Supplementary Information

Supplementary Figures

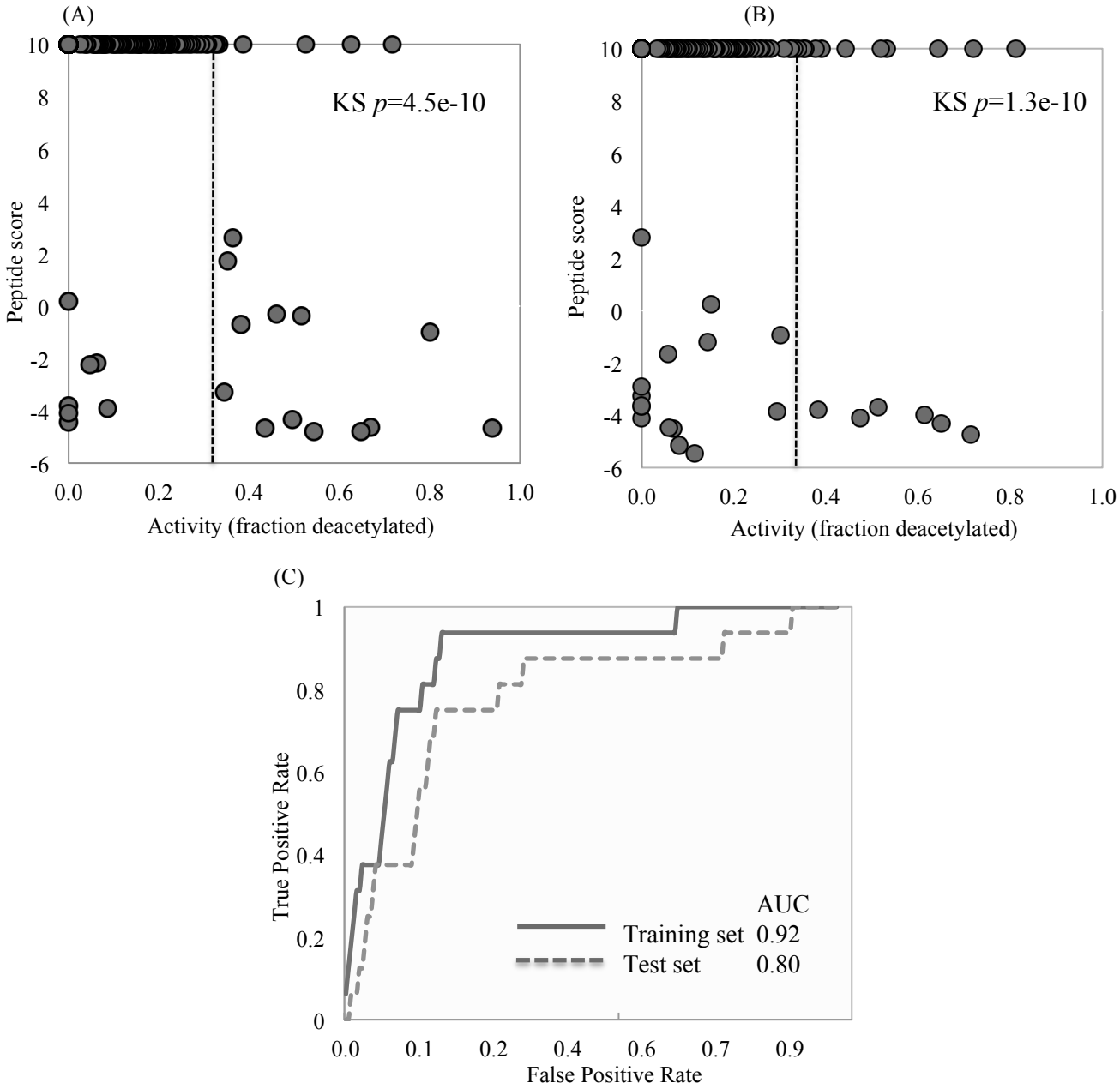

Figure S1. Related to Figure 2A and Table S2: Performance of the initial protocol on *Set_GX$_1$KacX$_2$GC*. The set was separated into two parts to calibrate **(A)** and validate **(B)** the protocol (see **Table S2**). **(A, B)** Activity *vs.* peptide score plots. Activity is presented as the fraction of deacetylated substrate after 30 minutes. A 0.34 deacetylation threshold distinguishes optimally substrates from non-substrates (Kolmogorov-Smirnov *p*-values of 4.5e-10 and 1.3e-05, respectively). **(C)** ROC plot for distinction of binders/non-binders according to the 0.34 deacetylation threshold (AUC values=0.92 and 0.80 for training [solid line] and test sets [dotted line], respectively).

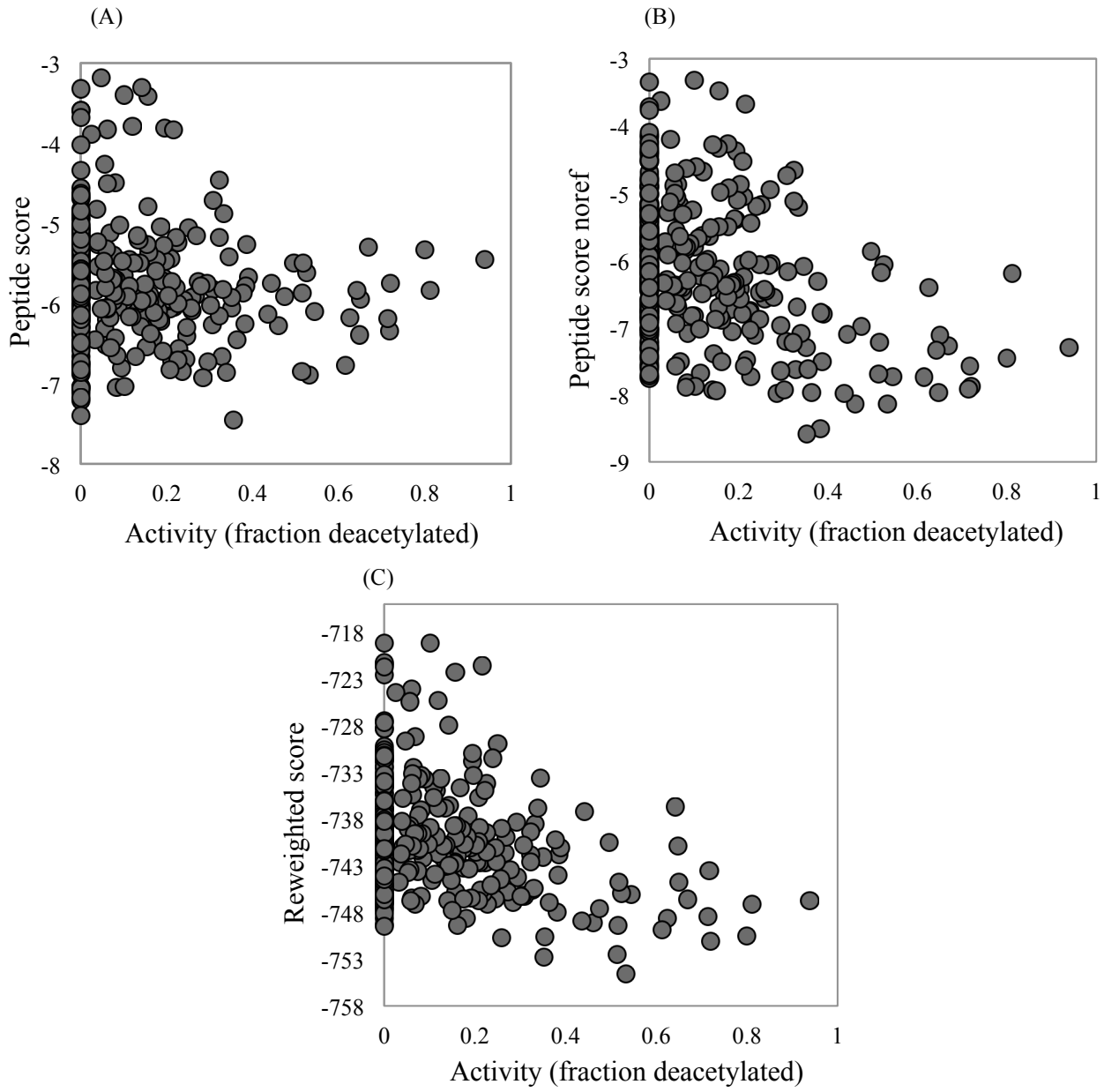

Figure S2. Related to Figure 2A: Performance of the optimized protocol (see Methods) on *Set_GX$_1$KacX$_2$GC* using different scoring measures. Plots of predicted binding (y-axis) *vs*. experimental activity (x-axis) are shown for the measures *Peptide score* **(A)**, *Peptide score noref* **(B)**, and *Reweighted score* **(C)**.

Supplementary Tables

Table S1. Related to Figure 1: Constraints used during structure optimization (derived from PDB id 2v5w)

Comment	Type of constraint	Constraint	Atoms involved	Type of function	#
Enforce critical double hydrogen bond D101-Kac backbone	atom pair[a]	2.2 – 3.2Å	D101 $O^{\delta 2}$: Kac N	BOUNDED[b]	1
	atom pair	2.5 - 3.5Å	D101 $O^{\delta 1}$: Coumarin N	BOUNDED	2
	dihedral[c]	5.46 radians	Kac N : Kac C^{α} : Kac C : Coumarine N	CIRCULAR HARMONIC[d]	3
Coordination of catalytic Zn ion	atom pair	2.8Å	D267 $O^{\delta 2}$ / D267 $O^{\delta 2}$: Kac O^{η}	HARMONIC[e]	4
	atom pair	3.7Å	D178 $O^{\delta 2}$: Kac O^{η}	HARMONIC	5
	atom pair	3.3Å	H180 $N^{\delta 1}$: Kac O^{η}	HARMONIC	6
Lock Kac side chain via stacking interactions and hydrogen bond	atom pair	3.0Å	G151 O : Kac N^{ζ}	HARMONIC	7
	atom pair	3.8Å	F152 $C^{\varepsilon 1}$ / F152 $C^{\varepsilon 2}$: Kac C^{δ}	HARMONIC	8
	atom pair	3.7Å	F208 $C^{\delta 1}$ / F208 $C^{\delta 2}$: Kac C^{γ}	HARMONIC	9

[a] Distance constraint between two atoms

[b] Bounded function: $$f(x) = \begin{cases} \left(\frac{x-lb}{sd}\right)^2 & \text{for} \quad x < lb \\ 0 & \text{for} \quad lb \le x \le ub \\ \left(\frac{x-ub}{sd}\right)^2 & \text{for} \quad ub < x \le ub + rswitch \cdot sd \\ \frac{1}{sd}\left(x - (ub + rswitch \cdot sd)\right) + \left(\frac{rswitch \cdot sd}{sd}\right)^2 & \text{for} \quad x > ub + rswitch \cdot sd \end{cases}$$, where *lb*, *ub* represent the lower and upper bounds, respectively, within which zero scoring penalty is applied; rswitch is set to 0.5

[c] Dihedral constraint between four atoms

[d] Circular harmonic function: $$f(x) = \left(\frac{NearestAngleRadians(x, x0) - x0}{sd}\right)^2$$, where *x0* represents the radian in the starting structure.

[e] Harmonic function: $$f(x) = \left(\frac{x - x0}{sd}\right)^2$$

A standard deviation of 0.2 was used in all constraint implementations.

Table S2. Related to Figure 2A: HDAC8 activity on peptides in *Set_GX$_1$KacX$_2$GC*. The fraction of substrate deacetylated after 30 minutes is given for each amino acid combination (except cysteine) at the residue preceding (X_1, rows) and following (X_2, columns) Kac (values >0.34 are in *italics*). The two subsets used for calibration/validation are colored in white/gray, and the peptide used for generating the starting template structure is highlighted in bold.

	G	**A**	**V**	**L**	**M**	**I**	**F**	**Y**	**W**	**S**	**T**	**P**	**N**	**Q**	**K**	**R**	**H**	**D**	**E**
G	0.0	0.0	0.1	*0.4*	0.0	0.1	*0.7*	0.1	0.2	0.2	0.2	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2
A	0.0	0.0	0.2	0.3	0.0	0.1	*0.6*	*0.4*	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.2	0.1
V	0.2	0.0	0.0	0.2	0.1	0.0	*0.6*	*0.4*	*0.5*	0.2	0.2	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0
L	0.0	0.0	0.1	0.2	0.0	0.1	*0.7*	0.2	*0.4*	*0.3*	0.2	0.0	0.0	0.2	0.0	0.1	0.0	0.0	0.0
M	0.0	0.1	0.0	0.0	0.1	0.0	*0.4*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
I	0.1	0.0	0.0	0.2	0.0	0.0	*0.7*	0.1	*0.5*	0.3	0.2	0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.1
F	0.2	0.0	0.1	0.1	0.2	0.2	*0.7*	*0.4*	*0.4*	0.3	0.3	0.1	0.0	0.3	0.0	0.0	0.0	0.0	0.0
Y	0.3	0.0	0.2	*0.4*	0.3	0.0	***0.9***	*0.7*	*0.8*	0.2	0.3	0.1	0.1	0.1	0.0	0.2	0.0	0.1	0.0
W	0.0	0.0	0.2	0.0	0.0	0.0	*0.4*	*0.7*	*0.4*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S	*0.4*	0.0	0.1	0.3	0.0	0.1	*0.6*	*0.5*	*0.5*	0.2	0.3	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0
T	0.2	0.0	0.1	0.3	0.0	0.0	*0.5*	0.1	*0.5*	0.2	0.3	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
P	0.0	0.0	0.0	0.1	0.0	0.1	*0.4*	0.1	0.1	0.3	0.2	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0
N	0.1	0.0	0.3	0.2	0.0	0.1	*0.5*	0.3	0.0	0.2	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2
Q	0.0	0.0	0.3	0.2	0.0	0.0	*0.5*	0.0	0.2	0.3	0.2	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.2
K	0.0	0.0	0.0	0.0	0.0	0.0	*0.5*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R	0.0	0.1	0.1	0.1	0.0	0.1	*0.8*	0.6	0.3	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.1	0.0
H	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D	0.0	0.0	0.3	0.3	0.0	0.0	*0.4*	0.0	0.2	0.2	0.1	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2
E	0.0	0.0	0.3	0.3	0.0	0.0	0.2	0.0	0.0	0.2	0.2	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.2

Table S3. Related to Figure 2B: HDAC8 activity on peptides in *Set_GRKacX$_2$X$_3$C*. The percentage of substrate deacetylated is given for each amino acid combination (except cysteine) at the two residues following Kac (X_2, rows; X_3, columns).

	G	**A**	**V**	**L**	**M**	**I**	**F**	**Y**	**W**	**S**	**T**	**P**	**N**	**Q**	**K**	**R**	**H**	**D**	**E**
G	3	11	3	12	3	0	3	3	3	1	0	0	0	0	3	3	0	0	0
A	3	3	3	0	21	0	3	3	3	0	2	0	0	0	11	4	0	0	3
V	3	3	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	3	3
L	11	22	0	11	0	0	3	3	3	0	0	0	0	0	3	0	0	0	0
M	3	0	0	0	0	1	11	3	1	0	0	0	0	0	0	0	0	0	0
I	3	11	0	11	0	0	3	0	3	0	0	0	0	0	4	0	0	0	0
F	97	95	80	69	69	69	69	80	80	68	32	80	80	11	50	80	32	11	3
Y	69	11	11	11	11	11	21	11	11	12	2	21	11	3	11	11	3	4	3
W	32	11	11	11	3	32	21	11	11	3	11	21	21	0	22	21	3	11	3
S	3	12	3	11	0	32	3	3	11	0	0	0	0	0	11	0	0	0	0
T	3	0	3	0	0	1	3	3	1	0	0	0	0	0	0	0	0	3	3
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

N	11	3	0	3	0	1	2	0	0	0	0	0	3	0	0	0	0	0	3
Q	0	3	0	2	0	1	0	3	3	2	0	0	0	0	3	0	0	0	3
K	11	3	0	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0	0
R	11	3	0	2	3	3	21	3	11	3	0	0	0	3	3	0	3	0	4
H	2	12	3	11	0	0	2	3	3	2	0	0	0	0	4	3	0	0	0
D	3	11	3	0	0	0	0	0	3	0	0	0	0	0	11	0	0	0	0
E	0	0	3	3	0	1	0	0	0	0	0	0	0	0	3	0	0	0	0

Table S4. Related to Table 2 and Figure 4C: Determination of *Novel_candidate_set2* to define optimal scoring measure: I_sc *vs*. peptide_sc_noref

Peptide	Protein (Uniprot ID)	Min. I_sc	Min. Pep_sc_ noref	Binding strength[a] I_sc/Pep_sc_ noref	k_{cat}/K_M
LS K_{747} FLR	CAD (P27708)	-21.6	-7.3	++/+	630
SF K_{55} YAW	EF1 alpha 1 (P68104)	-20.8	-10.2	++/++	1176
FG K_{1275}FSW	Znf 318 (Q5VUA4)	-20	-9.5	++/++	4826
TW K_{78} ANF	IRF2 (P14316)	-19.6	-7.9	++/++	77
IS K_{504} YDR	Msh6 (P52701)	-18.6	-8.2	++/++	83.5
SL K_{269} EFY	ALDHIII (P30838)	-18	-7.9	++/++	78
RL K_{713} YSQ	SMC1A (Q14683)	-16.8	-7.2	+/+	79.5
SG K_{149} YYY	CBX4 (O00257)	-16.8	-9.8	+/++	21
LG K_{1017}FRR	LARP1 (Q6PKG0)	-19.1	-6.1	++/-	349
KI K_{31} RLR	60S L7 (P18124)	-18.5	-5.1	++/--	38
SG K_{516} YFA	AMPD2 (Q01433)	-16.2	-9.2	-/++	20
FV K_{358} AFA	ALDH5A1 (P51649)	-14.7	-8.9	--/++	3
TF K_{8} GVD	HN1 (Q9UK76)	-15.4	-8.2	--/++	14
GG K_{11} AFG	AIF1 (P55008)	-14.5	-7.4	--/+	21
GI K_{371} PFL	anillin (Q9NQW6)	-7	-7.1	--/+	0
SQ K_{359} EED	SMARCC1 (Q92922)	-14.6	-6.3	--/-	NA
KG K_{878} DAE	N-CoR2 (Q9Y618)	-14.5	-5	--/--	36
RK K_{292} GEP	p53 (P04637)	-11.9	-4.9	--/--	10
SE K_{1970}PEK	DNA-PKcs (P78527)	-8	-2.8	--/--	5

[a] Strength measure: ++ = very strong binder (I_sc <= -18.0; Pep_sc_noref <= -7.5); + = moderate binder (I_sc <= -16.5; Pep_sc_noref <= -7.0); - = poor substrate (I_sc > -16.5; Pep_sc_noref > -7.0); -- =very poor substrate (I_sc > -16.0; Pep_sc_noref > -6.0)

Table S5. Related to Tables 1-3: Local structure of the substrate peptide within the context of the full protein (predicted, or derived from structure if crystallographic coordinates are available). For proteins without solved structures, probability for disorder was calculated using IUPred (Dosztanyi et al., 2005) and secondary structure was predicted using PSIPred (Jones, 1999).

Novel_candidate_set1

Peptide	Protein (Uniprot ID)	Disorder (IUPred [a])	Pdb Id-Chain	Secondary structure (PSIPred [b])
VS K_{350} GPF	KAT6A (Q92794)	0.7	-	CCCCC-CCCCCC-CCCCC
SF K_{769} SDQ	TRIM33 (Q9UPN9)	0.5	-	CCCCC-EECCCC-CEECC
VS K_{93} GTL	Histone H1.5 (P16401)	0.2	-	HHHHH-HHCCCE-CEEEC
DI K_{42} YPL	B5R (P00387)	Exposed in loop	1umk-A	TTTTT-TTCEEE-EEEEE
DH K_{604} TLY	KAT6A (Q92794)	Loop region; points inwards	2rc4-A	HHHCT-TTTCTT-TTTTT

SG K_{256} YDL	Alpha-enolase (P06733)	Exposed in loop	3b97-C	GGCEE-TTEETT-TTTTC
DS K_{1583} NAK	CBP (Q92793)	0.7	-	CCCCC-CCCCCC-CCCCC
KP K_{209} AAK	Histone H1.5 (P16401)	0.8	-	CCCCC-CCCCCC-CCCCC
PG K_{117} GVK	E2F1 (Q01094)	0.8	-	CCCCC-CCCCCC-CCCCC
FT K_{252} DHL	TRIM33 (Q9UPN9)	0.3	-	HCCCC-CCCCCE-CEEEE
KG K_{953} TAQ	TRIM33 (Q9UPN9)	Missing coordinates - high flexibility	3u5o-C	TTTTT-TTCTTT-TCCHH*
KK K_{1588} NNK	CBP (Q92793)	0.8	-	CCCCC-CCCCCC-CCCCC
II K_{593} DGE	AIFM1 (O95831)	Exposed at end of helix	4lii-A	HHHHH-HHHHCC-BTTTH
TQ K_{755} QEQ	Nucleoprotein TPR (P12270)	0.5	-	HHHHH-HHHHHH-HHHHH
GV K_{168} KVA	Histone H1.5 (P16401)	0.7	-	CCCCC-CCCCCC-CCCCC
DD K_{169} YTL	SPC25 (Q15005)	0.1	-	CCCCC-CCEEEE-EEEEE
SG K_{345} KGQ	SMARCC1 (Q92922)	0.8	-	CCCCC-CCCCCC-CCCCC
SF K_{109} LNK	Histone H1.5 (P16401)	0.5	-	CCCCE-EEEECC-CCCCC
NQ K_{748} LTA	Nucleoprotein TPR (P12270)	0.5	-	HHHHH-HHHHHH-HHHHH
VG K_{815} SVS	KAT6A (Q92794)	0.7	-	CCCCC-CCCCCC-CCCCC

[a] Disorder prediction using IUPRED (Dosztanyi et al., 2005): disordered (p >= 0.5), structured (p < 0.5)
[b] Secondary structure: previous 5 residues-peptide fragment- following 5 residues. Based on solved structure, or secondary structure prediction using PSIPRED (Jones, 1999)
* Missing coordinates within solved structure

Novel_candidate_set2

Peptides	Protein (Uniprot ID)	Disorder (IUPred)	Pdb Id-Chain	Secondary structure (PSIPred)
FG K_{1275}FSW	Znf 318 (Q5VUA4)	0.5	-	CCCCC-CCCCCC-CCCCC
SF K_{55} YAW	EF1 alpha 1 (P68104)	0.1	-	HHCCC-CEEHHH-HHHHH
LS K_{747} FLR	CAD (P27708)	0.1	-	ECCCC-HHHHCC-CCCCC
LG K_{1017}FRR	LARP1 (Q6PKG0)	Exposed near end of flexible helix	4zc4-A	HHHHH-HHHHCC-CCCCH
IS K_{504} YDR	Msh6 (P52701)	Exposed at end of short helix; Close to protein-DNA interface	2o8f-B	HHTTT-TCGGGC-CCCEE
RL K_{713} YSQ	SMC1A (Q14683)	0.5	-	HHHHH-HHHHHH-HHHHH
TW K_{78} ANF	IRF2 (P14316)	0.5	-	CCCHH-HHHHHH-HHHHH
SL K_{269} EFY	ALDHIII (P30838)	Exposed near end of helix	4l2o-G	HHHHH-HHHHHH-TTTGG
KI K_{31} RLR	60S L7 (P18124)	0.4		HHHHH-HHHHHH-HHHHH
KG K_{878} DAE	N-CoR2 (Q9Y618)	0.7	-	CCCCC-CCCCCC-CCCCC
SG K_{149} YYY	CBX4 (O00257)	0.7	-	CCCCC-CCCCCE-CEECC
GG K_{11} AFG	AIF1 (P55008)	Flexible region in NMR model	2g2b-A	TTTCC-CCCCGG-GHHHH
SG K_{516} YFA	AMPD2 (Q01433)	0.1	-	CCCCC-CHHHHH-HHHHH
TF K_{8} GVD	HN1 (Q9UK76)	0.8	-	CCEEE-ECCCCC-CCCCC
RK K_{292} GEP	p53 (P04637)	0.7	-	HHHHC-CCCCCC-CCCCC
SE K_{1970}PEK	DNA-PKcs (P78527)	0.0	-	EEEEE-CCCCCC-CCCEE
FV K_{358} AFA	ALDH5A1 (P51649)	Exposed in middle of helix	2w8r-A	GGHHH-HHHHHH-HHHHH
GI K_{371} PFL	anillin (Q9NQW6)	0.7	-	CCCCC-CCCCCC-CCCCC

Proteometargets_set3

Peptides[a]	Protein (Uniprot ID)	Disorder (IUPred)	Secondary structure (PSIPred*)
KL**IS** $\mathbf{K_{1808}}$ **FDK**L	ARID1A (O14497)	0.4	CCCCC-CCCCCC-CCEEE
ST**PV** $\mathbf{K_{292}}$ **FIS**R	CSRP2BP (Q9H8E8)	0.5	CCCCC-CCCCCC-CCCCC
RVIG**AK** $\mathbf{K_{106}}$ **DQY**	SMC3 (Q9UQE7)	0.1	EEEEE-CCCEEE-EECCE
KRI**LH** $\mathbf{K_{687}}$ **LLQ**N	NCOA3 (Q9Y6Q9)	0.5	HHHHH-HHHHHC-CCCCC
SK**IQ** $\mathbf{K_{3579}}$ **QLD**Q	MLL2 (O14686)	0.6	HHHHH-HHHHHH-HHHHH
KL**GG** $\mathbf{K_{1087}}$ **QRA**A	RAI1 (Q7Z5J4)	0.9	CHHHC-CCCCCC-CCCCC
KL**SG** $\mathbf{K_{167}}$ **EIN**G	SRSF5 (Q13243)	0.5	HHHHH-CCCCCC-CEEEE
LG**DG** $\mathbf{K_{387}}$ **MKS**	THRAP3 (Q9Y2W1)	0.6	CCCCC-CCCCCC-CCCCC
TE**IG** $\mathbf{K_{54}}$ **TLA**EK	ZRANB2 (O95218)	0.5	CCCCC-HHHHHC-CCCCC

[a] hexamer stretch of interest in shown in bold

* no solved structures were available for any of these targets in the regions of interest

Supplemental Experimental Procedures

- **Rosetta Run-line commands. Related to Methods:** In all runs, we used Rosetta version 3.4., with minor changes. All files are included as Appendix).

  1. Thread peptide sequences onto template, using the *Rosetta fixbb* protocol:

```
$ROSETTA_BIN/fixbb.linuxgccrelease -database $ROSETTA_DB -s template.pdb -resfile
threading_resfile -ex1 -ex2aro –use_input_sc  -nstruct 1 >design.log
```

  where the Rosetta executable ($ROSETTA_BIN/fixbb.linuxgccrelease) and the Rosetta database ($ROSETTA_DB) paths are provided. The input is provided using the '*-s template.pdb*' flag; threading instructions are defined in the resfile and provided using the '*–resfile threading_resfile*' flag. Increased rotamer sampling is enforced for $\chi 1$ and aromatic $\chi 2$ angles, using the '*–ex1 –ex2aro*' flags. One output structure is generated as instructed by the '*–nstruct 1*' flag. The output is written to the log file *design.log*.

  Given below is an example resfile to thread hexamer GYKFGC onto the chain B residues numbered 364 - 369.

```
NATRO
start
364    B PIKAA G  EX 1 EX 2 USE_INPUT_SC
365    B PIKAA Y  EX 1 EX 2 USE_INPUT_SC
366    B PIKAA K  EX 1 EX 2 USE_INPUT_SC
367    B PIKAA F  EX 1 EX 2 USE_INPUT_SC
368    B PIKAA G  EX 1 EX 2 USE_INPUT_SC
369    B PIKAA C  EX 1 EX 2 USE_INPUT_SC
```

  2. Optimize receptor structure before peptide docking: prepack threaded template using Rosetta FlexPepDock:

```
$ROSETTA_BIN/FlexPepDocking.linuxgccrelease -database $ROSETTA_DB -s
threaded_template.pdb -ex1 -ex2aro -use_input_sc -unboundrot unbound.pdb –
flexpep_prepack  -nstruct 1
```

  The flag '*–flexpep_prepack*' performs prepacking on the input provided using the '*–s threaded_template.pdb*' flag and generates one prepacked output as instructed by '*–nstruct 1*' flag. Input side-chain and unbound receptor side-chain coordinates are included in the rotamer library using the '*–use_input_sc*' and '*–unboundrot'* flags, respectively.

  3. Minimize structure using Rosetta FlexPepDock:

```
$ROSETTA_BIN/FlexPepDocking.linuxgccrelease -database $ROSETTA_DB -s
threaded_template_prepacked.pdb -flexPepDockingMinimizeOnly -cst_fa_file cons.cst
–cst_fa_weight 1.0 –nstruct 1
```

Minimization of the prepacked input threaded_template_prepacked.pdb is performed using the '-*flexPepDockingMinimizeOnly*' flag. The constraint file and corresponding weight addition to the scoring function are provided by '*–cst_fa_file*' and '*cst_fa_weight*' flags, respectively.

4. Refine structure using Rosetta FlexPepDock:

```
$ROSETTA_BIN/FlexPepDocking.linuxgccrelease -database $ROSETTA_DB -s
threaded_template_prepacked.pdb -ex1 -ex2aro -use_input_sc -unboundrot unbound.pdb
–pep_refine -cst_fa_file cons.cst –cst_fa_weight 1.0 -nstruct 200
```

Refinement of the prepacked input threaded_template_prepacked.pdb is performed (200 models are generated) using the '*–pep_refine*' flag. The constraint file and corresponding weight addition to the scoring function is provided by '*–cst_fa_file*' and '*cst_fa_weight*' flags, respectively.

Given below is the constraint file used to keep critical interations during the simulation (each constraint is numbered as in **Table S1**)

```
AtomPair OD1 101 N 367 BOUNDED 2.5 3.5 0.2 0.5 TAG              #1
AtomPair OD2 101 N 366 BOUNDED 2.2 3.2 0.2 0.5 TAG              #2
Dihedral N 366 CA 366 C 366 N 367 CIRCULARHARMONIC 5.463 0.2    #3
AmbiguousConstraint
AtomPair OD2 267 OH 366 HARMONIC 2.9 0.2                        #4
AtomPair OD1 267 OH 366 HARMONIC 2.9 0.2
END_AMBIGUOUS
AmbiguousConstraint
AtomPair OD2 164 OH 366 HARMONIC 3.7 0.2                        #5
AtomPair OD1 164 OH 366 HARMONIC 3.7 0.2
END_AMBIGUOUS

AtomPair ND1 166 OH 366 HARMONIC 3.3 0.2                        #6
AtomPair O   137 NZ 366 HARMONIC 3.0 0.2                        #7
AmbiguousConstraint
AtomPair CE1 138 CD 366 HARMONIC 3.8 0.2                        #8
AtomPair CE2 138 CD 366 HARMONIC 3.8 0.2
END_AMBIGUOUS
AmbiguousConstraint
AtomPair CD1 194 CG 366 HARMONIC 3.7 0.2                        #9
AtomPair CD2 194 CG 366 HARMONIC 3.7 0.2
END_AMBIGUOUS
```

HDAC8 expression and assay. Recombinant HDAC8 was expressed, purified and reconstituted with stoichiometric zinc, as previously described (Wolfson et al., 2014). Peptides were purchased from Sigma-Aldrich with acetylated N-termini as unpurified peptides and were solubilized in 50% acetonitrile or water and treated with Chelex resin before use. *Novel_candidate_set1* peptides contain a C-terminal carboxylate while *Novel_candidate_set2* peptides contain a C-terminal carboxamide. The initial rate for deacetylation of peptides (two concentrations ≥100 μM) catalyzed by 1 μM HDAC8 was determined by enzymatically coupling the formation of acetate to the production of NADH in a stopped assay format and quantified by fluorescence (ex. = 340 nm, em. = 460 nm). The values of k_{cat}/K_M were calculated from fitting a line to the substrate-dependence of the initial rate.

Prediction of local structural features. Related to Table S5. In order to assess the ability of the identified substrates to undergo deacetylation by HDAC8 within the context of the full protein, we further investigated the accessibility of the target lysine by HDAC8: (1) *Local accessibility*: For proteins with solved structures, we calculated the accessibility of the target lysine. For the rest, we calculated the *probability for being intrinsically disordered* (using *IUPred* (Dosztanyi et al., 2005)): such regions within a protein are generally accessible. (2) *Local secondary structure* can play an important role in the ability of the substrate peptide to bind to HDAC8 in a catalysis-competent conformation. In particular, the *cis* peptide backbone geometry necessary to allow hydrogen bonding of the acetylated lysine residue of the substrate to HDAC8 residue D101 (see **Figure 1**) indicates that a helical peptide might not be a good candidate for deacetylation. The secondary structure was extracted from the solved structure, or predicted using *PSIPred* secondary structure prediction (Jones, 1999).

References

DOSZTANYI, Z., CSIZMOK, V., TOMPA, P. & SIMON, I. 2005. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol,* 347**,** 827-39.

JONES, D. T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol,* 292**,** 195-202.