L1000 protocol, each well corresponds to a different perturbagen and these differential expression values are collapsed across replicate wells to yield a differential expression signature for each perturbagen. The signatures of different perturbagens can then be compared to identify those that result in similar or dissimilar transcriptional responses as previously described.[39,40,41] In particular, to understand the mechanism of action of PCI-34051, we sought to identify connections that persisted across multiple distinct tissue types reasoning such connections are most robust. To assess this, we did a meta-analysis of the CMap query results (i.e., lists of perturbations ordered by similarity to an input gene expression signature). The input signature consisted of the 100 most differentially expressed genes (50 up and 50 down) upon treatment with PCI-34051. We performed the same analysis with BRD3811 as well.

When queried across the L1000 database at lincscloud.org, this produced a ranked list of 476251 connections corresponding to 51385 unique perturbagens. To summarize, the query result is first grouped by cell line and perturbagen type (small molecule, gene knockdown, or overexpression). The connectivity scores are then normalized by dividing by the signed mean score of each group. The scores are converted to percentile ranks within each group. The perturbagens are then ranked according to the direction of connectivity. Positive connections are ranked highest, and negative connections ranked lowest. For each unique perturbagen, we considered the average percentile rank in the four cell lines for which the connection to the query was strongest.

Western Blotting. MCF7 cells were treated with compounds for 48 h, at which time, lysates were collected using RIPA buffer with added protease (Roche) and phosphatase (Roche) inhibitors. Electrophoresis was performed using NuPage 4-12% Bis-Tris gels (Invitrogen). Proteins were transferred to a nitrocellulose

membrane and probed using antibodies for p21 (Cell Signaling) and GAPDH (Cell Signalling). Chemiluminescence was induced by subsequent incubation with HRP-linked secondary antibodies (GE Healthcare UK Ltd.) and treatment of the membrane with the appropriate ECL solutions (Thermo Scientific). Visualization was accomplished using a ChemiDoc MP System (Bio-Rad), and the raw data files were converted to jpegs using ImageJ (NIH).

Additional References

29. Jaffe, J.D., Wang, Y., Chan, H. M., Zhang, J., Huether, R., Kryukov, G. V., Bhang, H. E., Taylor, J. E., Hu, M., Englund, N. P., Yan, F., Wang, Z., McDonald, E. R. 3rd, Wei, L., Ma. J., Easton, J., Yu, Z., deBeaumount, R., Gibaja, V., Venkatesan, K., Schlegel, R., Sellers, W. R., Keen, N., Liu, J., Caponigro, G., Barretina, J., Cooke, V. G., Mullighan, C. Carr, S. A., Downing, J. R., Garraway, L. A., Stegmeier, F. (2013) Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *et al. Nat. Genet. 45*, 1386-1391.

30. Zhang, Y.-L., Holson, E., Wagner, F. F., Fluorescent substrates for determining lysine modifying enzyme activity, 2013, Patent Application WO 2013067391.

31. Somoza, J. R., Skene, R. J., Katz, B. A., Mol, C., Ho, J. D., Jennings, A. J., Luong, C., Arvai, A., Buggy, J. J., Chi, E., Tang, J., Sang, B.-C., Verner, E., Wynands, R., Leahy, E. M., Dougan, D. R., Snell, G., Navre, M., Knuth, M. W., Swanson, R. V., McRee, D. E., Tari, L. W. (2004) Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases. *Structure 12*, 1325-1334.

32. Small-Molecule Drug Discovery Suite 2013-2: Glide, version 6.0, Schrödinger, LLC, New York, NY, 2013.

33. Friesner, R.A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., Mainz, D. T. (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem. 49*, 6177-6196.

34. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics 5*, 376-386.

35. Udeshi, N. D., Mertins, P., Svinkina, T., Carr, S. A. (2013) Large-scale identification of ubiquitination sites by mass spectrometry. *Nat. Protoc. 8*, 1950-1960.

36. Wolfson, N. A., Pitcairn, C. A., Sullivan, E. D., Joseph, C. G., Fierke, C. A., (2014) An enzyme-coupled assay measuring acetate production for profiling histone deacetylase specificity. *Anal. Biochem. 456*, 61-69.

37. Gantt, S. L., Gattis, S. G., Fierke, C.A . (2006) Catalytic activity and inhibition of human histone deacetylase 8 is dependent on the identity of the active site metal ion. *Biochemistry 45*, 6170-6178.

38. Peck, D., Crawford, E. D., Ross, K. N., Stegmaier, K., Golub, T. R., Lamb, J. (2006) A method for high-throughput gene expression signature analysis. *Genome Biol.* **7**, R61

39. Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I., C., Wrobel, M. J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., Golub, T. R. (2006) The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease *Science 313*, 1929–1935.

40. Lamb, J., (2007) The Connectivity Map: a new tool for biomedical research. *Nat. Rev. Cancer 7*, 54–60.

41. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., Mesirov, J. P. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Nat. Acad. Sci. 102*, 15545-15550.