Supporting Information

Novel Class IIa-Selective Histone Deacetylase Inhibitors Discovered Using an *in Silico* Virtual Screening Approach

Kai-Cheng Hsu¹, Chang-Yi Liu², Tony Eight Lin¹, Jui-Hua Hsieh³, Tzu-Ying Sung⁴, Hui-Ju Tseng⁵, Jinn-Moon Yang^{4,6} and Wei-Jan Huang^{2,7*}

¹Graduate Institute of Cancer Biology and Drug Discovery, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

²Graduate Institute of Pharmacognosy, Taipei Medical University, Taipei, Taiwan

³Kelly Government Solutions, Research Triangle Park, North Carolina, United States of America

⁴Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan

⁵Ph.D. Program in Biotechnology Research and Development, Taipei Medical University, Taipei, Taiwan

⁶Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan

⁷Ph.D Program in Biotechnology Research and Development, College of Pharmacy, Taipei Medical University

*Correspondence should be addressed to:

Dr. Wei-Jan Huang, Graduate Institute of Pharmacognosy, Taipei Medical University, No.250, Wu-hsing St., Taipei 11031, Taiwan; Tel: +886-2-27361661 ext. 6152; Fax: +886-2-27355276; E-mail: wjhuang@tmu.edu.tw

Analysis of ligand-protein interaction in HDAC5 and HDAC7

We compared the six inhibitor interactions with HDAC5 and HDAC7 (PDB ID: 3ZNS) to understand their different activities. Because the crystal structure of HDAC5 is unavailable, a homology model was used instead. Our analysis of the six compounds showed that they yield similar interactions with HDAC5 and HDAC4 (Table 2). A4291 (Supplementary Fig. 3A) and A159452 (Supplementary Fig. 3B) share similar moieties and binding activities. Their cyclohexane-1,2-diol moieties form coordinated bonds with the zinc ion, as well as hydrogen bond with the polar amino acid residue H833. Van der Waals interactions are formed with the hydrophobic L973 amino acid residue. Inhibitors A226640 (Supplementary Fig. 3C), A640973 (Supplementary Fig. 3D) and A364365 (Supplementary Fig. 3E) contain sulfonyl moieties that create coordinated bonds with the zinc ion. They also share interactions with residues H872 and G1004. Inhibitor 363007 (Supplementary Fig. 3F) contains a phosphoric moiety and forms hydrogen bonds to R710 and G1004, but hydrophobic bonds to H872 and F900.

Our IC₅₀ results showed the identified inhibitors having less inhibitory effect on HDAC7 (Table 2). Inhibitors A4291 (Supplementary Fig. 4A), A159452 (Supplementary Fig. 4B), A226640 (Supplementary Fig. 4C) and A640973 (Supplementary Fig. 4D) were unable to form interactions with the zinc ion within the catalytic site. In contrast, Table 2 displayed inhibitors A364365 (Supplementary Fig. 4E) and A363007 (Supplementary Fig. 4F) with weak inhibitory effects.

Finally, an interaction analysis was performed on HDAC9. The structure of HDAC9 is unavailable; therefore, a homology model was developed based on HDAC4. The interaction analysis of HDAC9 showed similar results to HDAC4 and HDAC5 (Supplementary Fig. 5A-F).



Supplementary Figure 1. The structures of 40 candidate compounds.



Supplementary Figure 2. Residue interactions of identified inhibitors.

(A-F) 2D diagram interactions between the identified inhibitors and cavity site residues of HDAC4 as developed from DS. Residues and bonds are colored by interaction types. (G) The heat map of interactions between cavity site residues and HDAC4. The highest interactions (4) are shaded red, whereas the lowest interactions (1) are shaded pink.



Supplementary Figure 3. Interaction analysis between identified inhibitors and HDAC5. The docking poses of six inhibitors in the binding site of HDAC5 (A-F) reveals the spatial conformations within the active site. Atoms were colored by type whereas lines (pink) represent HDAC5 protein residues. The zinc ion (grey sphere) is located within the active site, with dashed line indicating coordinate binding. Amino acid residues are listed as shown.



Supplementary Figure 4. Interaction analysis between identified inhibitors and HDAC7. The docking poses of six inhibitors in the binding site of HDAC7 (A-F) reveals the spatial conformations within the active site. Atoms were colored by type whereas lines (grey) represent HDAC7 protein residues. The zinc ion (grey sphere) is located within the active site, with dashed line indicating coordinate binding. Amino acid residues are listed as shown.



Supplementary Figure 5. Interaction analysis between identified inhibitors and HDAC9. The docking poses of six inhibitors in the binding site of HDAC9 (A-F) reveals the spatial conformations within the active site. Atoms were colored by type whereas lines (green) represent HDAC9 protein residues. The zinc ion (grey sphere) is located within the active site, with dashed line indicating coordinate binding. Amino acid residues are listed as shown.

HDAC1/1-482 HDAC2/1-488 HDAC3/1-428 HDAC4/1-1084 HDAC5/1-1122 HDAC6/1-1215 HDAC7/1-952 HDAC9/1-1011	15 Y DGD V19 16 Y DGD I20 9 Y DGD I20 9 Y DGD I20 97 - LE QQR I HQLR NY QASME AA G I P VSF GGHR PL SRAQS SP AS ATF P VS V 631 - SDAQ PL QPL QVY QA PL SLA 465 - P DQP VKH - LF T G VY D T F M601 465 - E P V L 451 - WE QQRL AG RL P R S T GD T V LL PL A Q G HR PL S RAQS SP AS A S V L P HP AS L S A PE P AS Q ARV L S S S E T P ART L P F T GL I Y D S VM528 574 QP F LE P T HT RALS VR QAP L AA V GMD - G L E KHRL VS RT HS SP AA S V L P HP A - MDR PL Q P - G S A T G I AY D P L M641
HDAC1/1-482 HDAC2/1-488 HDAC3/1-428 HDAC4/1-1084 HDAC5/1-1122 HDAC6/1-1215 HDAC7/1-952 HDAC9/1-1011	20 GNY YY GQGHPMKPH KIRMTHNL LLNY GLYR KME IY RPHKAN A EEMTKYHS DDY IKFLRS I R PDMSE Y SK QMQRFN V96 21 GNY YY GQGHPMKPH KIRMTHNL LLNY GLYR KME IY RPHKAT A EEMTKYHS DEY IKFLRS I R PDMSE Y SK QMQRFN V97 14 GNF HY GAGHPMKPH RLA LTHS LVLHY GLYK KMI VF KPY QA S QHDMC RFHSEDY IDFLQ RVS PTMMQGFTKSLNAFN V90 663 LKHQC TC GSSS HPEHA GRIQSIWS RLQET GLRGK CE CIRGRKATLDE IQTVHSE AHT - LLY GT PL N RQKLDSK KLL739 692 LKHQC NC GNTH VHPEHA GRIQSIWS RLQET GLLS NCERIRGRKATLDE IQTVHSE VHT - LLY GT SPL N RQKLDSK KLL739 490 MNH - CNLWD - SHHPE VPQRILRING RLGELGLA GRCLTLTPRPATEA ELLTCHSAE YV GHLRATEKKTRE LHRE S-353 529 LKHQC SC GDNS RHPEHA GRIQSIWS RLQET GLLS NCERIRGRKASLEE LQS VHSER HV - LLY GT NPL SRLKLDNG KLA605 642 LKHQC VC GNS THPEHA GRIQSIWS RLQET GLLNKCERIQGRKASLEE IQLVHSE HS - LLYGTNPL SRLKLDNF
HDAC1/1-482 HDAC2/1-488 HDAC3/1-428 HDAC4/1-1084 HDAC5/1-122 HDAC6/1-1215 HDAC6/1-1215 HDAC7/1-952 HDAC9/1-1011	97 GE D C P VF D G L F E F C Q L S T G G - S V A S A V K L N K Q Q T D I A V N WA G G L HHA K K S E A S G F C Y 152 98 GE D C P VF D G L F E F C Q L S T G G - S V A G A V K L N Q Q T D M A V N WA G G L HHA K K S E A S G F C Y 153 91 G D D C P VF P G L F E F C S K T G A - S L G A T Q L N K I C D I A I N WA G G L HHA K K S E A S G F C Y 153 91 G D C P VF P G L F E F C S K T G A - S L G A T Q L N K I C D I A I N WA G G L HHA K K S E A S G F C Y 154 740 G S L A S V F V R L P C G G V G V D S D T I W N E V H S A G A AR L A V G C V E L V F K V A T G E L K N G F A V V R P P G H HA E E S T A M G F C F 844 769 G P I S Q K V F V R L P C G G V G V D S D T I W N E V H S A G A AR L A V G C L L E L A F K V A S G E L K N G F A V V R P P G H HA E E S T A M G F C F 844 764 S F D S I Y I C P S T F A C A Q L A T G A A C R L V E A V L S G E V L N G F A V V R P P G H HA E E S T A M G F C F 844 766 G L L A Q R MF W L P C G G V Q V D D T I W N E L H S N A A R W A G S W T D L A F K V A S R E L K N G F A V V R P P G H HA E Q D A A C G F C F 681 719 G D D S Q K F F S L P C G G L G V D S D T I W N E L H S S G A A R MA Y G C V I E L A S K V A S G E L K N G F A V V R P P G H HA E E S T A M G F C F 744 719 G D D S Q K F F S L P C G G L G V D S D T I W N E L H S S G A A R MA Y G C V I E L A S K V A S G E L K N G F A V V R P P G H HA E E S T A M G F C F 744 719 G D D S Q K F F S L P C G G L G V D S D T I W N E L H S S G A A R MA Y G C V I E L A S K V A S G E L K N G F A V V R P P G H H A E E S T A M G F C F 744 719 G D D S Q K F F S S L P C G G L G V D S D T I W N E L H S S G A A R MA V G C V I E L A S K V A S G E L K N G F A V V R P P G H H A E E S T A M G F C F 744
HDAC1/1-482 HDAC2/1-488 HDAC3/1-428 HDAC4/1-1084 HDAC5/1-122 HDAC6/1-1215 HDAC6/1-1215 HDAC7/1-952 HDAC9/1-1011	153 VND IVLA ILELL KYH QK VLYID ID IHHGDG VEEAFYTTDRVM TV SFHKY GE - YFPG TGDL RD IGAGKGK YYA VN225 154 VND IVLA ILELL KYH QK VLYID ID IHHGDG VEEAFYTTDRVM TV SFHKY GE - YFPG TGDL RD IGAGKGK YA VN225 147 VND IVIG ILELL KYH QK VLYID ID IHHGDG VEEAFYTTDRVM TV SFHKY GN YFPG TGDL RD IGAGKGK YA VN225 155 FNS VA VAAKLLQ R - LS VSK IL IVD VD VHHGNG TQ QAFYSDPSVLYNSLHRYDD GN - FFPG SGAPE VGG GPG VG YN N921 162 FNS VA VAAK HA QT IS GHAL R IL IVD WDVHHGNG TQ QAFYSDPSVLYNSLHRYDD GN - FFPG - SGAPE VGG GG GVG YN N921 162 FNS VA 1A CR QLQ Q - SKASK IL IVD WD VHHGNG TQ QAFYSDPSVLYNSLHRYDHG T - FFPMG DE GASS QIG RAAGTG FTV N728 179 FNS VA 1A CR QLQ Q - SKASK IL IVD WD VHHGNG TQ QAFY DPSVLYNSLHRYDHG T - FFPG - SGAVE VG GG GG G GFF VN738 179 FNS VA 1A CR QLQ Q - LNISK IL IVD WD VHHGNG TQ QAFYADPSILYISLHRYDHG GN - FFPG - SGAVE VG GG GG GF VN738 179 FNS VA 11 CA VLRD Q - LNISK IL IVD WD VHHGNG TQ QAFYADPSILYISLHRYDE GN - FFPG - SGAVE VG GG GG GF VN738 170 FNS VA 11 CA VLRD Q - LNISK IL IVD WD VHHGNG TQ QAFYADPSILYISLHRYDE GN - FFPG - SGAVE CG GYN VN738 170 FNS VA 11 CA VLRD Q - LNISK IL VD WD VHHGNG TQ QAFYADPSILYISLHRYDE GN - FFPG - SGAVE CG GYN VN738 170 FNS VA 11 CA VLRD Q - LNISK IL VD WD VHGN TA TA TA VA TA TA NYLRD Q - LNISK IL VVNN TA TA NYLRD Q - LNISK IL VVNN TA TA NYLRD Q - LNISK IL VVNN TA TA NYLRD Q - LNISK IL VNN TA TA NYLRD Q - LNISK IL VNN TA NYN TA NYN TA TA NYLRD Q - LNISK IL VNN TA NYN TA NYN TA NYN TA NYN TA NYLRD Q - LNISK IL VNN TA NYN TA NYN TA NYN TA NYLRD Q - LNISK IL VNN TA NYN TA NYN TA NYN TA NYL TA NYLRD Q - LNISK IL VNN TA NYN T
HDAC1/1-482 HDAC2/1-488 HDAC3/1-428 HDAC4/1-1084 HDAC5/1-122 HDAC6/1-1215 HDAC6/1-952 HDAC9/1-1011	226 YPL RDG ID DES YEA IF KP VMSK VMEMF QPS AVVL QCGSDSL SGD RLGC FNL T INGHAKC VEF VKSFNL PMLML - G299 227 FPMRDG ID DES YGQ IF KP I ISK VMEMY QPS AVVL QCGADSL SGD RLGC FNL T VKGHAKC VEF VKTFNL PLLML - G300 221 VPL RDG ID DQS YKH FQV INQ VVD FYQPT CIVL QCGADSL SGD RLGC FNL T VKGHAKC VEV VKTFNL PLLML - G300 322 VVL GGU DP PMGDAE YL AAF KTVVMP IA SE FAP DVVL VS SGFD AVE GH T PL GGYNL SA KCF GYL TKQLMGL AGGR I VL AL972 324 VMVG GGU DP PI GDVE VL TAF KTVVMP IA SE FAP DVVL VS SGFD AVE GH T FL GGYNL SA KCF GYL TKQLMGL AGGR I VL AL972 325 VAWT GGV DP PI GDVE VL TAF KTVVMP IA HEFS PD VVL VS SGFD AVE GH T FL GGYNL SA KCF GYL TKQLMGL AGGR I VL AL972 326 VAWN GG PRMGDA DYL AAF KTVVMP IA HEFS PD VVL VS SGFD AVE GH T PL GGYNL SA CF GYL TKQLMTL A GGR VVL AL902 327 VAWN GGL DP PMGD VE VL TAF KTVVMP IA HEFS PD VVL VS AGFD AAR GD - PL GG CQVS PE GYA HL THL LMGL AS GR I IL 1178 329 VAWN GGL DP PMGD VE VL AF KTIVVMP I AREFS PD VVL VS AGFD AAR GD - PL GG CQVS PE GYA HL THL LMGL AS GR VVL AL902 321 VAWN GGL DP PMGD VE VL AF KTIVVMP I AKEFS PD VVL VS AGFD AAR GD - FL GG CQVS PE GYA HL THL LMGL AS GR VVL AL902 322 I AWT GGL DP PMGD VE YL AF KTIVVMP I AKEFS PD VVL VS AGFD AAF GH PAPL GGY WY AKCFGHL TKQLMTL A GG AVVL AL902 322 I AWT GGL DP PMGD VE YL AF KTIVVA VA KEFD PDMVL VS AGFD AAF GH PAPL GGYK VT AKCFGHL TKQLMTL ADGR VVL AL902
HDAC1/1-482 HDAC2/1-488 HDAC3/1-428 HDAC4/1-1084 HDAC5/1-1122 HDAC6/1-1215 HDAC7/1-952 HDAC9/1-1011	300 GGGYT IRNVAR CWTYE TAVALDTE IPNEL - PYNDYFE YF GPDF KLHIS PS - NMTNQNT NEYLEKIKQRLFENLRMLPHAP G378 301 GGGYT IRNVAR CWTYE TAVALDCE IPNEL - PYNDYFE YF GPDF KLHIS PS - NMTNQNT PEYMEKIKQRLFENLRMLPHAP G378 295 GGGYT IRNVAR CWTYE TAVALDCE IPNEL - PYNDYFE YF GPDF KLHIS PS - NMTNQNT PEYMEKIKQRLFENLRMLPHAP G379 295 GGGYT URNVAR CWTYE TAVALDCE IPNEL - PYNDYFE YF GPDF KLHIS PS - NMTNQNT PEYMEKIKQRLFENLRMLPHAP G379 295 GGGYT URNVAR CWTYE TAVALDCE IPNEL - PYNDYFE YF GPDF KLHIS PS - NMTNQNT PEYMEKIKQRLFENLRMLPHAP G379 295 GGGYT URNVAR CWTYE TAVALDCE IPNEL - PYNDYFE YF GPDF KLHIS PS - NMTNQNT PEYMEKIKQRLFENLRMLPHAP G379 295 GGGHDLTAI - CDASEACVS - ALLGNELDPLPENVLQQRPNA - NA - 1013 1003 EGGHDLTAI - CDASEACVS - ALLS VELQPLDEAVLQQKPNI - NA - 1043 779 EGGYNLTSI - SESMAACTR - SLLGDP - PPLLTLPRPLS - GA - 816 840 EGGHDLTAI - CDASEACVA - ALLGNEVDPLSE - 60K QKPNL - SA - 880 953 EGGHDLTAI - CDASEACVN - ALLGNEVDPLSE - 880 953 EGGHDLTAI - CDASEACVN - ALLGNEVPLS - 880

Supplementary Figure 6. Multiple sequence alignment of HDAC isozymes. The HDAC isozyme sequences was obtained and aligned. Amino acid residues are color coded and highlighted to show residue consensus at numbered position.



Supplementary Figure 7. SAHA interactions in HDAC isozymes.

(A) HDAC2 (green) and HDAC6 (yellow) complex was aligned and a surface model was produced. The tyrosine residue is labeled as shown. (B) The structures of class I (HDAC2, green), IIa (HDAC4, pink; HDAC7, blue), and IIb (HDAC6, yellow) are aligned. Inhibitor A4291, posed from HDAC4, is aligned within the active site. The grey sphere represents the zinc ion.





(A) Ranking order of the identified inhibitors across three programs. (B) Hit rate of different methods. The six inhibitors are docked and ranked in DS, LeadIT or iGEMDOCK. The hit rate is defined as I/T (%), where *I* is the number of the identified inhibitors among the *T* highest-ranking compounds The consensus score (red line) improves the ranking of the inhibitors against 27 false positives.