Structure-based prediction of KDAC6 substrates validated by enzymatic assay reveals determinants of promiscuity and detects new potential substrates

Julia K. Varga[a]*, Kelsey Diffley[b]*, Katherine R. Welker Leng[b], Carol A. Fierke[b,c], Ora Schueler-Furman[a]

[a] Department of Microbiology and Molecular Genetics, Institute for Medical Research Israel-Canada (IMRIC), Hebrew University, Faculty of Medicine, POB 12272, Jerusalem, 91120 Israel

[b] Department of Chemistry, University of Michigan, 930 North University Avenue, Ann Arbor, MI 48109, United States

[c] Department of Biochemistry, Brandeis University, 415 South Street, Waltham, MA 02453, United States

* Equal contribution

Julia K. Varga: julia.varga@mail.huji.ac.il

Kelsey Diffley: kdiffley@umich.edu

Katherine R. Welker Leng: krleng@umich.edu

Carol A. Fierke: carolfierke@brandeis.edu

Corresponding author:

- Ora Schueler-Furman, ora.furman-schueler@mail.huji.ac.il, +972-2-675-7094

Abbreviations

AUC: area under the curve

DD2: deacetylase domain 2

HAT: histone acetyltransferase

HDAC: histone deacetylase

KAT: acetyl-lysine acetyltransferase

KDAC: acetyl-lysine deacetylase

PSSM: position-specific scoring matrix

PTM: post-translational modifications

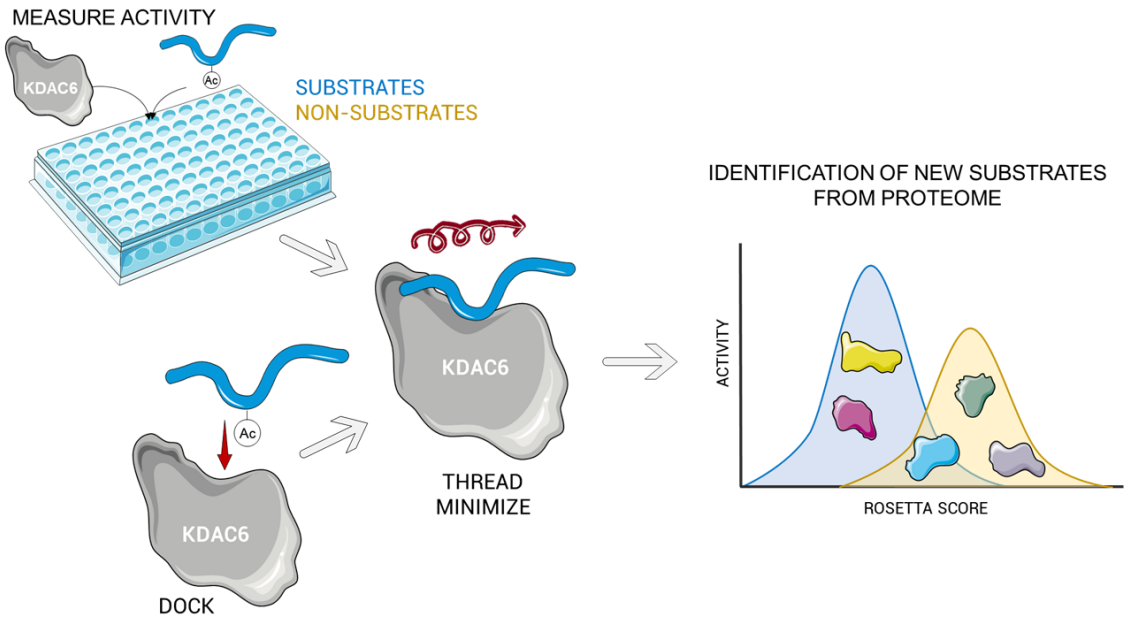
REU: Rosetta Energy Units

SLIM: short linear motif

Abstract

Histone deacetylases play important biological roles well beyond the deacetylation of histone tails, and therefore have recently been renamed to acetyl-lysine deacetylases (KDACs). In particular, KDAC6 is involved in multiple cellular processes such as apoptosis, cytoskeleton reorganization, and protein folding, affecting substrates such as □-tubulin, Hsp90 and cortactin proteins. We have applied a biochemical enzymatic assay to measure the activity of KDAC6 on a set of candidate unlabeled peptides. These served for the calibration of a structure-based substrate prediction protocol, Rosetta FlexPepBind, previously used for the successful substrate prediction of KDAC8 and other enzymes. The calibration process and comparison of the results between KDAC6 and KDAC8 highlighted structural differences that explain the already reported promiscuity of KDAC6. A proteome-wide screen of reported acetylation sites using our calibrated protocol together with the enzymatic assay provide new peptide substrates and avenues to novel potential functional regulatory roles of this promiscuous, multi-faceted enzyme.

Graphical abstract



Keywords

KDAC6, Rosetta FlexPepBind, substrate selectivity, acetate detection assay, deacetylation

1. Introduction

In order for cells and organisms to survive and adapt to different conditions, complex, tightly-controlled, and context-dependent regulation is crucial. Much of this regulation is achieved by post-translational modifications (PTM) that can change the behavior of a protein (as well as corresponding modifications of other macromolecules such as DNA and RNA). One of the major regulatory modifications is the acetylation and deacetylation of histones, which is a main component of the histone code dictating chromatin organization and transcriptional activity. Indeed, lysine acetylases (KATs) and deacetylases (KDACs) were originally termed histone acetylases and deacetylases (HATs and HDACs, respectively), when their activity on histones was originally discovered [1,2]. KDACs catalyze the removal of an acetyl group from the post-translational modification of acetyl-lysine in proteins. They are divided into two major groups, the zinc-dependent KDACs and the NAD^+-dependent KDACs also known as sirtuins. Zinc-dependent KDACs are further divided into subclasses based on their homology to yeast KDACs [3].

Lysine deacetylase 6 (KDAC6) is a class IIB deacetylase and is the only KDAC to contain two deacetylase domains of distinct specificities. The first domain specifically deacetylates acetylated C-terminal lysine residues, while the second shows a particularly broad substrate selectivity [4,5]. In addition, KDAC6 possesses a C-terminal ZnF-UBP ubiquitin binding domain, as well as a nuclear export signal and SE14 motif that anchors the protein to the cytoplasm [6,7]. There is evidence that KDAC6 catalyzes deacetylation of several cellular proteins involved in a variety of cellular processes. KDAC6 catalyzes deacetylation of α-tubulin and thereby regulates microtubule stability and cell motility [8]. Another characterized substrate, cortactin, binds to deacetylated actin filaments and participates in the fusion of lysosomes and autophagosomes [9]. The enzyme also plays a role in protein folding by regulating the activity of the Hsp90 chaperone protein via deacetylation [10,11]. In addition, KDAC6 is an important player in innate immunity, regulating the detection of pathogen

genomic material via deacetylation of retinoic acid inducible gene-I protein [12–14]. Furthermore, the ubiquitin binding domain of KDAC6 contributes to the formation of aggresomes by recognizing both misfolded proteins by binding to their ubiquitin tag and dynein motor proteins by binding to their dynein motor binding region to bring them together [15]. While the broad specificity of KDAC6 has been reported, a full understanding of the selectivity determinants is still lacking, as is a proper understanding of the underlying structural basis that makes this particular KDAC more promiscuous than others, such as KDAC8 [16].

Many of the enzymes that add or remove PTMs act on short linear motifs (SLIMs) that are often exposed. Therefore, their substrate selectivity may be approximated by short peptides that cover the region [17]. There are several *in vitro* methods that allow activity determination of a large variety of peptide substrate sequences including peptide arrays [18–23] and phage display experiments [24,25]. Peptide arrays have been used to probe KDAC substrate selectivity, and KDAC6 was found to be the most promiscuous amongst the tested KDAC enzymes [4,18,23].

Different types of prediction methods have been developed for finding putative substrates. Many sequence-based predictions find modification sites based on position-specific scoring matrices (PSSMs) or regular expressions [26] that are derived from a large set of substrates, often selected by high-throughput approaches as those described above. However, these do not account for possible interdependencies between amino acids at different positions in the substrate, nor do they consider secondary structure that might be important for recognition. Machine Learning-based approaches can be used for these aims (e.g. Hidden Markov Models [27] and naive Bayes [28]), but such approaches depend on considerable amounts of data. In fact, enzyme substrate patterns may not always adequately be depicted by a sequence-based description.

Structure-based methods can complement sequence-based methods, particularly in cases of non-canonical motifs, as we have previously shown for PTM enzymes using the Rosetta FlexPepBind protocol [29]. This approach assumes that the ability of the substrate local peptide sequence to bind in a catalysis-competent conformation is a main determinant of enzyme selectivity, and thus the binding energy of such enzyme-substrate complex structures can be taken as a proxy for substrate activity. The accuracy of the calibrated protocol can be estimated by applying it to an independent test set. New substrates can then be identified by applying the calibrated protocol to candidate peptides with unknown activity.

In this study we utilized an accurate biochemical assay that measures acetate production following deacetylation [30] to quantify the catalytic activity of KDAC6 for specific peptides and establish a gold standard set of peptide substrates. Based on these activities, we calibrated FlexPepBind to evaluate activity of potential substrates, as we have done in the past for other KDACs and PTM enzymes (e.g. KDAC8 and FTase [16,31]). Calibration revealed important structural differences between KDAC6 and KDAC8 that form the basis of the considerable difference in selectivity of these two deacetylases. Application of this method to screen the acetylome identified novel potential regulatory mechanisms based on KDAC6-dependent regulation. In the end, the combination of our structure-based approach that is based on accurate, *in vitro* biochemical measures of substrate activity, with the previously reported large-scale approaches leads to a better understanding of KDAC6 substrate selectivity.

2. Results

Accurate measure of KDAC6 substrate activity using a biochemical enzymatic assay

We used an enzyme-coupled acetate detection assay, or simply the 'acetate assay' (see Methods), developed and applied previously [16,30], to measure the catalytic efficiency (k_{cat}/K_M) of KDAC6. We measured catalysis of deacetylation of acetylated peptides with sequences taken from known substrate proteins, as well as a set of selected peptides with reported acetylation sites that were available to us from previous studies (**Table 1**). We used the second deacetylase domain (DD2) of KDAC6 for these experiments (from *Danio rerio,* which is more stable and has been shown to be a valid substitute for human KDAC6 [5]). To ensure an accurate determination of k_{cat}/K_M we measured KDAC6-catalyzed deacetylation at a minimum of four peptide concentrations, with at least two concentrations below K_M (**Figure 1**, squares). A total of 26 peptides which met these criteria were used as training set (**Table 1A).** Additionally, we included 16 peptides (set D-CAPPED, see below **Table 1B**) where the value of k_{cat} was measured accurately but the K_M value was lower than the limit of detection for the acetate assay (~10-20 μM substrate, circles in **Figure 1**) allowing determination of only a lower limit for the value of k_{cat}/K_M. The measured k_{cat}/K_M values span the range of three orders of magnitude. We defined a cutoff of $k_{cat}/K_M = 10^4\ M^{-1}s^{-1}$ to distinguish non-substrates from the rest. There were additional peptides tested that were not used in further analysis (longer than 6 amino acids or unreliably measured peptides, see **Table S1)**.

Structure-based computational prediction can identify most substrates of KDAC6

The calibration of FlexPepBind was performed using the k_{cat}/K_M values for KDAC6-catalyzed deacetylation of hexamer peptides (**Table 1A**, hereinafter referred to as D-TRAINING), using

an approach similar to the one previously applied to KDAC8 and FTase enzymes [16,31]. Here we provide a general outline of the calibration process. For details we refer to the Methods section. **Table 2** summarizes the different protocols evaluated in the calibration step (all results obtained with these protocols are provided in **Table S2**).

First, the KDAC6 substrate with the highest k_{cat}/K_M value (EGK$_{Ac}$FVR, derived from prelamin A) was docked using Rosetta FlexPepDock [32] into the binding pocket of a solved DD2 KDAC6 structure. Peptides from the D-TRAINING set were then threaded onto this template set and minimized (see Methods). The top-scoring model (according to reweighted score) was used as an estimate for their binding ability (Protocols BASIC_I and BASIC_R in **Table 2**). The performance of the protocol was evaluated based on the calculated binary distinction (area under the curve (AUC) values) and correlation between experimental values and Rosetta scores (Spearman correlation).

The use of this standard protocol resulted in low AUC values and no correlation between the experimental values of our dataset and the Rosetta scores (reweighted score; scoring according to interface scores provided worse results for all protocols tested, **Table S2**). We note that by using a similar protocol we were able to achieve excellent distinction in our previous study of KDAC8 substrate selectivity [16]. Examination of possible differences between the structures of KDAC6 and KDAC8 highlighted a significantly longer loop in KDAC6 (455-467) near the binding site that might rearrange for substrate binding (**Figure 2**). We therefore first ran the same protocol with receptor backbone minimization in the refinement step for template generation (corresponding protocols RMIN_I and RMIN_R). This allowed the loop to move away from the binding site, opening it up. However, this resulted only in a slight improvement of correlation to experimental values, while the binary distinction did not change much.

We proposed that additional receptor loop flexibility might improve the performance of our protocol. To test this, we modelled this loop (455-467) using the Rosetta kinematic loop

closure with fragments protocol [33] (see Methods). We picked the top-scoring model and repeated the above protocol (involving docking the best substrate peptide into the binding pocket and threading different candidate substrates onto top-scoring models selected by interface score and reweighted score (protocols RLOOP_I and RLOOP_R, respectively, see **Table 2A**). By using starting structures from this setup, we were able to obtain good performance for the D-TRAINING set (**Table 2B** and **Figure 3).** When selecting among a set of loop structures obtained from loop modeling (protocols RLOOPS_I and RLOOPS_R), the results did not improve further.

We first tested our protocols on dataset D-CAPPED, since our measurements indicated that these peptides are substrates, although accurate k_{cat}/K_M values could not be determined due to low K_M values (<50 μM) (**Table 1B**). We found that even though the RLOOP_I protocol performed best for D-TRAINING in terms of AUC and correlation values, it predicted that most of these peptides were non-substrates (**Figure 3A**, depicted with triangles). Therefore, to reduce the number of false negatives, we moved forward with protocol RLOOP_R, which showed similar performance on the D-TRAINING set (**Figure 3B,E** and **Table 2B**). To define substrates, we defined two Rosetta score cutoffs: a Rosetta score of -890 gave the best value by allowing maximum one false positive, while the stricter Rosetta score cutoff of -900 safely eliminated all false positive peptides in the D-TRAINING set.

Predictions on the human acetylome

To detect new potential KDAC6 substrates, we used our calibrated protocol (RLOOP_R) to screen the human acetylome (from PhosphoSitePlus [34], restricted to peptides annotated from low-throughput experiments) (**Figure 4A**). This screen detected only three peptides that scored better than the top-scoring peptide in the D-TRAINING set, and 242 and 600 peptides were classified as substrates by our strict and non-strict cutoffs (belonging to 141 and 257 proteins, respectively). In comparison to our previous study on KDAC8 specificity

[16], many more substrates were suggested for KDAC6 than for KDAC8, which agrees with the reported greater promiscuity observed for KDAC6.

Our screen of the known acetylome was able to recapitulate several previously identified substrates, including K118 of ATP-dependent RNA helicase DDX3X [35]score: -910,), K274 of the microtubule-associated protein tau [36](score: -901), K104 of K-Ras [37](score: -901), and K49 of β-catenin [38](score: -907). Moreover, we identified 30 proteins that have been previously reported to interact with KDAC6 (based on BioGRID [39]). For these, the interaction could be regulated further by the enzymatic removal of acetylation. In addition to these known substrates and protein interactors, we report several novel hits. Almost all of the acetylated sites in histone acetyltransferase p300 (EP300) are predicted to be putative substrates. Although the deacetylation of EP300 was previously linked to several other KDACs [40,41], these data suggest that KDAC6 could also be a potential regulator of this protein. This reveals an intriguing possibility of mutual crosstalk via PTMs, since EP300 has been shown to acetylate KDAC6 and by this to modify its ability to deacetylate tubulin [42]. A top predicted substrate is Nuclear factor NF-κB p105 subunit (NFKB1) at K441 (score: -911); acetylation of this site enhances the DNA binding of anti□inflammatory transcription factor p50 [43].

Reactome Pathway analysis [44] of the proteins belonging to the top 100 best scoring peptides identified several pathways which have been reported to be linked to KDAC6, such as aggrephagy [45], transcriptional regulation by TP53 [46], NOTCH1 signaling [47,48], EGFR signaling [49,50], MAPK signaling [51] and cytosolic sensors of pathogen-associated DNA [12,52]. However, none of these pathways were significantly enriched compared to their general association in the acetylated proteome.

We also compared the sequence logo created from these peptides to the one created from the training dataset (D-TRAINING), to ensure that we did not simply recreate the sequence specificity of that dataset (**Figure 4B-D**). Although there are some similarities, such as the

preference for glycine at position P_{-1} and for valine, tyrosine and methionine at P_{+2}, the sequence logo is different from the original database and shows more variability.

Finally, we also evaluated the pathway involvement of the proteins associated with the peptide substrates selected for D-TRAINING. Interestingly, most of the proteins harboring the peptides for which KDAC6 displayed the greatest deacetylase activity (>40,000 $M^{-1}s^{-1}$) either play a structural role or are associated with cellular structural elements (LMNA [53], MYO1G [54], ACTN1 [55], TUBA1A [8,56,57], have chaperone functions (GRP94 [58], HSP90 [10,11]), or are involved in the Rho signaling pathway (ARHGEF1 [59]).

Validating acetylome predictions

From the predicted substrates of the acetylome, we selected 10 peptides for additional validation. These peptides were measured using the same acetate assay as used for the D-TRAINING dataset (**Table 3**). Out of the 10 peptides, 9 were indeed measured to be substrates, albeit with poor correlation ($R = -0.41, p = 0.28$ for the correctly identified substrates, **Figure 3B;** with one outlier, $PCK_{Ac}EVD$ of NFAT5, see below**)**, highlighting both the predictive power of our model, as well as its restriction in ranking substrates for this promiscuous enzyme.

Novel potential regulatory functions of KDAC6

The two peptides with the greatest KDAC6-catalyzed deacetylation values (40,000 $M^{-1}s^{-1}$), derived from the proteins from EGFR and TARDBP, have been previously explored as direct substrates of KDAC6. Deacetylation of TARDBP, shown to be catalyzed by KDAC6, prevents TARDBP aggregation [60]. Acetylated TARDBP aggregates are found in patients with amyotrophic lateral sclerosis (ALS) [60]. Another peptide with high activity, IFI16 (25,000 $M^{-1}s^{-1}$), further connects KDAC6 to regulation of the immune response as IFI16

triggers host response to viral DNA, though the functional role of IFI16 acetylation is yet unknown [61]. In contrast, $PCK_{Ac}EVD$, a peptide of NFAT5 which was shown to co-localize with KDAC6 [62], turned out not to be a substrate by our criteria ($k_{cat}/K_M<10^4$ $M^{-1}s^{-1}$). In all of the datasets measured using the acetate assay, this peptide is one of two containing a cysteine, raising the possibility of the peptides dimerizing by forming disulfide-bridges that could affect its activity.

KDAC6 is known to regulate EGFR endocytic trafficking between the apical and basolateral membranes by regulating deacetylation of -tubulin and affecting its turnover rate [50,63]. Acetylation of EGFR has been shown to enhance its activity, an effect observed through treatment with KDAC inhibitors [64]. Two high-scoring peptides harboring residue K1179 and K1061 (scores: -914 and -908, respectively) are derived EGFR. We hypothesize that KDAC6 could play a role in the regulation of the protein level and activity not just by catalyzing deacetylation of EGFR interactors, but also by deacetylating EGFR itself. Some of the peptides that were efficient substrates are from proteins of the MAPK/ERK pathway which is activated by EGFR. KDAC6 deacetylates K-RAS at position K104, attenuating its transforming activity [37], and MAPK3/ERK1 at K72, impairing its kinase activity [65]. With our protocol, we were able to predict K104 of K-RAS, but not K72 of MAPK3/ERK1, as substrates for KDAC6 (scores: -901 and -860, respectively).

When looking at all the enzymatic data, our protocol reinforces KDAC6 involvement in cellular structure, MAPK signaling, aggrephagy, and the immune response. The identification of these additional potential substrates illustrates the need for further exploration of KDAC6 function within these pathways.

3. Discussion

In this study we have calibrated a structure-based protocol to characterize KDAC6 substrates, and applied this protocol to identify new potential deacetylation substrates. We

trained the method on a set of selected peptides for which we measured catalytic activity, validated the model on a set of independently measured peptides and applied it to the human acetylome. In the following we will summarize potential roles for the newly predicted substrates, compare the selectivity determinants of KDAC6 and KDAC8, and point to challenges in the accurate and comprehensive characterization of a promiscuous enzyme such as KDAC6.

Impact of predicted substrate deacetylation. Our substrate prediction model suggests novel aspects of KDAC6 deacetylase function. Deacetylation can have profound effects on protein function. Acetylation and ubiquitination reportedly compete for lysine residues: Acetylation can prevent lysine ubiquitination [66] or ubiquitin chain elongation [67] and consequently protect against proteasomal degradation. Therefore, we inspected the acetylated lysines suggested to be substrates by our screen, as well as their flanking regions, for additional reported post-translational modifications. Indeed, we found that among the 247 candidates that passed the strict score cutoff of -900, 88 (36%) undergo ubiquitination (data from PhosphoSitePlus [34]). For example, K1179 of EGFR, K451 of IFI16 or K441 of NFKB1 are both acetylated and ubiquitinated at the same position which likely affects the turnover of these proteins. Further exploration of the role of KDAC6, deacetylation, and ubiquitination of these proteins could illuminate important regulation of the involved biological pathways.

Differences between KDAC6 and KDAC8. KDAC6 has a much stronger catalytic efficiency and increased promiscuity towards peptide substrates *in vitro* compared to that of KDAC8 reported previously. For example, the fastest measured peptide substrate of KDAC8, ZNF318, has a k_{cat}/K_M value of 4,800 $M^{-1}s^{-1}$ [16] (a catalytic efficiency over 40-fold less than its fastest substrate, and therefore considered a non-substrate for KDAC6). We found that it was necessary to obtain accurate measurements of catalytic efficiency to develop our protocol, as the low K_M values of most of the peptides easily allow the k_{cat}/K_M value to be underestimated using measurements with a single substrate concentration.

Even with the assurance of accurate activity measurements, predicting the selectivity of KDAC6 proved to be much more challenging than its paralog KDAC8 for which we were able to obtain good predictions without introducing any backbone receptor flexibility [16]. We explored structural differences between these proteins that could explain this finding. Comparison of the KDAC6 (PDB ID: 5EFN [5]) and KDAC8 (PDB ID: 2V5W [68] structures (**Figure 2**) highlight two main differences in loops involved in forming the binding pocket that could lead to differences in binding selectivity. First, the residue that forms a hydrogen bond with acetylated lysine to position the substrate, D101 in KDAC8 and S531 in KDAC6, contacts the backbone at a similar position, but stems from a very different loop backbone. This loop participates in the formation of the pocket accommodating the residue preceding the acetylated lysine (P_{-1}). Indeed, in KDAC6 this pocket is considerably smaller, explaining the significant enrichment for glycine in the peptide libraries. This loop in KDAC8 also harbors Y100, a residue whose hydroxyl group forms a hydrogen bond with the peptide backbone in KDAC8, providing an additional recognition feature. Y100 also interacts with K19 from a neighboring loop, which holds the two loops together and the tyrosine in place. No residue corresponding to Y100 is found in KDAC6. Second, the loop located near the pocket that accommodates the residues trailing to the acetylated lysine (P_{+1-3}) of KDAC8 is significantly longer (12 residues in KDAC6 compared to 7 in KDAC8) and more hydrophobic in KDAC6. The larger size of the loop, together with its more non-polar character, suggest that the KDAC6 loop may be more flexible, allowing for adaptations of the binding groove, resulting in a more promiscuous binding pattern. Indeed, we showed here that the loop needs to move in order for our protocol to succeed, suggesting that this pocket is restructured for binding. Sampling of yet more loop structures, prior or together with the peptide structure optimization, can pave the way for application of this approach to more flexible receptors in the future. Remodeling of the loop while simultaneously docking a peptide to the binding pocket requires the calibration of a peptide docking protocol that can efficiently sample low energy peptide-loop conformation combinations.

Comparison to previous substrate studies. The family of KDACs has been extensively studied due to its importance in biological regulation. Beyond the few known substrates of KDAC6 mentioned above, substrate selectivity of human KDAC6 has been assessed at large scale in three key experimental studies (**Table S3**). Riester *et al.* used arrays of trimer peptides conjugated to 7-amino-4-methylcoumarin (AMC) (dataset D-3MER) [23]. The activity of the enzyme was detected by change in fluorescence. Schölz *et al.* applied specific inhibitors against several KDACs in cell lines and quantified the change in acetylated lysine sites by SILAC-MS (D-SILAC) [69]. In both studies, the experiments were run with at least 5 different KDACs, and both reached the conclusion that KDAC6 was, by far, the most promiscuous among the examined KDACs. Finally, Kutil *et al.* assessed KDAC6 deacetylation of 13-mer peptides synthesized on an array, measuring activity with a mixture of anti-acetyl lysine antibodies (dataset D-13MER) [4]. The overlap between substrates identified with these different experimental assays was minimal at the protein level [4]. To further probe these differences, we compared the substrates of these high-throughput methods to the substrates predicted using our protocol.

First, we computed two different types of sequence logos for the respective substrate sets, showing enrichment of certain residues compared to a proteome-level background and by using the non-substrate sets as background, respectively. Although there were similarities between the sequence logos of different datasets (e.g. the preference of glycine at the P_{-1} position, which our protocol also captured), there were significant differences as well (**Figure 5A-F**). A plot of peptides shared between D-SILAC and D-13MER also showed that there was a low degree of agreement between substrates sets from these two studies (**Figure 5G and Table S3**). Second, we used the computed PSSMs (**Figure S1**) to cross-score the datasets (using different peptide lengths) and calculated the correlations between the experimental values and PSSM scores (**Figure S2**). The highest correlation ($R = -0.47$) was found between the D-3MER dataset scored with the PSSM derived from itself and the experimental values of the D-TRAINING dataset scored by the PSSM of the D-13MER

dataset ($R = -0.40$). The low correlation between these calculated and measured values suggests that different experimental setups capture different features of this promiscuous enzyme's sequence selectivity. However, PSSMs treat every position in the peptide independently and without incorporating additional information, e.g. secondary structure and disorder, and may not be able to fully encompass substrate selectivity determinants, as also suggested by the poor correlations obtained even for predictions of substrate activity based on a PSSM derived from the same substrate set.

Disagreement between the multiple studies could be due to experimental differences. One potential difference is using only the DD2 domain of KDAC6 instead of the full-length protein. However, the studies discussed here (D-3MER [23], D-SILAC [4] and D-13MER [69] datasets) all used the full-length enzyme and still reported little overlap (5 common substrates between D-13MER and D-SILAC) among the identified substrates. Moreover, the varied size of the peptides and their flanking regions in the different experiments could also explain some of the variation. Among the peptides measured both in the D-13MER and D-SILAC datasets, there were 171 and 89 respectively whose core hexamer were the same, but their flanking regions differed (**Figure S3**). Closer inspection suggests that most of these variations could be explained by experimental measurement variations (outliers among the repeated experiments in case of D-13MER and protein inference differences for the D-SILAC dataset).

In vivo experiments will evaluate KDAC6 substrate selectivity in its physiological context, including potential cofactors, binding proteins and post-translational modifications of both the substrates and KDAC6. For example, KDAC6 displays different affinities toward tubulin dimers compared to assembled microtubules [70]. In *in vitro* experiments, peptides outside of their native environment might not fold into their native secondary structure that may be important for recognition [71]. Moreover, *in vivo* experiments can also detect downstream effects of treatments, therefore introducing bias into the results. Additionally, as with all high throughput experiments, they have a higher chance of amplifying noise. Furthermore,

antibodies widely used to detect or enrich for acetylation might not be sensitive or specific enough or have a bias for certain amino acids around the modification site [72], and mass spectrometry is biased to capture peptides where the positive charge is not removed by post-translational modifications like acetylation [73].

Conclusion. The successful construction of a structure-based model for a second KDAC isozyme indicates the utility and limitations of this method for predicting novel substrates. Structure-based modelling is dependent on substrate selectivity being predicted at the local sequence level, which was a challenge for the promiscuous KDAC6 in comparison to the selective KDAC8. The promiscuity of KDAC6 is particularly apparent in the limited agreement amongst the different large-scale substrate selectivity datasets. Our prediction model illuminated possible structural determinants of the broad selectivity of KDAC6 in comparison to KDAC8. Importantly, the present study provides for the first time accurately measured enzymatic activities on a large set of unlabeled peptides. The prediction model developed here allowed for the discovery of a potential set of novel substrates and novel avenues of substrate exploration for KDAC6.

4. Materials and Methods

Measuring enzyme activity

Reagents

High flow amylose resin was purchased from New England Biolabs and Ni-NTA agarose was purchased from Qiagen. Adenosine triphosphate (ATP), coenzyme A (CoA), NAD^+, NADH, L-malic acid, malate dehydrogenase (MDH), citrate synthase (CS), and mouse monoclonal anti-polyhistidine-alkaline phosphatase antibody were purchased from Sigma. N-terminally acetylated and C-terminally carboxylated, monoacetylated peptides were

purchased from Peptide 2.0 or Synthetic Biomolecules. 3% (v/v) acetic acid standard was purchased from RICCA Chemical. All other materials were purchased from Fisher at >95% purity unless noted otherwise.

KDAC6 Expression and Purification

The plasmid and protocol for the expression and purification of *Danio rerio* KDAC6 catalytic domain 2 (DD2) was generously provided by David Christianson (University of Pennsylvania). The expression construct was prepared previously by the Christianson lab by cloning the residues 440-798 of the *Danio rerio* KDAC6 gene into a modified pET28a(+) vector in frame with a TEV-protease cleavable N-terminal 6xHis-maltose binding protein (MBP) tag [5]. KDAC6 was expressed and purified as described with several alterations for expression optimization [5]. BL21(DE3) *E. coli* cells (Novagen 69450-3) were transformed with plasmid according to the protocol and plated on LB media-agar supplemented with 50 µg/mL kanamycin. Plates were incubated overnight at 37°C (16-18 hours), and a single colony was added to a LB media starter culture supplemented with 50 µg/mL kanamycin and incubated with shaking at 37°C for 16-18 hours. This overnight starter culture was diluted (1:200) into 2x-YT media supplemented with 50 µg/mL kanamycin and incubated at 37°C with shaking until the cell density reached an OD600=1. The cultures were then cooled to 18°C for one hour and supplemented with 100 µM $ZnSO_4$ and 500 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) to induce expression. The cultures were grown for an additional 16-18 hours with shaking at 18°C and harvested by centrifugation at 6,000 x g for 15 min at 4°C. Cell pellets were stored at -80°C. 1-mL pre- and post-induction samples were taken and tested for KDAC6 expression by polyhistidine western blot and activity using the commercial Fluor de Lys assay (Enzo Life Sciences).

Cell pellets were resuspended in running buffer (50 mM HEPES, pH 7.5, 300 mM KCl, 10% (v/v) glycerol and 1 mM TCEP) supplemented with protease inhibitor tablets (Pierce) at 2 mL/g cell pellet. The cells were lysed by three passages through a chilled microfluidizer

(Microfluidics) and centrifuged for 1 h at 26,000 x g at 4°C. Using an AKTA Pure FPLC (GE) running at 2 mL/min, the cleared lysate was loaded onto a 10-mL packed Ni-NTA column equilibrated with running buffer. The column was washed with 10 column volumes (CVs) of running buffer and 10 CVs of running buffer containing 30 mM imidazole, and the protein was eluted with 5 CVs elution buffer containing 500 mM imidazole. 8 mL fractions were collected and analyzed by SDS-PAGE and western blot, and fractions containing His-tagged KDAC6 were combined and loaded onto a 30-mL amylose column equilibrated with running buffer at 1 mL/min. The column was washed with 2 CVs running buffer and the protein was eluted with 5 CVs of running buffer supplemented with 20 mM maltose. Fractions containing KDAC6 were combined with His6x-TEV S219V protease (0.5 mg TEV protease/L culture), previously purified in-house [74] using a commercially purchased plasmid (Addgene plasmid pRK739), and dialyzed in 20K molecular weight cut-off (MWCO) dialysis cassettes against 200-fold running buffer containing 20 mM imidazole overnight at 4°C. After dialysis, the sample was loaded onto a 10-mL Ni-NTA column pre-equilibrated with running buffer containing 50mM imidazole at 2 mL/min. The column was washed with 5 CVs of 50 mM imidazole running buffer to elute cleaved KDAC. Non-cleaved KDAC6 and His-tagged TEV-protease were eluted with 20 CVs of a 50-500 mM linear imidazole gradient. Fractions containing cleaved KDAC6 were combined, concentrated to <2 mL, and loaded onto a 26/60 Sephacryl S200 size exclusion chromatography (SEC) column (GE) equilibrated with SEC/storage buffer (50 mM HEPES, pH 7.5, 100 mM KCl, 5% glycerol, and 1 mM TCEP) at 0.5 mL/min. Eluted peaks were tested for deacetylase activity, and active fractions were concentrated, aliquoted, flash frozen with liquid nitrogen, and stored at -80°C.

ACS Expression and Purification

The pHD4-ACS-TEV-His6x expression vector was prepared previously by inserting the ACS gene from a chitin-tagged acetyl-CoA synthetase plasmid Acs/pTYB1, a generous gift from Andrew Gulick (Hauptman-Woodward Institute), into a pET vector containing a His6x tag to

increase expression [16,30]. The pHD4-ACS-TEV-His6x construct was expressed and purified as previously described [30].

Coupled Acetate Detection Assay

The coupled acetate detection assay or simply the 'acetate assay' was performed as previously described with a few modifications [30]. Briefly, lyophilized peptides were re-suspended in water when possible or with minimal quantities of acid, base, or organic solvent to improve solubility. Peptide concentration was determined by one or more of the following methods: measuring A280 using the extinction coefficients if the peptide contained a tryptophan or tyrosine, using the fluorescamine assay if the peptide contained a free lysine [75], performing the bicinchoninic (BCA) assay using bovine serum albumin (BSA) as a standard, and determining concentration of acetate produced by complete deacetylation by KDAC6.

Reactions containing 10-2000 μM monoacetylated peptides in 1X KDAC6 assay buffer (50mM HEPES, pH 8.0, 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$) were initiated with 0.1-1 μM KDAC6 at 30°C. Timepoints, 60μL, were quenched with 5μL of 10% hydrochloric acid and kept on ice until assay completion (no more than 90 minutes). Timepoints were flash frozen with liquid nitrogen and stored at -80°C until work-up.

Coupled solution (50 mM HEPES, pH 8, 400 μM ATP, 10 μM NAD^+, 30 μM CoA, 0.07 U/μL CS, 0.04 U/μL MDH, 50 μM ACS, 100 mM NaCl, 3 mM KCl, 50 mM $MgCl_2$, and 2.5 mM L-malic acid) was prepared the day of the work-up and incubated at room temperature away from light for at least 25 minutes. Timepoints were quickly thawed and neutralized with 15 μL of freshly prepared and filtered 6% sodium bicarbonate. Neutralized timepoints and each acetate or NADH standard, 60μL, were added to 10 μL coupled solution (or 1X assay buffer for NADH standards) in a black, flat-bottomed, half-area, non-binding, 96-well plate (Corning No. 3686). The resulting NADH fluorescence (E_x=340 nm, E_m=460 nm) of standards and

timepoints was read on a PolarStar fluorescence plate reader at 1-3-minute increments until the signal reached equilibrium. The slopes of the acetate and NADH standard curves were compared to verify the coupled solution's activity. When possible, a positive control reaction for enzyme activity was included. Using the acetate standard curve, the fluorescence of each timepoint was converted to μM product, and the slopes of the linear portion of the reaction (<10%) were plotted against substrate concentration. Using GraphPad Prism, the Michaelis-Menten equation (**Equation 1**) was fit to the resulting dependence of the initial velocity on substrate concentration to determine the kinetic parameters k_{cat}/K_M, k_{cat}, and K_M. Standard error was calculated using GraphPad Prism analysis. A total of 50 peptides were tested. Of those, 26 were used in the training set and another 16 were used for validation of the protocol. 8 of the peptides were longer than 6 residues or their activity could not be measured precisely, therefore we did not use them for further analysis (**Table S1**).

Equation 1. $$\frac{v_0}{[E]} = \frac{k_{cat}}{K_M} * \frac{[S]}{\left(\frac{[S]}{K_M}+1\right)}$$

Calibration of FlexPepBind

Running FlexPepBind requires the creation of a starting structure to generate a template (or a set of templates, as in the present study) for threading peptides. We used the structure of the *Danio rerio* KDAC6 catalytic domain 2 from PDB (Protein Data Bank [76]) (DD2; PDB ID 5EFN) which was crystallized in a complex with a coumarine-linked trimer peptide substrate. In every Rosetta protocol described, we used Rosetta v2020.28. The protocol implemented in this study is similar to the one used in our previous study on KDAC8 specificity [16], and in the following we mainly highlight the differences.

To enforce a catalysis-competent binding conformation, we defined constraints that characterize substrate binding as revealed by the solved structures of KDAC6 bound to ligands. Constraints were defined for Rosetta runs as with KDAC8 [16]. These include: (1) interactions coordinating the proper binding of the Zn^{2+} ion required for enzymatic activity, (2)

interactions between the acetylated lysine side chain and the binding pocket, (3) a dihedral angle constraint in the peptide between residues 3 and 4 (*i.e.,* adjacent to the acetylated lysine in the modeled hexamers) to enforce a cis-peptide bond (see **Table S4** for further details). All of the distances between interacting residues were measured on structure PDB ID 5EFN.

The best substrate peptide (sequence: *EGK$_{Ac}$FVR*) was built in the binding pocket using the corresponding atoms of trichostatin A. The rest of the peptide was added in extended conformation. FlexPepDock was run on this structure with the constraints added, generating *nstruct*=1000 decoys with different setups (**Table 2**). The scoring function of ref2015 was used. In contrast to the KDAC8 study, we selected not only the top-scoring structure, but rather the top 5 structures, according to the measures *I_sc* or *reweighted_sc,* depending on the protocol (Protocols x_I and x_R, respectively). Every peptide of the training dataset was threaded onto these starting structures using the Rosetta *fixbb* protocol and running FlexPepDock with minimization only (see **Table S4 and S5** for runline commands and parameters). For each peptide sequence, the best score (*reweighted_sc)* among the 5 templates was used to reflect its substrate strength (final scoring according to *I_sc* resulted in inferior predictions and was therefore not used).

For protocols RBASIC_x & RMIN_x, the receptor backbone was not moved, or only minimized, respectively. Since these protocols did not yield satisfactory predictions, for protocols RLOOP_x and RLOOPS_x, the template was optimized using loop modeling prior to the peptide docking step. Loop modeling was performed with kinematic closure with fragments (Stein and Kortemme 2013) on the loop of residues 455-467 (**Figure 2)**. Fragments were generated as described in the fragment picker application manual (details in **Table S4 and S5**) and 1000 decoys were generated with the loop modeling protocol. The top (top2) scoring decoys were selected for protocol RLOOP(S)_x to accommodate for different possible loop and peptide conformations.

Comparison of datasets

The PSSMs for the comparison of different datasets were generated with PSSMSearch [77], using PSI-BLAST as the scoring method. Substrates were defined according to the thresholds applied in the reported studies, except for the D-3MER dataset, where no such threshold was provided. The substrates for this dataset were taken from the first five out of a total of 15 bins. The differential sequence logos were generated with the TwoSampleLogo standalone program [78], using default values. Correlation and AUC calculations and visualizations were created in R (v3.5.1), using *corrplot* (version 0.84) and *pROC* (v1.16.2, [79]) packages, respectively.

For the calculation of the scores using the PSSM of the D-3MER dataset, we used only the leading two amino acids before the acetylated lysine in the substrate. We also scored the peptides by only using values from D-SILAC and D-13MER PSSMs in P_{-1} and P_{-2} positions. In summary, every 3-mer peptide has three different scores and every 13-mer peptide has 7 different scores (4 for the full-length peptide based on D-SILAC, and D-13MER with 13 and 6 scored amino acids; and 3 for assessing the core 3-mers).

Common peptides from the D-13MER and D-SILAC dataset were selected for comparison of the two approaches. For defining substrates and non-substrates, we used the thresholds of 0.8-1.2 for H/L ratio in the D-SILAC dataset, and maximum 1 replica in which the peptide was identified as substrate for the D-13MER dataset. For substrates, the enrichment threshold was 2 and above and 3 experiments in which a peptide was identified, for the D-SILAC and D-13MER sets, respectively. We note the significant difference in the number of non-substrates in the training and test sets, which preclude a rigorous comparison of performance on these sets.

Running the calibrated protocol on the acetylome

The dataset of the human acetylome was extracted from the PhosphoSitePlus database (downloaded on 23/09/2020 [34]). Hexamer peptides around sites on human proteins with at least one low-throughput experiment to support were derived with two leading and three trailing residues around the modification site. Only peptides spanning a full hexamer (*i.e*., the modification is not at the termini) were selected. BioGrid data [39] for KDAC6 was downloaded on 11/09/2020 from database version 4.1.190. For pathway analysis, we used Reactome with Pathway Browser version 3.7 and database release 74. DAVID [80] and GOrilla [81] was used to GO annotation enrichment analysis between identified substrate and non-substrate proteins.

Accession numbers

PDB ID: **5EFN**, **2V5W**; UniProt ID: **Q9UBN7**, **F8W4B7**

Acknowledgements

This work was supported, in whole or in part, by the Israel Science Foundation, founded by the Israel Academy of Science and Humanities (grant number 717/2017 to O.S.-F.), the US-Israel Binational Science Foundation 2015207 (to O.S.-F.), and the National Institutes of Health (grant R-R01-GM-040602 to C.A.F.). J.K.V. acknowledges support from CM-SMP-KA107/426559/2019 and CM-SMP-SH/311264/2018 of the Tempus Mundi Foundation. K.D. and K.W.L acknowledge support from the Rackham Graduate School.

We thank Dr. David W. Christianson for providing us the KDAC6 construct and Dr. Nicholas J. Porter for assistance with KDAC6 purification. We are grateful to Dr. Cyril Barinka and Dr. Zsófia Kutil for sharing with us the data from their paper.

Elements of graphical abstract were drawn using figures from Servier Medical Art (https://smart.servier.com/)

Declaration of Interests

The authors declare no competing interests.

References

[1] Inoue A, Fujimoto D. Enzymatic deacetylation of histone. Biochem Biophys Res Commun 1969;36:146–50. doi:10.1016/0006-291x(69)90661-5.

[2] Allfrey VG, Faulkner R, Mirsky AE. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci USA 1964;51:786–94. doi:10.1073/pnas.51.5.786.

[3] Parbin S, Kar S, Shilpi A, Sengupta D, Deb M, Rath SK, et al. Histone deacetylases: a saga of perturbed acetylation homeostasis in cancer. J Histochem Cytochem 2014;62:11–33. doi:10.1369/0022155413506582.

[4] Kutil Z, Skultetyova L, Rauh D, Meleshin M, Snajdr I, Novakova Z, et al. The unraveling of substrate specificity of histone deacetylase 6 domains using acetylome peptide microarrays and peptide libraries. FASEB J 2019;33:4035–45. doi:10.1096/fj.201801680R.

[5] Hai Y, Christianson DW. Histone deacetylase 6 structure and molecular basis of catalysis and inhibition. Nat Chem Biol 2016;12:741–7. doi:10.1038/nchembio.2134.

[6] Li Y, Shin D, Kwon SH. Histone deacetylase 6 plays a role as a distinct regulator of diverse cellular processes. FEBS J 2013;280:775–93. doi:10.1111/febs.12079.

[7] Zheng K, Jiang Y, He Z, Kitazato K, Wang Y. Cellular defence or viral assist: the dilemma of HDAC6. J Gen Virol 2017;98:322–37. doi:10.1099/jgv.0.000679.

[8] Hubbert C, Guardiola A, Shao R, Kawaguchi Y, Ito A, Nixon A, et al. HDAC6 is a microtubule-associated deacetylase. Nature 2002;417:455–8. doi:10.1038/417455a.

[9] Yan J. Interplay between HDAC6 and its interacting partners: essential roles in the aggresome-autophagy pathway and neurodegenerative diseases. DNA Cell Biol 2014;33:567–80. doi:10.1089/dna.2013.2300.

[10] Bali P, Pranpat M, Bradner J, Balasis M, Fiskus W, Guo F, et al. Inhibition of histone deacetylase 6 acetylates and disrupts the chaperone function of heat shock protein 90: a novel basis for antileukemia activity of histone deacetylase inhibitors. J Biol Chem 2005;280:26729–34. doi:10.1074/jbc.C500186200.

[11] Kovacs JJ, Murphy PJM, Gaillard S, Zhao X, Wu J-T, Nicchitta CV, et al. HDAC6 regulates Hsp90 acetylation and chaperone-dependent activation of glucocorticoid receptor. Mol Cell 2005;18:601–7. doi:10.1016/j.molcel.2005.04.021.

[12] Moreno-Gonzalo O, Ramírez-Huesca M, Blas-Rus N, Cibrián D, Saiz ML, Jorge I, et al. HDAC6 controls innate immune and autophagy responses to TLR-mediated signalling by the intracellular bacteria Listeria monocytogenes. PLoS Pathog 2017;13:e1006799. doi:10.1371/journal.ppat.1006799.

[13] Choi SJ, Lee H-C, Kim J-H, Park SY, Kim T-H, Lee W-K, et al. HDAC6 regulates cellular viral RNA sensing by deacetylation of RIG-I. EMBO J 2016;35:429–42. doi:10.15252/embj.201592586.

[14] Liu HM, Jiang F, Loo YM, Hsu S, Hsiang T-Y, Marcotrigiano J, et al. Regulation of Retinoic Acid Inducible Gene-I (RIG-I) Activation by the Histone Deacetylase 6. EBioMedicine 2016;9:195–206. doi:10.1016/j.ebiom.2016.06.015.

[15] Kawaguchi Y, Kovacs JJ, McLaurin A, Vance JM, Ito A, Yao TP. The deacetylase HDAC6 regulates aggresome formation and cell viability in response to misfolded protein stress. Cell 2003;115:727–38. doi:10.1016/s0092-8674(03)00939-5.

[16] Alam N, Zimmerman L, Wolfson NA, Joseph CG, Fierke CA, Schueler-Furman O. Structure-Based Identification of HDAC8 Non-histone Substrates. Structure 2016;24:458–68. doi:10.1016/j.str.2016.02.002.
[17] Gibson TJ, Dinkel H, Van Roey K, Diella F. Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. Cell Commun Signal 2015;13:42. doi:10.1186/s12964-015-0121-y.
[18] Gurard-Levin ZA, Kim J, Mrksich M. Combining mass spectrometry and peptide arrays to profile the specificities of histone deacetylases. Chembiochem 2009;10:2159–61. doi:10.1002/cbic.200900417.
[19] Meyer NO, O'Donoghue AJ, Schulze-Gahmen U, Ravalin M, Moss SM, Winter MB, et al. Multiplex substrate profiling by mass spectrometry for kinases as a method for revealing quantitative substrate motifs. Anal Chem 2017;89:4550–8. doi:10.1021/acs.analchem.6b05002.
[20] Ivry SL, Meyer NO, Winter MB, Bohn MF, Knudsen GM, O'Donoghue AJ, et al. Global substrate specificity profiling of post-translational modifying enzymes. Protein Sci 2018;27:584–94. doi:10.1002/pro.3352.
[21] Wigington CP, Roy J, Damle NP, Yadav VK, Blikstad C, Resch E, et al. Systematic discovery of short linear motifs decodes calcineurin phosphatase signaling. Mol Cell 2020;79:342-358.e12. doi:10.1016/j.molcel.2020.06.029.
[22] Moreno-Yruela C, Bæk M, Vrsanova A-E, Schulte C, Maric HM, Olsen CA. Hydroxamic acid-modified peptide microarrays for profiling isozyme-selective interactions and inhibition of histone deacetylases. Nat Commun 2021;12:62. doi:10.1038/s41467-020-20250-9.
[23] Riester D, Hildmann C, Grünewald S, Beckers T, Schwienhorst A. Factors affecting the substrate specificity of histone deacetylases. Biochem Biophys Res Commun 2007;357:439–45. doi:10.1016/j.bbrc.2007.03.158.
[24] Ratnikov B, Cieplak P, Smith JW. High throughput substrate phage display for protease profiling. Methods Mol Biol 2009;539:93–114. doi:10.1007/978-1-60327-003-8_6.
[25] Guo Z, Wang X, Li H, Gao Y. Screening E3 substrates using a live phage display library. PLoS ONE 2013;8:e76622. doi:10.1371/journal.pone.0076622.
[26] Kumar M, Gouw M, Michael S, Sámano-Sánchez H, Pancsa R, Glavina J, et al. ELM-the eukaryotic linear motif resource in 2020. Nucleic Acids Res 2020;48:D296–306. doi:10.1093/nar/gkz1030.
[27] Weatheritt RJ, Luck K, Petsalaki E, Davey NE, Gibson TJ. The identification of short linear motif-mediated interfaces within the human interactome. Bioinformatics 2012;28:976–82. doi:10.1093/bioinformatics/bts072.
[28] Tallorin L, Wang J, Kim WE, Sahu S, Kosa NM, Yang P, et al. Discovering de novo peptide substrates for enzymes using machine learning. Nat Commun 2018;9:5253. doi:10.1038/s41467-018-07717-6.
[29] Alam N, Schueler-Furman O. Modeling Peptide-Protein Structure and Binding Using Monte Carlo Sampling Approaches: Rosetta FlexPepDock and FlexPepBind. Methods Mol Biol 2017;1561:139–69. doi:10.1007/978-1-4939-6798-8_9.
[30] Wolfson NA, Pitcairn CA, Sullivan ED, Joseph CG, Fierke CA. An enzyme-coupled assay measuring acetate production for profiling histone deacetylase specificity. Anal Biochem 2014;456:61–9. doi:10.1016/j.ab.2014.03.012.

[31] London N, Lamphear CL, Hougland JL, Fierke CA, Schueler-Furman O. Identification of a novel class of farnesylation targets by structure-based modeling of binding specificity. PLoS Comput Biol 2011;7:e1002170. doi:10.1371/journal.pcbi.1002170.
[32] London N, Raveh B, Cohen E, Fathi G, Schueler-Furman O. Rosetta FlexPepDock web server--high resolution modeling of peptide-protein interactions. Nucleic Acids Res 2011;39:W249-53. doi:10.1093/nar/gkr431.
[33] Stein A, Kortemme T. Improvements to robotics-inspired conformational sampling in rosetta. PLoS ONE 2013;8:e63090. doi:10.1371/journal.pone.0063090.
[34] Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res 2015;43:D512-20. doi:10.1093/nar/gku1267.
[35] Saito M, Hess D, Eglinger J, Fritsch AW, Kreysing M, Weinert BT, et al. Acetylation of intrinsically disordered regions regulates phase separation. Nat Chem Biol 2019;15:51–61. doi:10.1038/s41589-018-0180-7.
[36] Perez M, Santa-Maria I, Gomez de Barreda E, Zhu X, Cuadros R, Cabrero JR, et al. Tau--an inhibitor of deacetylase HDAC6 function. J Neurochem 2009;109:1756–66. doi:10.1111/j.1471-4159.2009.06102.x.
[37] Yang MH, Laurent G, Bause AS, Spang R, German N, Haigis MC, et al. HDAC6 and SIRT2 regulate the acetylation state and oncogenic activity of mutant K-RAS. Mol Cancer Res 2013;11:1072–7. doi:10.1158/1541-7786.MCR-13-0040-T.
[38] Iaconelli J, Xuan L, Karmacharya R. HDAC6 Modulates Signaling Pathways Relevant to Synaptic Biology and Neuronal Differentiation in Human Stem-Cell-Derived Neurons. Int J Mol Sci 2019;20. doi:10.3390/ijms20071605.
[39] Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, et al. The BioGRID interaction database: 2019 update. Nucleic Acids Res 2019;47:D529–41. doi:10.1093/nar/gky1079.
[40] Stiehl DP, Fath DM, Liang D, Jiang Y, Sang N. Histone deacetylase inhibitors synergize p300 autoacetylation that regulates its transactivation activity and complex formation. Cancer Res 2007;67:2256–64. doi:10.1158/0008-5472.CAN-06-3985.
[41] Black JC, Mosley A, Kitada T, Washburn M, Carey M. The SIRT2 deacetylase regulates autoacetylation of p300. Mol Cell 2008;32:449–55. doi:10.1016/j.molcel.2008.09.018.
[42] Han Y, Jeong HM, Jin Y-H, Kim Y-J, Jeong HG, Yeo C-Y, et al. Acetylation of histone deacetylase 6 by p300 attenuates its deacetylase activity. Biochem Biophys Res Commun 2009;383:88–92. doi:10.1016/j.bbrc.2009.03.147.
[43] Cartwright T, Perkins ND, L Wilson C. NFKB1: a suppressor of inflammation, ageing and cancer. FEBS J 2016;283:1812–22. doi:10.1111/febs.13627.
[44] Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Res 2014;42:D472-7. doi:10.1093/nar/gkt1102.
[45] Lamark T, Johansen T. Aggrephagy: selective disposal of protein aggregates by macroautophagy. Int J Cell Biol 2012;2012:736905. doi:10.1155/2012/736905.
[46] Ryu H-W, Shin D-H, Lee DH, Choi J, Han G, Lee KY, et al. HDAC6 deacetylates p53 at lysines 381/382 and differentially coordinates p53-induced apoptosis. Cancer Lett 2017;391:162–71. doi:10.1016/j.canlet.2017.01.033.
[47] Deskin B, Yin Q, Zhuang Y, Saito S, Shan B, Lasky JA. Inhibition of HDAC6 Attenuates Tumor Growth of Non-Small Cell Lung Cancer. Transl Oncol 2020;13:135–45. doi:10.1016/j.tranon.2019.11.001.

[48] Deskin B, Lasky J, Zhuang Y, Shan B. Requirement of HDAC6 for activation of Notch1 by TGF-β1. Sci Rep 2016;6:31086. doi:10.1038/srep31086.
[49] Wang Z, Tang F, Hu P, Wang Y, Gong J, Sun S, et al. HDAC6 promotes cell proliferation and confers resistance to gefitinib in lung adenocarcinoma. Oncol Rep 2016;36:589–97. doi:10.3892/or.2016.4811.
[50] Liu W, Fan LX, Zhou X, Sweeney WE, Avner ED, Li X. HDAC6 regulates epidermal growth factor receptor (EGFR) endocytic trafficking and degradation in renal epithelial cells. PLoS ONE 2012;7:e49418. doi:10.1371/journal.pone.0049418.
[51] Haakenson J, Wu J-Y, Xiang S, Williams KA, Bai W, Zhang X. HDAC6-Dependent Functions in Tumor Cells: Crossroad with the MAPK Pathways. Crit Rev Oncog 2015;20:65–81. doi:10.1615/critrevoncog.2014012484.
[52] Moreno-Gonzalo O, Mayor F, Sánchez-Madrid F. HDAC6 at crossroads of infection and innate immunity. Trends Immunol 2018;39:591–5. doi:10.1016/j.it.2018.05.004.
[53] Karoutas A, Szymanski W, Rausch T, Guhathakurta S, Rog-Zielinska EA, Peyronnet R, et al. The NSL complex maintains nuclear architecture stability via lamin A/C acetylation. Nat Cell Biol 2019;21:1248–60. doi:10.1038/s41556-019-0397-z.
[54] Olety B, Wälte M, Honnert U, Schillers H, Bähler M. Myosin 1G (Myo1G) is a haematopoietic specific myosin that localises to the plasma membrane and regulates cell elasticity. FEBS Lett 2010;584:493–9. doi:10.1016/j.febslet.2009.11.096.
[55] Oikonomou KG, Zachou K, Dalekos GN. Alpha-actinin: a multidisciplinary protein with important role in B-cell driven autoimmunity. Autoimmun Rev 2011;10:389–96. doi:10.1016/j.autrev.2010.12.009.
[56] Matsuyama A, Shimazu T, Sumida Y, Saito A, Yoshimatsu Y, Seigneurin-Berny D, et al. In vivo destabilization of dynamic microtubules by HDAC6-mediated deacetylation. EMBO J 2002;21:6820–31. doi:10.1093/emboj/cdf682.
[57] Zhao Z, Xu H, Gong W. Histone deacetylase 6 (HDAC6) is an independent deacetylase for alpha-tubulin. Protein Pept Lett 2010;17:555–8.
[58] Marzec M, Eletto D, Argon Y. GRP94: An HSP90-like protein specialized for protein folding and quality control in the endoplasmic reticulum. Biochim Biophys Acta 2012;1823:774–87. doi:10.1016/j.bbamcr.2011.10.013.
[59] Hart MJ, Sharma S, elMasry N, Qiu RG, McCabe P, Polakis P, et al. Identification of a novel guanine nucleotide exchange factor for the Rho GTPase. J Biol Chem 1996;271:25452–8. doi:10.1074/jbc.271.41.25452.
[60] Cohen TJ, Hwang AW, Restrepo CR, Yuan C-X, Trojanowski JQ, Lee VMY. An acetylation switch controls TDP-43 function and aggregation propensity. Nat Commun 2015;6:5845. doi:10.1038/ncomms6845.
[61] Guise AJ, Budayeva HG, Diner BA, Cristea IM. Histone deacetylases in herpesvirus replication and virus-stimulated host defense. Viruses 2013;5:1607–32. doi:10.3390/v5071607.
[62] Herbelet S, De Vlieghere E, Gonçalves A, De Paepe B, Schmidt K, Nys E, et al. Localization and Expression of Nuclear Factor of Activated T-Cells 5 in Myoblasts Exposed to Pro-inflammatory Cytokines or Hyperosmolar Stress and in Biopsies from Myositis Patients. Front Physiol 2018;9:126. doi:10.3389/fphys.2018.00126.
[63] Wang Z, Hu P, Tang F, Xie C. HDAC6-mediated EGFR stabilization and activation restrict cell response to sorafenib in non-small cell lung cancer cells. Med Oncol 2016;33:50. doi:10.1007/s12032-016-0765-5.

[64] Song H, Li C-W, Labaff AM, Lim S-O, Li L-Y, Kan S-F, et al. Acetylation of EGF receptor contributes to tumor cell resistance to histone deacetylase inhibitors. Biochem Biophys Res Commun 2011;404:68–73. doi:10.1016/j.bbrc.2010.11.064.

[65] Wu J-Y, Xiang S, Zhang M, Fang B, Huang H, Kwon OK, et al. Histone deacetylase 6 (HDAC6) deacetylates extracellular signal-regulated kinase 1 (ERK1) and thereby stimulates ERK1 activity. J Biol Chem 2018;293:1976–93. doi:10.1074/jbc.M117.795955.

[66] Caron C, Boyault C, Khochbin S. Regulatory cross-talk between lysine acetylation and ubiquitination: role in the control of protein stability. Bioessays 2005;27:408–15. doi:10.1002/bies.20210.

[67] Ohtake F, Saeki Y, Sakamoto K, Ohtake K, Nishikawa H, Tsuchiya H, et al. Ubiquitin acetylation inhibits polyubiquitin chain elongation. EMBO Rep 2015;16:192–201. doi:10.15252/embr.201439152.

[68] Vannini A, Volpari C, Gallinari P, Jones P, Mattu M, Carfí A, et al. Substrate binding to histone deacetylases as shown by the crystal structure of the HDAC8-substrate complex. EMBO Rep 2007;8:879–84. doi:10.1038/sj.embor.7401047.

[69] Schölz C, Weinert BT, Wagner SA, Beli P, Miyake Y, Qi J, et al. Acetylation site specificities of lysine deacetylase inhibitors in human cells. Nat Biotechnol 2015;33:415–23. doi:10.1038/nbt.3130.

[70] Skultetyova L, Ustinova K, Kutil Z, Novakova Z, Pavlicek J, Mikesova J, et al. Human histone deacetylase 6 shows strong preference for tubulin dimers over assembled microtubules. Sci Rep 2017;7:11547. doi:10.1038/s41598-017-11739-3.

[71] Katz C, Levy-Beladev L, Rotem-Bamberger S, Rito T, Rüdiger SGD, Friedler A. Studying protein-protein interactions using peptide arrays. Chem Soc Rev 2011;40:2131–45. doi:10.1039/c0cs00029a.

[72] Kim S-Y, Sim CK, Zhang Q, Tang H, Brunmeir R, Pan H, et al. An Alternative Strategy for Pan-acetyl-lysine Antibody Generation. PLoS ONE 2016;11:e0162528. doi:10.1371/journal.pone.0162528.

[73] Virág D, Dalmadi-Kiss B, Vékey K, Drahos L, Klebovich I, Antal I, et al. Current Trends in the Analysis of Post-translational Modifications. Chromatographia 2020;83:1–10. doi:10.1007/s10337-019-03796-9.

[74] Tropea JE, Cherry S, Waugh DS. Expression and purification of soluble His(6)-tagged TEV protease. Methods Mol Biol 2009;498:297–307. doi:10.1007/978-1-59745-196-3_19.

[75] Huang X, Hernick M. A fluorescence-based assay for measuring N-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside deacetylase activity. Anal Biochem 2011;414:278–81. doi:10.1016/j.ab.2011.04.001.

[76] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic Acids Res 2000;28:235–42. doi:10.1093/nar/28.1.235.

[77] Krystkowiak I, Manguy J, Davey NE. PSSMSearch: a server for modeling, visualization, proteome-wide discovery and annotation of protein motif specificity determinants. Nucleic Acids Res 2018;46:W235–41. doi:10.1093/nar/gky426.

[78] Vacic V, Iakoucheva LM, Radivojac P. Two Sample Logo: a graphical representation of the differences between two sets of sequence alignments. Bioinformatics 2006;22:1536–7. doi:10.1093/bioinformatics/btl151.

[79] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77. doi:10.1186/1471-2105-12-77.

[80] Jiao X, Sherman BT, Huang DW, Stephens R, Baseler MW, Lane HC, et al. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics 2012;28:1805–6. doi:10.1093/bioinformatics/bts251.

[81] Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 2009;10:48. doi:10.1186/1471-2105-10-48.

[82] Sui L, Huang R, Yu H, Zhang S, Li Z. Inhibition of HDAC6 by tubastatin A disrupts mouse oocyte meiosis via regulating histone modifications and mRNA expression. J Cell Physiol 2020;235:7030–42. doi:10.1002/jcp.29599.

[83] Zhang M, Xiang S, Joo H-Y, Wang L, Williams KA, Liu W, et al. HDAC6 deacetylates and ubiquitinates MSH2 to maintain proper levels of MutSα. Mol Cell 2014;55:31–46. doi:10.1016/j.molcel.2014.04.028.

Figure 1. Dependence of deacetylation rate on substrate concentration for two representative peptides catalyzed by KDAC6, measured using the acetate assay. The initial velocity for each substrate concentration was determined from a linear regression of a time course consisting of a minimum of three timepoints, standard error is shown. The kinetic parameters are determined from a nonlinear least square fit of the Michaelis-Menten equation to the data and are listed in **Table 1**. Black squares: example of data which met the criteria to produce accurate k_{cat}/K_M values. Open circles: example of overfit data resulting in calculation of a lower limit for k_{cat}/K_M, due to the K_M being lower than the detection limit of the assay.

Figure 2. Key differences between KDAC6 and KDAC8 in the coordination of the substrate. The active sites of KDAC6 (PDB ID 5EFN, orange) and KDAC8 (PDB ID 2V5W, teal) are overlaid. Two loops that distinguish the two KDACs are highlighted, as are specific residues S531 (KDAC6) and D101 (KDAC8) that coordinate hydrogen bonding to the backbone of the acetylated substrate lysine residue, as well as positions W459 (KDAC6) and Y100 (KDAC8). The hydroxamic acid inhibitors are shown in sticks, and the catalytic Zn^{2+} ion is shown in grey sphere. See Text for more details.

Figure 3. Performance of different protocols on the D-TRAINING set. (A-C) Correlation plots between experimental (y-axis, logarithmic scale) and predicted substrate activities (x-axis). **(A)** Protocol RLOOP_I (score cutoff: -870), **(B)** Protocol RLOOP_R (score cutoff: -890, strict score cutoff: -900), **(C)** Protocol RLOOPS_R (score cutoff: -936). Dashed line: experimental score cutoff value for substrate definition; dotted dashed line: protocol score cutoff value; dotted line: protocol strict cutoff; circles: D-TRAINING dataset; triangles: D-CAPPED - substrates for which no exact experimental value could be measured (see Text); rhombuses: validation dataset. **(D-E)** AUC plots for the three protocols (RLOOP_I, RLOOP_R and RLOOPS_R; solid, dashed and dotted line, respectively), reflecting the ability to distinguish between substrates and non-substrates. **(D)** D-TRAINING set (26 peptides) **(E)** D-TRAINING + D-CAPPED (26 + 16 peptides). See **Table 2** for protocol definitions.

Figure 4. Application of the calibrated protocol (RLOOP_R) to the acetylome to detect novel potential KDAC6 substrates. (A) Distribution of scores obtained for acetylated peptides (as annotated in the PhosphoSitePlus database). The bin for the best scoring peptide in the training set is highlighted in blue, and a green arrow marks the threshold for discriminating substrates defined for RLOOP_R (-890 and -900 for non-strict and strict score cutoffs, respectively). **(B-D)** Sequence logos of **(B)** measured peptides (D-TRAINING set) and **(C)** merged D-TRAINING and D-CAPPED datasets **(D)** top 100 peptides predicted by our protocol.

Figure 5. Comparison of substrate specificities of different experimental datasets. (A-C) Sequence logos made with PSSMSearch, using its default background containing human sequences for (A) D-13MER, (B) D-SILAC, and (C) D-3MER datasets. (D-F) Sequence logos made with Two Sample Logo using the non-substrate peptides as background for each dataset respectively for (D) D-13MER, (E) D-SILAC, and (F) D-3MER datasets. (G) Plot of substrate activities for peptides shared between D-13MER and D-SILAC datasets. S, NS: substrates and non-substrates, respectively.

A

Peptide	Protein (site of modification)	k_{cat}/K_M ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_M (μM)
EG(K-Ac)FVR	LMNA (K-450)	220,000 ± 40,000	4.2 ± 0.7	19 ± 4
SD(K-Ac)TIG	TUBA1A, TUBA4A, TUBA3E, TUBA1C, TUBA1B, TUBA4A (K-40) [8,56,57]	70,000 ± 30,000	2.1 ± 0.3	29 ± 15
AM(K-Ac)HRS	MYO1G (K-90)	63,000 ± 9,000	1.9 ± 0.2	30 ± 7
LA(K-Ac)EMK	CFAP157 (K-35)	60,000 ± 20,000	1.4 ± 0.2	20 ± 10
TG(K-Ac)TVA	TCOF1 (K-146)	50,000 ± 10,000	1.37 ± 0.09	27 ± 7
RT(K-Ac)SGD	ARHGEF1 (K-234)	47,000 ± 9,000	1.5 ± 0.2	32 ± 9
SQ(K-Ac)KTF	GRP94 (K-682)	44,000 ± 9,000	1.6 ± 0.1	36 ± 10
AG(K-Ac)RIA	DIP2A (K-962)	40,000 ± 10,000	1.9 ± 0.5	50 ± 30
YK(K-Ac)FYE	HSP90A (K-436) [10,11]	40,000 ± 10,000	1.8 ± 0.6	21 ± 9
YG(K-Ac)LRK	ACTN1 (K-195)	31,000 ± 4,000	6.0 ± 0.7	200 ± 80
EG(K-Ac)TNY	ZNF587 (K-209)	29,000 ± 7,000	2.0 ± 0.4	70 ± 30
QK(K-Ac)VKE	ZNF280D (K-32)	28,000 ± 6,000	1.8 ± 0.2	60 ± 20
YE(K-Ac)EKE	KIF5A, KIF5C (K-348); KIF5B (K-346)	27 ,000 ± 5,000	1.8 ± 0.1	60 ± 10
EG(K-Ac)TGE	ZKSCAN1 (K-310)	25,000 ± 4,000	1.07 ± 0.08	43 ± 9
LS(K-Ac)KSK	PARP1 (K-505)	24,000 ± 3,000	1.7 ± 0.2	70 ± 20
PA(K-Ac)ESP	ATBF1 (K-3416); treacle (K-904)	22,000 ± 4,000	1.08 ± 0.05	49 ± 9
AM(K-Ac)KIR	CDK1 (K-33)	16,000 ± 2,000	0.98 ± 0.09	60 ± 10
QY(K-Ac)KEL	LMNA (K-260)	13,200 ± 700	1.58 ± 0.08	120 ± 10
KT(K-Ac)PIW	HSP90A (K-294) [10,11]	12,000 ± 2,000	0.8 ± 0.1	60 ± 20
AH(K-Ac)RGS	*DPP3 (K-294); CBLL2 (K-107); RFPL4A (K-257)*	*9,000 ± 3,000*	*>1*	*>150*
KL(K-Ac)KKE	*MYH1 (K-1085)*	5,700 ± 1,100	0.50 ± 0.09	90 ± 30
EV(K-Ac)KMT	*MAP4 (K-847)*	5,600 ± 800	0.19 ± 0.02	35 ± 7
GY(K-Ac)KTK	*RPL4 (K-162)*	5,000 ± 2,000	2.3 ± 0.5	400 ± 200
AH(K-Ac)KSH	*S100A8 (K-84)*	4,700 ± 400	>0.5	>100
PL(K-Ac)KDR	*RPL3 (K-393)*	2,100 ± 200	>0.5	>200
SW(K-Ac)DGL	*ACTN2 (K-181)*	1,200 ± 300	>0.3	>200

B

YD(K-Ac)DEV	SIRT1 (K-430)	*>110,000*	0.28 ± 0.04	<25
ME(K-Ac)FKI	XPOT (K-627)	*>80,000*	2.0 ± 0.5	<25
ET(K-Ac)YRW	VDAC1 (K-61)	*>70,000*	3.7 ± 1.2	<50
GG(K-Ac)RVM	VPS35 (K-694)	*>60,000*	1.4 ± 0.2	<25
YD(K-Ac)LRK	ACTN4 (K-214); TUBG1 (K-397)	*>50,000*	1.2 ± 0.2	<25
HG(K-Ac)EVG	PCPB1 (K-23); PCPB1 (K-57)	*>40,000*	2.2 ± 0.5	<50
QG(K-Ac)SGN	MYO18B (K-401)	*>40,000*	1.1 ± 0.2	<25
ID(K-Ac)RTI	EEF1A1 (K-31); EEF1A2, EEF1A1P5 (K-36)	*>40,000*	1.0 ± 0.2	<25
GA(K-Ac)DEP	TCOF1 (K-1414)	*>40,000*	0.9 ± 0.2	<25
VS(K-Ac)RKL	UBA1 (K-1024); ARHGAP32 (K-674)	*>40,000*	1.0 ± 0.1	<25
YE(K-Ac)FRN	USP32 (K-145)	*>40,000*	1.1 ± 0.2	<25
MK(K-Ac)LKE	PASD1 (K-379)	*>30,000*	0.7 ± 0.1	<25
SQ(K-Ac)YKR	MART3 (K-473)	*>30,000**	1.3 ± 0.4*	<50*

CG(K-Ac)GLE	CSRP1 (K-151)	*>20,000*	1.1 ± 0.4	<50
SE(K-Ac)ILQ	TCOF1 (K-294)	*>20,000*	1.2 ± 0.2	<50
EG(K-Ac)GNG	EIF5 (K-28)	*>17,000*	0.84 ± 0.09	<50

Table 1. Activities of 47 selected peptides, as measured by the acetate assay. (A) Peptide measurements with precise k_{cat}/K_M values (D-TRAINING set). **(B)** Peptide substrates with k_{cat}/K_M values with lower limits: K_M values were below the detection limit (D-CAPPED set). If a peptide sequence appears in more than one protein, all proteins are mentioned. **bold**: strongest measured substrate used for template optimization; *italics:* non-substrates, as defined by the applied cutoff of $k_{cat}/K_M < 10^4$ $M^{-1}s^{-1}$; *calculated from 3 data points.

A

Protocol name	Starting structure selected by	Receptor backbone minimization	Loop modeling
BASIC_I	interface score	-	-
RMIN_I		✓	-
RLOOP_I		-	✓
RLOOPS_I		-	✓✓
BASIC_R	***reweighted score***	-	-
RMIN_R		✓	-
RLOOP_R		**-**	✓
RLOOPS_R		-	✓✓

B

Protocol	Threshold	Specificity	Sensitivity	MCC	Spearman correlation	AUC
RLOOP_I	-866	0.86 (0.86)	0.63 (0.66)	0.43 (0.38)	-0.61	0.79
RLOOP_R	**-891**	**0.86 (0.86)**	**0.53 (0.68)**	**0.34 (0.40)**	**-0.55**	**0.75**
RLOOPS_R	-936	0.86 (0.86)	0.58 (0.54)	0.39 (0.30)	-0.40	0.60

Table 2. FlexPepBind protocols evaluated in this study. (A) Definition of different protocols. RLOOP_R, the final protocol, is highlighted in bold and italicized. See Text for more details. Checkmarks indicate which protocol elements were used in each protocol. **(B)** Performance metrics of the top 3 protocols on the D-TRAINING dataset ($n = 26$). MCC: Matthew correlation coefficient, AUC: area under the ROC curve. Values for the dataset including both D-TRAINING and D-CAPPED peptides ($n = 26 + 16$) are indicated in parentheses.

Peptide	Protein (site of modification)	Rosetta reweighted score (REU)	k_{cat}/K_M ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_M (μM)
FP(K-Ac)EAK	EGFR (K-1179)	-913.74	40,000 ± 5,000	0.53 ± 0.02	13 ± 2
HS(K-Ac)GFG	TARDBP (K-145)	-909.19	40,000 ± 10,000	1.2 ± 0.3	30 ± 20
QA(K-Ac)SPP	MEF2C (K-239)	-905.57	26,000 ± 4,000	1.2 ± 0.2	40 ± 10
IS(K-Ac)MND	IFI16 (K-451)	-911.69	25,000 ± 9,000	0.48 ± 0.08	20 ± 10
SG(K-Ac)GKK	GATA1 (K-312)	-908.68	16,000 ± 3,000	0.31 ± 0.03	19 ± 6
PG(K-Ac)EEK	FOXM1 (K-440)	-914.77	15,000 ± 4,000	0.5 ± 0.1	35 ± 16
KG(K-Ac)QAE	HMGN1 (K-61)	-909.30	14,000 ± 3,000	0.36 ± 0.04	26 ± 8
NG(K-Ac)LTG	GAPDH (K-227)	-897.70	14,000 ± 1,000	0.36 ± 0.03	26 ± 4
MG(K-Ac)GVS	ENO1 (K-60)	-905.40	11,000 ± 1,000	0.49 ± 0.06	40 ± 10
PC(K-Ac)EVD	NFAT5 (K-282)	-923.56	4,800 ± 500	>0.2	>20

Table 3. Validating the predicted substrates of the acetylome. The peptides are sorted according to their experimentally measured KDAC6 substrate strength.

Figure 1.

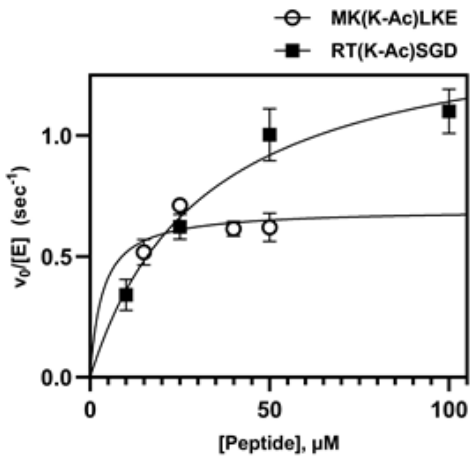


Figure 2.

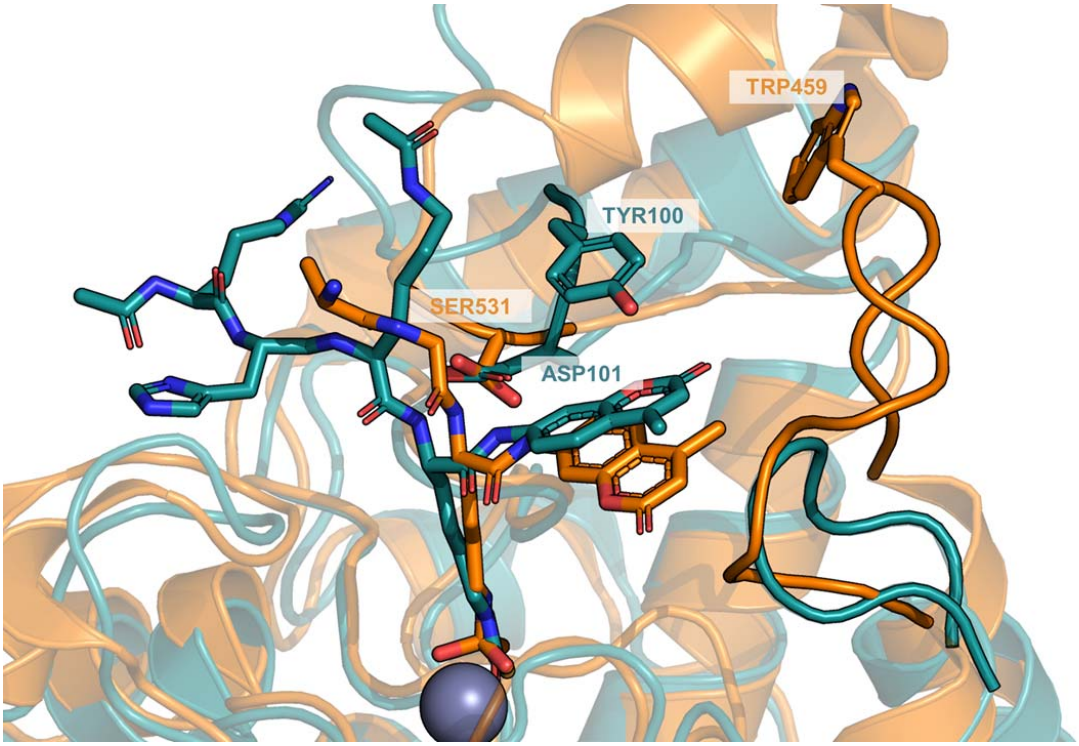


Figure 3.

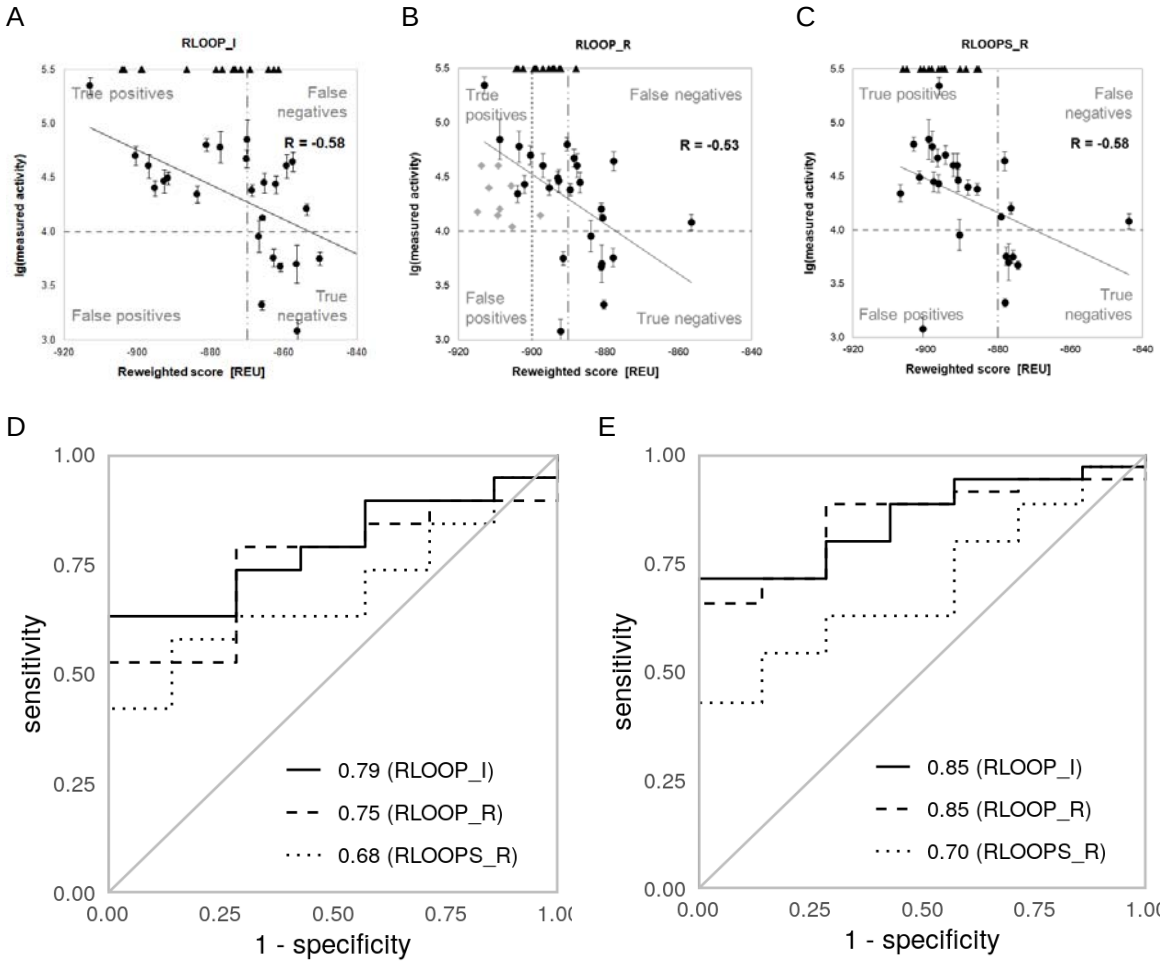

A
RLOOP_I
True positives
False negatives
R = -0.58
False positives
True negatives
lg(measured activity)
Reweighted score [REU]
B
RLOOP_R
True positives
False negatives
R = -0.53
False positives
True negatives
lg(measured activity)
Reweighted score [REU]
C
RLOOPS_R
True positives
False negatives
R = -0.58
False positives
True negatives
lg(measured activity)
Reweighted score [REU]
D
sensitivity
1 - specificity
0.79 (RLOOP_I)
0.75 (RLOOP_R)
0.68 (RLOOPS_R)
E
sensitivity
1 - specificity
0.85 (RLOOP_I)
0.85 (RLOOP_R)
0.70 (RLOOPS_R)

Figure 4.

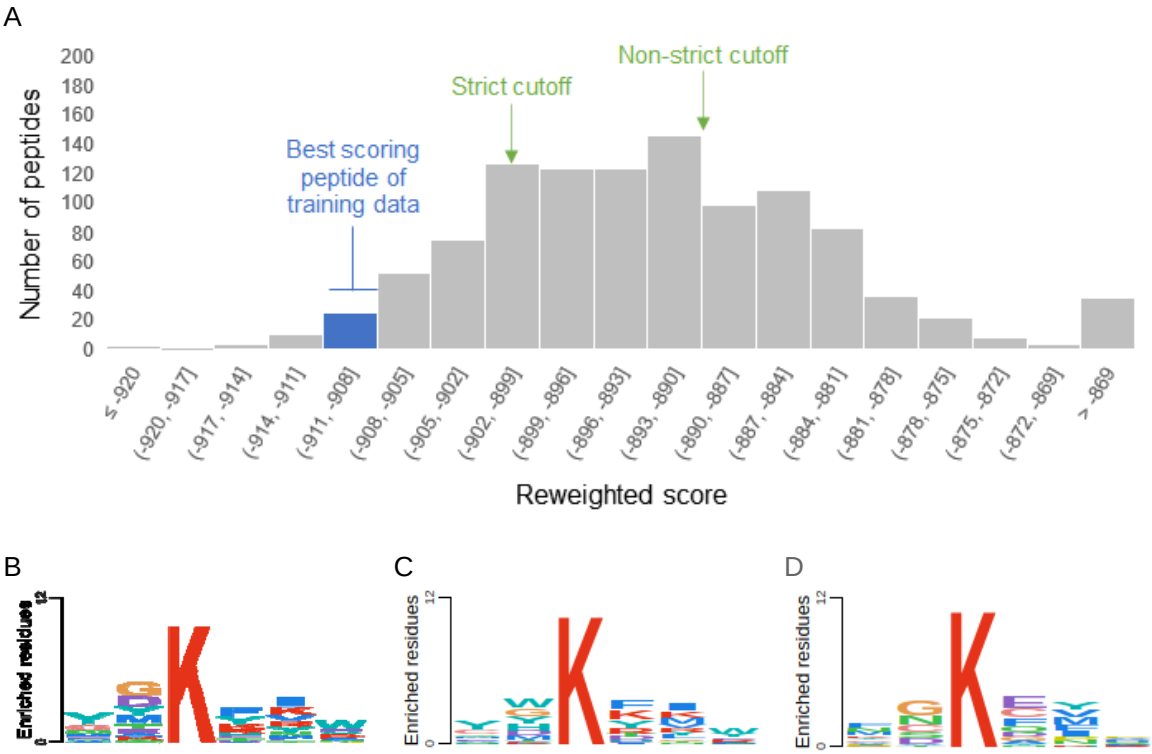
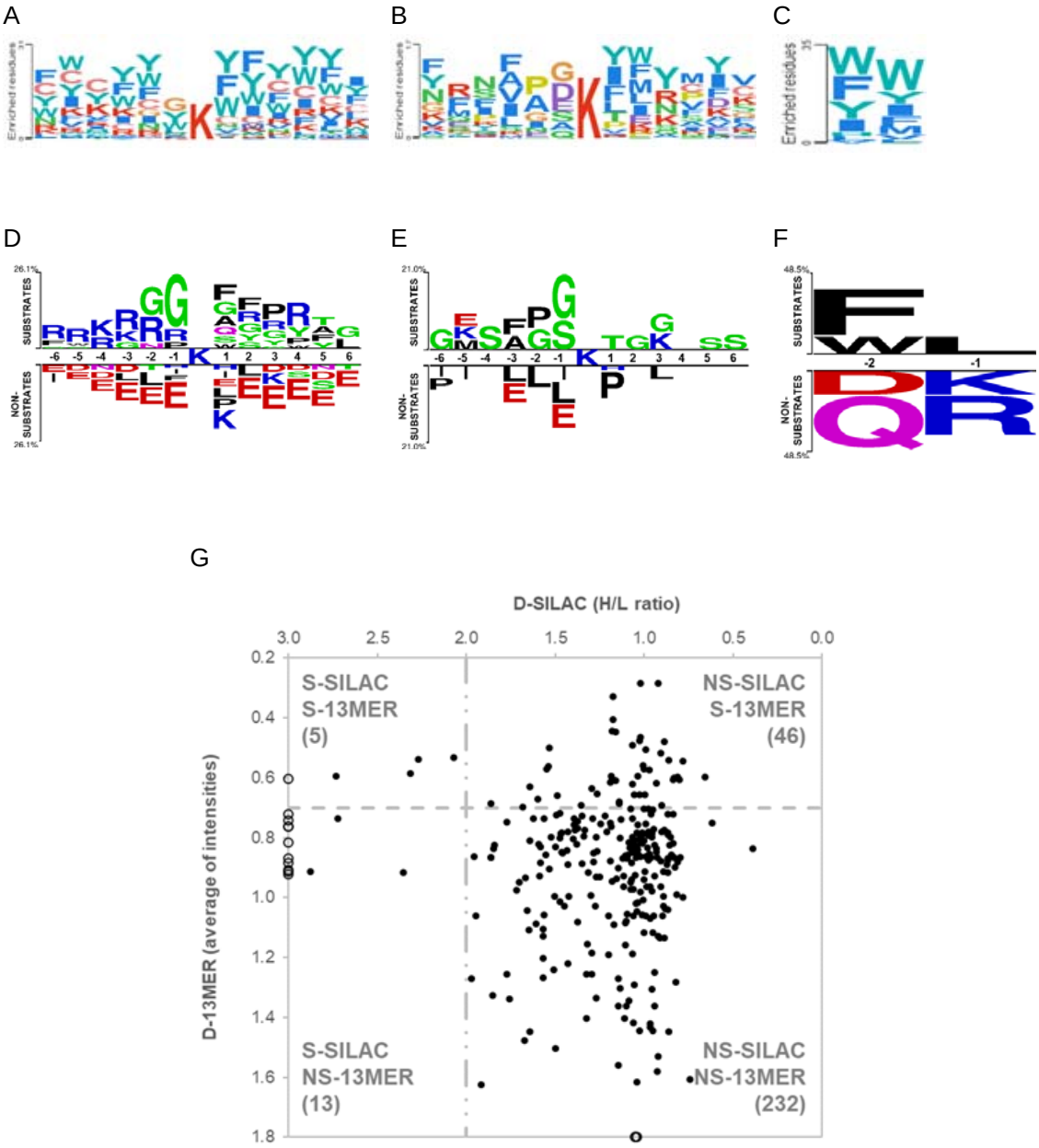

A
Number of peptides
200
180
160
140
120
100
80
60
40
20
0
Best scoring peptide of training data
Strict cutoff
Non-strict cutoff
≤ -920
(-920, -917]
(-917, -914]
(-914, -911]
(-911, -908]
(-908, -905]
(-905, -902]
(-902, -899]
(-899, -896]
(-896, -893]
(-893, -890]
(-890, -887]
(-887, -884]
(-884, -881]
(-881, -878]
(-878, -875]
(-875, -872]
(-872, -869]
> -869
Reweighted score
B
C
D
Enriched residues
12
0

Figure 5.



See supplementary tables and materials in Supplementary Materials.