Supplementary Materials

for

Structure-based prediction of HDAC6 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]

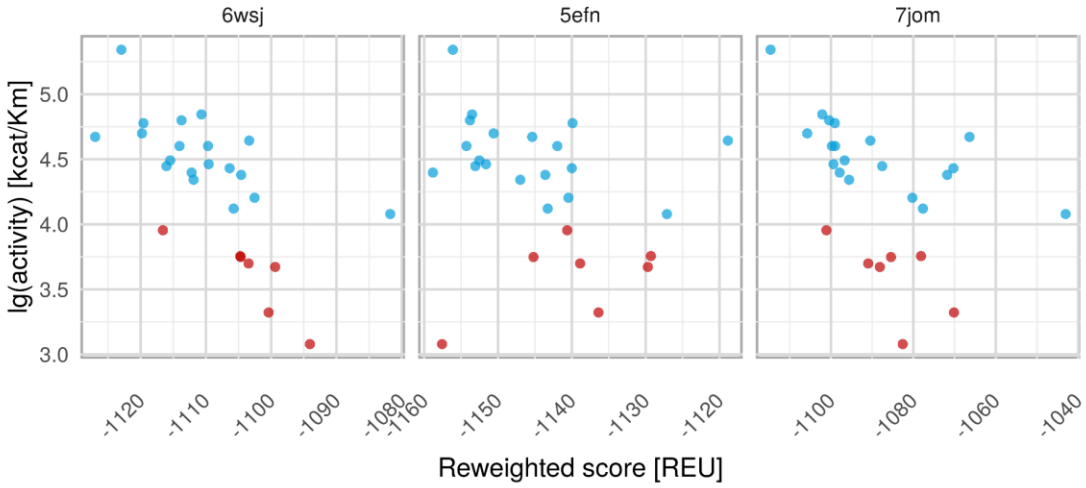
a Department of Microbiology and Molecular Genetics, Institute for Medical Research Israel-Canada (IMRIC), Hebrew University, Faculty of Medicine, POB 12272, Jerusalem, 91120 Israel

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

Contents



Supplementary Figure S1. Performance on different starting structures

Shown are results for the training set (D-TRAINING). The calibrated protocol was run on a structure of HDAC6 DD2 domain bound to (1) a cyclic peptide (6WSJ; same as **Figure 2A**, (2) a trimer peptide connected to coumarine (5EFN), and (3) a small molecule inhibitor (7JOM). (blue: substrates; red: non-substrates).

A

	1	2	3	4	5	6	7	8	9	10	11	12	13
C	-4.80	-4.38	-4.23	-4.53	-4.67	-5.75	0	-4.82	-5.07	-4.78	1.61	-4.79	1.90
P	-1.86	-0.84	1.31	-1.61	3.46	-4.91	0	1.25	0.53	-0.09	1.55	-4.56	-0.97
Q	-0.53	0.98	0.45	-0.69	-1.21	1.42	0	-2.06	-0.78	0.80	0.31	-2.41	-2.13
N	2.61	0.19	2.44	-4.16	0.86	-0.05	0	0.20	-0.58	1.98	0.43	-0.06	-2.68
T	-0.24	0.45	0.03	-0.49	0.15	-0.34	0	1.83	-1.83	-0.42	-0.65	-0.24	0.44
S	0.46	0.63	1.69	0.95	-0.74	1.95	0	-0.33	0.35	0.90	1.11	0.23	1.35
G	1.69	-0.69	-3.61	-0.42	1.69	3.73	0	-1.67	-2.30	-0.15	-1.69	-1.54	1.09
A	0.83	-0.79	0.46	2.81	3.11	1.45	0	-1.17	-3.06	0.61	0.83	0.99	0.91
V	1.04	0.03	-0.34	2.79	-0.68	-0.49	0	0.78	-0.02	-1.41	1.54	0.87	2.03
I	-0.89	-0.21	1.36	2.42	1.13	-5.98	0	3.50	1.93	-4.17	0.72	2.66	-0.47
L	-4.14	0.30	-0.85	1.79	-1.14	-6.44	0	1.91	2.02	-0.93	-0.29	0.67	-2.45
M	-0.56	1.15	-2.67	0.46	-3.65	-5.27	0	-2.58	2.61	-3.26	1.93	0.18	-0.35
F	2.69	2.45	1.52	3.36	-5.53	-6.95	0	3.00	2.63	-4.05	1.11	1.80	1.04
Y	2.63	-3.11	0.66	-4.06	-5.17	-6.05	0	3.60	-3.39	4.22	-3.36	3.11	0.33
W	-4.93	-4.71	-4.88	-5.67	-6.27	-7.02	0	-4.94	3.20	-4.77	-4.93	-5.09	-5.15
H	-3.20	-2.93	-0.54	-4.78	-4.18	-4.35	0	-4.02	-0.49	-2.65	-3.26	-3.67	-3.61
K	1.48	0.46	0.71	0.71	-0.68	-3.26	K	0.31	1.14	1.90	0.61	1.59	1.49
R	0.86	2.56	0.34	-1.22	-3.77	-2.39	0	0.00	0.63	2.48	-0.15	0.20	0.71
D	-0.11	-1.40	-0.24	-4.81	-0.33	3.34	0	0.36	-1.41	0.05	-0.76	1.73	-0.34
E	-0.39	0.79	0.25	-2.37	0.81	1.98	0	-2.43	1.55	0.01	-0.74	0.75	0.21

C

	1	2	3
C	-5.96	-4.89	0
P	-0.72	-0.74	0
Q	-4.64	-3.81	0
N	-5.62	0.22	0
T	-1.20	-1.09	0
S	-4.90	-1.42	0
G	-0.88	-0.25	0
A	-4.77	-0.86	0
V	1.05	1.75	0
I	4.58	4.13	0
L	1.35	1.15	0
M	-3.13	2.87	0
F	9.09	3.75	0
Y	6.67	4.62	0
W	10.58	11.55	0
H	1.54	-4.33	0
K	-0.40	-4.41	K
R	-5.13	-4.94	0
D	-5.94	-4.70	0
E	-1.18	-0.34	0

B

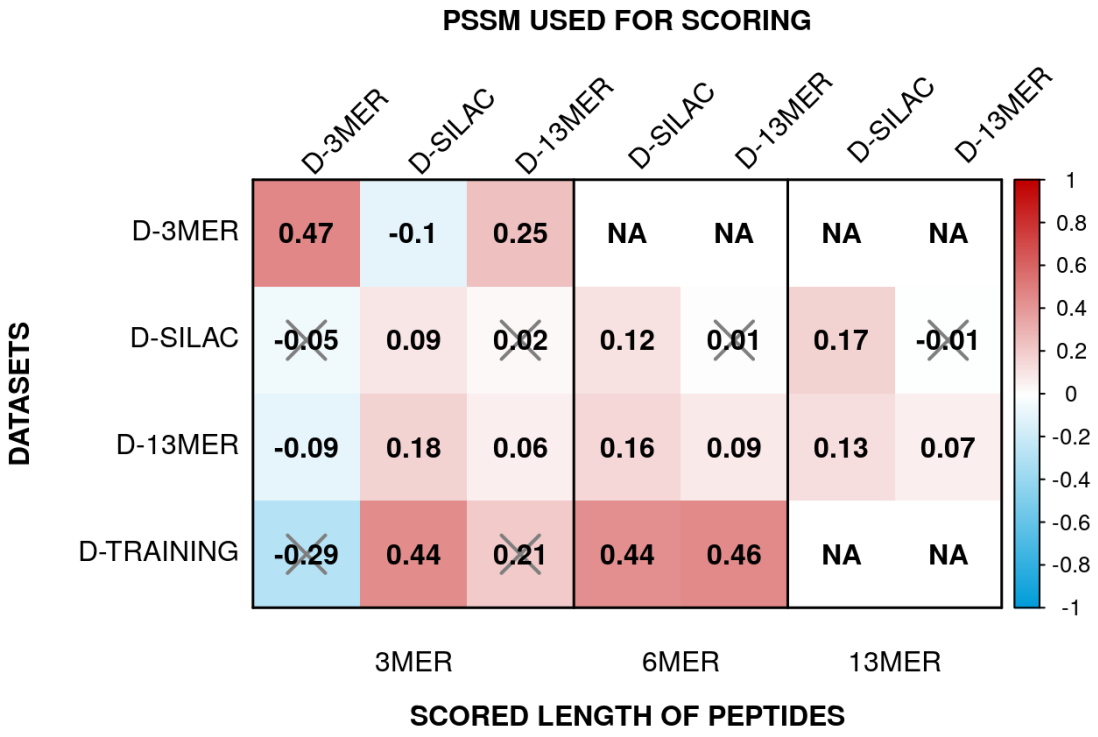
	1	2	3	4	5	6	7	8	9	10	11	12	13
C	4.16	4.65	4.43	2.17	3.12	-1.34	0	2.46	2.86	4.06	4.24	3.41	2.49
P	-1.50	-4.22	-1.59	-2.40	-4.44	-1.56	0	-5.10	-3.18	0.19	-1.64	-2.13	-1.59
Q	-0.61	-0.94	-1.31	-0.82	0.00	-0.87	0	0.11	-1.68	-1.22	-2.42	-1.74	-1.81
N	2.48	0.66	-1.23	-0.21	2.21	0.71	0	1.38	0.91	1.43	0.90	-1.47	-0.51
T	0.33	0.11	0.30	-0.97	-3.03	-1.56	0	-0.09	-1.50	-0.93	-1.16	0.98	-2.32
S	-1.89	-2.53	-1.70	-0.84	-1.67	-1.40	0	-1.50	-1.52	-1.54	-3.55	-4.35	-1.63
G	-0.68	-0.84	-0.84	-0.35	2.37	4.32	0	0.80	-0.28	-0.62	-0.01	-0.28	0.19
A	-0.21	0.68	-0.16	0.56	0.10	-0.68	0	0.82	-1.41	0.01	-1.02	0.58	0.16
V	0.01	1.81	1.41	0.51	0.51	-0.11	0	1.35	0.92	1.53	-0.01	2.25	0.77
I	-1.08	3.55	1.00	2.19	2.22	0.47	0	0.49	3.36	2.11	3.31	1.88	3.18
L	1.08	0.36	0.61	-0.54	0.08	0.86	0	-0.03	0.17	0.77	1.41	1.52	2.45
M	0.13	0.42	0.43	-0.95	-2.59	-0.14	0	-2.09	1.07	-0.43	0.30	0.81	0.16
F	4.64	2.04	1.28	4.29	4.18	0.44	0	7.33	7.41	3.67	2.63	5.09	3.01
Y	3.63	4.12	3.99	5.12	5.84	3.94	0	7.60	6.47	5.14	6.51	6.54	3.16
W	3.33	4.69	0.69	3.91	4.84	2.40	0	6.62	4.30	3.04	6.38	4.01	1.24
H	-2.38	-0.67	0.98	0.13	-1.20	-3.54	0	-1.72	-0.26	-1.69	-2.24	-2.09	0.98
K	1.57	2.62	3.65	2.93	1.04	-0.74	K	-3.17	0.64	0.89	2.10	2.60	2.43
R	1.59	1.50	1.04	1.77	1.23	0.30	0	-1.07	1.04	1.68	2.70	0.12	0.30
D	-0.77	-2.18	-1.50	-1.62	-0.31	-0.41	0	-0.99	-0.35	-1.09	-1.05	-3.06	-0.55
E	-1.74	-1.41	-1.81	-3.09	-3.75	-2.03	0	-1.43	-2.79	-4.20	-3.58	-3.16	-2.46

D

	1	2	3	4	5	6
C	-2.48	-2.79	0	-3.84	-2.77	-3.37
P	-2.20	-2.77	0	0.95	-3.81	-2.96
Q	0.00	1.08	0	0.22	-2.15	-0.95
N	-1.23	-1.35	0	-0.82	-2.78	-1.39
T	0.49	1.52	0	0.32	0.78	-1.34
S	0.61	0.06	0	0.33	-1.69	0.38
G	-2.02	2.09	0	-2.94	0.48	-2.35
A	0.81	0.78	0	-1.27	-1.45	0.62
V	-1.46	-1.80	0	-3.05	1.48	-2.89
I	-1.94	-2.42	0	-3.50	1.91	-3.35
L	0.13	-2.08	0	-3.17	-0.01	-2.92
M	-0.95	1.42	0	-2.30	1.61	-2.33
F	-3.04	-3.61	0	0.76	-2.24	1.49
Y	-2.72	-3.29	0	-0.56	-2.86	-0.35
W	-4.00	-3.82	0	-3.37	-4.19	3.04
H	-1.92	-2.46	0	1.39	-3.53	-2.21
K	0.65	-1.36	K	0.66	-1.76	0.70
R	0.83	-2.03	0	0.52	0.67	0.65
D	-1.56	-2.45	0	-1.20	-3.90	0.75
E	0.48	-1.44	0	0.63	-3.20	-0.97

Supplementary Figure S2. Position specific scoring matrices derived from different datasets

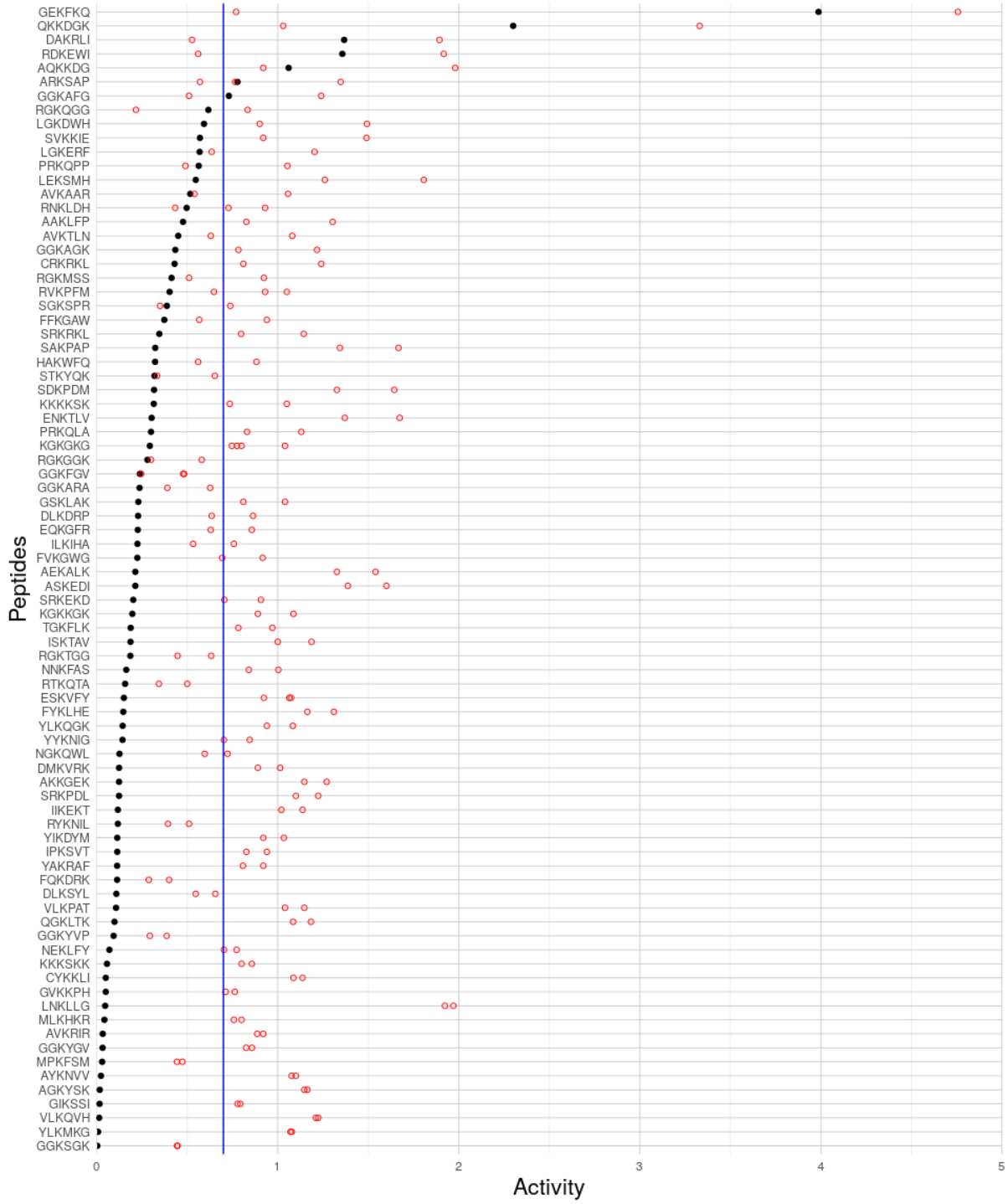
PSSMS were generated from **A)** D-SILAC, **B)** D-13MER, **C)** D-3MER, and **D)** D-TRAINING sets by applying PSSMSearch to the substrate list of each experiment (for definition of substrates, see **Supplementary Table S1**). (blue: enriched residues, red: depleted amino acids).



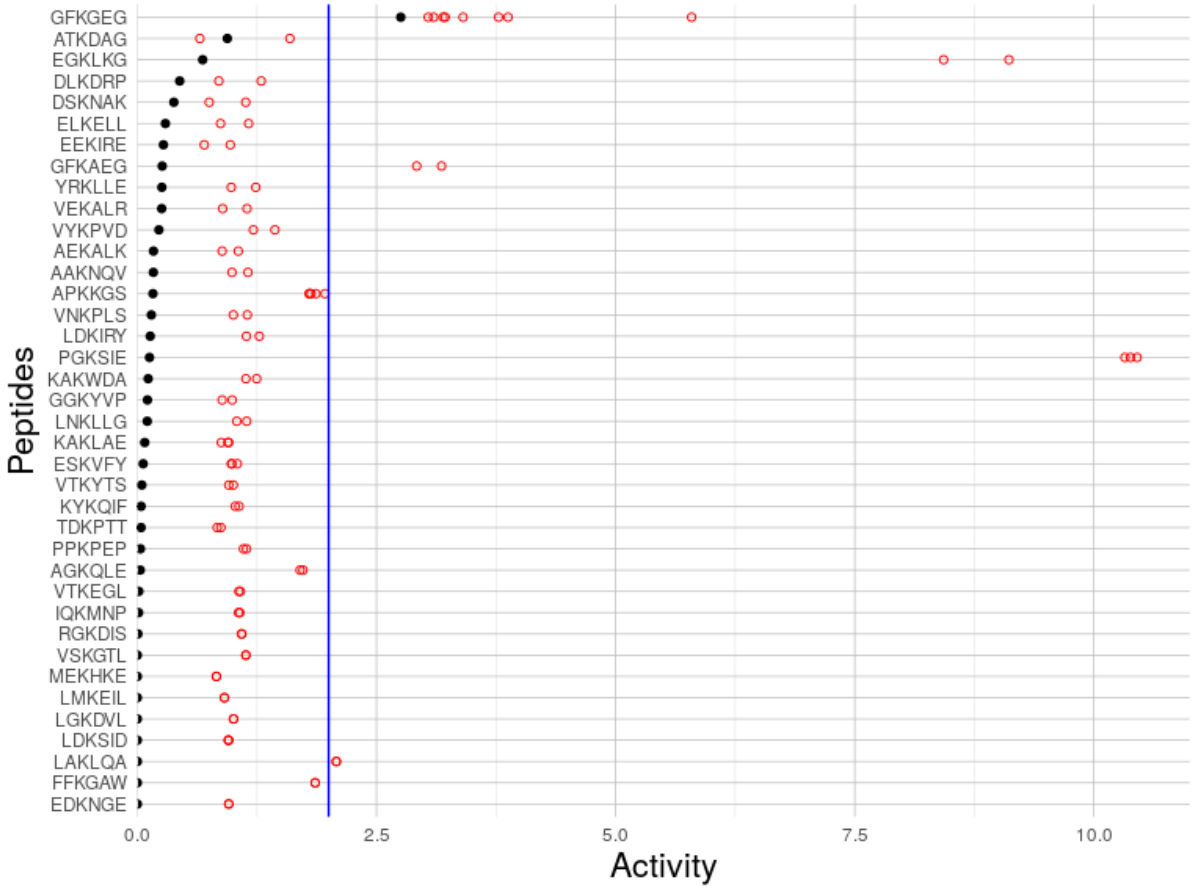
Supplementary Figure S3. PSSM scores and experimental values do not correlate.

Each dataset was cross-scored with every obtained PSSM depicted in **Supplementary Figure S2**. Three different lengths for scoring (3-, 6- and 13-mers) were used where applicable. D-3MER, D-SILAC and D-13MER labels at the top represent the dataset that the respective PSSM was derived from. 3MER, 6MER and 13MER labels indicate the number of amino acids used for scoring. X indicates non-significant correlations.

A


GEKFKQ
QKKDGK
DAKRLI
RDKEWI
AQKKDG
ARKSAP
GGKAFG
RGKQGG
LGKDWH
SVKKIE
LGKERF
PRKQPP
LEKSMH
AVKAAR
RNKLDH
AAKLFP
AVKTLN
GGKAGK
CRKRKL
RGKMSS
RVKPFM
SGKSPR
FFKGAW
SRKRKL
SAKPAP
HAKWFQ
STKYQK
SDKPDM
KKKKSK
ENKTLV
PRKQLA
KGKGKG
RGKGGK
GGKFGV
GGKARA
GSKLAK
DLKDRP
EQKGFR
ILKIHA
FVKGWG
AEKALK
ASKEDI
SRKEKD
KGKKGK
TGKFLK
ISKTAV
RGKTGG
NNKFAS
RTKQTA
ESKVFY
FYKLHE
YLKQGK
YYKNIG
NGKQWL
DMKVRK
AKKGEK
SRKPDL
IIKEKT
RYKNIL
YIKDYM
IPKSVT
YAKRAF
FQKDRK
DLKSYL
VLKPAT
QGKLTK
GGKYVP
NEKLFY
KKKSKK
CYKKLI
GVKKPH
LNKLLG
MLKHKR
AVKRIR
GGKYGV
MPKFSM
AYKNVV
AGKYSK
GIKSSI
VLKQVH
YLKMKG
GGKSGK
Peptides
0
1
2
3
4
5
Activity

B



Supplementary Figure S4. Range of measured activities for peptides sharing identical core hexamers with differing flanking regions.

A) D-13MER and **B)** D-SILAC datasets. (Black dots: range, red circles: measured values, blue line: threshold of substrate-non-substrate distinction) Related to **Figure 4**.

Supplementary Table S1. Datasets evaluated in this study. The unit of measurement of substrate activity and parameters used to define substrates are indicated. See Text for more details. Related to **Figure 4**.

Dataset [reference]	**N**[b] ***peptides***	**N*substrates***	**N*non-substrates***	**Cutoff for substrates**	**Cutoff for non-substrates**	**Unit**
D-TRAINING[a]	26	19	7	≧10^4	-	k_{cat}/K_M
D-CAPPED[a]	16	16	0	≧10^4	-	k_{cat}/K_M
D-13MER[1]	6797	395 (51)[c]	4127 (245)[c]	4 out of 4	1 out of 4	# experiments that identified a substrate
D-HPLC[1]	24	17	7	≧10^4	-	k_{cat}/K_M
D-SILAC[2]	929	63 (18)[c]	328 (278)[c]	≧2	≦1	ratio (H/L)
D-3MER[3]	361	35	33	≧70	≦21	- (intensity)

[a] D-TRAINING & D-CAPPED: present study (see **Table 1)**.

[b] *N* indicates the number of respective peptides

[c] In parentheses: peptides that were reported both in the D-13MER and D-SILAC sets.

Supplementary Table S2. Kinetic parameters of the H3 K14$_{Ac}$ peptide and full-length protein from which it was derived. Related to **Table 1.**

H3 K14ac substrate	k_{cat}/K_M ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_M (μM)
13-mer: RKSTGG(K-ac)APRKQL	140,000±10,000	4.1 ± 0.3	28 ± 5
Full length protein	80,000±40,000	1.4 ± 0.6	20 ± 10

Supplementary Table S3. Peptides measured for HDAC6 deacetylation in this study but not used for training or validation.

These peptides were not included due to poor measurement accuracy, or peptide length beyond 6 residues. *calculated from only 3 data points. Related to **Table 1**.

Peptide	Protein (site of modification)	k_{cat}/K_M ($M^{-1}s^{-1}$)	k_{cat} (s^{-1})	K_M (μM)	Rosetta reweighted score [REU]
ME(K-Ac)KKE	GBP7 (K-389)	3,300 ± 200*	>0.3*	>150*	-1,105
QD(K-Ac)PLR	CCDC86 (K-261)	>2,000*	0.11 ± 0.04*	<50*	-1,092
kgGA(K-Ac)RHR	H4K16 (K-16)[4]	70,000 ± 20,000	1.23 ± 0.09	17 ± 6	-1,106
SLG(K-Ac)DWHR	CRIP1 (K-22), CRIP1 (K-144)	31,000 ± 5,000	10 ± 2	300 ± 200	-1,113
rMF(K-Ac)QFNK	TRIM28 (K-770)	21,000 ± 1,000	>2	>200	-1,110
riIL(K-Ac)ASR	MSH2 (K-635)[5]	20,000 ± 3,000	4 ± 1	180 ± 80	-1,115

Supplementary Table S4. Performance of protocol on high-throughput datasets.

(MCC: Matthews correlation coefficient, AUC: area under the ROC curve)

Dataset	Specificity	Sensitivity	MCC	Spearman correlation	AUC
D-13MER	0.18	0.77	-0.05	-0.01	0.49
D-SILAC	0.19	0.89	0.05	-0.05	0.52

Supplementary Table S5. Files of Rosetta protocols.

Constraint File for docking and minimization measured on PDB ID 6WSJ. (Chain A: receptor, chain F is the peptide). Fixbb resfile is used for peptide threading.

Constraint file	Dihedral N 3F CA 3F C 3F N 4F CIRCULARHARMONIC -36.7 0.5 AtomPair OG 531A N 3F HARMONIC 2.9 0.2 AtomPair OD2 705A OH 3F HARMONIC 3.7 0.2 AtomPair OD1 705A OH 3F HARMONIC 5.2 0.2 AtomPair OD2 612A OH 3F HARMONIC 3.4 0.2 AtomPair OD1 612A OH 3F HARMONIC 3.7 0.2 AtomPair ND1 614A OH 3F HARMONIC 3.8 0.2 AtomPair O 582A NZ 3F HARMONIC 3.3 0.2 AtomPair CE1 583A CD 3F HARMONIC 4.0 0.2 AtomPair CE2 583A CD 3F HARMONIC 3.7 0.2 AtomPair CD2 643A CG 3F HARMONIC 4.0 0.2 AtomPair CD1 643A CG 3F HARMONIC 3.7 0.2
fixxb resfile	NATRO start 1 F PIKAA E EX 1 EX 2 USE_INPUT_SC 2 F PIKAA G EX 1 EX 2 USE_INPUT_SC 4 F PIKAA F EX 1 EX 2 USE_INPUT_SC 5 F PIKAA V EX 1 EX 2 USE_INPUT_SC 6 F PIKAA R EX 1 EX 2 USE_INPUT_SC

Supplementary Table S6. Commands and flags of Rosetta runs.

All commands were run with Rosetta v2020.28.

Prepack	`$ROSETTA_HOME/main/source/bin/FlexPepDocking.default.linuxgccrelease -s input/6wsj.pdb -ex1 -ex2aro -use_input_sc -flexpep_prepack -nstruct 1 -scorefile ppk.score.sc -flexpep_score_only -out:path:pdb input -out:path:score output -unboundrot input/5eem.pdb 6wsj.pdb`
Peptide docking	`$ROSETTA_HOME/main/source/bin/FlexPepDocking.mpi.linuxgccrelease -s input/6wsj.ppk.pdb -ex1 -ex2aro -use_input_sc -constraints:cst_fa_file input/constraints_6wsj.cst -constraints:cst_fa_weight 1.0 -unboundrot input/5eem.pdb input/6wsj.ppk.pdb -nstruct 250 -flexpep_score_only -scorefile refine.score.sc -out:path:pdb output -out:file:silent output/decoys.silent -out:file:silent_struct_type binary -out:path:score output -overwrite -flexPepDocking:pep_refine -lowres_preoptimize`  `(-min_receptor_bb flag was added in some protocols, as described in Methods)`
Peptide threading	`$ROSETTA_HOME/main/source/bin/fixbb.default.linuxgccrelease -database $ROSETTA_HOME/database -resfile resfile -s template.pdb -ex1 -ex2aro -use_input_sc -scorefile design.score.sc -nstruct 1 -unboundrot 5eem.pdb template.pdb`
Minimization	`$ROSETTA_HOME/main/source/bin/FlexPepDocking.mpi.linuxgccrelease -s start.ppk.pdb -ex1 -ex2 -ex3 -ex4 -constraints:cst_fa_file constraints.cst -constraints:cst_fa_weight 1.0 -scorefile min.score.sc -unboundrot 5eem.pdb start.ppk.pdb -flexPepDockingMinimizeOnly -flexpep_score_only` `(-min_receptor_bb flag was added in some protocols, as described in Methods)`

References

1. Kutil, Z. *et al.* The unraveling of substrate specificity of histone deacetylase 6 domains using acetylome peptide microarrays and peptide libraries. *FASEB J.* **33**, 4035–4045 (2019).
2. Schölz, C. *et al.* Acetylation site specificities of lysine deacetylase inhibitors in human cells. *Nat. Biotechnol.* **33**, 415–423 (2015).
3. Riester, D., Hildmann, C., Grünewald, S., Beckers, T. & Schwienhorst, A. Factors affecting the substrate specificity of histone deacetylases. *Biochem. Biophys. Res. Commun.* **357**, 439–445 (2007).
4. 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.* **235**, 7030–7042 (2020).
5. Zhang, M. *et al.* HDAC6 deacetylates and ubiquitinates MSH2 to maintain proper levels of MutSα. *Mol. Cell* **55**, 31–46 (2014).